KU LEUVEN

FACULTEIT WETENSCHAPPEN

Sensitivity to new physics in $t\bar{t}$ production at e^+e^- colliders at $\sqrt{s} = 365$ GeV

Olivier DUPON

Promotor: Prof. Riccardo Raabe
Affiliatie *KULeuven*

Co-promotor: Prof. Freya Blekman
Affiliatie *VUB*

Begeleider: Prof. Freya Blekman
Affiliatie *VUB*

Proefschrift ingediend tot het
behalen van de graad van
Master of Science in Physics

Academiejaar 2019-2020

Abstract

The top quark is sensitive to many new physics processes and is an important probe to model-independent testing of the Standard Model through its large mass and consequently dominant coupling with the electroweak symmetry breaking sector. The future circular collider FCC-ee is planned to produce top quark pairs at a center-of-momentum energy of 365 GeV and an integrated luminosity of 1.5 ab^{-1}. The isolation of $t\bar{t}$ events was achieved through the design of a neural network that differentiated the events from the dominant WW and HZ backgrounds using variables sensitive to the kinematic behaviour of top quark events. The resulting study examines the following New Physics scenarios: a scenario that considered a small modification to the SM $t\bar{t}$ production, and a scenario that considered the hypothetical existence of an additional W boson, called the W', that would modify $t\bar{t}$ production at lepton colliders.

Contents

Acknowledgments

References

Chapter 1

Introduction

The current landscape in the field of Particle Physics is a deeply puzzling one. The Standard Model(SM) has, in the last decades, proven remarkably successful in describing the various constituents of matter as well as the strong, the weak and electromagnetic interactions. It however fails to address some other aspects that have been observed in recent experiments and in the universe in general. This provides a fertile ground for a plethora of ideas to grow, attempting to enlarge our understanding by deeper, beyond the Standard Model theories ranging from Supersymmetry to Grand Unified Theory.

A example of such a growing idea is that of Lepton Universality Violation, which states that the leptons are not actually identical as the SM predicts. However, if this idea were to be correct, it could mean that various approaches considered in modern experiments were leptons were considered identical were wrong. This Master thesis centres around that problem. It considers a different approach to a lepton independent measurement of the semi-leptonic $t\bar{t}$ decay mode which in many experiments used the property were leptons were considered identical.

In the first chapter, a theoretical overview will be provided. A brief overview of the SM will be provided. Examples of its limitations and possible solutions thereof will be described. Additionally a short introduction into Lepton Universality Violation will be given as this forms the motivation behind this research project. This chapter will conclude with a short description of the hypothetical W' boson which is later used for study in this thesis.

The e^+-e^- collision considered for study in this research project were simulated considering them to be accelerated in the Future Circular Lepton Collider (FCC-ee). As such in chapter 3, a short overview is presented about the FCC. Additionally, the collision is expected to be detected by the conceptual detector, the ILD. As such a very brief description of this detector is also given.

In chapter 4, the framework Delphes will be introduced which was used to simulate the detector response of ILD detector. Here a short overview is provided, explaining the various steps Delphes takes to simulate a detector response.

The algorithm presented in this thesis, uses a machine learning method, a neural network, to optimise the separation of signal and background. As such chapter 5 provides an introduction to machine learning and the general theory behind neural networks.

The data analysis presented in chapter 6 represents every step taken in the optimisation of the algorithm. It starts with a description of the used signal samples and background samples. This is followed up by a description of the event selection. Here each variable used in the neural network will be introduced that was taken. In addition, direct cuts were also considered and will also be described here.
A section on the layout and performance of the considered neural network is also presented. The layout would provide the readers an idea of how the neural network is specifically set-up and the reasons behind its set-up. The performance then provides insight on how efficient it really is.
Finally this chapter concludes with a description of the results that are obtained after the neural network and direct cuts have been applied.

The thesis concludes in chapter 7 with an application of the algorithm when used for testing the sensitivity of the electron momentum in $t\bar{t}$ production. There are two applications considered in this thesis. The first application is a general sensitivity test of the SM generated semi-leptonic $t\bar{t}$ spectrum considering model-independent variations. The second sensitivity test considers together with the SM generated semi-leptonic $t\bar{t}$ sample, a sample that add the hypothetical W' boson.

Chapter 2

Theoretical Background

2.1 Overview of the Standard Model

At this point in time, matter and energy are often best understood in terms of kinematics of and the interactions between elementary particles. It is for this reason that physicists have combined the theories of special relativity, quantum mechanics and field theory to form what is known as quantum field theory. This theory allows for the study and prediction of said kinematics and interactions and lead to the construction of a set of fundamental laws and theories. Based on this a model was created by Weinberg et al.[61][36][57] which is commonly called the Standard Model (SM). It consists in a quantum field gauge theory based on the symmetry group $SU(3) \times SU(2) \times U(1)$ narrating the story of interacting particles. A visualization of the SM and the fundamental particles which it describes can be seen in figure 2.1[60].

In this section, a general overview of the SM will be given. The fundamental particles, the fermions and bosons, will be discussed and a additional focus on the importance of the top quark will be given as well. Note that we consider here elementary particles from the standard model. As such particles, such as mesons and hadrons and will not be discussed.

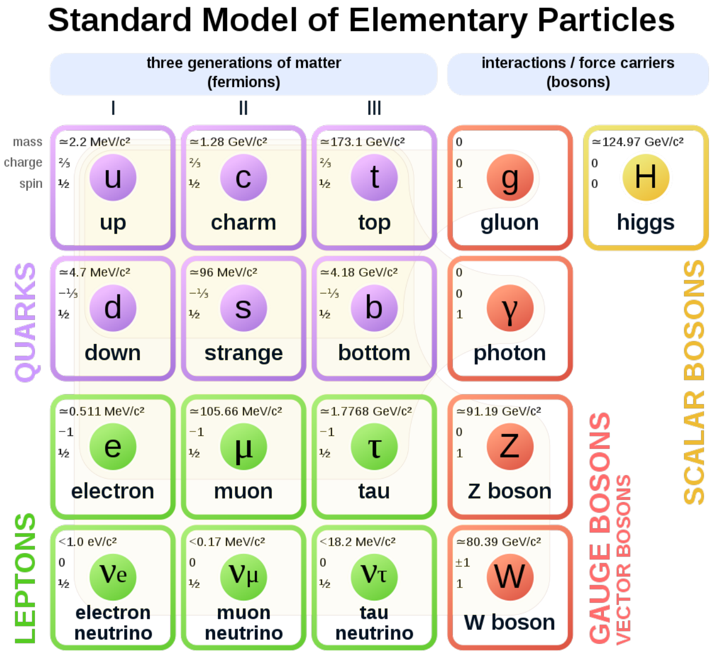


Figure 2.1: A visual overview of the standard model[60]

2.1.1 Forces and their Bosons

In the world there are four fundamental which according to the SM govern the interaction between fermions. These interactions are described by the exchange of a force-carrier particle known as bosons. These bosons are characterised by their integer quantum mechanical spin $(0, 1, 2, ...)$. The fundamental forces that are currently known are:

- The Electromagnetic Force: The electromagentic (EM) force was originally described in classical mechanics by the Maxwell equations. Later it was described by the quantum mechanics to be an interaction between charged particles through the exchange of the force carrier boson, the photon. It is a long-range force being the most dominant force over large distances at the microscopic scale.

- The Weak Nuclear Force: The weak force is probably most known for its role in radioactive decay. It carries out its interaction between particles through the exhange of its force-carrier bosons, the charged W boson and the neutral Z boson. As its name suggest it is quite a weak force and only acts small ranges of approximately $\sim 10^{-18}$

- The Strong Nuclear Force: As mentioned in previous section, quarks are held together to form hadrons. The force that holds these quarks together is the strong force, allowing interaction between these quarks through the exchange of the force-carrier boson, the gluon. However, besides the quarks it is not known to interact with any other particle. It is considered to be the strongest force in the universe, but it is limited to only a short range in which it allows interaction. This range is observed to be approximately $\sim 10^{-15}$.

- The Gravitational Force: Gravity is a force that everyone on the planet is influenced by and keeps humans from floating up to space. While it maybe the most well known force in the universe, it is also considered to be the weakest. It is observed to be about 10^{29} times weaker than the second weakest force the weak nuclear force. Aside from being the weakest, it is also possibly the least understood one. It is currently not described by the SM and has no observed corresponding boson. Models trying to expand the SM for this have called the hypothetical boson associated to gravity, the graviton. Contrast to gravity being the weakest, it is the most dominant force at the macropic scale and having infinite range.

These bosons are called gauge bosons and are often characterised by them carrying the forces of nature and by having a spin of 1. Aside of these gauge bosons, however, there is also the scalar boson, the Higgs boson. Unlike gauge bosons, scalar bosons have a spin of 0 and are currently not known to have a direct relation to the forces of nature. Discovered in 2012 by the ATLAS and CMS collaborations based on collisions in the LHC at CERN, the Higgs boson was a much desired extension to the SM due to its involvement in the masses of the gauge bosons and fermions. Before the introduction of the Higgs boson, the W bosons and the Z boson, could theoretically only be considered massless. However, emperical observations showed this to be definitely not the case, which led to theory behind the Higgs boson and its discovery later. The way this mass generation is described is that Higgs boson are generated within a field called the Brout-Englert-Higgs field [7] which would span the whole universe. It's through interactions with this field that gauge bosons and fermion would obtain their mass. The discovery of the Higgs boson was awarded with Nobel prize in 2013.

To conclude, it was said in the introduction to this section, that the SM is described in mathematical notation by the gauge symmetry group: $SU(3) \times SU(2) \times U(1)$. It is interesting to note that these groups represents three of its fundamental interactions given by the SM. The strong interaction is described by a Yang–Mills gauge theory with $SU(3)$ symmetry. The electroweak interaction, covering both the weak force and electromagnetic force, is described by a Yang–Mills gauge theory with the symmetry group $U(1) \times SU(2)_L$. Here the $U(1)$ group describes the interaction of the photon and the $SU(2)_L$ group describes the interactions with the W and Z bosons, where the L defines them as left-handed particles having their spin always aligned to their momentum. And finally, the Higgs boson is described as a complex scalar to the $SU(2)_L$ group. A sharp reader may note that the fundamental force of gravity is not mentioned among these, which is because of the fact the SM does not actually explain this fundamental force and does describe any interactions it could have with the SM particles. This is just one of the many limitations to SM and goes to show that it still is far from complete.

2.1.2 Leptons and Quarks

The world is made up out of matter, and matter itself is then made up out of particles. These elementary particles are called fermions. Fermions are characterised by their half odd integer quantum mechanical spin ($\pm\frac{1}{2}, \pm\frac{3}{2}$,...). They obey Fermi-Dirac statistics and additionally follow the Pauli exclusion principle, which says that no two fermions can occupy the exact same quantum at a given time. Fermions themselves are then again subdivided into leptons and quarks.

There are six types of leptons, known as flavours, grouped in three generations. Each generation consists out of a charged lepton, a neutral lepton called the neutrino and their corresponding anti-particles. An overview of the SM leptons are given in table 2.1. The three generation and their particles that are currently known are:

- **The electron:** The first generation of leptons considers probably the most well known particle in the world, the electron. The electron is most known as the particle responsible for electricity and as one the main constituents of atoms, alongside protons and neutron. The electric charge of an electron is $Q = -1.602176634 \times 10^{-19}$ C $= -1.q_e$ and a mass measured to be $m_e = 0.5109989$ MeV/c^2. Additionally in 1956 the electron-neutrino was discovered through experiments around β-decay performed by Cowan and Reines [34]. It is considered a neutral lepton due its lack of an electric charge. An interesting trait that is shown by neutrinos in general is that while the SM predicts them to be massless, they are not actually massless. Experiments performed by Fukuda et al. [35] showed that, although very small, neutrinos do exhibit a mass. This often considered as the very first signs of Beyond Standard Model (BSM) physics that were discovered.
  The anti-particles of these leptons are called the positron and electron anti-neutrino.
- **The muon:** The second generation of leptons considers the muon as its charged lepton. It was discovered through the study of cosmic rays and reported on in 1937[55]. It was observed to have the same properties as electron, but was measured to be almost 200 times heavier with a mass of $m_\mu = 105.6583745$ MeV/c^2. Few decades later in 1962, the second generation neutral lepton, the muon neutrino, was reported to be observed by Danby et al.[28] and its discovery earned to them the

Nobel prize of 1988.
The anti-particles of these leptons are called the anti-muon and muon anti-neutrino.

- **The tau lepton:** The third generation of leptons considers the charged lepton, the tau lepton. It was discovered through electron-positron collisions at the SLAC accelerator in the USA in 1975. As with the muon, it was observed to have the exact same properties as the electrons, but was measured to be more than 3000 times heavier with a mass of $m_\tau = 1776.86$ MeV/c^2. In addition, it is also the only lepton that has been observed to decay to either other leptons or even quarks. Its neutral lepton, the tau neutrino and its anti-particle, was discovered in 2000 in the DONUT experiment[41] at Fermilab in the USA.
  The anti-particles of these leptons are called the anti-tau lepton and tau anti-neutrino.

Leptons				
Generation	Particle	Charge[q$_e$]	Mass[MeV/c^2]	Lifetime[s]
1^{st} gen.	electron (e$^-$)	-1	$0.511 \pm 31 \times 10^{-9}$	stable
	e-neutrino (ν_e)	0	$< 10^{-9}$	unknown
2^{nd} gen.	muon (μ^-)	-1	$105.66 \pm 24 \times 10^{-6}$	$(2.197 \pm 22 \times 10^{-6}) \times 10^{-6}$
	μ-neutrino (ν_μ)	0	$< 10^{-9}$	unknown
3^{nd} gen.	tau (τ^-)	-1	1776.86 ± 0.12	$(290.3 \pm 0.5) \times 10^{-15}$
	τ-neutrino (ν_τ)	0	$< 10^{-9}$	unknown

Table 2.1: Summary of the leptons of the SM. Here their electric charge (in units of the elementary charge q$_e$), their of the currently best estimate of their mass and lifetime according to the Particle Data Group (PDG)[5] are given.

Similar to Leptons, there are six flavours of quarks also grouped in three generations. Quarks behave very different from leptons and are not observed to exist on their own. They are often[1] observed to form together creating other particles called hadrons. Hadrons can be divided into mesons, consisting out of a quark and anti-quark, examples of which are pion and kaons, or baryons, which are groups of three quarks, examples of which are the proton and neutron. An overview of these SM quarks are given in table 2.2. The three generations of quarks that are currently known are:

- **The up and down quark (first generation):** The first generation of quarks consider the up quark and the down quark, which are most known for being the constituents of the proton and the neutron, which make up the nucleus of an atom. They were discovered in 1969 at the SLAC accelerator in the USA during the deep

---

[1] top quarks are not observed to form into hadrons or mesons as will be discussed later on in section 2.1.3

inelastic scattering experiments[21][23] where electrons collided with protons and neutrons in an atom. Unlike leptons, quarks are not observed to have integer electric charges, instead an up quark is observed to have an electric charge of $Q = \frac{2}{3}q_e$ and the down quark is observed to have an electric charge of $Q = \frac{-1}{3}q_e$. Additionally the mass of the up quark is measured to be $m_u = 2.2$ MeV and for the down quark it is $m_d = 4.7$ MeV.

- **The charm and strange quark (second generation):** A second generation predicted to exist due to the existence of the kaon meson which was observed to be made up out of neither a up quark or a down quark. This quark, that made up the kaon[2], was called the strange quark, and was observed to have an electric charge of $Q = \frac{-1}{3}q_e$ and a mass of $m_s = 95$ MeV. Later in 1974, the second second generation quark was found due to the discovery of a charmed meson, the J/ψ meson [14]. This quark was observed to have an electric charge of $Q = \frac{2}{3}q_e$ and a mass of $m_c = 1.275$ GeV.

- **The top and bottom quark (third generation):** The third and final generation of quarks was first reported to be observed with the discovery of the Y meson in 1977 through observation of dimuon resonance at 9.5 GeV during 400 GeV proton-nucleus collisions.[37] The quark that made up this meson was called the bottom quark, and was observed to have an electric charge of $Q = \frac{-1}{3}q_e$ and a mass of $m_b = 4.18$ GeV. The final quark in the SM was discovered 1995 by the CDF and D∅ experiments at the Tevatron accelerator in Fermilab.[12] This quark was called the truth quark and was observed to have an electric charge of $Q = \frac{2}{3}q_e$ and a mass of approximately $m_t = 173.0$ GeV[5]. The top quark is considered to be major tool in testing the consistency of the SM, which will discussed further in depth in section 2.1.3.

---

[2] There are three possible kaon mesons depending on their composition. K^+: $u\bar{s}$, K^0: $s\bar{d}$ or $d\bar{s}$ and K^-: $s\bar{u}$

Quarks			
Generation	Particle	Charge[q$_e$]	Mass[MeV/c^2]
1^{st} gen.	up (u)	$\frac{2}{3}$	$2.2^{+0.5}_{-0.4}$
	down (d)	$\frac{-1}{3}$	$4.7^{+0.5}_{-0.3}$
2^{nd} gen.	charm (c)	$\frac{2}{3}$	1275^{+25}_{-35}
	strange (s)	$\frac{-1}{3}$	95^{+9}_{-3}
3^{nd} gen.	top (t)	$\frac{2}{3}$	173.0 ± 0.4
	bottom (b)	$\frac{-1}{3}$	4180^{+40}_{-30}

Table 2.2: Summary of the quarks of the SM. Here their electric charge (in units of the elementary charge q$_e$) and their of the currently best estimate of their mass according to the PDG[5] are given.

An interesting to note, is that quarks can transform into a quark of a different flavor through the weak interaction. This is called quark mixing. In order to describe this, quarks are generally expressed by two eigenstate bases. One which describes their state during the free propagation of the quark, the mass eigenstate, and the other describes the form in which they interact through the weak interaction, the weak eigenstate. In principle quark mixing occurs when transforming from the weak eigenstates to the mass eigenstates which are related by:

$$\begin{bmatrix} d_W \\ s_W \\ b_W \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d_m \\ s_m \\ b_m \end{bmatrix} = V_{CKM} \begin{bmatrix} d_m \\ s_m \\ b_m \end{bmatrix} \tag{2.1.1}$$

where V_{CKM} is the unitary Cabibbo–Kobayashi–Maskawa matrix and the square value of its matrix elements quantifies the probabilities for one flavor of quark to transform into another one through weak interaction. Currently the experimentally independent measured values of the matrix elements are given by[5]:

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97420 \pm 0.00021 & 0.2243 \pm 0.0005 & 0.00394 \pm 0.00036 \\ 0.218 \pm 0.004 & 0.997 \pm 0.017 & 0.0422 \pm 0.0008 \\ 0.0081 \pm 0.0005 & 0.0394 \pm 0.0023 & 1.019 \pm 0.025 \end{bmatrix} \tag{2.1.2}$$

2.1.3 The Top Quark

With its discovery in 1995 in the Tevatron Collider [12], the top quark remains one of the most important tools for testing the consistency of the SM and the study into New Physics. The reason for its importance as a tool can be related to three of its characteristics: its mass, life time and its decay behaviour, which will be shortly discussed in the following.

The width of a top quark is measured to be $1.41^{+0.19}_{-0.15}$ GeV[5], leading to a lifetime of 0.46×10^{-24} seconds. The reason why this value is of importance is related to its lack of hadronization.
The formation of baryons (protons, neutrons, ...) and mesons (pions, kaons, ...) happens through the process of hadronization where quarks will interact through the strong interaction and will be held together forming said particles. The characteristic time for the strong interaction to occur and thus hadronization as well is about 10^{-23} s. This can be seen from a little thought experiment. Suppose two particles are undergoing a strong interaction. The force-carrier for this interaction is the gluon which has to travel between particles in order to make this interaction happen. Since the gluon is considered to be massless it is expected to move at the speed of light, i.e. $\sim 10^8$ m/s. Taking into account that strong interactions occur at a distance range of about 10^{-15} m. It can be said that the time it takes for the strong interaction to occur is about 10^{-23} s. This means that hadronization only occurs for particles with a longer lifetime then 10^{-23}.
As seen above, this is not the case for the top quark, which allows it decay on its own. In fact, it is currently the only known quark which is able to decay on its own, as other quarks have hadronized by that point. This characteristic is one of the strongest features of a top quark as it enables it to transfer some of its properties, such as its direction of spin, directly to its decay products. Investigating its constituents then would allow the study of properties directly related to the top quark.

Aside from the fact that the top quark is currently the only known quark to decay on its own, it's also been found to decay exclusively to a negative charged quark and a W boson. Of these decays it's has been observed that the top quark decays to a bottom quark and a W boson $95.7 \pm 3.4\%$[5] of the times which is further supported by the CKM matrix as seen in equation 2.1.2 which shows that V_{tb} is close to unity, meaning that the top quark has a high probability of decaying into a bottom quark. This exclusive decay behaviour not only allows for a more simple reconstruction of said quark, but also led to the first precise measurement of the W boson mass in 2018 at the LHC [6], which is a force carrier of the weak interaction and one of the heaviest known particles in the SM.

The mass of the top quark as seen in table 2.2 is approximately 173.0 ± 0.4 GeV[5], about 40 times heavier than the bottom quark and the heaviest particle in the SM. Due to this large mass its coupling to the Higgs boson is the strongest of all SM particles. Therefore it is the best candidate in studying the Higgs boson particle.

Additionally, various models trying to predict NP often describe these phenomenon through the existence of new particle. Some these particles consider enormous masses of which some are predicted to be TeV scale. To prove or disprove these models some form of probe is needed that is sensitive to these possible NP. The top quark with its large mass is expected to be a lot more sensitive to large masses as it there is a good possibility for these models to decay into such a quark, making it an important probe in these areas. It is because of these possible applications, that precise measurements of the top quark mass & other properties have been performed and at this point have reached a precision of below 500 MeV.

Top quarks are mainly be produced[3] either singly or in quark-antiquark pairs. Top quarks that are produced singly do this through charged electroweak current processes (using W boson force carriers) and can be visualised through the feynmann diagrams in figure 2.2. The first type that is shown in figure 2.2a, is called the s channel production mode. This is the least common mode of production as it requires the intermediate W boson to decay into top quark which is more than twice its mass. The next type, seen in figure 2.2b, is called the t channel and is the most common type of top quark production. Finally, it is also possible for the top quark the be produced together with a W boson, as can be seen in figure 2.2c. These type of processes are denoted as the tW channel.

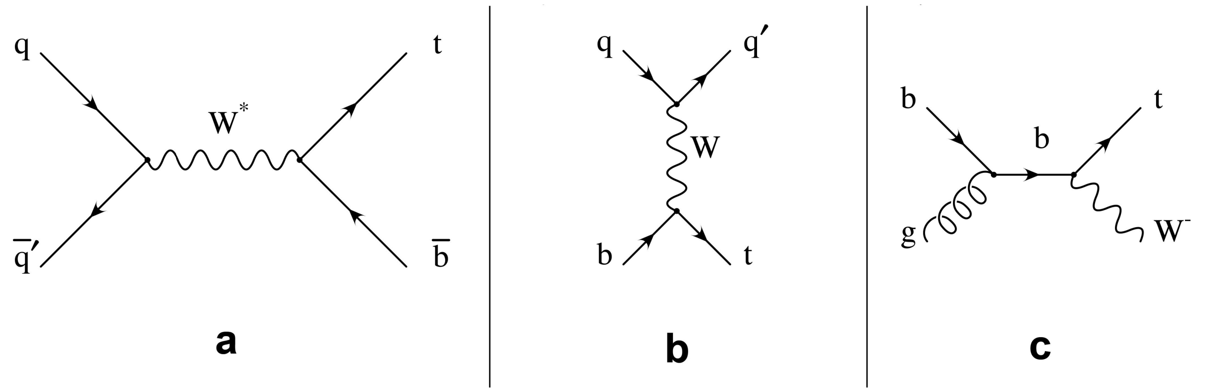


Figure 2.2: The feynman diagrams depicting singly top production. Note that the shown diagrams are limited to leading order.

Top at hadron colliders

Pair production is different for hadron collider and lepton collider. In hadron colliders, such as the LHC, top pair production processes occur through strong interaction and are visualised by the feynman diagrams seen in figure 2.3. These processes can be further sub-

---

[3] It should be mentioned that ATLAS recently observed evidence for $t\bar{t}t\bar{t}$ production[4]. However, it will not covered here.

divided into two types of strong interaction processes: gluon fusion and quark-antiquark annihilation. Gluon fusion, see figure 2.3a, occurs when two protons collide and in each one there is a gluon that has a lot of energy. It is these two energetic gluons that collide with each other that causes gluon fusion to occur. The second type, quark-antiquark annihilation as seen in figure 2.3b, takes place when a proton and anti-proton collide. Since protons are made up out of two up quarks and one down quark, an anti-proton is made up out of two anti-up quarks and one anti-down quark. If these up or down quarks where to collide with each other, a large amount would be released resulting in the possible creation of a top and anti-top quark pair.

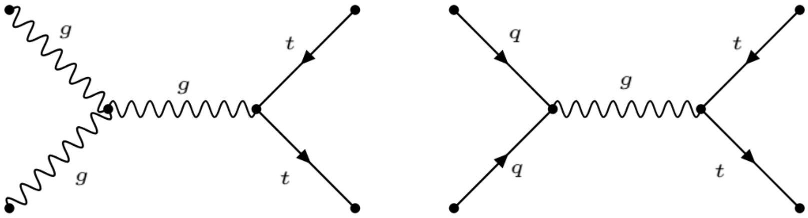


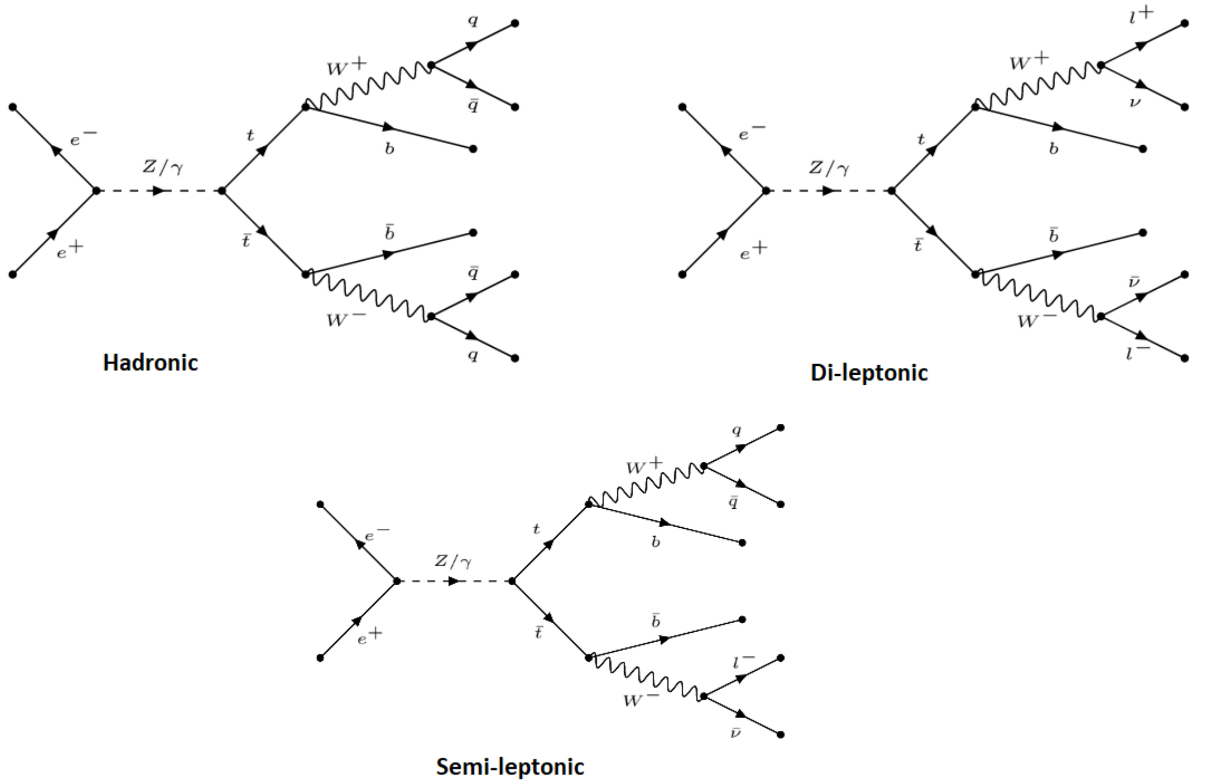
Figure 2.3: The feynman diagrams depicting the top pair production via strong interaction at hadron colliders. Note that the shown diagrams are limited to leading order.

Top at lepton colliders

In lepton colliders, the top and anti-top quark pair gets produced through electron-positrion collision. From these collisions, top pair production occurs through either a neutal charged electroweak interactions (using Z boson force carriers) or through electromagnetic interactions (using photons as force carriers). These are visualised in figure 2.4. When produced, this pair will further decay into two W bosons, bottom quark and an anti-bottom quark.

The decay of the top quark

When the top quark decays it has a probability of approximately 95.7 ± 3.4 to decay into b quark and a W boson. The remaining branches represent different quarks. Considering that the W boson also decays, there are three different final states that can be observed, depending on how the W boson decays, as seen in figure 2.4 and . First there is the di-leptonic final state where both W bosons decay into a lepton and a neutrino. The probability for a W boson to decay into a lepton and neutrino is approximately $10.86 \pm 0.09\%$. The probability where both W bosons then decay into a lepton and neutrino pair is about $\frac{1}{9} \approx 11.1\%$. Second there is the hadronic final state in which both W bosons decay into a quark and anti-quark pair. The W boson has a probability of 67.41±0.27%[5] to decay into a quark and anti-quark pair. Due to this the probability of a hadronic final state is approximately 45.44%. Finally, the last final state, the semi-leptonic final state, is where one W boson decay into a lepton and neutrino pair and where the other decays into a quark and anti-quark pair. By combining the chances related each possible decay a similar probability as for the hadronic final state is found and is approximately 43.92%. It is this final state that is the state of interest in this paper.



Hadronic

Di-leptonic

Semi-leptonic

Figure 2.4: The feynman diagrams depicting the top pair decay via electroweak interactions. Note that the shown diagrams are of leading order. These feynman diagrams were generated using madgraph.

2.2 Limitations of the Standard Model

Despite its tremendous success in describing all present measurements, the SM can only be regarded as the low-energy, effective, incarnation of a more global theory. It has at present still a lot of limitations that have been experimentally observed and of which the SM does not offer an explanation. Some of these limitations are:

- Dark matter and dark energy: According to astrophysical observation, there exist gravitational effects in the universe which do not follow currently accepted theory of gravity that considers current visible matter, baryonic matter. It is estimated that baryonic matter, only makes up about 5% of the universe. Because of this it is surmised that there exist another form of invisible[4] matter, called dark matter. Dark Matter is currently expected to make up approximately 27% of the universe. The remaining 68% is considered as Dark Energy, and while it is not understood at all, it is expected to be the cause of the accelerated expansion of the universe.

- Matter-antimatter asymmetry: Every physical object in our daily lives is known to be made up out of matter. The symmetry of the SM, however, would imply that there should be a same amount anti-matter in the universe. But when matter and anti-matter come close to each other they would annihilate one another. How then is it possible that there is such an abundance of matter observable when looking outside? This assymetry of matter and anti-matter can only be explained if there was some imbalance of matter in the early universe. The only way the SM would allow such an asymmetry is if there was a violation of the CP-symmetry. The CP-symmetry considers the charge–conjugation symmetry[5], C-symmetry, and parity symmetry[6], P-symmetry, to hold when applied one after another. This violation has already been observed for weak interactions.

- Gravity force-carrier has not yet been observed: As seen in section 2.1.1, the fundamental force of gravity considers a force carrier called the graviton. This particle, however, has not yet been experimentally verified and the SM does not predict any of its possible interactions with the other SM particles. It is considered to be a gaping hole in the SM and is the main focus of many models extending the SM.

- Neutrinos are not fully described: Aside from there experimental observations, neutrinos are very elusive particles. Initially they were considered to be massless particles, but in 1968 a deficit of neutrinos originating from the sun was observed. This deficit, called to solar neutrino problem, was later discovered to be due to neutrino oscillations which allows neutrinos to change their neutrino flavor as they travel, this is called neutrino mixing. This phenomenon was only possible if neutrinos considered a non-zero mass. According to the SM this then required them to have not only left-handed neutrinos, which were already observed, but also right-handed neutrinos, which to this point have never been observed.

- Hierarchy problem (Higgs mass): One of the strange things about the Higgs boson, is that while it is responsible for the mass of the SM particles, it has been found to be lighter than the top quark and many theorised hypothetical particles. This

---

[4] in other words, invisible to the electromagnetic spectrum

[5] The symmetry of transformations between particles and their anti-particles

[6] The symmetry associated with the inverse transformation of spatial coordinates

problem has been referred to as the Hierarchy problem. It is expected to be due to perturbative corrections which fine-tune the Higgs mass. These corrections are predicted to be at the scale at which new physics are expected to occur. A typical example of such a scale is the Planck scale which is approximately $\sim 10^{19}$ GeV.

2.3 Beyond the Standard Model

The various limitations to the SM represents suggests that there is still a deeper theory that would explain such inconsistencies. In this section, three of the most popular models that would explain these inconsistencies and extend the SM accordingly are given. These models are:

- Supersymmetry (SUSY)[47]: Probably the most well-known model that suggests NP is the model of supersymmetry. In simple terms, this model associates with each particle of the SM another group of particles with a one-half spin difference. This means that the supersymmetry associates to the fermions, which have a half-integer spins, another group of particles, with integer spins, called the sfermion. And for the gauge bosons, a group of half-integer spin particles called the gauginos would be associated to them.
  If any of these particles were to be observed, the SUSY model could provide many answers to current limitations to the SM, like for example dark matter as some SUSY particles are considered to be candidates for it.

- Grand Unified Theory (GUT): The Grand Unified Theory model hypothesizes that, in the early universe, the three gauge interactions of the SM, the electromagnetic, weak and strong interactions, were one unified force. Currently, experiments have confirmed that at high energies, the electromagnetic and weak interactions unify to what is known as the electroweak interaction.
  Some models based on the GUT model have been made that would explain the SM limitations to neutrino masses. An example of such a model is the $SO(10)$ model, where it is implied that the SM gauge symmetry group $SU(3) \times SU(2) \times U(1)$ was initially the gauge symmetry group $SO(10)$ in the early universe, and during the expansion broke into several symmetry groups only part of which are now the current SM gauge symmetry group. This model would then accomodate for the neutrino masses.

- The String Theory: A last but also popular model, is the String theory model. In comparison to other models, it takes a rather controversial approach. Currently, particles are considered to be point-like entities each occupying a specific point in space. String theory, on the other hand describes particles as strings instead of points. Many different elementary particles would then be described as certain vibrational modes of a string. It is then expected that interaction of particles are just strings interacting with each altering their vibrational modes. However, in order for string theory to be mathematically justified it requires at least 10 dimensions. This provided to be difficult to explain in our 4 dimensional universe. But if the SM is limited, maybe the assumption that the universe has only 4 dimension might be wrong as well.

One of the most promising explanations on the limitations of the SM it provides, is the explanation of the graviton. In string theory, it is suggested that like all other SM particles, that the graviton is just another vibrational mode of the string. It is then suggested to interact with the other particles on a higher dimension which is not currently visible.

2.4 Lepton Universality Violation

2.4.1 What is Lepton Universality

The Standard model of particle physics is expected to describe and predict the behaviour of all the known particles and forces. The SM further organises the twelve elementary fermions (and their antiparticles) known to date into three different families. According to the SM, the families in the leptonic part can be considered completely identical to one another except for the different strengths of their interactions with the Higgs field. This results in different masses for each particle.

This pattern is called “lepton universality” and has been over the last couple of decades been tested to sub-percent precision in electroweak interactions, and in transitions between light quarks. Examples of such measurements were made at the Large Electron–Positron (LEP) collider at CERN in decays of W and Z bosons and by the PIENU and NA62 fixed-target experiments in decays of pions and kaons. The results can be seen and compared to the SM in table 2.3. When LU is tested for the third family, very precise results were obtained for the partial widths of the Z boson. The results of the partial wirdths of the W boson, on the other hand, were considered significantly less accurate and caused tension with the SM prediction at the level of 2.6σ. Sadly, this is too much of an ambiguous sign to be considered the result of NP. However, LU has never been established to such a degree of precision in decays of heavy quarks. And this is where more regular signs of violation started to appear.

Ratio of partial widths	Experimental value	SM prediction
$\frac{BR(Z\to\mu^+\mu-)}{BR(Z\to e^+e^-)}$	1.0009 ± 0.0028	1
$\frac{BR(Z\to\tau^+\tau-)}{BR(Z\to e^+e^-)}$	1.0019 ± 0.0032	1
$\frac{BR(W\to e^-\bar{\nu_e})}{BR(W\to\mu^-\bar{\nu_\mu})}$	0.997 ± 0.010	1
$\frac{BR(W\to\tau^-\bar{\nu_\tau})}{BR(W\to e^-\bar{\nu_e})}$	1.063 ± 0.027	1
$\frac{BR(W\to\tau^-\bar{\nu_\tau})}{BR(W\to\mu^-\bar{\nu_\mu})}$	1.070 ± 0.026	1
$\frac{BR(\pi\to e\nu(\gamma))}{BR(\pi\to\mu\nu(\gamma))}$	$(1.230 \pm 0.004)\times10^{-4}$	$(1.2352 \pm 0.0002)\times10^{-4}$
$\frac{BR(K^\pm\to e^\pm\nu)}{BR(K^\pm\to\mu^\pm\nu)}$	$(2.488 \pm 0.009)\times10^{-5}$	$(2.477 \pm 0.001)\times10^{-5}$

Table 2.3: Summary of the test of LU performed by studying the decay of W and Z bosons, pions and Kaons. [18]

2.4.2 Signs of Violation

As mentioned in previous section, it was during the study of the decays of heavy quarks that the tension of the SM predictions and the experimental results. These results came in two different classes of decays, both of which considered the decay of a B meson, which is a hadron made up out of two quarks, one of which is an anti-bottom quark.

The first class of decay considers the decay of a bottom quark into a charm quark. In precision tests of the LU, semi-leptonic $b \to cl^-\bar{\nu_l}$ where l^- represents a lepton and $\bar{\nu_l}$ the corresponding neutrino, which for B mesons correspond to the decays: $B \to Dl^-\bar{\nu_l}$ and $B \to D^{(*)}l^-\bar{\nu_l}$ where the $D^{(*)}$ represents a D meson in an excited state, are considered. This decay class is visualised as a Feynman diagram as seen in figure 2.5. From this figure it can be seen that this type of decay is mediated by a W boson. Due to the previous tension caused by the decay of a W boson into a τ and because of massive nature of the τ which makes it more sensitive to NP effects[7], it was considered more interesting to perform measurements where the B meson decays to a τ lepton, referred to as semi-tauonic B decays. The quantity that is experimentally measured is the ratio of branching fractions R_D given by:

$$R_D = \frac{BR(B \to D\tau^-\bar{\nu_\tau})}{BR(B \to Dl'^-\bar{\nu_{l'}})}, \quad R_{D^{(*)}} = \frac{BR(B \to D^{(*)}\tau^-\bar{\nu_\tau})}{BR(B \to D^{(*)}l'^-\bar{\nu_{l'}})} \tag{2.4.1}$$

where l' represents an electron or a muon.

---

[7] NP particles are expected to be rather heavy due to the difficulty of measuring them

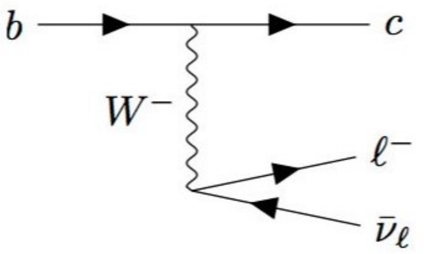


Figure 2.5: The feynman diagrams depicting the decay of b quark into a c quark through weak interaction.

These type of tests where performed by the BaBar experiment in the SLAC National Accelerator Laboratory at the Stanford University in the USA, by the Belle experiment at the High Energy Accelerator Research Organisation (KEK) in Japan, and by the Large Hadron Collider beauty (LHCb) experiment at the LHC in CERN. Results from these experiment can be found in table 2.4 together with SM predictions and are plotted in figure 2.6. Aside from the BaBar experiment results, while the measurements where consistently above the SM predictions, none of these measurements showed a result which caused tension with the SM prediction at a 3σ level. However, the combined world average of R_D and $R_{D^{(*)}}$ measurements gave a result which is in tension with the SM prediction at a 3.1σ level. This tension provides ample reason to further look into whether or not the LU is violated in semi-tauonic decays of B mesons.

Experiment	$R(D^{(*)})$	$R(D)$
BaBar (2012)[44]	0.332 ± 0.03	0.440 ± 0.072
Belle (2015)[39]	0.293 ± 0.041	0.375 ± 0.069
LHCb (2015)[8]	0.336 ± 0.04	/
Belle (2017)[38]	$0.270^{+0.044}_{-0.043}$	/
LHCb (2018)[9]	0.280 ± 0.034	/
Belle (2019)[33]	0.283 ± 0.023	0.307 ± 0.04
Average	0.298 ± 0.013	0.349 ± 0.031
SM	0.258 ± 0.005	0.299 ± 0.003

Table 2.4: Summary of the $R(D^{(*)})$ and $R(D)$ measurements performed by the Belle, BaBar and LHCb experiments. [3]

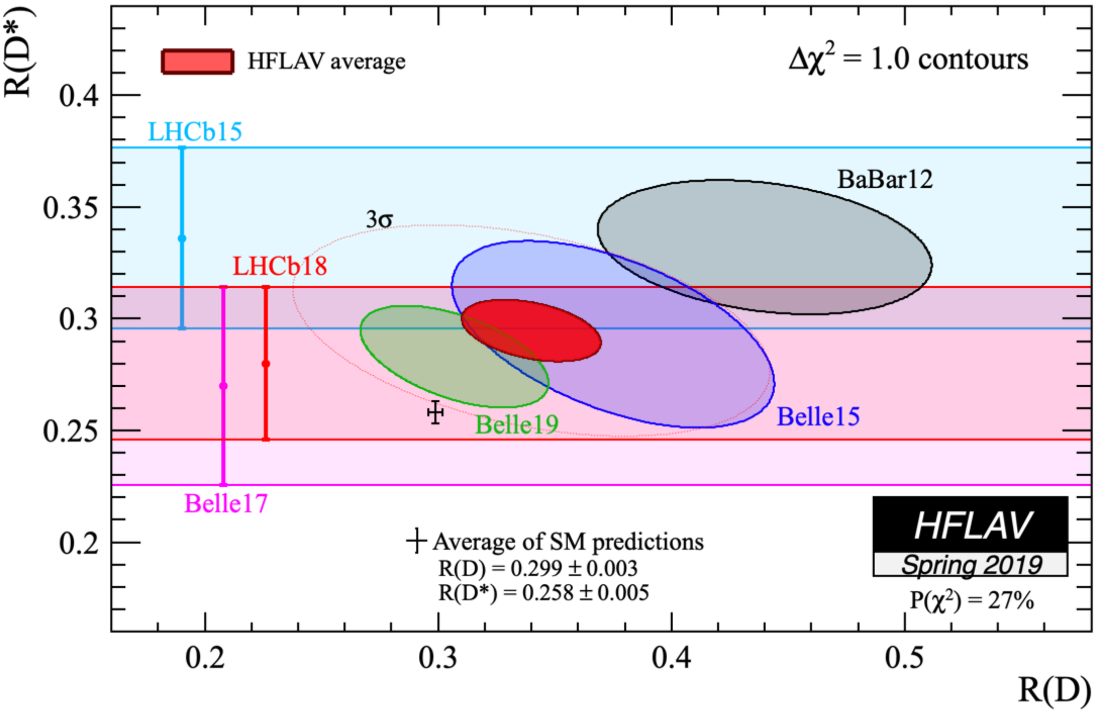


Figure 2.6: Average of all measurements of R_D and $R_{D^{(*)}}$ provided by the HFLAV group. The red ellipse shows the combined average and the data point is the SM prediction. [3]

The second class considers the decay of the bottom quark into the strange quark. Precision tests of the LU involving this class consider a loop-level transition of the type $b \to sl^+l^-$ where l represents the lepton, which for B mesons correspond to the decay: $B' \to Kl^+l^-$ and $B' \to K^{(*)}l^+l^-$ where B' can be either the B^0 and B^+ mesons and $K^{(*)}$ is an excited state of the K meson. This decay class is visualised as a Feynman diagram as seen in figure 2.7. From this figure it can be seen that this type of decay is mediated by a Z boson and loop where a virtual top quark is created through interaction with W boson in a loop. Due to the interaction being mediated by a Z boson which changes the type of quark, this interaction is called a Flavor-Changin Neutral Current (FCNC). FCNCs are forbidden in the SM at tree level and arise only at one loop and because of this they are highly suppressed. This makes them an ideal probe for NP as they have an enhanced sensitivity to them due to this suppression. As of this point, only the electron and muon have been observed in the decay of B meson to K meson. As such the quantity that is experimentally measured is the ratio of branching fractions R_K given by:

$$R_K = \frac{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma BR(B'\to K\mu^+\mu^-)}{dq^2}}{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma BR(B'\to Ke^+e^-)}{dq^2}}, \quad R_{K^{(*)}} = \frac{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma BR(B'\to K^{(*)}\mu^+\mu^-)}{dq^2}}{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma BR(B'\to K^{(*)}e^+e^-)}{dq^2}} \tag{2.4.2}$$

where q^2 is the invariant mass squared of the di-lepton system integrated between q^2_{min}

and q^2_{max}.

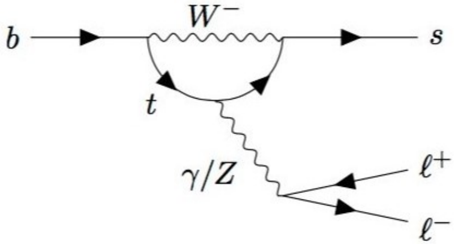


Figure 2.7: The feynman diagrams depicting the decay of b quark into a c quark through the electroweak interaction.

These types of experiments where again performed by the BaBar, Belle and LHCb experiments. Results that came out of these experiment can be found in table 2.5 and their ratio to the SM prediction is visualised in figure 2.8. From these it can be seen that the most recent measured values which are performed by the LHCb are consistently below the SM predicted value, decreasing even further as new experiments are performed. This could indicate that something might be interfering with the muonic decay amplitude. However, none of these results provided a strong enough tension with the SM prediction, the largest being of the level of 2.5σ. Nevertheless, these class of decays is still very difficult to study with many external factors with interfering with it. Hence this class of decays still warrants a reason to study even further.

Experiment	Kaon type	q^2 value	$R_{K,exp}/R_{K,SM}$
Belle (2009)	K		1.03 ± 0.20
Belle (2009)	$K^{(*)}$		0.83 ± 0.19
BaBar (2012)	K	$0.10 - 8.12$	$0.74^{+0.4}_{-0.32}$
BaBar (2012)	K	> 10.11	$1.43^{+0.66}_{-0.46}$
BaBar (2012)	$K^{(*)}$	$0.10 - 8.12$	$1.06^{+0.49}_{-0.34}$
BaBar (2012)	$K^{(*)}$	> 10.11	$1.18^{+0.56}_{-0.39}$
LHCb (2014)	K^+	$1.0 - 6.0$	$0.745^{+0.097}_{-0.082}$
LHCb (2017)	$K^{0(*)}$	$0.045 - 1.1$	$0.66^{+0.12}_{-0.06}$
LHCb (2017)	$K^{0(*)}$	$1.1 - 6.0$	$0.69^{+0.12}_{-0.09}$

Table 2.5: Summary of the R_K measurements performed by the Belle, BaBar and LHCb experiments.[19]

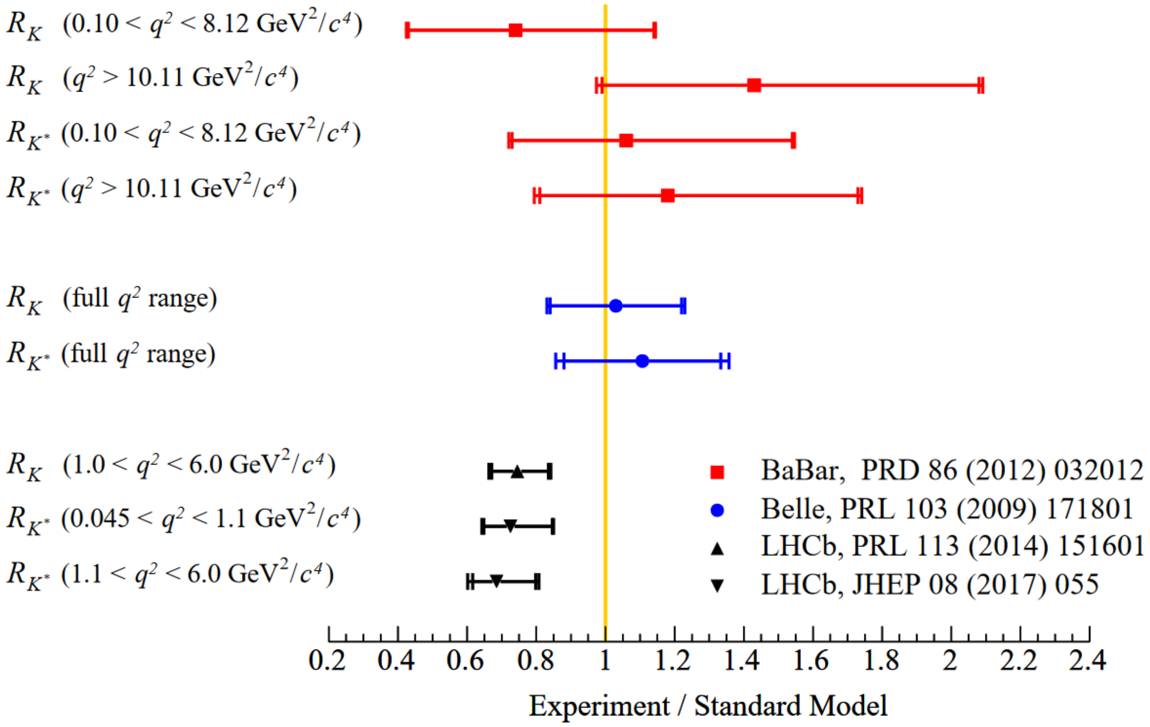


Figure 2.8: Average of all measurements of R_D and $R_{D^{(*)}}$ provided by the HFLAV group. The red ellipse shows the combined average and the data point is the SM prediction. [19]

2.4.3 Implications for the Standard Model

As of now, the signs that show violation are too ambiguous to make conclusive decisions about the whether or not the SM is wrong. However, if future studies would provide ample proof of such a violation than it would mean that New Physics are at play and that the SM has extended with those in mind. Currently, a few theoretical models that introduce new particles, have been made that would explain this violation. Amongst these examples the most prominent are: the existence of a leptoquark, which are particles that carry both lepton and quark quantum numbers and the existence of the hypothetical Z' boson, which is implied to be a Z boson with mass of a few TeV.

However, aside of the discovery of NP, a violation of lepton universality would bring into question the validity of previous measured results like the decay of the W and Z boson. These measurements showed a strong acceptance of the lepton universality and violation of said property would mean in the best scenario, that there were NP interaction happening in the background which were previously not detected. And in worst scenario, this could mean that approach that was used for measurement was wrong. It is the fear of the latter and the clear possibility of such violation of occuring that is the main motivation behind this paper. As the study in this paper is to use a lepton independent approach of detecting semi-leptonic $t\bar{t}$ decays as violation of LU could bring the leptonic approach into question.

2.5 The Hypothetical W' boson

Through the research in new physics that go beyond the Standard Model, the possibility of other particles that would explain the NP to the SM is often a recurring subject. One of such hypothesized new particles is the W' boson. A boson that, model independently, is completely the same as the W boson, with a spin of 1 and charge of ±1. It would decay even to the exact same particles. It's only difference, similar to the leptons, is that it expected to be massive. Various theories and experiments predict this boson's mass to be around the TeV scale. However, it has not been concluded yet that it is not possible for it to have a value around the GeV scale which is what in this study is considered.
Popular models that predict and explain the W' boson are for example: model based on the "left-right" symmetry theory, where unlike the W boson, which is described by the left-handed symmetry group $SU(2)_L$, the W' boson is described by the right-handed symmetry group $SU(2)_R$. In this theory, the W' boson is predicted to be of a right-handed nature, having his spin inversely aligned with its momentum. It is then expected to couple to hypothetical right-handed fermions and left-handed anti-fermions, completely opposite to the W boson.
Another popular type of model are those which are based on the theory of extra dimensions, like String theory. In these types of models, the W' boson is expected to differentiate in its interactions withing an extra dimension. Imagining an extra dimension is rather difficult with a humans 3d vision, but a simple thought might provide an idea of what is meant by it. The thought experiment requires you to imagine yourself on a tightrope. An normal human being would see this as an one-dimensional path. Now imagine a tiny ant on this tightrope. To the ant this is not only an one-dimensional path but it can also move rotationally along the wire. It is this additional movement that can be compared to the interactions within an extra dimension. It is only apparent on apparent on a microscopic scale and humans are still unable to perceive it. In this extra dimension, particles would interact with the fundamental force of gravity, which the SM fails to describe. It is there were the W' boson and W boson would differ from each other majorly.

Since its prediction, the ATLAS en CMS experiments have performed many simulations on the possible mass of the W' boson and its corresponding cross section through its decay from proton-proton collision. They have simulated these predictions up to a few TeV. An example of the results that were found through these simulations as performed by CMS can be seen in figure 2.9. Simulations later based on electron-positron collision at LEP-II[30] showed that the mass of the W' boson would had to have a lower limit of

$M_W > \frac{\sqrt{s}}{2} \approx 105$ GeV. However, due to large interest in the mass of the W' boson being around the TeV scale, experiment limit the lower limit mass of their simulated W' boson to be between 200 and 500 GeV[30]. As such not much research is done on low mass ranges.

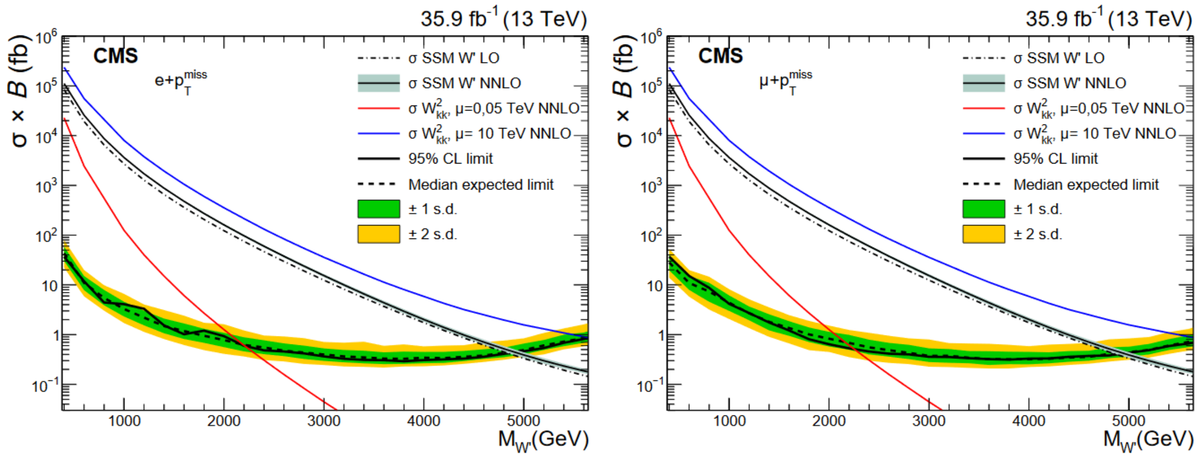


Figure 2.9: Expected (dashed line) and observed (solid line) 95%CL limits in the SSM interpretation for the electron (left) and muon (right) channels. The shaded bands represent the one and two standard deviation (s.d.) uncertainty bands. Also shown are theoretical cross sections for the Super Symmetry benchmark model (black with a grey band for the PDF uncertainties) and split-UED (red and blue solid lines) interpretations.[58]

Chapter 3

The e^+e^- collision at the FCC-ee with the ILD

3.1 The Future Accelerator Complex

CERN, the European Organization for Nuclear Research, is the world's leading facility in state-of-the-art fundamental research in particle physics. Founded in 1954 near Geneva, the facility has grown over the last decades, hosting numerous different particle physics experiments covering the various subjects within the field.
The most well known of these experiments is probably those using the worlds most powerfull particle accelerator, the Large Hadron Collider (LHC)[45], located deep underground. It lies in tunnel of about 27 km circumference in which protons are accelerated to cause proton-proton collisions at a center-of-mass energy of up to 13 TeV. Through many dedicated experiments performed at LHC, much progress was made in the study of SM processes and exploration of new physics. Such progress includes feats such as the discovery of Higgs boson, the study into dark matter, etc..
However, as various questions got answered, new questions arrived which proved well beyond the limits of even the LHC. The FCC-project, was proposed as an answer to these new questions. It aims to improve upon LHC, by pushing its potential of discovery through enhanced sensitivity, allowing the detection of phenomena too elusive for the LHC, and by increasing its mass range through higher precision and larger center-of-mass energy, allowing the study of massive particles well beyond the LHC. Illustrated in figure 3.1, are the current conceptual designs of the FCC accelerator. Lying in a tunnel of about 97.75 km, the accelator is planned, unlike the LHC, to not function as a hadron collider (FCC-hh), but also as a lepton collider (FCC-ee). In addition, the complementary design allows for a this type of collider which considers electron-proton collision (FCC-eh).

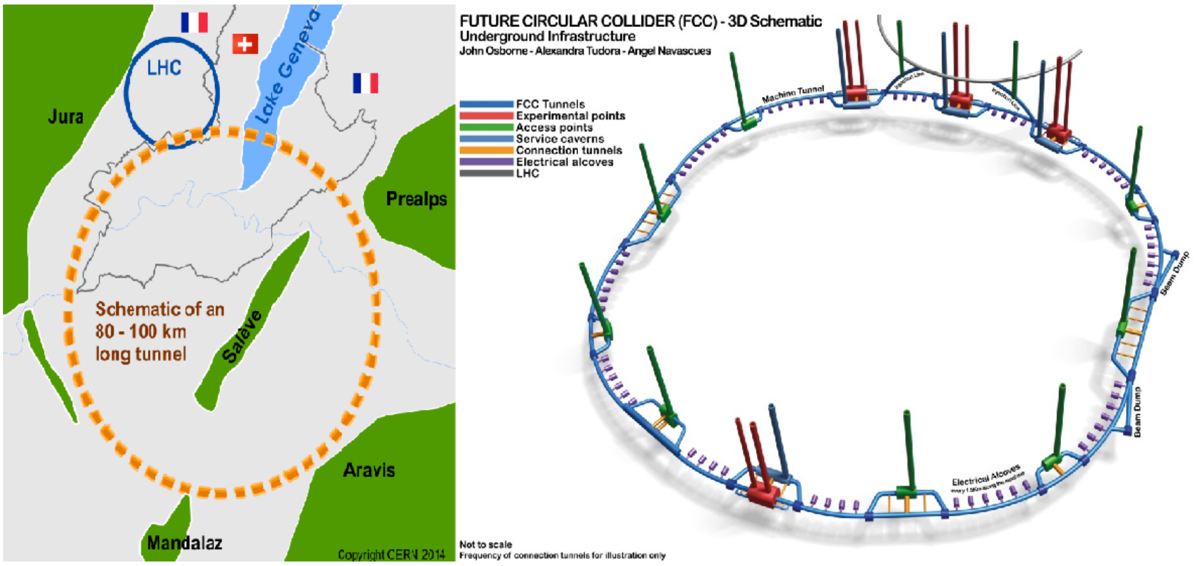


Figure 3.1: An shematic layout of the current expected design of FCC-ee complex where the SPS serves as the used PBR.[10]

3.1.1 The FCC collision experiments

The particles that flow in opposite direction within the accelerator are made to collide at two specific collision points. Each of these points are equipped with detectors that are able to detect the resulting particles that are created from such collisions. Through the detection of these resulting particles and their underlying physics, the detectors are capable of reconstructing the interesting physics phenomena that occurred during the collision.

Following this principle, the FCC is designed to be able to host three different type of collisions in each of these collision points:

- Electron-Positron collisions (FCC-ee)[10]: The FCC-ee is designed to provide electron-positron collisions at center-of-mass energies between 88 and 365 GeV and at high integrated luminosity.
  Currently, its expected to run for about 15 years after which it will be replaced by the FCC-hh. Its main proposed studies of interest consider the study of the properties of the Z boson (at 91 GeV) for 4 years, the W boson (at 161 GeV) for 2 years, the Higgs boson (at 240 GeV) for 3 years and the top quark (at 340-365 GeV) for the last 5 years. The latter of which is the research focus of this paper and thus this will be discussed next section.

- Hadron-Hadron collisions (FCC-hh)[11]: The FCC-hh its main feature is to provide proton-proton collisions at a center-of-mass energy up to 100 TeV and at an integrated luminosity of approximately 20 ab^{-1} in each of the collision points. This immense upgrade from the LHC's and HL-LHC's center-of-mass energy of 14 TeV, will allow for the study of completely uncharted territory. Currently the main research goals of the FCC-hh consider for example: very precise study of the Higgs potential, the study of dark matter of which LHC could only study a fraction due to its small production rates and direct searches of massive particles which were well beyond the reach of the LHC.

- Electron/Positron-Hadron collisions (FCC-eh)[11]: Additional to the FCC-ee and FCC-hh, the collider is designed for an optional collision between electrons of 60 GeV (obtained from an energy recovery linac) and the protons of 50 TeV form the FCC-hh. This results in a center-of-mass energy of 3.5 TeV.
  Due to the high energy environment in which these these collisions would take place, these collisions are intended to allow for high precision studies into the weak sector and Higgs mechanisms that would be overlooked in the FCC-hh. In addition, further prospects are made to the FCC-eh capability into producing new particles proposed in the theory of supersymmetry.

Since the subject of this paper revolves around electron-positron type collisions, only FCC-ee will be considered from this point on and FCC-hh and FCC-eh will not be further discussed in depth. Interested readers are instead referred to the FCC Conceptual Design Report volume 3. [11]

3.1.2 The design and expected performance of the FCC-ee

The FCC-ee, formerly known as TLEP, is a high-luminosity, high-precision e+e- circular collider envisioned in a new 80-100 km tunnel in the Geneva area.

The conceptual design[10][42] of the FCC-ee and the approach to accelerating the electrons and positrons to the desired level can be seen in figure 3.2. An experiment starts with the generation of two independent electron beams. One of these provides a low emittance e^- beam, which will be used in the collider, and the other provides a high emittance e^- beam, which will be directed towards an e^+ target in order to generate a sufficient e^+ beam. Both of these beams are injected into a linear particle accelerator (LINAC).

A linear particle accelerator is a long beamline containing a series of electrodes on which an oscillating voltage is applied. This in turn allows for an oscillating electric field which exerts force on the particles when they pass through, imparting energy to them by accelerating them. This specific LINAC is designed to be 250 m long, carrying both initial beams and e^+ target, and is intended to accelerate electrons and positrons from ~ 0 to 6 GeV.

Before injection into the pre-booster ring (PBR), the e^+ beam and the e^- beam consider an unmatched emittance due to their different sources and have to be matched such that they both collide with matching energy. For that reason the e^+ beam is fed into a damping ring (DR) where it is cooled for a short time and then injected into the LINAC at an energy of 1.54 GeV where it is once again accelerated and injected into the PBS together with the e^- beam.

The pre-booster ring (PBR), is a particle accelerator which is designed to accelerate particles to high levels of energies before injection into a collider or, in this case, into an additional booster ring which would accelerate it even further. Currently, it is intended for the PBR to accelerate the e^+- and e^- beams from 6 to 20 GeV. At the moment, two different options are considered as a pre-booster before the beams are transfered to the high-energy booster: using the existing Super Proton Synchrotron (SPS) or designing a completely new ring. The SPS is a current particle accelerator used as a booster for injection in the LHC.

Finally, after being accelerated in the PBS, both beams get injected into the Top-up Booster, which is a ring lying above the line where the actual collisions are to be measured. Here they get further accelerated from 20 GeV to the desired collision energy. When reached, these particles are injected from the top-up to the Electron-Positron collider where they will collide with eacht other at designated interaction points.

The FCC-ee programme will commence with a preparatory phase of 8 years, followed by the construction phase (all civil and technical infrastructure, machines and detectors including commissioning) lasting 10 years. A duration of 15 years is projected for the subsequent operation of the FCC-ee facility, to complete the currently envisaged physics programme. This makes a total of nearly 35 years for construction and operation of FCC-ee. [10]

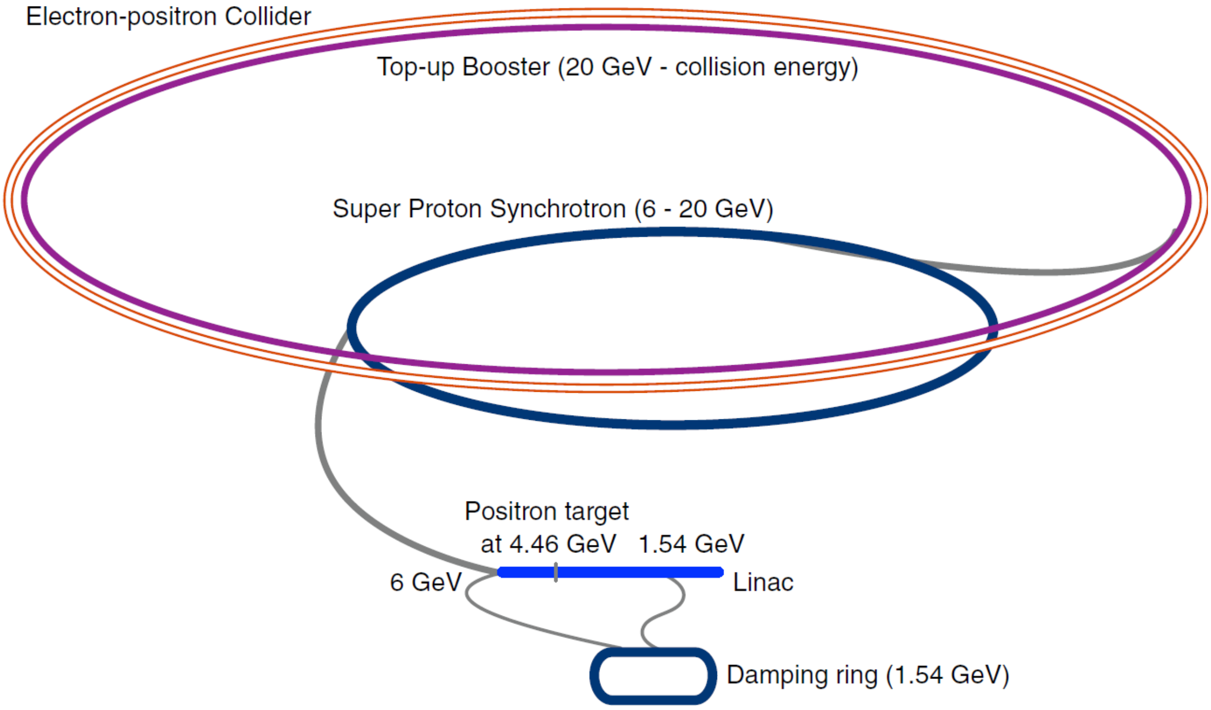


Figure 3.2: An shematic layout of the current expected design of FCC-ee complex where the SPS serves as the used PBR.[10]

Aside from the FCC-ee project, there are currently three other projects in the world have been started with interest in e^+e^- physics. As of now, three other e^+e^- collider designs are contemplated to study the properties of the Higgs boson and exploring extensions of the SM:

- International Linear Collider (ILC [16]): This is a proposed linear particle accelerator which is currently intended to be built in Japan. It is intended to be used for studying the Higgs boson with a centre-of-mass energy of 250 GeV.

- Compact Linear Collider (CLIC [51]): A second proposed linear particle accelerator which is to also be built at CERN. It is designed for studying the Higgs bosons and top quarks at centre-of-mass energies from 380 GeV up to 3 TeV.

- Circular Electron Positron Collider (CEPC [53]): Similar to FCC-ee, CEPC is a proposed circular collider to be built in China and which would span a 100 km. It is able to study the Z, the W, and the Higgs boson, with centre-of-mass energies from 90 to 250 GeV.

The baseline luminosities expected to be delivered at the ILC, CLIC, CEPC, and FCC-ee centre-of-mass energies are illustrated in Figure 3.3. From this figure it can be seen that while most projects aim to study the same areas they do so at completely different luminosities. It can be seen for example, that while CLIC and the FCC-ee both study the top quarks at 380 GeV and 365 GeV respectively, CLIC does so at a lower luminosity which is rather disadvantageous in comparison to the FCC-ee. However, it is the only future project that considers such a large range of centre-of-mass energies making it unique in its own field.

Another interesting example is the study of the Higgs boson for the various different projects. It can be seen in the figure that FCC-ee completely eclipses each project in luminosity even though they to similar centre-of-mass energies. This shows large advantage the FCC-ee poses for this study field.

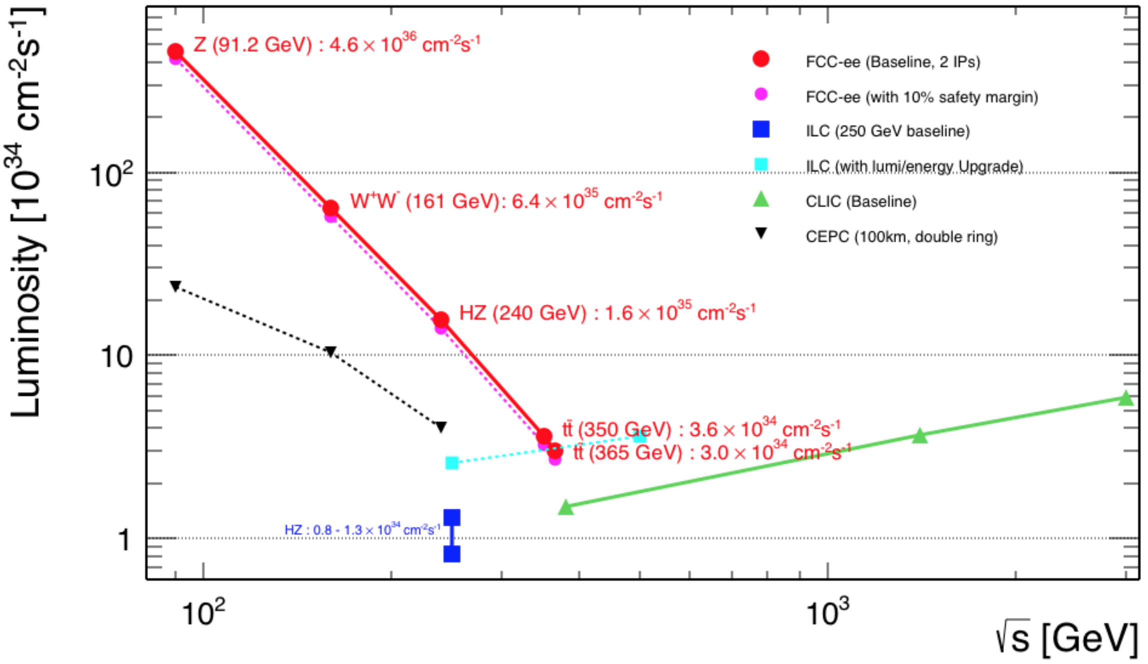


Figure 3.3: The expected baseline luminosity as a function of the centre-of-mass $\sqrt{s}$, for each of the four e^+e^- collider projects: ILC (blue square), CLIC (green upward triangles), CEPC (black downward triangles), and FCC-ee (red dots).[10]

3.2 The International Large Detector

The International Large Detector is a proposed concept for a detector at the International Linear Collider (ILC)[26][25]. The design of the ILD, as seen in figure 3.4, is aimed to be for a multi-purpose detector, combining excellent calorimetry and tracking to obtain the best possible overall event reconstruction. In addition, the combination would allow for the capability to reconstruct individual particles within jets for particle flow calorimetry.

The tracking devices that would allow such a feat consist of a Time Projection Chamber (TPC). A TPC is a particle detector that consists out of gas- or liquid-filled volume in an electric field with a collection system sensitive to electrons. Incoming particles will undergo ionization with the gas- or liquid particles inside the volume. The resulting electrons will then due to the electric field drift towards the collection system. Using the drift time, the spatial information and the type of interactions that could occur, the TPC is able to perform a three-dimensional reconstruction of a particle trajectory or interaction. The considered TPC is designed to provide up to 224 precise measurements along the track of a charged particle.

Additional tracking devices that supplement the TPC are a system silicon based tracking detectors, extending the angular coverage down to very small angles by measuring both in- and outside the TPC, and a silicon-pixel based vertex detector (VTX) which enables long lived particles such as b- and c-hadrons to be reconstructed.

To account for the excellent calorimetry, a tungsten absorber based electromagnetic calorimeter (ECAL), and a somewhat coarser steel based sampling hadronic calorimeter (HCAL) are currently being considered. However because various considerations are still being made regarding the technology used by these calorimeters, it will not be further discussed.

In this paper, even though it is not a planned detector to be used at the FCC-ee, the ILD's specifications were used for the detector response that were obtained from the simulated e^+e^- collision. The reason for this is due to the availability of the simulation data and the range at which the ILD performed, namely the proper coverage of the centre-of-mass range considered at the FCC-ee.

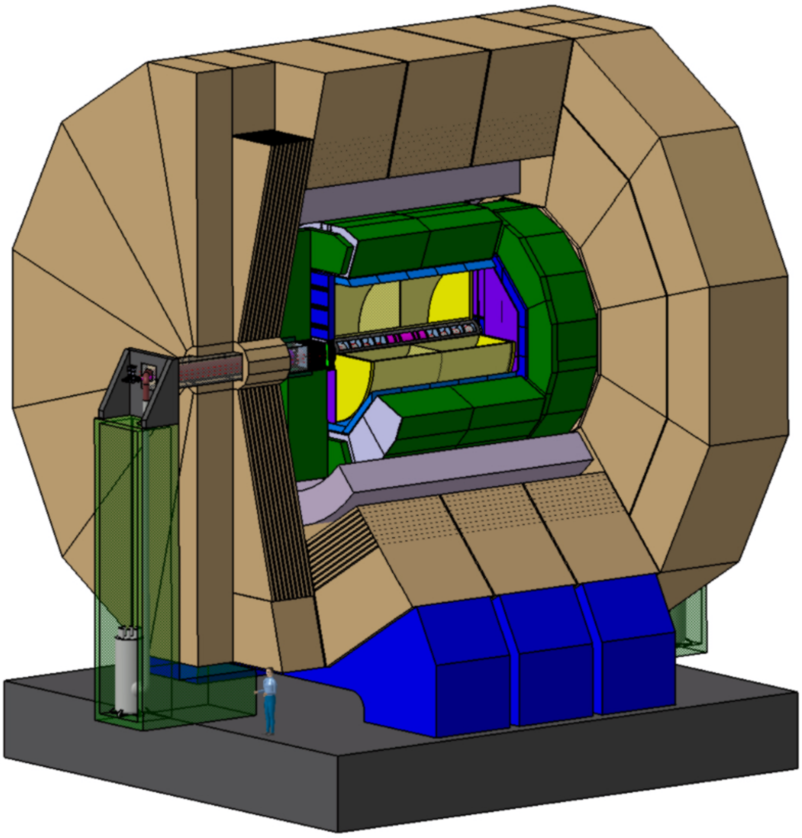


Figure 3.4: View of the ILD detector concept.[15]

Chapter 4

Detector Simulation with Delphes

In this study, the simulation of the detector that hypothetically could be placed in the FCC-ee is performed by the Delphes framework. Delphes is a C++ framework, performing a fast multipurpose detector response simulation. Currently it is capable of simulating the response of various detectors found at CERN, for example: LHC, ATLAS, CMS. In addition it also has the c
In this section a comprehensive, though non-exhaustive, description will be given to how Delphes performs its simulations and how these have been defined for ILD detector. Note that it is not the purpose of the paper to present Delphes. As such this section will be limited to present certain options taken for the Delphes simulation and a few interesting points which should be presented. Interested readers are referred to the paper[29] that the describes framework of Delphes 3.0 which has been used in this study and which served as the largest inspiration for most of this chapter.

4.1 Detector Simulation

The purpose of the Delphes framework is to simulate response of a detector. Delphes does this using a modular system. Here the input data is read and passed through a series of modules which are able to make alterations to this data. The data that passes through these modules are referred to as candidates which represent objects which are useful for the detector simulation. Examples of such objects include: jets, tracks, leptons, ...
The process of a detector response depends on the specifications associated to that specific detector. These specification are outlined into what is referred to as a *detector card*. In these cards it is described which modules are called and the limitations or specifications that are put upon them. A general detector simulation, though, can roughly be described by the following four steps[29]:

- Particle Propagation: As a first step, the simulated particles are put through an axial magnetic field. This magnetic field is localized within the associated detector volume and is parallel to the beam direction.
  Depending on the charge of the particles, their trajectory will differ. Neutral particles will follow a straight trajectory and charged particles will follow an heloicidal trajectory. These tracks are kept track of independently and will serve as an input of the calorimeter step.

- Calorimeters: When the sufficiently stable particles have been propagated in the magnetic field, they will go through the calorimeters. Delphes considers two types of calorimeters: the electromagnetic calorimeter (ECAL) and the hadron calorimeter (HCAL). ECAL is responsible for absorbing the energy of the leptons and photons and HCAL is responsible for absorbing the energy of the long-lived hadrons. The position on a calorimeter where a particle deposits energy is an important property to keep track of. This would allow proper differentiation of the particles. To simulate this, the calorimeter considers a segmented structure in the pseudorapidity, η and the azimuthal angle, ϕ. In other words positions on the calorimeter are defined by coordinates (η,ϕ). The segment on which energy is deposited is kept track of and later outputted as part of the calorimeter tower. This tower keeps track of a particle's energy that was deposited and the coordinates where it deposited it for both the ECAL and HCAL. It is tower that is outputted and used in the third step.
  Note that the tower defines the energy deposited on a specific segment. It does not differentiate between different particles depositing on the same segment and is defined in general as either E_{ECAL} or E_{HCAL}. To be able to differentiate a separate output is defined for energy deposited by a particle's track. This separate output defines the energy associated to the reconstructed[1] track left by a charged particle. It thus provides separate outputs $E_{ECAL,trk}$ and $E_{HCAL,trk}$ that defines the energy deposited by individual charged particles for which their was reconstructed. As such the tower outputs consider both charged and neutral particles, but the track outputs only discerns charged particles.
  It should be noted that not all leptons or hadrons are fully absorbed by the ECAL and HCAL and sometimes only fractions are deposited. This will not be explained in depth, as it is often set by user preference.

- Particle-Flow Reconstruction: One could argue that the track output of the calorimeter is sufficient for reconstructing particles as it has spatial information as well as its energy. However detector are far from perfect and can sometimes cause various systematic effect like smearing or pile-up. Using the outputs of the calorimeter, the

---

[1] It is the first step that performs this reconstruction.

particle flow can be defined by taking the difference of the outputs and the track outputs. This is given by the following formula:

$$\Delta_{ECAL} = E_{ECAL} - E_{ECAL,trk}$$
$$\Delta_{HCAL} = E_{HCAL} - E_{HCAL,trk}$$

By then taking the sum of the maximum positive values of these differences, one can get an idea of how energy is deposited in these segments of the calorimeter. This step is represented by the formula:

$$E^{eflow}_{Tower} = max(0, \Delta_{ECAL}) + max(0, \Delta_{HCAL}) \tag{4.1.1}$$

If it would show that $E^{eflow}_{Tower} > 0$ than that would mean that there is additional energy in these segments that could possibly to smearing caused by the calorimeter for example or by the presence of neutral particle. In other words the particle flow step provides a lot of information about what occured during the detection of the particles which is important to keep track of.

The last step is the object reconstruction using the four-momenta vectors created from the track outputs. This is step which outputs the used variables in the end. As such this will be discussed a bit more in depth in the following sections.

4.1.1 Charged Lepton Reconstruction

For charged lepton reconstruction only two leptons are consider: the electron and muon. Tau leptons are in Delphes not reconstructed as a separate object and are instead seen as a possible jet object. The reason for this is due to the particles quick decay time, which makes it in general difficult for detectors to detect these particles. As such a proper simulation does not do this either so that it can produce a realistic response.

Both leptons are only considered as a separate object if the data from the tracking system falls within an acceptance region defined in the card. For the ILD card this acceptance region was defined for both by their transverse momentum, p_T, and pseudorapidity, η. Electrons only got accepted if: $p_T > 10.0$ GeV and $|\eta| < 2.5$. For the muons this acceptance region was defined by: $p_T > 10.0$ GeV and $|\eta| < 2.4$, additionally when $p_T < 1000.0$ GeV the acceptance probability falls off with increasing p_T.

To conclude, it is often possible that from within a jet, an additional charged lepton can be created. For high energetic jets, these leptons can fall within the acceptance region and thus could cause problems when trying to discern different processes. Especially for studies which try to discern for example, semi-leptonic $t\bar{t}$ events from their leptonic and hadronic counterparts. As such before an accepted lepton is recognised as an object, it goes through a module which checks whether or not it is isolated. Since these unnecessary charged leptons originate from a jet, they ought to be withing a certain range of said jet. The module checks this for each charged lepton ($P = e^{\pm} and \mu^{\pm}$) through the isolation variable given by:

$$I(P) = \frac{\sum_{i \neq P}^{\Delta R < R, p_T(i) > p_{T,min}} p_T(i)}{p_T(P)} \tag{4.1.2}$$

where for the ILD card $p_{T,min} = 0.5$ GeV and where $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ defines a cone drawn around the considered charged particle with radius $R = 0.3$.

The charged lepton is then considered isolated if the isolation variable is below a specified threshold, I_{min} given in the ILD card. For electrons this was, $I_{min} = 0.12$ and for the muon this was, $I_{min} = 0.25$.

4.1.2 Jet Reconstruction

In both lepton and hadron colliders, it is often the case that the final state particles hadronise before detection, causing a long chain of particle productions which grouped together are referred to as a jet. Therefore jet reconstruction is of crucial importance. Delphes therefore employs several different widely used jet clustering algorithms in its framework such that proper algorithms can be used when required. In general, these algorithms try to predict which particles belong to a jet, through their distances between each other. There are two types of jet clustering algorithms which Delphes provides: Cone algorithms and Recombination algorithms.

Cone algorithms, as the name implies, uses the principle of spacial cones to recognise jets by associating together towers lying within a circle with a radius define by the detector card. The cone algorithms that Delphes provides are: CDF Jet Clusters[1], CDF MidPoint[20] and Seedless Infrared Safe Cone (SIScone)[56]. In this study a recombination algorithm was used, and therefore this type of algorithm will not be further explained. Interested readers are referred to the references provided with algorithms.

Recombination algorithms approaches finding jets by recombining/merging particle towers until they represent a proper jet. It does this by considering the distances between each tower, $d_{i,j}$ and their distance to the beam line, $d_{i,B}$ which depends on the tower's transverse momentum. The process checks the minimum distance, d_{min}, between both $d_{i,j}$ and $d_{i,B}$. If $d_{i,j}$ is the minimum distance, then the tower are merged together forming a new tower with their combined momentum. If $d_{i,B}$ is the minimum value then this tower will be considered the final jet and this tower will not be used anymore by the algorithm. This process repeats itself until there are no towers anymore. Each recombination algorithm considers a different definition for $d_{i,j}$ and $d_{i,B}$. The recombination algorithms that Delphes provides are given in table 4.1

Algorithm	$d_{i,j}$	$d_{i,B}$
Hadron Collider Simulations		
kt[31]	$min(p_{T,i}, p_{T,j})\frac{\Delta R^2_{i,j}}{R^2}$	$p^2_{T,i}$
Cambridge-Aachen[62]	$\frac{\Delta R^2_{i,j}}{R^2}$	1
anti-kt[24]	$min(\frac{1}{p_{T,i}}, \frac{1}{p_{T,j}})\frac{\Delta R^2_{i,j}}{R^2}$	$\frac{1}{p_{T,i}}$
Lepton Collider Simulations		
Durham[54]	$2min(E_i^2, E_j^2)(1 - \cos(\theta_{ij}))$	/
generalised kt	$min(E_i^{2p}, E_j^{2p})\frac{(1-\cos(\theta_{ij}))}{1-\cos(R)}$	E_i^{2p}
Valencia[22]	$min(E_i^{2\beta}, E_j^{2\beta})\frac{(1-\cos(\theta_{ij}))}{R^2}$	$p_T^{2\beta}$

Table 4.1: Summary of the jet algorithms native to FastJets and incorporated in Delphes. It should be noted that the Valencia algorithm is not native to the FastJets package, but is given due to it being used in this study.

Normally, the ILD detector card uses the anti-kt algorithm as its default. The reason for using a jet algorithm used in hadron collider simulations is because of the backgrounds in future leptons colliders. In lepton colliders like LEP, an environment is presented with essentially a negligible background. Studies of the $\gamma\gamma \rightarrow hadrons$ background showed a non-negligible background which can disturb the jet reconstruction. It was therefore chosen to use a hadron collider jet algorithm which are able to overcome these types of background as they have been optimised to perform for detectors in the LHC. In the end the anti-kt algorithm proved to be the most robust and is currently considered as the default jet algorithm for future lepton colliders.

In this study, however, it was chosen to use a different algorithm in this study, the Valencia algorithm. This algorithm was proposed as a lepton collider jet algorithm, but one that shares the robustness that the anti-kt algorithm has when it comes to the background. The Valencia algorithm was studied for both $e^-e^+ \rightarrow t\bar{t}$ and $e^-e^+ \rightarrow ZZ$ and has shown to be an improvement in both cases compared to the anti-kt algorithm and other lepton collider jet algorithms. Since this study mainly portrays the former process, this algorithm can be considered the best choice.

4.1.3 B-tagging

The identification of jets is an important topic in high energy collider experiments. B-tagging is a method that allows for the identification of jets that result from the hadronization of b quarks. Due to having the second largest mass of all the quarks, hadrons that originate from the hadronisation of b-quarks, tend to have a lot shorter lifetime then the lighter quark hadrons. These hadrons tend to decay within the jets originating from b-quarks and therefore which will cause the cone of these jets to be a lot wider as seen in figure 4.1.

Delphes simulates this method by checking for each reconstructed jet whether or not a generated b-quark is within a given angular distance, , to the jet axisto the jet axis defined by:

$$\Delta R = \sqrt{(\eta_{jet} - \eta_b)^2 + (\phi_{jet} - \phi_b)^2} \tag{4.1.3}$$

where η represents the pseudorapidity and ϕ azimuthal angle with respect to the jet axis. Naturally, the method itself is not perfect. A jet algorithm does not always perfectly reconstruct the jets and hadrons from c-quarks are also able to decay within its jet. Therefore there is a certain misidentification rate associated to the b-tagging of each detector and likewise to the results of the Delphes b-tagging module. The misidentification rates associated to ILD detector card can be seen in figure 4.2. From this plot it can be seen that the misidentification rate is plotted against the jet's transverse momentum. It should be noted that in the case of the b-quark, this is not misidentification, but rather the identification rate. The equation corresponding to each curve is given as follows, using the symbol ϕ to represent the misidentification rate:

$$\begin{aligned}
\phi_b &= \frac{25}{1 + 0.063 \cdot P_T} \cdot 0.85 \cdot \tanh\left(0.0025 \cdot P_T\right) \\
\phi_c &= \frac{1}{1 + 0.0013 \cdot P_T} \cdot 0.25 \cdot \tanh\left(0.018 \cdot P_T\right) \\
\phi_r &= 0.01 + 0.000038 \cdot P_T
\end{aligned}$$

where P_T represents the transverse momentum, ϕ_b the identification rate of b-quarks, ϕ_c the misidentification rate due to c-quarks and ϕ_r the misidentification rate to remaining other possible constituents like the up and down quark.

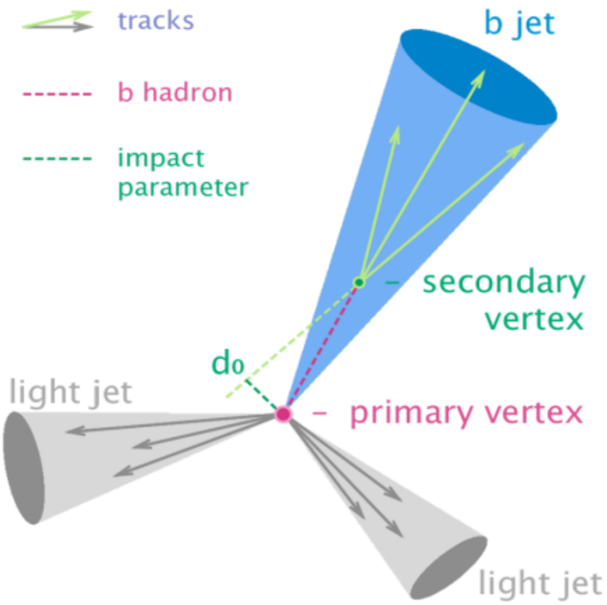


Figure 4.1: A visual representation of how b-tagging differentiates b-jets from other lighter jets. [50]

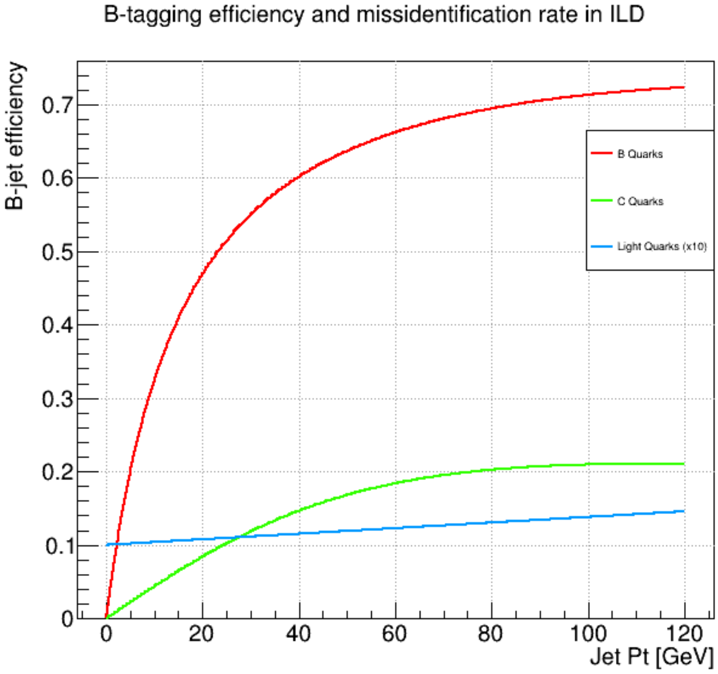


Figure 4.2: The b-tagging misidentification rate that is considered by the ILD detector card. The plot shows the probability proportional to the jet transverse momentum for a b-jet (red) to be properly tagged, a c-jet (green) to be misidentified as b-jet and remaining quarks (blue) to be misidentified as b-jet. Note that the probability that light quarks get misidentified is multiplied by 10 such that it's increasing nature is visible.

Chapter 5

Machine Learning for Data Analysis

In the modern age, the usage of Machine Learning methods has become a popular approach in various fields like for example in image recognition. This is largely due to rapidly improving computational power[1] in the world allowing for more detailed and precise algorithms to be run. Especially in the physics community, Machine Learning methods have shown to provide many advantages. In the field of High Energy Physics (HEP), for example, such algorithms have been employed regularly to differentiate signal and background events with an increased efficiency compared to traditional selections on individual observables.
Whereas neural networks are being used throughout this study, a more detailed explanation on their functionality is provided in the following sections. To start off there will be given a comprehensive, though non-exhaustive, explanation about what machine learning is and how it in a sense works. The reason for keeping this simplified, while being quite important in this field of study, is due to the fact that books can be written explaining the workings of machine learning and it is preferred to not diverge too much from the purpose of this thesis and its research. To do this a few examples of different machine learning approaches will be provided, illustrating how broad the subject of machine learning is. These steps also provide an ideal way to show the contrast between the used method of neural networks and different types of methods.
Afterwards a short introduction to the general working principle behind neural networks will be provided such that the reader is able to comprehend the neural network used in this study when it will be used later on. Many of the ideas in this chapter are were inspired by Ref. [32] and Ref. [49].

---

[1] In the past, machine learning was not very popular due to it requiring a lot CPU to work properly.

5.1 Introduction to Machine Learning

In particle physics the way to study particles that are invisible to the naked eye, is through their behaviour under various circumstances. Every particle has some property that governs some of their actions. When the behaviour of many particles is studied together, these actions become patterns. And it is the recognition of said patterns that allows physicists to associate properties to them. However, humans are limited in their physical ability to discern these patterns. They simply are not able to compare millions of data sets at the same time within a reasonable time. Now in the modern age, the usage of algorithms and computers that allow for the recognition of such patterns has become the benchmark of modern day particle physics. Currently a popular approach of this is through the usage of machine learning (ML).

ML revolves around the principle of teaching an algorithm to be able to recognize a set pattern. This is often done by providing it with base examples which contain this pattern. Afterwards the algorithm would be trained to make prediction on unseen data. This provides an adequate way of testing a theoretical proposed idea (example data) to be tested against empirical data (unseen data).

This explanation is a very basic definition of what ML actually entails. A traditional training process, however, can be described by the following three learning approaches:

- Supervised learning: Often the most simple way to check whether or not an exercise has been solved correctly is to compare it to a provided correct answer. Supervised learning applies this by requiring an input example containing both input data and the expected solution. A ML algorithm would then try to predict the correct answer from the given input data and compare it afterwards to the expected solution. Whether right or wrong, it will update its parameters to reach the correct answer more efficiently.

- Unsupervised learning: Often in data analysis a pattern can be overlooked or is just not immediately visible. Unsupervised learning takes input examples containing only input data and tries to discern patterns. It then reacts based on the presence of these patterns in provided input data.
  A typical application of such an approach is in the determination of probability density functions.

- Reinforcement learning: An approach often taken by psychologists and even parents, is to reward people when they make correct decisions and punish them when making incorrect decisions. Reinforcement learning, follows a similar approach. Like supervised learning, it requires an input example with input data and the associated solution. The algorithm, however, has no direct access to the solutions and instead receives a *reward* when making decisions during training that approach the correct solution. In other words, it similar to a trial-and-error approach.

To provide an example, consider the supervised learning approach for a moment. As said above, the algorithm receives a set input examples with a corresponding desired output. The machine begins the training phase by iteratively going over each given example in the training dataset. Such an iteration is called an epoch. At the start it will produce random output predictions that rarely match the desired output with high accuracy. However, after several epochs it will have adapted this output to be more and more like the desired output. To do this, the algorithm requires a loss function which describes how well the prediction matches the desired output. By minimizing this function, the algorithm can adapt its internal parametrization, such that its prediction slowly start to become accurate. There are many different loss functions and the choice of one depends mostly on the problem at hand. Often though, a loss function can be attributed to one of two categories: the regression losses and the classification losses. An example of a regression loss function is the common mean-square error loss function given by:

$$L_{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{n} \tag{5.1.1}$$

where y_i is the i^{th} desired output of the n examples and $\hat{y_i}$ is the predicted output for said example. As the name suggest, this loss function is ideal for training algorithms for which the target data is normally distributed around a mean value, and when it's important to penalize outliers extra much.
An example of a classification loss function is cross entropy loss function given by:

$$L_{cross-entropy} = -\sum_{i=1}^{C} y_i \log(\hat{y_i}) \tag{5.1.2}$$

where C represents the different classes in the problem. Here y_i represents the true label of the i^{th} input and $\hat{y_i}$ is the label predicted by the algorithm.

There are many different ways as how a ML algorithm can be parametrized. As such many different methods, or more often called models, have been created, each their own parametrization that brings both advantages and disadvantages. In this study, the method of neural networks was applied and will be introduced in the next section. However, a few examples of other possible methods that apply ML are:

- **Support Vector Machines (SVM)[27]**: The principle behind SVM algorithms is the creation of a sort of boundary between different classes. This boundary is referred to as a "*decision boundary*" and considers the largest distance between data points of different classes. The construct this the algorithm applies the use of kernels which measures the degree of similarity between the data points and allows for representation in higher dimensions

- **k-Nearest Neighbours (kNN)[17]**: If an event is surrounded by signal events, there is high chance it is also a signal event. This is the intuitive approach considered in the kNN algorithm. Let us say this algorithm is trained to classify two classes. Depending on the value of k, the algorithm measures the distances[2] between given data point and its first k closest neighbouring data points. It then classifies the data points on which class is more dominant.
- **Decision Tree[43]**: Possibly the most well recognised machine learning method is the decision tree. This algorithm puts the input data through a series of selection nodes which splits into a given number *branches.* The branch the given input data follows depends on the selection criteria. These selections continue until the data input has reached a *leaf* which classifies the given input.

5.2 Basics of Neural Networks

As the name suggests, neural networks (NN) are inspired by the human brain. This ML method was pioneered in 1943 by Warren McCulloch and Walter Pitts[48] who created the first computational model. Later on, in 1958 it was Frank Rosenblat[52] who first created an algorithm model for pattern recognition, the *perceptron.* This model was based on how information is transferred in the brain through its neurons. Initially, the subject of neural network was not a popular. It was in a sense, too early for its time, as there was not enough computational power for it to be useful. In the current day and age, however, computational power has increased significantly and the usage of neural networks has become an almost necessity in some field.

Like the human brain, it is convoluted system of neurons. In this system each neuron is connected to other sets of neurons and can transmit a signal over this connection, similar to the synapses in the human brain. For an artificial neural network this system would start from a set of input neurons, referred to as the *input layer.* Signals than pass from this input layer to the final layer, *output layer* which contains the final prediction made by the algorithm. To reach this output layer, the signal, originating form these input neurons, have to be transmitted to another layer containing several sets of neurons. It is in this layer where the algorithm performs its pattern recognition and is referred to

---

[2] Note that this algorithms can consider different metrics.

as the *hidden layer.* To understand how a neural network does this pattern recognition, it is interesting to look at how an input signal actually traverses this system.

The number of neurons in this layers depends on the problem at hand and its corresponding variables. For example, if this NN is to be trained for image recognition, the input layer consists out of neurons each representing a pixel or a square into which the picture is divided. Another example, which is applicable to this thesis, is that of recognising a specific particle collision process. Here the input could be the various properties represented by variables that are detected in the detector.
Let us consider N input neurons. These input neurons are all connected to a neuron of the hidden layer. Each of these connection to that neuron are then assigned a weight, w_i. In each of these neurons, the weighted sum of the connected inputs, x_i originating from the previous connected layer[3] is then calculated. At this point the calculated signal, $h_i^{(1)}$, in a single neuron of the hidden layer connected to the input layer would be given by:

$$h_i^{(1)} = \sum_{i=1}^{N} x_i \cdot w_i \tag{5.2.1}$$

This is nothing more than a simple linear combination of the input variables which will result in a output of a linear transformed vector with a dimension depending on the number of neurons in the output layer. This type of patter recognition might be useful when trying to derive linear relationships between the in- and output. To add non-linearity to the network, a function is added that determines the response of a neuron depending on the strength proportional to the weighed sum of the inputs. This function is called an *activation function*, σ, which refers to them only activating if the value of the weighted sum reaches a certain threshold (often chosen to be zero such that the sum is non-negative). Examples of such functions are the common heaviside step function sigmoid function and the rectified linear unit (RELU) of which the later will be introduced later. With this the calculated signal in the neuron is given as follows:

$$h_i^{(1)} = \sigma(\sum_{i=1}^{N} x_i \cdot w_i) \tag{5.2.2}$$

Often though, the activation function does not provide a correct threshold or maybe a different threshold is required for a specific neuron. For this, a bias, b is added to the weighted sum of each neuron such that it accounts for this possibility. With this the final

---

[3] This occurs between different sets within the hidden layer after the first set has been reached by the input layer.

calculated signal in the neuron is:

$$h_i^{(1)} = \sigma(\sum_{i=1}^{N} x_i \cdot w_i + b) \tag{5.2.3}$$

It is this signal than will be transferred to every neuron, each considering different weights and biases. During each epoch, these weights and biases are updated accordingly such that the loss function corresponding to this neural net is minimized. Due to this, the neurons in the hidden layer will start to correspond to certain patterns found in the input variables.

When the signal has traversed the hidden layer, it will finally be transferred to the output layer. Depending on the number of neurons in this layer, a neural net can be used for both regression problems, when it has a single neuron, and classification problems, when it has more than one neuron. It should be noted though that the signal transferred to the output node considers a different activation function. The reason for this is that the output neurons should not only keep track of the signal received by each previous neuron, it also should keep track of each of the other output neurons.

As of now, many different neural networks have been created, each serving their own purpose. It is therefore interesting to list off a few of these to show how these things can actually be used. Note that there will not be given an in depth analysis of these examples because a book can be written about each one of them. As such a few neural network examples are:

- Fully-connected: This is the most basic and common type of neural neural network and the one used in this study. As the name implies it is characterised by the fact each neuron between layers are connected.

- Reccurent: Recurrent Neural Networks (RNN) differ from other neural networks due to them having "*memory*". When using a RNN the output data from the neurons are used as feedback to the neurons of the previous layer. A previous run, in other words, provides an additional input data set to the hidden layer neurons in the next run. This feedback allows them to recognise sequential ordering pattern almost immediately.

  An common example of RNNs are Long short-term memory (LSTM) networks. These networks are common in speech recognition technology.

- Convolutional: Convolutional Neural Networks (CNN) are often applied in the field of image recognition. They differ from normal fully connected neural networks through the application of their hidden layers. In CNN's hidden layers are referred to as convolution layers and they differ from normal hidden layers in that they possess a 'filter'. As the definition implies, these filters are there emphasize certain properties when given inputs. These properties often include things that make objects on an image stand out. These can be for example: edges, corners, etc.. Let's consider a filter for an edge. If an image where to be inputted into a CNN, the pixels that make it up would be considered the input. When going through the convolution layers, the filter would convolve over these groups of pixels. Through the mathematical principle of convolution the result would be a new representation of the group of pixels that emphasize the edge pattern that filter has. As such through this principle the CNN is able to discern certain properties objects on an image. By doing for multiple properties and finding correlations between different properties of said objects, the CNN is able to reconstruct the image.

Chapter 6

Data Analysis

6.1 Event Preparation and Generation

As mentioned in the introduction, the purpose of this study is to simulate the detector response of a FCC-ee detector and to take a lepton independent approach for the isolation of semi-leptonic $t\bar{t}$ events. Naturally, if these events need to be isolated, they have to consider background to be isolated from. The main processes responsible for background to $t\bar{t}$ events are: $e^-e^+ \rightarrow W^-W^+$, $e^-e^+ \rightarrow ZZ$ and $e^-e^+ \rightarrow HZ$. Additionally, since semi-leptonic $t\bar{t}$ events are considered, the hadronic and di-leptonic decay of $\ddot{\bar{t}}$ also are considered to be background. The Feynman diagrams of these processes can be seen in figures 6.1-6.3 and their corresponding cross-sections can be found in table 6.1.

event generation (LO)			
Process	Cross-section [fb]	expected events	generated events
$e^-e^+ \rightarrow t\bar{t} \rightarrow b\bar{b}q\bar{q}l^{\pm}\nu_l$	134.40 ± 0.27	201600 ± 405	318864
$e^-e^+ \rightarrow t\bar{t} \rightarrow b\bar{b}l^+\nu_l l^-\bar{\nu}_l$	44.88 ± 0.073	67320 ± 110	76698
$e^-e^+ \rightarrow t\bar{t} \rightarrow b\bar{b}qq\bar{q}\bar{q}$	97.37 ± 0.22	146055 ± 330	331517
$e^-e^+ \rightarrow HZ$	117.30 ± 0.028	175950 ± 42	200000
$e^-e^+ \rightarrow ZZ$	642.8 ± 2.1	964200 ± 3150	1000000
$e^-e^+ \rightarrow W^-W^+$	10720 ± 29	16080000 ± 43500	8040000

Table 6.1: Summary of the cross section, expected events and actually generated amount of events for each process created during the e^+-e^- collision with an integrated luminosity of 1.5 ab^{-1}

Using Pythia8[59] and MadGraph[13] LO samples were generated for each of these processes. The size of the generated samples depended on the integrated luminosity, L_{int}, planned for the FCC-ee. Using the cross-section, σ, an expected amount events, N can be calculated using the following:

$$N = L_{int} \cdot \sigma \tag{6.1.1}$$

Recent plans[10] showed that the integrated luminosity for $t\bar{t}$ production at a centre-of-mass energy of $\sqrt{s} = 365$ GeV, is planned to be about 1.5 ab^{-1}. As such, the size of the generated samples was chosen to be comparable to expected amount events produced in the FCC-ee. Note that this decision excludes the WW background due to its enormous size. It was therefore chosen to generate half the expected amount in order save memory space and time[1].

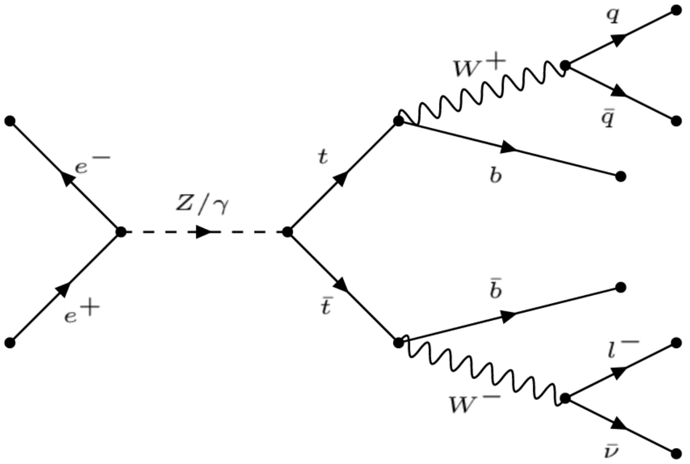


Figure 6.1: The feynman diagram representing the semi-leptonic $t\bar{t}$ process. In this study these are considered the signal events.

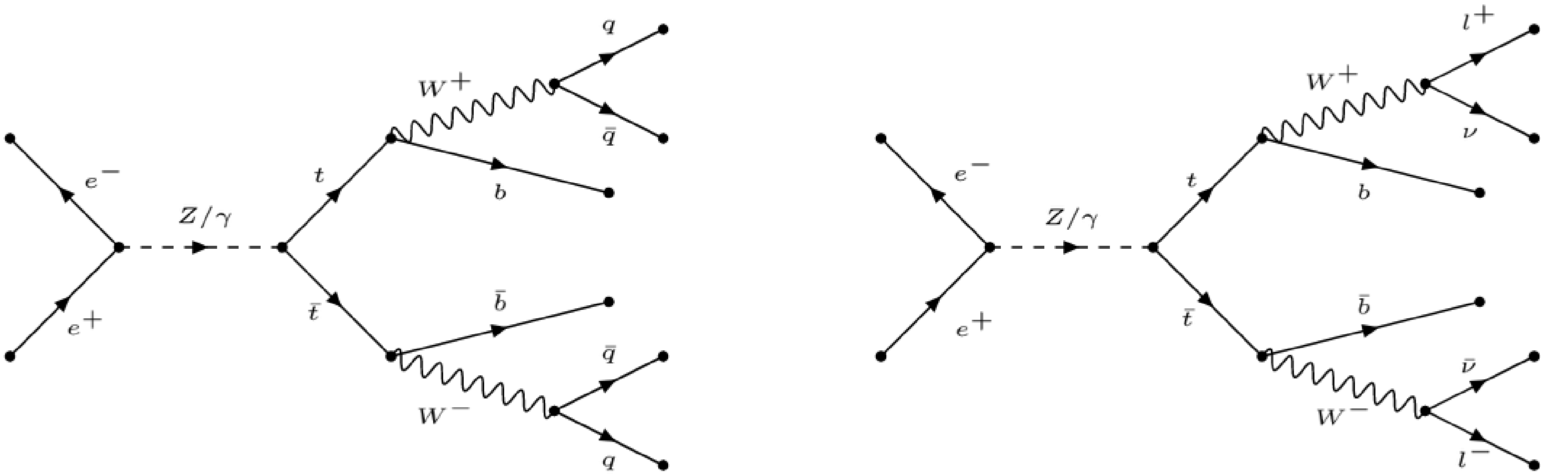


Figure 6.2: The feynman diagram representing the di-leptonic (right) and hadronic (left) $t\bar{t}$ processes. In this study these are considered the background events originating from the $t\bar{t}$ process.

---

[1] The time to generate these 8 million events already took around 3-4 days of a single core of the cluster. This does not consider possible re-generation attempts when errors were made.

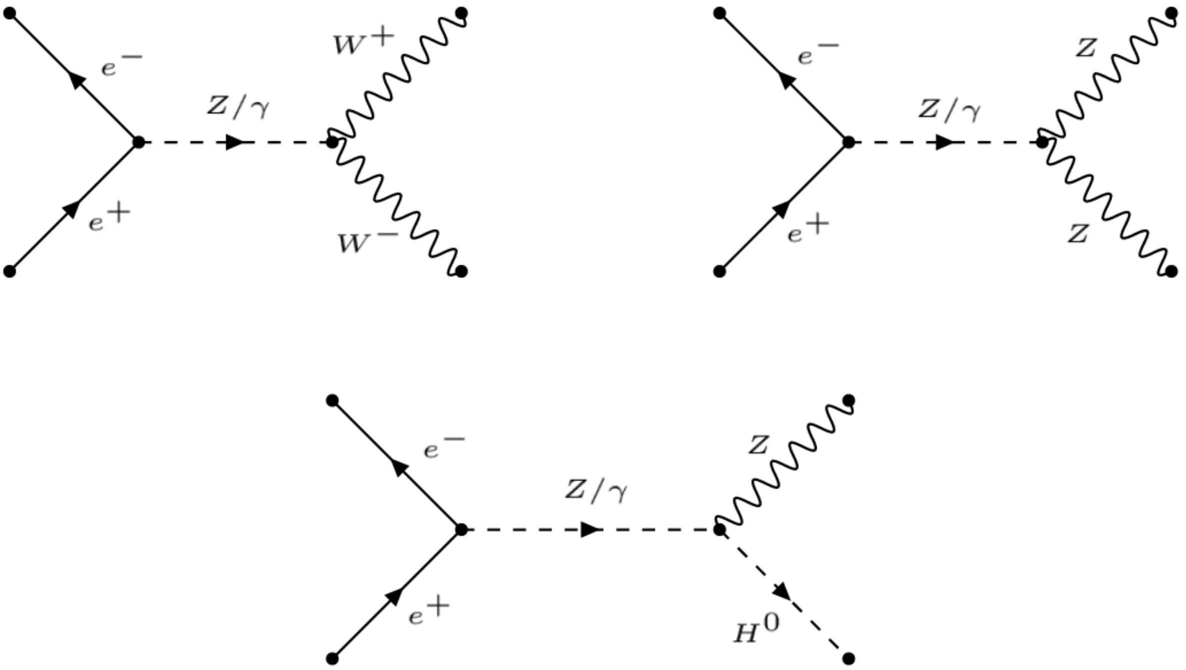


Figure 6.3: The feynman diagram representing the WW (upper left), ZZ (upper right) and HZ (lower) processes. In this study these are considered the background events directly generated during the e^+-e^- collision.

Additionally, as discussed in the introduction, additional samples are necesarry to perform studies of the hypothetical W' boson. For this a model that included this particle was provided by Jernej F. Kamenik et al.[40]. Using this model, 10 sets of semi-leptonic $t\bar{t}$ events, which considered both SM events and events containing W' bosons., were generated using MadGraph, each differing in the W' mass that was assumed during generations. The cross sections, as determined by MadGraph, and number of generated events are shown in table 6.2. Additionally, the cross section is given for events with only W''s. This should give an idea to the reader why the choice of masses was limited to 200 GeV as at this there is a rather small chance of such events to be generated compared to SM diagrams of $t\bar{t}$. For a visual reference these cross section can be compared to each other in figure 6.4. The quick reduction in cross section vs mass is due to the fact that the mass starts to exceed that of the top quark with a mass of 173.5 GeV. As such, the chance of the top quark decaying in something heavier than itself is rather small and depends largely on the Heisenberg Principle.

Event generation (LO) ($L_{int} = 1.5\text{ab}^{-1}$)				
W' Mass [GeV]	Cross-section [fb]	Cross-section for W' [fb]	expected events	generated events
110	259.11 ± 0.50	130.12 ± 0.25	388665 ± 746	300000
120	214.30 ± 0.43	85.49 ± 0.17	321450 ± 638	300000
130	181.26 ± 0.38	53.89 ± 0.11	271890 ± 536	300000
140	158.05 ± 0.29	30.51 ± 0.056	237075 ± 432	300000
150	142.04 ± 0.26	14.74 ± 0.027	213060 ± 384	300000
160	132.32 ± 0.25	5.01 ± 0.0094	198480 ± 372	300000
170	128.30 ± 0.24	1.10 ± 0.0021	192450 ± 366	300000
180	127.58 ± 0.25	0.43 ± 0.00084	191385 ± 369	300000
190	127.39 ± 0.25	0.27 ± 0.00052	191085 ± 372	300000
200	127.29 ± 0.25	0.17 ± 0.00033	190935 ± 375	300000

Table 6.2: Summary of the cross section, expected events and actually generated amount of events for the semi-leptonic $t\bar{t}$ processes that contain a W' boson for each of its mass considered during this study considering an integrated luminosity of 1.5 ab^{-1}.

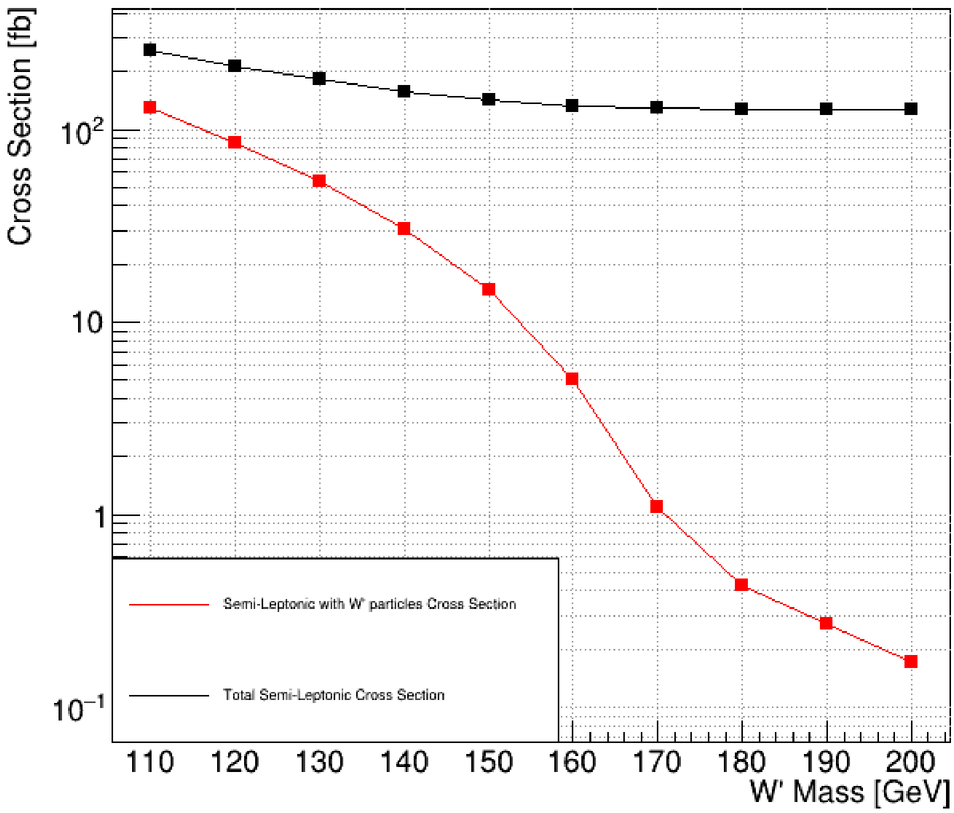


Figure 6.4: The cross sections of the semi-leptonic events generated using the W' model. The black line represents the cross section of every possible semi-leptonic $t\bar{t}$ events and the red line represent said events having at least a single W' boson.

6.2 Event Selection

Like fingerprints on a human, the way a particle interacts with its surroundings is distinctively different from each other depending on their various properties. However, there would not be any problem with backgrounds if this difference was apparent. Event selection to isolate a signature of interest is therefore a crucial subject in many collision experiments. There are many different approaches for selecting/isolating events, each with their advantages and disadvantages. A common, but simple example is taking involves one variable that distinguishes signal from background, puts cuts around the area where the signal is most dominant and rejects the signal and background with a hard cut. Therefore as soon as an event lies outside these limits it will get removed immediately independent of whether it is signal or background. This method, however, is only really advantageous when there are variables which clearly differentiate the background from the signal. When this is not the case, a single cut can remove an unnecessary amount of data. Additionally, direct cuts require optimisation of signal-to-noise ratio for each cut. Which for increasing number of properties can be quite difficult as the order in which cuts are taken can influence it.

In some cases, it might be a lot more useful to separate the background from signal based on the properties they show without removing them. This can be done by studying several properties of a process. Through such a study one could predict a pattern-like behaviour for which the background or signal should behave and make prediction on unseen samples based on that. This would then result in discriminating choice of the prediction to which end the signal can separated from the background.

In contrast to the direct cuts, this approach uses the information obtained from the properties to make predictions instead of simply removing them and requires only to be optimised for the separation based on this predictions instead of over multiple hard cuts which makes it a lot more efficient when considering multiple cuts.

In the case of event selection, the neural network was chosen to separate signal and background based on their patterns associated to several variables. These variables will presented in this section. However, not all variables are suited to be used as property for which the processes separated using a neural network. This is because some properties have a definite known value belonging to a signal event. If the neural network were to train upon such a variable it would try to guess the correct corresponding value to this property. This brings the possibility that it does so incorrectly. For example, the semi-leptonic $t\bar{t}$ signal events consider a definite single lepton in the final state. It could very well be that the neural network associates a different value to it which would be simply wrong. As such aside from the event selection through the neural network, three additional direct cuts were made. One before applying the neural network and two direct cuts after.

6.2.1 Pre-Selection

Before defining cuts based on the properties of the various processes that would allow a separation between signal and background, an initial pre-selection was performed. The semi-leptonic $t\bar{t}$ decay mode generates 4 quarks (2 b quarks and 2 random quarks) and 1 charged lepton and its corresponding neutrino. As discussed in section 4.1.1, Delphes recognises the charged lepton as a separate object, the neutrino as missing energy and the 4 would be recognised as 4 jets. As such an initial pre-selection requirement was that there were 4 jets reconstructed during the simulation. To not lose too much info the jet algorithm considered the exclusive clustering of 4 jets so it most cases, 4 jets would be detected. However, not all background are able to do this an example is the ZZ background if it decays purely leptonic. This would provide no constituents to make up a jet and as such can promptly removed from consideration.

6.2.2 Input variables for the neural network

In this study, the choice was made to train a neural network to recognise semi-leptonic $t\bar{t}$ events and differentiate them from background. Neural networks have been introduced in section 5.2, but for this neural network to be properly trained it requires characteristic properties/variables where the semi-leptonic $t\bar{t}$ events can be differentiated from the background. In this section these properties/variables will be introduced alongside physical explanation which motivates their choice and their difference with the background processes will be shown. There is only one constraint that was put upon the choice of variables, they were not to be related to a charged lepton. This was to ensure that event selection was completely lepton independent.

Of the following variables, the first four were variables provided by a previous thesis study performed by Matthias Mancini[46] at VUB. Of his considered variables these four proved to most applicable to this study. The remaining five variables were newly developed by the student himself for their advantage in this study. The considered variables used in the neural network are:

Invariant mass of the 4 jets (M_{4j}): A logical property to consider first is the invariant mass of the 4 jets. Note that this is not the sum of the masses of the four jets, but rather the mass associated to the four-vector sum of these jets. Because the jets of the WW and ZZ processes originate directly from particles that were created during the collision of the electron and the positrion, there is little chance of energy to be lost to external sources. Especially so when compared to the $t\bar{t}$ process which goes through a whole decay chain and generally does not solely produce 4 jets and loses energy to other particles. This is especially true for the semi-leptonic and di-leptonic decay mode where a lot of the energy is given to the charged leptons and their neutrinos which Delphes does not recognise as jets. As such these background processes are expected to provide an invariant mass close to the initial centre-of-momentum $\sqrt{s} = 365$ GeV.

This hypothesis proved to be correct as seen in figure 6.5. From the figure the HZ process was seen to have a similar behaviour as the WW and ZZ background. This is largely due to fact that the Higgs boson has one of its dominant decay modes is bb, which considers a branching ratio of 0.5824 ± 0.0016[5]. If the Z than also decays hadronically, which has branching ratio of 0.69911 ± 0.00056[5], the process would produce what Delphes would recognise as 4 jets. Similar to the WW and ZZ processes, the HZ would thus not lose much energy to other particles. Additionally, the peak at low mass values is also seen for the WW, ZZ and HZ processes. This low peak results due to fact that it is not a guarantee that these particles will decay into jets. As such the jet algorithm tries to generate jets from whatever it can find, within the event, that is not considered a lepton or a photon. Because of this, most of the energy will be lost which results in the lower peaks seen in figure 6.5.

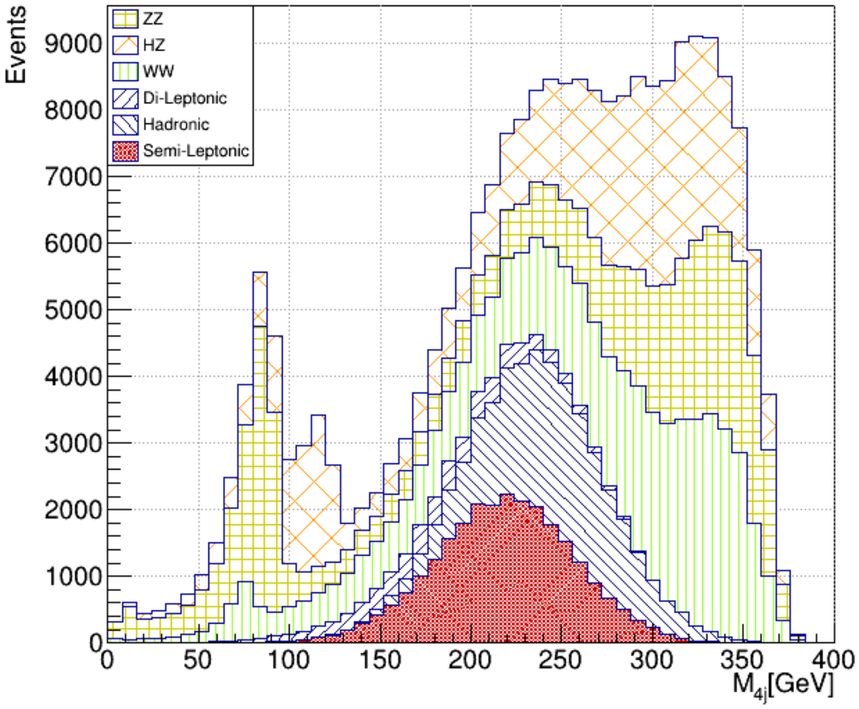


Figure 6.5: The invariant mass of 4 jets distributed for each generated process that occurs during e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}. For each process 60000 generated events were considered. In the case for the $t\bar{t}$ this amount is splitted among the di-leptonic, semi-leptonic and hadronic decay mode. The signal here is represented by the semi-leptonic distribution which is the red filled in histogram.

1^{st} & 2^{nd} **Minimum mass of 2 jets (M_{min2j}):** Processes like WW, ZZ, HZ and di-leptonic $t\bar{t}$ are not necessarily expected to generate 4 jets. It is very possible for them to decay purely leptonic. Because the jet algorithm uses exclusive clustering it enforces the existence of 4 jets. As such, jets resulting from processes that would not generate enough jets are created in some other manner. This is done often by splitting existing jets. The original jets originate from quarks or gluons. By looking at the minimum mass of 2 jets it is possible find back masses that are smaller than those from massive objects like the W and Z boson. Figure 6.6 and 6.7 prove this point clearly. The semi-leptonic and hadronic $t\bar{t}$ processes are not effected heavily due to them definitely generating more 4 quarks which result into jets. This is not the case for the other processes as seen in the figures peaking at low values.

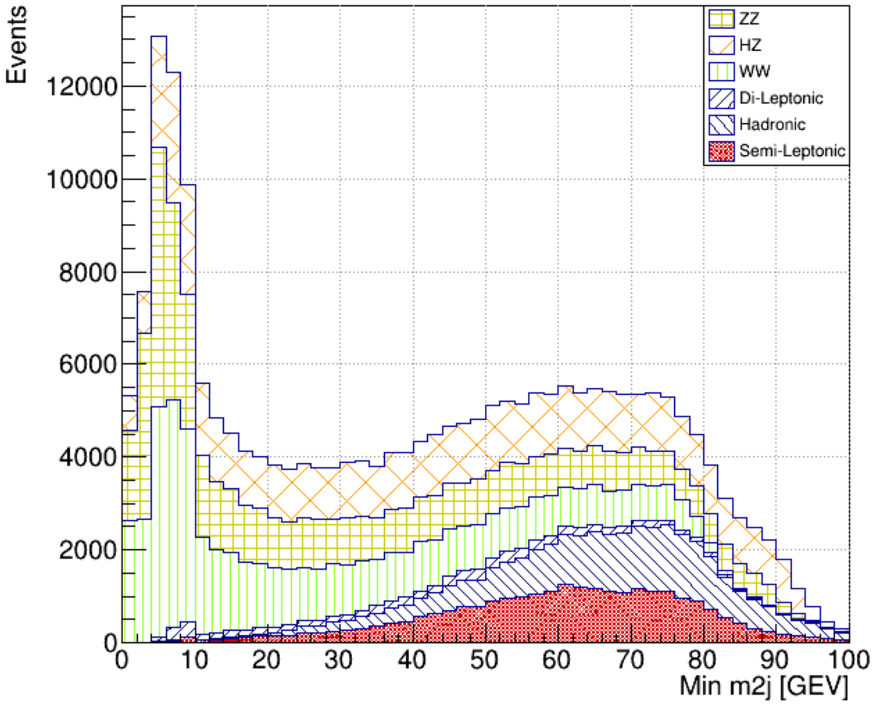


Figure 6.6: The 1^{st} minimum of the invariant mass of 2 jets distributed for each generated process that occurs during e^{+}-e^{-} collision at an integrated luminosity of 1.5 ab^{-1}. For each process 60000 generated events were considered. In the case for the $t\bar{t}$ this amount is splitted among the di-leptonic, semi-leptonic and hadronic decay mode. The signal here is represented by the semi-leptonic distribution which is the red filled in histogram.

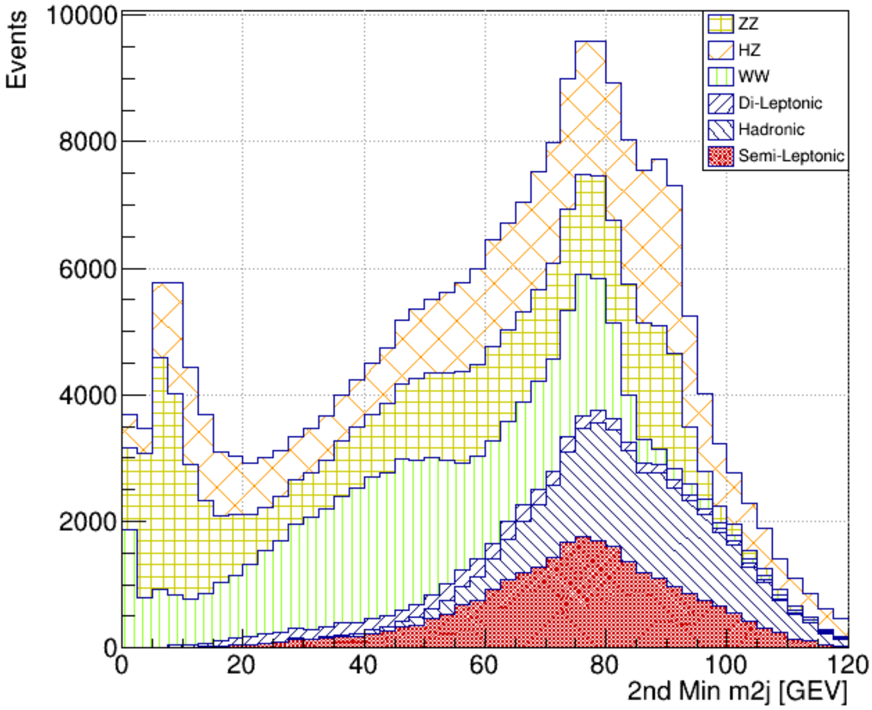


Figure 6.7: The 2^{nd} minimum of the invariant mass of 2 jets distributed for each generated process that occurs during e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}. For each process 60000 generated events were considered. In the case for the $t\bar{t}$ this amount is splitted among the di-leptonic, semi-leptonic and hadronic decay mode. The signal here is represented by the semi-leptonic distribution which is the red filled in histogram.

Missing Transverse Energy (MET): A logical property to consider is the MET. The largest contributions to missing energy are the neutrinos. Processes like the hadronic $t\bar{t}$ and ZZ production do not create any neutrinos and as such are expected to have low values. Furthermore processes like WW and ZZ will be a lot more sensitive because compared to $t\bar{t}$ processes they produce les energetic neutrinos so a lot less transverse momentum.

Figure 6.8 proves the usefulness of this property showing how well the hadronic, WW, ZZ and HZ processes peak at low MET values.

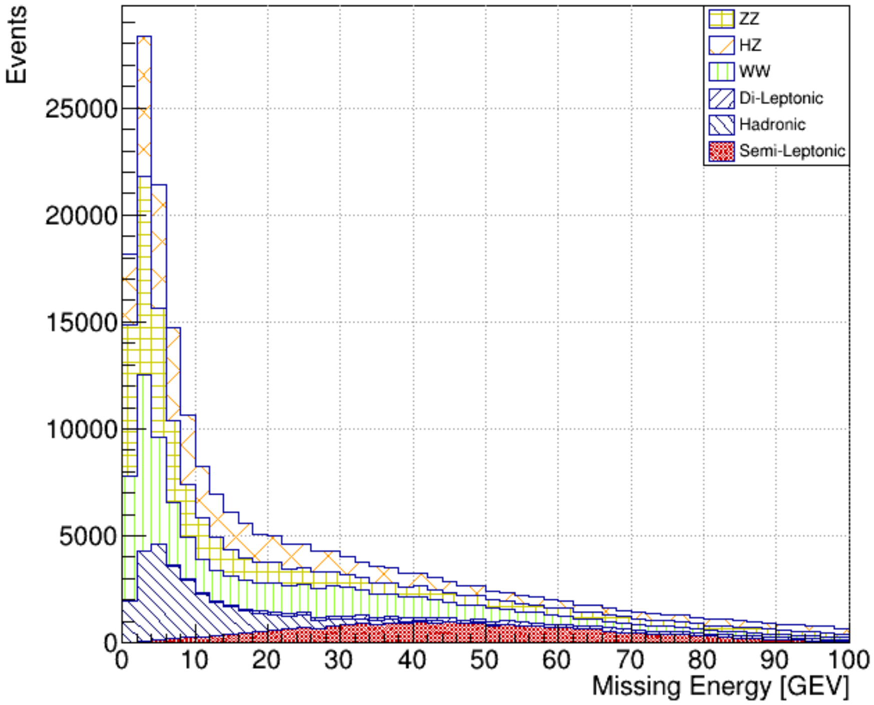


Figure 6.8: The missing transverse energy distribution for each generated process that occurs during e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}. For each process 60000 generated events were considered. In the case for the $t\bar{t}$ this amount is splitted among the di-leptonic, semi-leptonic and hadronic decay mode. The signal here is represented by the semi-leptonic distribution which is the red filled in histogram.

Minimum transverse energy of individual jets: It was earlier mentioned that the transverse momentum and energy of the jets from WW and ZZ processes is a lot smaller than compared to that of the $t\bar{t}$ processes. The reason for this is due to $t\bar{t}$ constituents transferring a lot more kinetic energy towards the transverse direction during their decay chain. As such this energy aimed more towards the transverse direction. In contrast, processes like WW and ZZ lose a lot less due their shorter decay chain and as such individual jets will generally consider a lot less transverse momentum than those of the $t\bar{t}$ processes.

Figure 6.9 proves this point. Here it can be seen that the WW, ZZ and HZ processes consider a lot smaller minimum transverse energy than those of the $t\bar{t}$ processes which are more normally distributed.

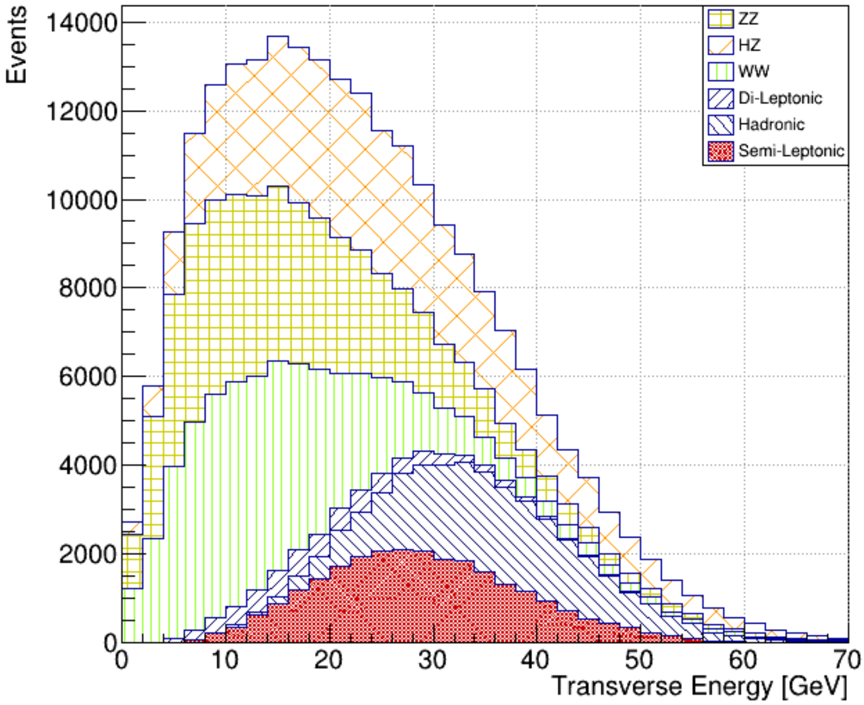


Figure 6.9: The minimal transverse energy of individual jets distribution for each generated process that occurs during e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}. For each process 60000 generated events were considered. In the case for the $t\bar{t}$ this amount is splitted among the di-leptonic, semi-leptonic and hadronic decay mode. The signal here is represented by the semi-leptonic distribution which is the red filled in histogram.

Maximum energy of individual jets: As explained earlier, processes like the WW and ZZ share energy among less constituents during its decay chain. As such it can be expected that the total energy of an individual jet will be a lot larger when compared to jets from the $t\bar{t}$ processes which have their energies more shared among their final constituents. This is shown to true as seen in figure 6.10.

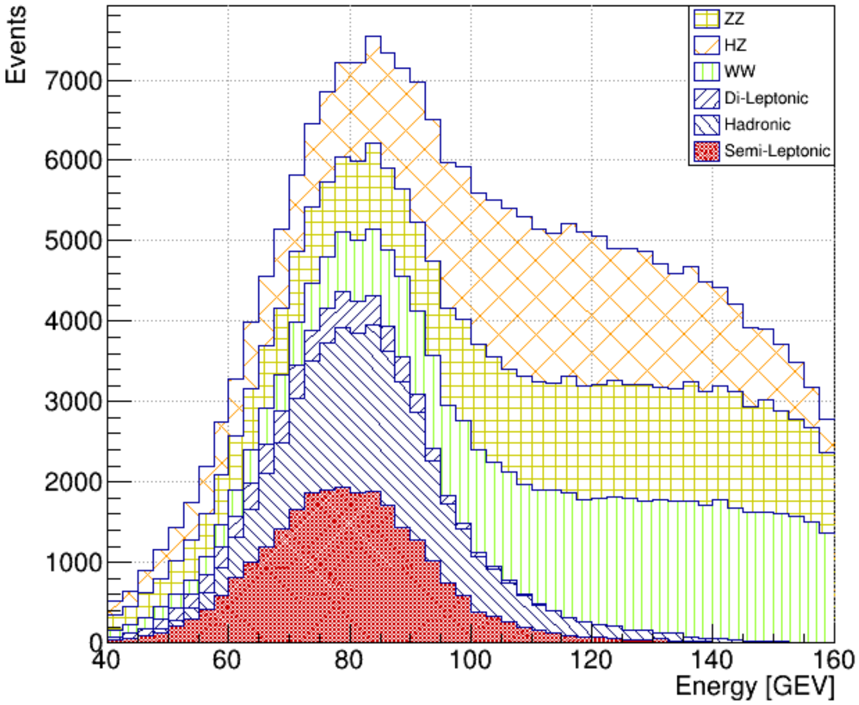


Figure 6.10: The maximum energy of individual jets distribution for each generated process that occurs during e^{+}-e^{-} collision at an integrated luminosity of 1.5 ab^{-1}. For each process 60000 generated events were considered. In the case for the $t\bar{t}$ this amount is splitted among the di-leptonic, semi-leptonic and hadronic decay mode. The signal here is represented by the semi-leptonic distribution which is the red filled in histogram.

Rapidity Mean (y_{mean}): In particle physics the rapidity, y, is a measure of a particle's velocity relative to the beam axis and as such is sensitive to particles being produced at threshold. It is therefore a useful variable to discern processes which differ in their momenta in specific directions. The property, the Rapitidy Mean, is given by the following equation:

$$y_{mean} = \frac{(y_{j1} + y_{j2} + y_{j3} + y_{j4})}{4}$$

where j_i represents one of the four jets that are detected. The choice of this variable was supported due to the fact the jets from the WW and ZZ processes are expected to have a larger momentum along the beam direction than the jets from the $t\bar{t}$ processes. The reason for this is because the particles from which the jets originate in the case of the WW and ZZ processes were created directly out of the collision between the electron and positron and the additional collision energy is converted into their kinetic energy. In the $t\bar{t}$ process, on the other hand, two top quarks are created at rest and carry little additional energy. These will in turn decay into a jet and a W boson distributing the momentum between those two. And the W boson will afterwards decay into two jets distributing the momentum even further. As such the mean rapidity spread over all the jets will generally be lower for the $t\bar{t}$ processes than for WW and ZZ processes which will manifest itself in the rapidity. This hypothesis proofed to be correct as can be seen in figure 6.11. It can be seen from this picture that the WW and ZZ processes are well discernible from the $t\bar{t}$ process.

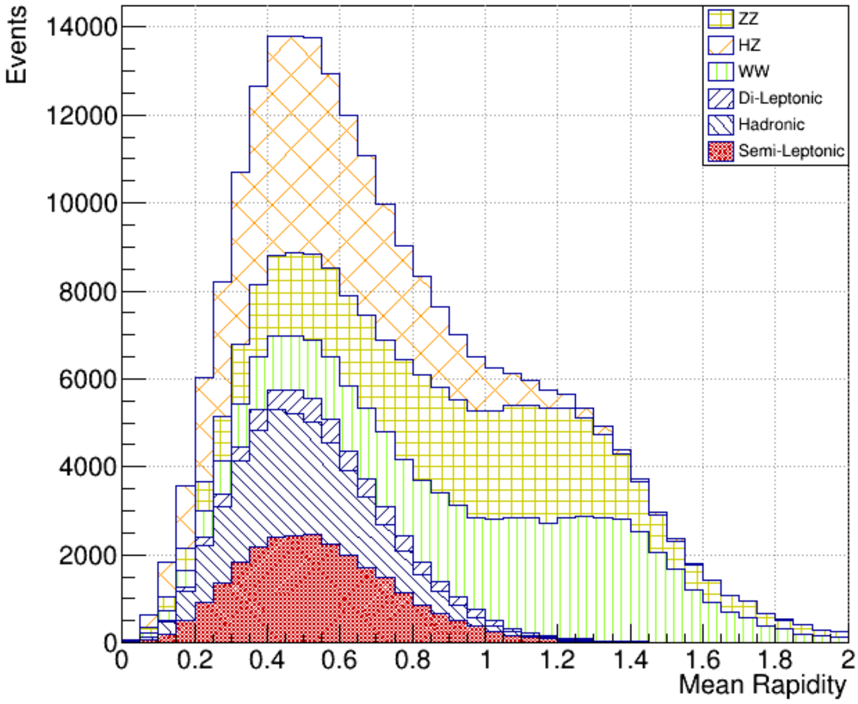


Figure 6.11: The Rapidity Mean property distribution for each generated process that occurs during e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}. For each process 60000 generated events were considered. In the case for the $t\bar{t}$ this amount is splitted among the di-leptonic, semi-leptonic and hadronic decay mode. The signal here is represented by the semi-leptonic distribution which is the red filled in histogram.

Rapidity Gap: As was discussed in the explanation of the rapidity mean property, processes like the WW and ZZ will in general have jets with a higher rapitidy compared to the jets of the $t\bar{t}$ processes. Similarly then, the maximum difference of rapidity between individual jets will also be a lot larger than for those background processes. The rapidity gap, y_{gap} is defined by the equation:

$$y_{gap} = min(|y_i - y_k|)$$

where i and k represent one of the four jets. Figure 6.12 shows the validity of this property where the process WW and ZZ have tail extending to larger values of the rapidity gap.

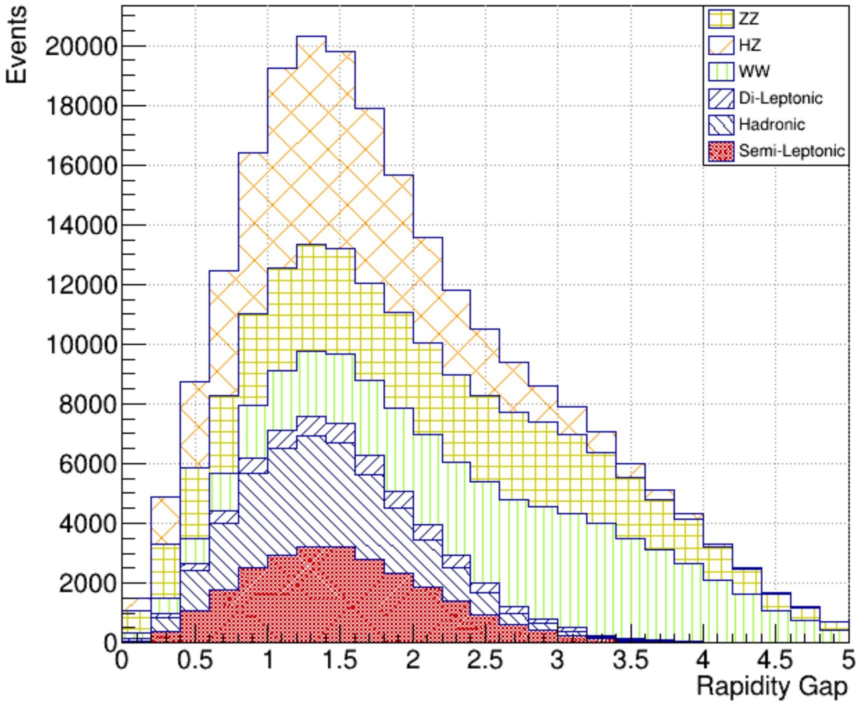


Figure 6.12: The Rapidity gap property distribution for each generated process that occurs during e^{+}-e^{-} collision at an integrated luminosity of 1.5 ab^{-1}. For each process 60000 generated events were considered. In the case for the $t\bar{t}$ this amount is splitted among the di-leptonic, semi-leptonic and hadronic decay mode. The signal here is represented by the semi-leptonic distribution which is the red filled in histogram.

The minimum difference of the invariant mass of 3 jets with that of the W boson mass:

As explained earlier, in this study exclusive clustering is used and creates four jets. WW processes are not expected to generate four jets and a lepton and as such to make up for this the jet algorithm splits existing jets into two, such to make up for this. As such W boson mass can possibly be determined by considering the invariant mass of the direct product of these split jets. As such taking the minimal difference between the W boson mass and the invariant mass of three jets seems the optimal choice for a property.

Figure 6.13 shows that the above argument is well supported as not only the WW, but also the ZZ and HZ processes are found to be sensitive to this. The former is probably to the Z boson mass being close to the W boson mass and the latter due to the fact that the Higgs boson is able to decay into two W bosons which is one of its dominant decay modes with branching ratio of 0.2137 ± 0.0099.

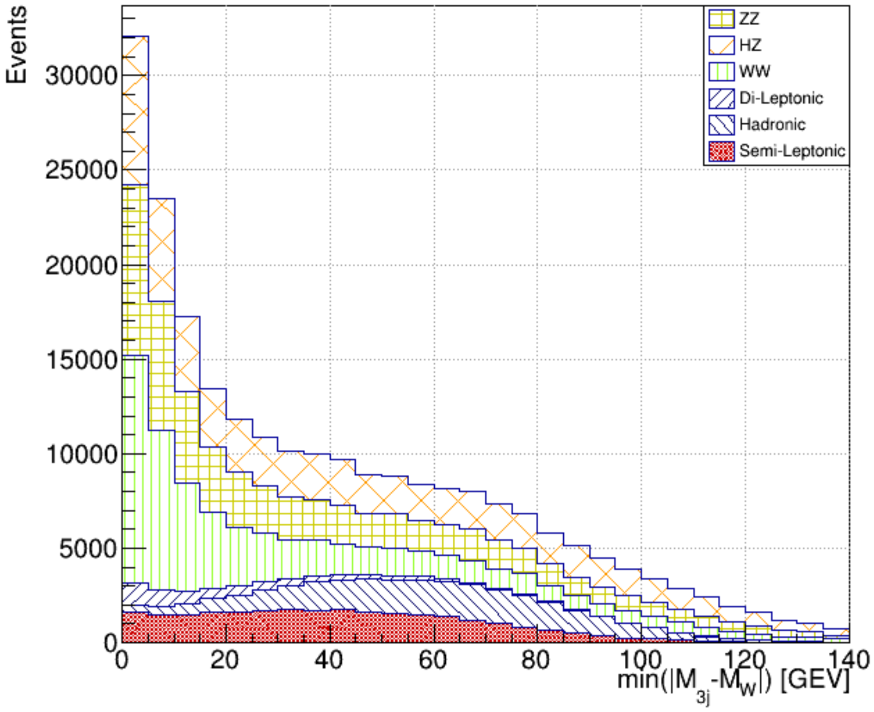


Figure 6.13: The minimum difference of the invariant mass of 3 jets with the W boson mass distributed for each generated process that occurs during e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}. For each process 60000 generated events were considered. In the case for the $t\bar{t}$ this amount is splitted among the di-leptonic, semi-leptonic and hadronic decay mode. The signal here is represented by the semi-leptonic distribution which is the red filled in histogram.

6.2.3 Direct Cuts

After the neural network has been trained using these variables, it will be easier to separate signal events from background events as the neural network will predict a discriminating value to them based on the patterns found in their variables. However, as explained in the beginning of this section two additional direct cuts were applied after the neural network to the results. The first variable, the b-tagging, is very useful for when considering $t\bar{t}$ events. This process almost definitely contains 2 b quarks in its final state. Additionally, detectors often employ b-tagging methods such that these quarks are identified. As such requiring a certain number b quarks to exist in the final state is good property to consider as backgrounds such as the WW process does not contain definite b quarks. Note

that this does not imply that it will be required for 2 b quarks to exist when taking the direct cut. This is a possible choice. However, b-tagging methods are far from perfect, as discussed in section 4.1.3, and this should be taken into account when considering such a cut and choosing the amount b quarks are required when taking the direct cut.
The second property, the presence of a lepton, has been chosen because this is what differentiates the semi-leptonic events the most from its hadronic and di-leptonic counterparts. One can argue that this does not make this algorithm lepton independent. However, this property relies only on the existence of a lepton, it does not rely on any of its properties. Also, the algorithm was first desired for studies into Lepton Universality Violation, which would require a lepton to properly study. As such requiring a lepton to exist is not detrimental to this algorithm at all. It should be noted that tau leptons are not under consideration here as Delphes does not isolate them properly. As such only the presence of a muon/anti-muon or electron/positron are considered.

6.3 Neural Network Layout & Performance

6.3.1 The Layout & Specifications

In the study represented by this paper, the choice was made to use neural network for the classification of the generated data. The type neural net used in this study was a fully-connected neural network. The reason for this was mainly due to the fact that neural nets like RNN and CNN are initially not suitable for simply discerning singal and background based on given input variables. As such a fully-connected neural network, which is the most common and basic neural network, was the ideal choice. Additionally, it was chosen to take a supervised learning approach. The reason for this choice is due to the fact that it desired to learn the specific patterns belonging to the background and signal. It is therefore better if the neural network knew which given input corresponds to which class in the end, as it could be possible that the neural network would find wrong patterns if unsupervised learning were to be used. A visual representation of the used neural network is given in figure 6.14.

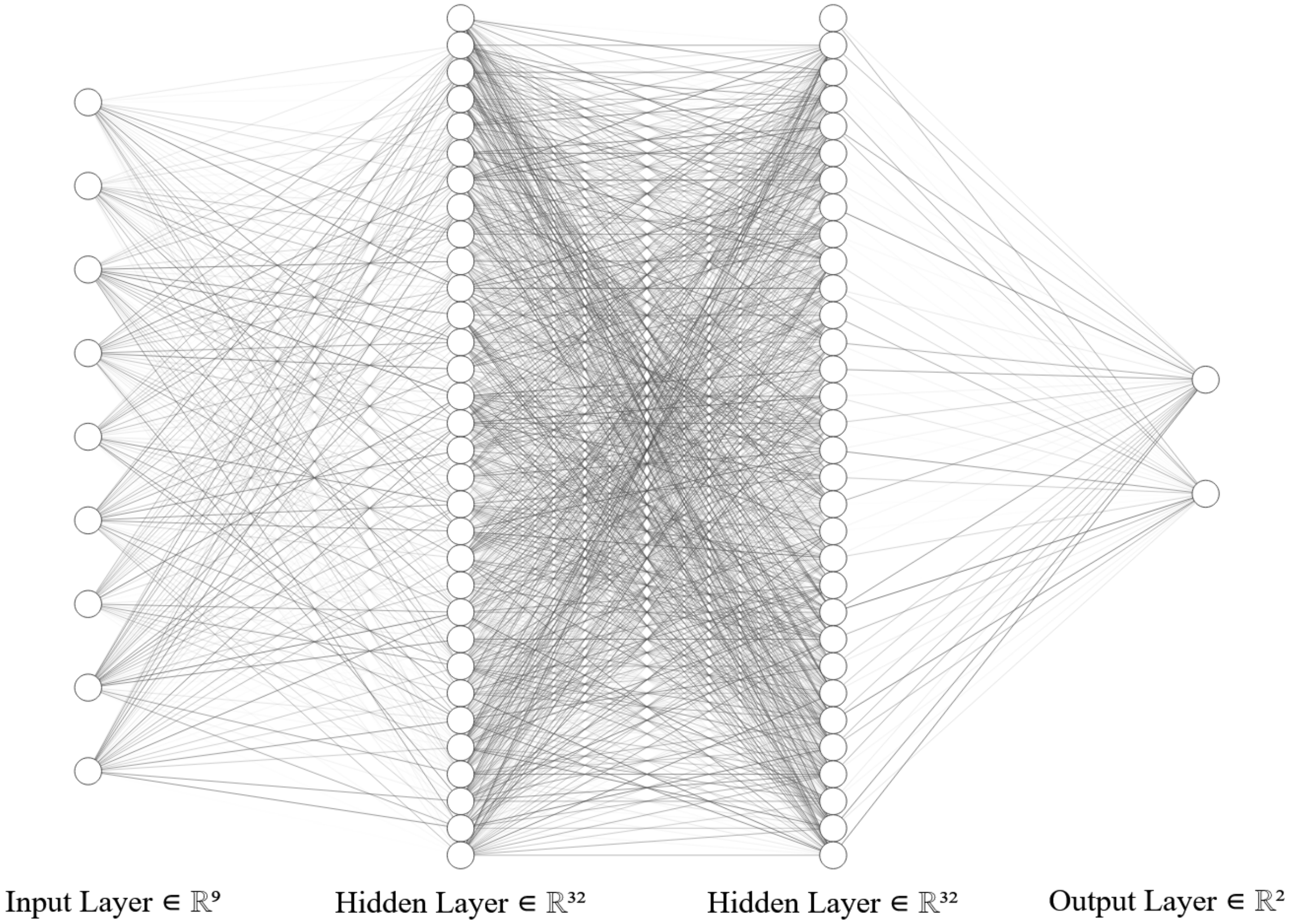


Figure 6.14: A schematic diagram representing the neural network used in this study.

As shown in the figure, the neural net considered an input layer with a set of 9 neurons, two hidden layers with a set of 32 neurons and an output layer with 2 neurons which classify the the input as either signal or background.
The amount of neurons in the input layer correspond to the 9 variables that were chosen to discriminate the semi-leptonic $t\bar{t}$ events from the background events. These variables have been introduced and discussed in the section 6.2.
The choice of hidden layer depends heavily on the resulting performance. Changing the amount of layers and neurons can heavily influence the training time and efficiency of a neural network. As such, in this study, the neural network training loss and accuracy where determined for different layers and neurons.
Initially, 2 layers were considered for the hidden layer. From this angle the optimal amount of neurons were determined. In figure 6.15 the loss (blue) and the accuracy (red) for increasing amount of neurons was determined. It can be seen from this figure that the optimal amount of neurons was around 32. As such it was decided to use this amount for

future training.

Like for the neurons, the loss (blue) and accuracy (red) from training a neural network were determined for different hidden layers. This can be seen in figure 6.16. For each these hidden layers 32 neurons were considered. It can be seen from the figure that accuracy and loss already start to converge at two hidden layers. As such; 2 hidden layers were decided to be the optimal choice for the future neural network.

It should be noted, that an additional test could have been performed testing different amounts of neurons for each different layer. This was not done for this study due to the quickly converging nature of the number of hidden layers and mostly because of time. Training a single neural network can take up to as short a 2 hours to as long as half a day. Future studies should consider this additional test as it could improve the neural network. For example employing deep learning methods could improve this approach significantly but training times can go up to weeks.

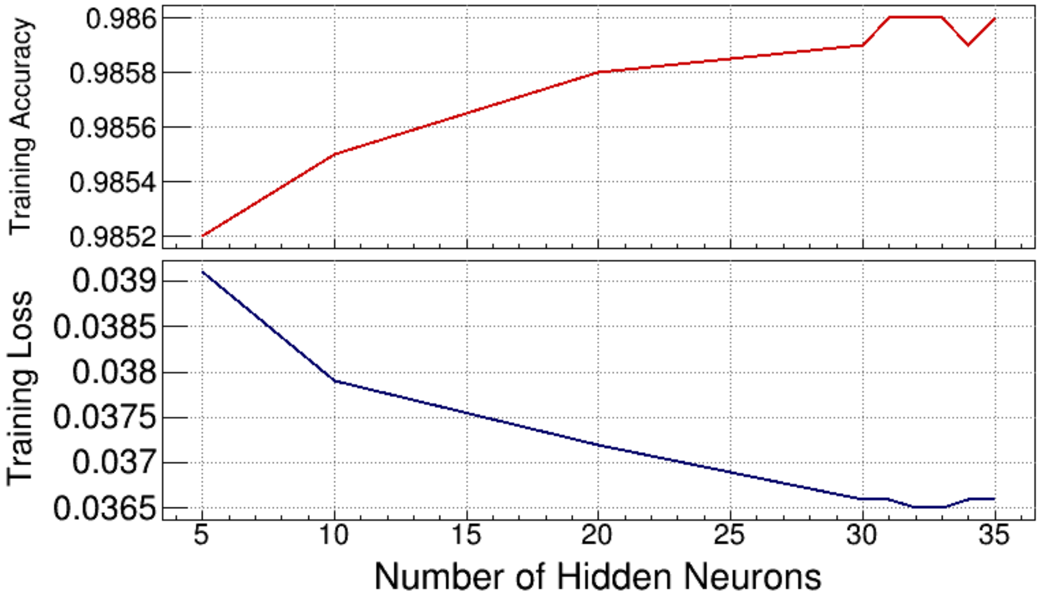


Figure 6.15: The different training accuracies (red) and losses (blue) attained by the training sample considering 2 hidden layers and when varying the number of hidden layer neurons.

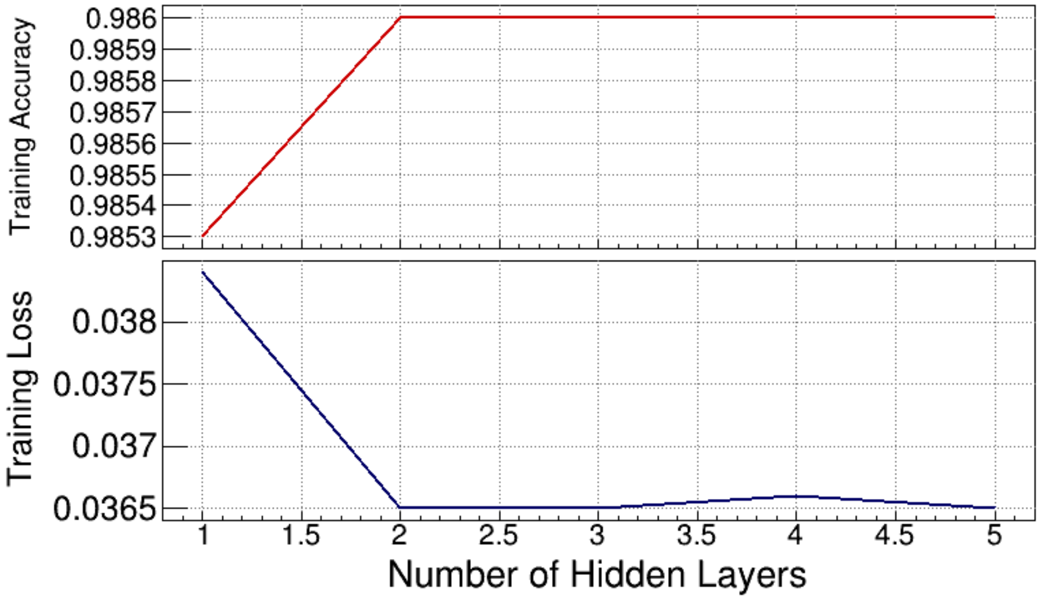


Figure 6.16: The different training accuracies (red) and losses (blue) attained by the training sample considering 32 hidden layer neurons and when varying the number of hidden layers.

Finally, for the output layer two neurons were considered. One which represented the signal, the semi-leptonic $t\bar{t}$ events, and one which considered the backgrounds. However, having a single an output node represent all of the possible backgrounds is very inefficient. The neural network would have a difficult time to discern what makes the background different from the signal as it does not really have consistent pattern associated to it. In other words, it is best to consider a single process to each output neuron. In the current study, the WW background was considered to represent the background neuron. There are specifically two reason for this. The first reason is the most obvious reason, it is by far the largest background process. As discussed in section 6.1, about 201750 of semi-leptonic $t\bar{t}$ events are expected based on the cross-section provided by MadGraph. The WW process on the other, expected an amount of 16080000 events. This is close to 80 times the amount of semi-leptonic $t\bar{t}$. This background, therefore, will dominate a lot of the detected events. It is therefore important that the neural net is able to remove as much of these events as possible.

The second reason, relies mostly on the fact that the other $t\bar{t}$ processes are very similar to single lepton $t\bar{t}$ decays and that the ZZ background as seen in section 6.2, behaves rather similar to the WW background and will probably be considered such a background by the neural net. Also note that the other $t\bar{t}$ processes will heavily be reduced by the direct lepton cut as introduced in section 6.2 and as such training for these processes will probably take more signal away than necessary. To conclude this reasoning, one could argue why the HZ background was not considered. It was in fact a point of consideration when optimising the neural network. However, the direct cuts later showed this need for training on the HZ background was unnecessary and as such has been promptly removed from consideration. This point will become more clear later.

As mention in section 5.2, every hidden layer needs an associated activation function, σ, such that the neurons can transfer a signal to the following layer. The activation function that connects the input layer and the hidden layers is the ReLU activation. ReLu, or rectified linear unit, is a function represented by the following formula:

$$\sigma(a) = max(0, a) \tag{6.3.1}$$

where a in this case is the linear combination of the inputs that go into the hidden layer neuron.

The activation function chosen to be connected to the output layer is the softmax activation function. The softmax function is a commonly used ouput activation function in the classification. Its advantage lies in that it outputs probability values between 0 and 1 for each of the output neurons. Additionally these values are computed in such a way that the sum total value in each output will be 1. This would later allow for the choice of a specific value of probability that allows most effective discrimination between signal and background. Additionally it comes in the form of an exponential and therefore will prodice on events that show are specific pattern. This will cause the difference between signal and background to be even more discernable. This function is represented by the following:

$$\sigma(a)_i = \frac{e^{a_i}}{\sum_{j=1}^{2} e^{a_j}} \tag{6.3.2}$$

where a_i represents the outputs from the final hidden layer neurons and the bottom of the fraction represents a normalization term which sums over the outcomes of each output neuron and ensures that the resulting term falls between 0 and 1.

6.3.2 Performance

The layout of the neural network was made such that it optimised performance. To check the performance of such a neural network they are tested against an independent test sample. This test sample is from the same source as the used training samples. However, unlike the training samples, the neural does not 'know' what the output should be for these samples. Hence they are referred to as test samples as they provide a genuine view on whether the training had any effect. In this study the test sample was about 20%

of the generated signal and background samples, the training sample was the remaining 80%.

The neural network was trained over 40 epochs and during every epoch the training and test samples were put through it. For each of these samples a value of *accuracy* and *loss* would be obtained. These two values are calculated after each epoch. The *accuracy* represents the prediction accuracy can make about the sample classes compared to their true class. This can as such be described as a percentage as well. The loss on the other hand is the final value of the loss function associated to this neural network. Note that this does not represent a percentage. In the case of the test sample these two values are often referred to as the *validationaccuracy* and the *validationloss*. To obtain an ideal performance two things are desired from both of these samples. First, both the loss and accuracy of the training and test samples are expected to converge to a specific value. If this were not the case, it would imply that the neural network can still reach a better precision by running more epochs. Lastly, accuracies and losses of the testing and training samples are expected to lie close to one another in the end. The reason for this, is that it is expected that results obtained by the training sample should reflect what the neural network would do to an actual sample. If one of these did not properly overlap it would mean that there is a case of underfitting or overfitting meaning that it either is not trained enough or it starts to see patterns in the training samples that do not exist in the testing samples. The latter case, can also occur when initially converged values start to increase for the case loss and decrease for the case of accuracy. This means that the neural network has been "overtrained".

These two criteria can easily visualised by plotting training curves. This curve shows either the accuracy or loss for both samples at each epoch. The curve belonging to this neural network can be seen in figure 6.17. From this plot, it can be seen that even before reaching 40 epochs that the samples already converged and overlapped converging at an accuracy value of 98.60% and a loss value of 3.64%.

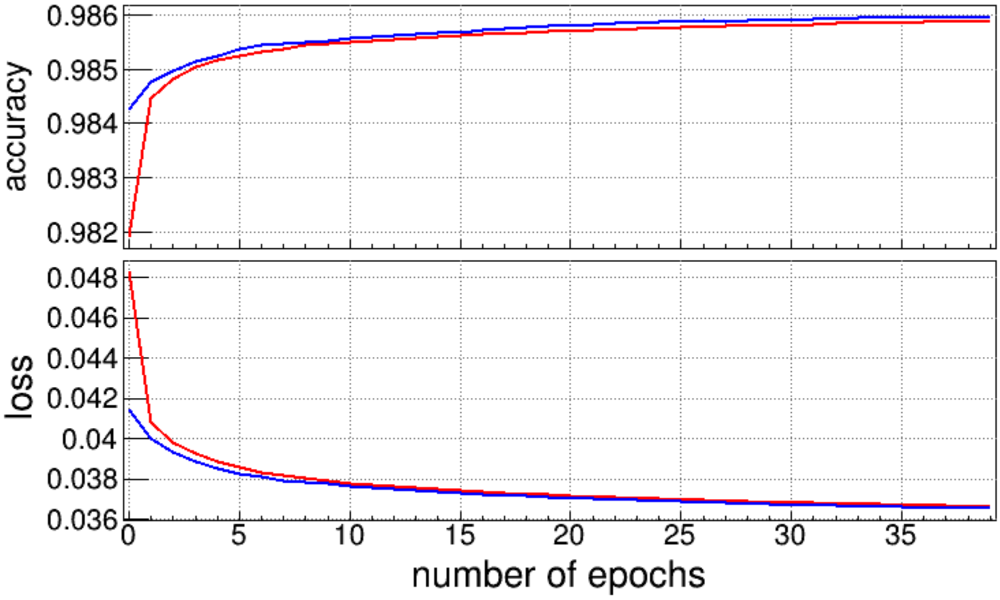


Figure 6.17: The training curve of the neural network. At each epoch the loss and accuracies are shown that both the test (blue) and training (red) samples obtain.

This training curve provides a good view on whether the neural network setup was achieved. It does not however, provide a view if the neural network is efficient. It should be noted that the size of the signal and the background samples used for training and testing is severely different. The training sample considered 80% of the generated samples. For the WW background this is about 6448000 events, and for the semi-leptonic $t\bar{t}$ sample this is about 240000 events. Similar for the test sample which combined the 20% the WW background contributed about 1612000 events and the semi-leptonic $t\bar{t}$ about 60000 events. As such it can very well be that the neural net is good in predicting the background, but lack in predicting the signal when using the test sample.

To account for this, two additional performance test were considered to see the effectiveness of the considered neural network and its used inputs. The first test was to see how the results of the test samples and training samples compared to one another. It is of course expected that they generally behave the same, but this is no guarantee. To compare these, the results of each output neuron from the training sample has been directly compared to the corresponding output of the test sample by overlapping there normalised histograms. By normalising and overlapping these histograms, one is able to see for example, whether or not the bins corresponding to a background signal overlap properly when considering WW events. If this were not the case this would be a clear sign of overfitting which was missed in the training curve. This test was performed for both the signal neuron as shown in figure 6.18a and the background neuron as shown in figure 6.18b. In these figures, the x-axis represents probability of the event being a signal event. Thus events with a probability close to 1 would be detected as signal while values close to 0 would be detected as background. From both these figure it can be seen that bins close to their respective expected probabilities overlap very well. However one could note that the background histogram bins do not overlap properly for probabilities close to 1 and that the bins of the signal histogram do not properly overlap for probabilities close to 0. In a sense this is not a problem as this could mean that the neural network is not good at predicting background events as signal and the inverse, which it was not meant to be good at. It could also mean that there is simply too little amount of background(signal) events that are predicted to be signal(background) such that it is easy for sudden deviations to occur. This further shows that the neural network performs adequately.

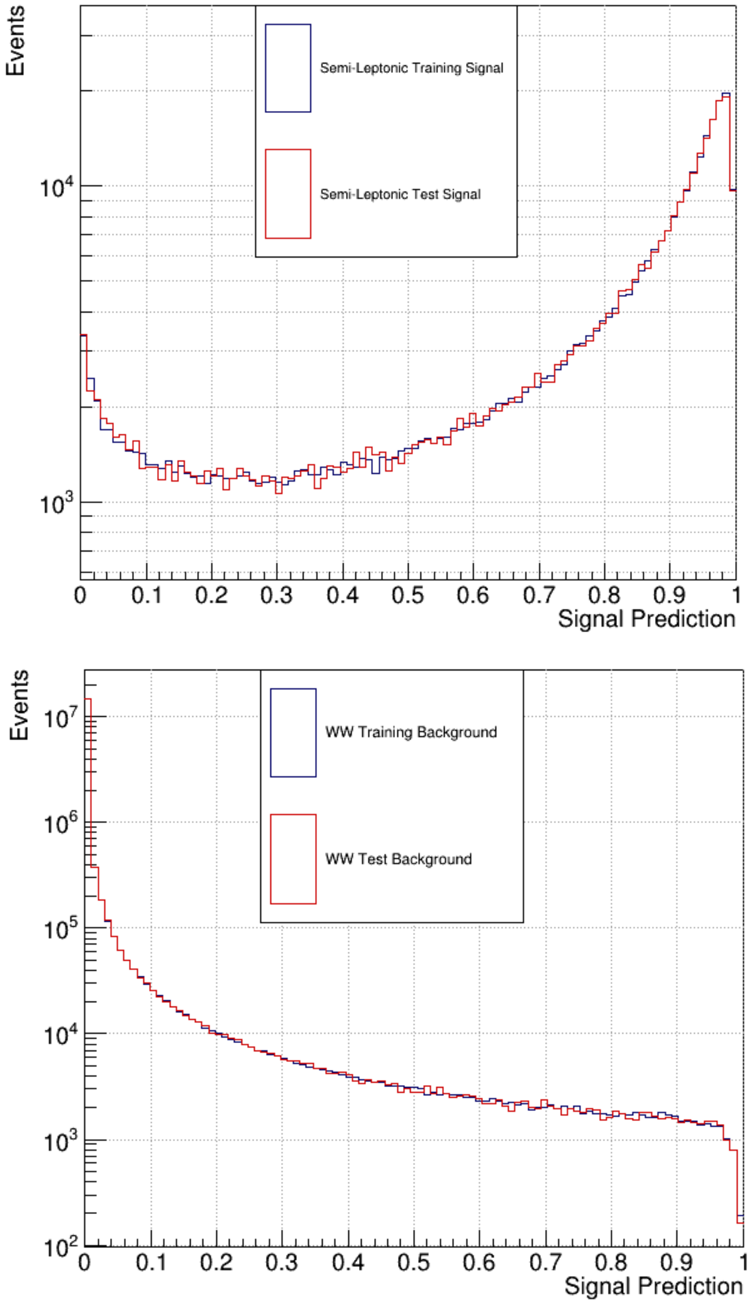


Figure 6.18: The probability distribution that represents the predictions made by the neural network for the signal sample (upper figure) and the background sample (lower figure). Here both the training and test sample are plotted as to study their overlap.

To conclude, using the results of the neural network, a probability value has to be chosen for which events with probabilities below that value are considered background and remover and probabilities above that value are considered signal events. This hard separation of events in to two separate classes results into so-called *hard labels* for the events. This decision can be right and wrong, but is often optimised in terms of the signal selection efficiency, ϵ_{TPR}, often referred to as the true positive efficiency, and the

background acceptance, ϵ_{FPR}, often referred to as the false positive efficiency. To optimise the probability these values are plotted against one another. This plot is referred to as a ROC curve. By then taking the integral corresponding to a maximised true positive efficiency and a minimised false positive efficiency, one is able to reach the most optimal value for this hard separation.
Because such a value is based on the efficiencies of separation, it is important that ROC curves of the training and test samples behave similarly. If this were not the case, then the neural network does not provide the most optimum probability value for which a hard separation could be made and has to undergo further training. For this reason the last performance test was done by comparing the ROC curves of both samples to each other. This resulted in figure 6.19. From this figure it can be seen that the largest deviations between the two samples are of the order of 10^{-4}. It can thus be concluded that the ROC graphs of these two samples overlap very well. Additionally, these ROC curves show already that the neural network provides a very good range for which a hard separation can be taken. For example, if it is chosen to consider a true positive efficiency of 0.8 it would only result in a false positive efficiency around the order of 10^{-3}. This shows that the signal and WW background can be well separated even if a probability value is considered close to 0.

From these various performance check of the neural network, it can be concluded that the used neural network works efficiently and provides good separation of signal and background. Of course further studies should be considered to improve upon this neural network as there are some points which could improve the separation significantly.

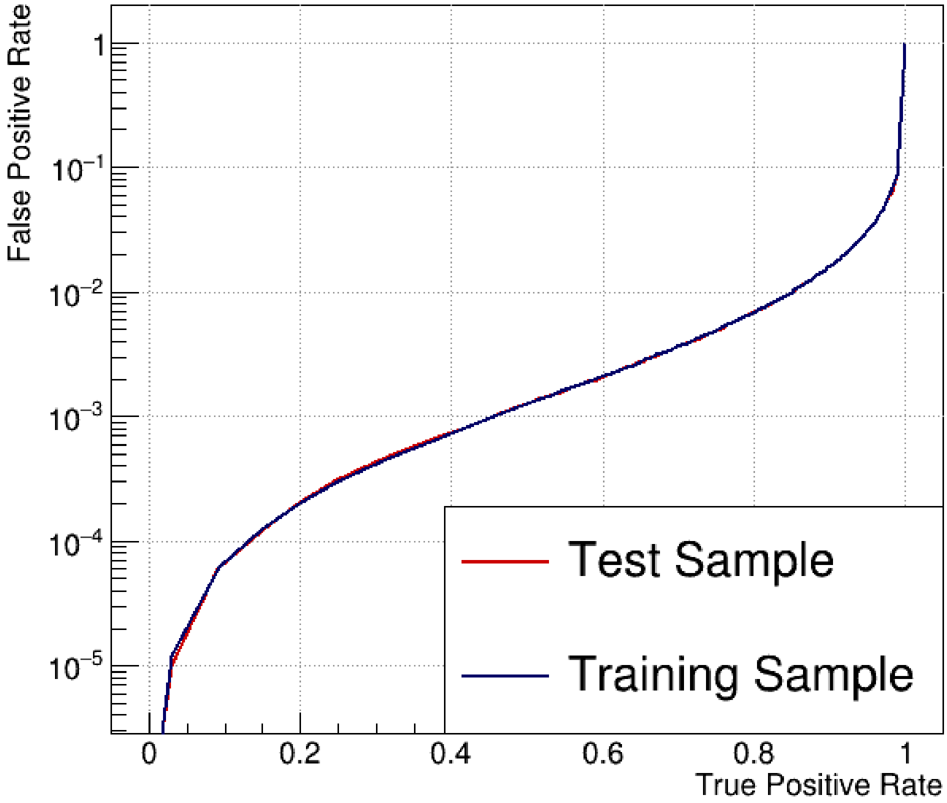


Figure 6.19: The ROC graph of the corresponding neural network performance. Here the red line represents the efficiencies related to the test sample, and the blue line the efficiencies related to the training sample.

6.4 Direct cuts & results

6.4.1 Neural Network Results

Using the trained neural network described in the previous section, each considered process, as presented in section 6.1, was put through this network in order to separate them from the signal. The output of the neural network is a probability of whether the provided event is more likely to be a signal event or a background event. By associating the probability of an event being signal to 1 and being background to 0, the following probability spectrum histogram shown in figure 6.20 can be made. To further study these separated events, a probability value needs to be chosen that would properly discriminate[2] background and signal. These probability values are always chosen such that they optimise the signal significance. The statistical significance of the signal is expressed as follows:

$$Sign = \frac{S}{\sqrt{S+B}} \tag{6.4.1}$$

---

[2] Because of this, this plot is often referred to as a discriminator plot.

where S and B represent the total signal and background events between the chosen probability value and 1 respectively. The probability value associated to the largest significance value then would consider the best signal to noise ratio of the provided events. Note though that equation 6.25 takes into account the statistical uncertainty and not the systematic uncertainty. A separate study in the future could aim carefully determine these systematic errors and further improve upon this algorithm.

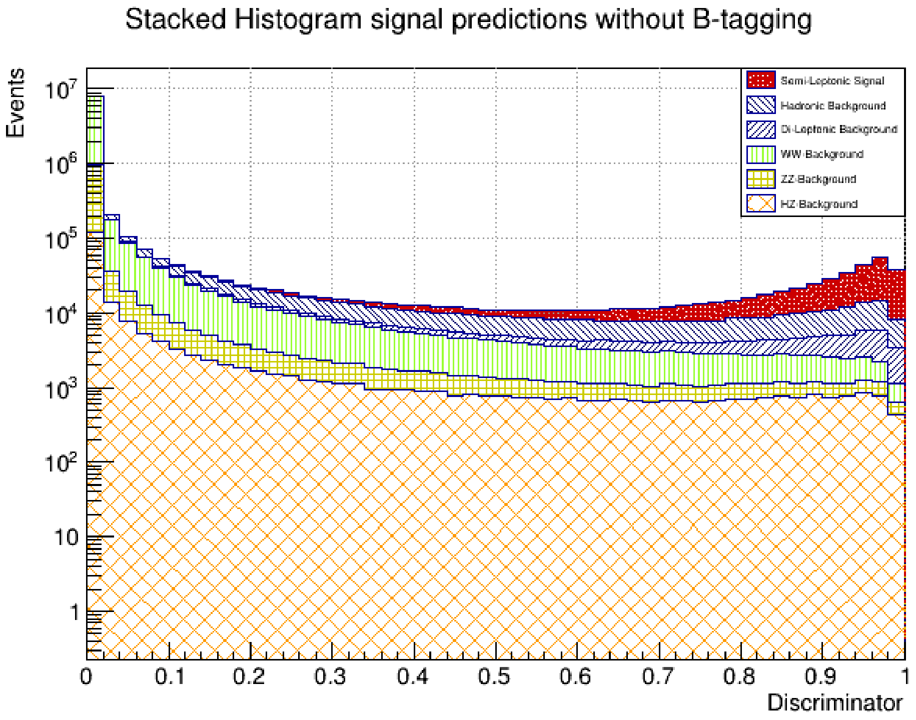


Figure 6.20: The probability spectrum that shows how the neural network discerned the various processes that would appear during e^+-e^- collision.

6.4.2 Direct Cuts

As mentioned in section 6.2, the neural network is not the only step taken in order to reduce the background noise. Additional direct cuts would be applied to the results. The properties for which these cuts are taken are: the b-tagging number and the amount of leptons in the final state. In this section the choice of how these cuts are represented are discussed. Additionally, the final optimal discriminator value that these cuts provide will determined and the amount of events that remain for each process will be shown. Giving the reader a clear view of how much signal this algorithm provides compared to the background.

B-tagging Cuts

By using Delphes its ability to simulate the b-tagging in the ILD detector, the b-tag associated to each event was kept track of. However, as discussed in section 4.1.3, the b-tagging algorithm is not perfect and carries a certain misidentification rate. Because of this different categories of b-tagging cuts were considered to compare which one would provide the highest significance. The categories that were considered were: $b - tag = 0$, $b - tag = 1$, $b - tag = 2$ and $b - tag \geq 1$. To provide a visual difference between the cuts, the probability spectrum associated to each of these cuts can be seen in figures 6.21-6.24. The best cut was chosen based on the associated optimal significance value. By choosing the significance value that is the largest of the four would provide a view on which one provides the best signal-to-noise ratio. The significance graph associated to these cuts can be seen in figure 6.25 where significance is plotted against the possible probability values that separate the signal from background based on the predictions provided by the neural network.

From the significance graph, it can be seen that the $b - tag \geq 1$ cut considers a maximal value of $Sign = 287.35$ which is a lot larger than those in the other categories by large margin. This means that the signal-to-noise ratio is at its most optimal for this category.

To get an idea of the actual number of events associated to a cut based solely on the b-tagging variable, table 6.3 shows the amount of events one would have left after such a cut.

Number of events ($L_{int} = 1.5\text{ab}^{-1}$)				
Process	b-tag = 0	b-tag = 1	b-tag = 2	b-tag $\geq$ 1
Semi-Leptonic $t\bar{t}$	34001 ± 151	91752 ± 248	64043 ± 207	163838 ± 332
Di-Leptonic $t\bar{t}$	9522 ± 91	29229 ± 160	24511 ± 147	57671 ± 225
Hadronic $t\bar{t}$	27204 ± 109	68256 ± 173	44443 ± 140	119449 ± 229
HZ	64273 ± 238	52021 ± 214	42177 ± 193	110508 ± 3112
ZZ	604283 ± 763	178471 ± 415	88296 ± 292	291679 ± 530
WW	13243360 ± 5147	1716598 ± 1853	95906 ± 438	1815332 ± 1905

Table 6.3: The remaining number of events after applying the associated b-tagging cut. The processes considered here are those created after the e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}.

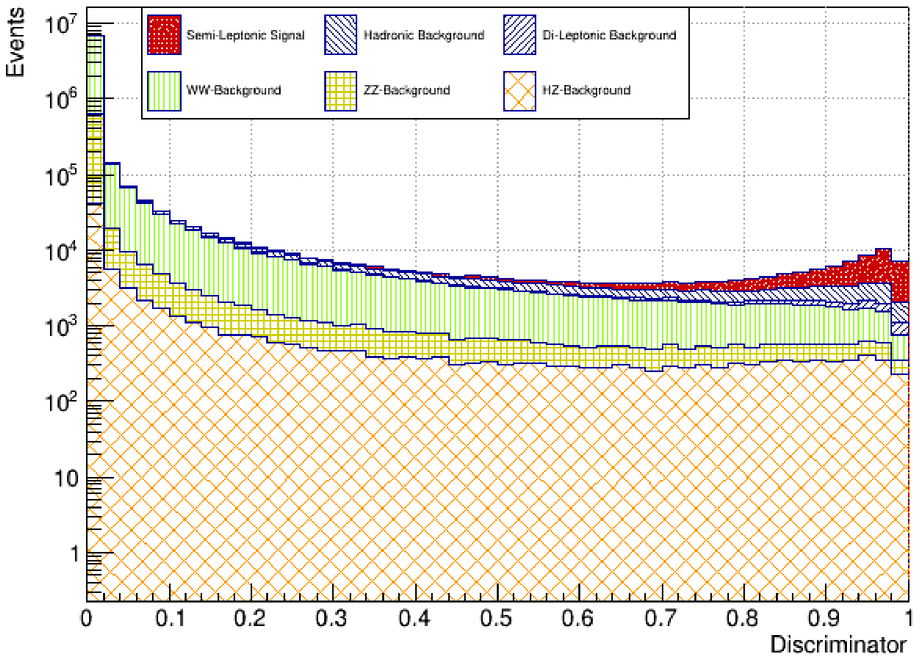


Figure 6.21: The probability spectrum of the processes after the going through the neural network and when a b-tag of 0 is required.

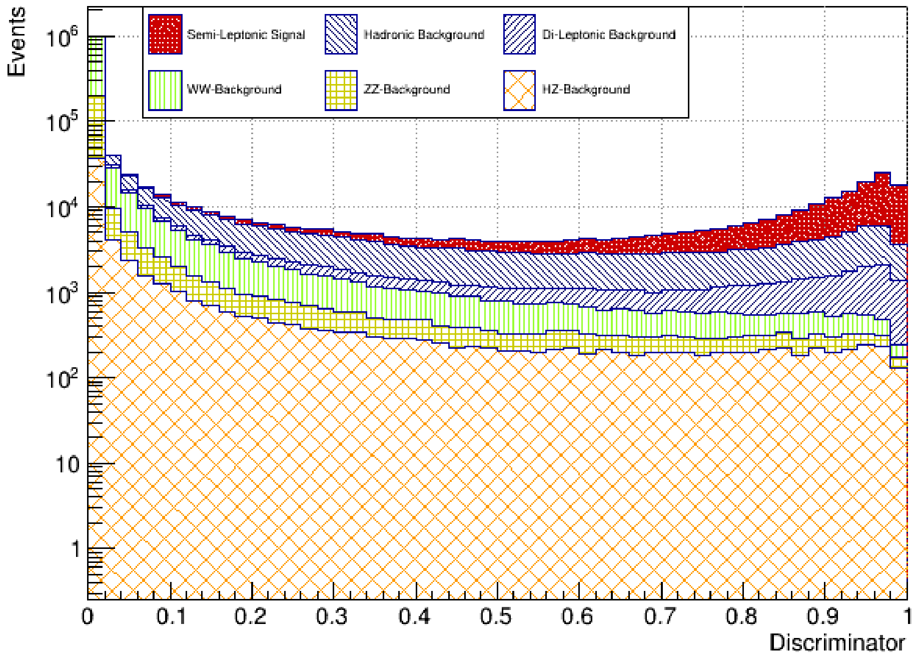


Figure 6.22: The probability spectrum of the processes after the going through the neural network and when a b-tag of 1 is required.

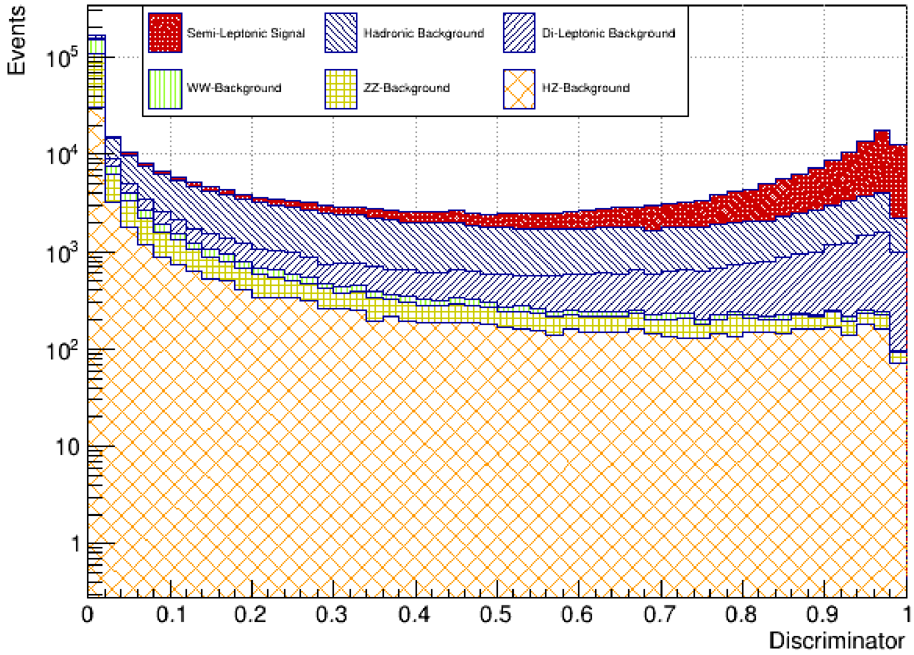


Figure 6.23: The probability spectrum of the processes after the going through the neural network and when a b-tag of 2 is required.

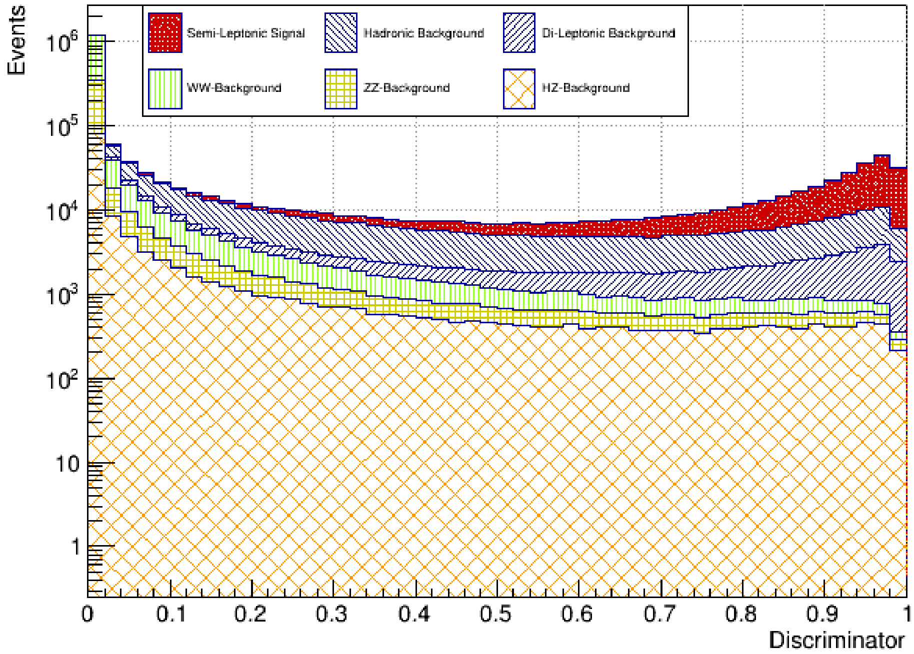


Figure 6.24: The probability spectrum of the processes after the going through the neural network and when a b-tag of larger or equal to 1 is required.

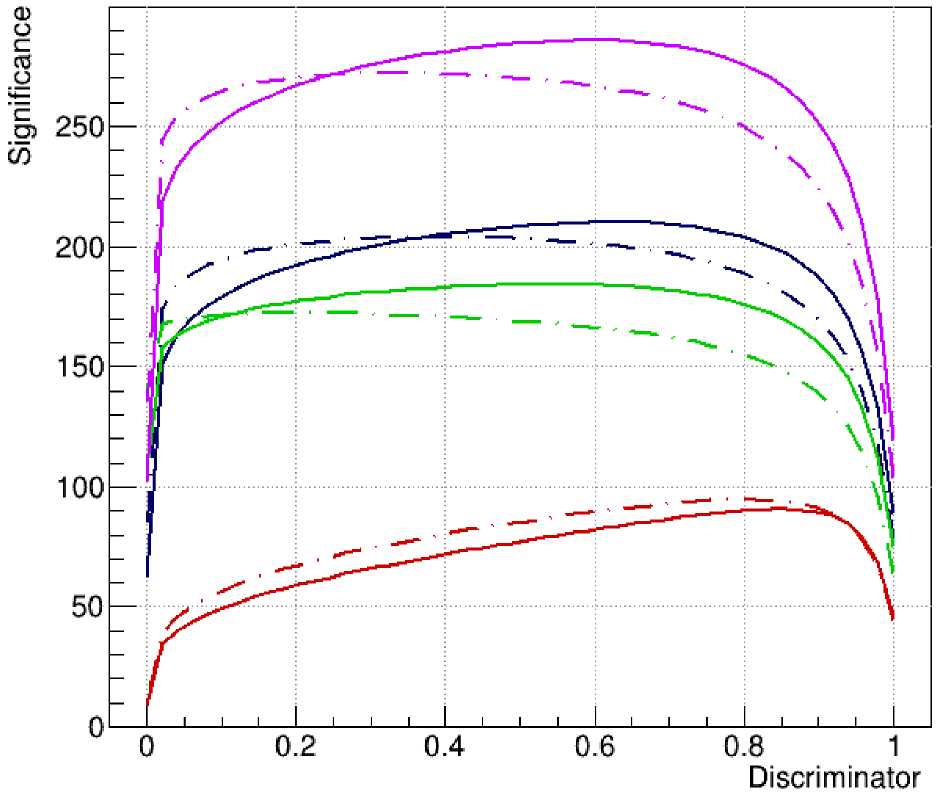


Figure 6.25: This plot shows the significance associated to each b-tagging cut (full lines) and the lepton cut (dashed lines) which follows up the b-tagging cut against their corresponding probability value or discriminator predicted by the neural network. The red line represents the b-tag = 1 category, the blue line the b-tag = 2 category, the green line the b-tag = 0 category and the violet line the b-tag ≥ 1 category.

Lepton Cuts

As explained in section 6.2, by enforcing the existence of a lepton in the final group of events, the lepton-independent nature is not lost to the used algorithm. Incidentally, the reason behind creating this algorithm is due to the fact Lepton Universality Violation, as discussed in section 2.4, hypothesises that the leptons should not be treated identically as the SM implies. To properly study this an algorithm is needed that does not takes these assumptions into account. However, in order to properly study, one would need to compare the behaviour of the resulting leptons against each other and thus it is required that they do exist in the end. Therefore, requiring the existence of only a single lepton, which the semi-leptonic $t\bar{t}$ process is expected to end with, is a good cut to finalise the algorithm. As such, the probability value which will be used as discriminator will be based on the optimal significance value calculated for this final cut.

Together with the chosen b-tagging cut and the lepton requirement, the final probability spectrum provided by the neural network can be seen in figures **??**-6.26. The optimal significance value was then chosen from the significance graph shown in 6.25. For a b-tagging of ≥ 1 category a probability value of 0.21 was determined to be optimal choice that provides the largest signal-to-noise ratio.

To get an idea of the actual number of events associated to the last cut, the amount before and after are present in table 6.4. Note that the b-tagging cut has already been applied at this point.

Number of events that were generated					
Semi-Leptonic $t\bar{t}$	Di-Leptonic $t\bar{t}$	Hadronic $t\bar{t}$	HZ	ZZ	WW
201600 ± 405	67320 ± 110	146055 ± 330	175950 ± 42	964200 ± 3150	16080000 ± 43500
Number of events after $N_{btag\geq 1}$					
163838 ± 332	57671 ± 225	119449 ± 229	110508 ± 3112	291679 ± 530	1815332 ± 1905
Number of events with the lepton cut applied					
100805 ± 260	27378 ± 155	400 ± 13	2 ± 1	14858 ± 120	458738 ± 958

Table 6.4: The remaining number of events after applying the lepton cut and b-tag ≥ 1 cut. Additionally the amount of events that were generated is also shown for comparison. The processes considered here are those created after the e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}.

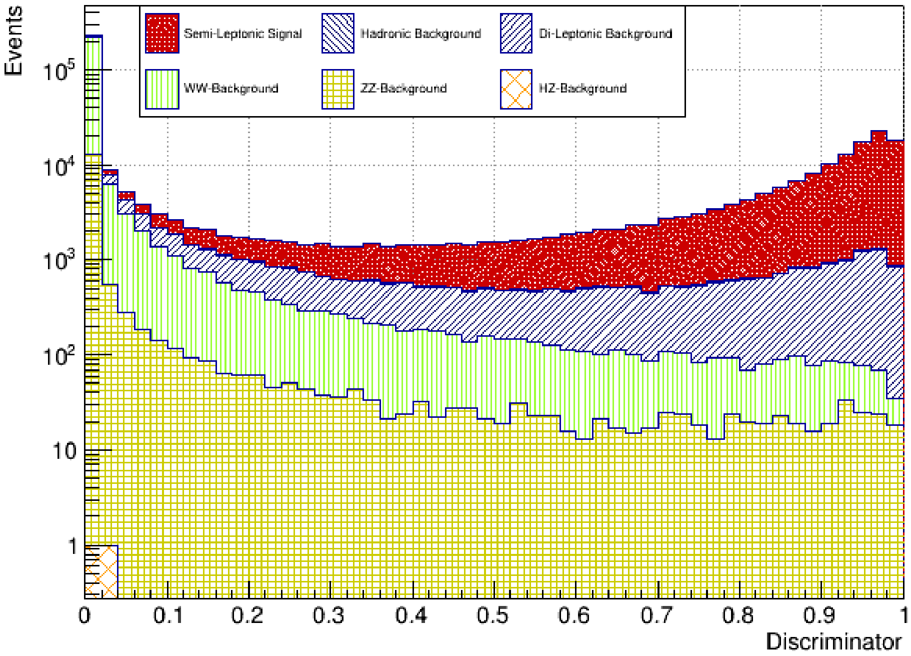


Figure 6.26: The probability spectrum of the processes after the going through the neural network and when a b-tag of larger or equal to 1 and a single lepton are required.

6.5 Final Result

Using the optimal probability discriminating value determined after the lepton cut, the final remaining of signal and background events could be determined. This is presented in table 6.5 against the the amount of event from which this started. It can be seen that after the cuts the signal retains about 49.81% of its original amount. Compared to this the Di-Leptonic background retains 22.93%, the Hadronic background retain 0.16%, the HZ background retain 0%, the ZZ background retain 0.080% and the WW background retain 0.088%. From this it can be learned that most of the background aside from di-leptonic get suppressed to even 1% of their original amount. In the end, the only background left to cause much noise are the Di-Leptonic and WW background which relative to the total events amounts to about 11.77% and 10.82% respectively.

Expected number of event without any cuts					
Semi-leptonic $t\bar{t}$	Di-leptonic $t\bar{t}$	Hadronic $t\bar{t}$	HZ	ZZ	WW
201600 ± 405	67320 ± 110	146055 ± 330	175950 ± 42	964200 ± 3150	16080000 ± 43500
Expected number of events after $N_{btag\geq 1}$ & $N_{Lep} = 1$					
100805 ± 260	27378 ± 155	400 ± 13	2 ± 1	14858 ± 120	458738 ± 958
Expected number of events after NN ≥ 0.21					
100516 ± 259	15435 ± 87	233 ± 8	0	752 ± 6	14184 ± 30

Table 6.5: The remaining number of events after applying the lepton cut and b-tag ≥ 1 cut. Additionally the amount of events that were generated is also shown for comparison. The processes considered here are those created after the e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}.

6.6 Tau Lepton Background

A proper detection of tau leptons comparable to its fellow leptons can considered a challenge for current detectors. As discussed in section 4.1.1 does not consider tau leptons in the isolation module. This means that in the end tau leptons will not be recognised by the program as leptons, but rather as jets. W boson have a chance of $11.38 \pm 0.21\%$ to decay in a tau lepton. Similarly, it has a chance of $10.71 \pm 0.16\%$ and $10.63 \pm 0.15\%$ to decay in a electron and muon respectively. This means that in general, there would be a rather large chance of the W bosons generated by the top quarks to decay into said leptons. Which can result into a lot of problematic background.

One could argue that the isolation variable should properly remove these types of background. However, using a Monte Carlo level tau-tagging variable, the semi-leptonic $t\bar{t}$ events corresponding to the decay of a tau lepton were kept track of before and after each direct cut. The resulting amount of events and percentages to the total can be seen in table 6.6. From this it can be seen that after the chosen direct cut of b-tagging ≥ 1

and requiring only 1 lepton, that there are still 9.86% of semi-leptonic events that do not correspond to the decay of the preferred lepton, in this case the electron, but come from τ decay instead. Because the sensitivity tests rely on the properties of the lepton that originate from the decay of the W boson, these tau lepton events were separated from the correct signal events and considered as additional background.

Total Semi-leptonic events	Semi-Leptonic events with Tau Leptons	percentage
100516 ± 256	9906 ± 17	9.86%

Table 6.6: The amount of semi-leptonic events for which the W boson decays into a tau lepton.

Chapter 7

Sensitivity of the $t\bar{t}$ Spectrum to New Physics

In this section variations to signal events were made and compared to the SM signal values through a statistical test. This test would then provide an idea of how sensitive $t\bar{t}$ production at lepton colliders is to NP. To this end, two BSM cases were studied to provide an overview of the sensitivity of the SM to NP. The first case considered general model independent deviations. This would provide a general overview of the sensitivity of the SM to variations that could occur due to NP. The second case considered a more specific model in BSM physics. Here the a hypothetical W' boson was added to the SM. In a similar approach model-independent variations were considered to provide an overview on the sensitivity of the SM including the W' boson.

For each of these cases a hypothesis test was performed. The null-hypothesis, H_0, for each test considered the associated SM model to be predict the observed distribution accurately. The alternate-hypothesis, H_1, for each case considered the model-independent deviations to not be able to be predicted by the SM. To perform this test, it was chosen to use a likelihood fit method with Poisson statistics. The fit used in this method would represent the scenario where the null-hypothesis is correct. The distribution which is fitted over considers the possible variations applied to it.

7.1 An initial remark

Before going into the sensitivity tests and the steps take for it, it is important to mention that from this point on, only the variations on signal events containing a electron are considered as opposed to just a lepton. The reason for this is due the large difference in behaviour from both the muon and the electron. Time constraints did not allow for both to be studied in depth at the same time. Future studies should be performed that also study the effects of the muon selection.

In the case of the SM generated events, table 7.1 shows the remaining amount of events when a electron cut is taken. Additionally, table 7.2 the same is shown for all the semi-leptonic $t\bar{t}$ events generated with the model considering the W' boson.

Number of events that were initially generated					
Semi-Leptonic $t\bar{t}$	Di-Leptonic $t\bar{t}$	Hadronic $t\bar{t}$	HZ	ZZ	WW
201600 ± 405	67320 ± 110	146055 ± 330	175950 ± 42	964200 ± 3150	16080000 ± 43500
Expected number of events after NN ≥ 0.21, $\mathrm{N}_{btag} \geq 1$ and $\mathrm{N}_{Lep} = 1$					
100516 ± 259	15435 ± 87	233 ± 8	0	752 ± 6	14184 ± 30
Number of events with cuts applied and forcing an electron					
45913 ± 92	8179 ± 13	77 ± 1	0	485 ± 2	4616 ± 12

Table 7.1: Summary of the amount of events after the cuts and forcing a single electron for each process created during e^+-e^- collision considering an integrated luminosity of 1.5 ab^{-1}.

Number of Events after NN ≥ 0.21, $\mathrm{N}_{btag} \geq 1$ and $\mathrm{N}_{Elec} = 1$				
W' Mass [GeV]	Cross-Section [fb]	Total events	Cross-Section for W' [fb]	Events with W'
110	259.11 ± 0.50	$85386 \pm 25 \times 10^{-2}$	130.12 ± 0.25	42495
120	214.30 ± 0.43	$69891 \pm 17 \times 10^{-2}$	85.49 ± 0.17	26976
130	181.26 ± 0.38	$58361 \pm 11 \times 10^{-2}$	53.89 ± 0.11	16254
140	158.05 ± 0.29	$51227 \pm 56 \times 10^{-2}$	$30.51 \pm 0.56 \times 10^{-1}$	8839
150	142.04 ± 0.26	$46917 \pm 27 \times 10^{-2}$	$14.74 \pm 0.27 \times 10^{-1}$	4113
160	132.32 ± 0.25	$44081 \pm 94 \times 10^{-3}$	$5.01 \pm 0.94 \times 10^{-2}$	1405
170	128.30 ± 0.24	$43408 \pm 21 \times 10^{-3}$	$1.10 \pm 0.21 \times 10^{-2}$	378
180	127.58 ± 0.25	$43189 \pm 84 \times 10^{-5}$	$0.43 \pm 0.84 \times 10^{-3}$	223
190	127.39 ± 0.25	$43239 \pm 52 \times 10^{-5}$	$0.27 \pm 0.52 \times 10^{-3}$	162
200	127.29 ± 0.25	$43069 \pm 33 \times 10^{-5}$	$0.17 \pm 0.33 \times 10^{-3}$	112

Table 7.2: Summary of the amount of events considering W' bosons after the cuts and forcing a single electron for each process created during e^+-e^- collision considering an integrated luminosity of 1.5 ab^{-1}.

7.2 Statistical test

In order to study NP, a property must be chosen that is the most sensitive to the considered NP. However, if NP was easily discerned then there would not be the need of such precision study. Often NP get detected through deviations of certain property values from SM values. However, due to the nature of particle collisions, no experimental outcomes are always the same even if the experiment is performed exactly the same. For this reason uncertainties are estimated within which the expected true values should lie.

As such the detection of NP is the size of deviation the detected property value's uncertainty has against that of SM calculated value. These are often referred to as σ-levels. In testing hypotheses a probability is often associated to the size of such a level. These probabilities allow us to state at which significance a hypothesis is rejected. In table 7.3, the first 5 common sigma levels are presented with their associated probability. From this table it can be seen that deviations around the 5σ level have probability of 0.0000002867 associated to them, which implies that this a 1 in 3.5 million of occurring. As such when deviations on that level occur than that means it is definite sign of NP having been observed. Additionally when detected values have a 3σ-level deviation, it is said that only then a possible sign of NP has occurred.

σ-level	Probability
1σ	0.1586552539
1.96σ	0.0249978951
3σ	0.0013498980
5σ	0.0000002867

Table 7.3: The probabilities associated to each of the first four considered σ-levels.

In this study it was chosen to use template fitting techniques. For the fit a likelihood fit using Poisson statistics was used through the usage of the software package *TFractionFitter* which is implemented within ROOT[2]. What follows is a very brief explanation of how this technique is applied.

Template fitting methods, in general, exploit the shape of a provided binned distribution for signal and background in combination with the number of events in each bin. To do this they require templates which would represent the expected signal and background distributions. By taking the sum of said templates, a linear combination is generated for each corresponding bin i. This sum is given by the following:

$$N_{Exp,i} = a_{Sign} \cdot N_{Sign,i} + a_{BG} \cdot N_{BG,i} \tag{7.2.1}$$

where $\mathrm{N}_{Exp,i}$ represents the number of events i^{th} bin of the distribution resulting from the fit, $\mathrm{N}_{Sign,i}$ the i^{th} bin of the signal distribution template, $\mathrm{N}_{BG,i}$ the i^{th} bin of the background distribution template. Additionally, a_{Sign} and a_{BG} represent the weights of each template to the sum. It are these weights that the fit will determine and can then be compared to a hypothesis.

During a fit, the program will calculate the likelihood value, $\mathcal{L}$, of the provided distribution fitted to the templates. Classically, the likelihood fit of a histogram is given by the following:

$$\mathcal{L}(N_{Obs,i}; N_{Exp,i}) = \prod_{i=1}^{n} p_i(N_{Obs,i}; N_{Exp,i}) \tag{7.2.2}$$

where $N_{Obs,i}$ is the number of observed events in bin i of the provided distribution and p_i represents the probability calculated from the associated statistics. In the case of *TFractionFitter* this is the Poisson Law given by:

$$p_i(O_i; E_i) = \frac{e^{-E_i} \cdot E_i^{O_i}}{O_i!} \tag{7.2.3}$$

After such a likelihood value has been calculated, the fitting procedure repeats the same calculation, but with different weights associated to the linear combinations of each bin. The newly calculated likelihood value is then compared to the previous after which the max likelihood value is taken as the preferred value. This process repeats itself until the likelihood value converges to a maximum after which the final value is considered to be the optimal to the given distribution.

The final result that *TFractionFitter* provides are fractional weights together with their uncertainties of each template that was initially provided. Additionally, the χ^2 of the fit can be calculated by the program through following:

$$\chi^2 = -2\log\left(\frac{\mathcal{L}(h_{Obs}; h_{Exp})}{\mathcal{L}(h_{Obs}; h_{Obs})}\right) \tag{7.2.4}$$

where h represents the associated distribution. Note that when the fit perfectly matches, a χ^2 value of 1 would be obtained.
To conclude, from these χ^2 values and the associated degrees of freedom used in the fit, a probability can be calculated which would define the consistency of the provided distribution to that of the templates. By considering the templates to represent the SM and the distribution the variation of SM which would contain NP, one can effectively determine the sensitivity of the SM to the presence of NP. This is, in the end, how the sensitivity was tested for the variations considered further in this chapter.

7.3 Sensitivity Probe

As seen in the previous section, to study NP it is important that properties used to discern NP have a high precision for their detected value and calculated SM value. Which in turn requires them to have probe that improves upon this.

In this study, the property under consideration for both sensitivity tests was the cross-section, σ_{Cross}. The probes under consideration for this property were the momentum of the lepton, P_{Lep}, and its transverse momentum, $P_{T,Lep}$. However not both were used and as such the two were compared to one another. The cross-section is calculated using the equation:

$$\sigma_{Cross} = \frac{N}{L_{int}} \tag{7.3.1}$$

where L_{int} is the integrated luminosity, which is 1.5 ab^{-1} for the FCC-ee, and N the total expected events. This provides an idea of how sensitive each probe is to small variations to the total produced events.

Two criteria were used to decide on which probe would be used. The first, most logical, criterion is that the probe must be sensitive in the desired region. The second criteria is that it is sensitive in the small momenta regions. The reason for this criteria is due to it being hypothesised that if NP were to be apparent, that it would do so in the low regions of momenta.

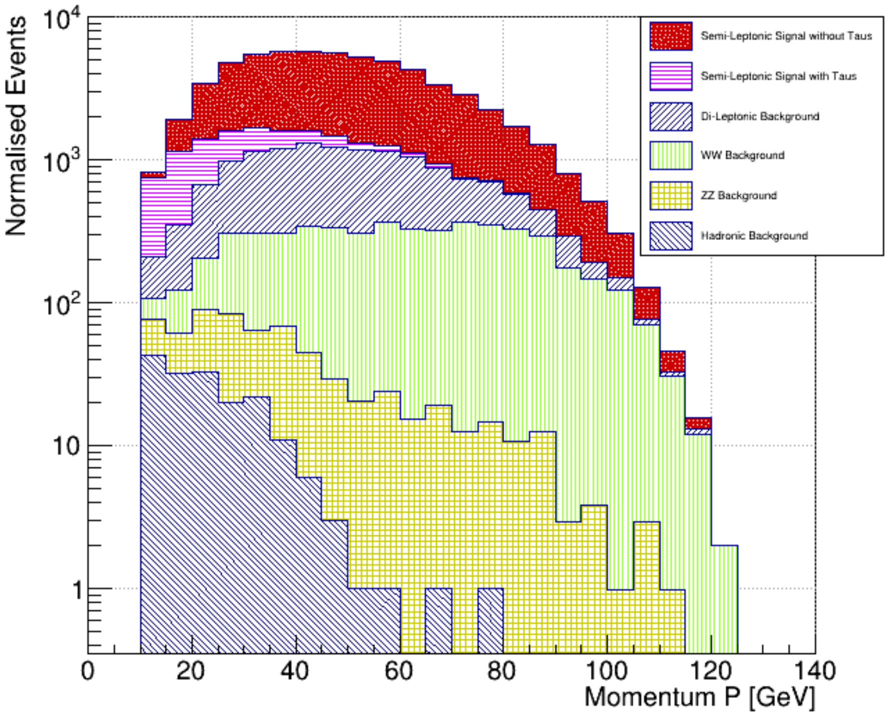


Figure 7.1: The stacked histogram of the P distribution of the electron for every process generated in the e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}.

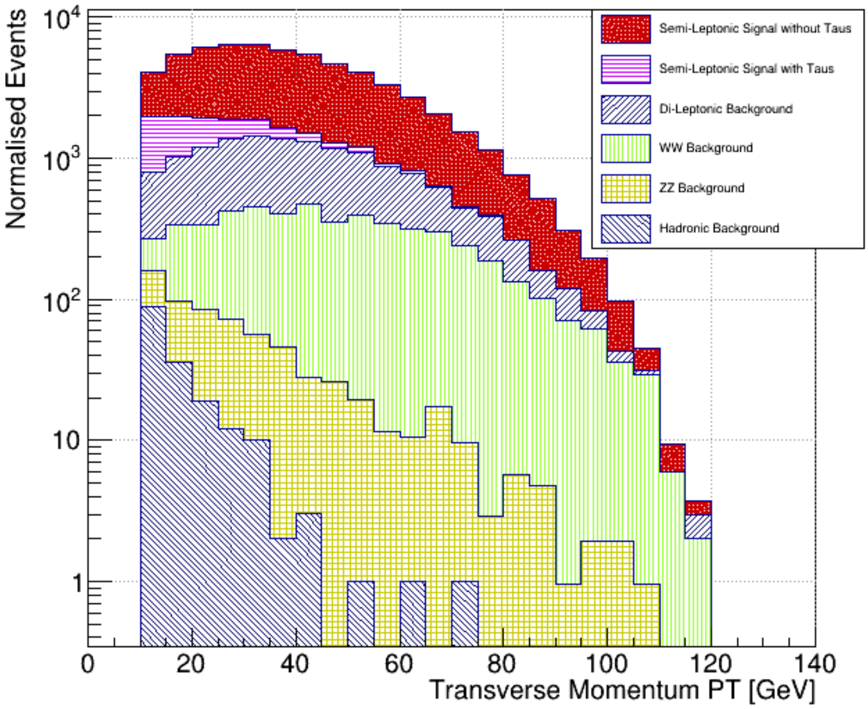


Figure 7.2: The stacked histogram of the PT distribution of the electron for every process generated in the e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}

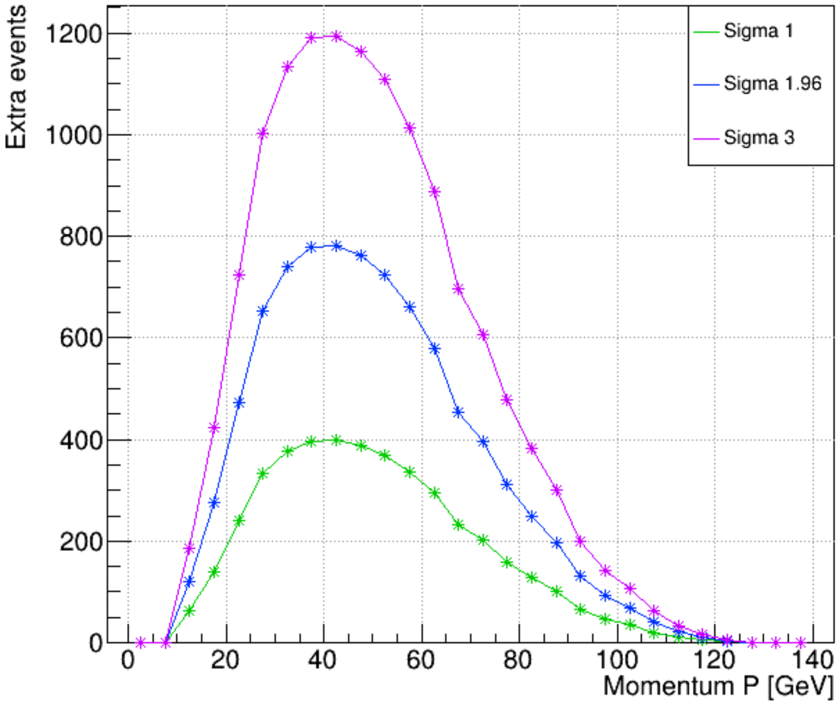


Figure 7.3: The plot shows the extra events that were calculated based on the statistical and systematic uncertainty associated to each momentum bin. The curves define extra signal events required to fall outside a specified CL when specific momentum region varied. In this plot the 1σ (68.7% CL), 1.96σ (95.4% CL) and 3σ (99.7% CL) are shown.

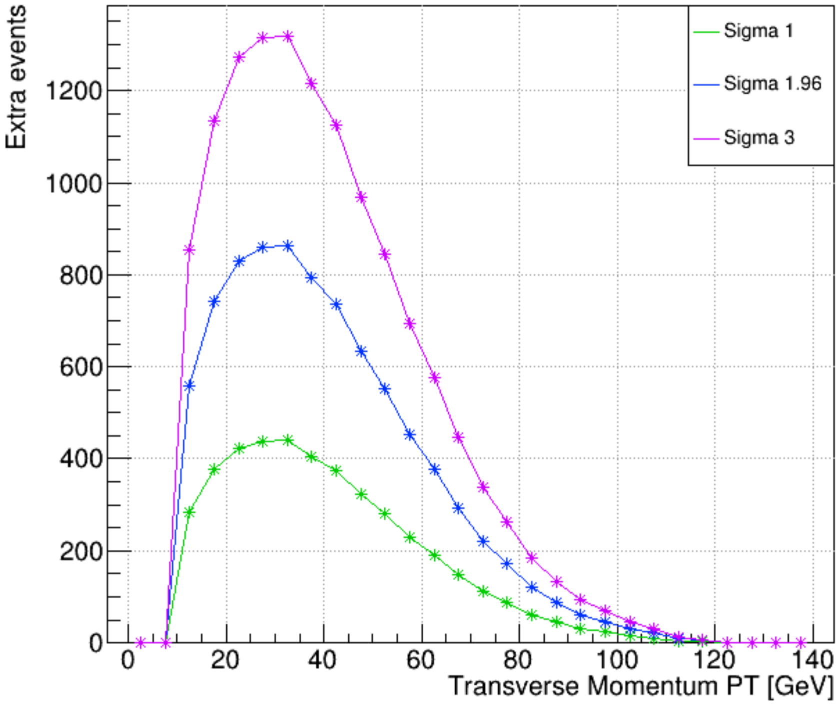


Figure 7.4: The plot shows the extra events that were calculated based on the statistical and systematic uncertainty associated to each transverse momentum bin. The curves define extra signal events required to fall outside a specified CL when specific transverse momentum region varied. In this plot the 1σ (68.7% CL), 1.96σ (95.4% CL) and 3σ (99.7% CL) are shown.

The momentum distribution of the lepton can be seen in figure 7.1 and the transverse momentum distribution of the lepton can be seen in figure 7.2. It should be noted that in this plot, the events originating from a W boson decaying to a tau lepton have been removed. The check for the criteria the statistical nature was checked for the lower regions of momenta. The reason for this, is because it would provide an idea of how much variation would still be within a given statistical confidence level of the SM sample. By checking the uncertainties associated to each confidence level, one is given an idea of how sensitive each region is to certain amounts of variations. The lower the uncertainty the more likely the SM momentum distribution is sensitive to relatively small variations compared to those of higher uncertainty.

For this the confidence levels of 68.7% (σ), 95.4%(1.96σ) and 99.7%(3σ) were considered. The uncertainties consider only the statistical uncertainty associated to each bin and the systematical uncertainty of 5% which was put upon the cross section calculated by MadGraph. From these values, confidence intervals (CL) could be calculated for each (transverse) momentum bin. By then using the expected integrated luminosity of 1.5 ab^{-1} and equation 7.3.1, it was possible to calculate the upper limits of the confidence intervals in terms of added extra events for each bin in the histogram. The results of these calculations are depicted in figure 7.3 for the P distribution and figure 7.4 for the P_T distribution. From these graphs it can visually be seen that the momentum distribution promotes a higher sensitivity to event variations in the lower value of the momentum when compared to the graph depicting the transverse momentum distribution. This is mainly due to the fact the electrons are selected with a transverse momentum < 10 GeV. Due to the transverse momentum being a component of the momentum it can only achieve lower values than the momentum itself. As such it will tend to gather a lot more events there than higher regions.

The uncertainties found in the CL graphs were studied for the six first bins. The results can be seen for the momenta in table 7.4 and for the transverse momenta in table 7.5. From these tables it can be seen that, for example, events with a momentum between 10-15 GeV only require a minimum 185 extra events 99.7% confidence that it is not withing the statistical uncertainty of the SM. Those of the same transverse momentum value require 854 extra events to lie outside the same CL making them a lot less sensitive. Additionally, for each momenta up to 40 GeV the amount extra events required to overcome the 99.7% confidence level is way lower than those of the transverse momenta. As such it can be concluded that the momentum provides a more sensitive probe into NP as it requires a lot less extra events to provide a clear signal that leads to the rejection of null-hypothesis. In conclusion, it was chosen to further study the sensitivity of the $t\bar{t}$ spectrum using the lepton momentum, P_{lep}, as a probe.

Extra events in the specified momentum region required			
Momentum Region [GeV]	68.7% CL (1σ)	95.4% CL (1.96σ)	99.7% CL (3σ)
10-15	62	121	185
15-20	141	276	423
20-25	242	474	725
25-30	334	654	1002
30-35	378	740	1133
35-40	397	778	1190

Table 7.4: Summary of the required extra events in a specified momentum region that defines the upper limit of the associated confidence level CL.

Extra events in the specified transverse momentum region required			
Transverse Momentum Region [GeV]	68.7% CL (1σ)	95.4% CL (1.96σ)	99.7% CL (3σ)
10-15	285	558	854
15-20	378	741	1134
20-25	424	831	1272
25-30	438	859	1315
30-35	440	862	1319
35-40	405	793	1214

Table 7.5: Summary of the required extra events in a specified transverse momentum region that defines the upper limit of the associated confidence level CL.

7.4 Model-independent study of the electron momentum in $t\bar{t}$ production

Because this study was made to improve upon detecting inconsistencies in lepton from $t\bar{t}$ production at lepton colliders, it is interesting to see when variations provide signs of NP. As explained in section 7.2 the maximal log likelihood was used as a statistical test to determine the degree of inconsistency that such small variation would bring, the electron momentum is used as a probe to study the cross-sections associated to the semi-leptonic $t\bar{t}$ events. The momentum distribution of the all the processes is the figure of merit. This distribution as expected from only the Standard Model is visualised in figure 7.5.

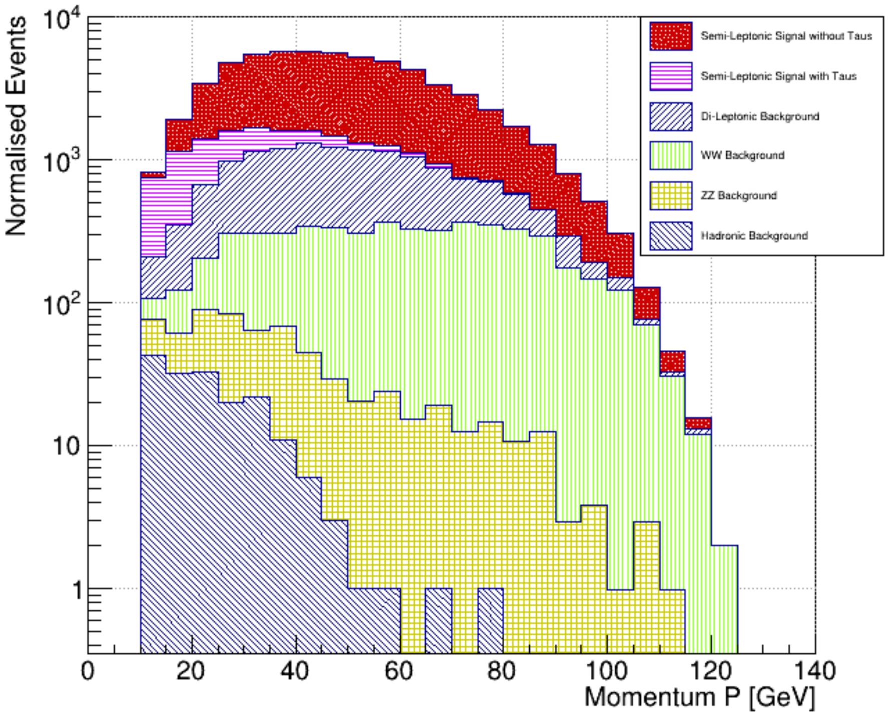


Figure 7.5: The stacked histogram of the P distribution of the electron for every process generated in the e^+-e^- collision at an integrated luminosity of 1.5 ab^{-1}.

Three different contributions were considered: the signal template, the $t\bar{t}$ background and the other non-$t\bar{t}$ backgrounds. These templates are visualised in figures 7.6, 7.7 and 7.8. These contributions represent the null-hypothesis, H_0, associated to the statistical test. As mentioned in the introduction, this null-hypothesis represents the claim that the SM predicts the cross section accurately.

The alternate hypothesis, H_1, on the hand states that there is NP that the SM does not predict. H_1 is represented by the variations on this SM distribution. Therefore an additional contribution is considered that represents these variations. This contribution comes in the form of extra events in a specific momentum bin on the signal template. The reason for picking a specific bin to add events to, is due to the fact that this would provide an idea of how sensitive the predicted SM distribution to possible NP around certain regions. By doing this for all the considered regions separately, this would provide a visual idea on how the sensitivity changes for increasing momentum.

As such, by considering the H_1 distribution as the "observed" distribution and the H_0 distribution as the "expected" distribution which the test uses as a fit, one is able to apply a proper hypothesis test on these possible variations. The variations applied to the events in the momentum bins ranged from 1 event to 1000 events with jumps of 20. Note that the jump choice was arbitrary and the reason to limit the variations to 1000 events will be later discussed. Additionally, the bins that were considered for variations, were limited to the bins associated to events with a lepton momentum between 75-80 GeV and lower. The reason for this can be seen in the signal template given in figure 7.6. In section 7.3 it was already explained briefly that the low momentum regions are of main interest. However, to get a good view of how the sensitivity changes over different regions momentum, it might be interesting to include regions which consider a large number of events. This region is mostly around 30-50 GeV. By then choosing to limit this to 75-80 GeV and lower no unnecessary points are included.

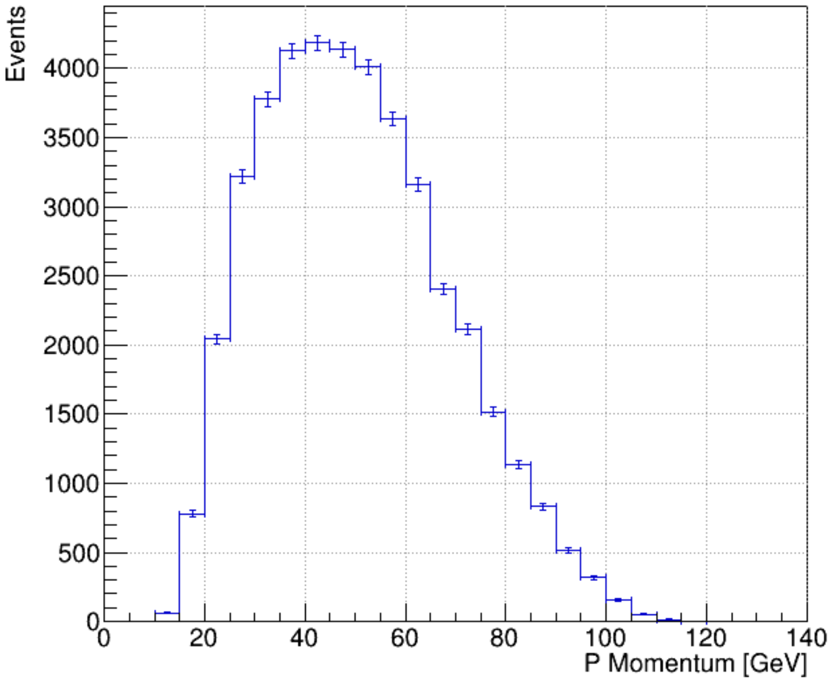


Figure 7.6: The momentum distribution of $t\bar{t} \rightarrow l + jets$. This is the template that would be used in the fitter.

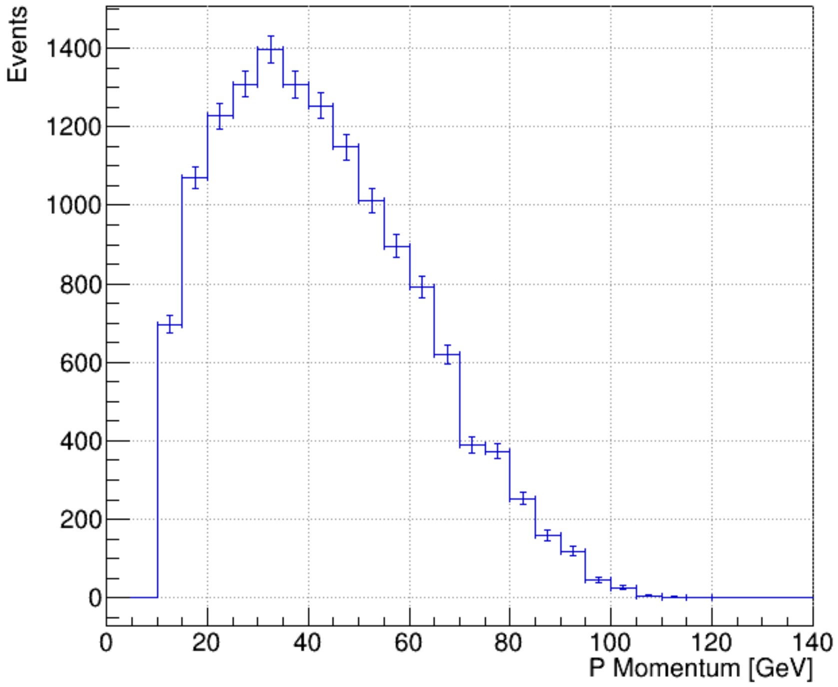


Figure 7.7: The momentum distribution of $t\bar{t} \rightarrow other$. This is the template that would be used in the fitter.

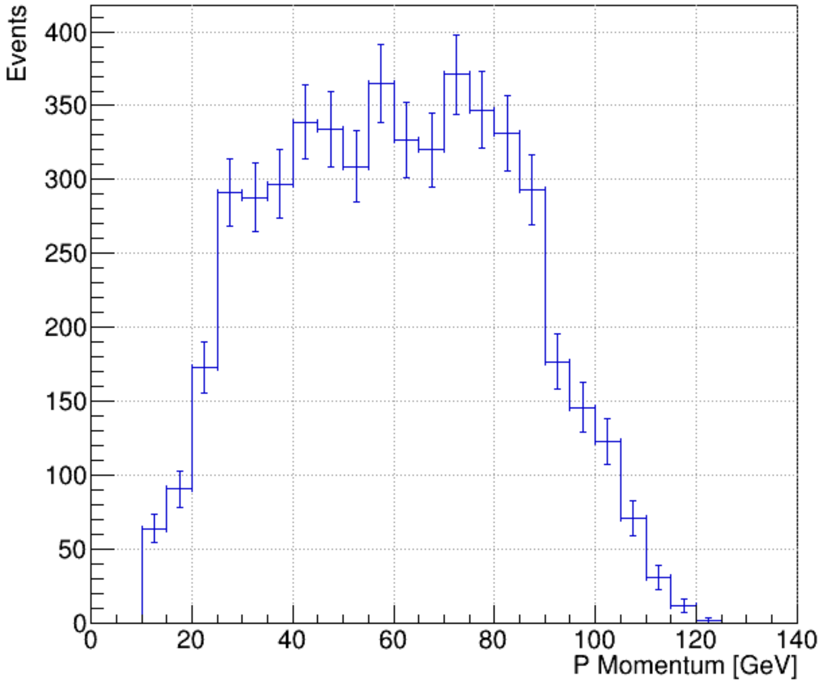


Figure 7.8: The momentum distribution of the remaining background (non-$t\bar{t}$) events. This is the template that would be used in the fitter.

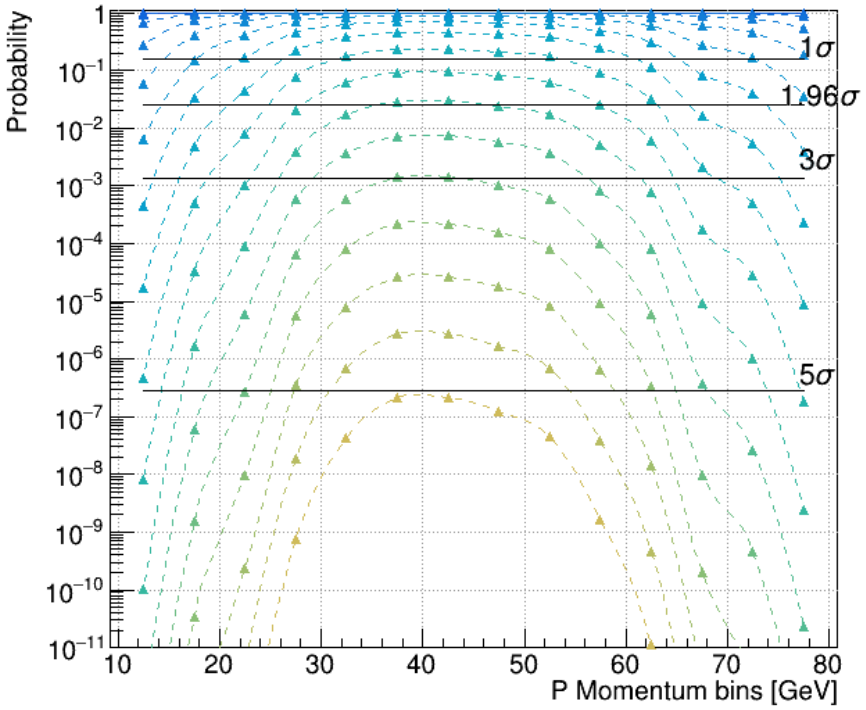


Figure 7.9: The plot shows the probability of every variation considered in every momentum bin as obtained from the likelihood ratio fit. The amount of variation is associated to color the point has going from no variation (dark blue) to a variation of 1000 events (yellow) with steps of 20. Additionally the appropriate σ-levels are shown.

From *TFractionFitter*, the χ^2 was obtained for each variation in each bin. By using the fact that this fits contained 29 degrees of freedom, the probability of the consistency of a hypothesis relative to the SM values was calculated and the results can be seen in figure 7.9. In this figure, every point represents an alternate hypothesis H_1 considering a certain amount of extra events in given momentum region. The probabilities that are given on the y-axis represent the consistency with the null hypothesis, H_0. It should really be stressed that these probabilities do not represent p-values of the hypotheses, but rather the probability which represents the associated the likelihood ratio test statistic value. As such by looking at the value of said probability, one can estimate to which significant level the H_0 can be rejected It is for this reason the plot also shows the various significant σ-levels such that a rejection relative to certain significant levels can be made. For example, points falling below the 3σ line represent H_1's that reject the H_0 at a significant level of 3σ.

The variations that each point represents are color coded based on the added extra events variation steps discussed above. The dark blue colors represent low variations starting from no added events, which in turn just represents a SM distribution. As this color becomes lighter the variations increase ending at a variation by adding 1000 events in a single momentum bin when the color becomes yellow. The best way to see this is through an example. Looking at the momentum region of 40-45 GeV, it can be seen that the lightest yellow point, considering a variation of 1000 added extra events, is the only point that considers 5σ-level significance. Because the variation steps where taken per 20 added extra events, going to points with higher probability in the same bin, corresponding to darkening of the color, will decrease number with 20 every point going up this bin. So by going up 4 points one reaches the first point that considers a 3σ-level significance. Since this goes up 4 this point would correspond to a H_1 which considers a variation of 920 extra events in the 40-45 GeV momentum region. Before calculating the cross sections for these variations, a few things can be learned from this graph.

By comparing the graph to the signal templates used for the fitting procedure, one should see that it carries a similar shape. It requires a the maximal amount of variations at momentum regions that contain the most events. This is rather logical due to the fact that here the uncertainty on the SM events is relatively larger compared to the low momentum regions of 10-20 GeV and the higher momentum regions of 70-80 GeV as seen in figure 7.6. This would imply, due to statistical uncertainty which the statistical test takes into account, that there is a need for larger variations in these specific momentum regions if they were to reject H_0 at certain significant levels. As such larger variations are required to give a clear signal of NP at momentum regions where the amount of events are already large. This is not the case in the lower momentum regions and high momentum regions where the number of events are relatively low. Here the statistical uncertainty is relatively low and therefore the SM calculations do not expect much extra events corresponding to these momenta values. As such these regions, as seen in figure 7.9, appear to reach higher significant levels for a lot smaller variation compared to the other regions. This makes these areas more likely to reject H_0 for small variations of the SM and as such more sensitive to NP.

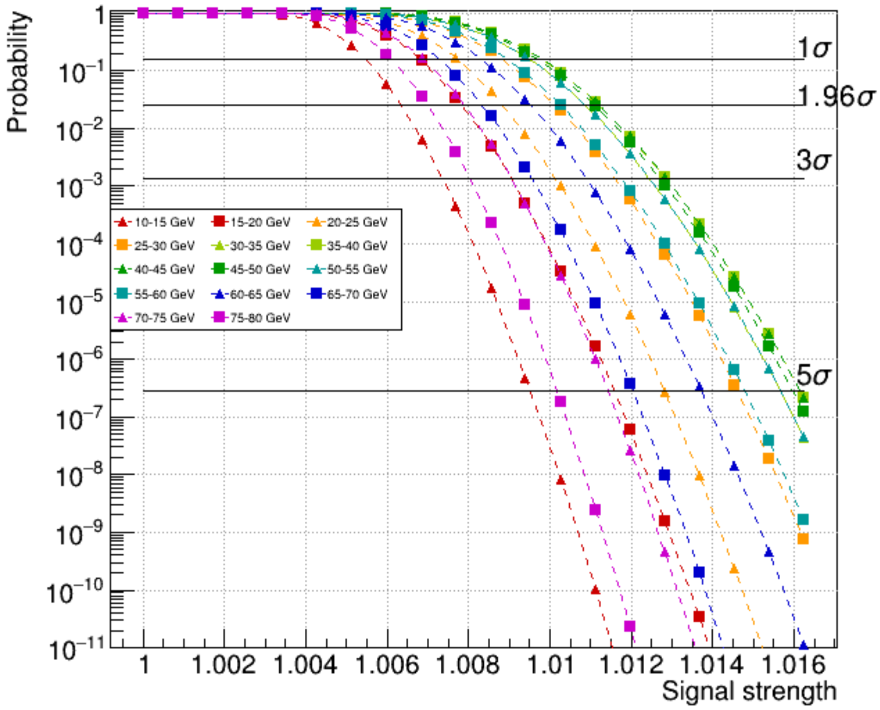


Figure 7.10: The plot shows the probability of the likelihood ratio to reject H_0 associated to a significance level which is represented by the horizontal lines in the plot. This is shown for each considered momentum bin.

Using the formula for the cross sections seen in equation 7.3.1, it was possible to determine each cross section associated a specific variation. The result of these calculations can be seen for most variations in figure 7.10 and for the specific σ-levels associated to each momentum region in table 7.6. Note that the x-axis in figure 7.10 does not represent the pure cross-section of each variation on the semi-leptonic $t\bar{t}$ signal, but rather its ratio to the SM generated cross section which considered no variations. As such it is better to refer to it as the signal strength. This would allow us to see by what fraction the original amount of events had to differentiate from its original value such that it reached a certain σ-level of inconsistency. The figure is also color coded. The colors represent momentum regions with values between multiples of 10, for example red represents the region between 10 and 20 GeV. The shape of each curve then represents which half of this region it covers. The triangles cover the lower half and the squares the upper half. For example the red triangles cover 10-15 GeV and the squares cover 15-20 GeV.

Similar to what already was seen from figure 7.9, is that the areas containing the maximal amount of events require by far the largest amount of variation to reach a specific σ-level significance. But additionally, it can be seen that the regions between 30 and 50 GeV require a similar amount of signal strength to achieve the associated σ-level significance to reject the H_0. This is quite important additional conclusion. It technically would imply that the minimum required amount of signal strength to reject the H_0 for a specific significance level can be defined by the amount of signal strength that the momentum regions of 30 and 50 GeV require. From this fact it can be concluded for example that to reject H_0 at a significance level of 3σ a signal strength of minimally 1.01286 is required.

signal strength associated to σ-levels				
Momentum region [GeV]	1σ-level	1.96σ-level	3σ-level	5σ-level
10-15	1.00558	1.00654	1.00757	1.00975
15-20	1.00681	1.00795	1.00924	1.01184
20-25	1.00772	1.00899	1.01022	1.01282
25-30	1.00891	1.01019	1.01177	1.01471
30-35	1.00952	1.01096	1.01261	1.01591
35-40	1.00986	1.01128	1.01286	1.01621
40-45	1.00986	1.01128	1.01286	1.01621
45-50	1.00975	1.01110	1.01277	1.01615
50-55	1.00952	1.01097	1.01261	1.01591
55-60	1.00903	1.01028	1.01187	1.01599
60-65	1.00836	1.00961	1.01103	1.01382
65-70	1.00736	1.00844	1.00972	1.01217
70-75	1.00690	1.00804	1.00925	1.01174
75-80	1.00616	1.00715	1.00830	1.01025

Table 7.6: A summary of the signal strengths associated to a specific σ-level for each considered momentum bin.

7.5 Sensitivity of the $t\bar{t}$ spectrum to a W' Boson

Aside from a general model independent variation of the SM, it is also interesting to perform a similar test for BSM theorised model. This study considered an additional sensitivity test which bases itself on such a theorised model. The model under consideration here, is the one where a W' boson is considered which has been introduced earlier in the theory section 2.5.

The tests were performed for 10 different W' mass hypotheses going from 110 GeV to 200 GeV. Every set was generated considering a mix of events that are predicted by the SM and events that contain W' bosons. The lower limit of the W' boson mass was chosen because experiments in LEP II showed that $M_W > 105$ GeV[30]. The upper limit was chosen because for values higher than 200 GeV there is no change in the $t\bar{t}$ determined cross section with respect to the SM. In figures 7.11 - 7.20 the P distributions are shown for every mass of the W' boson. Note that for increasing mass of a W' boson the W' yields are getting lower and lower.

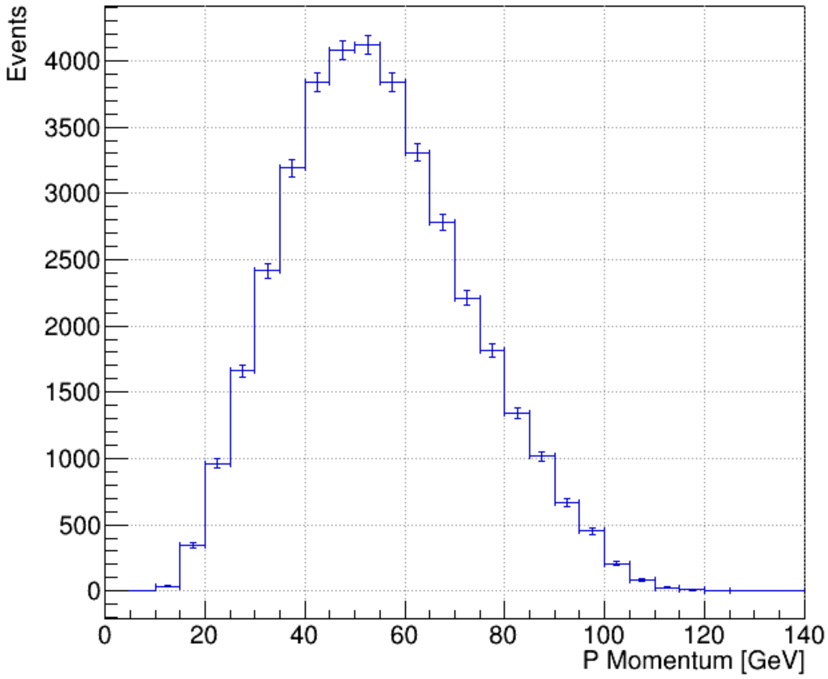


Figure 7.11: The momentum distribution of the semi-leptonic $t\bar{t}$ processes that contain a W' boson with a mass of 110 GeV.

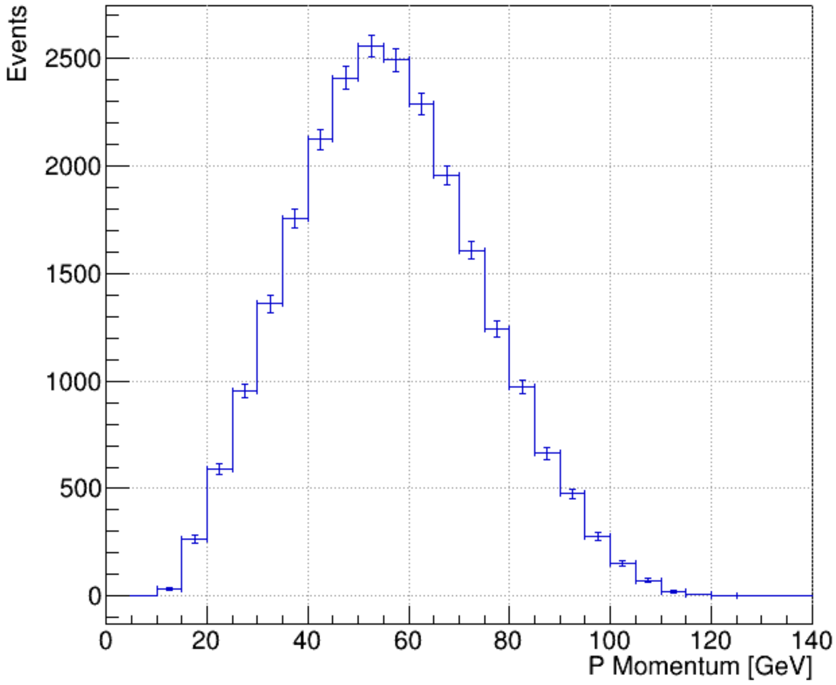


Figure 7.12: The momentum distribution of the semi-leptonic $t\bar{t}$ processes that contain a W' boson with a mass of 120 GeV.

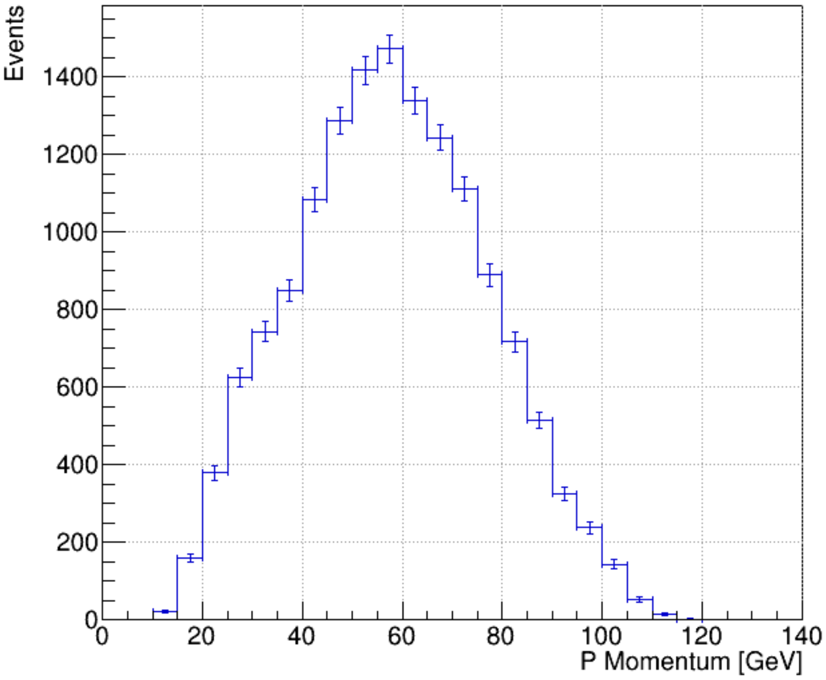


Figure 7.13: The momentum distribution of the semi-leptonic $t\bar{t}$ processes that contain a W' boson with a mass of 130 GeV.

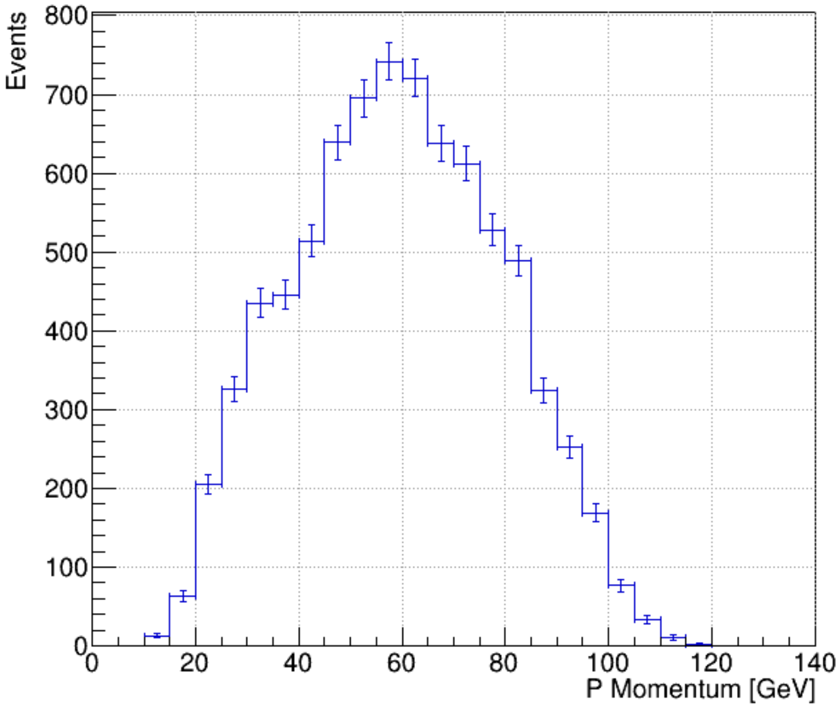


Figure 7.14: The momentum distribution of the semi-leptonic $t\bar{t}$ processes that contain a W' boson with a mass of 140 GeV.

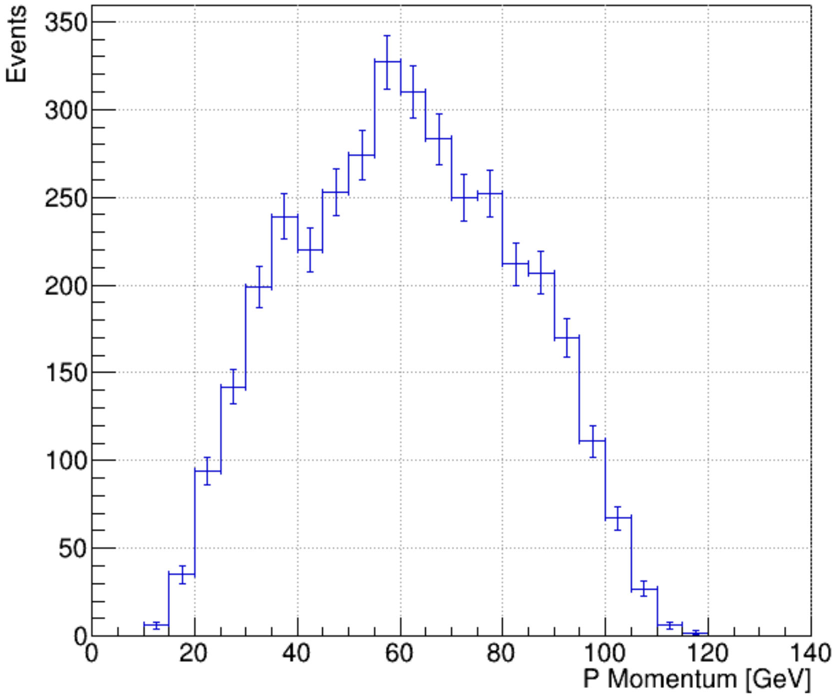


Figure 7.15: The momentum distribution of the semi-leptonic $t\bar{t}$ processes that contain a W' boson with a mass of 150 GeV.

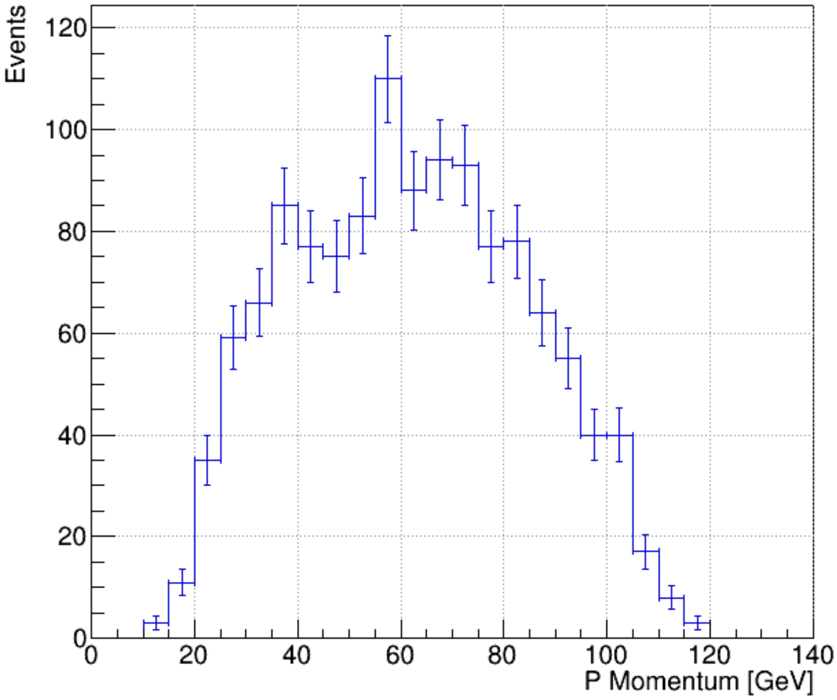


Figure 7.16: The momentum distribution of the semi-leptonic $t\bar{t}$ processes that contain a W' boson with a mass of 160 GeV.

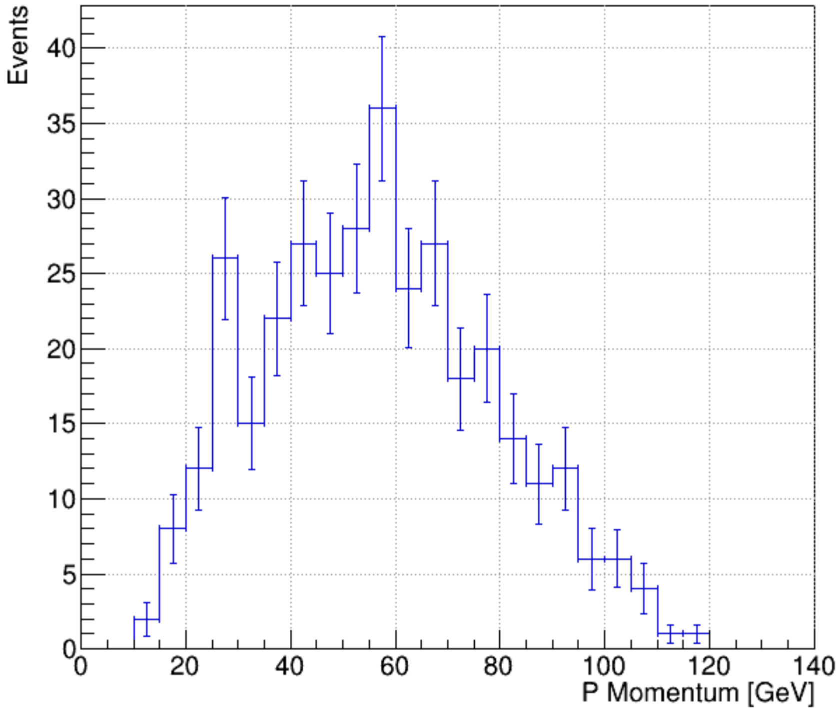


Figure 7.17: The momentum distribution of the semi-leptonic $t\bar{t}$ processes that contain a W' boson with a mass of 170 GeV.

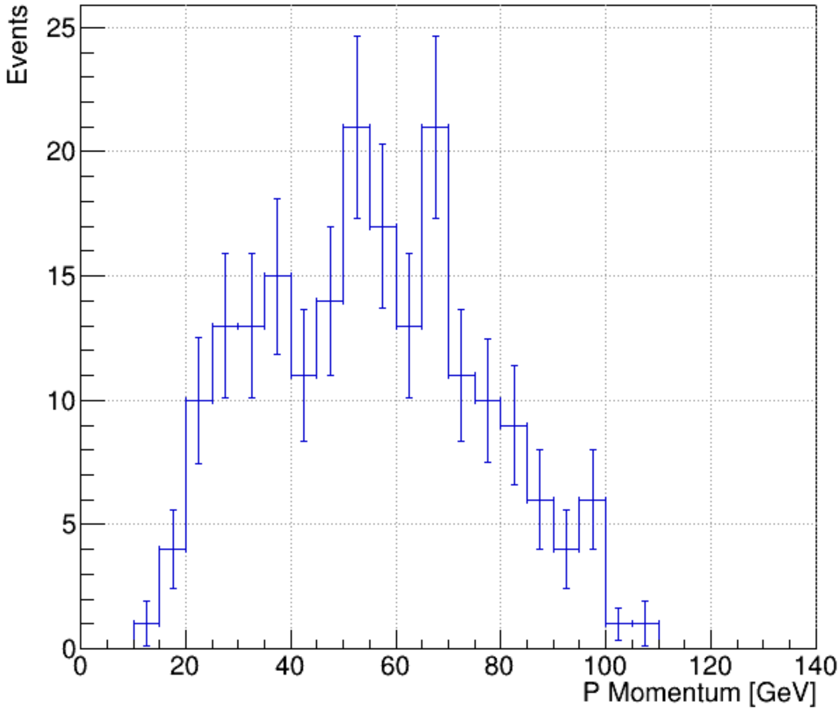


Figure 7.18: The momentum distribution of the semi-leptonic $t\bar{t}$ processes that contain a W' boson with a mass of 180 GeV.

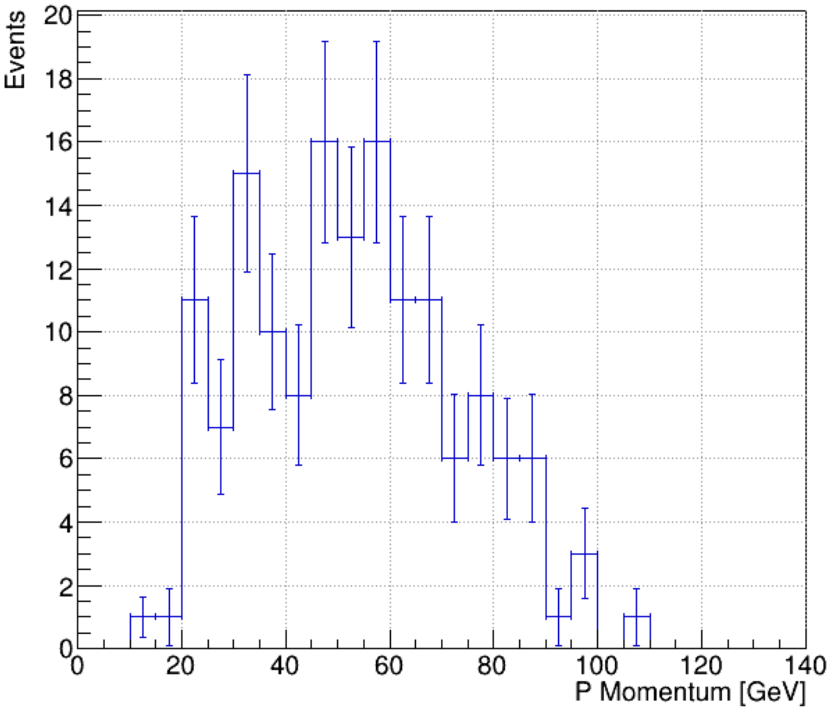


Figure 7.19: The momentum distribution of the semi-leptonic $t\bar{t}$ processes that contain a W' boson with a mass of 190 GeV.

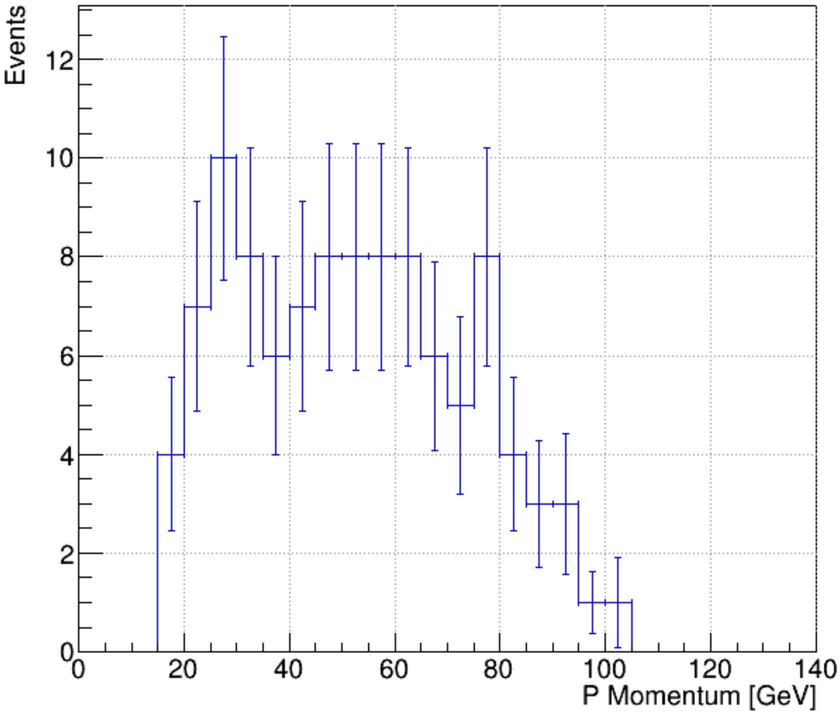


Figure 7.20: The momentum distribution of the semi-leptonic $t\bar{t}$ processes that contain a W' boson with a mass of 200 GeV.

The way the sensitivity was tested, runs along similar lines as was described in section 7.4. The main difference lies in the used new physics template which represents the additional contribution that describes the alternate hypothesis H_1. In this test, the NP has been defined to be originating from the W' boson and as such instead of varying selected areas of the considered probe, the extra events provided by the W' boson will be distributed over the P distribution. SM semi-leptonic $t\bar{t}$ events were separated from the events containing the W' boson and were rescaled such that it would reflect a SM generated process. The W' boson events would then be treated as additional BSM contributions to the SM model similar to those presented in section 7.4. To properly vary these contributions, so that the range of sensitivity can be visualised, the distributions are multiplied by constant value as to increase there presence. The considered variations in terms of the multiplications are: 0.2 to 1.3 with steps of 12, 1.4 to 2.6 with steps of 5, 3 to 6.6 with steps of 4 and to 10 to 90 with steps of 9. So, for example, when considering a multiplication of 1.4 to the W' mass of 120 GeV, the events in each bin shown in figure 7.12 is multiplied with a value of 1.4.
The reason behind each of these specific multiplication values where such that for each mass of W' boson the σ-level of significance could be reached.

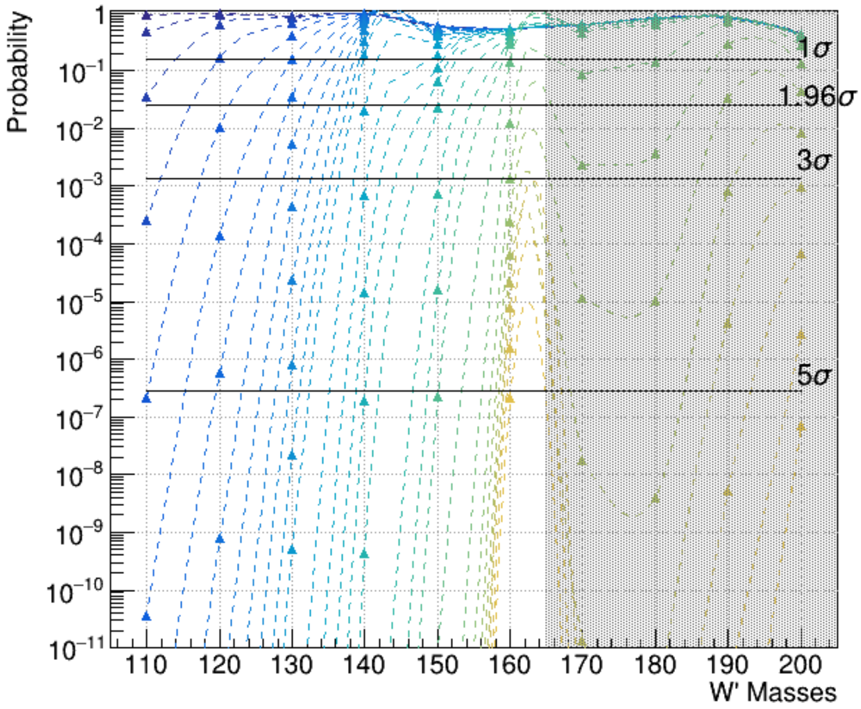


Figure 7.21: The plot shows the probability of every variation considered for every considered W' mass as obtained from the likelihood ratio fit. The amount of variation is associated to color the point has going from no variation (dark blue) to a variation where the W' distribution gets multiplied by a scale of 90 (yellow). Additionally the appropriate σ-levels are shown. The greyed out area corresponds to masses that are not statistically reliable.

Again, *TFractionFitter* was used to obtain the χ^2 of the hypothesis for each variation provided by each scale. By using the fact that these fits contained 29 degrees of freedom, the probability of the likelihood ratio of the statistical test was calculated and can be seen in figure 7.21. Similar to the same plot shown for SM variations this plot shows the various probability values for which can associate a certain significance level when rejecting the null-hypothesis H_0. However unlike the same plot shown for SM variations seen in the previous section, there is sensitivity that depends on the W' mass. The variation points are also color coded, with the dark blue color points representing multiplication scales like 0.2 at its darkest. The lighter points represent multiplication scales up to 90 which is represented by the most yellow color point.

It should be noticed that, unlike in the previous section, this plot shows a lot more chaotic behaviour. The reason for the chaotic behaviour can be seen in the templates of the W' events in figures 7.11 - 7.20. There is a certain shape that can be seen for the lower masses. This shape is very apparent due to the large amount of events, but as the W' mass increases this general shape will slowly start to disappear. It can be seen that from

W' of 160 GeV and higher this shape becomes difficult to discern as other peaks start to arise. As such at high masses the templates do not have enough statistics to be reliable when used in the statistical test. The physical reason for this occurring for increasing mass values is due to the fact that its mass get closer and closer to the mass value of top quark. This in turn makes it more difficult for it to decay to a W' boson. Even more due to the fact that the top quark gets created almost in rest. As such due to the unreliable statistical nature of the high mass W' contributions, it was decided that the masses of 170 GeV and above do not provide trustworthy results and have been omitted from further study and hypothesis testing.

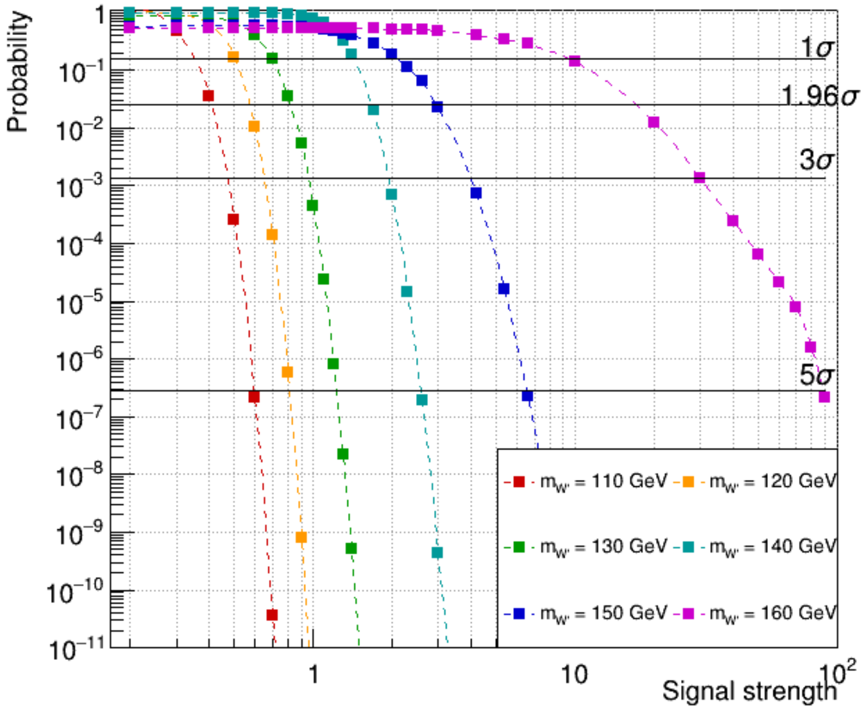


Figure 7.22: The plot shows the probability of the likelihood ratio to reject H_0 associated to a significance level which is represented by the horizontal lines in the plot. This is shown for each considered W' mass.

Following the determination of the probabilities of consistency, the cross section associated to each variation of the W' masses was calculated. Similar to previous these were plotted as the ratio to the the cross-section of the W' generation before it was varied, or in other words the *signal strength*. This would provide us an idea of how much it had to differentiate from its original value such that it reached a certain σ-level. The resulting plot can be seen in figure 7.22. Additionally the signal strengths associated to the 4 sigma levels are also given in table 7.7 for each considered W' mass. Similar to the plot in the previous section, the points are color coded with different color associated to each mass. From this plot it can be seen that the distance between the various cross sections associated to each mass starts to increase for increasing mass. This very logical due to the exponentially decreasing amount of generated events for increasing mass. Because of this masses like 160 GeV requires at least a increase of 90 times to reach a 5σ-level. This should give the reader a more clear view on why these specific scales where chosen. If they were chosen more equally spaced, these graphs would be filled with unnecessary data points which do not provide additional info.

Another interesting thing to see, is that for the mass of 110 GeV to reach 5σ-level it has to reach a cross-section that is lower than the BSM model predicts. What this means is that based on the provided model, a W' boson of mass 120 GeV or lower would already reject the H_0 at a significance level of 5σ even without varying its contribution. Additionally, for W' masses of 130 GeV or lower the H_0 would be rejected at a significance level of 3σ when considering no variations to the contributions of said W' masses.

In contrast, the higher masses of W' boson do require variations to its contribution in order to reach certain significance levels such that it can reject the H_0. In the case of W' masses between 130 GeV and 150 GeV these variations are less than 10 times the unvaried associated cross section to reach a significance level of 5σ that rejects the H_0. Same for the masses between 140 GeV and 150 GeV to reach a significance level of 3σ. The only extreme outlier as explained above is the W' mass of 160 GeV which requires at least variations of 10 times the unvaried associated cross section to reject the H_0 at significance levels that show signs of NP.

To conclude this section, in contrast to last section, the higher mass values do not say much about what minimum signal strength is required for other masses to reach a certain σ-level significance. This was the case for the SM variations due to them all being based on the same initial unvaried SM cross section. Each of these masses, however, consider a different unvaried cross section and as such each result of a specific mass does not imply anything about other masses. As such to reach a 3σ-level, which is a signal that NP can be present, the values presented in table 7.7 are the definite associated values.

Signal strengths associated to σ-levels				
W' mass [GeV]	1σ-level	1.96σ-level	3σ-level	5σ-level
110	0.37141	0.42952	0.49689	0.59997
120	0.50620	0.59094	0.68856	0.84999
130	0.69789	0.83530	0.98183	1.26676
140	1.44883	1.69076	1.98998	2.59796
150	2.11153	2.97639	4.16634	6.59557
160	9.56121	19.03056	30.06354	89.46868

Table 7.7: A summary of the signal strengths associated to a specific σ-level for each W' mass.

Chapter 8

Conclusions & Outlook

Although the Standard Model (SM) predicts that leptons aside from their mass act the same, studies into Lepton Universality Violation provided results that could bring this into doubt. In this thesis, a two-part research project was presented on e^+-e^- collisions that would generated in the Future Circular Lepton Collider (FCC-ee) at an expected integrated luminosity of 1.5 ab^{-1} and a centre-of-mass energy of $\sqrt{s} = 365$ GeV. For the first part, an algorithm was presented that tries to separate semi-leptonic $t\bar{t}$ events from the other generated backgrounds in e^+-e^- collision. Afterwards, the results of the algorithm was used to study the sensitivity of the $t\bar{t}$ spectrum to new physics which represented the second part of this research project

Because the Lepton Universality was brought into doubt, future studies into this subject are required to take a lepton independent approach when studying data. The presented algorithm in this thesis a first lepton independent algorithm that tries to isolate semi-leptonic $t\bar{t}$ events generated during e^+-e^- collisions simulations and detected by a yet conceptual detector, the ILD. The algorithm trained a neural network for 9 lepton independent properties and required the existence of a least a single b quark and only one lepton to optimise the separation of signal and background. The optimisation resulted in sample which consisted out of approximately 76.66% semi-leptonic $t\bar{t}$ events. Furthermore aside from the di-leptonic $t\bar{t}$ background which retained 22.93% of its original generated sample, every other considered background had their resulting number events reduced to less than 1% of their original amount.

In order to further improve the performance of this algorithm, further optimisation is required on the training of the neural network. This improvement can be done by considering other neural networks such as deep learning algorithms. Additionally, tests should be performed considering other detectors for which these samples were detected. Such detectors include the hypothetical IDEA and CLD detectors which are considered for the FCC-ee.

The research project concluded with an application of the presented algorithm studying the sensitivity of the semi-leptonic $t\bar{t}$ electron momentum to new physics. This study performed considering two cases and applied a likelihood ratio test for the hypothesis test. The first case considered SM generated samples and applied variations on the resulting signal sample obtained from the presented algorithm. From this it was concluded that the null-hypothesis could be rejected at a significance level of 3σ if the cross section of the model varied from the SM cross section to achieve a signal strength of 1.01286.
The second case considered SM generated samples mixed with samples considering the W' boson with masses between 110 and 200 GeV. Like the first case, for each of these masses a variation was applied to their generated samples. It was shown that the W' masses of 170 GeV and higher did not provide statistically reliable results and were omitted. It was then concluded for the masses of 130 GeV or lower that the null-hypothesis would be rejected at a significance level of 3σ even without considering variations on the generated samples. In contrast; for W' masses between 140 and 160 Gev it was concluded that they did need variations on the generated samples if they were to reject the null-hypothesis at a significance level of 3σ. For this masses between 140 and 150 GeV these variations were less then 10 times the original amount, but for a mass of 160 GeV a variation of at least 30 times the original amount was required.
These tests on sensitivity should in the future be extended towards the muon momentum. Additionally, a proper study into the systematics should be performed as to optimise the results achieved in these tests.

In conclusion, a new lepton independent algorithm has presented. An application of this algorithm has been performed in the study of the sensitivity of the electron momentum in $t\bar{t}$ production.

Acknowledgements

This master thesis presented me an amazing opportunity to work on something that has never been done before and allowed me to gain a lot of experience in the simulation field of particle physics. This can only be achieved within a well established scientific environment like the IIHE and with the help of many other researchers sharing the same passion for this topic. Therefore I would like to thank all of the members of the IIHE to welcome me and help me whenever needed.

A great deal of gratitude should be contributed to my supervisor Freya Blekman, whose motivation, inspiration, knowledge and mental support allowed me to work on these incredible topics even in the difficult times world currently is in. I would also like to give my gratitude to my promotor, Riccardo Raabe, providing me the opportunity to work on this topic at the VUB. And a lot of thanks should go the Post-Doctorate student at the VUB, Seth Moortgat, who helped me reach a good understanding behind the workings of neural networks and helped me set up the one used in this thesis.

Finally it should not be forgotten that working on a research project for such a period oftime can only be done efficiently with the support and help of family and friends. I considermyself lucky to have such a supportive and motivational group of people behind me

Bibliography

[1] Topology of three-jet events in $\overline{p}p$ collisions at $\sqrt{s} = 1.8$ tev. *Phys. Rev. D*, 45:1448–1458, Mar 1992.

[2] Root — an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(1):81 – 86, 1997. New Computing Techniques in Physics Research V.

[3] Average of r(d) and r(d*+) for spring 2019, 2019. URL:`https://hflav-eos.web.cern.ch/hflav-eos/semi/spring19/html/RDsDsstar/RDRDs.html`.

[4] Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2020-013, CERN, Geneva, Jun 2020.

[5] Particle data group. website, 2020. URL: `https://pdg.lbl.gov/`.

[6] Aad G. Abbott B. et al. Aaboud, M. Measurement of the w-boson mass in pp collisions at s=7tev with the atlas detector. *Eur. Phys. J.*, 78:110, 2018.

[7] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A.A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, Sep 2012.

[8] R. Aaij, B. Adeva, M. Adinolfi, A. Affolder, Z. Ajaltouni, S. Akar, J. Albrecht, F. Alessio, M. Alexander, S. Ali, and et al. Measurement of the ratio of branching fractionsb(b0d*+)/b(b0d*+). *Physical Review Letters*, 115(11), Sep 2015.

[9] R. Aaij, B. Adeva, M. Adinolfi, Z. Ajaltouni, S. Akar, J. Albrecht, F. Alessio, M. Alexander, A. Alfonso Albero, S. Ali, and et al. Test of lepton flavor universality by the measurement of the b0d*+ branching fraction using three-prong decays. *Physical Review D*, 97(7), Apr 2018.

[10] Abbrescia M. AbdusSalam S.S. et al. Abada, A. Fcc-ee: The lepton collider, 2019.

[11] Abbrescia M. AbdusSalam S.S. et al. Abada, A. Fcc-hh: The hadron collider, 2019.

[12] F. Abe, H. Akimoto, A. Akopian, M. G. Albrow, S. R. Amendolia, D. Amidei, J. Antos, C. Anway-Wiese, S. Aota, G. Apollinari, and et al. Observation of top quark production inppcollisions with the collider detector at fermilab. *Physical Review Letters*, 74(14):2626–2631, Apr 1995.

[13] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. Madgraph 5: going beyond. *Journal of High Energy Physics*, 2011(6), Jun 2011.

[14] J.E. Augustin et al. Discovery of a Narrow Resonance in e^+e^- Annihilation. *Phys. Rev. Lett.*, 33:1406–1408, 1974.

[15] Warmbein B. Progress on ilc detectors, 2019. URL:`http://newsline.linearcollider.org/readmore_20090917_ftr1.html`.

[16] Philip Bambade, Tim Barklow, Ties Behnke, Mikael Berggren, James Brau, Philip Burrows, Dmitri Denisov, Angeles Faus-Golfe, Brian Foster, Keisuke Fujii, Juan Fuster, Frank Gaede, Paul Grannis, Christophe Grojean, Andrew Hutton, Benno List, Jenny List, Shinichiro Michizono, Akiya Miyamoto, Olivier Napoly, Michael Peskin, Roman Poeschl, Frank Simon, Jan Strube, Junping Tian, Maksym Titov, Marcel Vos, Andrew White, Graham Wilson, Akira Yamamoto, Hitoshi Yamamoto, and Kaoru Yokoya. The international linear collider: A global project, 2019.

[17] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18(9):509–517, September 1975.

[18] S. Bifani. Lepton universality violation an experimental perspective, 2018. URL:`https://conference.ippp.dur.ac.uk/event/631/sessions/708/attachments/3020/3284/yeti18.pdf`.

[19] Simone Bifani, Sébastien Descotes-Genon, Antonio Romero Vidal, and Marie-Hélène Schune. Review of lepton universality tests in b decays. *Journal of Physics G: Nuclear and Particle Physics*, 46(2):023001, Dec 2018.

[20] Gerald C. Blazey, Jay R. Dittmann, Stephen D. Ellis, V. Daniel Elvira, K. Frame, S. Grinstein, Robert Hirosky, R. Piegaia, H. Schellman, R. Snihur, V. Sorin, and Dieter Zeppenfeld. Run ii jet physics: Proceedings of the run ii qcd and weak boson physics workshop, 2000.

[21] et al. Bloom, E. D. High-energy inelastic e p scattering at 6-degrees and 10-degrees. *Phys. Rev. Lett.*, 23:930–934, 1969.

[22] M. Boronat, J. Fuster, I. García, E. Ros, and M. Vos. A robust jet reconstruction algorithm for high-energy lepton colliders. *Physics Letters B*, 750:95–99, Nov 2015.

[23] et al. Breidenbach, M. Observed behavior of highly inelastic electron-proton scattering. *Phys. Rev. Lett.*, 23:935–939, 1969.

[24] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008.

[25] The ILD Collaboration. International large detector: Interim design report, 2020.

[26] The ILD Collaboration and contact Ties Behnke. The ild detector at the ilc, 2019.

[27] Vapnik V. Cortes C. Support-vector networks. *Machine Learning*, 20:273–297, 1995.

[28] et al. Danby, G. Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Phys. Rev. Lett.*, 9:36–44, 1962.

[29] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. Delphes 3: a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014(2), Feb 2014.

[30] Willocq S. Dobrescu, B. A. W-boson searches, 2019. URL:`https://pdg.lbl.gov/2020/reviews/rpp2020-rev-wprime-searches.pdf`.

[31] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, Oct 1993.

[32] O. Behnke et al. *Data Analysis in High Energy Physics.* Wiley-VCH, 2013.

[33] The Belle Collaboration et al. Measurement of $\mathcal{R}(d)$ and $\mathcal{R}(d^*)$ with a semileptonic tagging method, 2019.

[34] C. L. Cowan F. Reines. The neutrino. *Nature*, 178:446–449, 1956.

[35] Y. Fukuda, T. Hayakawa, E. Ichihara, K. Inoue, K. Ishihara, H. Ishino, Y. Itow, T. Kajita, J. Kameda, S. Kasuga, and et al. Evidence for oscillation of atmospheric neutrinos. *Physical Review Letters*, 81(8):1562–1567, Aug 1998.

[36] Gell-Mann M. Glashow, S. Gauge theories of vector particles. *Annals Phys.*, 15:437–460, 1961.

[37] S.W. Herb et al. Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, 1977.

[38] S. Hirose, T. Iijima, I. Adachi, K. Adamczyk, H. Aihara, S. Al Said, D.M. Asner, H. Atmacan, T. Aushev, R. Ayad, and et al. Measurement of the lepton polarization and r(d*) in the decay bd* with one-prong hadronic decays at belle. *Physical Review D*, 97(1), Jan 2018.

[39] M. Huschle, T. Kuhr, M. Heck, P. Goldenzweig, A. Abdesselam, I. Adachi, K. Adamczyk, H. Aihara, S. Al Said, K. Arinstein, and et al. Measurement of the branching ratio ofbd(*)relative tobd(*)decays with hadronic tagging at belle. *Physical Review D*, 92(7), Oct 2015.

[40] Jernej F. Kamenik, Andrey Katz, and Daniel Stolarski. On lepton flavor universality in top quark decays. *Journal of High Energy Physics*, 2019(1), Jan 2019.

[41] K. Kodama, N. Ushida, C. Andreopoulos, N. Saoulidou, G. Tzanakos, P. Yager, B. Baller, D. Boehnlein, W. Freeman, B. Lundberg, and et al. Observation of tau neutrino interactions. *Physics Letters B*, 504(3):218–224, Apr 2001.

[42] M. Koratzinos. The fcc-ee design study: luminosity and beam polarization, 2015.

[43] Breiman L. Random forests. *Machine Learning*, 45:5–32, 2001.

[44] J. P. Lees, V. Poireau, V. Tisserand, E. Grauges, A. Palano, G. Eigen, B. Stugu, D. N. Brown, L. T. Kerth, Yu. G. Kolomensky, and et al. Measurement of an excess ofbd(*)decays and implications for charged higgs bosons. *Physical Review D*, 88(7), Oct 2013.

[45] T. S. Pettersson LHC Study Group and P. Lefèvre. The large hadron collider: conceptual design. *Tech. Rep. CERN-AC-95-05-LHC*, 1995.

[46] Mancini M. The top quark pair cross-section atthe threshold at the fcc-ee, 2019. URL:`https://iihe.ac.be/sites/default/files/thesis-mathias-mancini-master-2019pdf/thesis-mathias-mancini-master-2019.pdf`.

[47] STEPHEN P. MARTIN. A supersymmetry primer. *Advanced Series on Directions in High Energy Physics*, page 1–98, Jul 1998.

[48] Pitts W. McCulloch W. S. A logical calculus of the ideas immanent in nervous activity, 1943.

[49] Porter F. C. Narsky I. *Statistical analysis techniques in particle physics.* Wiley-VCH, 2014.

[50] Charitos P. Deep learning and the quest for new physics at the lhc, 2017. URL:`https://ep-news.web.cern.ch/content/deep-learning-and-quest-new-physics-lhc`.

[51] A. Robson, P. N. Burrows, N. Catalan Lasheras, L. Linssen, M. Petric, D. Schulte, E. Sicking, S. Stapnes, and W. Wuensch. The compact linear e^+e^- collider (clic): Accelerator and detector, 2018.

[52] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65 6:386–408, 1958.

[53] Manqi Ruan. Higgs measurement at e+e- circular colliders, 2014.

[54] M. Olsson G. Turnockand B. Webber S. Catani, Y. L. Dokshitzer. New clustering algorithm for multi -jet cross-sections in e+e- annihilation, 1991.

[55] C. D. Anderson S. H. Neddermeyer. The neutrinonote on the nature of cosmic ray particles. *Phys. Rev.*, 51:884–886, 1937.

[56] Gavin P Salam and Grégory Soyez. A practical seedless infrared-safe cone jet algorithm. *Journal of High Energy Physics*, 2007(05):086–086, may 2007.

[57] Ward J. Salam, A. Weak and electromagnetic interactions. *Nuovo Cim.*, 11:568–577, 1959.

[58] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, and et al. Search for high-mass resonances in final states with a lepton and missing transverse momentum at $\sqrt{s} = 13$ tev. *Journal of High Energy Physics*, 2018(6), Jun 2018.

[59] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015.

[60] M. Thomson. *Modern Particle Physics.* Cambridge University Press, 2013.

[61] S. Weinberg. Charge symmetry of weak interactions. *Phys.Rev.*, 112:1375–137, 1958.

[62] M. Wobisch and T. Wengler. Hadronization corrections to jet cross sections in deep-inelastic scattering, 1999.

AFDELING
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be