FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

SCHOOL OF PHYSICS AND ASTRONOMY

BACHELOR OF SCIENCE PROJECT:

LHC OLYMPICS

A project submitted by Miguel Pardo for the degree of
BSc Physics in the Southampton University

Supervised by: Steve F. King

Contents

1. Introduction

The Standard Model [1] is the more complete physics theory we have nowadays to explain how every particle and force interacts, except for gravity. It consists of 4 particles who carry the forces, called bosons, and 12 particles, that form all the matter we know for now: 6 particles called quarks and another 6 called leptons. Although it has turned out to be the most precise theory in the history of physics it still lacks some phenomena understanding, which suggests that there must be a more complete theory. A lot of physicists around the world are developing new theories that could fill the gaps that the SM has, like supersymmetry (SUSY) [2] or string theory [3]. These theories require more energy that the one is being produced in the LHC [4] to be verified, but in a few years it can be achieved, and scientists have to be prepared to know what to look at when they accomplish that range of energies.

In the LHC, beams formed by protons travel through a giant loop until they smash against each other, reaching energies of 14 TeV [5]. This reproduces the early universe conditions, around a fraction of second after the Big Bang [5], and can help to discover new physics.

The first challenge related with LHC data is the storage and selection of particular or interesting events among all the massive data coming from the events. But even after that, there is still one problem remaining, and it is that the selected events only have information about the decaying products, and not about the original signal.

This is the reason why some years ago some training series, called The LHC Olympics [6], were created at CERN. These series consisted of some challenges for theoretical physicists in which they had to interpret some data sets, known as black boxes, that were simulations of the real data obtained in the LHC collisions. Scientists were encouraged to apply different techniques to these data files in order to reconstruct the events that had taken place and discuss their results. These challenges were also an effective training for physicists to prepare them for new physics discoveries, teaching them what to look for in this scenario.

In this project we reproduce the physics developed in the LHC Olympics. We analyse the data files using Microsoft Excel [7] and apply some algorithms written in VBA [8] to simulate different analysis techniques. Using the samples provided [9] we try to determine the W and Z bosons mass in order to check that the invariant and transverse mass methods [10] work properly. After this, we use these methods to determine the top quark mass through its decaying products and discuss the results obtained.

2. Theoretical background

2.1 Standard Model

2.1.1 Components

This theory was developed in the early 1970's and it has proved through the years and experiments to be the best explanation for now of how the subatomic world works.

According to the Standard Model, there are two types of elementary particles (depending on one property called spin): fermions, which form all the matter we observe in the universe (their spin is half-integer), and bosons, which carry the forces between the fermions (their spin is integer). All of them have their own anti-particle, with same mass but opposite electric charge, isospin, etc.

- Fermions:
  They are subdivided at the same time into two types of particles: quarks and leptons. Leptons partake in the weak and electromagnetic interactions, but they do not interact strongly and are not found inside the nucleus. Quarks partake in all interactions and are the basic constituents of the particles inside the nucleus and other unstable ones. There are 6 particles of each type and they are usually organised into three different groups, called families, that are formed by 2 quarks and 2 leptons, as we can see in Figure 2.1:
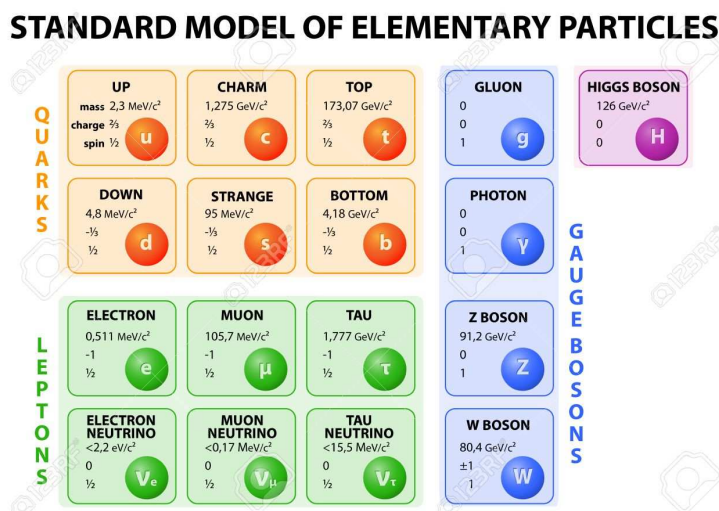
Figure 2.1: Components of the Standard Model and their properties [11].

The first family (formed by up and down quark, electron neutrino and electron) is the most stable one, and ordinary matter is compound of it. The following families are a "copy" of the first one in the following sense: particles that compose these families have the same properties (electric charge, spin, isospin, etc) as particles of the first family, but they are heavier, which makes them more unstable.

The heaviest particle of the SM is the top quark, with a mass of $173.1 \pm 0.6\,GeV$ [12], which makes it a very unstable and difficult to detect particle.

- Bosons:
  There are 4 fundamental interactions of nature, and each of them is associated with one gauge boson: the electromagnetic force is carried by photons, the weak nuclear force by W and Z bosons and the strong nuclear force by gluons. Although the model does not explain how gravity works, it has to provide a mechanism whereby particles gain their mass, and so there is a Higgs boson, which was discovered on 2012 at LHC [13].

2.1.2   Production process

As well as Top quarks, Z and W bosons are also very unstable, so we will not detect them in the collisions but their decaying products. However, according to SM, these particles have infinite different ways of decaying into more stable particles, each process with different probabilities. Moreover, at the LHC millions of collisions happen every second, most of them with different final state particles detected. Because of this, knowing which ones are the more probable decaying processes is crucial if we want to detect and reconstruct these kind of particles.

The SM provides accurate predictions about the probabilities of each decaying process, and these calculations have been far more simplified due to a poweful tool: Feynman diagrams [14]. Now we are going to take a look to the particles whose mass is going to be estimated in this project:

- Z boson [12]: this boson has enough mass to be unstable and it decays into other particles. Its main decay routes are: the leptonic decay, where the final state particles are two charged leptons, each of them with opposite charge ($l^+\,l^-$), or a pair neutrino-antineutrino ($\nu\bar{\nu}$); and the hadronic decay, with final state particles being a quark and an anti-quark of the same kind ($q\,\bar{q}$).
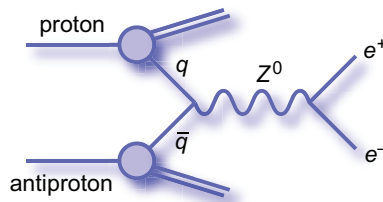

Figure 2.2: Example of production and decay of a Z^0 boson [15].

- W boson [12]: this charged boson (W^- or W^+) is also a relatively huge massive particle, so it undergoes further decay. The two main routes for decay of W^+ (W^-) are similar to the ones of the Z boson: the leptonic one, where it decays into a charged lepton l^+ (l^-) and same generation neutrino ν ($\bar{\nu}$); and the hadronic one, where it decays into a pair of light quark-antiquark of different kind ($q\,\bar{q}$).
  Although the pair $q\,\bar{q}$ is different for W than for Z bosons, these particles are not detected as only two, but as a spray of light hadrons (usually referred as a jet) due to QCD interactions, which makes the pair look like the same object (a jet) when is detected, no matters from which boson is coming from.
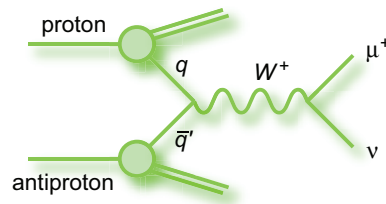
Figure 2.3: Example of production and decay of a W^+ boson [15].

- Top quark [12, 16]: This quark can be considered as a bare one, as its mean lifetime ($\sim 10^{-25}$) is relatively small compared with the hadronization time ($\sim 10^{-23}$), and so it decays before any QCD effects can affect it.
  The two main process for top quark pairs production are through the strong interaction: $90\,\%$ of the times comes from gluon fusion:
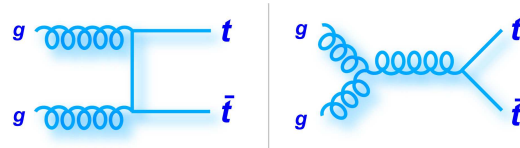
Figure 2.4: Examples of production of a pair $t\bar{t}$ through gluon fusion [15].

The remaining $10\,\%$ is due to quark-antiquark annihilation:
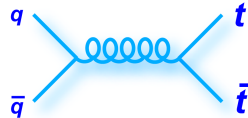
Figure 2.5: Example of production of a pair $t\bar{t}$ through $q\bar{q}$ annihilation [15].

The top quark t ($\bar{t}$) decays almost 100 % of the times into a W^+ (W^-) boson and a bottom quark b ($\bar{b}$). The bottom quark is also unstable and it fragments into a jet due to QCD interactions.

Figure 2.6: Examples of production and decay of a pair $t\bar{t}$. On the left: di-lepton decay route. On the right: hadronic decay route [15].

We have already seen the possible main decay products of W boson, so the three different decay routes for a top quark pair produced are:

– Di-lepton decay channel:
$$t\bar{t} \rightarrow b + \bar{b} + l^+ + l^- + \nu + \bar{\nu} \tag{2.1}$$
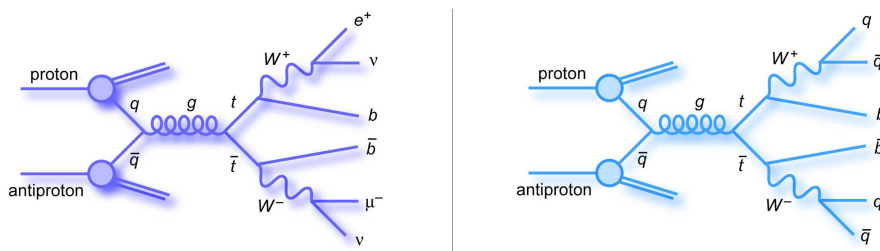
– Lepton+jets decay channel:
$$t\bar{t} \rightarrow b + \bar{b} + q + \bar{q} + l^{\pm} + \nu \tag{2.2}$$

– Hadronic decay channel:
$$t\bar{t} \rightarrow b + \bar{b} + q + q + \bar{q} + \bar{q} \tag{2.3}$$

2.2   Colliders and detectors

2.2.1   Collision geometry

In the LHC, the proton beams that collide travel in opposite directions along the beam-pipe, which corresponds to z-axis in Figure 2.7 [10].
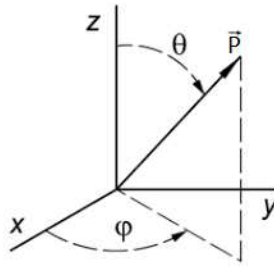
Figure 2.7: Collision geometry for an outgoing particle with momentum $\vec{P}$.

If the protons have any motion perpendicular to the beam-pipe, then they have a component in the transverse plane, and we can split any proton momentum $\vec{P}$ into two components: the one in the transverse plane $\vec{P^T}$ and the one along the beam-pipe $\vec{P^z}$:

$$\vec{P} = \vec{P^z} + \vec{P^T} \tag{2.4}$$

They can be denoted as transverse and longitudinal momentum and can be defined using the angle subtended by the beam-pipe and the total momentum vector:

$$\left|\vec{P^z}\right| = \left|\vec{P}\right| \cdot \sin\theta \tag{2.5}$$

$$\left|\vec{P^T}\right| = \left|\vec{P}\right| \cdot \cos\theta \tag{2.6}$$

However, as it simplifies calculations involving θ, it is more common to use a variable called pseudo-rapidity, that is defined as [10]:

$$\eta = -\ln\left|\tan\frac{\theta}{2}\right| \tag{2.7}$$

Figure 2.8: Projections of the transverse momentum along the x and y axis.

Taking a look to Figure 2.8 we see that is also possible to split the transverse momentum into its projections along the x and y axis using the azimuthal angle ϕ:

$$\left|\vec{P^x}\right| = \left|\vec{P^T}\right| \cdot \cos\phi \tag{2.8}$$

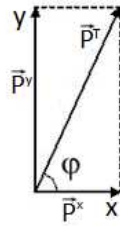$$\left|\vec{P^y}\right| = \left|\vec{P^T}\right| \cdot \sin\phi \tag{2.9}$$

2.2.2 Detectors

Detectors in particle accelerators are usually divided up to different components so that a particle will pass through them sequentially.

Fast-moving particles of a life-time $\tau > 10^{-10}\,s$ will still interact with the detector and they can be detected [10]. However, as we have seen for the W and Z bosons or the top quark in Section 2.1.2, most of the particles in the SM are not directly "seen" in the detectors. Instead, only their decay products, stable particles such as $p, \bar{p}, e^{\pm}, \gamma$, will show up as energy deposit or charge tracks in the different layers of the detector.

The layers are surrounding the beam-pipe, as we can see in Figure 2.9, and we can identify the particles depending on how they interact with each layer.
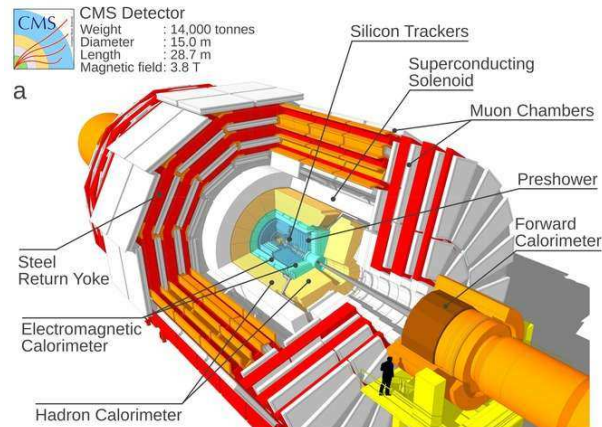
Figure 2.9: CMS detector scheme [17].

The innermost layer is usually embedded with a magnetic field causing the charge particles, either positive or negative, to curve in different directions. Depending on the degree of curvature we can determine the momentum of the tracking particle [5].

The next layers are usually calorimeters, like the electromagnetic and hadronic ones. Calorimeters are composed of different materials depending on which particle they are stopping, and are able to detect even neutral particles such as neutrons or photons (not visible in the tracking chamber).

The muon chamber is the outermost layer due to muons require passing through metres of dense material

before being stopped and detected.

Figure 2.10: ATLAS detector scheme [18].

Neutrinos and other weakly interacting particles are invisible to this type of detectors, so the only way to detect them is using our knowledge about particle collisions. In the collisions, the particles that interact are not really protons, but their constituent partons: quarks and gluons. Unfortunately we can only determine the momentum of the proton, and not how it is distributed among its partons. However, proton is motionless in the transverse plane, which implies that any transverse momentum of a parton has to be negligible compared to its longitudinal momen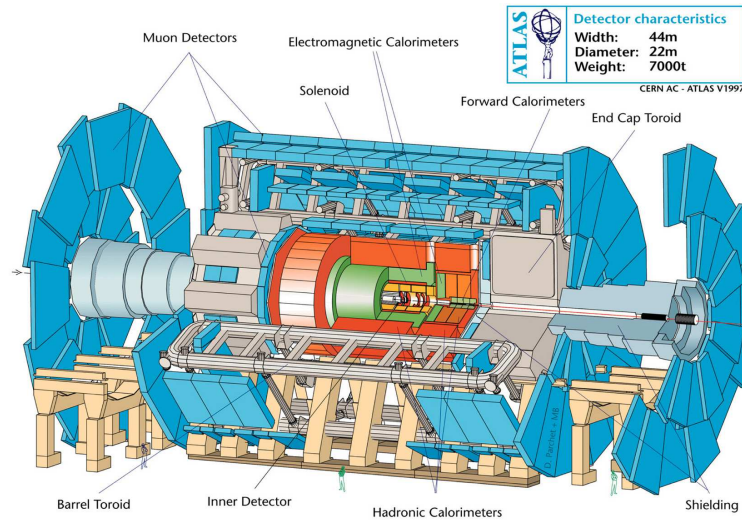tum and sets up collisions with no net transverse momentum. Hence, by adding the transverse momentum of all the final state particles detected and comparing it to zero, we can determine if there is some undetected particle with a missing transverse momentum often referred as Missing Energy [10].

2.2.3 Triggering

All missing energies or particles detected in a collision are described by a list of objects which is referred as an event. If we consider $p\bar{p}$ collisions, there is a continuous output of data coming from every event in real-time while the detector is running: sensors deliver data 30 million times per second at LHC experiments [5], which makes too many data to process and store (in e^+e^- collisions there are no such problems due to the lower reaction rate). Moreover, only several hundred collisions per second provide useful or interesting data. Thus, there is a process called triggering whereby the 30 million collisions per second are compared with a list of requirements. The events that do not match the requirements are thrown away [10], cutting the amount of data down to a manageable quantity that can be stored. Modern detectors for hadron colliders such as CMS or ATLAS at LHC typically adopt three levels of triggering [10, 19].

2.3 Data simulations

2.3.1 Black boxes

Data files used in the LHC Olympics are created using simulation programmes known as event generators, that simulate the creation, decay and detection of particles at colliders. Two examples of these programmes are PYTHIA [20] and Herwig [21]. The stages for the creation of these black boxes are:

1. Feynman diagrams are calculated for a particular short distance physics production process.

2. The evolution of this process is simulated, taking into account all decay routes.

3. The resulting final state particles are then processed by a software called PGS [22] that behaves as a detector and triggering takes place. The final output becomes the black box [19].

A black box file can contain different physics processes, although there are limitations on how many of them can be processed accurately. LHC Olympics black boxes do not contain any SM background [23], which makes Olympics unrealistic in this aspect.

2.3.2   Understanding black box files

We have used Microsoft Excel [7] as the analysis tool for our data files [9]. The overall appearance of a black box file when using this programme is shown in Figure 2.11:

	C	D	E	F	G	H	I	J	K
1		Number	Type	Pseudorapidity	Azimuthal angle	Transverse momentum	Invariant mass	Number of tracks	Bottom tagging
2		1	2	-0,931	5,391	28,00	1,00	0	0
3		2	2	0,034	2,847	67,02	-1,00	0	0
4		3	4	-1,431	0,278	25,44	5,96	3	0
5		4	4	-0,083	6,189	17,95	7,15	5	0
6		5	6	0,000	4,926	2,06	0,00	0	0
7		1	2	-0,105	4,669	135,64	1,00	2	0
8		2	2	-1,554	3,934	21,25	-1,00	0	0
9		3	3	0,298	3,674	13,59	-1,00	1	0
10		4	4	1,852	3,740	12,50	3,41	1	0
11		5	4	-2,918	0,710	18,82	7,72	0	0
12		6	6	0,000	1,455	144,98	0,00	0	0
13		1	2	0,302	1,967	41,92	-1,00	0	0
14		2	2	-0,128	1,796	16,91	-1,00	0	0
15		3	4	-1,918	4,341	22,14	3,79	1	0
16		4	4	3,486	1,664	15,88	3,11	0	0
17		5	6	0,000	5,322	62,96	0,00	0	0
18		1	2	1,801	2,921	24,74	1,00	0	0
19		2	2	0,235	0,241	80,72	-1,00	0	0
20		3	4	1,659	4,283	44,29	6,40	6	0
21		4	4	1,006	2,204	60,62	25,50	16	0
22		5	6	0,000	4,868	41,68	0,00	0	0
23		1	2	0,461	4,648	40,95	-1,00	1	0

Figure 2.11: Diboson sample [9] shown in MS Excel [7].

An event consists of a set of rows, like the one we have highlighted as an example. Each row of the data sheet describes a particle or decay product detected and some of its properties, like the collision angles we have talked about in Section 2.2.1 and others we will discuss now in Table 2.1:

Column name	Description
Number	Counts how many objects are in an event, making the start of a new one when its value is 1 again
Type	Labels the kind of object using the integers 0,1,2,3,4,6 refering to a photon, electron, muon, tau, jet and missing energy, respectively.
Pseudorapidity	Gives the pseudorapidity η of the object.
Azimuthal angle	Gives the azimuthal angle ϕ of the object (in radians).
Transverse momentum	Gives the value of the transverse momentum modulus $\left|\vec{P^T}\right|$ of the object.
Invariant mass	Gives the invariant mass in case the object is a jet or the charge of the particle otherwise.
Number of tracks	Gives the number of different particle tracks within a jet.
Bottom tagging	Its value is 1 if the object is a jet that comes from a charm or bottom quark, otherwise is 0.

Table 2.1: Description of the values of each column in the data box file [19].

Now we can interpret the event highlighted in Figure 2.11: one antimuon μ^+, one muon μ^-, 2 jets, none of them heavy, and missing energy.

Transverse momentum and invariant mass values are given in units such that the speed of light c is re-scaled to unity, which implies that these quantities are in units of GeV (Giga electron Volt) [24]. The missing energy does not have a defined pseudorapidity, as it cannot be calculated.

The resolution of the detector is determined by the one in the data file: 3 decimal places for azimuthal angle and pseudorapidity and 2 decimal places for invariant mass and transverse momentum.

3. Analysis methods

3.1 VB algorithms

After importing each data file [9] to MS Excel [7], we had to look for a software programme that enabled us to carry out different tasks related with these data files, such as select events we are interested in or calculate some variables like the invariant or transverse mass.

For this purpose we decided to use Visual Basic for Applications [8], a programming language that extends the functionality of Office applications . VBA allows us to create algorithms that can be run over the Excel data sheets. At the beginning we created some basic macros and algorithms to count how many objects of a kind are within the data file and to delete or copy data from one data sheet to another. Then we built some more complex ones to select or filter specific events and reconstruct properties of the decaying particles.

We will refer to some of the important macros and functions in this section and the followings, but we are not going to discuss them here. Instead a list of all the algorithms used is provided in Appendix A, where we describe them shortly.

3.2 Invariant mass method

The invariant mass is a variable that allows us to reconstruct short-living particles. If we have 2 particles and we know their momentum, then the invariant mass is defined as [10]:

$$M_{Inv}^2 = (\vec{p_1} + \vec{p_2})^2 \tag{3.1}$$

where we have used the four momentum vector, a common variable in high energy physics, constructed with the energy E and the usual three momentum vector $\vec{P}$, that reproduces the relativistic mass:

$$p^\mu p_\mu = \vec{p}^2 = E^2 - \left|\vec{P}\right|^2 = m^2 \tag{3.2}$$

The invariant mass can also be expressed as:

$$M_{Inv}^2 = (\vec{p_1} + \vec{p_2})^2 = \vec{p_1}^2 + \vec{p_2}^2 + 2\,\vec{p_1} \cdot \vec{p_2} = m_1^2 + m_2^2 + 2\left(E_1 E_2 - \vec{P_1} \cdot \vec{P_2}\right) \tag{3.3}$$

For N particles, Eq. (3.1) changes to:

$$M_{Inv}^2 = \left(\sum_{i=1}^N \vec{p_i}\right)^2 \tag{3.4}$$

which can be expressed as:

$$M_{Inv}^2 = (\vec{p_1} + \vec{p_2} + ... + \vec{p_N})^2 = \left(\sum_{i=1}^N E_i\right)^2 - \left(\sum_{i=1}^N \vec{P_i}\right)^2 \tag{3.5}$$

and remembering Eq. (2.4):

$$M_{Inv}^2 = \left(\sum_{i=1}^N E_i\right)^2 - \left(\sum_{i=1}^N \vec{P_i^z}\right)^2 - \left(\sum_{i=1}^N \vec{P_i^T}\right)^2 \tag{3.6}$$

If we use the massless limit [10], in which light quarks and leptons are considered to have negligible mass (this implies $E_i = \left|\vec{P_i}\right|$ from Eq. (3.2)), and consider that in proton-proton collisions there is no net transverse momentum:

$$M_{Inv}^2 = \left(\sum_{i=1}^{N} \left|\vec{P_i}\right|\right)^2 - \left(\sum_{i=1}^{N} \left|\vec{P_i^z}\right|\right)^2 \tag{3.7}$$

Finally, remembering Eq. (2.5) and (2.6):

$$M_{Inv}^2 = \left(\sum_{i=1}^{N} \frac{\left|\vec{P_i^T}\right|}{\sin\theta_i}\right)^2 - \left(\sum_{i=1}^{N} \frac{\left|\vec{P_i^T}\right|}{\tan\theta_i}\right)^2 \tag{3.8}$$

When we plot this invariant mass for one type of event we obtain a frequency distribution that can be described by the Breit-Wigner formula [10]. The peak of the distribution corresponds with the rest mass of the decaying particle and the width of the peak at half maximum Γ is determined by the lifetime of the particle, although the structure of the distribution can change due to detector inefficiencies.
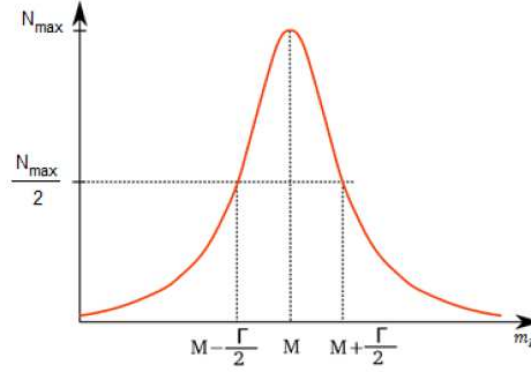

Figure 3.1: The Breit-Wigner resonance.

3.3   Transverse mass method

If we consider events that involve neutrinos, such as a W boson decaying into a lepton and a neutrino:

$$W^- \longrightarrow l^- + \nu \tag{3.9}$$

and now we calculate its invariant mass using Eq. (3.8):

$$M_{Inv}^2 = \left(\frac{\left|\vec{P_{l^-}^T}\right|}{\sin\theta_{l^-}} + \frac{\left|\vec{P_\nu^T}\right|}{\sin\theta_\nu}\right)^2 - \left(\frac{\left|\vec{P_{l^-}^T}\right|}{\tan\theta_{l^-}} + \frac{\left|\vec{P_\nu^T}\right|}{\tan\theta_\nu}\right)^2 \tag{3.10}$$

We can notice that there is a problem: the pseudorapidity of the neutrino η_ν is not determined, and so is the angle θ_ν.
To solve this we define the missing transverse momentum [10] by applying momentum conservation in the transverse plane:

$$\vec{P_{miss}^T} = -\sum_{visible} \vec{P^T} \tag{3.11}$$

The modulus of this momentum is the value given in the data sheet to an object labelled as missing energy. We define the transverse mass for a system of N particles as (ignoring the longitudinal momentum of the

system):

$$M_T^2 = \left(\sum_{i=1}^{N} E_i^T\right)^2 - \left(\sum_{i=1}^{N} \vec{P_i^T}\right)^2 \tag{3.12}$$

where the transverse energy E_i^T is given by:

$$E_i^T = \sqrt{\left|\vec{P_i^T}\right|^2 + m_i^2} \tag{3.13}$$

and it can be simplified using the massless limit [10] to:

$$E_i^T = \left|\vec{P_i^T}\right| \tag{3.14}$$

If we compare Eq. (3.6) with Eq. (3.12), we can see they are the same if we neglect the longitudinal component of the momentum, so this implies the following inequality:

$$0 \leq M_T \leq M_{Inv} \tag{3.15}$$

If we plot this transverse mass the same way as we talked about in Section 3.2 for invariant mass we should observe a frequency distribution with a sharply cut off at the rest mass of the decaying particle due to the resonance in the invariant mass distribution. However in practice we observe a cut off with a finite width due to the Γ width of the invariant mass.
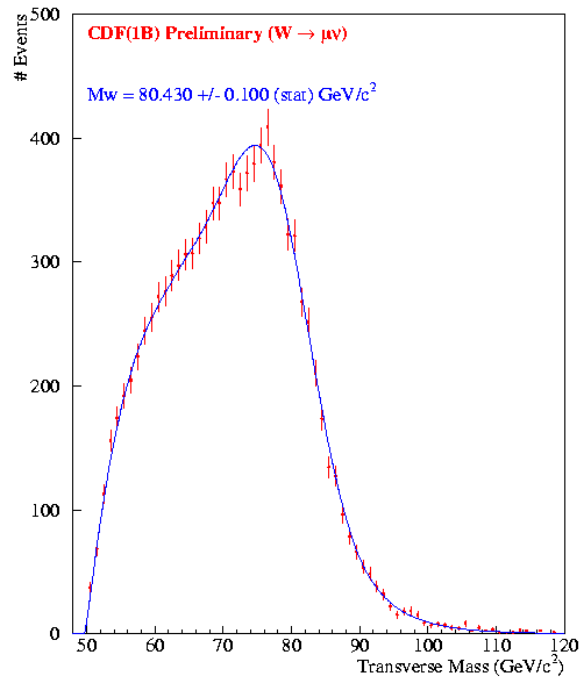

Figure 3.2: Transverse Mass distribution for $W \rightarrow \mu\nu$ observed in CDF.

4. Results

4.1 Diboson sample

For our analysis of Z boson and W boson masses, we used the data files $diboson\_1005.dat.0.gz$ and $diboson\_1005.dat.1.gz$ [9]. In these files, most of the registered events corresponded to collisions of the type: $W^+ W^-$, $Z^0 Z^0$ or $W^\pm Z^0$. The main possible final state products of these collisions are listed below in Table 4.1:

Type of collision	Decay route	Final state particles
$W^+ W^-$	Both hadronic	$(q + \bar{q} + q + \bar{q})$
	One hadronic and one leptonic	$(q + \bar{q} + l^\pm + \nu)$
	Both leptonic	$(l^+ + l^- + \nu + \nu)$
$Z^0 Z^0$	Both hadronic	$(q + \bar{q} + q + \bar{q})$
	One hadronic and one leptonic	$(q + \bar{q} + l^+ + l^-)$ or $(q + \bar{q} + \nu + \bar{\nu})$
	Both leptonic	$(l^+ + l^- + l^+ + l^-)$ or $(l^+ + l^- + \nu + \bar{\nu})$ or $(\nu + \bar{\nu} + \nu + \bar{\nu})$
$W^\pm Z^0$	Both hadronic	$(q + \bar{q} + q + \bar{q})$
	Hadronic $W^\pm$ and leptonic Z^0	$(q + \bar{q} + l^+ + l^-)$ or $(q + \bar{q} + \nu + \bar{\nu})$
	Leptonic $W^\pm$ and hadronic Z^0	$(l^\pm + \nu + q + \bar{q})$
	Both leptonic	$(l^\pm + \nu + l^+ + l^-)$ or $(l^\pm + \nu + \nu + \bar{\nu})$

Table 4.1: List of main possible decay routes for W and Z boson collisions.

4.1.1 Z boson

Taking a look to Table 4.1, we decided that the events we were going to look for were the ones with final state particles:

$$q + \bar{q} + l^+ + l^- \tag{4.1}$$

since these products could only come from a collision of the type $Z^0 Z^0$ or $W^\pm Z^0$, where the charged leptons came unmistakably from a Z boson. This implies that we could apply the invariant mass method to both of these leptons to reconstruct the Z boson mass. We could not choose a final state with different particles that the ones in Eq (4.1) since the other states produced ambiguity when it came to determine which particle came from which boson.

Then we selected the events we were interested in from our data files making use of the algorithm $JetStart$. To improve the statistics we did two different signatures: events where these two leptons were a pair of electron-positron ($e^- e^+$) and events where they were a pair muon-antimuon ($\mu^- \mu^+$).

The signature we used at first was the same as in Eq (4.1) i.e, we looked for events with two light jets and two leptons with opposite charge. However, there was not a single event in those data files with such detected objects, but all the events with these signatures had also some missing energy. This fact can be explained as detector inefficiencies, since the value of the transverse momentum for this missing energy was relatively small, as we can see in Figure 4.1:

D	E	F	G	H	I	J	K
Number	Type	Pseudorapidity	Azimuthal angle	Transverse momentum	Invariant mass	Number of tracks	Bottom tagging
1	2	0,927	0,608	224,16	-1,00	0	0
2	2	1,748	5,214	13,55	1,00	0	0
3	4	1,394	3,673	215,65	85,70	30	0
4	4	2,668	3,234	26,07	4,35	0	0
5	6	0,000	5,763	24,35	0,00	0	0
1	2	0,704	2,226	45,76	1,00	0	0
2	2	0,233	4,742	46,89	-1,00	0	0
3	4	3,864	0,590	31,49	5,41	0	0
4	4	2,617	5,660	18,94	3,17	0	0
5	6	0,000	4,868	5,43	0,00	0	0
1	2	-0,763	3,349	71,30	-1,00	0	0
2	2	-0,140	0,403	24,32	1,00	0	0
3	4	-0,347	5,545	50,99	8,30	5	0
4	4	-1,301	1,673	28,20	4,55	3	0
5	6	0,000	1,519	8,18	0,00	0	0
1	2	0,788	3,622	38,11	-1,00	0	0
2	2	1,490	0,224	9,43	1,00	1	0
3	4	1,003	5,062	10,78	3,06	2	0
4	4	2,167	1,647	14,73	5,25	1	0
5	6	0,000	0,448	22,33	0,00	0	0
1	2	-1,096	2,136	36,08	-1,00	0	0
2	2	-0,488	2,755	37,31	1,00	0	0

Figure 4.1: Data sheet with events of the type of Eq (4.1). The value of the transverse momentum is highlighted in orange.

We selected these events and then we calculated the invariant mass of the pair $e^- e^+$ or $\mu^- \mu^+$ using the algorithm $InvariantMass$ and $InvariantMassWxy$. The results obtained by the algorithm $InvariantMass$ were closer to the "real" value than the ones obtained by $InvariantMassWxy$, so we used the first one and plotted the distribution.

To plot the distribution we used the software programme SciDAVis [25], since it has really powerful tools for representing and analysing data. We obtained the following plots for a bin size of 1 GeV:
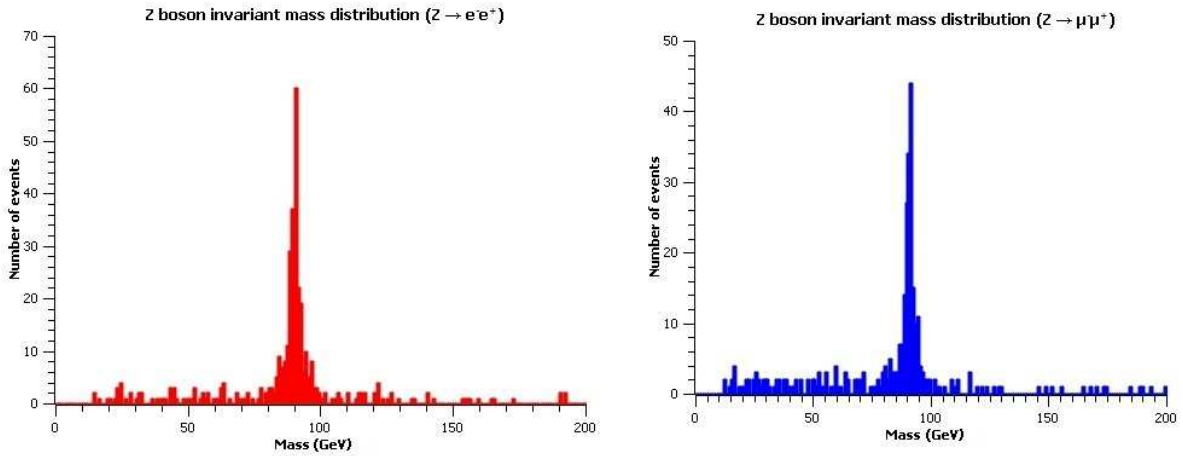

Figure 4.2: Z boson invariant mass frequency distributions from 0 to 200 GeV. On the left: distribution for $e^- e^+$. On the right: distribution for $\mu^- \mu^+$.
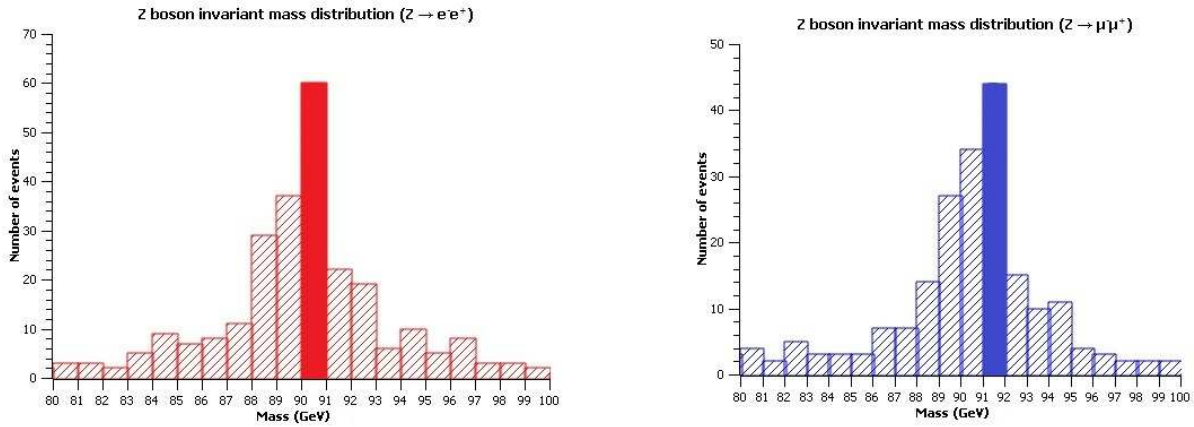
If we take a closer look to Figure 4.2:

Figure 4.3: Z boson invariant mass frequency distributions from 80 to 100 GeV. On the left: distribution for e^-e^+. On the right: distribution for $\mu^-\mu^+$. The peak of each distribution is highlighted.

From these graphics is easy to determine where is the peak of the distribution, which corresponds with the Z boson mass M_Z. We estimated the following values for it:

$$90 \leq M_Z^e \leq 91\,GeV \Rightarrow M_Z^e = 90.5 \pm 0.5\,GeV \longrightarrow \text{ for the } e^-e^+ \text{ decay.} \tag{4.2}$$

$$91 \leq M_Z^\mu \leq 92\,GeV \Rightarrow M_Z^\mu = 91.5 \pm 0.5\,GeV \longrightarrow \text{ for the } \mu^-\mu^+ \text{ decay.} \tag{4.3}$$

In order to check the reliability of the invariant mass method, we also tried to determine the mean lifetime of the Z boson using the Breit-Wigner formula [10]. This formula allows us to estimate the mean lifetime τ of the particle by measuring the width of the peak of its invariant mass distribution at half-maximum Γ:

$$\tau = \frac{\hbar}{\Gamma} \tag{4.4}$$

We estimated the following mean lifetimes, using the value of $\hbar = 6.58212 \cdot 10^{-16}\,eV \cdot s$ [26]:

$$\Gamma^e = 2 \cdot 10^9\,eV \Rightarrow \tau^e = 3.291 \cdot 10^{-25}\,s \longrightarrow \text{ for the } e^-e^+ \text{ decay.} \tag{4.5}$$

$$\Gamma^\mu = 3 \cdot 10^9\,eV \Rightarrow \tau^\mu = 2.194 \cdot 10^{-25}\,s \longrightarrow \text{ for the } \mu^-\mu^+ \text{ decay.} \tag{4.6}$$

4.1.2   W boson

Following the same analysis as in Section 4.1.1 for the Z boson, the events we used for the calculations of the W boson mass were the ones with the following final state products:

$$q + \bar{q} + l^\pm + \nu \tag{4.7}$$

In this case, as there was a neutrino among the other products, we applied the transverse mass method to the charged lepton and the neutrino in order to reconstruct the W boson mass. We chose as the charged lepton an electron or positron, because almost all the events in the diboson samples involving Eq (4.7) were of this kind and this way we had better statistics.

Applying the algorithm *JetStart* with a different signature to select the events we were interested in and the algorithm *TransverseMass* to calculate the transverse mass of the electron or positron and the neutrino, we obtained the following distribution:
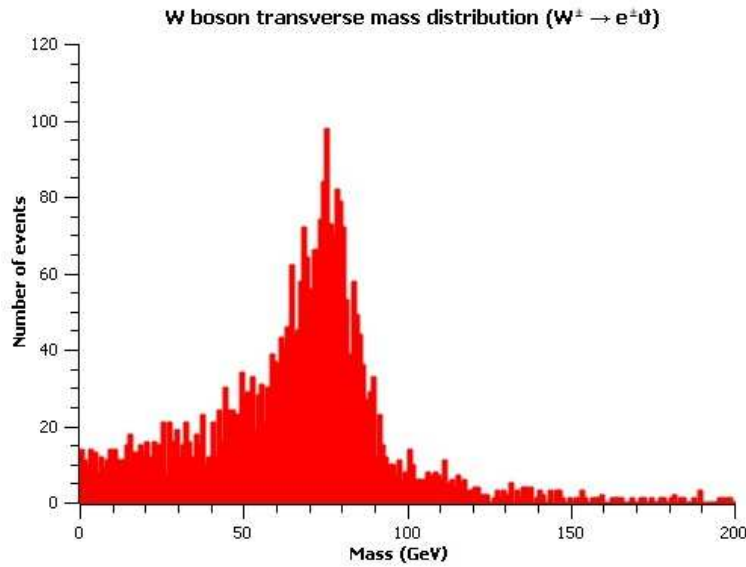
Figure 4.4: W boson transverse mass frequency distribution from 0 to 200 GeV.

In this case the bin size was also 1 GeV. Here it was harder to determine which one was the value for the W boson mass M_W, since instead of a sharply cut off it had some width. Taking a closer look to the cut off:
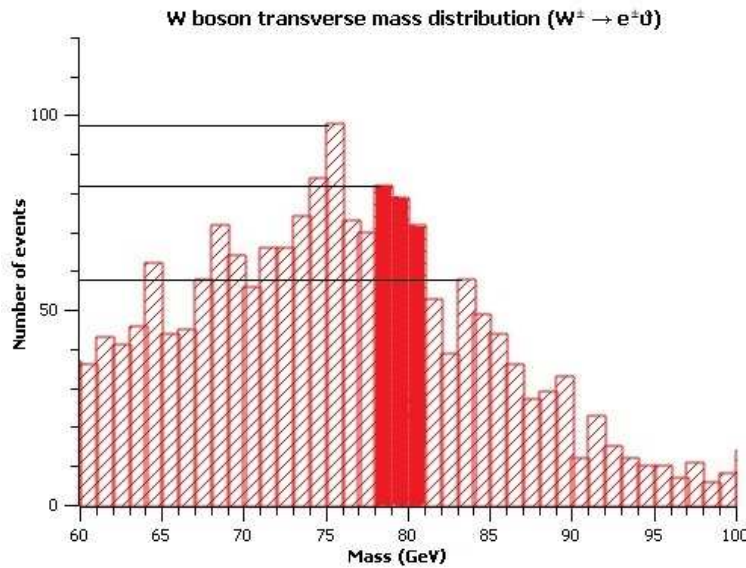

Figure 4.5: W boson transverse mass frequency distribution from 60 to 100 GeV. The bins highlighted correspond to the estimated W boson mass.

Looking to Figure 4.5, the first highlighted bin is the first peak after the maximum of the distribution that has far more number of events than the following peak. This bin could have been a good choice for the W mass, but the other highlighted bin values have quite similar values for the number of events, so we decided to select the three bins as the W boson mass, i.e:

$$78 \leq M_W \leq 81\,GeV \Rightarrow M_W = 79.5 \pm 1.5\,GeV \tag{4.8}$$

Miguel Pardo

4.2   Ttbar sample

In this section we tried to determine the top quark mass. For this purpose we used the following data files [9]:
$ttbar\_1005.dat.0.gz$, $ttbar\_1005.dat.1.gz$, $ttbar\_1005.dat.2.gz$, $ttbar\_1005.dat.3.gz$, $ttbar\_1005.dat.4.gz$,
$ttbar\_1005.dat.5.gz$, $ttbar\_1005.dat.6.gz$, $ttbar\_1005.dat.7.gz$, $ttbar\_1005.dat.8.gz$, $ttbar\_1005.dat.9.gz$,
$ttbar\_1005.dat.10.gz$, $ttbar\_1005.dat.11.gz$, $ttbar\_1005.dat.12.gz$, $ttbar\_1005.dat.13.gz$.
The general decay route of a pair $t\bar{t}$ is:

$$t\bar{t} \rightarrow W^+ + b + W^- + \bar{b} \tag{4.9}$$

Remembering the possible specific decay routes we have seen in Eq. (2.1), (2.2) and (2.3), the events we were going to look for were those ones that have followed a lepton+jets decay route. The reason is that this decay route was the only one that allowed us to unmistakably determine which objects correspond to which original particles, as we can see in Table 4.2:

Decay route name	Final state particles	Original particles determination problem
Di-lepton decay	$t\bar{t} \rightarrow b+\bar{b}+l^++l^-+\nu+\bar{\nu}$	Two neutrinos: ν and $\bar{\nu} \Rightarrow$ we only detect one object called missing energy, so we miss the information about both neutrinos.
Lepton+jets decay	$t\bar{t} \rightarrow b+b+q+\bar{q}+l^\pm+\nu$	No problem: $l^+ (l^-) + \nu$ comes from $W^+ (W^-)$ and $q + \bar{q}$ comes from $W^- (W^+)$.
Hadronic decay	$t\bar{t} \rightarrow b+b+q+q+\bar{q}+\bar{q}$	Two quarks and two antiquarks: detected as 4 jets, we cannot determine which pair $q\bar{q}$ comes from which W boson.

Table 4.2: Possible decay routes for top quark pair collisions.

To reconstruct the top quark mass, we selected events with the lepton+jets decay route signature using the algorithm *JetStart*. In each event, the detected objects were: a lepton ($e^\pm$, $\mu^\pm$, $\tau^\pm$), two light jets ($q\bar{q}$), two heavy jets ($b\bar{b}$) and missing energy (ν). Then we should have applied the invariant mass method to the pair $q\bar{q}$ and the bottom quark b, and the transverse mass method to the lepton $l^\pm$, the neutrino ν and the bottom antiquark $\bar{b}$. However, there was one problem: the pair $b\bar{b}$ is detected as just two heavy jets, so we don't know which one comes from the top quark t and which one from the top antiquark $\bar{t}$ i.e, we don't know which one use for the calculation of the invariant mass, and which one for the transverse mass.
In order to solve this, we designed the following strategy:
Using algorithm $InvTransvMassTOP$ we calculated for each event 4 values: 2 invariant masses M_{Inv_1}, M_{Inv_2} and 2 transverse masses M_{Trans_1}, M_{Trans_2}. If the event objects were labelled as: $l, j_1, j_2, hj_1, hj_2, me$ (lepton, light jet, light jet, heavy jet, heavy jet, missing energy; respectively), then the values calculated were:

Value calculated	Objects used
M_{Trans_1}	l, me, hj_1
M_{Inv_1}	j_1, j_2, hj_2
M_{Trans_2}	l, me, hj_2
M_{Inv_1}	j_1, j_2, hj_1

Table 4.3: Values calculated for each event using algorithm $InvTransvMassTOP$.

Then we had to select one of the two pair of values M_{Inv}, M_{Trans}, since only one of them represented the decay that has taken place. For this purpose we defined the variables:

$$R_1 = M_{Inv_1} - M_{Trans_1} \tag{4.10}$$

$$R_2 = M_{Inv_2} - M_{Trans_2} \tag{4.11}$$

As we have seen in Section 3.2 and 3.3, the invariant and transverse mass distributions have a peak at the rest mass of the reconstructed particle. This means that in most of the events the R which corresponds to the correct pair of values will be the smallest one, since its values are closer to each other.
Because of this we used algorithm $SelectTOP$ with R_1 and R_2 and we finally obtained the right value for the invariant and transverse mass of the top quark for each event.

4.2.1 Top quark transverse mass distribution

Plotting with SciDAVis [25] the transverse mass frequency distribution for a bin size of 1 GeV:
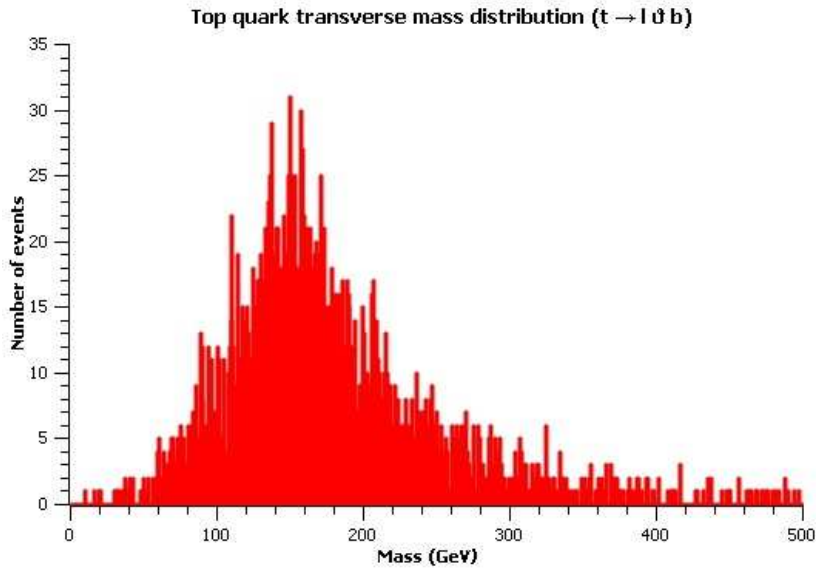

Figure 4.6: Top quark transverse mass frequency distribution from 0 to 500 GeV.

It was hard to determine where was the cut off, but taking a closer look and using the same approach as in Section 4.1.2 for the W boson mass, we estimated the top quark mass:
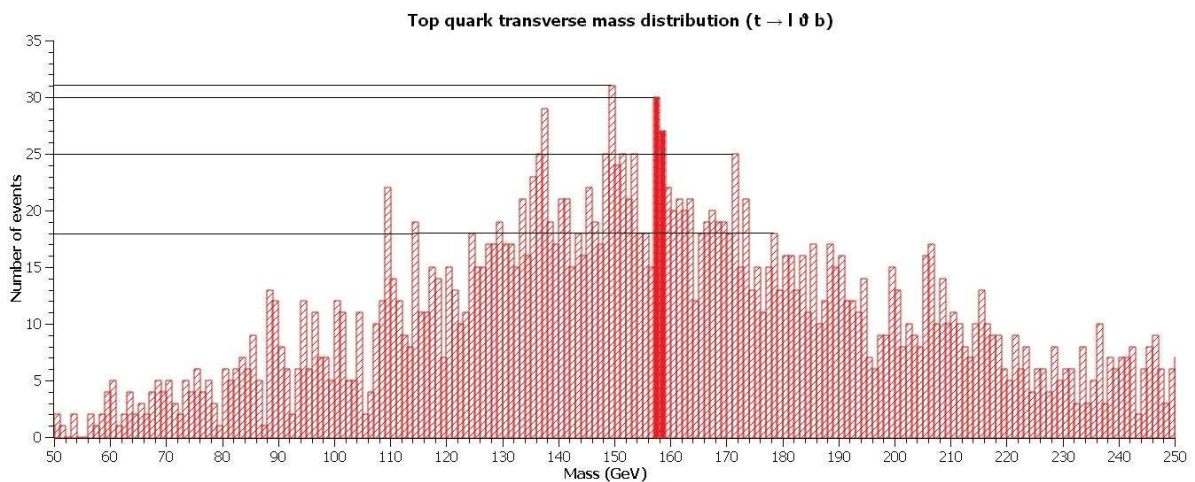

Figure 4.7: Top quark transverse mass frequency distribution from 50 to 250 GeV. The bins highlighted correspond to the estimated top quark mass.

We can see the difference in the number of events for each peak after the maximum of the distribution. After the bins highlighted, the peaks have far less number of events, so these bins were a good estimation of the top quark mass:

$$157 \leq M_t \leq 159\,GeV \Rightarrow M_t = 158 \pm 1\,GeV \tag{4.12}$$

However, if we took a bin size of 2.5 GeV and repeated the same procedure:
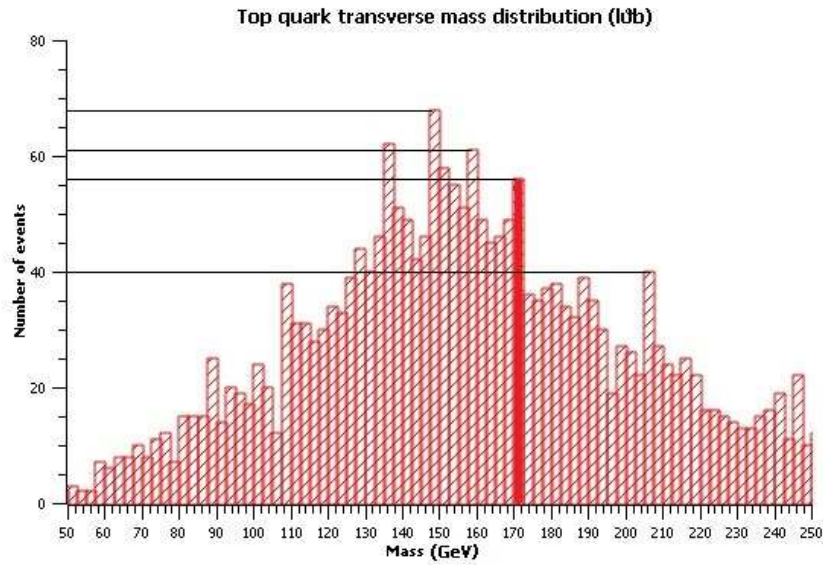

Figure 4.8: Top quark transverse mass frequency distribution from 50 to 250 GeV. The bin highlighted corresponds to the estimated top quark mass.

This time the difference between the number of events for each peak indicated us that the top quark mass was:

$$170 \leq M_t \leq 172.5\,GeV \Rightarrow M_t = 171.25 \pm 1.25\,GeV \tag{4.13}$$

4.2.2   Top quark invariant mass distribution

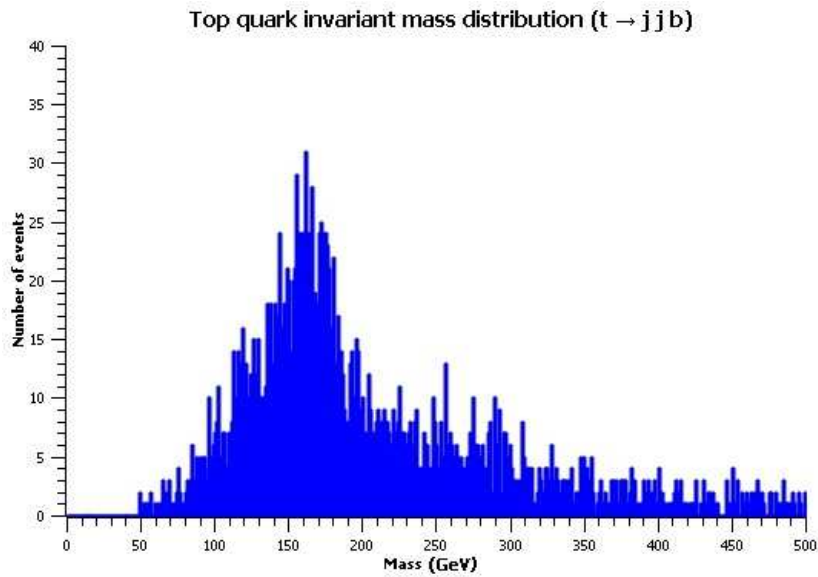We plotted the invariant mass distribution for a bin size of 1 GeV:

Figure 4.9: Top quark invariant mass frequency distribution from 0 to 500 GeV.

Here it was also difficult to estimate the exact peak of the distribution, so to help us we fitted the data to a Lorentzian function [27] of the type:

$$f(x) = y_0 + 2 \cdot \frac{A}{\pi} \cdot \frac{w}{4(x - x_c)^2 + w^2} \tag{4.14}$$
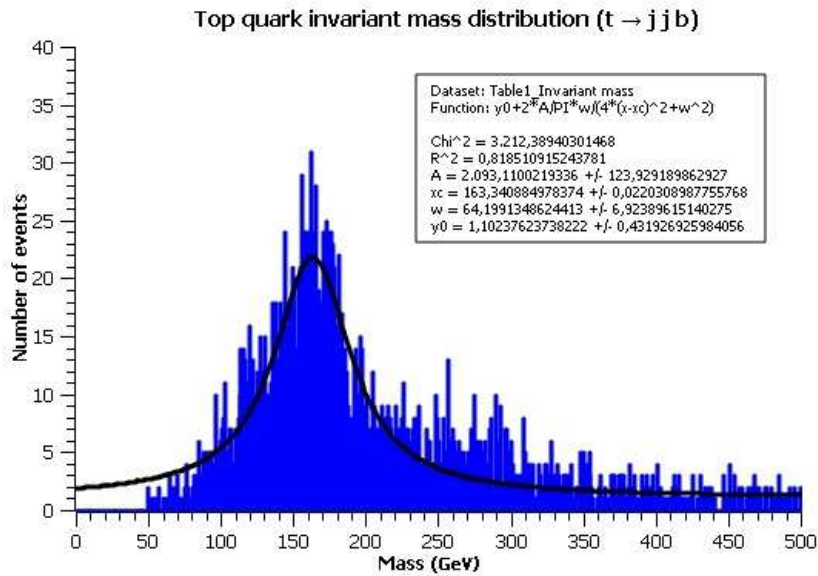

Figure 4.10: Top quark invariant mass frequency distribution from 0 to 500 GeV. The black line is the Lorentzian fit function.

This fit enabled us the following value for the peak of the distribution, which is the invariant top quark mass:

$$x_c = M_t = 163.34 \pm 0.02 \, GeV \tag{4.15}$$

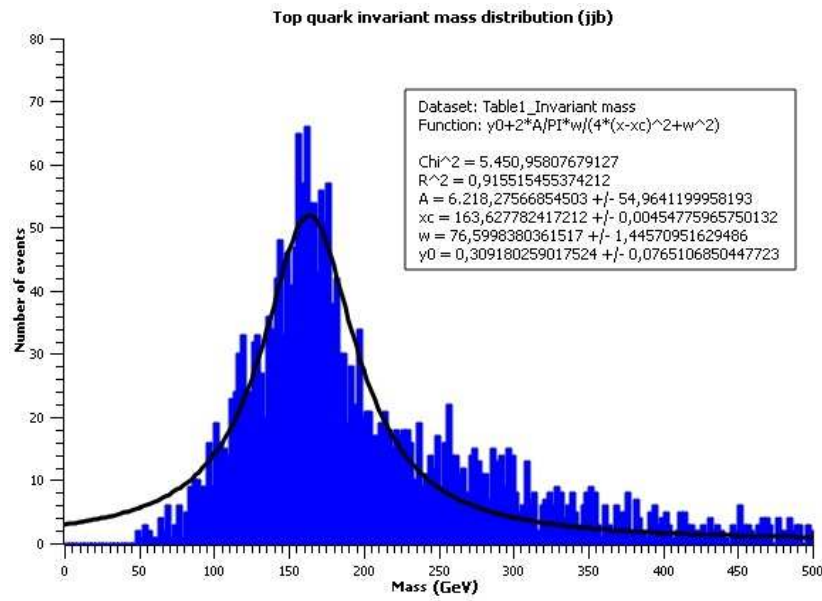We repeated the same procedure for a bin size of 2.5 GeV:

Figure 4.11: Top quark invariant mass frequency distribution from 0 to 500 GeV. The black line is the Lorentzian fit function.

$$x_c = M_t = 163.628 \pm 0.005 \, GeV \tag{4.16}$$

5. Conclusions

5.1   Z and W bosons

For the Z boson we obtained:

$$M_Z^e = 90.5 \pm 0.5 \, GeV \; \longrightarrow \;\; \text{for the } e^-e^+ \text{ decay.} \tag{5.1}$$

$$M_Z^\mu = 91.5 \pm 0.5 \, GeV \; \longrightarrow \;\; \text{for the } \mu^-\mu^+ \text{ decay.} \tag{5.2}$$

We can now average the results for each decay route:

$$M_Z = 91.0 \pm 0.5 \, GeV \tag{5.3}$$

To compare this last value we will use as a reference the data from [12] i.e, $M_Z^{[12]} = 91.1876 \pm 0.0021 \, GeV$. This value $M_Z^{[12]}$ is inside the interval of M_Z, so our obtained value is reliable.
Regarding the mean lifetime, we obtained:

$$\tau^e = 3.3 \cdot 10^{-25} \, s \; \longrightarrow \;\; \text{for the } e^-e^+ \text{ decay.} \tag{5.4}$$

$$\tau^\mu = 2.2 \cdot 10^{-25} \, s \; \longrightarrow \;\; \text{for the } \mu^-\mu^+ \text{ decay.} \tag{5.5}$$

So the mean value is:
$$\tau = 2.8 \cdot 10^{-25} \, s \tag{5.6}$$

The mean lifetime of the Z boson is [28]: $\tau \simeq 2.6 \cdot 10^{-25} \, s$, so our obtained value is not just in the same order of magnitude, but also quite close to it, which is a good signal.

In the case of the W boson:

$$M_W = 79.5 \pm 1.5 \, GeV \; \longrightarrow \;\; \text{for the } e^\pm \nu \text{ decay.} \tag{5.7}$$

The reference value for the W boson mass [12] is $M_W^{[12]} = 80.385 \pm 0.015 \, GeV$, which is inside the interval of M_W. In the case of the transverse mass we have more indetermination, as it was harder to determine the cut off of this distribution than the peak of the invariant mass one.

This way, applying techniques from the LHC Olympics such as the invariant and transverse mass methods to our data [9], we have obtained results which are in great accordance with others obtained in particle accelerators experiments. Hence this proves that these methods are reliable and we can apply them to reconstruct the mass of a particle that has decayed into two products.

5.2   Top quark

Our values for the top quark lepton+jets decay $t\bar{t} \to b + \bar{b} + q + \bar{q} + l^\pm + \nu$ were:

- Transverse mass method $t \to b + l^\pm + \nu$:

$$1 \text{ GeV bin size} \; \longrightarrow \; M_t = 158 \pm 1 \, GeV \tag{5.8}$$

$$2.5 \text{ GeV bin size} \; \longrightarrow \; M_t = 171.25 \pm 1.25 \, GeV \tag{5.9}$$

- Invariant mass method $t \rightarrow b + q + \bar{q}$:

$$1 \text{ GeV bin size } \longrightarrow M_t = 163.34 \pm 0.02 \, GeV \tag{5.10}$$

$$2.5 \text{ GeV bin size } \longrightarrow M_t = 163.628 \pm 0.005 \, GeV \tag{5.11}$$

The top quark mass reference value [12] is: $M_t^{[12]} = 173.1 \pm 0.6 \, GeV$.

In the case of the transverse mass values, they are quite different, with a discrepancy between them of 13.25 GeV. The reference value is not inside either of both intervals, and the first result differs in 8,7 %, although the second one only in 1,1 %. However, this dependence on the scale we are using to look at the distribution tells us that we cannot trust either of the results.

In the case of the invariant mass, both of the values obtained are quite similar between them, but far apart from the reference value, as they differ from it in around the 5,6 % of its value.

The discrepancies observed in the results of the top quark mass can be interpreted as a failure of these mass methods for events with more complexity than just 2 decay products. This can be due to the methods are not powerful enough to correctly predict the values of complex events, but also because our procedure to select the correct invariant and transverse mass for each event does not always work, just most of the times. The values obtained the rest of the times can influence in the statistics making the distributions shift to another value or widen them.

Looking to the dependence on the scale in the transverse mass values we can also infer that the results could be improved using more data, as the final number of selected events for the transverse and invariant mass was 2611. This number is really small if we compare it with the number of total events, since in each of the files [9] we had around 190,000 objects labelled.

Analysis of statistics for our samples [9] has shown that reconstruction of the top quark mass at hadron colliders is not as simple as it could seem when looking at its decay products. Reconstruction algorithms, detector inefficiencies or QCD effects makes the identification of resonances harder. Even for simpler events like the W and Z boson decays the methods we have used are not good enough to precisely determine their masses, as the values obtained have too much indetermination compared with the reference ones [12].

One way of continuing this project could be taking into account the SM background. At real colliders physics signals are swamped by this SM "noise", requiring more sophisticated analysis techniques to be developed. We could try to apply these techniques to more realistic data and learn about new challenges faced in the LHC.

Appendix A

VB functions and subroutines

Function or subroutine name	Description
Function $NObjects$	Counts how many objects of a specific type are in a given range
Function $Jets$	Counts how many light jets are in a given range
Function $HeavyJets$	Counts how many heavy jets are in a given range
Sub $CopyPaste$	Copies the first 15 cells of the selected row into another data sheet.
Sub $Delete$	Deletes the content of the given range of cells.
Sub $CopyPasteRange$	Executes $CopyPaste$ for each row of the given range.
Function $Signature$	Compares the number of objects of each type with selected values. If they are the same, the function returns TRUE, else FALSE.
Sub $SingleEventJets$	Counts the number of objects of each type of the given range making use of $NObjects$, $Jets$ and $HeavyJets$. Then $Signature$ is run and, if the result is TRUE, $CopyPasteRange$ is also run.
Sub $JetStart$	Executes $SingleEventJets$ for each event inside the given range.
Sub $InvariantMass$	Calculates the invariant mass (using Eq. (3.6)) of the selected type of objects for each event in the given range.
Sub $InvariantMassWxy$	Does the same as $InvariantMass$ but it calculates the invariant mass using Eq. (3.7).
Sub $TransverseMass$	Calculates the transverse mass (using Eq. (3.12)) of the selected type of objects for each event in the given range.
Sub $InvTransvMass$	Calculates the invariant mass (using Eq. (3.6)) of an electron or muon and a neutrino and the transverse mass (using Eq. (3.12)) of two light jets for each event in the given range.
Sub $InvTransvMassTOP$	Calculates the invariant mass (using Eq. (3.6)) of an electron or muon, a neutrino and a bottom jet; and the transverse mass (using Eq. (3.12)) of two light jets and another bottom jet. Then the bottom jets are exchanged and it calculates the invariant and transverse mass again. All of this is run for each event in the given range.
Sub $SelectTOP$	Compares two values and copies in another data sheet the smaller one for each row in the given range.

Table A.1: Description of each function and subroutine used to select and analyse data.

Bibliography

[1] CERN. Standard model. `https://home.cern/about/physics/standard-model`.

[2] CERN. Supersymmetry. `https://home.cern/about/physics/supersymmetry`.

[3] Patricia Schwarz. The official string theory web site. `http://superstringtheory.com/index.html`.

[4] CERN. The large hadron collider. `https://home.cern/topics/large-hadron-collider`.

[5] Communications Education and Outreach Group CERN. Lhc faq the guide. `http://cds.cern.ch/record/2255762/files/CERN-Brochure-2017-002-Eng.pdf`, February 2017.

[6] CERN. Lhc olympics. `https://public-archive.web.cern.ch/public-archive/en/Spotlight/SpotlightOlympics-en.html`.

[7] Microsoft. Excel 2016. `https://products.office.com/en-us/excel`.

[8] Ben Chinowsky. Getting started with vba in excel 2010. `https://msdn.microsoft.com/en-us/library/office/ee814737(v=office.14).aspx`.

[9] Steve F. King. Bsc datafiles. `https://www.dropbox.com/sh/rhzmvrlsc3ux9nw/AADguwdHTM8RrXOyGDhSYfYua?dl=0`.

[10] Tao Han. Collider phenomenology: Basic knowledge and techniques. `https://arxiv.org/pdf/hep-ph/0508097v1`, August 2005.

[11] Symmetry Rashmi Shivni. This is the closest thing we have to a master equation of the universe. `http://www.businessinsider.com/standard-model-physics-master-equation-2017-6?op=0`.

[12] C.Patrignani et al. (Particle Data Group). Chin. Phys. C, 40, 100001 (2016) and 2017 update.

[13] CERN. The higgs boson. `https://home.cern/topics/higgs-boson`.

[14] Kresimir Kumericki. Feynman diagrams for beginners. February 2016.

[15] Ann Heinson. Useful diagrams of top signals and backgrounds. `https://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html`.

[16] Tao Han. The top priority at the lhc. International Journal of Modern Physics A, 23(25):4107–4124, 2008.

[17] C. M. S. Collaboration and LHCb Collaboration. Observation of the rare bs0 + decay from the combined analysis of cms and lhcb data. Nature, 522(7554):68–72, June 2015.

[18] Xabier Cid Vidal and Ramon Cid Manzano. Atlas. `https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.atlas/idioma/en_GB`.

[19] LHC Olympics. Lhc olympics topics. `http://www.jthaler.net/olympicswiki/doku.php`.
login: olympian
password: blackbox.

[20] Stephen Mrenna et al. Pythia. `http://home.thep.lu.se/~torbjorn/Pythia.html`.

[21] Ken Arnold et al. Herwig 7. `https://herwig.hepforge.org/`.

[22] John Conway et al. Pgs 4. `http://conway.physics.ucdavis.edu/research/software/pgs/pgs4-general.htm`.

[23] LHC Olympics. Backgrounds are hard! `http://www.jthaler.net/olympicswiki/doku.php?id=lhc_olympics:backgrounds_hard#backgrounds_are_hard`.

[24] Dr. David P. Stern. Energetic particles. `https://www-spof.gsfc.nasa.gov/Education/wenpart1.html`.

[25] Russell Standish. Scidavis: Welcome. `http://scidavis.sourceforge.net/`.

[26] Peter J. Mohr, David B. Newell, and Barry N. Taylor. Codata recommended values of the fundamental physical constants: 2014. Rev. Mod. Phys., 88:035009, Sep 2016.

[27] WolframMathWorld. Lorentzian function. `http://mathworld.wolfram.com/LorentzianFunction.html`.

[28] Mohit Joshi. The Mysterious World of Fundamental Particles: Cosmic Beginnings. 2015.