Sapienza Università di Roma

Facoltà di Scienze Matematiche Fisiche e Naturali

Corso di Laurea in Fisica

Tesi di Laurea Specialistica

Ricerca del segnale $p\bar{p} \rightarrow WZ \rightarrow l\nu_l b\bar{b}$ con l'esperimento CDF al Tevatron

Candidata

Priscilla Pani

Matricola

1127491

Relatore

dott. Stefano Giagu

Correlatore

dott. Marco Rescigno

Anno Accademico 2009/2010

Faculty of Scienze Matematiche Fisiche e Naturali

Corso di Laurea in Fisica

Master Degree in Physics

Search for $p\bar{p} \rightarrow WZ \rightarrow l\nu_l b\bar{b}$ signal with the CDF experiment at Tevatron

Student

Priscilla Pani

Matriculation number

1127491

Supervisor

dott. Stefano Giagu

Supervisor

dott. Marco Rescigno

Academic Year 2009/2010

Contents

Introduction

The high energy physics has made huge steps forward the comprehension of the inner most nature of our universe and the matter we are composed of. The experimental discoveries, and the theories of the last 50 years that the experimental discoveries had confirmed or inspired, made possible to build a *theory of the interactions*. Weak interactions have been discovered and unified with the Electromagnetic ones in the Standard Model, which is the most widely experimentally tested and confirmed model of this century. The only prediction which is still unconfirmed is the existence of a particle, the Higgs boson, which provides particles with mass, interacting with them, in a spontaneous symmetry breakdown that doesn't violate the natural gauge symmetry of the Lagrangian of the system. One of the ways in which the Standard Model has been tested during the last 20 years is by accelerating e^+e^- (LEP) or $p\bar{p}$ (Tevatron) particles in a circular ring and colliding them inside a detector which is designed to reveal the final reaction products. We now have two operating hadron colliders in the world. The Tevatron at Fermilab laboratory of Chicago, collides protons against anti-protons since 1989 and has reached its maximum energy in the mass center of 1.96 TeV since 2001. It has collected approximately 7 fb^{-1} of data so far, that allowed important discoveries, as the top quark one, B_s mixing, precision measurements of some of the Standard Model free parameters, e.g. the W mass, and search for New Phenomena. The LHC at CERN in Geneva is a proton proton collider and has started the data acquisition in March 2010, at a center of mass energy of 7 TeV, thus beating the world record of the Tevatron. LHC however has not yet either the integrated luminosity nor the detailed understanding of the detectors to start investigating Higgs or di-boson production. The purpose of this work is to analyse the data of the CDF experiment at Tevatron to search for the associate production of a $W^\pm$ and Z gauge boson, looking for them in the lepton, neutrino plus jets final state, This process is predicted by the Standard Model but not revealed yet in this particular channel, both for its small cross section ($\sigma_{WW/WZ} \sim 16\ \mathsf{pb}^{-1}$) and for the huge backgrounds we have to deal with. The W^+W^-or $W^\pm Z$ in $l\ \bar{\nu}_l\ j\ j$ process has been measured for the first time in [4] and represents the starting point of this work. Our aim is to discriminate $W^\pm Z$ process from W^+W^-one requiring the decay of the Z boson in two b-quarks. The evidence of a peak on the invariant mass distribution will allow a tuning of the invariant mass resolution of b-jets. In addition, one of the main motivations for this quest is the similarity of this exactly predicted process with the $W^\pm$ H associate production signature, for which it represents a test of the searching tools and techniques, as long as an irreducible background that must be understood before such Higgs search is performed.

Part I

Theoretical and experimental issues

Chapter **1**

Theoretical overview

A brief introduction to the Standard Model is given in this first chapter, with a particular attention to the steps that brings to the introduction of the gauge bosons in the model. The $W^{\pm}$ and Z masses, widths and decay partial widths are summarized, and the di-boson associate production is discussed, in order to describe the $W^{\pm}Z$ in $l\bar{\nu}_l\, j\, j$ process that is the subject of this analysis and highlight the importance of this search, which is based on the $W^{\pm}Z$ signature similarity with the associate production of a light Higgs with a $W^{\pm}$ boson.

1.1 Standard Model and Gauge Bosons

In the '70s, three of the four natural forces: weak, electromagnetic and strong interactions were described in a single model, known as the Standard Model, whose predictive power has been proved in the recent years with unprecedented precision in particle physics. The most important example of the predictive power of the Standard Model is the gyromagnetic ratio of the electron, whose theoretical calculation has been experimentally proved to the 14th decimal digit. In this model, particles are described by fields. There are matter spinorial fields, that create and annihilate the consituent particles (fermions) and are described by the Dirac Lagrangian

$$\mathcal{L}_{\mathcal{D}} = \bar{\psi}\, \imath\, \partial\!\!\!/\, \psi \tag{1.1}$$

and the gauge bosons fields that propagate the interaction and are described as free fields by the Maxwell-Proca equation

$$\mathcal{L}_M = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 A_\mu A^\mu \qquad (F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu) \tag{1.2}$$

where m is the mass of the boson ($m = 0$ for the photon, electromagnetic interaction carrier)

In the SM theory, $SU(2)_L \otimes U(1)_Y$ is the minimal symmetry group to which the electroweak lagrangian should belong in order to describe by Noether theorem all the experimental conserved currents, i.e. the electromagnetic and the weak currents. It means that spinorial fields are described as doublets for the SU(2) group with their associate neutrinos, as eigenstates of chirality with -1 eigenvalue (*left-handed* eigenstates), one for each generation (e, μ, τ).

$$\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L \quad \begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}_L \quad \begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}_L \qquad\qquad (e)_R \quad (\mu)_R \quad (\tau)_R$$

Since Goldhaber has experimentally proved that neutrinos with positive chirality eigenvalues do not exists, the *right-handed* fermions hould be singlets for $SU(2)_L$. In this way Weinberg and Salam wrote the electroweak lagrangian and required a local gauge symmetry.

$$\begin{aligned}
\mathcal{L}_{SM} = & \sum_i^{family} \bar{\psi}_i(\imath\not{\partial} - m)\psi_i - g\sum_i^{family} \bar{\psi}_i \frac{m_i H}{2M_W}\psi_i - e\sum_i^{family} q_i\bar{\psi}_i\gamma^\mu\psi_i A^\mu \\
& - \frac{g}{2\sqrt{2}}\sum_i^{family} \bar{\psi}_i\gamma^\mu(1-\gamma_5)(T^+W_\mu^+ + T^-W_\mu^-)\psi_i \\
& - \frac{g}{2\cos\theta_W}\sum_i^{family} \bar{\psi}_i\gamma^\mu(g_V^i - g_A^i\gamma_5)\psi_i Z_\mu
\end{aligned} \tag{1.3}$$

This important requirement was due to Yang and Mills mechanism and is essential if we want a theory that respects the Einstein's special relativity. In fact a global phase change would mean an instantaneous propagation of the information (the interaction) in the whole space. The Yang and Mills theory implies the existence of $W^\pm$ and Z bosons, which are the propagators of the interaction through the space. These important components of the electroweak theory were discovered in 1983 at SPS of CERN, by UA1 and UA2

FERMIONS matter constituents spin = 1/2, 3/2, 5/2, ...

Leptons spin = 1/2		
Flavor	Mass GeV/c^2	Electric charge
ν_e electron neutrino	$<1\times10^{-8}$	0
e electron	0.000511	–1
ν_μ muon neutrino	<0.0002	0
μ muon	0.106	–1
ν_τ tau neutrino	<0.02	0
τ tau	1.7771	–1

Quarks spin = 1/2		
Flavor	Approx. Mass GeV/c^2	Electric charge
u up	0.003	2/3
d down	0.006	–1/3
c charm	1.3	2/3
s strange	0.1	–1/3
t top	175	2/3
b bottom	4.3	–1/3

Figure 1.1: The standard model constituents

experiments. Ever since milions of $W^{\pm}$s and Zs have been produced and detected, and their masses, widths and decay partial widths studied and compared with the theory's predictions.

1.2 Quarks and quantum cromodynamics

Strong interactions, as electroweak ones, are decribed within the Standard Model. The strong interaction is mediated by gluons and the strong charge is called *color*. The symmetry group of this interaction is $SU(3)$ and we can combine a color ($SU(3)$) and an anticolor ($\overline{SU(3)}$) to obtain an octect of gluons that carry color charge and a *white singlet* that has no physical evidence. The quarks are the particles that interacts by strong interaction. According to Standard Model they are divided into a *left-handed* doublet and a right-handed singlet as leptons and neutrinos:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L \qquad \begin{matrix} (u)_R & (d)_R & (c)_R \\ (s)_R & (t)_R & (b)_R \end{matrix}$$

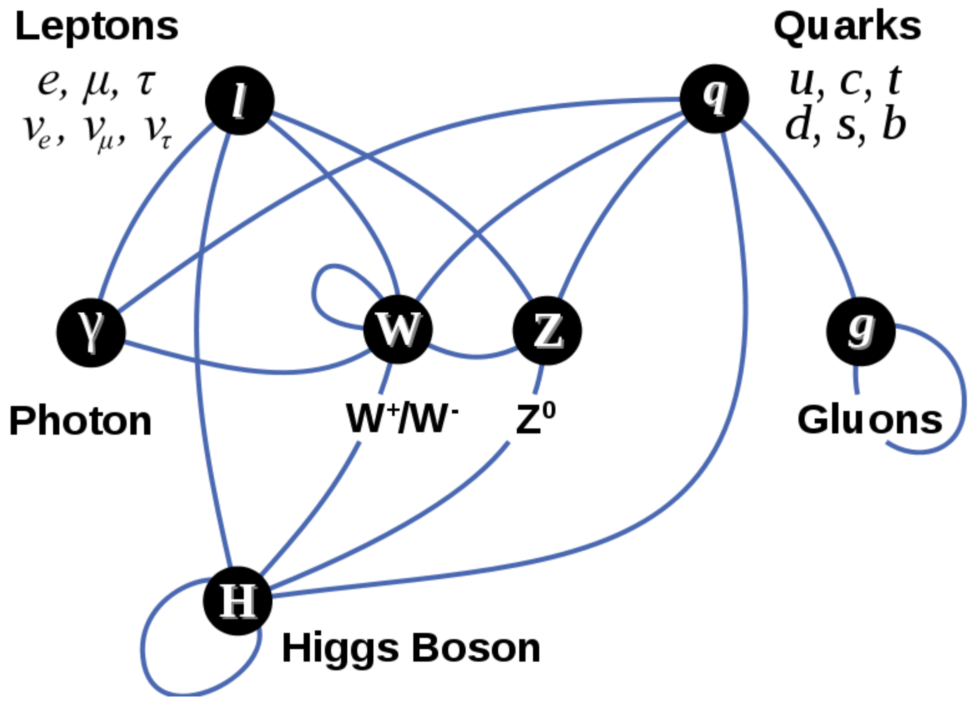


Figure 1.2: A diagram summarizing the tree-level interactions between elementary particles described in the Standard Model. Vertices (darkened circles) represent types of particles, and edges (blue arcs) connecting them represent interaction

The most important experimental feature of the quarks is that they carry color charge. Since the color charge has never been observed in nature, interactions and final states of the interaction must be singlets under $SU(3)$ color. This means that we can't observe *bare* quarks that have to bind into color neutral states called hadrons, and color is confined. In Fig. 1.3 we show a schematic example of the internal structure of the proton. When highly energetic quarks or gluons are produced in an high energy physics experiment a process called hadronization or showering takes place: after a quark-antiquark pair, or more in general a parton[1], is produced in an interaction, the potential between them, due to gluon exchange,

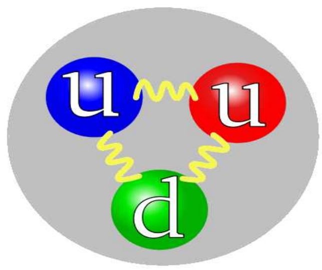


Figure 1.3: Schematic representation of the internal structur of the proton. Although quarks are coloured, the total color charge of the hadron is null

[1] with this word Feynmann originally calld that particle the proton is made of, which are both valence quarks, sea quarks and gluons

	m [GeV]			Γ [GeV]			Decay Modes [%]			
$W^{\pm}$	80.425	±	0.038	2.124	±	0.041	$l\bar{\nu}$	10.80	±	0.09
							hadrons	67.70	±	0.27
Z	91.1876	±	0.0021	2.4952	±	0.0023	l^+l^-	3.3658	±	0.0023
							hadrons	69.91	±	0.06
							invisible	20.00	±	0.06
							$b\bar{b}$	15.12	±	0.05

Table 1.1: Masses, widths and decays brancing ratios of $W^{\pm}$ and Z gauge bosons ([5])

tries to keep them together. When the strenght reaches a breaking point further quark-antiquark pairs are created and finally bind together with the original partons. This process involves a large number of interactions at different scales until the scale of hadrons is reached. The process is then essentially non-perturbative and not completely theoretically under control. The quarks could also radiate gluons that creates other $q\bar{q}$ pairs. The final state in which we observe the parton generated in the interaction is a collimated jet of (white) particles approximately in the direction of the original parton. We refer to the Analysis Tools Chapter (Ch. 3) for a description of the experimental techniques developed at CDF to deal with *jets*.

1.3 W and Z production and decay modes

$W^{\pm}$ and Z gauge bosons, as long as the photon (γ), are the carriers of the electroweak interaction and have a foundamental role within the Standard Model. W^+ is an electrically charged gauge boson that mediate the weak force, as its antiparticle W^-. The Z is neutral. They have spin 1, so obey to Bose-Einstein statistics and they are massive. We report in Table 1.3 their masses, widths and decay modes brancing ratios.

They have an inclusive cross section at Tevatron's center of mass energy ($\sqrt{s} = 1.96$ TeV) of:

$$\sigma_{W\to l_\nu \nu} = 2.5 \text{ nb}$$

$$\sigma_{Z\to l^+l^-} = 250 \text{ pb}$$

$W^{\pm}$ and Z decay immediately ($\tau = \frac{\hbar}{\Gamma} \approx 10^{-25}$ s) and are thus revealed by detecting their decay products The leptonic decay channel is usually the cleanest, expecially in

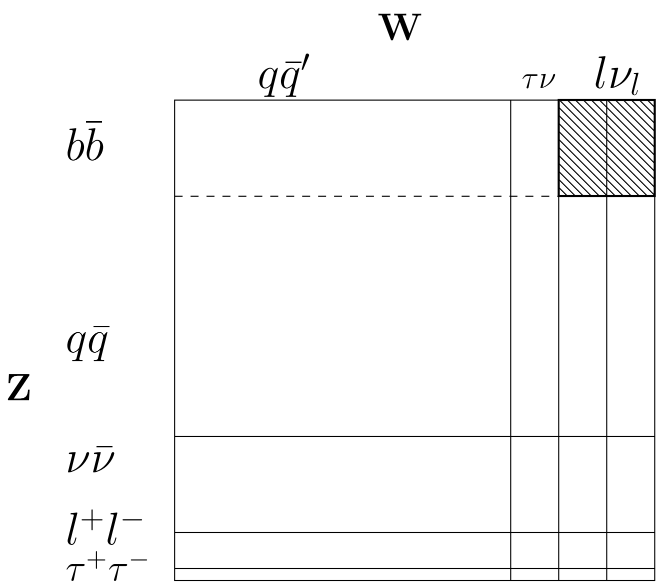


Figure 1.4: Figurative rapresentation of the combination of the hadronic/leptonic decay of a $W^{\pm}$ and a Z in associate production. The highlighted box gives a visual idea of the rate limitation of a $WZ \rightarrow l\bar{\nu}_l b\bar{b}$ search. The convention is that $l = e, \mu$

$\sqrt{s} = 2$ TeV ($p\bar{p}$)	W^+W^-	ZW^+or ZW^-	ZZ
Born level [pb]	10.0	1.46	1.22
NLO [pb]	13.0	1.95	1.56

Table 1.2: Leading and next to leading order calculation of di-boson asociate production cross sections. [6]

hadronic collisions as in CDF, since a background of high transverse momentum leptons can only be produced in Heavy Flavour hadron decays and thus significantly suppressed with respect to QCD jet production. However the $W^{\pm}$ decays into muon or electrons only 20 times out of 100, while the brancing ratio of the quarks decay channel is 70%. For the Z boson we have 6% for the leptonic decay modes[2] compared to 70% of the hadronic mode (the remaining 20% is the "invisible" decay mode, i.e. neutrinos).

One of the peculiarity of the Standar Model of the interactions is that is based on a non abelian theory. This implies that these bosons have auto-interactions, and vertexes with three and four gauge bosons are permitted, making possible a wide range of vector bosons associate production (W^+W^-, $W^{\pm}Z$, γZ, ZZ), that have a theoretical interest

[2]we always mean only $e^{\pm}$ and $\mu^{\pm}$ because the small lifetime of the $\tau^{\pm}$ doesn't permits his direct detection inside the detector and its dominant decay channel is again the hadronic one.

for the investigation of the non-abelian nature of the Standard Model. The cross section of the associate gauge bosons productions, predicted by the Standard Model are shown in Table 1.3 and in Figure 1.5, to highlight the first challenge of this processes observation: a small cross section.

The signal we are going to study is the associate production of a W and another vector boson, that could be either another $W^{\pm}$ or a Z if we require 2 jets in the final state, but could only be a Z , if the two jets are required to originate from a b quark.

We would like to give a visual explanation of what in practice means the choice of the decay mode in an associate production search in Picture 1.4. We have to chose a decay mode for each one of the gauge bosons, hence multiplicate the brancing ratios of the two of them. The requirement of both leptonic modes, even providing a clear signature, would mean a strong limitation in rate. However, the requirement of the leptonic decay of the $W^{\pm}$ and hadronic decay of the other gauge boson it is produced in association with, increase the expected events less than the expected background, which becomes overwhelming. This is why the first observation and cross section measurement of a di-boson production (W^+W^-, $W^{\pm}Z$, $Z\gamma$) at hadronic collider has been made in the last five years in the all lepton channel [7], [8] and [9].

1.4 The WZ channel

As stated before, the aim of this analysis is the search for the associate production of a $W^{\pm}$ and a Z. The decay channel we choose is $l\nu_l+jets$ and we further require that the two jets are originated by a b-quark to distinguish the hadronic decay of the Z produced in association with a W from the associate production of two $W^{\pm}$s, one of which hadronically decays.

In fact, all the cross section measurements of di-boson associate production performed so far at CDF in the lepton + jets channel are combined measurements of the W^+W^-+$W^{\pm}Z$ associate production. The resolution on the jet-jet invariant mass of the two leading jets of the event make impossible to distinguish the $W^{\pm}$ contribution, that is a gaussian centered in the $W^{\pm}$ mass, from the Z one, that is centered in the Z mass at only 10 GeV of distance.

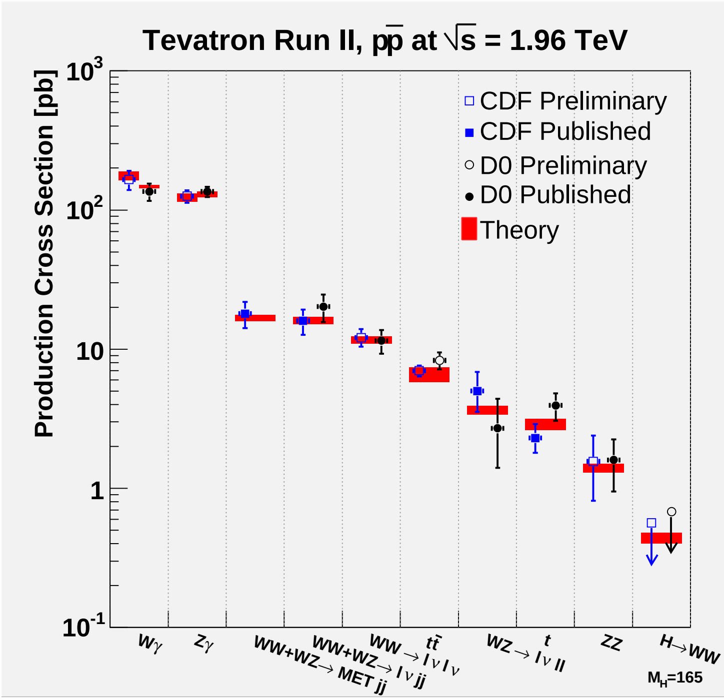


Figure 1.5: The most recent cross section mesurements (and upper limits) related to Higgs and gauge bosons, performed by CDF and DØ and compared to their theoretical prediction.

The most recent measurement of the $W^+W^-+W^\pm Z$ production cross section [4] is

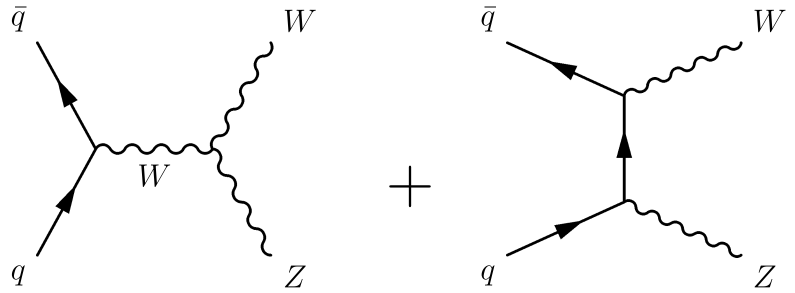
$$\sigma_{WW/WZ} = (16.0 \pm 3.3)\ \text{pb}$$

Some analysis have already been performed searching for the $WZ \to l\bar{\nu}_l b\bar{b}$ production, but an evidence has not been observed yet. The best upper limit of this process cross section, in the channel we are interested in, is

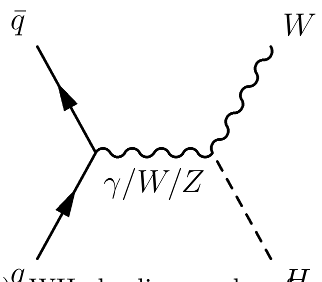
$$\sigma_{WZ\to l\bar{\nu}_l b\bar{b}} < 3.9 \cdot \sigma_{SM} \qquad @\ 95\%\ \text{C.L.}$$

estimated in [10] using a fit on a neural network output distribution for signal and backgrounds.

The $W^\pm Z$ production cross section measurement is a very interesting channel for many practical and theoretical aspects. First of all, it appears as the first resonance in b-jets invariant mass distribution, whose detection would permit an accurate calibration of the b-jets resolution and energy correction, direclty from data. For this reason, we do not implement any sophisticated technique, such as neural networks or multivariate discriminants, for the discrimination of the signal from the background. In fact the implementation of such tools, although enhancing the signal sensitivity, modifies the invariant



(a) $W^{\pm}Z$ leading order feynmann dyagrams



(b) WH leading order feynmann diagram

Figure 1.6: Comparison between the topology of $W^{\pm}Z$ and WH associate production processes

mass distribution of the selected events in a not well predictable way, making impossible a calibration of the b-jet resolution and energy correction with the Z peak. In addition, in this way our search is also a "model independent" search for a $b\bar{b}$ resonance, in the di-jet invariant mass distribution.

The $W^{\pm}Z$ production cross section measurement it's a very challenging measurement since this process has a cross section which is 3 order of magnitude smaller than the typical processes that have a similar signature (Tab. 1.3). The cross section of the $W^{\pm}Z$ associate production is (4.0 ± 0.7) pb, that, taking into account the brancing ratio of the $W \rightarrow l_{\nu}\nu$ (20%) and the $Z \rightarrow b\bar{b}$ (15%), where $l = e, \mu$, results in an effective cross section of

$$\sigma_{WZ \rightarrow l\bar{\nu}_l b\bar{b}} = \sigma_{WZ} \cdot \mathcal{BR}_{Z \rightarrow b\bar{b}} \cdot \mathcal{BR}_{W \rightarrow l_{\nu}\nu} = 0.13 \text{ pb} \tag{1.4}$$

So we are dealing with a very small signal that should be compared with a huge background (see Table 1.3). Despite this difficulty, this channel could provide important results as far as Triple Gauge Couplings are concerned, that would permit further confirmation of Standard Model predictions or could be a probe for New Physics.

However, the real importance of this channel is related to the strong similarity of the decay topology to the Higgs associate production, whose Feynmann diagram is shown in Figure 1.6, in comparison with the $WZ \to l\bar{\nu}_l b\bar{b}$ one.

Channel	$\sigma[\,\mathsf{pb}]$
$p\bar{p} \to W(\to l\bar{\nu}) + \mathcal{N}(\geqslant 0)jets$	~ 2066
$p\bar{p} \to Z(\to ll) + \mathcal{N}(\geqslant 0)jets$	187 ± 50
$p\bar{p} \to W(\to l\nu) + Q\bar{Q}\ \ (Q = c, b)$	22.7 ± 0.1
$p\bar{p} \to t\bar{t} + ...;\ t \to Wb$	7.4 ± 0.1
$p\bar{p} \to tb;\ t \to Wb$	2.86 ± 0.12

Table 1.3: Some example of background processes with their cross section, as calculated by ALPGEN. Refer to Section 5 for a more detailed explanation

This similarity means not only that the analysis techniques developed for $W^{\pm}Z$ are suitable for Higgs associate production too, but also that the $W^{\pm}Z$ associate production represents an irreducible background to the Higgs one. In fact, the tails of the $W^{\pm}Z$ peak in the invariant mass distribution can contaminate the signal of a light Higgs, since the energy resolution at Z peak is approximately 10 GeV.

The cross section of the $WH \to l\bar{\nu}_l b\bar{b}$ channel is five times smaller than the $WZ \to l\bar{\nu}_l b\bar{b}$ one (Eq. (1.7)) for an Higgs of mass $m_H = 120$ geV/c^2, therefore we expect that the $WZ \to l\bar{\nu}_l b\bar{b}$ signal will be simpler to be observed than the Higgs and a good knowledge of it is absolutely relevant for the Tevatron Higgs search perspective.

$$\sigma_{WH} = 0.16\ \mathsf{pb} \qquad \sigma_{WZ} = 4.0\ \mathsf{pb} \tag{1.5}$$

$$\mathcal{BR}_{H\to b\bar{b}} = 67.9\% \qquad \mathcal{BR}_{Z\to b\bar{b}} = 15.1\% \qquad (\mathcal{BR}_{W\to l\bar{\nu}_l} = 10.6\%) \tag{1.6}$$

$$\sigma_{WH\to l\bar{\nu}_l b\bar{b}} = 0.024\ \mathsf{pb} \qquad \sigma_{WZ\to l\bar{\nu}_l b\bar{b}} = 0.13\ \mathsf{pb} \tag{1.7}$$

Chapter **2**

Tevatron hadronic collider and CDF detector

The data for the analysis described in this thesis was collected with the Collider Detector at Fermilab (CDF) located at the Fermi National Accelerator Laboratory. In the following sections, a brief introduction to the Tevatron Collider and a description of the CDF Run II *detector are given.*

2.1 The Tevatron Collider

The Tevatron Collider [11] located at the Fermi National Accelerator Laboratory (Fermilab) in Batavia (Illinois, USA) is a proton-antiproton ($p\bar{p}$) collider with a center-of-mass energy of 1.96 TeV. As shown in figure 2.1, this complex has five major accelerators and storage rings used in successive steps, as is explained in detail below, to produce, store and accelerate the particles up to 980 GeV. This huge and complex apparatus was commissioned in 1983 as the first large scale superconducting synchrotron in the world. The first $p\bar{p}$ occurred in 1983 and, since then, various periods of collider operations alternate with fixed-target operations or shut-down periods for upgrades of the machine. Each period or

Tevatron collider operations is conventionally identified as a Run [1] The present anaysis uses the data collected in Run II. The performance of the collider is evaluated in terms of two key parameters:

- $\sqrt{s}$: the center of mass energy, which means the energy avalaible in the collision; this parameter defines the accessible phase-space of the final states of the reaction as well the mass of the particles that can be created. Until March 2010 Tevatron was the most energetic collider of the world. Now the world record belongs to LHC at Cern.

- $\mathcal{L}$: the integrated luminosity; this is the coefficient of proportionality between the number of events of a process and its cross section as in Eq. 2.1.

$$N = \mathcal{L} \cdot \sigma \tag{2.1}$$

2.1.1 Protons and antiprotons production

The acceleration cycle starts with the production of protons from ionized hydrogen atoms H^-, which are accelerated to 750 KeV by a Cockroft-Walton electrostatic accelerator. Pre-accelerated hydrogen ions are then injected into the Linac where they are accelerated up to 400 MeV by passing through a 150 m long chain of radio-frequency (RF) accelerator cavities. To obtain protons, the H^- ions are passed through a carbon foil which strips their electrons off. Inside the Booster the protons are merged into bunches and accelerated up to an energy of 8 GeV prior to entering the Main Injector. In the Main Injector, a synchrotron with a circumference of 3 km, the proton bunches are accelerated further to an energy of 150 GeV and coalesced[2] together before injection into the Tevatron.

The production of the antiproton beam is significantly more complicated. This process is the main hindrance to the increase of the Luminosity of the Tevatron. The cycle starts with extracting a 120 GeV proton beam from the Main Injector onto a stainless steel target. This process produces a variety of different particles, among which are antiprotons [3]. The particles come off the target at many different angles and they are focused into

[1] the Run shoul;d not be confused with the run, defined in CDF as a continuous period of data-taking in approximately constant detector and beam conditions

[2] coalescing is the process of merging proton bunches into one dense, high density, bunch

[3] The production rate, for 8 GeV antiprotons, is about $18\bar{p}/10^6 p$

a beam line with a Lithium lens. In order to select only the antiprotons, the beam is sent through a pulsed magnet which acts as a charge-mass spectrometer. The produced antiprotons are then injected into the Debuncher, an 8 GeV synchrotron, which reduces the spread in the energy distribution of the antiprotons. After that, the antiproton beam is directed into the Accumulator, a storage ring in the Antiproton Source, where the antiprotons are stored at an energy of 8 GeV and stacked to 10^{12} particles per bunch. The antiproton bunches are then injected into the Main Injector and accelerated to 150 GeV.

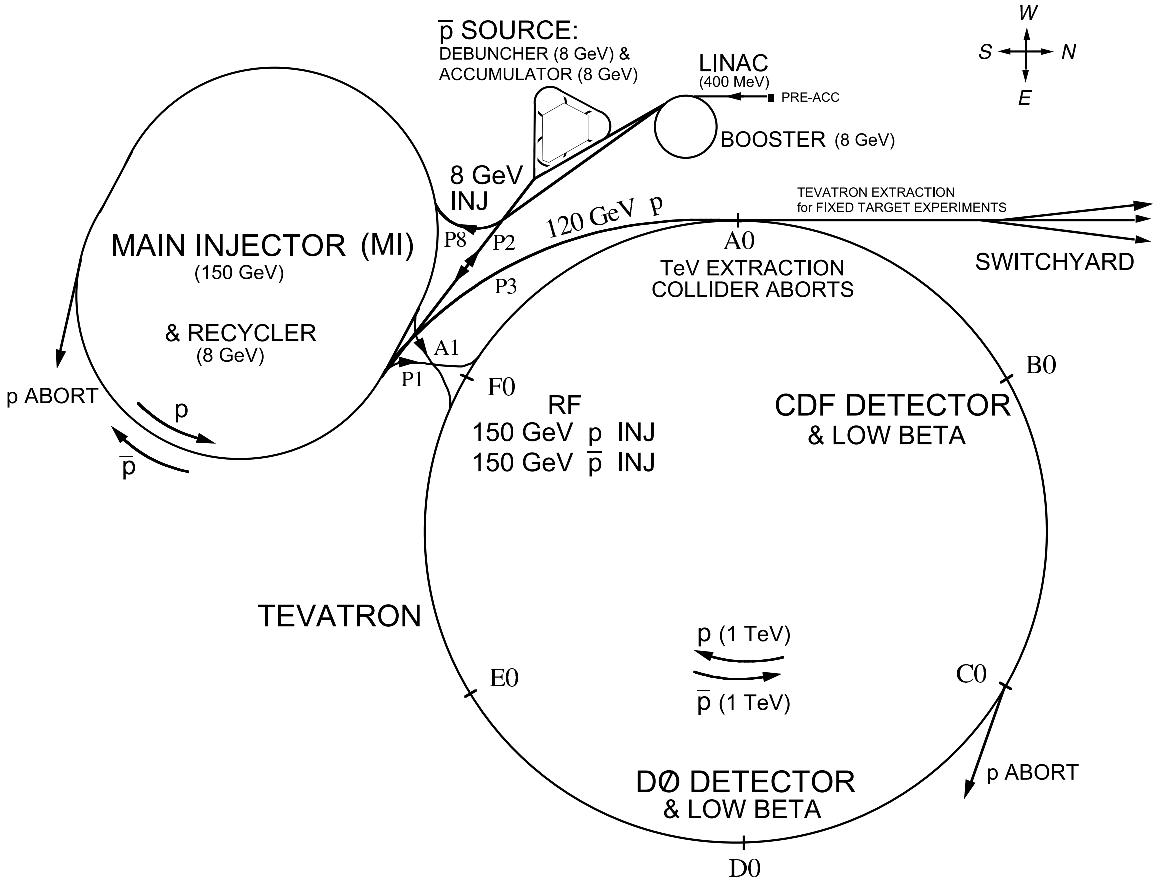


Figure 2.1: The Tevatron Collider Chain at Fermilab.

Finally, 36 proton and antiproton bunches are inserted into the Tevatron, a double acceleration ring of 1 km of radius, where their energy is increased up to 980 GeV in approximately 10 seconds. Proton and antiproton bunches circulate around the Tevatron in opposite directions guided by superconducting magnets and where their orbits cross at the two collision points, BØ and DØ, where CDF II and DØ detectors are respectevly located. Once that the maximum energy is reached, the luminosity have to be maximized. and the beam collimated as much as possible using high-power quadrupole magnets ("low-β squeezer") insalled on the beam pipe at either side of the detectors in order to reduce the transversal spread of the beam and both avoid the detectors damages caused by tha beam halo and increase the collision rate in the interaction region. When the beam reaches the stable condition, with an approximately gaussian distribution on the transverse plane of $\sigma_{x,y} \approx 25 - 30\ \mu\mathsf{m}$ and a bunch longitudinal dimension of $\approx 60 - 70\ \mathsf{cm}$ ($\sigma_z = 30\ \mathsf{cm}$), the detectors are switched on and the data-taking starts. space0.3cm In the absence of a crossing angle or position offset, the luminosity at the CDF or DØ is given by the expression:

$$L = \frac{f_{bc} N_b N_p N_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)} F\left(\frac{\sigma_l}{\beta^*}\right), \tag{2.2}$$

where f_{bc} is the revolution frequency, N_b is the number of bunches, $N_{p(\bar{p})}$ is the number of protons (antiprotons) per bunch, and $\sigma_{p(\bar{p})}$ is the transverse and longitudinal rms proton (antiproton) beam size at the interaction point.

F is a form factor with a complicated dependence on beta function, β^*, and the bunch length, σ_l. The beta function is a measure of the beam width, and it is proportional to the beam's x and y extent in phase space. Table 2.1 shows the design Run II accelerator parameters [12].

Parameter	Run II
number of bunches (N_b)	36
revolution frequency [MHz] (f_{bc})	1.7
bunch rms [m] σ_l	0.37
bunch spacing [ns]	396
protons/bunch (N_p)	2.7×10^{11}
antiprotons/bunch ($N_{\bar{p}}$)	3.0×10^{10}
total antiprotons	1.1×10^{12}
β^* [cm]	35

Table 2.1: Accelerator parameters for Run II configuration.

Figure 2.2 and 2.3 show, respectively, the evolution in the integrated luminosity, defined as $\mathcal{L} = \int L\, dt$, and the instantaneous luminosity delivered by Tevatron since the machine was turned on up to March 2010. The progressive increase in the integrated luminosity and the continuous records in the instantaneous luminosity prove the good performance of the accelerator.

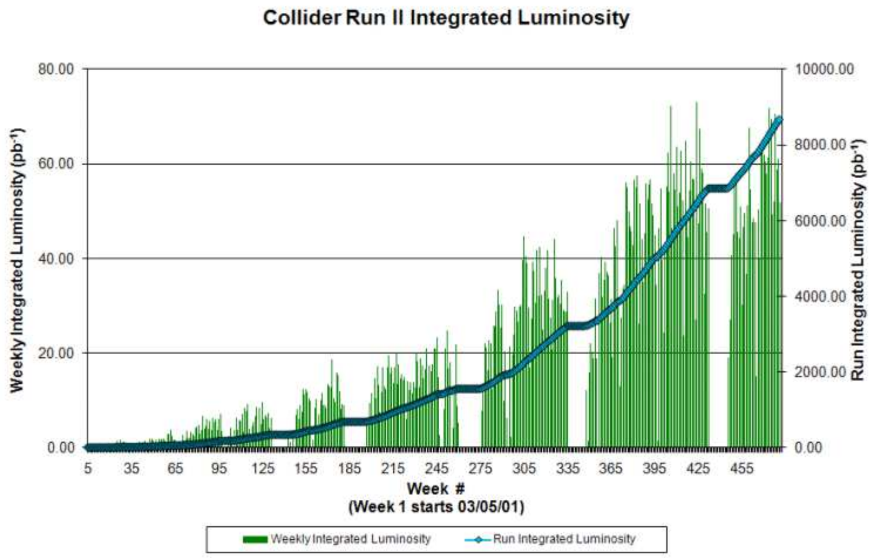


Figure 2.2: Tevatron Collider Run II Integrated Luminosity. The vertical green bar shows each week's total luminosity as measured in pb^{-1}. The diamond connected line displays the integrated luminosity.

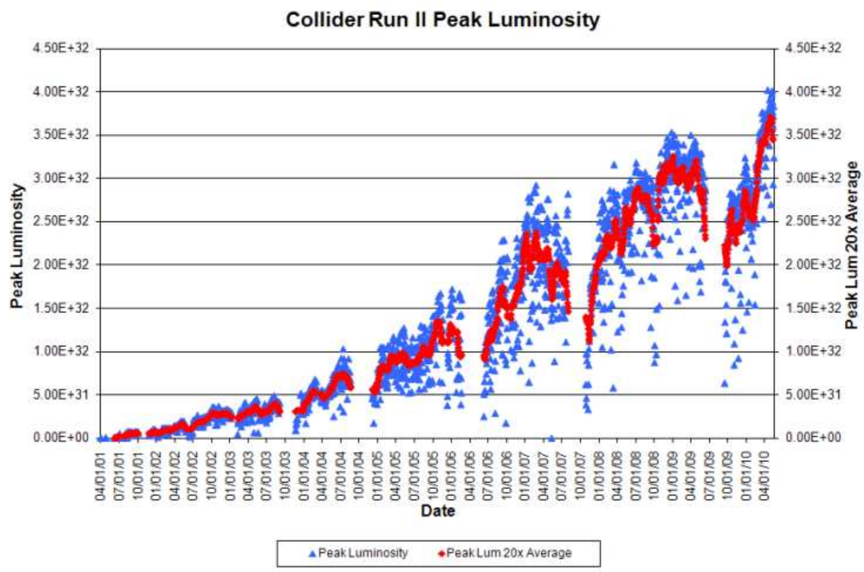


Figure 2.3: Tevatron Collider Run II Peak Luminosity. The blue squares show the peak luminosity at the beginning of each store and the red triangle displays a point representing the last 20 peak values averaged together.

2.2 The Collider Detector at Fermilab (CDF)

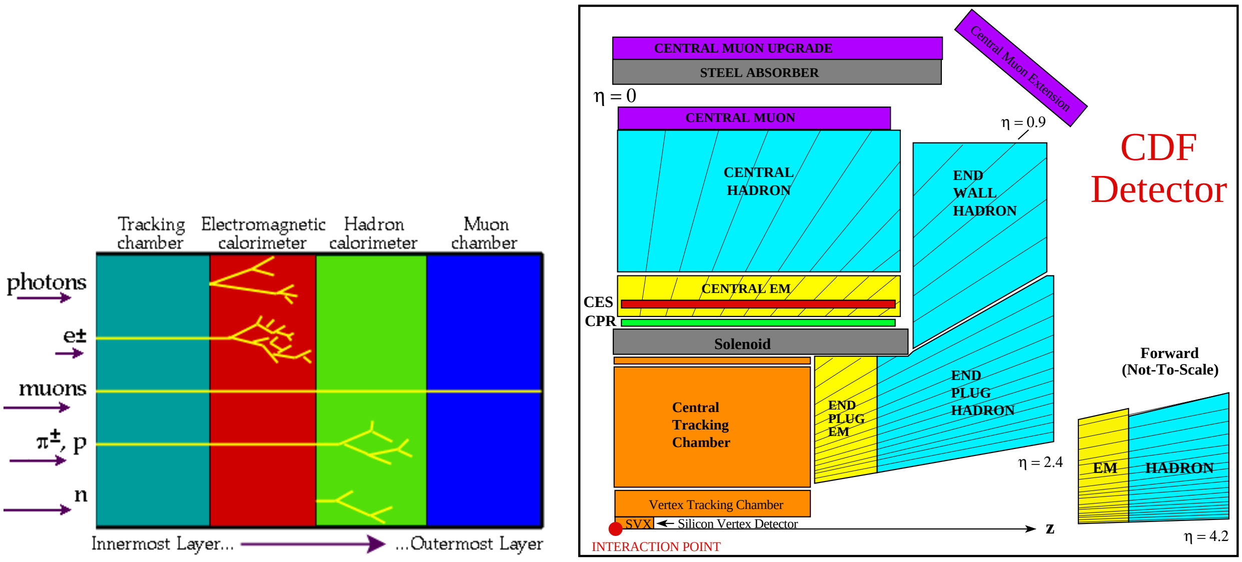
The CDF Run II detector [13] is an azimuthally and forward-backward symmetric multi-purpose apparatus installed at the BØ interaction point (see Fig 2.1) of the Tevatron collider and designed to determine energy, momentum and whenever possible, the identity of a broad range of particles produced in the $p\bar{p}$ collisions at the Tevatron. The original CDF detector, commissioned in 1985, was upgraded and modified during the years. Its most extensive upgrade started in 1995 and led up to the current detector whose operation is generally referred to as Run II. This detector is in operation since 2001 and its essential features are:

- A tracking system, that provides a measurement of the charged particle momenta, event z vertex position and detects secondary vertices.
- A Time-of-Flight system, to identify charged particles.
- A non-compensated calorimeter system, with the purpose of measuring the energy of charged and neutral particles produced in the interaction.
- Drift chambers and scintillators to muon detection.

These components are assembled at different radial distances in CDF (Fig. 2.5), in the so called "onion structure" and permits to obtain varied information of the revealed particles, that are combined for the identification, as shown in Figure 2.4. The tracking system is contained in a superconducting solenoid, 1.5 m in radius and 4.8 m in length, which generates a 1.4 T magnetic field parallel to the beam axis, in order to deflect charged particles to measure their momentum. Calorimetrers and muon systems are all outside the solenoid.

2.3 Reference frame

As already stated, the CDF detector is approximately colindrically symmetric around the beam axis. Its geometry can be described both in cartesian and in cylindrical coordinates. The left-handed carthesian system is centered on the nominal interaction point with the $\hat{z}$ axis pointing in the direction of the protin beam and the $\hat{x}$ axis on the Tevatron plane,



(a) performing particle identification with the different components of the detector

(b) radial composition of the detector

Figure 2.4: The so called “onion structure” of the detector permits to combine different informations in order to perform particle identification.

pointing radially outside. The cylindrical coordinates are the azimuthal angle ϕ ($\phi = 0$ on the $\hat{x}$ direction) and the polar angle θ ($\theta = 0$ along the positive $\hat{z}$ axis). Since the total momentum is usually not balanced in a $p\bar{p}$ interaction, since each parton involved carries a variable fraction of the (anti)proton momentum, in this kind of environment is customary to use a variable invariant under $\hat{z}$ boosts instead of the non-invariant azimuthal angle θ. This variable is the rapidity and is defined as:

$$Y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \tag{2.3}$$

It's limit in case of massles particles is the pseudorapidity. This variable doesn't depends any more on the momentum along $\hat{z}$ axis, which is usually unknown, and it's a function of the sole polar angle:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \tag{2.4}$$

The pseudorapidity is commonly used to identify different detector regions according to their position respect to the beamline and interaction vertex position, as shown in Figure 2.4(b).

For the same reason that lead us to define the boost-invariant rapidity, every variable defined in CDF, such as energy, momentum, etc ... has its corresponding projection in the transerve plane, denoted with a T in subscript, which is the only plane in which the event is (theoretically) balanced.

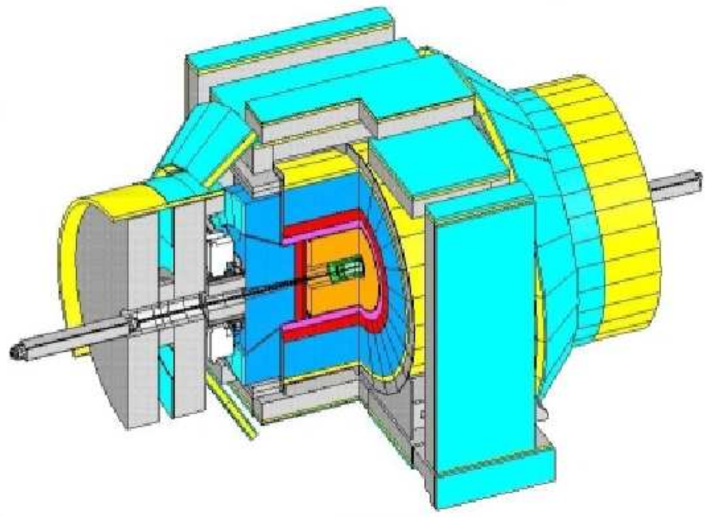


Figure 2.5: Isometric view of the CDF Run II detector.

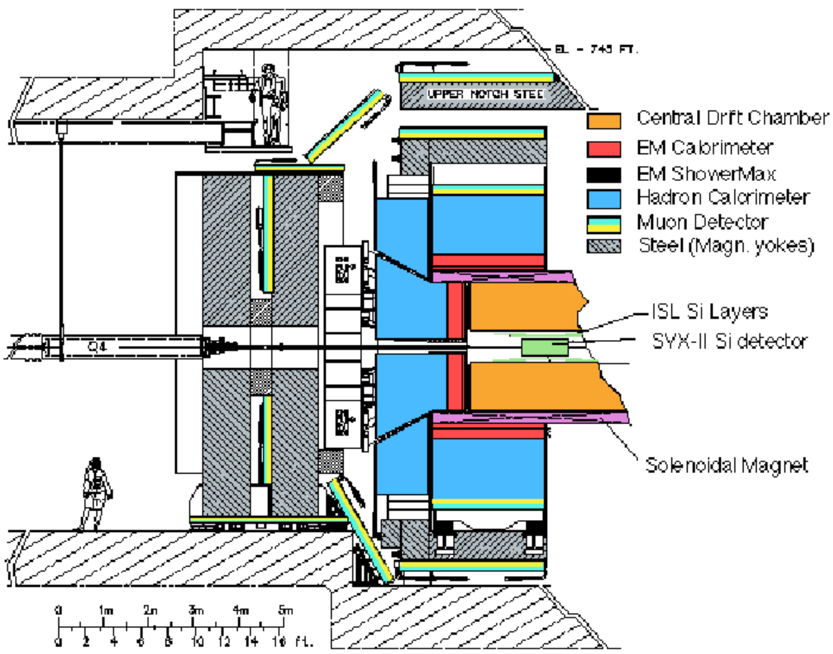


Figure 2.6: $r \times \eta$ side view of the CDF Run II detector.

2.3.1 Tracking and Time of Flight systems

The heart of CDF analysis technique is an high efficient and precise tracking system. From large to small radii the tracking system is composed of a huge drift chamber and a silicon inner tracker. Both these components are plunged in a solenoidal magnetic field of 1.4 T; the escape threshold for a particle in this field is $p_T > 0.3$ GeV.

Inner tracker The inner tracker (Fig. 2.7) is a silicon microstrip detector [14], which must be radiation-hard due its proximity to the beam and determines the impact parameter resolution. It extends from a radius of r = 1.5 cm from the beam line to r = 28 cm, covering $|\eta| < 2$ and has eight layers in a barrel geometry. The innermost layer is a single-sided silicon microstrip detector called Layer 00 which provides position measurement only in the $r \times \phi$ plane and and improves significantly the impact parameter resolution especially at low pt, due to its proximity to the beam line, and its overall low mass on all its lenght.

The first five layers after the Layer 00 constitute the Silicon Vertex Detector (SVXII) and the two outer layers comprise the Intermediate Silicon Layers system (ISL). These seven layers are made of double-sided silicon sensors, giving $r \times \phi$ and z position information. The best position resolution achieved is 9 μm in SVXII and the impact parameter resolution, including Layer 00, arrives to 40 μmm at $p_T > 3$ GeV/c. The impact parameter is calculated as the point of closest approach to the beam position, then it's uncertanty includes 30 μm of beam width. Therefore the resolution on the impact parameter due to the silicon detector performances are approximately 25 μm

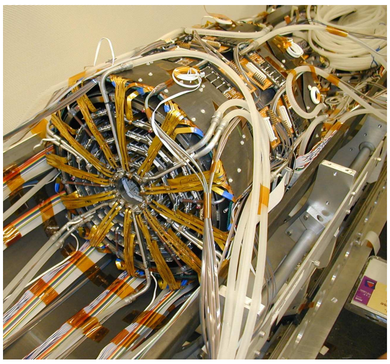


Figure 2.7: Photo of the SVX II silicon tracker

Outer tracker Surrounding the silicon detector is the Central Outer Tracker (COT) [15], showed in Figure 2.8, that is the anchor of the CDF Run II tracking system.

It is a 3.1 m long cylindrical drift chamber that covers the radial range from 40 to 137 cm

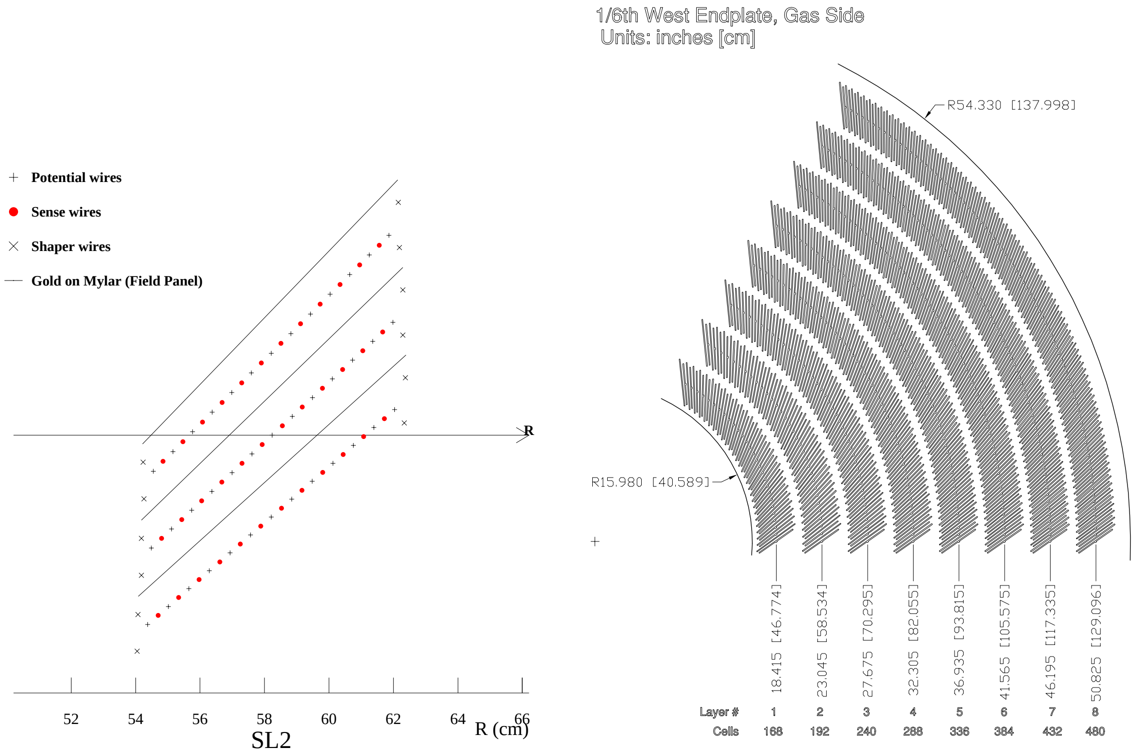
($|\eta| < 1$). The COT contains 96 sense wire layers, which are radially grouped into eight "superlayers", as inferred from the end plate section shown in figure 2.3.1.

Each superlayer is divided in ϕ into "supercells", and each supercell has 12 sense wires and a maximum drift distance that is approximately the same for all superlayers. Therefore, the number of supercells in a given superlayer scales approximately with the radius of the superlayer. The entire COT contains 30,240 sense wires. Approximately half the wires run along the z direction ("axial"). The other half are strung at a small angle (2°) with respect to the z direction ("stereo"). The combination of the axial and stereo information allows us to measure the z positions. Particles originated from the interaction point, which have $|\eta| < 1$, pass through all 8 superlayers of the COT.

Figure 2.8: Photo of the COT drift chamber

The supercell layout, shown in figure 2.3.1 for superlayer 2, consists of a wire plane containing sense and potential wires, for field shaping and a field (or cathode) sheet on either side. Both the sense and potential wires are 40 μm diameter gold plated tungsten. The field sheet is 6.35 μm thick Mylar with vapor-deposited gold on both sides. Each field sheet is shared with the neighboring supercell.

The COT is filled with an Argon-Ethane gas mixture and Isopropyl alcohol (49.5:49.5:1). The mixture is chosen to have a constant drift velocity, approximately 50 μm/ns across the cell width and the small content of isopropyl alcohol is intended to reduce the aging due to the ios deposition on the wires. When a charged particle passes through, the gas is ionized. Electrons drift toward the sense wires. Due to the magnetic field that the COT is immersed in, electrons drift at a Lorentz angle of 35°. The supercell is tilted by 35° with respect to the radial direction to compensate for this effect. The momentum resolution of the tracks in the COT chamber depends on the p_T and is measured to be approximately 0.15% GeV/c^{-1}, with corresponding hit resolution of about 140 μm [16]. In addition to the measurement of the charged particle momenta, the COT is used to identify particles, with $p_T > 2$ GeV, based on dE/dx measurements.



(a) Layout of wires in a COT supercell

(b) Layout of wire planes on a COT endplate

Figure 2.9: Layout of the COT supercell and endplate

Time of flight Just outside the tracking system, CDF II has a Time of Flight (TOF) detector [17]. It is a barrel of scintillator almost 3 m long located at 140 cm from the beam line with a total of 216 bars, each covering 1.7^o in ϕ and pseudorapidity range $|\eta| < 1$. Particle identification is achieved by measuring the time of arrival of a particle at the scintillators with respect to the collision time. Thus, combining the measured time-of-flight and the momentum and path length, measured by the tracking system, the mass of the particle can then determined. The resolution in the time-of-flight measurement is ≈ 100 ps and it provides at least two standard deviation separation between $K^{\pm}$ and $\pi^{\pm}$ for momenta p < 1.6 GeV/c.

As a summary, figure 2.10 illustrates the Tracking and Time of Flight systems.

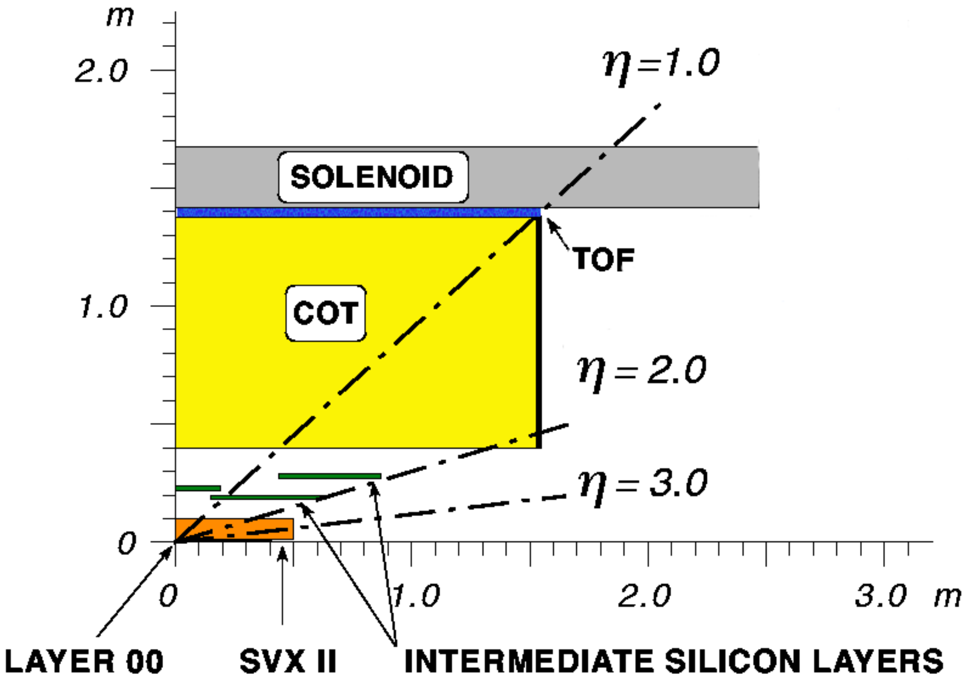


Figure 2.10: The CDF II tracker layout showing the different subdetector systems.

2.3.2 Calorimeter system

Surrounding the CDF tracking volume, outside of the solenoid coil, there is the calorimeter system. The different calorimeters that compose the system are scintillator-based detectors and segmented in projective towers (or wedges), in $\eta \times \phi$ space, that point to the interaction region. The total coverage of the system is 2π in ϕ and about $|\eta| < 3.64$ units in pseudorapidity and they are designeto to absorb up to $\sim 98\%$.

The calorimeter system is divided in two regions: central and plug. The central calorimeter covers the region $|\eta| < 1.1$ and is split into two halves at $|\eta| = 0$. The forward plug calorimeters cover the angular range corresponding to $1.1 < |\eta| < 3.64$, as it is shown in figure 2.11. Due to this structure two "gap" regions are found at $|\eta| = 0$ and $|\eta| \sim 1.1$.

CPR2: the central Preshower System This detector component is located just outside the solenoid coil. It is a scintillator layer that acts as a central pre-Radiation detector (CCR) for electrons and photons and provides a clear signature of the electromagnetic showers initiated in the solenoid coil. A Central crack Radiation detector extends the preshower to the mechanically intrigued regions between the calorimeter wedges, improv-

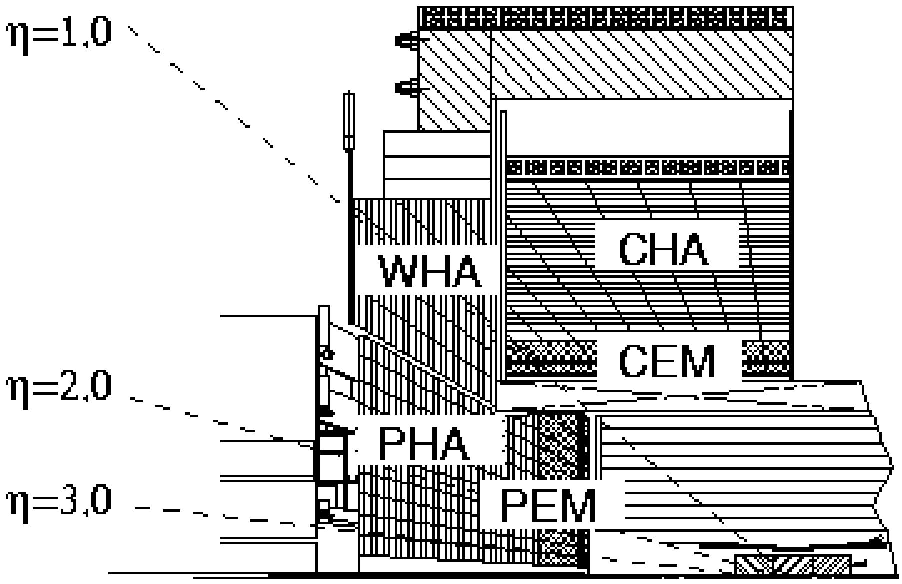


Figure 2.11: Elevation view of 1/4 of the CDF detector showering the components of the CDF calorimeter: CEM, CHA, WHA, PEM and PHA.

ing the jet energy resolution.

CES: the central shower maximum This detector component is located inside the electromagnetic calorimeter after 8 layers of lead ($\sim 5.9\ X_0$), in the position where usually the the electromagnetic shower reaches its maximum width. It is composed of proportional chambers that measure the local released ionisation projected in the two transverse directions. The CES resolution is about 1 cm in z and about 1 mm in $r \cdot \phi$.

2.3.2.1 Central Calorimeters

The central calorimeters consist of 478 towers, each one is 15^o in azimuth by about 0.11 in pseudorapidity. Each wedge consists of an electromagnetic component backed by a hadronic section. In the central electromagnetic calorimeter (CEM) [18], the scintillators are interleaved with lead layers. The total material has a depth of 18 radiation lengths (X_0) [4]. The central hadronic section (CHA) [19] has alternative layers of steel

[4]The radiation length X_0 describes the characteristic amount of matter transversed, for high-energy electrons to lose all but $1/e$ of its energy by bremsstrahlung, which is equivalent to $\frac{7}{9}$ of the length of the mean free path for pair e^+e^- production of high-energy photons. The average energy loss due to bremsstrahlung for an electron of energy E is related to the radiation length by $\left(\frac{dE}{dx}\right)_{brems} = -\frac{E}{X_0}$ and the probability for an electron pair to be created by a high-energy photon is $\frac{7}{9}X_0$.

and scintillator and is 4.7 interaction lengths deep (λ_0) [5]. The endwall hadron calorimeter (WHA), with similar construction to CHA, is located with half of the detector behind the CEM/CHA and the other half behind the plug calorimeter. The function of the WHA detector is to provide a hadronic coverage in the region $0.9 < |\eta| < 1.3$.

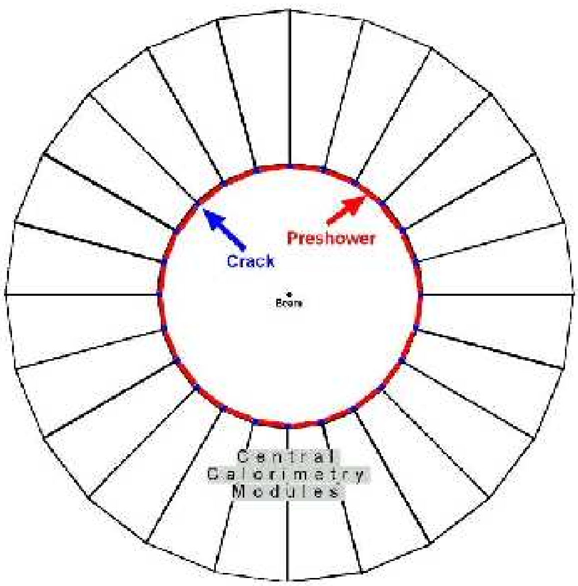


Figure 2.12: Calorimetry schematic picture

In the central calorimeter the light from the scintillator is redirected by two wavelength shifting (WLS) fibers, which are located on the ϕ surface between wedges covering the same pseudorapidity region, up through the lightguides into two phototubes (PMTs) per tower.

The energy resolution for each section was measured in the testbeam and, for a perpendicular incident beam, it can be parameterized as:

$$(\sigma/E)^2 = (\sigma_1/\sqrt{E})^2 + (\sigma_2)^2, \quad (2.5)$$

where the first term comes from sampling fluctuations and the photostatistics of PMTs, and the second term comes from the non-uniform response of the calorimeter. In the CEM, the energy resolution for high energy electrons and photons is $\frac{\sigma(E_T)}{E_T} = \frac{13.5\%}{\sqrt{E_T}} \oplus 1.5\%$, where E_T=$E\sin\theta$ being θ the beam incident angle. Charged pions were used to obtain the energy resolution in the CHA and WHA detectors that are $\frac{\sigma(E_T)}{E_T} = \frac{50\%}{\sqrt{E_T}} \oplus 3\%$ and $\frac{\sigma(E_T)}{E_T} = \frac{75\%}{\sqrt{E_T}} \oplus 4\%$, respectively.

2.3.2.2 Plug Calorimeters

One of the major components upgraded for the Run II was the plug calorimeter [20]. The new plug calorimeters are built with the same technology as the central components and replace the Run I gas calorimeters in the forward region. The $\eta \times \phi$ seg-

[5] An interaction length is the average distance a particle will travel before interacting with a nucleus: $\lambda = \frac{A}{\rho\sigma N_A}$, where A is the atomic weight, ρ is the material density, σ is the cross section and N_A is the Avogadro's number.

mentation depends on the tower pseudorapidity coverage. For towers in the region $|\eta| < 2.1$, the segmentation is 7.5^o in ϕ and from 0.1 to 0.16 in the pseudorapidity direction. For more forward wedges, the segmentation changes to 15^o in ϕ and about 0.2 to 0.6 in η.

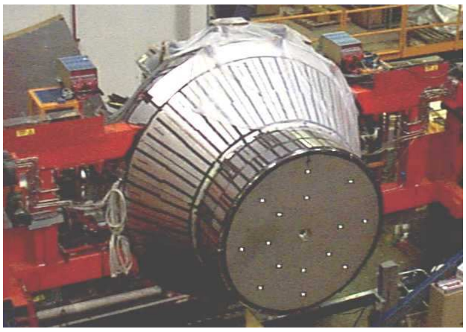


Figure 2.13: Photo of one of the plugs

As in the central calorimeters, each wedge consists of an electromagnetic (PEM) and a hadronic section (PHA). The PEM, with 23 layers composed of lead and scintillator, has a total thickness of about 21 X_0 . The PHA is a steel/scintillator device with a depth of about 7 λ_0. In both sections the scintillator tiles are read out by WLS fibers embedded in the scintillator. The WLS fibers carry the light out to PMTs tubes located on the back plane of each endplug. Unlike the central calorimeters, each tower is only read out by one PMT.

Testbeam measurements determined that the energy resolution of the PEM for electrons and photons is $\frac{\sigma}{E} = \frac{16\%}{\sqrt{E}} \oplus 1\%$. The PHA energy resolution is $\frac{\sigma}{E} = \frac{80\%}{\sqrt{E}} \oplus 5\%$ for charged pions that do not interact in the electromagnetic component. Table 2.2 summarizes the calorimeter subsystems and their characteristics.

Calorimeter	Coverage	Thickness	Energy resolution (E expressed in GeV)
CEM	$\|\eta\| < 1.1$	18 X_0	$\frac{13.5\%}{\sqrt{E_T}} \oplus 2\%$
CHA	$\|\eta\| < 0.9$	4.7 λ_0	$\frac{50\%}{\sqrt{E_T}} \oplus 3\%$
WHA	$0.9 < \|\eta\| < 1.3$	4.7 λ_0	$\frac{75\%}{\sqrt{E_T}} \oplus 4\%$
PEM	$1.1 < \|\eta\| < 3.6$	21 X_0, 1 λ_0	$\frac{16\%}{\sqrt{E}} \oplus 1\%$
PHA	$1.2 < \|\eta\| < 3.6$	7 λ_0	$\frac{80\%}{\sqrt{E}} \oplus 5\%$

Table 2.2: CDF II Calorimeter subsystems and characteristics. The energy resolution for the EM calorimeter is given for a single incident electron and that for the hadronic calorimeter for a single incident pion.

The central and forward parts of the calorimeter have their own shower profile detectors: shower maximum and preshower detectors. The Central Shower Maximum (CES) and the Plug Shower Maximum (PES) are positioned at about 6 X_0, while the Central Preradiator

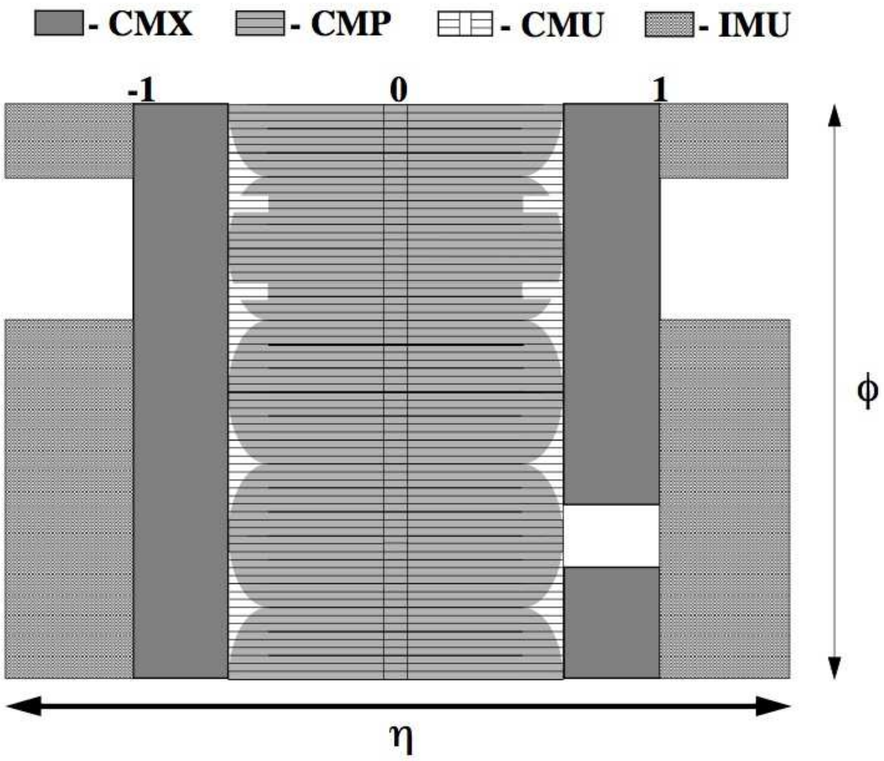


Figure 2.14: The η/φ coverage of the muon system. The shape is irregular because of the obstruction by systems such as cryo pipes or structural elements.

(CPR) and the Plug Preradiator (PPR) are located at the inner face of the calorimeters. These detectors help on particle identification, separating $e^{\pm}$, γs and $\pi^0 s$.

2.3.3 Muons system

The muon system (Fig. ??), which consists of sets of drift chambers and scintillators, is installed beyond the calorimetry system as the radially outermost component of CDF Run II detector (r$\sim$3.5 m). The muon system [21, 22] is divided into different subsystems, that cover the pseudorapidity range $|\eta| < 2.0$: the Central Muon Detector (CMU), the Central Muon Upgrade Detector (CMP/CSP), the Central Muon Extension Detector (CMX/CSX) and the Intermediate Muon Detector (IMU). They are very important elements of the detector as trigger elements as well as in the offline analysis of muons events. The z and ϕ coordinates of the muon candidate are often provided by the chambers while the scintillator detectors are used for triggering and spurious signal rejection.

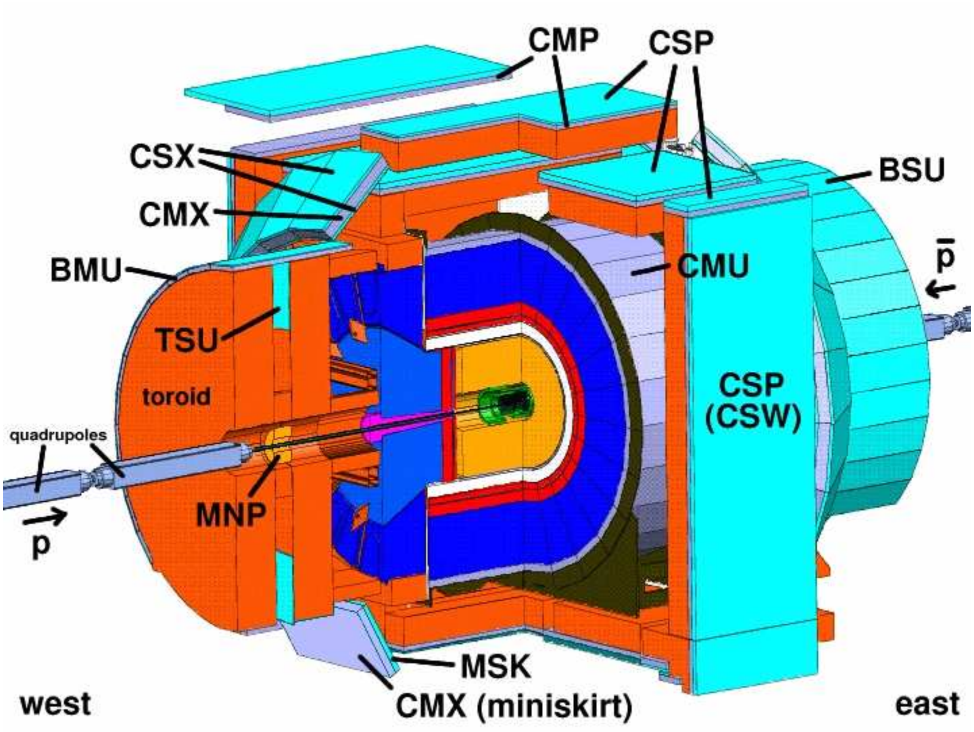


Figure 2.15: Scheme of the muon system

2.3.3.1 CMU

The Central Muon chambers (CMU) is a set of four layered drift chamber sandwiches housed on the back of wedges inside the central calorimeter shells covering the region $|\eta| < 0.6$ (see Fig 2.16). CMU is largely unchanged from Run I except for the fact that it operates now in proportional mode rather than in limited-streamer mode. The minimum detection energy for this system is ~ 1.4 GeV.

2.3.3.2 CMP

The Central Muon uPgrade (CMP) consist of a 4-layer sandwich of wire chamber operated in proportional mode and covering most of the $|\eta| < 0.6$ region (see Fig 2.16). Unlike mostly of the CDF components, this detector is not cylindrically-shaped, but box-like, since CMP uses the magnet return yoke steel as an absorber. On the outer surface of CMP, a scintillation layer, the Central Scintillator Upgrade (CSP), measures the muons trasversal time. The system CMU/CMP, which is called CMUP, detects muons having a minimum energy of ~ 3 GeV.

2.3.3.3 CMX

The muon extension CMX is a large system of drift chambers-scintillator sandwices arranged in two truncated conical arches detached from the mail CDF detector to cover the region $0.6 < |\eta| < 1.0$ and detects muons of minimum energy of $\sim$ 2 GeV. Due to main detector frame structure, some region of this subdetector are characterized by a peculiar geometry, as shown in Figure 2.17. In particular, there ar two subdetectors of the CMX apparatus, that have been added to provide a better covering of some holes due to cables and electronics. This subdetectors are the Keystone and the Miniskirt and have different performances than the rest of the CMX apparatus, so should be threated separately as far as trigger and selection efficiency is concerned (Sec. 3.2.2). Both CMX and CMUP detected muons will be used in this analysis. For the different peculiarities of these detectors, the two muon samples will not be merged till the very end of the analysis.

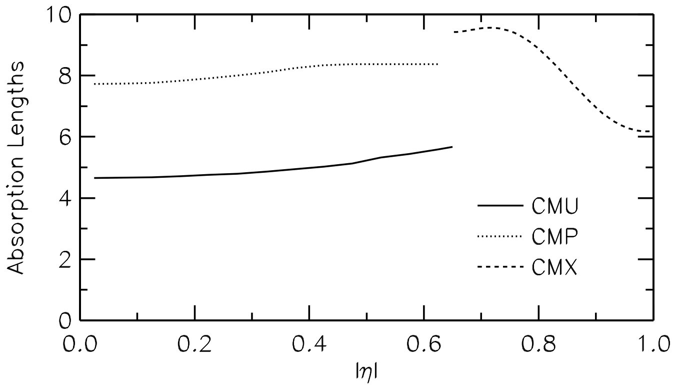


Figure 2.16: Number of absorption lengths as a function of pseudorapidity averaged over azimuthal acceptance of the CMU, CMP and CMX systems.

2.3.3.4 IMU

Muons in a more foreward region, at $1.0 < |\eta| < 1.5$, are detected by the Intermediate Muon Extension (IMU) on the back of the Plug Calorimeters (see Sec. 2.3.2.2).

2.4 Cherenkov Luminosity Counters

In CDF, the luminosity is one of the most important source of systematic indetermination in every measurement. it is inferred using gas Cherenkov counters (CLC) [23] located in the pseudorapidity region $3.7 < |\eta| < 4.7$, which measure the average number of inelastic interaction in a certain period. Each module consists of 48 thin, gas-filled, Cherenkov

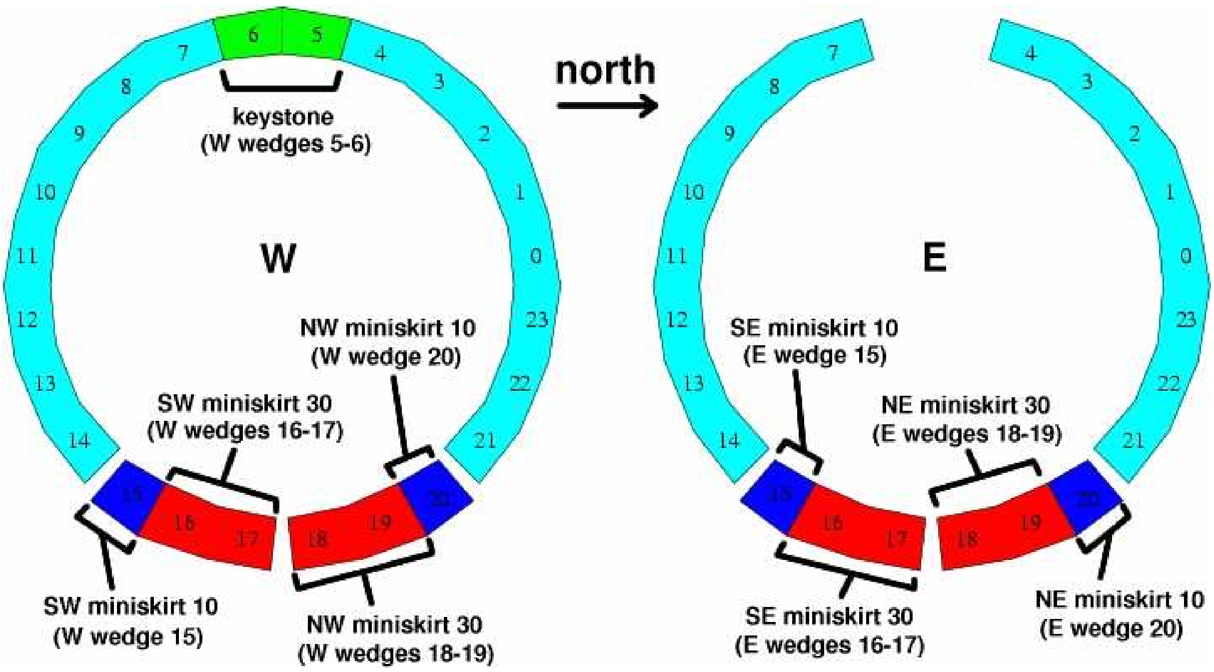


Figure 2.17: CMX subdetector scheme. The *arches*, *miniskirt* and *keystone* zones are shown.

.

counters. The counters are arranged around the beam pipe in three concentric layers, with 16 counters each, and pointing to the center of the interaction region. The cones in the two outer layers are about 180 cm long and the inner layer counters, closer to the beam pipe, have a length of 110 cm. The Cherenkov light is detected with photomultiplier tubes.

2.4.1 Measurement of the luminosity

The average number of primary interactions, μ, is related to the instantaneous luminosity, $\mathcal{L}$, by the expression:

$$\mu \cdot f_{bc} = \sigma_{tot} \cdot \mathcal{L} \tag{2.6}$$

where f_{bc} is the bunch crossings frequency at Tevatron, on average 1.7 MHz for 36×36 bunch operations, and σ_{tot} is the total $p\bar{p}$ cross section.

Since the CLC is not sensitive at all to the elastic component of the $p\bar{p}$ scattering, the equation 2.6 can be rewritten using the inelastic cross section, σ_{in}, as:

$$\mathcal{L} = \frac{\mu \cdot f_{bc}}{\sigma_{in}} \tag{2.7}$$

where now μ is the average number of inelastic $p\bar{p}$ interactions.

Different sources of uncertainties have been taken into account to evaluate the systematic uncertainties on the luminosity measurement [24]. The dominated contributions are related to the detector simulation and the event generator used, and have been evaluated to be about 3%. The total systematic uncertainty in the CLC luminosity measurements is 5.8%, which includes uncertainties on the measurement (4.2%) and on the inelastic cross section value (4%).

2.5 Trigger and Data Acquisition

The average interaction rate at the Tevatron is 1.7 MHz for 36 $\times$ 36 bunches. In fact, the actual interaction rate is higher because the bunches circulate in three trains of 12 bunches in each group spaced 396 ns which leads to a crossing rate of 2.53 MHz. The interaction rate is orders of magnitude higher than the maximum rate that the data acquisition system can handle. Furthermore, the majority of collisions are not of interest. This leads to implementation of a trigger system that preselects events online and decides if the corresponding event information is written to tape or discarded.

The CDF trigger system consists of three trigger levels, see figures 2.18 and 2.19, where the first two levels are hardware based and the third one is a processor farm. The decisions taken by the system are based on increasingly more complex event information. The two hardware levels are monitored and controlled by the Trigger Supervisor Interface (TSI), which distributes signals from the different sections of the trigger and DAQ system, a global clock and bunch crossing signal.

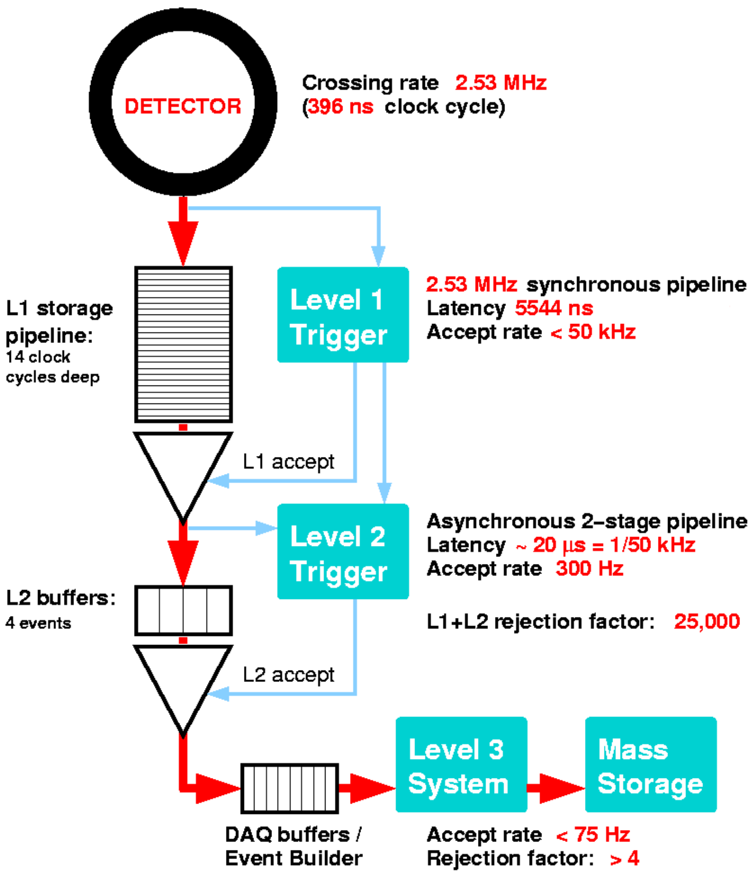


Figure 2.18: Block diagram showing the global trigger and DAQ systems at CDF II.

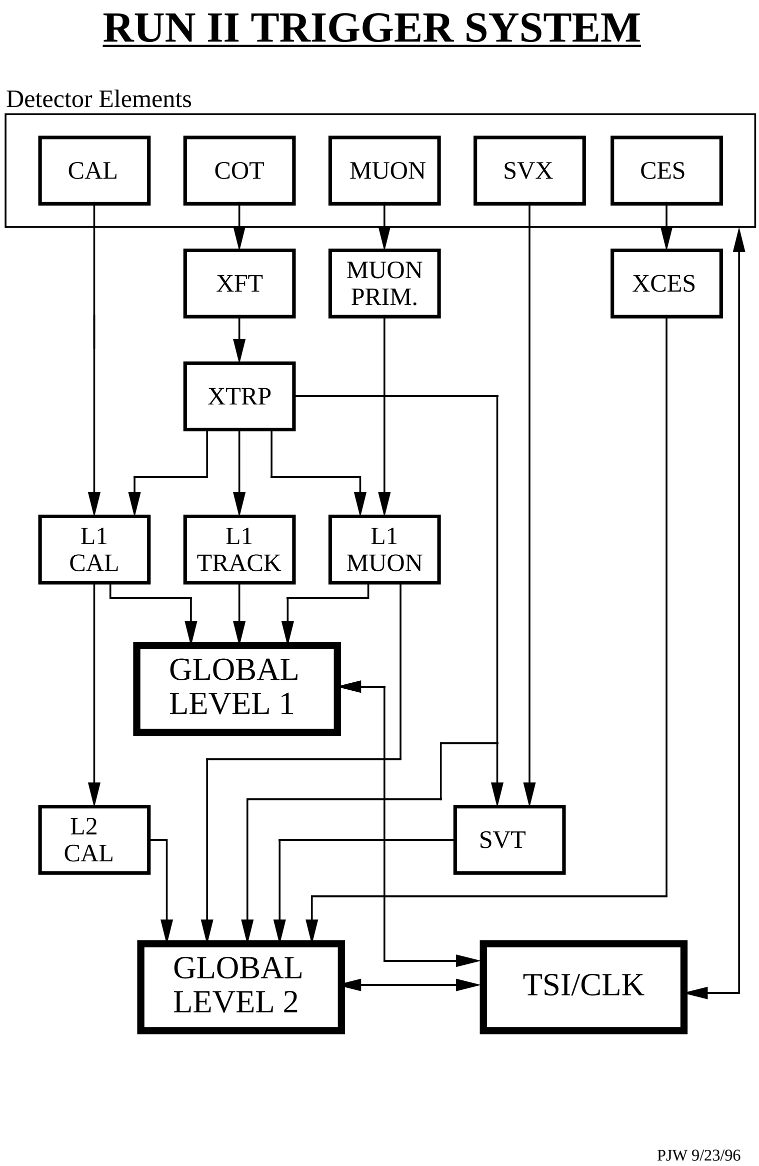


Figure 2.19: Block diagram showing the Level 1 and Level 2 trigger systems.

2.5.1 Level 1 trigger

The Level 1 trigger is a synchronous system with an event read and a decision made every beam crossing. The depth of the L1 decision pipeline is approximately 4 μs (L1 latency). The L1 buffer must be at least as deep as this processing pipeline or the data associated with a particular L1 decision would be lost before the decision is made. The L1 buffer is 14 crossings deep (5544 ns at 396 ns bunch spacing) to provide a margin for unanticipated increases in L1 latency. The Level 1 reduces the event rates from 2.53 MHz to less than 50 kHz.

The Level 1 hardware consists of three parallel processing streams which feed inputs of the Global Level 1 decision unit. One stream finds calorimeter based objects (L1 CAL), another finds muons (L1 MUON), while the third one finds tracks in the COT (L1 TRACK). Since the muons and the calorimeter based objects require the presence of a track pointing at the corresponding outer detector element, the tracks must be sent to the calorimeter and muon streams as well as the track only stream.

- The L1 CAL calorimeter trigger is employed to detect electrons, photons, jets, total transverse energy and missing transverse energy, E_T^{miss}. The calorimeter triggers are divided into two types: object triggers (electron, photons and jets) and global triggers ($\sum E_T$ and E_T^{miss}). The calorimeter towers are summed into trigger towers of 15^o in ϕ and by approximately 0.2 in η. Therefore, the calorimeter is divided in 24 x 24 towers in $\eta \times \phi$ space [25]. The object triggers are formed by applying thresholds to individual calorimeter trigger towers, while thresholds for the global triggers are applied after summing energies from all towers.

- The L1 TRACK trigger is designed to detect tracks on the COT. An eXtremely Fast Tracker (XFT) [26] uses hits from 4 axial layers of the COT to find tracks with a p_T greater than some threshold ($\sim$ 2 GeV/c). The resulting track list is sent to the extrapolation box (XTRP)[27] that distributes the tracks to the Level 1 and Level 2 trigger subsystems.

- L1 MUON system uses muon primitives, generated from various muon detector elements, and XFT tracks extrapolated to the muon chambers by the XTRP to form muon trigger objects. For the scintillators of the muon system, the primitives are derived from single hits or coincidences of hits. In the case of the wire chambers, the primitives are obtained from patterns of hits on projective wire with the requirement that the difference in the arrival times of signals be less than a present threshold. This maximum allowed time difference imposes a minimum p_T requirement for hits from a single tracks.

Finally, the Global Level 1 makes the L1 trigger decision based on the quantity of each trigger object passed to it.

2.5.2 Level 2 trigger

The Level 2 trigger is an asynchronous system which processes events that have received a L1 accept in FIFO (First In, First Out) manner. It is structured as a two stage pipeline with data buffering at the input of each stage. The first stage is based on dedicated hardware processor which assembles information from a particular section of the detector. The second stage consists of a programmable processors operating on lists of objects generated by the first stage. Each of the L2 stages is expected to take approximately 10 μs giving a latency of approximately 20 μs. The L2 buffers provide a storage of four events. After the Level 2, the event rate is reduced to about 300 Hz.

In addition of the trigger primitives generated for L1, data for the L2 come from the shower maximum strip chambers in the central calorimeter and the $r \times \phi$ strips of the SVX II. There are three hardware systems generating primitives at Level 2: Level 2 cluster finder (L2CAL), shower maximum strip chambers in the central calorimeter (XCES) and the Silicon Vertex Tracker (SVT).

- The L2CAL hardware carries out the hardware cluster finder functions. It receives trigger tower energies from the L1 CAL and applies seed and 'shoulder" thresholds for cluster finding. It is basically designed for jet triggers. More details about the cluster finder algorithm in section **??**.

- The shower maximum detector provides a much better spacial resolution than a calorimeter wedge. The XCES boards perform sum of the energy on groups of four adjacent CES wires and compare them to a threshold (around 4 GeV). This information is matched to XFT tracks to generate a Level 2 trigger. This trigger hardware provides a significant reduction in combinatorial background for electrons and photons.

- Silicon Vertex Tracker [28] uses hits from the $r \times \phi$ strips of the SVX II and tracks from the XFT to find tracks in SVX II. SVT improves on the XFT resolution for ϕ and p_T and adds a measurement of the track impact parameter d_0. Hereby the efficiency and resolution are comparable to those of the offline track reconstruction. The SVT enables triggering on displaced tracks, that have a large impact parameter d_0.

2.5.3 Level 3 trigger

When an event is accepted by the Level 2 trigger, its data become available for readout distributed over a couple of hundred of VME Readout Buffers (VRBs). The event has to be assembled from pieces of data from the L2 system into complete events, this is the purpose of the Event Builder. It is divided into 16 sub-farms, each consisting of 12-16 processor nodes. Once the event is built, it is sent to one place in the Level 3 farm. The Level 3 trigger reconstructs the event following given algorithms. These algorithms take advantage of the full detector information and improved resolution not available to the lower trigger levels. This includes a full 3-dimensional track reconstruction and tight matching of tracks to calorimeter and muon-system information. Events that satisfy the Level 3 trigger requirements are then transfered onward to the Consumer Server/Data Logger (CSL) system for storage first on disk and later on tape. The average processing time per event in Level 3 is on the order of one second. The Level 3 leads to a further reduction in the output rate, a roughly 50 Hz.
A set of requirements that an event has to fulfill at Level 1, Level 2 and Level 3 constitutes a trigger path. The CDF II trigger system implements about 150 trigger paths. An event will be accepted if it passes the requirements of any one of these paths and, depending of the trigger path, it will be stored in a trigger dataset. A complete description of the different datasets at CDF Run II can be found in [29].
In addition to impose the trigger requirements to select out interesting physics events, trigger can be prescaled in the different levels. To prescale means to accept only a predetermined fraction of events selected by a given trigger path.

2.5.4 Trigger paths used in the analysis

Since the W decays in a high energetic lepton, we have chosen the high transverse momentum triggers for electrons (CEM) and muons (CMUP and CMX), among the several CDF triggers. The selections applied in each one of the three trigger levels by these triggers are shown in Tables 2.3 and 2.4 for electrons and muons respectively. The electron trigger (at Level 3) requires a calorimetric cluster with $E_T > 18$ GeV, matched to a track with $P_T > 9$ GeV/c. A further condition on hadronic to electromagnetic deposited energy is required: Had/Em < 0.125. Muon trigger paths are more complicated, since some runs have been excluded from analysis because of CMXdetector malfunction. The general requirement is a COT track matched to muon chambers track segment with $P_T > 18$ GeV. Notice, that in the case of muon trigger in the forward region (CMX) for a certain period of data taking period, a special trigger reguiring a muon and an energetic lepton is used that allow us to avoid a prescale factor that was necessary in order to keep that trigger rate at a resaonable level at the highest instantenous luminosity. The additional inefficiency due to the jet requirement has been estimated as negligeable for the kind of jet selection we will apply in the analysis. In the latest period of data takging, instead, improvements to the hardware muon trigger at L1 allowed to use an unprescaled CMX trigger without the jet requirement.

The paths used in this analysis are:

- **CEM**: *ELECTRON_TRIGGER_18*
- **CMUP**:
  - run $\leq$ 229763: *MUON_CMUP_18_V* || *MUON_CMUP_18_L2_PT15V*
  - run $>$ 229763 : *MUON_CMUP18_V*
- **CMX**:
  - run $\leq$ 200272: *MUON_CMX18_V* || *MUON_CMX18_L2_PT15_V*
  - 200272 $<$ run $\leq$ 226194: *MUON_CMX18_L2_PT15_V* || *MUON_CMX18_L2_PT15_LUMI_200_V*
  - 226194 $<$ run $\leq$ 257201: *MUON_CMX18_&_JET10_V* || *MUON_CMX18_&_JET10_LUMI_270_V* || *MUON_CMX18_&_JET10_DPS_V*

Level 1	ELECTRON ET CENTRAL = 8 GeV HAD EM CENTRAL = .125 NUMBER = 1 XFT CHARGE = 0 XFT LAYERS = 4 XFT PT = 8.34 GeV/c
Level 2	EM ET ⩾ 16 GeV HAD EM RATIO ⩽ .125 real NUM ELECTRONS = 1 integer TRACK PT ⩾>= 8 GeV/c DCAS HIGH EM CENTRAL SEED = 8 GeV DCAS HIGH EM CENTRAL SHOULDER = 7.5 GeV DCAS HIGH EM FORWARD SEED = 8 GeV DCAS HIGH EM FORWARD SHOULDER = 7.5 GeV DCAS HIGH EM PLUG SEED = 8 GeV DCAS HIGH EM PLUG SHOULDER = 7.5 GeV
Level 3	run1SpikeKiller CalorimetryModule v1 globalCT HL2 CT TrackingModule v3 DoAxialHistogram = true LinkAxialSegments = false MaxSeedCurvature = 0.008 0.008 0.008 0 0 0 0 0 MergeMethod = HL MinAcceptHits = 20 20 15 48 48 48 48 48 CesThres 0.1 CentralStripClusterModule v1 CprSeedThres200 CprClusterModule v1 EmClustEt2 v1 EmClusterModule v2 PEMMethod = BFPEM clusterEMEtMin = 2 seedEMEtMin = 2 cesLshr CdfEmObjectModule v2 allowSetVertexZ = false allowTrackLshr = false electron18Central v2 L3EMFilterModule v2 CalorRegion = 0 cenEt = 18.0 cenHadEm = 0.125 cenTrackPt = 9.0 nEmObj = 1

Table 2.3: Selection requirements for Electron Central 18 trigger from level 1 to level 3

– run > 257201: *MUON_CMX18_V*

Trigger Level	CMUP	CMX
Level 1	stub CMP min PT $>$ 3 GeV/c	PRESCALE FACTOR = 1 integer
	stub CMP num of layers $>$ 2 integer	
	stub CSP gate width = 9999 ns	
	stub CSP to CMP mtch window = 9999 integer	
	CMU high PT stub threshold = 6 GeV/c	$\geqslant$ 6 GeV/c
	CMU high PT track $\geqslant$ 4 GeV/c	$\geqslant$ 8 GeV/c
Level 2	L2 AUTO DUMMY PARAMETER = 1 integer	
Level 3	run1SpikeKiller CalorimetryModule v1	
	globalCT HL2 CT TrackingModule v3	
	DoAxialHistogram = true	
	LinkAxialSegments = false	
	MaxSeedCurvature = 0.008 0.008 0.008 0 0 0 0 0	
	MergeMethod = HL	
	MinAcceptHits = 20 20 15 48 48 48 48 48	
	cmu default CMU DtoEModule v1	
	cmp default CMP DtoEModule v1	
	cmx default CMX DtoEModule v1	
	stub default MuonStubModule v1	
	maxCSX4 CMX EtoSModule v1	
	link default MuonLinkerModule v1	
	cmpDx = 10	cmxDx = 10
	cmuDx = 10	
	minPt = 18.0	minPt = 18.0
	nMuon = 1	nMuon = 1
	selectCMUP = true	selectCMX = true

Table 2.4: Selection requirements for Muon Central 18 trigger from level 1 to level 3

Chapter **3**

High tranverse momentum physics and analysis tools

The particles generated in proton-antiproton collisions are studied by the signals detected in the subdetectors that compose the CDF II experiment. Each of these particles produce a "physical object" of which we measure the properties (such as direction, quadrimomentum) to infer the ones of the particle linked to it. In this way neutrinos are detected as missing momentum in the transverse plane, electrons are calorimetric deposit matched to a track, muons are hits in the muon chambers and quarks are collimated bunches of particles (i.e. jets). In this section we are going to give all the relevant information for the particle recontruction and the corrections that needs to be applied in order to take into account trigger efficencies, different performances between the real detector and its simulation, jet energy and missing energy corrections and so on, trying to give a general view of the event recontruction methods.

The interesting events in hadronic colliders are the ones in which partons inside the protons interact. These events are characterised by a high momentum component on the transverse plane, unlike the more common scattering events that involve a small tranfer of momentum. In CDF jargon we call this class of events "high P_T physics" to distinguish from both soft or diffractive physics (minimum bias) and Heavy Flavour Physics, for both of which specific (and different from the high-P_T ones) analysis and reconstruction tools have been developed in the past. Usually the events are energy-balanced on the transverse plane, but not along the beam direction, due to the unknown component along the beam axis of the center of mass frame of the hard interaction. However, events that involve a neutrino, that escapes out without interacting in the detector, have a large amount of unbalancement in the transverse plane (referred as *missing energy*), that is proportional to the neutrino energy. It's straightforward to understand that any mis-measurement of the particles (and in particular jets) energies would turn out in an unbalancement in the transverse energy, mimicking a neutrino.

However, no one of the particles produced in the interaction or other particles decays is directly seen. Each particle is recontructed according to signals in the subdetectors. Hence an electron is a calorimetric deposit matched to a track and a muon is an hit in the muon chambers. We are going to give a description of each physical object we need to reconstruct and recognise for the purpose of this analysis. These physical objects are electrons, muons, jets and *missing energy*.

3.1 Electrons definition

The electron is substantially characterised by an electromagnetic deposit in the calorimeter and a matched track in the tracker (COT and silicon). The first reconstruction step for electrons is already made at Level 2 (Sec. 2.5.2) of the CDF trigger where the electromagnetic clustering is performed starting from the most energetic tower of the electromagnetic calorimeter (seed tower) and including the adjoining towers above a certain energy threshold. At Level 3 of the trigger and offline, there is a more sophisticated reconstruction of the clusters, where the energy threshold is lowered and the energy loss in the hadronic calorimeter is compared to the electromagnetic one, in order to distinguish hadrons that start the shower in the electromagnetic calorimeter from electrons (and photons) that are completely absorbed in it. The second requirement for an electron is that the calorimetric deposit is matched to a track in the inner tracker. Each track is extrapolated to the plane of the central shower maximum subdetector (CES, Sec. 2.3.2) plane position, assuming an elicoidal trajectory, and the track with the highest P_T within a certain distance to the center of the principal tower of the cluster is chosen.

The electrons selected for this analysis are the ones that are identified in the central region of the detector. It is then required that this *central* electrons have a minimum transverse energy of 20 GeV, a smaller minimum transverse momentum of 10 GeV, to account for bremsstrahlung radiation, and they have to belong to the fiducial region of the CES. This means they are within 21 cm from the central tower of the cluster on the $r - \phi$ plane and between 9 and 230 cm along the z axis. The other requirements that define a *"central tight electron (CEM)"* are summarized in Tables 3.1 and 3.2 and involves other geometrical cuts and calorimetric and tracker requirements.

It is worthwhile a closer examination of the isolation requirement imposed on a lepton. This variable is a very important discriminant to distinguish leptons, i.e. electrons and muons, generated within a jet, for example due to a leptonic decay of a B or D meson, and the leptons that are produced in the primary interaction, as the decay products of gauge bosons we are looking for. In fact, the isolation, defined in Eq (3.1), permits to measure the calorimetric/tracker activity around the candidate, comparing the candidate energy with the sum of the energies within a certain angular distance from the candidate electron.

$$Isol = \frac{1}{E_T}\left(\sum_{R<0.4} E_T^i - E_T^e\right) \tag{3.1}$$

where E_T^i is the transverse energy of the i^{th} tower, E_T^e is the transverse energy deposited in the tower crossed by the track and E_T is the electron transverse energy. The sum is performed over all the towers inside a cone with a radius $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$ around the track direction, as shown in figure 3.1.

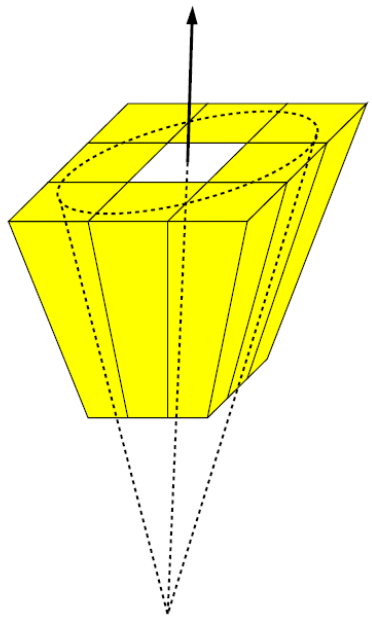


Figure 3.1: Calorimetric isolation of a candidate electron, evaluated considering the energy deposits of the towers included in a $R = 0.4$ cone.

As a last consideration, we would like to highlight that all the electron requirements could be divided into two cathegories: *kinematics* and *identification* requirements, summarized in Tabs 3.1 and 3.2 respectively. This division will be used in section 6.1 for the definition of a background enriched electron candidate sample.

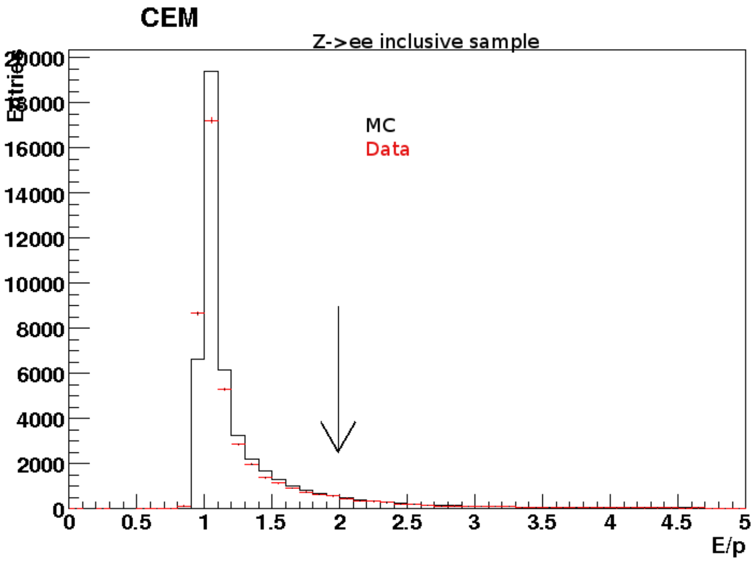


Figure 3.2: Small shift of electron energy in data respect to the MC reconstruction

3.1.1 Correction to the electron energy

The energy of the electrons, after all calibrations, show a small shift respect to the MC reconstruction. This shift is visible in the electron energy distribution of Figure 3.2 and is also in the Z mass peak reconstruction by electron pairs. Some studies have been performed in [30] to calculate the scale factor that permits to match the Z mass peak to the expected value of 91 GeV/c^2, for both MC and data. The resultant scale factors are $Sf = 1.005$ for electrons in data and $Sf = 0.995$ in MC and are applied as in Eq. (3.2) to obtain the electron corrected energy (E_T^{corr}) from the measured one (E_T^{meas})

$$E_T^{corr} = Sf \cdot E_T^{meas} \tag{3.2}$$

3.1.2 Electron trigger and selection efficiencies

The estimation of the trigger efficiency is very important for all the analysis in which a number of measured events is used to perform a cross section measurement, since the trigger is not simulated in the Monte Carlo samples. For the same reason, we need to correct any differences in the selection efficiency between data and MC simulation, to be able to derive from MC the effects of the selection cuts on the data sample.

Variable	**Description**	**Cut Value**
kinematics requirements		
Region	part of the detector that detects the electron	central
Track	a matched track in the inner tracker	yes
Iso/E_T^e	Isolation to transverse energy ratio, measures the activity around the candidate electron $$Iso = \sum_{within\ \Delta R<0.4} E_T \ - E_T^e$$	≤ 0.1
E_T	Transverse energy of the cluster	> 20 GeV
P_T	Transverse momentum of the electron track	> 10 GeV/c
Track $\lvert Z_0 \rvert$	Position along the z axis of the point of closest approach of the track to the beam line	≤ 60 cm
E/P	Comparison between the energy of the electromagnetic cluster and the track momentum, to assure that a right match was done. This cut is not required if the $P_t \geq 50$ GeV/c since the momentum resolution above that threshold is too low in the COT	≤ 2
Fiducial	The electron is recontructed in a region within 21 cm from the central tower of the cluster on the $r-\phi$ plane and between 9 and 230 cm along the z axis of the CES	yes

Table 3.1: kinematic requirements for the *tight* electrons selected for this analysis

3.1.2.1 Trigger efficiency

The standard method adopted by CDF to measure the trigger efficiency exploits an *unbiased* data sample, acquired with an independent trigger. The trigger *ELECTRON_CENTRAL18*, used for the present analysis, exploits both tracking and calorimetric information, and the corresponding contributions to the trigger efficiency can be evaluated separately.

The tracking efficiency can be evaluated in a data sample acquired with a trigger path which implements the same calorimeter requests of the *ELECTRON_CENTRAL18*, and has no requests on tracking quantities.

The calorimetric efficiency can be evaluated in the *tight* electrons sample acquired in an independent trigger. Due to the structure of the tower clustering algorithm implemented in the level 2 of this trigger path, the calorimetric efficiency is a function of the electron transverse energy and it has been evaluated for each period of data.

The average trigger efficiency for all the periods used in this analysis is approximately 88% for an electron of 25 GeV of energy.

Variable	Description	Cut Value
identification requirements		
Had/Em	Electromagnetic to hadronic energy ratio, in order to distiguish electrons, that loss all their energy into the electromagnetic calorimeter, from jets, that have a considerable amount of losses in the hadronic one. The cut depends from the cluster energy to reduce the correlation between the cut efficiency and the energy itself	$\leq 0.055 + 0.00043 \cdot E_{em}(\text{ GeV})$
Signed CES ΔX	Particle charge per distance between calorimeter centroid and track, in order to make use of the CES good resolution to verify the matching between the cluster and the track in the $r-\phi$ plane. The cut is asymmetric to take into account brem radiation and multiplied by charge to account the fact that positrons and electrons are deflected in opposite directions	$-3.0 \leq q\Delta X \leq 1.5$
CES ΔZ	Distance between calorimeter centroid and track in the $r-z$ plane, in order to make use of the CES good resolution to verify the matching between the cluster and the track	< 3 cm
Lshr	It provides a comparison between the electromagnetic cluster shape of the candidate electron and the test-beam one	< 0.2
CES χ^2_{strip}	χ^2 of a fit on the 11 strips of a CES cluster, considering the total energy of the cluster	$\leq$10

Table 3.2: identification requirements for the *tight* electrons selected for this analysis

3.1.2.2 Scale factor on selection efficiency

The procedure used by CDF collaboration to evaluate the selection efficiency of the different electron categories is based on a very pure $Z \rightarrow e^+e^-$ sample. These events are identified through the reconstruction of a pair of candidate electrons with invariant mass in a narrow window around the Z mass ($76 - 106$ GeV/c^2), with the first one satisfying very tight identification cuts. The second electron is then exploited for the evaluation of the selection efficiencies for the different set of cuts.

Any disagreement between data and Monte Carlo simulation can reflect in a different value of the selection efficiencies. The standard correction procedure adopted by the CDF collaboration relies on the evaluation of *scale factors* to reconcile the selection efficiencies measured in MC simulation with the ones measured in data samples:

$$SF = \frac{\epsilon_{sel}^{Data}}{\epsilon_{sel}^{MC}} \tag{3.3}$$

This *scale factors* are evaluated for non-overlapping categories, therefore the *tight* electrons are removed from the *loose* sample and are calculated for each period.

Only the scale factor is needed for this analysis, since the selection efficiency is obtained appling the analysis cuts on MC samples after correcting for the scale factor to make reliable the MC estimate of selection efficiency. The average scale factor on selection efficiency that is applied for this analysis is 98%.

3.2 Muons definition

The identification of a muon candidate is based on the reconstruction of a track with an associate calorimetric deposit compatible with a *minimum ionizing particle*. Further information can be added by the matching of the track with the track segment reconstructed by the CMUP ($|\eta| < 0.6$) or the CMX ($0.6 < |\eta| < 1$) detector. This two detectors have different performances, in particular the two detectors have different amounts of material in front of them, so this two samples have been analyzed separately before merging. The reconstructed quantities, and the relative cuts, used to select the muons are explained in Table 3.3. They are essentially composed of a minimum transverse momentum requirement $P_T > 20$ GeV, some quality requirements on the matched track, and the isolation of the candidate, that increases the separation between isolated muons coming from vector bosons decay and muons produced in a semileptonic decay of a hadron. All the consideration of electrons' isolation, in the previous section, stands for muons too, with the obvious change $E_T \rightarrow P_T$.

3.2.1 Correction to the muon momentum

The Monte Carlo simulation of the the momentum reconstruction of the muon does not fully represent the muons momentum in data, predicting a momentum resolution better than the one really obtained with data. For this reason we need to apply a smearing of the muon momentum in MC events, multiplicating each component of the momentum to

Variable	Description	Cut Value
both CMUP *and* CMX *muons*		
P_T	Transverse momentum of the electron	> 20 GeV/c
E_{em}	Energy deposited in the electromagnetic calorimeter, it's a function of the muon momentum	$< 2 + \max\big(0, (P_T - 100)0.0115\big)$ GeV
E_{had}	Energy deposited in the hadronic calorimeter, it's a function of the muon momentum	$< 6 + \max\big(0, (P_T - 100)0.028\big)$ GeV
Track $\|Z_0\|$	Position along the z axis of the point of closest approach of the track to the beam line	≤ 60 cm
$\mathcal{N}$ COT hits	Number of hits in the COT detector	> 0
Iso$/P_T^\mu$	Isolation to transverse energy ratio, measures the activity around the candidate electron $$Iso = \sum_{within\ \Delta R<0.4} E_T \ - E_T^\mu$$	≤ 0.1
Tracksislhits d_0	The impact parameter of the track in the transverse plane $(r - \phi)$, corrected after the offline reconstruction of the beam line position, in case the track has silicon hits	0.02
Tracknoslhits d_0	Impact parameter for tracks without silicon hits attached	0.2
$TrkAxSeg$	The number of axial super-layers with at least 5 hits	> 2
$TrkStSeg$	The number of stereo super-layers, with at least 5 hits	> 2
only CMUP *muons*		
Δx_{CMP}	Separation between the track segment in the CMU detector and the track extrapolated to the detector plane	< 5 cm
Δx_{CMU}	Separation between the track segment in the CMP detector and the track extrapolated to the detector plane	< 7 cm
only CMX *muons*		
Δx_{CMX}	Separation between the track segment in the CMX detector and the track extrapolated to the detector plane	< 6 cm
ρ_{COT}	The distance from the beam line at which the track crosses one of the endcap planes of the COT	> 140 cm

Table 3.3: Summary of the cuts used to select muon candidates with a *stub* in the CMUP or CMX sub-detectors

a number that is randomly generated according to a gaussian distribution of mean 1 and width 0.024. In other words, being g a randomly generated number from the smaring gaussian distribution, the muon corrected momentum P_T^{corr} is defined as

$$P_T^{corr} = g \cdot P_T^{raw} \tag{3.4}$$

3.2.2 Muon trigger and selection efficiencies

3.2.2.1 trigger efficiency

The procedure used by CDF collaboration to evaluate the trigger efficiency of the high-P_T muon triggers is based on a very pure $Z \rightarrow \mu^+\mu^-$ sample. These events are identified through the reconstruction of a pair of identified CMUP or CMX muons with invariant mass in a narrow window around the Z mass ($76 - 106$ GeV/c^2), and with $|z_0^{(1)} - z_0^{(2)}| < 4cm$. Furthermore at least one muon must satisfy the trigger requests. The other muon is then exploited for the evaluation of the trigger efficiencies for the high-P_T trigger paths.

In the case of muons, average trigger efficiencies have been used, for the small variation of this value according to data periods. It value is 89% for CMUP and assumes two different values for CMX in case the muon has been detected by Miniskirt or Keystone subdetectors (Sec. 2.3.3.3) or not, which are 87% and 93% respectively.

3.2.2.2 Scale factor on selection efficiency

As in the case of electrons, selection efficiencies are evaluated

on pure $Z \rightarrow \mu^+\mu^-$ sample. The events are identified through the reconstruction of a pair of identified muons with invariant mass in a narrow window around the Z mass ($76 - 106$ GeV), and with $|z_0^{(1)} - z_0^{(2)}| < 4cm$. The first muon must satisfy the CMUP or CMX requests reported in Table 3.3 and the other muon is exploited for the evaluation of the selection efficiencies of the different set of selection cuts.

Then a scale factor is calculate to take into account any possible disagreement between data and Monte Carlo simulation that can reflects in a different value of the selection efficiencies. The standard correction procedure adopted by the CDF collaboration relies on the evaluation of *scale factors* to reconcile the selection efficiencies measured in MC simulation with the ones measured in data samples:

$$SF = \frac{\epsilon_{sel}^{Data}}{\epsilon_{sel}^{MC}} \tag{3.5}$$

The average scale factor on selection efficiency that is applied for this analysis is 92% for CMUP. As far as CMX is concerned, two different corrective factors have to be taken into account, the first one is for muons detected by Miniskirt or Keystone subdetectors (Sec. 2.3.3.3), and is 98%, the second one is for the rest of the CMX subdetector, and values 88%.

3.3 Jet definition

The color confinement property of QCD processes leads to a potential between a $q\bar{q}$ pair, that increase with separation leading the production of more $q\bar{q}$ pair to be a more energetically favoured condition. In this way quarks produced in a hard scattering interaction will generate, in the hadronization process, a bunch of collimated hadrons with null color charge approximately in the direction of the original parton. This *jets* are the physical objects that we can measure and we have to deal with, in order to infer informations about the quark that have originated it. In fact, *bare* quarks have never been detected.

An approximate representation of the steps of a jet production is given in Figure 3.3. The parton generated in the interaction go through the hadronization process, generating a bunch of particles, the jet, that interacts in the detector. Jets must be defined by clustering algorithms, and the algorithms are designed such that the jets clustered from the complex structure of objects in each event accurately represent the physical properties of the partons originated from the hard scattering. We starts from the definition of a *jet reconstruction algorithm*, which is a “recipe” to selects particles or whatever has a quadrimomentum, either in the calorimeter or in the tracker to belong to a jet. There are basically two kind of jet reconstruction algorithm: the ones with seed and the ones without seed; to the first cathegory belongs the Cone Algorithm that we are going to use in this analysis. The next step to bring back to the original parton physical quantities is the Jet Energy correction, which consist of applying some corrections to the energy associated with the jet, in order to bring us back to the energy, and the direction, of the parton that has originated the jet.

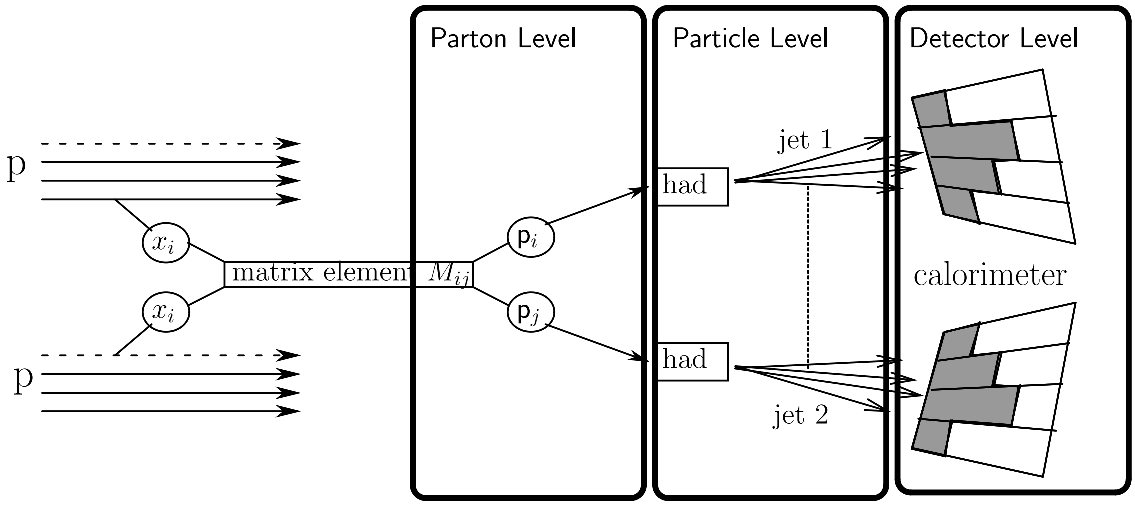


Figure 3.3: The production of quarks in an hard scattering interaction and their detection as jets

The CDF jet energy corrections [31] are applied in five consecutive steps, referred as "correction levels". Levels 2 and 3 do not exist any more for historical reasons. These corrections have been defined in order to accomodate different effects that can distort the measured jet energy, such as, response of the calorimeter to different particles, non-linearity response of the calorimeter to the particle energies, un-instrumented regions of the detector, spectator interactions, and energy radiated outside the jet clustering algorithm. In the following, it's reported a brief description of each level correction. From Level 5 to Level 7 the energy is referred as *absolute*, since all the detector dependencies are corrected and the measured energy can be compared to other experiments.

Level 0 Calibration: *it sets the calorimeter energy scale*

Level 1 Pseudo-rapidity dependence: *it is applied to raw jet energies measured in the calorimeter to make jet energy uniform along η. It gives the "Detector Level energy".*

Level 4 Multiple Interactions: *it corrects for the energy that falls inside the jet cone due to different $p\bar{p}$ interactions during the same bunch crossing. This correction substracts this contribution in average and is derived from minimum bias data and it is parameterized as a function of the number of interaction vertices in the event.*

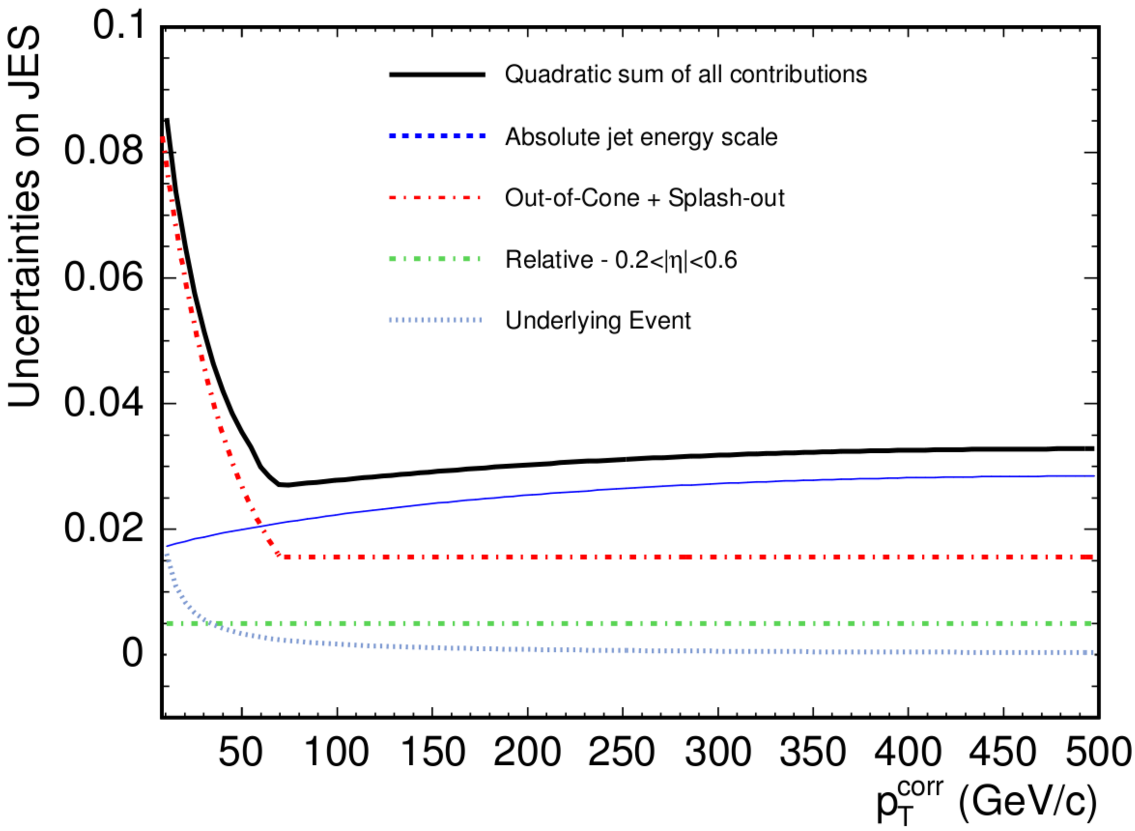


Figure 3.4: Systematic uncertainties for the Jet Energy Scale correction energy [31]

Level 5 Absolute: *it corrects the jet energy measured in the calorimeter for any non-linearity and energy loss in the un-instrumented regions of each calorimeter. The jet energy measured is corrected to the* P_T *sum of the particles within the cone around the parton direction which matched the jet direction with* $\Delta R < 0.4$ *and so is referred as "Particle Level energy".*

Level 6 Underlying Event: *subtract to the particle level jet the energy associated with the particles produced by the spectator partons in a hard collision event.*

Level 7 Out-of-Cone: *it corrects the particle-level energy for leakage of radiation outside the clustering cone used for jet definition, taking the jet energy back to "Parent Parton energy".*

The JES is one of the most important source of systematic uncertanty. A summary of the contribution of each correction level is shown in Figure 3.4.

3.3.1 The Cone Algorithm

The algorithm of jet reconstruction used for this analysis is the Cone Algorithm with $R = 0.4$. According to this reconstruction algorithm the jets are cones of fixed radius R in (η, ϕ) space:

$$R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$$

and the cone formation proceed through seeds, that are calorimeter towers above a given energy threshold. This characteristic has the advantage of speeding up the computation for jets reconstruction, although causes the algorithm to not fulfill all the theoretical requirements that guarantee a well-behavedness of the algorithm itself. However, this is important only for very low P_T jets and the algorithm is suitable for the reconstruction of jets of the energy of the “high-P_T physics” analyses. The algorithm can be summarized in the following steps, in its application in the calorimeter:

I. all the towers have to be sorted according to their energy;

II. the most energetic tower is used as a seed for the algorithm;

III. all the towers within a cone of ray R centered in the seed are selected;

IV. the E_T weighted centroid of the cone, in (η, ϕ) space, is then calculated, using the (η, ϕ) coordinates of each tower.

V. this centroid is used as a seed to reiterate the procedure until a convergence is obtained;

VI. after a cluster is defined as a jet, the procedure starts again considering the next seed, in order of energy, that is not yet associated to a jet.

3.4 B-tagging

Among the jets, a particular attention should be given to the ones originated by heavy flavour quarks, and in particular b-quarks. Their importance lies in the fact that b-jets are a fertile ground to investigate both low and high transvers momentum physical issues; to the former belong the investigations of the flavour sector of the Standard Model, among the latter, we have the measurement of the top quark properties, tests on QCD and, last but not least, they have a prime role in the search for Standard Model Higgs, whose decay channel, in the case of a sufficently light Higgs, is predicted to be in a b-quark pair.

B-hadrons, that are the color-neutral bound states of a b valence quark and one or two lighter valence quark (such as c, s, d, or u), are produced in a variety of processes in $p\bar{p}$ interactions, and their total cross section is approximately $\sim 100\ \mu\mathsf{b}$ at Tevatron.

The peculiarity that makes possible to distiguish the B-jets, i.e. the jets that have been originated or at least contain a B hadron inside, is the long lifetime (~ 1.5 s) of these hadrons, the large mass (~ 5 GeV) and their high brancing ratio for semileptonic decays ($\sim 20\%$). Another important information is that most of the not-semileptonic decays of the b quarks involve a charm, because of the Cabibbo-suppression of the other quark decays, according to the pattern $b \rightarrow c \rightarrow$LF.

For the b-jets identification, *B-tagger* algorithms have been developed at CDF, in order to exploit the B hadrons features to identify high-P_T jets originated by b-quarks. If a jet is recognised as containig a B hadron it is said that the jet is TAGGED. Three main taggers have been developed at CDF for the b-jets identification:

* SecVtx algorithm [32], [33]

* JetProbability algorithm [34], [35]

* RomaTagger neural network [36]

The **SecVtx** algorithm relies on the recontruction of a secondary vertex in the jet cone, due to the decay of the B hadron after ~ 1.5 ps. Two versions of this algorithm are in use, one optimized for higher efficiency (LOOSE), the other optimized for higher purity (TIGHT). The **JetProbability** algorithm is based on the possibility of assigning to each track a "probability" of coming from the primary vertex based on its impact parameter signed with respect to the jet axis. Combining the probability for the well-identified tracks in a jet it is possible to evaluate a probability for the jet itself ("JetProbability") to be composed by particles consistent with coming from the primary vertex. The distribution of this probability is, by construction, flat for jets originated by light quarks, and is peaked at small values for b and c jets. b-jets are tipically tagged by requiring that the value of the JetProbability output is less than a given threshold (values typically

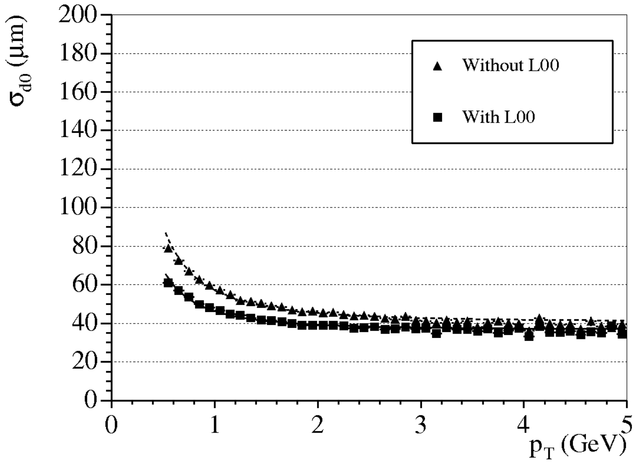


Figure 3.5: Resolution on the primary vertex recontruction, with and without taking into account the information from the most inner layer of the silicon detector, the Layer 00 [37]

used are 0.01 and 0.05). The **Roma-Tagger** is a neural network developed to exploit as much information as possible to identify the flavour of the quarks originating a jet. The vertexing algorithm is able to reconstruct not only one but several vertexes inside the jet cone, to take advantage of the frequent decay patterns with secondary and tertiary vertexes, while a chain of Neural Networks makes possible to combine all the available information in a single discriminant.

The only *B-tagger* used in this analysis is the SexVtx algorithm, for which is due a more detailed explanation.

3.4.1 SecVtx tagging algorithm

Thanks to the CDF traking system, it is possible to have a very high resolution on the tracks impact parameter, d_0, which is the distance between the point of closest approach of the track to the Primary Vertex (PV) and the primary vertex itself, as shown in Fig. 3.5.

This permits to recontruct the primary vertex of the interaction with a resolution that ranges from 10 to 30 μm in the plane transverse to the beam direction and the secondary vertices with a resolution of order 30 μm, depending mostly on the number of tracks used in the recontruction.

This high resolution permits to improve the resolution on the decay lenght measurement and exploit the long lifetime of B hadrons in the SecVtx identification algorithm. In fact,

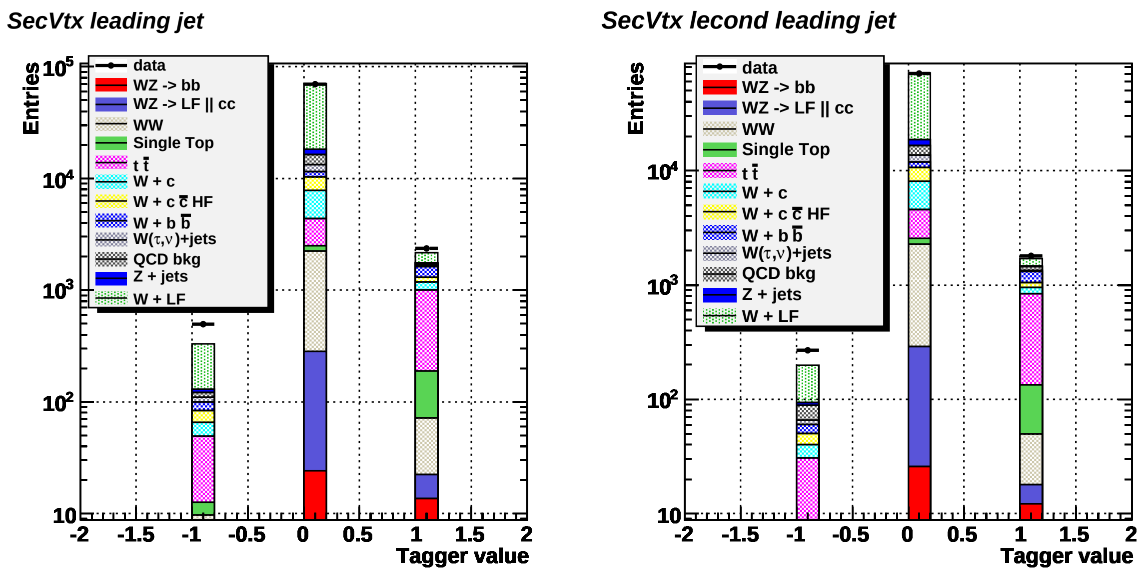


Figure 3.6: Distribution of the SecVtx TIGHT tagger value for the two leading jets of the event. A null value is UNTAGGED, a positive value is a TIGHT TAG and a negative value is a MIS-TAG. Electrons and muons are combined in this plot.

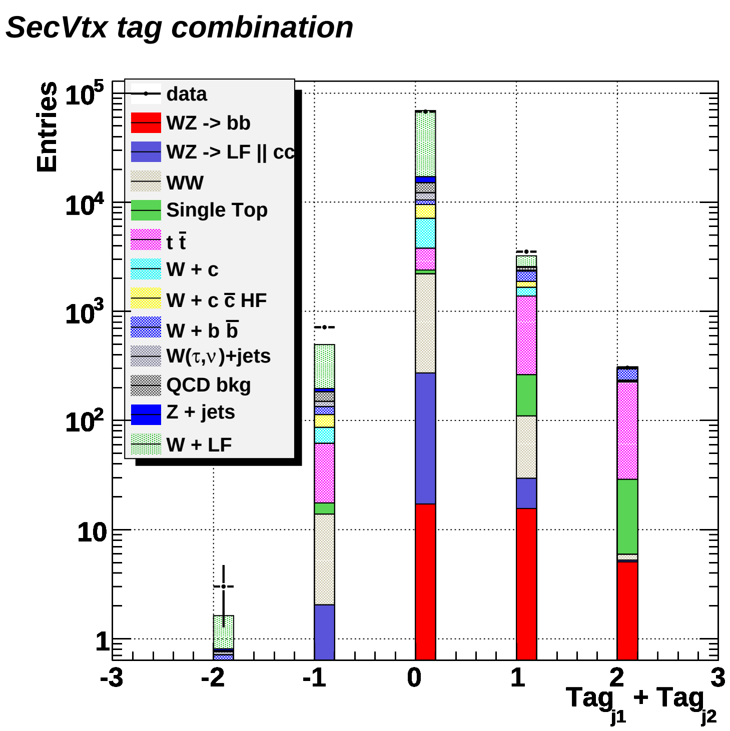


Figure 3.7: Distribution of the sum of the SecVtx TIGHT tagger value of the two leading jets of the event. A value of 2 means a double tag in the event (2TT), a value of 1 means a single tag (1T). Null and negative values represent UNTAGGED and MISTAGGED events respectively.

the B lifetime is approximately 1.5 ps, which corresponds to a lenght of $c\tau = 450$ μm; furthermore, folding in the momentum spectrum of the Bs, the mean decay lenght is on the order of a few mm, so most of the times, the B decay appears as a secondary vertex, i.e. a vertex displaced from the primary one. SecVtx uses tracks with $P_T > 1$ GeV and an impact parameter not compatible with zero as seeds to reconstruct secondary vertices. If any have been found, it uses the χ^2 of the recontruction and its significance ($\frac{L_{xy}}{\sigma_L}$) to decide if the jet is tagged, i.e. assigning a positive, unitary value on a specific variable. As an example, the SecVtx tag variable distribution is shown in Figures 3.6 and 3.7 for the *pretag* sample, that is the sample of events with 2 JETS that pass all the analysis cuts without the b-tagging requirement. The SecVtx LOOSE and TIGHT algorithms efficiencies are shown in Figure 3.9.

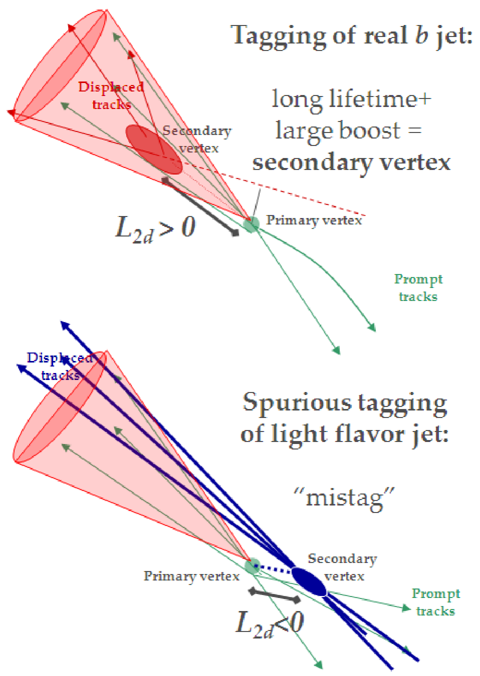
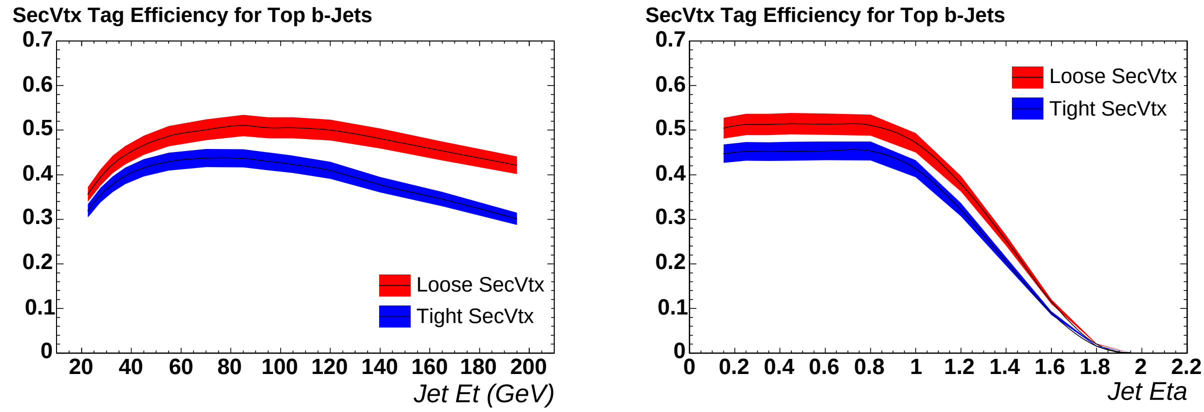


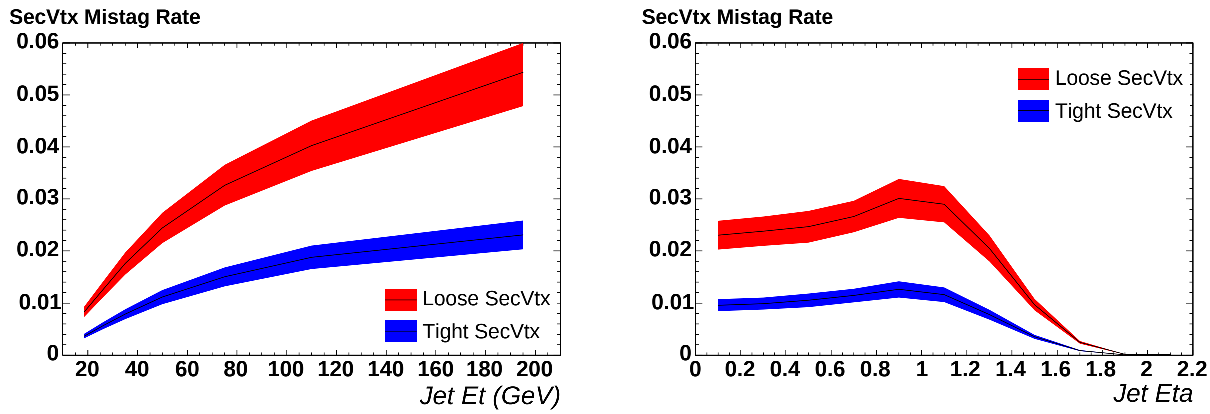
Figure 3.8: Schematic representation of the precence of a displaced vertex in the direction of the jet or in the opposite direction, due to a B-jet or a mistag respectively

In the *pretag* sample most of the jets are originated from light quarks and a considerable amount of them are MIS-TAGGED, that is SecVtx algorithm finds a significantly displaced vertex though no b quark were inside. The *mistag* rate and can be at first order estimated using the negative value of the SecVtx tagging variable. A *negative tag* is defined when the identified secondary vertex is well separated from the primary one, but lies on the "wrong" side of the primary vertex with respect to the jet direction. As schematically represented in Figure 3.8, in a secondary vertex produced by a B decay, the jet direction $\hat{j}$ and the flight part of the decaying particle $\vec{d}$ have the same direction, so that $\vec{t} = \hat{j} \cdot \vec{d} > 0$. In a sample of jets with no lifetime, the distribution of $\vec{t}$ is approximately symmetric around 0, and an apparent flight path is equally likely to appear as a positive or negative lifetime. Hence, the position distribution of the displaced vertex is approximately simmetric around the primary vertex and negative tags can be used to give an estimation of the positive *mis-tag* rate for LIGHT-FLAVOUR jets.



(a) (b)

Figure 3.9: The SecVtx TIGHT and LOOSE algorithms tag efficiencies estimated from Monte Carlo simulation



(a) (b)

Figure 3.10: The false positive tag rate (including asymmetry corrections) at CDF with the SecVtx algorithm. These are have been measured from inclusive jet data.

The mistag rate as a function of the jet energy and pseudo-rapidity measured in an inclusive jet sample is shown in Figure 3.10. This technique has been made more sophisticated in the development of the *Mistag Matrix* technique, that is detailing described in Section 6.2.2.

3.5 Neutrinos and *missing energy*

The *missing energy* of the event is defined by performing a vectorial sum of the calorimeter measured transverse energy, using the event primary vertex position for the calculation

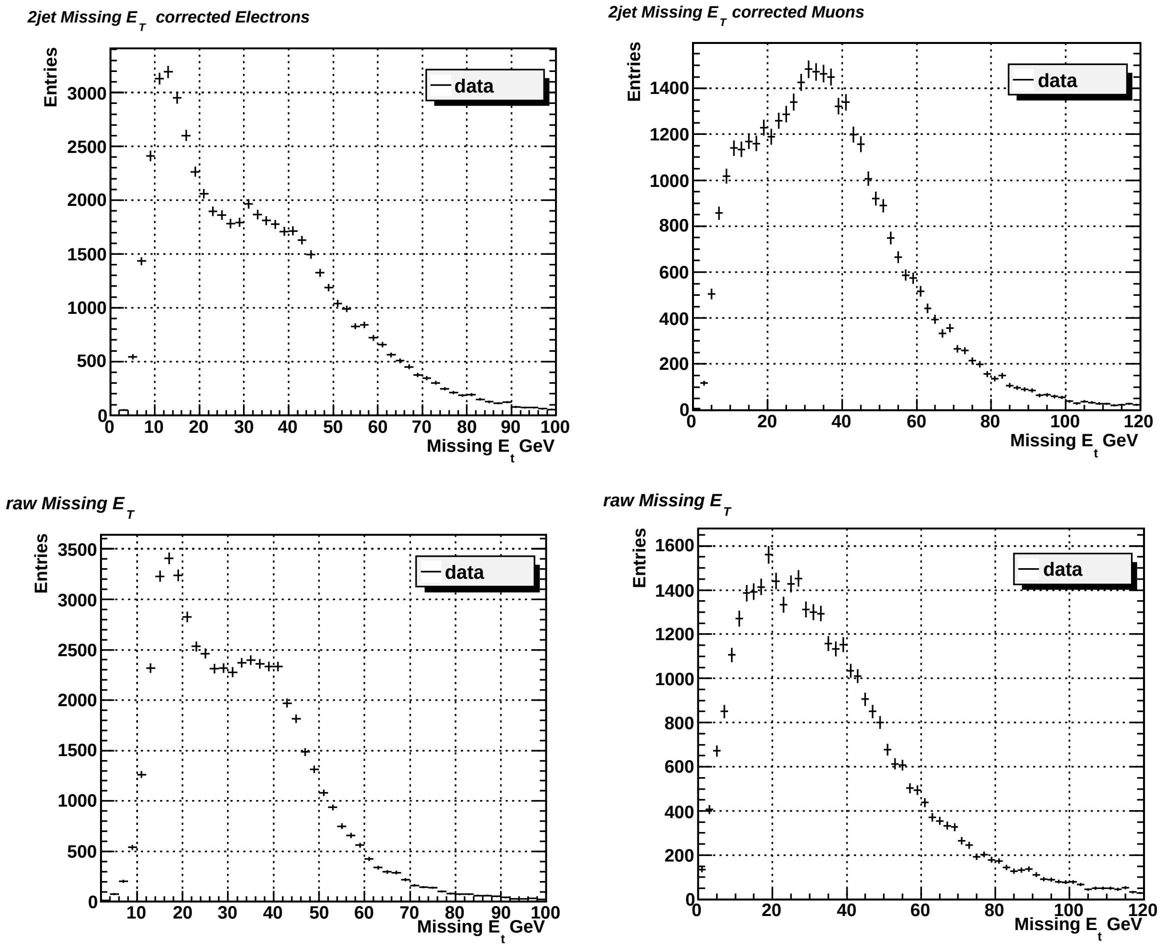


Figure 3.11: Comparison between the raw *missing energy* and the corrected one, for 2 JETS events; both electrons and muons are shown

of the calorimetric tower directions, on a tower by tower basis and is the characteristic signature of a neutrino, that escapes from the detector without interacting. It needs to be corrected with two major effects, following the CDF standard prescription ([38]). Muons, that have only ionization energy losses in the calorimeter, and jets, whose raw measured energy within the jet-cone is systematically shifted from the hadron's one.

For muons, we can naively say that we need to add the momentum of each muon that pass certain standard cuts ([38]) to the $\not{E}_T$, after subtracting the energy that the muon has deposited into the calorimeter, according to Equations (3.6).

$$\not{E}_T = \sqrt{(\not{E}_x^{raw} - \Delta\not{E}_x^{\mu})^2 + (\not{E}_y^{raw} - \Delta\not{E}_y^{\mu})^2} \tag{3.6a}$$

$$\Delta \not{E}_x^{\mu} = P_x^{\mu}\left(1 - \frac{\mu_{CAL}}{|P^{\mu}|}\right) \qquad \Delta \not{E}_x^{\mu} = P_x^{\mu}\left(1 - \frac{\mu_{CAL}}{|P^{\mu}|}\right) \tag{3.6b}$$

where $\not{E}_x^{raw}$ is the x-component of the raw *missing energy*, μ_{CAL} is the calorimeter energy deposited by the muon and P_x^{μ} is its momentum x-component. As far as jets is concerned, the difference between the corrected and the raw jet energy is subtracted to the $\not{E}_T^{raw}$, according to Equations (3.7), for each jet, reconstructed with a cone algorithm, with hadron level corrected energy greater than 15 GeV (see Sec. 3.3).

$$\not{E}_T = \sqrt{(\not{E}_x^{raw} - \Delta\not{E}_x^{jet})^2 + (\not{E}_y^{raw} - \Delta\not{E}_y^{jet})^2} \tag{3.7a}$$

$$\Delta\not{E}_x^{jet} = E_x^{corr} - E_x^{raw} \qquad \Delta\not{E}_y^{jet} = E_y^{corr} - E_y^{raw} \tag{3.7b}$$

where $\not{E}_x^{raw}$ is the x-component of the raw *missing energy*, E_x^{corr} and E_x^{raw} is the x-component of the jet energy, respectevely corrected and raw.

In Figure 3.11 is shown the *missing energy* distribution of our muons and electrons datasets, both without any correction (*raw*) and after muon and jet corrections, when 2 JETS in the event are required. Focusing our attention on the electrons distribution, two peaks are clearly visible in the trend: the one at smaller values of $\not{E}_T$ is due to QCD events, where the *missing energy* is not related to the presence of a neutrino, while the shoulder around $\not{E}_T = 40$ GeV is due to the real W events. With the corrections, small $\not{E}_T$ values, due to jets fakes, are corrected and lowered and high $\not{E}_T$ values are increased. At the end of the procedure there is a better discrimination between the two peaks, for both electrons and muons.

Chapter **4**

Signal and Background modeling: Monte Carlo simulation

One of the most important techniques used to study and parametrize the processes produced in $p\bar{p}$ collisions are Monte Carlo simulations. In particular some of the processes simulated for this analysis needs the use of two MC *generators matched together, i.e.* ALPGEN *for the parton level generation and* PYTHIA *for the showering. This matching leads to an overlapping in phase space of events that belongs to different, and in principle indipendent samples. The method used to remove this double counting is analyzed and discussed in this chapter.*

4.1 Monte Carlo generators

The interpretation of data from high energy physics particle colliders and their use to extract measurements on foundamental physical parameters often heavily relies on the theoretical modelling of the physical processes and detailed simulation of the interactions of particles with detectors - we refer to this as Monte Carlo since the current knowledge of QCD and electroweak interactions is implemented using numerical MC techniques -. In recent years a number of tools have been developed to enable an increasingly more precise description of the final states resulting from high energy collisions.

The main goal of a MC event generator is to provide a complete picture of the large multiplicity of particles in which consist the outcome of an hard interaction, whether it is a simple scattering at large angle of some of the hadron's elementary constituents or their annihilation into resonances or a combination of the two. It is required to provide the description of the particles types and momenta on event-by-event basis.

The foundamental idea behind the simulation of hadron-hadron collisions is the "factorization", the possibility of splitting the overall collision into separate and sequential phases, approximately independent. In particular, factorization allows to decouple the complexity of the proton structure and of the final state hadrons formation from the elementary hard interaction among parton constituents. In other terms the proton structure, made of valence quarks that are held together by a continuos exchange of gluons, the hard interaction between the constituents of the protons that collides, and the hadronization of the final quarks to bound into color neutral states, are threated as 3 well separated steps of the whole interaction. This is possible since, defined as Q the scale of the the hard interaction, its time frame is so short ($\frac{1}{Q}$) that the interaction of the quark involved in the scattering with the rest of the quark can be neglected, being impossible for the struck quark to negotiate with its partners a coherent response to the external perturbation, while it is kicked away. After the interaction, the final partons get through a phase in which they emit radiation until an exchange equilibrium is reached again and the memory of the hard process has been lost. At this moment the hadronization process takes over, nearby partons merge into color singlets and the initial hadrons fragments are recombined leading to the underlying event final states.

The Monte Carlo generators can be divided into two main cathegories:

1 Parton-Level generators

2 Parton-Shower generators

The MC generators used to produce the MC samples used in this analysis are ALPGEN ([1]) and PYTHIA ([39]). The first one belongs to the Parton-Level typology, while the second one can generate both the matrix element interaction and the parton shower. Since PYTHIA is not able to deal with more than 2 partons in the final state, ALPGEN is needed for such processes, such as W + $\mathcal{N}p$, that involve more partons in the final state. ALPGEN can produce up to $\mathcal{N}p = 4$. Therefore PYTHIA has been used to produce the showering when coupled to ALPGEN, for that processes where a more precise calculation of the matrix alement was needed, and the sole PYTHIA generation for others, such as top sngle top and di-boson production.

4.1.1 W+jets with ALPGEN ([1])

ALPGEN has been used to generate the W+jets processes to the parton-level. It has been matched to PYTHIA to produce the showering and the hadronization. In a nutshell, the parton level generation can be summarized in the following steps:

I. some initial parameters of the interaction, such as quark masses, jet multiplicity and, in case, rapidity and P_T cuts are defined.

II. a first set of phase space integration cycles is performed, in order to explore the cross section distribution in phase space and among the possible contributing subprocesses. A subprocess, a phase space point, the flavour configuration, spin and color of each parton are randomly assigned and the matrix element calculation is performed.

III. Since the information about the weighting is obtained during the previos step, at this point there is a map of the cross section distribution among phase space and subprocesses, that will be used in subsequent iterations.

For W+jets, the subprocesses considered include all configuration with up to 2 light quark pairs. As a default, the following cuts to the kinematic configurations among the generated events are applied :

$$P_T^j > 15 \text{ GeV} \qquad |\eta_j| < 3 \qquad \Delta R > 0.4 \tag{4.1}$$

$$P_T^{hf} > 8 \text{ GeV} \qquad |\eta_j^{hf}| < 3 \qquad \Delta R^{hf} > 0.4 \tag{4.2}$$

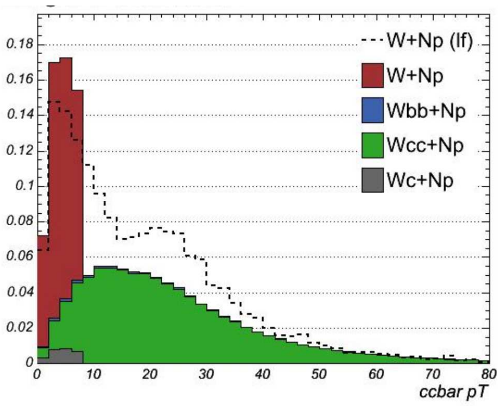
The samples generated by ALPGEN are W+$\mathcal{N}p$, W+$Q\bar{Q}$+$\mathcal{N}p$, W+c+$\mathcal{N}p$, where $Q = c, b$ and $\mathcal{N}p$ is the number of extra partons generated, with $\mathcal{N} = 1, 2, (3, 4$ for LF sample). After the matrix element calculation, ALPGEN is interfaced to PYTHIA for the showering. However, the fact that these two programs acts independently, induces an overlapping in the phase space of events between the generated samples. In fact, since PYTHIA can generate extra jets due to gluon emission or partons at large angle ($\Delta R > 0.7$), it can produce events with the same jet multiplicity and kinematics of ALPGEN. Several algorithms have been developed to avoid this double counting in case of light jets. The one implemented in the CDF MC samples is the *matching conditions* (MLM) of Michelangelo Mangano [40], that suppres the production of PYTHIA extra jet in the sample W+$\mathcal{N}p$ to avoid a double counting in the W+$(\mathcal{N}+1)$ p. For $\mathcal{N} \geqslant 4$ the sample is inclusive, so no suppression is needed.

However, there are no implemented algorithm to avoid the overlap in phase space between heavy flavour samples and the W+$\mathcal{N}p$ one. In fact, it is possible to generate, for example, W + $b\bar{b}$ + 1 jet from the W + $bb + 1p$ sample or from the W + $1p$ when an extra gluon produces a bottom pair. This overlap is not a physical issue, but just an accident induced by the use of two independent generators for two factorized phases of the interaction, an should be removed, according to the method explained in the next section. In fact, this double counting involves heavy flavours, whose modelling is very important in such analyses, like the present one, that require b-tagging.

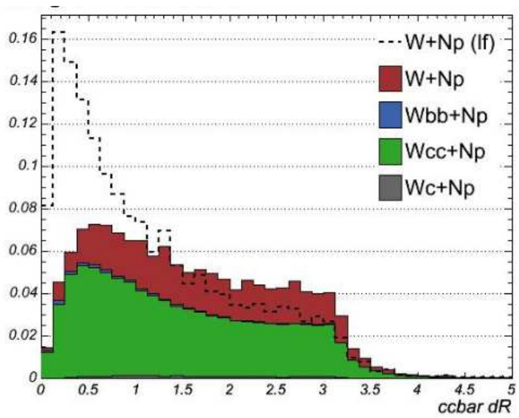
4.2 ALPGEN+PYTHIA overlap removal

The simplest way to perform the removal of the overlap on heavy flavour quark generation between the showering (performed by PYTHIA) and the parton level generation (due to ALPGEN) is to enforce appropriate heavy flavour countents in dedicated heavy flavour samples. It consist of allowing, in each sample, heavy flavours from the showering only if they are lighter than the primary generated partons (i.e. charm pairs are allowed in the W+$b\bar{b}$ sample, but not viceversa) or when they fail the kinematic filter used at generation, i.e. $P_T < 8$ GeV. In the CDF literature, this method is referred as “kinematic removal”. In Figure 4.1(a) is shown an example of the transverse momentum distribution that we obtain for a charm pair after this kind of removal is applied, taken from [41]. Since the distribution suffers of some discontinuity in the connection between the charm pair contribution of W+lf and W+$c\bar{c}$, an alternative method is proposed in the note [41].

The overlap removal method proposed in [41] and applied for this analysis is a “jet-based” method that bases the removal choice on recontructed quantities, like jets. The



(a) Transverse momentum distribution from a charm pair after kinematic removal is applied (coloured histogram).



(b) Angular distance ditribution of two charm pairs. The dotted line, that is the shower, show an higher rate at small angles, while the matrix element products have the opposite behaviour.

Figure 4.1: This two plots are shown in [41] to highlight a different behaviour between shower generated heavy querks and ALPGEN's ones

idea is that the showering and the matrix element generations are characterized by a very different ΔR distribution of heavy quarks pairs. As shown in Figure 4.1(b), the showering has an higher rate for collinear pairs, as expected from gluon splitting processes and because of ALPGEN generator cuts.

Defining an heavy flavour jet as in Definition 1, and reconstructing the jet as described in Sec. 3.3, the *jet-based* method prescrives to veto events from ALPGEN where the matrix element heavy flavour quarks wind up in the same jet, and to remove events that involves heavy flavour quarks generated by the showering, when only one of the quark pair is inside one jet cone.

Definition 1 *a b-jet is a bottom hadron (with PDG [5] code 5xx or 5xxx) within a ΔR cone of 0.4 about the jet axis; a c-jet is a non-b-jet containing a charm hadron (with PDG [5] code 4xx or 4xxx) in its jet cone.*

The application of the method resolves in removing "not-appropriated" heavy flavour events in each dedicated heavy flavour sample, e.g. removing bottom and charm pairs from the light flavour sample, bottom pairs from the W+$c\bar{c}$ sample, etc ... , only when they do not belong to the same recontructed jet. The removal of ALPGEN events is already guarantee by the ΔR cut at generation level, with the small exception of events in which the jet is well balanced between the pair of quarks, and the distance between them is $0.4 < d < 0.8$, which means that pass the selection at generation level, but wind up in the same jet, so should be removed according to the method applied. This case is considered an effect of the second order in this analysis, since also the differences between the *jet-based* and *kinematic* overlap removal methods are small. The distribution of the kinematic variables of heavy flavour quark pairs obtained with the *jet-based* method, compared to the *kinematic* method, produce an higher contribution of heavy flavours from the light flavour sample, and a smoother, hence more reasonable, distribution, as far as ΔR (Fig 4.2) and P_T (Fig 4.3). Notice that ΔR distribution is directly related with the m_{jj} distribution that we will use in the following to estimate signal events.

In conclusion, since the *jet based* method permits to obtain a more realistic momentum and ΔR distributions and is the one used to estimate the data driven *k-factor* for the correction of the heavy flavour content of MC samples (Sec 6.2), we applied the second method to our analysis.

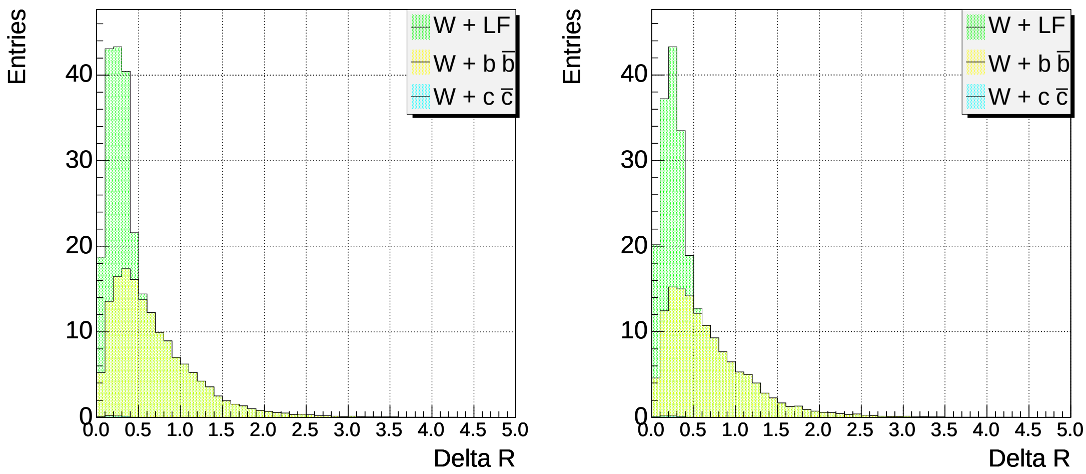
4.3 Normalization of Monte Carlo samples

We will use the MC simulation to obtain a template of each considered background of this analysis, appropriately normalized to the data luminosity and to all that factors and efficiencies that are needed to compare the MC simulation to the data sample. We will use these normalizations as the starting point to performe a fit on the invariant mass distribution to finally estimate the $W^{\pm}Z$ events on data.

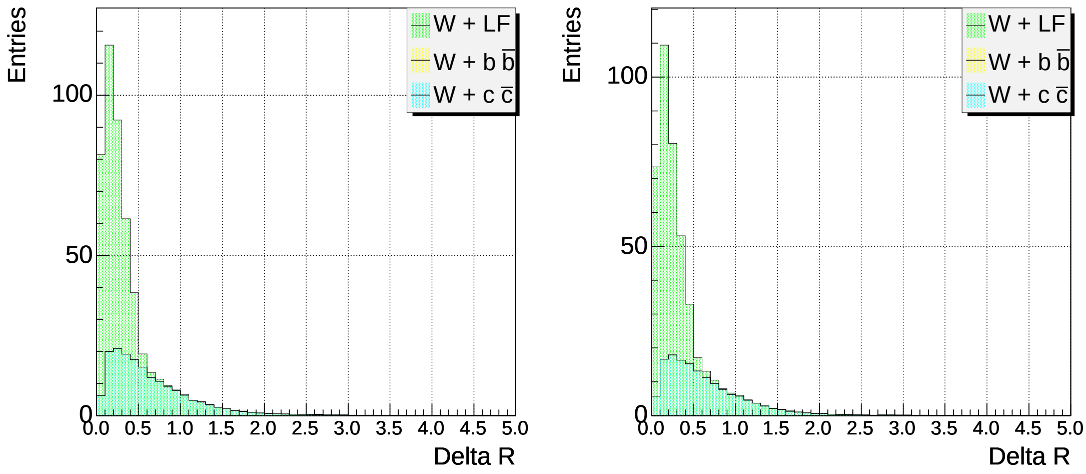
4.3.1 Cross section and luminosity normalization

Every MC sample is generated independently from each other. This means that they need to be normalized according to their cross section, the number of generated events and the total integrated luminosity of the dataset. Furthermore, the number of removed events according to the method described in Sec 4.2 needs to be subtracted to the total of the events generated to perform the normalization of the samples.

In Equation (6.1) is shown the formula used to obtain the number of events of a generic



(a) $Q = b$



(b) $Q = c$

Figure 4.2: Distribution of ΔR between the two heavy flavour quark pairs ($Q\bar{Q}$) of the event, after the "jet-based" double counting removal is applied.

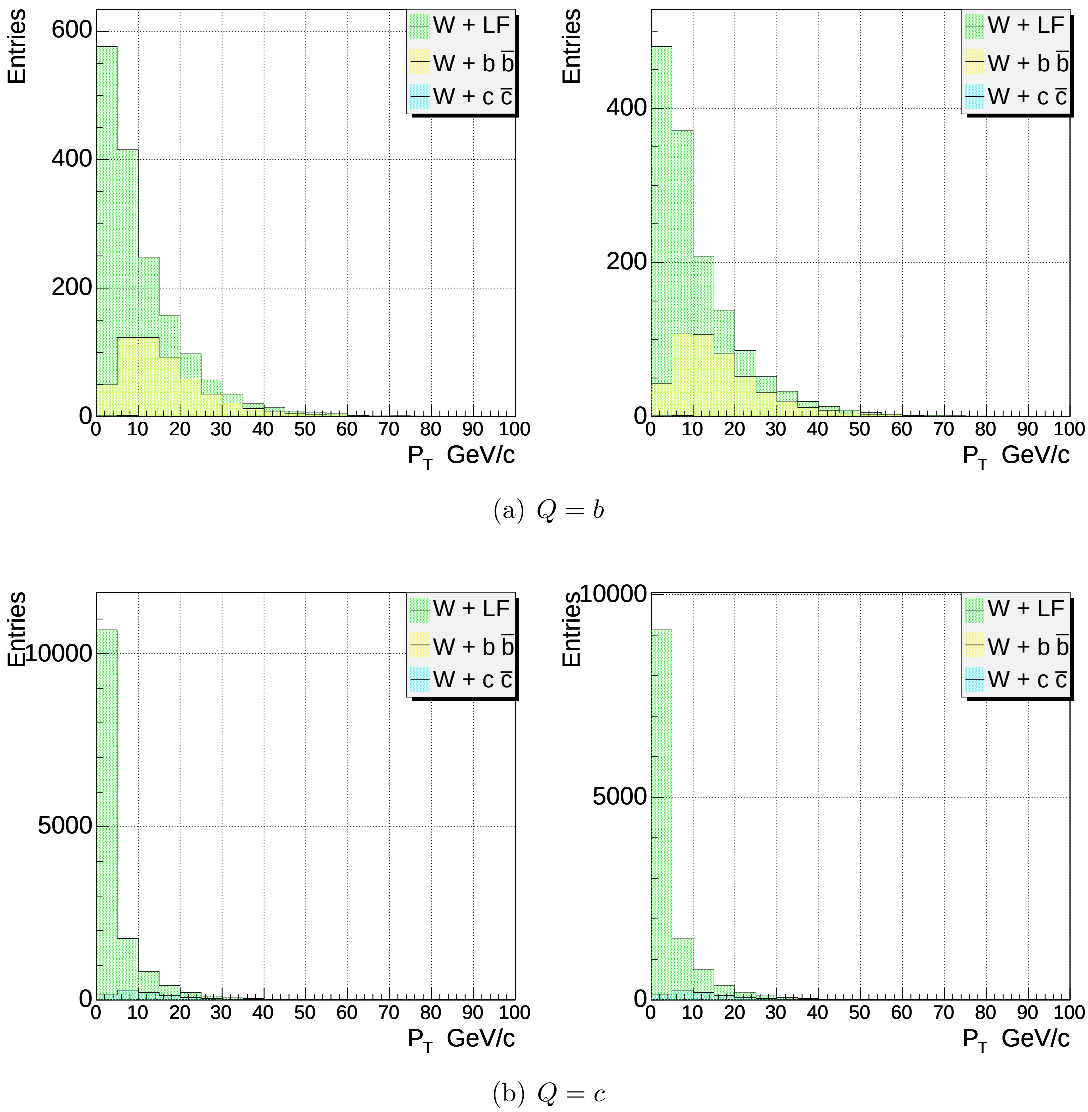


Figure 4.3: Distribution of the combined P_T of the two heavy flavour quark pairs ($Q\bar{Q}$) of the event, after the “jet-based” double counting removal is applied.

$p\bar{p} \rightarrow X$ MC sample normalized to the data luminosity.

$$\mathcal{N}orm\ \mathcal{F}actor = \mathcal{L}_{data} \cdot \frac{\sigma_{p\bar{p}\rightarrow X}}{\mathcal{N}_{gen} \cdot f_{rem}}\ k\text{-}factor \tag{4.3a}$$

$$f_{rem} = \frac{\mathcal{N}_{gen} - \mathcal{N}_{removed}}{\mathcal{N}_{gen}} \tag{4.3b}$$

where f_{rem} is the factor that takes into account the ALPGEN+PYTHIA double conting removal, $\mathcal{L}_{data}$ is the total integrated luminosity of the dataset, $\sigma_{p\bar{p}\rightarrow X}$ the cross section of the considered process and $\mathcal{N}_{gen}$ the number of generated events in that sample. The *k-factor* takes into account next-to-leading corrections and is ~ 1.4 for ALPGEN generated samples.

4.3.2 Trigger efficiency and scale factors

The MC sample that we are going to use in this analysis involve a complete simulation of the collisions and the detector, although the trigger system is not accounted. For this reason we give a weight to each MC event, according to the trigger chosen for data and the relative trigger efficiency described in Sections 3.1.2 and 3.2.2 for electrons and muons respectively. Furthermore a scale factor should be applied to correct MC and data different efficiency for electrons and muons identification cuts, that is also described in the above stated sections. The final weight, calculated for each event, is

$$\mathsf{event\ weight} = \epsilon_{trigg} \cdot \epsilon_{SF} \cdot \epsilon_{zvtx} \tag{4.4}$$

where ϵ_{trigg} is the trigger efficiency, ϵ_{SF} is the selection scale factor and ϵ_{zvtx} is the efficiency on the request that the high-P_T lepton has $|Z_0| < 60$ cm and is due to the large Tevatron beam spot in the longitudinal direction. It is the same for electrons and muons and have a constant value of 0.975.

Chapter 5

Search for $WZ \to l\bar{\nu}_l b\bar{b}$

The purpose of this analysis is to search for an evidence of the $WZ \to l\bar{\nu}_l b\bar{b}$ process in the jets invariant mass distribution. The characteristic signature of our signal is an energetic lepton plus missing energy plus two b-jets. We are going to describe which processes mostly contribute to background and which selection cuts are applied in order to enhance the signal acceptance in our sample, minimizing the background. First the leptonic W selection is described. Then the jets requirement are discussed, with a particular attention to the minimum energy required for the jets and the different kind of b-tagging that can be applied to the event. Each b-tagging cathegory is treated as a separate sample, indipendent from each other, in order to perform separate analysis and obtain enhanced sensitivity. In the end we summarize some other cuts to reduce background contamination and make a brief discussion on the importance of the background shape, besides statistics, to enhance signal sensitivity.

The aim of this analysis is a search for an evidence of the $WZ \to l\bar{\nu}_l b\bar{b}$ process using the jets invariant mass distribution, in events characterized by a $W \to l_\nu \nu$. The principal feature of our signal is the precence of a W boson which decays in a lepton and a neutrino. It is identified by the presence of a well isolated and high energetic lepton plus a large amount of *missing energy*, since the neutrino doesn't interact in the detector. The backgound processes which produce a W boson signature can be classified into two cathegories: QCD and *W-like* events. The latter is represented by electroweak bosons and top production which produce real leptons plus *missing energy* in the final state.

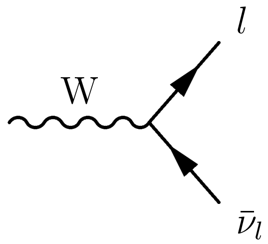


Figure 5.1: Leptonic decay of a W boson

The second one, whose estimation is described in Sec. 6.1, is represented by QCD events where a jet *fakes* an electron (or less frequantly a muon), that is the jet is reconstructed as a track associated with electromagnetic deposit and fulfills the electron requirements (Sec. 3.1), and the *missing energy* is produced by a mismeasurement of the jets energy or by a second jet that escapes through a crack in the calorimeter or the beam pipe.

The signature peculiarity of our signal are the 2 b-jets in which the Z boson is required to decay. This requirement produces a relative enhancement of some background that are very small in the sample with all the analysis cuts except the b-tagging requirement. This sample will be referred to in the following as the *pretag* sample. As an example, we can mention top pair and single top production, that become dominant in the b-tagged samples and represent a considerable background to our search.

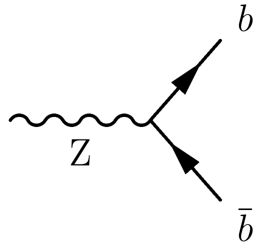


Figure 5.2: Decay of a Z boson in bottom quarks

5.1 Background processes

W+$\mathcal{N}_{jets}$ (light flavour jets) The main background is the production of W boson in association with multiple high-Et jets:

$$p\bar{p} \to W^{\pm} \to l^{\pm}\nu_l + \mathcal{N}_{jets} \qquad (\mathcal{N}_j \geqslant 2) \tag{5.1}$$

W bosons are produced in ppbar interaction mostly by Drell-Yan type process where a quark from a proton and antiquark from an antiproton annihilate in a $W^{\pm}$ boson. QCD radiation from the colliding partons easily produce multiple high-pt parton in association with the W boson. There are several feynamn diagrams leading to this final state but the detailed discussion of this pheonomenological issue are beyond the scope of this work. The inclusive cross section of W+$\mathcal{N}_j$ with $\mathcal{N}_j \geqslant 0$ is $\sim$ 2.1 nb ([1]), that is already multiplicated by leptonic W decay brancing ratio but still need to be summed over the lepton generations, although we are only interested to that processes that involve at leat

2 jets in the final state. To give an idea, the selection efficiency of this sample requiring at least two jets with $E_T > 20$ and $|\eta| < 2$ is $\sim 35\%$.

However, the jets produced in association with a W are originate by LIGHT-FLAVOUR quarks, therfore the most contribution of this sample is due to MIS-TAG, that is b-tagging algorithm recontructs a displaced vertex but the jet was originated mostly by a light flavour quark.

W+$\mathcal{N}_{jets}$ (heavy flavour jets) The above cross section do not include the production of heavy flavor jets. These are separate process and different cross section prediction exists. The inclusive production of an heavy flavour quark in association with a W boson is 1.5% ($\sim$ 32.6 pb) of the W+$\mathcal{N}_j$, and are the main background when b-tagging is required. The heavy flavour components are

$$p\bar{p} \to W^{\pm} \to l^{\pm}\nu_l + b\bar{b} \qquad (\sim 4.2 \text{ pb}) \tag{5.2}$$

$$p\bar{p} \to W^{\pm} \to l^{\pm}\nu_l + c\bar{c} \qquad (\sim 7.4 \text{ pb}) \tag{5.3}$$

$$p\bar{p} \to W^{\pm} \to l^{\pm}\nu_l + c \qquad (\sim 21.1 \text{ pb}) \tag{5.4}$$

where the quoted cross sections are the theoretical values calculated by ALPGEN and are already multiplied by W brancing ratio, although still need to be summed over the leptons generations.

Z+$\mathcal{N}_{jets}$ (heavy and light flavour jets) The second background, in order of cross section production, is the production of multiple jets in association with a Z boson

$$p\bar{p} \to Z \to l^{+}\bar{l}^{-} + \mathcal{N}_{jets} \qquad (\mathcal{N}_j \geqslant 0) \tag{5.5}$$

where a lepton escapes into the beam line or is not well identified and mismeasurement of jet energy fakes the *missing energy*. However, we'll see in the next subsection that there are ways to suppress this background contamination, both for LF and HF jets.

Top pairs, single top and di-boson producion The last two kinds of *W-like* background processes are those, with cross section of the same order of magnitude of our signal (or a bit more), that have final states very similar to the signal, with a W and 2 b-jets.

These processes are top pair and single top production and di-boson associate production (WW or ZZ), whose reaction are shown in Eq. (5.6a), (5.6b) and (5.7)

$$p\bar{p} \rightarrow t\bar{t} \qquad (\sim 7.5 \text{ pb}) \tag{5.6a}$$

$$p\bar{p} \rightarrow t\bar{b} \qquad (\sim 2.9 \text{ pb}) \tag{5.6b}$$

$$p\bar{p} \rightarrow WW \rightarrow \ l\ \nu_l\ j\ j \qquad (\sim 12.4 \text{ pb}) \tag{5.7}$$

The top backgroundshave an event topology that is almost the one of the WZ.However, the TOP PAIR production is characterised by much more jets in the final state that will give an handle to discriminate signal from this background, though the s-channel of single top production has exactly the same signature of WZ and is hardly distinguishable. As far as the other di-boson processes is concerned since W do not decay in b-quark, WW associate production would poorly contaminate our data, while ZZ will contaminate our data in case one of the leptons misses detector acceptance.

5.2 Initial selections of our sample

5.2.1 Trigger and good run list requirements

The first requirement on data is that they belong to a run that is in a *good run list*, which means to require that all the subdetectors relevant to the analysis were checked to be well functioning in each considered run. In particular for this analysis the Silicon Good Run List version 29 with logic (1,1,4,1), that requires operating the following subdetectors, that are important for our analysis:

- showermax (Sec. 2.3.2) and Calorimeters, for electrons identification
- inner tracker (both SVX and ISL, Sec. 2.3.1) for b-tagging
- central muons detectors (CMU, CMP, and CMX, Sec. 2.3.3), excluding the CMX before a certain run (150145), for muon selection.

The second pre-selection applied to the data is the trigger requirement. For this analysis the high transverse momentum triggers for electrons (CEM) and muons (CMUP and CMX) has been chosen, to enrich the data sample of leptonic Ws at trigger level. The complete description of the trigger paths is given in Section 2.5.4.

5.2.2 The data set

The entire dataset used in this analysis, after trigger and *good run list* application, adds up to 4.7 ± 0.3 fb^{-1} of data, which have been divided into 25 periods for electrons and muons to take into account different efficiencies or detector response due to a variety of reasons, including the ageing of the various subdetectors.

As far as the Monte Carlo simulation, we have considered all the backgrounds described at the beginning of this chapter, with the relative cross section at generation levels.

5.3 $W \to l_\nu \nu$ selection

The triggered lepton has a requirement of $E_T \geqslant 18$ GeV or $P_T \geqslant 18$ GeV/c in case of an electron or muon respectively. We require offline an electron with $E_T \geqslant 20$ GeV, that fulfills the definition of a *Central* and *Tight* described in Sec. 3.1. The muon is required to be a *Tight* CMUP or a *Tight* CMX, as defined in Sec. 3.2, with $P_T \geqslant 20$ GeV/c. It is verified, for each candidate lepton, that the z coordinate of its track is consistent with the Primary Vertex position within 5 cm along z axis, in order to assure that it belongs to the primary interaction of the bunch crossing.

A considerable unbalancement of the energy on the transverse plane is the sign of a neutrino The corrected $\not\!\!E_T$ in the event is then required to satisfy

$$\not\!\!E_T \geqslant 25 \text{ GeV}$$

in order to reject most of the QCD background, due to mismeasurement of the energy when jets are present. In Figure. 5.3 it is possible to notice the trend of the missing energy distribution when an high P_T lepton and 2 jets are present in the event. Two peaks are present, one at low values of $\not\!\!E_T$, due to QCD background, the other around 40 GeV mostly due to events with a real neutrino. The $\not\!\!E_T$ threshold permits to reject most of

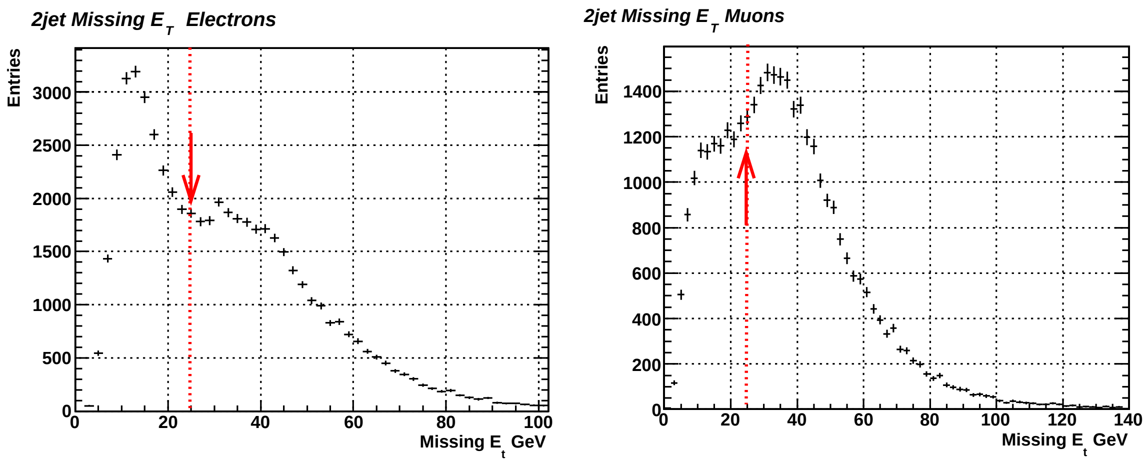


Figure 5.3: Distribution of the *missing energy* in events with an energetic lepton, for electrons and muons respectively

the QCD background. An estimate of the remaining contribution will be discussed in Sec. 6.1.

The reconstructed transverse invariant mass of the W[1], calculated as in Equation (5.8), is required to satisfy $M_T^W \geqslant 30$ GeV/c^2 and $M_T^W \geqslant 20$ GeV/c^2 for electrons and muons respectively. For muons the requirement can be looser, in order to take advantage of the lower background to increase the acceptance.

$$M_T^W = \sqrt{2 \cdot \left(E_T^l \cdot Sf\right) \cdot \not{E}_T \cdot (1 - \cos(\Delta\Phi_{l,\nu}))} \qquad (l = e, \mu) \tag{5.8}$$

where E_T^l stands for either the electron energy or the muon momentum, while Sf stands for the correction to the electron energy discussed in Sec. 3.1.1 or the muon momentum smearing described in Sec. 3.2.1. Both this correction have been introduced to take into account a disagreement in MC and data energy recontruction or calibration.

5.3.1 Z background rejections and cosmic ray vetos

Few events vetos have been introduced to enhance the purity of the inclusive W sample. Firstly the standard cosmic rays and conversion vetos are applied. The former rejects muon candidates from cosmic ray, which requires the muons track to be within a fiducial region around the interaction point and the muon chamber hit time compatible with the

[1]Since the event is not longitudinally balanced, because of the momentum of the partons inside the colliding protons, the neutrino longitudinal momentum is unknown

bunch crossing; the latter rejects lepton candidates produced by photon conversion in the beam pipe or tracker. In addition, only for muons candidate in the data sample is required, as quality cut, that the χ^2 of the muon track reconstruction in the tracker is less than 2.3. in order to reject ions and kaons decay in flight.

Next we apply a *Z-veto*. This veto requires that selected leptons do not have an invariant mass with any isolated track of the event in the Z-mass region ($66 < M_{ll} < 116$ GeV/c) and permits to reduce this kind of background by 51% for electrons and 31% for muons. Furthermore events with two *tight* leptons, either electrons or muons are also rejected. With this selection we can reduce the Z+jets background to less than 2%

5.4 Jets selection

First of all, we need to require at least 2 jets in the event, since we are going to perform a $l_\nu\ \nu\ j\ j$ control analysis before applying the tagging requirement. The jets are recontructed as explained in Section 3.3 and the energy is corrected to hadron level (level 5). The η of each selectected jet is required to be:

$$\eta_j \leqslant 2.4 \tag{5.9}$$

to exclude jets in the very forward, poorly instrumented region in order to minimize soft interaction products, and

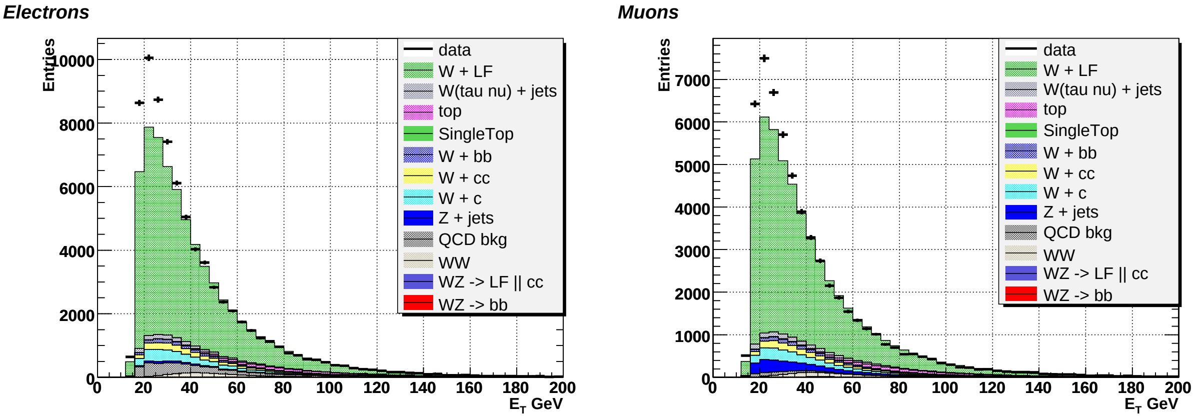
$$|\eta_{j1} - \eta_{j2}| \leqslant 2.5 \tag{5.10}$$

that rejectc some background of back-to-back without affecting the signal acceptance.

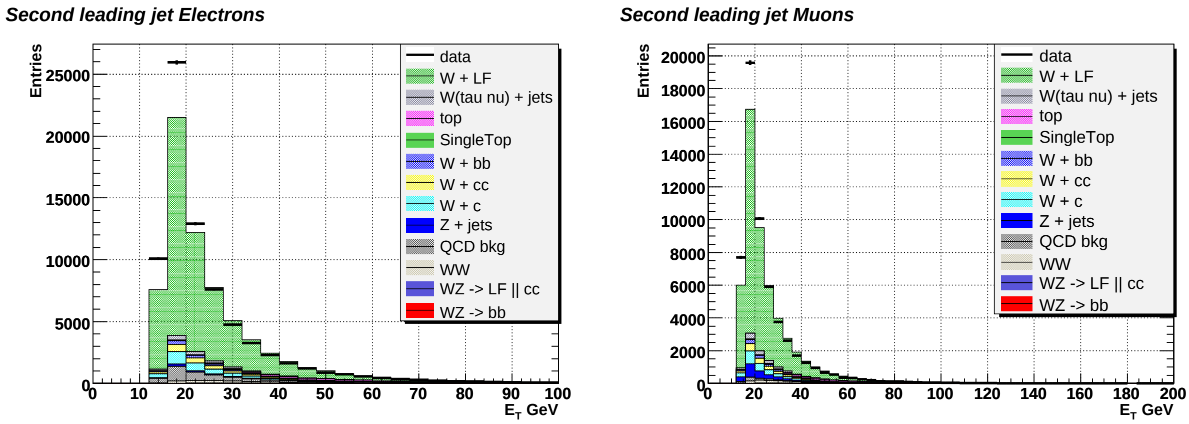
We show the energy distribution of the leading and the second leading jet in events with more than 2 jets above a certain threshold of energy (15 GeV), in Figure 5.4: it's straighforward to notice that the agreement of jet E_T distribution between data and MC improves at $E_T \sim 20$ GeV We decided to apply a symmetric energetic cut on the leading jets requiring, for each selected jet of the event:

$$E_T^1 \geqslant 20 \text{ GeV} \qquad E_T^2 \geqslant 20 \text{ GeV} \tag{5.11}$$

This energy threshold not only permits to obtain a better agreement between data and MC, but reduces the W+jets contamination of the sample, not much affecting the signal



(a) Transverse energy of the leading jet



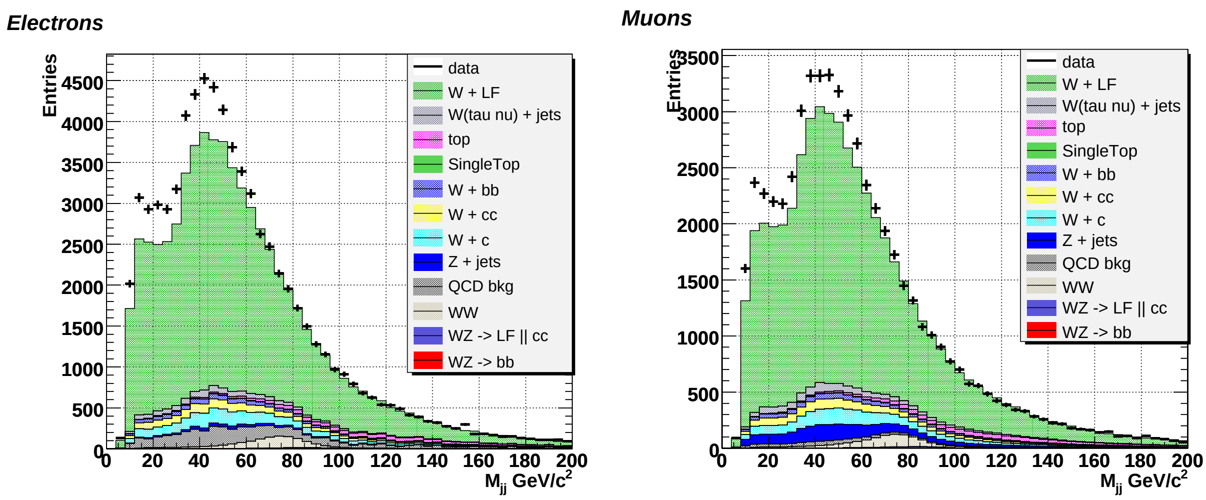
(b) Transverse energy of the second leading jet

Figure 5.4: Transverse jet energy of the two leading jets of the event, requiring $E_{T1}, E_{T2} >$ 15 GeV. The MC samples are normalized according to luminosity, as described in Sec. 4.3

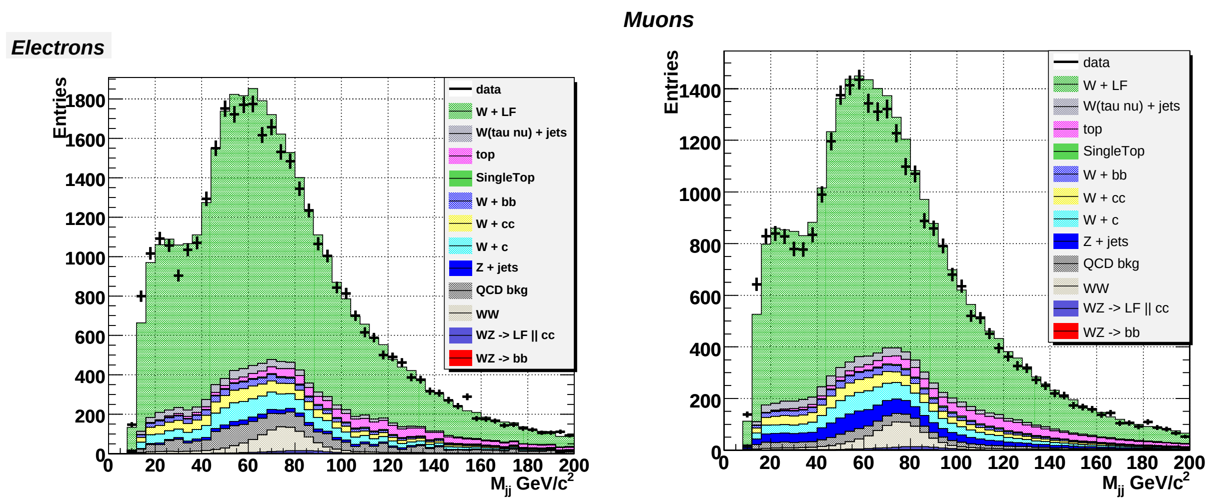
acceptance, as we show in Figure 5.5, where the two leading jets invariant mass distribution is shown.

5.4.1 Tag requirement to identify b-jets

The purpose of our search is the Z in the $b\bar{b}$ channel produced in association with a W that decays leptonically. As described in Sec. 3.4, the main feature of a jet originated from a b quark is the presence of a displaced vertex inside the jet, due to the long lifetime of the B mesons. SecVtX TIGHT and LOOSE TAG algorithm are used to recognize the b-jets in the event. Only the flavour of the two leading jets is investigated, even though events with more than 2 jets are accepted, for sake of simplicity and the fact that it's unlikely



(a) 2 leading jets $E_T \geqslant 15$ GeV



(b) 2 leading jets $E_T \geqslant 20$ GeV

Figure 5.5: Effect of an higher threshold in jet energies on the two leading jets invariant mass distribution. The MC samples are normalized according to luminosity, as described in Sec. 4.3. It's straighforward to notice how this cut reduces background. All other cuts applied.

that the b-jet is not one of the two most enegetic jets in a signal event, since the extra jet can only be produced as initial or final state radiation (ISR, FSR). Assuming that a jet tagged with the TIGHT algorithm fulfill also the LOOSE requirements, the possible combinations are the following:

I. 2TT: both the two most energetic jets are tagged as b-jets by the SecVtX TIGHT TAG algorithm

II. 2TL: one jet is tagged by the LOOSE TAG algorithm only (it doesn't met the TIGHT requirements), the other is tagged by the TIGHT TAG one.

III. 1T: one jet is tagged by the TIGHT TAG algorithm and the other jet is UNTAGGED.

IV. 2LL: both jets are tagged by the LOOSE TAG algorithm only.

V. 1L: one jet is tagged by the LOOSE TAG algorithm only and the other jet is UNTAGGED.

VI. UNTAGGED: both jet are UNTAGGED.

Since each of these combinations has different statistics and different signal to noise ratio, we decided to treat them into separate cathegories (samples) that are indipendent by construction. In this way an event will belong only to one cathegory and the analysis can be independently performed on each of them, combining only the final results. From here to the end we will refer to each tagging combination with the names in the above list, while the total of events before introducing the tagging requirement will be refered to as the *pre-tag* sample.

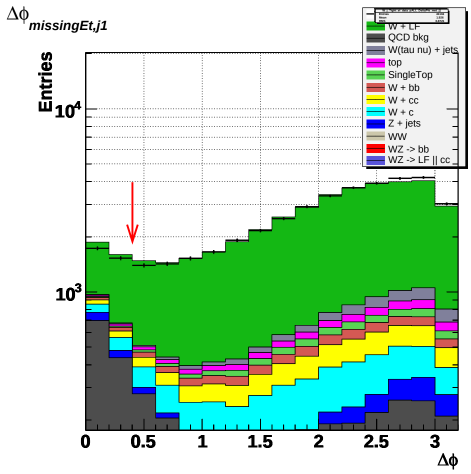
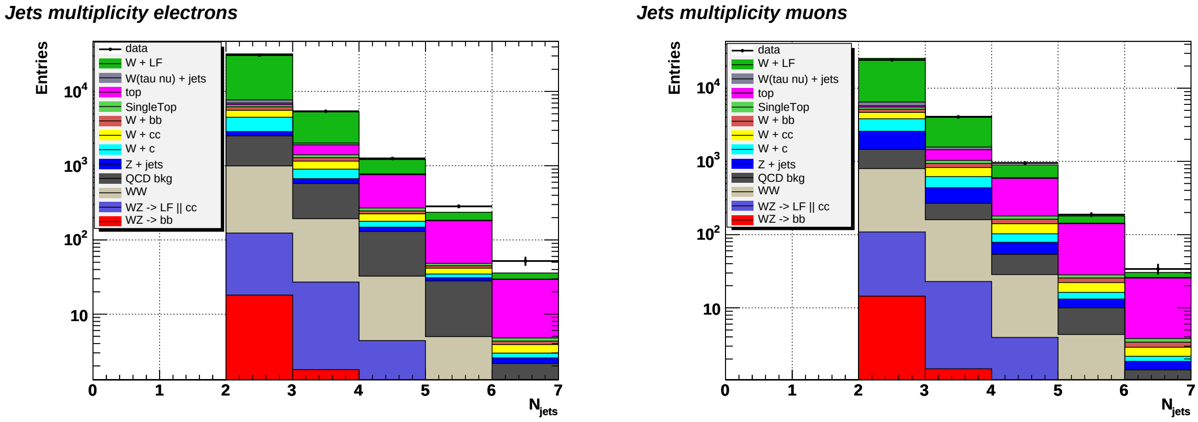


Figure 5.6: $\Delta\phi_{\not{E}_T, j_1}$ distribution for electrons. The cut value is indicated by the red arrow



(a) Electrons (b) Muons

Figure 5.7: Jet multiplicity in events, after appling all the analysis cuts. The top process increase with jet multiplicity, contrary to the signal.

5.5 Further QCD and top rejection

Two more "standard" cuts for this kind of analysis needs to be added to the above list:

I. $\Delta\phi_{\not{E}_T, j_1} \geqslant 0.4$

II. $\mathcal{N}_j \leqslant 2$

The first cut is needed to reduce *non-W* (QCD) backgorund contribution. In figure 5.6 is shown the $\Delta\phi_{\not{E}_T, j_1}$ distribution; especially for electrons there is a reduction of 15-20% of QCD fraction with this requirement.

The second cut is on the number of jets of the event that fulfill the jets requirements ($\mathcal{N}_j$). The jet multiplicity distribution, shown in Figure 5.7, highlight that the major contribution to $\mathcal{N}_j > 3$ bins is due to top pair production and has a little signal contribution. For this reason we decided to discard events with more than 3 selected jets, and divide the analysis for $\mathcal{N}_j = 2$ and $\mathcal{N}_j = 3$, for the different signal to noise ratio of this two samples. The 2-jets bin will demostrate higher sensitivity and, for sake of semplicity, will be the only one used for this preliminary analysis.

The requirements of this analysis are summarized in Table 5.1.

Variable	Cut
N *tight* leptons	1
$\not{E}_T$	$\geqslant 25$ GeV
M_W^e	$\geqslant 30$ GeV/c^2
M_W^μ	$\geqslant 20$ GeV/c^2
$\mathcal{N}_j$	$= 2$
$\Delta\eta_{j_1,j_2}$	$\leqslant 2.5$
$\Delta\phi_{\not{E}_T,j_1}$	$\geqslant 0.4$
Jet selection:	
E_T^{jet}	$\geqslant 20$ GeV
η^{jet}	$\leqslant 2.4$

Table 5.1: Summary of the analysis cuts

5.5.0.1 Preliminary sensitivity studies

We show a first and rough sensitivity study in Table 5.2, calculated with the figure of merit defined in Equation (5.12).

$$\mathcal{S} = \frac{\mathcal{N}_{signal}}{\sqrt{\mathcal{N}_{background} + \mathcal{N}_{signal}}} \tag{5.12}$$

It's easy to notice that the studied cathegories has very different expected sensitivities, as reasonably expected. In fact we expect that the 2 b-jets of our signal are reconstructed and identified by the TIGHT SecVtX algorithm, and we expect a better signal to noise ratio in the 2 jet multiplicity bin, since some background are characterised by an higher jet multiplicity. Tab 5.2 shows that the most sensitive cathegories are the 1 and 2 TIGHT TAGS, that however suffer of low statistics. Since the other cathegories have much less sensitivity that these two we will concentrate our search only to the 1 and 2 TIGHT TAG samples postponing the study of the LOOSE samples at a later stage because taking into account these cathegories now would greatly complicate the analysis without a reasonable gain of sensitivity.

5.6 Considerations on the invariant mass shape

The evidence of a signal, especially a small one, in not only related to the statistic significance, i.e. the signal to noise ratio, but also to the difference between the background and the signal shapes. As evident in Fig 5.5(b), our invariant mass distribution has a

Cathegory	Sensitivity			
	Electrons		Muons	
	2 JETS	3 JETS	2 JETS	3 JETS
pretag	0.11	0.025	0.092	0.024
1T	0.22	0.034	0.19	0.034
2TT	0.28	0.043	0.25	0.040
2TL	0.14	0.024	0.15	0.020
2LL	0.028	0.0041	0.047	0.0077
1L	0.054	0.010	0.048	0.010

Table 5.2: A preliminary sensitivity study of the different tagging cathegories, divided for 2 and 3 jet multiplicity bins.

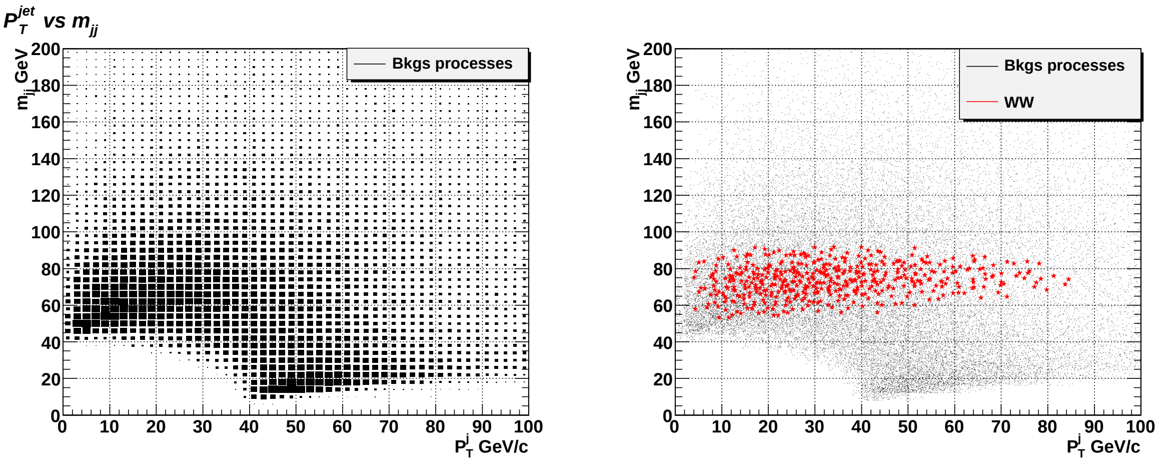


Figure 5.8: P_T^j and di-jet invariant mass correlation for background only and background compared to signal. Only MC simulation.

bump that starts around 40 GeV, extending also above our signal peak. This peak is kinematically correlated to the minimum jet energy requirement, the angular distance of the jets in the transverse plane ($\Delta\phi$) and the combined transverse momentum of the two jets (P_T^j), as shown in Figures 5.9 and 5.8. In particular we are going to explain how a cut on the last kinematic variable alter the shape of the invariant mass distribution, obtaining a smoother backgorund on which a small signal would be more clearly visible by eye, at the price of 40% of statistics reduction.

The peaking structure starting at 40 GeV in invariant mass distribution is due, as stated above, to the jet energy threshold of 20 GeV. In fact, this requirement gives rise to two different thresholds in the jets invariant mass distribution , the first one, at $m_{jj} \sim 20$ GeV, is for almost collinear jets ($\Delta\phi \sim 0.5$, infact we use jet cone of 0.4), where

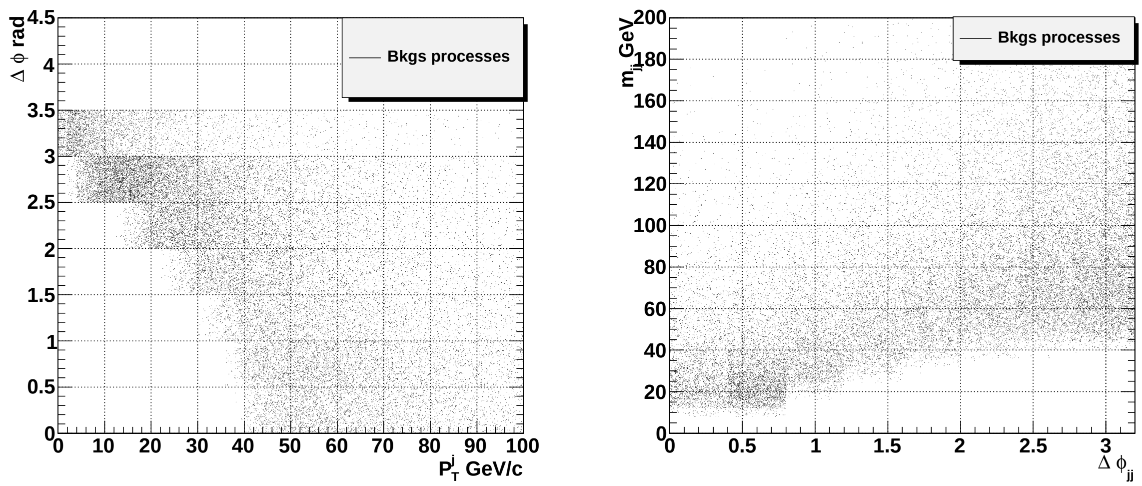


Figure 5.9: $\Delta\phi_{jj}$ correlation with P_T^j and di-jet invariant mass. Only MC simulation.

the invariant mass is minimum and the combined P_T^j is maximum. The second one is at $m_{jj} \sim 40$ GeV, for back to back jets ($\Delta\phi \sim \pi$), where the invariant mass is maximum.

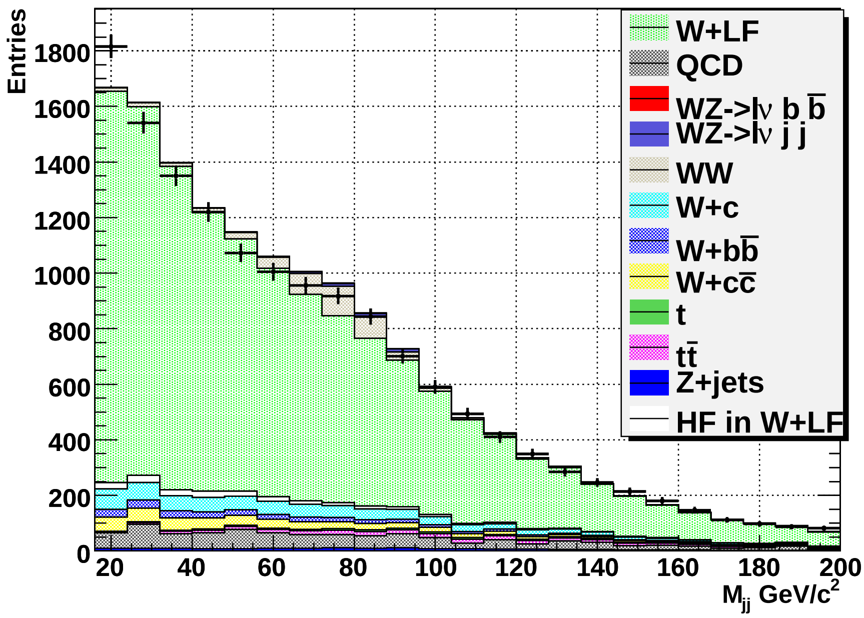
The threshold peak is just on the top the di-boson peak, both in the *pretag* and *tagged* distributions, and its is possible to demonstrate (using pseudo-experiments) that this kind of shape strongly reduces the signal sensitivity. There are two possible choices:

I. lower the jets E_T thresholds

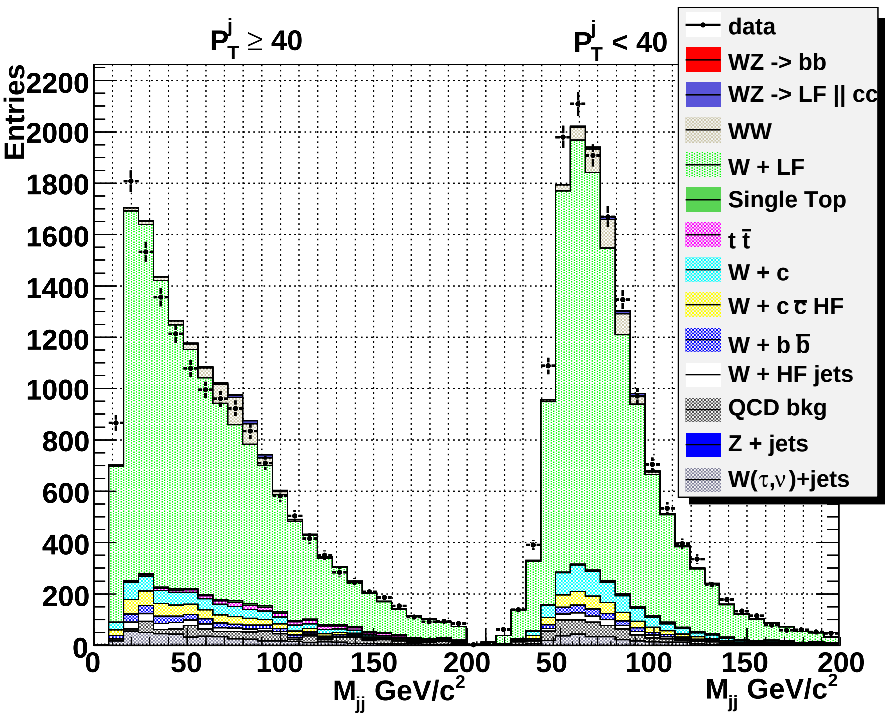
II. cut in the P_T^j distriution in order to reduce the peak ($P_T^j \geqslant 40$ GeV).

The first choice produce a backgroun peak at lower invariant mass. However this turned out not to be a good choice since, as we have shown in Fig 5.5(a) the disagreement between data and MC is larger at low jet energy. Additionally with this cut we suffer from significantly lower signal to noise ratio. The second choice (Fig 5.10(a)) leads to a signal to noise ratio lowering in turn but provides a very smooth background shape that compensate the statistic loss in the total fit sensitivity.

However, the smallness of our signal's cross section and the unavoidable inefficiency of b-tagging suggest us not to further reduce statistics. For this reason we decided to separate the events that fulfill the P_T^j requirement from the events that do not fulfill it, although performing the analysis on both samples. The Figure 5.10(b) show the invariant mass distribution that will be used in this analysis, in the case of pre-tag events, where



(a) Invariant mass distribution of the two leading jets of the event, requiring a combined transverse momentum $P_T^j \geqslant 40$ GeV. No fit is performed and the normalization is derived from MC as described in Sec. 4.3 and from QCD fit and *Method 2*as in Sec. 6.2 and 6.1. Electron sample



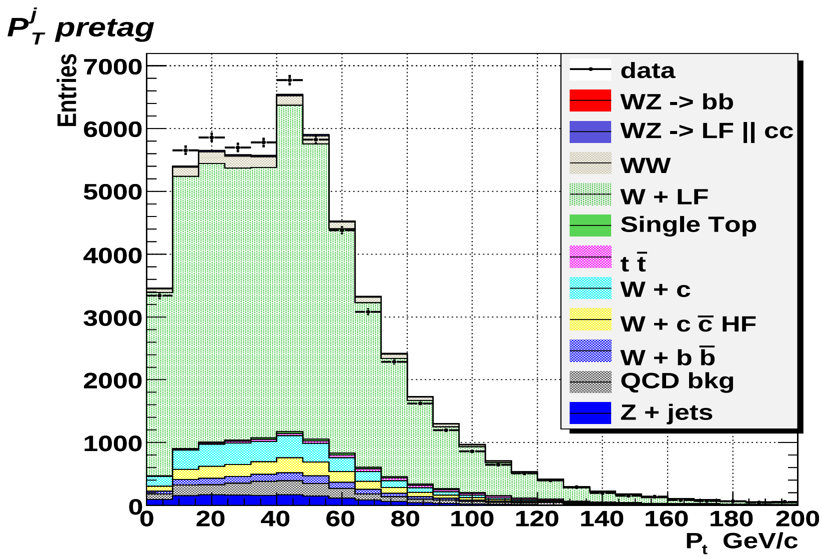
(b) Invariant mass distribution of the two leading jets of the event. No fit is performed and the normalization is derived from MC as described in Sec. 4.3 and from QCD fit and *Method 2*as in Sec. 6.2 and 6.1. The left part of the histogram is composed of events that fulfill the requirement of a combined transverse momentum $P_T^j \geqslant 40$ GeV, the right part is composed of the ones that do not fulfill that requirements. Electron sample

Figure 5.10:

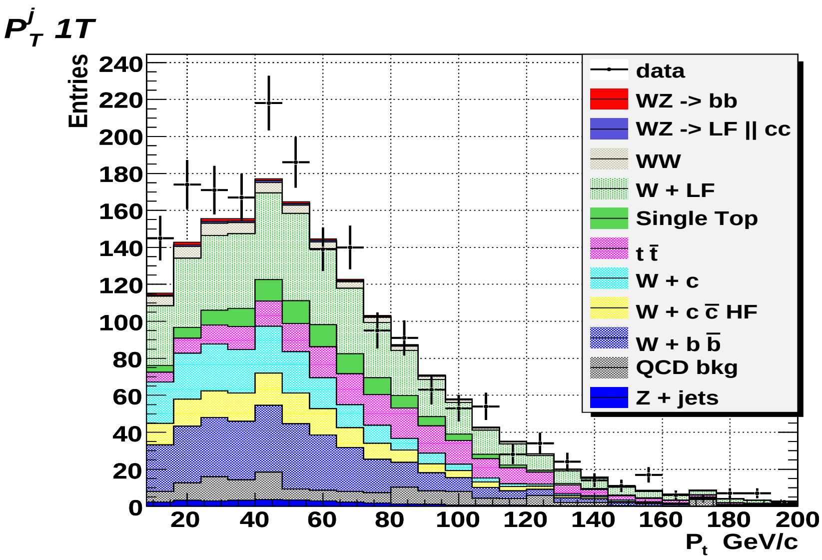
the first 100 bins of the histogram are filled with events that fulfill the $P_T^j \geqslant 40$ GeV requirement, while the other 100 are filled with the invariant mass of events that do not fulfill this requirement. In this way we both exploit statistics and shape, being able to simply evaluate which one will be the best choice for our specific case.

5.7 Di-jet momentum distribution

In order to verify a good agreement with MC and data, we show, in Figure 5.11, the distribution of the combined jet momentum of 2 JETS events, for *pretag* and 1T samples. As described in the previous section a cut on this variable can improve the background shape smoothness.



(a) *pretag*



(b) 1T

Figure 5.11: P^j_T distribution for the *pretag* and 1T samples. No fit is performed and the normalization is derived from MC as described in Sec. 4.3 and from QCD fit as described in 6.1.

Part II

Data analysis

Chapter **6**

Backgrounds normalizations

The background processes for our signal are many, but belongs to two main typologies: the W-like *processes, that involves a real W in the event and the* non-W*, also referred to as QCD. Among those that belong to the former kind, some of them are well known and can be estimated by a reliable* MC *simulation and normaliized to their well calculated or experimentally measured cross section, such as processes that involve top pair and single top production, while other processes are still not well understood and needs to be normalized with differents methods. The purpose of this section is to describe how some backgrounds, for which the* MC*simulation do not exists or is not completely reliable, are normalized. In the following section we are describing the* anti-electrons *and* non-isolated muons *methods for inferring QCD backgrounds and normalizations in the electrons and muons samples respectively and the so called* Method 2 *for the* $W{+}\mathcal{N}_j$ *background normalization in the tagged samples.*

6.1 QCD background estimate

As emphasised in Sec. 5, one of the relevant background of this analysis are QCD events where a jet fakes a lepton (most of the time an electron) and there is an amount of *missing energy* due to the escape of another jet through a calorimeter crack or the beam pipe or due to a mismeasurement of the jet energy. In this analysis we used the so called *anti-electrons* and *non-isolated muons* methods, introduced in CDF and tested in [38], to estimate this *non-W* background in the electron and muon samples respectively.

This method infers the QCD background fraction from a fit on the *missing energy* distribution. The general idea is to define a background dominated sample that reproduce the behaviour of the QCD background in the data sample. Assuming that QCD events are mostly characterized by a low amount of *missing energy*, compared to real $W \rightarrow l_\nu \nu$ events end assuming that the shape of the tail in $\not\!\!E_T$ in this background enhanced sample reproduces that in the real background sample, we can use this background enriched sample as a template to fit from the *missing energy* distribution the fraction of QCD events that pass our selections. The *W-like* template, that involves all the other backgrounds of this analysis, is obtained from MC simulation and normalized according to efficiencies and integrated luminosity as described in Sec 4.3.

6.1.1 *anti-electrons* sample

The key ingredient of this method is the way in which a QCD enriched sample that is resonably suitable to describe our *non-W* background is defined.
As far as electron is concerned, that are more likely to be faked by jets than muons, in a detector like CDF, we take events where neither *tight electrons* or muons are found and there is an "electron" that pass all the "kinematics" requirements of a *tight electron* but fails a number (we require more than two) of "identification" requirements (The electron cut variables are summarized in Tab. 6.1). We define such a sample of fake dominated electrons on *anti-electrons* sample. We refer to section 3.1 for the cut values and explanations of these variables. In this way our template is heavily enriched of fake electrons (because of the reversion of ID cuts), but is kinematically similar to the *tight electrons* sample of which we want to infer the contamination. The definition of a sample that kinematically behaves as our background is the starting point to properly normalize it. For this purpose a χ^2-fit on the *missing energy* distribution of data, selected without any *missing energy* requirement, is performed. The MC template is used to describe the *W-like* events, while the *anti-electrons* samlple is used as *non-W* events templates; the corresponding fraction to the total are variated in the fit to minimize the binned χ^2, using MINUIT, neglecting statistical uncertainties in templates.

The fits ar shown here for the inclusive W and for $W + \mathcal{N}_|$ samples ($\mathcal{N}_| = 2, 3$), for

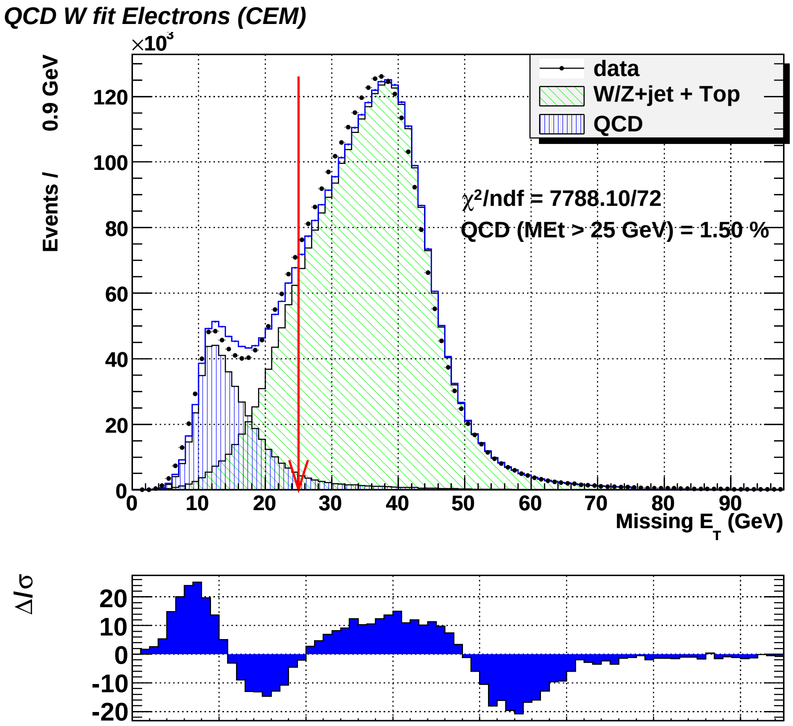


Figure 6.1: electrons QCD background estimation by the fit in *missing energy* distribution for the inclusive W sample ($\mathcal{N}_j \geqslant 0$). All the analysis cuts are applied

the *pretag* events and each tagging cathegory. As a further precaution we applied all the selection cuts of our analysis on M_T^W, $\Delta\phi$ described in Ch. **??**, to the templates before performing the fit, in order to increase the confidence that the kinematics of the QCD background is well reproduced.

The fit have been performed for the inclusive W sample ($\mathcal{N}_| \geqslant 0$), Fig 6.1, and for the $W + 2jets$ sample, Fig 6.2, The 3 JETS bin multiplicity is shown in the Appendix A just for the sake of completeness, because of its small statistic and the fact that it is not going to be used for our analysis.

The fit on the inclusive W sample (Fig 6.1) show a systematic disagreement between the shapes used in the fit and the data sample, confirmed by the bad $\frac{\chi^2}{ndf}$ and the residual trend shown in the same figure of the fit. This not so small disagreement was also in the analysis where the antielectron method was proposed and tested ([38]), The result on QCD fraction in that work is 1.60 ± 0.07, in agreement with our result of 1.50 ± 0.02, where the uncertainty is the combination of the statistical estimated from the fit and the systematic calculated in [38].

kinematics	central	fiducial	E_T	P_T	Z_0	E/P	Isolation
ID	Had/Em	χ^2	L_{shr}	CES ΔX	CES ΔZ		

Table 6.1: "identification" and "kinematic" requirements on *tight electrons* at CDF. The QCD template of *anti-electrons* method is obtained by reversing the cut on a number of electrons identification variables.

Otherwise the *pretag* and 1 TIGHT TAG $W + 2jets$ samples, Fig 6.2, show a better agreement in shape, proving that the *anti-electrons* sample is much more suitable in this case for the QCD background estimation. This is confirmed by a better $\frac{\chi^2}{ndf}$ of $\frac{215}{72}$ and $\frac{88}{67}$ respectively.

6.1.2 *non-isolated muons* sample

As far as muons is concerned, the QCD enriched sample used to describe our data is composed of events in which a candidate W includes a *non-isolated muons* defined as a muon with more than 0.2 of isolation:

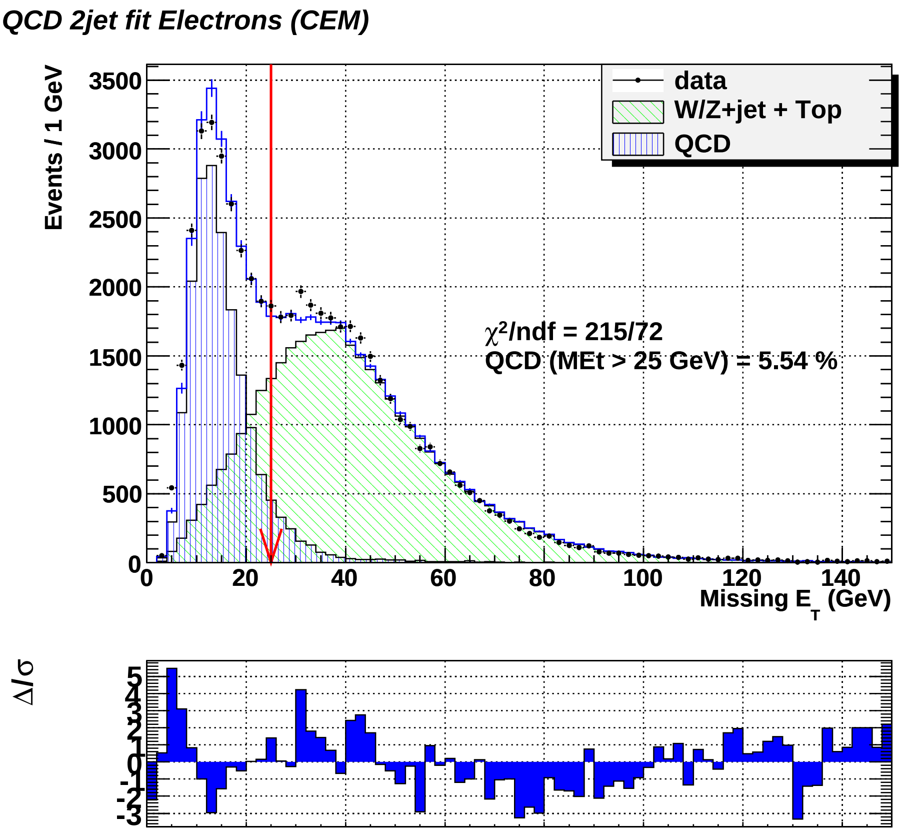
$$I = \frac{1}{P_T^\mu} \sum_{within\ \Delta R<0.4} E_T \; - E_T^\mu > 0.2$$

The W-like events shape is estimated from Monte Carlo simulation.

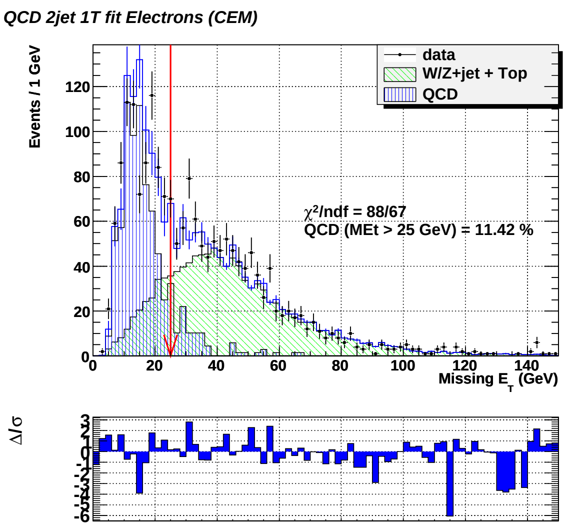
All other requirements on the muons are applied, in order to make this sample kinematically similar to our QCD background. As in the case of electrons (Sec 6.1.1), this sample is used to fit the *missing energy* distribution of data, with MC template used for the *W-like* events description. CMUP and CMX muons have been analysed separately.

For the sake of completeness we show in Figure A.2 the fit performed also for the 3 JETS bin multiplicity.

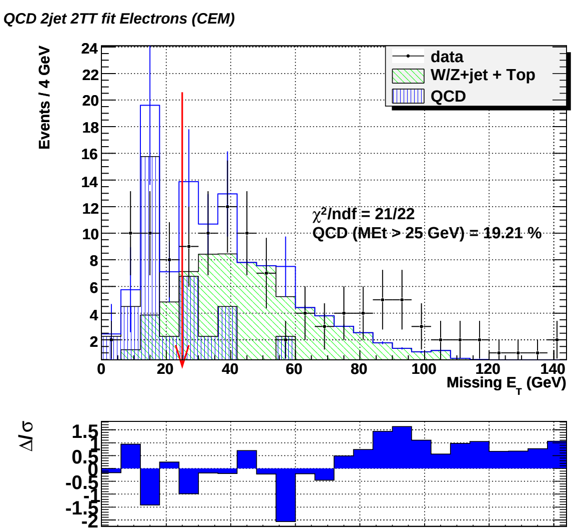
The fit on the inclusive W sample shown in Figure 6.3 presents a systematic difference in shape between the templates used for the the fit on $\not{E}_T$ distribution and the data, as in the electron sample. However, the residuals study show in this case a smaller difference. In the *pretag* and tagged $W + 2jets$ sample it's straightforward to notice that there is a smaller contamination of QCD events in this sample, compared to the electrons' one. The CMX 2TT sample suffers of very low statistics. For this reason, having observed that the QCD fraction of the muon tagged samples is compatible between CMX and CMUP muons, we decided to merge the two samples for the 2TT cathegory in order to improve our sensitivity to the QCD 2TT muons fraction (Figure 6.4(e)).



(a) *pretag*



(b) 1T



(c) 2TT

Figure 6.2: Electrons QCD background estimation by the fit in *missing energy* distribution for the *pretag*, 1 TIGHT TAG and 1TIGHT TAG samples, with only 2 selected jets in the event. All the analysis cuts are applied

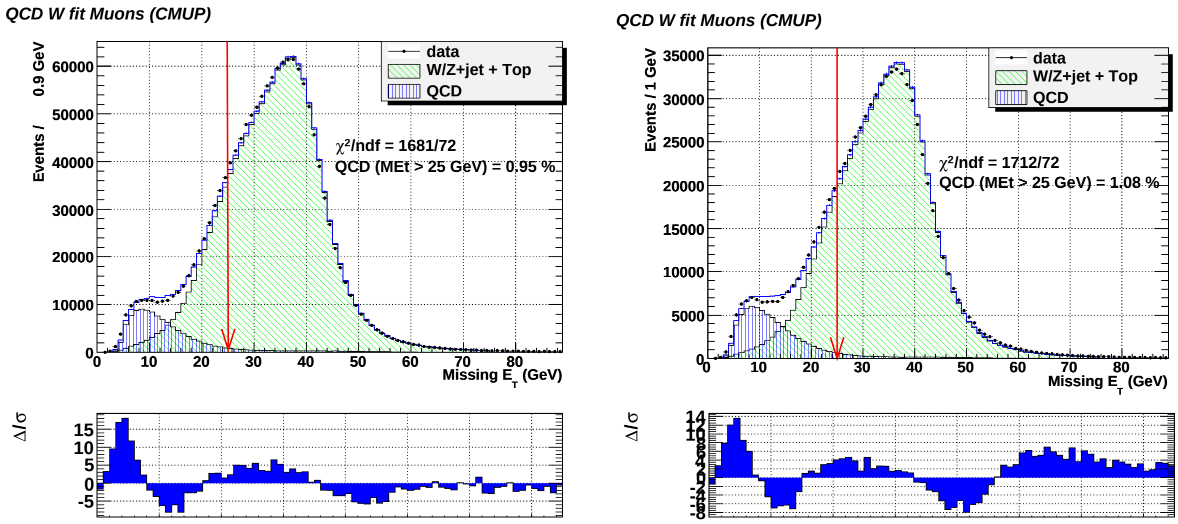
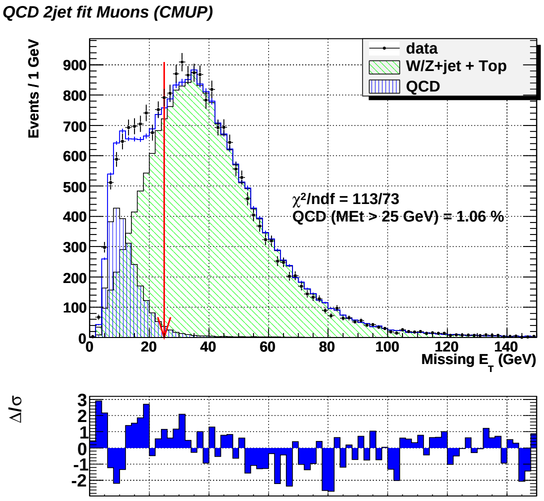


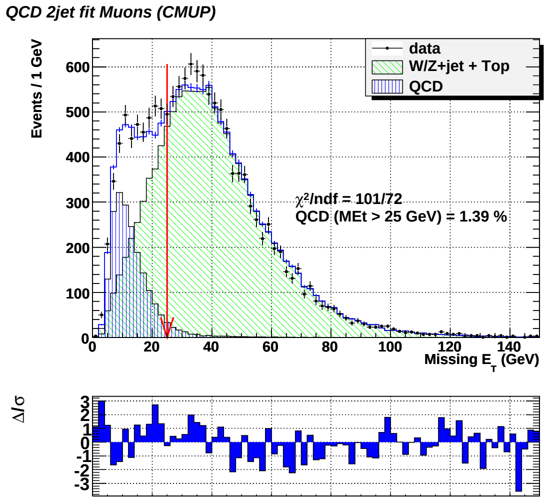
Figure 6.3: Muons QCD background estimation by the fit in *missing energy* distribution for the inclusive W sample ($\mathcal{N}_j \geqslant 0$). All the analysis cuts are applied

6.1.3 QCD fractions results

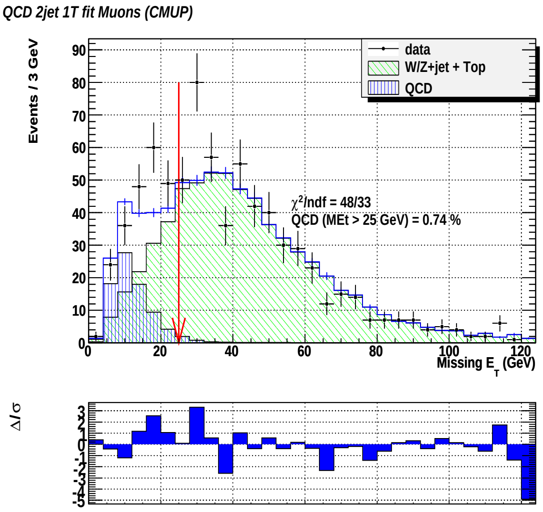
We report in Table 6.2 the QCD fractions resulting from the fit, with their statistical error, in the whole *missing energy* range. In Table 6.3 we show the QCD fraction for $\not\!\!E_T > 25$ GeV. The systematic uncertainty of these fractions is estimated performing the fit with different binning and $\not\!\!E_T$ranges and taking as systematic error the largest deviation to the fit results ([2]). Our systematic uncertainty is 25%. This means that systematics is the larger uncertainty for the QCD fraction for *pretag* and 1T sample. Otherwise the statistical uncertainty for 2TT sample is 40% for both muons and electrons, therefore we associate to this sample a 50% of overall uncertanty, combining, as indipendent, the systematic and the fit uncertanties. In the di-jet invariant mass distributio, the QCD template of each cathegory used in this analysis is normalized to the fraction of Tab. 6.3 multiplied for the integral of the data of that cathegory. The di-jet invariant mass template used is again the *anti-electrons* and *non-isolated muons* sample for electrons and muons respectively, $\not\!\!E_T$ cut applied. The only addition is that, for the evident low statistic of the tagged samples, *Mistag Matrix* has been applied to the *pretag anti-electrons* and *non-isolated muons* samples, to account their contribution to the tagged cathegories, in order to obtain a smoother template for QCD.



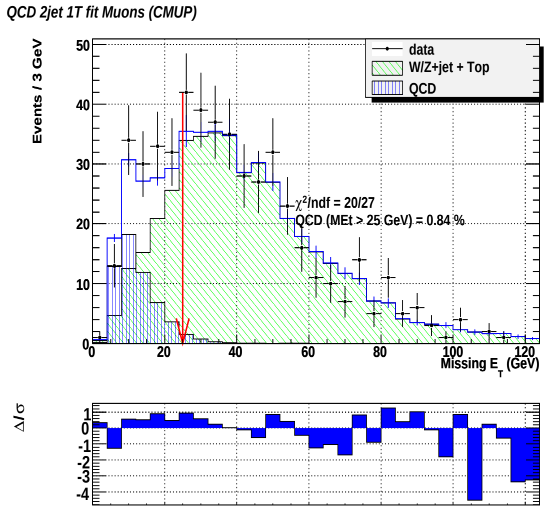
(a) *pretag* CMUP



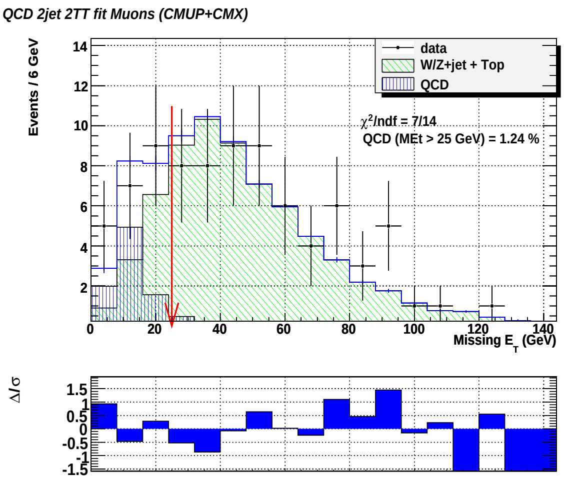
(b) *pretag* CMX



(c) 1T CMUP



(d) 1T CMX



(e) 2TT CMUP+CMX

Figure 6.4: Muons QCD background estimation by the fit in *missing energy* distribution for the *pretag*, 1 TIGHT TAGand 1TIGHT TAG samples, with only 2 selected jets in the event. All the analysis cuts are applied

2 JETS			
pretag CEM	0.3341	±	0.0036
pretag CMUP	0.1183	±	0.0036
pretag CMX	0.1397	±	0.0046
1T CEM	0.412	±	0.022
1T CMUP	0.112	±	0.017
1T CMX	0.114	±	0.022
2TT CEM	0.35	±	0.12
2TT CMUP	0.199	±	0.077
2TT CMX	0.12	±	0.17
2TT CMUP+CMX	0.12	±	0.05

Table 6.2: The QCD fractions resulting from the fit, with their statistical error, in the whole *missing energy* range, for 2 and 3 JETS events.

2 JETS							
pretag CEM	5.54%	±	0.05	(stat)	±	1.4	(syst)
pretag CMUP	1.1%	±	0.3	(stat)	±	0.3	(syst)
pretag CMX	1.39%	±	0.04	(stat)	±	0.4	(syst)
1T CEM	11.4%	±	0.6	(stat)	±	2.9	(syst)
1T CMUP	0.7%	±	0.1	(stat)	±	0.2	(syst)
1T CMX	0.8%	±	0.2	(stat)	±	0.2	(syst)
2TT CEM	9 %	±	4	(stat)	±	2	(syst)
2TT CMUP+CMX	1.2%	±	0.5	(stat)	±	0.3	(syst)

Table 6.3: The QCD fractions resulting from the fit calculated for $\not\!\!E_T \geqslant 25$ GeV, for 2 and 3 JETS events. The systematic uncertainty is 25%

6.2 The *Method 2*[2] for W+jets background normalization

The $W + \mathcal{N}_j$ production ($\mathcal{N}_j \geqslant 0$) is still a not well understood background.

The least understood background process involved in this analysis is the $W + \mathcal{N}_j$ production ($\mathcal{N}_j \geqslant 0$). This process is simulated by ALPGEN ([1]) as far as the matrix element interaction is concerned, and PYTHIA ([39]) for the parton showering. Although many steps forward has been done in QCD processes simulation and parametrization, the MC simulation are not completely trustable for the $W + \mathcal{N}_j$ production In fact, as we can see from Figure 6.5, there is a factor of 1.4 between the experimental measurement and the ALPGEN MC simulation, that tough is applied to our sample, is not enough to obtain a good agreement between data and MC.In addition, the experimental measurement for

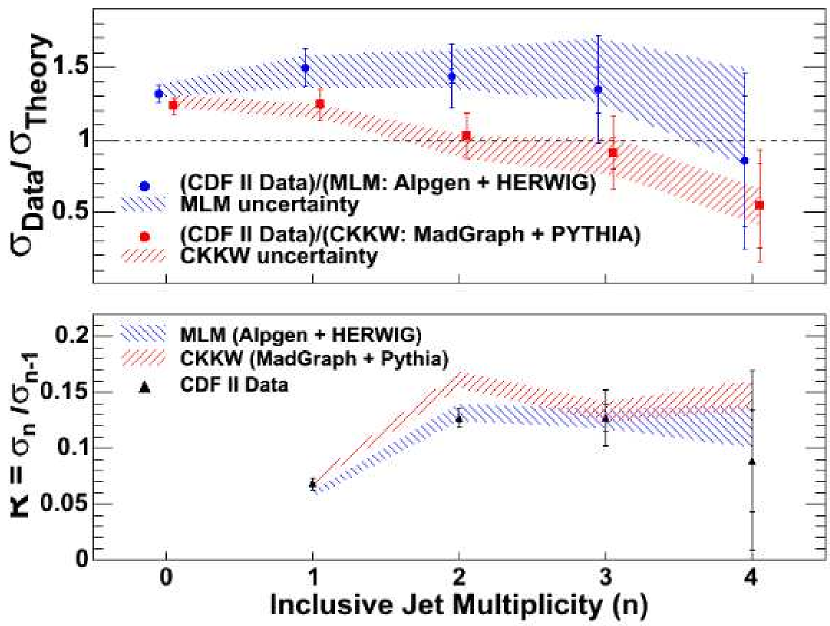


Figure 6.5: Ratio of the meaured inclusive cross section of W+jets processes to the Monte Carlo prediction as a function of jet multiplicity [42]

W+HF is not fully trustable, since the small statistic used for the measurement.

For all these reasons, althernative methods have been developing in CDF to estimate the $W + \mathcal{N}_j$ normalization when SecVtx tagging is involved. The method applied in this analysis is called *Method 2* and is described in [2]. It has widely been applied for many published results such as top pair and single top cross section measurements ([33], [43], [44]) and Higgs ([45]) and other di-boson ([10]) searches.

The main feature of this *Method 2* is to be strongly data driven, since both the total *pretag* $W + \mathcal{N}_j$ normalization and the correction to the heavy flavour fraction of MC are estimated from data. The method consist of assuming that all background processes contributing to our *pretag* sample are known with sufficient precision that will allow calculate the normalization of each sample and subtract it from the *pretag* dataset. At the end of this procedure, the total of remaining data is the value to which W+LIGHT-FLAVOUR sample is normalized. Thisi is the starting point to normalize also the tagged sample. The *Mistag Matrix* (Sec. 6.2.2) is used to evaluate W+LF mistagging and tagging efficiencies for W+HEAVY-FLAVOUR processes are taken from MC and corrected with a *k-factor* that is estimated by data ([41])

6.2.1 Description of the *Method 2*

The search method for this analysis is to look for a resonance in the di-jet invariant mass distribution of our dataset. For this purpose we are going to estimate, as described in the previous sections, a template of the invariant mass distribution of each one of our backgrounds and fit their relative fractions to the data. However, we first need to correcly normalize each sample, since we constrain to the predicted fraction each template with a gaussian constrain whose width is proportional to the uncertainty we have on that normalization. This is why we are going to use *Method 2* for estimating $W + \mathcal{N}_j$ normalization both in the *pretag* sample, that we are using as control reference, and in the 1 and 2 TIGHT TAGs samples, on which we perform our measurement.

The first step of the *Method 2* is to take from MC simulation electroweak, top pair and single top processes and normalize them using the theoretical cross section, the luminosity and the MC derived efficiencies and acceptances [1]. We do this normalization for both tag and *pretag* samples, as in equation (6.1)

$$\mathcal{N}_{p\bar{p}\to X} = \sigma_{p\bar{p}\to X} \cdot \epsilon \cdot \alpha \cdot \mathcal{L} \cdot (\epsilon_{tag} \cdot S_f) \tag{6.1}$$

where $\mathcal{N}_{p\bar{p}\to X}$ is the number of events to which each template is normalized, σ is the (theoretical or experimental) cross section of that process, ϵ is the efficiency and α is the acceptance; $\mathcal{L}$ is the integrated luminosity of the dataset.

TAGGING	Scale Factor (S_f)
1T	0.95 $\pm$ 0.04
2TT	0.90 $\pm$ 0.06
2TL	0.94 $\pm$ 0.06
1L	0.99 $\pm$ 0.05
2LL	0.98 $\pm$ 0.07

Table 6.4: Scale Factors for SecVtx LOOSE and TIGHT tags algorithms

ϵ_{tag} and S_f factors are applied only in case of a tagged sample. They are the tagging efficiency (MC derived) and the SecVtx scale factor. The latter is the factor for which the

[1] Efficiency, ϵ, is defined as the ratio between the events that fulfill our requirements and the total of generated events. Acceptance, α, is the ratio between the events that are revealed by our detector (for its geometry) and the total of generated events.

MC have to be corrected to compensate for a slight overestimation of the tagging rate in simulation.

The scale factors used in this analysis are shown in Table 6.4 for each tagging cathegory of SecVtX and have been calculated in [46] for the TIGHT and LOOSE taggings. These values have been combined to account for 2 tags requirements in our samples:

$$Sf_{2TT} = Sf_{1T}^2$$
$$Sf_{2LL} = Sf_{1L}^2$$
$$Sf_{2TL} = Sf_{1L} \cdot Sf_{1T}$$

The QCD (*non-W*) fraction estimation is the next step of this method. The template is obtained from *anti-electrons* and *non-isolated muons* samples and normalized with the fractions estimated in Sec. 6.1:

$$\mathcal{N}_{QCD} = \mathcal{N}_{data} \cdot \mathcal{F}_{QCD} \tag{6.2}$$

Electroweak, top and QCD contributions are subtracted from the total of *pretag* events, obtaining, directly from the dataset, the total of $W + \mathcal{N}_j$ events of our sample. Then the heavy flavour content of this *pretag* sample needs to be estimated. We define as "heavy flavour event" each event where a Q ora a pair of Q is found in one or both leading jets, $Q = b, \bar{b}, c, \bar{c}$. The "heavy flavour samples" considered are

I. $W + b\bar{b}$

II. $W + c\bar{c}$

III. $W + c$

IV. The heavy flavour contribution of the W+LF sample, i.e. heavy flavours produced in the parton showering simulated by PYTHIA in that sample, which was not completely removed by the "jet based overlap removal" (Sec. 4.2). This "heavy flavour from light flavour" template has been produced relying on MC truth information.

The heavy flavour content ($\mathcal{F}^{HF}$) of each W+$\mathcal{N}_j$ sample, that pass our analysis cuts, has been calculated using the MC truth. We show the fraction that we have obtained in Table 6.5, for all jet multiplicities, that are in good agreement with the ones calculated

(a)

Electrons					Muons				
$\mathcal{N}_j$	1B	2B	1C	2C	$\mathcal{N}_j$	1B	2B	1C	2C
2 JETS	0.013	0.0081	0.085	0.013	2 JETS	0.014	0.008	0.084	0.013
3 JETS	0.022	0.0061	0.095	0.012	3 JETS	0.026	0.0062	0.095	0.012
$\geqslant$ 4JET	0.026	0.0042	0.098	0.012	$\geqslant$ 4JET	0.037	0.0058	0.11	0.015

(b)

Sample	2 JETS	3 JETS
$W + b\bar{b}$	0.019	0.028
$W + c$	0.050	0.043
$W + c\bar{c}$	0.033	0.0485
W+LF	0.016	0.0195

Table 6.5: (a) Heavy Flavour fractions estimated by MC truth for all jet multiplicities (b) Heavy Flavour fraction for each sample, for both electrons and muons are assumed the same fractions

in [41]. In Tab. 6.5 (b) are shown the fractions for each heavy flavour sample. Since the differences between muons and electrons can be ascribed to statistics, the final heavy flavour fraction are calculated mediating on the two samples. These fraction needs to be multiplied for a factor that corrects them to better describe the data. This *k-factor* has been calculated in [41] and been found to value $k = 1.0 \pm 0.3$ and have been calibrated from the comparison of data and MC simulation in a generic QCD sample: the uncertainty takes into account the extrapolation of the information to the $W + \mathcal{N}_j$ sample. Actually this factor do not modifies the MC derived fractions, but introduces a lange uncertainty in the estimation of the heavy flavour content of our sample.

Then we have, for the *pretag* sample:

$$\mathcal{N}_{W+\mathcal{N}_j} = \mathcal{N}_{data} - (\mathcal{N}_{QCD} - \mathcal{N}_{t\bar{t}} - \mathcal{N}_{ST} - \mathcal{N}_{ew}) \tag{6.3a}$$

$$\mathcal{N}_{W+b\bar{b}} = \mathcal{N}_{W+\mathcal{N}_j} \cdot \mathcal{F}^{HF}_{Wbb} \cdot k \tag{6.3b}$$

$$\mathcal{N}_{W+c\bar{c}} = \mathcal{N}_{W+\mathcal{N}_j} \cdot \mathcal{F}^{HF}_{Wcc} \cdot k \tag{6.3c}$$

$$\mathcal{N}_{W+c} = \mathcal{N}_{W+\mathcal{N}_j} \cdot \mathcal{F}^{HF}_{Wc} \cdot k \tag{6.3d}$$

$$\mathcal{N}_{W+shower\ hf} = \mathcal{N}_{W+\mathcal{N}_j} \cdot \mathcal{F}^{HF}_{Wlf} \cdot k \tag{6.3e}$$

The Equations (6.3), that show how to calculate the normalization of W+LF and W+HF sample according to *Method 2*, can be turned into the normalization for the tagged samples just by multiplying each sample for the MC derived ϵ_{tag} (and the SecVtx S_f) as in Equations (6.4),

$$\mathcal{N}^{1T}_{W+QQ} = \mathcal{N}^{pretag}_{W+QQ} \cdot \epsilon^{1T}_{tag} \cdot S^{1T}_{f} \tag{6.4a}$$

$$\mathcal{N}^{2TT}_{W+QQ} = \mathcal{N}^{pretag}_{W+QQ} \cdot \epsilon^{2TT}_{tag} \cdot S^{2TT}_{f} \tag{6.4b}$$

where QQ stands for $b\bar{b}$, $c\bar{c}$, c and LF.

The last background process that still needs to be taken into account is contribution of the W+LFmistags. This process, infact, do not involve any bottom or charm quarks in the event, so enters the tagged sample only for a secondary vertex mistakenly reconstructed when poorly reconstructed tracks produce a fake vertex away from the origin. The MC poorly describes the mistagging rate in the W+LF sample, so a *Mistag Matrix* has been developed at CDF to derive from a LF-jet's kinematic the probability of being mistagged.

6.2.2 The Mistag Matrix [3]

The SecVtx tag algorithms relyy on the identification of vertex reconstructed from the tracks within a jet which is significantly displaced with respect to an event's primary vertex. As traking is a complex phenomenon, difficulties in its modelling lead to tagging efficiencies generally being inaccurate in Monte Carlo models.

In a high statistics sample, with both tags and mistags, a resonable estimate of the mistag rate is given by the negative tag rate [2] Unfortunately this is not our case, since our is a modestly sized sample in which the number of NEGATIVE TAGs is to small to give a reliable estimate of the mistag rate in the POSITIVE TAG sample.

The *Mistag Matrix* provides a way to extrapolate into our sample the average mistag rate, i.e. the rate of light flavour jet that are identified as b-jets, measured in very large

[2]a NEGATIVE TAGis defined when the displaced vertex is reconstructed behind the primary vertex, taking as positive the jet direction. For a more exhaustive explanation see Sec. 3.4

inclusive jet sample. In this sample the negative mistag rate has been measured and parametrized as a function of six kinematic variables:

I. jet E_T [GeV]: transverse energy

II. $\mathcal{N}_{trks}$: number of tracks per jet

III. $|\eta_{jet}|$: pseudorapidity

IV. n_{PV}: number of Primary Vertexes (PVs)

V. $\sum E_T$: of all jets in the event

VI. Z_{PV}: position of the PV on the z-axis

Then the NEGATIVE TAG rate is converted into an estimate POSITIVE MISTAG rate by applying two factors: the first one, known as α, is used to correct for the asimmetry between positive and negative taggings and the second one, β, is to take into account the fact that the *Mistag Matrix* was built on an inclusive jet sample that is dominated by LF but contains a small contamination of HF jets. After this procedure, it is possible to extract from *Mistag Matrix* the probability, for each jet in the event, of being a LF jet identified as a b-jet. We compute the *"total mistag probability of the event"* combining the tag probabilities of the 2 leading jets, with the formulas shown in Table 6.6 and this value was used as a weight for the event filling the appropriate histogram of invariant mass. TIn those formulas, $j1$ and $j2$ are the leading and second leading jets of the event, *loose* and *tight* stands for the two SecVtx algorithms. Therefore $P(j2 : loose\&\&notTight)$, for example, is the probability that the second leading jet is tagged from the LOOSE algorithm but not from the TIGHT one.

6.2.2.1 The W+lf template and heavy flavours corrections

The *Mistag Matrix* has been used to obtain a data-driven estimate of the mistagging in the 1 and 2 TIGHT TAGs samples. Starting from a simple interpretation, our *pretag* sample is completely dominated by events that involve a W and 2 light flavour jets. Therefore is reasonable to use the *Mistag Matrix* to give to each event a total probability that the two leading jets would be (mis)tagged, one or both of them, as in Tab. 6.6. Such total

Tag cathegory	Formula
2 TIGHT TAG	$P(j1 : tight) * P(j2 : tight)$
1 LOOSE1 TIGHT TAG	$P(j1 : tight) * P(j2 : loose \ \&\& \ notTight)$
1 TIGHT TAG	$P(j1 : tight) * P(j2 : notLoose) +$ symmetric formula
2 LOOSE TAG	$P(j1 : loose \ \&\& \ notTight) * P(j2 : loose \ \&\& \ notTight)$
1 LOOSE TAG	$P(j1 : loose \ \&\& \ notTight) * P(j2 : notLoose) +$ symmetric formula

Table 6.6: Formulas for the combination of the *tag probability* of light jets obtained by Mistag Matrix of the two leading jets of the event, for each tagging cathegory. $P(j1 : loose)$ means the probability of the leading jet to be loose tagged.

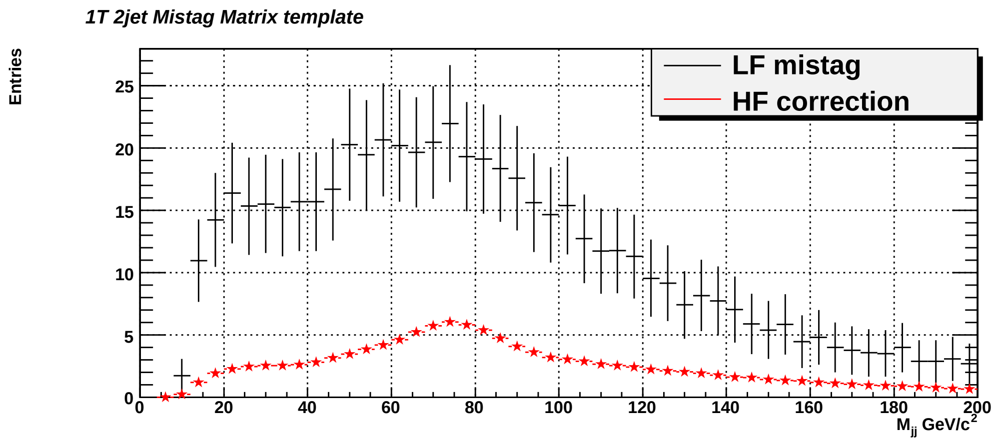


Figure 6.6: Comparison between the W+LF template, obtained applying *Mistag Matrix* to the data sample and the estimated correction for HF content in the data sample, estimated by MC simulation

probability is used to weight the event filling the invariant mass distribution, in order to obtain the light flavour content of the tagged sample, due to mistag.

However, the assumption that the whole *pretag* sample is composed of only light flavour jets is not correct, since we know that there is an amount of background processes, such as top and W+HF, that become relevant if we require tagging and are contaminating our estimation of the mistag template. Hence we have estimated from MC simulation the effect of the *Mistag Matrix* on the MC samples that involve heavy flavour jets. The template that results from the use of the *Mistag Matrix* to calculate a mistag probability on heavy flavour jets is subtracted from the light flavour template, as a correction to the presence of heavy flavours in the *pretag* sample.

In Figure 6.6 we show the light flavour template estimated by *Mistag Matrix* before corrections and the correction for heavy flavour we subtract to this template. It's important to notice that our signal gives a considerable contribution to the mistag template, if we do not apply this heavy flavour correction, especially in the signal region.

In Figure 6.2.2.1 there is an example of the effect of the usage of Mistag Matrix on the W + jet Monte Carlo sample, after the correction for heavy flavour is applied. The most important effect is the 2 TIGHT TAG cathegory, for which the *Mistag Matrix* permits to obtain a really smoother shape than the one we obtain applying the tagging requirement to the LF Monte Carlo. For 1T cathegory MC reasonably agrees and we take *Mistag Matrix* template to follow a more data driven approach for uniformity with the 2TT sample.

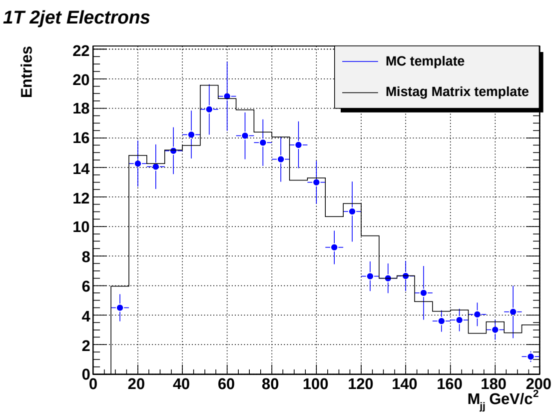
6.2.3 *Method 2* summary

Summary steps of *Method 2*:

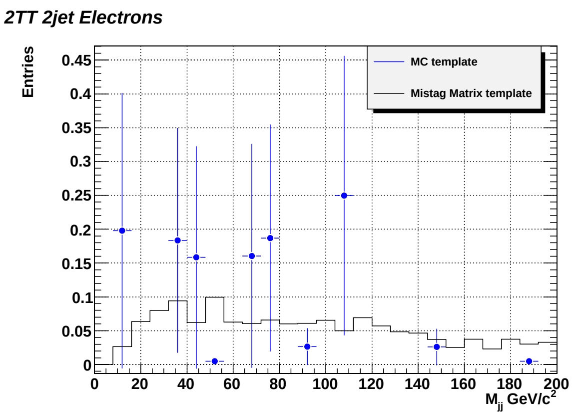
I. the MC templates of all backgrounds except for W+jets are properly normalized for the acceptance, efficiency, luminosity and cross section. The QCD is normalized using the fraction from the Missing Et fit.

II. the following montecarlo templates are subtracted from the data sample:

- Top pairs



(a) 1T



(b) 2TT

Figure 6.7: Comparison between MC template and *Mistag Matrix* template for 1 and 2 TIGHT TAGs samples, 2 jets in the event, for electrons. The correction for heavy flavour has been applied to the template.

- Single Top
- Z+jets
- di-bosons (WW/WZ)
- QCD

for which the montecarlo is considered reliable or has been differently estimated (as QCD). Now we have the number of W+jets pretag directly from data ($\mathcal{N}^{pretag}_{Wjets}$).

III. We calculate from the montecarlo truth the fraction of the W+jets events which contains heavy flavours, that is the fraction of event which have at least one c or b inside one of the two leading jets. This fraction has been calculated separately for W+bb, W+cc, W+c and W+lf samples. The last fraction is used to normalize the "heavy flavour from light flavour" that accounts for the heavy flavour contribution of the W+LF sample.

IV. each W+jets sample is rescaled to the number of W+jets pretag multiplied by the heavy flavour fraction of that sample:

$$\mathcal{N}_{W+q\bar{q}} = \mathcal{N}_{W+\mathcal{N}_j} \cdot \mathcal{F}^{HF}_{Wbb} \cdot k \qquad\qquad (q\bar{q} = b\bar{b},\ c\bar{c},\ c,\ hf\ in\ lf)$$

we will use this number of events for each pretag W + heavy flavour jets sample, obtained in this way, to normalize the tagged samples.

V. Now we have the pretag sample normalized on data.

VI. As far as the tagging cathegory is concerned, each W + heavy flavour jets sample is normalized to the number of the events pretag, prevously estimated, multiplied by the tagging efficiency from montecarlo and the SecVtx appropriate scale factor. For example

$$\mathcal{N}^{q\bar{q}}_{1tag} = \mathcal{N}^{q\bar{q}}_{pretag} \cdot S^{1T}_{f} \cdot \frac{\mathcal{N}^{q\bar{q}}_{mc,1T}}{\mathcal{N}^{q\bar{q}}_{mc}}$$

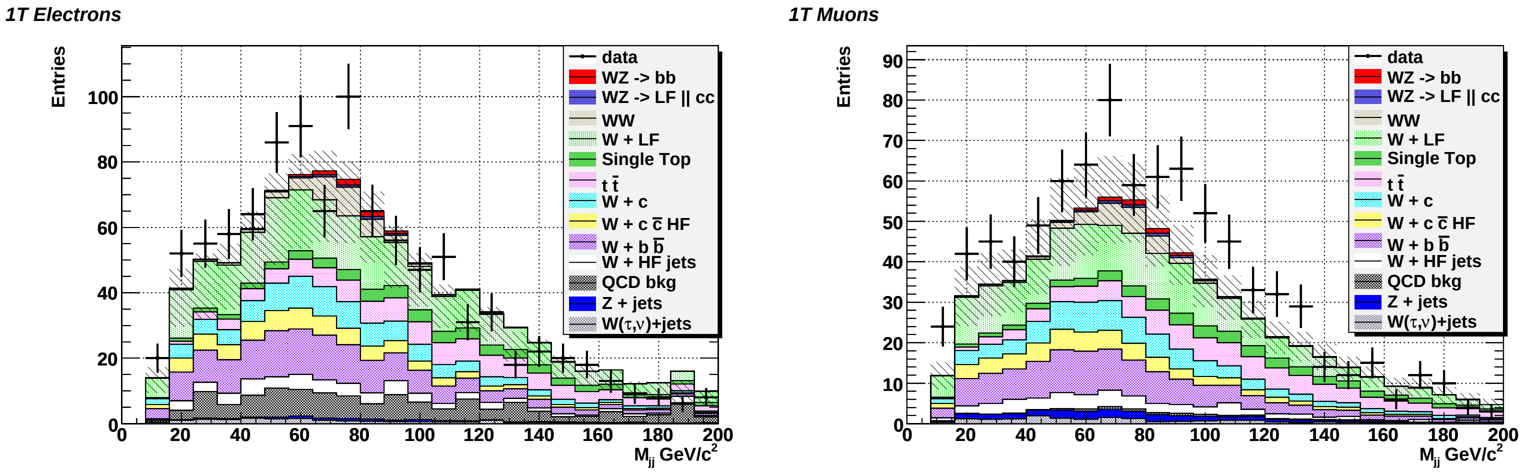
VII. the only sample which is not from montecarlo is the W+LF sample. It is estimated applying the mistag matrix to the pretag data, correcting the fact that the sample contains a small amount of heavy flavours by MC. The sample obtained from mistag matrix do not need any further normalization.

		Electrons				***Muons***			
sample/cathegory		*pretag*	1T	2TT	2TL	*pretag*	1T	2TT	2TL
$W+lf$	(±10%)	23800	251	1.3	3	18500	180	1	2
$W+b\bar{b}$	(±30%)	512	149	22	10	400	113	17	7
$W+c\bar{c}$	(±30%)	889	67	1.4	2	694	53	1	1
$W+c$	(±30%)	1350	95	0.7	0.7	1050	75	0.6	0.7
HF in $W+lf$	(±30%)	431	55	0.5	0.2	337	43	0.01	0.7
top pair	(±6%)	268	103	29	12	220	85	24	10
single top	(±6%)	133	53.3	9	3	105	42	7	3
W^+W^-	(±6%)	865	33.4	0.3	0.3	683	26	0.2	0.2
$WZ \rightarrow l\nu lf$	(±6%)	106	5	0.07	0.1	94	4	0.05	0.06
$WZ \rightarrow l\bar{\nu}_l b\bar{b}$	(±6%)	18	7	2.4	0.8	14	5	2	0.8
Z + jets	(±6%)	335	8	0.3	0.2	1080	22	1	0.6
QCD	(±25%)	1630	114	7	1	654	116	7	1
TOT MC	(∼ ± 15%)	30340	939	74	33	23830	764	61	27
data	(4.9 fb^{-1})	30200	997	84	43	23800	854	55	38

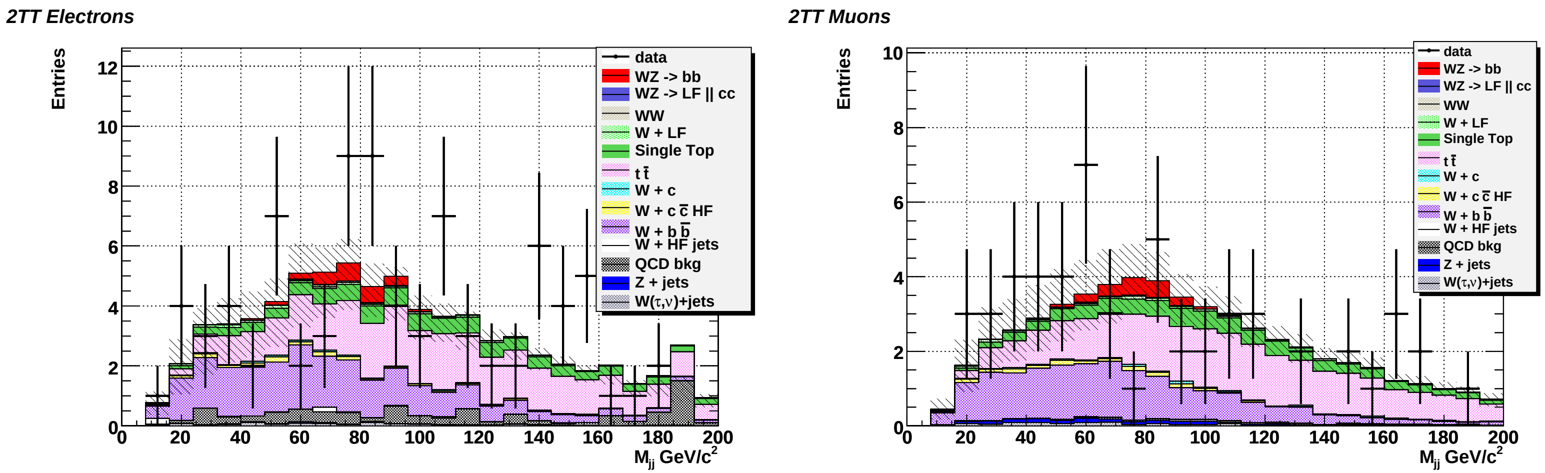
Figure 6.8: Resultant events after *Method 2* application for the 2 JETS bin multiplicity

6.2.4 *Method 2* results

In Table 6.8 we report the resultant events for each sample, divided for jet multiplicity bin and *pretag* and tags samples. In order to make the reader more confident that the *Method 2* has been correcly used, we have compared our results with the ones published by the $WH \rightarrow l\bar{\nu}_l b\bar{b}$ analysis [45], founding a reasonable agreement between them. The small differences are ascrivible to small selection differences. This samples normalization will be the start point for our fit to the inveriant mass distribution that we are going to explain in detail in the next chapter. The resultant di-jet invariant mass distribution are shown in Figures 6.9 for electrons and muons samples. In these figures the systematic bands are drawn, in order to highlight that data and *Method 2*-normalized MCs are compatibles within systematic uncertainties. The fact that systematics are correlated among the bins, justify that all the experimental points in 1T histogram are on the superior margin of the systematic band.



(a) 1T



(b) 2TT

Figure 6.9: Invariant mass distribution normalized by *Method 2*. Systematic bands are drawn in the figure, to take into account the 30% of uncertainty on W+HF normalization and 25% on QCD

Chapter 7

Cross section limit calculation

In this chapter a binned likelihood fit is performed on the di-jet invariant mass distribution of two indipendent samples: one composed of events where only one jet is tagged by SecVtx, and the other where both the leading jets of the event are b-tagged. The fitting tools are preliminary tested on the pretag *sample, to estimate the W^+W^-+$W^\pm Z$ cross section, that is found in agreement with published results. Since no evidence of the $W^\pm Z$ signal is found in the tagged samples, an upper limit is estimated for this process cross section.*

7.1 Tools and statistical procedures

7.1.1 Extended binned likelihood definition

The technique used to estimate the cross section (or an upper limit of that) of the $WZ \rightarrow l\bar{\nu}_l b\bar{b}$ process is a fit of the di-jet invariant mass distribution. The low statistics of the tagged samples suggest us to perform a likelihood fit, that is binned since the background shapes are not simply parametrizable.

The fit parameters are the fractions of the signal and of each background process to the total of the events. The shapes in the di-jet invariant mass distribution of this backgrounds that are used to fit the data histogram are called "templates" and are taken either from MC simulation or from data driven procedures, as described in Chapter 6. We define a likelihood that is the product over the histogram bins of Poisson distributions

(that gives the probability of observing x events in one bin when the expected value μ is our parameter)

$$L = \prod_{bin} P(x|\mu(f_1, f_2, f_3..., f_s)) = \prod_{bin} \frac{e^{-\mu(f_i)}\mu(f_i)^x}{x!} \tag{7.1}$$

where $f_1, f_2, f_3, ...$ are the fractions of the background processes that we indend to estimate from the fit (see next subsection), and f_s is the signal one. The bin expected value $\mu(f_i)$ is the sum of the background and signal events in each bin. The $\log L$, more than the likelihood itself, is used, for mathematical convenience, in the minimization process.

$$\begin{aligned} \log L &= \sum_{bin} -\mu(f_i) \quad +x \cdot \sum_{bin} \log \mu(f_i) - \sum_{bin} \log x! \\ &= \mathcal{N}_{tot} \qquad\quad +x \cdot \sum_{bin} \log \mu(f_i) - \sum_{bin} \log x! \end{aligned} \tag{7.2}$$

The last term of Eq (7.2) do not depend on fitting parameters, but only on experimental data, therefore we are allowed to neglect it in the minimization. $\mathcal{N}_{tot}$ is the overall normalization and explicitely appears in the $\log L$ definition[1]. The most reasonable choice is then to define it as a fit parameter, in addition to $n-1$ fractions of the n considered backgrounds and signal templates. The nth fraction is obtained imposing that $\sum_i f_i = 1$.

7.1.2 Gaussian constraints

The normalization of the backgrounds templates that we are going to variate in the fit as parameters are all known with a certain theretical or experimental precision. It is possible to use this information to increase the signal sensitivity, implementing a constrain to the parameters that estimate the fraction of these backgrounds.
The likelihood and its logarithm are then defined as

$$L' = L \cdot \prod_i g_i$$

$$\log L' = \log L + \sum_i \log g_i$$

where g_i is the gaussian constrain to the ith background and

$$\log g_i = -\frac{(f_i - f_i^c)^2}{{\sigma_i}^2} \tag{7.3}$$

[1]such defined likelihood is known as “extended”, since the overall normalization is allowed to fluctuate

The width σ_i of the constrain is the uncertanty on the considered parameter and the central value f_i^c is obtained from the normalization of that particular process, either by MC or with one of the data-driven methods discussed in Ch. 6.

7.1.3 Procedure for computing limits

Since we do not expect to have the sensitivity to observe a Standard Model signal, we have to define the procedure followed to set an upper limit to the cross section of our signal ([47]). The starting point for the definition of frequentist confidence limits is the definition a test statistic, that is a single real number which is a function of the experimental outcome. It is chosen to maximize the separation between the outcomes expected when a signal is present and those expected when there is only background contribution. Usually the optimal choice for this test statistic is the likelihood ratio ([48])

$$Q = \frac{P(data|\mathcal{H}_1)}{P(data|\mathcal{H}_0)} \tag{7.4}$$

where $\mathcal{H}_1$ is a model including the signal we intend to exclude, and $\mathcal{H}_0$ is the null, only background, hypothesis. In practical terms Q has to be calculated as the ratio between the likelihood obtained performing a fit on the dataset, with background and signal templates, and the likelihood obtained performing the fit without the signal template.

In terms of the more intuitive χ^2 function, that is related to the likelihood from the equation

$$\chi^2 = -2 \log L \tag{7.5}$$

this is equivalent to

$$2 \log Q_{data} = \chi^2(data|\mathcal{H}_0) - \chi^2(data|\mathcal{H}_1) = \Delta\chi^2 \tag{7.6}$$

Such $\Delta\chi^2$ is positive defined and express how better the signal hypothesis describes the data, compared to the background only one.

In order to set a confidence limit, an appropriate[2] number of pseudo-experiments are generated via Monte Carlo. Each pseudo-experiment is randomly generated with a signal

[2]appropriate, in comparison to what we have to exclude: for a probability of 10^-2, one thousand pseudo-experiment is enough, though in the more general case of a 5σ exclusion, a larger number of pseudo-experiments is needed.

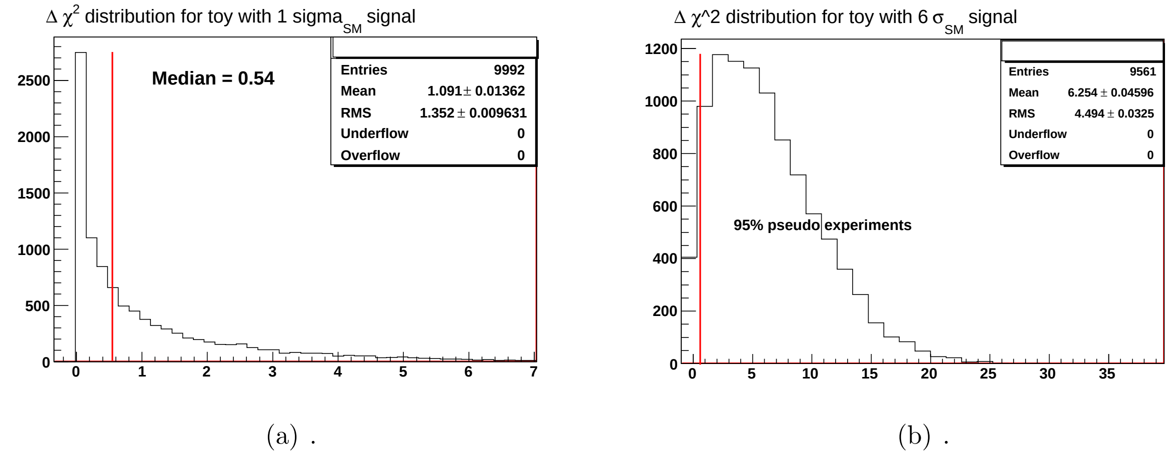


Figure 7.1: $\Delta\chi^2$ distribution based on the 2TT sample, electrons. On the left the pseudo-experiments have been generated with a Standard Model signal. On the right the signal was generated six times bigger.

hypothesis, from the fit templates, and the total number of events is fluctuated according to a Poisson distribution whose expected value is the dataset total number of events.

The likelihood ratio Q is calculated for each pseudo-experiment, and the distribution of $2\log Q = \Delta\chi^2$ of all the generated samples is obtained. As an example, we show the distribution obtained for 10000 pseudo-experiments of the 2TT sample[3] in Figure 7.1(a). In order to exclude the signal hypothesis, at a certain Confidence Level (CL), we require

$$CL = P_{\mathcal{H}1}(Q \leqslant Q_{data}) = P_{\mathcal{H}1}(\Delta\chi^2 \geqslant \Delta\chi^2_{data}) \tag{7.7}$$

which means to integrate the $\Delta\chi^2$ normalized distribution from 0 to $\Delta\chi^2_{data}$ and require that this integral is equal to $1 - C.L.$. Setting a limit on a certain value of the signal cross section, at a certain Confidence Level (CL), for example 95%, means to find that value of signal cross section for which 95% of pseudo-experiments have a $\Delta\chi^2$ higher than what we measured for data. This means that in case of such signal cross section, there is just 0.05% of possibility that the measured $\Delta\chi^2_{data}$ is just a background fluctuation. In practice, we need to build the $\Delta\chi^2$ distribution of pseudo-experiments varying the cross section of the signal hypothesis until the CL defined in Eq (7.7) is equal to the value we have chosen as a reference (usually 0.95).

[3]The tagged sample in which both the jets are identified as b-jets, i.e. tagged.

7.1.4 Expected limit calculation

The typical way to evaluate the sensitivity of different fit variants (that will be described in Sec 7.3.1) is to calculate the expected upper limits on the signal cross section; furthermore we can also verify the consistency of our final result with an expected upper limit. Pseudo-experiments, as described in the precedent section, have been used for this purpose. A significative number of pseudo-experiments is generated with the standard model signal hypothesis, and the median of the $\Delta\chi^2$ distribution of this ensamble is taken as the "expected" $\Delta\chi^2$. The procedure of the limit computation is then applied, using this $\Delta\chi^2_{MC}$ instead of the $\Delta\chi^2_{data}$ estimated from data. When 95% of the pseudo-experiments have an higher $\Delta\chi^2$ than the "expected" one, we have found the expected cross section upper limit. An example of two $\Delta\chi^2$ distribution, for Standard Model signal hypothesis and a signal cross section hypothesis six times greather than the Standard Model value are shown in Figures 7.1(a) and 7.1(b) respectively. In the first figure it's shown the median of the distribution, which is drawn in the second figure too, in order to highlight the trend of the $\Delta\chi^2$ distribution when varying the signal cross section.

Last but not least, confidence bands (1σ-range) on the expected limit have to be estimated, in order to compare it with the final result. For this purpose, we consider the $\Delta\chi^2$ values of the SM signal hypothesis distribution for which the normalized integral is 16% ($\Delta\chi^2_{-1\sigma}$) and 84% ($\Delta\chi^2_{+1\sigma}$) respectively. We reiterate the limit computation procedure using these two $\Delta\chi^2$ values instead of the $\Delta\chi^2_{MC}$ (median of the ditribution) and take the difference between the limits as an expected limit 1σ-range.

7.2 Fitter and templates validation on *pretag* data

We use the *pretag* sample for the validation of our fitting tool, analysis and templates. In fact, the W^+W^-+$W^\pm Z$ signal is evident in this sample and has already been measured, as a resonance in the jet invariant mass distribution, in [4]. The W^+W^-+$W^\pm Z$ yield will be estimated from the *pretag* sample and the results will be compared to the published ones, in order to confirm the correctness of our analysis procedure.

The fit will be performed on the sample where the requirement of $P_T^j > 40$ GeV, described in Section 5.6, is applied, as done in [4]. In addition the sample with 2 and 3 jets will be merged, in order to have more similarity with the above mentioned analysis.

The fit free parameters are:

- W + jets
- top background
- QCD

The Z+jet sample will be fixed to its standard model value, being a small fraction of the total, to limit the number of free parameters in the fit. The top and the QCD templates have been constrained with a gaussian of width equal to 10% and 25% respectively of their initial normalization value. The former uncertainty has been taken from the experimental measurement of top processes; the latter is due to the systematic uncertanty on QCD normalization described in Sec. 6.1.

The fit results are shown in Figure 7.2 for electrons and muon separately. The events of WW+WZ signal estimated are shown in Table 7.1, where are compared with the standard model expected values.

In order to analyse the statistical properties of the fits, *pulls* have been calculated for each of them as in Equation (7.8).

$$P = \frac{\mathcal{N}_{exp} - \mathcal{N}_{fit}}{\epsilon_{fit}} \tag{7.8}$$

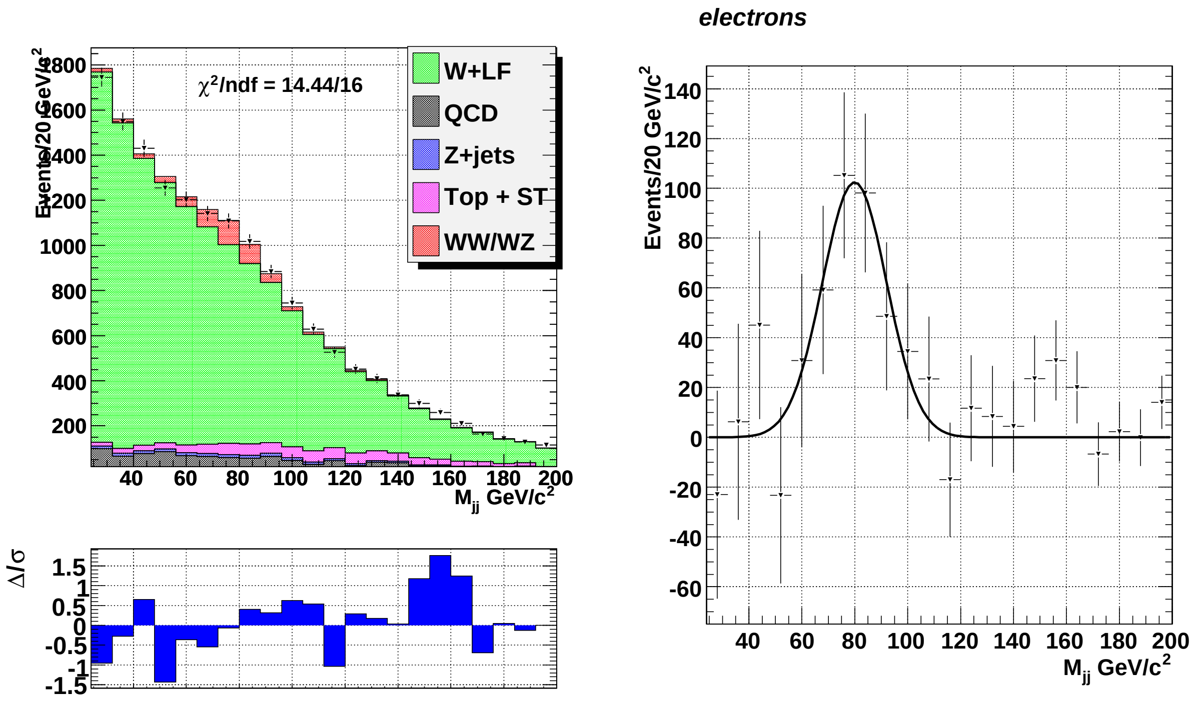
where $\mathcal{N}_{exp}$ is the number of signal expected events, i.e. the signal hypothesis of the pseudo-experiments, $\mathcal{N}_{fit}$ are the signal events exstimated by the fit and ϵ_{fit} the associated uncertainty. If the P distribution is a gaussian with null mean and unitary variance, means that our fits are not biased and the errors are well estimated.

The *pulls* in Figure 7.3 show a gaussian distribution with expected value zero and unitary variance, demostrating that the fit procedure is not biased and errors are well estimated. The resultant cross section for electrons is

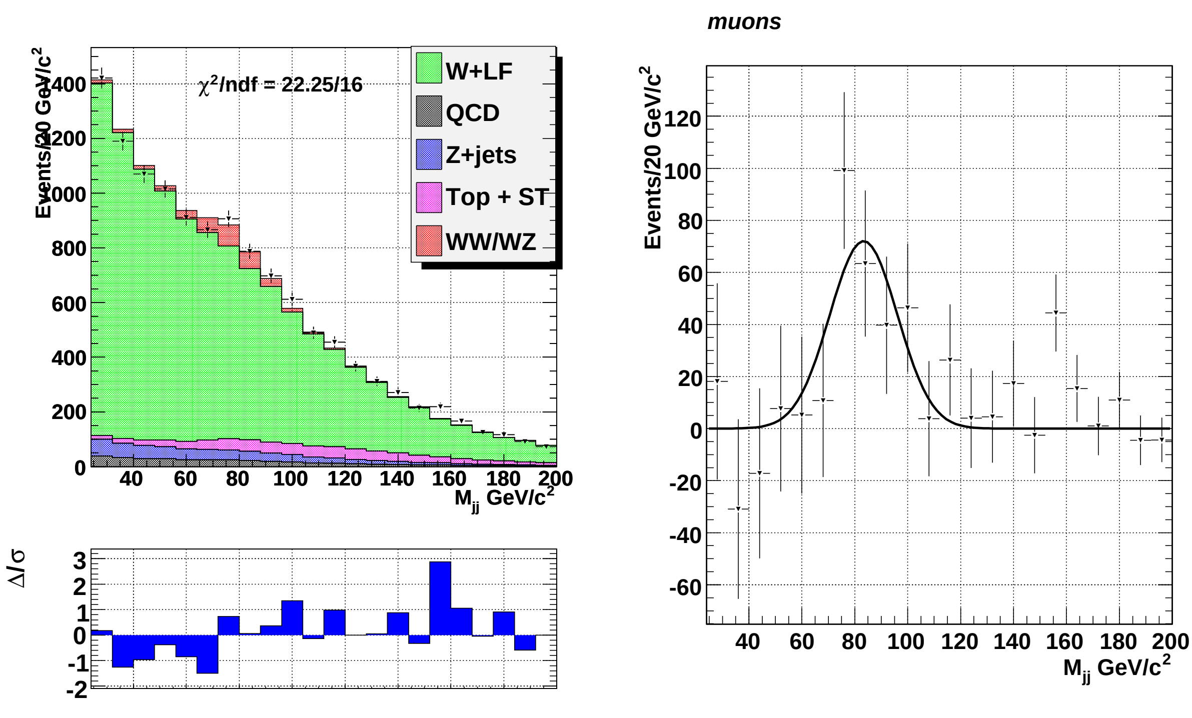
$$\sigma^{e}_{WW/WZ} = 13.0 \pm 3.5(stat) \pm 0.8(lumi) \tag{7.9}$$

that is in good agreement with the published result ([4]) of

$$\sigma^{e}_{WW/WZ} = 13.5 \pm 4.4(stat) \pm 0.8(lumi) \pm 1.7(syst) \tag{7.10}$$



(a) electrons

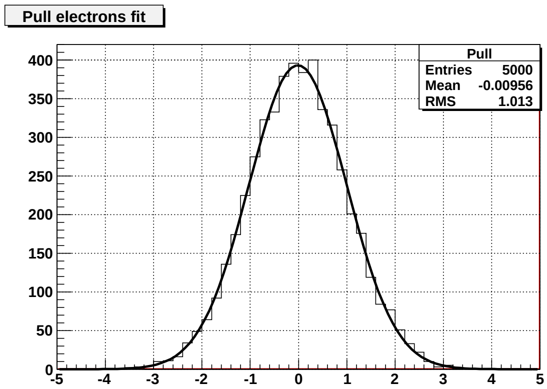


(b) muons

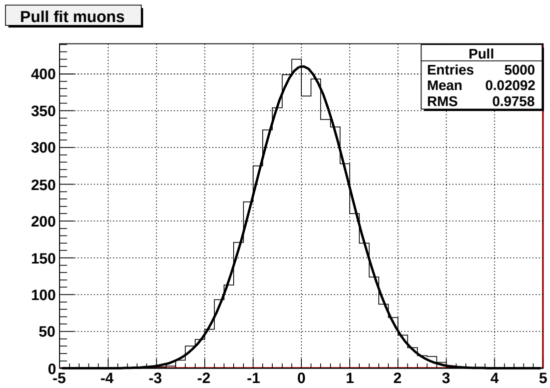
Figure 7.2: Fit results on the *pretag* $2+3$ jets sample for muon and electrons. On the left it's shown the fit on the invariant mass distribution with its residuals; on the right there is the signal, obtained from the subtraction of all the backgrounds, with the fit resulting normalizations, from data.

	fit result			SM prediction		
electrons	496	±	135	614	±	25
muons	358	±	121	504	±	22

Table 7.1: Number of WW+WZ events estimated by fit on the invariant mass distribution for *pretag* 2 and 3 jets events combined. They are compared, and found compatible, with the Standard Model expected values.



(a) electrons



(b) muons

Figure 7.3: Pull on fit results on the *pretag*, $2 + 3$ jets sample for muon and electrons. They have a gaussian distribution with expected value zero and unitary variance, demostrating that the fit procedure is not biased and errors are well estimated.

As far as muons is concern, we find a cross section of

$$\sigma^{\mu}_{WW/WZ} = 11.4 \pm 3.8(stat) \pm 0.7(lumi) \tag{7.11}$$

that is compatible with the electron measurement as well as the Standard Model cross section. The published muons result is:

$$\sigma^{e}_{WW/WZ} = 23.5 \pm 4.9(stat) \pm 0.8(lumi) \pm 3.1(syst) \tag{7.12}$$

We ascribe the small agreement of the muon published result with the electrons published result, and with our muon result, to statistical fluctuation.

7.3 Fit preliminaries on tagged data

7.3.1 Invariant mass distribution shape

As analized in Section 5.6, the smoothness of the background shape is, in principle, important as much as the signal to noise ratio for the sample sensitivity. In that section we have explained that a cut on the combined P_T^j of the jets can improve the smoothness of the background distribution, having a better "visibility" of the signal bump, at the price of 40% reduction on signal statistics.

For this reason we supposed that in order to exploit both the shape and the statistics could be reasonable to fit at once two histograms of invariant mass distribution, for events that fulfill and not fulfill the $P_T^j > 40$ GeV cut. In practice a "double" invariant mass distribution histogram will be build to simplify the fitting procedure. In order to prove which choice is the best, among the "double" invariant mass distribution, the m_{jj} distribution with the P_T^j cut and the one without that, we will perform an expected sensitivity study on each of them. From here to the end, the following notation will be adopted:

type A It's the "double" jets invariant mass distribution: the left histogram is filled with events that fulfill the $P_T^j > 40$ GeV requirement and the right with the complementary sample of Figure 7.4(a);

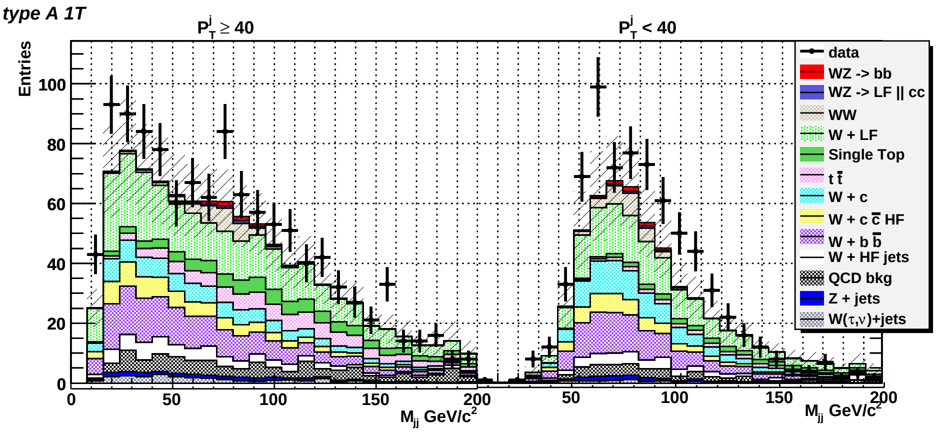
type B It's the invariant mass distribution of only that events passing the P_T^j cut of Figure 7.4(b);

type C It's the invariant mass distribution of selected events, without any requirement on P_T^j, of Figure 7.4(c).

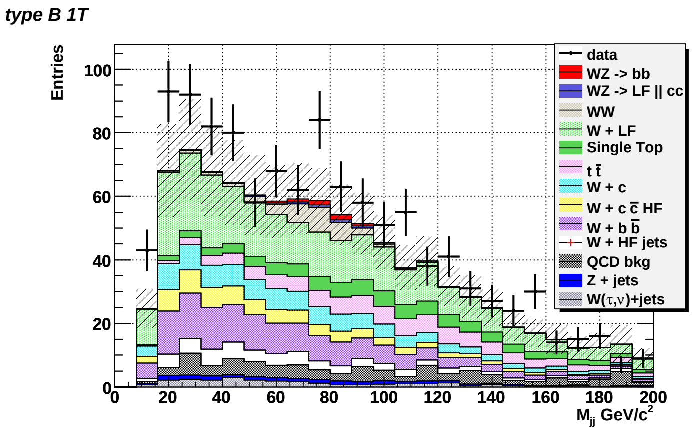
Figures 7.4(a), 7.4(b) and 7.4(c) are normalized by *Method 2* (see Sec. refM2), and the systematic bands are drawn in order to highlight that data and *Method 2*-normalized MCs are compatibles within systematic uncertainties. The fact that systematics are correlated among the bins, justify that all the experimental points are on the superior margin of the systematic band. On the other hand, the fact that there isn't a perfect agreement between the central values of data and MC samples, allow us to re-evaluate the $W + jets$ normalization in the fit.

7.4 Template choice

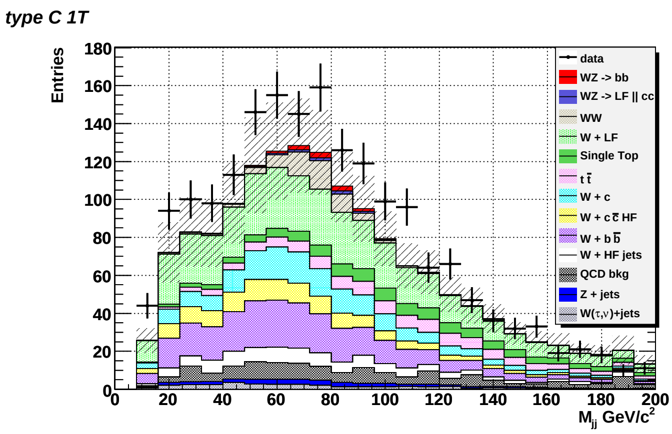
In Section 5 we have seen that we have to deal with a considerable number of background processes and each of them has different kinematic properties, so a different di-jet invariant



(a) type A



(b) type B



(c) type C

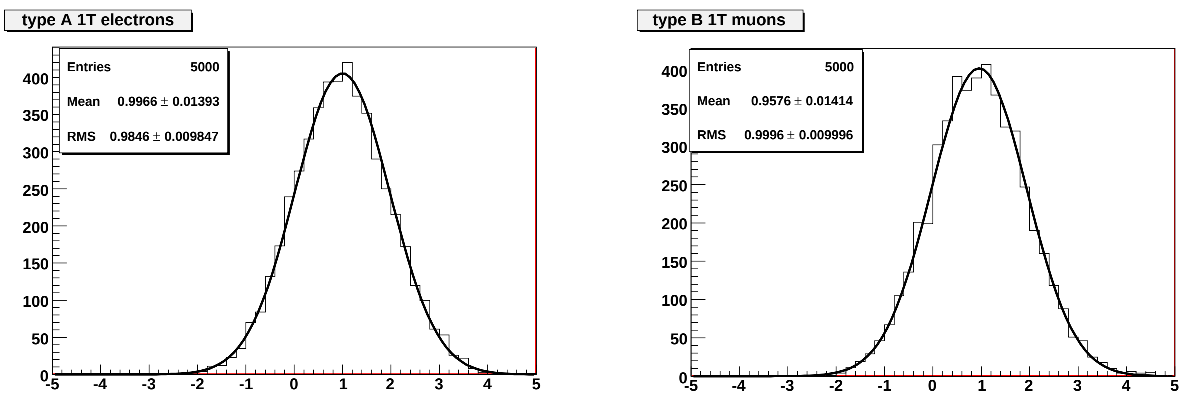
Figure 7.4: Variants of fit to the invariant mass shape, applying the $P_T^j > 40$ GeV requirement (typeB) or not (typeC) and a m_{jj} distribution build to fit the two samples, that fulfill or not the requirement, at once (typeA).

mass shape and contribution to the selected samples for this analysis, i.e. 1 and 2 TIGHT TAG samples. The first important step toward the fit of our samples is to choose which are the processes whose template will be fitted separately and which ones will be merged for their shape similarity. This choice will be driven both by the intention to derive as much information as possible on background normalization from the shape of the invariant mass distribution and the necessity of few free parameters to guarantee the fit convergence.

In Fig 7.6 and 7.7 are shown all the templates of the background processes, normalized, of 1 and 2 TIGHT TAG samples respectively. Electrons and muons shapes are compared, and found to be, as expected, very similar with the only exception of the QCD background.

Furthermore, in Figure 7.8 we compare some of the samples that looks more similar. The $W + Q\bar{Q}$ samples ($Q = b, c$) have the same shape, and can be merged together. The small differences in shape between the top pair and single top production templats only in the 1T sample are neglectable, coonsidering the fact that they are not dominant in that sample (11% and 5% of the total respectively). As far as $W + c$ (single c), it differs from $W + c\bar{c}$, as well as $W + b\bar{b}$, only at small invariant mass, but we decided that it's not enough to justify one more free parameter in this fit, so we performe the fit with only one template for W+HF.

The last question is whether to merge the W+LF template with the W+HF one. This would mean in particular to fix the two samples ratio, that is justified in the 2TT sam-



(a) 1Telectrons (b) 2TTmuons

Figure 7.5: type A, *pulls* resultant from the fit performed keeping separated the W+HF and W+LF templates on the 1T sample. It's clear that the fit cannot deal with this additional free parameter.

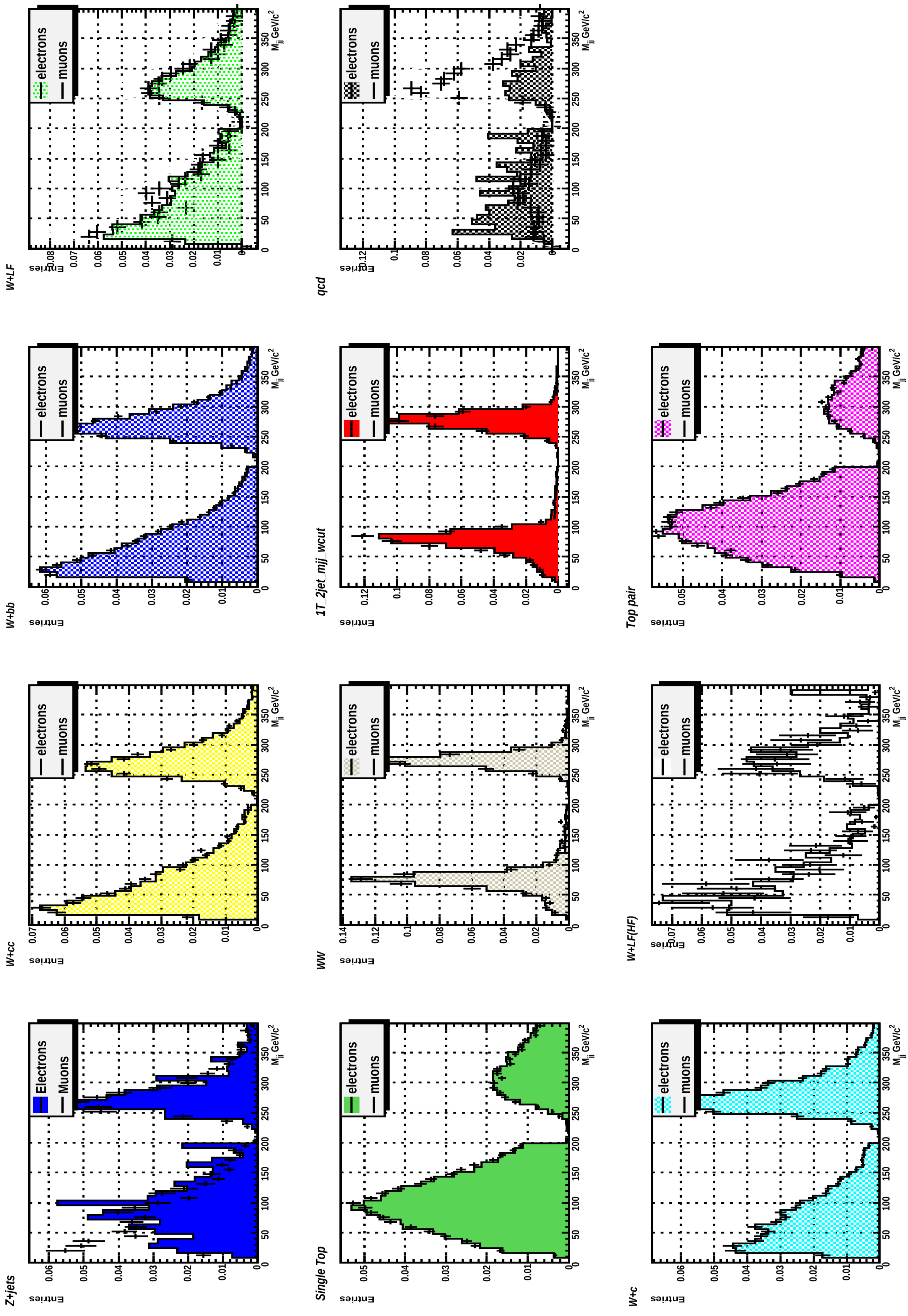


Figure 7.6: Normalized templates of each background process of the 1T sample for the invariant mass distribution of type A(see Sec. 7.3.1). Electrons and muons are compared.

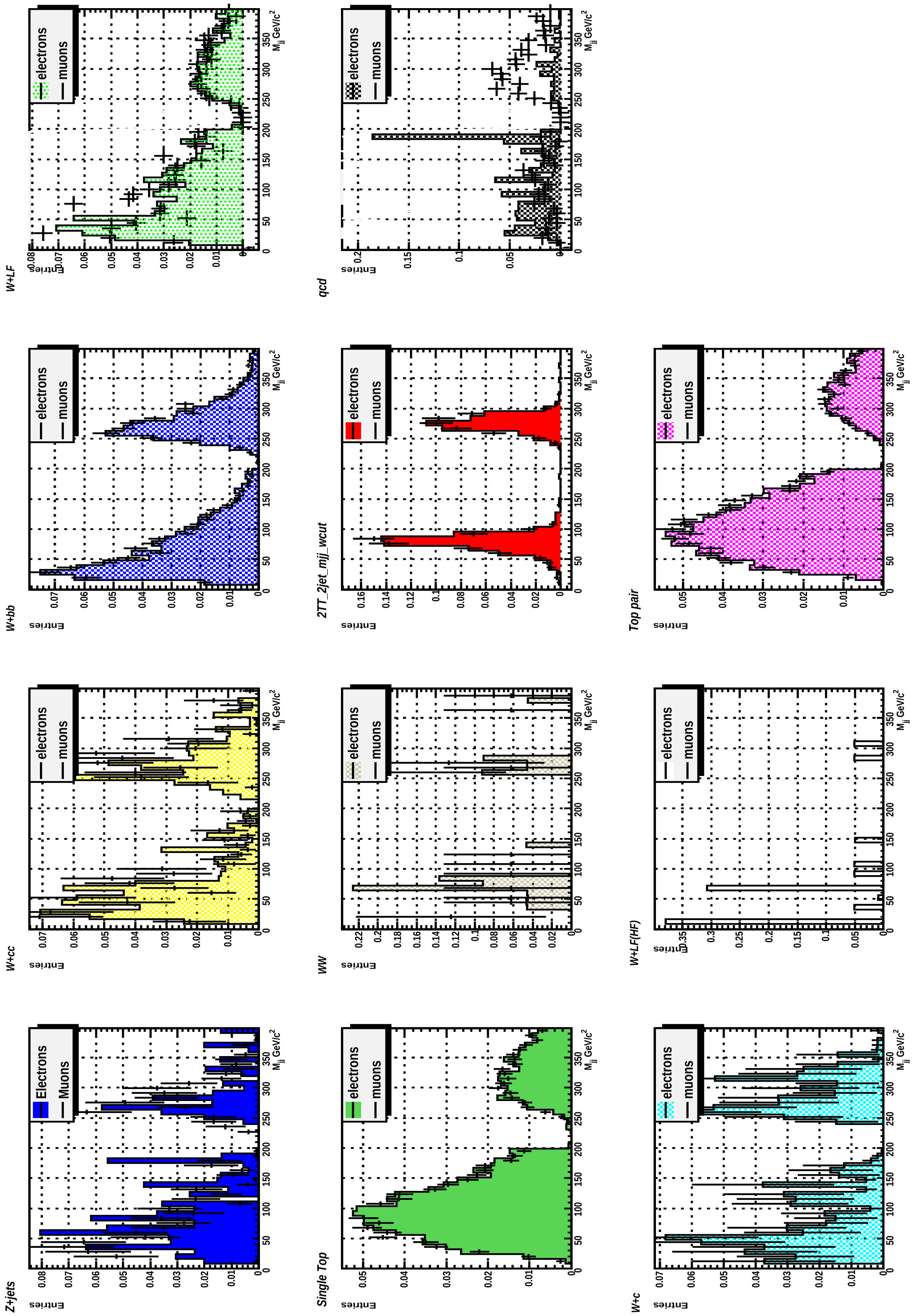
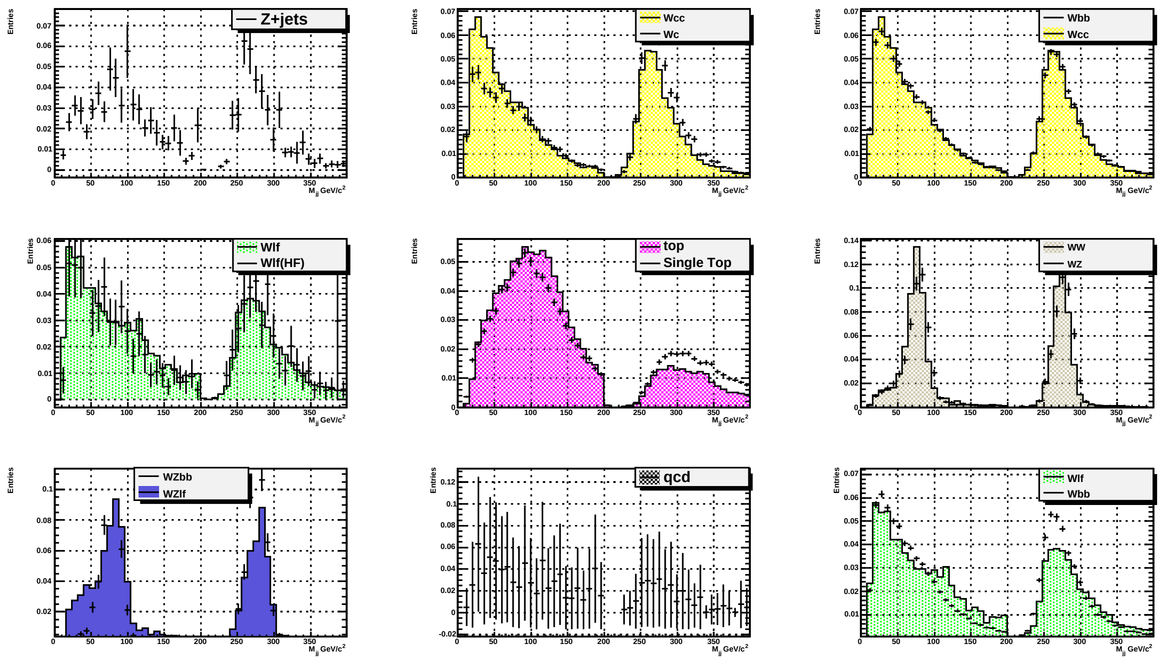
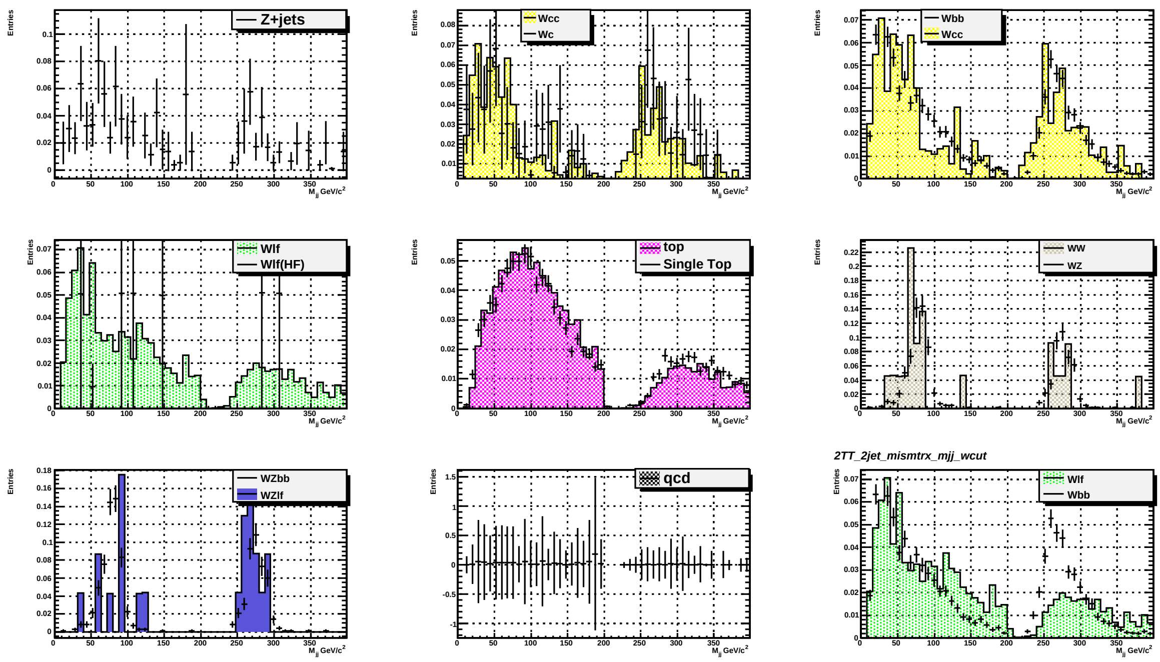


Figure 7.7: Normalized templates of each background process of the 2TT sample for the invariant mass distribution of type A(see Sec. 7.3.1). Electrons and muons are compared.



(a) 1T



(b) 2TT

Figure 7.8: Normalized templates of each background process of the tagged samples for the invariant mass distribution of `type A`(see Sec. 7.3.1)

	$P(\chi^2)$	$\mathcal{N} \cdot \sigma_{SM}$ top	$\mathcal{N}$signal fit			SM	$\mathcal{N}\sigma_{SM}$ expected limit
1T sample							
type A el	0.0294	0.92(±0.10)	20	±	32	(11)	6.3
type A mu	0.0557	1.18(±0.12)	-43	±	30	(9)	6.0
type B el	0.0055	0.99(±0.10)	7	±	21	(6)	8.2
type B mu	0.8283	1.23(±0.13)	-17	±	20	(5)	7.2
type C el	0.0096	0.95(±0.11)	19	±	32	(11)	6.3
type C mu	0.2343	1.25(±0.13)	9	±	31	(9)	6.2
2TT sample							
type A el	0.229	1.14 (±0.16)	5.7	±	6.8	(2.4)	6.1
type A mu	0.457	1.00 (±0.16)	-4.1	±	5.3	(1.9)	7.3
type B el	0.189	1.13 (±0.18)	1.2	±	4.7	(1.4)	7.5
type B mu	0.295	0.97 (±0.19)	-0.8	±	4.0	(1.1)	8.9
type C el	0.030	1.14 (±0.16)	6.9	±	6.6	(2.5)	6.4
type C mu	0.166	0.96 (±0.16)	-3.5	±	5.4	(1.9)	7.2

Table 7.2: fit results for the two tagging cathegories and all the fit types, separated for electrons and muons.

ple, where the W+HF is dominant and W+LF is only 2%. Otherwise the equivalent contribution of these two samples in the 1T sample and the high uncertainty on both normalizations ($\sim$ 20% and 30% respectively, Sec 6.2) would suggest, in this case, to separately estimate their fraction by fit. However, their shapes do not differ enough and the introduction of one more free fit parameter leads to a biased fit, as visible in the *pulls* in Figure 7.5.

To summarize our conclusions, the free parameters of our fit will be the fraction to the total of the following processes:

- W + jets (LF+HF)
- top pairs and single top
- QCD
- WW
- WZ (signal)

The Z+jet sample is fixed to its standard model value, being a small fraction of the total, to limit the number of free parameters in the fit.

7.5 Fits to the tagged samples

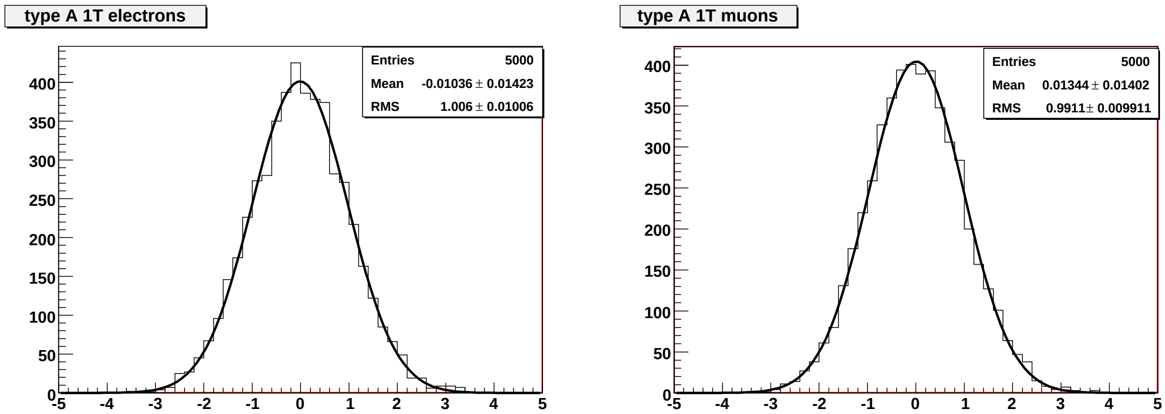
The fits on the tagged samples is a likelihood binned fit. The logarithm of the likelihood, defined in Sec. 7.1.1, has been maximized using the MINUIT minimization package [49]. First we performed the fits for separated electrons and muons samples. The first 20 GeV of the invariant mass distribution are not considered in the fit to the 1T sample, as in the *pretag* analysis, since the Monte Carlo seems to not well reproduce the the data trend. The starting point of the fit parameters are the background fractions estimated from the MC for top, WW and Z+jets processes. As far as QCQ and W+jets the normalization are the ones described in Chapter 6. Some gaussian costrains have been implemented to make use of these normalization to enhance signal sensitivity. The gaussian width is equal to 10% of the initial normalization value for WW and top processes, and is the experimental uncertainty on their cross sections.

The QCD constrain width is equal to 25% of the initial normalization due to the systematic uncertainty on this template, described in Sec. 6.1. Both 1 and 2 TIGHT TAG samples have been fitted, for types A, B and C and the χ^2 probability of each fit, as well as the estimated number of signal events have been reported in Table 7.2. In this table we show the fitted cross section of top processes, expressed in top SM cross section units, with the relative error. Since they are all compatible with 1 within the errors, we are more confident on the results of our fit.

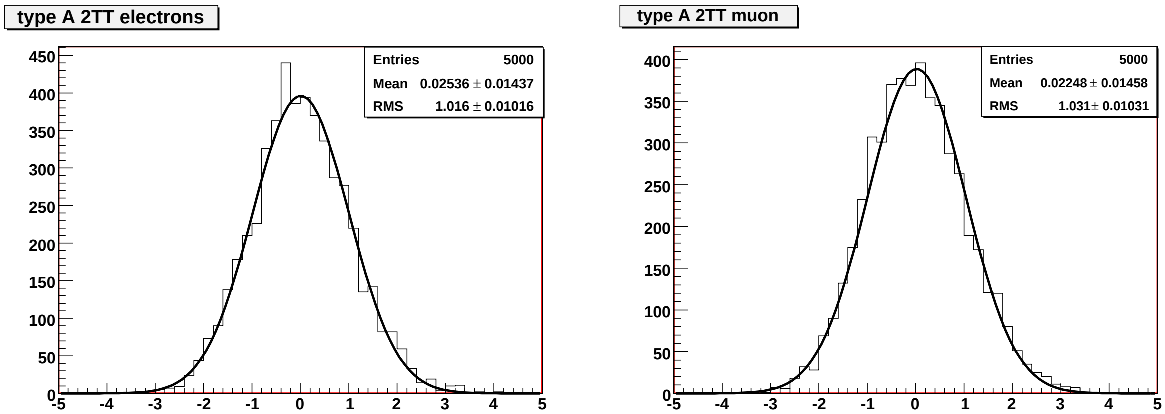
The expected limit, as described in Sec 7.1.4, has been calculated in order to decide which fit typology is the most sensitive for the limit calculation. An ensamble of 5000 pseudo-experiments has been generated, for each type and tagging cathegory, in order to estimate the median $\Delta\chi^2$ from MC. Such value has been then used to set a limit at 95% of confidence level.

From Tab. 7.2 it results that there is small difference between type A and type C fits, while the expected limit of type B fit is higher, demonstrating that the strong reduction of statistic in that sample is not compensated by a smoother background shape.

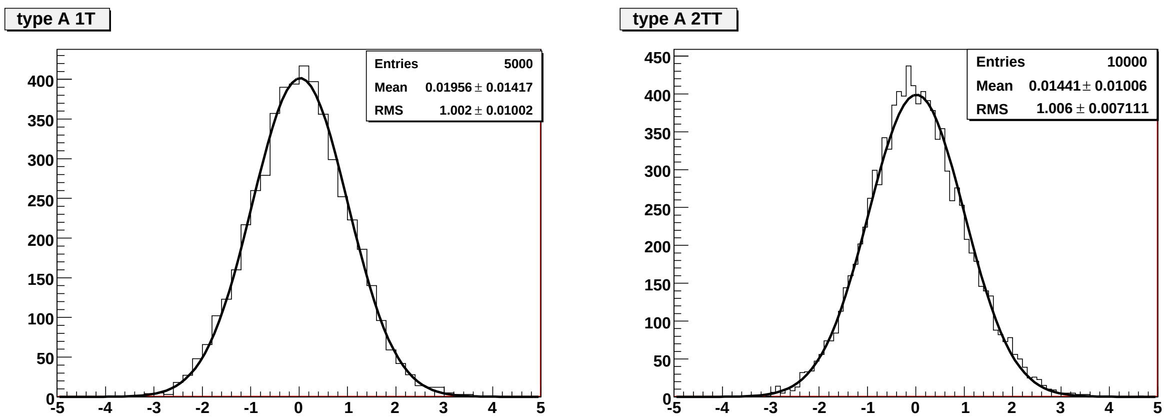
Secondly, the electrons and muons samples have been merged for the two tagged samples. The QCD template of muons and electrons are not merged for the shape difference highlighted in Sec. 7.4 The fit results for each type and the expected and measured limits are shown in Table 7.3. In the case of combined electrons and muons the expected limit



(a) 1T



(b) 2TT



(c) muons and electrons combined

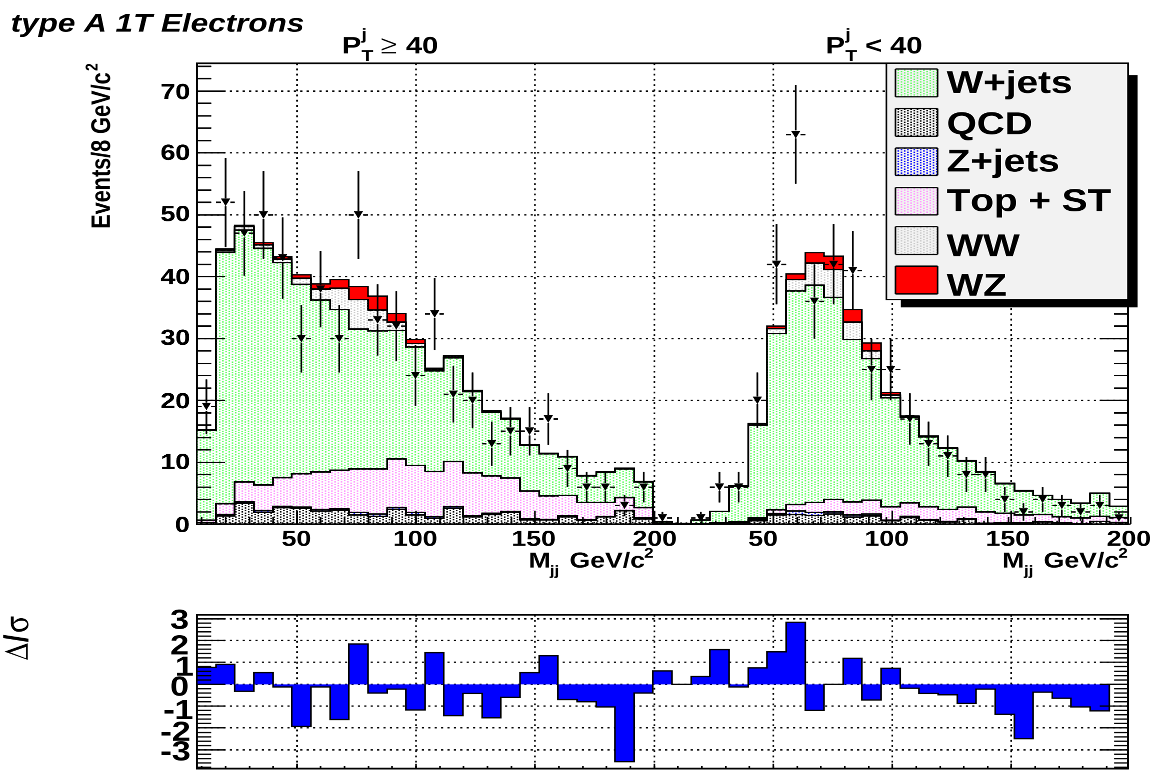
Figure 7.9: Pull on fit results on the type A tagged samples. Separate and combined muons and electrons sample are shown.

	P(χ^2)	$\mathcal{N} \cdot \sigma_{SM}$ top	$\mathcal{N}$signal fit			SM	$\mathcal{N}\sigma_{SM}$ limit expected	measured
1T sample								
type A	0.012	0.93(0.10)	18	±	44	(20.7)	4.5	3.7
type B	0.073	0.95(0.10)	-22	±	30	(11.5)	6.0	6.2
type C	0.017	0.92(0.10)	14	±	45	(20.7)	5.0	3.8
2TT sample								
type A	0.245	1.06(0.13)	2.6	±	8.7	(4.42)	4.8	3.5
type B	0.292	1.05(0.15)	0.6	±	6.2	(2.50)	6.0	3.5
type C	0.233	1.05(0.13)	3.3	±	8.6	(4.42)	4.8	3.9

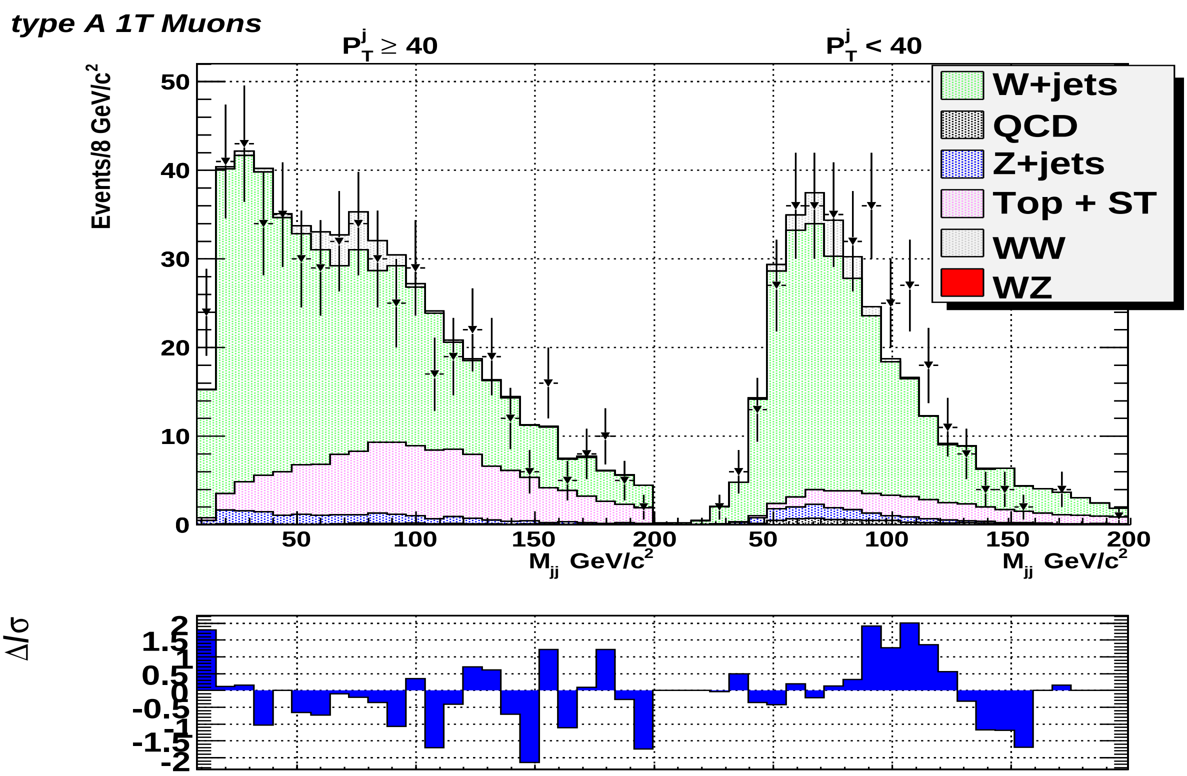
Table 7.3: Fit results for the two tagging cathegories, for muons and electrons combined samples. Expected and measured cross section upper limits are shown in the table.

differences are even smaller than in the separate fits, and type A seems better than type C only in the 1T sample. The χ^2 probability are very similar for the three fit typologies, hence they do not represent a matter of discrimination. However, we think that the type A fit variant has more potentiality than type C, since, for example, with more statistics could be possible to independently variate the background templates in the sample that fulfill and do not the P_T^j cut. For this reason we are going to choose the type A fit variation for the evaluation of a cross section upper limit of our signal.

The fit results for type A separate muons and electrons are show in Figures 7.10 and 7.11 for 1 and 2 TIGHT TAG samples respectively. The final results on WZ cross section upper limit is given in the next chapter. The pulls for this fit in Figure 7.9 prove that it is unbiased and that the errors are well estimated. The pulls for the fit on the muon and electrons combined samples are shown in the same figure and show a gaussian distribution too.

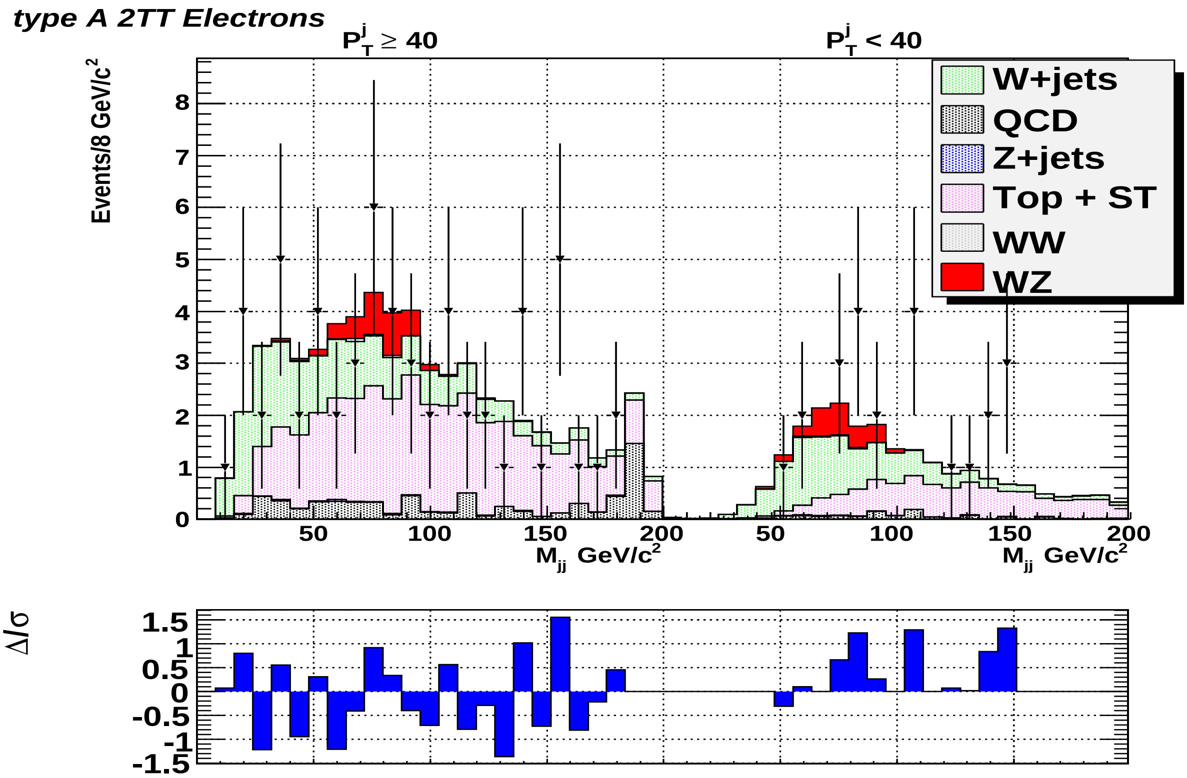


(a) electrons

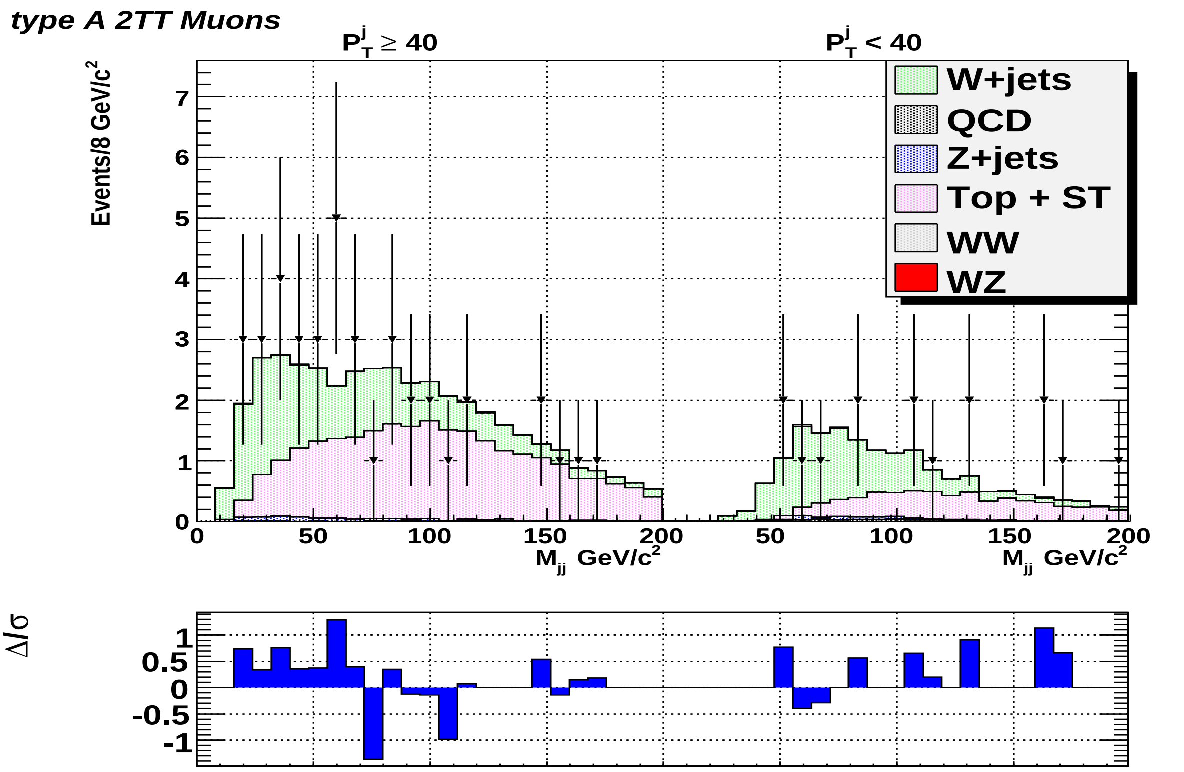


(b) muons

Figure 7.10: type Afit results on the 1Tsamples for muon and electrons.



(a) electrons



(b) muons

Figure 7.11: type Afit results on the 2TTsamples for muon and electrons.

Part III

Conclusions

Chapter **8**

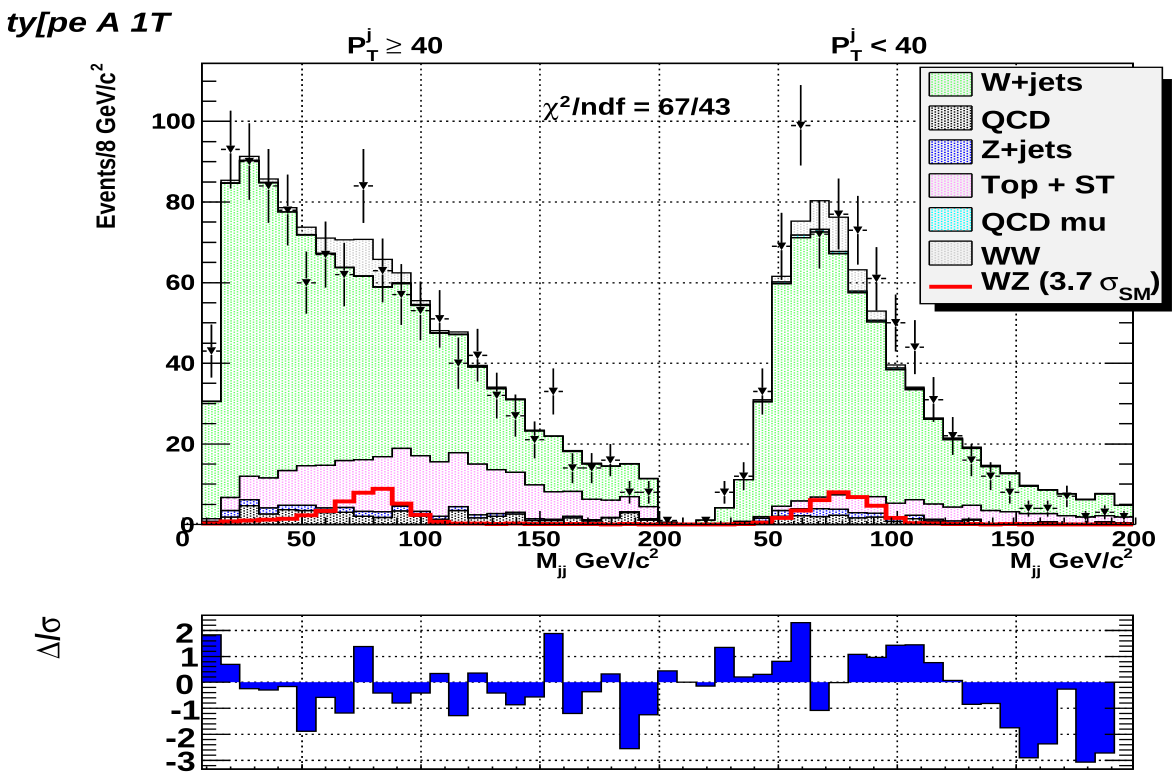
Result and conclusions

The final result on $W^{\pm}Z$ cross section upper limit is given in this chapter. An alternative method, based on a work of Feldman and Cousins, is used to combine the results for the two tagged samples. Systematic uncertainties, that are not accounted for yet, are listed and described, as long as the further possible improvements that could be implemented in this analysis in order to enhance the sensitivity to the signal. Finally, the results on $WZ \rightarrow l\bar{\nu}_l b\bar{b}$ cross section limit are compared with $WH \rightarrow l\bar{\nu}_l b\bar{b}$ ones

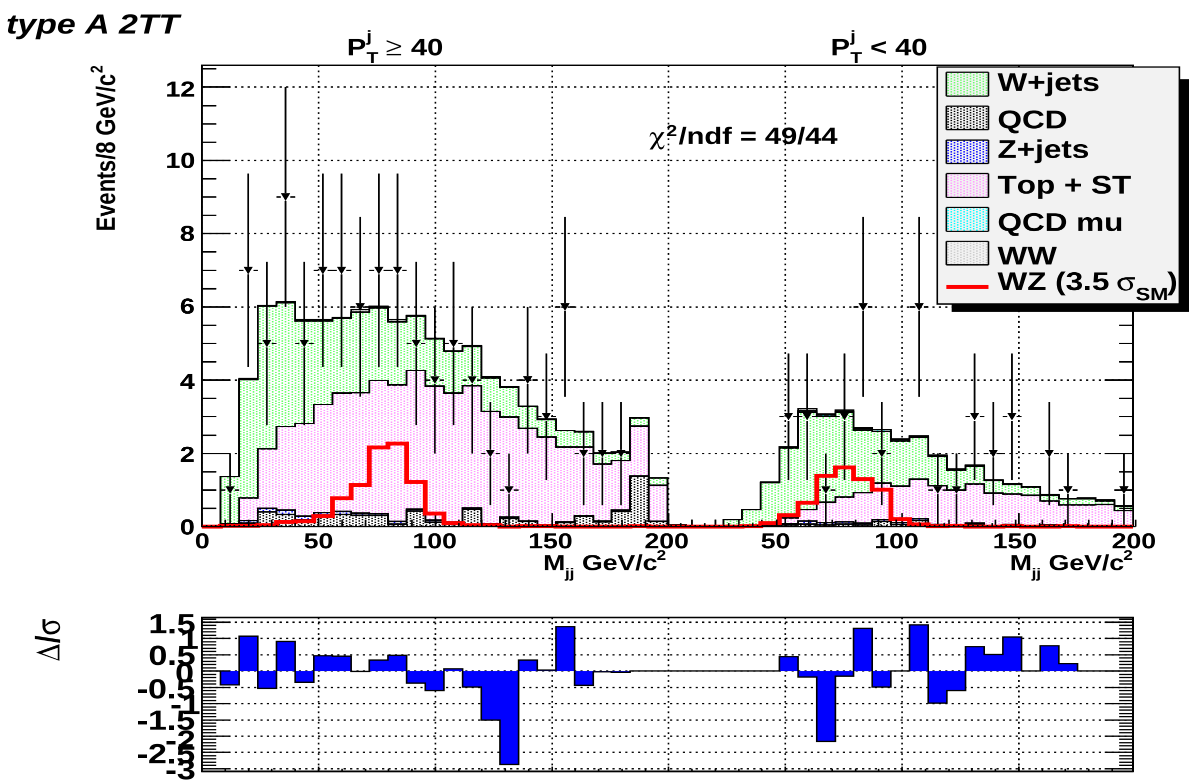
8.1 Final results on upper limits to $WZ \rightarrow l\bar{\nu}_l b\bar{b}$ cross section

The aim of this analysis is to look for an evidence of the $WZ \rightarrow l\bar{\nu}_l b\bar{b}$ process as a resonance in the di-jet invariant mass distribution of events with tagged jets. The dataset have been divided into two sample, according to the number of jets identified as b-jets by the SecVtx algorithm at the TIGHT working point (Sec. 3.4) in the event: 1T (1 TIGHT TAG) and 2TT (2 TIGHT TAG). Only events with two selected jets are considered and muons and electrons sample have been merged at this point. However, no evidence of the signal, or any other resonance in the di-jet invariant mass distribution, has been measured and an upper limit on the $WZ \rightarrow l\bar{\nu}_l b\bar{b}$ process cross section have been evaluated.

A the end of the previous chapter, a limit sensitivity study and some consideration lead



(a) 1T



(b) 2TT

Figure 8.1: type Afit results on the tagged samples for combined muon and electrons. The signal is not included in the background shape, but is rescaled to the cross section upper limit at 95% of C.L.

use chose the type A (Sec. 7.3.1) fit variant, as the invariant mass distribution on which perform our limit estimation. This distribution involves two complementary samples, composed of events that fulfill or not fulfill the $P_T^j > 40$ GeV requirement (for further details on the nature of this cut, see Sec. 5.6).

The expected limits for the tagged sample, with their expected limit 1σ-range, calculated as described in Sec. 7.1.4, are

$$\sigma_{WZ}^{1T} < 4.5 \cdot \sigma_{SM} \qquad (@95\%\ \text{C.L.}) \qquad [3.1 \div 6.1] \tag{8.1}$$

$$\sigma_{WZ}^{2TT} < 4.8 \cdot \sigma_{SM} \qquad (@95\%\ \text{C.L.}) \qquad [3.5 \div 9.0] \tag{8.2}$$

The resultant limit @ 95% of Confidence Level, for 1T sample is

$$\sigma_{WZ}^{1T} < 3.7 \cdot \sigma_{SM} \qquad @95\%\ \text{C.L.} \tag{8.3}$$

From the 2TT sample it results a cross section limit of

$$\sigma_{WZ}^{2TT} < 3.5 \cdot \sigma_{SM} \qquad @95\%\ \text{C.L.} \tag{8.4}$$

Both the limits are compatible with the expected values within the expected limit 1σ-range.

cathegory	$\mathcal{N}$signal fit			SM
type A 1T	18	±	44	(20.7)
type A 2TT	2.6	±	8.7	(4.42)

Table 8.1: Summary of fit results for combined electrons and muons, type A fit variant

Until the appropriate tools for a 1T-2TT joint fit will be developed, the the 2TT result should be taken as the result of this analysis, being the lower limit. It is compatible with the expected limit within the expected limit 1σ-range. In Figure 8.1 we show the fit results. The MC shape in those figures do not include the signal template, that is rescaled to the cross section on which we set the limit and drawn not-stacked on the graph.

8.1.1 Feldman and Cousins approach for an estimation of a combined upper limit

By virtue of the gaussian *pulls* of our fits (Fig. 7.9), we are allowed to apply and althernative procedure for limit calculation. malized to its error. This procedure takes advantage of the Feldman and Cousins study on statistical analysis of small signals [50], and their parametrization of the confidence intervals for the mean μ of a gaussian as a function of the measured mean x_0, nor It is possible to use their parametrization (Tab. X in [50]) to estimate the cross section upper limit from the signal events obtained by fit of Tab. 7.3, summarized in Table 8.1 for convenience, and calculate a joint limit for the 1T-2TT samples.

We obtain the following limits for the tagged samples:

$$\sigma_{WZ}^{1T} < 5.0 \cdot \sigma_{SM} \qquad \text{@95\% C.L.}$$
$$\sigma_{WZ}^{2TT} < 4.5 \cdot \sigma_{SM} \qquad \text{@95\% C.L.}$$

that are are slightly higher than the limits reported in Section 7.1.3 but well within the 1 sigma range. The number of signal events of each sample is then converted in a cross section measurement and combined with a weighted mean. The cross section combined result is

$$\sigma_{WZ} = 2.9 \pm 5.8 \text{ pb} \tag{8.5}$$

that leads to a combined upper limit, in the Feldman-Cousin approach, of

$$\sigma_{WZ} < 3.6 \cdot \sigma_{SM} \qquad \text{@95\% C.L.} \tag{8.6}$$

In conclusion, the final cross section upper limit on $W^{\pm}Z$ associate production is

$$\sigma_{p\bar{p}\to WZ} < 14.0 \text{ pb} \qquad \text{@95\% C.L.} \tag{8.7}$$

8.2 Future perspectives

There are many ways in which this analysis can be improved. Hence, we would like to highlight the most important points on which this analysis shoud be imporved.

The first important point is that the cross section limit presented do not account for systematic uncertainties yet.

The main systematics that should be analysed soon are listed in the following. For each of them, an invariant mass template should be derived and used to perform the limit calculation. The higher upper limit will be taken as the final result.

- the contribution of the Jet Energy Scale (JES) on the shape of the fit templates; it is evaluated performing the fit with the templates obtained variating the JES within the experimental uncertainty quoted in [31].
- uncertainties of trigger and lepton identification efficiencies.
- the uncertainty on the tagging efficiencies and on the *Mistag Matrix* parametrization; the latter is obtained variating the probability value from *Mistag Matrix* for each event up and down with the quoted systematic.
- $W + jets$ (LF and HF) shapes obtained from varying the factorization scale ($Q^2 = (M_W/2)^2$,$Q^2 = (2M_W)^2$) in theoretical calculations.
- the uncertanty of the integrated lumonosity of the dataset (6%);
- the possible contribution of the QCD template choice (*anti-electrons* and *non-isolated muons* samples); this particular systematic has to be evaluated changing the number of required ID cut fails on *anti-electrons* sample and requiring an isolation greather than 0.4 for the *non-isolated muons* one.

Further improvements can be applied to the analysis with the purpose of better exploiting all the potential sensitivity of the analized dataset. We list the ones that we consider the most important:

- the implementation of a combined limit calculation between the two tagging cathegories.
- the analysis cut optimization for our signal. Further kinematics variables, such as the total transverse energy in the detector (Ht) and the number of loose jets in the event, can be taken into account to further reduce the top and single top background.
- the LOOSE tag of SecVtX, that has demonstrated a small potential sensitivity to the signal should be sobstituted with a loose cut on the JetProbability tagging algorithm.
- Roma-Tagger tagging algorithm's sensitivity should be studied, in order to verify if it it more sensitive than JetProbability and SecVtX.

- further leptons cathegories can be included in the analysis to increase the signal acceptance.
- implementation of invariant mass resolution improvements, such as the possibility of recognize jets, in 3 JETS!events, originated from gluon radiation and combine them to obtain the invariant mass of the Z.
- an independent measurement on the W+HF cross section could add a constrain and reduce the uncertanties to the fit.

Improving in such way the analysis, and adding the luminosity that CDF has already taken at this moment ($L \sim 7.0$ fb^{-1}) it will be probably soon possible to finally measure the WZ signal in the lepton plus jets decay channel.

8.3 Comparison with the $WH \to l\bar{\nu}_l b\bar{b}$ results

As last conclusion, we compare our cross section upper limit for $WZ \to l\bar{\nu}_l b\bar{b}$ process with the the CDF $WH \to l\bar{\nu}_l b\bar{b}$ results obtained with an equivalent integrated luminosity. In [51] the resulting upper limit on Higgs associate production, for $m_H = 120$ GeV is:

$$\sigma_{WH} < 3.9 \cdot \sigma_{SM} \quad @95\% \;\; \text{C.L} \qquad (4.7 \cdot \sigma_{SM} \quad \text{expected}) \tag{8.8}$$

where more tagging and leptons cathegories than the ones we have considered contribute to this limit result, being our analysis still at a preliminary level.Despite a cross section five times greather, our limit on $W^{\pm}Z$ is comparable with the Higgs result, as well as the number of expected events. In fact we expect approximately 25$\pm$5 $W^{\pm}Z$ events combining the 1 and 2 TIGHT TAG cathegories, while the expected yield for $WH \to l\bar{\nu}_l b\bar{b}$ estimated in [51] is just a factor two less: 12.2 ± 1.1 events, for the same two cathegories. This is due to different acceptances and efficiencies between the jets produced in the Z decay and the more energetic ones produced in the Higgs decay, and demostrate that is not straightforward that the $WZ \to l\bar{\nu}_l b\bar{b}$ signal is a less challenging search than the Higgs one and will be observed before, on the contrary they will probably be observed together.

Bibliography

[1] Michelangelo L. Mangano, Fulvio Piccinini, Antonio D. Polosa, Mauro Moretti, and Roberto Pittau. Alpgen, a generator for hard multiparton processes in hadronic collisions. *Journal of High Energy Physics*, 2003(07):001, 2003.

[2] J. Adelman, S. Budd, R. Dong, P. andErbacher, S. Gringstein, D. Hare, K. Lannon, E. Palencia, C. Plager, J. Slaunwhite, D. Sherman, T. Schwarz, V. Sorin, B. Stelzer, and R. Waalny. Method ii for you. *CDF internal note 9185*, 2008.

[3] J. Freeman, T. Junk, E. Palencia, and S. Gringstein. Secvtx mistag matrices for 2.2/fb of data up to p13. *CDF internal note*, 2008.

[4] T. Aaltonen and *et al.* Measurement of the $ww + wz$ production cross section using the $lepton + jets$ final state at cdf ii. *Phys. Rev. Lett.*, 104(10):101801, Mar 2010.

[5] Claude Amsler, M Doser, M Antonelli, and D M et al Asner. Review of particle physics, 2008-2009. review of particle properties, 2008-2009. *Phys. Lett. B*, 667(1-5):1–6, 2008.

[6] J. M. Campbell and R. K. Ellis. Update on vector boson pair production at hadron colliders. *Phys. Rev. D*, 60(11):113006, Nov 1999.

[7] V. M. Abazov and *et al.* Measurement of the ww production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ *tev*. *Phys. Rev. Lett.*, 94(15):151801, Apr 2005.

[8] CDF Collaboration. Measurements on the wz production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ tev using 5.9 1/fb of cdf run ii data. *CDF public note 10176*, 2010.

[9] V. M. Abazov and *et al.* Measurement of the $z\gamma \to \nu\nu$ production cross section and limits on anomalous $zz\gamma$ and $z\gamma\gamma$ couplings in $p\bar{p}$ collisions at $s = 1.96 tev$. *Phys. Rev. Lett.*, 102(20):201802, May 2009.

[10] J. Keung, E. Thomson, T. Schwarz, and C. New. Search for teh standard model z boson production in association with $w^{\pm}$ boson using 4.3 1/fb. *CDF public note 10044*, 2010.

[11] Fermilab beam division. http://www-bd.fnal.gov.

[12] Fermilab beam division. Run ii handbook.

[13] R. Blair *et al.* The cdf-ii detector: Technical design report. *FERMILAB-PUB-96/390-E (1996).*

[14] A. Still *et al.* Svx-ii: Cdf run ii silicon tracking projects. *Nucl. Instrum. Methods, A 447, 1-8 (2000).*

[15] T.Affolder *et al.* Cot central outer tracker. *Nucl. Instrum. Methods, A 526, 249 (2004).*

[16] http://www-cdf.fnal.gov/internal/detectors/parameters.html.

[17] D. Acosta *et al.* A time-of-flight detector in cdf-ii. *Nucl. Instrum. Methods, A 518, 605-608 (2004).*

[18] L. Balka *et al.* The cdf central electromagnetic calorimeter. *Nucl. Instrum. Methods, A 267, 272 (1988).*

[19] S. Bertolucci *et al.* The cdf central and endwall hadron calorimeter. *Nucl. Instrum. Methods, A 267, 301 (1988).*

[20] M. Albrow *et al.* The cdf plug upgrade electromagnetic calorimeter: test beam results. *Nucl. Instrum. Methods, A 480, 524 (2002).*

[21] G. Ascoli *et al.* Cdf central muon detector. *Nucl. Instrum. Methods, A 268, 33 (1988).*

[22] K. Byrum *et al.* The cdf forward muon system. *Nucl. Instrum. Methods, A 268, 46 (1988).*

[23] D. Acosta *et al.* The cdf luminosity monitor. *Nucl. Instrum. Methods, A 461, 540 (2001).*

[24] S. Jundariani *et al.* Luminosity uncertainty for run 2 up until august 2004. *CDF/ANA/7446*, 2005.

[25] H.J. Frisch and P.J. Wilson. Trigger tower organization and summing in $\eta - \phi$ space for run ii and beyond. *cdf public note 2045*, 2001.

[26] http://www.physics.ohio state.edu/~hughes/xft/. The xft system webpage.

[27] http://web.hep.uiuc.edu/engin/cdf/xtrp/. The xtrp webpage.

[28] I. Vila. Performance and first physics results of the svt trigger at cdfii. *arXiv:hep-ph/0307165*, 2003.

[29] The Trigger and Datasets Working Group. Run ii trigger table and datasets plan. *CDF public note 4718*, 2001.

[30] D. Hare, E. Halkiadakis, and T. Spreitzer. Electron id efficiency and scale factor for winter 2007 analyses. *CDF internal note 8614*, 2006.

[31] A. Bhatti and *et al.* Determination of the jet energy scale at the collider detector at fermilab. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 566(2):375 – 412, 2006.

[32] D. Acosta and *et al.* Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $s = 1.96tev$ using lepton + jets events with secondary vertex b -tagging. *Phys. Rev. D*, 71(5):052003, Mar 2005.

[33] D. Acosta and *et el.* Measurement of the $t\bar{t}$production cross section in $p\bar{p}$ collisions at $s = 1.96tev$ using lepton + jets events with secondary vertex b -tagging. *Phys. Rev. D*, 71(5):052003, Mar 2005.

[34] D. Acosta, V. Necula, P. Pfeiffer, M. Schmitt, A. Sukhanov, D. Tsybychev, and S.M Wang. Introduction to run ii jet probability heavy flavour tagger. *CDF public note 6315.*

[35] A. Abulencia and *et al.* Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $s = 1.96 tev$ using $lepton + jets$ events with jet probability b-tagging. *Phys. Rev. D*, 74(7):072006, Oct 2006.

[36] C. Ferrazza and D. Jeans. A new b and c jet identification algorithm. *CDF public note 8451*, 2006.

[37] Daniel Jeans for the CDF collaboration. B tagging at cdf. *EPJ manuscript.*

[38] B. Cooper and A. Messina. Estimation of the background to w$\rightarrow$ $e\nu + n\ jets$ events. *CDF public internal note 7760*, 2005.

[39] Torbjrn Sjstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.

[40] Stefan Hoche et al. Matching parton showers and matrix elements. 2006.

[41] Franklin, Grinstein, Guimaraes da Costa, Lannon, Schwarz, Sherman, Taffard, and Zaw. Heavy-flavour content of the w+jets sample. *CDF internal note 8765*, 2007.

[42] CDF Collaboration. Measurement of the cross section for the production of a w boson in association with jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ tev. *CDF public internal note 8525*, 2006.

[43] A. Abulencia and *et el.* Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $s = 1.96 tev$. *Phys. Rev. Lett.*, 97(8):082004, Aug 2006.

[44] T. Aaltonen and *et el.* Measurement of the single-top-quark production cross section at cdf. *Phys. Rev. Lett.*, 101(25):252001, Dec 2008.

[45] The CDF Collaboration. Search for standard model higgs boson production in association with a w boson at cdf. *CDF public note 10172, submitted to PRD*, 2008.

[46] John Freeman. Summer 2009 secvtx scale factorscalculated using the electron method through period 22. *CDF public internal note 9848*, 2009.

[47] Tom Junk. Sensitivity, exclusion and discovery with small signals, large backgrounds and large systematic uncertainties. *CDF public note 8128*, 2007.

[48] Tom Junk. Frequentist exclusion and discovery of signals in combined searches with large siystematic uncertainties on the background. *CDF public note 6525*, 2003.

[49] F. James. Minuit function minimization and error analysis. *v94.1 Reference Manual*.

[50] Gary J. Feldman and Robert D. Cousins. Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57(7):3873–3889, Apr 1998.

[51] The CDF Collaboration. Search for standard model higgs boson production in association with a w boson using matrix element techniques with 4.8/fb of cdf data. *CDF public note 10068*, 2010.

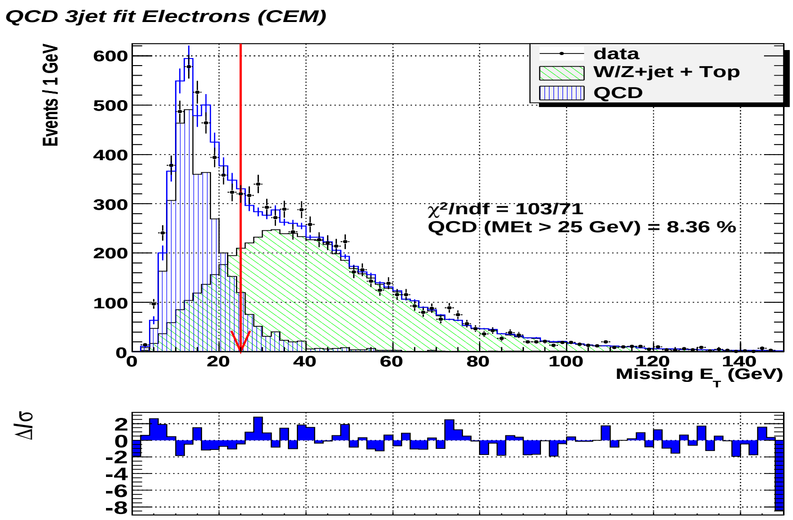
List of Figures

List of Tables

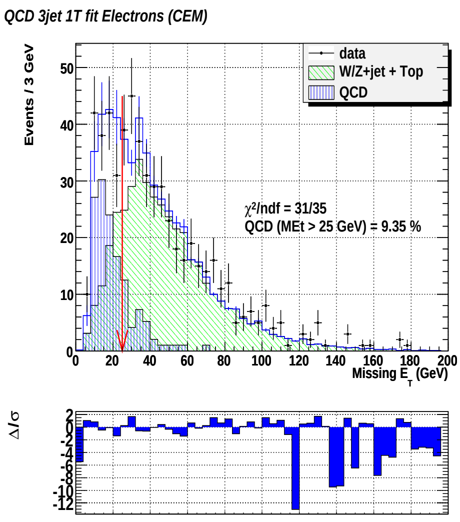
Appendix **A**

QCD fits for W + 3jets

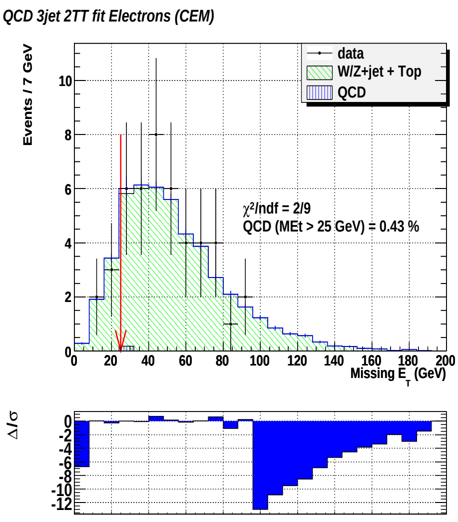
For sake of completeness we show, in the fillowing figures, the results for the QCD estimation in electrons and muons sample for 3 jets events.



(a) *pretag*

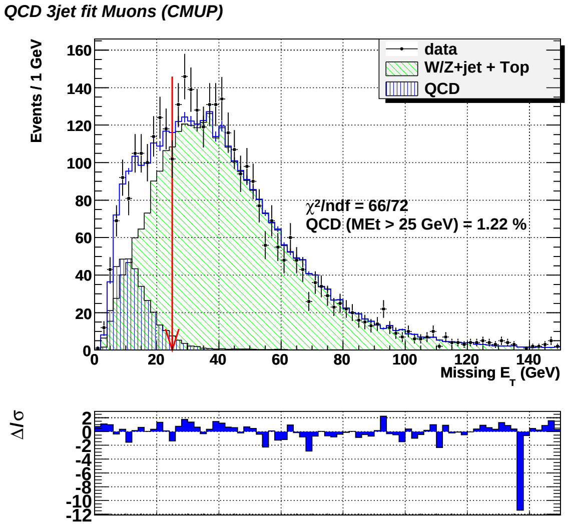


(b) 1T

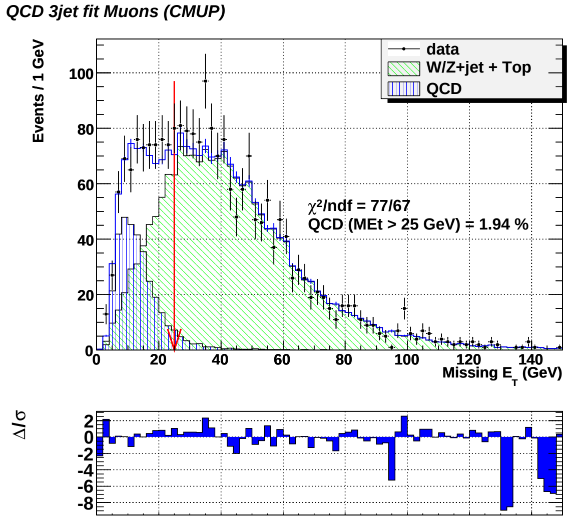


(c) 2TT

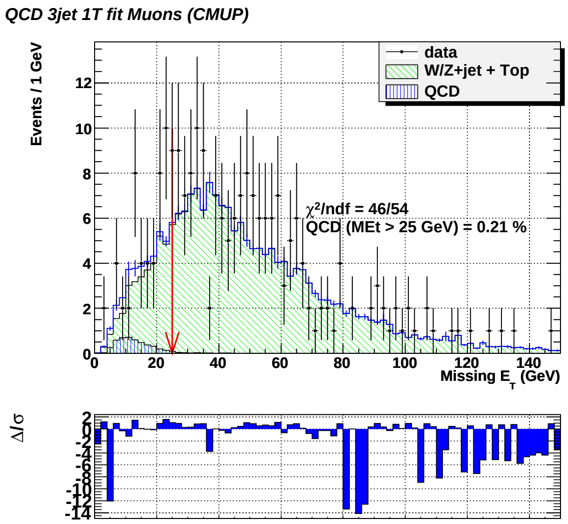
Figure A.1: electrons QCD background estimation by the fit in *missing energy* distribution for the *pretag*, 1 TIGHT TAGand 1TIGHT TAG samples, with only 3 selected jets in the event. All the analysis cuts are applied



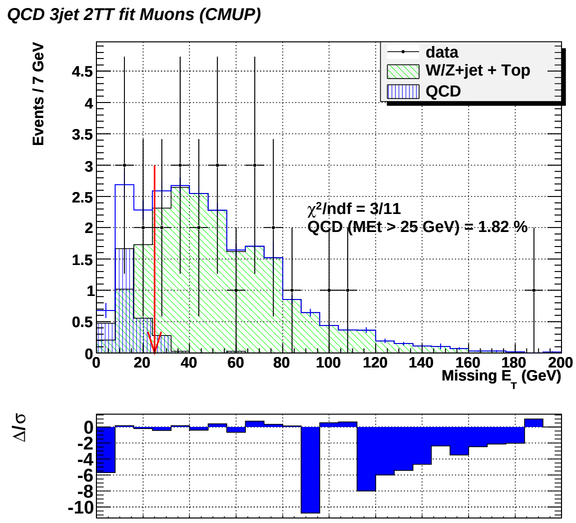
(a) *pretag* CMUP



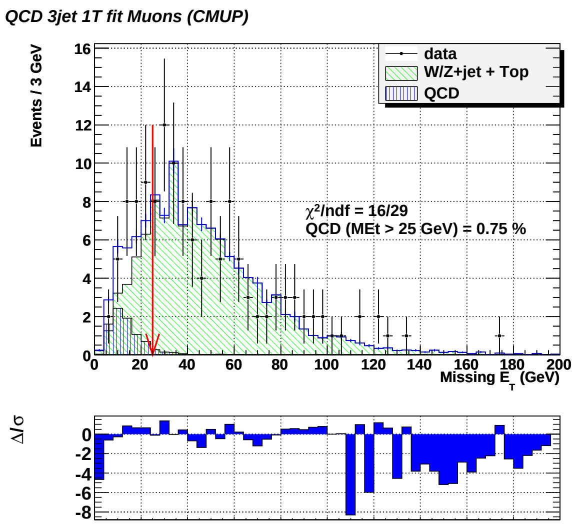
(b) *pretag* CMX



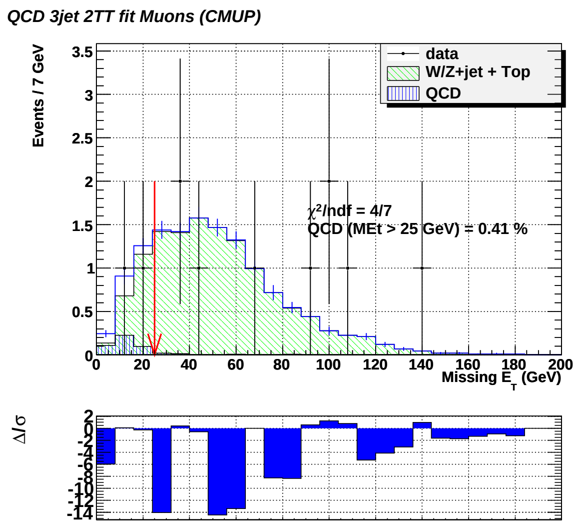
(c) 1T CMUP



(d) 2TT CMUP



(e) 1T CMX



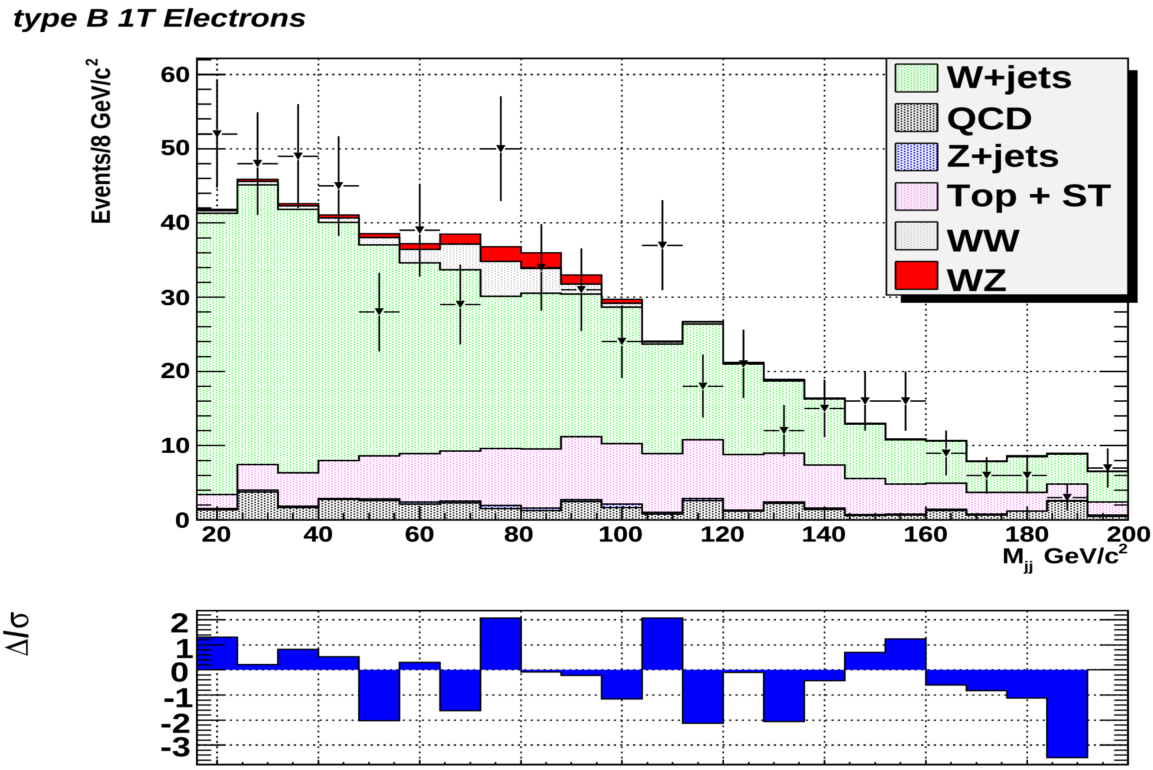
(f) 2TT CMX

Figure A.2: Muons QCD background estimation by the fit in *missing energy* distribution for the *pretag*, 1 TIGHT TAGand 1TIGHT TAG samples, with only 3 selected jets in the event. All the analysis cuts are applied

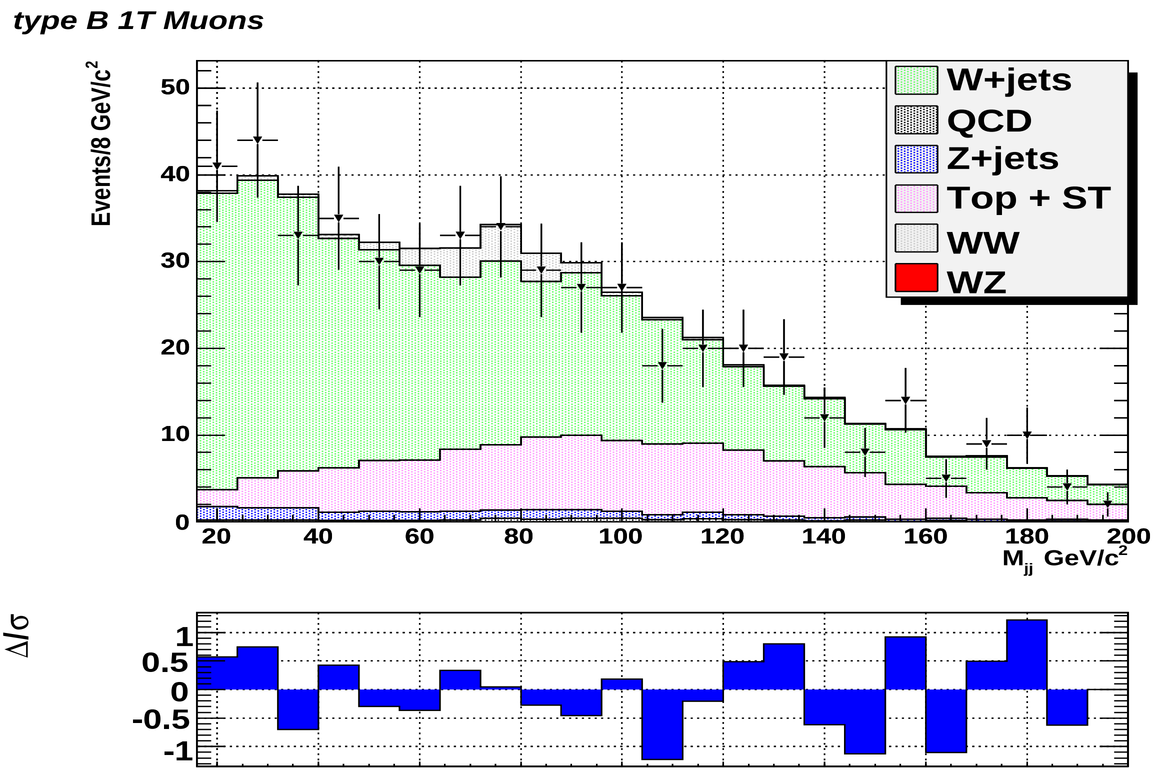
Appendix B

Further fits to the tagged samples

We show the fit results for all the type of histograms analysed in Sec 7.5 and 8.1, for electrons and muons combined and in separate samples.

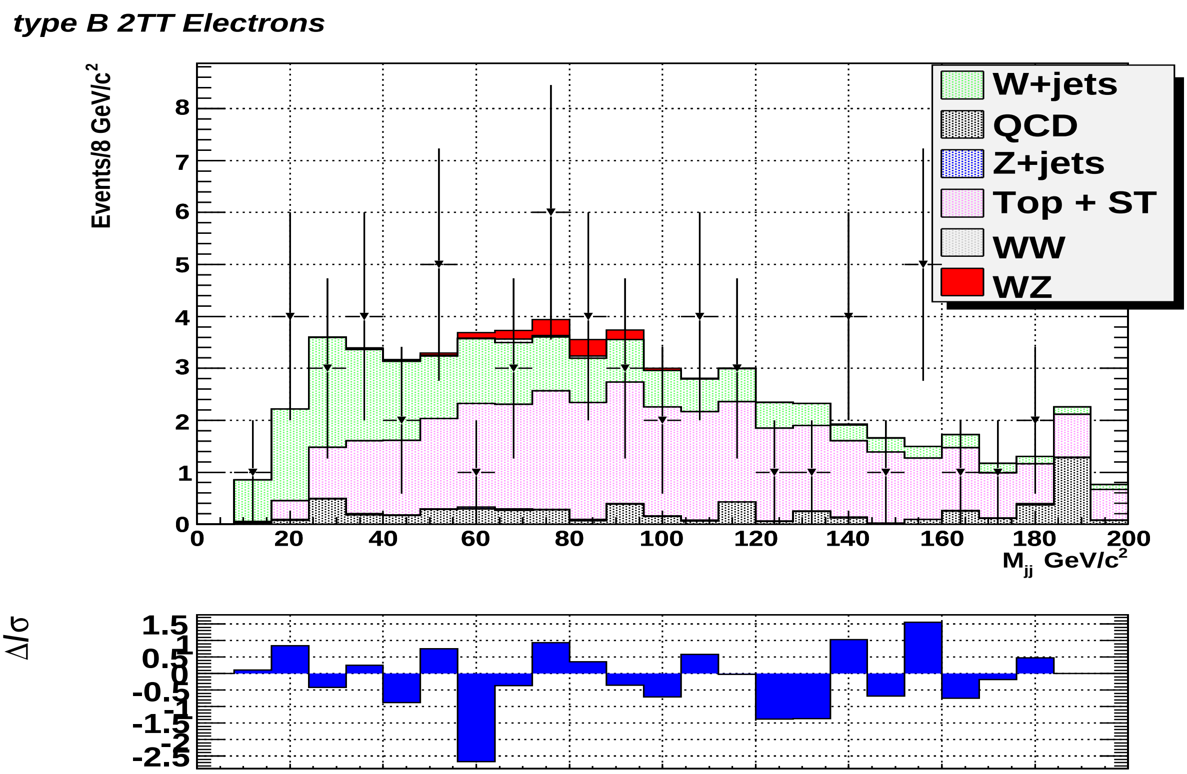


(a) electrons

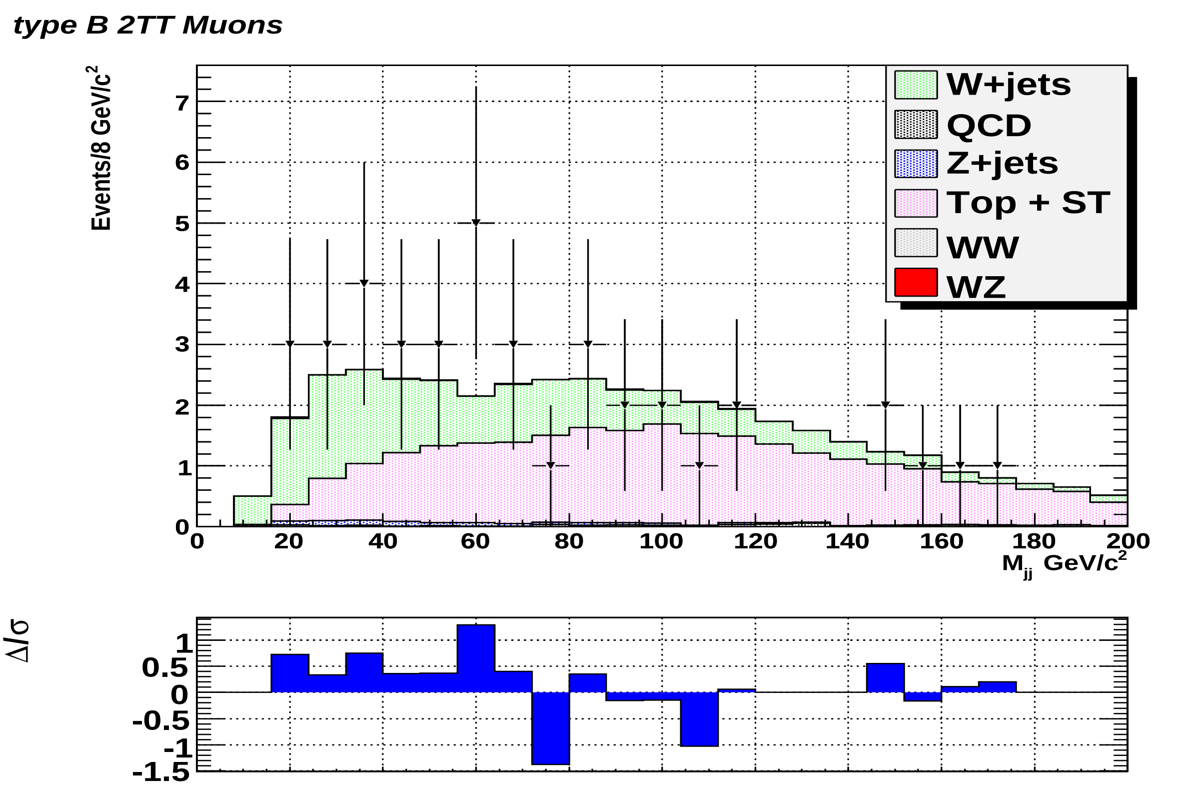


(b) muons

Figure B.1: type Bfit results on the 1Tsamples for muon and electrons.

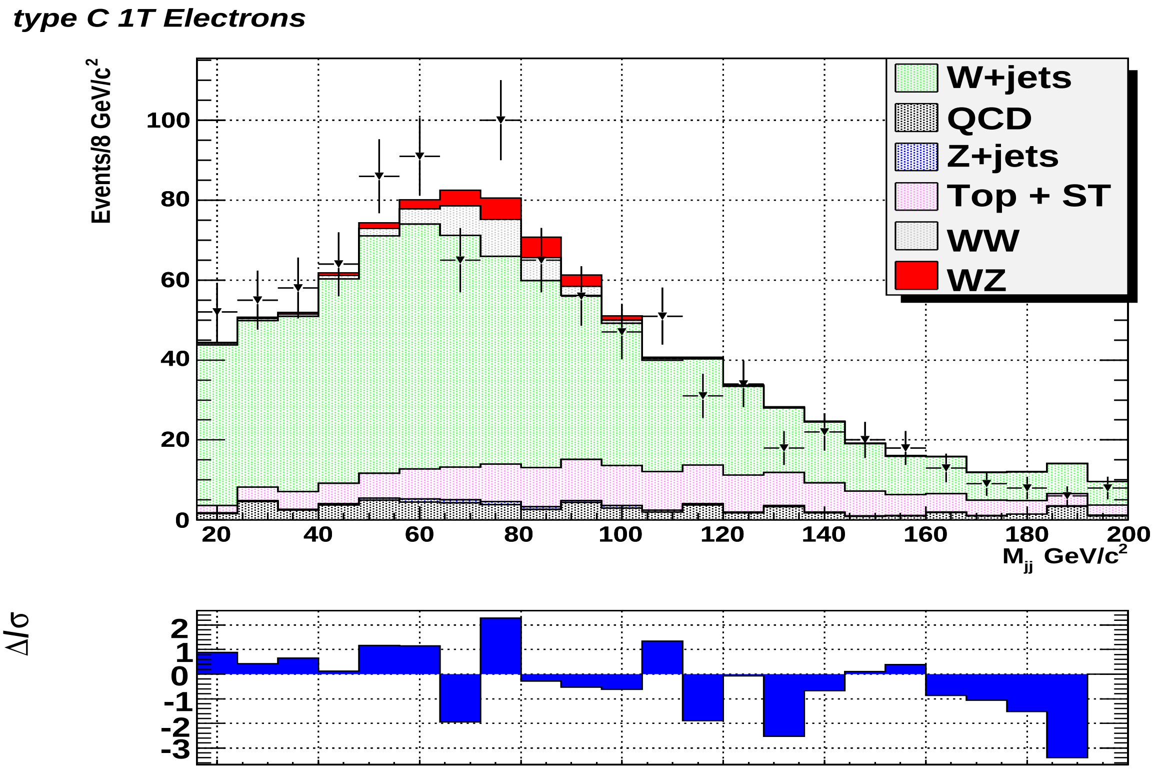


(a) electrons

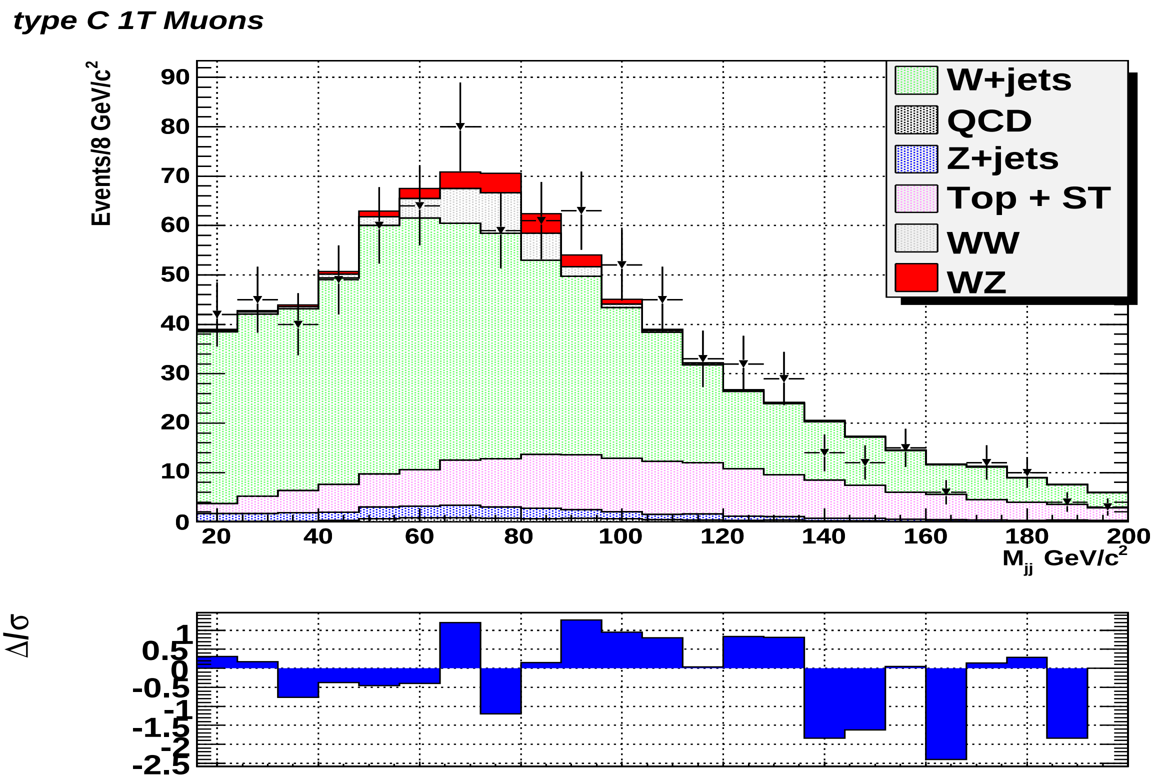


(b) muons

Figure B.2: type Bfit results on the 2TTsamples for muon and electrons.

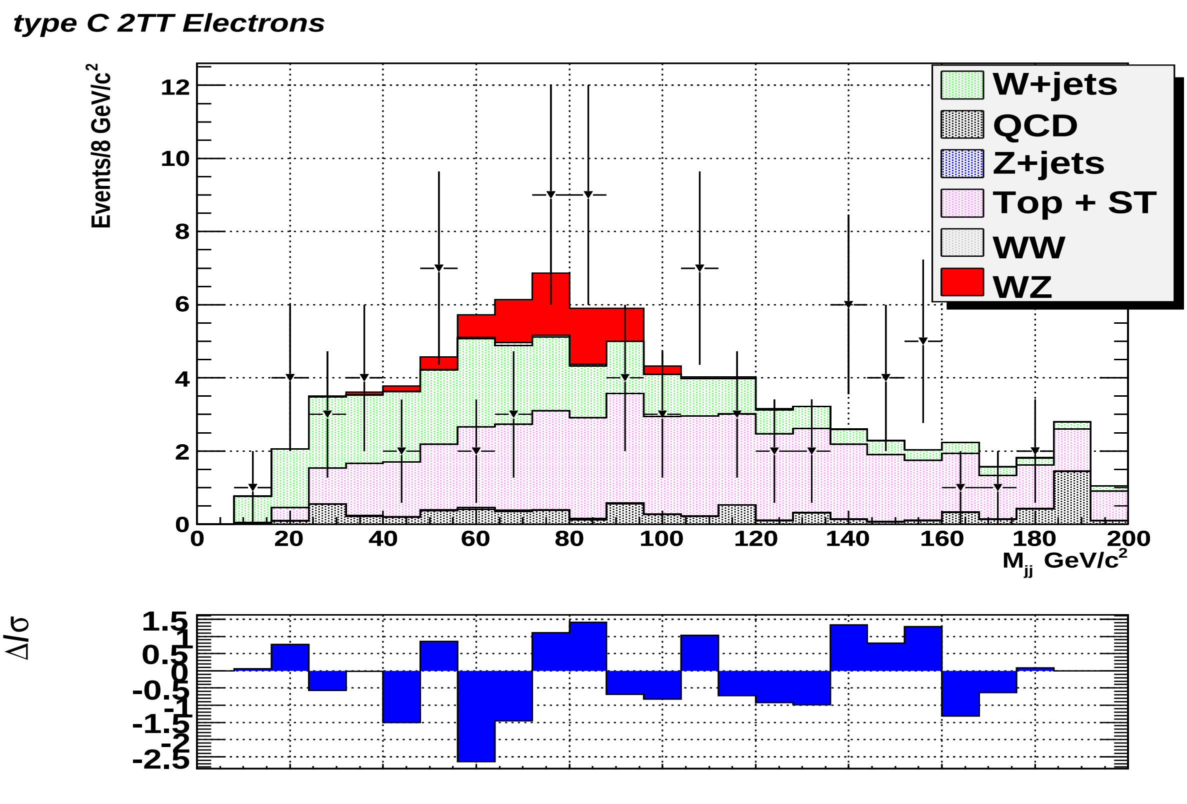


(a) electrons

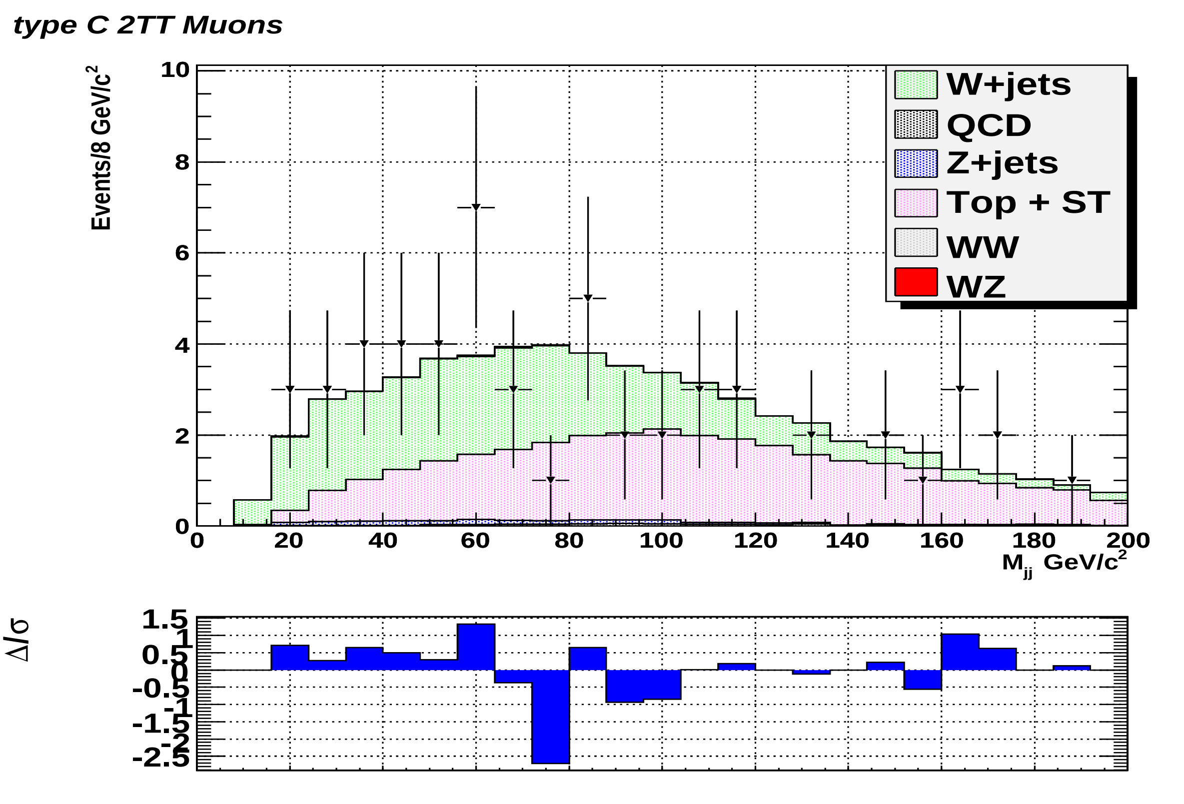


(b) muons

Figure B.3: type Cfit results on the 1Tsamples for muon and electrons.

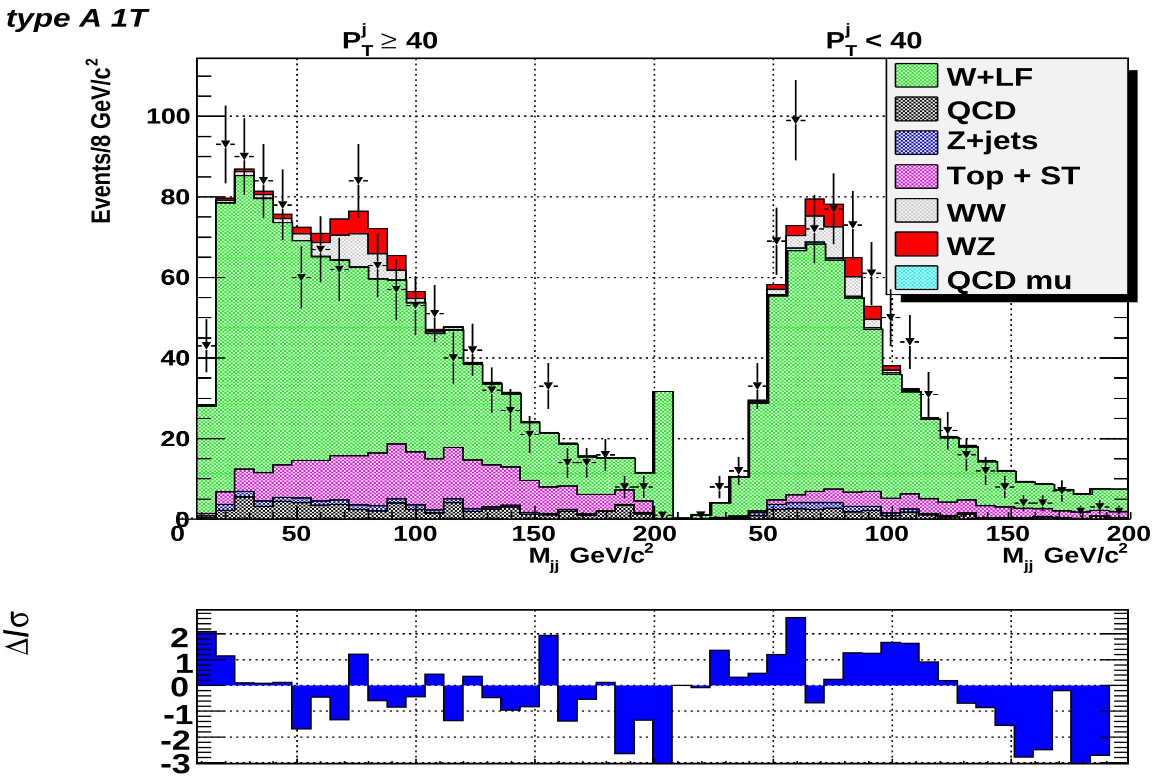


(a) electrons

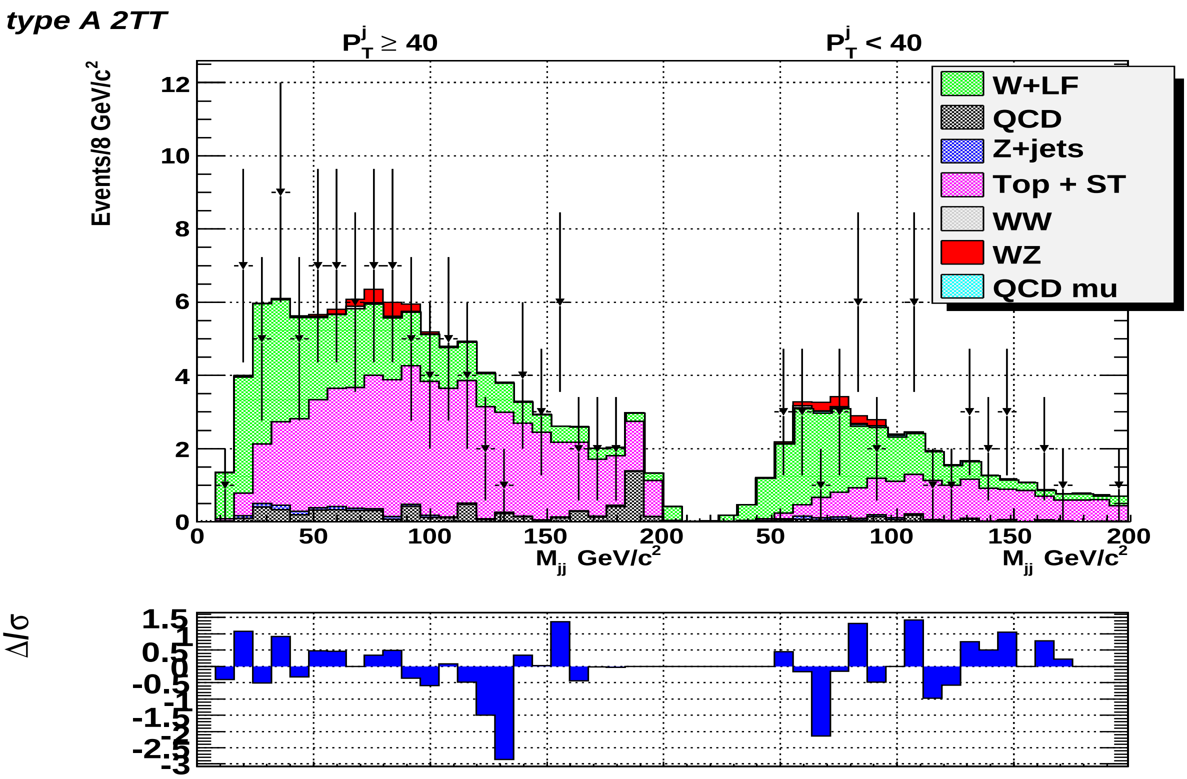


(b) muons

Figure B.4: type Cfit results on the 2TTsamples for muon and electrons.

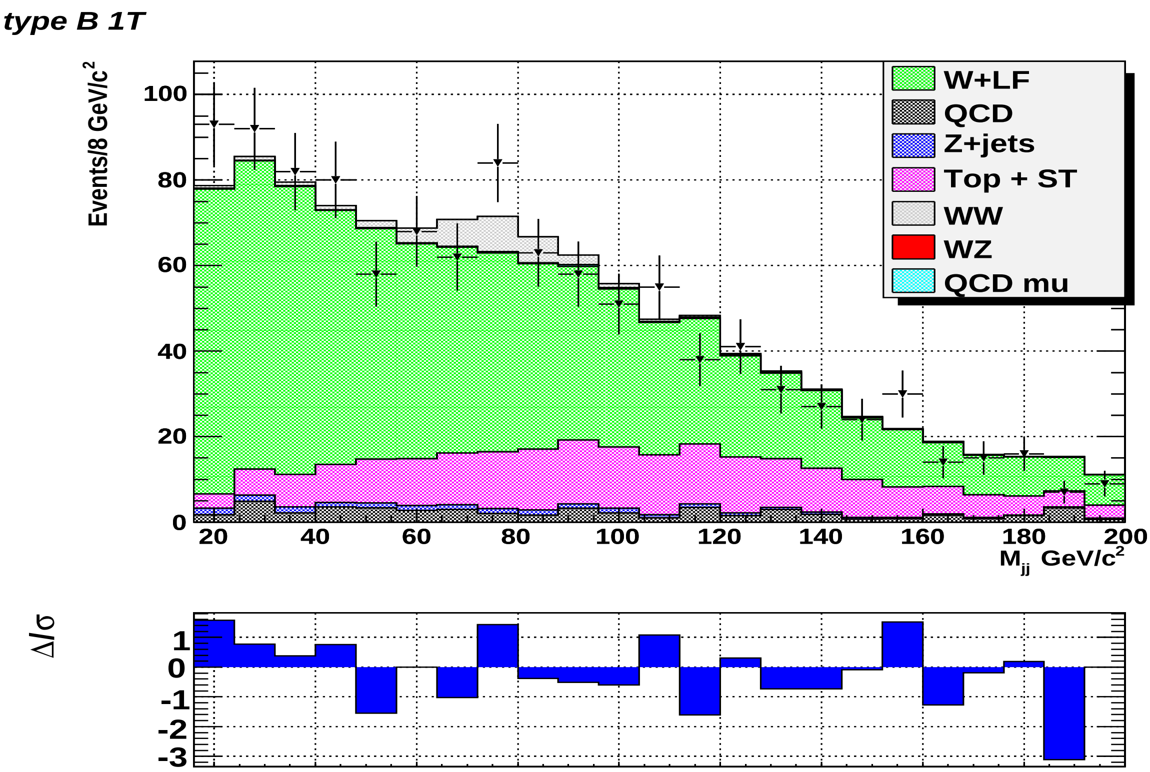


(a) 1T

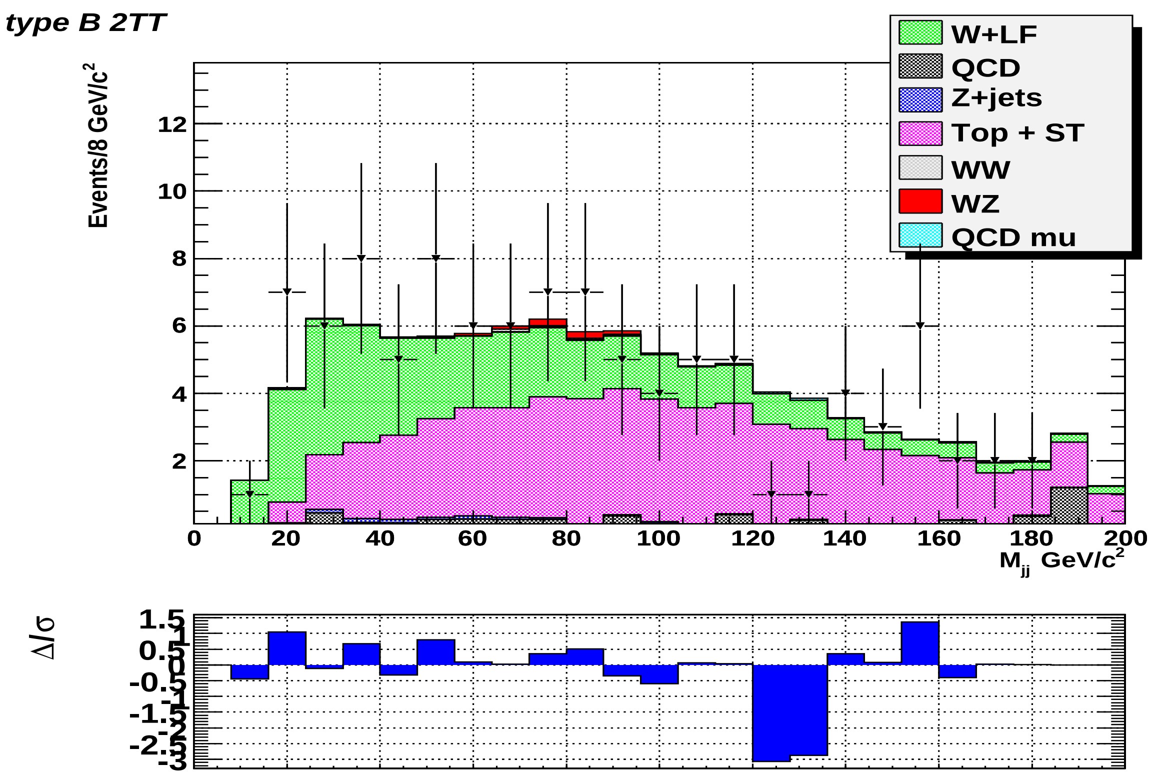


(b) 2TT

Figure B.5: type Afit results on the tagged samples for combined muon and electrons.

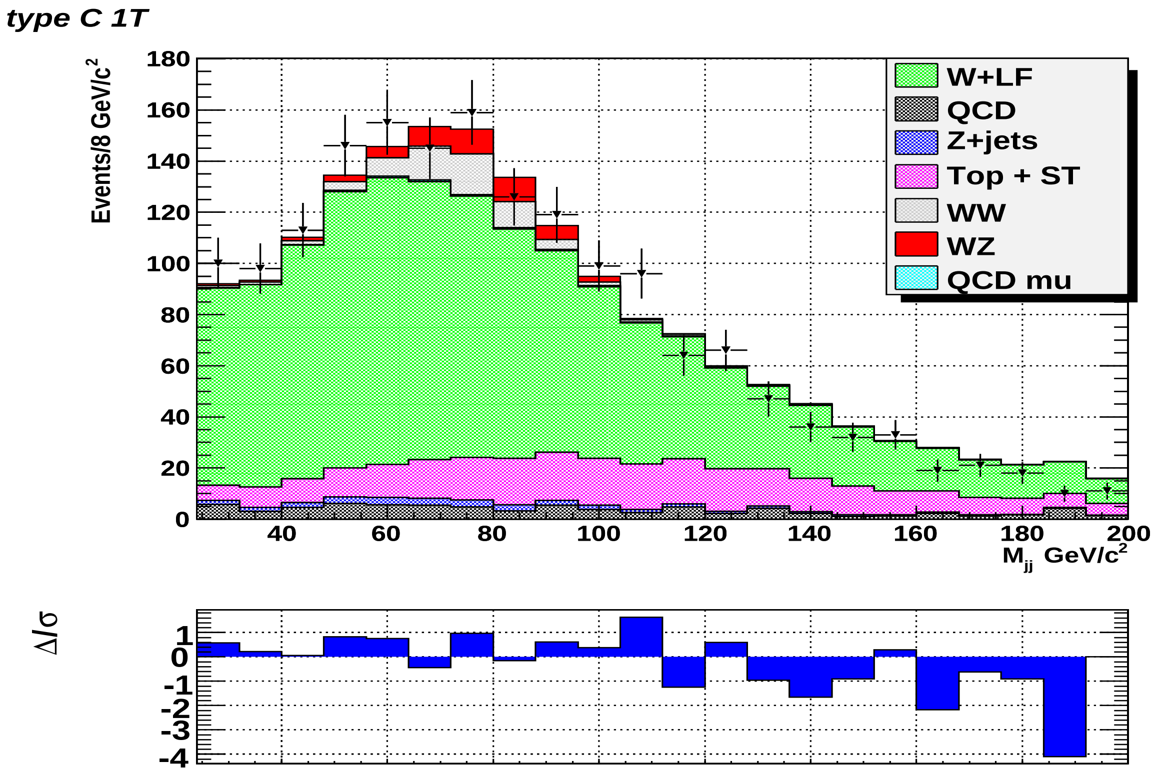


(a) 1T

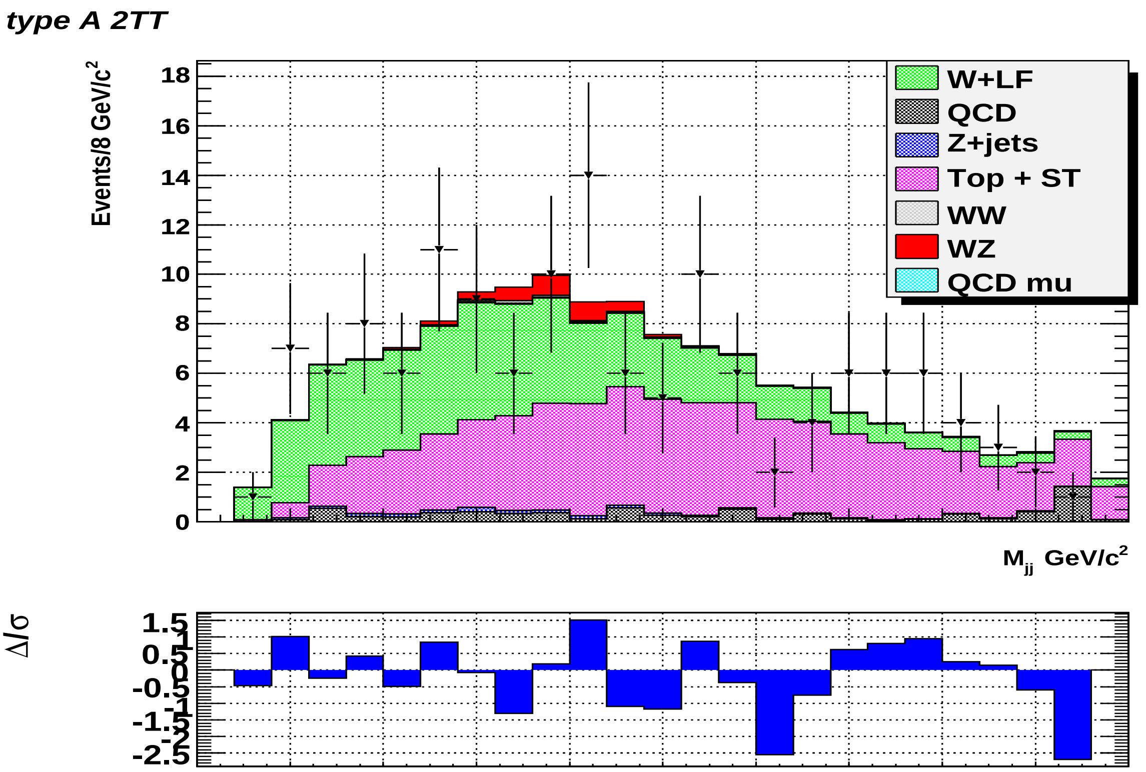


(b) 2TT

Figure B.6: type Bfit results on the tagged samples for combined muon and electrons.



(a) 1T



(b) 2TT

Figure B.7: type Cfit results on the tagged samples for combined muon and electrons.