Measurement of the Supersymmetric top quark mass at CLIC e^-e^+ collider.

Georgios Billis

August 18, 2017

Supervisor: Prof. Victoria Martin

MSc in Theoretical Physics
The University of Edinburgh
2017

Abstract

In this project I measure the mass of the Supersymmetric top quark in CLIC experiment at $\sqrt{s} = 3$ TeV and for an integrated luminosity $\mathcal{L}_{int}$=3000 fb^{-1} in e^- e^+ collisions. I assume the following decay for the top squark $\tilde{t} \to t\ \tilde{\chi}_1^0$, investigating the channel: $e^-e^+ \to \tilde{t}\bar{\tilde{t}} \to t\tilde{\chi}_1^0\bar{t}\tilde{\chi}_1^0 \to W^+W^-b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0$. The major background to this signal comes from top quark pair production $e^-e^+ \to t\bar{t}$ while other decays of the top squarks were used to model backgrounds from SUSY processes. For a generated Monte Carlo mass $M_{nom} = m_{\tilde{t}} = 845$ GeV/c^2, the top squark mass was measured $m_{\tilde{t}} = 839 \pm 11$ GeV/c^2 using the Boosted Descision Trees Multivariate Analysis and $m_{\tilde{t}} = 861 \pm 15$ GeV/c^2 using the Gradient Boosted Descision Trees Multivariate Analysis.

Declaration

I declare that this project is a product of my own personal work.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor Prof. Victoria J. Martin not only for her help, her patience and her deep and important insights on the subject, but also introducing me to a field that so far I was nothing more than a stranger.

I also owe my gratitude to Alan Taylor (UoE) for his comments and his technical help throughout this project.

Finally I would like to thank the CLICdp team that helped me. Especially Dr. Philipp Roloff, Dr. Marco Petric, Dr. Ignacio Garcia and Dr. Andre Sailer (all based at CERN).

Chapter 1

Introduction

On July 2012 the discovery of the Higgs boson was announced at CERN's Large Hadron Collider (LHC) [2] [3]. This marked the begining of a new era for experimental High Energy Physics motivating the design of new experiments for further and deeper exploration of the Higgs boson itself but also, a huge marathon addressing the questions/problems that arose with it.

One of these proposed experiments is Compact Linear Collider (CLIC), a high-luminosity linear e^- e^+ collider. It is designed for a staging scenario of three main centre of mass energies at $\sqrt{s}$ = 380 GeV, $\sqrt{s}$ = 1.5 TeV and $\sqrt{s}$ = 3 TeV targeting optimal physics output based on the current landscape [4]. The main difference of CLIC with LHC is that in the latter, protons collide which are non fundamental particles as they consist of quarks and gluons bound alltogether. One of the disadvantages of LHC is the inability to know beforehand the initial state of the colliding particles as quarks exist in a "sea of gluons" making it impossible to know their momenta. This sets some restrictions with resprect to the precision that it can probe various observables.

On the other hand, CLIC is designed to invastigate the interactions of elementary particles, an important advantage since initial states of the colliding particles is known. In addition, LHC suffers from energy loss due to Synchrotron radiation since protons are accelerated in a 27 km circular accelerator whereas CLIC, being linear does not.

The main targets of CLIC are dependent of the energy stage [4]. In the first stage it will focus on prescision standard model physics such as Higgs and top quark measurements and in the two subsequent, among others, there will be searches for new physics. One of the theories that aspires to give solutions to many of the problems of Standard Model is Supersymmetry (SUSY). In

this theory every particle has a Supersymmetric partner that differs in the spin by 1/2, thus relating bosons with fermions and fermions with bosons [5].

In the Minimal Supersymmetric Standard Model, the top squarks $\tilde{t}$ can decay into a top quark t and a dark matter candidate, the neutralino $\tilde{\chi}_1^0$. In this project I will use multivariate analysis to discriminate between signal and background with the goal to measure the top squark mass at $\sqrt{s} = 3$ TeV in the CLIC accelerator environment.

Chapter 2

CLIC

2.1 Outline of the experiment

CLIC is a proposed e^- e^+ linear collider optimised to perform in three centre of mass stages at $\sqrt{s} = 380$ GeV, $\sqrt{s} = 1.5$ TeV and $\sqrt{s} = 3$ TeV. The purpose of the different energy stages is to fully exploit its scientific potential including precision measurements and searches for physics Beyond the Standar Model.

Specifically, at $\sqrt{s} = 380$ GeV and with an integrated luminosity of $\mathcal{L}_{int}$ = 500 fb^{-1}, prescision measurements can be made in the Higgs and the top quark sector [4]. At this enegy stage the Higgsstrahlung process ($e^+e^- \rightarrow ZH$) alongside the WW fusion ($e^+e^- \rightarrow H\nu_e\bar{\nu}_e$) are the dominant and can shed light to properties of the Higgs boson in a model independent way. In CLIC there is also a dedicated program for top quark physics. This is because it is one of the most important particles in SM because it couples strongest to the Higgs field due to its mass but also it has an important role in many of the BSM scenarios. Furthermore at the next two energy stages, investigations will be done with respect to proposed scenarios for physics BSM with most importantly Supersymmetry at $\sqrt{s} = 3$ TeV and with $\mathcal{L}_{int} = 3000$ fb^{-1}. This is because CLIC has the potential for direct particle detection up to the kinematic limit of $\sqrt{s}/2$ for pair production. Also through indirect detection of observables that are sensitive to BSM scenarios, precision measurements can be made, and comparison with the SM expectations can yield unexpected divergencies, taking advantage of the full energy potential.

Given the linear nature of CLIC, there are no energy losses induced by Synchrotron radiation which appears in circular colliders, but due Beamsstrahlung radiation. As the colliding bunches get closer to the vertex, the strong elec-

tromagnetic fields (up to 10 Tesla) created by the opposing beam, cause deflection of the partices trajectories resulting to emit Synchrotron radiation. The effect is energy-dependent with huge impact at higher energies [6] as it can be seen in figure 2.1 [7].

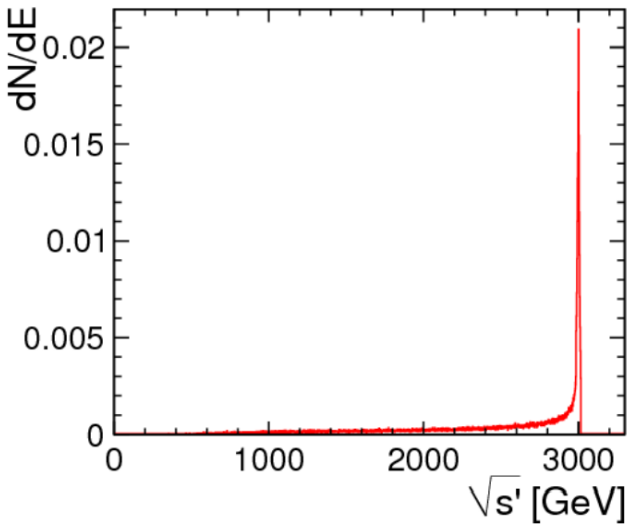


Figure 2.1: The luminosity spectrum for CLIC operating at $\sqrt{s} = 3$ TeV, where $\sqrt{s'}$ is the effective centre-of-mass energy after beamstrahlung and initial state radiation.

CLIC is designed to operate for seven, five and six years respectively in each energy stage while the upgrading periods will last two years. The total run will last for 22 years with the following table summarising the luminosities achieved in each stage:

$\sqrt{s}$ (TeV)	$\mathcal{L}_{int}$
0.38	500 fb^{-1}
1.5	1.5 ab^{-1}
3	3 ab^{-1}

Table 2.1:

Furthermore, the experiment has two detector concepts: CLIC SiD and CLIC ILD. Both have been developed for the International Linear Collinder but it has been adapted according to the needs of CLIC. Both detectors have strong central solenoid magnets creating an axial magnetic field of 5T for CLIC SiD and 4T for CLI ILD [7].

Chapter 3

Supersymmetry

3.1 Motivation

The model that describes elementary particle physics is the Standard Model [8] [9]. It involves matter fields (fermions) and vector gauge bosons (force carriers) that they stem from the fact that the SM Lagrangian is subject to $SU(3) \otimes SU(2) \otimes U(1)$ symmetries. An important part of this theory is the existence of the Higgs field which appeared during the unification of the electromagnetic and the weak force. It is known for the fact that it "gives" mass to the elementary particles (besides photons and gluons which they are protected by the $U(1)$, $SU(3)$ unbreakable symmetries respectively) via Spontaneous Symmetry Breaking.

It is true that even with the great success that the SM has been probed does not consist a flawless theory. Many parts of it are still to be answered or contain significant controversies. One example is the Hierarchy problem [10]. In the electroweak sector of the SM there is a parameter that has the dimensions of energy, the vacum expectation value (VEV) of the Higgs field which phenomenologicaly is

$$\upsilon \approx 246 \quad \text{GeV} \tag{3.1}$$

The importance of this parameter is that it sets the masses of the the theory and as well as the Higgs boson itself as it is

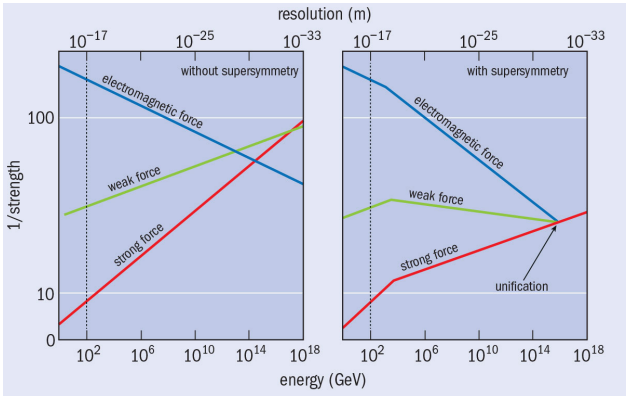
$$\mathrm{M}_H = \upsilon \sqrt{\frac{\lambda}{2}} \tag{3.2}$$

where λ is the constant of the Higgs self interaction. The problem arises when we try compute higher order corrections to the mass of the Higgs field. The self energy of the higgs boson has a term of the form

$$\int^{\Lambda} d^4k \quad f(k, \text{external momenta}) \tag{3.3}$$

where Λ represents the scale of the new physics such as the effects of quantum gravity. In dimensional regularisation, when $\Lambda \to \infty$ the renormalisability of the theory assures that there is no inconsistency; but if we take into account a scale such as the quantum gravity, then $\Lambda \to M_P \approx 1.2 \times 10^{19}$GeV which is the Plank mass. In that case the additive quantum corrections for the Higgs mass become too big dragging the value of the Higgs mass up. Because the VEV of the Higgs boson has a fixed phenomenologicaly value, one way to circumvent this problem is by relating bosons with fermions. This constitutes a solution since in the latter, the algebraic terms for their self energies come with a minus sign giving the desired cancelations without any "fine tuning" from our side. This results into the SUSY theory stabilising the Hierarchy $M_H \ll M_P$ with the constraint that the SUSY particles should be visible at a scale not too much greater than 1-10 TeV.

One other feature that makes SUSY attractive is the fact that it includes the convergence of the couplings, something that it is known as Grand Unification. From the point of view of SM, something like that is unable to be achieved since two of the couplings decrease with energy (weak and strong coupling) whereas one (electromagnetic) increases. In the Minimal Supersymmetric Standard Model (MSSM), by the inclusion of the supersymmertric particles (sparticles) such unification is actually achievable as it can be seen in the following figure *(Physics World, October 2014 - IOP publishing)*:



3.2 SUSY Phenomenology

Inside SUSY every boson is related to a fermionic supersymmetric partner and every fermion has a bosonic one. The fermion superpartners are called sfermions and differ in the spin value by 1/2 whereas the boson particles which are the force carries are related to a spin 1/2 particles, the gauginos. In the following table I show the correspondance between SM and SUSY gauge particles.

Standard Model		SUSY	
Photon	γ	Photino	$\tilde{\gamma}$
Gluon	g	Gluino	$\tilde{g}$
Z Boson	Z	Zino	$\tilde{Z}$
W Boson	$W^{\pm}$	Wino	$\tilde{W}^{\pm}$

Table 3.1:

The framework of this project is the Minimal Supersymmetric Standard Model.

Both particles and sparticles belong to the same super-multiplet having the same mass and the same quantum numbers as their SM partners but with different spin. Because no such degenerate fermion-boson pair exists in nature, then it is deduced that SUSY must be a broken symmetry. In order to constitute a solid solution to the Hierarchy problem, the difference in the masses of the particles and their supersymmetric partners must be of the order of $\mathcal{O}(1TeV)$ [11].

The MSSM includes the sleptons $\tilde{\ell}^{\pm}$, the sneutrinos $\tilde{\nu}_l$, the squarks $\tilde{q}$, the gluinos $\tilde{g}$, two pair of charginos $\tilde{\chi}_i^{\pm}$ where $i = 1, 2$, four neutralinos $\tilde{\chi}_i^0$, $i = 1, ..., 4$ and five Higgsinos $\tilde{h}^0$, $\tilde{H}^0$, $\tilde{A}^0$,$\tilde{H}^{\pm}$. In Standard Model, for the Spontaneous Symmetry Breaking one Higgs doublet is needed with 4 degrees of freedom (complex). For the SUSY model on the other hand, two such doublets are required with a total of eight degrees of freedom. The supersymmertric partners of the SM Higgs, the Higgsinos, mix with the Winos $\tilde{W}^{\pm}$ and the Binos $\tilde{B}$ in order to form the neutralinos $\tilde{\chi}_i^0$ and the charginos $\tilde{\chi}_i^{\pm}$.

It is known that within the SM the fermions carry chirality denoted by L or R whether they are left or right respectively, according to the way they transform under the symmetry group $SL(2, \mathbb{C})$. Thus given the SUSY relation of fermions-bosons that means that the sfermions (bosonic superpartners), come into doublets of the form (SM particle, SUSY partner), i.e.:

$(f_i, \tilde{f}_i)$, where i=L,R. The sfermions mix further to make the eigenstates of the mass matrix. This mixing is expected to be strong for third generation sfermions since the Yukawa couplings can be large. More precisely in our case the top squark mixing effects can be large because of the large mass of the top quark.

The following matrix is the mixing matrix expressed in the $(\tilde{f}_L, \tilde{f}_R)$ basis:

$$\mathcal{M}^2_{\tilde{f}} = \begin{pmatrix} M^2_{\tilde{f}_L} & \alpha_f m_f \\ \alpha_f m_f & M^2_{\tilde{f}_R} \end{pmatrix}$$

where the $M^2_{\tilde{f}_L}$ are terms that depend on the mass of the quark, the charge and the third component of the weak isospin [12]. The off-diagonal elements α_f, are dependent on the soft SUSY breaking trilinear scalar coupling parameters.

In this project I will focus on the top squarks where the mixing $\tilde{t}_R - \tilde{t}_L$ is important due to the large top quark mass. By computing the eigenvalues of the previous matrix and for the right flavour, we obtain:

$$m^2_{\tilde{t}_{1,2}} = \frac{1}{2}(M^2_{\tilde{t}_L} + M^2_{\tilde{t}_R} \mp \sqrt{(M^2_{\tilde{t}_L} - M^2_{\tilde{t}_R})^2 + 4m_t^2\alpha_t^2}) \tag{3.5}$$

It is obvious that the $m^2_{\tilde{t}_1}$ will have the lowest mass, something that makes it probably the lightest squark.

3.3 Results from LHC for SUSY

The search for indications for Supersymmetry is not new as LHC has been optimized and focused among other to the direct or indirect detection of sparticles [13]. The majority of the cases that have been studied, consider as main decay channel of the squarks the following:

$$BR(\tilde{q} \to q\chi_i^{0,\pm}) = 100\% \tag{3.6}$$

So far thought no indication of SUSY has been found in LHC but through each analysis the exclusion limits for the SUSY observables become bigger and bigger. This can be seen in the figure 3.1 that summarises the latest exclusion limits for the top squark mass at $\sqrt{s} = 8/13$ GeV in proton proton

collisions [14]. Whereas in the figure 3.2 are the latest exclusion limits in ATLAS for p-p collisions at 13 TeV in the gluino - stop plane.

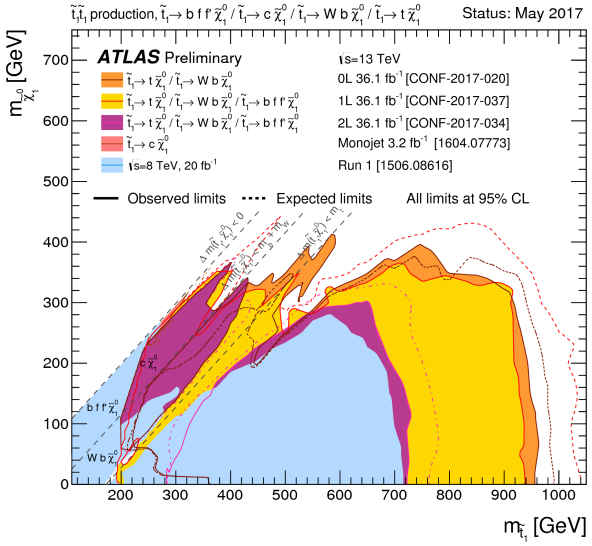


Figure 3.1: Mass exclusion limits at stop-neutralino plane (combined summary plot from ATLAS twiki).

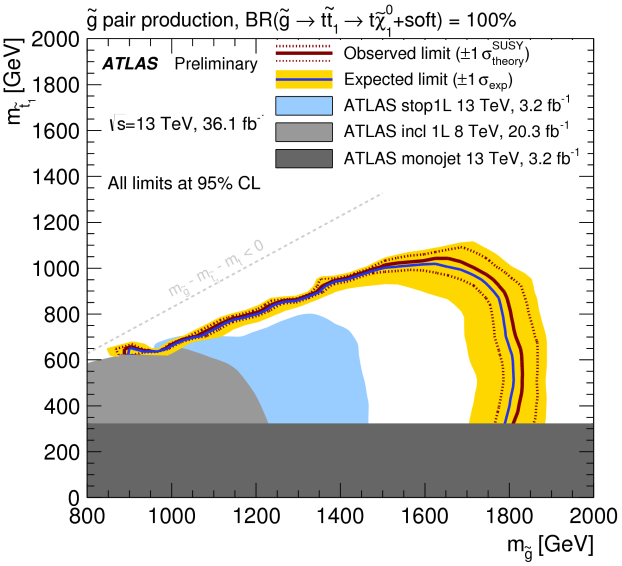


Figure 3.2: Mass exclusion limits at stop-gluino plane [1].

Chapter 4

Methods

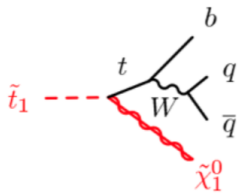
4.1 Introduction

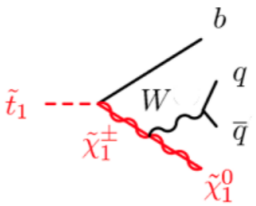
In this chapter I will describe the exact conditions and methods that were followed in this project for the measurement of the stop quark mass at the CLIC experiment.

I investigate the production of a pair of top squarks at $\sqrt{s} = 3$ TeV and assuming an integrated luminosity of $\mathcal{L}_{int} = 3000$ fb^{-1}. The channel of interest is the following:

$$e^-e^+ \to \tilde{t}\bar{\tilde{t}} \to t\tilde{\chi}_1^0\bar{t}\tilde{\chi}_1^0 \to W^+W^-b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 \tag{4.1}$$

and for this model the mass of the top squark is $m_{\tilde{t}} = 845$ GeV (denoted as M_{nom} later in the text) whereas the mass of the lightest neutralino is $m_{\tilde{\chi}_1^0}$= 357 GeV. This is not the only possible decay of a top squark but it is the dominant one. In what follows I show the top three branching ratios of the top squark decay and two of the corresponding Feynmann diagrams:





1. $\text{BR}(\tilde{t} \to t\tilde{\chi}_1^0) = 52.4\ \%$
2. $\text{BR}(\tilde{t} \to b\tilde{\chi}_1^\pm) = 34.1\ \%$
3. $\text{BR}(\tilde{t} \to t\tilde{\chi}_2^0) = 13.2\ \%$

Also, the branching ratios for the decay of the W boson to either a hadronic final state or a leptonic one are:

- $\text{BR}(W \to q\bar{q}) = 67.8\ \%$
- $\text{BR}(W \to l\nu_l) = 32.2\ \%$

In the following I summarise the possible decays of the stops indicating the branching pair fraction for each one:

- $e^-e^+ \to \tilde{t}\bar{\tilde{t}} \to t\tilde{\chi}_1^0\tilde{\chi}_1^\pm b \to W^+W^-b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 \quad (35.8\%)$
- $e^-e^+ \to \tilde{t}\bar{\tilde{t}} \to t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0 \to W^+W^-b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 \quad (27.4\%)$
- $e^-e^+ \to \tilde{t}\bar{\tilde{t}} \to t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_2^0 \to W^+W^-b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 h \quad (13.7\%)$
- $e^-e^+ \to \tilde{t}\bar{\tilde{t}} \to b\bar{b}\tilde{\chi}_1^+\tilde{\chi}_1^- \to W^+W^-b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 \quad (11.6\%)$
- $e^-e^+ \to \tilde{t}\bar{\tilde{t}} \to t\tilde{\chi}_2^0\tilde{\chi}_1^\pm b \to W^+W^-b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 h \quad (9\%)$

In this project I investigate the first channel with the W boson decaying into hadrons. This results into having 6 jets in the final products. The channel explicitly is:
$e^-e^+ \to \tilde{t}\bar{\tilde{t}} \to t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0 \to W^+W^-b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 \to q\bar{q}q\bar{q}b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0$

4.2 Measurement of SUSY particle masses

There are various ways that have been developed for the calculation of the squark masses at electron-positron linear colliders. In this chapter I present one that will not be used in my analysis but sketch the potencial and thus the importance of the aforementioned colliders regarding these kind of calculations.

The main topologies that are considered are those that the gluino is heavier than the squarks and thus the dominant production and decay channel is:

$$e^- e^+ \to \tilde{q}\tilde{\bar{q}} \to q\bar{q}\chi_1^0\chi_1^0 \tag{4.2}$$

It needs to be mentioned that a powerfull advantage of the CLIC experiment is the possibility of achieving polarisation up to 80% helping to reduce the WW background and the ability to create right or left squarks depending on the needs of the physical analysis.

4.2.1 Modified Invariant Mass

This technique has been developed initially for LHC experiment where the energy that is used to make the squark - antisquark pair is not known. The calculation was achieved by creating the variable M_C which is related to the standard invariant mass but it's invariance comes from contra linear boosts of equal magnitude. The squark mass is given by:

$$m_{\tilde{q}} = \frac{1}{2}\big(M_C^{max} + \sqrt{(M_C^{max})^2 + 4m_\chi^2}\big) \tag{4.3}$$

where

$$M_C^2 = 2(E_{q,1}E_{q,2} + \vec{p}_{q,1} \cdot \vec{p}_{q,2}) \tag{4.4}$$

and

$$M_C^{max} = \frac{m_{\tilde{q}}^2 - m_\chi^2}{m_{\tilde{q}}} \tag{4.5}$$

The power off this kind of calculation comes from the fact that no knowledge for the exact center of mass energy is needed (In comparison with the first method described in [12]), something that is not purely known since the

effects of initial state radiation and the Beamstrahlung can distort it significantly resulting to even bigger uncertainties. Note that this method assumes that the mass of the neutralino is known and that the quark is massless and therefore unsuitable for the mass measurement of the top squark.

4.3 Minimisaton of χ^2 Method

In this chapter I describe the procedure that I followed until the measurement of the top squark mass. This involved vertex finding, jet clustering, flavour tagging and the calculation of the χ^2 for each template of the top squark mass.

4.3.1 Vertex Finding

The point that particles collide in a collider experiment is called primary vertex or interaction point (IP). After the collision, it is possible that some of the particles will decay further away from the IP. The point that this decay takes place is called secondary vertex. Every analysis that involves jet clustering procedure (and not only) has as starting point the vertex finding. This involves an algorithm that classifies the tracks of the particles into vertex candidates (vertex finding) and "passes it on" to the vertex fitting algorithm. In turn, the latter, takes the vertex candidates and outputs a list of vertices with their position using the χ^2 or a related statistical method. Additionally, it yields a set of parameters that are related to each particular production point.

The reasons of why an accurate vertex finding is important are:

- Information can be exctracted about short lived particles such as their decay length.
- There can be improvement with respect to momentum estimate of the particles involved in each vertex.
- Reconstruction can be achieved for short lived particles by its decay products by fitting them into one particular vertex.

For the needs of this projecct I used the package of LCFIplus which is an improved software of the already existing one LCFIVertex. The important difference between the two is that LCFIplus employs the vertex finding before

the jet clustering (in comparison to LCFIVertex which follows the exact opposite procedure) with the goal to protect the secondary vertices from breaking up and grouped into different jets [15]. This is an important step towards better performance since the bottom jets can be identified from their sufficiently large distance between the primary vertex and their decay vertex. Additionally, the importance of this step to be performed before jet clustering, is that hadrons that contain bottom or charm quarks have sizeable lifetimes and they can be measured inside the detector [15].

That said, the vertex finding was the first step towards obtaining the final data for analysis.

4.3.2 Jet Clustering

Jets are an indespensible part of high energy physics. They come from the manifestation of strong nuclear as free quarks or gluons hadronize. They are visual structures that are measured in the detectors and they serve the need of approaching the initial parton that they originated. Throught the years various algorithms have been developed for the clustering of such objects differing in the way that they approach such procedure. Two of the most famous are the sequential recombination algorithms and cone algorithms. In this projecct I use the a sequential recombination algorithm, the k_t algorithm which is a longitudinaly invariant algorithm firstly designed for hadron colliders [16] [17].

The k_t involves a measure of distance d_{ij} for all the pair of particles i, j:

$$d_{ij} = d_{ji} = min(p_{Ti}^2, p_{Tj}^2)\frac{\Delta R_{ij}^2}{R^2} \tag{4.6}$$

where p_{Ti} is the transverse momentum of i^{th} particle with respect to beam direction and $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ with y being the rapitdity, ϕ being the azimuth angle and R the jet radius fixed in this project as $R = 0.7$.

The basic idea behind this algorithm is that it calculates the d_{ij} with the smallest value and then it "merges" the i, j into a single object with momentum $p_i + p_j$. This procedure is repeated until the quantity d_{ij} is above some threshold and all the particles that are left are the event's jet [18].

For this project the jet clustering procedure is stoped when the number of jets reaches the number 6, since in the final products there are 6 jets (see section 4.1).

4.3.3 Flavour Tagging

Flavour Tagging is the procedure in which a pseudoprobability is assigned to each jet regarding the flavour of the parton that the jet originates from. In this project, there are six jets in the final products two of them originating from b quarks.

The flavour tagging algorithm belongs in the LCFIplus package [15]. It makes use of the TMVA package which in turn uses multivariate classifiers. A set of input variables are defined for each jet and they are passed to the Gradient Boosted Descision Trees (a certain type of classifiers) which they are trained and then applied to the jets of the physical process. Those input variables are constituents of the jets such as tracks and secondary vertices.

4.3.4 Classification Method: *Boosted Descision Trees*

For the calculation of the top squark mass I made use of the TMVA package which provides a multivariate analysis platform in order (for this project) to make separation between the signal and the background. Two of them were employed, used and compared: the *Boosted Descision Trees* (BDT) and the *Gradient Boosted Descision Trees* (GBDT).

In this chapter I describe the Boosted Descision Trees method of classification and cite the results of this investigation.

Background on BDT's

Before I explain what BDT's are, it would be beneficial to sketch what Descision Trees are.

Suppose we start with signal and background events described by a set of variables such as (in HEP) transverse momenta, missing energy, etc. Then TMVA sorts all events by each variable and we choose the one that gives the best separation between signal and background. This is how the first node is made called root node. Then, in the two branches that are made we repeat this procedure applying the best cut on each node until the splitting stops based on some criterion. That is how the terminal node is reached which is called a leaf [19]. This results into a descision tree where all nodes, from root to leaf node, separate (cut) the input data into signal and background based on some criteria.

The "learning" phase where the nodes are created is called training phase.

After this, where the nodes have been created, they are tested with a (not necessarily) new data set in order to check their efficiency. This is called the testing phase.

One *issue* that arises for Descision Trees is their instability in statistical fluctuations. If two of the input variables have similar separation power then a fluctuation of the input sample may cause the selection of one of the two variables disregarding the other, something that would not happen if the fluctuation was not present. This results in a totaly different tree structure.

A way to circumvent this problem is to employ many different trees whose depth (number of nodes of each tree) is small. They are called weak learners. The classification of the event comes from the majority vote of these weak learners (boosting). It is important to be noted that each tree is trained with the same sample of input variables. That way not only we reach better results in the possible presence of statistical fluctuations but we also avoid the overtrainning of the trees.

Separation of Signal and Background

The products of the particle collision can be separated as signal and background. The channel of interest is denoted as SUSY signal, i.e.:

$e^- e^+ \to \tilde{t}\bar{\tilde{t}} \to t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0 \to W^+W^- b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 \to q\bar{q}q\bar{q}b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0$

The SUSY background is any other SUSY process that involves top squark pair production but the final products are not the ones of interest, and the SM background which does not involve any supersymmetric process.

Before the training and testing phase it is possible to impose some constraints to the input kinematic variables. They are called preselection cuts.

To this end, using the TMVA package in ROOT and trainning 1000 trees with the preselection cuts:

- $E_{visible} < 2$ TeV
- $E_{top1,2} < 1.2$ TeV

where $E_{visible}$ is the total energy of the jets and $E_{top,i}$ is the energy of the i^{th} top quark. 99% of the signal events pass the preselection cuts whereas for the background 70%.

In what follows I enumerate the input variables in the TMVA that they are used for trainning and testing the classifiers. It is important to notice

that the sample with which the BDT's was trainned and tested are different. Some of those variables can be seen in figures 4.3 - 4.6.

1. The the reconstructed masses of the top quark candidates
2. The reconstructed masses of the W boson candidates
3. The missing transverse momentum p_T (transversity with respect to the beam line)
4. The $\Delta\phi$ angle between the $t\bar{t}$ quarks
5. The ΔR between the W candidates and the b jets
6. The $\cos\theta_{top}$. We expect the top quarks created by purely SM processes to be more boosted than the ones that involved SUSY processes
7. Thrust:
$$T = max_{\hat{n}} \frac{\sum_i |\vec{p}_i \cdot \hat{n}_i|}{\sum_i |\vec{p}_i|} \tag{4.7}$$
where $\hat{n}$ is the unit vector that maximises the ratio of the sums. The sums run over all the particles that emerge from the collision
8. The three highest b tags
9. The three highest c tags
10. θ_{miss}, which is defined to be the polar angle of the missing momentum
11. Oblatness
12. Sphericity (defined in [20])
13. Aplanarity (defined in [20])
14. The number of particles detected
15. $y_{3,4}, y_{4,5}, y_{5,6}, y_{6,7}, y_{7,8}$: The jet transition values defined from the following formula:
$$y_{i,j} = \min(E_i^2, E_j^2) \frac{1 - (\cos\theta_i - \cos\theta_j)}{s} \tag{4.8}$$
where i, j jets are chosen such that y gets the minimum value

The TMVA trains the classifiers with respect to the input variables, listed above, on generated samples of the signal and backgrounds. The TMVA creates "weight files" which in turn are applied to statistically independent samples of the signal and background processes (test phase). Essentially what TMVA achieves is reducing the multidimensional phase space of variables into one variable, known as BDT cut.

As shown in the following graph, the value of the BDT cut that was chosen is $cut_{BDT} = -0.036$ due to the fact that it has the maximum statistical significance $\frac{S}{\sqrt{S+B}}$=5.68, where S is the number of signal events and B is the number of background events.

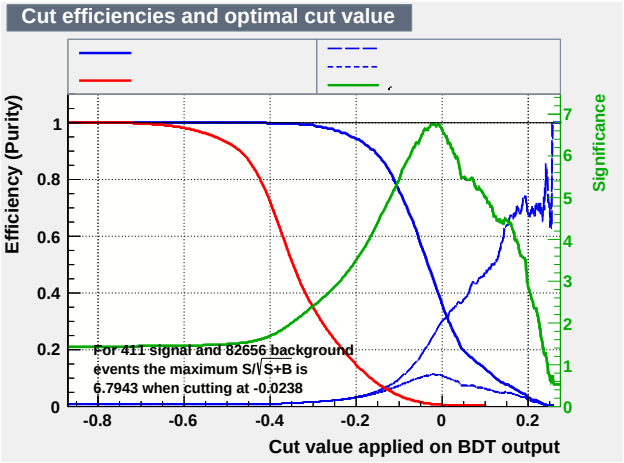


Figure 4.1: The green line represents the statistical significance $\frac{S}{\sqrt{S+B}}$ where as it can be seen it reaches the greatest value at -0.0238. The cut is also known as descision boundary and defines the critical region.

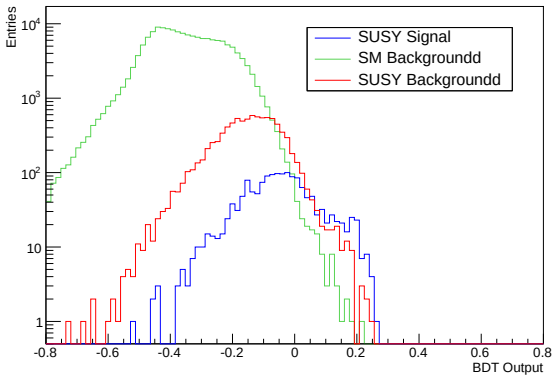


Figure 4.2: The BDT output for the signal and the two backgrounds. The y axis has a log scale.

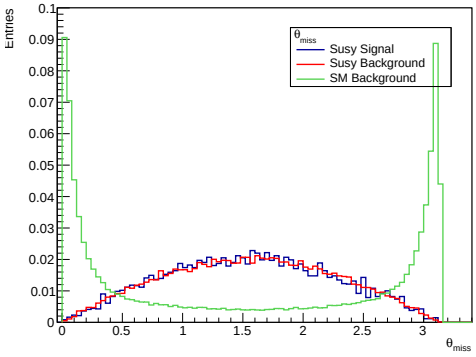


Figure 4.3: Polar angle of the missing momentum of the jets

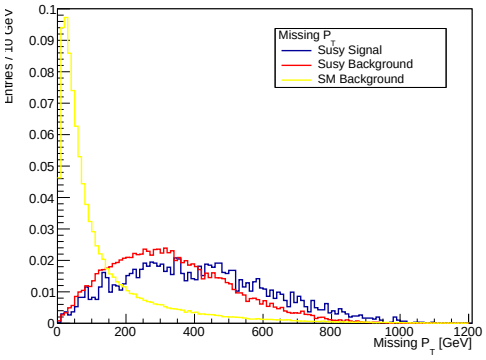


Figure 4.4: Missing transverse momentum

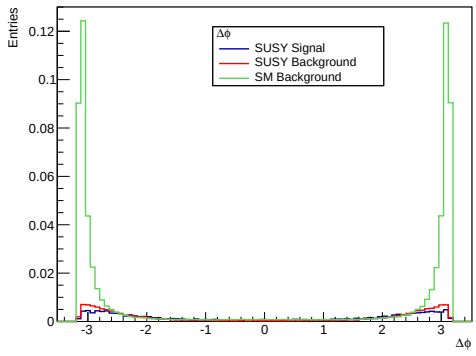


Figure 4.5: The angle between the reconstructed tops

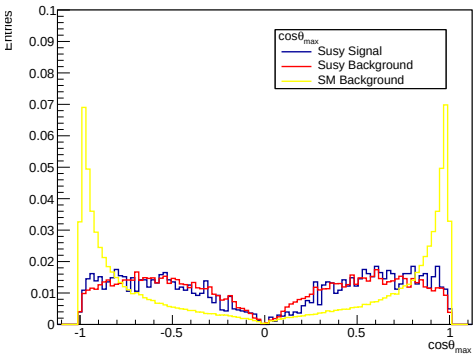


Figure 4.6: $\cos\theta_{top}$

After the training phase, the TMVA yields which variables where the most important in the classification procedure where (ordered from the most significant):

1. θ_{miss}
2. missing p_T
3. $\Delta\phi$ as the angle between the $t\bar{t}$
4. $cos\theta_{top}$

The above histograms show the aforementioned variables as they were given for training and testing.

Energy Spectrum of the Reconstructed Top quark

In the histogram below the energy spectrum of the top quark can be seen for the signal events. This variable was given to the TMVA for training and testing but it was not among the important ones. It is clear that the maximum value of the energy is 1.2 TeV something that explains why this was also set as a preselection cut.

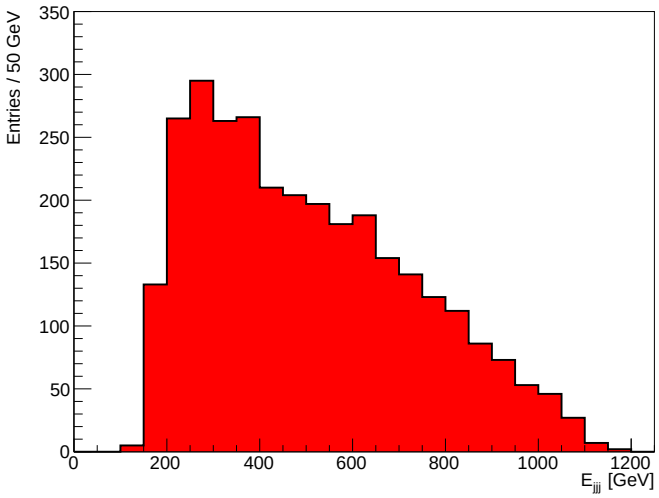


Figure 4.7: The energy of the reconstructed top quark candidate.

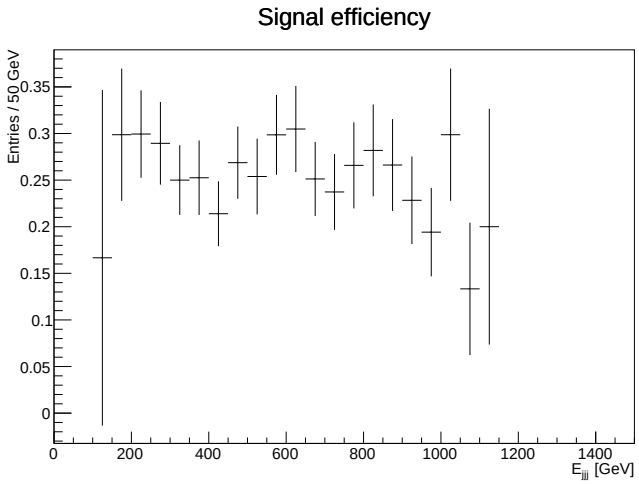


Figure 4.8: The signal efficiency with the associated errors.

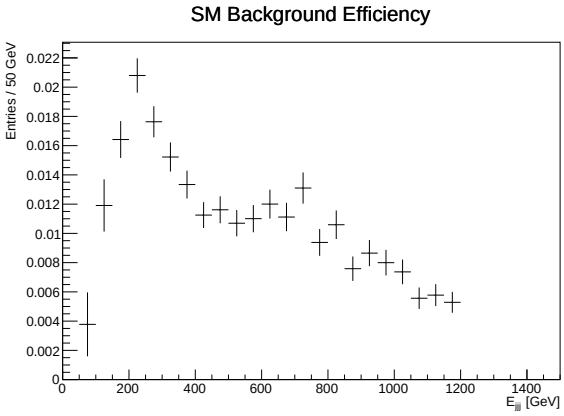


Figure 4.9: The Standard Model Background Efficiency

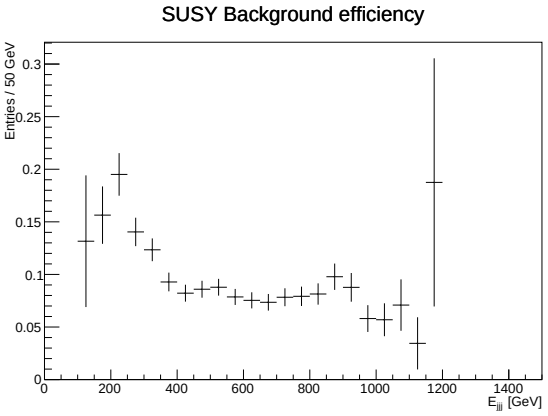


Figure 4.10: The SUSY Background Efficiency

In the above histogram it can be seen the signal efficiency (with errors) as with respect to the energy of the three jets E_{jjj}. It is defined to be the ratio of the number of events that passed the cut_{BDT} over the total number of the signal events.

For completenes, the above diagrams depict the SM background efficiency and the SUSY background efficiency.

Effectiveness of BDT's

In the following histograms we can see the number of events of the signal and the two backgrounds with respect of the E_{jjj} before and after the cut_{BDT}. As it can be seen, the Standard Model Background as well as the SUSY Background have reduced significantly.

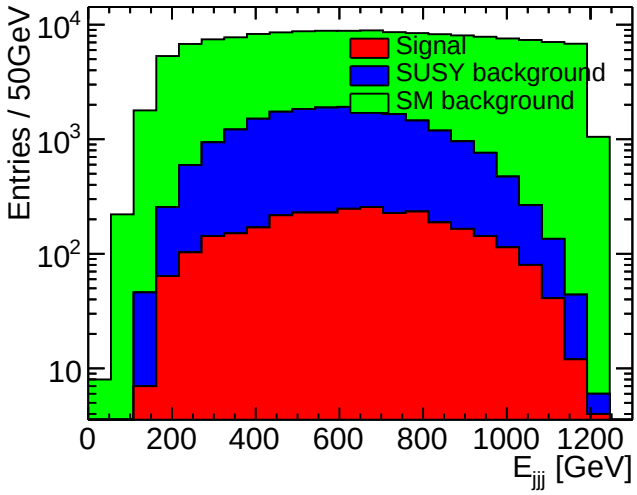


Figure 4.11: Signal, SM background and SUSY background before the TMVA application. It is a stacked histogram with log y axis.

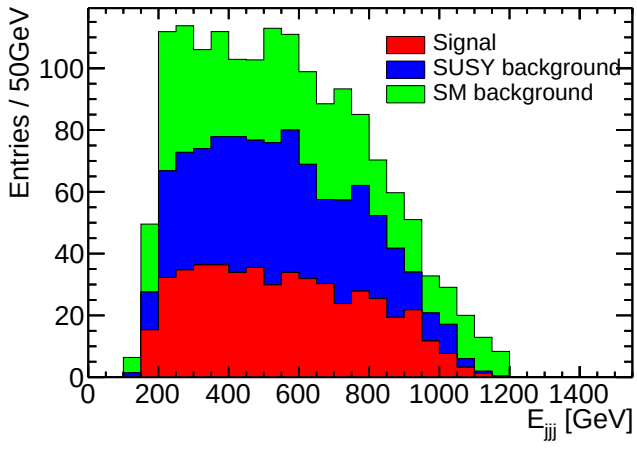


Figure 4.12: Signal, SM background and SUSY background after the TMVA application, again stacked.

The following table summarises the number of events that passed the pre-selection cuts, before and after the TMVA cut.

BDT	Before TMVA (N_{events})	After TMVA (N_{events})	%
Signal	3005	493	16
SUSY background	17628	558	3.1
SM background	113494	528	0.4

Table 4.1:

Mass Measurements Results

For the calculation of the top squark mass I made use of the minimisation of the χ^2 method. In this method we calculate the χ^2 value for different masses of the top squark, and the minimum of the fitted polynomial yields the actuall mass.

Concretely, the formula for the calculation of the χ^2 is:

$$\chi^2 = \sum_i \frac{(n_{data,i} - n_{template,i})^2}{\sigma^2_{data,i} + \sigma^2_{template,i}} \tag{4.9}$$

where n_{data} are the number of events in the energy spectrum of the three jets after the TMVA (figure 4.12) and $n_{template}$ are the number of events for each different template, while the sum runs over all the bins i of the histogram. A χ^2 was calculated for each different template, summarised in the table below.

$M_{template}$	σ (fb)
M_{nom}−200 GeV	2.8
M_{nom}−100 GeV	2.2
M_{nom}−50 GeV	1.9
M_{nom}	1.65
M_{nom}+50 GeV	1.4
M_{nom}+100 GeV	1.2
M_{nom}+200 GeV	0.8

Table 4.2:

The mass was found to be $m_{\tilde{t}} = 839$ GeV compaired to the mass of the Monte Carlo sample which is $m_{\tilde{t}} = 845$ GeV/c^2.

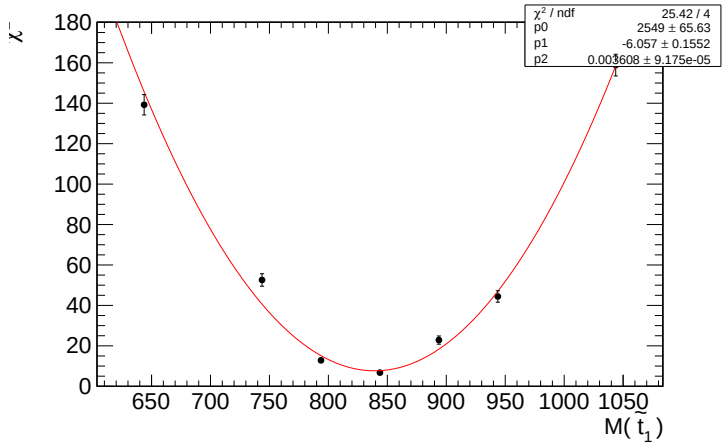


Figure 4.13: The χ^2 versus the different top squark mass templates. The minimum is at 839 GeV/c^2.

In order to calculate the statistical uncertainty off the mass we employed a *toy Monte Carlo Method*.

In this method, I smeared each bin of the energy spectrum of the three jets (Figure 4.12), with a Gaussian of deviation $\sqrt{n_{data,i}}$ and mean zero. Then the mass calculation was repeated. This procedure was done 5000 times. The deviation of the Gaussian was chosen to have that value since $\sqrt{n_{data,i}}$ is the error of each bin content i. The result was an uncertainty of the mass 11 GeV.

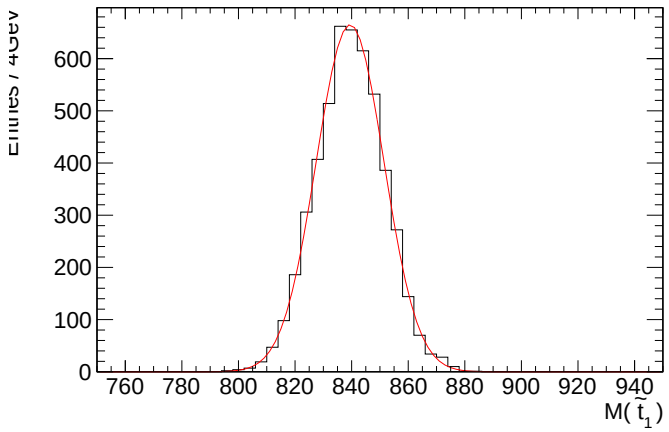


Figure 4.14: The mass result after 5000 different measurements. A toy Monte Carlo method was used that yielded an uncertainty of 11 GeV.

4.3.5 Results - Summary of the BDT method

Using the Boosted Descision Trees and with the method of minimisation of the χ^2 for different templates, I found the mass of the supersymmetric top quark to be

$$m_{\tilde{t}} = 839 \pm 11 \text{GeV}/c^2 \tag{4.10}$$

The nominal mass is M_{nom}= 845 GeV/c^2, my measurement is one sigma inside this range, taking into account the aforementioned uncertainty at CLIC for integrated luminosity $\mathcal{L}_{int}$=3000 fb^{-1}.

4.3.6 Classification Method: *Gradient Boosted Descision Trees*

In this chapter I will describe the mass calculation for the top squark using the Gradient Boosted Decision Trees (GBDT) as a method of classification for discrimination between signal and background. Furthermore, comparison will be made between BDT's and GBDT's.

Background on GBDT's

Gradient Boosting is a different method for classification. The goal, as in the BDT's, is to create a prediction model. The main difference from the classification method described in chapter 4.3.4 is that the prediction model is built in an iterative way allowing its own optimisation by the calculation of errors using a different loss function than the one used for the BDT's.

Again like all the boosting methods, it combines a set of weak learners which is defined to be the trees that are "short", i.e. that they have small number of nodes. In the first phase -training phase- a rough approximation for the prediction model is made. Subsequently, the the error of the model and the actual expected values is calculated (residual) and it is added in the previous model making it a slightly better. This procedure is repeated various times until a good description is reached.

The name gradient comes from the observation that the residuals that each time are added are just the negative gradient of a squared error loss function [19]. The mathematical formulation of both BDT's and GBDT's lies beyond the scope of this project but can be seen in [21], [19].

Training and testing phase

In this part of the procedure I "fed" the TMVA with the exact same variables as I did for the BDT's and imposed the exact same preselection cuts. The purpose of this is to make a comparison of the two methods in the end. For this method 1000 trees were also employed.

The preselection cuts were:

- $E_{visible} < 2$ TeV
- $E_{top1,2} < 1.2$ TeV

The list of the input variables can be seen in chapter 4.3.4 (p.16).

Based again, to the graph that has the efficiency vs the cut_{GBDT}, the chosen value for the cut was the one that maximised the statistical significance $\frac{S}{\sqrt{S+B}}$=7.68, where S is the signal and B is the background: $cut_{GBDT} = -0.0606$. It needs to be noted that the statistical significance achieved through this method is almost one sigma (standar deviation) higher than that of the BDT method of classification.

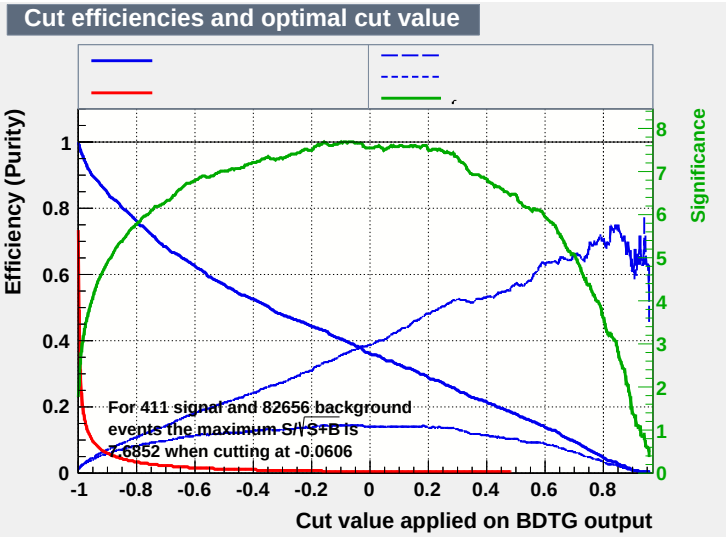


Figure 4.15: The green line represents the statistical significance $\frac{S}{\sqrt{S+B}}$ where as it can be seen it reaches the greatest value (7.68) at -0.0606. The cut is also known as descision boundary and defines the critical region.

The variables that TMVA found most important are exactly the same with those of the BDT method, namely:

1. θ_{miss} : the polar angle of the missing momentum of the jets.
2. p_T : the missing transverse momentum.
3. $\Delta\phi$: the angle between the reconstructed tops.
4. $\cos\theta_{top}$.

In the following stacked histogram can be seen the post GBDT reduction of background and in the table above, the number of events that passed (for both signal and background) are denoted.

GBDT	Before TMVA (N_{events})	After TMVA (N_{events})	%
Signal	3005	455	15
SUSY background	17628	330	1.8
SM background	113494	198	0.1

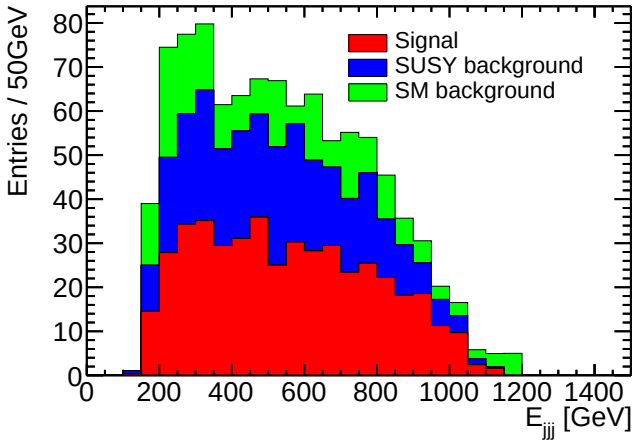


Figure 4.16: Energy of the three jets after the cut_{GBDT}.

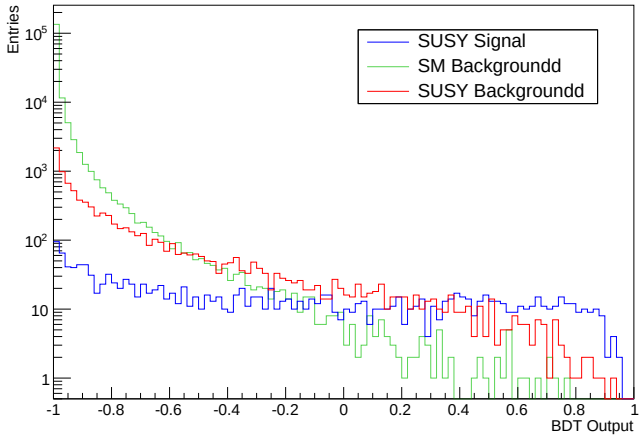


Figure 4.17: The GBDT output for the signal (blue), SUSY background (red) and SM background (green).

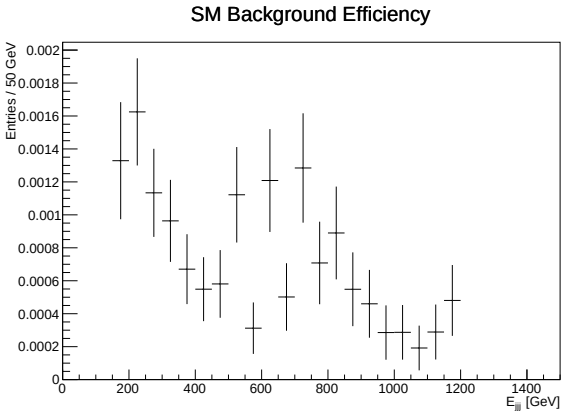


Figure 4.18: The Standard Model Background Efficiency for the GBDT method.

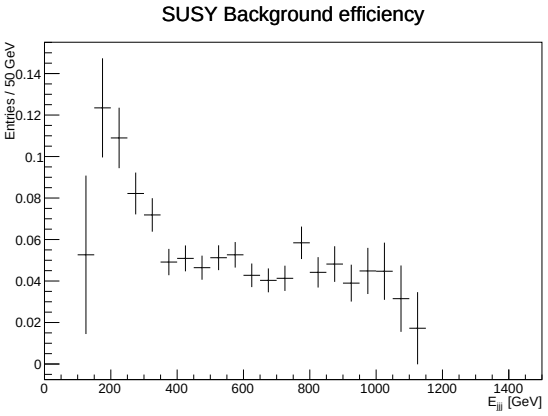


Figure 4.19: The SUSY Background Efficiency for the GBDT method.

The above histograms show the effciency for both backgrounds whereas the one below, for the signal.

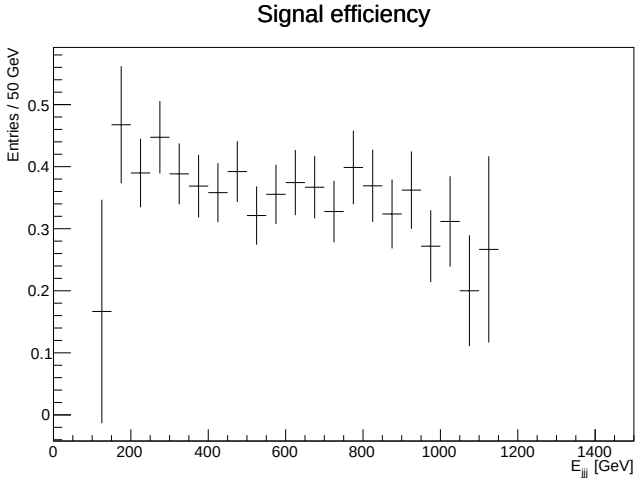


Figure 4.20: The signal efficiency for the GBDT method.

Using again the method of minimisation of the χ^2 for different templates (table in p.22) of the top squark mass (formula 4.9), I calculated the mass to be:

$$m_{\tilde{t}} = 811\text{GeV}/c^2 \tag{4.11}$$

again using a toy Monte Carlo method I smeared each bin of the energy spectrum of the three jets with a Gaussian of deviation $\sqrt{n_{data}}$ and mean zero, and I repeated the mass calculation. This procedure was done 5000 times. The resulting uncertainty was 15 GeV.

Both calculations can be seen in the following graphs where in the first the minimisation of the χ^2 can be seen and in the second the calculation of the uncertainty.

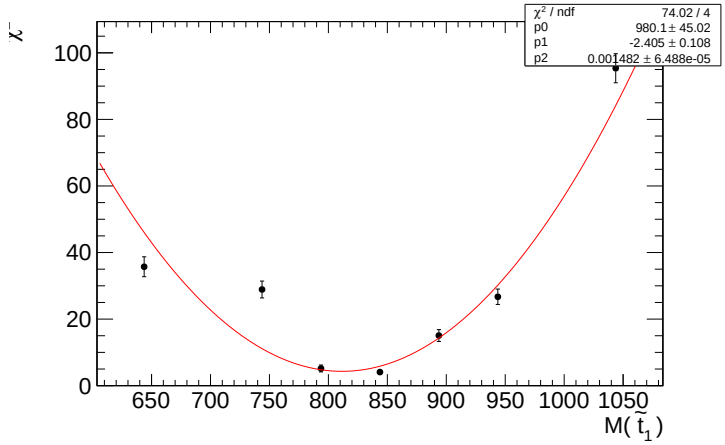


Figure 4.21: The χ^2 versus the different mass templates. The minimum is at 811 GeV/c^2.

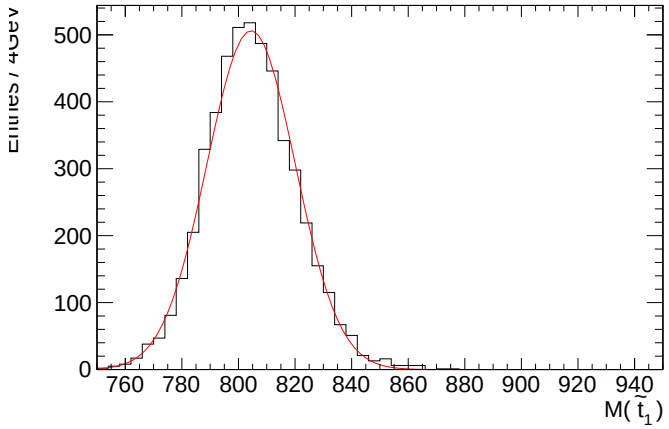


Figure 4.22: The resulting uncertaity from the toy Monte Carlo method is 15 GeV.

4.3.7 Results - Summary of the GBDT method

Using the Gradient Boosted Descision Trees and with the method of minimisation of the χ^2 for different mass templates, I found the mass of the supersymmetric top quark to be:

$$m_{\tilde{t}} = 811 \pm 15 \quad \text{GeV}/c^2 \tag{4.12}$$

The nominal mass is $M_{nom} = 845$ GeV/c^2, my calculation is two sigma inside this range taking into account the uncertainty calculated through the toy MC method at CLIC for an integrated luminosity $\mathcal{L}_{int}$=3000 fb^{-1}.

4.4 Comparison of BDT - GBDT methods/Results

In this section I am going to compare the performance as well as the results of the two methods followed in this project.

For the BDT method I set the number of weak learners (trees) to be 1000 and the maximum depth of the tree to be 3. Same considarations took place also for the the GBDT method.

The following table summarises the basic characteristics of the two methods:

Method	Stat. Efficiency	$m_{\tilde{t}}$	N_{signal}
BDT	6.79	839 $\pm$ 11 GeV/c^2	493
GBDT	7.68	811 $\pm$ 15 GeV/c^2	455

Table 4.3:

As it can be seen the GBDT method achieves higher statistical significance in comparison with the BDT method but higher uncertainty with a difference of 4 GeV. This can be attributed to the fact that the proposed cuts from the TMVA are mainly addressed for particle discoveries rather than for measuring an observable. That means that for the optimal precision of the TMVA one has to "play" with the value of the cut -primarily around the proposed value - in order to extract the lowest uncertainty possible for the mass calculation. Furthermore as it can be seen in the above table, the number of events after the TMVA cut, in the case of GBDT are almost 40 less. Even thought that they sound few, as a fraction they constitute 8% of the N_{signal}, number not at all negligible especially when we deal with a

few hundreads of events. Additionally another interesting thing is the importance of these 40 more events in the case of BDT (for the mass and the uncertainty calculation). This is clear because in the BDT the number of events of SM background (N_{events}=528) in total are far more than those in the GBDT method (N_{events}=198), and at first sight it would be expected that the GBDT has better potential than the BDT. But the mass and uncertainty measurements proved to be the other way round giving evidence to the fact that more signal events have more impact than less background.

Similar contradiction produces the careful examination of the signal efficiency plots(4.8 for BDT and 4.20 for GBDT). Both of them (taking into account the errors) tend to have an almost constant trend but the efficiency of GBDT is higher.

Another interesting observation is the dependence of the discrimination strength of the classifiers with respect to the number of trees. Before I set the number of trees to be 1000, my initial calculation for the mass was with 200 trees, something that yielded an uncertainty of 19 GeV. The reason for that is clear if the following histogram is taken into account:

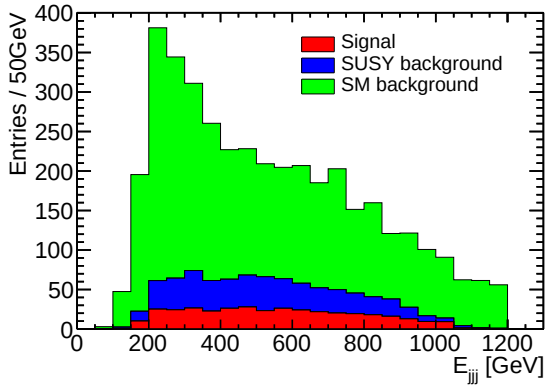


Figure 4.23: The signal and the two backgrounds after the BDT cuts for the case of 200 trees.

If the above histogram is compared with the corresponding one for 1000 trees (Figure 4.12), then it is clear that the SM background has reduced significantly. This is the reason why we get a significance of 11 GeV, i.e. 4 GeV less.

Although, apart from the "tunning" issues, the fact that the BDT's performance is superior is something that the experimental community knew. From 2005, similar studies were made for the MiniBoone experiment and it was found that BDT's (on Adaboost mode) outperform other methods [22].

Chapter 5

Conclusions

The mass of the supersymmetric top quark was calculated for the CLIC experiment at $\sqrt{s} = 3$ TeV and for an integrated luminosity $\mathcal{L}_{int} = 3000$ fb^{-1} for a nominal mass of $M_{nom} = 845$ GeV/c^2.

Using the Boosted Decision Trees multivariate method of classification it was found to be

$$m_{\tilde{t}} = 839 \pm 11 \text{GeV}/c^2 \tag{5.1}$$

whereas using the Gradient Boosted Decision Trees method of classification it was found to be

$$m_{\tilde{t}} = 811 \pm 15 \text{GeV}/c^2 \tag{5.2}$$

Althought the GBDT's at first seem to be the best candidate for a more precise mass calculation, the BDT's proved to be the ones that offered a more precise one. That is because the resulting number of signal events in the BDT method are more than the ones in GBDT implying that having stronger signal than better separation between signal-background is preferable in the case of the mass calculation of the $\tilde{t}$ in CLIC experiment.

Having in hand the mass of the top squark calculated with the BDT method along with the Monte Carlo mass of the neutralino, comparison can be made with figure 3.1 of ATLAS. As it can be seen the set of these masses ($m_{\tilde{t}}$ = 839 GeV/c^2, $m_{\tilde{\chi}_1^0}$ = 357 GeV/c^2) belong in the already excluded limits of ATLAS experiment. Besides that, the potencial of CLIC collider for a precise measurement of the mass is not ruled out but on the opposite stressed, given the potencial of low uncertainty on the mass measurement.

Summing up, I demonstrated that it is possible the CLIC experiment to make a precise measurement of the top squark mass with significantly low

uncertainty for the scenario that the mass of the $\tilde{t}$ is $m_{\tilde{t}} = 845$ GeV/c^2 and the mass of the neutralino being $m_{\tilde{\chi}_1^0} = 357$ GeV/c^2 investigating the channel:

$$e^- e^+ \to \tilde{t}\bar{\tilde{t}} \to t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0 \to W^+W^-b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 \to q\bar{q}q\bar{q}b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0$$

Chapter 6

Further Work

As further work on this project the following can be mentioned:

- As I investigated the following channel

$$e^- e^+ \to \tilde{t}\bar{\tilde{t}} \to t\tilde{\chi}_1^0 \bar{t}\tilde{\chi}_1^0 \to W^+W^- b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 \tag{6.1}$$

  the decay of the W boson can be hadronic or leptonic. Isolating and investigating the cases that $W \to l\nu_l$,i.e. the leptonic channel a mass measurement can be achieved.

- Another possibility is to look also for other decays of the top squarks such as

$$e^- e^+ \to \tilde{t}\bar{\tilde{t}} \to t\tilde{\chi}_1^0 \chi_1^\pm b \to W^+W^- b\bar{b}\tilde{\chi}_1^0\tilde{\chi}_1^0 \tag{6.2}$$

- Besides looking for other decays, work can be done on taking into account on the mass measurement all the possible stop decays denoted in page 11. In that case, since there will be no SUSY background and work has to be done in order to simulate a new one involving other interactions.

- Work also can be done for finding the ideal cut of the BDT classification method as well the most important variables that affect the procedure. One possible way to do such thing is by firstly including all the variables in the TMVA and after the training/testing phase to plot the backgroung rejection versus the signal efficiency. Then, repeat the procedure for all the variables excluding one and plot again the same graph. That way, step by step, fewer and fewer variables will be taken into account by TMVA and where a huge difference in the plot

of backgroung rejection versus signal efficiency is seen, then this will be an indication that this variable is an important one.

Bibliography

[1] ATLAS collaboration et al. Search for a scalar partner of the top quark in the jets+ etmiss final state at s= 13 tev with the atlas detector. ATLAS-CONF-2017-020.

[2] Georges Aad, T Abajyan, B Abbott, J Abdallah, S Abdel Khalek, AA Abdelalim, O Abdinov, R Aben, B Abi, M Abolins, 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, 2012.

[3] Peter W Higgs. Spontaneous symmetry breakdown without massless bosons. *Physical Review*, 145(4):1156, 1966.

[4] The CLIC, MJ Boland, U Felzmann, PJ Giansiracusa, TG Lucas, RP Rassool, C Balazs, TK Charles, K Afanaciev, I Emeliantchik, et al. Updated baseline for a staged compact linear collider. *arXiv preprint arXiv:1608.07537*, 2016.

[5] Adel Bilal. Introduction to supersymmetry. *arXiv preprint hep-th/0101055*, 2001.

[6] Giovanni Bonvicini, E Gero, R Frey, W Koska, C Field, N Phinney, and A Minten. First observation of beamstrahlung. *Physical review letters*, 62(20):2381, 1989.

[7] H Abramowicz, A Abusleme, K Afanaciev, N Alipour Tehrani, C Balázs, Y Benhammou, M Benoit, B Bilki, J-J Blaising, MJ Boland, et al. Higgs physics at the clic electron–positron linear collider. *The European Physical Journal C*, 77(7):475, 2017.

[8] Ta-Pei Cheng, Ling-Fong Li, and Ta-Pei Cheng. Gauge theory of elementary particle physics. 1984.

[9] Steven Weinberg. A model of leptons. *Physical review letters*, 19(21):1264, 1967.

[10] Ian JR Aitchison. Supersymmetry and the mssm: An elementary introduction. *arXiv preprint hep-ph/0505105*, 2005.

[11] Yorikiyo Nagashima. *Beyond the standard model of elementary particle physics*. John Wiley & Sons, 2014.

[12] Frank Simon. Techniques and prospects for light-flavored squark mass measurements at a multi-tev e+ e- collider. Technical report, 2010.

[13] Serguei Chatrchyan, Vardan Khachatryan, Albert M Sirunyan, Armen Tumasyan, Wolfgang Adam, Thomas Bergauer, Marko Dragicevic, Janos Erö, Christian Fabjan, Markus Friedl, et al. Search for supersymmetry at the lhc in events with jets and missing transverse energy. *Physical review letters*, 107(22):221804, 2011.

[14] Morad Aaboud, G Aad, B Abbott, J Abdallah, O Abdinov, B Abeloos, R Aben, OS AbouZeid, NL Abraham, H Abramowicz, et al. Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in s= 13 tev p p collisions with the atlas detector. *Physical Review D*, 94(5):052009, 2016.

[15] Taikan Suehara and Tomohiko Tanabe. Lcfiplus: A framework for jet analysis in linear collider studies. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 808:109–116, 2016.

[16] Stefano Catani, Yu L Dokshitzer, Michael H Seymour, and Bryan R Webber. Longitudinally-invariant k-clustering algorithms for hadron-hadron collisions. *Nuclear Physics B*, 406(1-2):187–224, 1993.

[17] Stephen D Ellis and Davison E Soper. Successive combination jet algorithm for hadron collisions. *Physical Review D*, 48(7):3160, 1993.

[18] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. Fastjet user manual. *The European Physical Journal C*, 72(3):1896, 2012.

[19] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, Helge Voss, M Backes, T Carli, O Cohen, A Christov, et al. Tmva-toolkit for multivariate data analysis. *arXiv preprint physics/0703039*, 2007.

[20] Chunhui Chen. New approach to identifying boosted hadronically decaying particles using jet substructure in its center-of-mass frame. *Physical Review D*, 85(3):034007, 2012.

[21] Robert E Schapire. A brief introduction to boosting. In *Ijcai*, volume 99, pages 1401–1406, 1999.

[22] Hai-Jun Yang, Byron P Roe, and Ji Zhu. Studies of boosted decision trees for miniboone particle identification. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 555(1):370–385, 2005.