Alma Mater Studiorum · University of Bologna

School of Science
Department of Physics and Astronomy
Master Degree in Physics

Estimation of the SM background for search of electroweakinos with an isolated lepton, jets and E_T^{miss} at $\sqrt{s} = 13$ TeV using the ATLAS detector

Supervisor:
Prof. Alberto Cervelli

Co-supervisor:
Dr. Kazuki Todome

Candidate:
Ishan Vyas

Academic Year 2020/2021

The first thing to realize about physics ... is its extraordinary indirectness.... For physics is not about the real world, it is about "abstractions" from the real world, and this is what makes it so scientific.... Theoretical physics runs merrily along with these unreal abstractions, but its conclusions are checked, at every possible point, by experiments.

-Anthony Standen

Dedicated to overcoming all the ups and downs, the distance and soon to starting a new life with my best friend, my girlfriend and now my wife, Kalindi.

Contents

Abstract

Supersymmetric extension of the Standard Model of particle physics addresses most of the SM open questions. One of the predictions of the model is that the lightest supersymmetric particle (LSP) must have mass of the $\mathcal{O}(1$ TeV) (electroweak scale) which is within the reach of the current accelerators. The present work is part of a larger analysis which aims to search for the pair-produced chargino and neutralino, or two charginos using the final state of single lepton coming from W boson decay, light jets produced from W/Z decays and, the missing transverse energy expected due to escape of the LSP.

First part of the analysis quantifies the kinematic agreement among the region definitions used estimate the SM background yields for Monte Carlo simulated samples containing electroweak decay of top quark. The yield variation (interpreted as systematic error) obtained during the extrapolation of the background to the region of signal events is less than 10%, thus providing a strong argument in the favor of the employed kinematic selection conditions. The second part of the analysis uses the associated production of two bosons to evaluate the impact of uncertainties arising from the non-perturbative dynamics of the model used to simulate the diboson decay channels. The results of our analysis show that the impact of mismodelling due to theoretical uncertainties are well within the acceptable range of 40% and thus, the estimated background along with the uncertainties can be taken as input for the likelihood fit and the subsequent steps of the analysis.

Introduction

Supersymmetric extension of the Standard Model of particle physics addresses most of the SM open questions by associating a fermion/boson to each of the observed boson/fermion in the Standard Model. One of the predictions of the model is that the lightest supersymmetric particle (LSP) must have mass of the $\mathcal{O}(1$ TeV) (electroweak scale). The present-day accelerators are capable of reaching such energies and hence, the SUSY events in principle, should be observed with the current energies or near-future upgrades.

The first chapter introduces the strong theoretical foundation of SUSY, introduces a few open questions and how SUSY can addresses them. The second chapter gives a comprehensive sketch to the Large Hadron Collider (LHC) and the ATLAS detector. The detector components are also explained.

The third chapter starts with the statistical analysis tools used throughout the thesis and introduces the simplified model of electroweak 1L SUSY channel. Using the simplified model we aim to search for the pair-produced chargino and neutralino, or two charginos using the final state of single lepton coming from W boson decay, light jets produced from W/Z decays and, the missing transverse energy expected due to escape of the LSP. The possible supersymmetric events are hidden in a large number of SM events which act as background for the SUSY searches. Various SM backgrounds are introduced and techniques to estimate them in the MC event samples is discussed.

Chapter 4 contains the analysis, calculations and results. The first part of the analysis quantifies the kinematic agreement among the region definitions used estimate the SM background yields for Monte Carlo simulated samples containing electroweak decay of top quark. The yield variation which are interpreted as systematic error are obtained during the extrapolation of the background to the region of signal events. The second part of the analysis uses the associated production of two bosons to evaluate the impact of uncertainties arising from the non-perturbative dynamics of the model used to simulate the diboson decay channels. The final chapter adds concluding remarks for the analysis.

Chapter 1

Physics beyond the Standard Model

The Standard Model of Particle Physics is formulated within the framework of *Quantum Field Theory*. The idea is that the building blocks of nature are not particles, but fields that span our infinite Universe. There is a field associated with every particle and these particles are merely disturbances in the field. Until now, the QFT framework has been extremely successful in explaining the baryonic matter and subatomic processes in our Universe.

1.1 Standard Model

The Standard Model is a phenomenological model that comprises 12 matter particles, 4 particles of interaction, and one Higgs boson as shown in figure 1.1. Mathematically, the theory of the Standard Model can be written as $SU(3)_C \times SU(2)_L \times U(1)_Y$. The groups are associated with each of the subatomic interactions that we observe. $\mathrm{SU}(3)_C$ describes the strong interactions due to color charges, $\mathrm{SU}(2)_L$ represents the weak interaction that is responsible for β-Decay and $\mathrm{U}(1)_Y$ represents electromagnetism.

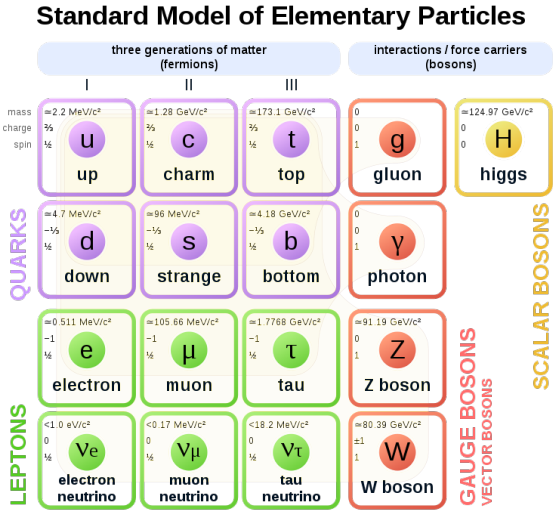


Figure 1.1: The fundamental particles of the Standard Model along with the mass, charge and spin values provided by the Particle Data Group[1]

While constructing the SM of particle physics, gravitational effects are not taken into consideration and hence are decoupled from the model. At the order of ~ 1 *fm* (roughly the size of a nucleus), the effects of gravitational interaction between the subatomic particles are negligible compared to the rest of the forces. Hence, the SM is successful in explaining the quantum phenomena. The elementary particles mentioned in figure 1.1 can be divided in two categories, *Fermions* and *Bosons*. Fermions are particles with half-integer spin and bosons carry integer spins. Based on this classification, all the quarks and leptons are fermions because of their $\frac{1}{2}$ spin values, whereas, vector bosons and the Higgs boson carry spin 1 and spin 0 respectively. The vector bosons are responsible for interaction among the matter particles(quarks and leptons).

1.2 Problems with the Standard Model

The SM has been remarkably successful at predicting the outcome of particle physics experiments. The SM predicted the existence and properties of W and Z bosons, gluon, top and charm quarks before these particles were observed. However, the theory has a few shortcomings, some of which are mentioned here.

The Hierarchy Problem

- *There is an enormous energy gap between Plank scale M_P and the Electro-Weak Scale M_{EW}*

This is the famous *hierarchy problem*[2][3] which is related to the concept of *numerical naturalness* as defined by Philip Nelson[4]. The Plank scale and the ElectroWeak scales are of the order of 10^{18} GeV and 10^2 GeV respectively. Taking the ratio of the two, we see that:

$$\frac{M_P}{M_{EW}} = \frac{10^{18}}{10^2} \approx 10^{16} GeV$$

The Plank scale M_P is the energy scale beyond which, the Standard Model is incapable of explaining the physics. The renormalization of electroweak forces at energies as low as 10^2 GeV suggests that there must be new physics somewhere between the 16 orders of magnitude that preserves the naturalness of the theory. An upper bound of M_P on the SM suggests that SM is only ab effective theory and can only describe physics up to energy scales of Λ_{UV}, also known as the ultraviolet cutoff. In the case of SM, $\Lambda_{UV} = M_P \approx 10^{18}$ GeV .

Higgs Mass Problem

The Higgs mass problem stems from the fact that ultraviolet cutoff for SM is enormously large. According to QFT, mass of a particle comes from it's bare mass and all the radiative corrections due to higher order Feynman diagrams. So, the mass of the Higgs is given as[5]:

$$m_h^2 = m_{h0}^2 + \kappa \delta m_h^2 \tag{1.1}$$

Where, m_{h0}^2 is the bare mass of Higgs and,

$$\delta m_h^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{UV}^2 + \mathcal{O}\left(m_f^2 \log \frac{\Lambda_{UV}^2}{m_f}\right), \tag{1.2}$$

with λ_f as the Yukawa coupling between the higgs and the fermion. As mentioned before, gravity is incompatible with the Standard Model[6]. This means that $M_P \approx 10^{18} GeV$ in natural units is the upper limit or Λ_{UV} of the theory. Solely based on Eq. 1.2, we see that for ultraviolet cutoff to be $\approx 10^{18}$ GeV, the radiative corrections will dominate and we can expect the Higgs mass to be proportional to the Plank mass:

$$m_H \propto M_P \tag{1.3}$$

But, the observed Higgs mass is at the order of 10^2 GeV, the order of weak scale or:

$$m_H \propto G_F^{-1/2} \tag{1.4}$$

Where G_F is the Fermi constant. Eq. 1.4 has been experimentally verified which leads to κ in Eq. 1.1 to be of $\mathcal{O}(\sqrt{G_F} M_P)^{-1} \approx 10^{-16}$. If the value of κ were anything but close to 10^{-16}, the anthropic principle states that our universe would not be able to sustain life as we know it today.

Such a specific value of κ introduces a *fine-tuning* in our theory leading us to a less natural explanation of our observable universe. In order to avoid the fine-tuning, new physics needs to be introduced somewhere between 10^2 GeV and 10^{18} GeV in order to reduce the value of ultraviolet cutoff for the Standard Model. Let us look at the radiative corrections once again. For the term $\mathcal{O}\left(m_f^2 \log(\Lambda_{UV}^2/m_f)\right)$ in Eq 1.2, the variable of interest is m_f. Taking m_f^2 to be as large as it can get(the mass of top quark), which is again of the order of electro-weak scale($\approx 10^2$ GeV). The correction becomes,

$$\frac{\delta m_h^2}{m_h} \propto \frac{m_t^2}{m_h} \log \frac{\Lambda_{UV}}{m_t}. \tag{1.5}$$

That is, assuming that we find a way to cancel out the first term in Eq. 1.2. A simple order of magnitude calculation of the above equation tells us that,

$$\frac{\delta m_h}{m_h} \approx 10 \tag{1.6}$$

given the m_h and m_t at electroweak scale and $\Lambda_{UV} \approx M_P$. To put it simply, if we can make the $\mathcal{O}(\Lambda_{UV}^2)$ term go away, the observed mass of Higgs boson would be within an order of magnitude of the expected mass computation value, making the observed relation in Eq. 1.4, the correct one.

So now, all we need to do is to reduce the first term of Eq. 1.2. The only way to do that would be to bring Λ_{UV} down to the electroweak scale. An elegant way of solving the problem is to introduce a new symmetry that would protect the Higgs mass from the higher-order corrections. This is the principle of supersymmetry

(SUSY). The idea is similar to introducing chiral symmetry in QED which forbids the self-energy diagrams involving photons responsible for large mass corrections to fermions[7].

Apart from SUSY, other proposed solutions include *extra-dimensions model* where the energy scale of gravity is of $\mathcal{O}$(TeV) but since the other forces don't work in the extra dimensions, the energy scale appears to be much larger[8]. There is also a composite Higgs hypothesis which states that the Higgs constituents are responsible for dynamically generating the mass that eliminates the fine-tuning problem[9]. However, as we will see in section 1.3, including the principle of SUSY in the Standard Model does much more than solving the Higgs mass problem.

Dark Matter

- *Galaxies spin faster than we can explain*

Unlike subatomic physics which began in the late 1800s, laws describing our vast Universe were known much earlier. Kepler's laws of planetary motion helped us understand and predict the orbits of astronomical objects. Kepler's third law suggests a relation between the rotational speed of an object at a distance r from the center of the galaxy and the mass distribution $m(r)$ of the galaxy as:

$$v(r) \propto \sqrt{\frac{m(r)}{r}}. \tag{1.7}$$

Long-standing observations of light emitted by luminous objects in a galaxy and their mass-to-luminosity ratios suggested that most of the mass of a spiral galaxy is located at the center. Within the central region, the total mass enclosed in an orbit would grow with the volume enclosed by the orbit. Whereas outside the center, the luminous mass can be neglected and the enclosed mass is approximately constant[10].

$$v(r) \propto \begin{cases} r, & \text{near the center,} \\ \frac{1}{\sqrt{r}}, & \text{outside the center.} \end{cases} \tag{1.8}$$

This was believed to be true until the 20th century. The advent of radio-wave astronomy allowed us to study Doppler shifts in electromagnetic radiations emitted by galactic matter. The observations do not match with the results derived from Eq. 1.8. As described in figure 1.2, the rotation curves show that luminous galactic matter has roughly constant velocity outside the central part of the galaxy. This implies that a large portion of the mass of a galaxy is non-luminous or "dark" matter[11].

The phenomena of *Gravitational Lensing*, one of the most remarkable implications of general relativity gives another strong hint for presence of the dark matter. Gravitational Lensing states that even though the light is massless, the path traveled by light from the source to the observer can still be affected, or bent by the gravitational field of objects in the way. Analyzing data from distant light sources in our Universe, we should in principle, be able to infer the mass distribution in the path from source to our observation point. An example of such a study is the

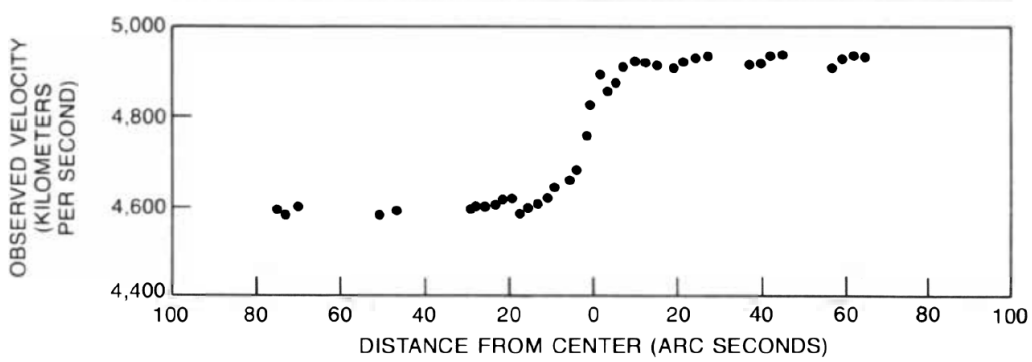


Figure 1.2: Rotational Curve for the galaxy NGC 2998 in Ursa Major cluster, measured from H-alpha emission spectrum[12].

Bullet cluster consisting of two merging galaxies. As seen in figure 1.3, the distribution of luminous matter is peaked at the center of the system, as expected in a collision but, the mass distribution obtained from gravitational lensing suggests two peaks(contours) on either side of the collision center. Since dark matter is hypothesized to not interact via strong or electromagnetic forces, it passes straight through the collision center as opposed to the luminous matter. The presence of non-luminous matter peaks gives a strong argument in the favour of the existence of dark matter in the colliding clusters[13].

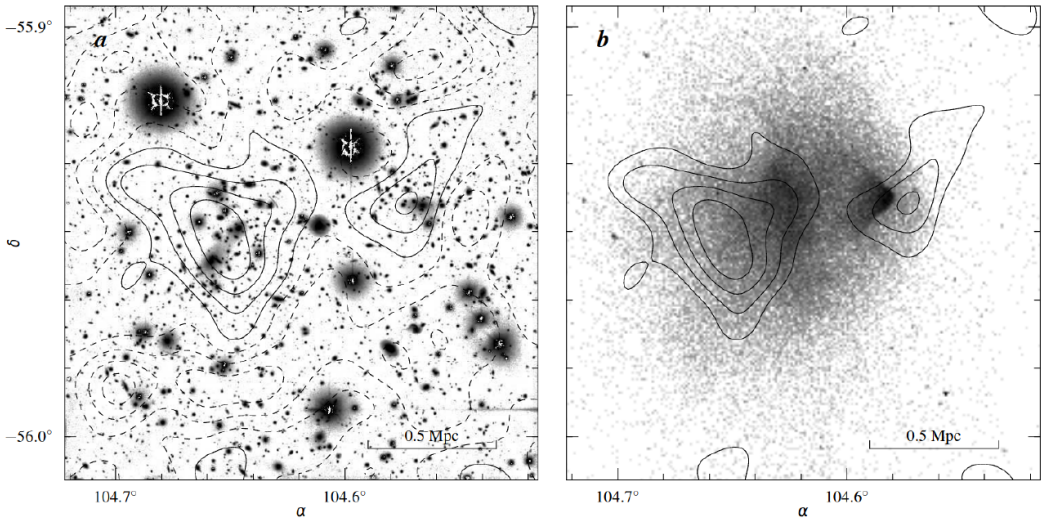


Figure 1.3: Overlaying mass contours from Gravitational lensing in visible *(left)* and in x-ray spectrum *(right)* from Bullet clusters, showing the mass distribution[13].

Another piece of evidence supporting the existence of dark matter comes from the anisotropy of the cosmic microwave background(CMB). CMB originated at a time in the early universe when the density of our Universe was decreasing with expansion, allowing photons to escape from the Big Bang plasma. Before escaping, the photons were confined to areas of high density, a measurement of temperature fluctuations in the CMB allows us to create a map of baryonic structures of the early

Universe. Measurements of CMB via satellites show peaks at higher-order moments indicating the presence of neutral massive particles that allowed the formation of structures via gravitational interaction[10], shown in figure 1.4.

The Λ_{CDM} model, also known as the *Standard Model of Cosmology* is able to describe the structure of CMB by including non-relativistic uncharged matter, or *Cold Dark Matter*.[1] The Λ_{CDM} model suggests that baryonic matter in our Universe consists of only 26.6%, and the rest is made up of the hypothesized cold dark matter[14].

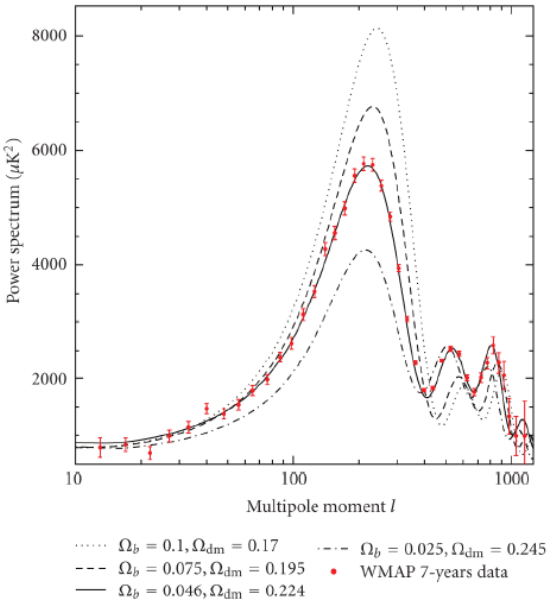


Figure 1.4: WMAP seven-year data[15] along with various baryonic (Ω_b) and dark matter (Ω_{dm}) fractions showing the effects on CMB anisotropy[10].

WIMP, or Weakly Interacting Massive Particle is a prime dark matter[2] candidate[17]. It may seem that SM neutrinos would be good dark matter candidates given that:

- they do not interact via strong and electromagnetic forces
- discovery of neutrino oscillations makes neutrinos massive.

However, the small upper bound on neutrino mass makes them relativistic quite easily and are not viable candidates for cold dark matter. Thus, the WIMP solution to the dark matter problem points towards a phenomenon beyond the Standard Model(BSM) physics discussed further on in section 1.3.

[1] Models with relativistic dark matter have failed to describe the clustering of matter with varying scales in our Universe

[2] Alternate solutions for effects of dark matter include modification of Newtonian dynamics or *MOND*[16]. However, MOND and similar theories that modify mathematical description of gravity have historically been less successful compared to the Λ_{CDM} model that assumes the existence of dark matter

1.3 Supersymmetric Extension of the Standard Model

Supersymmetry(SUSY) is an extension of the space-time symmetries that treat boson and fermions equally[18]. The idea originated in the late 60s and early 70s independently in the Soviet Union and the West and became a hot topic because the new symmetry could potentially solve the naturalness problem in the Higgs sector as discussed in section 1.2.

Overview

Similar to spacetime and gauge symmetries, SUSY must be manifested in the Lagrangian of the theory. Particles related by SUSY transformation can be arranged in *supermultiplets* with the following properties:

- All the members of a supermultiplet must share the same properties except for spins that must differ by 1/2 among the bosons and the fermions;
- When a supermultiplet contains a spin 1/2 fermion and a spin 1 boson, both the chiral components of fermion must transform according to the same gauge group as the boson;
- Within a supermultiplet, the total number of bosonic and fermionic degrees of freedom must match exactly.

As a consequence of the first property, the formation of supermultiplets containing only the SM particles is forbidden. A rather important result following the second property is that the supersymmetric partners undergo the same interaction as their SM counterparts. Additionally, by following the third property, the simplest possible supermultiplet with an SM fermion must contain two scalar fields(or one complex field) and is called *chiral supermultiplet*.

Chiral supermultiplet consisting of a fermion, ψ and two scalar bosons ϕ_1 and ϕ_2 is enough to build a basic supersymmetric theory called the Wess-Zumino model[19]. The Wess-Zumino Lagrangian excluding the auxillary fields[3] can be written as follows:

$$\mathcal{L}_{WZ} = \frac{1}{2}\Big(i\bar{\psi}\partial\!\!\!/\psi + (\partial_u\phi_1)^2 + (\partial_u\phi_2)^2\Big) \tag{1.9}$$

with SUSY transformations of the fields as,

$$\begin{aligned} \psi &\rightarrow \psi + \partial\!\!\!/\phi_1\alpha + i\partial\!\!\!/\phi_2\alpha, \\ \phi_1 &\rightarrow \phi_1 + i\bar{\alpha}\gamma^5\psi, \\ \phi_2 &\rightarrow \phi_2 - i\bar{\alpha}\psi, \end{aligned} \tag{1.10}$$

where α is the parameter of transformation, a Grassmann number(anti-commuting spinor) that ensures dimensionality. The Wess-Zumino Lagrangian of Eq. 1.9 is

[3] Auxillary fields don't correspond to physical degrees of freedom

invariant under these SUSY transformations on-shell.

Similar to fermions, the SM vector bosons are placed alonside their spin 1/2 superpartners, *gauginos* and the group is called *vector supermultiplet*. A supersymmetric lagrangian with a vector supermultiplet will be of the form;

$$\mathcal{L}_v = -\frac{1}{2}F^a_{\mu\nu}F^{a,\mu\nu} + \lambda^{a,\dagger}\bar{\sigma}^\mu\Big(\partial_\mu\lambda^a + g\sum_{b,c} f^{abc}A^b_\mu\lambda^c\Big) \tag{1.11}$$

, where A_μ is vector boson with $F_{\mu\nu}$ as kinetic term. λ is the gaugino and σ^μ is a Grassmann number with a, b and c indices running over all the generators of the corresponding gauge group. The SUSY transformations for vector multiplet that mixes vector bosons and sfermions are as follows:

$$\begin{aligned} A^a_\mu &\to A^a_\mu - \frac{1}{\sqrt{2}}\Big(\epsilon^\dagger\bar{\sigma}_\mu\lambda^a + \lambda^{a,\dagger}\bar{\sigma}_\mu\epsilon\Big), \\ \lambda^a &\to \lambda^a\frac{i}{2\sqrt{2}}\Big(\sigma^\mu\bar{\sigma}^\nu\epsilon\Big)F^a_{\mu\nu}. \end{aligned} \tag{1.12}$$

Here, ϵ in the paramter of transformation and the transformations mentioned in Eq. 1.10 mixes the gauginos and the SM bosons giving a supersymmetric Lagrangian. Let us define Q as a generator of the SUSY transformations such that;

$$\begin{aligned} Q\,|f\rangle &= |b\rangle, \\ Q\,|b\rangle &= |f\rangle, \end{aligned}$$

where, $|f\rangle$ and $|b\rangle$ represent fermionic and bosonic states respectively. In order for Q to be a SUSY generator, it must be an anti-commuting spinor as formally, the properties "bosonic" and"fermionic" are part of graded algebra($\mathbb{Z}_2 = \{1, -1\}$) with even degrees(grading 1) of the algebra are "bosonic" and odd degrees(grading -1) are "fermionic". Following the grading, if I need to convert an odd(fermionic) to even(bosonic) grading or vice versa, the defining property of the generator needs to be odd, or fermionic.

Now that we have established Q to be a spinor, its hermitian conjugate $Q^\dagger$ must also be a fermionic symmetry operator. Both Q and $Q^\dagger$ change the spin angular quantum number of state to which they are applied, making SUSY a spacetime symmetry and an extension of the Poincare symmetry. Coleman and Mandula in 1967 with their *no-go theorem*[20] stated that the symmetry group of a spacetime consistent with QFT must only contain Poincare group along with internal symmetry group which would be devastating for SUSY transformations if it were not for the Haag-Lopuszanski-Sohnius extension[21] which showed that the possible symmetries of a consistent 4-dimensional quantum field theory can consist of a non-trivial supersymmetric extension of Poincare group along with the internal symmetries if the additional generators are fermionic in nature. Hence, the symmetry operators are spinors not by coincidence, rather it is the only possibility! Obeying the no-go theorem and simultaneously allowing for parity-violating interactions observed in the SM, the SUSY generators satisfy the following relations;

$$\begin{aligned} \{Q, Q^\dagger\} &= P^\mu, \\ \{Q, Q\} &= \{Q^\dagger, Q\dagger\} = 0, \\ [P^\mu, Q] &= [P^\mu, Q^\dagger] = 0, \end{aligned} \tag{1.13}$$

where P^μ is the four-momentum generator of spacetime translations.

Names		**spin 0**	**spin 1/2**	$SU(3)_C \times SU(2)_L \times U(1)_Y$
squarks, quarks $\times$ 3 families	Q	$\left(\tilde{u}_L \quad \tilde{d}_L\right)$	$\left(u_L \quad d_L\right)$	$\left(\mathbf{3}, \mathbf{2}, \frac{1}{6}\right)$
	$\bar{u}$	$\tilde{u}_R^*$	u_R	$\left(\bar{\mathbf{3}}, \mathbf{2}, -\frac{2}{3}\right)$
	$\bar{d}$	$\tilde{d}_R^*$	d_R	$\left(\bar{\mathbf{3}}, \mathbf{2}, \frac{1}{3}\right)$
sleptons, leptons $\times$ 3 families	L	$\left(\tilde{\nu} \quad \tilde{e}_L\right)$	$\left(\nu_L \quad e_L\right)$	$\left(\mathbf{1}, \mathbf{2}, -\frac{1}{2}\right)$
	$\bar{e}$	$\tilde{e}_R^*$	e_R	$\left(\bar{\mathbf{1}}, \mathbf{1}, 1\right)$
Higgs, higgsinos	H_u	$\left(H_u^+ \quad H_u^0\right)$	$\left(\tilde{H}_u^+ \quad \tilde{H}_u^0\right)$	$\left(\mathbf{1}, \mathbf{2}, \frac{1}{2}\right)$
	H_d	$\left(H_d^+ \quad H_d^0\right)$	$\left(\tilde{H}_d^0 \quad \tilde{H}_d^-\right)$	$\left(\mathbf{1}, \mathbf{2}, -\frac{1}{2}\right)$

Table 1.1: Chiral supermultiplets in the MSSM[18]

Minimal Supersymmetric Standard Model

Minimal Supersymmetric Standard Model(or *MSSM*) is the minimal extension of the standard model employing only the chiral and vector supermultiplets discussed before. The scalar partners of SM fermions are prefixed by *s* to their name(e.g. *stop*, *selectron*, ...). Whereas, spin-1/2 partners of SM bosons get a suffix *-ino* to their name (e.g. *higgsino*, *Wino,* ...). Symbols denoting sparticles are the same as their SM counterparts except for a tilde (e.g. $\tilde{u}$, $\tilde{e}$, ...). According to the third property, a supermultiplet must contain equal number of bosonic and fermionic degrees of freedom which can be seen in tables 1.1 and 1.2.

Names	**spin 0**	**spin 1/2**	$SU(3)_C \times SU(2)_L \times U(1)_Y$
gluino, gluon	$\tilde{g}$	g	$\left(\mathbf{8}, \mathbf{1}, 0\right)$
winos, W bosons	$\tilde{W}^\pm \; \tilde{W}^0$	$W^\pm \; W^0$	$\left(\mathbf{1}, \mathbf{3}, 0\right)$
bino, B bosons	$\tilde{B}^0$	B^0	$\left(\mathbf{1}, \mathbf{1}, 0\right)$

Table 1.2: Vector supermultiplets in the MSSM[18]

The chiral supermultiplets in table 1.1 consist of Weyl fermions with two spin helicity states and two real scalar fields with one degree of freedom each. The chiral supermultiplet includes SM fermions along with their spin-0 partners squarks ($\tilde{q}$) and sleptons ($\tilde{l}$). The left and right-handed fermions are part of different supermultiplets and hence, have their own superpartners(e.g. $\tilde{e}_L$ and $\tilde{e}_R$). Similarly, vector supermultiplets in table 1.2 consist of spin-1 vector bosons with 2 helicity states and

their spin-1/2 superpartners with two helicity states called *gauginos*.

The Higgs boson lives in the chiral multiplet as it is a scalar boson. However, 2 Higgs doublets are required in order to be able to give mass to both u and d quarks, resulting in two $SU(2)_L$ doublets with $Y = \pm 1/2$ called H_u and H_d respectively. In all, Higgs doublets generate 8 degrees of freedom, 3 of which are used to generate masses for electroweak gauge bosons. The remaining lead to 5 different Higgs bosons, two CP even states h^0 and H^0, one CP odd state A^0, and two charged bosons $H^{\pm}$. Where h^0 is the SM Higgs boson.

After electroweak symmetry breaking, MSSM predicts particles with no equivalence in the Standard Model. The two charged winos ($\tilde{W}^{\pm}$) and two charged higgsinos ($\tilde{H}_u^+$, $\tilde{H}_d^-$) form two charged mass eigenstates called *charginos* ($\tilde{\chi}_1^{\pm}$, $\tilde{\chi}_2^{\pm}$). The neutral wino ($\tilde{W}^0$) and bino ($\tilde{B}^0$) mix with the remaining neutral higgsinos ($\tilde{H}_u^0$, $\tilde{H}_d^0$) to give four neutral mass eigen states called *neutralinos* ($\tilde{\chi}_1^0$, $\tilde{\chi}_2^0$, $\tilde{\chi}_3^0$ and $\tilde{\chi}_4^0$ increasing in mass).

Higgs Mass Solution

One of the basic predictions of SUSY is the presence of particles heavier than top quark. If we consider a heavy scalar particle m_S in Eq. 1.2,

$$\delta m_{h,s}^2 = \frac{\lambda_S}{16\pi^2}\Lambda_{UV}^2 + \mathcal{O}\left(m_S^2 \log \frac{\Lambda_{UV}^2}{m_S}\right). \tag{1.14}$$

Notice the sign difference in the first term with respect to Eq. 1.2,

$$\delta m_{h,f}^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{UV}^2 + \mathcal{O}\left(m_f^2 \log \frac{\Lambda_{UV}^2}{m_f}\right), \tag{1.15}$$

this is an effect of supersymmetric extension of the Standard Model. According to the Minimal Supersymmetric Extension of the Standard Model(MSSM) each of the associated scalars of the fermions that are the members of the chiral supermultiplet also contributes to the Higgs mass along with the Standard Model fermions. Hence, we have to include contributions from both Eq. 1.14 and Eq. 1.15. Comparing the $\mathcal{O}(\Lambda_{UV}^2)$ terms from the two equations, we see that,

- they have opposite signs,
- they differ by a factor of 2 if ($\lambda_S = |\lambda_f|^2$)[4]

At the one-loop level, each chiral supermultiplet will contribute two additional diagrams along with the top contribution as mentioned in figure 1.5. The $\mathcal{O}(\Lambda_{UV}^2)$ contributions from the three diagrams sum up to zero. This is the result of the new symmetry introduced in MSSM and hence, the results hold at all perturbative orders and for all the fermions.

[4]This relation holds true in unbroken and broken SUSY

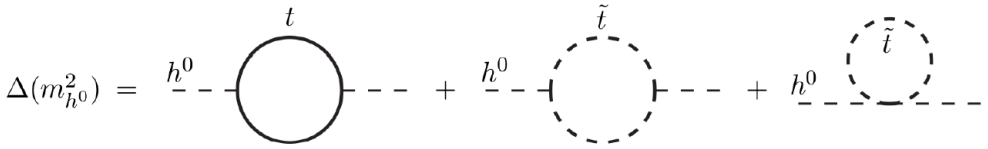


Figure 1.5: One-loop contributions from top quark supermultiplet to Higgs mass[18]. Notice the signs of each diagram according to Eq. 1.14 and Eq. 1.15.

Introducing SUSY breaking will make the scalar mass contribution as the dominant correction. To keep fine-tuning at a minimum, scalar masses should not be too far above the EW scale. Since the highest mass fermions are top $\approx$ 175 GeV and bottom $\approx$ 4 GeV, stop and sbottom should cause the majority of the remaining corrections to the Higgs mass.

SUSY Breaking and R-parity Conservation

From Eq. 1.13, we can see that SUSY generators commute with spacetime translations, hence, they will also commute with the squared mass operator $-P^2$. In the Minimal Supersymmetric Standard Model(MSSM), the defining properties of a supermultiplet suggest that all the members of a supermultiplet should have the same properties except for spin. Hence, both the super-particle states of a supermultiplet must have the same eigenvalue under $-P^2$. Furthermore, SUSY generators commute with generators of gauge transformation hence supersymmetric particles, or *sparticles* must have the same electric charge, weak isospin, and color charge as the SM particle in the supermultiplet. This is clearly not the case as we haven't detected any sparticles in the energy range of SM particles. SUSY needs to be broken for superpartners to be heavier than the SM particles. However, SUSY needs to break in a way that it doesn't affect the cancellation of quadratic divergences in calculating the Higgs mass mentioned in section 1.2.

In MSSM, this is achieved via *soft susy breaking*;

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{soft}, \tag{1.16}$$

Where, $\mathcal{L}_{\text{SUSY}}$ is part of the Lagrangian that is invariant under supersymmetric transformations and $\mathcal{L}_{soft}$ is responsible for giving masses to sparticles in the TeV range by breaking SUSY[18].

Renormalisability and gauge invariance of superpotential in its most general form, allow for terms that violate Baryon number (B) and Lepton number (L). Since these violation are directly related to proton decay[22], R-parity is introduced as an additional symmetry to get rid of the L- and B- violating terms;

$$P_R = (-1)^{3(B-L)+2s}. \tag{1.17}$$

It can be easily verified that R-parity is positive for SM particles and negative for sparticles. Along with suppressing L- and B- violating terms, R-parity implies

that SUSY particles can only be produced in pairs and each vertex in MSSM must contain an even number of SUSY particles. Given that sparticle cannot decay solely into an SM particle, the lightest supersymmetric particle(or *LSP*) must be stable. Lightest Neutralino ($\tilde{\chi}_1^0$) discussed earlier can be a possible LSP. Considering that LSP only interacts weakly, $\tilde{\chi}_1^0$ can be an excellent candidate for dark matter. Along with that, LSP or $\tilde{\chi}_1^0$ must follow additional properties in order for the extended model to be natural. Considering the electroweak scale, strong upper limits can be placed on the mass of $\tilde{\chi}_1^0$ no greater than 2.5 TeV for a wino-type and 1.8 TeV for a Higgsino-type[23].

1.4 Experimental Status

With 100+ free parameters on top of the SM, the MSSM phase space is beyond comprehension to be considered in its entirety. Some cleverness is required to organize constraints during the searches. By applying a series of constraints that are designed to discard models that are not phenomenologically viable, the parameter space can be reduced to 19 called phenomenological MSSM or *p-MSSM* approach. The constraints used to reduce the parameters are as follows[24]:

- The parameters arising from soft SUSY breaking must be real that avoids any new sources of CP violation.
- No additional sources of flavor-changing neutral currents (FCNC).
- The first two generations of sfermions must have degenerate masses.

Along with the electroweak symmetry breaking conditions, these constraints reduce the number of parameters to the following 19[24]:

- $\tan\beta$: the ratio of vacuum expectation values (VEVs) of the two Higgs fields.
- M_A: The mass of Higgs boson carrying odd CP.
- μ: the Higgs-higgsino mass parameter.
- M_1, M_2, M_3: mass parameters for bino, wino and gluino.
- $m_{\tilde{q}}$, $m_{\tilde{u}_R}$, $m_{\tilde{d}_R}$, $m_{\tilde{l}}$, $m_{\tilde{e}_R}$: degenerate masses of first and second generation sfermions.
- $m_{\tilde{Q}}$, $m_{\tilde{t}_R}$, $m_{\tilde{b}_R}$, $m_{\tilde{L}}$, $m_{\tilde{\tau}_R}$: degenerate masses of first and second generation sfermions.

Although the pMSSM approach is quite useful, the statistical tests employed during the process are quite resource-intensive. Moreover, we wish to have a model that can be employed generally to all the possible SUSY producing channels hence we are inclined towards Simplified Models that are more suitable for individual searches.

Simplified Models

Simplified models[25][26] is a way to reduce the parameter space to a more manageable level. In the model, all the SUSY particles that do not participate in the decay are set to be kinematically inaccessible. Masses of the produced SUSY particles are considered as free parameters. The simplified model mentioned in figure 1.6 can be further classified in two categories based on the intermediate SUSY particles produced from the proton collisions;

- Electroweak SUSY Decay
- Strong SUSY Decay

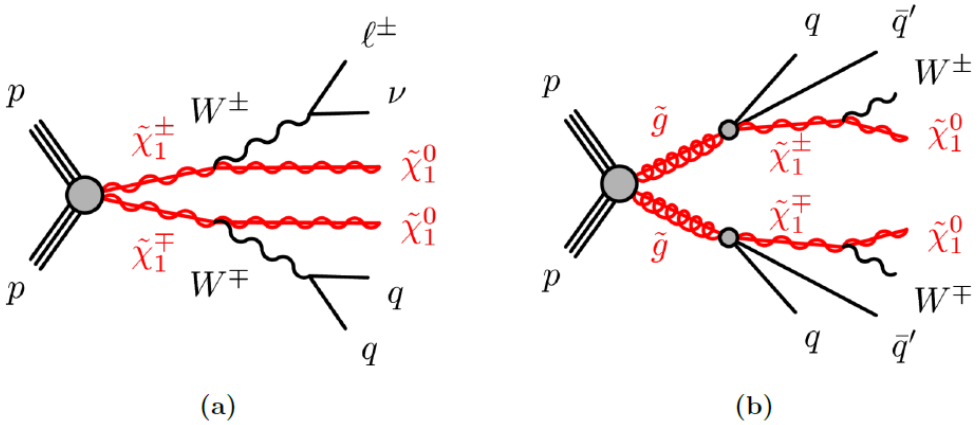


Figure 1.6: The electroweak SUSY decay (a) and the strong SUSY decay (b).

In figure 1.6(a), proton collisions produce a pair of charginos ($\tilde{\chi}_1^{\pm}$) which can further decay into vector bosons and lightest neutralino ($\tilde{\chi}_1^0$), the LSP. Since charginos are a superposition of charged Winos ($\tilde{W}^{\pm}$) and charged higgsinos ($\tilde{H}_u^+$, $\tilde{H}_d^-$) as mentioned in section 1.3, they can only be formed via electroweak process from proton collisions and hence, the name. On the contrary, in figure 1.6(b), a pair of gluinos ($\tilde{g}$) are produced from proton collisions which further decay into next-to leading order charginos ($\tilde{\chi}_1^{\pm}$) and hadronic jets. The charginos furhter decay into lightest neutralinos ($\tilde{\chi}_1^0$) and charged vector bosons. Since gluinos or squarks are produced from the proton collisions, the decay process is termed as strong SUSY decay.

In both the processes, the final signatures that reach our detector comprise the decay products of vector bosons along with LSP in the form of large missing transverse energy (E_T^{miss}). One important point to note here is that the R-parity discussed in section 1.3 is conserved at every vertex of the simplified model. It is worth mentioning that in the electroweak SUSY decay, charginos ($\tilde{\chi}_1^{\pm}$) are not the only possible decay products, a higher mass neutralino ($\tilde{\chi}_2^0$) can also produced along with chargino ($\tilde{\chi}_1^{\pm}$) as observed in figure 1.8, but more on that later.

Lepton signatures in the final decay products are always easier to detect compared to relativistic jets. Hence 1L chargino decays which consist of one $W^{\pm}$ boson

decaying into a lepton (electron or muon) and a neutrino and, the other $W^{\pm}$ decaying hadronically are of particular interest. QCD always exhibits a higher cross-section compared to the electroweak process and hence, the strong SUSY decay, also called the *gluino one-step* model seems to be the obvious choice and there are groups working on it, but our focus has been towards the electroweak 1L SUSY decay for reasons explained further.

Electroweak 1L SUSY decay

The naturalness discussed in the section 1.2 suggests that if the masses of lightest and next to lightest supersymmetric particles (LSP and NLSP) are too large as compared to the top mass, some fine-tuning would be required in the theory in order to explain the gap. Considering gluinos ($\tilde{g}$) or squarks ($\tilde{q}$) as the NLSP or NNLSP would mean that their masses mustn't be too far from the Electroweak scale. Analysis of gluino one-step model have put strong exclusion limits on gluino ($\tilde{g}$) and stop ($\tilde{t}$) masses[27]. Considering this, the 1L EW-SUSY model for chargino/neutralino decay is more favorable despite the lower cross-sections. Furthermore, in the recent years of analysis, enough statistical data has been gathered to counterbalance the low cross-sections making the electroweak sector more interesting at the moment.

Another advantage of EW decay is the lower jet multiplicity. Strong channel predominantly involves multiple hadronic showers at $\tilde{g}$-$\tilde{\chi}_1^{\pm}$ and the $W^{\pm}$-l decay vertices as can be seen in figure 1.6. Since LHC was built as a QCD machine, the detectors are prone to spot the high energy jets which act as a background for SUSY detection, discussed further in section 3.6.

The heavy neutralino NSLP, $\tilde{\chi}_2^0$ and the LSP $\tilde{\chi}_1^0$ produced during the p-p collision are superpositions of Wino ($\tilde{W}^0$), Bino ($\tilde{B}^0$) and Higgsino ($\tilde{H}_u^0$, $\tilde{H}_d^0$) depending on the following mass scenarios where M_1, M_2 and μ are masses of Bino, Wino and Higgsinos respectively:

1. **Bino-Like LSP:**
   (a) $M_1 < M_2 < \mu$: $\tilde{\chi}_1^0$ Bino-like LSP; $\tilde{\chi}_2^0$ Wino-like NLSP.
   Wino-Bino mixing with a decoupled Higgsino by taking $|\mu| \gg M_1, M_2$.
   (b) $M_1 < \mu < M_2$: $\tilde{\chi}_1^0$ Bino-like LSP; $\tilde{\chi}_{2,3}^0$ Higgsino-like NLSPs.
   Bino-Higgsino mixing with a decoupled Wino by taking $M_2 \gg M_1, \mu$.
2. **Wino-Like LSP:**
   (a) $M_2 < M_1 < \mu$: $\tilde{\chi}_1^0$ Wino-like LSP; $\tilde{\chi}_2^0$ Bino-like NLSP.
   Wino-Bino mixing with a decoupled Higgsino by taking $|\mu| \gg M_1, M_2$.
   (b) $M_2 < \mu < M_1$: $\tilde{\chi}_1^0$ Wino-like LSP; $\tilde{\chi}_{2,3}^0$ Higgsino-like NLSPs.
   Wino-Higgsino mixing with a decoupled Bino by taking $M_1 \gg M_2, \mu$.
3. **Higgsino-Like LSP:**
   (a) $\mu < M_2 < M_1$: $\tilde{\chi}_{1,2}^0$ Higgsino-like LSPs; $\tilde{\chi}_3^0$ Wino-like NLSP.
   Wino-Higgsino mixing with a decoupled Bino by taking $M_1 \gg M_2, \mu$.

(b) $\mu < M_1 < M_2$: $\tilde{\chi}^0_{1,2}$ Higgsino-like LSPs; $\tilde{\chi}^0_3$ Bino-like NLSP.
Bino-Higgsino mixing with a decoupled Wino by taking $M_2 \gg M_1, \mu$.

Assuming s-channel Drell-Yan (DY) processes with W/Z exchanges

$$pp \longrightarrow \tilde{\chi}^+_i \tilde{\chi}^-_j X, \quad \tilde{\chi}^\pm_i \tilde{\chi}^0_j X, \quad \tilde{\chi}^0_i \tilde{\chi}^0_j X$$

where X generally denotes hadronic remnants of protons, dominant processes are those involving two Wino-like or two Higgsino-like states since their coupling to W and Z are unsuppressed. The Electroweakino pair production via W-exchange has the largest cross-section due to a large $\mathrm{SU(2)_L}$ coupling.

figure 1.7(a) shows pair production cross sections for Bino-like LSP and Wino-like NLSPs (Scenario 1a). With unsupressed $\mathrm{SU(2)_L}$ couplings, leading production channel are Wino-like NLSPs ($\tilde{\chi}^\pm_1$ and $\tilde{\chi}^0_2$)

$$pp \longrightarrow \tilde{\chi}^\pm_1 \tilde{\chi}^0_2 X, \quad \tilde{\chi}^\pm_1 \tilde{\chi}^\mp_1 X.$$

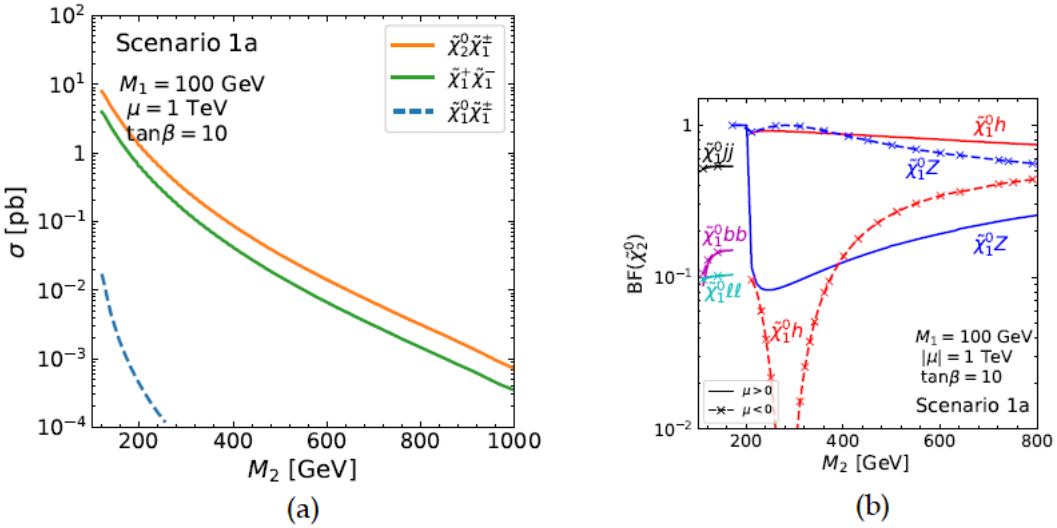


Figure 1.7: (a) Cross sections for production and (b) Decay branching fractions for Wino-like NLSPs taking LSP mass parameter to be 100 GeV, heaviest mass parameter at 1 TeV and $\tan\beta = 10$[28]

As shown in the figure 1.7(a) the cross sections can be of order 1 pb to 1 fm for $M_2 \sim 200$ GeV to 800 GeV. Although kinematically favourable, the Bino-like LSP productions of $\tilde{\chi}^0_1\tilde{\chi}^\pm_1$ and $\tilde{\chi}^0_1\tilde{\chi}^0_2$ are highly suppressed due to Bino-Wino mixing. The Wino NLSPs decay to the LSP $\tilde{\chi}^0_1$ and the SM partners via a mixture of Higgsino states. The branching fraction $\mathrm{BF}(\tilde{\chi}^\pm_1 \to \tilde{\chi}^0_1 W^\pm)$ is 100% but for the case of Wino-like neutralino $\tilde{\chi}^0_2$, there are two competing channels once kinematically accessible,

$$\tilde{\chi}^0_2 \longrightarrow Z\tilde{\chi}^0_1, \quad h\tilde{\chi}^0_1.$$

Depending on the flavor structures of charginos and neutralinos, their cross-section of production and decay branching fraction changes. For example, the decay

branching fractions for Wino-like $\tilde{\chi}_2^0$ are shown in figure 1.7(b) which suggests that depending on the mass of Wino, M_2 and whether μ is positive or negative, the branching fraction changes drastically. Similar analysis can be performed for all the scenarios mentioned above. As both the competing decay scenarios through h and Z boson are possible, considering a branching fraction of 100% in each state helps create a generalized model for understanding the underlying particle properties without positing additional assumptions on the physics of decay. This is the aim of the simplified model approach.

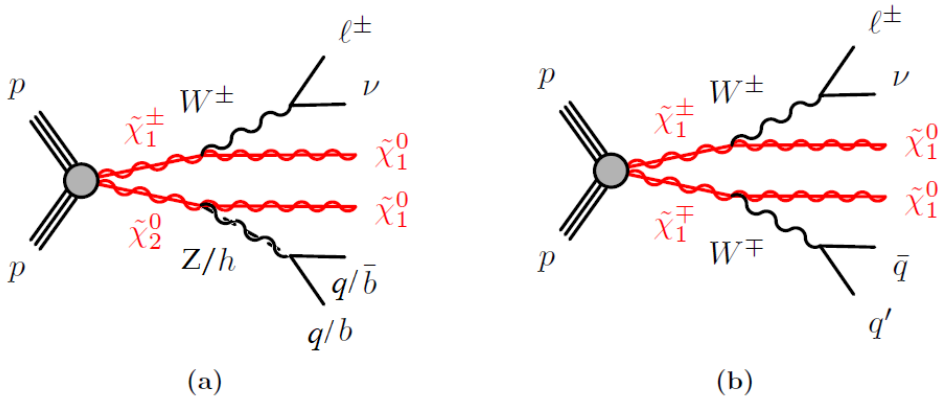


Figure 1.8: Figure (a) shows a specific case of simplified electroweak 1-lepton channel with neutralino decaying into Z boson or Higgs, further into $b\bar{b}$ and a lepton as decay signatures. Figure (b) shows the generic WW model for electroweak SUSY decay previously shown in figure 1.6(a)

Previous analysis of 1-lepton strong channel has shown that 1L final state is a powerful channel for the search of SUSY particles. Figure 1.8 shows a complete model of 1L electroweak SUSY decay inclusive of all possible decay channels used in this analysis.

1.5 Summary

The chapter gave a theoretical background that leads us to search for supersymmetric particles. Starting from the Standard Model and its limitations, we extended the usual spacetime symmetries to include the exchange of fermions and bosons following the principle of supersymmetry (SUSY). We constructed the SUSY Lagrangian using the Wess-Zumino model and grouped the Standard Model particles together with their superpartners into chiral or vector multiplets. We've then discussed the Minimal Supersymmetric Standard Model (MSSM) which makes the hierarchy and dark matter problems potentially solvable. Moreover, considering the simplified SUSY models, we've given our reasons for inclination towards the search for ElectroWeak 1L SUSY events.

With the theoretical support, let us switch our focus to experimental physics, starting with the description of the current most powerful particle accelerator: the

Large Hadron Collider, or LHC, a machine that could lead to observations of supersymmetric particles.

Chapter 2

ATLAS and the LHC

Conseil Européen pour la Recherche Nucléaire, (*CERN*) is a Europe's joint venture in science and one of the largest research centers in the world[29]. Located on the French-Swiss border near Geneva, CERN was founded in 1954 with the main research focus in the area of particle physics.

2.1 Large Hadron Collider

The Large Hadron Collider (*LHC*) is the largest particle accelerator at CERN[30]. The tunnel housing the LHC was originally built for the Large Electron-Positron Collider (LEP) and lies between 175m and 50m below the surface of the earth with a tilt of 1.4% and a circumference of 26.7km[30]. The accelerator is composed of eight arcs and eight straight sections. The arcs are filled with a total of 1232 dipole magnets that keeps the trajectory of the particle beam aligned around the ring[30]. The center-of-mass energy of the accelerator($\sqrt{s}$) is limited by the strength of the dipole magnets. To reach a $\sqrt{s} = 14$ TeV or 7 TeV per beam, the required dipole magnetic field value is 8.3 T[31]. Such high magnetic fields can only be achieved by using superconductive magnets operating at 1.9 K[31]. Alongside dipole magnets, LHC contains quadrupole magnets for beam-shaping, and other magnets for beam-injection and -dumping, resulting in a total of 9593 magnets[30]. The flow of the magnetic fields in the LHC pipes is shown in figure 2.1.

The eight straight sections serve as insertion regions (IR) either for detectors or for the machine hardware of the collider. IR4 contains 16 RF cavities responsible for accelerating the particles. RF cavities accelerate the particles periodically in bunches of length $\approx$ 7.5m. Each beam carries a total of 2808 bunches[30].

The opposing beams are brought to a collision at four of the eight IRs called Interaction Points (IPs). ATLAS (A toroidal detector) is located at IP1 while IP5 which is located on the opposite side of the ring houses Compact Muon Solenoid (CMS) detector. The remaining IPs 2 and 8 located next to ATLAS comprises of the Large Hadron Collider beauty (LHCb) and A Large Ion Collider Experiment (ALICE) as shown in figure2.2(b). ATLAS and CMS are designed to be general-purpose experiments whereas, the LHCb and ALICE specialize in b-physics and heavy-ion physics respectively. Apart from the four main experiments mentioned

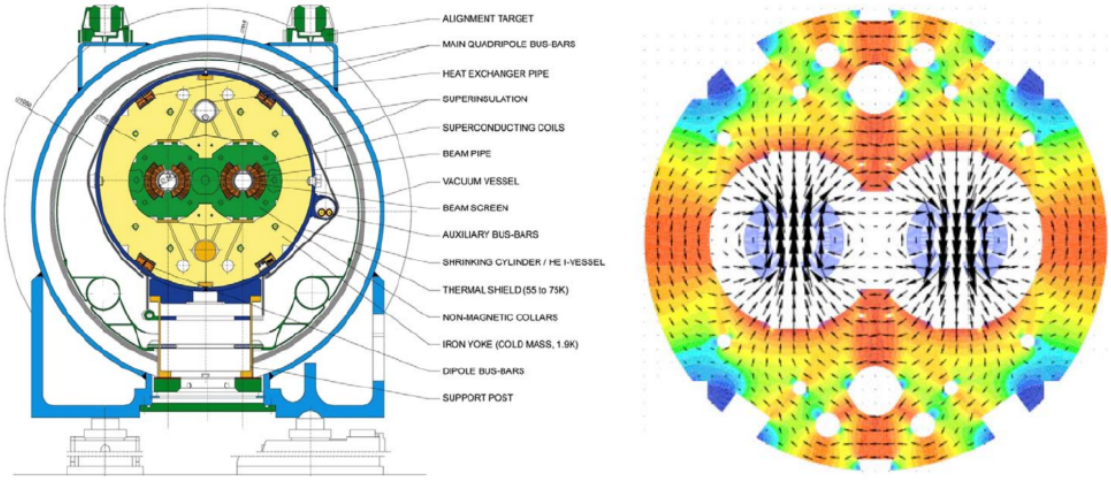


Figure 2.1: (a) Cross-section[31] and (b) the magnetic field of LHC dipole system[32]

above, there are three more experiments making use of the beams of LHC namely; the Large Hadron Collider forward (LHCf), the Total Elastic and Diffractive Cross Section Measurement (TOTEM), and the Monopole and Exotics Detector at the LHC (MoEDAL).

Before being injected into the LHC, particles are first accelerated by the CERN accelerator complex shown in figure 2.2(a). The particles start at the linear accelerator 2 (LINAC 2), go through the BOOSTER ring, and are further accelerated in the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS) before finally reaching the LHC with an energy of 450 GeV. From there-after, LHC is responsible for the acceleration to reach the desired energy for collision. In the first proton-proton run (RUN-1), the LHC reached beam energy of 3.5 TeV(2010-2011) and 4 TeV (2012)[33]. The LHC was given an upgrade before the RUN-2 which boosted the beam energy to 6.5 TeV ($\sqrt{s} = 13$ TeV) in 2015[34], just short of the designed center-of-mass energy of $\sqrt{s} = 14$ TeV[30].

2.2 The ATLAS Experiment

The ATLAS detector (figure 2.3 is a general-purpose LHC detector located IP1, it is 44m long and 25m high and sits roughly 100m below the surface[35].

ATLAS Coordinate System

The origin of the coordinate system used to describe events in the ATLAS detector is at the interaction point. The z axis is longitudinal to the beam, while the x and y axes point to the center of the LHC ring and upwards respectively. The transverse $x - y$ plane is of particular interest since the initial momentum along the z axis is

zero. The azimuthal angle in the transverse plane (around z axis) is denoted by ϕ. Massive objects can be localized in the direction of the beam using their rapidity, Υ;

$$\Upsilon = \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right) \tag{2.1}$$

.

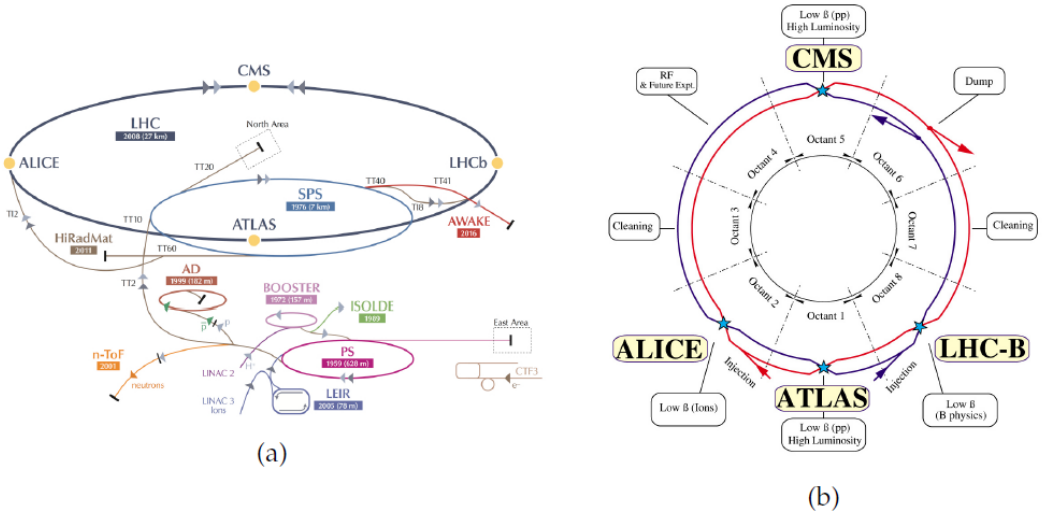


Figure 2.2: (a) the CERN accelerator complex[36] and (b) the LHC layout[37].

For ultra-relativistic particles where the particle's mass is small as compared to its energy,

$$E \approx |\overrightarrow{p_z}| \implies p_z \approx E \cos\theta. \tag{2.2}$$

Substituting in Eq. 2.1, we get rapidity depending only on angle θ between the three-momentum vector ($\overrightarrow{p_z}$) and the z axis;

$$\Upsilon \approx \eta = -\log \left(\tan \left(\frac{\theta}{2} \right) \right), \tag{2.3}$$

where η is called the *pseudo-rapidity*. Pseudorapidity has the advantage to be independent of energy and momentum calibration of the detected objects, and to be in direct correspondence with the polar angle θ. Distance between two objects in the detector are often quantified in η-θ space by the dimensionless variable ΔR:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{2.4}$$

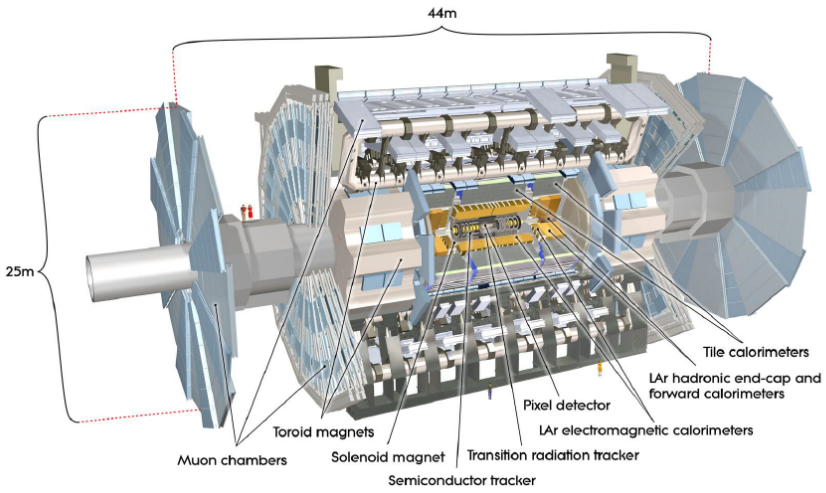


Figure 2.3: The ATLAS detector[38]

Inner Detector

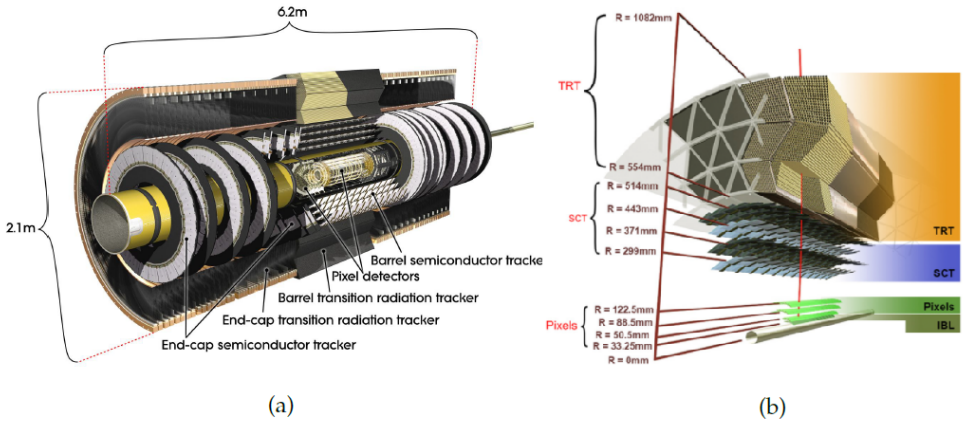
The ATLAS inner detector (figure 2.4) is composed of 3 sub-detectors installed closest to the beam pipe that is responsible for track reconstruction of charged particles in $|\eta| < 2.5$ region. The Inner Detector (ID) is enclosed within a 2T axial magnetic field that forces the charged products of the interaction to move in a helical path away from the collision point. The radius of curvature of the helix will be proportional to the momenta of the particles[14] and hence, measuring the radius can give a measurement of the particle momenta.

$$R \propto \frac{p \cos \lambda}{B} \tag{2.5}$$

Here B is the magnetic field, p is the momenta, λ is the angle of helical trajectory and R is the radius of curvature. In order to have a high accuracy in the calculation of particle momenta, an accurate measurement of the track's curvature must be obtained.

The innermost layer of the ID is composed of silicon pixel detectors. Originally the pixel detector consisted of three layers built for resilience towards the high-radiation environment close to the beam pipe while giving high spatial resolution for vertex identification. Before the RUN-2 of LHC, a new innermost pixel layer Insertable B-Layer (IBL) was added[39]. The pixel detector provides a spatial resolution of 10 μm in bending plane (r-ϕ) and 115 μm in the longitudinal (z) direction[35].

The second subsystem if the ID named Semi-Conducting Trackers (SCT), provide an additional source of particle tracking with a spatial resolution of 17 μm and 580 μm in the r-ϕ and the z direction respectively. Unlike the pixel detectors, silicon strips only provide a 1D measurement and hence they are assembled in pairs laid on top of each other to provide two-dimensional hit information. Although the measurements are less precise compared to the pixel detector, four additional at a much lower cost provides a good cost-benefit compromise.



(a) (b)

Figure 2.4: (a)Longitudinal[40] and (b) Radial[41] cross-sections of the ATLAS inner detector

Third and the largest of the ID subsystems is the Transition Radiation Tracker (TRT) that uses 4mm wide drift tubes for particle tracking. Tubes are oriented parallel to the beam pipe and are filled with a gas mixture[35]. The transitioning charged particles release EM radiation that interacts with the gas inside the tube and produces a cascade that can be detected. This provides additional electron identification in $|\eta| < 2$ region[35].

The Calorimeters

The ATLAS ID provides a precise estimation of the particle momenta however, it doesn't give particle identification. Moreover, ID is only sensitive to charged particles hence neutral hadronic bound states do not get detected in ID. These shortcomings are covered by two types of calorimeters employing the technique of sampling calorimetry;

- Electromagnetic Calorimeter (ECal)
- Hadronic Calorimeter (HCal)

Both the calorimeters are comprised of alternate layers of active and absorbing material. The traversing particles deposit their energy in the absorbing material and in the process create particle showers. The following active material detects the shape and intensity of the produced showers.

The ECal uses lead plates as absorbing material and liquid Argon as active material. ECal is located next to the inner detector and as the name suggests, it is responsible for measuring energies of electromagnetically interacting particles. It has a thickness of > 22 radiation lenghts[1] in the barrel and > 24 radiation lengths in the end-caps[35]. The energy resolution for electron in ECal is measured at σ_E/E $\sim 10\%/\sqrt{E} \oplus 170$ MeV/E $\oplus$ 0.7%.

[1]defined as the length after which the energy of the traversing particle reduces to $1/e$ of the initial energy

The HCal consists of a steel-scintillator barrel, copper/liquid Argon end caps, and copper-tungsten/liquid Argon forward calorimeters giving a total resolution in $|\eta| < 4.9$ region. HCals typically have a worse energy resolution when compared with the ECals since hadronic cascades are complex showers compared to the single incident charged particles in ECals. The energy resolution of HCal using pion test beam is $\sigma_E/\mathrm{E} \sim 50\%/\sqrt{E} \oplus 6\%$.

The Muon Spectrometer

The energy lost by a particle in a certain region is determined by the stopping power ($-dE/dx$) which depends on the material density and the $\beta\gamma = p/m$ factor of the inbound particle. In p-p collisions, muons are produced from the decay of W/Z or top quarks and hence have an energy range in GeV. As can be seen in figure 2.5, the stopping power of muon has a minimum in the GeV energy range therefore, muons pass through calorimeters losing a minimal amount of energy. Taking advantage of this, a muon spectrometer is installed surrounding the hadronic calorimeter. It comprises four types of detectors; Monitored Drift Tube (MDT) chambers are responsible for obtaining tracking information with a precision of 35 μm per chamber[35]. Cathode Strip Chambers (CSC), which are multi-wire proportional chambers used in regions with uneven magnetic fields and high muon rates. The Resistive Plate Chambers (RPC) in the barrel and Thin Gap Chambers (TGC) in the end-caps together make the muon trigger system. The muon reconstruction efficiency is measured to be close to 99% over most of the covered phase space ($|\eta| < 2.5$ and $5 < p_T < 100$ GeV)[42]. The p_T resolution in data and simulation agree to better than 5% for most of the η range.

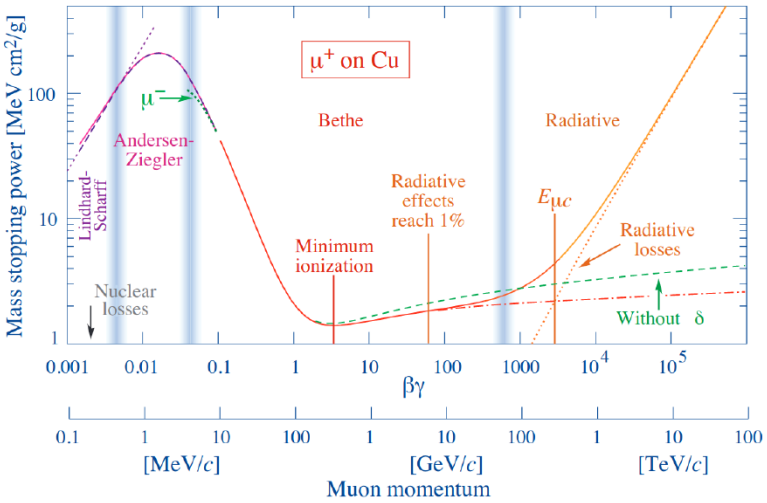


Figure 2.5: Stopping power of muon in copper

The Trigger and Data Acquisition System

The LHC has bunch spacing of 25 ns that corresponds to an event rate of 40 MHz[35]. With a mean event size of ≈ 1.3 MB, the rate at which ATLAS data can be obtained

amounts to ≈ 60 TB/s. Bandwidth and storage consideration forced us to develop an intermediate triggering system that brings the readout rate down to an acceptable level.

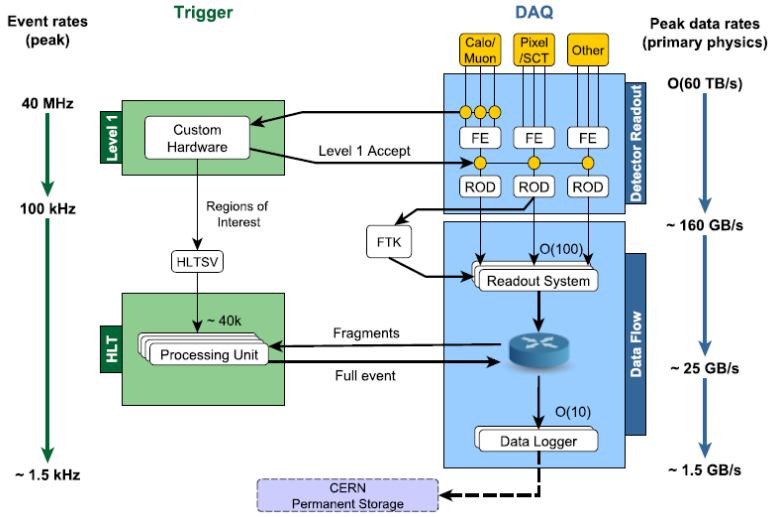


Figure 2.6: The ATLAS Trigger and Data Acquisition System[43]

The ATLAS Trigger and Data Acquisition (TDAQ) system shown in figure 2.6 consists of level 1 (L1) and High-Level Trigger (HLT) farm. L1 is implemented in the hardware that identified the Region of Interest (ROI) in η-ϕ space and reduces the data rates to 100 kHz or 160 GB/s. These ROIs are sent to HLT farm housed on an array of computers situated near the detector that performs complex event analysis and brings the readout rate to 1.5 kHz or 1.5GB/s.

Instantaneous beam luminosity gradually reduces with time due to collision losses and beam quality degradation. This can be seen in figure 2.7. In the beginning, the raw rates of some triggers are too high for the system to deal leading to "dead time" during which events are lost. To mitigate the loss of data, some triggers are *pre-scaled*; a pre-scaled trigger will drop a fraction of the acceptable events at a rate corresponding to the pre-scaled factor which gradually reduces to zero.

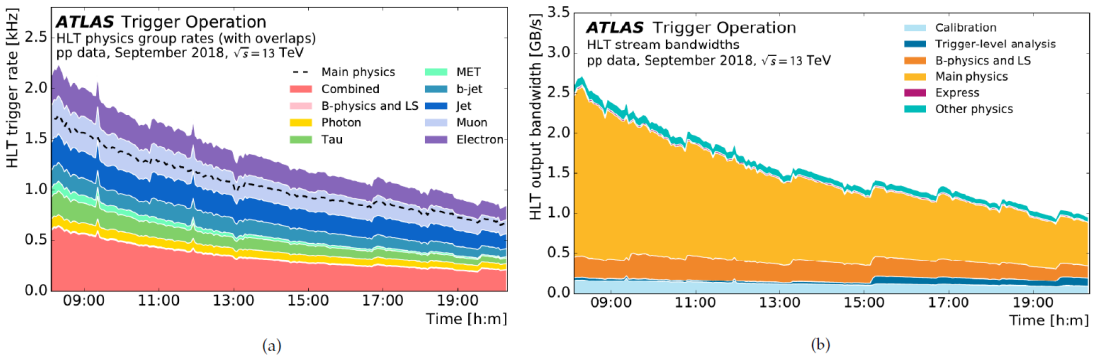


Figure 2.7: HLT (a)trigger and (b) data rates showing luminosity drops with time and discontinuities corresponding to changes in trigger pre-scaled factors[44]

Both L1 and HLT trigger select events based on physics observables such as single, high p_T leptons or jets, or energetic photons. Triggers based on missing transverse energy E_T^{miss} are of particular interest for R-parity conserving (RPC) SUSY. As discussed in section 3.6 of chapter **??**, the E_T^{miss} of an RPC SUSY event is largely due to massive neutralinos which can be used to apply lower bounds on E_T^{miss} of the readout events. figure 2.8 shows almost 100% efficiency for $E_T^{miss} > 200$ GeV.

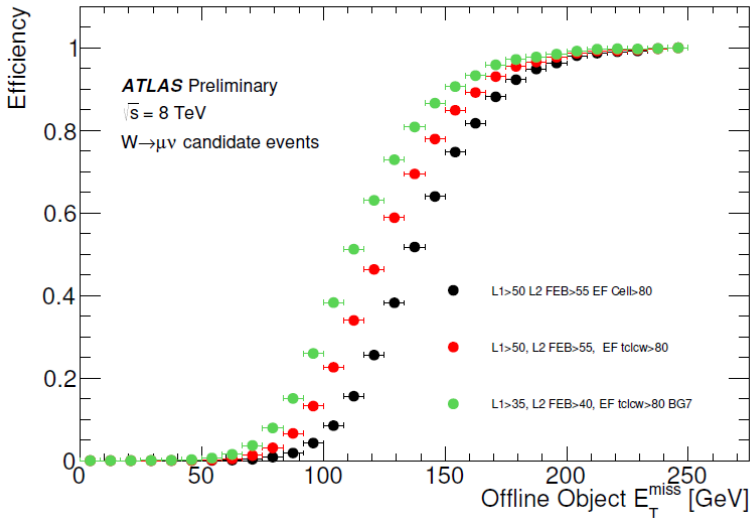


Figure 2.8: E_T^{miss} trigger efficiency calibrated to detect RPC SUSY events[45]

2.3 Summary

We started our discussion with an overview of LHC and then moved on to discuss one of the eight LHC detectors, ATLAS which is of primary interest for this thesis. ATLAS consists of:

- a tracking detector (Inner Detector) that enables estimation of particle momenta and reconstruction of interaction vertices.
- an array of electromagnetic and hadronic calorimeters that measure particle energies
- a muon spectrometer that provides triggering and estimation of muon momenta
- a sophisticated trigger and data acquisition system that reduces the data and event rates to a more manageable level.

The upcoming chapter discusses the importance of Monte Carlo simulation in our analysis.

Chapter 3

Analysis Strategy

After an overview of the LHC and the ATLAS detector, this section focuses on how the data is collected and introduces the statistical methods and the analysis strategy adopted for this thesis. The idea is to have a blind analysis where the collected data will only be used at the final step of the analysis to find possible SUSY events, and hence we require an alternate source with large number of collision events which can be used for modelling and testing our methods of investigation. This source is provided by the Monte Carlo simulation.

3.1 Data

The raw number of events recorded by the ATLAS detector is expressed in the form of the integrated luminosity as follows:

$$N = \sigma \mathcal{L} = \sigma \int L \, dt, \tag{3.1}$$

where N is the total number of raw events, σ is the cross-section of the processes in the study and L is the instantaneous luminosity of the LHC at the interaction point IP1[31]. The data available for the analysis is in general less than the number of events N obtained here because of the detector limitations and hence, are termed as raw events. The instantaneous luminosity is given by

$$L = \frac{N_b^2 n_b f_{col} \gamma_r}{4\pi \epsilon_n \beta^*} F, \tag{3.2}$$

where N_b is the mean number of particles in a bunch, n_b denotes number of bunches in the beam, f_{col} is the collision frequency, γ_r is the Lorentz factor, ϵ_n is the normalized transverse beam emmitance and β^* is the beta function at the collision point. F is a geometric factor that indicates the reduction in luminosity due to crossing angle at the interaction point. The maximum possible instantaneous luminosity by design at the IP1 (ATLAS) for p-p collisions is $L = 10^{34}\text{cm}^{-2}\text{s}^{-1}$[31].

The data essentially comprises of events with variables that store values of kinematic observables associated with the collision. The aim is to find events with possible sparticles as final decay states. Most of the data collected from p-p collisions will contain events defined by the Standard Model of particle physics. These

events serve as background and sensitivity to signal events is increased by accurately estimating these SM events in a sample of events similar to the data collected. This sample is obtained using Monte Carlo simulation.

3.2 Monte Carlo simulation

Monte Carlo (MC) simulations are computational algorithms that use the repeated random sampling to estimate multi-dimensional integrals which cannot be solved analytically[46]. In principle, the technique can be used to solve any problem that has a probabilistic interpretation and hence is quite suitable for particle physics as many aspects of the subject are connected to probability distributions. MC simulated datasets are used to model the SM backgrounds and evaluate signal selection efficiencies and yields.

All the simulated samples were produced using the ATLAS simulation infrastructure [47] and GEANT4[48]. POWHEG-BOX v2 was used to produce $t\bar{t}$[49][50][51][52] and single top[53][54] samples. Whereas, SHERPA 2.2.1 and 2.2.2[55] were used to produce diboson and multiboson samples used in this analysis. A general-purpose MC generator simulates a *p*-*p* collision at LHC in the following steps[56]:

I Hard scattering

II Parton shower

III Hadronisation

IV Underlying event

V Unstable particle decays

An event corresponds to the instance when two protons moving at ultra-relativistic speeds collide. The *hard scattering* comprises of the simulated processes with the largest momentum transfers in an event. Before collision, the two protons are travelling in opposite directions so the center-of-mass momentum of the system is near to zero. Hence, the large momentum transfers occur in the transverse plane to conserve the center-of-mass momentum after the collision.

The high energy *p*-*p* collision essntially destroys the protons in the process to create *parton showers* that originates from initial and final state radiations(ISR, FSR). Similar to EM radiation in QED, partons carrying color charges radiate gluons, however, unlike photons, gluons themselves carry a color charge, and hence, the emitted gluons can cause new secondary radiation. Every radiation step distributes the energy of the parent gluon among the product particles and hence, the showering process ends with the release of soft gluons[1]. For all the samples showered with PYTHIA[57], the EvtGen v 1.2.0[58] program was used to simulate the properties of the bottom-hadron decays.

[1]Low momentum compared to other particles

Parton shower starts from high momentum or hard gluons and continues till the point where perturbative QCD breaks down, the emission of soft glouns which lead to parton hadronization. The *hadronisation* process form colour neutral hadrons using non-perturbative models such as the cluster model[14] used in the HERWIG generator[59] or the Lund model[14] used in the PYTHIA generator[60].

While observing the collision of interest, it maybe possible that there are additional *p*-*p* collisions that may occur in the same bunch crossing or in the bunch just before and after the collision of interest. To account for such events during the hard scatter, additional soft *underlying events*, also called *in-time and out-of-time pile-up* are generated with PYTHIA 8.186[57] using the NNPDF2.3LO set of PDFs[61] and the A3 tune[62]. The simulated events were reconstructed with the same algorithms as those used for data.

In the final step, unstable hadrons further decay in stable particles which are actually observed by the detectors as the collision signatures.

The SUSY signal samples are generated using the MADGRAPH5_aMC@NLO v2.6.2[63] and PYTHIA 8.230 with the A14[64] set of tuned parameters for the modelling of parton showering (PS), hadronisation and underlying event

The MC simulated events need to be as close to the observed data as possible. To accomplish this, the predicted MC events N_{pred} are normalized to the total cross-section σ_{tot} of the processes of interest. Furthermore, since the MC simulations are obtained during or before the data collection, N_{pred} is scaled using the integrated luminosity and sum of event weights w. The weights are for example used when including next-to-leading order (NLO) corrections in MC generation, or when using event filters during simulation that select only interesting events to increase computer efficiency

$$N_{pred} = \frac{\sigma_{tot} \int L\, dt}{\sum_{i \in generated} w_i} \cdot \sum_{i \in selected} w_i. \tag{3.3}$$

The statistical uncertainty given by the finite size of the simulated sample is given by the square root of sum of squared weights,

$$\Delta N_{pred} = \sqrt{\sum_{i \in selected} w_i^2} \tag{3.4}$$

.

3.3 Event Reconstruction

Hadronic jets, muons and electrons are basic signatures any collision event. These objects are reconstructed for every event in the simulation. There are two identification levels associated to the lepton and jets reconstruction: 'baseline' and 'signal'. Baseline selections are loose identification criteria and hence gives high efficiency

in object identification. Baseline leptons and jets are used in computing missing transverse momentum and in resolving possible reconstruction ambiguities.

Signal leptons and jets are a subset of baseline objects with tighter quality requirements. A list of tracking and calorimeter based variables are used to select signal leptons by discriminating against semi-leptonic heavy flavour decays and jets misidentified as leptons.

Possible electron candidates are reconstructed from energy deposits in electromagnetic calorimeter cross-verified with charged-particle tracks in the Inner-Detector (ID). Baseline electrons need to have transverse momentum $p_T > 7$ GeV and pseudorapidity $|\eta| < 2.47$. Muon candidates are reconstructed from matching the tracks in ID and Muon Spectrometer, and are further refined using a global fit that uses hits from both subdetectors[65]. Baseline muons need to satisfy $p_T > 6$ GeV and $|\eta| < 2.7$. Both signal electrons and muons are defined with tighter pseudo-rapidity and impact parameter requirements[65].

Jets are reconstructed from three-dimensional topological energy clusters in the calorimeters. The baseline jets are selected in the region $|\eta| < 4.5$ and $p_T > 20$ GeV whereas the signal jets have stricter requirements, $|\eta| < 2.8$ and $p_T > 30$ GeV. Jets containing b-hadroms are identified as *b-tagged* jets.

An object may sometimes be categorized as of more than one type. For example, a muon-like electron maybe identified as both a muon and an electron. Overlap procedures are used to treat such reconstruction ambiguities and achieve a one-to-one object categorization. Finally, missing transverse momentum $\mathbf{p}_T^{\text{miss}}$ with magnitude E_T^{miss} is calculated as a negative vectorial sum of transverse momentum of all baseline reconstructed objects and the soft terms. The soft term includes all the tracks observed in the tracker but not matched to any reconstructed physics objects. The tracks that are not associated to the collision of interest are not considered in the E_T^{miss} calculation to suppress the effect of pile-up.

3.4 Test Statistic, p-value and the Likelihood fit

The aim of our analysis is to find SUSY events in the collision data collected from the ATLAS detector. However, we use the MC simulated events in its entirety to develop the best possible methods of finding the signal events. We start by defining a signal region (SR) where we aim to find signal events and then we use the statistical methods to enhance its sensitivity towards the signal events.

The first tool at our disposal is $p-value$, defined as the probability of measuring a certain outcome of the experiment which is at least as rare as the observed outcome. Hence, p-value represents a cumulative probability of all the possible events which have probability of observation less or equal to the possible SUSY events.

Next is the *test statistic* q which quantifies results of data observations by combining them in a single value. A more extreme outcome leads to higher value of test

statistic, hence by definition,

$$p = \int_{q_{obs}}^{\infty} f(q|\boldsymbol{\xi})\, dq \tag{3.5}$$

where, distribution of test statistic function f with model assumptions $\boldsymbol{\xi}$ must be known or determined from pseudo-experiments. p-value is converted in standard deviations from Gaussian distribution, the *significance*,

$$Z = \Phi^{-1}(1-p), \tag{3.6}$$

where Φ^{-1} is inverse of cumulative distribution of a unit Gaussian.

To put the Z value in perspective, a discovery that is based on rejecting a background hypothesis in particle physics is not claimed for $Z < 5$, while significance level of $Z = 3$ counts as evidence. So the challenge now is to calculate and improve the significance with limited statistics. The statistics are employed on MC sample but we cannot simply estimate signal events from the signal region as the SRs are filled with SM events that have similar signatures as the possible SUSY events.

To reduce the effect of the background in the signal region, we first need to estimate it. We define a Control Region (CR) within the sample of simulated events that is rich in SM events. We then estimate the SM background in the control region. Ideally, a control region should be chosen such that it is devoid of any signal events and must comprise entirely of the events governed by the background hypothesis, which in our case, must comprise of events explained by the Standard Model of particle physics.

The SM background estimated in the control region is to be extrapolated to the signal region. But before that, the accuracy of the estimation needs to be assessed by extrapolating it to buffer region, which share kinematics similar to the signal region, called the validation region (VR).

The statistical model employed here uses binned *likelihood* fits where a bin is assigned to each of the kinematic regions of the analysis. Since our analysis is essentially a counting experiment, the events counted must be independent of each other. Hence, every kinematic region gets its own Poisson term. The Poisson probability to observe N events with expected number λ is,

$$P(N/\lambda) = \frac{\lambda^N e^{-\lambda}}{N!}. \tag{3.7}$$

The complete Likelihood plot is the obtained by taking product of the Poisson distributions for every defined kinematic region. This requires the bins to be statistically independent which is ensured while choosing the selection conditions.The expectation λ_i in every bin i is parameterized by introducing the signal strength parameter μ_{sig} as

$$\lambda_i = \mu_{sig} s_i + b_i, \tag{3.8}$$

where, for each bin i, b_i is the number of expected background events and s_i is the expected signal events. μ_{sig} is used as a free parameter and as can be seen

from the equation above, $\mu_{sig} = 0$ and $\mu_{sig} = 1$ corresponds to background-only and signal+background hypothesis respectively. The expected signal and background yields depend on nuisance parameters $\boldsymbol{b}$[2], which are directly constrained by data observations in the CRs and are termed as $\boldsymbol{\mu_{bkg}}$. The other nuisance parameters describe the systematic uncertainties $\boldsymbol{\theta}$.

Schematically, the total likelihood function L is the product of possion distributions and is written as

$$L(\mathbf{n}, \boldsymbol{\theta}^0 | \mu_{sig}, \mathbf{b}, \boldsymbol{\theta}) = P(n_S | \lambda_S(\mu_{sig}, \mathbf{b}, \boldsymbol{\theta})) \times \prod_{i \in bins} P(n_i | \lambda_i(\mu_{sig}, \mathbf{b}, \boldsymbol{\theta})) \times C_{syst}(\boldsymbol{\theta}^0, \boldsymbol{\theta}) \tag{3.9}$$

where the expected value λ_i (number of predicted events) are the functions of expected signal events $\mathbf{s}$, and background events $\mathbf{b}$, nuisance parameters $\boldsymbol{\mu_{sig}}$, $\boldsymbol{\theta}$ and the POI μ_{sig}. The likelihood function is composed of two parts as seen in 3.9,

- $P(n_S | \lambda_S(\mu_{sig}, \mathbf{b}, \boldsymbol{\theta})) \times \prod_{i \in bins} P(n_i | \lambda_i(\mu_{sig}, \mathbf{b}, \boldsymbol{\theta}))$, is the first part of the likelihood fit that represents a combined fit of all the SM backgrounds to the data in the signal and eventually to the control region. The background events estimated individually for MC samples such as top, diboson, etc are simultaneously fitted to observed data using the scale factors. These scale factors are obtained via background extrapolations from control to signal regions in the MC sample.

- The second half of the equation 3.9 is relates to the systematic uncertainties associated with the event estimation. In general there are three sources of errors in our analysis:

  - The first is the statistical error arising from the finite size of the MC sample. Since MC is a numerical analysis, the statistical errors relate to the size of the event sampling. Larger the data, smaller is the associated uncertainty. For performance issues these uncertainties are included only if the standard deviations are above a configurable threshold set to 5% otherwise they are neglected.

  - Errors may arise in the background extrapolation if the region definitions do not share similar kinematics. Among the regions, a variation in kinematics can lead to discrepancies in event distributions in the available decay channels. Such errors can be solved by modifying the region definitions.

  - Theoretical uncertainties are associated to our limited understanding of the underlying dynamics of the processes. Due to a incomplete knowledge of the physics of p-p collisions, the processes may not be perfectly simulated and hence, can lead to uncertainties. As we learn more about the fundamental interactions between subatomic particles, we can better develop the physics and reduce these systematic errors. In the current analysis, the theoretical uncertainties are estimated to show the present-day limitations in the MC sampling.

[2] $\boldsymbol{b} = (b_1, b_2, ..., ba_N)$

3.5 1-Lepton Decay channels

The simplified model used for the analysis is described in fig. 3.1. The model considers electroweak production of lightest chargino $\tilde{\chi}_1^{\pm}$ along with the next-to-lightest neutralino $\tilde{\chi}_2^0$ or another chargino $\tilde{\chi}_1^{\pm}$. The chargino further decays into an LSP and a W boson that in turn decays into a lepton (electron or a muon) and a neutrino thereby producing a final state with a single lepton. Whereas, the neutralino $\tilde{\chi}_2^0$ decays into LSP and a Z boson that can further decay hadronically or produce a pair of leptons. However, for our target model, we need Z to decay hadronically to produce an isolated lepton in the final state. This model is called the WZ model, with signatures similar to the WW generic model shown in fig.3.1(b). There is another possibility for the $\tilde{\chi}_2^0$ to decay into the Higgs boson which further decays in a pair of b-quarks as $h \longrightarrow b\bar{b}$. Since the presence of two b-tagged jets can be a rather specific criteria, the dedicated analysis needs to be done separately and is carried out by other ATLAS groups.

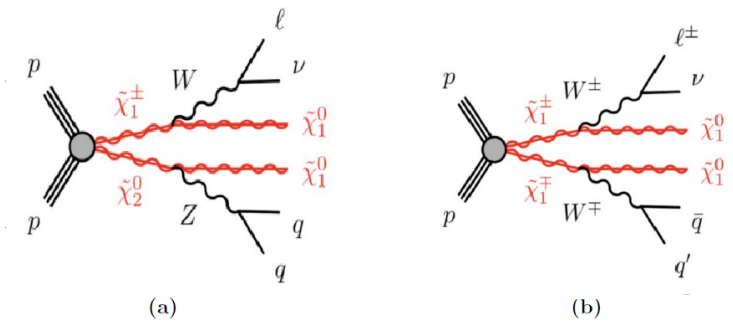


Figure 3.1: Figure (a) shows production of next-to-lightest neutralino $\tilde{\chi}_2^0$ along with chargino $\tilde{\chi}_1^{\pm}$ decaying into jets and single lepton. Figure (b) shows the generic WW model for electroweak SUSY decay previously shown in fig1.6(a).

3.6 Standard Model Backgrounds

Focusing on the 1L decay channel reduces the majority of the QCD multijet backgrounds however, numerous SM processes can yield signatures consisting of exactly 1-lepton, hadronic jets, and missing transverse energy (E_T^{miss}), similar to possible SUSY final states. An exhaustive list of all the such processes is shown in fig. 3.2.

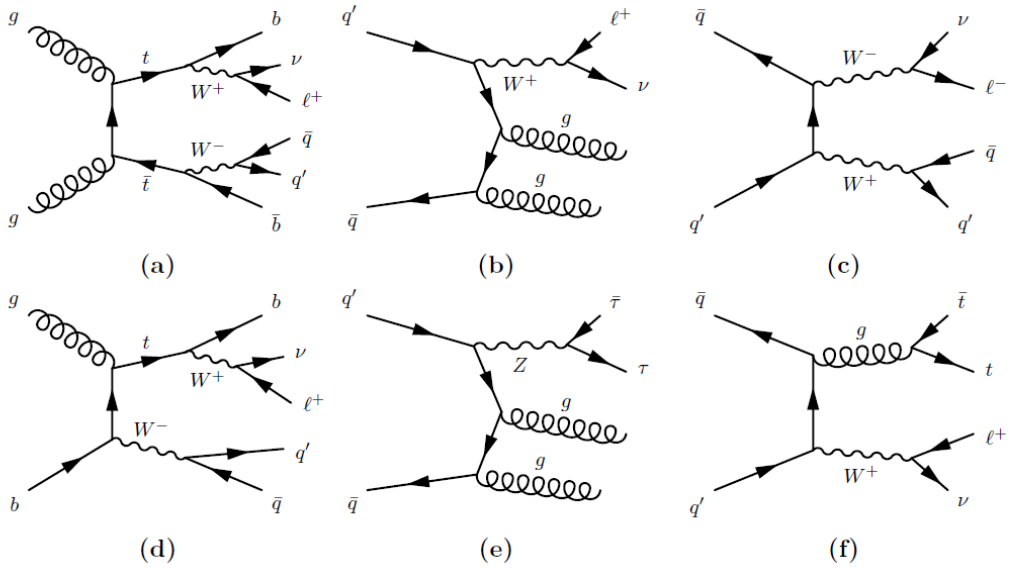


Figure 3.2: Feynman diagrams of SM processes sharing the same final states as the 1L electroweak SUSY channel with (a) $t\bar{t}$, (b) W + jets, (c) diboson, (d) singletop, (e) Z + jets, (f) $t\bar{t} + V$ intermediate products.

- **Singletop and $t\bar{t}$**

One of the largest contributors of the SM background for 1L SUSY events, t quark production from proton collision is also one of the most common processes at LHC. The top quark is the only quark that is massive enough to be kinematically allowed to decay in a real W boson. As shown in fig. 3.2(a) and 3.2(d), the t-quark decays in W bosons which further decay hadronically or leptonically. For our simplified model, we want one of the W bosons to decay leptonically and the other to produce jets.

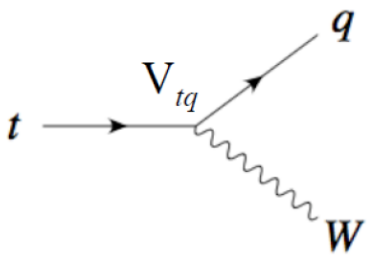


Figure 3.3: Top quark decaying in a real W boson and a quark.

Another reason for interest in electroweak decay of top quark comes from the CKM value of the tbW vertex. As shown in fig. 3.3, normally for the product quark q can be any quark but the CKM value of the q to be a bottom quark ≈ 1. Hence, top quark must almost always produce a b quark in electroweak decays. The b-tagging algorithms can pick up signatures of EW top decay and hence, 2 bjets for a $t\bar{t}$ and one for singletop can act as strong selection conditions for estimating the SM background associated to t-quark.

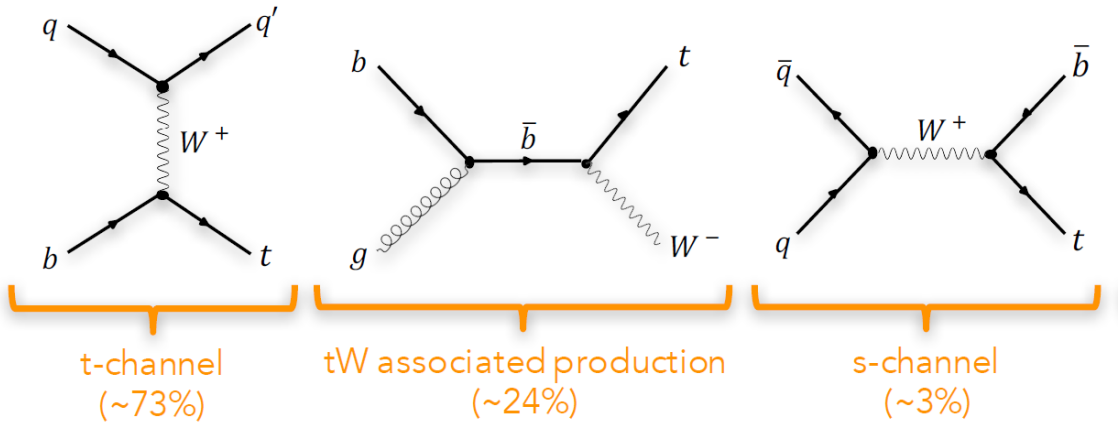


Figure 3.4: s, t and Wt channels of singletop production in electroweak SUSY model[66].

Furthermore, the p-p collisions producing top quarks, can decay via three possible channels as shown in fig. 3.4 with Wt-channel the only one producing a real W boson. Event distributions among the single top decay channels will be crucial for our analysis as discussed in detail in chapter 4.

• Multiboson and Diboson

The W and Z bosons produced are not directly measurable in the detector due to their short lifetimes and hence, are reconstructed from their decay products that can be leptons + missing energy, or hadronic jets. Due to enormous hadronic activity at the hadron collider, most of the analyses for multiboson production only consider leptonic decay modes of the vector bosons since the final state leptons can be detected and measured quite well.

The branching ratios of W and Z decaying leptonically is less compared to a hadronic decay,

$$BR(W \longrightarrow \sum q\bar{q}') \approx BR(Z \longrightarrow \sum q\bar{q}) = \mathcal{O}(70\%) \tag{3.10}$$

$$BR(W^+ \longrightarrow \sum \ell^+\nu) = \mathcal{O}(30\%) \quad \text{and} \quad BR(Z \longrightarrow \sum \ell^+\ell^-) = \mathcal{O}(10\%) \tag{3.11}$$

However, for one-lepton analysis, only one of the vector bosons need to decay leptonically, hence the cross-sections of semileptonic decay modes are large enough to be significant. Also, the jets from the other on-shell vector boson can be identified by summing up their invariant mass since the mass of vector bosons is known with high accuracy.

The figure 3.2(c) only shows one of the possible decay modes however, diboson can undergo,

- Fully leptonic decay, where both the vector bosons decay leptonically. The branching fractions for the same can be seen in equation 3.11.
- Semi-leptonic decay, where one of the boson decays hadronically, while the other decays leptonic. The branching fraction for semi-leptonic decays are larger compared to full leptonic decays.
- W+jets or loop-induced ggVV, which are next to leading order processes and involve loops.

For our analysis, we categorize these decays according to the number of charged leptons in final states and hence, the diboson sample is divided in 3 catagories,

- Diboson with ≥ 2 charged leptons, include all the processes that generate two or more charged leptons in the final state.
- Diboson with 1 charged lepton, is associated to exactly one charged lepton in the final state.
- Other Diboson processes with no charged leptons, can be processes with jets or neutrinos in the final state.

Since ours is a 1L analysis, ideally we shouldn't care about channels with diboson $\geq$ 2L or diboson with no charged leptons. But detectors are not perfect and sometimes, jets can be misconstructed as leptons or soft leptons that may not be reconstructed at all. To cover all the possibilities, we estimate SM background from all three of the above mentioned channels.

- **W/Z + Jets**

Backgrounds associated to $W/Z+$ jets are shown in figure 3.2(b) and 3.2(e). The vector bosons in these channels decay leptonically alongside ultra-relativistic gluons that may produce parton showers. The Z boson decays into a pair of charged leptons whereas, W decays in a charged lepton and a neutrino in the form of missing transverse energy. These are prominent background events for one-lepton and two-lepton SUSY events.

- **$t\bar{t} + V$**

As shown in figure 3.2(f), $t\bar{t}$ can be produced with an electroweak boson V, where V can be either W or Z boson. For the case of Beyond Standard Model (BSM) Physics, $t\bar{t}W$ is of interest as final states give an isolated lepton if the $t\bar{t}$ decay hadronically.

3.7 Observables for electroweak SUSY channel

Let us take a closer look at the Electroweak SUSY channel. First step is to formulate a signal region using the kinematic variables associated to the MC events. Signal region is where we wish to see the SUSY signals, however, it is filled with the SM background with signatures similar to SUSY events. In order to estimate the background, we define a control region which is kinematically similar to the signal region but is selected to be far from the SR to have a region which only filled with SM events. The region definitions for control region are chosen so that they favor SM events. To validate the estimated background from the CR, validation regions (VRs) are defined to be similar to signal regions but without any signal contamination. The background is extrapolated to the validation regions and the scaling factors are evaluated that help reduce the background and increase the sensitivity of the signal region.

To formulate the signal region definitions,we start with constraints applied by our target model that are common for the entire sample such as variables that simulate lepton triggers used in the detectors that sense the presence of lepton in the final state, and constraints applied for obtaining single-lepton final states. These constraints are called *preselection conditions*. The event distribution overlaying each of the kinematic variables are obtained with preselection conditions to better understand the impact of these variables in our counting experiment.

Boosted Objects

At center-of-mass energies as high as $\sqrt{s} = 13$ TeV, heavy particles such as W, Z, and t quark are often produced with large transverse momentum and hence their decay products are Lorentz boosted. The boosted object decay products are collimated to the momentum direction in the detector rest frame. At high p_T, decay products of a hadronically decaying object merge in a single, energetic large radius jet (*large-R jet*), called boosted jets. Boosted jets are reconstructed using the anti-k_t algorithm. Since boosted jets have substructures that are different from the ones initiated by a single parton, we need to have two sets of selection conditions on the kinematic variables to distinguish between the events that contain Lorentz boosted jets and the ones that do not. These region definitions are identified as boosted and resolved[3] jet conditions.

[3] relatively low p_T parton showers

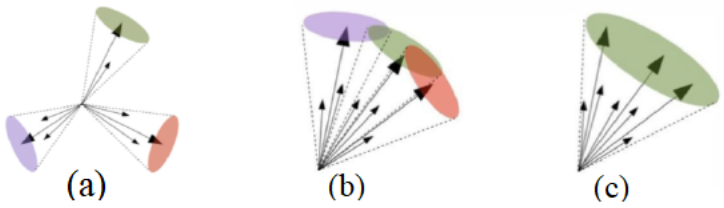


Figure 3.5: (a) Low-p_T and (b) High-p_T jets associated to hadronic decays. (c) shows the same high-p_T configuration as before, using a large-R jet reconstruction[67].

Selection conditions for the case of boosted kinematic regions used in the estimation of SM background include at least one large-R jets and their momenta. Additionally, the fatjets can lead to either Z or W boson via electroweak decay. Tagging algorithms are used to identify whether the jets are $W-$ or $Z-$ tagged. The lower bound for jet momentum is kept at 250 GeV as shown in figure 3.6. For the resolved jets, there must be no presence of large-R jets and hence, there are no corresponding constraints on the resolved jets kinematic conditions. These events can be identified in the figure 3.6 as the peak at large-R jet Pt = 0 GeV.

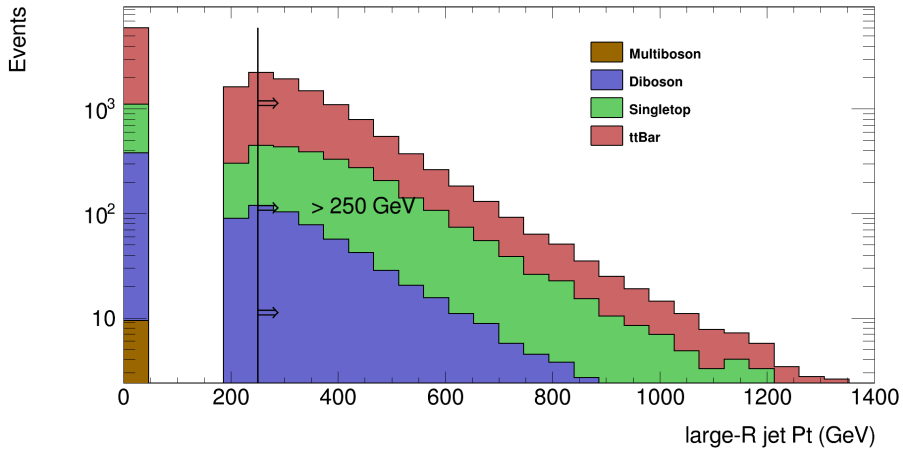


Figure 3.6: MC events overlaying reconstructed large radius jets, with a cut at Pt = 250 GeV.

Lepton and Jet Momenta

The SUSY particles considered in the decay channel usually have higher masses than the SM particles. In the cases where the mass difference between LSP ($\tilde{\chi}_1^0$) and NLSP ($\tilde{\chi}_1^\pm$) is large, the decay products will have higher momenta(*hard*) compared to many SM processes. Hence, the transverse momenta of the final state lepton and jets are promising observables for identifying processes different from the SM. The figure 3.7 shows the distribution of events with growing lepton momentum. Majority of the events in all the samples contain final state leptons with $p_T > 25$ GeV. This is chosen because single lepton trigger activates only for leptons with $p_T \approx 25$ GeV or above.

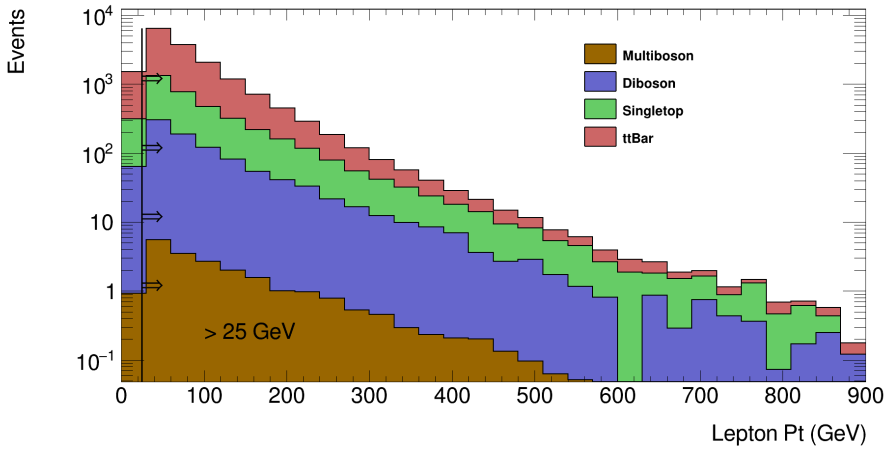


Figure 3.7: MC events overlaying reconstructed lepton momentum, with a cut at $p_T = 25$ GeV.

Number of jets N_{jet}

The final state of the WW model explicitly produces two jets originating from the decay of the W boson. We account for a third jet in the final state due to Initial or Final State Radiation (*ISR and FSR*) and hence, selection condition for events to have final states with two or three jets can prevent unwanted signal losses. However, angular distance of ultra-relativistic jets can sometimes be small enough for them to be indistinguishable, as is the case for Lorentz boosted jets. Hence, < 3 jets is used to select SM backgrounds from the control region of boosted jet events and $2 - 3$ jets are used for events with resolved jets.

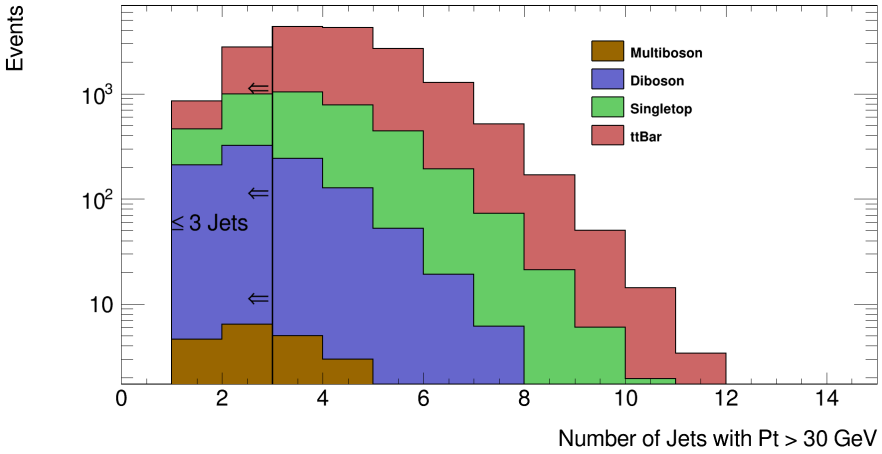


Figure 3.8: MC events overlaying the number of reconstructed hadronic jets with energies of at least 30 GeV, with a cut at 3 Jets.

As shown in figure 3.8, there are SM events with more than 3 jets in our sample but since our model can only produce at most 3 jets, events with large number of jets can be dropped using the above mentioned selection conditions.

Number of b-tagged jets $N_{b-\text{jet}}$

As seen in section 3.6, final states of SM events containing t-quark tend to contain bottom quarks as $t \to Wb$. Since, bottom quarks can be identified using the b-tagging algorithms, for the SM background events associated to t-quark, selection condition of events having at least 1 b-tagged jet is quite important. Since, b-quark is not observed in any other decay channels except for t-decay in our target model, having exactly zero b-tagged jets can be important in identifying the SM background with diboson and multiboson decay channels.

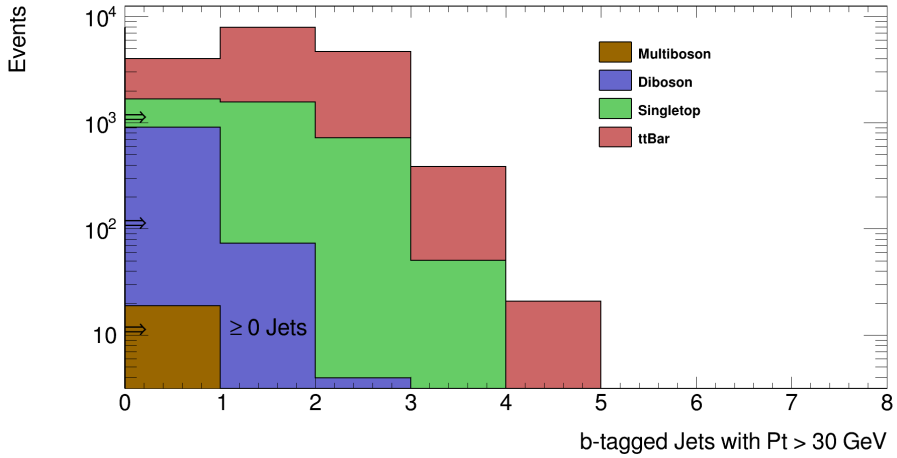


Figure 3.9: MC events overlaying the number of reconstructed b-tagged jets with energies of at least 30 GeV, with a cut of at least 1 b-Jet.

Missing transverse energy $\mathrm{E}_T^{\mathrm{miss}}$

In SM, $\mathrm{E}_T^{\mathrm{miss}}$ mainly originates from neutrinos escaping the detector or mismeasurements in object reconstruction. For the SUSY decay channel, two neutralinos escape the detector in addition to a neutrino leading to much larger $\mathrm{E}_T^{\mathrm{miss}}$ compared to the SM processes. A lower bound on $\mathrm{E}_T^{\mathrm{miss}}$ can be used as a preselection condition to distinguish between SUSY and SM processes. This bound is set at 200 GeV for 1L analysis. As shown in the figure 3.10, there are events where missing transverse energies can go upto 700 or 800 GeV.

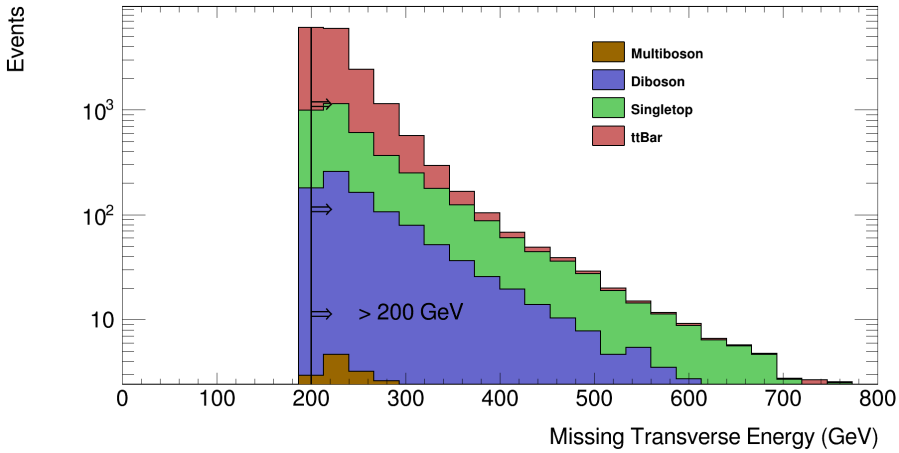


Figure 3.10: MC events overlaying the missing transverse energy, with a cut of at least 200 GeV.

Another reason for accepting only events with $E_T^{miss} > 200$ GeV is to avoid including QCD multijet background during the MC simulation. Pure multijet background can only appear in the one-lepton final states via false reconstruction of a lepton, called fake lepton. Since, a reliable MC simulation of QCD dynamics is inherently difficult, it can be avoided altogether by selecting $E_T^{miss} > 200$ GeV. As can be seen in figure 3.11, multijet events reduce to negligible for $E_T^{miss} > 150$ GeV.

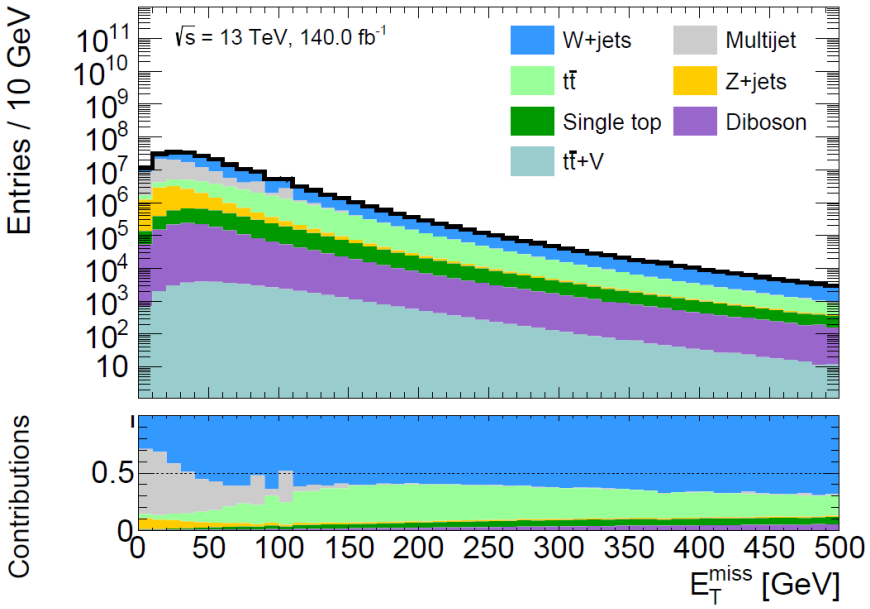


Figure 3.11: In the top part of the plot, the multijet background can be seen in grey. The lower pad shows the relative contribution from different backgrounds. A preselection of exactly one isolated lepton and at least 2 jets is applied.[68]

E_T^{miss} significance

At over 1 billion collisions every second, finite detector resolution can sometimes create mismeasurements in particle reconstruction or detection, this is called *fake*

E_T^{miss}. Detectors also have blind spots where particles can escape detection contributing to fake E_T^{miss}. E_T^{miss} significance gives an estimate of events with fake E_T^{miss} in the form of a ratio between the measured E_T^{miss} with resolution fluctuations and if E_T^{miss} were observed in the ideal case. Both E_T^{miss} significance and E_T^{miss} are used as discriminating observables as it is not clear a priori which one will perform better. For signal, E_T^{miss} is constructed by high p_T objects whereas, for backgrounds, it is constructed using low p_T objects. Hence, signal events tend to show high E_T^{miss} significance, whereas, the background events show a lower value. Hence, for the case of control regions, E_T^{miss} significance is capped at 14, but for signal regions, higher values are used.

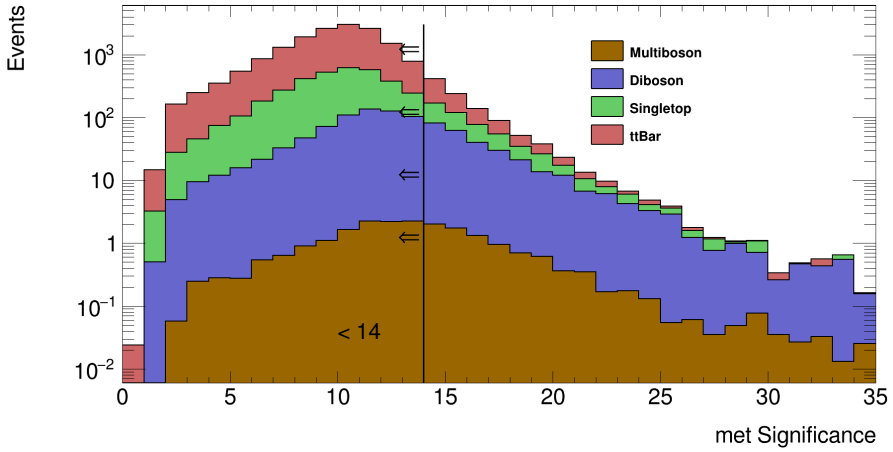


Figure 3.12: MC events overlaying the missing transverse energy Significance, with a cut of at most 14 for top control region.

Transverse mass m_T

Transverse mass is a powerful observable for a 1-lepton channel. Its computation is similar to the invariant mass, however, observables such as missing transverse momentum p_T^{miss} and transverse momentum of lepton p_T^l are used.

$$m_T = \sqrt{2p_T^l E_T^{miss}\Big(1 - cos\Big[\Delta\phi\Big(p_T^l, p_T^{miss}\Big)\Big]\Big)} \tag{3.12}$$

In the SM processes, the source of m_T is only from the E_T^{miss} of neutrinos, but in SUSY processes, m_T values are more spread out. More the mass difference between $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^{\pm}$, a larger fraction of the E_T^{miss} will be carried by the neutralinos in the form of kinetic energy, and m_T will provide better efficiency in rejecting the background. Hence, for background estimation, low constraints of 50 GeV $< m_T <$ 80 GeV are put on transverse mass as SM events must not contain neutralinos. For validation and signal regions, higher transverse masses are used.

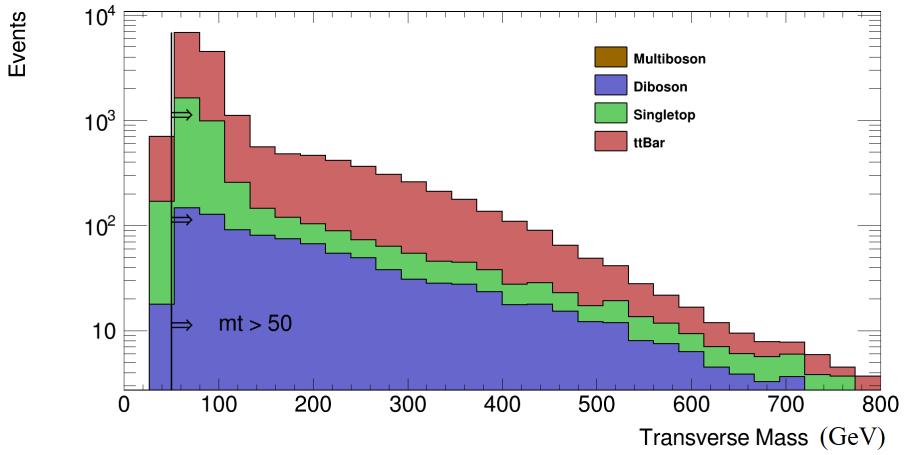


Figure 3.13: MC events overlaying the transverse mass, with a cut of at least 50 GeV.

Reconstructed mass of dijet system m_{jj}

In SUSY events, the primary source of jets is the hadronic decay of the W/Z boson. If the two jets are considered as a single system and the mass of the combined dijet is calculated, it should peak at around the W/Z boson mass. A vector boson mass value cut of $70 - 105$ GeV is applied as shown in figure 3.14 for selecting signal events and eliminating backgrounds where the sources of the jets are not the vector bosons.

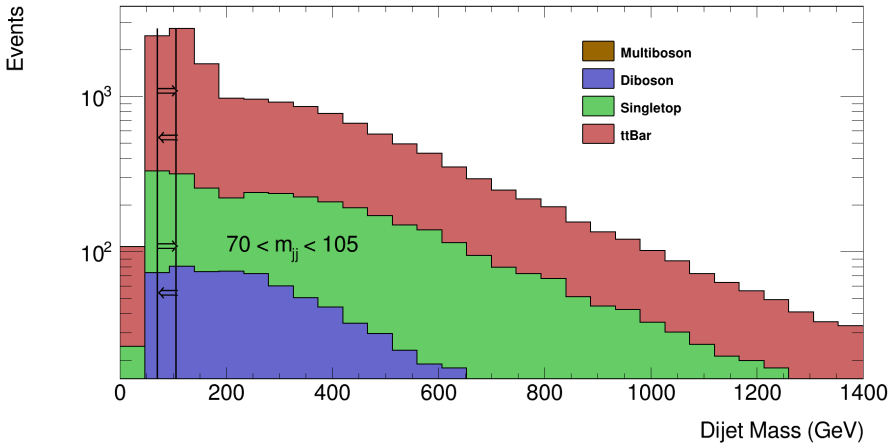


Figure 3.14: MC events overlaying the mass of the dijet system originating from one of the vector bosons, with a cut of 70-105 GeV.

Angular distance between E_T^{miss} and the lepton $\Delta\phi(\ell, E_T^{miss})$

In SM processes E_T^{miss} and lepton originate from a single source, W through $W \to l\nu$ whereas in the SUSY events, the angular distribution $\Delta\phi$ is more evenly spread out as neutralinos are responsible for a larger fraction of E_T^{miss}. Hence, for enhancing SM contributions, an upper limit of $\Delta\phi(\ell, E_T^{miss}) < 2.9$ is applied to the signal region.

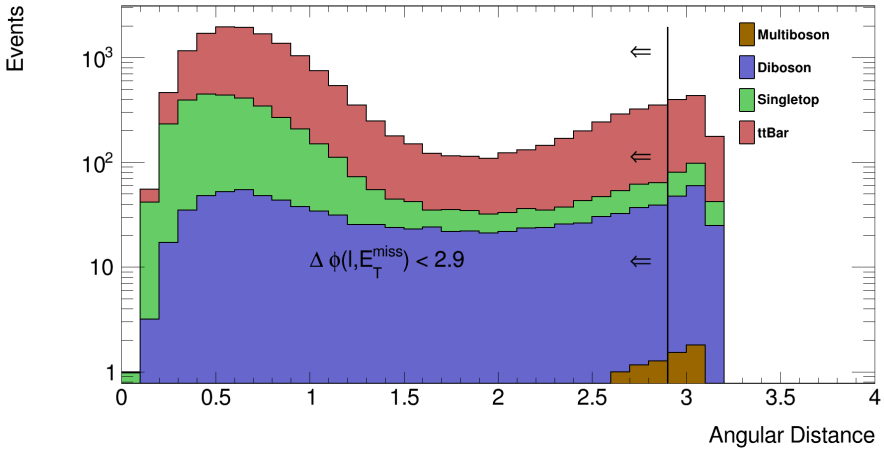


Figure 3.15: MC events overlaying the $\Delta\phi(\ell, E_T^{\text{miss}})$, with a cut-value of at most 2.9.

Signal Region Definitions and Events

Based on section 3.7, a signal region (SR) is formulated which is mentioned in tables 3.1 and 3.2. These kinematic conditions are used to identify the signal events in the MC sample. The region definitions differ for selecting events with boosted and resolved jets, and also according to whether WW or WZ bosons are produced in the decay channels. Large values of m_T are chosen for events with well-resolved jets to favor signal events whereas, for events with boosted jets, large E_T^{miss} significance conditions are chosen along with ranges in transverse mass.

Table 3.1: Signal Region conditions for WW model

Variable	Cuts		Cuts		
	SRLM Resolved	SRHM Resolved	SRLM Boosted	SRMM Boosted	SRHM Boosted
N_{lep}	1			1	
p_T^{ℓ}[GeV]	> 25			> 25	
$N_{\text{jet}}(p_T^{\text{jets}} > 30$ GeV)	2–3			≤ 3	
$N_{\text{b-jet}}(p_T^{\text{jets}} > 30$ GeV)	0			0	
E_T^{miss}[GeV]	> 200			> 200	
$\Delta\phi(\ell, \text{E}_T^{\text{miss}})$	< 2.8			< 2.9	
m_{jj}[GeV]	70–105			–	
$N_{\text{large-Rjets}}$	0			≥ 1	
W-tagged large-R jet	–			yes	
$p_T^{\text{large-Rjet}}$[GeV]	–			> 300	
E_T^{miss} significance	–			> 14	
m_T[GeV]	200–380	> 380	120–240	240–360	> 360

Table 3.2: Signal Region conditions for WZ model

Variable	Cuts		Cuts		
	SRLM Resolved	SRHM Resolved	SRLM Boosted	SRMM Boosted	SRHM Boosted
N_{lep}	1			1	
p_T^{ℓ}[GeV]	> 25			> 25	
$N_{\text{jet}}(p_T^{\text{jets}} > 30$ GeV)	2–3			≤ 3	
$N_{\text{b-jet}}(p_T^{\text{jets}} > 30$ GeV)	0			≤ 2	
E_T^{miss}[GeV]	> 200			> 200	
$\Delta\phi(\ell, \text{E}_T^{\text{miss}})$	< 2.8			< 2.6	
m_{jj}[GeV]	70–105			–	
$N_{\text{large-Rjets}}$	0			≥ 1	
Z-tagged large-R jet	–			yes	
$p_T^{\text{large-Rjet}}$[GeV]	–			> 250	
E_T^{miss} significance	–			> 15	
m_T[GeV]	200–380	> 380	120–240	240–420	> 420

According to the above mentioned SR region definitions, the events are estimated for Multiboson, Diboson, single top and $t\bar{t}$ samples with luminosity of 138.96516 fb^{-1}, summarized in table 3.3. These events are a mixture of signal events along with SM events having similar final states as the target model for SUSY events.

Table 3.3: Events estimated in different samples according to the signal region definitions described in tables 3.1 and 3.2.

Signal Regions	MultiBoson	DiBoson ≥2L	DiBoson 1L	DiBoson 0L	DiBoson Total	Singletop	$t\bar{t}$
SRLMres	0.14 ± 0.03	26.34 ± 1.15	32.68 ± 1.55	0.18 ± 0.11	59.20 ± 2.80	5.78 ± 0.76	50.07 ± 1.94
SRHMres	0.04 ± 0.02	8.82 ± 0.60	7.81 ± 0.58	-0.01 ± 0.01	16.61 ± 1.19	1.47 ± 0.37	8.96 ± 0.73
Z-SRLMboost	0.02 ± 0.02	1.14 ± 0.24	3.32 ± 0.43	0.00 ± 0.00	4.46 ± 0.67	0.54 ± 0.19	1.08 ± 0.12
Z-SRMMboost	0.00 ± 0.00	1.38 ± 0.14	2.08 ± 0.20	0.00 ± 0.00	3.47 ± 0.34	0.77 ± 0.28	0.76 ± 0.18
Z-SRHMboost	0.01 ± 0.01	0.42 ± 0.07	0.70 ± 0.12	0.00 ± 0.00	1.12 ± 0.19	0.40 ± 0.20	0.11 ± 0.04
W-SRLMboost	0.00 ± 0.00	1.26 ± 0.23	3.40 ± 0.58	0.00 ± 0.00	4.65 ± 0.81	0.30 ± 0.17	0.57 ± 0.09
W-SRMMboost	0.00 ± 0.00	1.04 ± 0.15	1.31 ± 0.13	0.00 ± 0.00	2.35 ± 0.28	0.23 ± 0.16	0.20 ± 0.05
W-SRHMboost	0.01 ± 0.01	0.50 ± 0.08	0.65 ± 0.09	0.00 ± 0.00	1.15 ± 0.16	0.03 ± 0.02	0.07 ± 0.02

The diboson sample is divided according to whether events have exactly zero, exactly one and two or more charged leptons in the final states. Ours is a one-lepton analysis and ideally, we shouldn't care about the diboson with zero or multiple leptons in the final state. However, while collecting the collision data, sometimes jets can be misidentified as leptons which can redefine a 1L event to a 2L event. Also, sometimes leptons are not identified all together making it a zero lepton event. To account for events with all such misidentified jets and unidentified leptons, we divide the diboson sample in three. The events estimated for diboson sample in table 3.3 show a negligible contribution from the channel with exactly zero leptons as compared to diboson channels with at least 1 lepton and t-quark channels. Hence, the events with misidentified jets coming from the diboson 0L sample do not affect our analysis.

We also observe a negligible contribution from the Multiboson sample. A joint production of three vector bosons is a rare process in the Standard Model. Hence even a large error in the sampling of the multiboson events will not have any significant impact on our analysis. For these reasons, we can identify the major background contributors as diboson channels with at least one-lepton in final state, and events with t-quarks. Hence, top and diboson channels with charged leptons will play a crucial in our analysis.

Control and Validation Regions

To increase the signal sensitivity of SRs, we need an estimate of the SM background events in the signal regions. We do this by defining a control region (CR) which is kinematically similar to the SRs. We then estimate the SM background in the control regions. Control regions are designed to be far enough from the SRs so that we don't loose any signal events, but need to be kinematically similar to the signal regions. The SM background from the control region is then extrapolated to a buffer region which is similar to the signal region but doesn't contain any signal events. This is the validation region (VR). The extrapolations are verified in the VRs before

they are propagated to the signal regions. Control and Validation region definitions are chosen to be specific to the backgrounds that need to be estimated. Hence, the tables 3.4 and 3.5 are used to estimate SM events with t-quark decay channels in the control and validation regions of the MC samples with an isolated lepton, jets and large E_T^{miss}.

Control region conditions are specific to the backgrounds that need to be estimated. The region definitions are divided in boosted and resolved regions as was the case for signal region definitions. Two validation regions are chosen for events with boosted jets. VR1 is chosen for events with low E_T^{miss} significance and a high transverse mass, whereas, VR2 is chosen for high E_T^{miss} and low m_T.

Table 3.4: Top Boosted Regions

Variable	Cuts		
	CR	**VR1**	**VR2**
N_{lep}	1	1	1
p_T^{ℓ}[GeV]	> 25	> 25	> 25
$N_{\text{jet}}(p_T^{\text{jets}} > 30$ GeV)	≤ 3	≤ 3	≤ 3
$N_{\text{b-jet}}(p_T^{\text{jets}} > 30$ GeV)	> 0	> 0	> 0
E_T^{miss}[GeV]	> 200	> 200	> 200
$\Delta\phi(\ell, E_T^{\text{miss}})$	< 2.9	< 2.9	< 2.9
E_T^{miss} significance	< 14	< 14	> 14
$N_{\text{large-Rjets}}$	≥ 1	≥ 1	≥ 1
W-tagged large-R jet	yes	yes	yes
$p_T^{\text{large-Rjet}}$[GeV]	> 250	> 250	> 250
m_T[GeV]	50−80	> 80	50−120

Table 3.5: Top Resolved Regions

Variable	Cuts	
	CR	**VR**
N_{lep}	1	
p_T^{ℓ}[GeV]	> 25	
$N_{\text{jet}}(p_T^{\text{jets}} > 30$ GeV)	$2 \leq N_{\text{jet}} \leq 3$	
$N_{\text{b-jet}}(p_T^{\text{jets}} > 30$ GeV)	> 0	
m_{jj}[GeV]	70−105	
E_T^{miss}[GeV]	> 200	
$\Delta\phi(\ell, E_T^{\text{miss}})$	< 2.8	
$N_{\text{large-Rjets}}$	0	
m_T[GeV]	50−80	80−120

The event estimates for top control and validation regions which are mainly SM backgrounds events, are shown in table 3.6. As can be seen in the table, the events in singletop and $t\bar{t}$ are significantly more compared to other samples which should be the case since the top region definitions are supposed to favor the events to top quark processes.

Table 3.6: SM background estimates for control and validation regions of t-quark definitions.

top regions	MultiBoson	DiBoson ≥2L	DiBoson 1L	DiBoson 0L	DiBoson Total	Singletop	$t\bar{t}$
TCRres	0.00 ± 0.00	0.44 ± 0.09	16.50 ± 1.75	0.00 ± 0.00	16.94 ± 1.84	84.28 ± 2.88	611.60 ± 6.63
TCRboost	0.00 ± 0.00	0.59 ± 0.11	30.85 ± 2.61	0.00 ± 0.00	31.43 ± 2.71	237.94 ± 4.96	800.32 ± 6.45
TVRres	0.00 ± 0.00	1.45 ± 0.20	9.72 ± 1.13	-0.05 ± 0.05	11.12 ± 1.38	65.79 ± 2.58	532.22 ± 7.01
TVR1boost	0.00 ± 0.00	0.82 ± 0.20	10.43 ± 2.80	0.00 ± 0.00	11.25 ± 3.00	152.17 ± 4.06	430.38 ± 5.22
TVR2boost	0.00 ± 0.00	0.13 ± 0.06	7.81 ± 1.38	0.00 ± 0.00	7.94 ± 1.45	23.95 ± 1.35	45.28 ± 0.80

The background estimations are additionally supported by cutflow tables for all the region definitions. A cutflow table is a tabular progression of kinematic constraints put forth for event selection. Essentially cutflow tables are used to observe the effects on the event selection as the kinematic variables are applied. The tables 3.7 and 3.8 are cutflow tables for boosted top control and resolved top control regions. As seen in the tables, the effect of every kinematic condition can be evaluated

and compared for different samples.

For example, if we observe the constraints for the presence of b-jets (highlighted in blue), both for boosted and resolved regions, a majority of the events are rejected in the case of diboson. Whereas, the effect is not so significant in the case of t backgrounds such as singletop and $t\bar{t}$. This is because the electroweak decay of t quark almost always produces a b-quark, as discussed in the section **??** whereas, the diboson 1 lepton channels do not produce b-quarks, and the only contribution stems from misidentified jets. The SM event estimates (highlighted in yellow) are shown at the bottom of a cutflow table after all the necessary constraints are applied. The cutflow tables are also an important tool for debugging purposes and are used as a reference for the final background estimates.

Table 3.7: Cutflow table for control region of boosted jets showing the effect of kinematic conditions on event selection, with the b-jet condition highlighted in blue and the final yields highlighted in yellow.

TCRboost	Singletop	$t\bar{t}$	Multiboson	Diboson
Preselection	31674.08 ± 53.62	168205.71 ± 93.21	25.63 ± 0.50	21371.42 ± 112.83
<= 3 jets	15815.74 ± 37.74	64542.85 ± 61.89	19.01 ± 0.42	15518.78 ± 109.72
at least one b-jet	10844.67 ± 31.25	48453.73 ± 53.44	1.84 ± 0.15	1178.70 ± 12.96
$\Delta\phi(l, E_T^{miss}) < 2.9$	10225.09 ± 30.23	42602.63 ± 50.08	1.45 ± 0.13	1114.79 ± 12.74
E_T^{miss} significance < 14	9119.50 ± 28.79	38510.47 ± 48.64	0.76 ± 0.09	892.59 ± 11.92
at least one large-R jet	6879.94 ± 25.21	24564.93 ± 35.61	0.53 ± 0.07	691.37 ± 10.88
W-tagged large-R jet	608.53 ± 7.99	2351.38 ± 12.09	0.00 ± 0.00	72.48 ± 4.56
large-R jet $p_T > 250$ GeV	504.31 ± 7.29	1585.26 ± 9.35	0.00 ± 0.00	52.20 ± 4.15
$50 < m_T < 80$	302.51 ± 5.58	1026.50 ± 7.25	0.00 ± 0.00	37.65 ± 2.76

Table 3.8: Cutflow table for control region of resolved jets showing the effect of kinematic conditions on event selection, with the b-jet condition highlighted in blue and the final yields highlighted in yellow.

TCRres	Singletop	$t\bar{t}$	Multiboson	Diboson
Preselection	31674.08 ± 53.62	168205.71 ± 93.21	25.63 ± 0.50	21371.42 ± 112.83
$2-3$ jets	14102.57 ± 35.86	60276.48 ± 59.54	13.68 ± 0.37	11948.83 ± 101.10
at least one b-jet	10240.38 ± 30.36	45944.49 ± 51.81	1.62 ± 0.14	1057.18 ± 11.58
$\Delta\phi(l, E_T^{miss}) < 2.8$	9566.98 ± 29.24	38985.34 ± 47.67	1.28 ± 0.13	992.12 ± 11.42
no large-R jet	2187.27 ± 13.70	13375.30 ± 31.85	0.35 ± 0.07	215.65 ± 4.81
$70 <= mjj <= 105$	232.28 ± 4.79	1968.40 ± 12.54	0.06 ± 0.04	37.23 ± 2.27
$50 < m_T < 80$	113.40 ± 3.32	854.22 ± 7.75	0.00 ± 0.00	18.84 ± 1.86

To represent a region definition in fewer letters, the region names are abbreviated. For example, the region definitions belonging to control region of top boosted/resolved sectors are written as *TCRboost/TCRres*. Similarly for signal regions, we have *SRLMres/SRHMres* for resolved regions and for boosted regions, depending on whether the region is W/Z-tagged, we have *W-SRLMboost* or *Z-SRLMboost* and so on. The cutflow tables pertaining to the kinematic region definitions for t-quark backgrounds are detailed in the section A.1.

A similar approach is developed for estimating SM events with diboson production channels. The control and validation region definitions are shown in tables 3.9 and 3.10 for boosted and resolved hadronic jets respectively. The only difference between selection conditions for top and diboson background is that there are no b-tagged jets in the diboson definitions. As bottom quarks are exclusively produced from top quarks via electroweak decay, there should be no traces of b-quarks in the diboson channels which is signified by taking the $N_{\text{b-jet}} = 0$.

Table 3.9: Diboson Boosted Regions

Variable	Cuts		
	CR	**VR1**	**VR2**
N_{lep}	1	1	1
p_T^{ℓ}[GeV]	> 25	> 25	> 25
$N_{\text{jet}}(p_T^{\text{jets}} > 30$ GeV)	≤ 3	≤ 3	≤ 3
$N_{\text{b-jet}}(p_T^{\text{jets}} > 30$ GeV)	0	0	0
E_T^{miss}[GeV]	> 200	> 200	> 200
$\Delta\phi(\ell, \text{E}_T^{\text{miss}})$	< 2.9	< 2.9	< 2.9
E_T^{miss} significance	< 14	< 14	> 14
$N_{\text{large-Rjets}}$	≥ 1	≥ 1	≥ 1
W-tagged large-R jet	yes	yes	yes
$p_T^{\text{large-Rjet}}$[GeV]	> 250	> 250	> 250
m_T[GeV]	50−80	> 80	50−120

Table 3.10: Diboson Resolved Regions

Variable	Cuts	
	CR	**VR**
N_{lep}	1	
p_T^{ℓ}[GeV]	> 25	
$N_{\text{jet}}(p_T^{\text{jets}} > 30$ GeV)	$2 \leq N_{\text{jet}} \leq 3$	
$N_{\text{b-jet}}(p_T^{\text{jets}} > 30$ GeV)	0	
m_{jj}[GeV]	70−105	
E_T^{miss}[GeV]	> 200	
$\Delta\phi(\ell, \text{E}_T^{\text{miss}})$	< 2.8	
E_T^{miss} significance	> 12	
$N_{\text{large-Rjets}}$	0	
m_T[GeV]	50−80	80−120

In the data taken by the ATLAS detector, sometimes diboson channels with two lepton final states maybe misindentified to be a 1L state. To account such misidentifications in the MC samples, additional selection conditions mentioned in table 3.11 are used with exactly two real leptons. Together with two leptons, the following constraints are added:

- m_{jj} veto, m_{jj} represents the invariant mass of the dijet system. As mentioned in section 3.7, m_{jj} conditions are primarily used to select jets originating from vector bosons, as opposed to other sources. An m_{jj} veto of 75-95 GeV is exactly the opposite, as it will select events with dijet masses that are not in the range of 75-95 GeV.

- $m_{\ell,\ell}$, is the invariant mass of the two leptons identified in the final states. This is used to select the events where the two leptons originated from a Z boson.

Table 3.11: Diboson 2L Conditions

Variable	Cuts	
	CR	**VR**
N_{lep}	2	
p_T^{ℓ}[GeV]	> 25	
$N_{\text{jet}}(p_T^{\text{jets}} > 30$ GeV)	≤ 3	
$N_{\text{b-jet}}(p_T^{\text{jets}} > 30$ GeV)	0	
E_T^{miss}[GeV]	> 200	
$\Delta\phi(\ell, \text{E}_T^{\text{miss}})$	< 2.9	
$m_{\ell\ell}$[GeV]	70−100	
m_{jj}veto [GeV]	75−95	
E_T^{miss} significance	> 12	> 10
m_T[GeV]	50−200	200−350

The SM background estimations for 1L and 2L diboson regions are summarised in table 3.12. Although the SM backgrounds are estimated here for the control and validation regions of table 3.11, they are not used for the 1L likelihood fit to the data. The fit is carried out in the 2L analysis but the scaling factor obtained from the fit is propagated to our analysis and hence is important to estimate the associated background.

Table 3.12: SM background estimates for control and validation region of diboson nature.

Diboson Regions	MultiBoson	DiBoson ≥2L	DiBoson 1L	DiBoson 0L	DiBoson Total	Singletop	$t\bar{t}$
WDB1LCRres	0.00 ± 0.00	2.00 ± 0.26	46.09 ± 3.58	0.00 ± 0.00	48.09 ± 3.84	15.10 ± 1.16	71.14 ± 1.70
WDB1LCRboost	0.00 ± 0.00	7.65 ± 0.56	379.26 ± 21.00	0.00 ± 0.00	386.91 ± 21.56	188.66 ± 4.38	508.29 ± 5.23
DB2LCR	0.48 ± 0.07	125.21 ± 1.89	0.07 ± 0.07	0.00 ± 0.00	125.28 ± 1.96	16.04 ± 1.14	79.45 ± 1.74
WDB1LVRres	0.06 ± 0.02	11.18 ± 0.58	40.10 ± 2.81	0.00 ± 0.00	51.28 ± 3.39	10.16 ± 1.00	59.06 ± 1.87
WDB1LVR1boost	0.03 ± 0.01	11.43 ± 0.49	140.36 ± 6.27	0.00 ± 0.00	151.79 ± 6.77	97.31 ± 3.24	277.24 ± 4.32
WD1LVR2boost	0.00 ± 0.00	2.46 ± 0.31	78.29 ± 5.06	0.00 ± 0.00	80.74 ± 5.37	19.70 ± 1.24	39.76 ± 0.76
DB2LVR	0.90 ± 0.10	207.38 ± 2.73	0.01 ± 0.01	0.00 ± 0.00	207.38 ± 2.74	14.82 ± 1.16	88.80 ± 2.20

Formulating the signal region and estimating the corresponding events have helped us identify the focal point for our analysis. Most of the SM event contributions to the production of electroweakinos come from the diboson and the top electroweak decay channels. Our signal region is well-defined using the kinematic distributions from section 3.7, however, the control and validation region definitions need to be verified. From section 3.6, we can observe that t-quark has three possible decay channels. SM event yields from t-quark samples can give insights about the extent of kinematic similarity between the well-defined signal regions and, the possibly inconclusive definitions of control and validation regions.

The diboson background events provide us with an opportunity to estimate the theoretical uncertainties arising due to the limited knowledge of the physics governing the diboson production. The systematic uncertainties corresponding to the diboson samples are evaluated and summarised as part of the analysis for this thesis.

Chapter 4

Background Uncertainty Evaluation for 1L Analysis

The previous chapters support our search for electroweakinos. We started by giving a strong theoretical foundation that the open questions about the Standard Model of Particle Physics and predicts supersymmetric particles not too far from the electroweak scale. We then looked at the present-day collider and ATLAS detector capabilities for observing the lightest stable supersymmetric particles (LSPs) in the form of large E_T^{miss}. Later we defined the statistical methods that act as basis of our analyses, and the importance of achieving a one-one agreement between the data collected and the samples generated from the Monte Carlo simulations. The particle productions via strong channels are possible however strong exclusion limits are placed on the lower bound of their masses, hence we focused our search on electroweakinos using the targeted WW/WZ simplified model.

4.1 Overview

The statistical analysis of the SM background estimation strongly relies on the MC generated events. Once the SR, CR, and VR are defined a described in section 3.7, the most prominent SM backgrounds, in our case top and diboson production, are estimated with the help of the information on the event yields in the CR. The MC is fitted simultaneously in the dedicated CR in order to ensure a perfect match between data and MC in the CR. In this way a scaling factor for each of the background of interest is set, and the same scaling factor is then used to scale the MC in VR and SR. The VR is then checked to evaluate the agreement of data and MC: since data in VR is expected to show the same behaviour as data in the SR, it provides a good test for the solidity of our procedure. If the matching between data and MC in the VR is considered acceptable, only at that point we look for data in the signal region in order to perform our test for SUSY detection.

As stated in the previous chapter the two important sources of background for possible SUSY events characterized by a final state with an isolated lepton, hadronic jets and E_T^{miss} come from the t-quark and WW WZ, and ZZ production channels. The scaling factors used to evaluate the background in the signal regions is prone to

a number of errors ranging from the poorly defined control regions, to errors arising in the MC simulation from our incomplete knowledge of the physics of the process. The associated inaccuracies are interpreted as systematic errors and are the main focus of this thesis.

The data in the signal regions is not considered before the validation of the background estimation to avoid observer bias. Hence, the analysis done up to this point is a blind analysis. For this reason the understanding of the errors and uncertainties associated to the background modelling before the unblinding is crucial for this analysis. After unblinding, the scaled background signal region estimates are compared with the data present in the signal regions. If the data and the scaled signal region estimates differ by a Z value of more than 3, the analysis is considered as evidence for SUSY observation.

As described in section 3.7, events containing at least one real t-quarks can be produced through $s-$, $t-$ and Wt$-$ channels and inclusively evaluated as mentioned in section 3.6. However we assume that all the production modes share the same scaling factor despite the fact that due to the mixture of production modes, there is a potential discrepancy in the behaviour of CR and SR. Therefore, to assess the stability of the scaling factor used for the t-quark background we need to verify how kinematically similar the control regions are as compared to the signal and validation regions. To assess the uncertainty introduced by the difference in the kinematic characteristics of the final states, the yield variations in each of the t-quark production channels are evaluated for each of the SR/VR/CR. These variations are a direct consequence of using a single scale factor for three different physical processes and after being quantitatively assessed we will evaluate the possible systematic error associated to the choice of using a single scale factor.

Another prominent background stems from diboson production, as concluded in 3.7. The diboson events arising from p-p collisions have unique characteristics that are utilized to calculated variations arising from our limited understanding about the physics of the process. The underlying dynamics leading to diboson production are governed by the strong interaction and hence, the systematic uncertainties arise from non-perturbative effects. These uncertainties directly point towards how accurately the diboson samples are generated. This lack of knowledge about the real physics process introduces a possible error in the scale factor evaluation. In the following we will assess the impact of our poor understanding of the diboson production dynamics in the scaling factor evaluation, and we will evaluate the systematic error associated to the model used to simulate our samples.

4.2 Yield Variations in t-quark Backgrounds

In this section we will present the procedure to evaluate the systematic uncertainty arising from considering all the events with t-quarks to have the same scale factor. The first step is to produce the SM background associated with t-quark MC samples using the region definitions described in section 3.7. The background events corresponding to all the region definitions mentioned in the section 3.7 are shown

the table 4.1. The luminosity used for this analysis is 139 fb^{-1}. The yields in $s-$, $t-$ and Wt$-$ channels are shown for the singletop sample.

Table 4.1: Singletop and $t\bar{t}$ background along with the event yields in the t-quark decay channels for all the region definitions from section 3.7.

Region	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
SRLMres	0.00 ± 0.00	0.00 ± 0.00	7.90 ± 0.89	7.90 ± 0.89	69.40 ± 2.24
SRHMres	0.00 ± 0.00	0.00 ± 0.00	2.32 ± 0.46	2.32 ± 0.46	11.42 ± 0.82
Z-SRLMboost	0.00 ± 0.00	0.00 ± 0.00	0.68 ± 0.19	0.68 ± 0.19	1.72 ± 0.20
Z-SRMMboost	0.00 ± 0.00	0.00 ± 0.00	0.97 ± 0.30	0.97 ± 0.30	0.98 ± 0.19
Z-SRHMboost	0.00 ± 0.00	0.00 ± 0.00	0.43 ± 0.20	0.43 ± 0.20	0.16 ± 0.04
W-SRLMboost	0.00 ± 0.00	0.00 ± 0.00	0.37 ± 0.17	0.37 ± 0.17	0.76 ± 0.10
W-SRMMboost	0.00 ± 0.00	0.00 ± 0.00	0.33 ± 0.18	0.33 ± 0.18	0.24 ± 0.06
W-SRHMboost	0.00 ± 0.00	0.00 ± 0.00	0.05 ± 0.02	0.05 ± 0.02	0.10 ± 0.03
TCRboost	1.62 ± 0.24	6.59 ± 0.70	294.30 ± 5.53	302.51 ± 6.47	1026.50 ± 7.25
TVR1boost	1.01 ± 0.18	3.88 ± 0.46	196.91 ± 4.66	201.80 ± 5.30	558.76 ± 5.91
TVR2boost	0.04 ± 0.04	0.89 ± 0.24	30.85 ± 1.50	31.78 ± 1.78	61.17 ± 0.93
TCRres	1.11 ± 0.19	20.34 ± 1.09	91.95 ± 3.13	113.40 ± 4.41	854.22 ± 7.75
TVRres	1.10 ± 0.19	16.17 ± 0.97	68.97 ± 2.78	86.25 ± 3.94	759.64 ± 8.37

The table 4.1 clearly suggests that Wt-channel is the dominant channel for production of top events with the desired final signatures for all the regions. The signal regions are shown in the first eight rows of the table, whereas the bottom part shows the control and the validation regions. From the table 4.1, we observe that the signal regions have no events in $s-$ and $t-$ channel top quark yields whereas, that is clearly not the case for control and validation regions. This difference in composition of the singletop yields between signal and control/validation regions can be source of observable bias. To have a qualitative assessment event distribution in different production modes of top for CR, VR and SR, we introduce the N-1 plots. The N-1 plots are one-dimensional plots event distribution plots. If there are N kinematic variables in a region definition, the events are constrained by N-1 of those variables and the distribution is obtained overlaying the complete range of the remaining variable. Hence, for each of the N variables, there is a corresponding N-1 plot. The difference in the shape can lead to systematic effects which can affect the scale factor for SM yields arising from the singletop sample.

Normalised N-1 Plots

To evaluate systematic error arising due to observed yield variations in t-quark production channels for CR and VR, the first step is to generate *N-1 plots*. For our analysis, we require normalised N-1 plots. As the dominant channel is the Wt-channel, we consider the yields from Wt-channel as y_w and the rest as y_o which comprise of $s-$ and $t-$ channel yields. Our purpose is to check the difference in the shape. Therefore N-1 plot for y_w and y_o are normalised w.r.t. the SM background yields in respective channels.

The figure 4.1 shows normalised N-1 plots for the resolved top control region of table 3.5. The green and blue dotted line represent the normalised N-1 plots for y_o and y_w yields respectively. If the normalised plots would coincide, the effect of yield variations would have been exactly zero however, there are discrepancies observed such as, in the top right plot in the figure 4.1. This could lead to high systematic effects which can affect the scale factor. To understand if the singletop sample suffers with observable bias, we evaluate the systematic effects in the following section.

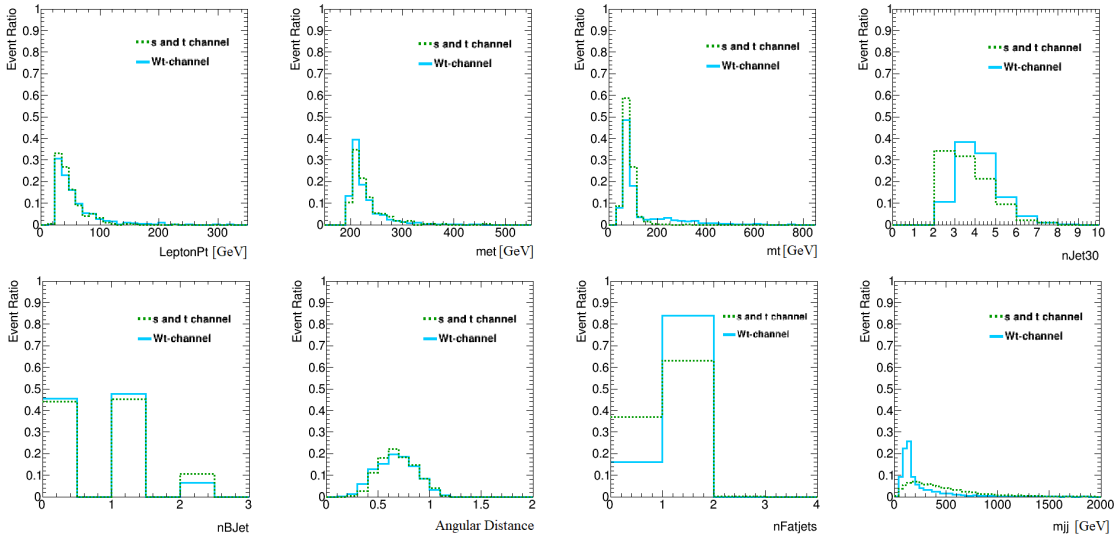


Figure 4.1: Normalised plots overlaying all the kinematic variables involved in defining resolved top control region, with green representing y_o events and blue representing y_w events.

Calculation and Results

Two important parameters involved in estimating systematic effects are;

- Difference in the normalised yield for y_w and y_o at each bin of the histograms.

The more similar the plots are, the less would be the effect of y_o on y_w. Hence the difference in the shape is directly proportional to the yield variations.

- Total $y_o + y_w$ events.

From this analysis, we aim to understand the effect of y_o events on the scale factor which is associated to the total SM background, hence the $y_o + y_w$ events will be the reference for calculating the percentage contamination. Note here that the term *percentage contamination* is used quite loosely and it refers to the systematic effect in percentage.

With that said, the equation to calculate the effect of y_o-channels on total yields in percentage is given as,

$$\text{Systematic Error(x)\%} = \frac{(\text{shape\_difference(x)} \times \text{total\_background\_events}) \times 100}{(\text{total\_signal\_events} + \text{total\_background\_events})}. \tag{4.1}$$

Where,

$$\text{shape\_difference(x)} = \left\| \frac{y_w(x)}{\text{total\_}y_w\text{\_events}} - \frac{y_o(x)}{\text{total\_}y_o\text{\_events}} \right\| \tag{4.2}$$

and $y_w(x)$ and $y_o(x)$ refers to values of y_w and y_o evaluated at each bin for fig. 4.1. Hence the shape_difference(x) is the difference in the normalised plots for y_w and y_o, evaluated at each bin x.

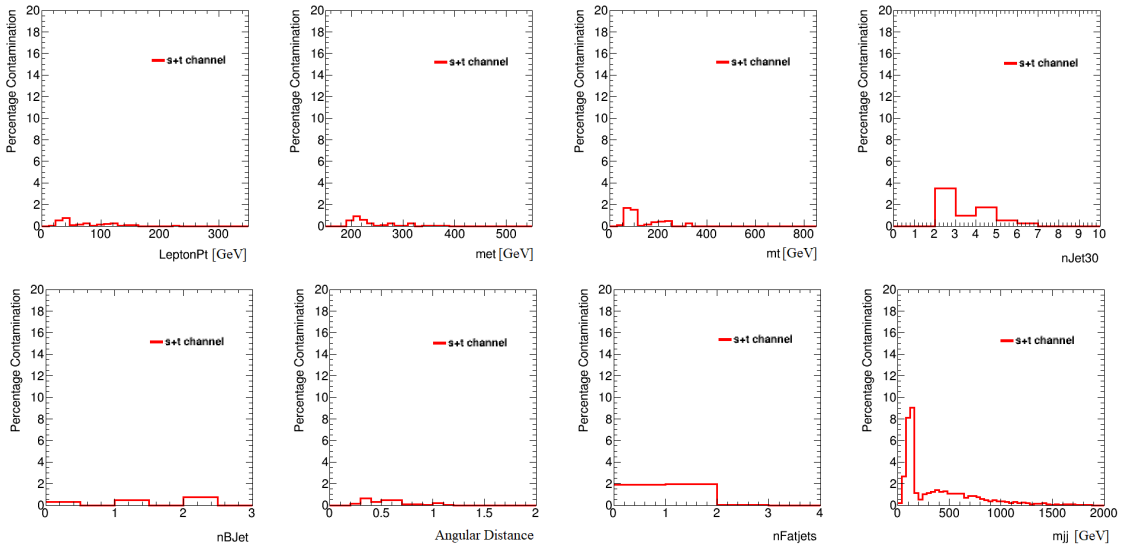


Figure 4.2: Systematic error in percentage corresponding to the N-1 plots of figure 4.1 for the resolved top control region.

Using the equations 4.1 and 4.2, the systematic error overlaying the kinematic variables used for resolved top control region definitions are evaluated and shown in figure 4.2. In the figure, we can observe the effect of the the difference in the shape of the different production modes of the top quark. In bottom-right plot overlaying the dijet mass, there is a relatively large systematic effect compared to the rest of the plots in the figure. To understand further the case of the plot overlaying the mass of the dijet system shown in figure 4.2, we convert the plot in a tabular form 4.2. The highlighted dijet ranges show a maximum systematic uncertainties of 8.08% and 9.07% for the dijet mass range of 120-160 GeV and 160-200 GeV respectively. However, the dijet mass selection condition in table 3.5 suggests that only the events with $m_{jj} \in (70, 105)$ GeV are considered for estimating the SM background which would correspond to the bins 40-80 GeV and 80-120 GeV in table 4.2 with the systematic error of $< 3\%$.

Table 4.2: Bin-wise systematic error for N-1 plot overlaying m_{jj} for the resolved top region along with % of total B and S event contribution for the respective dijet mass ranges.

Bin	Systematic Error	B yields	B yields (%)	S yields	S yields (%)
0 - 40	0.00%	0.00	0.00%	0.00	0.00%
40 - 80	0.20%	2.76	0.57%	0.73	0.14%
80 - 120	2.66%	18.27	3.76%	49.91	9.34%
120-160	8.08%	26.48	5.45%	119.67	22.40%
160-200	9.07%	32.14	6.61%	136.95	25.64%
200 - 240	1.13%	33.82	6.96%	49.83	9.33%
240 - 280	0.54%	33.58	6.91%	30.82	5.77%
280 - 320	1.01%	30.33	6.24%	21.97	4.11%
320 - 360	1.14%	27.77	5.71%	17.69	3.31%
360 - 400	1.24%	28.22	5.81%	17.08	3.20%
400 - 440	1.43%	26.07	5.36%	12.63	2.36%
440 - 480	1.22%	23.78	4.89%	12.50	2.34%
480 - 520	1.29%	21.10	4.34%	8.75	1.64%
520 - 560	1.10%	18.83	3.87%	8.33	1.56%
560 - 600	1.09%	18.47	3.80%	8.10	1.52%
600 - 640	1.08%	15.25	3.14%	4.64	0.87%
640 - 680	1.10%	15.31	3.15%	4.54	0.85%
680 - 720	0.71%	10.94	2.25%	4.02	0.75%
720 - 760	0.86%	11.54	2.37%	3.01	0.56%
760 - 800	0.88%	10.74	2.21%	1.90	0.36%
800 - 840	0.75%	10.64	2.19%	3.24	0.61%
840 - 880	0.56%	7.71	1.59%	2.15	0.40%
880 - 920	0.45%	6.61	1.36%	2.21	0.41%
920 - 960	0.34%	5.57	1.15%	2.35	0.44%
960 - 1000	0.47%	6.43	1.32%	1.83	0.34%
1000 - 1040	0.37%	4.75	0.98%	1.08	0.20%
1040 - 1080	0.36%	4.01	0.83%	0.35	0.07%
1080 - 1120	0.29%	3.84	0.79%	0.92	0.17%
1120 - 1160	0.19%	3.03	0.62%	1.20	0.23%
1160 - 1200	0.29%	3.51	0.72%	0.62	0.12%
1200 - 1240	0.19%	2.15	0.44%	0.21	0.04%
1240 - 1280	0.24%	2.91	0.60%	0.55	0.10%
1280 - 1320	0.20%	2.17	0.45%	0.15	0.03%
1320 - 1360	0.11%	1.77	0.36%	0.69	0.13%
1360 - 1400	0.14%	1.97	0.41%	0.61	0.11%
1400 - 1440	0.14%	1.59	0.33%	0.18	0.03%
1440 - 1480	0.18%	2.08	0.43%	0.22	0.04%
1480 - 1520	0.10%	1.32	0.27%	0.29	0.05%
1520 - 1560	0.05%	1.05	0.22%	0.63	0.12%
1560 - 1600	0.08%	1.12	0.23%	0.30	0.06%
1600 - 1640	0.06%	0.77	0.16%	0.20	0.04%
1640 - 1680	0.08%	0.94	0.19%	0.10	0.02%
1680 - 1720	0.07%	0.94	0.19%	0.29	0.05%
1720 - 1760	0.08%	1.01	0.21%	0.23	0.04%
1760 - 1800	0.04%	0.54	0.11%	0.09	0.02%
1800 - 1840	0.05%	0.63	0.13%	0.10	0.02%
1840 - 1880	0.03%	0.44	0.09%	0.13	0.02%
1880 - 1920	0.02%	0.29	0.06%	0.13	0.02%
1920 - 1960	0.00%	0.43	0.09%	0.00	0.00%
1960 - 2000	0.00%	0.46	0.10%	0.00	0.00%

For other plots in the figure 4.2, the systematic error evaluated is well under 5%. Since all the systematic uncertainties are $< 5\%$, the corresponding yield variations are not significant enough and hence, the resolved top control region defined in table 3.5 show kinematic agreement of upwards of 95% with the signal region definitions.

A similar analysis was performed for the remaining control/validation regions and the results are summarized in the section A.2. The systematic effects obtained in all the regions are less than 5% which is smaller than the statistical errors of MC simulation. Therefore, we can safely conclude that they are negligible.

4.3 Theoretical Uncertainties in diboson Backgrounds

In this section, the SM background associated to the different event samples are estimated. The focus is then shifted to the diboson channel as the physics governing the diboson production is not entirely known, hence the MC simulations of diboson sample have inherent uncertainties. These uncertainties are evaluated for the diboson sample with one or more leptons.

diboson and t-channel yields

The first part of the analysis focuses on the MC simulated samples for the multiboson, diboson, singletop and $t\bar{t}$ production. The SM backgrounds associated to these samples are estimated using the region definitions described in the section 3.7 with the luminosity 138.96516 fb^{-1}. One important point to mention here is that the background measurements shown in the table 4.3 were carried out with two additional conditions;

- nLep_combiBase, is a combination condition for the number of baseline leptons. The number of combination baseline leptons are chosen to be < 3.
- nLep_combiBaseHighPt, is a combination condition for the number of baseline leptons with large momenta, chosen to be < 2.

Table 4.3: SM background estimated for all different samples using the selection conditions mentioned in 3.7

Region	MultiBoson	DiBoson ≥2L	DiBoson 1L	DiBoson 0L	DiBoson Total	Singletop	$t\bar{t}$
SRLMres	0.14 ± 0.03	26.34 ± 1.15	32.68 ± 1.55	0.18 ± 0.11	59.20 ± 2.80	5.78 ± 0.76	50.07 ± 1.94
SRHMres	0.04 ± 0.02	8.82 ± 0.60	7.81 ± 0.58	-0.01 ± 0.01	16.61 ± 1.19	1.47 ± 0.37	8.96 ± 0.73
Z-SRLMboost	0.02 ± 0.02	1.14 ± 0.24	3.32 ± 0.43	0.00 ± 0.00	4.46 ± 0.67	0.54 ± 0.19	1.08 ± 0.12
Z-SRMMboost	0.00 ± 0.00	1.38 ± 0.14	2.08 ± 0.20	0.00 ± 0.00	3.47 ± 0.34	0.77 ± 0.28	0.76 ± 0.18
Z-SRHMboost	0.01 ± 0.01	0.42 ± 0.07	0.70 ± 0.12	0.00 ± 0.00	1.12 ± 0.19	0.40 ± 0.20	0.11 ± 0.04
W-SRLMboost	0.00 ± 0.00	1.26 ± 0.23	3.40 ± 0.58	0.00 ± 0.00	4.65 ± 0.81	0.30 ± 0.17	0.57 ± 0.09
W-SRMMboost	0.00 ± 0.00	1.04 ± 0.15	1.31 ± 0.13	0.00 ± 0.00	2.35 ± 0.28	0.23 ± 0.16	0.20 ± 0.05
W-SRHMboost	0.01 ± 0.01	0.50 ± 0.08	0.65 ± 0.09	0.00 ± 0.00	1.15 ± 0.16	0.03 ± 0.02	0.07 ± 0.02
DB2LCR	0.48 ± 0.07	125.21 ± 1.89	0.07 ± 0.07	0.00 ± 0.00	125.28 ± 1.96	16.04 ± 1.14	79.45 ± 1.74
DB2LVR	0.90 ± 0.10	207.38 ± 2.73	0.01 ± 0.01	0.00 ± 0.00	207.38 ± 2.74	14.82 ± 1.16	88.80 ± 2.20
WDB1LCRboost	0.00 ± 0.00	7.65 ± 0.56	379.26 ± 21.00	0.00 ± 0.00	386.91 ± 21.56	188.66 ± 4.38	508.29 ± 5.23
WDB1LVR1boost	0.03 ± 0.01	11.43 ± 0.49	140.36 ± 6.27	0.00 ± 0.00	151.79 ± 6.77	97.31 ± 3.24	277.24 ± 4.32
WD1LVR2boost	0.00 ± 0.00	2.46 ± 0.31	78.29 ± 5.06	0.00 ± 0.00	80.74 ± 5.37	19.70 ± 1.24	39.76 ± 0.76
WDB1LCRres	0.00 ± 0.00	2.00 ± 0.26	46.09 ± 3.58	0.00 ± 0.00	48.09 ± 3.84	15.10 ± 1.16	71.14 ± 1.70
WDB1LVRres	0.06 ± 0.02	11.18 ± 0.58	40.10 ± 2.81	0.00 ± 0.00	51.28 ± 3.39	10.16 ± 1.00	59.06 ± 1.87
TCRboost	0.00 ± 0.00	0.59 ± 0.11	30.85 ± 2.61	0.00 ± 0.00	31.43 ± 2.71	237.94 ± 4.96	800.32 ± 6.45
TVR1boost	0.00 ± 0.00	0.82 ± 0.20	10.43 ± 2.80	0.00 ± 0.00	11.25 ± 3.00	152.17 ± 4.06	430.38 ± 5.22
TVR2boost	0.00 ± 0.00	0.13 ± 0.06	7.81 ± 1.38	0.00 ± 0.00	7.94 ± 1.45	23.95 ± 1.35	45.28 ± 0.80
TCRres	0.00 ± 0.00	0.44 ± 0.09	16.50 ± 1.75	0.00 ± 0.00	16.94 ± 1.84	84.28 ± 2.88	611.60 ± 6.63
TVRres	0.00 ± 0.00	1.45 ± 0.20	9.72 ± 1.13	-0.05 ± 0.05	11.12 ± 1.38	65.79 ± 2.58	532.22 ± 7.01

These conditions are applied to make all the regions mutually orthogonal. Orthogonal regions are important at subsequent stages of the analysis when different channels are combined to get a likelihood fit. Since the fit is performed with

backgrounds associated to all the possible samples simultaneously, the condition of orthogonality ensures that the scaling factor obtained for different backgrounds are independent of each other. Hence, to make sure that the errors from one region don't propagate to other regions, the region definitions are bound by the combination requirements. Also, the combination conditions are not applied to the diboson region definitions with two leptons mentioned in table 3.11, rather, the preselection conditions for choosing exactly two leptons are applied, i.e. two base leptons and one signal leptons are chosen.

For the reasons discussed in 3.7, we can neglect the diboson sample with exactly zero charged lepton, and the multiboson sample. Out of the remaining samples in the table 4.3, t-quark samples have already been analysed in the section 4.2, and as discussed in section 4.1, the diboson decay channels have interesting characteristics that allow us to calculate the theoretical uncertainties associated to the production process. Hence, we will focus on the diboson samples with at least one charged lepton in the final state.

Theoretical Uncertainties

Alongside the standard deviations involved in the MC simulation of the SM background, shown in table 4.3, the diboson events are prone to theoretical uncertainties which are;

- Variations in Parton Distribution Function (PDF).
- Scale Variations

PDF Variations

Parton Distribution Functions (PDFs) give the probability to find partons (quarks and gluons) in a hadron as a function of the fraction x of the proton's momentum carried by the parton. They are experimentally determined from a short-distance scattering of the partons. In the past, experimental uncertainties in PDF fits were largely due to a lack of high-precision experimental measurements at LHC. This is no longer the case and PDFs in kinematic regions can be measured with high precision. These variations arise due to non-perturbative effects at high energy ($\sqrt{s}$ = 13 TeV) p-p collisions.

Scale Variations

Cross-section calculations for strong processes lead to two types of divergences;

- Ultraviolet (UV) divergences arise due to large loop momentum in the Feynman diagrams representing the amplitude.
- Infrared (IR) divergences which appear because (i) either virtual or real particle can reach zero momentum, (ii) a massless particle radiates another massless particle.

The UV divergences are cured by introducing a renormalization scale μ_R dependency of the strong coupling α_S. The IR divergences in case (i) cancels out due to the Kinoshita–Lee–Nauenberg theorem or KLN theorem[69]. IR divergences in case (ii) are cured by introducing factorization scale μ_F dependency PDF and fragmentation functions. μ_R and μ_F are spurious parameters and physical observables should ideally not depend on them. Since cross-section measurements are perturbative, in principle one could try and sum up the entire perturbation series but is practically impossible. Nevertheless, including more terms would decrease the observable dependence on these scales.

The worst case would be to consider only the leading-order contributions since then, the observables will be monotonous functions of each of these scales. It turns out that next to leading-order corrections give a leftover dependence and not even next-next-leading order entirely cures the problem. There is no ideal solution and hence uncertainties arise due to μ_R and μ_F scale factors. Given that scale factors are a function of the center-of-mass energy of the process (in our case, $\sqrt{s} = 13$ TeV), the best procedure would be to study how the theoretical predictions vary with variations in the scale factors. The general notion is to compute for μ, 2μ and $\mu/2$, where μ represents μ_R and μ_F.

PDF and scale variations vary for each event in the MC simulation. The scale variations are of 6 types;

- μ_R0.5-μ_F0.5
- μ_R0.5-μ_F1
- μ_R1-μ_F0.5
- μ_R1-μ_F2
- μ_R2-μ_F1
- μ_R2-μ_F2

Calculation and Results

Scale and PDF variations are calculated via one of the two equations. For Diboson with one of more charged leptons, uncertainty is estimated as follows:

$$\text{uncertainty} = \frac{\text{tf\_syst}}{\text{tf\_nom}} - 1 \tag{4.3}$$

Where,

$$\text{tf\_syst (Region)} = \frac{\text{syst\_yield (Region)}}{\text{syst\_yield (Control)}} - 1 \tag{4.4}$$

$$\text{tf\_nom (Region)} = \frac{\text{nom\_yield (Region)}}{\text{nom\_yield (Control)}} - 1 \tag{4.5}$$

For the single lepton region definitions in 3.7, boosted and resolved diboson control region conditions are taken as control (reference) for the boosted and resolved jet region definitions respectively. Hence, while calculating tf_syst and tf_nom values for regions containing 1L conditions, the systematic yields corresponding to the control will be WDB1LCRboost for boosted jets and WDB1LCRres for the resolved jets. For additional conditions, i.e. DB2LCR and DB2LVR, the DB2LCR is taken

as the reference control region. Eq. 4.3, along with eq. 4.4 and 4.5 are used for the Diboson samples with exactly one charged lepton or > 2 leptons. For the rest of the samples, a more simple definition is used to estimate the uncertainty,

$$\text{uncertainty} = \frac{\text{syst\_yield}}{\text{nom\_yield}} - 1. \tag{4.6}$$

nom_yield corresponds to the SM background estimated in table 4.3, whereas, syst_yield are calculated by factoring in an additional weight. Since our main focus is on the diboson channel with at least one lepton, we will not be needing the Eq. 4.6. However, during the analysis, systematic yields and uncertainties were evaluated also for the backgrounds such as W + jets and ttV using Eq. 4.6 for cross-checks and debugging purposes. All the results associated with the estimation of systematic yields and uncertainties for different MC samples can be found in section 4.3 of appendix **??**.

Estimation of Scale Variations

The scale uncertainties associated with Diboson with one or more charged leptons are shown in tables 4.4 and 4.5. Almost all of the uncertainties are $< 15\%$ with a few exceptions above 20%. The maximum associated scale uncertainty was found to be $\sim 30\%$. This is larger than the MC statistical error associated with the simulation, hence scale uncertainties play an important role in the likelihood fit, Eq. 3.9.

Table 4.4: Scale Uncertainties corresponding to Diboson sample with exaclty one charged lepton

Diboson 1L	$\mu_R 0.5 - \mu_F 0.5$	$\mu_R 0.5 - \mu_F 1$	$\mu_R 1 - \mu_F 0.5$	$\mu_R 1 - \mu_F 2$	$\mu_R 2 \mu_F 1$	$\mu_R 2 \mu_F 2$
SRLMres	0.1503	0.1207	-0.0697	-0.0169	-0.1699	-0.1747
SRHMres	0.2375	0.2341	-0.0060	-0.0073	-0.1603	-0.1596
Z-SRLMboost	0.0691	0.0602	0.0122	0.0814	-0.0552	0.0084
Z-SRMMboost	0.0686	0.0493	0.0244	0.0728	-0.0591	-0.0031
Z-SRHMboost	0.0499	0.0497	0.0209	0.0734	-0.0390	0.0204
W-SRLMboost	0.0152	0.0232	0.0053	0.0880	-0.0257	0.0440
W-SRMMboost	0.1103	0.0873	0.0268	0.0687	-0.0753	-0.0258
W-SRHMboost	0.0556	0.0604	0.0270	0.0677	-0.0524	-0.0018
DB2LCR	0.3306	0.2019	0.0986	-0.0096	-0.1842	-0.1959
DB2LVR	0.2462	-0.0166	0.2208	-0.1137	0.0001	-0.0931
WDB1LCRboost	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
WDB1LVR1boost	-0.1123	-0.0993	-0.0035	0.0903	0.0846	0.1660
WD1LVR2boost	-0.1741	-0.1428	-0.0096	0.1056	0.0994	0.1920
WDB1LCRres	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
WDB1LVRres	0.1372	0.1181	0.0070	-0.0095	-0.0796	-0.0828
TCRboost	-0.1574	-0.1540	-0.0143	0.0959	0.1145	0.2026
TVR1boost	-0.2671	-0.2189	-0.0325	0.1324	0.1704	0.2925
TVR2boost	-0.1475	-0.1018	-0.0271	0.1063	0.0770	0.1675
TCRres	0.0600	0.0624	-0.0059	0.0042	-0.0396	-0.0343
TVRres	0.1700	0.1397	0.0154	-0.0146	-0.0849	-0.0919

Table 4.5: Scale Uncertainties corresponding to Diboson sample with two or more charged lepton

Diboson ≥2	$\mu_R 0.5 - \mu_F 0.5$	$\mu_R 0.5 - \mu_F 1$	$\mu_R 1 - \mu_F 0.5$	$\mu_R 1 - \mu_F 2$	$\mu_R 2 \mu_F 1$	$\mu_R 2 \mu_F 2$
SRLMres	0.0036	0.0116	-0.0147	0.0086	-0.0181	-0.0024
SRHMres	0.0005	0.0326	-0.0291	0.0339	-0.0271	0.0040
Z-SRLMboost	-0.0063	-0.0344	0.0235	-0.0204	0.0226	0.0027
Z-SRMMboost	-0.0142	-0.0136	0.0036	0.0009	0.0217	0.0197
Z-SRHMboost	-0.0132	-0.0137	0.0025	-0.0004	0.0117	0.0104
W-SRLMboost	0.0146	-0.0114	0.0249	-0.0194	0.0151	-0.0035
W-SRMMboost	-0.0015	-0.0082	0.0099	-0.0053	0.0160	0.0086
W-SRHMboost	0.0090	-0.0094	0.0222	-0.0163	0.0177	-0.0013
DB2LCR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DB2LVR	-0.0181	-0.0101	-0.0103	0.0091	0.0065	0.0185
WDB1LCRboost	-0.0396	-0.0287	-0.0106	0.0142	0.0280	0.0412
WDB1LVR1boost	-0.0036	-0.0044	0.0011	0.0004	0.0036	0.0044
WD1LVR2boost	-0.0226	-0.0249	-0.0001	0.0011	0.0224	0.0215
WDB1LCRres	-0.0093	-0.0238	0.0148	-0.0083	0.0290	0.0197
WDB1LVRres	0.0110	0.0013	-0.0128	-0.0050	-0.0161	-0.0161
TCRboost	-0.0655	-0.0634	-0.0024	0.0063	0.0381	0.0452
TVR1boost	-0.0220	-0.0006	-0.0199	0.0063	0.0097	0.0151
TVR2boost	-0.0172	-0.0279	0.0099	0.0022	0.0281	0.0311
TCRres	-0.0683	-0.0315	-0.0320	0.0317	0.0315	0.0601
TVRres	0.0416	0.0516	-0.0108	0.0161	-0.0402	-0.0246

Estimation of PDF Variations

There are at least 100 different PDF variations in the diboson samples but only the one that has the largest absolute value is taken to account for maximum possible uncertainty. The 3 largest absolute values for PDF variations for Diboson 1L and Diboson > 2L are shown in tables 4.6 and 4.7 respectively.

Table 4.6: Three largest PDF variations of Diboson 1L sample. Uncertainties > 20% are highlighted in gray and uncertainties > 100% in blue.

Diboson 1L	PDFs	Uncertainties	Diboson 1L	PDFs	Uncertainties
SRLMres	PDF261077	0.1993	WDB1LCRboost	MEWeight	0.0000
SRLMres	PDF261020	0.1873	WDB1LCRboost	PDF261000	0.0000
SRLMres	PDF261062	0.0722	WDB1LCRboost	PDF261000	0.0000
SRHMres	PDF261077	0.1648	WDB1LVR1boost	PDF261042	-5551.2637
SRHMres	PDF261020	0.1491	WDB1LVR1boost	PDF261079	0.2455
SRHMres	PDF261099	0.1364	WDB1LVR1boost	PDF261073	0.0863
Z-SRLMboost	PDF261058	0.0834	WD1LVR2boost	PDF261049	0.0830
Z-SRLMboost	PDF261053	0.0754	WD1LVR2boost	PDF261055	0.0799
Z-SRLMboost	PDF261016	0.0735	WD1LVR2boost	PDF261052	0.0788
Z-SRMMboost	PDF261058	0.0785	WDB1LCRres	MEWeight	0.0000
Z-SRMMboost	PDF261034	0.0752	WDB1LCRres	PDF261000	0.0000
Z-SRMMboost	PDF261090	0.0736	WDB1LCRres	PDF261000	0.0000
Z-SRHMboost	PDF261060	0.0684	WDB1LVRres	PDF261077	0.2001
Z-SRHMboost	PDF261003	0.0610	WDB1LVRres	PDF261020	0.1905
Z-SRHMboost	PDF261055	0.0592	WDB1LVRres	PDF261073	-0.0799
W-SRLMboost	PDF261090	0.0722	TCRboost	PDF261079	0.0718
W-SRLMboost	PDF261030	0.0698	TCRboost	PDF261017	0.0695
W-SRLMboost	PDF261009	0.0685	TCRboost	PDF261050	0.0683
W-SRMMboost	PDF261090	0.0959	TVR1boost	PDF261075	0.0903
W-SRMMboost	PDF261016	0.0750	TVR1boost	PDF261043	0.0868
W-SRMMboost	PDF261058	0.0743	TVR1boost	PDF261034	0.0775
W-SRHMboost	PDF261058	0.1183	TVR2boost	PDF261090	0.0733
W-SRHMboost	PDF261009	0.0963	TVR2boost	PDF261055	0.0656
W-SRHMboost	PDF261034	0.0836	TVR2boost	PDF261091	0.0636
DB2LCR	PDF261016	0.1016	TCRres	PDF261020	0.2224
DB2LCR	PDF261030	0.1014	TCRres	PDF261077	0.1911
DB2LCR	PDF261070	0.0990	TCRres	PDF261090	0.1033
DB2LVR	PDF261016	0.1952	TVRres	PDF261058	0.6117
DB2LVR	PDF261053	0.1815	TVRres	PDF261054	0.3719
DB2LVR	PDF261009	0.1696	TVRres	PDF261024	0.2753

Table 4.7: Three largest PDF variations of Diboson ≥ 2 sample. Uncertainties $> 20\%$ are highlighted in gray and uncertainties $> 100\%$ in blue.

Diboson ≥2L	PDFs	Uncertainties	Diboson ≥2L	PDFs	Uncertainties
SRLMres	PDF261012	-0.0678	WDB1LCRboost	PDF261012	-0.1417
SRLMres	PDF261090	-0.0274	WDB1LCRboost	PDF261077	-0.1037
SRLMres	PDF261016	-0.0269	WDB1LCRboost	PDF261096	-0.0713
SRHMres	PDF261062	-60.6839	WDB1LVR1boost	PDF261012	-0.0915
SRHMres	PDF261056	-0.3181	WDB1LVR1boost	PDF261073	0.0428
SRHMres	PDF261036	-0.1783	WDB1LVR1boost	PDF261082	0.0323
Z-SRLMboost	PDF261090	-0.0824	WD1LVR2boost	PDF261062	-0.0843
Z-SRLMboost	PDF261062	-0.0792	WD1LVR2boost	PDF261019	-0.0694
Z-SRLMboost	PDF261056	-0.0657	WD1LVR2boost	PDF261049	0.0591
Z-SRMMboost	PDF261099	0.0817	WDB1LCRres	PDF261056	-0.1598
Z-SRMMboost	PDF261077	0.0693	WDB1LCRres	PDF261068	-0.1421
Z-SRMMboost	PDF261001	0.0607	WDB1LCRres	PDF261072	0.1312
Z-SRHMboost	PDF261056	-0.3715	WDB1LVRres	PDF261012	-0.0908
Z-SRHMboost	PDF261030	0.2238	WDB1LVRres	PDF261016	-0.0330
Z-SRHMboost	PDF261062	-0.2130	WDB1LVRres	PDF261049	-0.0296
W-SRLMboost	PDF261012	-0.0812	TCRboost	PDF261012	-0.0662
W-SRLMboost	PDF261003	0.0346	TCRboost	PDF261072	0.0390
W-SRLMboost	PDF261049	0.0343	TCRboost	PDF261020	0.0335
W-SRMMboost	PDF261012	-0.0867	TVR1boost	PDF261012	-0.0653
W-SRMMboost	PDF261090	-0.0254	TVR1boost	PDF261033	0.0347
W-SRMMboost	PDF261037	-0.0240	TVR1boost	PDF261095	-0.0216
W-SRHMboost	PDF261012	-0.0984	TVR2boost	PDF261012	-0.0693
W-SRHMboost	PDF261061	-0.0444	TVR2boost	PDF261016	-0.0299
W-SRHMboost	PDF261072	0.0340	TVR2boost	PDF261045	-0.0266
DB2LCR	MEWeight	0.0000	TCRres	PDF261062	-0.1981
DB2LCR	PDF261000	0.0000	TCRres	PDF261019	-0.1833
DB2LCR	PDF261000	0.0000	TCRres	PDF261053	0.1778
DB2LVR	PDF261072	-0.2319	TVRres	PDF261012	-0.0774
DB2LVR	PDF261012	-0.0797	TVRres	PDF261056	0.0493
DB2LVR	PDF261016	-0.0145	TVRres	PDF261061	-0.0476

The rows highlighted in blue correspond to uncertainties $\gg 100\%$ which are unexpectedly large compared to other regions, and hence a further analysis is required to find the cause of the anomaly. A thorough inspection of the weights associated with the events contributing to the highlighted uncertainties revealed that there were a few events with anomalously large weight values. The effect of these large weights on the total weight of the event is shown in table 4.8 where the mean of total weights associated with an event with and without the systematic weights are taken for reference.

Table 4.8: Unexpectedly large systematic weights associated to 1L and 2L Diboson sample, compared with mean weights in the sample.

Diboson2L(SRHMres)	PDFs	Weights w/o generator weight	Total Weight
Avg Event	PDF261062	0.0106	0.0105
Anomalous Event	PDF261062	-0.0090	-522.6090

Diboson1L(WDB1LVR1boost)	PDFs	Weight w/o generator weight	Total Weight
Avg Event	PDF261042	0.0344	0.0346
Anomalous Event	PDF261042	-0.1571	-749243.3125

Following the above findings, and the fact that we observe a few PDF variations that are $< 100\%$ but $> 20\%$ highlighted in gray in tables 4.6 and 4.7, one can always question whether there are more events containing significantly large weights that may have altered the uncertainties to have values larger than actual. A follow-up analysis was performed for Diboson 1L and Diboson $\geq$2L samples to fetch unexpectedly large weights in the regions with uncertainties $> 20\%$. The findings are summarised in table 4.9, where the weighted contribution (%) denotes the contribution of the sample in question to the total SM background obtained from the region. For example, the contribution of the Diboson 2L sample obtained by using the selection conditions for SRHMres constitutes 32.61% of the background comprising of Diboson, multiboson, singletop, and $t\bar{t}$ obtained from the same selection conditions. Weighted contributions are important as they give us an idea of the effect of uncertainties.

Table 4.9: Large weight values corresponding to uncertainties $> 20\%$.

Diboson $\geq$2L	Weighted Contribution (%)	PDFs	Generator Weight	Uncertainty
SRHMres	32.61%	PDF261056	320.3471	-0.3181
DB2LVR	66.68%	PDF261072	15927.1172	-0.2319

Diboson1L	Weighted Contribution (%)	PDFs	Generator Weight	Uncertainty
WDB1LVR1boost	26.67%	PDF261079	-163.5695	0.2455
WDB1LCRres	34.31%	PDF261073	126.2110	–

The weights of the two events with uncertainties $\gg 100\%$ shown in table 4.8 were set to unity. These events were verified to be no special and thus the associated unexpectedly large weight was corrected to unity. Redefining the weights of two events doesn't affect the weighted SM background yields, however, the resulting uncertainties reduced to well within the acceptable range as shown in table 4.10. Hence, it was concluded that the an unknown cause in MC generator assigned a ridiculously large weights to the two events which justifies the procedure here.

Table 4.10: Comparison of uncertainties and yields before and after setting the anomalous weights to unity for SRHMres in Diboson 2L and WDB1LVR1boost in Diboson 1L samples.

Region	Generator Weight	Previous Yields	New Yields	Previous Uncertainties	New Uncertainties
2L SRHMres	57992.7773	-513.6083	9.2113	-60.6839	0.0449
1L WDB1LVR1boost	4768614.0000	-749101.625	141.5644	-5551.2637	0.0489

Interpretation for likelihood fit

The final step of this work is to summarise the scale variations and the largest absolute PDF uncertainties for a likelihood fit of the SM background. The uncertainties from diboson 1L and diboson $\geq$ 2L samples are divided into four systematic groups for each region,

I *Renormalisation*, that includes limited knowledge of the dependency of the strong force coefficient on renormalisation factor. i.e. $\mu_R 2 \mu_F 1$ and $\mu_R 0.5 \mu_F 1$

II *Factorisation*, that includes limited knowledge of the dependency of the strong force coefficient on factorisation factor i.e. $\mu_R 1 \mu_F 2$ and $\mu_R 1 \mu_F 0.5$

III *Renormal_Factor*, is used to include the combined effect of renormalisation and factorisation factors comprising of $\mu_R 2 \mu_F 2$ and $\mu_R 0.5 \mu_F 0.5$

IV *PDF*, represents the non-perturbative effects of the strong force. It contains the envelop of largest absolute PDF uncertainty in the region.

For Renormalisation, Factorisation, and Renormal_Factor, the larger of the two-scale uncertainties in each of these groups is denoted as *up* and the lower one as *down*. For the case of PDF uncertainties, the largest available uncertainty is enveloped, hence, the largest absolute value with a positive sign is denoted as *up* and with a negative sign as *down*. The variations for Diboson 1L and Diboson $\geq$2L samples are summarised in tables 4.11 and 4.12.

The likelihood fit is the subsequent step in the analysis and the data shown in tables 4.11 and 4.12 have already been transferred to the ATLAS Collaboration. We should expect the results of the background-only likelihood fit in the near future.

Table 4.11: PDF and scale variations required for likelihood plot of the Diboson 1L background.

Diboson1L	Systematic	up	down
SRLMresolved	Renormalisation	1.+0.1207	1.-0.1699
	Factorisation	1.-0.0169	1.-0.0697
	Renormal_Factor	1.+0.1503	1.-0.1747
	PDF	1.+0.1993	1.-0.1993
SRHMresolved	Renormalisation	1.+0.2341	1.-0.1603
	Factorisation	1.-0.0060	1.-0.0073
	Renormal_Factor	1.+0.2375	1.-0.1596
	PDF	1.+0.1648	1.-0.1648
SRLMboostedWZ	Renormalisation	1.+0.0602	1.-0.0552
	Factorisation	1.+0.0814	1.+0.0122
	Renormal_Factor	1.+0.0691	1.+0.0084
	PDF	1.+0.0834	1.-0.0834
SRMMboostedWZ	Renormalisation	1.+0.0493	1.-0.0591
	Factorisation	1.+0.0728	1.+0.0244
	Renormal_Factor	1.+0.0686	1.-0.0031
	PDF	1.+0.0785	1.-0.0785
SRHMboostedWZ	Renormalisation	1.+0.0497	1.-0.0390
	Factorisation	1.+0.0734	1.+0.0209
	Renormal_Factor	1.+0.0499	1.+0.0204
	PDF	1.+0.0684	1.-0.0684
SRLMboostedWW	Renormalisation	1.+0.0232	1.-0.0257
	Factorisation	1.+0.0880	1.+0.0053
	Renormal_Factor	1.+0.0440	1.+0.0152
	PDF	1.+0.0722	1.-0.0722
SRMMboostedWW	Renormalisation	1.+0.0873	1.-0.0753
	Factorisation	1.+0.0687	1.+0.0268
	Renormal_Factor	1.+0.1103	1.-0.0258
	PDF	1.+0.0959	1.-0.0959
SRHMboostedWW	Renormalisation	1.+0.0604	1.-0.0524
	Factorisation	1.+0.0677	1.+0.0270
	Renormal_Factor	1.+0.0556	1.-0.0018
	PDF	1.+0.1183	1.-0.1183
DB2LCR	Renormalisation	1.+0.2019	1.-0.1842
	Factorisation	1.+0.0986	1.-0.0096
	Renormal_Factor	1.+0.3306	1.-0.1959
	PDF	1.+0.1016	1.-0.1016
DB2LVR	Renormalisation	1.+0.0001	1.-0.0166
	Factorisation	1.+0.2208	1.-0.1137
	Renormal_Factor	1.+0.2462	1.-0.0931
	PDF	1.+0.1952	1.-0.1952

Diboson1L	Systematic	up	down
WDB1LCRboosted	Renormalisation	1.+0.0000	1.+0.0000
	Factorisation	1.+0.0000	1.+0.0000
	Renormal_Factor	1.+0.0000	1.+0.0000
	PDF	1.+0.0000	1.-0.0000
WDB1LVR1boosted	Renormalisation	1.+0.0846	1.-0.0993
	Factorisation	1.+0.0903	1.-0.0035
	Renormal_Factor	1.+0.1660	1.-0.1123
	PDF	1.+0.2455	1.-0.2455
WDB1LVR2boosted	Renormalisation	1.+0.0994	1.-0.1428
	Factorisation	1.+0.1056	1.-0.0096
	Renormal_Factor	1.+0.1920	1.-0.1741
	PDF	1.+0.0830	1.-0.0830
WBD1LCRresolved	Renormalisation	1.+0.0000	1.+0.0000
	Factorisation	1.+0.0000	1.+0.0000
	Renormal_Factor	1.+0.0000	1.+0.0000
	PDF	1.+0.0000	1.-0.0000
WDB1LVRresolved	Renormalisation	1.+0.1181	1.-0.0796
	Factorisation	1.+0.0070	1.-0.0095
	Renormal_Factor	1.+0.1372	1.-0.0828
	PDF	1.+0.2001	1.-0.2001
TCRboosted	Renormalisation	1.+0.1145	1.-0.1540
	Factorisation	1.+0.0959	1.-0.0143
	Renormal_Factor	1.+0.2026	1.-0.1574
	PDF	1.+0.0718	1.-0.0718
TVR1boosted	Renormalisation	1.+0.1704	1.-0.2189
	Factorisation	1.+0.1324	1.-0.0325
	Renormal_Factor	1.+0.2925	1.-0.2671
	PDF	1.+0.0903	1.-0.0903
TVR2boosted	Renormalisation	1.+0.0770	1.-0.1018
	Factorisation	1.+0.1063	1.-0.0271
	Renormal_Factor	1.+0.1675	1.-0.1475
	PDF	1.+0.0733	1.-0.0733
TCRresolved	Renormalisation	1.+0.0624	1.-0.0396
	Factorisation	1.+0.0042	1.-0.0059
	Renormal_Factor	1.+0.0600	1.-0.0343
	PDF	1.+0.2224	1.-0.2224
TVRresolved	Renormalisation	1.+0.1397	1.-0.0849
	Factorisation	1.+0.0154	1.-0.0146
	Renormal_Factor	1.+0.1700	1.-0.0919
	PDF	1.+0.6117	1.-0.6117

Table 4.12: PDF and scale variations required for likelihood plot of the Diboson $\geq$ 2L background.

Diboson≥2L	Systematic	up	down
SRLMresolved	Renormalisation	1.+0.0116	1.-0.0181
	Factorisation	1.+0.0086	1.-0.0147
	Renormal_Factor	1.+0.0036	1.-0.0024
	PDF	1.+0.0678	1.-0.0678
SRHMresolved	Renormalisation	1.+0.0326	1.-0.0271
	Factorisation	1.+0.0339	1.-0.0291
	Renormal_Factor	1.+0.0040	1.+0.0005
	PDF	1.+0.3181	1.-0.3181
SRLMboostedWZ	Renormalisation	1.+0.0226	1.-0.0344
	Factorisation	1.+0.0235	1.-0.0204
	Renormal_Factor	1.+0.0027	1.-0.0063
	PDF	1.+0.0824	1.-0.0824
SRMMboostedWZ	Renormalisation	1.+0.0217	1.-0.0136
	Factorisation	1.+0.0036	1.+0.0009
	Renormal_Factor	1.+0.0197	1.-0.0142
	PDF	1.+0.0817	1.-0.0817
SRHMboostedWZ	Renormalisation	1.+0.0117	1.-0.0137
	Factorisation	1.+0.0025	1.-0.0004
	Renormal_Factor	1.+0.0104	1.-0.0132
	PDF	1.+0.3715	1.-0.3715
SRLMboostedWW	Renormalisation	1.+0.0151	1.-0.0114
	Factorisation	1.+0.0249	1.-0.0194
	Renormal_Factor	1.+0.0146	1.-0.0035
	PDF	1.+0.0812	1.-0.0812
SRMMboostedWW	Renormalisation	1.+0.0160	1.-0.0082
	Factorisation	1.+0.0099	1.-0.0053
	Renormal_Factor	1.+0.0086	1.-0.0015
	PDF	1.+0.0867	1.-0.0867
SRHMboostedWW	Renormalisation	1.+0.0177	1.-0.0094
	Factorisation	1.+0.0222	1.-0.0163
	Renormal_Factor	1.+0.0090	1.-0.0013
	PDF	1.+0.0984	1.-0.0984
DB2LCR	Renormalisation	1.+0.0000	1.+0.0000
	Factorisation	1.+0.0000	1.+0.0000
	Renormal_Factor	1.+0.0000	1.+0.0000
	PDF	1.+0.0000	1.-0.0000
DB2LVR	Renormalisation	1.+0.0065	1.-0.0101
	Factorisation	1.+0.0091	1.-0.0103
	Renormal_Factor	1.+0.0185	1.-0.0181
	PDF	1.+0.2319	1.-0.2319

Diboson1L	Systematic	up	down
WDB1LCRboosted	Renormalisation	1.+0.0280	1.-0.0287
	Factorisation	1.+0.0142	1.-0.0106
	Renormal_Factor	1.+0.0412	1.-0.0396
	PDF	1.+0.1417	1.-0.1417
WDB1LVR1boosted	Renormalisation	1.+0.0036	1.-0.0044
	Factorisation	1.+0.0011	1.+0.0004
	Renormal_Factor	1.+0.0044	1.-0.0036
	PDF	1.+0.0915	1.-0.0915
WDB1LVR2boosted	Renormalisation	1.+0.0224	1.-0.0249
	Factorisation	1.+0.0011	1.-0.0001
	Renormal_Factor	1.+0.0215	1.-0.0226
	PDF	1.+0.0843	1.-0.0843
WBD1LCRresolved	Renormalisation	1.+0.0290	1.-0.0238
	Factorisation	1.+0.0148	1.-0.0083
	Renormal_Factor	1.+0.0197	1.-0.0093
	PDF	1.+0.1598	1.-0.1598
WDB1LVRresolved	Renormalisation	1.+0.0013	1.-0.0161
	Factorisation	1.-0.0050	1.-0.0128
	Renormal_Factor	1.+0.0110	1.-0.0161
	PDF	1.+0.0908	1.-0.0908
TCRboosted	Renormalisation	1.+0.0381	1.-0.0634
	Factorisation	1.+0.0063	1.-0.0024
	Renormal_Factor	1.+0.0452	1.-0.0655
	PDF	1.+0.0662	1.-0.0662
TVR1boosted	Renormalisation	1.+0.0097	1.-0.0006
	Factorisation	1.+0.0063	1.-0.0199
	Renormal_Factor	1.+0.0151	1.-0.0220
	PDF	1.+0.0653	1.-0.0653
TVR2boosted	Renormalisation	1.+0.0281	1.-0.0279
	Factorisation	1.+0.0099	1.+0.0022
	Renormal_Factor	1.+0.0311	1.-0.0172
	PDF	1.+0.0693	1.-0.0693
TCRresolved	Renormalisation	1.+0.0315	1.-0.0315
	Factorisation	1.+0.0317	1.-0.0320
	Renormal_Factor	1.+0.0601	1.-0.0683
	PDF	1.+0.1981	1.-0.1981
TVRresolved	Renormalisation	1.+0.0516	1.-0.0402
	Factorisation	1.+0.0161	1.-0.0108
	Renormal_Factor	1.+0.0416	1.-0.0246
	PDF	1.+0.0774	1.-0.0774

Chapter 5

Concluding Remarks

5.1 SM yield variation in t-quark samples

The first study presented in this work is aimed at quantifying the kinematic agreements between the signal and control/validation regions for the electroweak one-lepton decay channels. The analysis focuses mainly on the singletop MC samples that are simulated with a mix of three independent production modes of the t-quark namely, namely, the $s-$, $t-$ and Wt$-$ channels. The estimated SM background yields in the production modes of top quark showed a discrepancy in the control/validation and the signal region. We want to check if there is any observable bias due to the mixing in the singletop sample when the kinematic region definitions are used to estimate the SM background. To do this, we quantify the yield variations by evaluating the systematic effect which was observed to be less than 5%. Hence, we can conclude that the defined control/validation regions are kinematically similar enough to be accepted for the later stage of the analyses, and the yield variations for different productions channels are negligible.

5.2 Theoretical Uncertainties in diboson sample

The second part of the work is aimed at evaluating the uncertainties associated with the limited understanding of the physics of the diboson production. These uncertainties arise due to divergences and non-perturbative effects in the calculation of parton distribution functions (PDFs) which play an important role in statistically evaluating the observed data. In this thesis, we work with the diboson decay channels with at least one charged lepton. The largest of the uncertainties evaluated for scale and PDF variations were observed to be $\sim 30\%$ and $\sim 40\%$ respectively. This error is larger than the associated MC error in SM background yields for the respective regions. Hence, the estimated theoretical uncertainties provide a new limit for the estimated background events. A better approximation of the error increases the signal region sensitivity to possible SUSY signals.

5.3 Remarks

The two different analyses mentioned above stems from two different ideas, aimed towards increasing the sensitivity of our signal model to detect SUSY events. In the first part, the three independent top production channels are mixed together and events are evaluated as a combined MC sample, singletop. Instead of having three different scale factors, the mixing allows us to simplify our analysis by having a single scaling factor for three channels of production. This simplification is possible at the expense of systematic effects that were quantitatively evaluated to be negligible.

The intention behind the second analysis is different. We are adding the theoretical uncertainties in our estimation of the SM backgrounds as a cost for our poor understanding of the physics leading to the diboson production. Instead of simplifying the model, we are including the uncertainties because it is important to understand the effects of unknown physics in the analysis and to evaluate how large are the affects of inclusion. The results show that the scale and PDF uncertainties are acceptable and hence, including the uncertainties make our analysis strategy more accurate.

Bibliography

[1] P.A. Zyla *et al.*[Particle Data Group]. Review of Particle Physics. *PTEP*, 083C01(8), 2020.

[2] Leonard Susskind. Dynamics of spontaneous symmetry breaking in the weinberg-salam theory. *Phys. Rev. D*, 20:2619–2625, Nov 1979.

[3] G.'t Hooft. *Naturalness, Chiral Symmetry, and Spontaneous Chiral Symmetry Breaking*, pages 135–157. Springer US, Boston, MA, 1980.

[4] Philip Nelson. Naturalness in theoretical physics: Internal constraints on theories, especially the requirement of naturalness, play a pivotal role in physics. *American Scientist*, 73(1):60–67, 1985.

[5] Gian Francesco Giudice. Naturally speaking: The naturalness criterion and physics at the lhc. *Perspectives on LHC Physics*, page 155–178, Jun 2008.

[6] Cliff P. Burgess. Quantum gravity in everyday life: General relativity as an effective field theory. *Living Reviews in Relativity*, 7(1), Apr 2004.

[7] H. Baer and X. Tata. *Weak scale supersymmetry: From superfields to scattering events*. Cambridge University Press, 5 2006.

[8] Gian F. Giudice, Riccardo Rattazzi, and James D. Wells. Quantum gravity and extra dimensions at high-energy colliders. *Nuclear Physics B*, 544(1-2):3–38, Apr 1999.

[9] Ben Gripaios, Marco Nardecchia, and Tevong You. On the structure of anomalous composite higgs models. *The European Physical Journal C*, 77(1), Jan 2017.

[10] Katherine Garrett and Gintaras Duda. Dark matter: A primer. *Advances in Astronomy*, 2011:1–22, 2011.

[11] Yoshiaki Sofue and Vera Rubin. Rotation curves of spiral galaxies. *Annual Review of Astronomy and Astrophysics*, 39(1):137–174, 2001.

[12] V. C. Rubin. Dark matter in spiral galaxies. *Scientific American*, 248:96–106, June 1983.

[13] M. Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, W. Forman, C. Jones, S. Murray, and W. Tucker. Direct constraints on the dark matter self-interaction cross section from the merging galaxy cluster 1e 065756. *The Astrophysical Journal*, 606(2):819–824, May 2004.

[14] M. Tanabashi *et al.* Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[15] N. Jarosik *et al.* SEVEN-YEAR WILKINSON MICROWAVE ANISOTROPY PROBE (WMAP) OBSERVATIONS: SKY MAPS, SYSTEMATIC ERRORS, AND BASIC RESULTS. *The Astrophysical Journal Supplement Series*, 192(2):14, jan 2011.

[16] M. Milgrom. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. , 270:365–370, July 1983.

[17] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: evidence, candidates and constraints. *Physics Reports*, 405(5):279–390, 2005.

[18] STEPHEN P. MARTIN. A supersymmetry primer. *Advanced Series on Directions in High Energy Physics*, pages 98–100, Jul 1998.

[19] J. Wess and B. Zumino. Supergauge transformations in four dimensions. *Nuclear Physics B*, 70(1):39–50, 1974.

[20] Sidney Coleman and Jeffrey Mandula. All possible symmetries of the s matrix. *Phys. Rev.*, 159:1251–1256, Jul 1967.

[21] Rudolf Haag, Jan T. Łopuszański, and Martin Sohnius. All possible generators of supersymmetries of the s-matrix. *Nuclear Physics B*, 88(2):257–274, 1975.

[22] K. Abe *et al.* Search for proton decay via $p \to e^+\pi^0$ and $p \to \mu^+\pi^0$ in 0.31 megaton $\cdot$ years exposure of the super-kamiokande water cherenkov detector. *Phys. Rev. D*, 95:012004, Jan 2017.

[23] Rudy C. Gilmore. Mass limits on neutralino dark matter. *Physical Review D*, 76(4), Aug 2007.

[24] A. Djouadi et al. The minimal supersymmetric standard model: Group summary report, 1999.

[25] Johan Alwall, Philip C. Schuster, and Natalia Toro. Simplified models for a first characterization of new physics at the lhc. *Physical Review D*, 79(7), Apr 2009.

[26] Daniele Alves, Nima Arkani-Hamed, Sanjay Arora, Yang Bai, Matthew Baumgart, Joshua Berger, Matthew Buckley, Bart Butler, Spencer Chang, Hsin-Chia Cheng, and et al. Simplified models for lhc new physics searches. *Journal of Physics G: Nuclear and Particle Physics*, 39(10):105005, Sep 2012.

[27] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, S.H. Abidi, O.S. AbouZeid, N.L. Abraham, H. Abramowicz, H. Abreu, and et al. Search for squarks and gluinos in events with an isolated lepton, jets, and missing transverse momentum at s=13tev with the atlas detector. *Physical Review D*, 96(11), Dec 2017.

[28] Anadi Canepa, Tao Han, and Xing Wang. The search for electroweakinos. *Annual Review of Nuclear and Particle Science*, 70(1):425–454, Oct 2020.

[29] About CERN. Jan 2012.

[30] LHC Guide. Mar 2017.

[31] Lyndon R Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001. 164 p, 2008. This report is an abridged version of the LHC Design Report (CERN-2004-003).

[32] J-P Burnet. Putting it into Practice. page 13 p, Jul 2016. 13 pages, contribution to the 2014 CAS - CERN Accelerator School: Power Converters, Baden, Switzerland, 7-14 May 2014.

[33] **ATLAS** Collaboration. ATLAS Public Results - Luminosity Public Results Run 1. Accessed: May 2021.

[34] **ATLAS** Collaboration. ATLAS Public Results - Luminosity Public Results Run 2. Accessed: May 2021.

[35] The ATLAS Collaboration. The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation*, 3(08):S08003–S08003, aug 2008.

[36] Julie Haffner. The CERN accelerator complex. Complexe des accélérateurs du CERN. Oct 2013. General Photo.

[37] Jean-Luc Caron. LHC layout. Schema general du LHC. AC Collection. Legacy of AC. Pictures from 1992 to 2002., Sep 1997.

[38] Joao Pequenao. Computer generated image of the whole ATLAS detector. Mar 2008.

[39] M Capeans, G Darbo, K Einsweiller, M Elsing, T Flick, M Garcia-Sciveres, C Gemme, H Pernegger, O Rohne, and R Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical report, Sep 2010.

[40] Joao Pequenao. Computer generated image of the ATLAS inner detector. Mar 2008.

[41] **ATLAS** Collaboration. Alignment of the ATLAS Inner Detector in Run 2. *Eur. Phys. J.*, C 80:1194 p, 2020.

[42] Garelli, Nicoletta. Performance of the atlas detector in run-2. *EPJ Web Conf.*, 164:01021, 2017.

[43] ATLAS *Approved DAQ Plots*.

[44] **ATLAS** Collaboration. ATLAS Public Results - Trigger Operation 1. Accessed: May 2021.

[45] Performance of the ATLAS global transverse-momentum triggers at $\sqrt{s} = 8$ TeV. Technical report, CERN, Geneva, Apr 2018.

[46] Stefan Weinzierl. Introduction to monte carlo methods, 2000.

[47] G. Aad, B. Abbott, J. Abdallah, A. A. Abdelalim, A. Abdesselam, O. Abdinov, B. Abi, M. Abolins, H. Abramowicz, and et al. The atlas simulation infrastructure. *The European Physical Journal C*, 70(3):823–874, Sep 2010.

[48] **ATLAS** Collaboration. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[49] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box. *Journal of High Energy Physics*, 2010(6), Jun 2010.

[50] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. *Journal of High Energy Physics*, 2007(11):070–070, Nov 2007.

[51] Stefano Frixione, Giovanni Ridolfi, and Paolo Nason. A positive-weight next-to-leading-order monte carlo for heavy flavour hadroproduction. *Journal of High Energy Physics*, 2007(09):126–126, Sep 2007.

[52] Paolo Nason. A new method for combining nlo qcd with shower monte carlo algorithms. *Journal of High Energy Physics*, 2004(11):040–040, Nov 2004.

[53] Emanuele Re. Single-top wt-channel production matched with parton showers using the powheg method. *The European Physical Journal C*, 71(2), Feb 2011.

[54] Rikkert Frederix, Emanuele Re, and Paolo Torrielli. Single-top t-channel hadroproduction in the four-flavour scheme with powheg and amc@nlo. *Journal of High Energy Physics*, 2012(9), Sep 2012.

[55] Enrico Bothmann, Gurpreet Singh Chahal, Stefan Höche, Johannes Krause, Frank Krauss, Silvan Kuttimalai, Sebastian Liebschner, Davide Napoletano, Marek Schönherr, Holger Schulz, and et al. Event generation with sherpa 2.2. *SciPost Physics*, 7(3), Sep 2019.

[56] Michael H. Seymour and Marilyn Marx. Monte carlo event generators, 2013.

[57] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11):852–867, 2008.

[58] David J. Lange. The evtgen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1):152–155, 2001. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[59] Johannes Bellm, Stefan Gieseke, David Grellscheid, Simon Plätzer, Michael Rauch, Christian Reuschle, Peter Richardson, Peter Schichtel, Michael H. Seymour, Andrzej Siódmok, and et al. Herwig 7.0/herwig++ 3.0 release note. *The European Physical Journal C*, 76(4), Apr 2016.

[60] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026–026, May 2006.

[61] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, José I. Latorre, Juan Rojo, and Maria Ubiali. Parton distributions with lhc data. *Nuclear Physics B*, 867(2):244–289, 2013.

[62] The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model. Technical report, CERN, Geneva, Aug 2016. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-017.

[63] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), Jul 2014.

[64] ATLAS Pythia 8 tunes to 7 TeV data. Technical report, CERN, Geneva, Nov 2014. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-021.

[65] G. Aad, B. Abbott, D.C. Abbott, A. Abed Abud, K. Abeling, D.K. Abhayasinghe, S.H. Abidi, O.S. AbouZeid, N.L. Abraham, H. Abramowicz, and et al. Electron and photon performance measurements with the atlas detector using the 2015–2017 lhc proton-proton collision data. *Journal of Instrumentation*, 14(12):P12006–P12006, Dec 2019.

[66] Geoffrey Gilles. Top quark production at the LHC. Apr 2018.

[67] Silvia Biondi. Study of the associated production of the Higgs boson with a top quark pair in a boosted regime in the ATLAS experiment at LHC., Jan 2017. Presented 20 Mar 2017.

[68] Eric Schanet. Sensitivity studies for a search for charginos in final states with one lepton with the atlas detector and interpretation of a search for gluinos and squarks in the pmssm. **Master Thesis**, *Munich U.*, 2018.

[69] Abdullah Khalil and W. A. Horowitz. A complete diagrammatic implementation of the kinoshita-lee-nauenberg theorem at next-to-leading order, 2017.

Appendices

Appendix A

A.1 Cutflow Tables for top background

In order to understand better about the impact of kinematic variables on the event selection, cutflow tables are used. These tables are region specific unlike the breakdown tables that give a concise summary of all the regions in a single table. During the analysis anomalies can be easily identified using cutflow tables and hence they are also used for debugging purposes. This section consists of cutflow tables related to every region definition used during the estimation of the SM background consisting of singletop and $t\bar{t}$ events from section 4.2 of chapter 4.

Table A.1: SRLMres Cutflow table for top and $t\bar{t}$ background

SRLMres	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31674.08 ± 53.62	159587.91 ± 88.06
$2-3$ jets	317.89 ± 3.29	3551.97 ± 16.46	10232.71 ± 31.69	14102.57 ± 51.43	57123.45 ± 56.22
no b-jet	43.89 ± 1.22	968.98 ± 8.83	2849.33 ± 16.86	3862.19 ± 26.91	13556.08 ± 27.64
$\Delta\phi(l, E_T^{miss}) < 2.8$	43.46 ± 1.21	966.87 ± 8.82	2664.84 ± 16.30	3675.17 ± 26.33	12399.37 ± 26.45
no large-R jet	8.44 ± 0.52	339.78 ± 5.79	575.36 ± 7.70	923.58 ± 14.01	4234.09 ± 17.16
$70 <= mjj <= 105$	0.67 ± 0.14	27.23 ± 1.27	136.88 ± 3.81	164.77 ± 5.22	931.43 ± 8.33
$200 < m_T <= 380$	0.00 ± 0.00	0.00 ± 0.00	7.90 ± 0.89	7.90 ± 0.89	69.40 ± 2.24

Table A.2: SRHMres Cutflow table for top and $t\bar{t}$ background

SRHMres	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31674.08 ± 53.62	159587.91 ± 88.06
$2-3$ jets	317.89 ± 3.29	3551.97 ± 16.46	10232.71 ± 31.69	14102.57 ± 51.43	57123.45 ± 56.22
no b-jet	43.89 ± 1.22	968.98 ± 8.83	2849.33 ± 16.86	3862.19 ± 26.91	13556.08 ± 27.64
$\Delta\phi(l, E_T^{miss}) < 2.8$	43.46 ± 1.21	966.87 ± 8.82	2664.84 ± 16.30	3675.17 ± 26.33	12399.37 ± 26.45
no large-R jet	8.44 ± 0.52	339.78 ± 5.79	575.36 ± 7.70	923.58 ± 14.01	4234.09 ± 17.16
$70 <= mjj <= 105$	0.67 ± 0.14	27.23 ± 1.27	136.88 ± 3.81	164.77 ± 5.22	931.43 ± 8.33
$m_T > 380$	0.00 ± 0.00	0.00 ± 0.00	2.32 ± 0.46	2.32 ± 0.46	11.42 ± 0.82

Table A.3: Z-SRLMboost Cutflow table for top and $t\bar{t}$ background

Z-SRLMboost	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31754.42 ± 53.68	168752.51 ± 93.38
<= 3 jets	375.71 ± 3.57	4407.85 ± 18.05	11080.08 ± 33.01	15863.65 ± 54.63	64756.65 ± 62.00
$N_{\text{b-jet}} \leq 2$	373.96 ± 3.56	4398.14 ± 18.03	11016.86 ± 32.92	15788.97 ± 54.51	64325.43 ± 61.82
$\Delta\phi(l, E_T^{miss}) < 2.6$	369.75 ± 3.54	4372.14 ± 17.96	9495.79 ± 30.52	14237.68 ± 52.02	50516.84 ± 54.45
E_T^{miss} significance > 15	10.87 ± 0.61	70.00 ± 2.23	656.23 ± 6.88	737.10 ± 9.72	1657.12 ± 5.90
at least one large-R jet	10.04 ± 0.59	53.24 ± 1.97	562.04 ± 6.27	625.32 ± 8.83	1178.45 ± 3.91
Z-tagged large-R jet	0.06 ± 0.04	0.62 ± 0.20	30.00 ± 1.51	30.68 ± 1.75	54.76 ± 0.88
large-R jet $p_T > 250$ GeV	0.06 ± 0.04	0.62 ± 0.20	28.44 ± 1.46	29.12 ± 1.71	50.04 ± 0.77
$120 < m_T < 240$	0.00 ± 0.00	0.00 ± 0.00	0.68 ± 0.19	0.68 ± 0.19	1.72 ± 0.20

Table A.4: Z-SRMMboost Cutflow table for top and $t\bar{t}$ background

Z-SRMMboost	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31754.42 ± 53.68	168752.51 ± 93.38
<= 3 jets	375.71 ± 3.57	4407.85 ± 18.05	11080.08 ± 33.01	15863.65 ± 54.63	64756.65 ± 62.00
$N_{\text{b-jet}} \leq 2$	373.96 ± 3.56	4398.14 ± 18.03	11016.86 ± 32.92	15788.97 ± 54.51	64325.43 ± 61.82
$\Delta\phi(l, E_T^{miss}) < 2.6$	369.75 ± 3.54	4372.14 ± 17.96	9495.79 ± 30.52	14237.68 ± 52.02	50516.84 ± 54.45
E_T^{miss} significance > 15	10.87 ± 0.61	70.00 ± 2.23	656.23 ± 6.88	737.10 ± 9.72	1657.12 ± 5.90
at least one large-R jet	10.04 ± 0.59	53.24 ± 1.97	562.04 ± 6.27	625.32 ± 8.83	1178.45 ± 3.91
Z-tagged large-R jet	0.06 ± 0.04	0.62 ± 0.20	30.00 ± 1.51	30.68 ± 1.75	54.76 ± 0.88
large-R jet $p_T > 250$ GeV	0.06 ± 0.04	0.62 ± 0.20	28.44 ± 1.46	29.12 ± 1.71	50.04 ± 0.77
$240 < m_T < 420$	0.00 ± 0.00	0.00 ± 0.00	0.97 ± 0.30	0.97 ± 0.30	0.98 ± 0.19

Table A.5: Z-SRHMboost Cutflow table for top and $t\bar{t}$ background

Z-SRHMboost	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31754.42 ± 53.68	168752.51 ± 93.38
<= 3 jets	375.71 ± 3.57	4407.85 ± 18.05	11080.08 ± 33.01	15863.65 ± 54.63	64756.65 ± 62.00
$N_{\text{b-jet}} \leq 2$	373.96 ± 3.56	4398.14 ± 18.03	11016.86 ± 32.92	15788.97 ± 54.51	64325.43 ± 61.82
$\Delta\phi(l, E_T^{miss}) < 2.6$	369.75 ± 3.54	4372.14 ± 17.96	9495.79 ± 30.52	14237.68 ± 52.02	50516.84 ± 54.45
E_T^{miss} significance > 15	10.87 ± 0.61	70.00 ± 2.23	656.23 ± 6.88	737.10 ± 9.72	1657.12 ± 5.90
at least one large-R jet	10.04 ± 0.59	53.24 ± 1.97	562.04 ± 6.27	625.32 ± 8.83	1178.45 ± 3.91
Z-tagged large-R jet	0.06 ± 0.04	0.62 ± 0.20	30.00 ± 1.51	30.68 ± 1.75	54.76 ± 0.88
large-R jet $p_T > 250$ GeV	0.06 ± 0.04	0.62 ± 0.20	28.44 ± 1.46	29.12 ± 1.71	50.04 ± 0.77
$m_T > 420$	0.00 ± 0.00	0.00 ± 0.00	0.43 ± 0.20	0.43 ± 0.20	0.16 ± 0.04

Table A.6: W-SRLMboost Cutflow table for top and $t\bar{t}$ background

W-SRLMboost	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31754.42 ± 53.68	168752.51 ± 93.38
<= 3 jets	375.71 ± 3.57	4407.85 ± 18.05	11080.08 ± 33.01	15863.65 ± 54.63	64756.65 ± 62.00
No b-jet	60.71 ± 1.42	1631.85 ± 10.88	3295.42 ± 18.13	4987.98 ± 30.44	16147.69 ± 31.28
$\Delta\phi(l, E_T^{miss}) < 2.9$	60.36 ± 1.42	1629.92 ± 10.87	3062.47 ± 17.47	4752.76 ± 29.76	14872.36 ± 30.01
E_T^{miss} significance > 14	4.66 ± 0.40	66.42 ± 2.10	344.69 ± 5.21	415.77 ± 7.71	1296.19 ± 6.41
at least one large-R jet	4.08 ± 0.38	45.04 ± 1.75	237.49 ± 4.14	286.61 ± 6.26	691.03 ± 3.40
W-tagged large-R jet	0.00 ± 0.00	1.07 ± 0.27	26.08 ± 1.44	27.16 ± 1.70	58.83 ± 0.98
large-R jet $p_T > 300$ GeV	0.00 ± 0.00	0.76 ± 0.23	22.52 ± 1.34	23.28 ± 1.57	42.75 ± 0.75
$120 < m_T < 240$	0.00 ± 0.00	0.00 ± 0.00	0.37 ± 0.17	0.37 ± 0.17	0.76 ± 0.10

Table A.7: W-SRMMboost Cutflow table for top and $t\bar{t}$ background

W-SRMMboost	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31754.42 ± 53.68	168752.51 ± 93.38
<= 3 jets	375.71 ± 3.57	4407.85 ± 18.05	11080.08 ± 33.01	15863.65 ± 54.63	64756.65 ± 62.00
No b-jet	60.71 ± 1.42	1631.85 ± 10.88	3295.42 ± 18.13	4987.98 ± 30.44	16147.69 ± 31.28
$\Delta\phi(l, E_T^{miss}) < 2.9$	60.36 ± 1.42	1629.92 ± 10.87	3062.47 ± 17.47	4752.76 ± 29.76	14872.36 ± 30.01
E_T^{miss} significance > 14	4.66 ± 0.40	66.42 ± 2.10	344.69 ± 5.21	415.77 ± 7.71	1296.19 ± 6.41
at least one large-R jet	4.08 ± 0.38	45.04 ± 1.75	237.49 ± 4.14	286.61 ± 6.26	691.03 ± 3.40
W-tagged large-R jet	0.00 ± 0.00	1.07 ± 0.27	26.08 ± 1.44	27.15 ± 1.70	58.83 ± 0.98
large-R jet $p_T > 300$ GeV	0.00 ± 0.00	0.76 ± 0.23	22.52 ± 1.34	23.28 ± 1.57	42.75 ± 0.75
$240 < m_T < 360$	0.00 ± 0.00	0.00 ± 0.00	0.33 ± 0.18	0.33 ± 0.18	0.24 ± 0.06

Table A.8: W-SRHMboost Cutflow table for top and $t\bar{t}$ background

W-SRHMboost	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31754.42 ± 53.68	168752.51 ± 93.38
<= 3 jets	375.71 ± 3.57	4407.85 ± 18.05	11080.08 ± 33.01	15863.65 ± 54.63	64756.65 ± 62.00
No b-jet	60.71 ± 1.42	1631.85 ± 10.88	3295.42 ± 18.13	4987.98 ± 30.44	16147.69 ± 31.28
$\Delta\phi(l, E_T^{miss}) < 2.9$	60.36 ± 1.42	1629.92 ± 10.87	3062.47 ± 17.47	4752.76 ± 29.76	14872.36 ± 30.01
E_T^{miss} significance > 14	4.66 ± 0.40	66.42 ± 2.10	344.69 ± 5.21	415.77 ± 7.71	1296.19 ± 6.41
at least one large-R jet	4.08 ± 0.38	45.04 ± 1.75	237.49 ± 4.14	286.61 ± 6.26	691.03 ± 3.40
W-tagged large-R jet	0.00 ± 0.00	1.07 ± 0.27	26.08 ± 1.44	27.16 ± 1.70	58.83 ± 0.98
large-R jet $p_T > 300$ GeV	0.00 ± 0.00	0.76 ± 0.23	22.52 ± 1.34	23.28 ± 1.57	42.75 ± 0.75
$360 < m_T < 420$	0.00 ± 0.00	0.00 ± 0.00	0.05 ± 0.02	0.05 ± 0.02	0.10 ± 0.03

Table A.9: TCRboost Cutflow table for top and $t\bar{t}$ background

TCRboost	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31674.08 ± 53.62	168205.72 ± 93.21
<= 3 jets	373.63 ± 3.56	4386.67 ± 18.01	11055.44 ± 32.97	15815.74 ± 54.54	64542.85 ± 61.89
at least one b-jet	313.50 ± 3.26	2762.56 ± 14.37	7768.60 ± 27.56	10844.67 ± 45.19	48453.73 ± 53.44
$\Delta\phi(l, E_T^{miss}) < 2.9$	311.46 ± 3.25	2749.55 ± 14.31	7164.08 ± 26.43	10225.09 ± 44.00	42602.63 ± 50.08
E_T^{miss} significance < 14	294.38 ± 3.16	2653.67 ± 14.07	6171.45 ± 24.92	9119.50 ± 42.15	38510.47 ± 48.64
at least one large-R jet	261.31 ± 2.99	1818.08 ± 12.15	4800.55 ± 21.88	6879.94 ± 37.03	24564.93 ± 35.61
W-tagged large-R jet	3.13 ± 0.33	13.92 ± 0.95	591.48 ± 7.93	608.53 ± 9.21	2351.38 ± 12.09
large-R jet $p_T > 250$ GeV	2.63 ± 0.31	10.46 ± 0.84	491.22 ± 7.23	504.31 ± 8.37	1585.26 ± 9.35
$50 < m_T < 80$	1.62 ± 0.24	6.59 ± 0.70	294.30 ± 5.53	302.51 ± 6.47	1026.50 ± 7.25

Table A.10: TVR1boost Cutflow table for top and $t\bar{t}$ background

TVR1boost	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31674.08 ± 53.62	168205.72 ± 93.21
<= 3 jets	373.63 ± 3.56	4386.67 ± 18.01	11055.44 ± 32.97	15815.74 ± 54.54	64542.85 ± 61.89
at least one b-jet	313.50 ± 3.26	2762.56 ± 14.37	7768.60 ± 27.56	10844.67 ± 45.19	48453.73 ± 53.44
$\Delta\phi(l, E_T^{miss}) < 2.9$	311.46 ± 3.25	2749.55 ± 14.31	7164.08 ± 26.43	10225.09 ± 44.00	42602.63 ± 50.08
E_T^{miss} significance < 14	294.38 ± 3.16	2653.67 ± 14.07	6171.45 ± 24.92	9119.50 ± 42.15	38510.47 ± 48.64
at least one large-R jet	261.31 ± 2.99	1818.08 ± 12.15	4800.55 ± 21.88	6879.94 ± 37.03	24564.93 ± 35.61
W-tagged large-R jet	3.13 ± 0.33	13.92 ± 0.95	591.48 ± 7.93	608.53 ± 9.21	2351.38 ± 12.09
large-R jet $p_T > 250$ GeV	2.63 ± 0.31	10.46 ± 0.84	491.22 ± 7.23	504.31 ± 8.37	1585.26 ± 9.35
$m_T > 80$	1.01 ± 0.18	3.88 ± 0.46	196.91 ± 4.66	201.80 ± 5.30	558.76 ± 5.91

Table A.11: TVR2boost Cutflow table for top and $t\bar{t}$ background

TVR2boost	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31674.08 ± 53.62	168205.72 ± 93.21
<= 3 jets	373.63 ± 3.56	4386.67 ± 18.01	11055.44 ± 32.97	15815.74 ± 54.54	64542.85 ± 61.89
at least one b-jet	313.50 ± 3.26	2762.56 ± 14.37	7768.60 ± 27.56	10844.67 ± 45.19	48453.73 ± 53.44
$\Delta\phi(l, E_T^{miss}) < 2.9$	311.46 ± 3.25	2749.55 ± 14.31	7164.08 ± 26.43	10225.09 ± 44.00	42602.63 ± 50.08
E_T^{miss} significance > 14	17.07 ± 0.76	95.88 ± 2.61	992.63 ± 8.82	1105.58 ± 12.19	4092.16 ± 11.93
at least one large-R jet	15.55 ± 0.73	70.69 ± 2.29	716.25 ± 7.21	802.49 ± 10.23	1929.22 ± 5.70
W-tagged large-R jet	0.07 ± 0.05	0.95 ± 0.25	37.78 ± 1.68	38.80 ± 1.98	81.16 ± 1.21
large-R jet $p_T > 250$ GeV	0.04 ± 0.04	0.89 ± 0.24	34.08 ± 1.60	35.01 ± 1.87	66.37 ± 1.02
$50 < m_T < 120$	0.04 ± 0.04	0.89 ± 0.24	30.85 ± 1.50	31.78 ± 1.78	61.17 ± 0.93

Table A.12: TCRres Cutflow table for top and $t\bar{t}$ background

TCRres	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31674.08 ± 53.62	159587.91 ± 88.06
$2-3$ jets	317.89 ± 3.29	3551.97 ± 16.46	10232.71 ± 31.69	14102.57 ± 51.43	57123.45 ± 56.22
at least one b-jet	274.00 ± 3.06	2582.99 ± 13.89	7383.38 ± 26.83	10240.38 ± 43.77	43567.37 ± 48.95
$\Delta\phi(l, E_T^{miss}) < 2.8$	271.34 ± 3.04	2567.26 ± 13.82	6728.38 ± 25.59	9566.98 ± 42.45	36959.87 ± 45.02
no large-R jet	30.04 ± 0.97	781.11 ± 6.86	1376.13 ± 11.82	2187.27 ± 19.66	12657.54 ± 30.07
$70 <= mjj <= 105$	2.21 ± 0.27	36.67 ± 1.46	193.39 ± 4.56	232.28 ± 6.28	1867.97 ± 11.87
$50 < m_T < 80$	1.11 ± 0.19	20.34 ± 1.09	91.95 ± 3.13	113.40 ± 4.41	854.22 ± 7.75

Table A.13: TVRres Cutflow table for top and $t\bar{t}$ background

TVRres	s-channel	t-channel	Wt-channel	Singletop	$t\bar{t}$
Preselection	-	-	-	31674.08 ± 53.62	159587.91 ± 88.06
$2-3$ jets	317.89 ± 3.29	3551.97 ± 16.46	10232.71 ± 31.69	14102.57 ± 51.43	57123.45 ± 56.22
at least one b-jet	274.00 ± 3.06	2582.99 ± 13.89	7383.38 ± 26.83	10240.38 ± 43.77	43567.37 ± 48.95
$\Delta\phi(l, E_T^{miss}) < 2.8$	271.34 ± 3.04	2567.26 ± 13.82	6728.38 ± 25.59	9566.98 ± 42.45	36959.87 ± 45.02
no large-R jet	30.04 ± 0.97	781.11 ± 6.86	1376.13 ± 11.82	2187.27 ± 19.66	12657.54 ± 30.07
$70 <= mjj <= 105$	2.21 ± 0.27	36.67 ± 1.46	193.39 ± 4.56	232.28 ± 6.28	1867.97 ± 11.87
$80 < m_T < 200$	1.10 ± 0.19	16.17 ± 0.97	68.97 ± 2.78	86.25 ± 3.94	759.64 ± 8.37

A.2 N-1 Plots for singletop sample

This section contains N-1 Plots combined with the respective histograms showing the effect of control and validation regions not being Wt-pure as the signal regions for both the boosted and resolved jets. The plots on the top part of the image shows the normalised[1] N-1 plots overlaying the respective kinematic variables, with $s-$ and $t-$ channel contributions shown in green and the Wt-channel contributions are shown in blue.

The plots at the bottom part of the image quantitatively depicts the effect of $s-$ and $t-$ channel contributions in the form of percentage contamination overlaying the same kinematic variable as the one in the histogram immediately above it in the same figure.

TCRboost

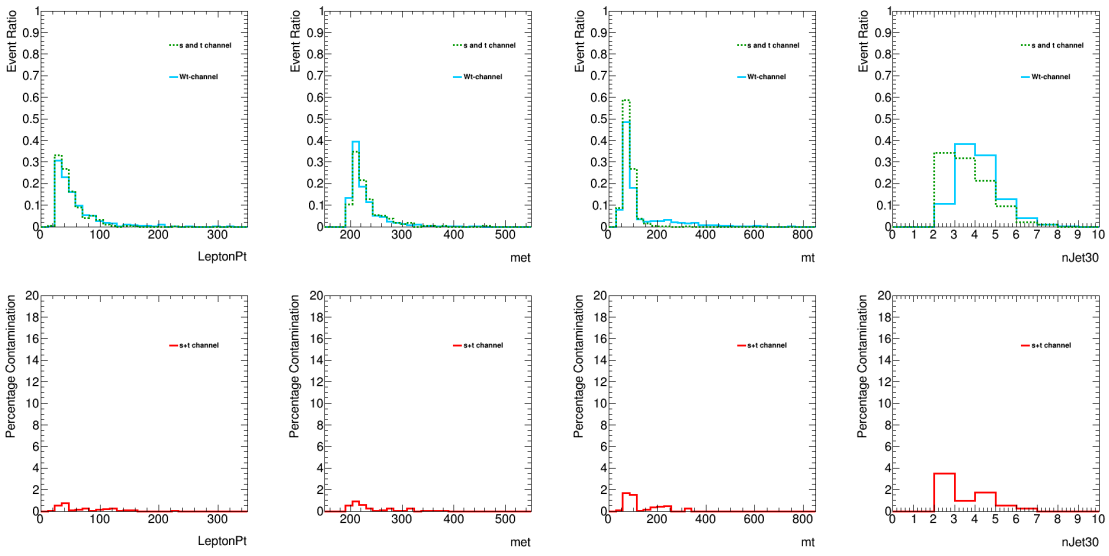


Figure A.1: Normalised N-1 Plots and the percentage contribution from $s-$ and $t-$ channel overlaying kinematic variables of control region of the boosted jets.

[1] w.r.t the integrated B-events for $s-$ and $t-$ channel, integrated S-event for Wt-channel

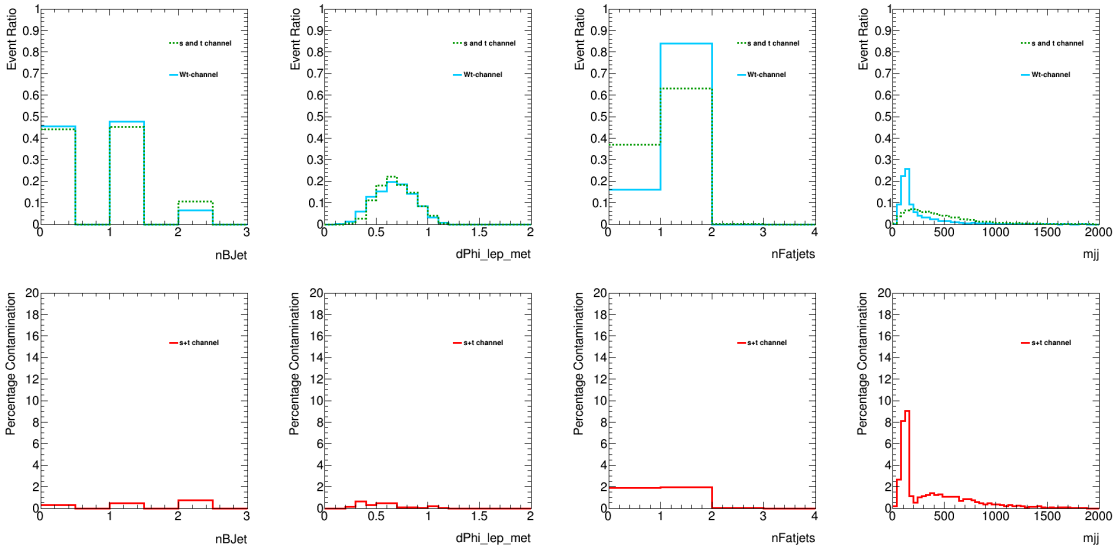


Figure A.2: Normalised N-1 Plots and the percentage contribution from $s-$ and $t-$ channel overlaying kinematic variables of control region of the boosted jets.

TVR1boost

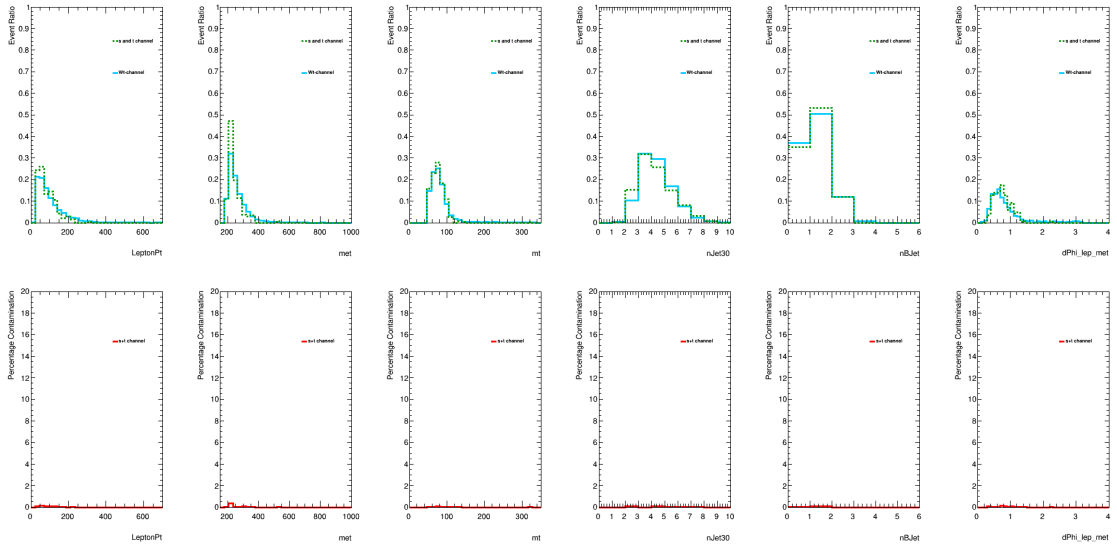


Figure A.3: Normalised N-1 Plots and the percentage contribution from $s-$ and $t-$ channel overlaying kinematic variables of validation region 1 of the boosted jets.

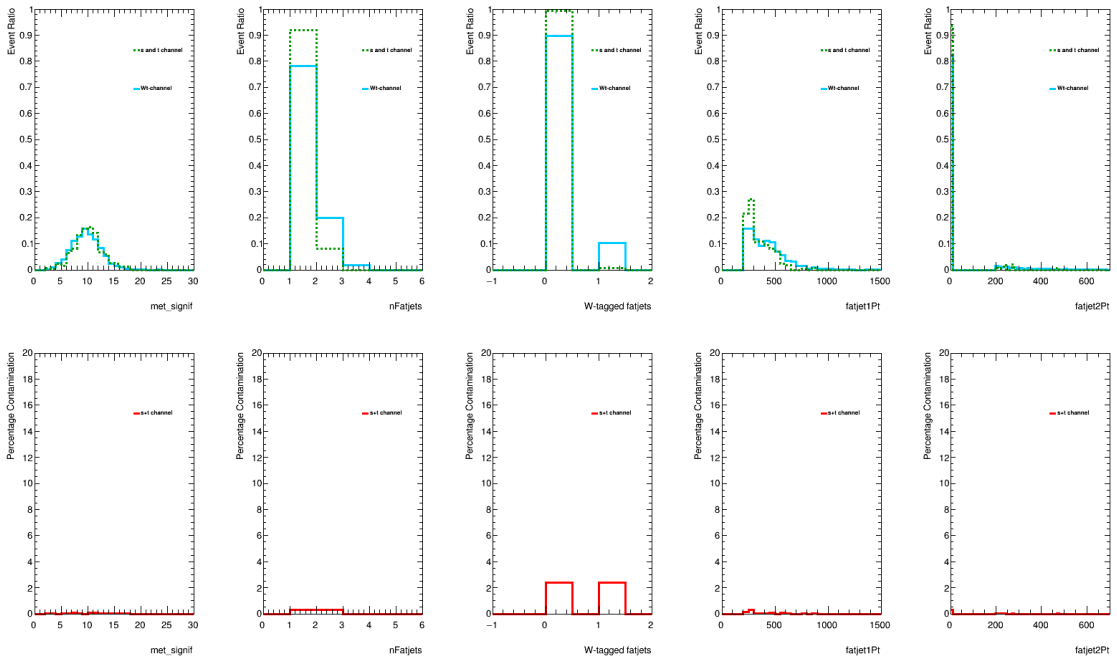


Figure A.4: Normalised N-1 Plots and the percentage contribution from $s-$ and $t-$ channel overlaying kinematic variables of validation region 1 of the boosted jets.

TVR2boost

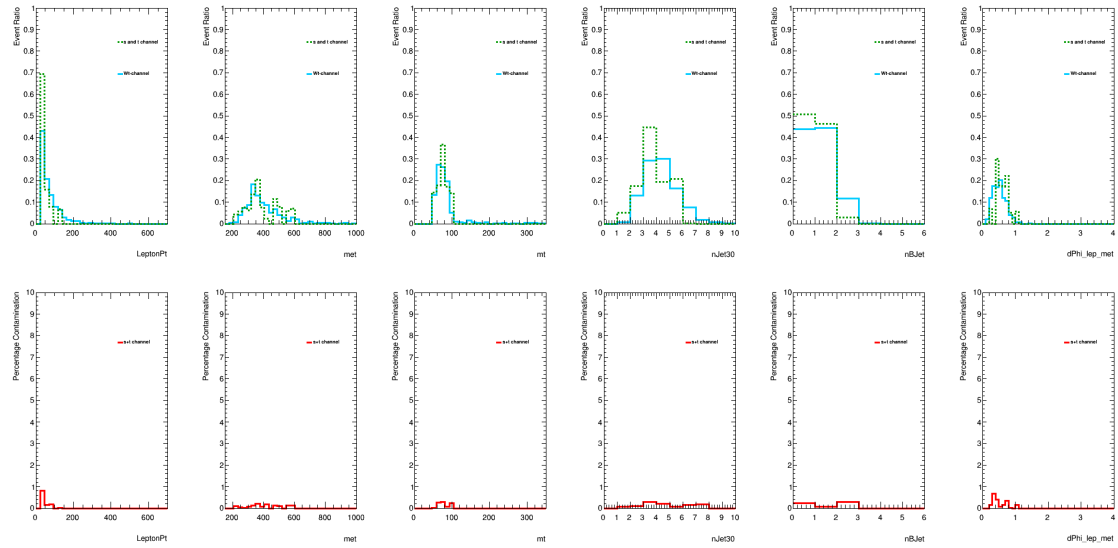


Figure A.5: Normalised N-1 Plots and the percentage contribution from $s-$ and $t-$ channel overlaying kinematic variables of validation region 2 of the boosted jets.

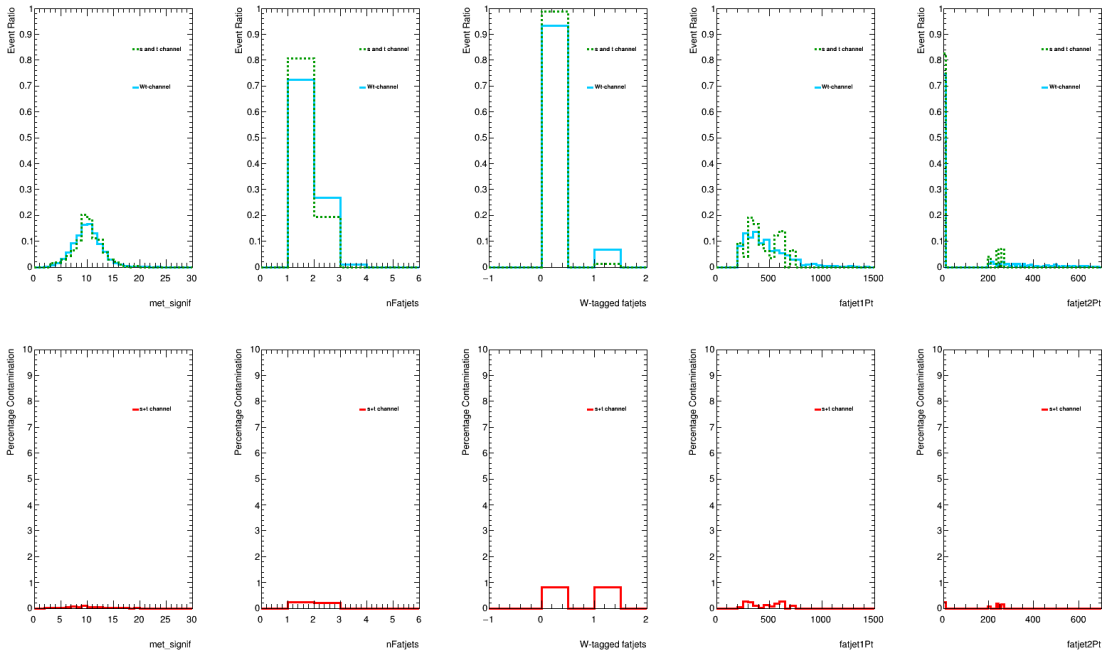


Figure A.6: Normalised N-1 Plots and the percentage contribution from $s-$ and $t-$ channel overlaying kinematic variables of validation region 2 of the boosted jets.

TCRres

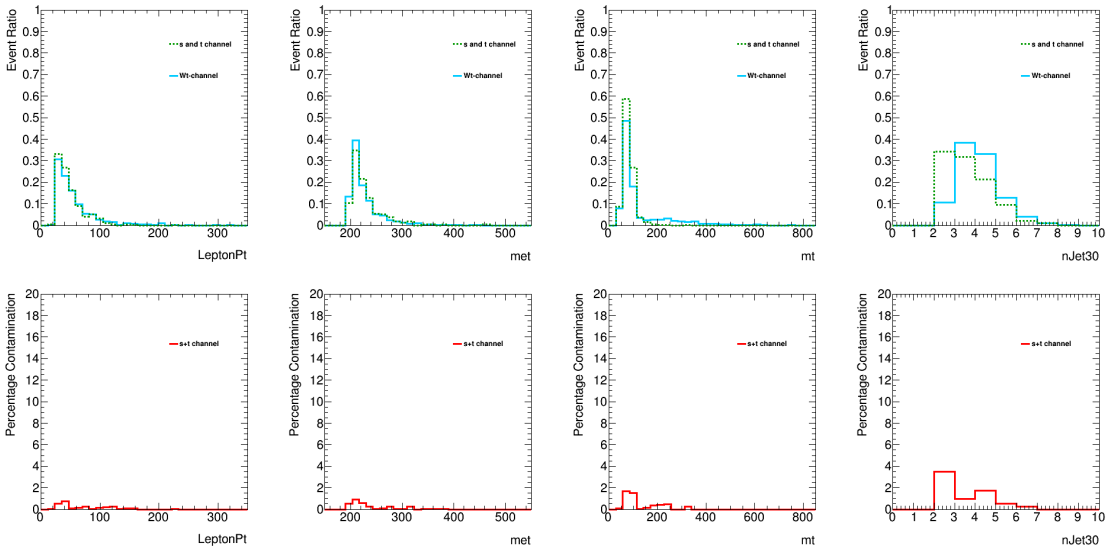


Figure A.7: Normalised N-1 Plots and the percentage contribution from $s-$ and $t-$ channel overlaying kinematic variables of control region of the resolved jets.

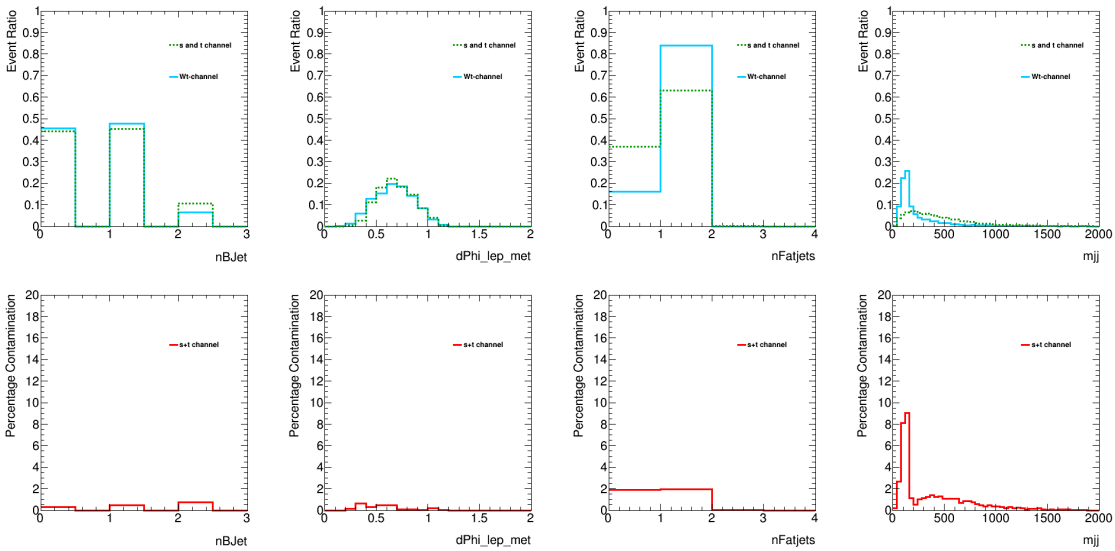


Figure A.8: Normalised N-1 Plots and the percentage contribution from $s-$ and $t-$ channel overlaying kinematic variables of control region of the resolved jets.

TVRres

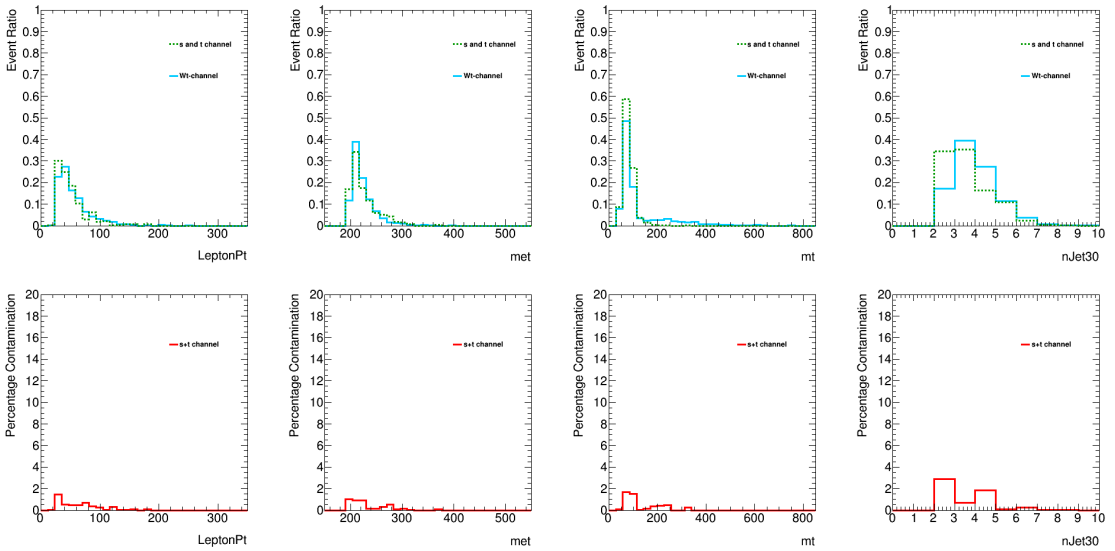


Figure A.9: Normalised N-1 Plots and the percentage contribution from $s-$ and $t-$ channel overlaying kinematic variables of validation region of the resolved jets.

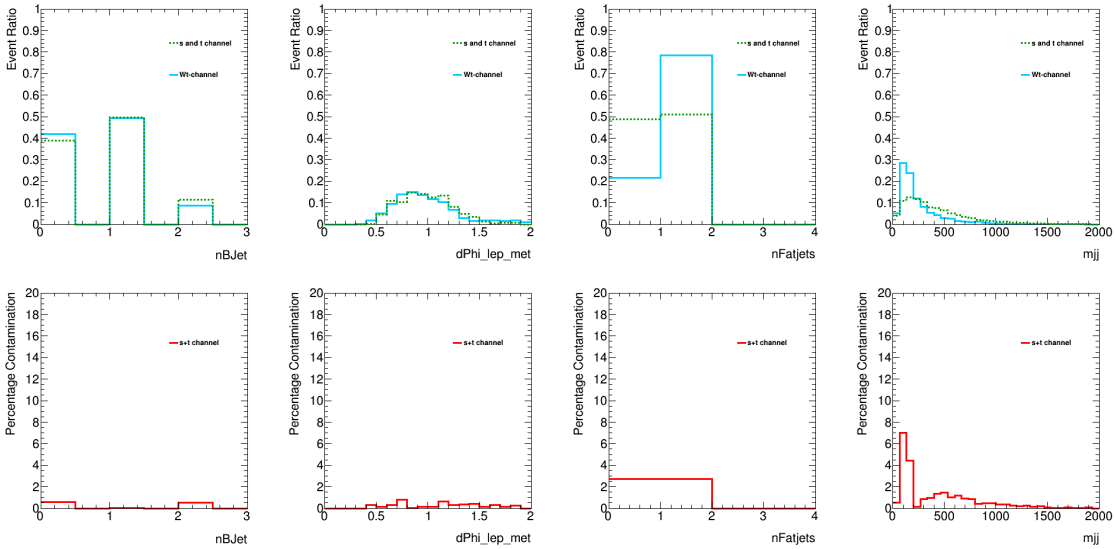


Figure A.10: Normalised N-1 Plots and the percentage contribution from $s-$ and $t-$ channel overlaying kinematic variables of validation region of the resolved jets.