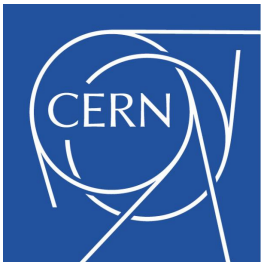
THE UNIVERSITY *of* EDINBURGH



Jet Reconstruction and Kinematic Fitting of the Top Quark Pair Production at CLIC at $\sqrt{s} = 3$ TeV

Master Thesis - 2015

Théo Galy-Fajou

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Supervisors:
Dr. Victoria J. Martin, *University of Edinburgh*
Pr. Aurelio Bay, *EPFL*

CERN-THESIS-2015-288
03/02/2015

Abstract

Top quark physics, due to its possible link with new physics, is a critical topic now that the Standard Model has been experimentally verified. A complete method to reconstruct top quarks pairs at the proposed Compact LInear Collider project is presented here. In this study, MC generated events of $e^+e^- \to t\bar{t}$ have been used to tune and optimize algorithms in order to reconstruct faithfully the decay products of the top quarks. An emphasis is made on the flavour identification of the jets since it is critical to identify correctly identify the jets to remove most of the background. The reconstructed jets are fitted to the topology with the KLFitter algorithms that have been adapted for CLIC. Using a multi-variable analysis, it finds the best permutation of jets with the best set of parameters using the kinematics of the event. The results of this technique applied on a sample of 49500 $e^+e^- \to t\bar{t}$ events (corresponding to 850 fb^{-1} at $\sqrt{s} = 3$ TeV) is presented here.

Contents

Chapter 1

Introduction

The discovery of the Higgs particle in 2012 was the final confirmation of the validity of the Standard Model (SM). However this is far from being the end of story for particle physics! Some observations still do not fit the predictions of the SM like, for example, the neutrino oscillations [1] or the existence of dark matter [2], and nothing proves that there are not more fundamental particles to discover. Moreover, many particles like the Higgs boson or the top quark, need much more precise measurements of their properties. These particles have only been observed in a hadron collider, able to reach high energies (the LHC will be able to reach 14TeV), but lacking of precision, for example the different top quark mass measurements are shown on figure 1.1. To overcome those weaknesses, new experiments and new tools are needed. Many new collider projects are under developments: some circular, like the Future Circular Collider (FCC) [4] at CERN and the Circular Electron-Positron Collider (CEPC) [5] in China, and other linear, like the International Linear Collider (ILC) [6]. But the one that interest us here is the Compact LInear Collider (CLIC) [7] developed at CERN. This linear electron-positron collider reaching energies up to 3 TeV produces almost no low energy QCD radiation, has a well-defined center of mass energy and also contains new detector technologies based on high-granularity for an improved detection resolution. The main aims of CLIC are to improve the measurements on the properties of the top quark and the Higgs boson, and to discover and eventually measure any new particles.

The measurement of the top quark properties is not only about having highly accurate values for the PDG[8]! Its important mass makes it an actor in many high energy physics processes. For example it most strongly couples to the Higgs field, and it could also couple to other Higgs fields belonging to any BSM(Beyond the Standard Model) theories.

The top pole mass is also critical in a specific BSM theory: the electro-weak false vacuum [9]. A false vacuum is a metastable sector of space that appears to be a perturbative vacuum, but is unstable due to tunneling that could lead to a lower energy state. Quantum field theory allows us to know if we live in a meta stable world, i.e. if this false vacuum exists, from the value of the Higgs boson's mass and the top quark's pole mass, as shown on figure 1.2. The pole mass is a theoretical value, is slightly different from the mass measured at hadron colliders and can be measured with precision in a lepton collider. As one can see if the unstable vacuum is almost excluded, there is still uncertainty if the vacuum is stable or meta-stable, a better measurement of the top pole mass would give a clearer answer to the question.

In order to make those precise measurements, the best reconstruction possible is needed. This study aims at reconstructing the top quark and antiquark, mass and kinematics as well as possible using Monte Carlo simulated data. It would also provide a background for the search of scalar top quarks as predicted by supersymmetric models, for CLIC the model chosen is mSUGRA model III, that have a very similar topology signature with the SM tops which would be then the main background.

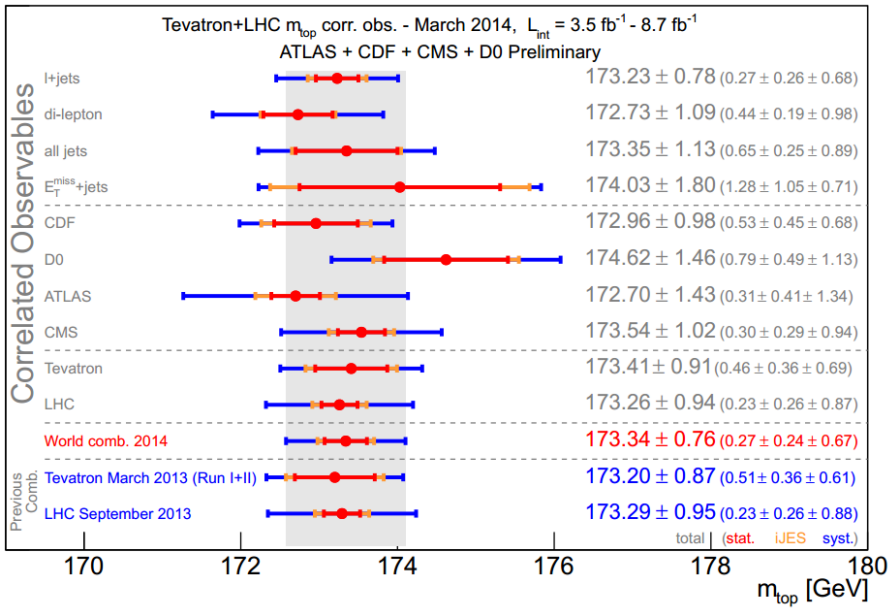
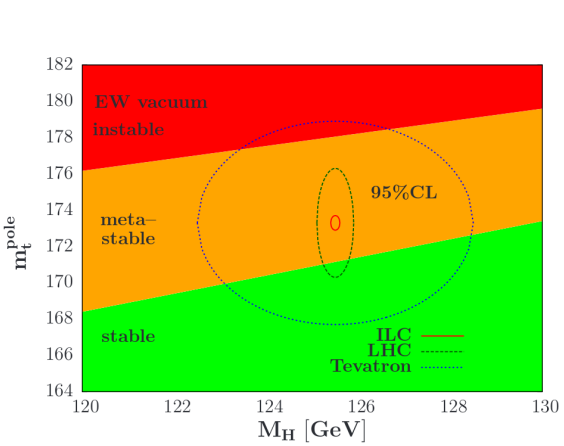


Figure 1.1: Measurement of the top quark mass in CDF, D0, ATLAS and CMS, from [3]

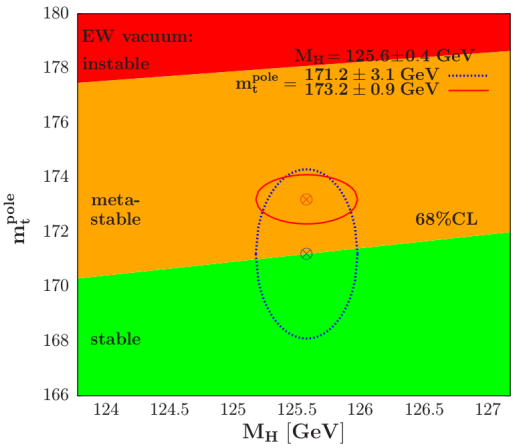
The CLIC study expects to measure the top quark mass, making measurements at 500 GeV, with a precision of better than 100MeV with 100fb^{-1}. [7]

In chapter 2, the underlying physical phenomena and processes of the top quark pair production are detailed and explained. The tools to collide and observe and analyse this process, the CLIC accelerator and detector, are then described in chapter 3. The raw output from the detector need to be treated and interpreted through a series of different algorithms, defined in chapter 4. The reconstructed jets obtained from this analysis are then fitted kinematically in a specific topology using the KLFitter tool [10] adapted for CLIC and described in chapter 5. The analysis of the results over a MC sample are finally done in chapter 6.

This work has been done at the end of 2014, as the CLIC project is still in conceptual development. Some of the informations announced here have possibly changed, especially concerning the detector design, however the results expected should not vary dramatically.



(a) 2σ ellipse at LHC, ILC (predicted) and Tevatron



(b) 1σ ellipses at Tevatron ($m_t^{pole} = 173.2$ GeV) and (updated) LHC ($m_t^{pole} = 171.2$ GeV)

Figure 1.2: $[M_H, m_t^{pole}]$ planes, with state of the electroweak vacuum, figures from [9], CLIC results should approach or be better than the ILC results.

Chapter 2

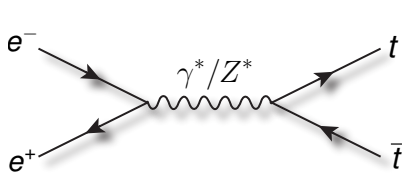
Overview of the top quark physics

2.1 Top quark pair production

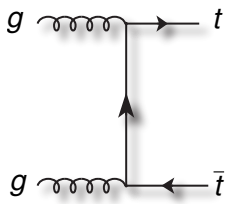
The top quark with its $173.21 \pm 0.51 \pm 0.71$ GeV [8] is the heaviest of the known quarks and was discovered in 1995 by the CDF and DØ experiments at Tevatron. Twenty years later, the precision on its mass, its production cross section and other properties such as its spin or its charge have still room for improvement. This is mainly due to the fact that the colliders able to reach the high-energy needed for the production of a pair of top quarks (around 350 GeV) are hadron colliders (Tevatron, LHC) and lack of precision.

The way the tops are produced is completely different in a hadron collider and a lepton collider as shown in figure 2.1: In a hadron collider the top quark pairs are produced from gluons emitted by the colliding protons. In addition to the important QCD (quantum chromodynamic) background it can generate, it is not possible to know exactly the center of mass energy of the interaction.

In a leptonic collider, tops are produced through an electroweak process: the electron and positron will annihilate and all the energy will be used to create the top quark pair and give them energy. It is not only possible to know the center of mass energy of the process accurately with a minimum QCD radiation, but the beams can also be polarized to study the electromagnetic properties of the top quark since the helicity of the mother particle is transmitted to its daughters.



(a) Lepton collider



(b) Hadron collider

Figure 2.1: Production mode of top quark pairs in hadronic and leptonic colliders

The other particularity of the top quark production in an electro-weak process is that the cross-section $\sigma(e^-e^+ \to t\bar{t})$ varies significantly when the energy of the process is closed to the top quark pair mass (~ 350 GeV)[11] as shown on figure 2.2. The measurement of the peak with a high precision would give very interesting measurement of top coupling values, and a precise measurement of the top pole mass, it would also be an easy way to measure its cross-section since it would be the major production.

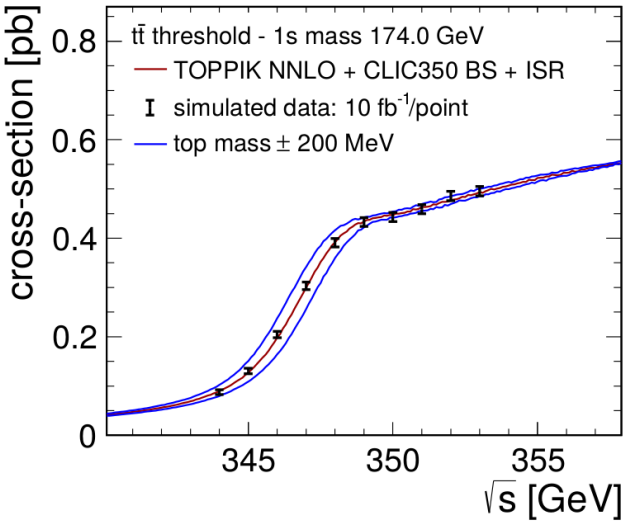


Figure 2.2: Background-subtracted cross-section for 10 fb^{-1} per data point, together with the cross- section for the generator mass of 174 GeV as well as for a shift in mass of $\pm$ 200 MeV from [7]

2.2 Event Signature

As explained previously, in an electro-weak process all the energy goes to the top quark pair. In the most simple model, the electron and the positron arrive with equal and opposite momenta (which is a pretty good approximation), the top quark and top antiquark are produced back-to-back with a non-boosted center of mass. In the case of $\sqrt{s} = 3$ TeV each top would have an energy of 1.5 TeV. In reality, because of the initial state radiation as well as the beamstrahlung, this picture is inaccurate and the energy spectrum is quite large.

With its lifetime of 5×10^{-25}s [8], the top decays before entering the detector therefore it cannot be observed directly. It is possible however to reconstruct it through its decay products. The Standard Model predicts that the top quark can have quite exotic decays like $t \to u\ \gamma/Z$ or $t \to c\ \gamma/Z$ but the probability to have such events is completely insignificant and they have never been observed yet.

The main decay mode of the top quark is into a W and a quark with negative charge (b, s, d), but since the CKM coefficient $V_{tb} \approx 0.99$ and $m_{\text{top}} > m_{\text{W}} + m_b$ the branching ratio $\frac{\Gamma_{t\to Wb}}{\Gamma_t} > 0.99$ [8] is completely dominant. This decay is shown on figure 2.3.

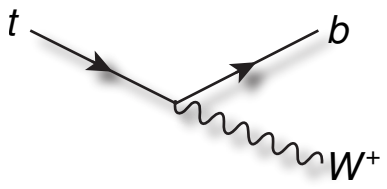


Figure 2.3: Main decay mode of the top quark

If the b quark has a sufficient lifetime to hadronize and to be detected, it is not the case for the W which will decay directly (its lifetime is $3.076 \times 10^{-25} s$ [8]). The possible decays mode of the W^- with their branching ratio are shown on table 2.1, for the W^+ the conjugates are taken but the branching ratios are the same. For the hadrons, which are $q_i\overline{q_j}$ pairs, all quark

combinations are possible, but because of the CKM matrix coefficients, the most probable ones are the pairs containing quarks and antiquarks from the same generation (e.g. $\overline{u}d, \overline{c}s$)

Mode	Fraction(Γ_i/Γ)
$e^-\overline{\nu_e}$	$10.75 \pm 0.13\%$
$\mu^-\overline{\nu_\mu}$	$10.57 \pm 0.15\%$
$\tau^-\overline{\nu_\tau}$	$11.25 \pm 0.20\%$
hadrons	$67.27 \pm 0.27\%$

Table 2.1: Main decay modes of the W^- (take the conjugate for W^+)

When looking at all the possible decay combinations of the two Ws, three main channels can be obtained : fully hadronic when both Ws produce hadrons, semileptonic when one W produces leptons and the other W produces hadrons, and dileptonic when both Ws produce leptons. The channels are illustrated on figure 2.4.

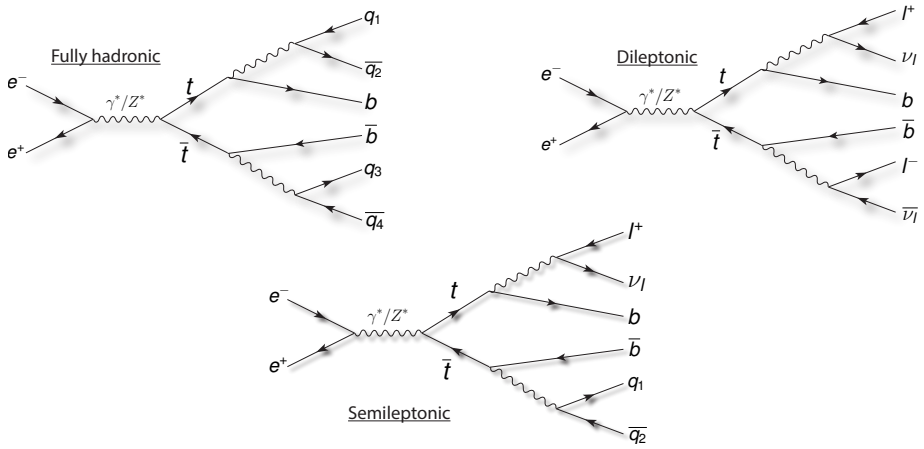


Figure 2.4: Three main channels for top quark pair decay

The main background for the top quark pair production consists in the single top production $e^-e^+ \to W^+W^- \to t\bar{b}W^-$, with one possible feynman diagram shown on figure 2.5, but there is also the process $e^-e^+ \to W^+W^-\ \gamma/Z/H$ with $\gamma/Z/H \to b\bar{b}$. The cross sections computed with MADGRAPH [12] are shown on table 2.2. The cross-sections at $\sqrt{s} = 3$ TeV for the $WW\gamma/Z/H$ production are relatively small when $\gamma/Z/H$ gives a pair of b quarks, making the identification of the two b jet in the event crucial for the reconstruction. These processes have exactly the same particle signature than the top quark pair production, but the topology of the events is not. The kinematics are completely different since the top quark and their decay products are produced back to back in the center of mass frame whereas in the background the Ws are produced to back to back. When reconstructing the top kinematics, a mass difference would clearly appear.

One have to notice that, when considering the initial state radiation as well as the beamstrahlung effect, the range of energies becomes bigger. The cross-section of the top quark pair cross-section decreases with energy, so with a wider range of energies, the cross section is bigger. The effective cross section at CLIC is 59fb, the other cross-sections have not been computed but are expected to be slightly bigger as well.

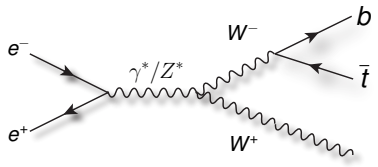


Figure 2.5: Feynamn diagram of the single top quark production in a lepton collider

Process	Cross-Section [fb]
$e^+e^- \to t\bar{t}$	19.15
$e^+e^- \to t\bar{b}W^-$	33.6
$e^+e^- \to W^+W^-\gamma$	27.67
with $\gamma \to b\bar{b}$	0.19
$e^+e^- \to W^+W^-Z$	32.77
with $Z \to b\bar{b}$	4.64
$e^+e^- \to W^+W^-H$	1.12
with $H \to b\bar{b}$	0.85

Table 2.2: Cross-sections for signal and background at $\sqrt{s} = 3$ TeV

2.3 Observables

There are five possible type of particles coming out of the top decays : b quark, light quark, electron, muon and neutrino. They all have different behaviour in the detector but one common point for the quarks is that they will produce jets. A jet is a narrow cone of hadrons and other particles produced by the hadronization of a quark. Because of QCD confinement [13], particles carrying a color charge, such as quarks, cannot exist in free form. Therefore they fragment into hadrons before they can be directly detected, becoming jets.

The different signatures are detailed here:

2.3.1 light quarks (u, d, s, c)

The light quarks hadronize immediately after having been produced. Mainly pions, kaons, protons and neutrons are formed. These particles decay again, producing even more particles making a cascade of secondary particles resulting in a hadronic shower. The jets they produce are coming directly from the primary vertex, the interaction point.

2.3.2 b quark

The b quark instantly hadronizes into a b (possibly excited) meson or a b baryon for example : $B^{\pm}, B^0, B_s^0,\ B_c^+$ for the mesons and $\Lambda_b^0, \Xi_b^0, \Xi_b^-, \Omega_b^-$ for the baryons, depending on the quark it couples to. They all have a lifetime of an order 10^{-12}s [8] which allow them to travel a short distance before decaying, but not enough to reach the detector. They produce a specific secondary vertex, which can be reconstructed. The b quark is also a lot more massive than the lighter quarks and this can be used to differentiate it when reconstructing the particle through kinematics. The b quark decays through the weak interaction with $b \to W^-c$ as the dominant channel.

2.3.3 Charged leptons ($e^{\pm}$ and $\mu^{\pm}$)

Charged leptons behave very differently from the quarks in the detector. As they are not interacting at all with the strong force, they can exist alone and do not hadronize, moreover muons and electrons have a different behaviour in the detector.

Muons have a lifetime of 2.19×10^{-6}s [8], but considering that they are highly energetic when produced by a W boson, and that they do not emit much Bremsstrahlung they are able to travel out of the detector [14]. Particularly thick detector layers are usually put on the outside to detect the muon trace, triggering in this way their presence, also muons leave a significantly different trace than the electrons in the calorimeters (figure 4.4b).

Electrons are completely stable, their lifetime is over $4.6 \times 10^{26}y$ but they do not travel as much as the muons. As they transverse the detector they will emit Bremsstrahlung radiation, producing a lot of photons and losing energy quickly. Some of the photons will also create electron-positron pairs which will also emit Bremsstrahlung. This cascade of repeated events creates an electromagnetic shower that is usually stopped by the electromagnetic calorimeter.

2.3.4 Neutrinos

High-energy neutrinos are produced though the Ws decays. Ws are of course produced through the top quark decay but also through the τ and b decays but with lower energies. Neutrinos do not decay and only interact with the matter through the weak force, it is therefore impossible to detect them directly in a detector since they do not leave any trace. They will appear in the form of a missing momentum when looking at the total momentum. For the top quark reconstruction The four-momentum of the neutrino can be fitted using the information of the charged lepton and fixing the W mass or simply using it as a free parameter.

2.4 Analysis selection

The goal for top physics at CLIC is to make high-precision measurement of the top quark properties: some top events are then inappropriate to use due to a large amount of missing information. This is due to the neutrinos which create a missing four-momentum. If there is only one neutrino in the event, it can be easily reconstructed without too much error but with two neutrinos the uncertainty become a lot larger. Those neutrinos can be produced through the W decays as explained previously but also from the τ decays. The τ has a very short lifetime of 2.9×10^{-13}s [8] similar to the b quark and decay dominantly into a W and ν_τ. This behaviour is problematic since it creates a similar topology but with more missing energy and a displaced vertex, making it more complicated to reconstruct correctly the event.

Dileptonic events (with two neutrinos coming from Ws) are rejected as well as semileptonic event producing a τ, this selection is done at the MC generator level but can be realized using a Tau finder processor. From now on "charged leptons" will only refer to electrons and muons.

This leaves two workable channels with different topologies:

- **Fully Hadronic:** The event is characterized by two b-jets, four other jets and low missing transverse momentum.
- **Semileptonic:** The event has also two b-jets, two other lighter jets, one isolated track corresponding to a high-energy lepton and finally an important missing transverse momentum due to the neutrino

The ratio of the different channel is illustrated in details on figure 2.6.

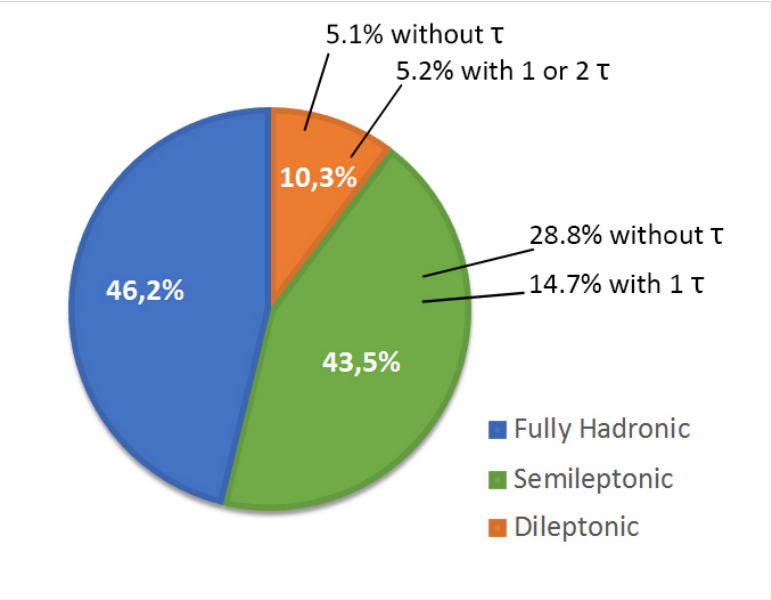


Figure 2.6: Ratio of the different channel of the top quark pair production

Chapter 3

CLIC

To detect and produce top quark events a particle accelerator and a detector are needed, as explained previously the idea is to use a lepton collider. In this study, the collider used is CLIC. CLIC (Compact LInear Collider), is a new linear electron-positron collider project developed at CERN. Two different accelerator models are currently designed to reach a maximum center of mass energy $\sqrt{s}$ of 500 GeV or 3 TeV. This study is only focused on the 3 TeV model at an energy of 3 Tev as it involves a lot more possibilities for new physics discovery but the methods are almost the same for other ranges of energy. The collider would also be able to go through different energy stages : 350 GeV, 1400 GeV and 3000 GeV.

The conceptual layout of CLIC is shown on figure 3.1. The collider is still in a conception status, it also has to be approved and chosen by the particle physics community to be the next generation collider so the location of the accelerator is not defined but it would probably be close to CERN.

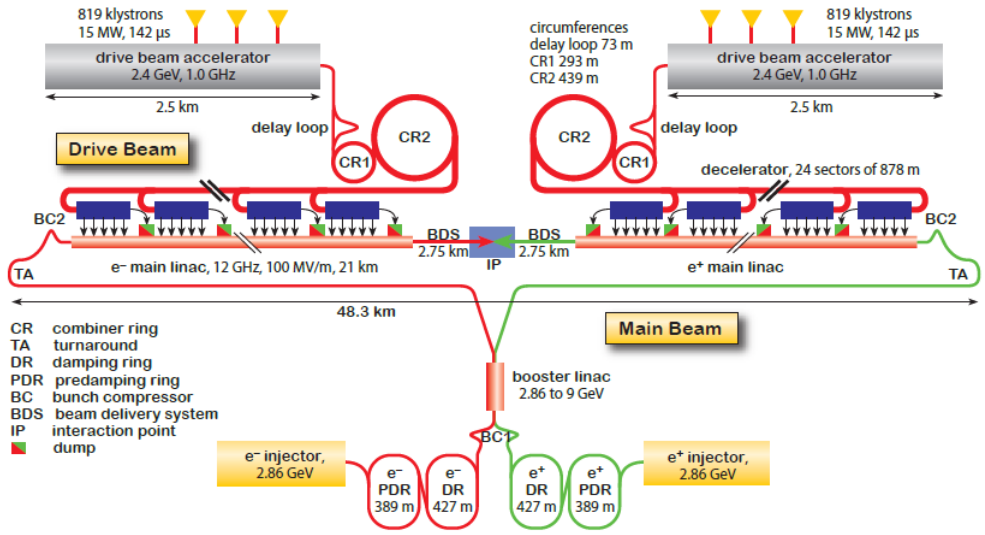


Figure 3.1: Overview of the CLIC layout from [7]

3.1 The accelerator

On the layout one can see that the length of the accelerator is about 48km, making it twice longer than the LHC and bigger than any already built accelerator. But the size is just one of the many huge challenges of this project.

First, reaching a 3 TeV center of mass energy with a linear setup requires an incredibly high acceleration that the usual RF cavities used at LHC cannot produce. In order to achieve this energy level, the positron and electron beam are separately injected in a predamping ring, go through a damping ring and a bunch compressor before being accelerated from 2.86 GeV to 9 GeV through a booster linear accelerator (linac). The beams are brought to the main accelerator and after reaching the turnaround they are accelerated to the desired energy of 3 TeV through the main linac. The main linac is composed of about 7000 100 MV/m 12 GHz accelerating structures with novel low energy beam-driven acceleration technique. The beams are finally cleaned by collimation and compressed to a very small size at collision.

Another critical challenge linked again to the linear structure of the collider is the luminosity of the collider. The beams cross each other only once after being accelerated and on one unique interaction point. This reduces considerably the number of events produced. In order to offset it, the beam size is considerably more focused and therefore its density increases leading to higher cross-section and a higher probability to get collisions at every passing. This is achieved through a first damping after injecting the two beams and finally by focusing the beams just before the interaction point. The objective is to reach a size of 40 nm horizontally and 1 nm vertically (compared to 16000 nm×16000 nm for LHC).

An additional problem is the beamstrahlung emission from the colliding particles that will first reduce their energy and leads to a wider luminosity spectrum but also produce an additional background in the detector.

3.2 The detector

The current design for the CLIC detector includes two different detectors working on a same point. Only one at a time can be used but they would be moved a few times a year through a push-pull system. Those detectors, based on the ILC detector designs, are the International Large Detector (ILD) and the Silicon Detector (SiD) which has the particularity to have all its tracking system made of silicon. This study has been only focusing on ILD but all the different methods shown are easily adaptable to SiD. The layouts of the CLIC_ILD is described on the figure 3.2.

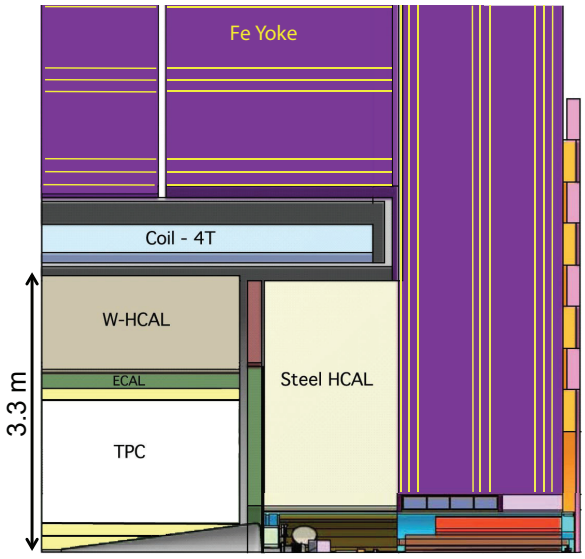


Figure 3.2: Longitudinal cross-section of the top quadrant of ILD from [7]

The first layer of ILD is a Vertex detector (VTX) made of three double layers of pixelated sil-

icon detector integrated into the tracking system. Since it is the closest layer to the interaction, the inner radius of the detector is about 31mm to avoid any hits from the e^-e^+ background. It has mainly a role of achieving optimal vertex reconstruction and flavour tagging, in order to do so it must have particularly high space resolution, cover almost all the geometry of the detector and time tag at at an order of 10^{-8} s.

The second layer is a large Time Projection Chamber (TPC) complemented with a small inner silicon tracker and a silicon envelope. The TPC design is based a lightweight field cage with a fine central high-voltage cathode, read out by Micro-Pattern Gas Detector. The goal of this layer is to track the different particles and reconstruct as well as possible their four-momentum. This is achieved by a high granularity particle flow calorimetry, allowing to reach a jet energy resolution of $\sim 3.5\%$.

The Electromagnetic CALorimeter (ECAL) uses silicon-tungsten sampling calorimeters optimised for particle flow and an emphasis has been made on the separation of adjacent electromagnetic showers. The calorimeter itself consists of 5.1×5.1 mm^2 pads in 30 longitudinal layers, where the rear ten absorber layers are twice thicker than the front twenty. This detector layer stops electrons and photons and can therefore help to differentiate hadronic showers from electro-magnetic showers.

The Hadron CALorimeter (HCAL), the goal of which is to stop all the remaining jets, uses tungsten as absorber material for the barrel. It makes it able to stop the high-energy showers while limiting the diameter of the detector and therefore of the solenoid. It is made of 1 cm thick absorber tungsten layers and 5 mm thick active layers. The active layers considered are scintillator tiles coupled with SiPMs (Silicon Photon Multipliers) and gaseous devices with different amplification structures but the final approach has not been decided yet.

To measure charged particle momentums and energy, a high-magnetic field of 4T is applied in the detector through a solenoid magnet. Superconducting solenoid coils are considered to be used as they have proven their efficiency in the CMS and ATLAS experiments at the LHC. The magnetic flux is returned through an iron yoke.

The muon system is integrated in the iron yoke and is made out of nine track sensitive layers of either glass RPC (Resistive Plate Chamber) or scintillators. It does not only enhance the detection of the muons but also works as a tail catcher for showers happening late in the calorimeters, helping to identify the high-energy hadrons.

Chapter 4

Jets Reconstruction

From the hits and the tracks detected in the detector to the complete interpretation of the event, many reconstruction processes have to take place. The complete software is illustrated in figure 4.1.

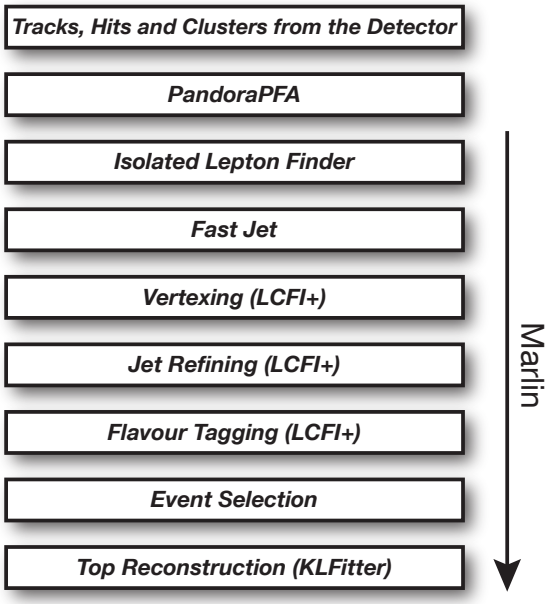


Figure 4.1: Algorithm of the top quark pair reconstruction

For this study $e^-e^+ \to t\bar{t}$ with $\sqrt{s} = 3$ TeV events have been generated. They went through a Monte Carlo (MC) simulation for the creation of the truth particles using PYTHIA [15]. Then the physical reaction of the detector to the event generated has also been simulated through a MC simulation with GEANT4 [16]. The sample files contain the particles information at the MC generator level as well as the tracks reconstructed by PANDORA using the simulated hits, cluster and tracks in the detector.

4.1 PANDORA

The first step is to interpret the different signals of the detector and combine them in the right way. This is particularly complex when two jets are very close, a common case at high energy, and have to be distinguished. The PANDORA Particle Flow Analysis (PFA) is a package of

more than 60 different algorithms adapted specifically to the CLIC detectors (ILD and SiD) and especially the high-granularity calorimeters. Instead of focusing on the energy reconstruction, PANDORA aims at give the best direction reconstruction. Particle flow reconstruction is facilitated by the high granularity of the calorimeter. The PANDORA algorithms uses the information of the tracker and continue the track reconstruction in the calorimeter, the energy being secondary. This way the resolution on the charged particles, representing 60% of the particles, is considerably improved.

PANDORA also filters the tracks and hits coming from the background and those coming from the interaction by imposing cuts on the transverse momentum and the detection time. The goal is to remove all the $\gamma\gamma \rightarrow$ *hadrons* background due to the beam interaction with the detector [17]. Different level of criteria are applied to give more or less restricted collections of particles. A more detailed description of the selection criteria is shown on the table 4.1 and the illustration of the effect of the cut is shown in figure 4.2

	Loose Configuration		Tight Configuration	
Region	**p_T range [GeV]**	**Time [ns]**	**p_T range [GeV]**	**Time [ns]**
Photons				
Central	$0.75 \leq p_T < 4.0$	$t < 2.0$	$1.0 \leq p_T < 4.0$	$t < 2.0$
$\lvert\cos(\theta)\rvert \leq 0.975$	$0 \leq p_T < 0.75$	$t < 2.0$	$0.2 \leq p_T < 1.0$	$t < 1.0$
Forward	$0.75 \leq p_T < 4.0$	$t < 2.0$	$1.0 \leq p_T < 4.0$	$t < 2.0$
$\lvert\cos(\theta)\rvert \leq 0.975$	$0 \leq p_T < 0.75$	$t < 1.0$	$0.2 \leq p_T < 1.0$	$t < 1.0$
Neutral hadrons				
Central	$0.75 \leq p_T < 8.0$	$t < 2.5$	$1.0 \leq p_T < 8.0$	$t < 2.5$
$\lvert\cos(\theta)\rvert \leq 0.975$	$0 \leq p_T < 0.75$	$t < 1.5$	$0.5 \leq p_T < 1.0$	$t < 1.5$
Forward	$0.75 \leq p_T < 8.0$	$t < 2.5$	$1.0 \leq p_T < 8.0$	$t < 1.5$
$\lvert\cos(\theta)\rvert \leq 0.975$	$0 \leq p_T < 0.75$	$t < 1.5$	$0.5 \leq p_T < 1.0$	$t < 1.0$
Charged particles				
All	$0.75 \leq p_T < 4.0$	$t < 3.0$	$1.0 \leq p_T < 4.0$	$t < 2.0$
	$0 \leq p_T < 0.75$	$t < 1.5$	$0 \leq p_T < 1.0$	$t < 1.0$

Table 4.1: Loose and tight selection criteria used to reject $\gamma\gamma$ →hadron background in ILD (from [18]).

4.2 Isolated Lepton Processor Finder

Once the particles are reconstructed the next step is to categorize the event, by trying to find isolated leptons. This is done by using both a tau finder and the isolated lepton finder. In this study, the tau finder has not been used, but the tau events have been rejected at the truth level.

In the case of a semileptonic event, one of the W will decay into a charged lepton/antilepton and a antineutrino/neutrino. Muons and electrons can be easily identified in the detector and there are methods to identify taus.

A set of tuned parameters is used to distinguish an isolated lepton track among all the other tracks: the most specific property of the isolated lepton is its isolation, meaning that there are no or few tracks around the lepton track. Therefore one of the parameter looked at is the cone energy, i.e. the sum of the energy's track located inside a cone of an angle θ centred on the lepton track, compared to the track energy.

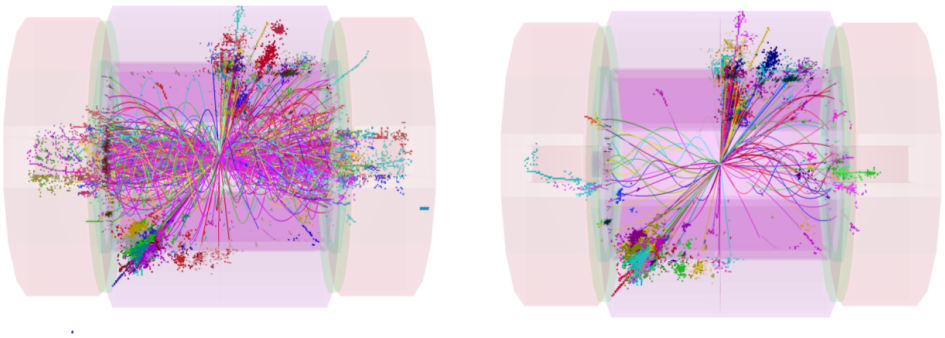


Figure 4.2: Reconstructed particles for a time window of 10 ns (100 ns in HCAL barrel) in a simulated $e^- + e^+ \to H^+H^- \to t\bar{b}b\bar{t}$ event in the CLIC ILD detector. The figure on the left represent the raw reconstruction and the figure on the right is after using the tight cut selection (from [18].

Another important attribute is that the isolated lepton is high-energy. The impact parameters, D_0, Z_0 and R_0 detailed in figure 4.6, can also be useful to distinguish efficiently the lepton track from soft leptons creatd in the hadronic and electromagnetic showers.

Finally electrons and muons have a very different behaviour in the calorimeters, while the electrons stop in the ECAL, muons only lose a fraction of energy in the HCAL. All those parameters have been tuned empirically, on a sample of 49500 $e^-e^+ \to t\bar{t}$ events, in order to optimize the efficiency defined by

$$\text{eff} = N_{\text{semileptonic event with exactly 1 letpton}}/N_{\text{semileptonic events}}, \tag{4.1}$$

while keeping the highest purity possible, defined by

$$P = N_{\text{true leptons}}/N_{\text{leptons found}} \tag{4.2}$$

The selection process simply selects tracks, by applying criteria on the previous cited parameters. The results on the training sample are shown on figure 4.3 and table 4.2 while the empirically tuned cuts are detailed in table 4.3.

There are many parameters involved, but the most critical cuts concern the cone energy, through two different cuts: the size of the cone, defined by $\cos\theta$ and the polynomial used to reject high-energy non isolated track as shown on 4.4a. Also there is a clear distinction between muon and electrons behaviour in the calorimeters as shown on 4.4b. The comparison for the other parameters are shown in appendix on figure 8.1.

This process is crucial when using data only as it determines the channel studied, defining the topology and therefore the number of jets to reconstruct by the next algorithms. Here, the efficiency is $83.5 \pm 0.3\%$ and the purity is about $93.5 \pm 0.1\%$. While the efficiency is not perfect, mainly because of some leptons that have a low energy or that are hidden in a hadronic shower, the purity is very high and the misidentified event has low chance to pass the selection cut applied afterwards concerning the jets momentum and total missing energy. However this process is rather simple and could probably improved.

4.3 FastJet

The FastJet Processor [19] aim is to reconstruct a defined number of jets using the longitudinally invariant k_T jet algorithm. It assembles tracks together if their angular distance

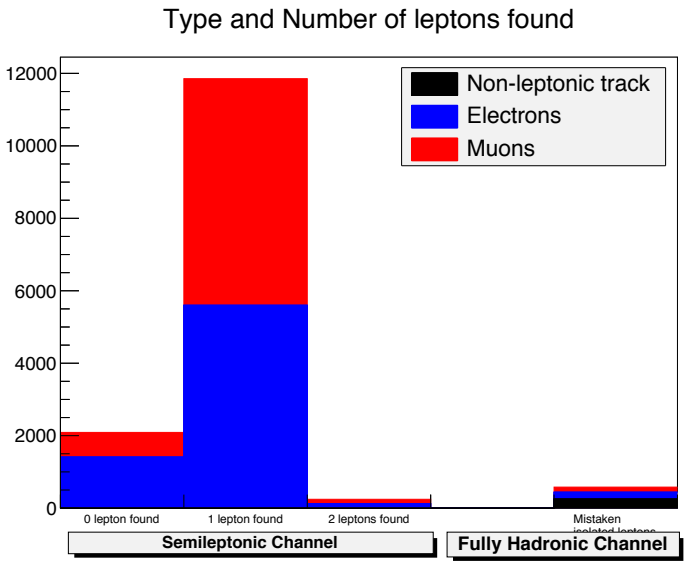


Figure 4.3: Number of leptons detected in the different channels and type of mistaken isolated lepton (using PANDORA PID information) for the semileptonic channel. On the left the colours represent the lepton that should be found in the event, on the right it represent the type of particle that has been mistaken for an isolated lepton

$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\theta)^2}$ is smaller than a given value (depending of the type and energy of the event, here 0.7) until the number of left tracks has reached a value specified by the user. However if a track is too close to the beam axis, it will be "removed" from the collection but its momentum will not be considered as missing as it will be merged to an artificial "beam jet". This process is fast and simple but has the inconvenience to not take into account the vertices of the event. It can be used alone to have an quick overview of the event or, and this is the case here, to remove all tracks too close to the beam. It completes the $\gamma\gamma \rightarrow$ hadrons background cleaning already made by PANDORA. and also removes tracks which would be misleading in the jet and vertex reconstruction.

4.4 LCFI+

LCFI+ [20], standing for Linear Collider Flavour Idenficiation, is a dedicated software developed by the linear collider community specifically for the ILC and CLIC detectors. It aims at reconstructing the jets and the vertices of the event to finally attribute flavour weights to the jets. The software is made of different processes with different uses, the ones used in this study are specified and explained:

4.4.1 Vertexing

Knowing the position of the primary and secondary vertices, described in figure 4.5, is not only crucial to determine the impact parameters of the different jets, it can also helps to reconstruct more accurately the different jets. The impact parameters are defined here as D_0 the transversal component, Z_0 the longitudinal component and $R_0 = \sqrt{D_0^2 + Z_0^2}$. D_0 and Z_0 are illustrated on the figure 4.6.

For the primary vertex finder a tear-down method is used while for the secondary vertices a build-up method is prefered [21]. In both methods Minuit2 [22] is used to minimize χ^2, the value computed by the distance between each track and the supposed vertex point, divided by

Total Number of events	49500	
Fully Hadronic Channel		22843 Events
Mistaken Isolated Electrons	Mistaken Isolated Muons	Other tracks mistaken
141	187	252
Semileptonic Channel:	**Electronic**	**Muonic**
# events	7133	7056
0 lepton found	1410	680
1 lepton found	5600	6255
2 leptons found	123	121
efficiency	78.5%	88.6%
Total Efficiency	83.5%	
Purity	93.5%	

Table 4.2: Result of the Isolated Lepton Processor applied on a sample of 49500 $e^+e^- \to t\bar{t}$ events. Purity and efficiency are defined in the text.

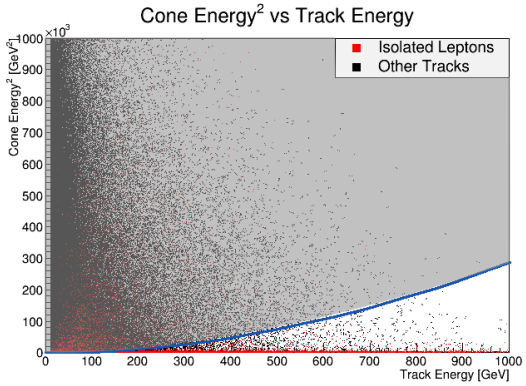
Cut Parameter	Value(s)
$\cos\theta$ (min)	0.999
E_{cone} (min/max) [GeV]	0/5000
E_{track} (min/max) [GeV]	20/1000
$E^2_{\text{cone}} < A \cdot E^2_{\text{track}} + B \cdot E_{\text{track}} + C$, $A/B/C$	0.3/1/100
D_0 (min/max) [mm]	0/0.05
Z_0 (min/max) [mm]	0/0.05
R_0 (min/max) [mm]	0/0.05
$e^-_{\text{deposit}} : \frac{E_{\text{ECAL}}}{(E_{\text{ECAL}}+E_{\text{HCAL}})}$ (min/max)	0.9/1
$\mu^-_{\text{deposit}} : \frac{E_{\text{ECAL}}}{(E_{\text{ECAL}}+E_{\text{HCAL}})}$ (min/max)	0/1

Table 4.3: Tuned parameters for optimal isolated lepton finding

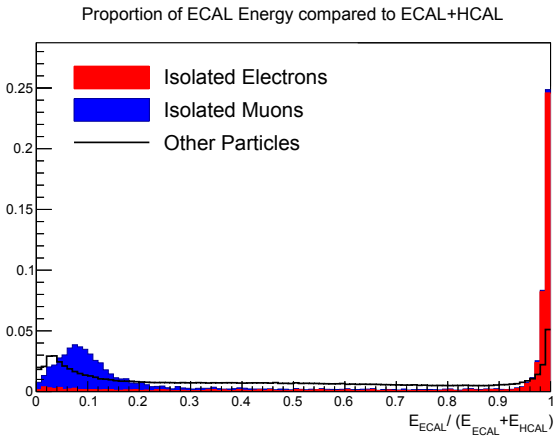
the track covariance matrix trace, summed for all tracks which are being considered for the vertex.

The tear-down method consists of taking all the tracks in the event and make it the list of the primary track candidates. The initial vertex point is taken as the closest point to all tracks without considering the errors, then the loop starts:
The track with the biggest χ^2 contribution is removed from the list. The vertex is fitted again with the new list of tracks. This procedure is repeated until the χ^2 value is less then 25.

The build-up method has an opposite way of working. First, all the tracks removed from the primary track candidates, the non-primary tracks, are taken. Pairs of tracks are made requiring that the χ^2 value is less than 9 for each vertex pair. Selection criteria on the momentum of the tracks are also applied on the created vertices in order to improve the quality of the reconstruction. The vertices created this way become the initial vertices. The non-primary tracks are looped over again to test against every of these initial vertices. For each try χ^2 is computed again and, if the resulting vertex pass all the cuts cited earlier, it is retained. At the end, resulting vertices are checked to remove duplicate vertices and multiple uses of tracks.



(a) Cone Energy squared compared to the Track Energy with cutting polynomial as described in table 4.3



(b) Energy deposited in ECAL over energy deposited in both calorimeters for isolated electrons and muons compared to the other tracks

Figure 4.4: Parameters compared for all tracks and isolated lepton track

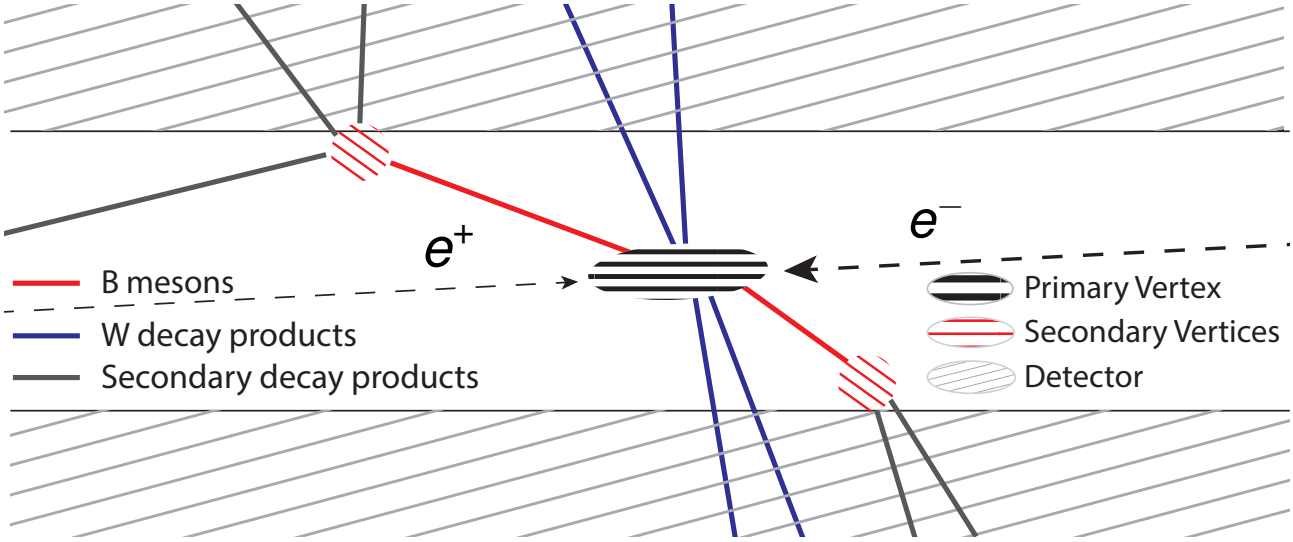


Figure 4.5: Representation of the vertices in a fully hadronic event

The secondary vertices are passing another selection round which aims at reducing V^0 and fake vertices: the vertices with a mass close to the $K^0_{S/L}$ and Λ^0 ones are rejected. These two particles are strange baryons and will fake secondary vertices by decaying into charged particles. Which would induce error for the jet reconstruction and the flavour identification (especially of the charms). Also vertices that are too far (>30mm) or too close (<0.3mm) from the primary vertex are also eliminated.

4.4.2 Refining Jets

While most reconstruction methods use the jets to reconstruct the vertices, it is the opposite in LCFI+. It improves the splitting efficiency between the jets which are too close. The k_T-Durham algorithm[19, 22] is used to form the jets. It is pretty similar to the kt-algorithm detailed earlier in FastJet except that the angular distance is replaced by

$$d_{ij} = 2\min\left(E_i^2, E_j^2\right)(1-\cos\theta_{ij}) \tag{4.3}$$

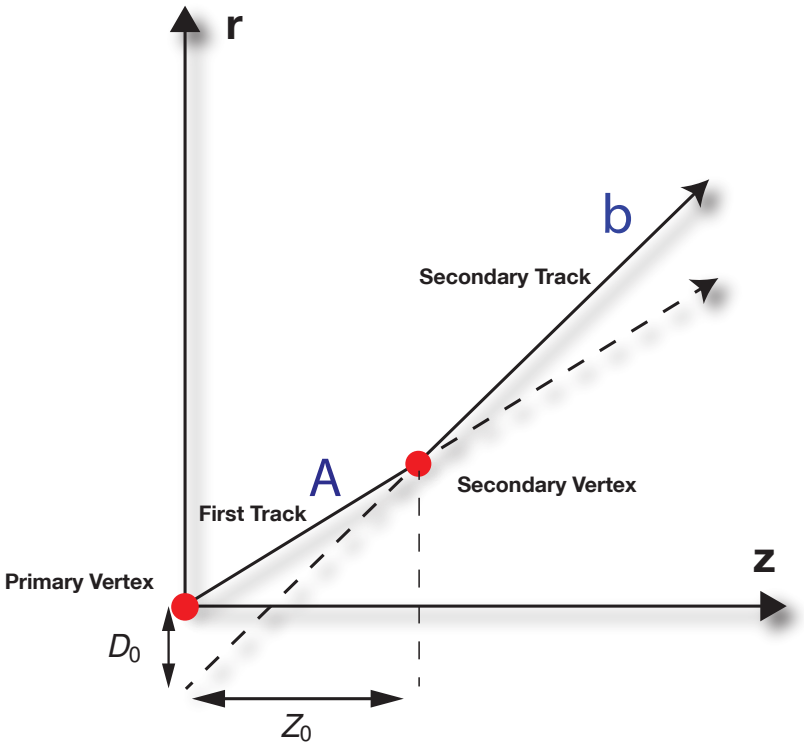


Figure 4.6: Impact parameters of the track b decaying from a particle A

Also in order to separate correctly the jets one from another, penalty terms are added to split up correctly the secondary tracks.

4.4.3 b - and c-Tagging

Quark flavour tagging is the identification of the flavour of the quark initiating the jet. It is unfortunately impossible to identify lighter quarks (u, d, s), since they are indistinguishable one from another. It is however possible to identify charm (c-tagging) and bottom quarks (b-tagging). The c-tagging is unfortunately very difficult to do due to the lower mass of the charm quark, in this study the efficiency of the c-tagging is too low and therefore is not used.

The tagging of the jets is done by passing a selection of parameters (including the impact parameters that have been computed through the vertexing) through a Boosted Decision Tree (BDT) [23]. When a jet containing a set of PID (Particle Identification) values goes through a BDT, it is submitted to a series of interdependent decisions. For every parameter, a decision is made in function of a splitting value, leading to another decision depending on another parameter. For each step, the decision will give a probability on the identification of the particle flavour. At the end of the tree the combination of these probabilities will give a weight on the b- or c-tagging of the particle between 0 and 1.

The BDT is "trained" by using an independent sample of $e^+e^- \to Z\nu\nu$ events. The Z boson can decay into any possible pair of particle-antiparticle except for the top quark. The hadronic channel of the Z are selected and once the jets are reconstructed (using the same vertices and jets clustering algorithms) NTuples trees are created, storing all the information about the different jets, and categorized as signal or background in function of their flavour. These NTuple trees are used for the decision tree creation: an analysis is done to defihe best order of parameters to use in the decision as well as the splitting values. At each decision there is a

proportion of signal events and background events used to define the purity $P = \frac{\sum_s W_s}{\sum_s W_s + \sum_b W_b}$ where $W_{s/b}$ is a weight attributed to each signal/background event. When testing an event, at the end of the tree if the $P < 0.5$ it is considered as background and if $P > 0.5$ it is considered as a signal. That is where the boost comes in: when an event from the training sample is misidentified with the first tree, its weight is increased or "boosted" and the process of creating a tree is repeated until an error threshold is reached. The tree is finally tested to check the tagging efficiency, the results is shown on figure 4.7. This figure shows how much background is kept in function of how much signal we want. For example, to keep 90% of the b-jets, the cut used would keep (using the function shown on the figure) 50% of the charm background and 15% of the lighter flavours background.

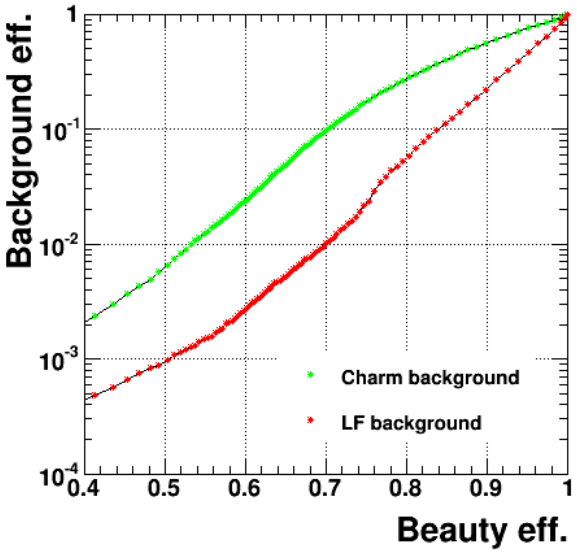


Figure 4.7: b-tagging tested on a training sample

4.5 Event Selection

Since the reconstruction of the jets is not fully efficient, the result needs to be checked before pursuing the kinematic fitting. A series of specific cut is applied for each kind of event as detailed in table 4.4. The cut put on the transverse momentum is directly aimed at removing badly reconstructed jets while the $|\eta|$, the pseudo-rapidity $\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$, cut helps to remove jets whose direction would be too close to the beam jets.

Channel	**Fully Hadronic**		
Reconstructed Jets	$N = 6$	$\lvert p_T\rvert > 20$ GeV	$\lvert\eta\rvert < 2.5$
Channel	**Semileptonic**		
Reconstructed Jets	$N = 4$	$\lvert p_T\rvert > 20$ GeV	$\lvert\eta\rvert < 2.5$
Missing Transverse Energy		MET>20 GeV	
Leptons (e/μ)	$N = 1$	$\lvert p_T\rvert > 20$ GeV	$\lvert\eta\rvert < 2.5$

Table 4.4: Selection cuts on the reconstructed fully hadronic and semileptonic events

To summarize the final reconstruction output, we categorized the events in function of their channel using the MC generator information and rejected the non-workable ones. The isolated lepton for the semileptonic events have been identified, the particles in the detector have been filtered to remove the beam jet radiation background. The primary and secondary vertices have been localized and a defined number of jets has been reconstructed using the vertices and the tracks. A tagging probability has been attributed to all the jets to distinguish the light quarks from the b quarks, using the jet and vertices informations. Finally events have been checked in function of the reconstructed information.

Chapter 5

Top Quark Pair Reconstruction

Assuming that the jets have been correctly reconstructed the last task is to correctly fit them to the topology (see chapter 2). The Kinematic Likelihood Fitter (KLFitter) [10] is a framework developed at LHC, it has been developed for the top quark pair reconstruction, but its design allow to adapt it easily for other processes and topologies. Originally developed for the LHC and more specifically the ATLAS framework (working with D3PD files), it has been modified and rewritten in some parts for this study to work with the CLIC framework.

5.1 KLFitter

The fitting of the jets consists in associating correctly each jet to a quark in the topology. Statistically, since there are N reconstructed jets to associate with N quarks, there are $N!$ possible permutations (24 for the semileptonic channel and 720 for the fully hadronic channel). However since the light quarks are undistinguishable only through the kinematics and since the top quarks are not numbered the number of permutations reduces respectively to 12 and 90.

In order to find the best permutation, the KLFitter's algorithm will work out what is the most probable using the kinematics, the b tagging previously computed but also the resolution of the detector, by maximising a likelihood function. Instead of directly using the measured four-momentum of the reconstructed jets to reconstruct the top quarks, KLFitter will perform a multi-variant analysis to find the kinematics variables that will both give the best possible invariant masses for the top quarks and the W boson, and be coherent with the measured variables. The methods depends directly of the topology of the event, for example for the semileptonic case, the likelihood function is defined by:

$$\begin{aligned} L = & B\left(m_{q_1q_2q_3}|m_{\text{top}},\Gamma_{\text{top}}\right)\cdot B\left(m_{q_1q_2}|m_{\text{W}},\Gamma_{\text{W}}\right) \\ & \times B\left(m_{q_4l\nu}|m_{\text{top}},\Gamma_{\text{top}}\right)\cdot B\left(m_{l\nu}|m_{W},\Gamma_{W}\right) \\ & \times \sum_{i=1}^{4} W_{\text{jet}}\left(E_{\text{jet,i}}^{\text{meas}}|E_{\text{jet,i}}\right)\cdot W\left(E_{l}^{\text{meas}}|E_{l}\right) \\ & \times W_{\text{miss}}\left(E_{x}^{\text{miss}}|p_{x}^{\nu}\right)\cdot W_{\text{miss}}\left(E_{y}^{\text{miss}}|p_{y}^{\nu}\right) \end{aligned} \tag{5.1}$$

There are eight free parameters in this function: the top mass, the jet and lepton energies $E_{\text{jet,i}}/E_l$ and the neutrino transverse momentum components p_x^ν and p_y^ν. $\Gamma_{\text{top}}, m_{\text{W}}$ and Γ_{W} are fixed constants whereas $E_{\text{jet,i}}^{\text{meas}}, E_l^{\text{meas}}$ and $E_{x/y}^{\text{miss}}$ are the measured energies of the jets, of the lepton and the missing transverse energy. $m_{q_1q_2q_3}, m_{q_1q_2}$, $m_{q_4l\nu}$ and $m_{l\nu}$ are respectively the invariant masses of the hadronic top and W and the leptonic top and W. The z component of the neutrino momentum can be either fixed by solving the equation $m_W^2 = (p_\nu + p_l)^2$ and taking the most probable solution (with the maximum likelihood) if there are two solutions, setting it

to $p_z^\nu = 0$ GeV if there are none, it is possible, however, to leave it as a free parameter. The top mass is not fixed but the relation $m_{\text{top hadronic}} = m_{\text{top leptonic}}$ puts a constraint on it. However for other purposes than the measurement of the top quark mass, it is possible to artificially fix the top quark mass to a certain value. B is the relativistic Breit-Wigner function:

$$B(E|M,\Gamma) = \frac{2\sqrt{2}M\Gamma}{\pi\sqrt{M^2+\sqrt{M^2(M^2+\Gamma^2)}}} \frac{\sqrt{M^2(M^2+\Gamma^2)}}{(E^2-M^2)^2+M^2\Gamma^2} \tag{5.2}$$

Finally W and W_{miss} are the transfer functions, expressing how likely the energy used in the parameter set will be coherent with the measured energy. The real parametrisation of these functions and their uses is detailed in the section 5.4.

For the semileptonic channel, it is only possible to do the analysis for one type of lepton at a time. It is not an issue since one would expect this differentiation in a data analysis as well.

5.2 Bayesian Analysis

KLFitter is based on Bayesian Analysis Toolkit (BAT) [24], a collection of algorithm solving problems using Bayes' theorem:

$$P(\vec{\lambda}|\vec{D}) = \frac{P(\vec{D}|\vec{\lambda})P_0(\vec{\lambda})}{\int P(\vec{D}|\vec{\lambda})P_0(\vec{\lambda})d\vec{\lambda}} \tag{5.3}$$

This theorem says that the probability for a parameter set $\vec{\lambda}$ given some data $\vec{D}$, i.e. the posterior probability, is proportional to the probability of data given parameters, called the prior probability. The probability is then normalized by integrating the numerator over the allowed region of $\vec{\lambda}$. This theorem gives very powerful methods since it allows to predict probabilities using past data. In our case $P(\vec{D}|\vec{\lambda})$ represents the combination of the transfer functions while $P_0(\vec{\lambda})$ is the combination of Breit-Wigner functions, $P(\vec{\lambda}|\vec{D})$ representing the normalized likelihood, described in equation 5.1.

Using different tools like CUBA, MINUIT [22] or RooStats, the BAT can determine the parameter set giving the best likelihood. The default tool used to optimize the likelihood is MINUIT with the minimization method MIGRAD. One must be careful with the ranges allowed for the parameters. They are determined here by the σs of the different transfer functions.

5.3 Use of b-tagging

Not only the b-tagging is essential for removing the background but it also helps to improve the reconstruction efficiency. The b-tagging can used in KLFitter through different methods (these are the original names):

- **Veto**: The jets associated to light quarks must not be b-tagged
- **VetoLight**: The jets associated to b quarks must be b-tagged
- **VetoBoth**: Combination of the two previous vetoes.

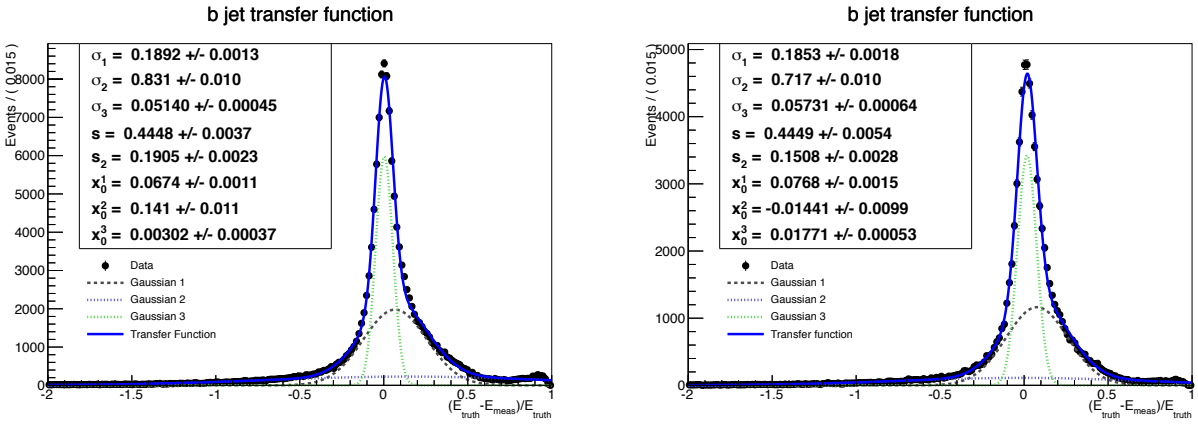
The term b-tagged means that the b flavour weight of the jet is over a threshold defined by the user. The veto works such that if the permutation does not respect the cuts required its probability is set to 0. However, it is possible to directly remove the permutations not passing the cuts to save computation time, this option is called NoFit. The number of permutation left when using the VetoLight or the VetoBoth method is 2 for the semileptonic channel and 8 for the fully hadronic channel.

5.4 Transfer Functions

The transfer functions introduced previously are functions that makes a link between the jet energies and the energy of the particles at parton-level. This difference of energy is due to the detector resolution, to the reconstruction mistakes or to the background radiation. These functions are derived from the distribution of $\left(E^{\text{truth}} - E^{\text{meas}}\right)/E^{\text{truth}}$, evaluated on enough $e^-e^+ \to t\bar{t}$ event to get significant statistics. In the original KLFitter, designed for LHC experiments double gaussian transfer functions are used for jets, electrons and muons, a simple gaussian transfer function is used for the missing energy and the transverse momentum is used instead of the energy for the muons. However due to the difference of resolution with CLIC, some changes were necessary: triple gaussian functions are used for jets and electrons resolution, a double gaussian is used for the muons using the energy resolution and simple gaussian for the missing energy has been replaced by a Crystal Ball function added to a gaussian [25] (gaussian with a tail). The N-gaussian are defined as

$$f(x) = \frac{1}{\sqrt{2\pi}\sum_i^N s_i\sigma_i}\left(\sum_i^N s_i \exp\left(\frac{-\left(x-x_0^i\right)^2}{2\sigma_i^2}\right)\right), \qquad s_N = \left(1 - \sum_i^{N-1} s_i\right)$$

Where x is defined by $\frac{E^{\text{truth}}-E^{\text{meas}}}{E^{\text{truth}}}$, and x_0^1, x_0^2, σ_1,σ_2, s are the double gaussian constants found by fitting a training sample. The functions are binned in function of the jet's $|\eta|$. In this work the bins used are $[0; 0.8[$, $[0.8; 1.5[$, $[1.5; 2.5[$, jets with $|\eta| > 2.5$ would not pass the selection cuts. An example on how the binning make a difference is shown on figure 5.1. A tool exists to fit directly the distributions : TFTool [26], its main feature is to have fitting parameters directly depending of the parton energy, unfortunately it has been developed for LHC data, in the form of D3PD root files. Instead of adapting this tool to the CLIC framework, a more simple fitting has been made using RooFit [27], ignoring then the energy dependence of the transfer functions parameters. It is an approximation that can be improved later by adapting the TFTool for CLIC. The transfer functions have been fitted on a sample of 49500 $e^+e^- \to t\bar{t}$ events, they are plotted in the appendix in figure 8.2 for the b-jets, 8.3 for the light jets, 8.4 for the missing energy, 8.5 for the electrons and 8.6 for the muons.



(a) $|\eta| < 0.8$ (b) $0.8 \leq |\eta| < 1.5$

Figure 5.1: Binned transfer functions for b jets using a data sample of 49500 $e^+e^- \to t\bar{t}$ events.

Chapter 6

Results and analysis

On table 6.1, the number of events used for this sample as well as the number of event passing the selection cut and the number of events with jets matching the particle at truth-level (MC generator level) are displayed both for the fully hadronic and the semileptonic channel, as well as the number of events passing the different b-tag cuts. An event is "matched" means that every jet has been associated independently to a quark at the MC generator-level, where associated means that the angular distance ΔR satisfy the condition $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$.

Channel	**Fully Hadronic**		**Semileptonic** Electron		Muon	
MC Events	22801		7133		7056	
Selected	17539	76.9%	4398	61.7%	4964	70.4%
Veto	17237	75.6%	4340	60.8%	4902	69.5%
VetoLight	6193	27.2%	1498	21.0%	1674	23.7%
VetoBoth	5891	25.7%	1440	20.2%	1612	22.8%
Matched	3612	15.8%	1349	18.9%	1564	22.2%
Veto	3575	15.6%	1339	18.8%	1534	21.7%
VetoLight	1430	6.3%	557	7.8%	624	8.8%
VetoBoth	1393	6.1%	547	7.6%	614	8.7%

Table 6.1: Number of events for fully hadronic and semileptonic channel before and after the different cuts

6.1 b-Tagging results

The validity of the b-tagging is checked by looking at the flavour tag weight of the jets that could be associated to b quarks using the information at the truth level. On figure 6.1, the weight of the two bests candidates as well as the weight of the other jets is shown. As one can see, on both channels, the flavour tagging is very accurate. The tail can be attributed to the method of the identification of the best candidates, if the reconstruction efficiency is not good enough, it deteriorates the tagging quality.

From this plots the cutting value to decide if a jet is "b-tagged" has been set to 0.8. With this cut, it keeps $61.9 \pm 0.7\%$ of the jets identified as b jets and removes $97.9 \pm 0.6\%$ of the other jets for the semileptonic channel and keeps $61.2 \pm 0.7\%$ of the jets identified as b jets and remove $98.3 \pm 0.5\%$ of the other jets for the fully hadronic channel.

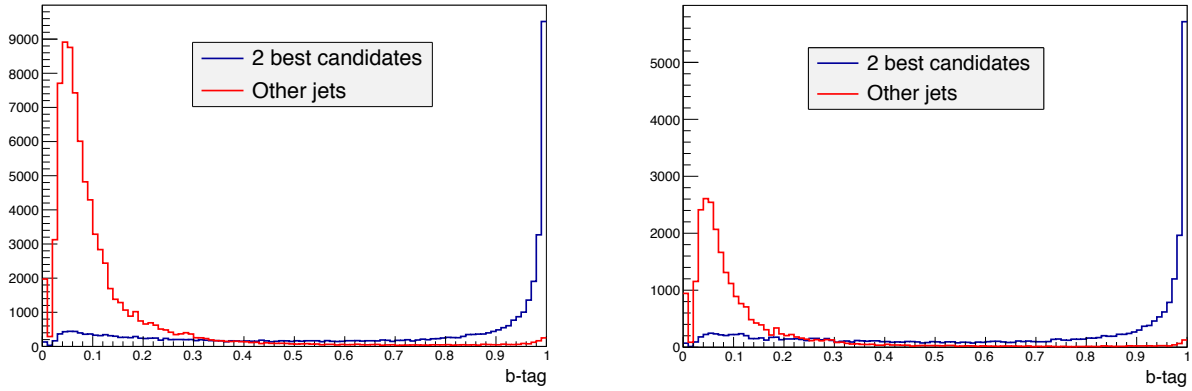


Figure 6.1: *b*-tagging of the two best b-jet candidates (using truth information) compared to the other jets, for the fully hadronic case (left) and the semileptonic case (right)

6.2 Mass Reconstruction

The final goal of this work is to measure the top quark mass. The mass has been left as a free parameter for the computation, and the results are shown on figure 6.2 for the fully hadronic channel and on figure 6.3 for the semileptonic channel. A peak is found around $m_{\text{top}} = 173$ GeV as expected, however the tail going heavier masses is quite important. On the plots one can see that the use of the *b*-tagging reduce slightly the width of the peak and reduces the tail size. The reconstruction quality however, has a much more important effect, showing the importance of the jet reconstruction. Having the correct permutation does not improve considerably the result, but since the VetoBoth method has been used the permutation of the matched events is usually correctly found.

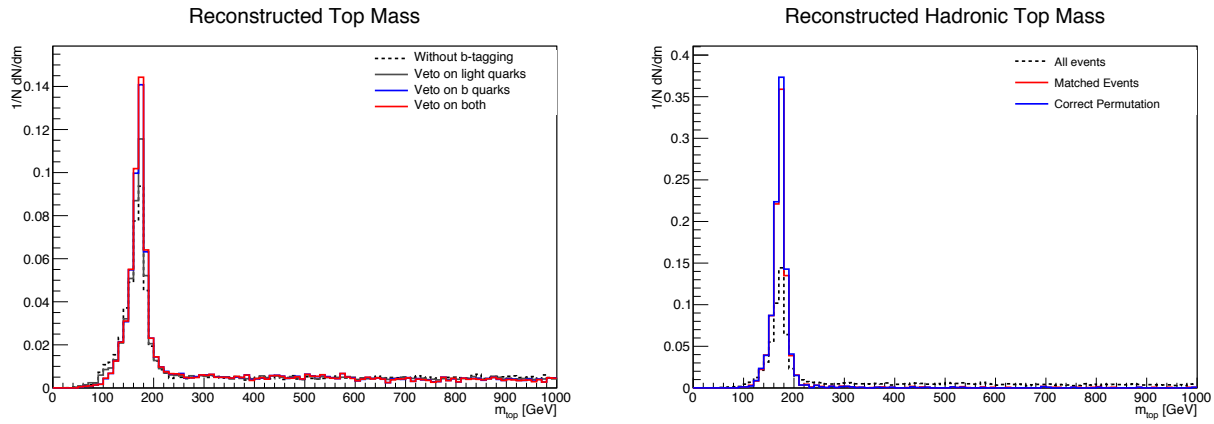


Figure 6.2: Reconstructed top quark mass for the fully hadronic channel. On the left the effects of the b-tagging methods. On the right the effects on the efficiency of the reconstruction quality with the VetoBoth method.

On the semileptonic events plots on figure 6.3, one can see a peak at 1 TeV, this is a non-physical effect due to the KLFitter algorithm, which limits the top mass to 1 TeV. This problem should be corrected in the future. Also the tail is a lot more important for semileptonic events than for fully hadronic events. This is linked to the fact that semileptonic events have a lot of missing energy, and therefore of missing information, leading to a deteriorated reconstruction, as well. Also it appears to make no difference if the top is on the leptonic side or on the hadronic

side.

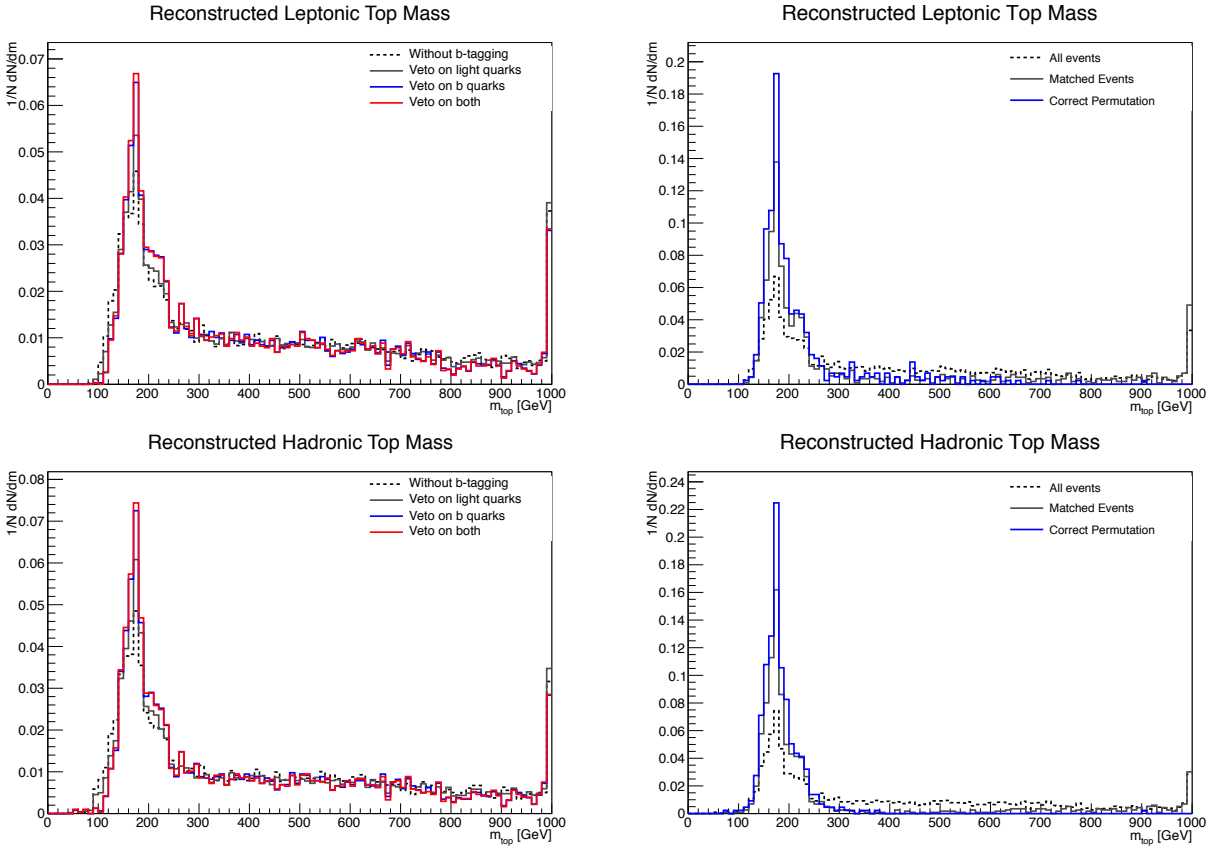


Figure 6.3: Reconstructed top quark mass for the semileptonic channel. Top row represents the leptonic top quark mass while bottom row represents the hadronic top quark mass. On the left column the effects of the b-tagging methods while on the right the effects on the efficiency of the reconstruction while using the VetoBoth method

On figure 6.4, the the two top quark masses computed for the fully hadronic channel are compared. As one can see, when the top quark mass is left as a free parameter, both top tends to have the same mass. When the mass is not fixed instead of a diagonal two lines are observed. When the KLFitter cannot attribute the top quark mass to both sides of the event, it gives priority to one top quark at the cost of the other top quark accuracy. A similar behaviour is present for the floating mass case.

On figure 6.5, the energy of the two top quarks from the same event in the fully hadronic channel is compared. The result is far from a perfect picture where both top quarks shares the same amount of energy. There are even surprising results where the total energy is less than 600 GeV or more than 3.6 TeV. This diffusion is particularly bounded to the initial state radiation and Bremsstrahlung effect on the beams, widening the energy spectrum, but can also be attributed to the hazard of the reconstruction.

Plots for the reconstructed mass for the fixed mass did not have any interest since they just represent a peak at the chosen value with a low tail for badly reconstructed events. The same goes for the W masses.

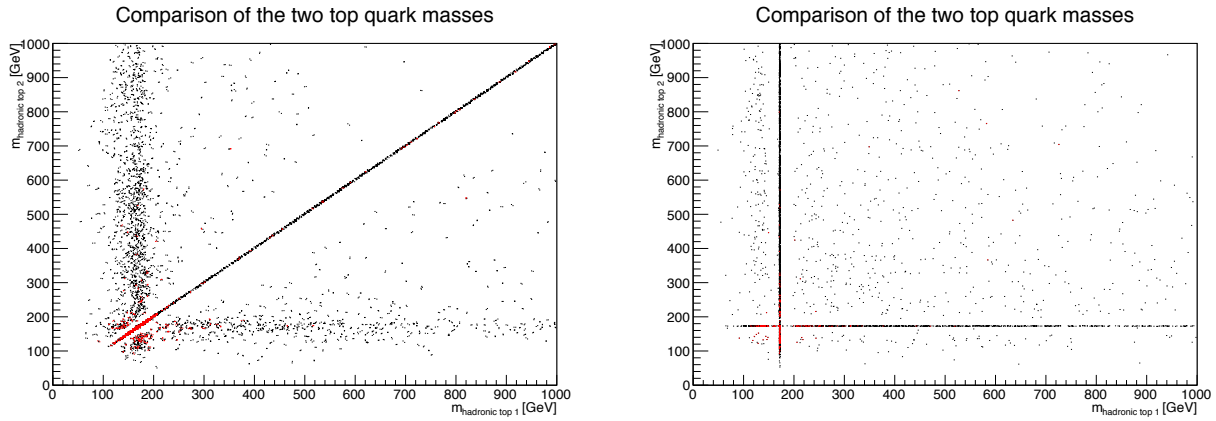


Figure 6.4: Reconstructed top quark masses compared to each other in the fully hadronic case while using the VetoBoth method, black dots represent the selected events while red is for the matched events. Top quark masses are left as float on left and are fixed both to $m_{\text{top}} = 173$ GeV on right

6.3 Pairing efficiency

The pairing efficiency (called reconstruction efficiency in [10]) represents the accuracy of the fitting, i.e. if the jets identified as W decays products are matching the W decays products particle at the MC generator level, and same goes for the b jets. This efficiency can only be reconstructed when all the reconstructed jets have been matched with the particles at the MC generator level. The results for the semileptonic channel and fully hadronic channel are detailed in table 6.1, the separated efficiencies for events with an electron or a muon is in the appendix on table 8.1.

		Pairing efficiency [%]				
Channel	Method	Overall	W	b	b-tag|b-jets	b-tag|light jets
Fully Hadronic	Likelihood	44.0 ± 0.8	64.0 ± 0.8	64.2 ± 0.8	35.6 ± 0.6	13.8 ± 0.4
	+Veto	73.4 ± 0.7	84.3 ± 0.6	84.4 ± 0.6	62.4 ± 0.5	0 ± 0
	+VetoLight	92.8 ± 0.7	94.7 ± 0.6	94.8 ± 0.5	100 ± 0	0.6 ± 0.1
	+VetoBoth	93.5 ± 0.7	95.0 ± 0.6	95.2 ± 0.6	100.0 ± 0	0 ± 0
		Overall	W_{had}	b_{had}	b_{lep}	b-tag|b-jets	b-tag|light jets
Semileptonic	Likelihood	21.2 ± 0.8	29.0 ± 0.8	33.3 ± 0.9	45.1 ± 0.9	46.2 ± 0.6	18.0 ± 0.4
	+Veto	32.1 ± 0.9	42.1 ± 0.9	39.9 ± 0.9	48.1 ± 1.3	62.7 ± 0.6	0 ± 0
	+VetoLight	37.4 ± 1.4	48.9 ± 1.4	42.0 ± 1.4	48.7 ± 1.4	100 ± 0	0.8 ± 0.1
	+VetoBoth	37.5 ± 1.4	49.2 ± 1.5	42.0 ± 1.5	48.8 ± 1.5	100 ± 0	0 ± 0

Table 6.2: Pairing efficiencies for different methods and channels

On figure 6.6, the reconstruction efficiencies for the fully hadronic channel is illustrated with the top quark mass left as float and fixed. As one can see the fixing the mass does not affect the reconstruction efficiency at all.

On figure 6.7, the difference between the reconstruction efficiencies between the semileptonic events with an electron and a muon is illustrated. First, one can see that the type of lepton in the semileptonic event does not make a difference, and it goes the same for the mass reconstruction.

The reconstruction efficiencies are a lot lower than the fully hadronic case. This could be due to the higher amount of missing energy, misleading the fitter. But there is probably another

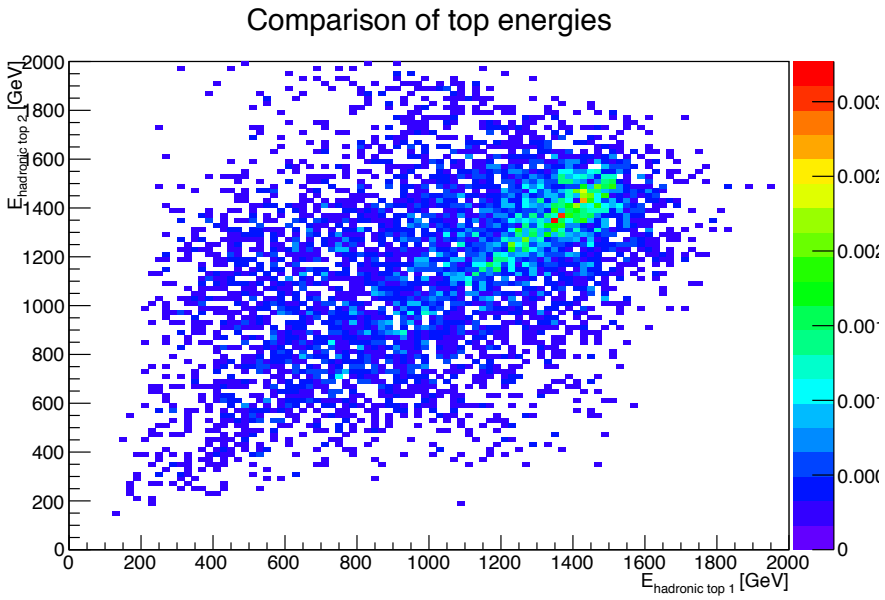


Figure 6.5: Top quark energies compared to each other for the fully hadronic channel using the VetoBoth method

internal error of the algorithm that still needs to be fixed, a lot higher reconstruction efficiencies are expected.

One can notice that the reconstruction efficiency of the leptonic b quark is slightly better than the hadronic one for all methods. This is expected since they belong to different tops and have completely different kinematics.

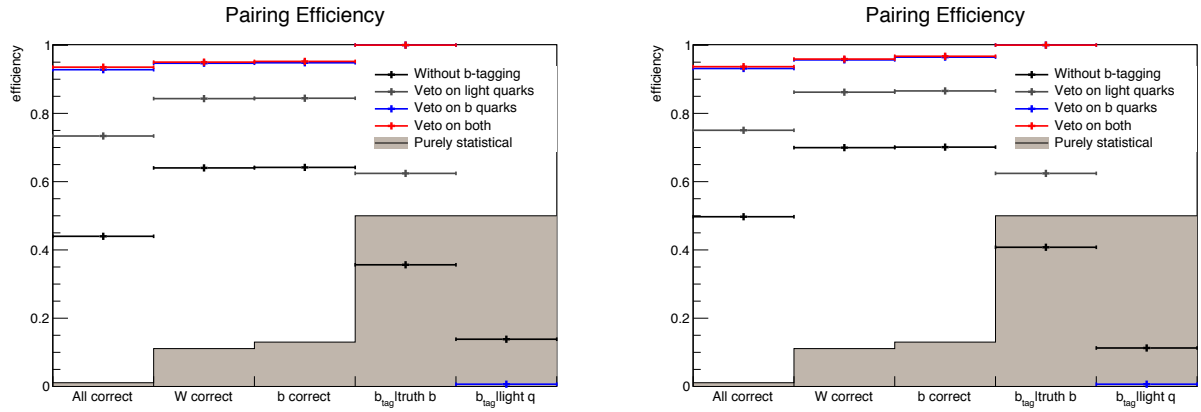


Figure 6.6: Reconstruction efficiency for the fully hadronic channel with top mass fixed (left) and left a a float (right)

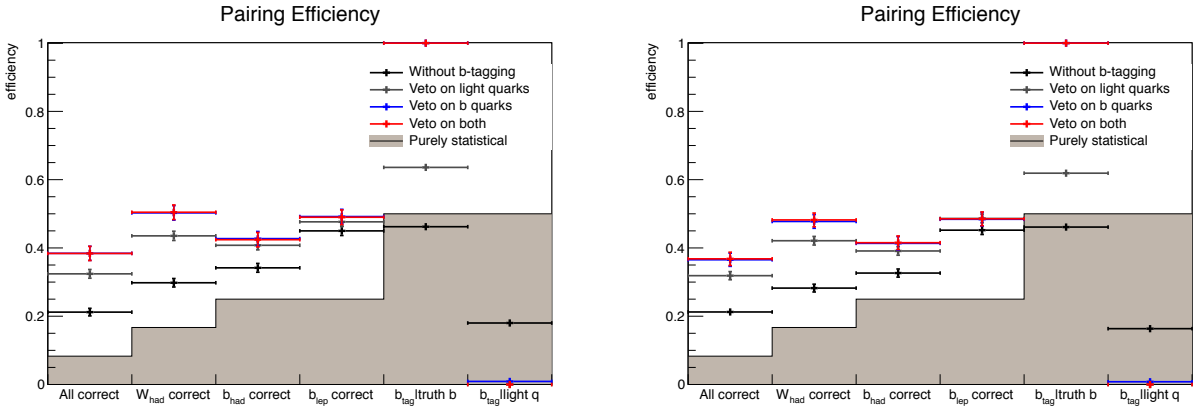


Figure 6.7: Reconstruction efficiency for the semileptonic channel, with top mass left as a float, for events with an electron(left) and events with a muon (right)

Chapter 7

Conclusion

In this work, the full reconstruction of the top quark in the fully hadronic channel $e^+e^- \to t\bar{t} \to bbqqqq$ and semileptonic channel$e^+e^- \to t\bar{t} \to bbqql\nu$ has been accomplished on a sample of 49500 $e^+e^- \to t\bar{t}$ events. It has been done by tuning the reconstruction algorithms to give the best jet reconstruction possible and by the adaptation of the KLFitter framework to the CLIC environment to fit the jets to the topology of the event. However only 15% to 22% of the events have a completely satisfying jet reconstruction. The flavour tagging is having better performances, identifying more than 60% of the b jets.

For the matched events the KLFitter algorithm reconstructs almost perfectly the fully hadronic topology when using the b-tagging, and the resulting mass spectrum is very satisfying. However it is not the case for the semileptonic channel, where even if the topology is more complicated to reconstruct due to the missing energy, there is still a lot of possibilities for improvements. The transfer functions has been also defined on statistically small sample and without the energy dependency required by the KLFitter. This approximation worked well for the fully hadronic events, but new events fitted with the TFTool which needs to be adapted would be a big improvement.

It would be now interesting to try the algorithm on background samples, especially to the single top quark production which have a close production cross-section in this range of energy. The single top quark production have the same particle signature than the top quark pair production, but with completely different kinematics. It would be interesting to see the efficiency of the KLFitter for separating the two processes. It would also be a great improvement to define likelihood functions adapted to the single top quark topology and use the results as a complement to the top quark pair production analysis.

The measurement of the top quark properties at $\sqrt{s} = 3$ TeV with the top quark pair production is also probably not the best choice. The production cross-section of the top quark pair would be a lot higher at $\sqrt{s} = 1.3$ TeV, $\sqrt{s} = 500$ GeV or even better at $\sqrt{s} = 350$ GeV. However the KLFitter is energy independent, this tool could be easily used on lower energies as long as transfer functions are made available.

It has been shown that fixing the mass of the top quark for the simulation was not changing much the performance of the fitter. This feature could be essential for the stop squark research $e^+e^- \to \tilde{t_1}\bar{\tilde{t_1}}$. Since the signature is the same except for an enormous amount of missing energy, the KLFitter could be directly used on these events to extract a signal: there are different methods to compute the mass of the stop squark using for example the top energy and fixing the mass of the neutralino.

The results leave a lot of openings, but the main achievement has been to make the KLFitter framework available to all the CLIC (and ILC) users for their analysis. A documentation for it will be produced soon.

Bibliography

[1] Michele Maltoni M.C. Gonzalez-Garcia. Phenomenology with Massive Neutrinos. *Phys. Rept.*, 460:1–129, 2008. arXiv:0704.1800.

[2] V. Trimble. Existence and nature of dark matter in the universe. *Rev. of Astronomy and Astrophysics*, 25:425–472, 1987.

[3] CDF ATLAS, D0 and CMS Collaborations. First combination of Tevatron and LHC measurements of the top-quark mass. arXiv:1403.4427.

[4] Future circular collider study. URL: https://espace2013.cern.ch/fcc/.

[5] Circular electron positron collider. URL: http://cepc.ihep.ac.cn.

[6] ILC Collaboration. Ilc technical design report. Technical report. URL: http://www.linearcollider.org/ILC/Publications/Technical-Design-Report.

[7] edited by P.Lebrun, L.Linssen, A.Lucaci-Timoce, D.Schulte, F.Simon, S.Stapnes, N.Toge, H.Weerts, and J.Wells. The CLIC Programme: towards a staged e+ e- Linear Collider exploring the Terascale, CLIC Conceptual Design Report. Technical report, CERN, 2012. URL: http://clic-study.web.cern.ch/.

[8] K.A. Olive et al. (Particle Data Group). *Chin. Phys.*, C38, 2014. URL: http://pdg.lbl.gov.

[9] A. Djouadi & S.Moch S. Alkhin. The top quark and higgs boson masses and the stability of the electroweak vacuum. *Phys. Lett.*, B716:214, 2012. arXiv:1207.0980v3.

[10] Kevin Kroeninger Johannes Erdmann, Stefan Guindon, Boris Lemmer, Olaf Nackenhorst, Arnulf Quadt, and Philipp Stolte. A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework. *Nuclear Instruments and Methods in Physics Research*, Section A 748:18–25, 2014. arXiv:1312.5595v2.

[11] T.Teubner A.H.Hoang. Top Quark Pair Production close to Threshold: Top Mass, Width and Momentum Distribution. *Phys.Rev.*, D60, 1999.

[12] F. Maltoni and T. Stelzer. MadEvent: Automatic event generation with MadGraph. *JHEP*, 02, 2003. arXiv:hep-ph/0208156.

[13] W.-Y. Pauchy Hwang T.-Y. Wu. *Relativistic quantum mechanics and quantum fields*, 1991.

[14] W. Heitler H.A.Bethe. On the stopping of fast particles and on the creation of positive electrons. *Proc. Phys. Soc. Lond.*, 146:83–112, 1934.

[15] S. Mrenna T. Sjostrand and P. Skands. *PYTHIA 6.4, Physics and Manual*, 2006.

[16] Geant4 official website. URL: http://geant4.cern.ch/.

[17] Michael E. Peskin Pisin Chen, Timothy L. Barklow. Hadron production in gamma-gamma collisions as a background for e+e- linear colliders. *Phys.Rev.*, D49:3209–3227, 1994.

[18] M.A. Thomson J.S. Marshall, A. Münnich. Performance of Particle Flow Calorimetry at CLIC. `arXiv:1209.4039`.

[19] Gregory Soyez Matteo Cacciari, Gavin P. Salam. Fastjet user manual. `arXiv:1111.6097`.

[20] Lcfiplus main webpage. URL: `https://confluence.slac.stanford.edu/display/ilc/LCFIPlus`.

[21] S. Yamashita T. Suehara, T. Tanabe. Improved jet clustering algorithm with vertex information for multi-bottom final states. *Technology and Instrumentation in Particle Physics*, 2011.

[22] M. Olsson S. Catani, Y. L. Dokshitzer, G. Turnock, and B. R. Webber. New clustering algorithm for multi - jet cross-sections in e+ e- annihilation. *Phys. Lett.*, B269:432–438, 1991.

[23] J. Zhu B. P. Roe, H.-J. Yang, Y. Liu, I. Stancu, and G. McGregor. Boosted Decision Trees as an Alternative to Artificial Neural Networks for Particle Identification. *Nucl.Instrum.Meth.*, A543:577–584, 2005.

[24] D.Kollar A.Caldwell and K. Kröninger. BAT - The Bayesian Analysis Toolkit. *Comp. Phys. Comm.*, 180:2197, 2009.

[25] T. Skwarnicki. A Study of the Radiative Cascade Transitions between the Upsilon-Prime and Upsilon Resonances. DESY-F31-86-02.

[26] Tftool repository. URL: `https://svnweb.cern.ch/trac/atlasinst/browser/Institutes/Goettingen/tftool`.

[27] D. Kirkby W. Verkerke. The roofit toolkit for data modeling. `arXiv:physics/0306116`.

Chapter 8

Appendix

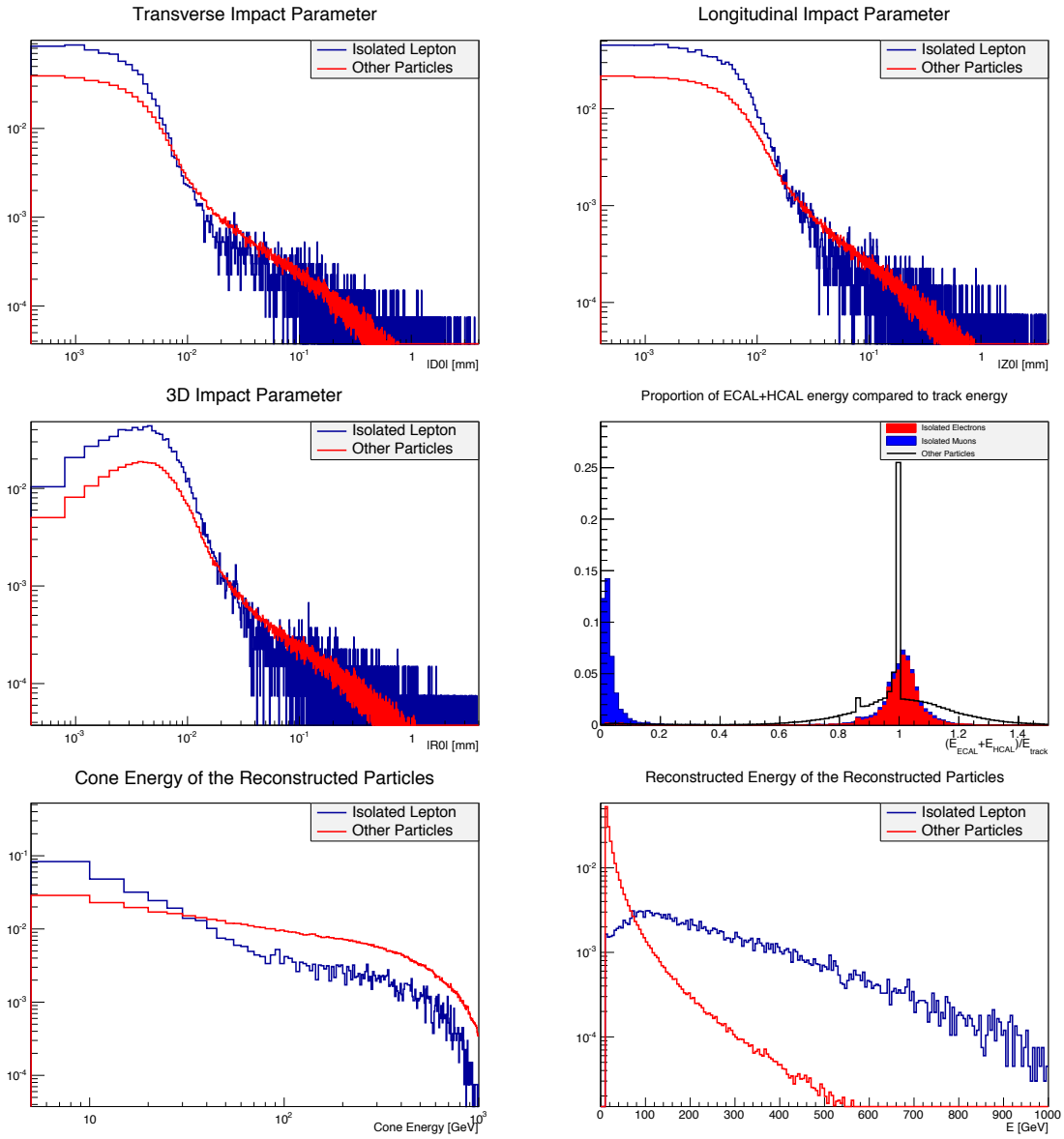
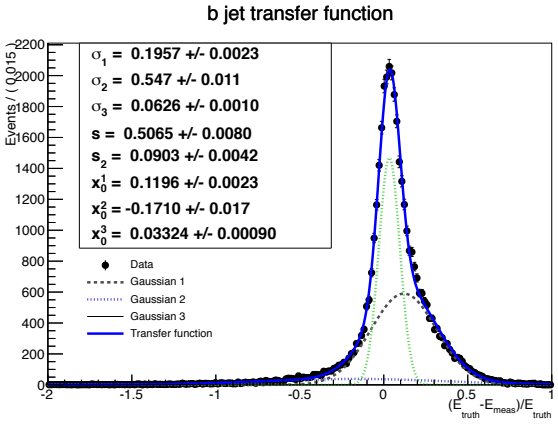
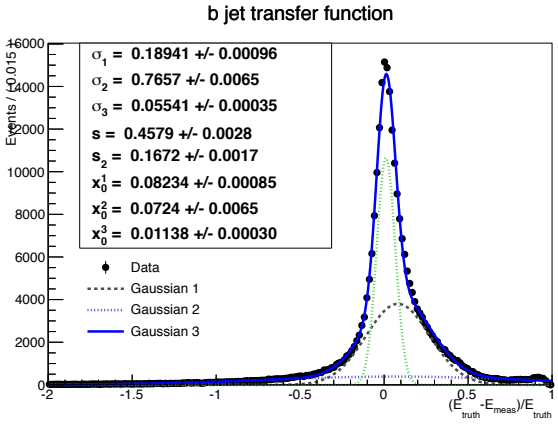


Figure 8.1: Impact Parameters and Energies of all the tracks compared to the isolated leptons for $\cos\theta = 0.999$

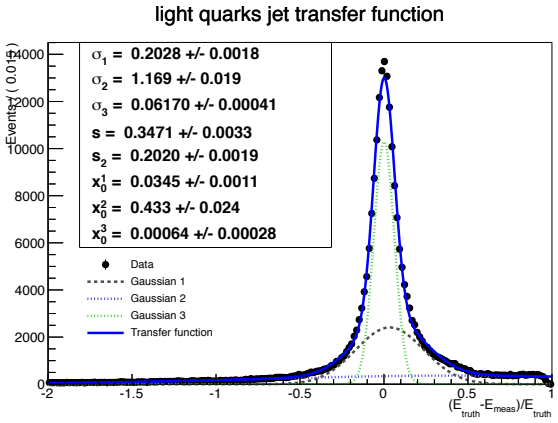


(a) $1.5 < |\eta| < 2.5$

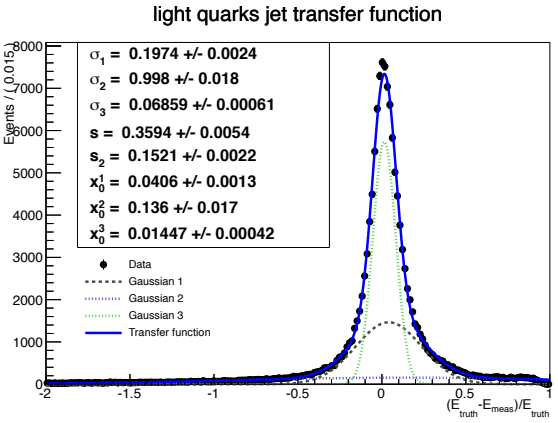


(b) $\forall|\eta| < 2.5$

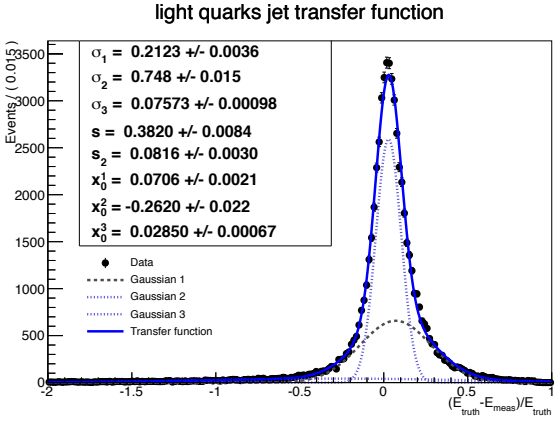
Figure 8.2: Transfer function for b quarks



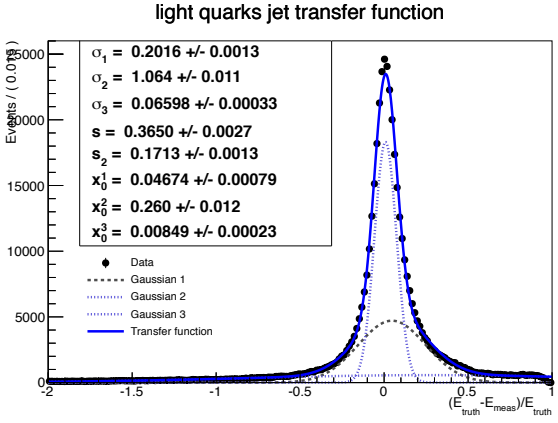
(a) $|\eta| < 0.8$



(b) $0.8 < |\eta| < 1.5$



(c) $1.5 < |\eta| < 2.5$



(d) $\forall|\eta| < 2.5$

Figure 8.3: Transfer function for light quarks (u, d, s, c)

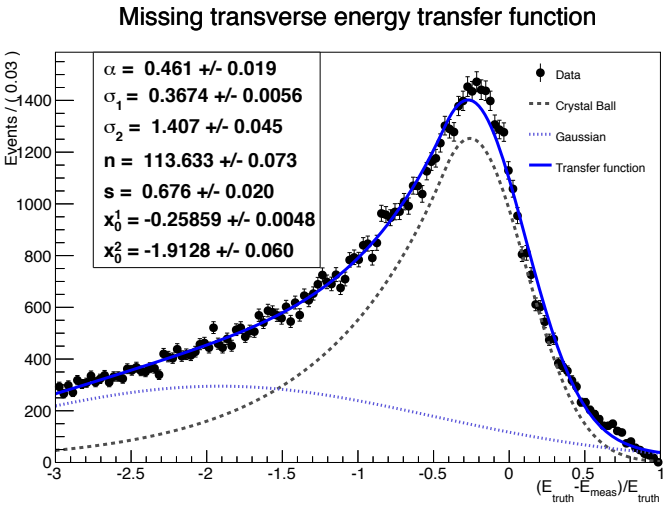
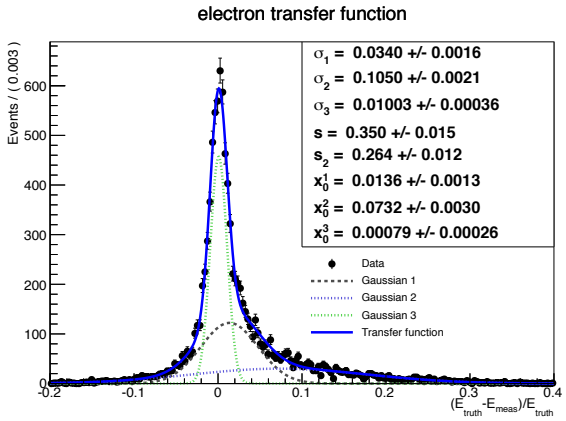
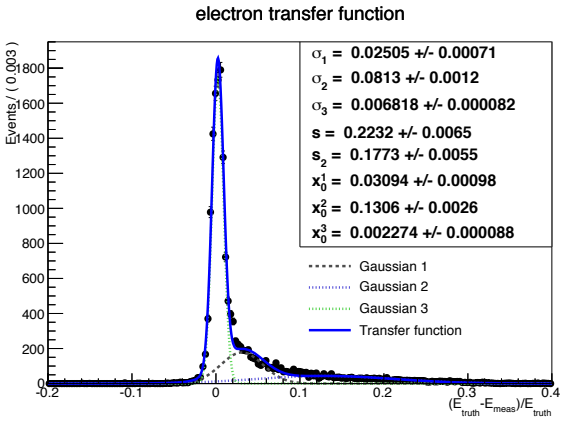


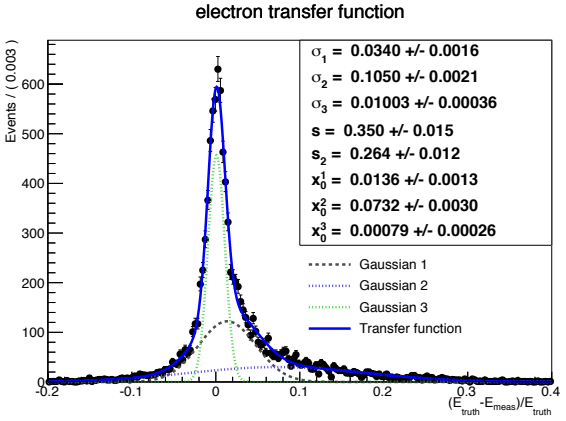
Figure 8.4: Transfer function for missing transverse energy



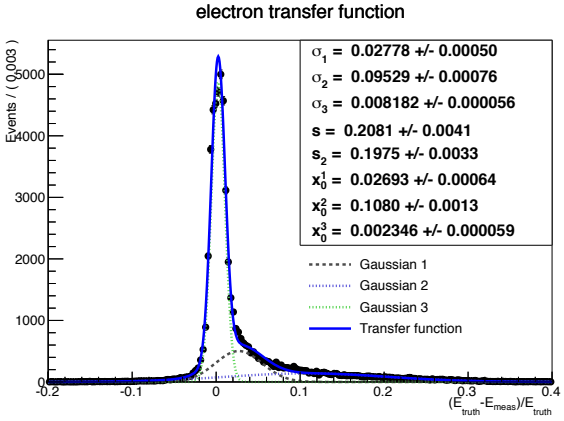
(a) $|\eta| < 0.8$



(b) $0.8 < |\eta| < 1.5$

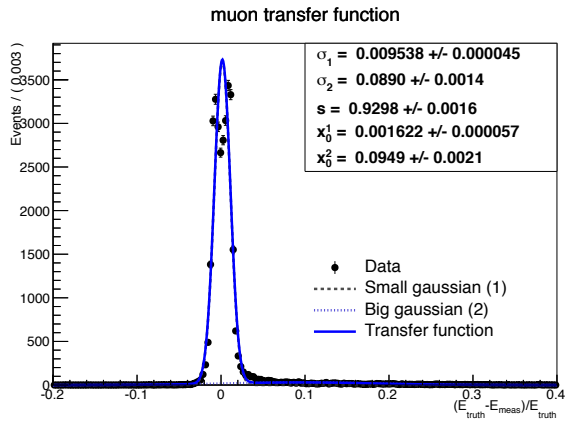


(c) $1.5 < |\eta| < 2.5$

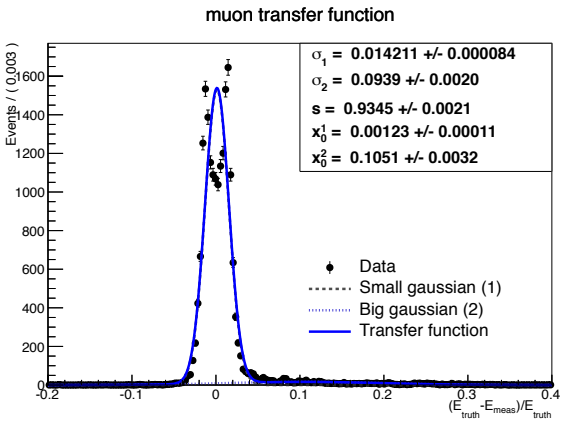


(d) $\forall|\eta| < 2.5$

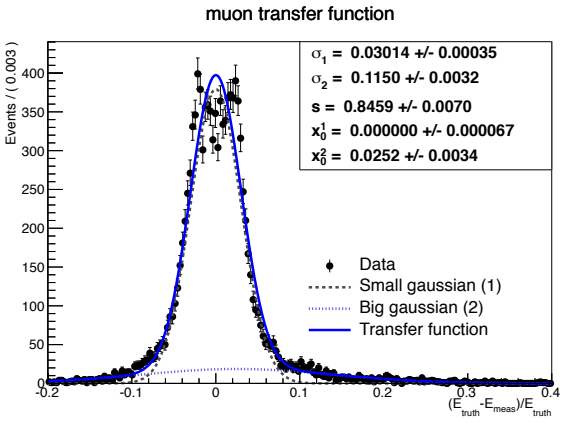
Figure 8.5: Transfer function for electrons



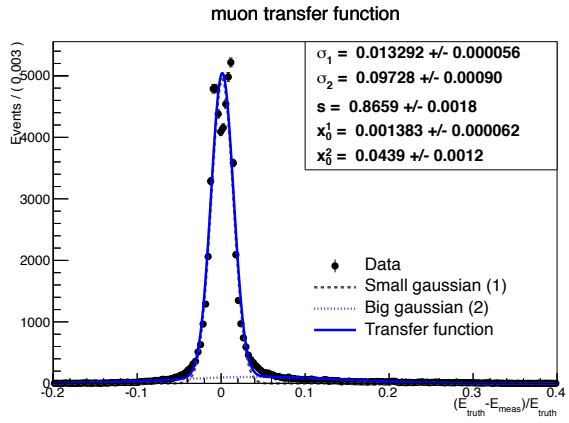
(a) $|\eta| < 0.8$



(b) $0.8 < |\eta| < 1.5$



(c) $1.5 < |\eta| < 2.5$



(d) $\forall|\eta| < 2.5$

Figure 8.6: Transfer function for muons

		Reconstruction efficiency [%]					
Lepton	Method	Overall	W_{had}	b_{had}	b_{lep}	b-tag|b-jets	b-tag|light jets
Electron	Likelihood	21.2 ± 1.1	29.7 ± 1.2	34.1 ± 1.3	45.0 ± 1.35	46.2 ± 0.9	18.0 ± 0.7
	Veto	32.4 ± 1.3	43.5 ± 1.3	40.7 ± 1.3	47.6 ± 1.3	63.6 ± 0.9	0 ± 0
	VetoLight	38.4 ± 2.0	50.2 ± 2.1	42.7 ± 2.1	49.2 ± 2.1	100 ± 0	0.9 ± 0.2
	VetoBoth	38.9 ± 2.1	50.5 ± 2.1	42.4 ± 2.1	49.0 ± 2.1	100 ± 0	0 ± 0
Muon	Likelihood	21.2 ± 1.0	28.2 ± 1.1	32.6 ± 1.2	45.2 ± 1.2	46.1 ± 0.9	16.3 ± 0.7
	Veto	31.9 ± 1.2	42.1 ± 1.2	39.1 ± 1.2	48.6 ± 1.2	$62. \pm 0.8$	0 ± 0
	VetoLight	36.5 ± 1.9	47.8 ± 2.0	41.3 ± 2.0	48.4 ± 2.0	100 ± 0	0.2 ± 0.2
	VetoBoth	36.8 ± 1.9	48.2 ± 2.0	41.5 ± 2.0	48.5 ± 2.0	100 ± 0	0 ± 0

Table 8.1: Detailed reconstruction efficiency of the semileptonic channel for events with a lepton and events with a muon