Università degli Studi di Siena

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI

Corso di Laurea Specialistica in Fisica Sperimentale

Tesi di Laurea Specialistica

A multivariate approach for the extraction of $WW/WZ \to l\nu + Jets$ events at CDF II.

Candidato:

Januscia Duchini

Relatore:

Prof. Maria Agnese Ciocci

Correlatore:

Dott. Pierluigi Catastini

Controrelatore:

Prof. Angelo Scribano

Anno Accademico 2008–2009

Contents

Introduction

The Standard Model, the theory that describes the known particles and the way they interact, has proven very successful. It has been tested experimentally, and its predictions have been confirmed. However one important particle, the Higgs boson, is missing and some basic question have no answer yet.

The fundamental principles of nature can be studied using the particle accelerator for reach necessary energy to produce particles and detectors to detect them. Data used for the study presented in this thesis are collected by CDF II, one of the two detectors placed at Tevatron in USA.

In this thesis we describe a multivariate approach for the study of the process WW/WZ decaying in charged lepton, neutrino and jets, produced in $p\bar{p}$ collisions.

$WW/WZ \to \ell\bar{\nu}_\ell + jets$ process has been observed recently at CDF (2009) using two complementary methods. The corresponding cross section was measured to be $\sigma_{WW+WZ} = 16.0 \pm 3.3$ pb [1], in agreement with the standard model prediction of 16.1 ± 3.1 pb.

The aim of the study presented in this thesis is to improve the statistical sensitivity to the $WW/WZ \to \ell\bar{\nu}_\ell + jets$ signal estimator of [1, 2, 29] using two multivariate techniques. A such improvement translate in a more precise measurement of the cross section $p\bar{p} \to WW/WZ$ of [1, 2, 29].

This study starts from a 1-dimensional fit to the dijet invariant mass, (one of the two methods mentioned above [1, 2, 29]).

In this thesis we use only the muon channel for the leptonic decay of the W boson.

The signature of WW/WZ is reconstructed requiring one W boson decaying leptonically, with an high p_T muon and missing transverse energy (neutrino), and at least 2 jets for the hadronically decaying W or Z boson.

We determined kinematical quantities discriminating the signal from the background using simulated samples, then we used these quantities to train two discriminant methods: the

Boosted Decision Tree (BDT) and the Projective Likelihood (Likelihood).

After that we built 2-dimensional histograms for signal and backgrounds (as expected from simulated samples) and data. On one axis of each histogram is reported the jet jet invariant mass distribution (M_{jj}), on the other is reported the discriminant method output. We then perform a 2-dimentional fit on data histogram to extract the signal fraction (one of the parameter determined by the fit) using the expected distributions to parameterized the fit.

An interesting feature of this approach is that we simultaneously look for a peak in the M_{jj} and add kinematical informations of the full event, while the previous analysis performs an 1-dimentional fit on the Mjj distribution exploiting only the different properties of one variable for signal and background.

Before applying the fit procedure on data we validate and asses the performances of the 2-dimensional and 1-dimensional (on M_{jj}) fits using pseudo experiments.

The statistical sensitivity on the signal extraction, corresponds to an improvement of the 21% for BDT and of the 18% for the Likelihood with respect to the 1-dimentional M_{jj}, as determined using the pseudo experiments.

These results are also confirmed when applying the fits (2-dimensional for BDT, 2-dimensional for Likelihood and 1-dimensional on M_{jj}) on the data sample.

The techniques developed in this analysis are also applicable in the Higgs search.

For Higgs boson masses below 135 GeV/c^2, $b\bar{b}$ pairs is the main decay mode at Tevatron energies and the associated production with a W boson provide one of the most sensitive channel in this mass range. Moreover, since the final state of $WW/WZ \to \ell\bar{\nu}_\ell + jets$ is topologically similar to associated production of a Higgs boson with a W, it shares the same challenge of separating a small signal from a large background.

In Chapter 1 are described the theory and the motivations behind this analysis. In Chapter 2 are described the CDF II detector and Tevatron's performances. In Chapter 3 are described the reconstruction and the requirements for the selection of the $WW/WZ \to \mu\nu + jets$ candidate events on data. Next chapter (Chapter 4) is dedicated to the realistic Monte Carlo, used to generate, simulate and reconstruct the signal and the background samples for this analysis using the same requirements of Chapter 3. Chapter 5 explains the two multivariate analysis methods and the way we use them. In the same Chapter we validate and evaluate the performances of the fits through the pseudo-experiments. And finally in Chapter 6 we show the fits results on data.

Chapter 1

Theoretical Overview and Motivations

All experimental data from high energy experiments that has been collected so far can be accounted for by the so-called Standard Model (SM) of particles and their interactions. This was for the first time presented in a single report in 1974 [3]. Since then, the SM has been extended and their prediction tested at the smallest scale that have so far been explored and all (but one, the Higgs boson) of the particles that this theory predicts have thus far been discovered experimentally. It should be noted that the SM is not accepted as an accurate description of how the universe operates at its most fundamental level since it not includes the gravity; it represents only our best current knowledge.

The experiments at the Fermilab Tevatron proton antiproton collider and at the CERN new powerful proton proton collider LHC [4], are in an excellent position to give conclusive answers at many open questions of fundamental physics, as the existence of the Higgs boson (the only fundamental particle in the SM which has yet to be experimentally verified), the origin of the CP asymmetry and existence of physics beyond SM.

The following chapter provides an outline of this model and its workings and the theoretical motivation for the measurement of $WW/WZ \to l\nu + jets$.

1.1 The Standard Model

The SM of particle physics is a combination of theoretical constructs in the framework of the relativistic quantum field theory and experimentally measured quantities that taken together provide an attempt to describe the fundamental particles of matter and their interaction.

It describes three of the four known fundamental forces of nature: strong interactions, electromagnetic interactions, and weak interactions. The fourth known force, being gravity, is far weaker (roughly 40 orders of magnitude smaller than the strong nuclear force) and is not expected to contribute significantly to the physical processes which are of current interest in high energy particle physics. The SM is based on the gauge simmetry group $\mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ which is invariant under local gauge transformations. The C is a reminder that $\mathrm{SU}(3)$ represents the symmetry group of the colored strong interactions of quantum chromodynamics (QCD). The L indicates that the $\mathrm{SU}(2)$ group contains left-handed weak isospin doublets and the Y is a reminder that the $\mathrm{U}(1)$ group contains the right-handed weak hypercharge singlets. Together, the $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ groups govern the now unified electroweak force.

There are two main classifications of particles which the SM describes. These are the spin-1/2 fermions that are the constituents of normal matter and the integer spin bosons which are the mediators of the strong and electroweak forces. Particles in the SM acquire mass via interactions with a so-called Higgs field which is a result of a spontaneously broken symmetry arising in the $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ electroweak sector. The predicted Higgs boson resulting from this broken symmetry is the only particle in the SM which has yet to be experimentally verified.

1.1.1 The fundamental particles and the interactions

The fundamental fermions are six quarks and six leptons, which can be categorized into 3 distinct "generations" according to their mass hierarchy (Table 1.1).

Matter particles have antimatter equivalents, with opposite charge. The quarks carry an additional charge to the electric charge that is called color charge. There are three colors, and the corresponding anticolors. Collections of quarks and antiquarks form composite particles, known as hadrons. The hadrons are colorless and are divided into mesons (quark-antiquark states) and baryons (three quark states).

Each of the three interactions (forces) among the fundamental fermions (electromag-

	Particle Name	Mass (GeV/c^2)	Charge	Interaction
Leptons	electron (e)	511×10^{-6}	+1	EM, Weak
	e neutrino (ν_e)	$< 2.3 \times 10^{-6}$	0	Weak
	muon (μ)	106×10^{-3}	+1	EM,Weak
	μ neutrino(ν_μ)	$< 0.17 \times 10^{-3}$	0	Weak
	tau (τ)	1.78	+1	EM,Weak
	τ neutrino(ν_τ)	$< 15.5 \times 10^{-3}$	0	Weak
Quarks	up (u)	$(1.5 \div 3.3) \times 10^{-3}$	+2/3	Strong,EM, Weak
	down (d)	$(3.5 \div 6.0) \times 10^{-3}$	−1/3	Strong,EM, Weak
	charm (c)	$1.27^{+0.07}_{-0.11}$	+2/3	Strong,EM, Weak
	strange (s)	$(70 \div 130) \times 10^{-3}$	−1/3	Strong,EM, Weak
	top (t)	171.3 ± 2.1	+2/3	Strong,EM, Weak
	bottom (b)	$4.20^{+0.17}_{-0.07}$	−1/3	Strong,EM, Weak

Table 1.1: *The masses of the elementary particles in the Standard Model. The measurements are reported from Reference [5]. Why these are the masses of the fermions and why the mass of the top quark is so different from the other five quarks are two of the questions that the Standard Model leaves unanswered.*

netic, weak and strong) is mediated by the exchange of integer-spin particles called Gauge bosons The gauge bosons comprise eight colored gluons of the strong interactions, the photon of the electromagnetic interactions and the W^+ , W^- and Z of the weak interactions (Table 1.2). A diagram picturing the three exchange interactions is shown in Fig. 1.1 along with the corresponding effective coupling constants (α_s, α, α_W).

The gluon coupling is proportional to the color charge C_q and to the coupling constant α_s. This is similar to the situation in electrodynamics, where the coupling is proportional to the electric charge e_q and to the fine structure constant α. However, unlike in QED, the force carriers in QCD are colored, hence self-coupled. As a result of this self-interaction the strong force increases linearly with distance, making quarks tightly bound inside hadrons. The impossibility of separating color charges, such as individual quarks and gluons, is called color confinement. So far, no free quarks or gluons have been observed. They occur only in bound states which are color-neutral. The color charge is conserved. Only a color-neutral pair of color-anticolor quarks can be created in a collision. If in the final state quark and antiquark have large energies, color confinement degrades their momentum by

	Particle Name	Mass (GeV/c^2)	Charge	Interaction
Gauge Bosons	gluon (g)	0	0	Strong
	photon (γ)	0	0	EM
	W boson ($W^{\pm}$)	80.40 $\pm$ 0.03	$\pm$1	Weak
	Z boson	91.188$\pm$ 0.002	0	Weak

Table 1.2: *The gauge bosons of the Standard Model.*

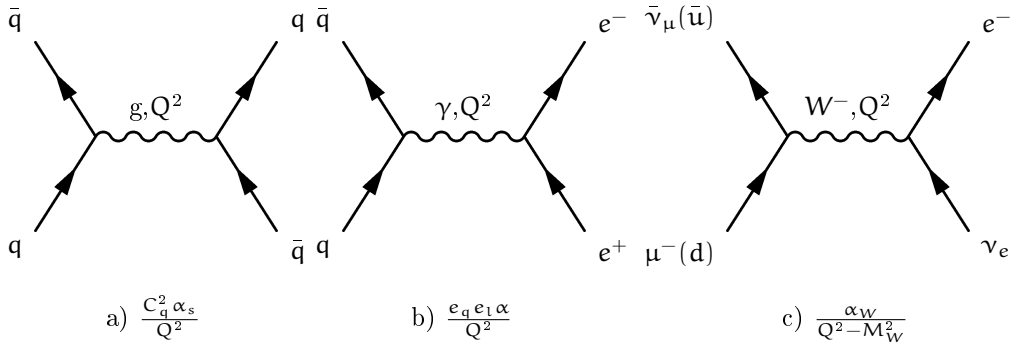


Figure 1.1: *Example of Feynman diagrams for (a) strong, (b) electromagnetic and (c) weak interactions.* Q^2 *is the square of 4-momentum transferred between the particles.*

radiating gluons or quark (parton) pairs. The new partons are approximately collinear with the original parton and combine into mesons or baryons in such a way that a spray of color-less particles is observed which move close to the same direction. This process is referred to as parton *fragmentation*, and the spray of collimated particles is called a jet. Energetic gluons trying to escape the interaction region undergo the same fragmentation process.

The coupling constant α_s in QCD is a function of transferred momentum Q^2. α_s decreases with increasing Q^2 and vanishes asymptotically. This leads to the property of QCD called *asymptotic freedom* and allows calculating the strong interaction cross sections perturbatively at high momentum transfer. This is often the case of collisions at the Tevatron, where it is possible to calculate interaction cross sections as perturbative expansions. However, in the process of jets formation the particle energies in the fragmentation process become successively smaller and perturbative QCD is no longer applicable. Phenomenological models are usually applied in order to describe completely jet features.

1.2 Electroweak Sector

Within the Standard Model, all the particles and the forces between them are described by a set of principles and equations[1]. The electromagnetic and weak interactions are combined into a unified theory, the Electroweak theory, specified by the $SU(2)_L \times U(1)_Y$ gauge symmetry. In this symmetry the lepton and quark doublets have left-handed weak isospin T and weak hypercharge Y that relates the electric charge Q to the third component of T, T_3 by

$$Q = T_3 + Y/2 \tag{1.1}$$

The $SU(2)_L \times U(1)_Y$ gauge symmetry introduces 4 fields; three connected with the weak isospin $(W^1_\mu, W^2_\mu, W^3_\mu)$ and one connected with the weak hypercharge (B_μ).
Linear combinations of these fields produce the physical photon (A_μ) and the weak bosons:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \pm W^2_\mu)$$

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\vartheta_W & \sin\vartheta_W \\ -\sin\vartheta_W & \cos\vartheta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix} \tag{1.2}$$

The weak mixing angle ϑ_W connects the coupling constants of the $SU(2)_L$ and $U(1)_Y$ interactions (g and g' respectively) to the electric charge:

$$g = \frac{e}{\sin\vartheta_W} \qquad g' = \frac{e}{\cos\vartheta_W} \tag{1.3}$$

In the Standard Model, the field A_μ (photon) has to be massless; in the QED Lagrangian the addition of a mass term $(\frac{1}{2}m^2 A_\mu A^\mu)$ is prohibited by gauge invariance. Gauge invariance of $SU(3)_C$ impose massless gluons. On the other hand, gauge invariance of $SU(2)_L \times U(1)_Y$ similarly implies massless weak bosons and fermions (all fermions, since all SM fermions have electroweak interaction). Despite the aesthetic qualities of the theory, reality causes problems. The resulting weak bosons (W and Z) and fermions do not quite connect with experimental results. The W and Z have both been measured to be very massive particles, but including terms in the Lagrangian to describe these masses destroys the $SU(2)_L \times U(1)_Y$ symmetry of the theory. The most accepted solution to this problem is known as the Higgs mechanism. The Higgs mechanism introduces an additional

[1] The work presented in this thesis is focused on the Electroweak sector of the Standard Model, so this will be discussed further in this chapter. Extensive discussions of the Electroweak theory can be found in textbooks. The ones used for this chapter are References [6] and [7]

term into the Lagrangian that represents two scalar fields along with an associated potential energy and then rearranges the Lagrangian such that the $SU(2)_L \times U(1)_Y$ symmetry is broken, or hidden. That which is left behind is a Lagrangian describing the electroweak interaction in terms of three massive weak bosons ($W^{\pm}$ and Z) and a massless photon (γ) plus an additional spin zero particle known as the Higgs Boson. The theory gives no accurate prediction for the mass of this Higgs boson itself and it has to date avoided experimental detection.

Gauge Boson masses

When breaking the $SU(2)_L \times U(1)_Y$ symmetry is required that a residual $U(1)$ symmetry remains for the electromagnetic interaction and this symmetry gives a massless boson. The remaining three gauge bosons must acquire mass in this process.
This is achieved by introducing four independent scalar fields that transform as an $SU(2)$ doublet:

$$\begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix} \tag{1.4}$$

where v represents the vacuum expectation value of the field φ and h (the Higgs field) excitations above this minimum. This process keeps the photon massless, while the weak bosons acquire masses determined by the vacuum expectation value and the coupling constants of $SU(2)$ and $U(1)$:

$$M_W = \frac{gv}{2} \qquad\qquad M_Z = \sqrt{(g^2 + g'^2)}\frac{v}{2} \tag{1.5}$$

The relation (at tree level) between the boson masses is given using the weak mixing angle:

$$\cos^2 \vartheta_W = \frac{M_W^2}{M_Z^2}. \tag{1.6}$$

Triple Gauge Coupling

To summarize the dynamics in the EWK, the Lagrangian is a useful formulation for the reason that it is independent of the coordinate system we choose. Hence, the Lagrangian of the symmetry group $SU(2)_L \times U(1)_Y$ can be separately into four terms as below [8]:

$$\mathscr{L}_{SM} = \mathscr{L}_{Gauge\_Higgs} + \mathscr{L}_{Gauge} + \mathscr{L}_{Gauge\_Fermion} + \mathscr{L}_{Higgs\_Fermion} \tag{1.7}$$

The $\mathscr{L}_{Gauge\_Higgs}$ term illustrates the interactions between gauge bosons and Higgs bosons, including the mass of gauge bosons. The second term, $\mathscr{L}_{Gauge}$, describes the

kinetic energy and self coupling of gauge bosons. The third term represents the interaction between fermions and gauge bosons and the last one is for the coupling between fermions and the Higgs bosons that explain how fermions get their mass. The couplings between three gauge bosons are included in the $\mathscr{L}_{\mathrm{Gauge}}$ and the field tensors of EWK are:

$$\begin{aligned} W_{\mu\nu} &= \partial_\mu W_\nu - \partial_\nu W_\mu + g W_\mu \times W_\nu \\ &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\varepsilon_{abc} W^b_\mu W^c_\nu \end{aligned} \tag{1.8}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.9}$$

and consequently $\mathscr{L}_{\mathrm{Gauge}}$ can be written:

$$\mathscr{L}_{\mathrm{Gauge}} = -\frac{1}{4} W_{\mu\nu} W^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.10}$$

Taking into account that

$$\begin{aligned} W^1_\mu &= \frac{W^+_\mu + W^-_\mu}{\sqrt{2}} \\ W^2_\mu &= \frac{W^+_\mu - W^-_\mu}{\sqrt{2}} \\ W^3_\mu &= \cos\vartheta_W Z_\mu + \sin\vartheta_W A_\mu \\ B_\mu &= -\sin\vartheta_W Z_\mu + \cos\vartheta_W A_\mu \end{aligned} \tag{1.11}$$

we get from the first term of the Lagrangian up to quartic interaction vertices between charged bosons or charged and neutral bosons (W^+W^-Z, $W^+W^-\gamma$, W^+W^-ZZ, $W^+W^-Z\gamma$, $W^+W^-\gamma\gamma$, $W^+W^-W^+W^-$). The second term of the Lagrangian, that includes only neutral bosons, involves in a vertex no more than two bosons. Triple or quartic interaction vertices of only neutral bosons (e.g. ZZZ, ZZγ, Z$\gamma\gamma$) are absent in the Standard Model.

1.3 The WW and WZ associated Production

In this thesis we study the associates production of Dibosons WW and WZ. The contributing tree-level Feynman diagrams are shown in Fig 1.2.

Direct WW and WZ production and boson triple gauge couplings occur in the s−channel as shown in (a) and (b). t−channel production is also present ((c) and (d)). Cancellations

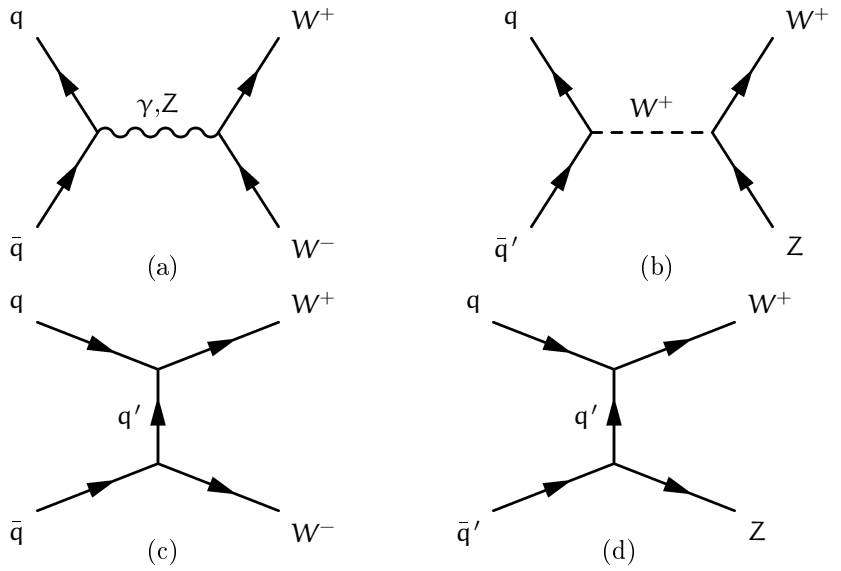


Figure 1.2: *WW and WZ production Feynman diagrams at three level.*

in the s− and t− channel diagrams result in low cross sections in the Standard Model. The *Next to Leading Order* (NLO) theoretical cross sections for the WW and WZ production at $\sqrt{s}$ =1.96 TeV are [9]:

$$\sigma_{WW} = 12.4 \pm 0.8 \text{ pb} \tag{1.12}$$
$$\sigma_{WZ} = 4.0 \pm 0.3 \text{ pb}$$

Possible experimental deviations of the WZ and WZ cross-sections from the Standard Model theoretical predictions are indications of non-standard triple ($WW(Z,\gamma)$ or $WZ(W)$) couplings. These are called *anomalous Triple Gauge Couplings* (*aTGCs*) [11, 12].

Moreover the Tevatron (and then LHC) is favorable for such measurement, compared to LEP [13] first because at an e^+e^- collider $e^+e^- \rightarrow W^{\pm}Z$ is forbidden by charge conservation, then because the production cross section of these processes increases with the available center of mass energy.

1.4 Diboson decay modes

The statistics of diboson events, which can be observed, depends on the chosen decay mode of each boson. Moreover diboson production signals have to be observed above particle

backgrounds. In Table. 1.3 are reported the $W^{\pm}$, Z and $W^{+}W^{-}$,WZ decay modes and their branching ratios.

$W^{\pm}$ decay modes	Branching Ratio (%) ([5])
$\ell^{\pm}\nu_{\ell}$	32.40 ± 0.27
hadrons	67.60 ± 0.27
Z^{0} decay modes	Branching Ratio (%) ([5])
$\ell^{+}\ell^{-}$	10.0974 ± 0.0069
invisible	20.00 ± 0.06
hadrons	69.91 ± 0.06

$W^{+}W^{-}$ decay modes	Branching Ratio (%)
$\ell^{+}\nu_{\ell}\ \ell'^{+}\nu_{\ell'}$	10.5
$\ell^{+}\nu_{\ell}$+hadrons	43.8
hadrons	45.7
$W^{\pm}Z^{0}$ decay modes	Branching Ratio (%)
$\ell^{\pm}\nu_{\ell}\ \ell'^{+}\ell'^{-}$	3.3
$\ell^{\pm}\nu_{\ell}\ \nu\nu$	6.5
$\ell^{\pm}\nu_{\ell}$+hadrons	22.6
hadrons+$\ell'^{+}\ell'^{-}$	6.8
hadrons+$\nu\nu$	13.5
hadrons	47.3

Table 1.3: *The $W^{\pm}Z$ and WW and WZ decay modes and branching ratios. The boson branching ratios are known with good precision, we therefore consider the precision in the diboson branching ratios negligible. In this table, ℓ stands for a lepton, an electron, a muon or a tau. Experimentally, taus are hard to reconstruct.*

The leptonic decay mode

Diboson production in the fully leptonic decay channel is rare because of low branching ratios. However it provides essentially background free signals. This is in fact the channel in which the WW and WZ production has been observed at hadron colliders, and the

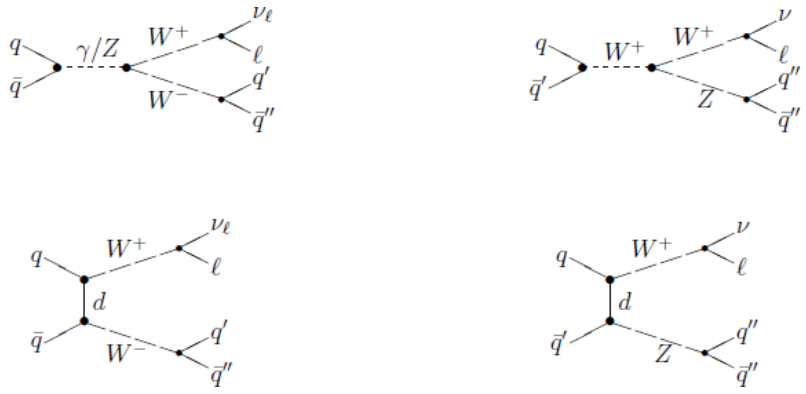


Figure 1.3: *Feynman diagrams for WW and WZ production in the semi-leptonic decay channel; there are similar diagrams for* $W^- \to \ell\nu_\ell$ *and* $W^- \to \ell q\bar{q}$

Standard Model predictions for the cross section have been verified [14].

The semi-leptonic decay mode

The WW and WZ production in the charged lepton, neutrino plus jets decay channel will be studied in this thesis. The Feynman diagrams for this decay mode are shown in Fig. 1.3. The hadronically decaying $W(W \to jj)$ can not be differentiated from a hadronically decaying $Z(Z \to jj)$ due to the limited jet energy resolution[2]. We therefore study the WW and WZ production together.

This decay mode has been observed recently at hadron colliders due to the large $W + jets$ background. The cross section of $W + jets$ production at $\sqrt{s} = 1.96$ TeV, where the W decays leptonically is of the order of 3 nb[15], resulting in a signal over background ratio that is smaller than 1%, making therefore the signal very difficult to observe. However, the branching ratio for this decay mode is very significant, making this channel favorable for aTGC studies.

The hadronic decay mode

Purely hadronic decays of gauge bosons are dominated by large four-jet QCD background, they are therefore practically unobservable at hadron colliders.

[2]When the quarks hadronize inside the detector, a collimated flow of high energy hadrons is generated. This is defined as a jet.

1.5 Motivation of our study

We study the production of events containing a W boson that decays leptonically ($W \to \ell\nu_\ell$[3]) in association with a W/Z boson that decays hadronically ($W \to q\bar{q}'$ or $Z \to q\bar{q}$). The studies are performed on data collected by the CDF II detector at Tevatron.

A primary motivation for studying diboson physics is that their production and interactions provide a fundamental test in electroweak sector of the SM. Diboson production can be studied by measuring their interactions and their production cross sections via trilinear gauge-boson couplings (TGC) [16, 17, 18]. The deviation of a TGC or production cross section from the values predicted by the SM would be an indication of new physics beyond the SM and could give a clue about the mechanism responsible for electroweak symmetry breaking.

The chosen decay channel has a large background, however this decay mode has a larger branching ratio than the cleaner fully leptonic decay mode. Our decay mode can be more sensitive to the anomalous TGC if we are able to disentangle signal from background events.

Given that the signal over background is initially very small (smaller than 1%), new and powerful statistic tecniques need to be used and tested to obtain optimal separation between signal decay mode and background events.

For Higgs boson masses below 135 GeV/c^2, $b\bar{b}$ pairs is the main decay mode [19] and at Tevatron energies associated production with a W boson (with $W \to \ell\nu_\ell$) provide one of the most sensitive channel in this mass range [28].

The most promising channel for low mass Higgs discovering is topologically similar to $WW/WZ \to \ell\nu_\ell + \text{jets}$ (see Fig. 1.4)

In both cases, the final state particles are a lepton and neutrino from the decay of a W boson and a quark-antiquark pair from the decay of either the Higgs or a weak gauge boson (W or Z). One consequence of this similarity is that $W^+W^-/W^\pm Z \to \ell\nu_\ell q\bar{q}$ is an important background for these Higgs searches. Making this direct measurement of diboson production supplies an in situ measurement of the size of this background. The Higgs boson searches and this analysis share the same challenge of separating a small signal from a large background. Of course, this analysis has the advantage that the signal is in several times larger and that the W and Z masses are known. Additionally,

[3] through this thesis we use $W \to \ell\nu_\ell$ for $W^+ \to \ell^+\nu_\ell$ or $W^- \to \ell^-\bar{\nu}_\ell$. Moreover, as we'll specify later, we reconstruct only $W \to \mu\nu_\mu$ leptonic decay modes.

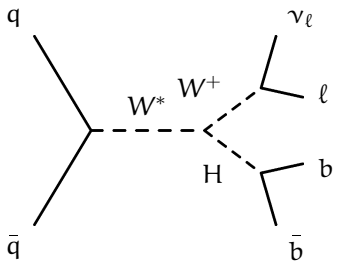


Figure 1.4: *A Higgs produced in association with a W (Higgs-Strahlung). This is a process topologically similar to the WW production in the semi-leptonic decay channel.*

the ability to discover the Higgs boson (and to observe with high significance final state $W^+W^-/W^\pm Z \to \ell\nu_\ell q\bar{q}$) depends largely on how precisely a resonance in the dijet invariant mass is reconstructed. A measurement of $W^+W^-/W^\pm Z \to \ell\nu_\ell q\bar{q}$ would determine the actual resolution of the measurement of a dijet resonance. Weak diboson production is also a significant background for high mass SM Higgs boson ($M_H \gtrsim 140$ GeV/c^2), in which the search focuses on $H \to W^+W^-$ decays. As in the low mass Higgs scenario, both the magnitude and the kinematics of diboson production impact the power of the search. In summary one can say that a measurement of $W^+W^-/W^\pm Z \to \ell\nu_\ell q\bar{q}$ production provides a "*standard candle*" with which to calibrate and optimize many of the techniques used in SM Higgs searches. The event selection for this search shares most of the trigger, Monte Carlo (MC) simulation, and normalization methods of the Higgs boson search. The multivariate event classification schemes, used in this thesis, that are becoming increasingly popular in Higgs boson searches can also be checked using a known signal. Finally, the statistical techniques used for the entire SM Higgs mass range can be tested on this known signal, providing opportunities for optimization.

1.6 Review of the Tevatron Diboson measurements

In this section we present a brief review of measurement of diboson production *WW*,*WZ* and *ZZ* at the Tevatron. The purpose is to establish the general status of the field, and to summarize the experimental issues relevant to the analysis presented in this thesis.

1.6.1 Overview of measurements

The CDF and DØ experiments have recorded about 7 fb^{-1} each of integrated $p\bar{p}$ luminosity up to January 2010. Recent measurement of most diboson channels are based on few fb^{-1}

of data. We present last updated measurements from CDF and DØ experiments in order to establish measurements precisions.

Representative measurements of production cross sections in diboson channels are presented in Table.1.4. In addition to the measurement (σ_{data}), the used integrated luminosity (L) and the theoretical cross section (σ_{theo}) of diboson associated production [9] are reported.

For the WW and WZ channels the signals are at the level of hundreds of events [20, 21, 22, 23], with a significant signal just starting to appear in the ZZ channel, where just few events were observed [24, 25].

	Process	**Source**	**L** $[fb^{-1}]$	**σ(data)[pb]** $\pm(stat) \pm (sys) \pm (lum)$	**σ(theory)** **[pb]**
1.	$W^+W^- \to \ell\nu\ell\nu$	CDF[20]	3.6	$12.1 \pm 0.9(stat)^{+1.6}_{-1.4}(sys)$	12.0 ± 0.7
	$\ell = e, \mu$	DØ[21]	1	$11.5 \pm 2.1(stat+sys) \pm 0.7(lumi)$	"
2.	$W^{\pm}Z \to \ell\nu\ell\ell$	CDF[22]	1.1	$5.0^{+1.8}_{1.6}(total)$	3.7 ± 0.3
	$\ell = e, \mu$	DØ[23]	1.0	$2.7^{+1.7}_{-1.3}(total)$	"
3.	$ZZ \to \ell\ell\ell\ell$	CDF[24]	1.9	$1.4^{+0.7}_{-0.6}(stat+sys)$	1.4 ± 0.1
	$\ell = e, \mu$	DØ[25]	1	1.71 ± 0.15	"
4.	$WW + WZ + ZZ$ $\to \ell\nu + jets + \nu\nu + jets$	CDF[26]	3.5	$18.0 \pm 2.8 \pm 2.4 \pm 1.1$	16.8 ± 0.5
5.	$WW + WZ$	CDF[1, 27]	2.7	$17.7 \pm 3.1(stat) \pm 2.4(sys)$	16.1 ± 0.9
	$\to \ell\nu + jets$	CDF[1, 2, 29]	3.9	$14.4 \pm 3.1(stat) \pm 2.2(sys)$	"

Table 1.4: *Summary of Tevatron* $p\bar{p}$ *diboson production cross sections. Total production cross-section are quoted*

As shown in Table.1.4, there is in general good agreement between the measured cross-sections and theory predictions, but with the exception of process 1, in Table.1.4, all measurements are dominated by statistical uncertainty and the precision of the measurement is still far from the teoretical prediction.

In this analysis we propose and test a multivariate approach to improve the statistical precision on the extraction of signal events, of reference [1, 2, 29] (process 5 of the Table.1.4), this measurement is dominated by the statistical uncertainty due to the extraction of signal events.

Chapter 2

The Tevatron Collider and the CDF II Detector

The data used in this thesis were collected by the upgraded Collider Detector (CDF II) at the Fermilab Tevatron Collider. This chapter provides a general description of the complex infrastructure, accelerator and detector, involved in producing and in recording the data sample and focus on elements having a crucial role in the diboson analysis.

2.1 Tevatron Collider

The Tevatron collider is a circular proton synchrotron 1 Km radius, and it is the last and the highest energy accelerating stage of a complex system of accelerators, storage rings, and transfer lines located at the Fermi National Accelerator Laboratory (FNAL or Fermilab), about 50 km west from Chicago, Illinois, United States.

While operating in collider mode, it collides bunches of protons spaced by 396 ns circulating clockwise against similar bunches of antiprotons accelerated counter-clockwise both at energies of 980 GeV, with a center-of-mass energy of 1.96 TeV. The collisions take place at two interaction points, conventionally named B$\emptyset$ and D$\emptyset$ (Fig. 2.1), in the center of the two detector CDF II and D$\emptyset$ II.

The Tevatron collider will be the world highest energy accelerator until the CERN Large Hadron Collider will start operating. It provides collisions of antiprotons with protons

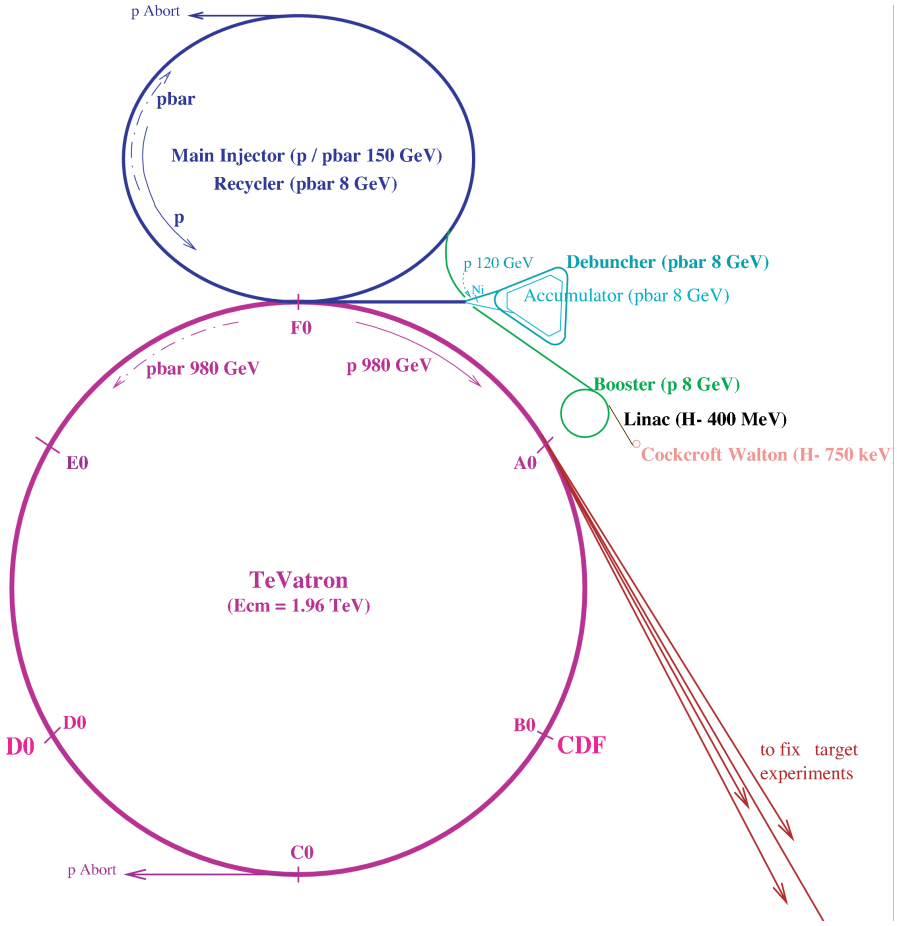


Figure 2.1: *The Fermilab accelerator chain.*

at a center-of-mass energy of 1.96 TeV. The Tevatron, an underground circular proton synchrotron, is the last stage of a system of accelerators, storage rings, and transfer lines located at the Fermi National Accelerator Laboratory (FNAL or Fermilab), about 50 km west from Chicago, Illinois, United States.

The Tevatron started operating in 1975 as the first superconducting synchrotron, the first $p\bar{p}$ collision occurred in 1985 and since the year 2002 it operates only in the collider mode. Between 1997 and 2001, both the accelerator complex and the collider detectors underwent major upgrades, mainly aimed at increasing the instantaneous luminosity (proportional to the increasing rate for a given final state, see Section 2.1.1) of the accelerator. The upgraded machine accelerates 36×36 bunches of protons and antiprotons, whereas the previous version of the accelerator operated with only 6×6 bunches. Consequently,

the time between two consecutive bunch crossings has been decreased from $3.5\mu s$ for the previous version to 396 ns for the current collider. By the fall of 2005, the anti-protons production chain has been further improved introducing the electron cooling [31] in the Recycler (Section 2.2). In the year 2009 the $3.8 \times 10^{32}\ cm^{-2}s^{-1}$ peak luminosity has been reached. Then, until the end of the Tevatron operations, the peak luminosity is expected to be $3 \times 10^{32}\ cm^{-2}s^{-1}$, with an expected total integrated luminosity of $9 - 10\ fb^{-1}$ at the end of 2010 (Fig. 2.2).

The new accelerator configuration required significant upgrades of the CDF II detector to ensure a maximum response in a time shorter than the time between two consecutive beam crossings.

Thanks to the Tevatron high luminosity (Section 2.1.1) and to the upgraded detector (Section 2.2), large samples of data have been collected by the CDF II detector and are now available for the analysis. Currently the cross sections of interesting processes are of the same order of magnitude for Tevatron and LHC experiments (since the available center-of-mass energy is nearly the same), while the instantaneous luminosity is several order of magnitude higher for the Tevatron experiments, leaving room for potential discoveries before the LHC reaches the design performances.

The data sample used in this thesis has been collected by the CDF II detector between March 2002 and December 2008, and it corresponds to an integrated luminosity of $3.9\ fb^{-1}$[1].

In the following sections, we will describe first how the protons and antiprotons beams are produced, accelerated to the final energy of 980 GeV and collided (Section 2.2) and then the Tevatron performances (Section 2.1.2).

2.1.1 $p\bar{p}$ Production

In order to create the particle beams Fermilab uses a series of accelerators. Fig. 2.1 shows the paths taken by protons and antiprotons from the initial acceleration to collision in the Tevatron.

For the purpose of the present analysis, we will describe the procedure for obtaining a continuous period of collider operation using the same collection of protons and antiprotons, called a *store*.

[1] The analysis described has been made with $3.9\ fb^{-1}$ of data but the update of the analysis on all the available sample of data is in progress at the time of the writing.

The Cockcroft-Walton [32] pre-accelerator provides the first stage of the acceleration. Inside this device, hydrogen gas is ionized to create H^- ions, which are accelerated to 750 keV of kinetic energy. Next, the H^- ions enter a linear accelerator (Linac) [33] , approximately 150 m long, where they are accelerated to 400 MeV. The acceleration in the Linac is done by a series of "kicks" from Radio Frequency (RF) cavities. The oscillating electric field of the RF cavities groups the ions into bunches.

The 400 MeV H^- ions are then injected into the Booster, a circular synchrotron [34] with a diameter of 7.45 m. A carbon foil strips the electrons from the H^- ions at injection, leaving bare protons. The intensity of the protons beam is increased by injecting new protons into the same orbit as the circulating ones. The protons are accelerated from 400 MeV to 8 GeV by a series of "kicks" applied by RF cavities. Each turn around the Booster, the protons accrue about 500 keV of kinetic energy.

Protons are extracted from the Booster into the Main Injector [34], which operates at 53 MHz. The Main Injector accelerates protons from 8 GeV to 150 GeV before injection into the Tevatron, it produces the 120 GeV protons used for antiprotons production, it receives antiprotons from the Antiproton Source and accelerates them to 150 GeV for injection into the Tevatron, and finally, it injects protons and antiprotons in the Tevatron.

The Main Injector replaced the Main Ring accelerator which was situated in the Tevatron tunnel. The Main Injector is capable of containing larger proton currents than its predecessor, which results in a higher rate of antiprotons production. The Main Injector tunnel also houses the Antiproton Recycler. Not all antiprotons in a given store are used for the collisions. Recycling the unused antiprotons and reusing them in the next store significantly reduces the stacking time. The task of the Antiproton Recycler is to receive antiprotons from a Tevatron store, cool them and re-integrate them into the stack, so that they can be used in the next store. To produce antiprotons, 120 GeV protons from the Main Injector are directed into a nichel target. In the collisions, about 20 antiprotons are produced per one million protons, with a mean kinetic energy of 8 GeV. The antiprotons are focused by a lithium lens and separated from other particle species by a pulsed magnet. The antiprotons are stored in the Accumulator ring. Once a sufficient number of antiprotons have been produced, they are sent to the Main Injector and accelerated to 150 GeV. Finally both, the protons and antiprotons, are injected into the Tevatron. The protons and antiprotons travel around the Tevatron, which uses superconducting dipole magnets, each one producing a field of 4.2 T, to bend the beams around the ring, in opposite directions.

The beams are kept apart with electrostatic separators and arranged in 3 trains of 12 bunches with 396 ns separation. The two beams are focused by quadrupole magnets that reduce the beam size increasing the luminosity (see Section 2.1.2) in the two interaction points (low β points), where the center of CDF II and D$\emptyset$ II are respectively located.

2.1.2 Perfomances of Tevatron

The performances of the Tevatron collider are evaluated in terms of two key parameters: the available center-of-mass energy, $\sqrt{s}$, and the instantaneous luminosity, $\mathcal{L}$. The former defines the accessible phase-space for the production of resonances in the final states. The latter is the coefficient of proportionality between the rate of a given process and its cross-section σ.

For physics studies the integrated luminosity, $\int \mathcal{L} dt$, is the quantity of interest, since for a specific process of cross-section σ, the number of events that are generated in a specific time interval are:

$$N = \sigma \times \int \mathcal{L} dt \tag{2.1}$$

To increase the instantaneous luminosity and so the integrated one, means to increase the number of events of the processes we are interested in.

Assuming an ideal head-on $p\bar{p}$ collision with no crossing angle between the beams, the instantaneous luminosity is approximated by:

$$\mathcal{L} = \frac{f \cdot B \cdot N_p \cdot N_{\bar{p}} \cdot F}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)} \tag{2.2}$$

where f is the revolution frequency (typically 47.7 kHz), B the number of bunches in each beam (36 bunches), $N_p(N_{\bar{p}})$ is the number of protons (antiprotons) in a bunch (3.3×10^{11} protons and 3.6×10^{11} anti-protons at the injection) and $\sigma_p(\sigma_{\bar{p}})$ is the r.m.s. beam size at the collision points (at CDF II, ~20 μm). The beam width and its evolution around the accelerator can affect the collisions rate, and this dependence is expressed in the form factor F (typically 0.72). While collisions are taking place the luminosity decreases because of the beam-gas and beam-halo interactions. In the meantime, antiproton production and storage continues. When the antiproton stack is sufficiently large ($\simeq 10^{12}$ antiprotons) and the circulating beams are degraded (~ 14 hours) the detector high-voltages are switched off and the store is dumped. The beam is extracted via a switch-yard and sent to an absorption zone. Beam abortion can occur also accidentally when the temperature of a superconducting magnet shift above the critical value and the

magnet quenches destroying the orbit of the beams. The typical time between the end of a store and the beginning of collisions of the next one is typically 2 hr.

At CDF II, the instantaneous luminosity at the beginning of the store typically exceeds 10^{32} collisions/(cm^2 s) (Fig. 2.2) with the peak initial luminosity currently being at 3×10^{32} collisions/(cm^2 s). The integrated luminosity is expressed in units of (cross-section)$^{-1}$. At CDF II, there have been so far (end 2009) more than 7 fb^{-1} of integrated luminosity delivered and more than about 6 fb^{-1} are already on tape (Fig. 2.2). At the end 2010, approximately 9-10 fb^{-1} of data are expected to be delivered at CDF.

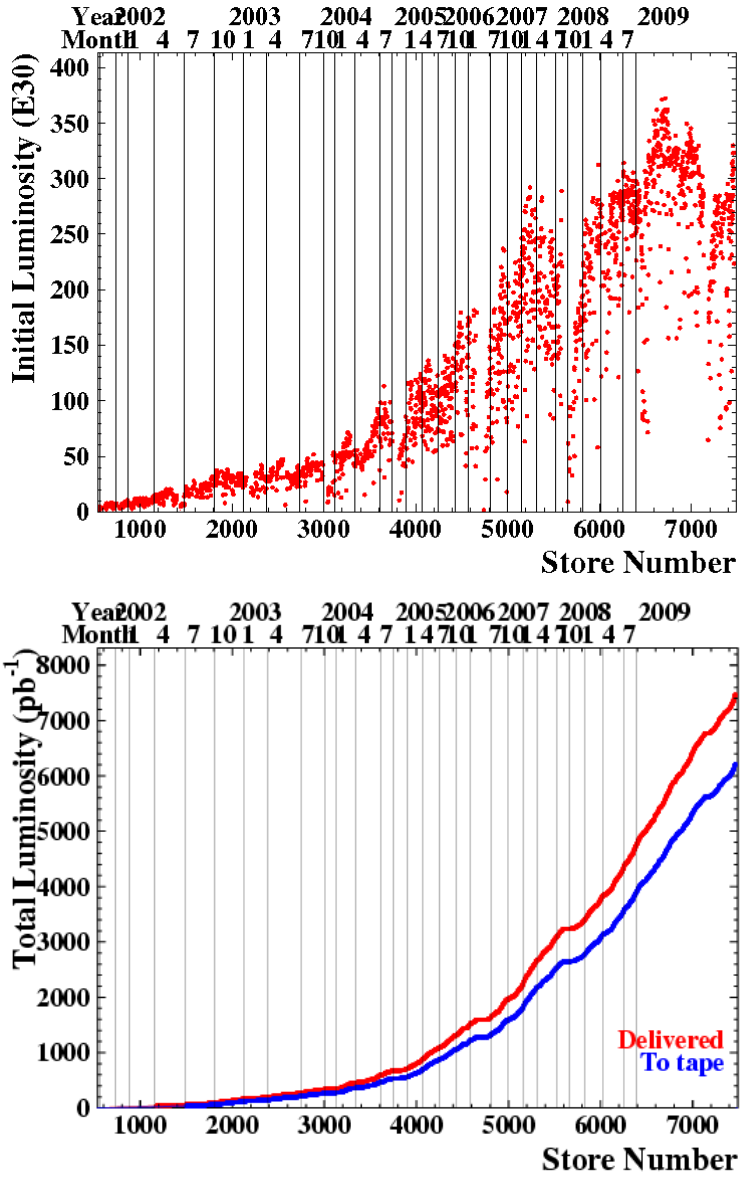


Figure 2.2: *Top: Peak luminosity as a function of time, measured using the store number. Bottom: Integrated luminosity as a function of time, measured using the store number. The integrated luminsosity recorded on tape is reduced by the data quality requirements.*

2.2 CDF II Detector

The upgraded Collider Detector at Fermilab (CDF II) is a complex but general purpose detector designed to investigate interesting events that are produced in $p\bar{p}$ collision at $\sqrt{s}$ = 1.96 TeV.

It is a large solenoidal magnetic spectrometer surrounded by almost full coverage, projective calorimeters and fine-grained muon detectors. CDF II [35, 36] (see Fig. 2.3) has a cylindrical geometry with axial and forward-backward symmetry. In the coordinate system of the CDF detector, the direction of the proton beam is defined as the $+z$ (East) direction. The $+x$ and $+y$ axis are chosen to be outward and upward from the Tevatron ring. The polar angle ϑ is measured from the z axis and the azimuthal angle φ is measured from the $+x$ axis. The pseudorapidity η is defined in terms of the angle relative to the beam axis, $\eta = -\ln(\tan(\vartheta/2))$.

The detector is installed at the B$\emptyset$ interaction point of the Tevatron (see Fig. 2.1). It comprises a number of coaxial sub-detectors that provide different information by which it is possible to determine energy, momentum and in a number of cases, nature of a broad range of particles produced in 1.96 TeV $p\bar{p}$ collisions:

- a tracking system composed by three silicon microstrip trackers (L00, SVX II and ISL, from inner to outer radii) and an open-cell drift chamber (COT) housed inside a superconducting solenoid providing a 1.4 T magnetic field. In this system the trajectories (helices) of charged particles are reconstructed, and the momentum and charge of particles are extracted from the bending of the tracks.

- a time of flight detector, radially outside the COT for particle identification up to momenta of few GeV.

- a set of calorimetric detectors located outside the magnet, divided into two sections, an electromagnetic (**EM**) section for electrons and photons, followed by a hadron (**HAD**) section to measure the energy of charged and neutral hadrons and jets.

- a system of muon chambers plus scintillators, used to track and identify muons that pass through the calorimeters interacting as minimum-ionizing-particles (`MIP`) .

- two small angle spectrometers in the very forward and backward regions with respect to the main detector for specialized studies of diffraction processes;

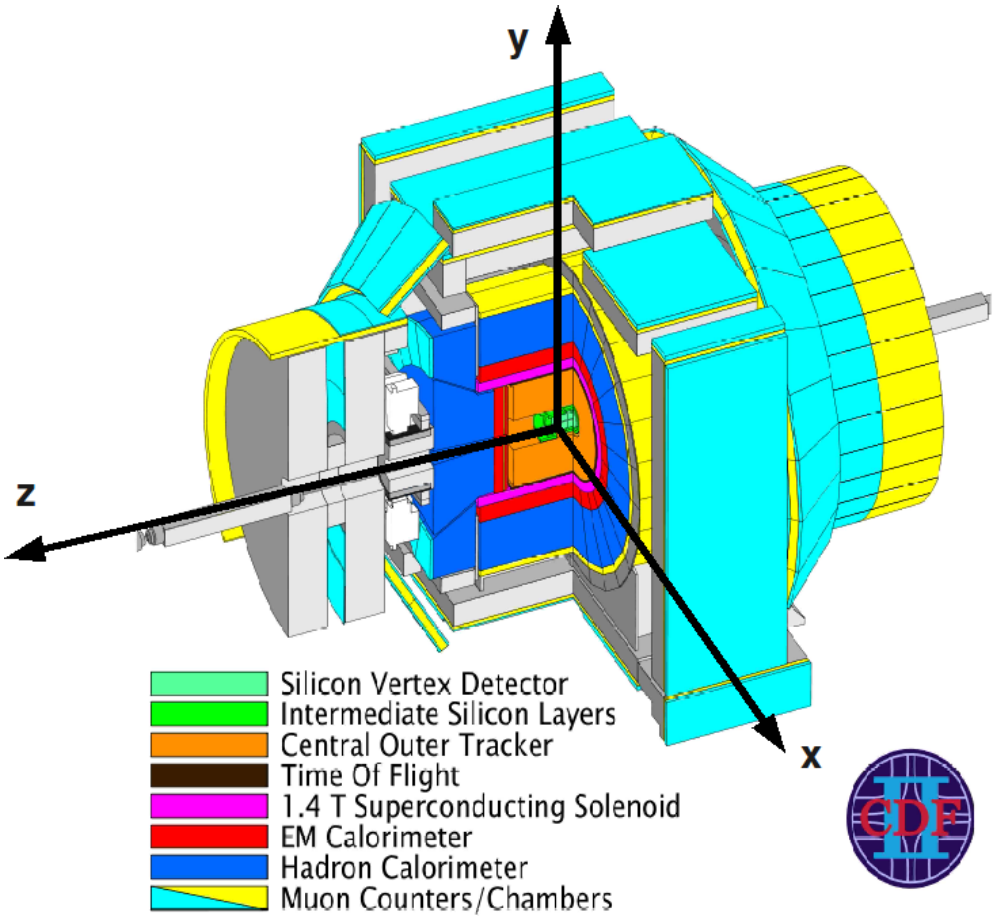


Figure 2.3: *The CDF II detector with a quadrant cut to expose the different subdetectors.*

- luminosity monitors devoted to the instantaneous luminosity measurement, necessary to predict event yields and monitoring the instantaneous luminosity critical to detector operation.

Moreover, every 396 ns a proton and anti-proton bunches collide in the middle of the beampipe which lies at the center of the detector. At present time it is not possible to record all of the data which would have to be read out every 396 ns, and an online hardware triggering system quickly looks at some of the data for each event and decides if it is an interesting enough event to keep or should be discarded. This significantly reduces the amount of data which must be read out from the detector and written to storage media, discarding the uninteresting events.

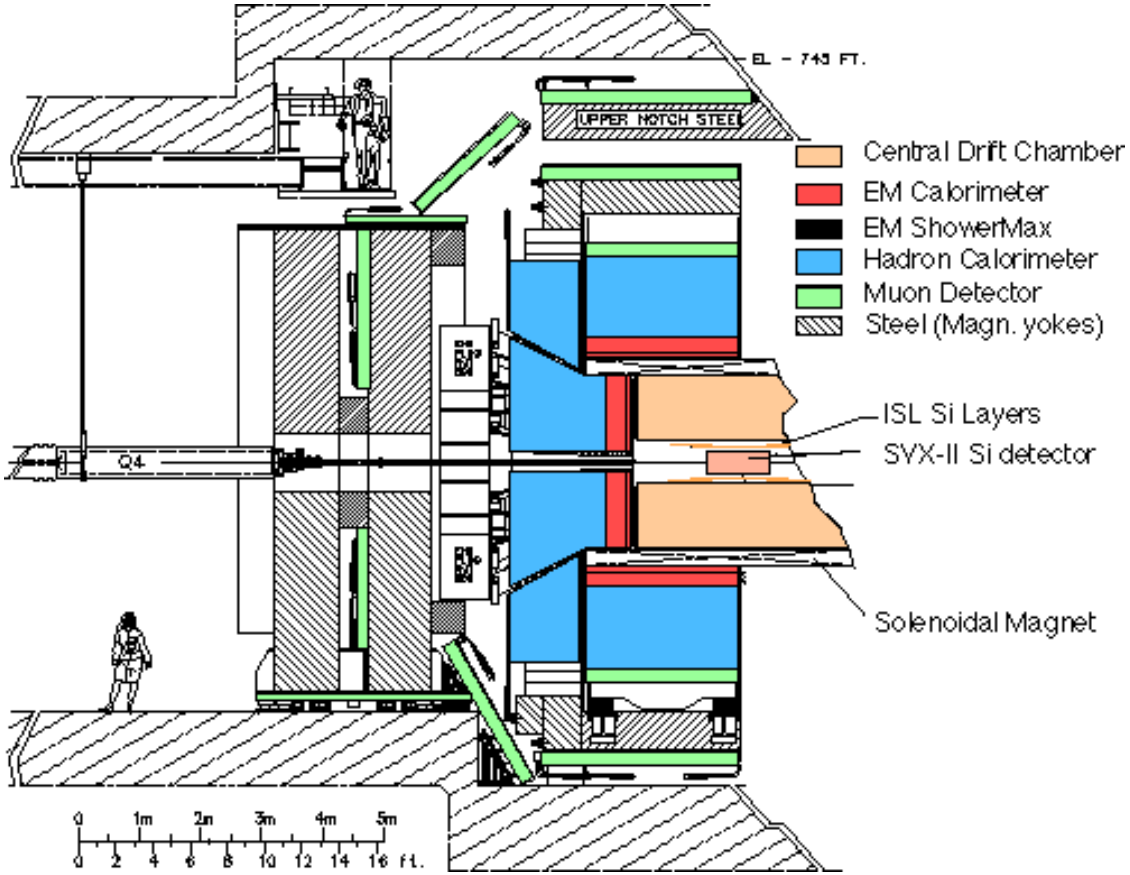


Figure 2.4: *Elevation view of one half of the CDF RunII detector: The TOF and the small angle detectors are not pictured, The Central Muon Extension (CEX) and the forward muon detectors are clearly visible on the left side.*

Fig. 2.4 shows a half side view of the CDF II detector.

The relevant components of the detector to this analysis are discussed along with the CDF trigger system in the remainder of this chapter.

A more detailed description of the CDF II detector can be found in [37] and in specific references cited there for each sub-detector.

2.2.1 Standard Definitions in CDF

Because of its barrel-like shape, the CDF II Detector uses a cylindrical coordinate system (r, φ, z) with the origin at the center of the detector and the z-axis along the nominal direction of the protons beam. r is the radial distance from the origin and φ the azimuthal angle. The $r - \varphi$ plane is called the *transverse plane*, since is perpendicular to the beam line. The polar angle, ϑ, is the angle relative to the z-axis.
An alternative way of expressing ϑ is the pseudorapidity η[2], defined as:

$$\eta \equiv -\ln \tan\left(\frac{\vartheta}{2}\right) \tag{2.3}$$

The coverage of each CDF II detector sub-system will be described using combinations of η, r, φ and z. Charged particles moving through a homogeneous solenoidal magnetic field along the z direction follow helical trajectories. To uniquely parameterize a helix in three dimensions, five parameters are needed: $C, \cot\vartheta, d_0, \varphi_0$ and z_0. The projection of the helix on the $r - \varphi$ plane is a circle. C is the signed curvature of the circle, defined as $C \equiv \frac{sign(Q)}{2\rho}$, where ρ is the radius of the circle and the charge of the particle (Q) determines the sign of C. Positive charged tracks curve counterclockwise in the $r - \varphi$ plane when looking along the z direction and negative charged tracks bend clockwise. The transverse momentum, p_T , depends on the curvature C, on the magnetic field (B_{magnet}), and on the charge of the particle through the following formula: $p_T = Q \cdot \frac{B_{magnet}}{2C}$. Since ϑ is the angle between the z axis and the particle momentum, $\cot\vartheta = \frac{p_z}{p_T}$, where p_z is the z component of the particle momentum.
The last three parameters, d_0, φ_0 and z_0, are the r, φ and z coordinates of the point of closest approach of the helix to the beam line. See Fig. 2.5 for the definition of d_0 and φ_0. d_0 is a signed variable:

[2] It is convenient because it is a purely geometrical quantity related to the polar angle, and, at the same time, it is a Lorentz invariant under a boost in z direction in ultra-relativistic limit. Since a lot of the physics at CDF is approximately ultra-relativistic, this variable is useful to describe the geometry of the decay.

$$d_0 = Q \cdot \left(\sqrt{x_0^2 + y_0^2} - \rho \right), \tag{2.4}$$

where (x_0, y_0) is the center of the helix circle in the $r - \varphi$ plane. Fig. 2.6 shows the definition of d_0 sign.

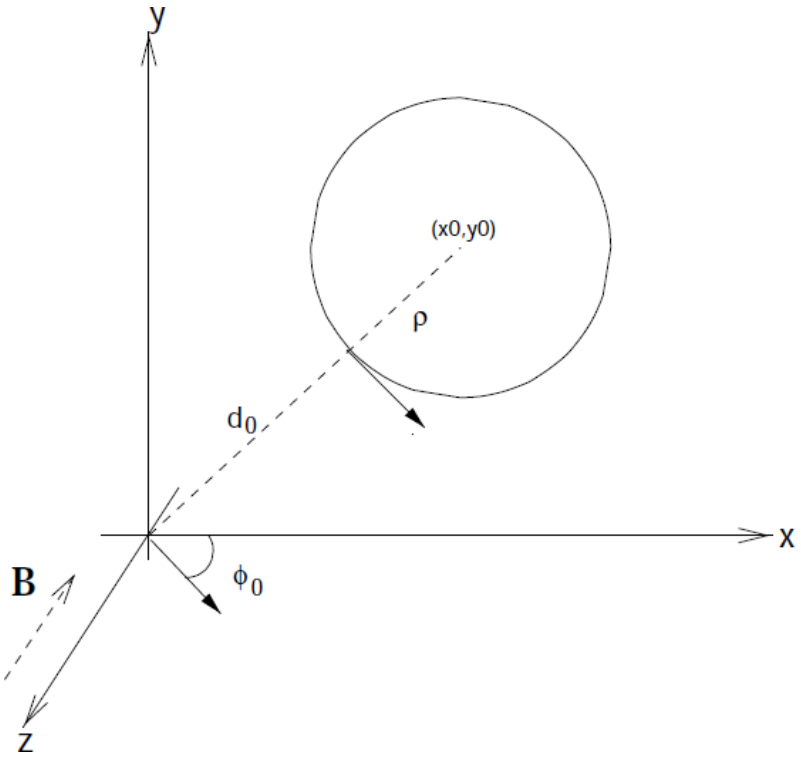


Figure 2.5: *Definition of the* d_0 *and* φ_0 *coordinates.*

2.2.2 The Tracking System

Three-dimensional charged particle tracking is achieved through an integrated system consisting of three silicon inner subdetectors and a large outer drift-chamber, all contained in a superconducting solenoid. The 1.4 T magnetic field and the 136 cm total lever arm provide excellent tracking performances.

We will describe this system, shown in Fig. 2.7, starting from the device closest to the beam and moving outwards.

The innermost tracking device is a silicon detector, which consists of three subdetectors that cover the region $|\eta| < 2$ and 2π of azimuthal angle. The first layer of silicon sensors,

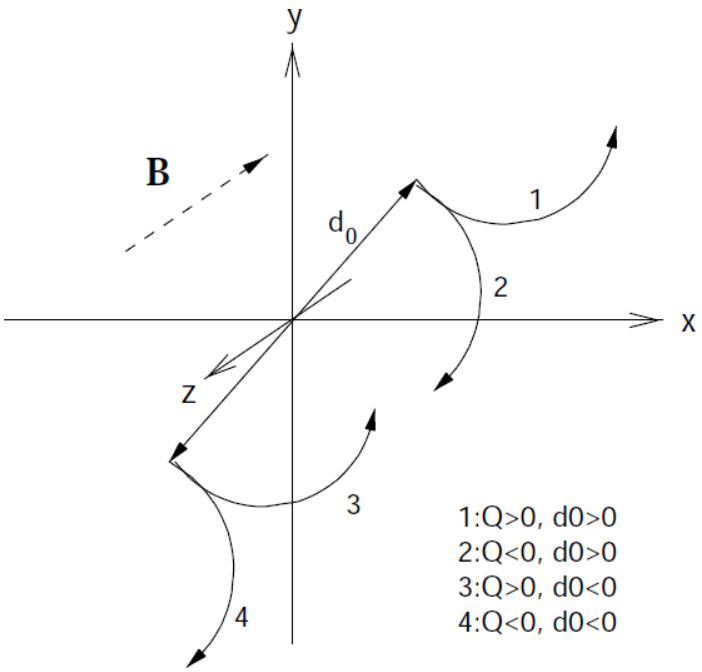


Figure 2.6: *Tracks of particles with positive/negative charge and positive/negative impact parameters.*

called Layer 00 (L00) [38], is installed directly onto the beryllium beam pipe, with the sensors at radii 1.35 and 1.62 cm from the beam. The beam pipe is made of beryllium because this metal has the best mechanical qualities with the lowest nuclear interaction cross section.

The L00 is followed by SVX II [39], made of five concentric layers of silicon sensors located at radii between 2.45 and 10.6 cm. The Intemediate Silicon Layers (ISL) [40] are the outermost silicon detectors, with one layer of sensors at a radius of 22 cm in the central region and two layers at radii 20 and 28 cm in the forward region.

Surrounding the silicon detectors is the Central Outer Tracker (COT) [41], a 3.1 m long cylindrical open-cell drift chamber covering the volume between 43.4 cm and 132.3 cm of radius and $|\eta| < 1$.

In the central region ($|\eta| \lesssim 1$), seven silicon samplings (one in the (r, φ) view and six in the (r, φ, z) view), and 96 chamber samplings (48 (r, φ) plus 48 (r, z)) are available between 1.6 and 132 cm. In the forward and backward regions ($1 \lesssim |\eta| \lesssim 2$), 8 silicon samplings (one in the (r, φ) view and seven in the (r, φ, z) view) are available between 1.6 and 29 cm, along with partial information from the chamber.

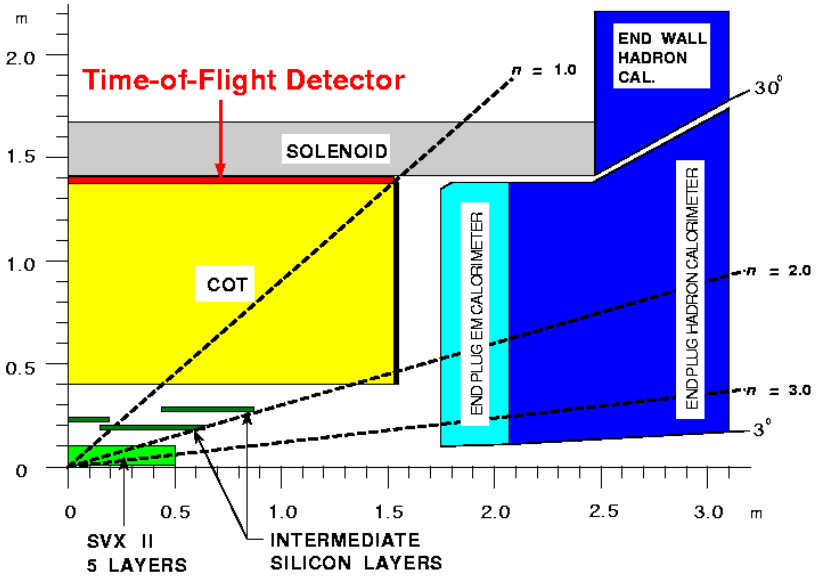


Figure 2.7: *Elevation view of one quadrant of the inner portion of the CDF II detector showing the tracking volume surrounded by the solenoid and the forward calorimeters.*

The high number of samplings over the 88 cm lever arm of the chamber ensure precise determination of the five tracks parameters in the central region. The chamber provides also track seeds for pattern-recognition in silicon.

The Silicon Detectors

The silicon strip detectors [42] at CDF II provide a precise determination of the particle trajectory close to the beam line. The impact parameter resolution measured in the transverse plane is of 27 μm. A silicon detector is fundamentally a reverse-biased p-n junction. When a charged particle passes through the detector material, it causes ionization. For a semiconductor, this means that electron-hole pairs are produced. Electrons drift towards the anode, and holes drift towards the cathode, where the charge is gathered. The amount of charge is, to first order, proportional to the path length traversed in the detector material by the charged particle.

By segmenting the p or n side of the junction into "strips" and reading out the charge deposition separately on every strip, we obtain sensitivity to the position of the charged particle. All the CDF II silicon detectors are implemented as microstrip detectors. The

typical distance between two strips is about 60 μm. Charge deposition from a single particle passing through the silicon sensor can be read out on one or more strips. This charge deposition is used to determine the hit position in the direction perpendicular to the strips.

There are two types of microstrip detectors: single and double-sided. In single-sided detectors only one (p) side of the junction is segmented into strips. Double-sided detectors have both sides of the junction segmented into strips. The benefit of double-sided detectors is that while one (p) side has strips parallel to the z direction, providing $r-\varphi$ position measurements, the (n) side can have strips at an angle (stereo angle) with respect to the z direction, and can provide z position information.

The innermost layer, L00, is made of single-sided silicon sensors, placed on the beam-pipe at radii, alternating in φ, of 1.35 or 1.62 cm from the beam, which provide only $r-\varphi$ measurements, but also, being only at $\approx$1.5 cm from the interaction point, it provides the best resolution on the transverse impact parameter of charged tracks and recovers the degradation in resolution on the primary or secondary vertices position (produced inside L00) due to multiple scattering of low momentum tracks, which is particularly significant on the SVXII read-out electronics and cooling system, installed within the tracking volume.

For SVX II, the core of the silicon tracker, made of double sided silicon sensor, four silicon sensors are assembled into a "ladder" structure which is 29 cm long. The readout electronics are mounted directly to the surface of the silicon sensor at each end of the ladder. The ladders are organized in an approximately cylindrical configuration, creating "barrels". A SVX II barrel is segmented into 12 wedges, each covering approximately 30° in φ, for each wedge there are five layers. Each layer provides one axial measurement on one side and a measurement at the stereo angle on the other side (see Tab.2.1).

The resolution on the single hit is 12 μm. There are three SVX II barrels, mounted adjacent to each other along the z-axis, as shown in Fig.2.8, covering the nominal interaction region at the center of the CDF II Detector. The longitudinal coverage (left) and the cross section view (right) of the silicon detector subsystem are shown in Fig.2.9.

The ISL is made of double-sided silicon sensors and it provides up to two additional tracking layers depending on pseudorapidity (Fig.2.9, left). In particular ISL provides a higher tracking efficiency by connecting tracks in SVX with the ones in COT and allows to extend tracking beyond the COT limit ($|\eta| < 1$), and up to $|\eta| < 2$.
The total amount of material (in units of $g \cdot cm^{-2}$) in the silicon system, averaged over

Property	Layer 0	Layer 1	Layer 2	Layer 3	Layer 4
number of φ strips	256	384	640	768	869
number of z strips	256	576	640	512	869
stereo angle	90°	90°	+1.2°	90°	−1.2°
φ strip pitch [μm]	60	62	60	60	65
z strip pitch [μm]	141	125.5	60	141	65
active width [mm]	15.30	23.75	38.34	46.02	58.18
active length [mm]	72.43	72.43	72.38	72.43	72.43

Table 2.1: *Relevant parameters for the layout of the sensors of the five SVX-II layers.*

φ and z, varies roughly as 0.1 $X_0/\sin(\vartheta)$ in the $|\eta| \leqslant 1$ region[3], and roughly doubles in $1 \leqslant |\eta| \leqslant 2$ because of the presence of cables, cooling bulk-heads, and portions of the support frame. The average amount of energy loss for a charged MIP particle crossing the detector at 90° is ~9 MeV. All the silicon detectors are used in the Offline track reconstruction algorithm.

Central Outer Tracker

The main tracker at CDF is the Central Outer Tracker (COT) [41], a cylindrical multi-wire open-cell drift chamber that provides tracking for charged particle with transverse momentum (p_T) as low as 400 MeV/c in the central pseudo-rapidity region ($|\eta| \lesssim 1$, see sec. 2.2.2). It gives an accurate information in the $r-\varphi$ plane for the measurement of the transverse momentum, and substantially less accurate information in the $r-z$ plane for the measurement of the z component, p_z.
The COT contains 96 sense wire layers, which are radially grouped into eight "superlayers". This can be seen from the end plate section shown in Fig.2.10.

Each superlayer is divided into cells, and each cell contains 12 sense wires. The maximum drift distance is approximately the same for all superlayers. Therefore, the number of cells in a given superlayer scales approximately with the radius of the superlayer. The entire COT contains 30240 sense wires spanning the entire length of the detector in z. Approximately half the wires run along z direction ("axial"). The other half are strung at

[3]The symbol X_0 indicates the radiation length in units of $g \cdot cm^{-2}$. The amount of material offered by the tracking system has to be minimal to reduce the photon convertions into electron-positron pairs and the multiple scattering affecting the reconstruction of charged tracks[43].

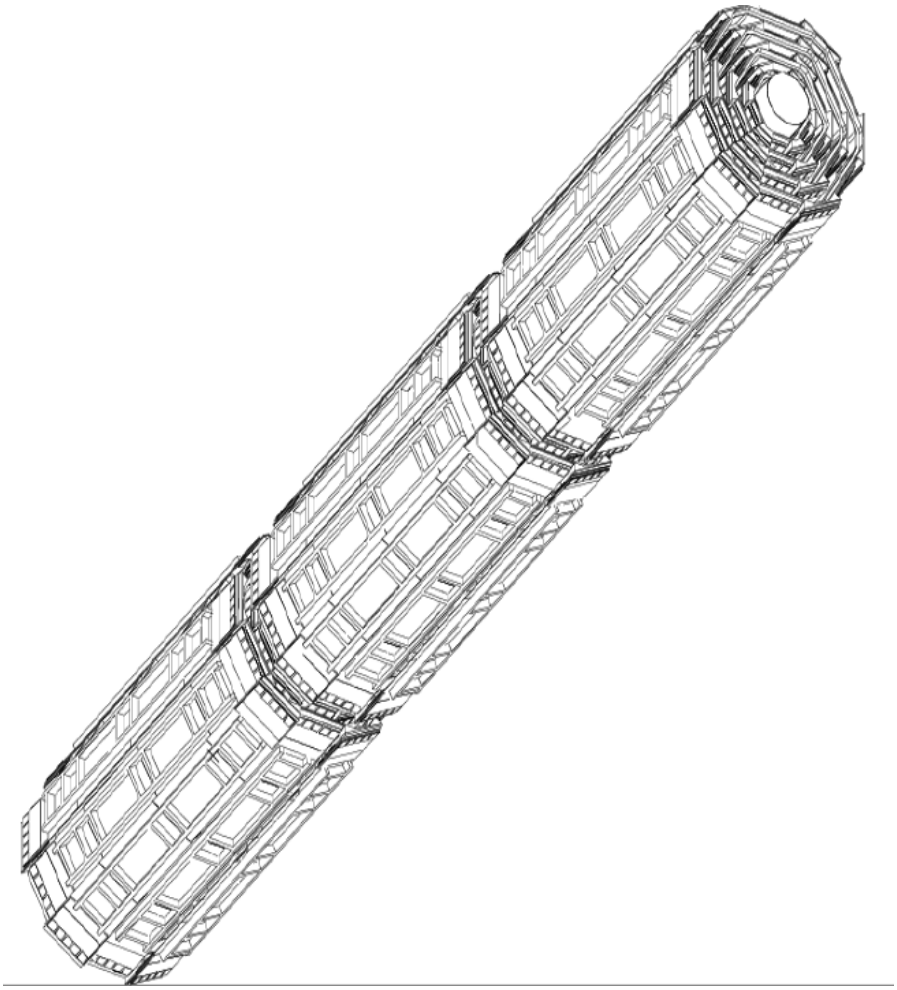


Figure 2.8: *Isometric view of three SVX II barrels.*

a small angle (2°) with respect to the z direction ("stereo"). This allows to perform track reconstruction in the $r-z$ plane. The active volume of the COT begins at a radius of 43.4 cm from the beamline and extends out to a radius of 132.3 cm. The chamber is 310 cm long. Particles originating from the interaction point with $|\eta| < 1$ pass through all the 8 superlayers of the COT. The cell layout, shown in Fig.2.11 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 of the cell. Both the sense and potential wires are 40 μm diameter gold plated tungsten wires. The field sheet is 6.35 μm thick mylar with vapor-deposited gold on both sides. Each field sheet is shared with the neighboring cell.

The COT is filled with an Argon-Ethane gas mixture (50:50). The gas mixture is

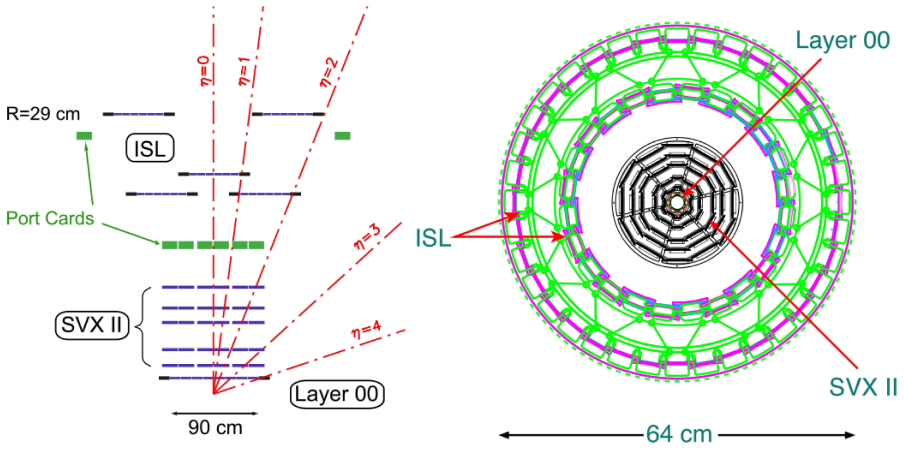


Figure 2.9: *Longitudinal coverage of the silicon trackers (left) and cross section view of the integrated* SVXII-*ISL tracking system (right).*

chosen to have a constant drift velocity across the cell width.

When a charged particle passes through the detector volume, the gas is ionized. Electrons drift towards the nearest sense wire. The electric field in a cylindrical system grows exponentially with decreasing radius. As a consequence, an avalanche multiplication of charge happens inside the high electric field region, in the vicinity of the wire, due to electron-atom collisions. The resulting charge reaches the wire and this so-called "hit" is read out by electronics. The avalanche discharge provides a gain of $\sim 10^4$. The maximum electron drift is approximately 100 ns, well smaller than the inter-bunch spacing 396 ns, providing the read-out and processing of the COT data available for the Level 1 trigger. Due to the magnetic field electrons drift at a Lorentz angle of $\sim 35^\circ$ with respect to the radius. The cell is tilted by $\sim 35^\circ$ with respect to the radial direction to compensate for this effect.

The analog pulses from the 30,240 sense wires flow to preamplifiers where are amplified and shaped. The discriminated differential output encodes charge information in its width to be used for dE/dx measures and leading edge to the arrival time information. Both are fed to a TDC which records them in 1 ns bins. After calibrating the width, variations due to the COT geometry, to the path length of the associated track, to the gas gain differences for the 96 wires, the Landau associated to the track is determined, using the amount of the charge collected (in nanosecond) for each hit along the track path length. From the Landau the energy loss is measured and used for particle identification. The

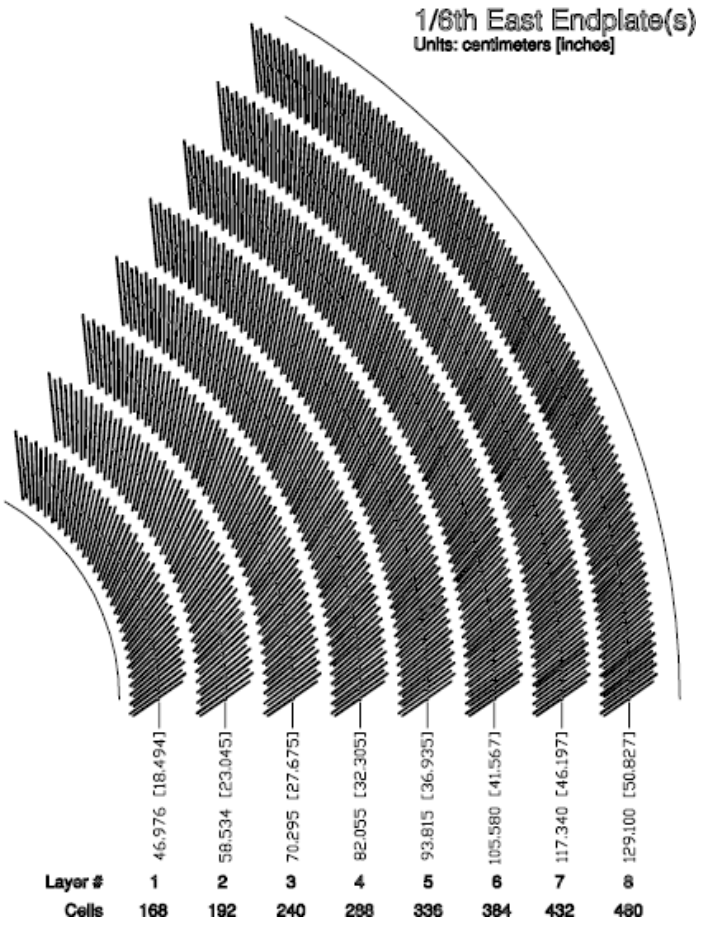


Figure 2.10: *Layout of the wire planes on a COT endplate.*

TDC boards contain also the buffer where the data are stored while waiting for the events to be accepted by the trigger. The TDC auxiliary card catch hits for the **eXtremely Fast Tracker (XFT)** track trigger processor (Section3.2). Hit times are later processed by pattern recognition (tracking) software to form helical tracks. The hit resolution of the COT is about 140 μm. The transverse momentum resolution has been measured using cosmic ray events to be:

$$\sigma_{p_T}/p_T^2 = 0.0017[\mathrm{GeV/c}]^{-1} \tag{2.5}$$

Tracking algorithms reconstruct particle trajectories (helixes) that best correspond to the observed hits. Reconstructed trajectories are referred to as "tracks".

The COT tracking system is a crucial element in this thesis, since allows the identification of the leptons (electrons or muons) in the central region. Leptons (muons or

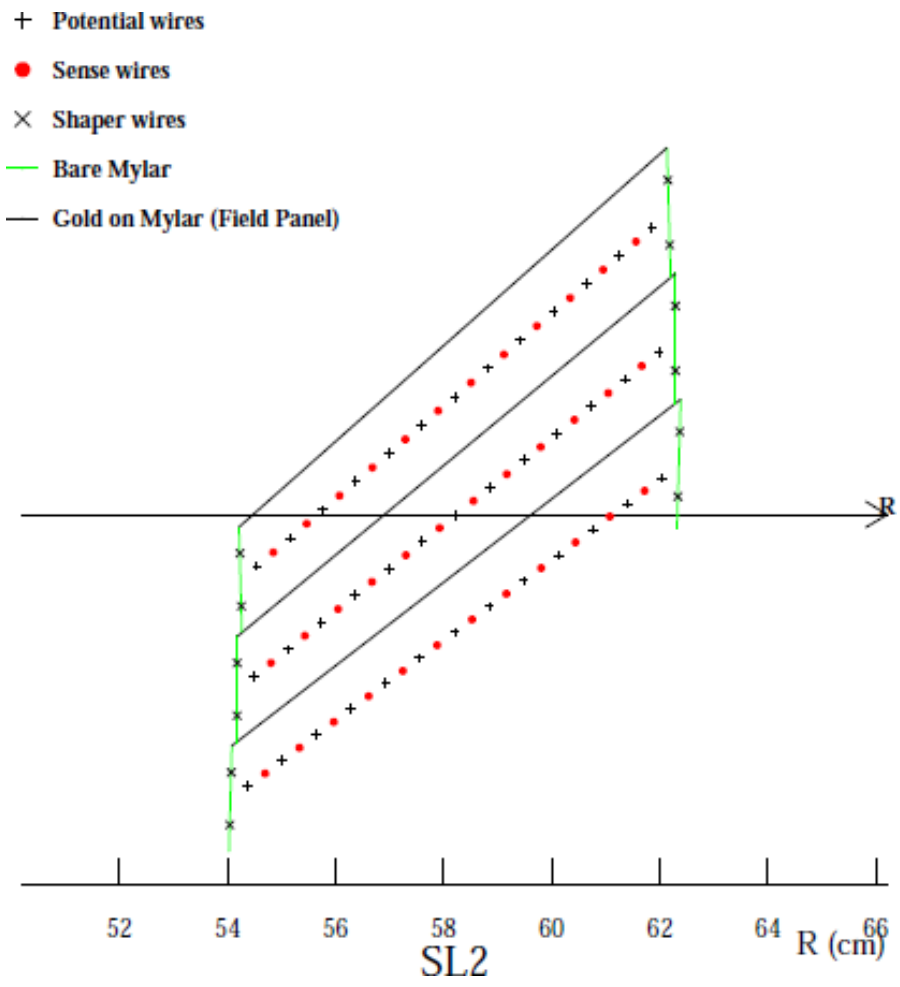


Figure 2.11: *Layout of the wires in a COT cell.*

electrons) candidates are selected by Level 1 trigger by matching the tracks in the tracking system to energy clusters in the electromagnetic calorimeter or to the segment of track reconstructed in the muon chambers (Section3.3.2).

2.2.3 The Calorimeters System

Located immediately outside the solenoid, the calorimeter system at CDF covers a solid angle of nearly 4π around $p\bar{p}$ interaction point up to $|\eta_{det}| \lesssim 3.6$[4]. It measures the energy flow from hadrons, electrons, or photons, using "shower" sampling [44] based on layers of high-Z passive absorber interspaced with layers of plastic scintillator.

[4]However, between detector sections there are regions ("cracks") where the response is poor.

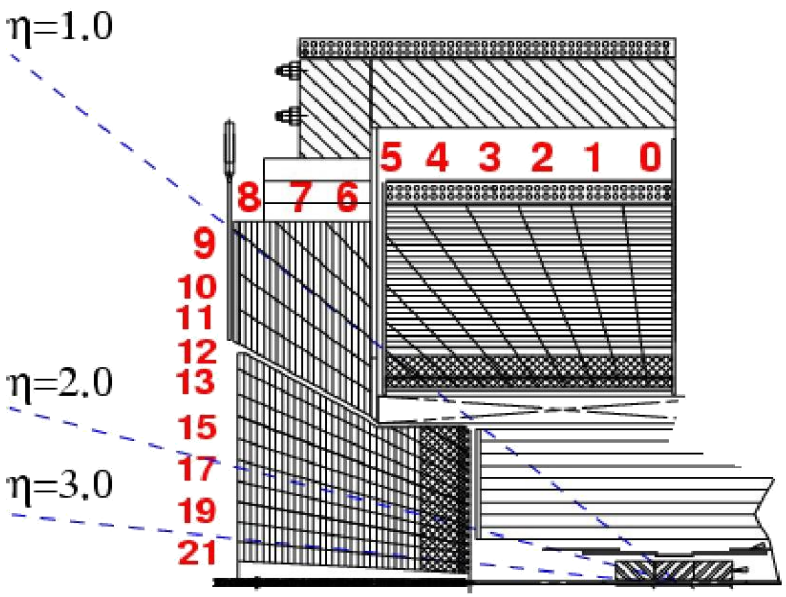


Figure 2.12: *A schematic drawing of the calorimeter. The central region extends up to* $\eta = 1$.

Electrons and photons traversing a calorimeter will interact with the heavy metal which creates a shower of electrons and photons. Similarly, hadrons will shower and deposit their energy in the calorimeters as well (though they will typically travel through more material). This shower excites atoms in the scintillator which then emit photons as they return to their ground state. These photons are then amplified by photomultiplier tubes and the amplitude gives a measurement of the energy deposition.
The CDF calorimeter is divided into two physical sections: the central and the forward end-plug region. The calorimeter in the central region ($\eta < 1.1$) is referred to as the "Central Calorimeter" and the calorimeter in the forward end plug region ($1.1 < |\eta| < 3.6$) as the "Plug Calorimeter". All the calorimeters are segmented in pseudorapidity and azimuth, with a projective tower geometry pointing to the interaction point. In-depth segmentation of each tower consists of two independent compartments: the *electromagnetic (EM)* and the *hadronic (HA)* components optimized to better react respectively to electromagnetic and hadronic interacting particles. Each tower component is read out independently, providing also spatial information of the particle detection, in a geometry shown in Fig.2.12 .

The EM towers samples the energy deposit by EM showers or MIP The HA towers

samples the energy deposit by hadronic shower or MIP. Different fractions of energy release in the two compartments distinguish photons and electrons from hadrons. A segmentation uniform in pseudorapidity instead of the polar angle is chosen, as in high energy collisions the number of charged particles per unit of rapidity is approximately constant in the central region.

Central Electromagnetic Calorimeter (CEM, CES, CPR)

In the $|\eta| \leqslant 1.1$ region, the central electromagnetic calorimeter (CEM) appears as an hollow cylinder occupying the radial region between 173 and 208 cm [45]. It consists of 31 layers of polystyrene scintillator interleaved with layers of lead clad in aluminum and it is split into 12 azimuthal 15°-wide sectors (see Fig.2.13)

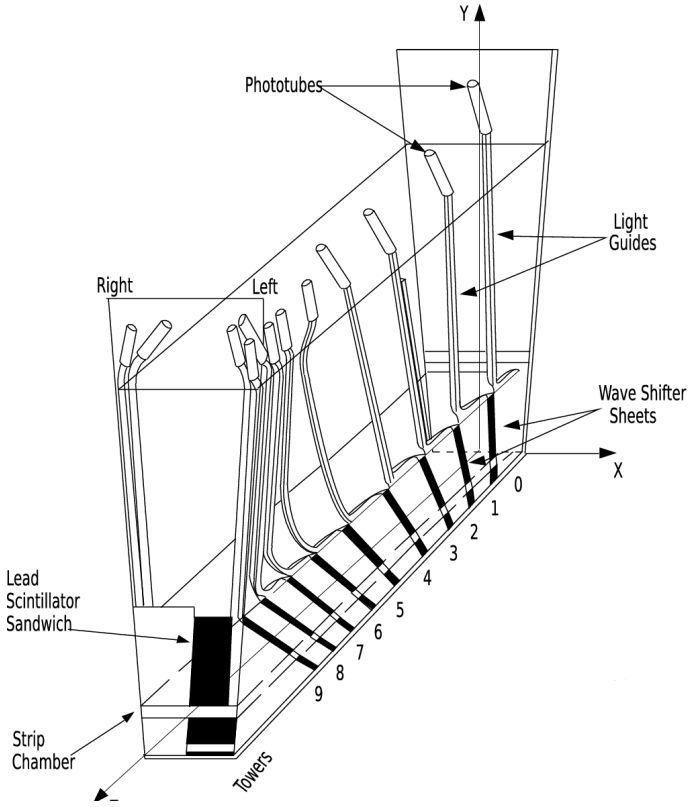


Figure 2.13: *A wedge of the CEM and its components.*

Each sector is divided into ten η towers ($\Delta\eta \times \Delta\varphi \approx 0.11 \times 15°$ per tower). To maintain a constant thickness in X_0, compensating the $\sin(\vartheta)$ variation from tower to tower, some lead layers are replaced with increasing amounts of acrylic as a function of

η.[5] The blue light from each tower is collected, wave-length shifted into green light by sheets of acrylic plastic placed on the azimuthal tower boundaries, and guided to two phototubes per tower. The two outer towers in one wedge are missing to allow accessing the solenoid for check and repairs if needed. The total number of instrumented towers is 478.

At a radial depth of $5.9\ X_0$, where the peak of shower development is typically located, an array of multi-wire proportional chambers measures the transverse shower shape with 2.0 mm resolution (for 50 GeV electrons). In Run I a layer of multi-wire proportional chambers was located in a 5 cm gap between the outer surface of the solenoid and the first layer of the calorimeter to monitor photon conversions started in the tracker material or in the solenoid("preshower detector"). During the fall 2004 shut-down, this system was replaced with a finely segmented layer of scintillator tiles [46].

The total thickness of the electromagnetic section corresponds to approximately $19\ X_0$ ($\sim 1\ \lambda_{\rm int}$, where $\lambda_{\rm int}$ is the pion nuclear absorption length in units of g cm^{-2}), the energy resolution is[6]:

$$\frac{\sigma_E}{E} = \frac{13.5\%}{\sqrt{E_T}} \oplus 2\% \tag{2.6}$$

Central and Wall Hadronic Calorimeters (CHA, WHA)

The central hadronic calorimeter (CHA) surrounds the CEM covering the region $|\eta| < 0.9$ and consists of steel layers sampled each 2.5 cm by scintillator. Filling a space between the CHA and the forward plug hadronic calorimeter (PHA) two calorimeter rings cover the gap between CHA and PHA in the region $0.7 < |\eta_{\rm det}| < 1.3$, the wall hadronic calorimeter (WHA), which continues the tower structure of the CHA but with reduced sampling each 5.0 cm. Like the electromagnetic calorimeters, the hadronic calorimeters are read out using waveshifting lightguides and phototubes. Each central wedge is segmented into nine η towers matching in size and position the electromagnetic towers, for 384 towers in total. The end-wall section has six additional η towers, three of which matching the outer central

[5] The number of lead layers varies from 30 in the innermost ($|\eta| \approx 0.06$) tower to 20 in the outermost ($|\eta| \approx 1.0$).

[6] The first term is called the "stochastic" term and derives from the intrinsic fluctuations of the shower sampling process and of the PMT photo-electron yield. The second term, added in quadrature, depends on the calorimeter non-uniformities and on the uncertainty of the calibrations. Energies are in GeV.

hadronic towers (see fig. 2.4) for a total number of 288 towers. A central hadronic tower is constructed of 32 layers of steel absorber, 2.5 cm thick, alternating radially with 1.0 cm-thick acrylic scintillator. It is composed by two parts joining at z=0. The technology of the WHA is similar to CHA one, but contain only 15 layers of 5.1 cm-thick absorber.

The total thickness of the hadronic section is approximately constant in the $|\eta| < 1.3$ region and corresponds to approximately 4.5 λ_{int}. The total number of projective towers of CHA+WHA is 12, out of which 6 are entirely contained in the CHA, 3 are entirely contained in the WHA and 3 are shared between the two. Each tower element is read by 2 photomultipliers. The resolutions on energy of CHA and WHA found in test beam measurements (response to single pions) are ([47]):

$$\mathrm{CHA}: \frac{\sigma_E}{E} \approx \frac{50\%}{\sqrt{E_T}} \oplus 3\% \ , \ \mathit{WHA}: \frac{\sigma_E}{E} \approx \frac{75\%}{\sqrt{E_T}} \oplus 4\%. \tag{2.7}$$

where $E_T = E \sin\vartheta$ is the transverse energy.

Plug Electromagnetic Calorimeter (PEM, PES, PPR)

The electromagnetic coverage is extended in the region $1.10 \lesssim |\eta| \lesssim 3.64$ by a separate scintillating tile calorimeter (see 2.14) [48].

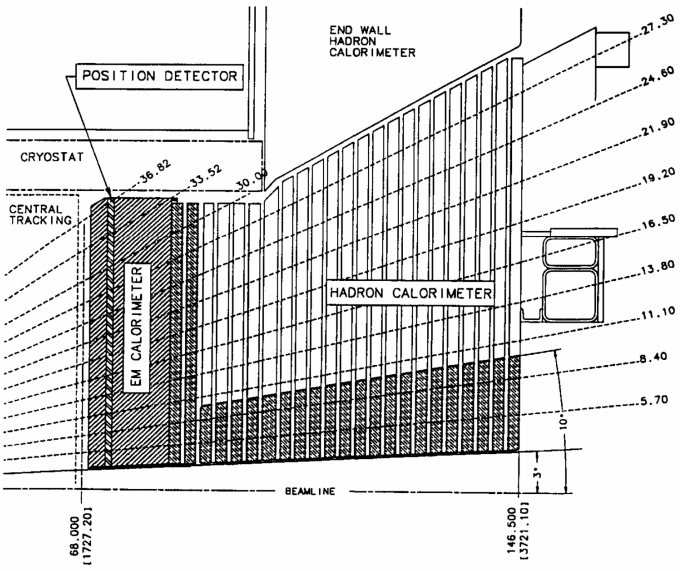


Figure 2.14: *Elevation view of one quarter of the plug calorimeter.*

The PEM calorimeter has a similar structure as the CEM: 22 layers of 4.5 mm thick

lead alternate with 22 layers of 4 mm thick scintillator. The PEM tower segmentation is 7.5° in φ for $|\eta| \lesssim 2.11$ and 15° for $2.11 < |\eta| < 3.6$. The segmentation in η can be understood by an inspection of Fig.2.14. Each scintillator tile is coupled to a different PMT, except for the first layer which is a 1 cm thick plane of scintillator bars read by a multi-anode PMT and acting as a preshower detector (PPR). The total thickness of the PEM section corresponds to approximately 21 X_0 (1 $\lambda_{\rm int}$), for an energy resolution of:

$$\frac{\sigma_E}{E} = \frac{16\%}{\sqrt{E_T}} \oplus 1\% \tag{2.8}$$

Also the PEM is equipped with a shower maximum detector (PES), made of three planes of scintillator strips rotated by 60° and providing a spatial resolution of about 1 mm on the shower maximum location.

Plug Hadronic Calorimeter (PHA)

The Plug Hadron calorimeter (PHA) is located behind the PEM [48] and has the same tower segmentation (2.14). The technology is the same as of CHA, with 23 layers alternating 2 cm thick steel absorber and 6 mm thick scintillator, for a total amount of material corresponding to 7 $\lambda_{\rm int}$. Its resolution in single pion test beam was found to be ([48]):

$$\text{PHA}: \frac{\sigma_E}{E} \approx \frac{80\%}{\sqrt{E_T}} \oplus 5\% \tag{2.9}$$

2.2.4 Muon Detectors

Outside of all other sub detectors is the CDF muon system (Fig. 2.15).

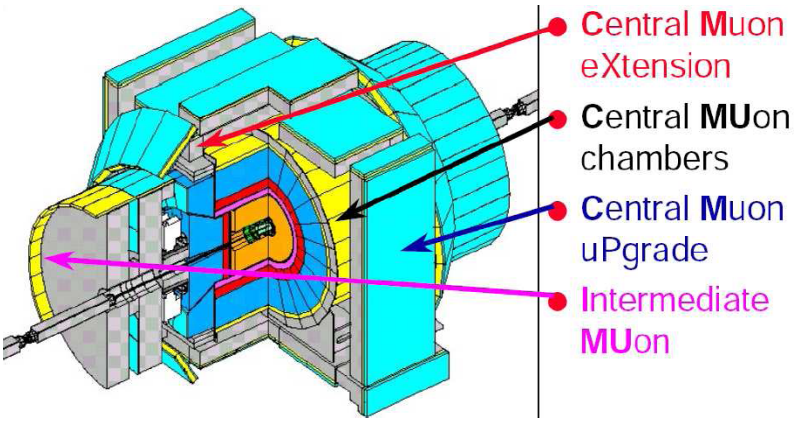


Figure 2.15: *Muons detector.*

High p_T muons at CDF are identified by taking advantage of the fact that they are minimum ionizing particles that loose only modest quantities of energy when passing through large amounts of matter. Muon detection is made by positioning drift chambers behind the calorimeters and in some cases behind additional shielding. Short track segments ("*stubs*") are reconstructed from the hits in these detectors and then matched to tracks in the tracking chamber. The majority of particles produced in $p\bar{p}$ collisions reaching these detectors are muons with a relatively small contamination from hadronic particles.

CMU and CMP

The muon detectors are grouped into different sets according to the region they cover (Fig. 2.16).

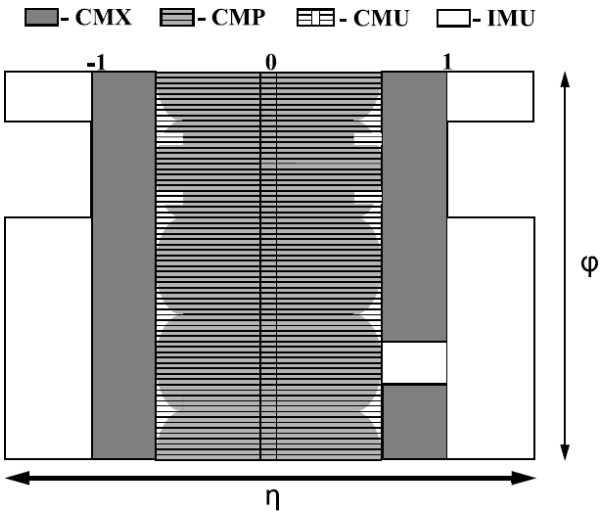


Figure 2.16: *Coverage of muon detectors in the* $\eta \times \varphi$ *space.*

The Central Muon detectors (CMU) are a set of 144 drift chamber modules of 16 cells each. These cells are 266 cm $\times$ 2.68 cm $\times$ 6.35 cm wide with a single 50 μm steel wire at their center (Fig. 2.17). Scintillation counters are sandwiched to the chamber to help finding the φ and z position of the hit and matching the muon to a COT track. The maximum drift time for CMU is 800 ns. Each scintillator covers two single drift chamber (see Fig.2.18).

In the $|\eta| < 0.65$ region an additional set of muon chambers, the Central Muon Upgrade detectors (CMP), surrounds the CMU detectors in the $|\eta| < 0.65$ region behind a 60 cm

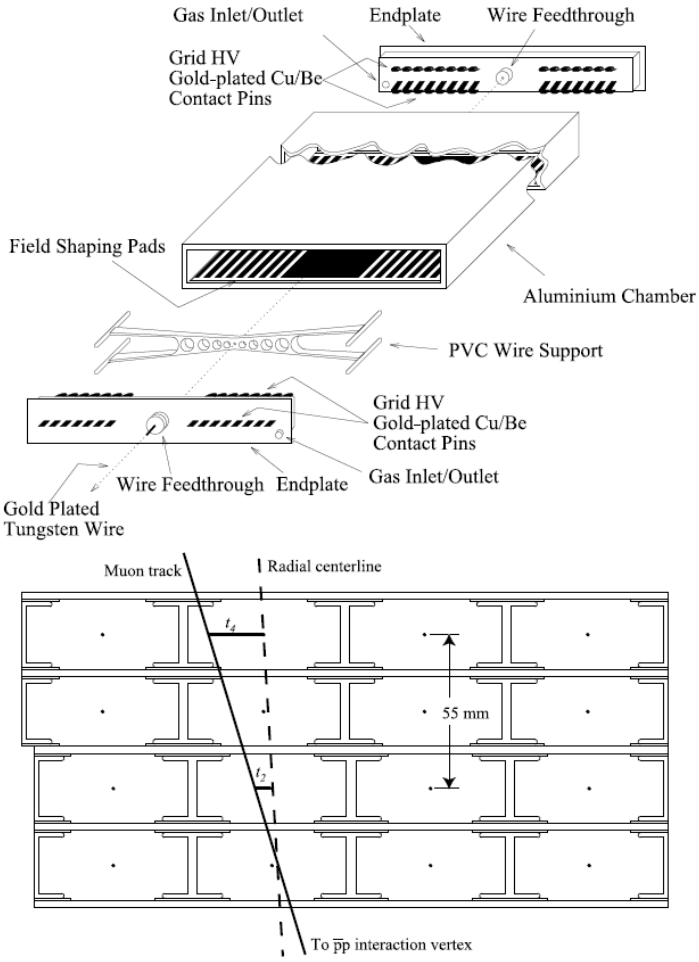


Figure 2.17: *Exploded view of CMP/CMX/IMU tubes (top) and cross section of CMU layers (bottom).*

thick steel absorber. As the CMU, the CMP are rectangular chambers stacked in four layers (Fig. 2.17), typically 640 cm $\times$ 15 cm $\times$ 2.5 cm in size. The first and fourth layer have different cell width in order to remove left-right ambiguities in position finding and help in triggering on muons. CMP chambers are sandwiched to scintillator layers (CSP) on the outermost side, for the identification of the bunch crossing. Scintillators are colored in light blue in Fig. 2.3, while drift chambers are yellow. The maximum drift time in CMP is 1.4 μs.

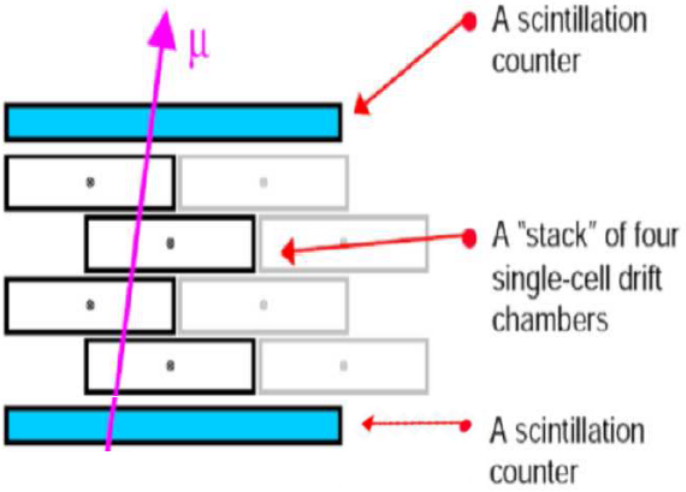


Figure 2.18: *Muon chambers system*

CMX

A set of muon detectors arranged in a truncated conical shape around the plug calorimeters, the Central Muon Extension (CMX), provides muon identification in the $0.65 < |\eta| < 1$ region. Their structure is analogous to the CMP, with a different cell length (180 cm) and scintillators on both sides (CSX).

IMU

The CDF detector is shielded in the forward regions by two pairs of steel toroids (Fig. 2.3), around which the Intermediate Muon detectors (IMU) are laid. These are scintillation counter/drift chamber sandwiches similar to CSP/CMP and CSX/CMX, arranged in azimuthal rings[7] to cover the $1 < |\eta| < 1.5$ region. The IMU tubes (Barrel Muon Chambers - BMU) are 8.4 cm wide and 363 cm long and are coupled to scintillators (BSU). There is a scintillator layer between the two toroids of each pair (TSU), laying on the transverse plane and covering the $1.3 < |\eta| < 2$ region.

[7] The barrels are not complete: they surround the toroids for 3/4 of their circumference, in the lower part they meet the floor, as in Fig. 2.16.

2.2.5 Cherenkov Luminosity Counters

The luminosity ($\mathcal{L}$) is inferred from the average number of inelastic interactions per bunch crossing ($\bar{N}$) according to $\bar{N} \times f_{b.c.} = \sigma_{p\bar{p}-in-} \times \varepsilon \times \mathcal{L}$, where the bunch-crossing frequency ($f_{b.c.}$) is precisely known from the Tevatron RF, $\sigma_{p\bar{p}-in} = 59.3 \pm 2.3 mb$ is the inelastic $p\bar{p}$ cross-section resulting from the averaged CDF and E811 luminosity-independent measurements at $\sqrt{s} = 1.8$ TeV, and extrapolated to $\sqrt{s} = 1.96$ TeV, and ε is the efficiency to detect an inelastic scattering.

The Cherenkov Luminosity Counters (CLC) are two separate modules, covering the $3.7 < |\eta| < 4.7$ range symmetrically in the forward and backward regions. Each module consists of 48 thin, 110-180 cm long, conical, isobutane-filled Cherenkov counters. They are arranged around the beampipe in three concentric layers and point to the nominal interaction region. The base of each cone, 6-8 cm in diameter and located at the furthest extremity from the interaction region, contains a 15 conical mirror that collects the light into a PMT, partially shielded from the solenoidal magnetic field. Isobutane guarantees high refraction index and good transparency for ultraviolet photons. With a Cherenkov angle $\vartheta_C = 3.4°$, the momentum thresholds for light emission are 9.3 MeV/c for electrons and 2.6 GeV/c for charged pions. Prompt charged particles from the $p\bar{p}$ interaction are likely to traverse the full counter length, thus generating large signals and allowing discrimination from the smaller signals of particles emitter at the same angle due to the beam halo or to secondary interactions. In addition, the signal amplitude distribution shows distinct peaks for different particle multiplicities entering the counters. This allows a measurement of $\bar{N}$ with 4.4% relative uncertainty in the luminosity range $10^{31} \leqslant \mathcal{L} \leqslant 10^{32}$ $cm^{-2}s^{-1}$. This accuracy, combined with the 4% relative uncertainty on the inelastic $p\bar{p}$ cross-section, results in an instantaneous luminosity measured with 5.9% relative uncertainty.

2.3 Trigger and Data Acquisition System

The 396 ns $p\bar{p}$ bunch crossing rate at the Tevatron implies a 2.53 MHz collision rate inside the detector[8].

Since the read-out of the entire detector needs about 2 ms, after the acquisition of one event, another approximately 5,000 interactions would occur and remain unrecorded.

[8] At an istantaneous luminosity of 1×10^{32} $cm^{-2}s^{-1}$, which is the typical order of magnitude for Tevatron luminosity, and for a total inelastic cross section of ≈60 mb, roughly 2 inelastic collisions occur per bunch crossing on average

Clearly this is unacceptable. The detector front-end electronics must be designed as to solve this problem and reduce the event loss to a few percents. The percentage of events which are rejected solely because the trigger is busy processing previous events is referred to as trigger deadtime. On the other hand, the average size of the information associated to each event from the $\mathcal{O}(10^6)$ total CDF II channels is 140 kbytes. Even in case of deadtime-less read-out of the detector, in order to record all events an approximate throughput and storage rate of 350 Gbyte/s would be needed, largely beyond the possibilities of currently available technology. However, since the cross-sections of most interesting processes are several order of magnitude (from 10^3 to 10^{12}) times smaller than the inelastic $\mathrm{p\bar{p}}$ cross-section, the above problems may be overcome with an on-line preselection of the most interesting events[9].This is the task of the trigger system, which evaluates on-line the information provided by the detector and discards the uninteresting events. The CDF II trigger is a three-level system that selectively reduces the acquisition rate, with virtually no deadtime, i. e., keeping each event in the trigger memory for a time sufficient to allow for a trigger decision without inhibiting acquisition of the following events (see fig. 2.19). Each level receives the accepted event from the previous one and, making use of detector information of increasing complexity and with more time for processing, applies a logical "OR" of several programmable selection criteria to make its decision.

Level 1 (L1)

A trigger divided in three stages does not remove the problem to deal with the Tevatron crossing rate of more than 2.53 MHz, the problem is now rerouted to the L1 stage. In order to avoid deadtime caused by the trigger processing time, the L1 has to sustain the clock of the Tevatron. In a complex detector as CDF II, it is unconceivable an effective trigger architecture able to process data and make a decision in less than 396 ns. The impasse is overcome with a fully pipelined front-end electronics for the whole detector. The signal of each channel is stored, every 396 ns, in a buffer of a 42-cell long pipeline. This means that the L1 has 396×42 ns $\simeq$ 16 μs to make its decision before the content of the buffer is deleted. The actual latency of the L1 is 5.5 μs, as was designed for a crossing time of 132 ns.

At L1 a synchronous system of custom-designed hardware process a simplified subset of

[9] As examples,the $\mathrm{b\bar{b}}$ production cross section is $\sim 10^3$ times smaller than the generic $\mathrm{p\bar{p}}$ inelastic one, when the Diboson is 10^9 time smaller.

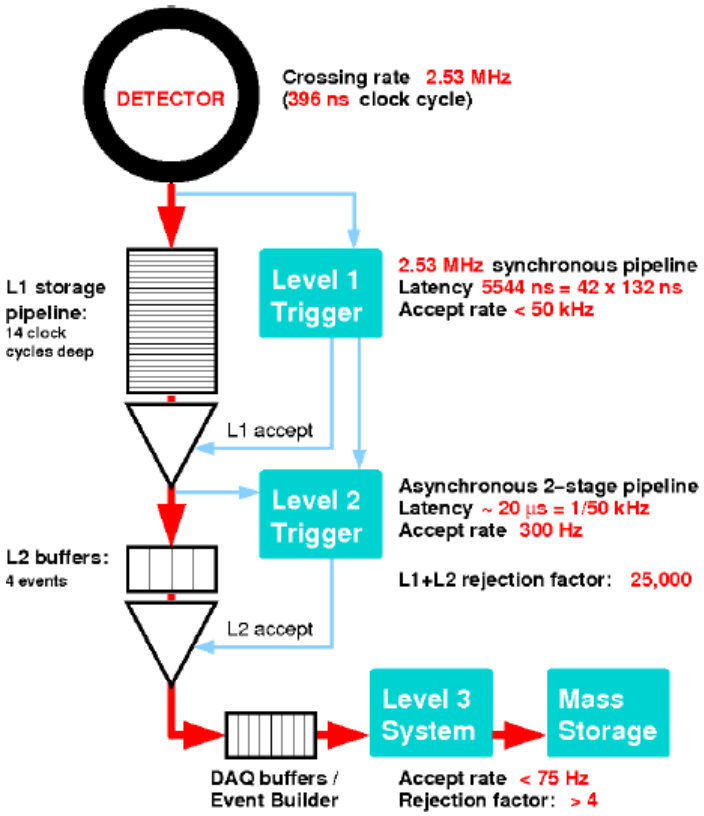


Figure 2.19: *Diagram of the CDF II trigger architecture*

data in three parallel streams to reconstruct coarse information from the calorimeters (total energy and presence of single towers over threshold), the COT (two-dimensional tracks in the transverse plane), and the muon system (muon stubs in the CMU, CMX, and CMP chambers). A decision stage combines the information from these low-resolution physics objects, called "primitives", into more sophisticated objects, e. g., track primitives are matched with muon stubs, or tower primitives, to form muon, electron, or jet objects, which are subjected to basic selections.

Currently are implemented about 56 different L1 combinations of requirements with an output rate of 18 KHz.

Level 2 (L2)

The L2 performs two subsequent operations. The Event building produces in output the event as reconstructed with L2 detector information, and the Decision combines outputs

from L1 and L2 to evaluate whether to flag or not the event for Level-3 processing.
The Event building process is done in parallel. Calorimetric information is used to perform clustering and identification of hadronic jets. Simultaneously, the Silicon Vertex Trigger (SVT), a dedicated processor, combines the COT track informations with SVXII hits. It measure the track parameters in the transverse plane $(x - y)$ with almost offline level quality for tracks with $p_T > 2$ GeV/c. The key improvement of the SVT track is the measurement of the track impact parameter.

The event building has 10 μs to complete its task. In the Decision stage some selection algorithms, customized for each different combinations of requirement, run on four dedicated CPUs and process the available information from L1 and L2 in less than 10 μs. The maximum latency of L2 is 20 μs for each event. The current number of different combinations of requirements at L2 are about 116 and the output rate is about 300 Hz.

Level 3 (L3)

This stage is implemented exclusively by software. About 400 commercial processors running in parallel reconstruct the event provided by L2 at full detector resolution. L3 codes are very similar to the offline reconstruction codes. About 140 trigger paths are implemented at L3. Moreover L3 distributes the information to on-line monitoring consumers and data logger programs. The L3 decision to write on tape happens after the full reconstruction of the event is completed and the integrity of its data is checked in less than 10 ms. Typical size for an event is 150 kbyte. Maximum storage rate is about 20 Mbyte/s. At L3 the number of different combinations of requirements is about 185. The available output rate is 75 Hz (40% tracking, 30% jet and photon, and 30% lepton).

Chapter 3

Event Selection and Reconstruction

The disobon production in the semileptonic decay channel, contains signatures of leptons, neutrinos and quarks. Inside the detector, the charged leptons are reconstructed using the tracking system, the calorimeter and muon chambers; the quarks, that hadronize, generate jets of particles that leave their signatures in the calorimeter (Fig.3.1).

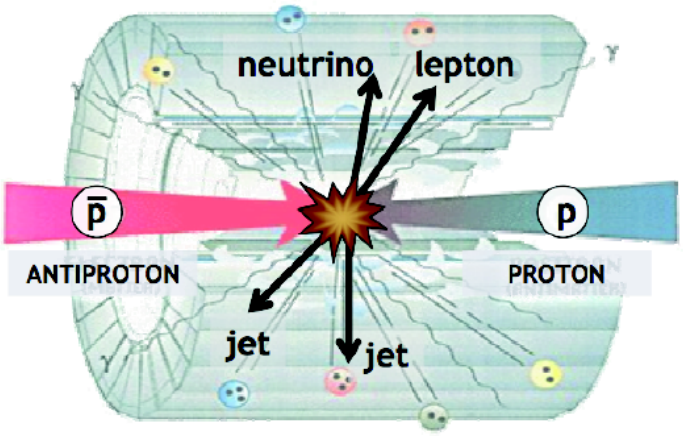


Figure 3.1: *The semi-leptonic signature of the WW/WZ production: a lepton, a neutrino (identified as missing transverse energy) and jets, generated by the quark hadronization.*

The neutrino does not leave any signature in the detector. Because the total transverse momentum should be null (in the detector frame), it is reconstructed and identified, by the energy imbalance in the transverse detector plane as measured by calorimeter (missing E_T, $\not\!\!E_T$).

In this chapter we will describe the reconstruction of the objects we'll use in the analysis, including the requirements applied during data taking by the trigger system and the final candidate event selection criteria. At the moment we reconstruct the W boson decaying leptonically only in the muon channel.

A schematic description of the objects reconstruction used in this study is shown on Fig.3.2

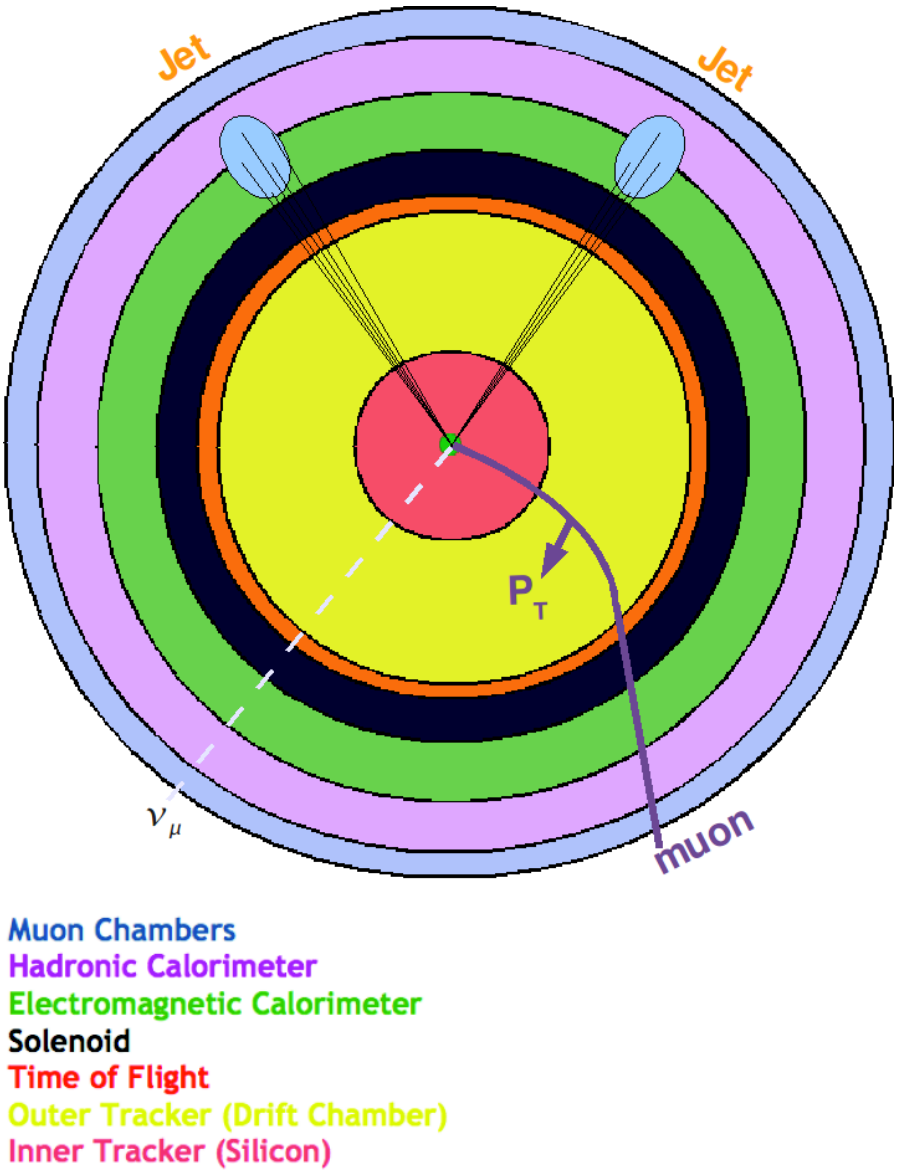


Figure 3.2: *A schematic description of the reconstruction, in the transverse plane of the CDF detector, of the objects used in the analysis: muons, neutrinos and jets.*

3.1 General

3.1.1 Data Sample

We reconstructed the objects we'll use in the analysis in data events collected by the CMUP18 and CMX18 trigger paths described in Section 3.2.1, requiring a central muon with $p_T > 18$ GeV/c. Good runs, i.e. runs with no detector malfunction during operation, are required. After selecting good runs, the data sample corresponds to an integrated luminosity of 3.9 fb^{-1}.

Events have been reconstructed using Gen6 version 6.1.4 of the offline software.

3.1.2 The signal definition

The decay channel of this analysis contains three objects: the **charged muon**, the **neutrino** (missing transverse energy, $\not\!\!E_T$), and the **jets**.

The theoretical cross sections for this decay mode are ([9], [10]):

$$\sigma_{WW} \times \mathrm{Br}(W \to \mu\nu_\mu, W \to jj) = (12.4 \pm 0.8)\mathrm{pb} \times 0.073 = 0.905 \pm 0.058 \ \mathrm{pb}$$

$$\sigma_{WZ} \times \mathrm{Br}(W \to \mu\nu_\mu, Z \to jj) = (4.0 \pm 0.3)\mathrm{pb} \times 0.076 = 0.304 \pm 0.023 \ \mathrm{pb}$$

The total cross section in the Standard Model of the signature we are looking for is therefore

$$\sigma_{WW/WZ} \times \mathrm{Br}(W \to \mu\nu_\mu, W/Z \to jj) = (1.21 \pm 0.081) \ \mathrm{pb}$$

3.1.3 The background definition

There are several processes that result in the same final state topology (1 charged muon, $\not\!\!E_T$ and jets) as the diboson production, and thus are backgrounds to this search. The background processes that are taken into account are:

- $p\bar{p} \to W(\to \mu\nu_\mu) + \mathrm{jets}$;
  In this case, generic QCD jets fake a hadronically decaying W.
- $p\bar{p} \to W(\to \tau\nu_\tau) + X$;
  A tau is detected as an electron or a muon.
- $p\bar{p} \to Z(\to \mu\mu) + X$;
  One of the two leptons is not reconstructed, resulting in larger missing transverse energy (e.g. because it fell in un-instrumented areas of the detector).

- QCD processes;
  A QCD jet fakes a lepton.

- $p\bar{p} \rightarrow t\bar{t} + X, t \rightarrow Wb$;
  The W (either leptonic or hadronic) comes from $t\bar{t}$ production.

The largest of these backgrounds is the $(W \rightarrow \mu\nu) + jets$. To have an order of magnitude of the above processes cross section see Table.4.1 and Table.4.2.

The objects reconstruction

In order to study the signal and the background of the process of interest using the CDF detector data, the information from the detector needs to be processed and interpreted into physical objects (**muons, neutrinos, jets**). The first step of the processing is performed online, via the CDF data acquisition and trigger system. Once the interesting data, containing muon, jet candidates and $\not\!\!E_T$ information, are stored on disks, offline selection algorithms are used to reconstruct the final objects that will be used in the analysis.

3.2 Trigger Requirements

At the trigger level, one attempts to select signal events with high efficiency while keeping the trigger accept rate low, rejecting many background events. It is possible to meet these criteria in this analysis by exploiting only one of the characteristics of the signature: the presence of centrally produced muons with high transverse momentum (p_T).

As mentioned, a muon is a minimum ionizing particle that leaves a track in the COT, which deposite very little energy in the electromagnetic and hadronic calorimeters. The COT tracking, calorimeter, and muon detectors information is available at Level 1. Evidence for the presence of muons is obtained by looking for aligned hits in the CMU and CMX that are consistent in arrival time. Hits in the CMP are reconstructed based on the hits found in the CMU. Jets (see Section3.3.4) and electrons are reconstructed as a total hadronic and electromagnetic calorimeter energy deposited in a trigger tower (defined as two physical towers adjacent in η) above a given threshold. The sum of the energy deposited in the calorimeters and the missing transverse energy (see Section3.3.6) are also computed. Limited charged particle tracking is performed by the eXtremely Fast Tracker (XFT), which identifies tracks with p_T of the order of 1.5 GeV/c in the $r-\varphi$ view using

the information provided by four axial superlayers of the COT. Track segments are formed in each superlayer that are later linked together to form track candidates. A charged track passing through an axial superlayer will generate a maximum of 12 hits, the presence of 10 or 11 hits per track segment (depending on the data taking period) and the presence of track segments in each four axial superlayers are required. A Track Extrapolation Unit (XTRP) extrapolates the track candidates to the calorimeters and muon detectors to form electron and muon candidates. This information is collected together with the calorimeter trigger level information to make the Level 1 decision.

The L2 is made of two main pieces of hardware: the Secondary Vertex Trigger (SVT) and the cluster finder hardware. The SVT uses silicon hit information, in conjunction with the XFT tracks, to trigger on the presence of tracks from displaced vertices. The cluster finder forms cluster of energies from neighboring calorimeter towers as defined by the Level 1 trigger, to form more sophisticated electron and jet candidates. The data from these systems as well as the data used for the Level 1 decision are sent to the Level 2 to make the decision, if accepted, are sent to the L3.

At L3 the full detector information is available to reconstruct particles. In the following we describe the trigger requirements used in this analysis.

3.2.1 Central Muon Trigger

Two trigger paths are used for the high p_T central muon selection. One finds the muons in the CMU and CMP ($|\eta| < 0.6$) and is called MUON_CMUP_18 and the other in the CMX ($0.6 < |\eta| < 1.0$) and is called MUON_CMX_18.

The trigger MUON_CMUP_18 requests are:

- L1
  - Hits in one or more layers of the CMU are found.
  - 3 or 4 hits in the CMP are found and required to be consistent with hits in the CMU.
  - An XFT track ($p_T > 4$ GeV/c) with at least 11 hits on each axial superlayer matching in the $r - \varphi$ plane the hits found in the CMU/CMP.
- L2

- an XFT track with $p_T > 8$ GeV/c not necessarily matching the muon hits.

- L3

  - A fully reconstructed COT track with $p_T > 18$ GeV/c matching a track segment (stub) in the CMU and in the CMP. The match require that the distance in the $r - \varphi$ plane, between the COT track extrapolated to the muon hit segment, and the position of the reconstructed muon hit, satisfy $|\Delta x_{CMU}| \leqslant 10$ cm for CMU hits, and $|\Delta x_{CMP}| \leqslant 20$ cm for CMP hits.

The trigger MUON_CMX_18 requests are:

- L1

  - CMX hits matching the central muon extension scintillator hits.
  - An XFT track ($p_T > 8$ GeV/c) with at least 11 hits on each axial superlayer matching in the $r - \varphi$ plane the hits found in the CMX.

- L2

  - No requirements.

- L3

  - A fully reconstructed COT track with $p_T > 18$ GeV/c matching a track segment (stub) in the CMX ($|\Delta x_{CMU}| \leqslant 10$ cm).

3.3 Offline Reconstruction

The offline reconstruction concerns the events that have been selected by any of the triggers mentioned above: MUON_CMUP_18, and/or MUON_CMX_18. In these events, the objects of the analysis will be reconstructed. A first step in the reconstruction of the objects we are interested in for this analysis is the reconstruction of a track that is essential for the lepton identification.

3.3.1 Track Reconstruction

The ability to detect and reconstruct charged particle trajectories is essential for lepton identification and momentum reconstruction. Multiple-track reconstruction allows to identify the vertices where either the $p\bar{p}$ interaction took place (primary vertex) or the decay

of a long-lived particle took place (secondary or displaced vertex). Also, a precise measurement of the track impact parameter, allows the identification of heavy-flavor quarks in jets. As explained in the previous sections, charged particles leave small charge depositions as they pass through the tracking system. By following, or "tracking", these depositions, pattern recognition algorithms can reconstruct the charged particle original trajectory (helix) that best correspond to the observed hits. Reconstructed trajectories are referred to as "tracks".

There are several algorithms used to reconstruct tracks in the CDF II tracking system. The central muons are reconstructed by the "Outside-In" algorithms which use the information from the COT chambers and SVX. The name of this group of algorithms suggests that the track is followed from the outside of the tracking system inwards.

Outside-In Algorithm

The Outside-In is the most reliable of CDF tracking algorithms as it is based on COT coverage that extends up to $\eta = 1$, the central region fully covered by 96 sampling planes of wires arranged in 8 COT superlayers. The helical track, when projected onto the $r - \varphi$ plane, is a circle. This simplifies pattern recognition, so the first step of pattern recognition in the COT looks for circular paths in the axial superlayers of the COT. Cells in the axial superlayers are searched for sets of 4 or more hits that can be fit to a straight line. These sets are called "segments". Once segments are found, there are two approaches to track finding ("segment linking" and "histogram linking" algorithms [49]). One approach is to link togheter the segment which are consistent with lying tangent to a common circle. The other approach is to constrain its circular fit to the beamline. Once a circular path is found in the $r - \varphi$ plane, segments and hits in the stereo superlayer are added depending on their proximity to the circular fit. This results in a three-dimensional track fit. Typically, if one algorithm fails to reconstruct a track, the other algorithm will not. This results in a high track reconstruction efficiency (~95%) in the COT for tracks which pass through all 8 superlayers ($p_T \geqslant 400$ MeV/c). The track reconstruction efficiency mostly depends on how many tracks are reconstructed in the event. If there are many tracks close to each other, hits from one track can shadow hits from the other track, resulting in efficiency losses. Once a track is reconstructed in the COT, it is extrapolated inward to the silicon system. Based on the estimated errors on the track parameters, a three dimensional "road" is formed around the extrapolated track. Starting from the outermost layer, and working

inwards, silicon hits found inside the road are added to the track. As a hits gets added, the road gets narrowed according to the knowledge of the updated track parameters. Reducing the width of the road reduces the chance of adding a wrong hit to the track, and also reduces the computation time. In the first pass of this algorithm, axial hits are added. In the second pass, hits with stereo information are added to the track.

3.3.2 Muon Reconstruction

The only lepton we use in this study is the muon.
High energy muons cross the calorimetric system as minimum ionizing particles. They are identified by a track inside COT, a deposit of a MIP inside EM and HAD calorimeters and an aligned track segment (stub) in the outer muon chambers. Muons can be faked by cosmic rays, by π, forward leptonically decays in the tracker and by hadrons not showering inside the calorimeters ("punch-through"). A number of software filters are used to reject the cosmics, and computed corrections are applied to the raw muon flux to account for hadronic decay and punch-through.

Muon Track Requirements

The central muons candidates have good track segments on three axial and three stereo superlayers. A good track segment is defined as containing at least 5 hits out of the possible 12 in the superlayer. In order to have good COT tracking efficiency we require the z_0[1] of the track to be less than 60 cm away from the center of the CDF detector. To remove background events from cosmic rays and pion and kaon in flight decays, the tracks are required to originate from the center of the detector in the $x-y$, requiring a low impact parameter (d_0). The cosmic background is mostly reduced by requiring the track hit timing information to be consistent with the bunch crossing for that event. Finally, the radius at which the track exits the COT is required to be greater than 140 cm to avoid regions where the muon reconstruction is not well-understood.

Muon CMU CMP CMX and Calorimeter Requirements

Muons are identified by either a charged track matched to a stub in the central (CMU and CMP) or extended (CMX) muon chambers. The stubs are formed by hits in the CMU,

[1] z_0 is z coordinate of the closest point of the helix to the beam line (see Section2.2.1)

CMP and CMX and the extrapolated tracks are required to match the position of the stub in $r-\varphi$ within 7, 5 and 6 cm in the CMU, CMP and CMX respectively.

The energy deposited in the calorimeter by the candidate track is required to be consistent with that of minimum ionizing particles to remove fake hadronic particles.

As part of the muon stub selection, we apply cuts that ensure that the tracks reconstructed with the COT, point at regions of the muon chambers with high hit reconstruction efficiency. The projected position of the track in the $r-\varphi$ and z axes are compared to the muon chamber edges position. A negative value of the fiducial distance indicated respectively as $x-fidx$ and $z-fidz$ means that the track is projected to the inside of a chamber [50]. We want reconstruct the $W \rightarrow \mu\nu_\mu$ candidate decays. In these decays

Event variable	Cuts
COT Ax hits / Ax Seg	$\geqslant 5/\geqslant 3$
COT St hits / St Seg	$\geqslant 5/\geqslant 3$
Track $\lvert z_0\rvert$	< 60cm
Track $\lvert d_0\rvert$	0.2 cm (if no silicon hits attached by OI) 0.02 cm (if silicon hits attached by OI)
COT exit radius	$>$ 140 cm
$\lvert\Delta x_{CMU}\rvert$	$<$ 7.0 cm
$\lvert\Delta x_{CMP}\rvert$	$<$ 5.0 cm
$\lvert\Delta x_{CMX}\rvert$	$<$ 6.0 cm (CMX)
E_{EM}	$< 2 + Max(0, 0.0115(P_T - 100))$ GeV
E_{HAD}	$< 6 + Max(0, 0.028(P_T - 100))$GeV
CMU Fid	$x - fid_x < 0$ cm, $z - fid_z < 0$ cm
CMP Fid	$x - fid_x < 0$ cm, $z - fid_z <$ -3 cm
CMX Fid	$x - fid_x < 0$ cm, $z - fid_z <$ -3 cm
Iso	$\leqslant 0.1$
Track P_T	> 20 GeV/c

Table 3.1: *Central Muon Selection [52]. The Fid labels refer to the cuts on the fiducial distances of the tracks from the muon chambers.*

the muon is isolated. The isolation variable [51] *Iso* is defined as the ratio the calorimeter energy deposited in a cone of radius $\Delta R = 0.4$ around the lepton direction to the

muon energy (~muon momentum) and is required to satisfy Iso $\leqslant 1\%$. We use the muon momentum in the ratio. The isolation requirement is meant to reject muons from semileptonic decay of heavy flavor hadrons and muons faked by hadrons: indeed in both cases, the muon candidate is produced in a jet environment and tends to be less isolated than a lepton produced by a W boson decay.

3.3.3 Jet Identification

According to QCD, partons composing the (anti)proton can be treated perturbatively as free particles if they are stuck by an external probe[2] with sufficiently large momentum transfer ("hard scattering"). Scattered partons convert into color-less hadrons when entering the physical world. This process is called parton "hadronization" or parton "showering" and produces a collimated cluster of stable particles named "jet".

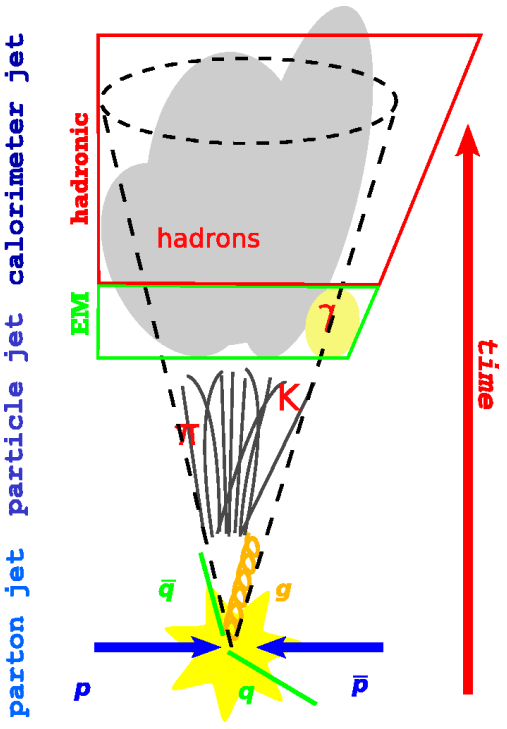


Figure 3.3: *A parton originating from a hard scattering hadronizes and generates a collimated spray of particles, a jet.*

A jet approximately retains the total momentum and direction of the initial parton (for

[2] I.e. a lepton or a parton belonging to another hadron.

a pictorial representation see Fig. 3.3). Because of the intrigued multistep relationship relating jets to primary partons one must be careful in their treatment and definition since any inappropriate handling would cause a systematic error or worsen the resolution in the parton momentum measurement.

From an experimenter's point of view a jet is defined as a large energy deposit in a localized area of the detector (see Fig. 3.4). The challenge of a physics analysis is to recover from detector information the initial energy, momentum and, possibly, the nature of the parton produced in the original interaction.

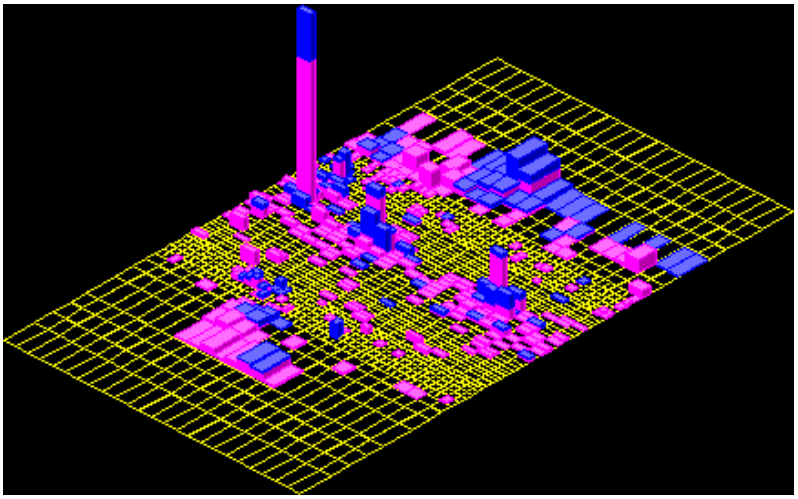


Figure 3.4: *Calorimetric deposit in the* $\eta - \varphi$ *plane as represented in the CDF event display of a typical event. EM deposits are red while HAD deposits are blue. The height of the signals are proportional to the deposit energy.*

The jet's information carries can be degraded on three levels, for non-accounted physical reasons and for instrumental features of the detectors:

1. the quark or the gluon from which the jet is generated corresponds to the "parton level": whether the jet is generated by a quark or by a gluon is not known and their parton emission in the slowing-down process is different;

2. the particles produced in the evolution of the jet correspond to the "particle (or hadron) level": the details of particle production the fragmentation process are unknown;

3. the signal pulses of the calorimeters employed to identify the jet correspond to the

"calorimeter level": the calorimeter response to hadrons is only approximately known and the calorimeter signals are subject to measurement errors.

A universally valid way of defining a jet does not exist because there is no control on how the hadronization process takes place: the experimentalist's task is to concieve an algorithm that allows the extrapolation of the parton properties from the calorimeter jet ones.

The reconstruction algorithm, also called *clustering algorithm*, tries to label a set of energy depositions in the calorimeters as "jets". It can be implemented in different ways, according to physical considerations and practical convenience. However, any reconstruction algorithm must satisfy some requirements which are independent on the actual implementation. These ideal attributes of a clustering algorithm are [53]:

- correct treatment of jet properties by a conceptual point of view ("theoretical attributes")
  1. it should not be "infrared sensitive", which means that jets originating from different partons should not be merged because of soft radiation in the region between them;
  2. it should not have any "collinear sensitivity", which means that jet identification and energy should not depend on how jet prongs are distributed on the calorimeter cells;
  3. the reconstructed transverse energy should not be sensitive to longitudinal Lorentz boosts;
  4. the algorithm should give the same results if applied at any of the three levels of jet evolution (parton, hadron and calorimiter level);
  5. it should be implemented in a computational environment with reasonable easiness;
- application flexibility to experimental conditions ("experimental attributes")
  1. it should be able to reconstruct jets in detectors of different design;
  2. it should not degrade the instrumental resolution;
  3. its performances should be stable against multiple interactions in the same bunch crossing;

4. it should be optimized in terms of computing time;
5. it should identify all interesting jets;
6. it should allow an easy calibration of the jet properties;
7. it should be applicable in a wide range of jet multiplicity, space distribution and energy;
8. it should be easily applicable.

3.3.4 Jet Reconstruction by CDF Cone Algorithm

A jet, which we assume to be massless, results in multiple tower clusters in both the Electromagnetic and Hadronic calorimeter and multiple tracks. In order to associate the particles (hadrons) with jets, we use a selection process that we call the jet algorithm. Several algorithms exist to define a jet object all of which are based on the same goal of bounding the jet within a portion of the detector to provide a measurement of its energy while minimising the amount of energy from the rest of the event that is included. Cone algorithms group together all calorimeter towers whose center (or, at particle level, all particles whose trajectories) fall within a cone of given radius in the $\eta - \varphi$ space. The official jet algorithm at CDF is a cone algorithm called `JetClu` [54]:

- each calorimeter tower is identified by a vector in the $r \times \eta \times \varphi$ space, that joins the origin of the coordinate frame with the center of the calorimetric tower[3], whose E is the magnitude of the vector;
- all towers with $E_T = E^{EM} \cdot \sin\vartheta^{EM} + E^{H} \cdot \sin\vartheta^{H} > 1$ GeV are marked as precluster seeds and ordered in decreasing E_T;
- the seeds contained in a 9-towers square centered on the highest E_T seed are grouped into a precluster, the centroid of the precluster is evaluated;
- a cone of fixed radius[4] is drawn centered on the precluster centroid and all towers with $E_T > 0.1$ GeV falling inside the cone are added to the precluster and its centroid is recalculated;

[3] The center of a tower is identified by the geometrical η and φ center of the tower and at a $6 \cdot X_0$ for the EM calorimeters or $1.5 \cdot \lambda$ for the hadron ones: ϑ^{EM} (ϑ^{H}) is the polar angle of the vector pointing to the geometrical center of the EM (hadron) compartment of the calorimeter tower.

[4] At CDF the standard cone radii are 0.4, 0.7, 1.0: the choice depends on the jet multiplicity of the events (usually the smallest R is chosen for events with many jets) and on the features of the analysis. In this analysis we use a cone with 0.4 radius.

- this procedure is iterated until the list of the towers included in the cluster does not change any more;
- these stable candidate jets can overlap and must be merged or split according to the amount of the shared E_T:
  - if the shared E_T is greater than 75% of the less energetic jet E_T, the two jets are merged and the centroid of the jet is recalculated;
  - if not, the shared towers are assigned to the closest jet in the $\eta - \varphi$ space.
- The final cluster is a jet, defined by the tower list.
- The transverse energy of the jet is given by:

$$E_{T\ jet}^{uncorr} \sqrt{\left(\sum_i E_i \sin(\vartheta_i)\cos(\varphi_i)\right)^2 + \left(\sum_i E_i \sin(\vartheta_i)\sin(\varphi_i)\right)^2} \qquad (3.1)$$

  were E_i is the total energy in tower i of the tower list. This quantity represents the energy deposited in the cluster, and does not include any corrections.

3.3.5 Jet Energy Corrections

In CDF there are 7 levels of jet corrections that can be applied to the jet energies. Depending on the use of the jet objects by the different analyses, different levels of corrections are applied. In our analysis we use the first five levels of corrections. A dedicated group in CDF, the jet energy group, determined the energy correction to scale the measured energy of the jet energy back to the energy of the final state particle level jet . Additionally, there are corrections to associate the measured jet energy to the parent parton energy, so that direct comparison to the theory can be made [54].

The CDF jet energy corrections are divided into different levels to accommodate 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. Depending on the physics analyses, a subset of these corrections can be applied.

Below, the jet corrections are summarized:

- "Level 0" or *Online/Offine Calibrations*; This correction sets the calorimeter energy scale. The effect is small (~3%).

- "Level 1" or *Eta-dependent*; It is applied to raw jet energies measured in the calorimeter to make jet energy uniform along η. It is an effect that clearly depends on the detector geometry and varies with η. Variations are as large as 20%. The uncertainty in the correction is η and p_T dependent and varies from 0.1% to 3%.

- "Level 4"[5] or *Multiple Interactions*; The energy from different $p\bar{p}$ interactions during the same bunch crossing falls inside the jet cluster, increasing the energy of the measured jet. This correction substracts this contribution in average. The correction is derived from minimum bias data and it is parameterized as a function of the number of vertices in the event. The correction is small (less than 1%).

- "Level 5" or *Absolute Corrections*; Corrects the jet energy measured in the calorimeter for any non-linearity and energy loss in the un-instrumented regions of each calorimeter. The effect is p_T dependent. The correction factor is about 1.35 for p_T of 20 GeV and reduces to 1.1 for high-p_T jets. The uncertainty is also p_T dependent.

- "Level 6' or *Underlying Event Corrections*; The underlying event (UE) is defined as the energy associated with the spectator partons in a hard collision event. Depending on the details of the particular analysis, this energy needs to be subtracted from the particle-level jet energy. A factor 1.6 is used to take this into account. The uncertainty is 30% of the underlying event correction.

- "Level 7' or *Out of Cone Corrections*; 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 correction is cone size dependent and jet p_T dependent. The uncertainty is also p_T dependent.

The corrections Level 6 and 7 bring the jet energy scale to parton level energy scale. The rest of the corrections bring the jet energy to particle level energy scale. The quantification of the corrections and their uncertainties described above refer to a jet cone size of 0.4.

[5] The corrections "Level 2" (time dependence of calorimeter photomultipliers) and "Level 3" (Run I-Run II differences) were used in early runs. Currently they are covered in other corrections and not applied anymore.

3.3.6 Neutrino Reconstruction and Identification

The neutrino does not leave any signature in the detector. It is reconstructed and identified, by the energy imbalance in the transverse detector plane. The transverse plane is the only plane where a $p\bar{p}$ collision can be fully reconstructed by CDF detector, thanks to the calorimeters coverage extended up to $\eta = 3.6$.

Due to momentum conservation the event is expected to have a null total transverse momentum.

The presence of a neutrino creates a large apparent imbalance in the total transverse energy as measured by the calorimeters. This imbalance is measured by combining the information from every calorimeter tower and defining the event missing transverse energy $\not\!\!E_T$ (which is, actually missing transverse momentum)[6]:

$$\vec{\not\!\!E}_T^{\,raw} \equiv -\sum_i \vec{E}_T^i \tag{3.2}$$

where $\vec{E}_T^i$ is a vector with magnitude equal to the transverse energy collected by the i-th calorimeter tower and pointing from the interaction vertex to the center of the tower. The sum involves all towers with total energy above 0.1 GeV in the region $|\eta| < 3.6$.

The $\not\!\!E_T$ is corrected for the muons that do not leave much of their energy in the calorimeter. It is also corrected for the jets, which are taken into account in the $\not\!\!E_T$ calculation without corrections, but in a tower by tower basis. For the jet corrections in the $\not\!\!E_T$, the uncorrected jet energy is subtracted by the $\not\!\!E_T$ and the corrected up to level 5 (detector level) jet energy is added:

$$\not\!\!E_T = \not\!\!E_T^{raw} - \sum_{i=1}^{N\ Jets} E_{T_i}^{uncorr.} + \sum_{i=1}^{N\ Jets} E_{T_i}^{corr.} \tag{3.3}$$

The effect of the corrections is significant, especially in events with high jet multiplicity and muons, and it affects not only the $\not\!\!E_T$ absolute value but also the shape.

[6] It is conventional in hadron collider experiments to speak of an object's energy as if it were a vector like the momentum. This is to designate the source of the measurement, the magnetic spectrometer for momentum and the calorimeter for energy. Since the detected particles are traveling close to the speed of light, the magnitude of energy and momentum are the same. Thus, a particle's E_T vector has magnitude $E \sin\vartheta$ and is directed along its transverse momentum vector. if the sum of all particles is zero, then the sum of the E_T vectors will also be zero and there will be no "missing E_T"

3.4 Event Selection

In order to study the semileptonic decays of WW and WZ production, we select events that have a candidate diboson, where a W boson decays in $\mu\nu_\mu$ and the other boson W/Z decays in a jet pair ($W/Z \rightarrow j_1\ j_2$). Our requirements are:

1. $W \rightarrow \mu\nu$

   - Exactly one central muon. The definition of this object is given in Table 3.1.
   - $\not\!\!E_T > 25$ GeV, because below that value the background dominates, as demonstrated in [55]. $\not\!\!E_T$ measured in the calorimeters is corrected taking into account the energy carried off by the muons. An additional correction to $\not\!\!E_T$ is for Jets with $E_T^{uncorr} > 8$ GeV.
   - The leptonic W transverse mass ($M_T(W \rightarrow \mu\nu)$[7]) has to be greater than 30 GeV/c^2. The requirements on $W \rightarrow \mu\nu$ candidate are summarized on Table.3.2

Event variable	Cuts
$\not\!\!E_T$	> 25 GeV
Muon correction	Yes
Jet Correction	$E_T^{uncorr} > 8$ GeV, then corrected L5
$M_T(W \rightarrow \mu\nu)$	> 30 GeV/c^2

Table 3.2: $W \rightarrow \mu\nu$ *Selection.*

2. $W/Z \rightarrow$ jet jet

   - At least 2 jets with $E_T^{jet} > 20$ GeV each.
     The jets are reconstructed using JETCLU04 and corrected at level 7;
   - $\Delta\eta_{j1j2} < 2.5$, $|\eta| < 2.4$ and $p_T > 40$ GeV/c [1] are required in order to achieve good data-MC The requirement on $W/Z \rightarrow j_1\ j_2$ candidate are summarized on Table 3.3:

[7] M_T is the mass on the transverse plane. In this specific case is defined as $M_T(W \rightarrow \mu\nu) = 2p_T^\mu \cdot \not\!\!E_T(\cos\Delta\varphi(\vec{p}_T^{\,\mu}, \vec{\not\!\!E}_T))$

Algoritm	JetClu
Cone	0.4
Correction	Level 7
E_T^{jet}	> 20 GeV
$\|\eta^{jet}\|$	< 2.4
N_{Jets}	$\geqslant 2$
$\Delta\eta(\vec{j_1}\vec{j_2})$	< 2.5

Table 3.3: $W/Z \rightarrow j_1\ j_2$ *Selection.*

Chapter 4

The Monte Carlo

To study physics processes, like in our case the semi-leptonic decays of the WW and WZ production, it is essential to know not only the characteristics of the signal signature but also the contributions of the background processes. The signatures of interest (both signal and background) are initially studied using Monte Carlo events, simulated events that by construction have the same average behavior and the same fluctuations as the data.
In this chapter, the techniques for the Monte Carlo generation are described and the samples used for this analysis are presented. Using the Monte Carlo, the event selection of this thesis was validated in reference [29] by comparing the expectations to the data.

4.1 Monte Carlo at High Energy Physics

To first approximation all processes have a simple structure at the level of interactions between the fundamental objects of nature. However corrections make this simple picture of two quarks producing at the final state two quarks and two lepton, become much more complex; indeed, instead of a 4-particle final state, hundreds of final particles are produced [56]. There is many correction to apply, one of those is the bremsstrahlung-type modifications (e.g. emission of additional final state particles, like photons or gluons). The photon emission process has a sizable effect in electron final states, while gluon emission, because of the large strong coupling, may cause a large flux of particles in the final state. This is called the *parton shower*.

Another very important correction comes from the confined nature of quarks and gluons. The structure of the incoming hadrons as well as the complicated hadronization

process of the quarks and gluons, are not well described in the perturbative language. The chain of processes subsequent to hadronization (e.g. the fragmentation) are very complicated and their description is only based on models.
Monte Carlo generators simplify the complexity problem, by factorizing it into a number of components, each of which can be handled accurately. They form events with the same average behavior and the same fluctuations as in data. This is done by selecting all relevant variables according to the desired probability distributions, ensuring randomness of the final events. The most challenging step in the Monte Carlo generation is the modeling of the perturbative corrections and effects like the fragmentation.

4.1.1 Monte Carlo Generators

Within the CDF collaboration, Pythia ([56]) is the Monte Carlo generator most widely used for the simulation of the electroweak processes. Pythia generates events at Leading Order (LO) approximation and uses the Parton Shower approach to the perturbative corrections modeling, except explicit matrix-element-inspired corrections to the parton shower in specific processes (e.g. inclusive single boson production). This approach offers a good description of the electroweak processes. However, if one is interested in large multiplicity of QCD jets, the description that Pythia provides isn't accurate. For such processes there is the need of a Matrix Element approach. ALPGEN ([57]) is '*A collection of codes for the generation of multi-parton processes in hadronic collisions*' ; it is a Monte Carlo generator that provides Matrix Element evaluation at Leading Order approximation and interfaced with Pythia for the Parton Shower implementation.

The MonteCarlo samples are generated with Alpgen v2.10 prime and PYTHIA v. 6.325 for showering.

4.1.2 Simulated Samples

The most significant background to the WW and WZ search in the lepton plus jets decay channel consists of W plus $jets$ events [1] where the leptonically decaying W boson is produced in association with jets that mimic a hadronically decaying W or Z. Smaller but non-negligible backgrounds come from QCD multi-jet (where one jet mimics a lepton signature), $Z + jets$, $t\bar{t}$, and single top production. Other, less significant backgrounds originate from events with large transverse energy due to the Drell-Yan process ($p\bar{p} \rightarrow \ell^+\ell^- X$), where one of the two leptons is not reconstructed.

The diboson signals as well as the $t\bar{t}$ and single top backgrounds are simulated using the PYTHIA event generator [56]. The W+jets and Z+jets backgrounds are simulated using the tree-level event generator ALPGEN [57], with an interface to PYTHIA providing parton showering and hadronization. QCD multi-jet events are modeled using data with loosened lepton selection criteria (see Section.5.2). All other signal and background processes are modeled using event generators and a geant-based CDF II detector simulation.

W + jets using ALPGEN+Pythia

The case of the $W+$jets is only a specific one, but it is a good example of an ALPGEN+Pythia sample; it is in any case a significant background for the signature of interest in this thesis, but also other various interesting signatures, including Higgs signatures. The good modeling of this process is therefore essential. The different subprocesses included in the $W+$jets calculation are listed in Fig. 4.1.

jproc	subprocess	jproc	subprocess	jproc	subprocess
1	$q\bar{q}' \to W$	2	$qg \to q'W$	3	$gq \to q'W$
4	$gg \to q\bar{q}'W$	5	$q\bar{q}' \to Wq''\bar{q}''$	6	$qq'' \to Wq'q''$
7	$q''q \to Wq'q''$	8	$q\bar{q} \to Wq'\bar{q}''$	9	$q\bar{q}' \to Wq\bar{q}$
10	$\bar{q}'q \to Wq\bar{q}$	11	$q\bar{q} \to Wq\bar{q}'$	12	$q\bar{q} \to Wq'\bar{q}$
13	$qq \to Wqq'$	14	$qq' \to Wqq$	15	$qq' \to Wq'q'$
16	$qg \to Wq'q''\bar{q}''$	17	$gq \to Wq'q''\bar{q}''$	18	$qg \to Wqq\bar{q}'$
19	$qg \to Wq'q\bar{q}$	20	$gq \to Wqq\bar{q}'$	21	$gq \to Wq'q\bar{q}$
22	$qg \to Wq'q'\bar{q}'$	23	$gq \to Wq'q'q'$	24	$gg \to Wq\bar{q}'q''\bar{q}''$
25	$gg \to Wq\bar{q}q\bar{q}'$				

Figure 4.1: *Subprocesses included in the* $W(\to \mu\nu_\mu)+$jets *code. Additional final state gluons are not explicitely shown here but are included in the calculations. [57]*

The W +jets sample is composed of sub-samples divided according to the parton multiplicity ($W+0$p, $W+1$p, ..., $W+n$p). For a given multiplicity κ, parton-level configurations are generated with partons constrained by a minimum p_T. PYTHIA is then used for the jet showering and using a jet algorithm, jets are formed at generator level (there are no detector effects applied yet). The generated κ partons are matched with jets in ΔR[1]. A jet can only be matched to one parton. If all κ partons are matched with jets, then the event is kept, else it is discarded. By following this algorithm, the

[1] $\Delta R = \sqrt{(\Delta\eta^2 + \Delta\varphi^2)}$

samples that are generated for each multiplicity will be inclusive samples for this specific multiplicity. Exclusive samples are formed by requiring that the number of reconstructed jets (at generator level) be equal to the number of partons κ. By merging exclusive jet samples we obtain an inclusive sample that contains all jet multiplicity[2]. At CDF the W+jets sample consists of 5 sub-samples. All are exclusive in multiplicity, except the last one that is inclusive ($W + 4\mathrm{p}$).

4.1.3 Signal and Background Modeling

The modeling of the signal and background processes of interest is performed using MC with the exception of the QCD background. The various simulated samples used in the analysis are listed in Table 4.1 and Table.4.2.

For each sample i, the cross section ($\sigma^{\mathrm{i}}_{\mathrm{theor}}$) of the process and the number of generated events ($\mathrm{N}^{\mathrm{i}}_{\mathrm{gen}}$) mentioned in this table can be used to evaluate the $weight$ used for scaling MC sample to the data since:

$$weight = \frac{\mathrm{L} \times \sigma^{\mathrm{i}}_{\mathrm{theor}}}{\mathrm{N}^{\mathrm{i}}_{\mathrm{gen}}} \tag{4.1}$$

where L is the experimental integrated luminosity used in this study.

Alpgen generates different samples for heavy flavour, $\mathrm{Z} + \mathrm{b\bar{b}} + \mathrm{Np}$ and $\mathrm{Z} + \mathrm{c\bar{c}} + \mathrm{Np}$. In the light flavour sample Alpgen generates quark up, down, strange and charm without mass, while, during the showering, Pythia generates all five flavours with mass. It is then necessary to remove the double counting between the Z+Np sample that contains b's and c's from the Pythia showering and the heavy flavour sample. A way to handle this overlap is to remove events that have jets that come from a *b* or a *c* quark with a bottom or charm hadron inside the cone of the jet (0.4 in our case) in the light sample. Then, we take the c sample and remove events that have jets with a b inside the cone of the jet. We don't have to remove anything from the b sample because a cut in the $\mathrm{p_T}$ of the jet removes all the charm. After this procedure we can simply add all the samples togheter. The only background that is not extracted by MC is the QCD contribution (see Section5.2). The other sources are extimated by the MC samples of Table.4.1 and Table.4.2.

[2] The procedure described is called "MLM prescription"

Process	σ_{theor}[pb]	Sample	N_{gen}
$WW \rightarrow \mu\nu + jets$	12.40×0.073	itopww	166795
$WZ \rightarrow \mu\nu + jets$	3.70×0.076	itopwz	175301
$W(\mu\nu) + 0p$	1800	ptopw5	5017218
$W(\mu\nu) + 1p$	225	ptopw6	5003166
$W(\mu\nu) + 2p$	35.3	ptop7w	1002804
$W(\mu\nu) + 3p$	5.59	ptop8w	1013373
$W(\mu\nu) + \geqslant 4p$	1.03	ptop9w	988545
$W(\tau\nu) + 0p$	1800	utopw0	4885557
$W(\tau\nu) + 1p$	225	utopw1	4987134
$W(\tau\nu) + 2p$	35.3	utop2w	92398
$W(\tau\nu) + 3p$	5.59	utop3w	1008221
$W(\tau\nu) + \geqslant 4p$	1.03	utop4w	186494
$W(\mu\nu) + b\bar{b} + 0p$	2.98	btop5w	1524880
$W(\mu\nu) + b\bar{b} + 1p$	0.888	btop6w	1508029
$W(\mu\nu) + b\bar{b} + \geqslant 2p$	0.287	btop7w	1506613
$W(\mu\nu) + c\bar{c} + 0p$	5.	ctop5w	1985033
$W(\mu\nu) + c\bar{c} + 1p$	1.79	ctop6w	1979810
$W(\mu\nu) + c\bar{c} + \geqslant 2p$	0.628	ctop7w	1970504
$Z(\mu\mu) + 0p$	158	ztopp5	2665104
$Z(\mu\mu) + 1p$	21.60	ztopp6	2664729
$Z(\mu\mu) + 2p$	3.47	ztop7p	530843
$Z(\mu\mu) + 3p$	0.548	ztop8p	536159
$Z(\mu\mu) + \geqslant 4p$	0.0992	ztop9p	536159
$Z(\mu\mu) + b\bar{b} + 0p$	0.511	ztopb5	437329
$Z(\mu\mu) + b\bar{b} + 1p$	0.134	ztopb6	494480
$Z(\mu\mu) + b\bar{b} + \geqslant 2p$	0.0385	ztopb7	478485
$Z(\mu\mu) + c\bar{c} + 0p$	1.08	ztopc5	671375
$Z(\mu\mu) + c\bar{c} + 1p$	0.331	ztopc6	663431
$Z(\mu\mu) + c\bar{c} + \geqslant 2p$	0.107	ztopc7	705108
$Z(\tau\tau) + 0p$	158.	ztopt3	5860164
$Z(\tau\tau) + 1p$	21.5	ztopt4	5864300
$Z(\tau\tau) + \geqslant 2p$	4.14	ztopt2	2273221
$t\bar{t}$	6.7	ttkt75	5445003

Table 4.1: *Summary of simulated samples. The samples names are the CDF descriptions.*

Process	σ_{theor}**[pb]**	**Sample**	N_{gen}
for $m_{\ell\ell} = [20, 75]$GeV			
DY($\tau\tau$) + 0p	160	xtopt0	1136851
DY($\tau\tau$) + 1p	8.3	xtopt1	1153959
DY($\tau\tau$)+ $\geqslant$ 2p	1.82	xtopt2	2270345
for $m_{\ell\ell} = [105, 800]$GeV			
DY($\tau\tau$) + 0p	4.07	zttt0h	268428
DY($\tau\tau$) + 1p	0.707	zttt1h	268428
DY($\tau\tau$) + 2p	0.117	zttt2h	263291
DY($\tau\tau$) + 3p	0.0185	zttt3h	268428
DY($\tau\tau$)+ $\geqslant$ 4p	0.0033	zttt4h	56398
for $m_{\ell\ell} = [20, 75]$GeV			
DY($\mu\mu$) + 0p	160.	xtop5p	524357
DY($\mu\mu$) + 1p	8.3900	xtop6p	530696
DY($\mu\mu$) + 2p	1.6100	xtop7p	525769
DY($\mu\mu$) + 3p	0.2330	xtop8p	524697
DY($\mu\mu$)+ $\geqslant$ 4p	0.0398	xtop9p	529635
for $m_{\ell\ell} = [105, 800]$GeV			
DY($\mu\mu$) + 0p	4.0700	ytop5p	536159
DY($\mu\mu$) + 1p	0.7060	ytop6p	529581
DY($\mu\mu$) + 2p	0.1170	ytop7p	531006
DY($\mu\mu$) + 3p	0.0185	ytop8p	520531
DY($\mu\mu$)+ $\geqslant$ 4p	0.0033	ytop9p	527838

Table 4.2: *Summary of simulated samples. The samples names are the CDF descriptions.*

Chapter 5

Analysis Description

5.1 Analysis Methodology Overview

The aim of this study is to improve the statistical sensitivity of signal fraction estimator for the diboson WW/WZ in $\ell\nu_\ell + jets$ analysis of [1, 2, 29], where lepton is a muon or an electron, using multivariate techniques.
This study originates from the results of [1, 2, 29], subject of a PhD thesis that will be discussed at end of 2010.
The same analysis framework and the same reconstructed diboson candidate events on data and simulated samples of [1, 2, 29], resumed respectively in Chapter3 and 4, are used.

We will test the improvement due to the new techniques using only the muon channel for the W decaying leptonically.
In [1, 2, 29], in order to extract the diboson signal on data, a 1-dimensional fit to the invariant mass distribution of the two jets (M_{jj}) is performed.
The multivariate techniques used in this thesis, take advantage from exploiting additional kinematical quantities of the diboson candidate to optimally discriminate the signal from the background events.
At first, we look for additional kinematic variables of the WW/WZ candidate event that show signal/background separation power and whose Monte Carlo (MC) modeling can be considered satisfactory (Section 5.3.1). Then, we use the selected separating variables to train multivariate discriminants (MVD, Section 5.3.4) using simulated samples for signal

and backgrounds. As discriminants we consider, Boosted Decision Trees (BDT, Section 5.3.2) and Projected Likelihood Estimator (Likelihood, Section 5.3.3). We then apply the MVD weights to each sample (also to data) and we have a discriminant output (MVDO) for each sample (Signal, QCD, Top and EWK) and for data.

Each MVDO and the M_{jj} are used to build 2-dimensional distributions (templates) of MVDO versus M_{jj}.

These templates are then used in a 2-dimensional fit on data to extract the fraction of signal events, that is the number of interest of this analysis. Before applying the fit procedure on data we validate and asses the performances of the 2-dimensional and 1-dimensional (on M_{jj}) fits using pseudo-experiments.

5.2 Starting Point

Our analysis begins from the results of [1, 2, 29] where, after the reconstruction of the diboson candidate (resumed in Section 3.4), the signal fraction is extracted from the data, by performing a χ^2 fit to the dijet invariant mass spectrum (M_{jj}), evaluated from the two most energetic jets (leading jets).
In the following we used the same analysis framework of [29]. The templates of M_{jj} distribution are constructed using the simulated samples of Table.4.1 and Table.4.2. These are for signal:

- WW/WZ

for backgrounds:

- $W(\mu\nu_\mu) + \text{jets}$,
- $W(\tau\nu_\tau) + \text{jets}$,
- $W + \text{Heavy Flavour}$ ($W+\text{HF}$) [$W \to \mu\nu_\mu + b\bar{b} + \text{jets}$ and $W \to \mu\nu_\mu + c\bar{c} + \text{jets}$],
- $Z(\mu^+\mu^-) + \text{jets}$,
- $Z(\tau^+\tau^-) + \text{jets}$,
- $Z + \text{Heavy Flavour}$ ($Z+\text{HF}$) [$Z \to \mu\mu + b\bar{b} + \text{jets}$ and $Z \to \mu\mu + c\bar{c} + \text{jets}$],
- $\text{Drell} - \text{Yan}$ (DY) [$\mu^+\mu^- + \text{jets}$ and $\tau^+\tau^- + \text{jets}$]
- $t\bar{t}$

The QCD background ($p\bar{p} \rightarrow jets$) template is extracted from data using the high isolation muon sample [29]($Iso > 0.2$, see Section 3.3.2).
The normalization of the QCD template is estimated by fitting the $\not\!\!E_T$ spectrum in data to the sum of all contributing processes, where the QCD and W+jets normalizations float in the fit [29].

In this study samples are further grouped in four categories:

- **Signal**: WW/WZ
- **EWK**: $W(\mu\nu_\mu)$+jets, $W(\tau\nu_\tau)$+jets, $Z(e^+e^-)$+jets, $Z(\mu^+\mu^-)$+jets, $Z(\tau^+\tau^-)$+jets, $W+$HF, $Z+$HF, DY
- **Top**: $t\bar{t}$
- **QCD**

In Fig.5.1 is reported the Dijet invariant mass distribution of reconstructed $W/Z \rightarrow jj$ candidates, with overlayed the expected stacked Monte Carlo contributes and the QCD background. Here the contributes of signal and background from simulated sample are evaluated using the weighting described in Section 4.1.3 while for the QCD background, we use the procedure mentioned above. The data MC agreement is good.

5.3 The Multivariate Analysis

Multivariate analysis (**MVA**) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest. Multivariate analysis is the best way to summarize a data tables with many variables by creating a few new variables containing most of the information. These new variables are then used for problem solving and display, i.e., classification, relationships, control charts, and more.

In this analysis we use the *BDT* (Section 5.3.2) method and the *Likelihood* (Section 5.3.3) method implemented, along with several other classifiers, in the **Toolkit for MultiVariate Analysis (TMVA)**, available in the ROOT package[59].
In the TMVA framework, an analysis procedure is typically organized in two indipendent stages, the *training* and the *application*.

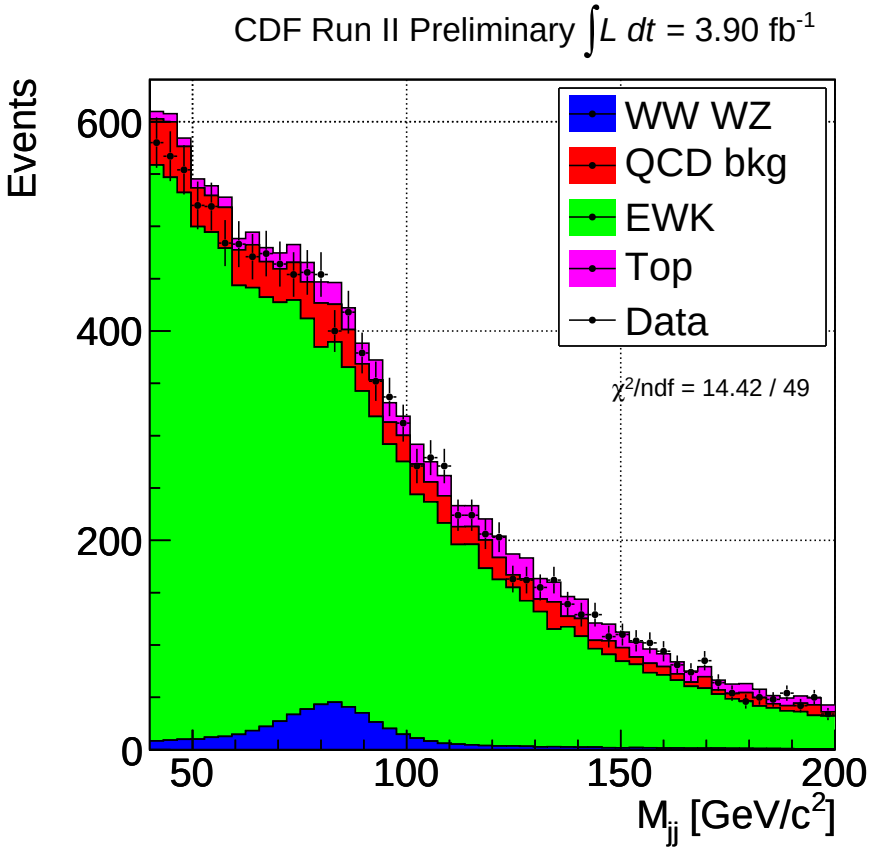


Figure 5.1: M_{jj} *distribution on data with overlayed the expected stacked contributes from MC and QCD*

In the training stage the user defines a *Factory* object, which allows one to book and configure the desired multivariate classifier, specify the training datasets and register the input variables to be used for discrimination. The *Factory* then calls several functions for the training, testing and evaluation of the classifier and stores the final results called **weight** files.

In the application stage the user defines a *Reader* object, where the results of the training are retrieved and applied to a dataset of unknown composition: each event is thus analyzed and given a score, this score is the Multivariate Discriminant Output (MVDO) of the used method.

In the *weight* files also a large variety of control and performance plots are stored, which can be displayed via a graphical user interface (GUI), by running a set of ROOT macros. More details about TMVA structure and all configuration parameters can be found in the

User Guide [58].

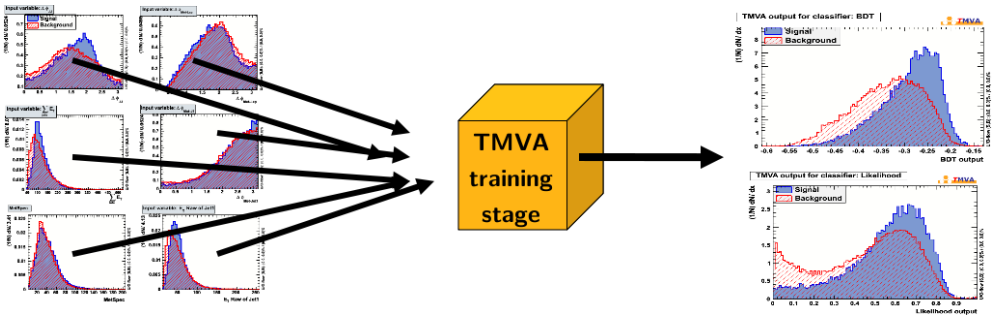


Figure 5.2: *TMVA framework: training stage*

5.3.1 Separating Variables

First step in a multivariate analysis to be applyed in diboson studies is determine an ensemble of kinematical quantities having a satisfactory balance between signal/background discriminating power and agreement between data and MC expectation. We investigate several variables to be used in the MVD.

As mentioned in the Section 5.1 the M_{jj} will be used as one of the axis of the 2-dimensional fit and is not included in the variables used to determine the MVDO, the other axis. The additional variables we consider are divided in 3 categories:

- the variables related to the leptonically decaying W:
  - $M_T(W \to \mu\nu_\mu)$, invariant transverse mass of W,
  - $P_T(W \to \mu\nu_\mu)$, momentum on transverse plane of the W,
  - $\not{E}_T$, missing E_T in the transverse plane,
  - $\Delta\varphi_{\not{E}_T-\ell}$, angle on transverse plane between $\vec{\not{E}}_T$ and muon direction,
- the variables related to the both bosons candidates:
  - $\Delta\varphi_{\not{E}_T-j1}$, angle on transverse plane between $\vec{\not{E}}_T$ and leading jet direction,
  - $\Delta\varphi_{\not{E}_T-j2}$, angle on transverse plane between $\vec{\not{E}}_T$ and second leading jet direction,
  - $\not{E}_T$ *spec*, is contructed using an algorithm. If the lepton or the leading jet or the second leading jet have an angle smaller than $\pi/2$ in the transverse plane,

the minimum of $\not\!\!E_T \sin(\widehat{\vec{\not\!\!E}_T \; \vec{part}})$[1] is evaluated for each $part$, where $part$ can be lepton, jet_1 or jet_2.

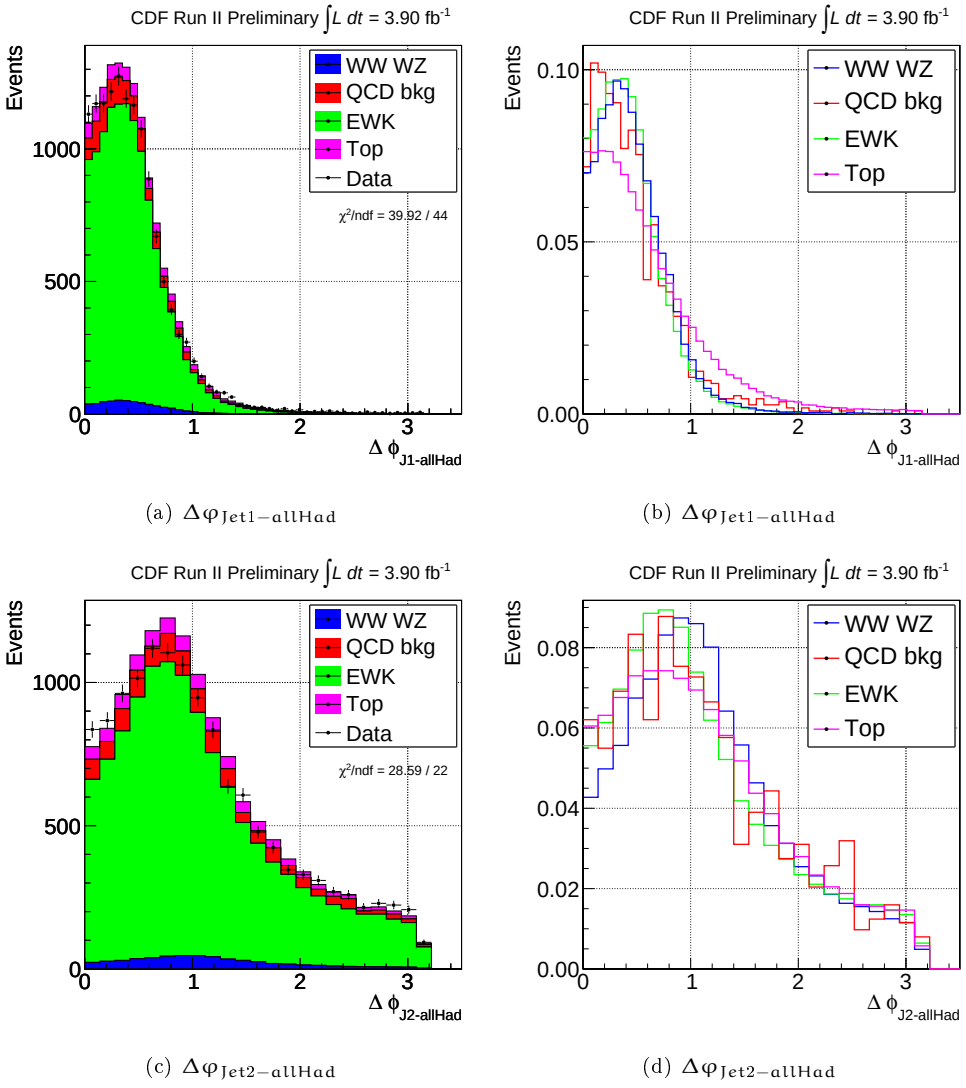
$\not\!\!E_T$ $spec$ is the minimum between $\not\!\!E_T$ and the evaluated value.

- the ones related to the hadronically decaying W/Z:
  - $E_{T\ j1}^{raw}$, transverse energy of the leading jet,
  - $E_{T\ j2}^{raw}$, transverse energy of the second leading jet,
  - η_{J1}, η of the leading jet,
  - η_{J2}, η of the second leading jet,
  - $\Delta\varphi_{j1j2}$, angle on transverse plane between two leading jets directions,
  - $\Delta\varphi_{J_1-allHad}$, angle on transverse plane between the leading jet and the vectorial sum of all the other jets (second leading excluded),
  - $\Delta\varphi_{J_2-allHad}$, angle on transverse plane between the second leading jet and the vectorial sum of all the other jets (leading excluded),
  - $\cos\vartheta^*_{j1}$, cosine of the angle between hadronically decaying W direction and the leading jet on the center of mass system.

Next step is to analize our variable to understand if we could use all variables set or we have to choose a subgroup.

Fig.5.3-5.10 show the variables distributions.
In each figure on left is reported the variable distribution from data with superimposed the corresponding MC (EWK, Top and Signal) and the QCD background stacked expected contributes. On the right is reported the variable's normalized MC contributions (EWK, Top and Signal) and QCD background.

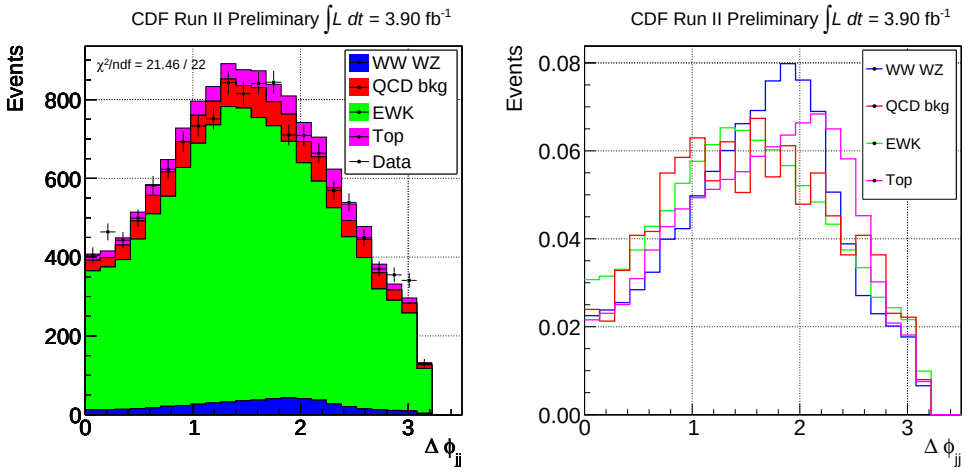
[1] If the lepton or one of the two leading jets is badly recontructed in energy, the $\vec{\not\!\!E}_T$ is along the direction of this object.



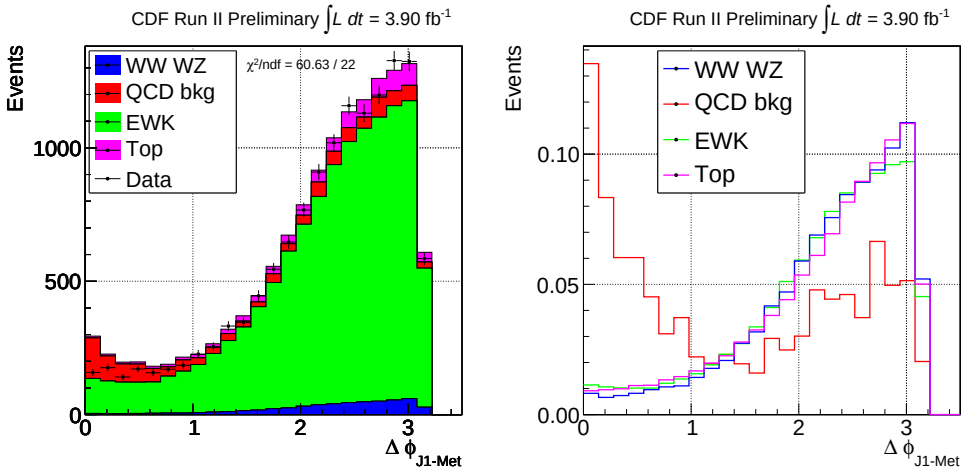
(a) $\Delta\varphi_{\mathrm{Jet1-allHad}}$ (b) $\Delta\varphi_{\mathrm{Jet1-allHad}}$

(c) $\Delta\varphi_{\mathrm{Jet2-allHad}}$ (d) $\Delta\varphi_{\mathrm{Jet2-allHad}}$

Figure 5.3: *Left plots: in each plot is reported the variable distribution from data with overlayed the corresponding MC and QCD stacked expected contributes. Right plots: in each plot is reported, for each variable, the normalised distributions of MC (EWK, Top and Signal) and QCD background.*

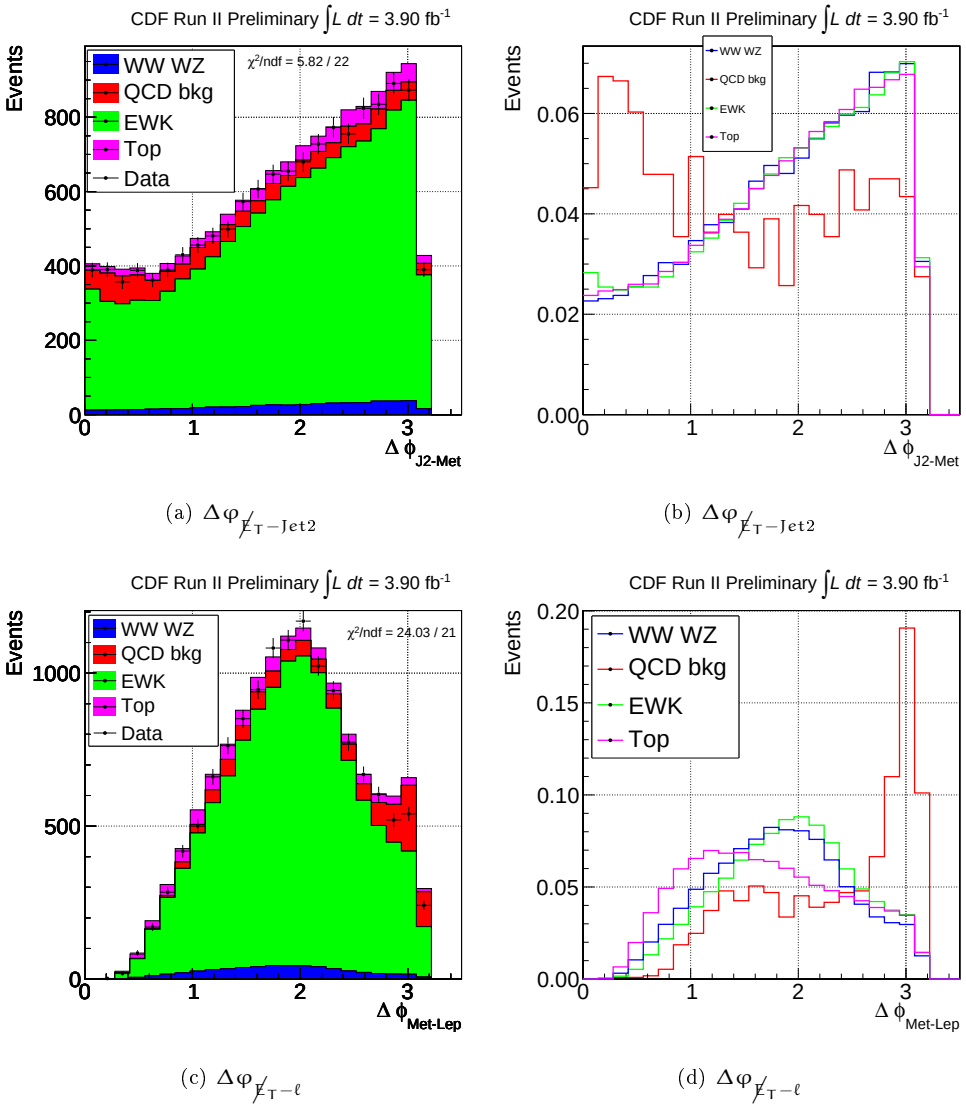


(a) $\Delta\varphi_{\mathrm{Jet1-Jet2}}$ (b) $\Delta\varphi_{\mathrm{Jet1-Jet2}}$



(c) $\Delta\varphi_{\not{E}_T-\mathrm{Jet1}}$ (d) $\Delta\varphi_{\not{E}_T-\mathrm{Jet1}}$

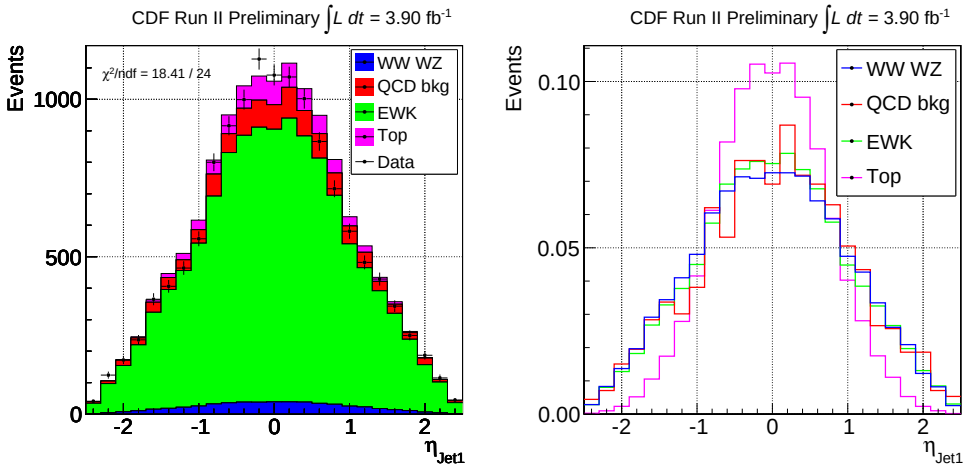
Figure 5.4: *Left plots: in each plot is reported the variable distribution from data with overlayed the corresponding MC and QCD stacked expected contributes. Right plots: in each plot is reported, for each variable, the normalised distributions of MC (EWK, Top and Signal) and QCD background.*



(a) $\Delta\varphi_{\not\!\!E_T-\mathrm{Jet2}}$ (b) $\Delta\varphi_{\not\!\!E_T-\mathrm{Jet2}}$

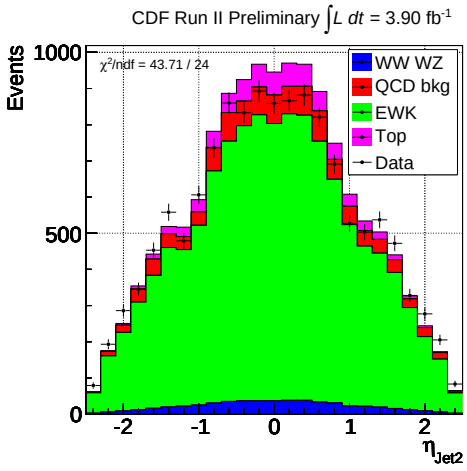
(c) $\Delta\varphi_{\not\!\!E_T-\ell}$ (d) $\Delta\varphi_{\not\!\!E_T-\ell}$

Figure 5.5: *Left plots: in each plot is reported the variable distribution from data with overlayed the corresponding MC and QCD stacked expected contributes. Right plots: in each plot is reported, for each variable, the normalised distributions of MC (EWK, Top and Signal) and QCD background.*

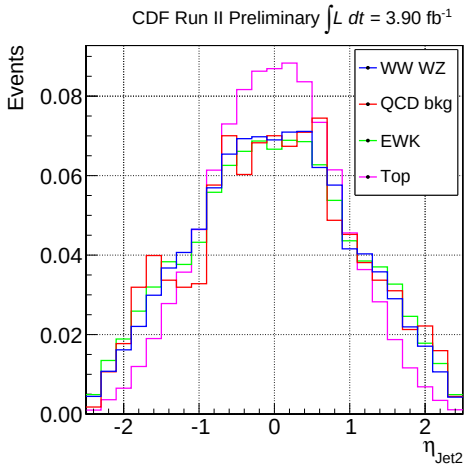


(a) η_{Jet1}

(b) η_{Jet1}

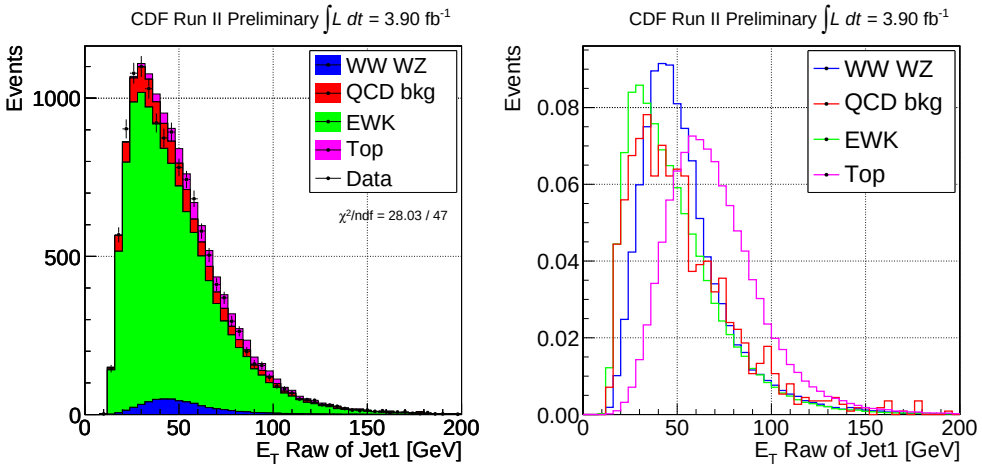


(c) η_{Jet2}



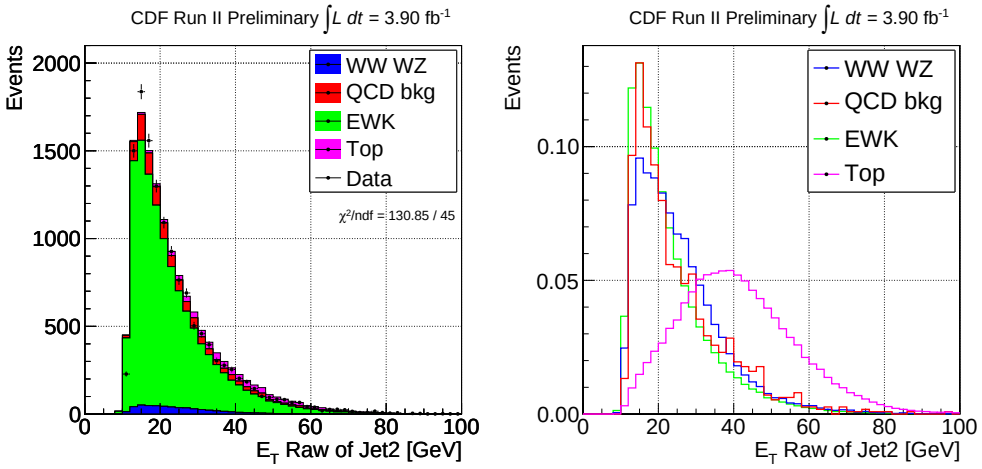
(d) η_{Jet2}

Figure 5.6: *Left plots: in each plot is reported the variable distribution from data with overlayed the corresponding MC and QCD stacked expected contributes. Right plots: in each plot is reported, for each variable, the normalised distributions of MC (EWK, Top and Signal) and QCD background.*



(a) E_T of Jet1 Raw

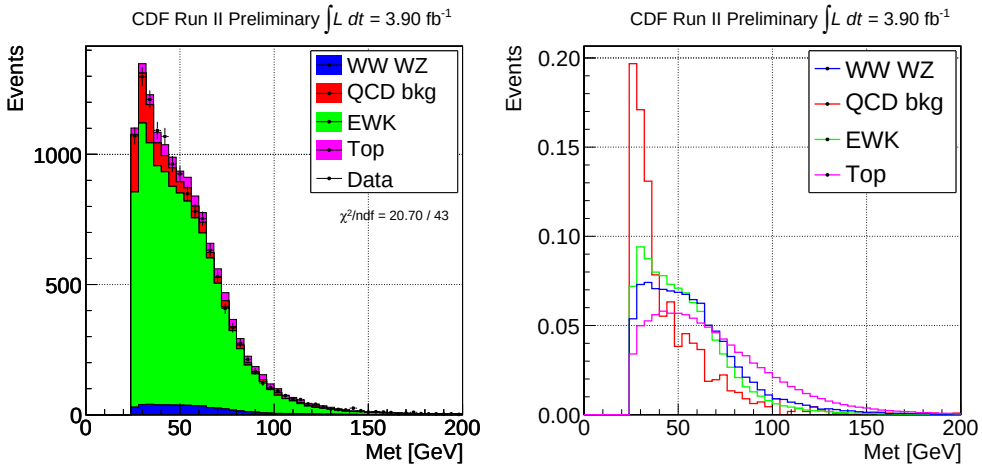
(b) E_T of Jet1 Raw



(c) E_T of Jet2 Raw

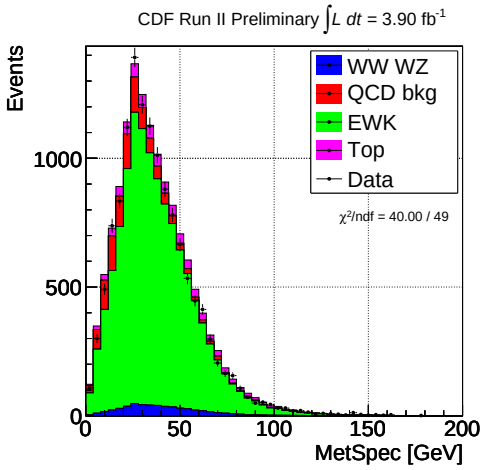
(d) E_T of Jet2 Raw

Figure 5.7: *Left plots: in each plot is reported the variable distribution from data with overlayed the corresponding MC and QCD stacked expected contributes. Right plots: in each plot is reported, for each variable, the normalised distributions of MC (EWK, Top and Signal) and QCD background.*

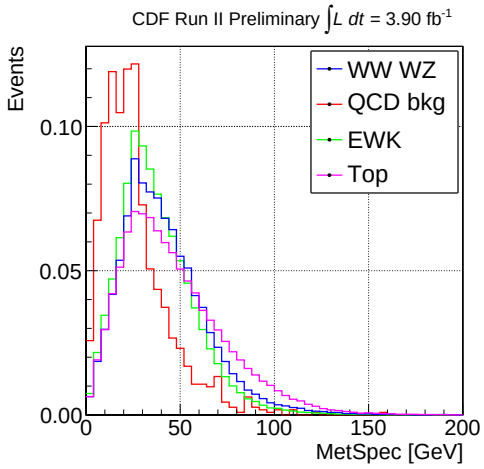


(a) $\not\!\!E_T$

(b) $\not\!\!E_T$

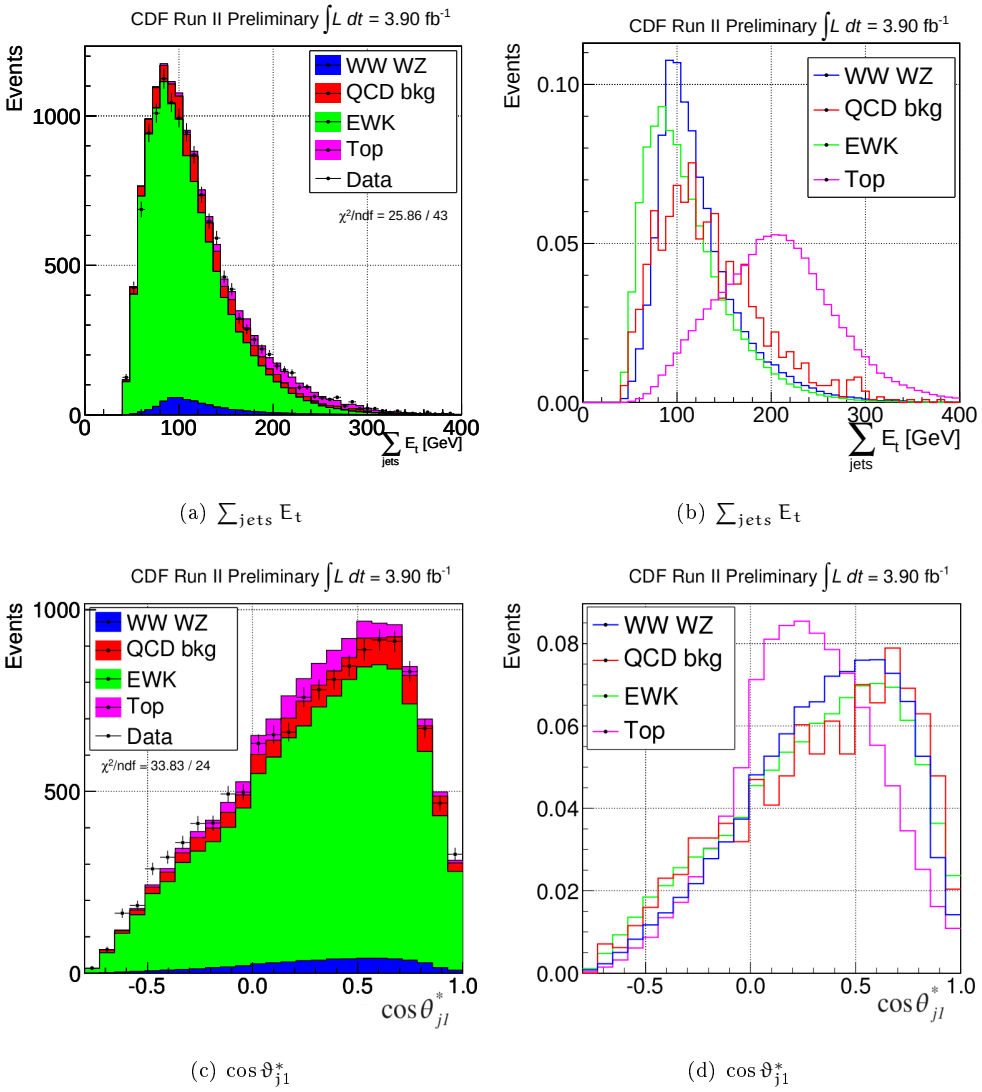


(c) $\not\!\!E_T$ Special



(d) $\not\!\!E_T$ Special

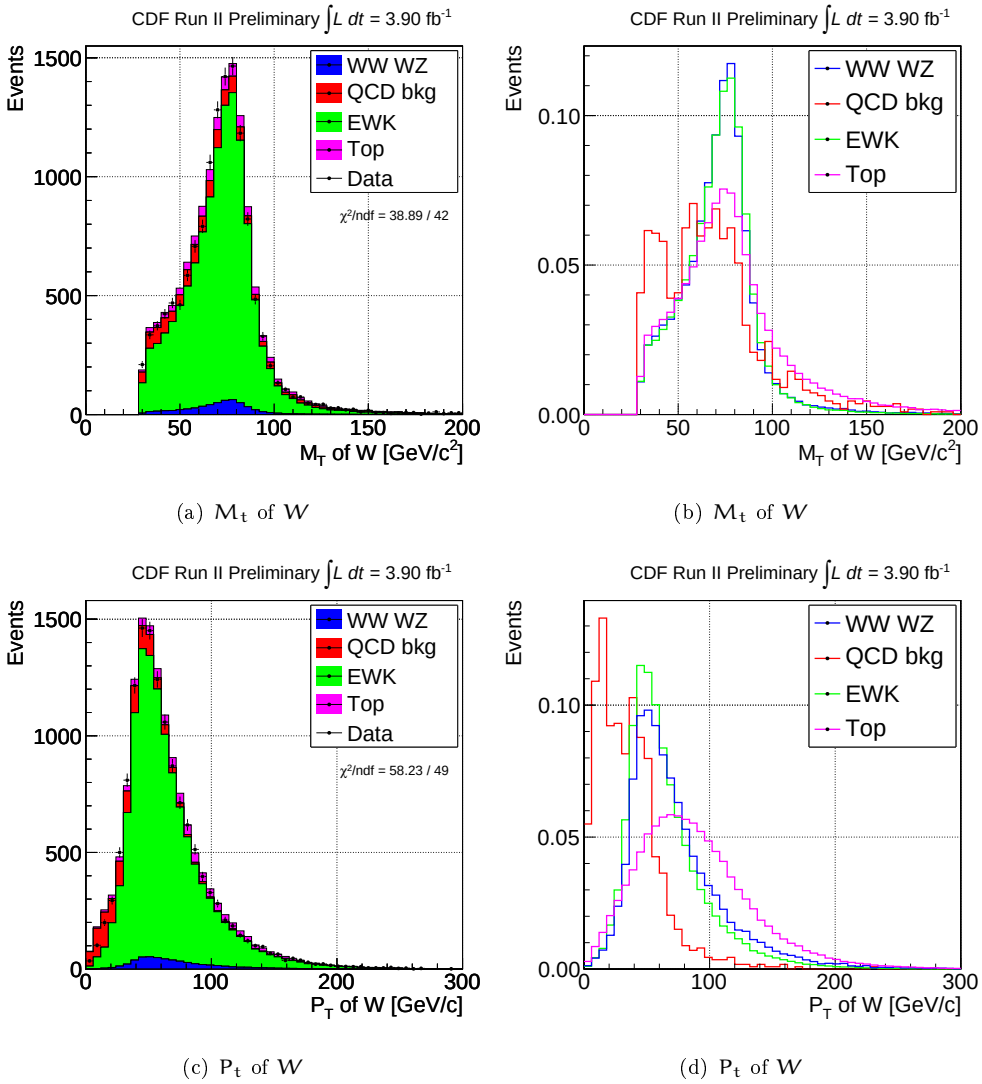
Figure 5.8: *Left plots: in each plot is reported the variable distribution from data with overlayed the corresponding MC and QCD stacked expected contributes. Right plots: in each plot is reported, for each variable, the normalised distributions of MC (EWK, Top and Signal) and QCD background.*



(a) $\sum_{\text{jets}} E_t$ (b) $\sum_{\text{jets}} E_t$

(c) $\cos\vartheta^*_{j1}$ (d) $\cos\vartheta^*_{j1}$

Figure 5.9: *Left plots: in each plot is reported the variable distribution from data with overlayed the corresponding MC and QCD stacked expected contributes. Right plots: in each plot is reported, for each variable, the normalised distributions of MC (EWK, Top and Signal) and QCD background.*



(a) M_t of W

(b) M_t of W

(c) P_t of W

(d) P_t of W

Figure 5.10: *Left plots: in each plot is reported the variable distribution from data with overlayed the corresponding MC and QCD stacked expected contributes. Right plots: in each plot is reported, for each variable, the normalised distributions of MC (EWK, Top and Signal) and QCD background.*

Left plots show a reasonable agreement between data and expected contributes. The right plots let us to understand the signal/background discriminating power of each variable.

Also if there is a reasonable agreement between data and expected distributions, we need variables with separating power as strong as possible.

Because of that we reject variables that don't have well separation power, the variables that don't add new information and the redundant variables. So, among the variables previously shown, we select the following subgroup:

Separating Variables
$\Delta\varphi_{J_1 J_2}$
$\Delta\varphi_{\not{E}_T - lep}$
$\sum_{jets} E_t$
$\Delta\varphi_{\not{E}_T - J_1}$
$\not{E}_T Spec$
$E_{T\ j1}^{raw}$
$E_{T\ j2}^{raw}$
η_{J_1}
$\cos\vartheta_{j1}^{*}$

Table 5.1: *List of separating variables used for discriminant analysis*

5.3.2 Boosted Decision Trees Method

A *decision tree* [60] is a sequence of rooted binary splits, performed using a set of discriminating variables. A sketch of the underlying logic is presented in Fig. 5.11. Given a training sample made of known signal and background events, repeated decision are performed: at each *node* the variable and the split value which give the best separation are selected and two classes (or child nodes) are created. When a predefined criterium is met, splitting stops and the terminal nodes are called *leafs* and tagged as signal **(S)** or background **(B)** according to their purity P, defined as the weighted signal fraction of the training sample:

$$P = \frac{\sum_s W_s}{\sum_s W_s + \sum_b W_b} \tag{5.1}$$

where W_s and W_b are the weights assigned to signal and background input events. If an event lands on a background leaf, it is given a score of -1, while if it lands on a signal leaf, it is given a score of +1.

The *boosting* consists in the creation of several trees, a *forest*: training events which were misclassified in the n-th tree, have their weight increased in the (n+1)-th tree. The result of increasing the weights is that these events become more important and it is more difficult for the new tree to misclassify them again: in this way the algorithm "learns" from previous errors. For each event the final score is then given by the weighted average of the different tree outputs.

The boosting procedure stabilizes the response of the decision trees and makes the algorithm more robust and less sensible to fluctuations in the samples. However, in some cases this feature is not enough to avoid the occurence of *overtraining*, a seeming increase in the classification performance evaluated with the training sample, with respect to the real performance measured on the independent test one.

Overtraining affects a multivariate method if too many parameters need to be adjusted to too few data points. BDT is particular sensitive to this effect, when splitting is extended to nodes with very poor statistics. In this case a decision tree must be *pruned*.

Pruning is the process of cutting back a tree from the bottom, by removing final statistically insignificant nodes. This procedure is performed after the growth of the tree to its maximum size, because it has been demonstrated that some splits may appear useless at one node but may lead to good splits in subsequent steps.

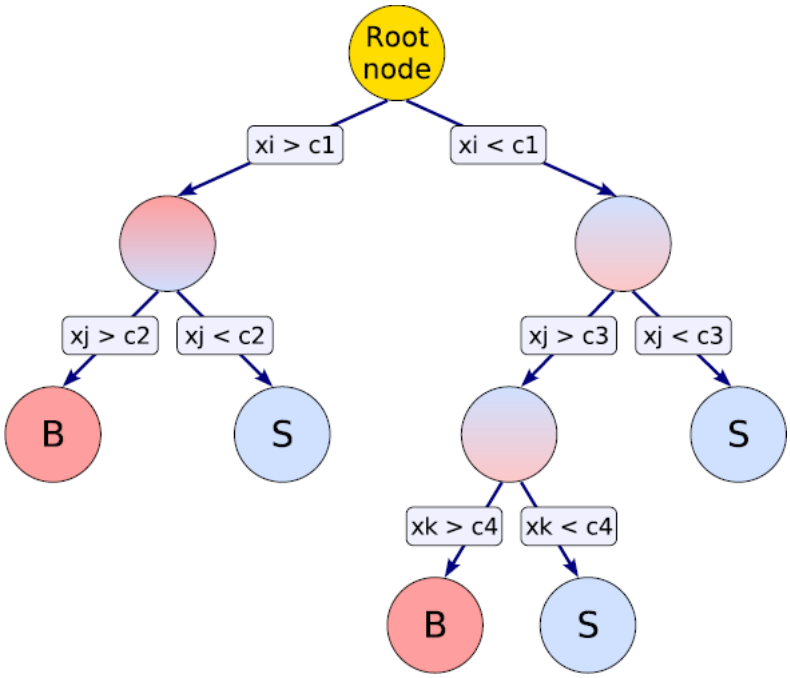


Figure 5.11: *An example picture of a decision tree. At each node, the most discriminating variable* x_i *is chosen and the cut which gives the best separation between signal and background is applied. When the predefined criterium is reached, splitting stops and each leaf is labelled as* ***S*** *or* ***B***.

5.3.3 Projective Likelihood Estimator Method

The method of maximum likelihood [58] consists of building a model out of probability density functions (PDF) that reproduces the input variables for signal and background. For a given event, the likelihood for being of signal type is obtained by multiplying the signal probability densities of all input variables, and normalising this by the sum of the signal and background likelihoods. Correlations among the variables are ignored.

The likelihood ratio $y_{\mathcal{L}}(i)$ for event i is defined by

$$y_{\mathcal{L}}(i) = \frac{\mathcal{L}_S(i)}{\mathcal{L}_S(i) + \mathcal{L}_B(i)}, \tag{5.2}$$

where

$$\mathcal{L}_{S(B)}(i) = \prod_{k=1}^{n_{var}} p_{S(B),k}(x_k(i)), \tag{5.3}$$

and where $p_{S(B),k}$ is the signal (background) PDF for the kth input variable x_k. The PDFs are normalised

$$\int_{-\infty}^{+\infty} p_{S(B),k}(x_k)dx_k = 1, \qquad \forall k. \tag{5.4}$$

It can be shown that in absence of correlations between input variables, the ratio (5.2) provides optimal signal from background separation for the given set of input variables.

Since the parametric form of the PDFs is generally unknown. In the TMVA framework the PDF shapes are empirically approximated from the training data by non parametric functions.
The present likelihood implementation does not provide a ranking of the input variables.

The performance of the classifier relies on the accuracy of the likelihood model. Because high fidelity PDF estimates are mandatory, sufficient training statistics is required to populate the tails of the distributions. The neglect of correlations between input variables in the model (5.3), often leads to a diminution of the discrimination performance.

5.3.4 Training and Application

In the TMVA framework, an analysis procedure is typically organized in two independent stages, the training and the application. The training stage of the Projective method is represented by the estimation of the probability density function of each input variable. The BDT training stage represents the definition of the binary splits rules, subsequent pruning and boosting of the trees.

In the training stage, our signal is represented by the WW/WZ simulated sample, while as background we consider all the MC contributions (EWK + Top) with the exception of the QCD sample. As previously mentioned, the relative normalisations are taken from the MC expectation.

The separating variables of Table. 5.1 are used as input to the two discriminants. Their normalized distribution, for signal and background, are shown on Fig. 5.12.

One of the TMVA framework output is the correlation matrix of variables separate for signal and background. This correlation matrix is reported in Fig.5.13 for signal (left) and background (right).

For each classifier the training stage returns the signal and the background superimposed distributions (Fig. 5.14) from which we can understand our separating power.

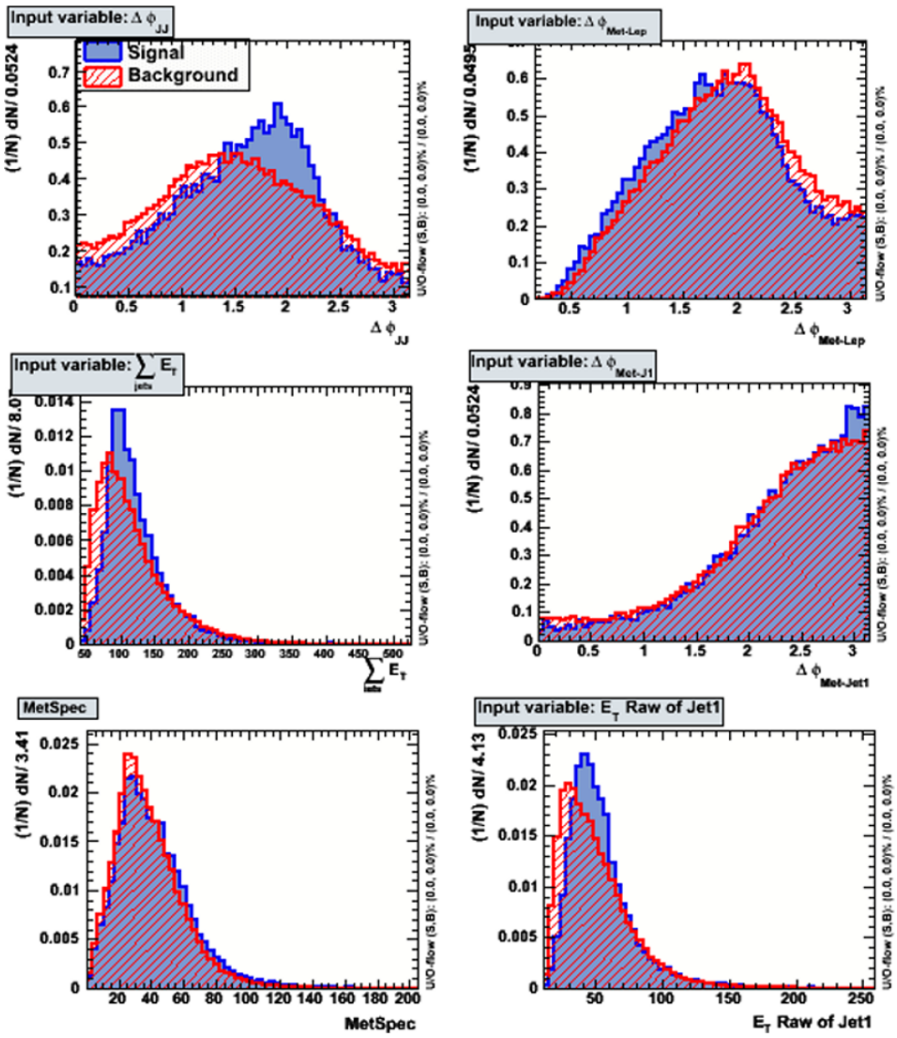


Figure 5.12: *Training input variables Signal (Blue) and Background (Red).*

The training stage returns also weights files (one for each discriminant), these are used on each sample in the application stage.

In the application stage for each sample (Signal, EWK, Top and also QCD) the separating variables are analized in each event and a score is given, this score is the Multivariate Discriminant Output of the used method (BDT or Likelihood).

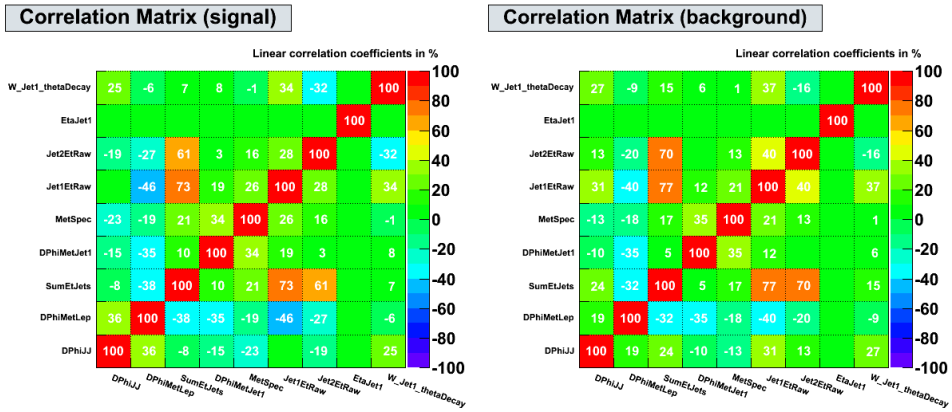
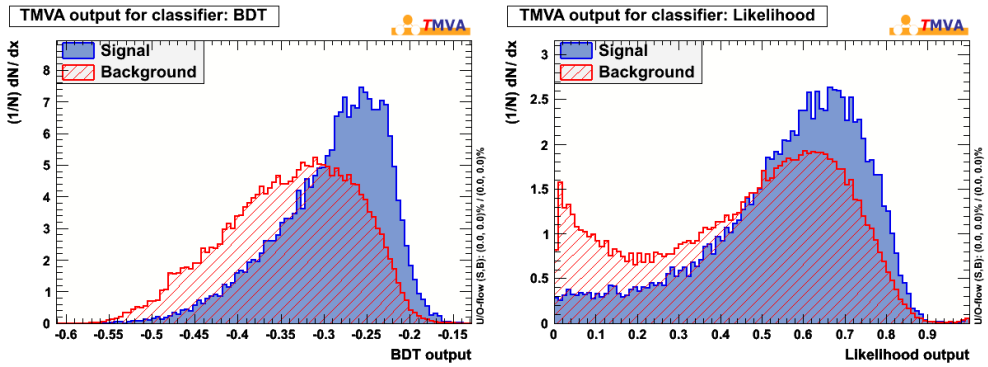


Figure 5.13: *Variables correlation matrix for Signal (left) and Background (right). Linear correlation coefficients are in %.*



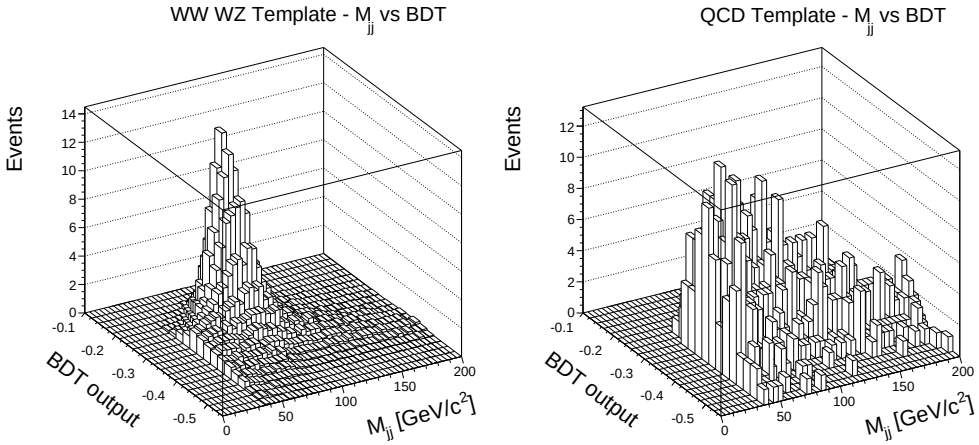
(a) BDT output (b) Likelihood output

Figure 5.14: *Training output for BDT (left) and Likelihood (right).*

5.3.5 Templates

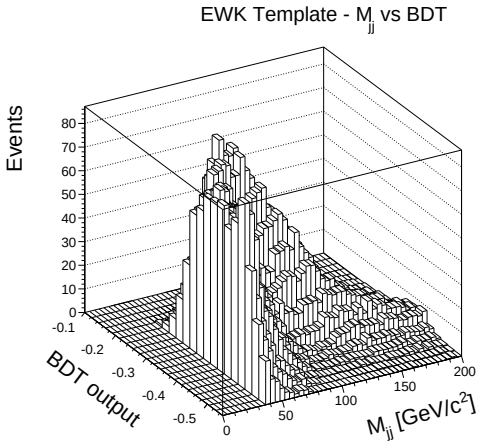
For each sample the discriminant output MVDO and M_{jj} are used to build a 2-dimensional distribution (template) of MVDO versus M_{jj}.

Fig.5.15 shows the templates of MVDO versus M_{jj} for MC *WW/WZ* signal (top left), QCD background (top right), MC EWK (bottom left) and MC Top (bottom right) using BDT method after applying th training weights. Fig.5.16 shows the same templates using Likelihood method. Fig.5.17 shows the templates for the data sample of *WW/WZ* candidates on left using BDT and on right using Likelihood method.

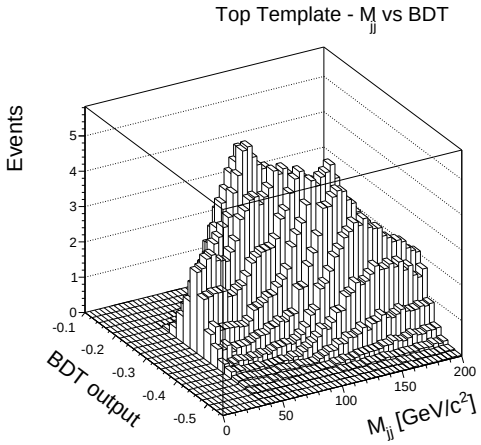


(a) Template WW WZ

(b) Template QCD bkg



(c) Template EWK

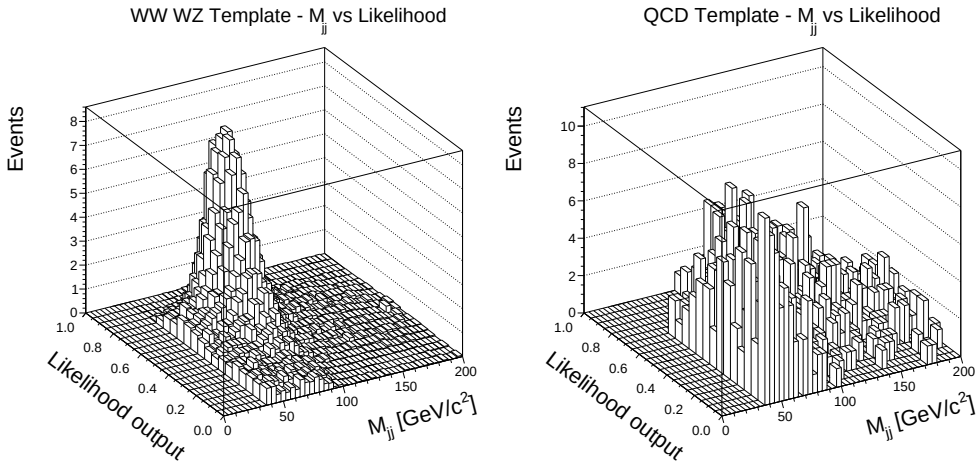


(d) Template Top

Figure 5.15: *MC and QCD templates for BDT.*

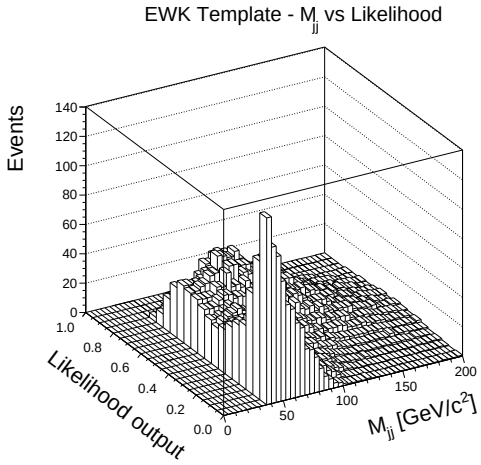
At first, we look at data-expected contributions agreement and we consider projections along x and y axis. In addition, we also check sliced projections on M_{jj} and sliced projections on the MVDs. Fig. 5.18(b) shows the projection of the BDT discriminant along M_{jj} , Fig. 5.18(a) shows M_{jj} projection along BDT discriminant. The same is shown for Likelihood, so Fig. 6.2(b) shows the projection of the Likelihood discriminant along M_{jj} , Fig. 6.2(a) shows M_{jj} projection along Likelihood discriminant.

Fig. 5.20 shows the sliced projections along BDT, while Fig.5.21 show the sliced pro-

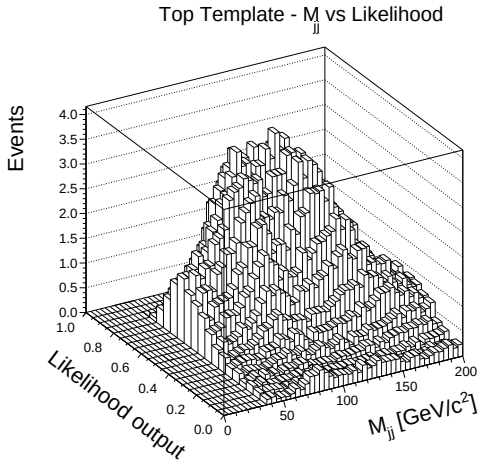


(a) Template WW WZ

(b) Template QCD bkg



(c) Template EWK

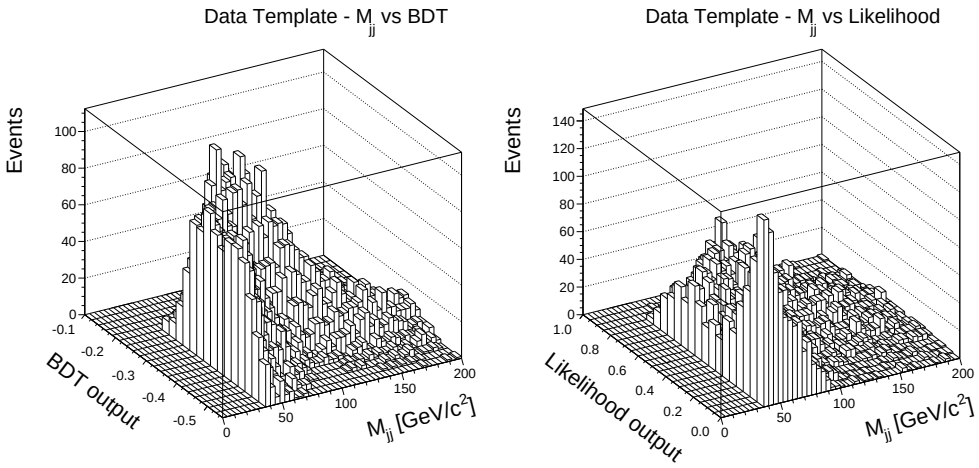


(d) Template Top

Figure 5.16: *MC and QCD templates for Likelihood.*

jections along M_{jj} for the BDT discriminant. Fig. 5.22 shows the sliced projections along Likelihood. Fig. 5.23 shows the sliced projections along M_{jj} for the Likelihood discriminant.

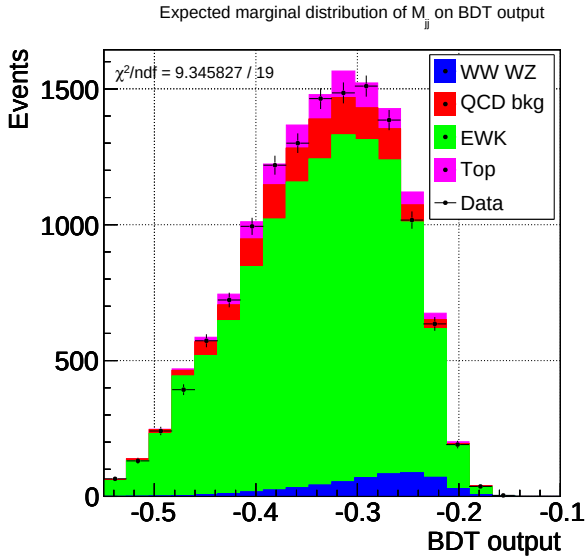
The agreement data-expected contribution in Fig.5.18-5.23 seems rather satisfactory both for BDT and Likelihood methods.



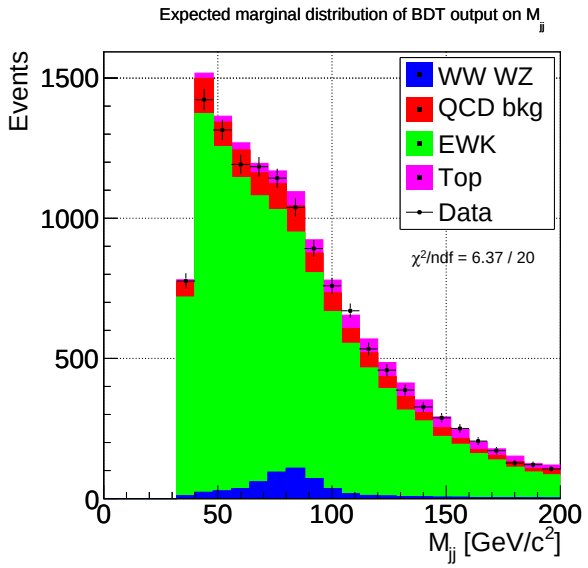
(a) Template Data

(b) Template Data

Figure 5.17: *Templates for Data sample for BDT (left) and Likelihood (right).*

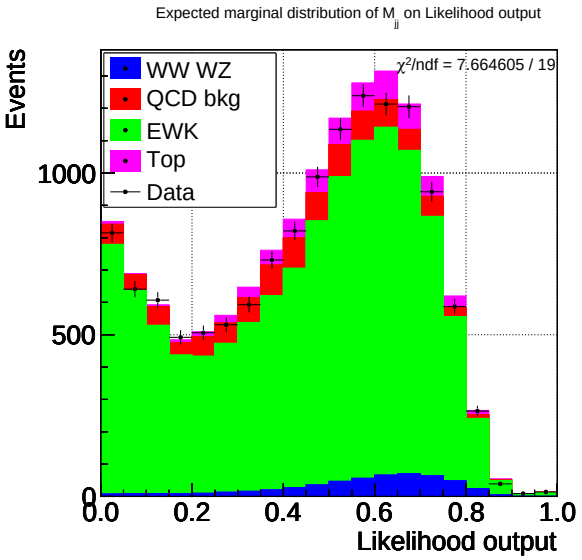


(a) Projection of M_{jj} along BDT output

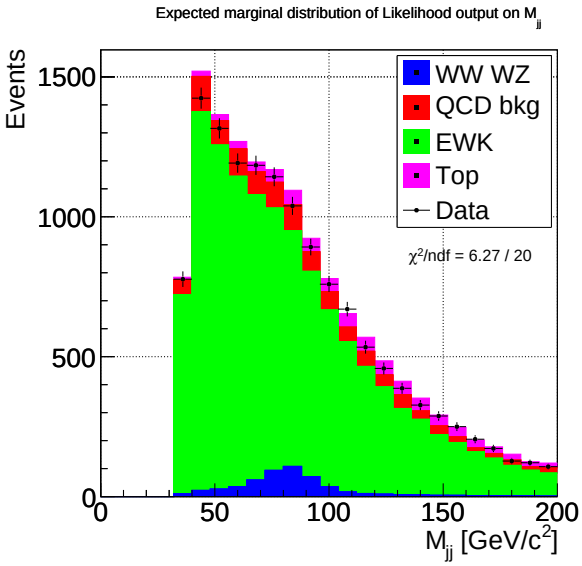


(b) Projection of BDT output along M_{jj}

Figure 5.18: *Projection of the 2-dimensional templates along BDT (top) and* M_{jj} *(bottom). MC and QCD expected stacked distributions with data overlayed.*



(a) Projection of M_{jj} along Likelihood output



(b) Projection of Likelihood output along M_{jj}

Figure 5.19: *Projection of the 2-dimensional templates along Likelihood (top) and* M_{jj} *(bottom).MC and QCD expected stacked distributions with data overlayed.*

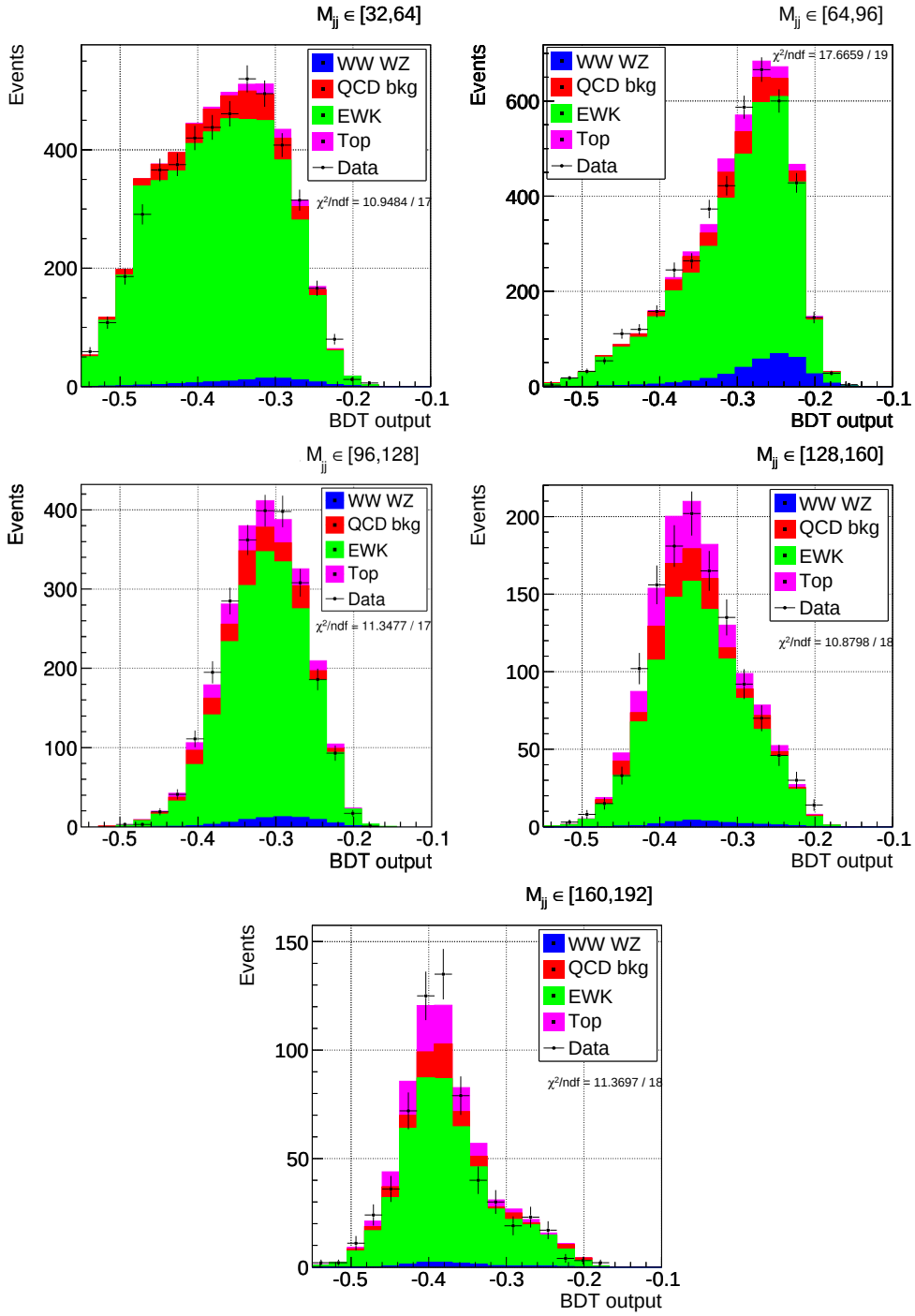


Figure 5.20: *Sliced projection of the 2-dimensional templates along BDT. MC and QCD expected stacked distributions with data overlayed.*

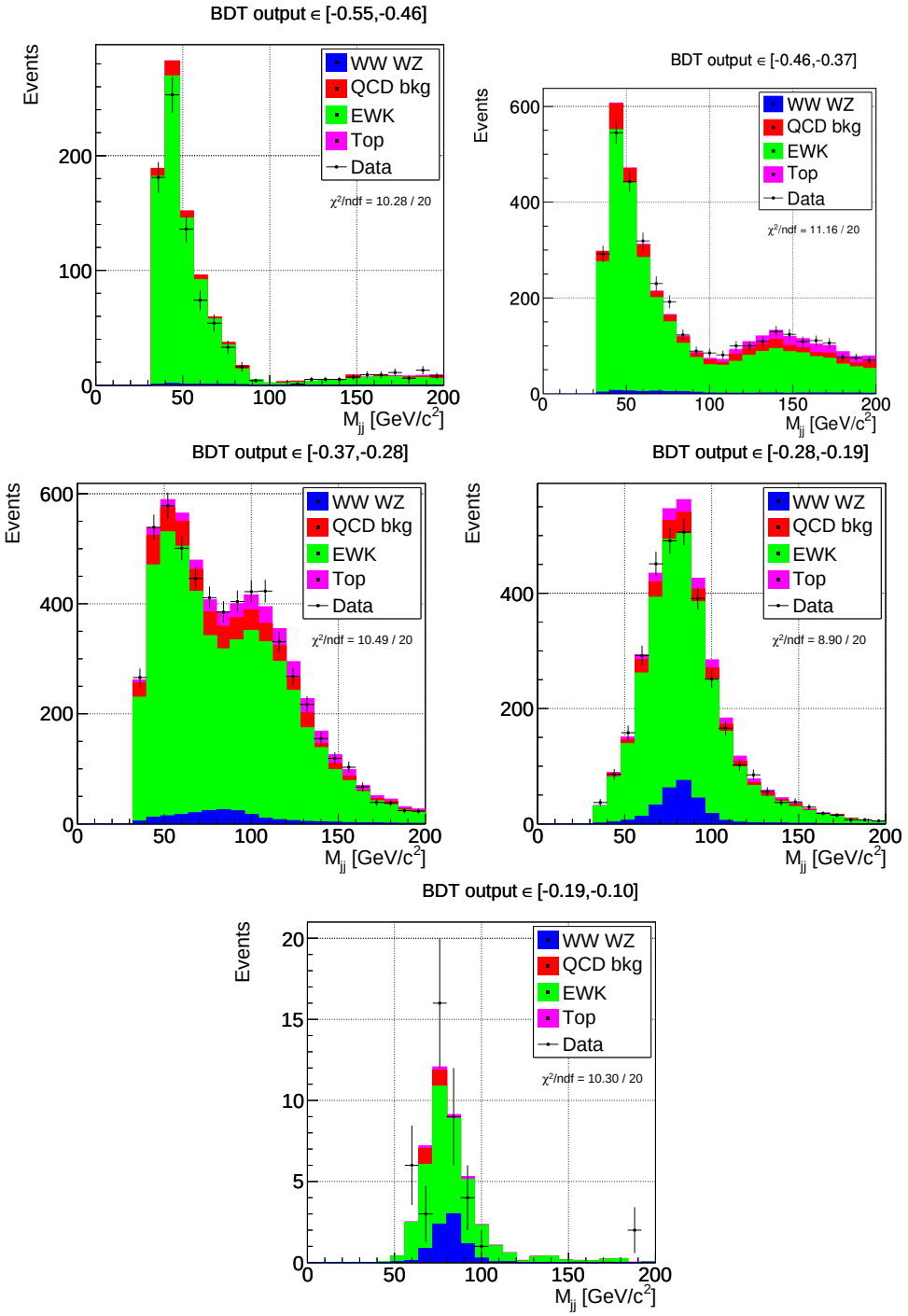


Figure 5.21: *Sliced projection of the 2-dimensional templates along* M_{jj} *for BDT. MC and QCD expected stacked distributions with data overlayed.*

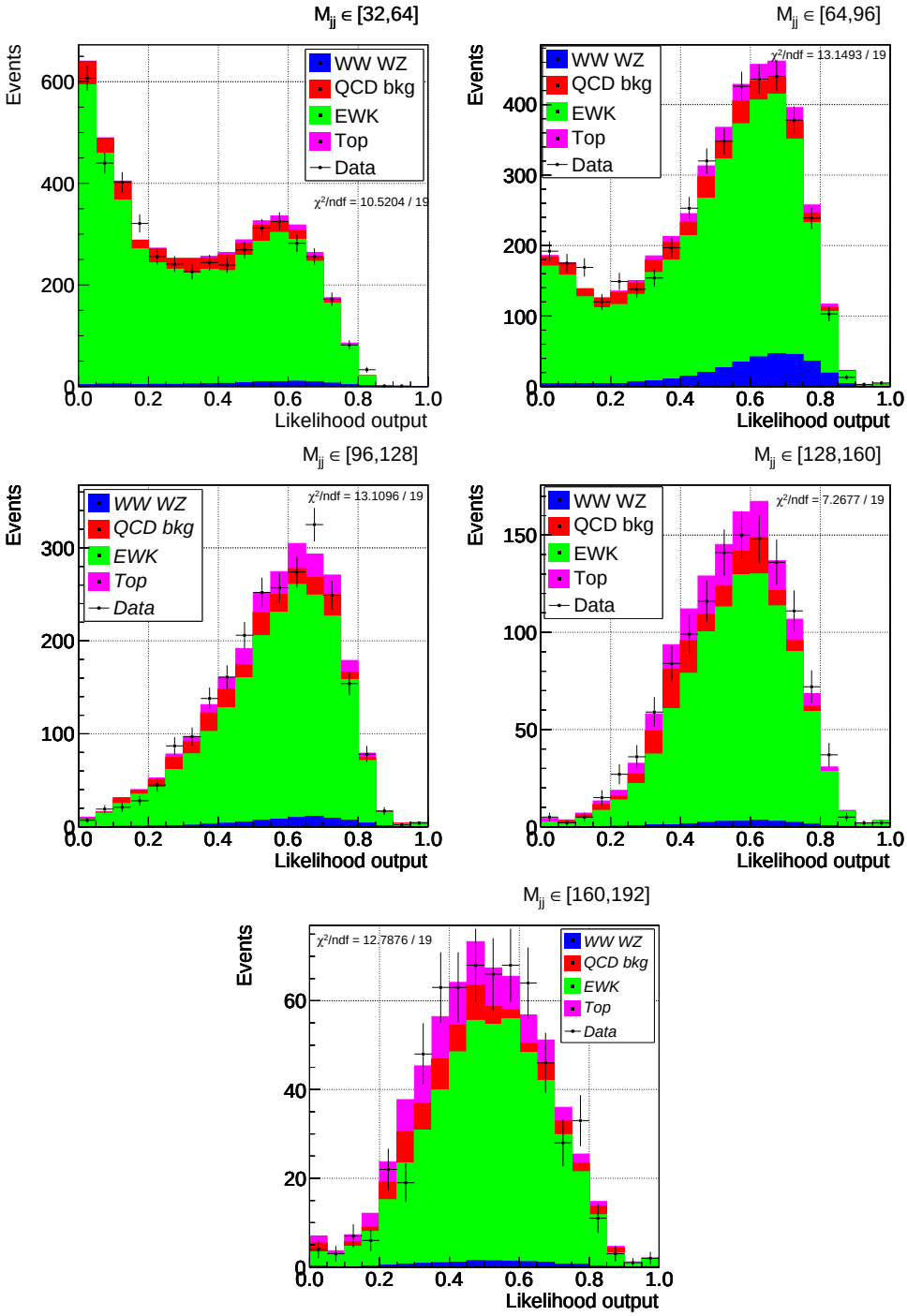


Figure 5.22: *Sliced projection of the 2-dimensional templates along Likelihood. MC and QCD expected stacked distributions with data overlayed.*

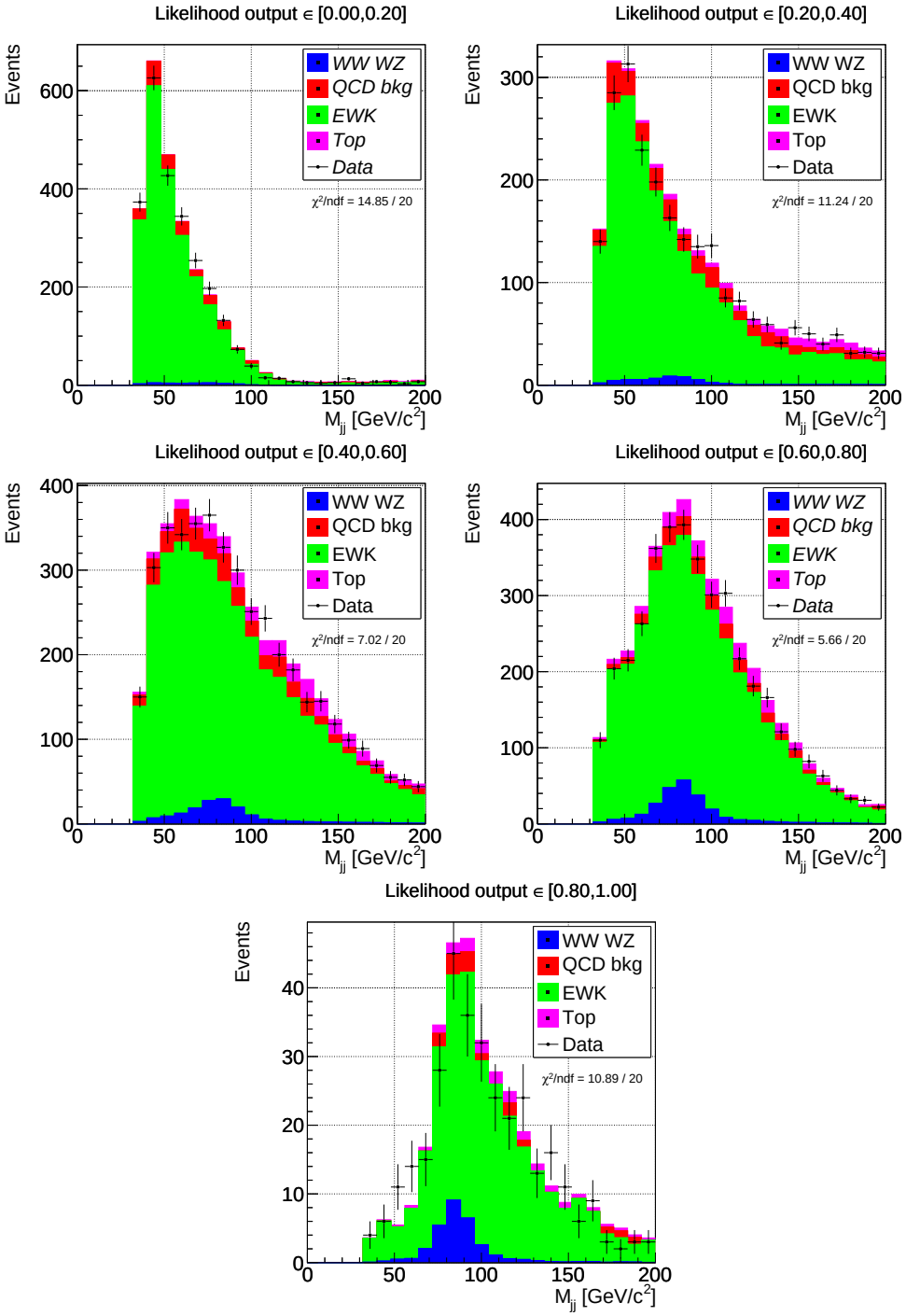


Figure 5.23: *Sliced projection of the 2-dimensional templates along* M_{jj} *for Likelihood. MC and QCD expected stacked distributions with data overlayed.*

5.4 2-Dimensional fitting procedure

The aim of this study is to increase signal/background separation, compared to the 1-dimensional fit to the dijet invariant mass. In the previous sections, we built a multivariate discriminant output. We then defineed 2-dimensional templates to estimate the probability density function in the $\mathrm{MVDO} - \mathrm{M_{jj}}$ space for each component. Our goal is perform a binned fit to data using the 2-dimensional templates. To this purpose, we use Fraction Fitter, available in ROOT (see Appendix. A). Fraction Fitter fits template histograms to data taking into account both data and Monte Carlo statistical uncertainties. The method is based on a binned maximum likelihood fit and the Monte Carlo templates are varied within statistics, leading to additional contribution to the overall likelihood. The algorithm is introduced and describes in [61].
The Fraction Fitter is used separately for BDT and Likelihood.

The input of the fitter are the 2-dimensional templates determined in Section 5.3.5. The output are the fraction of signal events ($\mathrm{f_{sig}}$) the fraction of Top and EWK ($\mathrm{f_{top}}$,$\mathrm{f_{ewk}}$) with the corresponding statistical uncertainties. The fraction of QCD background ($\mathrm{f_{qcd}}$) is fixed as estimated from data [29].
We first validate and asses the performance of our fit using pseudo-experiments and then we apply the 2-dimensional fit on data to extract the fraction of signal events.

5.4.1 Fitter validation and performances

The implemented 2-dimensional fit procedure of MVDO versus $\mathrm{M_{jj}}$ distributions, described above, will be tested and their performances will be evaluated using pseudo-experiments. The results obtained from the pseudo-experiments for the 2-dimensional fit will be then compared with the 1-dimensional fit results of the $\mathrm{M_{jj}}$ distribution using the same pseudo-experiment. Our goal is to determine the improvement of the statistical uncertainty on the extraction of the signal fraction.

5.4.2 Pseudo-experiments

We run indipendently 10000 pseudo-experiments. In each of them the signal fraction ($\mathrm{f_{sig}}$), the QCD fraction ($\mathrm{f_{qcd}}$) and the Top fraction ($\mathrm{f_{top}}$), from the expected composition data sample, are used ad input of the pseudo experiments generation (Table.5.2).

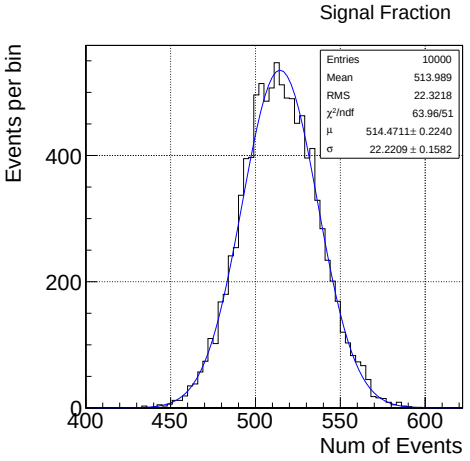
In each pseudo-experiment the number of signal ($\mathrm{N_{sig}^{pseudo}}$), QCD ($\mathrm{N_{qcd}^{pseudo}}$) and

Top (N_{top}^{pseudo}) events are extracted according to a binomial distribution. Each binomial distribution have as parameters the probability of a single success and the total number of trials. We consider as first parameter the fraction of events (f_{sig}, f_{qcd}, f_{top}) and the second the total number of events (N_{TOT}).
The fraction of EWK events (f_{ewk}) is calculated as $f_{ewk} = 1 - f_{sig} - f_{qcd} - f_{top}$.

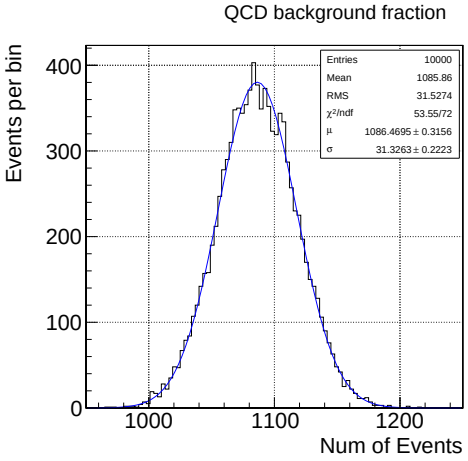
Sample	**Fraction**	**Expected Number of events**
MC Signal	$f_{sig} = 0.038$	$N_{sig} = 508$
QCD background	$f_{qcd} = 0.081$	$N_{qcd} = 1083$
MC Top	$f_{top} = 0.052$	$N_{top} = 695$
MC EWK	$f_{ewk} = 0.83$	$N_{ewk} = 11086$
Total Number of Events		$N_{TOT} = 13372$

Table 5.2: *Expected composition of data sample. First tree fractions and the number of total events, are the inputs of the pseudo-experiments.* $N_{ewk} = N_{TOT} - N_{sig} - N_{qcd} - N_{top}$

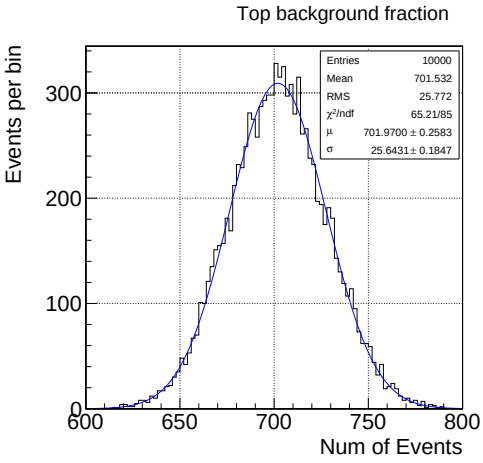
Fig.5.24 shows the distributions of the number of events for each sample for the 10000 pseudo-experiments. The mean value of each distribution correspond to the expected number of events of each sample of Table.5.2.



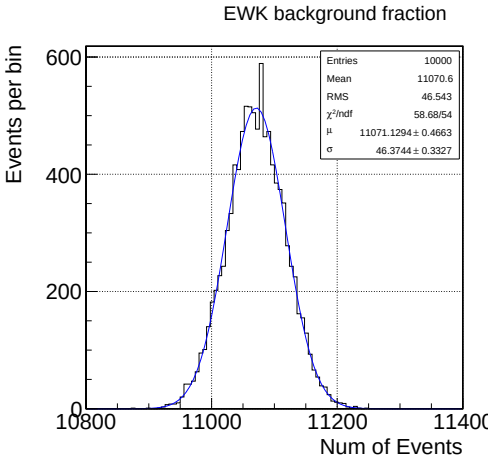
(a) Signal events distribution



(b) QCD events distribution



(c) Top events distribution



(d) Ewk events distribution

Figure 5.24: Pseudo experiments

5.4.3 Test and performance of 2-dimensional fit

In each pseudo-experiment the 2-dimensional fit is done on the templates samples of Table.5.2 respectively for BDT and Likelihood, and the fractions of samples are extracted. To evaluate and asses the performance of BDT and Likelihood on the extraction of signal fraction we use the residuals and pulls distributions.

The residual of the signal fraction is defined for each pseudo-experiment as the difference between the true input value (f_{sig}, Table.5.2) and the value retuned from the fit (f_{sig}^{fit}):

$$Residual = f_{sig} - f_{sig}^{fit}. \tag{5.5}$$

The Pull of the signal fraction is defined as the difference between true input value and fitted value over the uncertainty of the fitted value (returned by the fit):

$$Pull = \frac{f_{sig} - f_{sig}^{fit}}{\sigma_{f\ sig}^{fit}}. \tag{5.6}$$

The residuals distribution allows to determine the statistical uncertainty on measurement of the signal fraction on data. This distribution is expected to be parametrized by a Gaussian distribution with null mean value and σ corresponding to the estimated statistical uncertainty on the fraction extraction on data sample. In Fig.5.25 is reported the residual distributions of BDT (left) and Likelihood (right).

The Gaussian fit (superimposed on distributions, Fig.5.25 and Fig.5.26) gives a statistical uncertainty of $\sigma_{f\ sig}^{BDT} = 0.0087$ for BDT and $\sigma_{f\ sig}^{Likelihood} = 0.009$ for Likelihood methods and both have mean value compatible with zero.

If the used procedure is unbiased and the statistical uncertainty are properly evaluated by the fit, the pulls distributions are expected to be parametrized by a Gaussian distribution with null mean value and $\sigma = 1$. In Fig.5.26 is reported the pulls distribution for BDT (left) and Likelihood (right). The Gaussian fits (superimposed on distributions) has a σ compatibles with 1, and a small bias (mean value slightly different from zero).

We can conclude that the 2-dimensional fit procedure work properly for both methods. The performances of the two methods are resumed on Table.5.3.

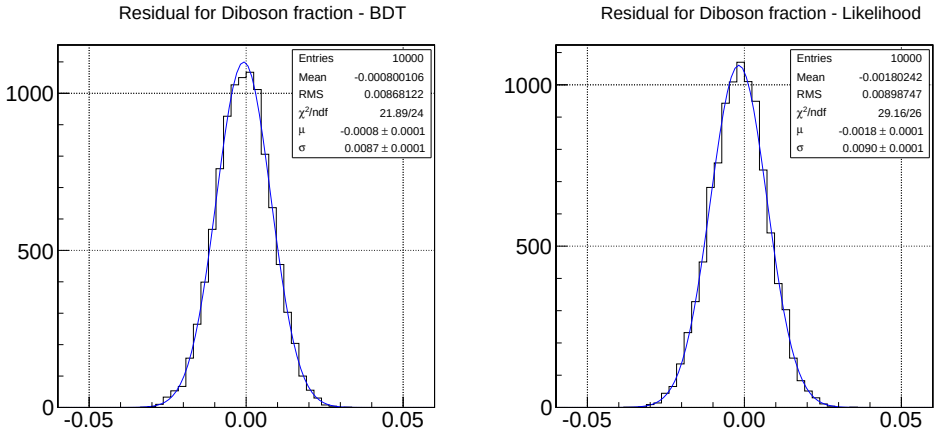


Figure 5.25: *Residuals distribution for BDT (left) and Likelihood (right)*

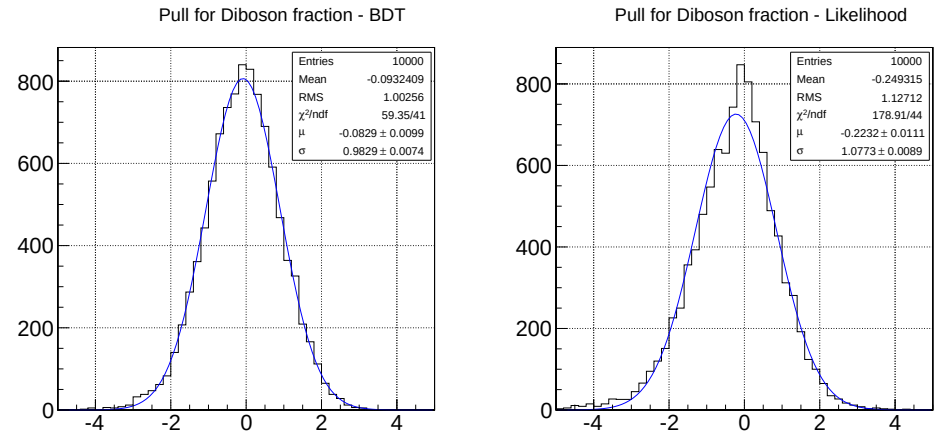


Figure 5.26: *Pulls distribution for BDT (left) and Likelihood (right)*

5.4.4 Test and performance of 1-dimensional fit

To test and evaluate the performances of 1-dimensional fit to the M_{jj} distribution we used the same pseudo-experiments procedure of the 2-dimensional fit. For each pseudo-experiments the M_{jj} distribution is fitted using Fraction Fitter.

The residuals (left) and pulls (right) distributions are shown on Fig.5.27.

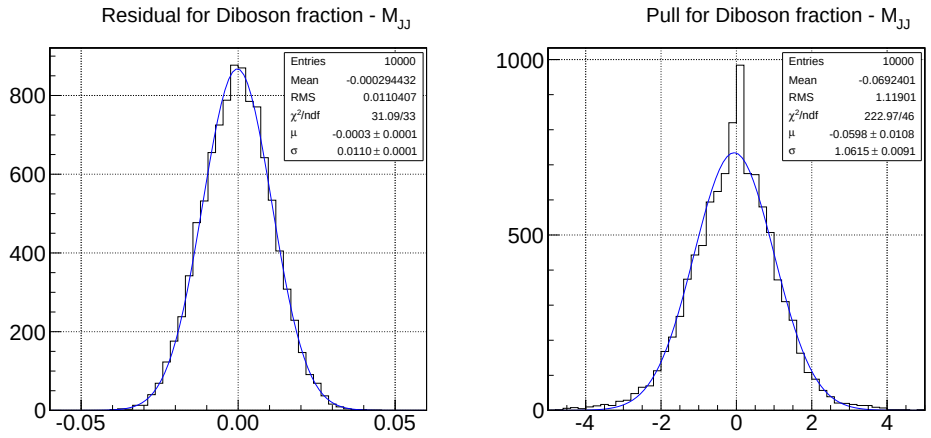


Figure 5.27: *Residuals and Pulls distribution of the Signal fraction*

The residual gaussian fit (superimposed on distribution) gives a statistical uncertainty of $\sigma^{M\ jj}_{f\ sig} = 0.011$ and a mean value compatible with zero. The pull Gaussian fit (superimposed on distribution) has a mean value and σ compatible respectively with zero and 1.

We can conclude that the 1-dimensional fit procedure on M_{jj} work properly. The performances are resumed on Table.5.3.

5.4.5 Results on fitter validation

The 2-dimensional fitting procedure leads to an improvement on signal fraction statistical uncertainty[2] of 21% for 2-dimensional fit using the BDT method and of 18% using the Likelihood method compared with the 1-dimensional fit on M_{jj} distribution.

Also, the improvement of BDT method is only 3.3% compared with the Likelihood

[2] To evaluate the improvement we used $(\sigma^{M\ jj}_{f\ sig} - \sigma^{method}_{f\ sig})/\sigma^{M\ jj}_{f\ sig}$ formula, where method means BDT or Likelihood (Table.5.3, for the values of the statistical uncertainties.)

	Method	**Statistical uncertainty** $\sigma_{f\ sig}$
2-dimensional fit	BDT	0.0087
	Likelihood	0.009
1-dimensional fit	M_{jj}	0.011

Table 5.3: *Statistical uncertainties on signal fraction using different methods*

method[3].

Since we used pseudo-experiments and in each of them the input value is the same for each method, the assed improvements[4] are also improvements of the relative statistical precision on the signal fraction extraction.

In Table.5.3 and Table.5.4 are resumed pur results.

Improvement on the extraction of signal fraction

i	j	$\frac{\sigma_j - \sigma_i}{\sigma_j}$
BDT	M_{jj}	21%
Likelihood	M_{jj}	18%
BDT	Likelihood	3.3%

Table 5.4: *Improvements on the extraction of signal fraction*

[3] In this case we used $(\sigma_{f\ sig}^{Likelihood} - \sigma_{f\ sig}^{BDT})/\sigma_{f\ sig}^{Likelihood}$ formula. see Table.5.3 for the values of the statistical uncertainties

[4] As example, in the BDT-Likelihood comparison we have $\frac{(\sigma_{f\ sig}^{Likelihood}/f_{sig}^{pseudo}) - (\sigma_{f\ sig}^{BDT}/f_{sig}^{pseudo})}{(\sigma_{f\ sig}^{Likelihood}/f_{sig}^{pseudo})}$.

Chapter 6

Extraction of signal fraction on Data

In this chapter the data sample of $WW/WZ \rightarrow \mu\nu + jets$ (see Chapter.3 and Section.5.3) candidates is used.
The results of the 2-dimensional fit on data templates for BDT and Likelihood methods is shown. We also show the results of 1-dimensional fit on data applied to the M_{jj} distribution.

6.1 2-dimensional fit results

The fit is done for BDT and Likelihood methods using the corresponding MVDO versus M_{jj} templates of MC signal, QCD background, MC Top and MC EWK.
In the fit the QCD fraction is fixed (see Section 5.3), while the others ($f_{sig}, f_{top}, f_{ewk}$) are left floating with the constraint that fraction's sum have to be equal to 1[1].

On Table.6.1 are reported the results of the 2-dimensional fit on data, on the extraction of fractions, for BDT and Likelihood. BDT results are compatible with Likelihood ones.

Also the statistical uncertainty on each fractions is compatible with the expected one (see Table.5.3).

[1] This constraint is fixed by the fitter (FractionFitter) itself.

BDT	
f_{sig}	0.0359 ± 0.009
f_{top}	0.0624 ± 0.0096
f_{ewk}	0.816 ± 0.015
Likelihood	
f_{sig}	0.0379 ± 0.008
f_{top}	0.0438 ± 0.0098
f_{ewk}	0.832 ± 0.015

Table 6.1: *2-dimensional fit on data: fractions results for BDT and Likelihood*

The projection of the fit results along MVDO and along M_{jj} are respectively shown for BDT on Fig.6.1 and for Likelihood on Fig.6.2 with overlayed the stacked contribute from each source (Signal, QCD, Top and EWK).

In Fig.6.3 and Fig.6.4 for BDT method and in Fig.6.5 and Fig.6.6 for Likelihood method the projections of the fit results are done in slices respectively of M_{jj} and MVDO.

In general there is a good agreement between the prediction of our fit and the experimental data in all distributions.

6.1.1 1-dimensional fit results

In this case only the 1-dimensional M_{jj} distribution is fitted using the corresponding templates of the samples described above and with the same prescriptions about the parameters of the fit.

In Table.6.2 are resported the results on the extraction of the fractions of the 1-dimensional fit for M_{jj}.

f_{sig}	0.046 ± 0.011
f_{top}	0.051 ± 0.011
f_{ewk}	0.822 ± 0.018

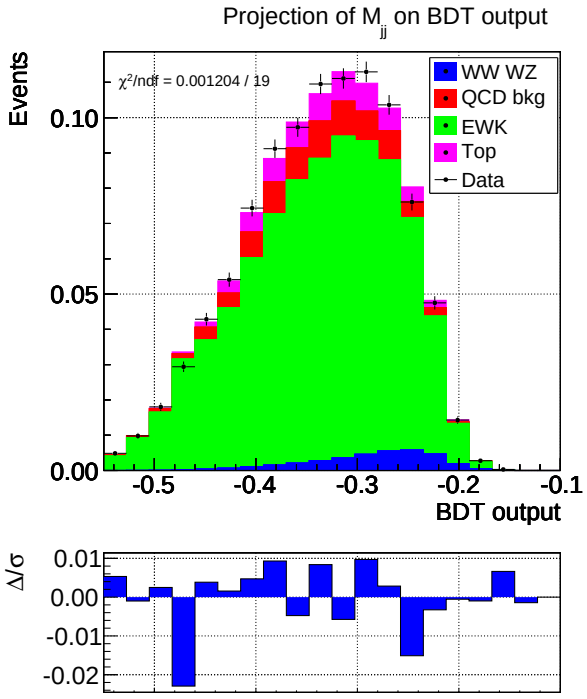
Table 6.2: *1-dimentional fit results on* M_{jj} *distribution*

These results are compatible both with BDT and Likelihood 2-dimensional fit (Table.6.1). The statistical uncertainty on the fraction of each fraction is compatible with the

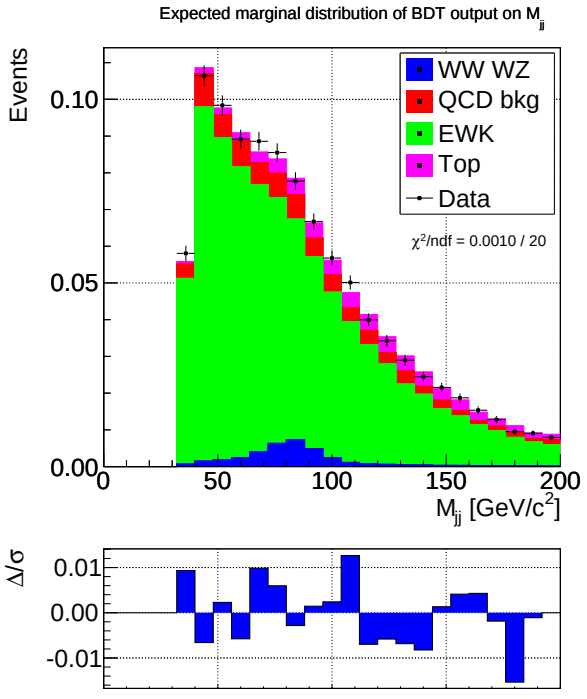
expected one(Table.5.3).

In Fig. 6.7 are shown the fit results on data for M_{jj}.

In general there is a good agreement between the prediction of our fit and the experimental data.

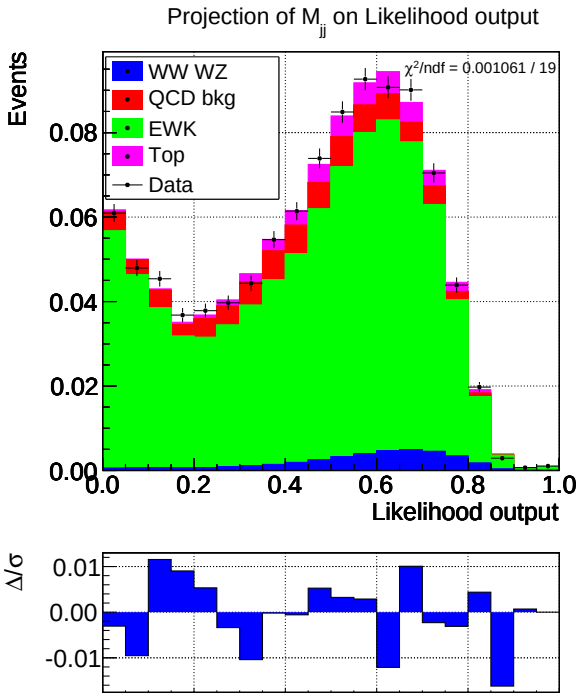


(a)

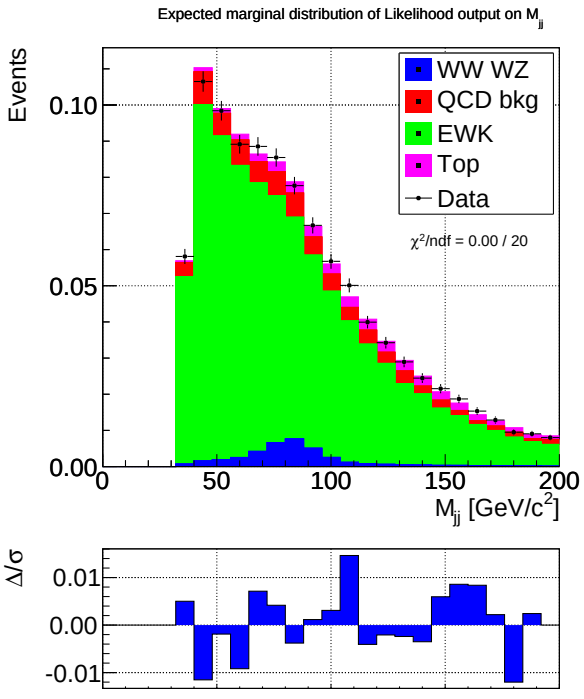


(b)

Figure 6.1: *2-dimensional fit on data: fit projection along MVDO (top) and along* M_{jj} *(bottom) for BDT method. MC and QCD extracted stacked distributions with data overlayed.*



(a)



(b)

Figure 6.2: *2-dimensional fit on data: fit projection along MVDO (top) and along* M_{jj} *(bottom) for Likelihood method. MC and QCD extracted stacked distributions with data overlayed.*

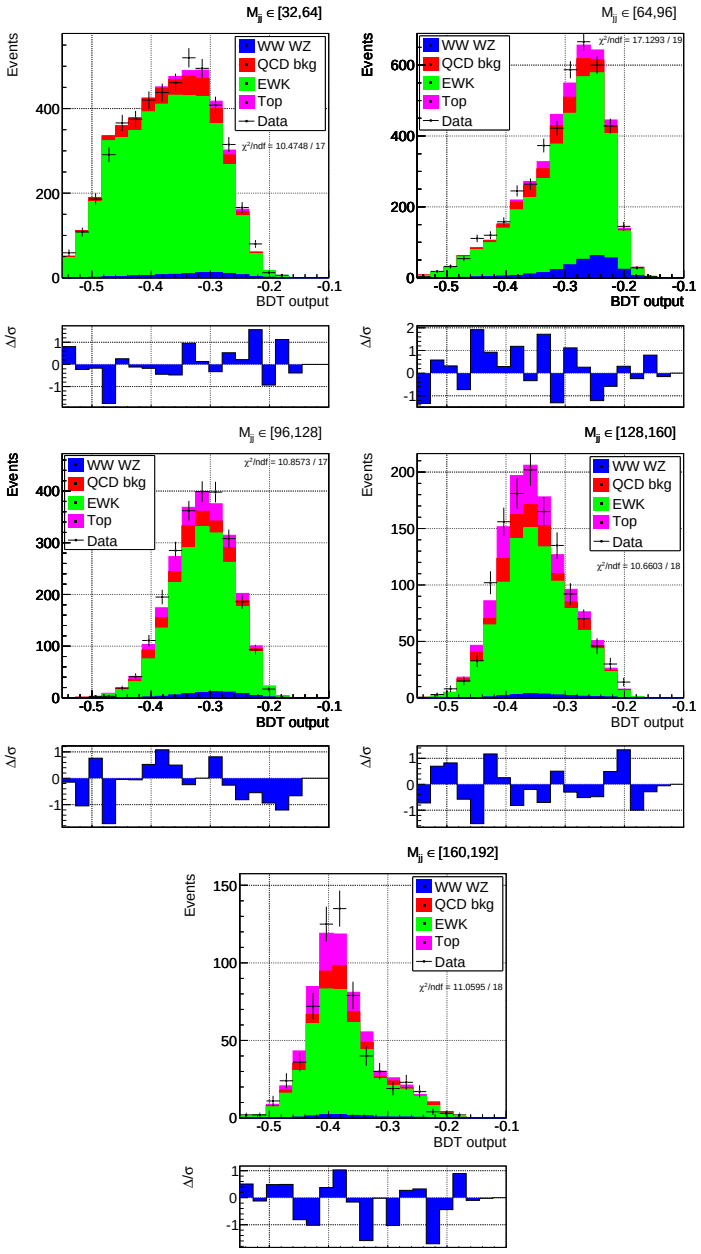


Figure 6.3: *2-dimensional fit on data: Sliced* M_{jj} *projections along MVDO for BDT. MC and QCD extracted stacked distributions with data overlayed.*

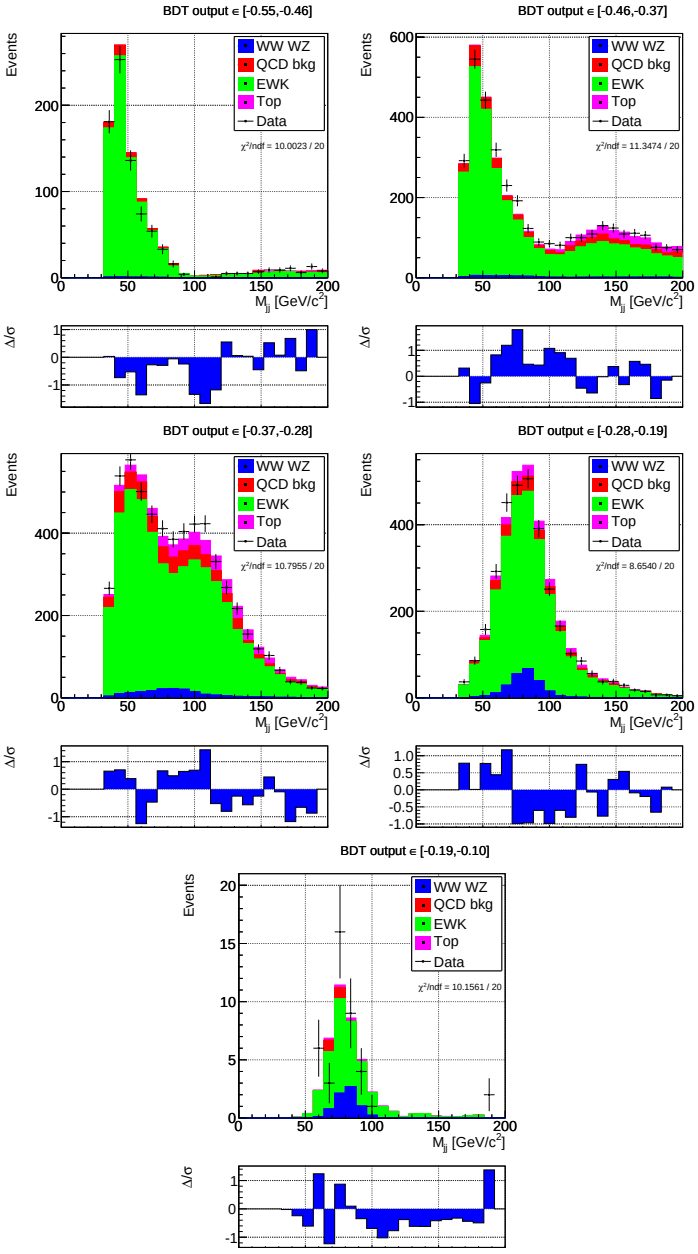


Figure 6.4: *2-dimensional fit on data: Sliced MVDO projections along* M_{jj} *for BDT. MC and QCD extracted stacked distributions with data overlayed.*

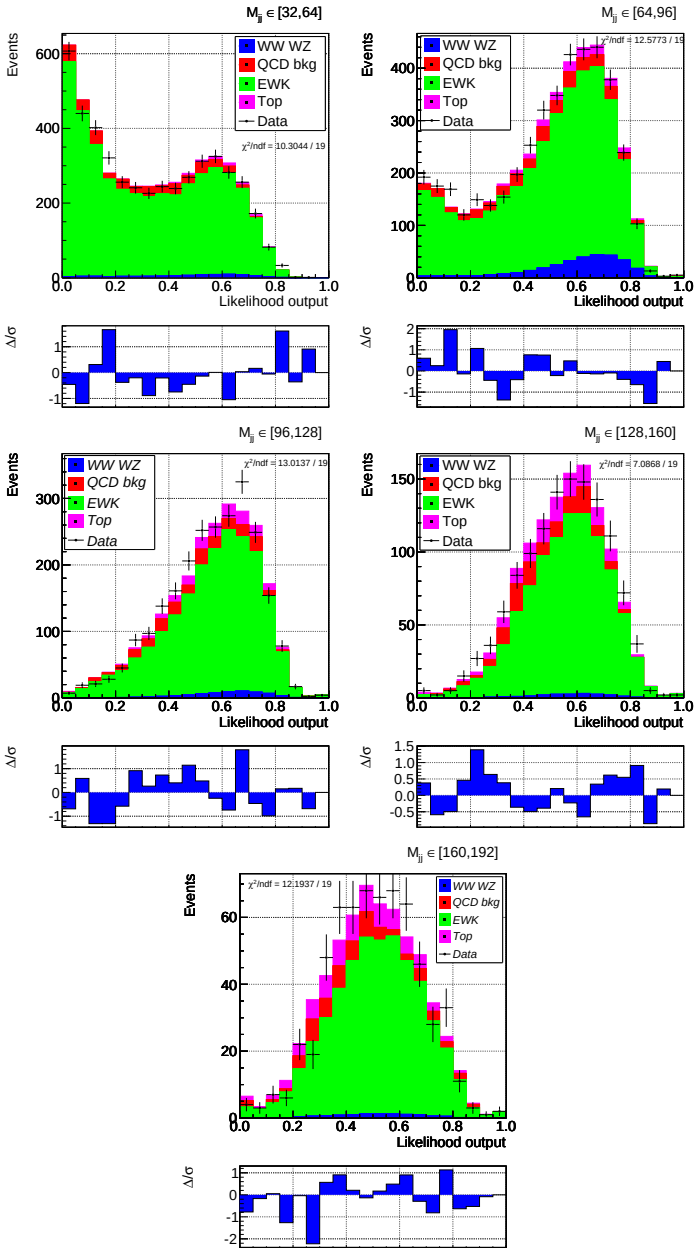


Figure 6.5: *2-dimensional fit on data: Sliced* M_{jj} *projections along MVDO for Likelihood. MC and QCD extracted stacked distributions with data overlayed.*

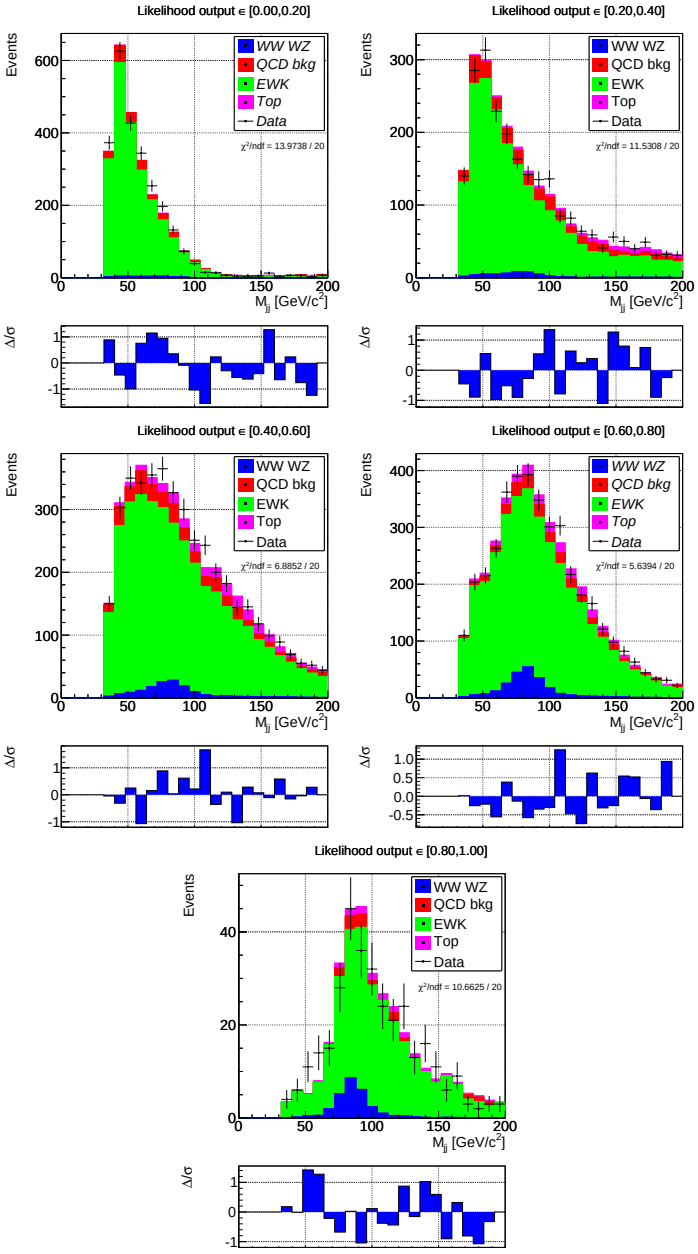


Figure 6.6: *2-dimensional fit on data: Sliced MVDO projections along* M_{jj} *for Likelihood. MC and QCD extracted stacked distributions with data overlayed.*

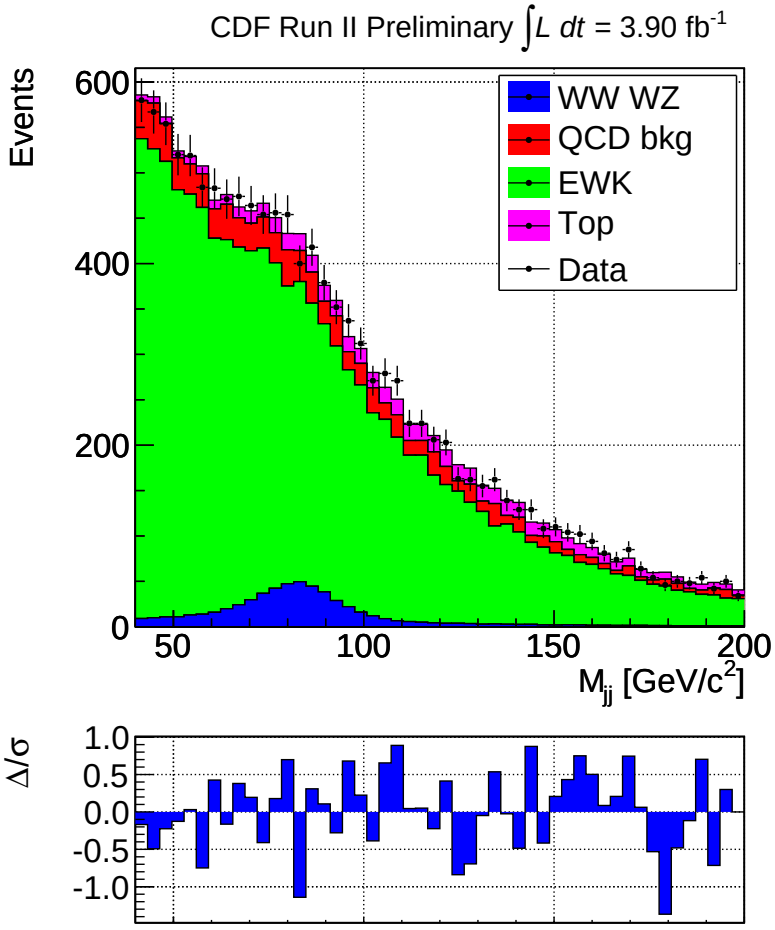


Figure 6.7: *1-dimensional fit of* M_{jj} *distribution*

Appendix A

The Fitter

The Fitter we use to perform all our fit procedure was the TFractionFitter available in ROOT package.

The virtue of this fit is that it takes into account both data and Monte Carlo statistical uncertainties. The way in which this is done is through a standard Likelihood fit using Poisson statistics.

The algoritm on which is based this fitter is explained below. When you have experimental data and you wish to determine the proportion X_j of the different sources in the data; if d_i is the number of events in the real data fall into bin i and f_i is the predicted number of events in the bin, given by the strenghts X_j, and a_{ji} is the number of Monte Carlo events from source j in the bin i, we have:

$$f_i = N_D \sum_{j=1}^{m} \frac{X_j a_{ij}}{N_j} \tag{A.1}$$

where N_D is the total number in the data sample and N_j is the total number in each MC source,

$$N_D = \sum_{i=1}^{n} d_i \qquad N_j = \sum_{i=1}^{n} a_{ji}. \tag{A.2}$$

We could rewrite the A.1 as

$$f_i = \sum_{j=1}^{m} x_j a_{ij} \tag{A.3}$$

where the normalisation factors are incorporate in the strenght factors $x_j = N_D X_j / N_j$. Using the Poisson distribution for the d_i one can write the probability for observing a particular d_i as

$$e^{-f_i} \frac{f_i^{d_i}}{d_i!} \tag{A.4}$$

and the estimates of the proportions x_j are found maximising, as usual, the logarithm of the likelihood:

$$\ln \mathcal{L} = \sum_{i=1}^{n} (d_i \ln f_i) - f_i \tag{A.5}$$

This accounts correctly for the small numbers of data events in the bins, but this does not account for the fact that the Monte Carlo samples used may also be finite size, leading to statistical fluctuation int the a_{ji}. The correct methodology to incorporate the MC statistics consist to consider for each source and in each bin some (unknown) expected number of events A_{ji} so that the eq. A.3 becomes

$$f_i = \sum_{j=1}^{m} x_j A_{ij} \tag{A.6}$$

so the total likelihood is now:

$$\ln \mathcal{L} = \sum_{i=1}^{n} (d_i \ln f_i) - f_i + \sum_{i=1}^{n} \sum_{j=1}^{m} (a_{ji} \ln A_{ji}) - A_{ji}. \tag{A.7}$$

Now it became a maximisation problem in $m \times (n+1)$ unknowns.

The solution of this problem is to find the m variables x_j by solving the m equations iteratively.

So this ROOT package is implemented by maximise $\ln \mathcal{L}$ itself with respect to the x_j, using a package such as MINUIT [?]. Some special care needs to be taken in the case of bins with zero content.

Conclusions and Perspective

In this thesis I described and tested a multivariate approach, to improve the statistical sensitivity in the extraction of signal events in the analysis of $p\bar{p} \rightarrow WW/WZ$ in lepton neutrino plus jets, of [1, 2, 29].

I tested two different discriminant methods in this approach: the Boosted Decision Tree and the Projective Likelihood Estimator.

To test the improvement of both methods, I used only the muon channel for the W decaying leptonically (in reference [1, 2, 29] also the electron channel is used) and the analysis is on 3.9 fb^{-1} of data collected by CDF II experiment.

My results show an improvement of 21% for BDT and of 18% when compared with standard method of [1, 2, 29].

A multivariate analysis based on BDT, if applied in the ongoing analysis (on 4.9 fb^{-1}), would improve the signal extraction sensitivity, and as consequence of that, the statistical precision in the measurement of the diboson production cross section.

In any case, the systematics due to the multivariate approach and the overall gain in the measurements of the above mentioned cross section has to be evaluted before being applied.

Increase the sensitivity to diboson events is crucial to the Higgs search program at the Tevatron because the diboson events are an irreducible background to the Higgs discovery principal signature (lepton neutrino + $b\bar{b}$ jets) in associated production with a W.

But also, same technique (BDT) can be used in the Higgs search adding separating variables to tag b-jets from Higgs decay.

Bibliography

[1] The CDF Collaboration, *Measurement of the $WW + WZ$ Production Cross Section Using the Lepton+Jets Final State at CDF II*, (2009), arXiv:0911.4449v1

[2] Diboson Analysis Website, http://www-cdf.fnal.gov/physics/ewk/2009/jjlnu/index.html

[3] F. J. Hasert *et al.* (1974). "Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment". *Nuclear Physics B* **73**: 1.

[4] LHC Design Report

[5] *Review of Particle Physics*, W.-M.Yao *et al*, Journal of Physics G 33. 1 (2006).

[6] F.Halzen, A.Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, John Wiley & Sons, 1984.

[7] Peter Renton, *Electroweak Interactions, An Introduction to the Physics of Quarks and Leptons*,Cambridge University Press.

[8] *Modern elementary particle physics*, Gordon Kane, (1993)Addison-Wesley.

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

[10] *Review of Particle Physics*, W.-M.Yao et al, Journal of Physics G 33. 1 (2006).

[11] Hagiwara *et al.*, *Low energy effects of new interactions in the electroweak boson sector, K*. Physical Review D, Vol.48, Num.5, 1 September 1993.

[12] John Ellison and Jose Wudka,*Study of Trilinear Gauge-Boson Couplings at the Tevatron Collider*, Annu. Rev. Nucl. Part. Sci. 1998. 48:33-80.

[13] The LEP Collaborations, *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, arXiv:hep-ex/0612034v2,2006.

[14] M. Neubauer, for the CDF and DØ Collaborations, arXiv:hep-ex/0605066v2,2006.

[15] CDF Collaboration, *Measurements of inclusive W and Z cross sections in* $\mathrm{p\overline{p}}$ *collisions at* $\sqrt{\mathrm{s}} = 1.96\,\mathrm{TeV}$, J. Phys. G: Nucl. Part. Phys. 34(2007) 2457-2544

[16] F. Abe et al. (CDF Collaboration), *"Evidence for* W^+W^- *Production in* $\mathrm{p\bar{p}}$ *Collisions at* $\sqrt{\mathrm{s}}$*= 1.8 TeV"*, Phys. Rev. Lett. **78**, 4536 (1997).

[17] B. Abbott *et al.* (D0 Collaboration), *"Studies of WW and WZ production and limits on anomalous* $WW\gamma$ *and WWZ couplings"*, Phys. Rev. **D 60**, 072002 (1999).

[18] K. Hagiwara, J. Woodside, and D. Zeppenfeld, *"Measuring the WWZ coupling at the Fermilab Tevatron"*, Phys. Rev. **D 41**, 2113 (1990).

[19] A.Djounadi, J.Kalinowsky and M.Spira, *Comput. Phys. Comm.* **108** (1998) 56.

[20] CDF Collaboration, *Measurement of the* W^+W^- *Production Cross Section and Search for Anomalous* $WW\gamma$ *and WWZ Couplings in* $\mathrm{p\bar{p}}$ *Collisions at* $\sqrt{\mathrm{s}} = 1.96$ *TeV*, arXiv:0912.4500v1

[21] DØ Collaboration,*Measurement of the WW Production Corss Section with Dilepton Final State in* $\mathrm{p\bar{p}}$ *Collisions at* $\sqrt{\mathrm{s}} = 1.96$ *TeV and Limits on Anomalous Trilinear Gauge Couplings*, Phys. Rev. Lett. 103, (2009) 191801.

[22] CDF Collaboration, *Observation of WZ Production*, Phys. Rev. Lett. 98, 161801 (2007).

[23] DØ Collaboration, *Measurement of the* $\mathrm{p\bar{p}} \rightarrow$ WZ+X *cross section at* $\sqrt{\mathrm{s}} = 1.96$ *TeV and limits on WWZ trilinear gauge couplings*, Phys. Rev. D 76, 111104(R) (2007).

[24] CDF Collaboration, *Strong Evidence for ZZ Production in* $\mathrm{p\bar{p}}$ *Collisions at* $\sqrt{\mathrm{s}} = 1.96$ *TeV*, Phys. Rev. Lett. 100, 201801 (2008).

[25] DØ Collaboration, *Search for ZZ and* $Z\gamma^*$ *production in p-barp collisions at* $\sqrt{(\mathrm{s})} = 1.96$ *TeV and limits on anomalous ZZZ and* $ZZ\gamma^*$ *couplings*, arXiv:0712.0599v1 (2007).

[26] CDF Collaboration, *First Observation of Vector Boson Pairs in a Hadronic Final State at the Tevatron Collider*, Phys. Rev. Lett. 103, 091803 (2009).

[27] B.Alvarez, F.Canelli, R.Eusebi, C.Group, M.Hurwitz, B.Casal Larana, E.Palencia, B.Stelzer *Observation of WW+WZ Production in lvjj Channel in L=2.7 fb^{-1} of CDF Run II Data Using a Matrix Element Technique*

[28] M. Carena, J. S. Conway, H. E. Haber, J. D. Hobbs, *et al.*, *Report of the Higgs Working Group of the Tevatron Run 2 SUSY/Higgs Workshop*, (2000),arXiv:hep-ph/0010338v2

[29] A.Annovi, P.Catastini, V.Cavaliere, M.A.Ciocci, J.Duchini, P.Mastrandrea, M.Rescigno, A.Sfyrla, Search for Diboson production in $WW/WZ \rightarrow l\nu + Jets$ usign 3.9 fb^{-1}, CDF Notes 9807.

[30] G. Flanagan *et al.* (CDF Collaboration), *"Observation of Diboson Production in* $\not\!\!E_T$ + jets *channel"*, May 28th 2009, to be pubblished on arXiv:0905.4714.

[31] Alexander Shemyakin et al. [Electron Cooling Collaboratio], *Electron cooling in the recycler ring*, FERMILAB-CONF-06-147-AD (2006)

[32] G. Aubrecht et al.*A teachers Guide to the Nuclear Science Wall Chart*, Contemporary Physics Education Project, (2003)

[33] C. W. Schmidt, *The Fermilab 400-MeV Linac upgrade*, FERMILAB-CONF- 93-111(1993).

[34] *Run II Handbook*, Fermilab Beams Division.

[35] The CDF Collaboration:*The CDF II Detector Technical Design Report* FERMILAB-PUB-96/390-E, October, 1996 The CDF Collabration: *Summary of the CDF RunII Detector Parameters*

[36] The CDF Collaboration:*Proposal for Ehnancement of the CDF II Detector: an Inner Silicon Layer and a Time of Flight Detector* FERMILAB-PROPOSAL-909 (and updates), October 23, 1998

[37] R. Blair *et al.* [CDF Collaboration], *The CDFII Detector: Technical Design Report*, FERMILAB-Pub-96/390-E (1996).

[38] C. S. Hill, *Initial experience with the CDF layer ØØ silicon detector* , Nucl. Instrum. Methods **A511,118,2003**, ()

[39] A. Sill:*CDF Run II silicon tracking projects* Nucl. Instrum. Methods **A447, 1, 2000**, (). C. Kruse:*Silicon Detector Upgrades for the Tevatron Run 2* FERMILAB-CONF-02/253-E. Published Proceedings 31st International Conference on High Energy Physics (ICHEP 2002), Amsterdam, The Netherlands, July 24-31, 2002

[40] A. Affolder *et al.*, *Status report of the intermediate silicon layers detector at CDF II*, Nucl. Instrum. Methods **A485,6,2002**, ()

[41] A. Affolder *et al.*, *CDF Central Outer Tracker*,Nucl. Instrum. Methods **A526, 249, 2004**, ()

[42] C.S.Hill et al., *Operational experience and performance of the CDF II silicon detector*, Nucl. Instrum. Methods, A530, 1-6 (2004)

[43] C. Amsler *et al.*, *Particle data group*, Phys. Lett. **B 667**, 1 (2008)

[44] R. D. Erbacher, *Calorimetry in CDF Run 2*, FERMILAB-CONF-02/251-E. Published Proceedings 31st International Conference on High Energy Physics (ICHEP 2002), Amsterdam, The Netherlands, July 24-31, 2002; S. Kuhlmann *et al.*,*The CDF calorimeter upgrade for Run IIb*, Nucl. Instrum. Methods **A518,39,2004**, () L. Breccia *et al.*,*Test of 2000 phototubes for the CDF endplug calorimeter upgrade*, Nucl. Instrum. Methods **A532,575,2004**, ()

[45] L. Balka *et al.*, *The CDF central electromagnetic calorimeter*, Nucl. Instrum. Methods **A267,272,1988**, ()

S. R. Hahn *et al.*, *Calibration systems for the CDF central electromagnetic calorimeter*, Nucl. Instrum. Methods **A267, 351,1988**, ()

[46] M. Gallinaro, *A New Scintillator Tile/Fiber Preshower Detector for the CDF Central Calorimeter*, IEEE Trans. Nucl. Sci. **52,879,2005**, , ([]physics/0411056].

[47] S. Bertolucci *et al.*, *The CDF central and endwall hadron calorimeter*, Nucl. Instrum. Methods **A267, 301, 1988**, ()

[48] M. Albrow *et al.*,*The CDF plug upgrade electromagnetic calorimeter: test beam results*,Nucl. Instrum. Methods **A480, 524,2002**, () M. Albrow *et al.*,*A preshower*

detector for the CDF Plug Upgrade: test beam results,Nucl. Instrum. Methods **A431, 104,1999**, () G. Apollinari *et al.*,*Shower maximum detector for the CDF plug upgrade calorimeter*,Nucl. Instrum. Methods **A412, 515,1998**, () M. Albrow *et al.*,*Intercalibration of the longitudinal segments of a calorimeter system*,Nucl. Instrum. Methods **A487, 381, 2002**, ()

[49] P. Azzi *et al.*, *Histogram tracking in the COT*, CDF note 5562, February 21 (2001).

[50] D.Amidei ,et al. *Measurements of sigma*$*$B($W \to \mu\nu$), *sigma*$*$B($Z \to \mu\mu$) *and the ratio R using CDF Run II Data*, CDF Note 6711.

[51] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D 71, 032001 (2005)

[52] E.Lipeles, M.Neubauer, S.C.Hsu, F.Wurthwein, *Lepton ID for Multilepton Diboson Analyses,* CDF Note 8538, University of California, San Diego

[53] The CDF Collaboration, *Topology of three-jet events in* $\mathrm{p\bar{p}}$ *collisions at* $\sqrt{s} = 1.8$ *TeV* , Phys. Rev. D $45, 1448, 1992$

[54] JetClu Algoritm

[55] Anna Sfyrla, *Search for WW and WZ production in electron-neutrino plus dijet final state using 1.29 fb*^{1} *of data*, Fermilab Thesis (2008).

[56] T. Sjostrand *et al.*, *High-Energy-Physics Event Generation with PYTHIA 6.1*, Comp. Phys. Commun. 135, 238 (2001); Pythia home page: http://www.thep.lu.se/ torbjorn/Pythia.html.

[57] M.L. Mangano *et al.*, *ALPGEN, a generator for hard multiparton processes in hadronic collisions* JHEP 0307:001, (2003).

[58] A.Hocker *et al.*, Toolkit for Multivariate Data Analysis with ROOT, arXiv physics/0703039.

[59] ROOT: *An Object Oriented Analysis Framework*,http://root.cern.ch

[60] Roe, B. P.; Yang, H.-J.; Zhu, J., *Boosted Decision Trees, A Powerful Event Classifier*, STATISTICAL PROBLEMS IN PARTICLE PHYSICS, ASTROPHYSICS AND COSMOLOGY (pp 139-142)

[61] R. Barlow and C. Beeston, Comp. Phys. Comm. 77 (1993) 219-228