Università degli Studi di Pisa

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI

Corso di Laurea Magistrale in Fisica

TESI DI LAUREA MAGISTRALE

A Novel Technique to Reconstruct the Mass of the Associated Z boson Decaying into Jets in W + Z and Z + Z Events with Lepton(s), Missing Transverse Energy and Three Jets.

Candidato:
Caterina Vernieri

Relatore:
Prof. Giorgio Bellettini

Correlatore:
Dr. Marco Trovato

Anno Accademico 2010-2011

A me,
a mamma e papà
e anche a te Fede.

Contents

Introduction

The study of associated WZ boson production with a lepton and a neutrino signaling the W, and a $b\bar{b}$-pair in the final state is important since the event topology of this process is the same as expected for WH associated production of a W and the Standard Model light-Higgs boson ($M_H \lesssim 135$ GeV). Thus, the investigation of the process $W^{\pm}Z \to \ell\nu b\bar{b}$ whose rate can be accurately predicted, allows to calibrate and optimize many of techniques used in the SM Higgs search and provides a "standard candle" for that crucially important search. In addition, WZ associated production generates a significant background for low mass Higgs Boson searches with H decaying into a $b\bar{b}$ pair.
At the Tevatron, the process $WH \to Wb\bar{b}$ has an expected $\sigma \cdot BR$ [1] about five times lower than $WZ \to Wb\bar{b}$ for $m_H \simeq 120$ GeV/c^2. Therefore, observing that process would be a benchmark for the even more difficult search for the light Higgs in the $WH \to Wb\bar{b}$ process.

Observing associated WZ production at the Tevatron in the channel $WZ \to \ell\nu b\bar{b}$ is extremely difficult for two main reasons.
The event rate is extremely low. A WZ production cross section of $\sim 3,22$ pb [1] together with a $Z \to b\bar{b}$ branching ratio of $\sim 15\%$ [2] provides about 50 fb in the $WZ \to \ell\nu b\bar{b}$ channel. With a trigger and kinematical selection efficiency of the order of a few %, one expects a handful of events per fb^{-1} of integrated luminosity.
This statement remains valid even if the few ZZ events with leptonic decay of one Z are included in the acceptance.
A standard kinematical cut requests exactly two high energy jets (i.e. $E_T > 20$ GeV) in the candidate sample. Simulations show that if a third energetic jet is allowed the signal acceptance is increased by about 1/3. Therefore, it would be very important to be able to extract a $Z \to b\bar{b}$ signal also in events with more than two high energy jets.
A second difficulty is that the signal to background ratio is very poor, due primarily to the contribution of associated production of W and incoherent jets. Optimal dijet mass resolution is of utmost importance for discriminating this background, since a fit to the invariant mass distribution of the two jets, associated to the hadronic decay of Z, is used to disentangle the diboson signal from the backgrounds in the candidate data sample.

[1] BR being the Branching ratios of $Z/H \to b\bar{b}$

In this thesis, we present a search for WZ/ZZ in events with a lepton(s), missing transverse energy and jets. Besides looking at the sample where two exclusive jets are found, we investigate the sample with 3 jets where about the 33% of the signal events lie.
In WZ events, additional jets may be initiated by gluon(s) radiated by the interacting partons (*initial state radiation*, ISR) or by the Z-decay products (*final state radiation*, FSR). FSR jets should legitimately be included in the reconstructed Z-mass.
However, the presence of either ISR or FSR jets in a 3-jets events confuses the choice of the jet system to be attributed to Z decay. In this sample the invariant mass of the two E_T-leading jets would normally be chosen to reconstruct the Z boson. To improve both the mass resolution and the sensitivity of the search we describe an alternative procedure to reconstruct the Z-invariant mass. Improving the resolution in such a sample means choosing the correct jet combination for building the Z mass.

My thesis work has been to investigate at generator level a sample of simulated CDF WZ events for finding a means to determine the origin of the extra jet and the right jet combination to be chosen for the best reconstruction of the Z mass. This is attempted for the first time in CDF [3, 4, 5].
Four different Neural Networks (NNs) have been trained: NN_{12}, NN_{13}, NN_{23} and NN_{123}. These NNs should make us to be able to decide event by event which among of the 4 four possible different combinations can be used for building the Z-mass in the three jets sample. If one jet is due to ISR, we expect one of the MJ1J2, MJ1J3, MJ2J3 combinations to be correct, while if one jet is due to FSR the choice should be MJ1J2J3. NNs combine kinematical information and some tools developed by CDF Collaboration for distinguishing gluon-like and b-like jets from light-flavored jets [80, 90].

Based on the response of the four NNs, we determine the most likely jet combination for building the Z mass in each event. The method allow to use a different combination from J1J2 in about 50% of cases.
To qualify the potential of the method we have studied an experimental data sample accepting events with a leptonically decaying W and 3 large transverse momentum jets, as in the studies of the simulated WZ sample. The selection cuts accept jets of all flavors (*pretag* sample), and all diboson events including WW besides WZ, ZZ may pass the cuts. We estimate the probability at three standard deviations level to extract an inclusive diboson signal in the 3-jets sample alone (P3σ). After our procedure for building the Z mass is applied, P3σ is about 4 times greater than when building the Z mass "by default" with the two E_T leading jets.

However, in order to discriminate against the WW contribution we apply our technique considering WZ/ZZ as the signal. By investigating the three jets region the sensitivity increases although only a modest improvement is reached. Still, our technique allows including the three jets sample in the WZ/ZZ analyses in order to increase acceptance and sensitivity in the search for the hadronically decaying Z-boson.

Chapter 1

Standard Model and Diboson Physics

The Standard Model (SM) summarizes the current best understanding of elementary particle physics. Its picture of particles and interactions has been tested successfully in a variety of experiments in the last 50 years. In spite of this remarkable success the search for the fundamental laws governing the microscopic world is still far from being completed. There are still many open questions concerning primarily the long list of parameters of the model whose value has no dynamic justification. At the present time the activity of high energy physicists is in fact devoted to searching for new phenomena beyond the Standard Model. The best hint for new physics presently comes from the recent experimental evidence for neutrino oscillations [7], which implies that neutrinos have a very small mass, whose deeper origin is suspected to be related to new physics.
In the Standard Model itself there is still a missing piece, since it predicts a particle, the Higgs boson, which plays a crucial role in assigning mass to the fundamental bosons and fermions, which has not yet been observed. It is hoped that this missing piece will soon be discovered and the model completed consistently.
Looking much beyond the Standard Model, there is a very substantial reason for believing that we are still far from a true understanding of the fundamental laws of Nature. This is because gravity cannot be included in the model, since general relativity appears to be incompatible with quantum field theory [8].

1.1 The Standard Model

The Standard Model is a gauge field theory which incorporates both quantum mechanics and Einstein's theory of special relativity in the attempt to describe the fundamental particles and their interactions [9]. The SM describes three of the four known fundamental forces of nature: strong interactions, electromagnetic interactions and weak interactions. The fourth force, the gravity, is far weaker[1] and is not expected to contribute significantly to the physical processes which are of current interest in high energy particle physics. The energy scale at which quantum gravity effects are expected to become important is of the order of 10^{19} GeV

[1] roughly 40 orders of magnitude smaller than the strong nuclear force

and are of no practical interest in accelerator experiment.
In the SM all fundamental interactions are derived from the requirement of local gauge invariance of the Lagrangian. The gauge transformations belong to the unitary group:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$$

The underscript C is a reminder that $SU(3)$ represents the symmetry group of the colored strong interactions of quantum chromodynamics (QCD). L indicates that the $SU(2)$ group contains left-handed weak isospin doublets and Y is a reminder that the $U(1)$ group contains the right-handed weak hypercharge singlets. These quantities are connected to the electric charge Q through the Gell-Mann-Nishijima relation,

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

Together, the $SU(2)_L \otimes U(1)_Y$ groups govern the unified electroweak force.
The elementary particles are representations of the symmetry group. There are two families of particles: 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. A brief summary of these particles is given below in Table 1.1 and 1.2. Electroweak gauge symmetry predicts massless W, Z and γ bosons.The most accredited theory to give mass to the W and Z bosons is based on the interaction with the so-called Higgs field which is a result of a spontaneously broken symmetry arising in the $SU(2)_L \otimes U(1)_Y$ electroweak sector. The predicted Higgs boson resulting from this broken symmetry is supposed to generate the fermion masses as well. However, it is the only particle in the SM which has yet to be discovered.

1.1.1 Fermions

There are 12 fundamental fermions and their correspondent anti-particles, 6 *quarks* and 6 *leptons*. Quarks are subject to strong interactions as well as to electroweak interactions, while leptons are colorless and are subject only to the electroweak force.

Quarks

Quarks are the constituents of composite particles such as the proton and neutron and can combine to form other unstable hadrons which are bound state of 2 or 3 quarks called mesons and baryons respectively. There are three families (*flavors*) of quark doublets comprising six quarks plus their anti-quarks: up, down, charm, strange, top and bottom. According to SM they are divided into left-handed doublets and right-handed singlets:

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

Quarks have fractional electric charge and a quantum property called *color* similar to the electric charge of electromagnetism. Color is the source of their strong interaction, and comes in three states conventionally named red r, green g and blue b. The charges and the masses of the quarks are listed in Table 1.1.

Flavour	Charge (Q)	Mass [GeV/c^2]	Interaction
u	$+2/3$	$(1.5-3.3)\cdot 10^{-3}$	Strong, EM, Weak
d	$-1/3$	$(3.5-6.0)\cdot 10^{-3}$	Strong, EM, Weak
c	$+2/3$	$1.27^{+0.07}_{-0.11}$	Strong, EM, Weak
s	$-1/3$	$104^{+26}_{-34}\cdot 10^{-3}$	Strong, EM, Weak
t	$+2/3$	171.2 ± 2.1	Strong, EM, Weak
b	$-1/3$	$4.2^{+0.17}_{-0.07}$	Strong, EM, Weak

Table 1.1: Observed properties of the quarks of the SM [2]

Leptons

There are six types of leptons (plus their anti-particles) in the SM. These are the electron, muon, tau (e, μ, τ) and their respective neutrinos (ν_e, ν_μ, ν_τ).
Leptons are described as doublets of the $SU(2)_L$ group with their associate neutrinos, as *left-handed* eigenstates (with -1 eigenvalue chirality), one for each generation. As the Goldhaber et al. experiment [10] has proved, neutrinos with positive chirality eigenvalues do not exist. This is the reason why the *right-handed* fermions in the SM are singlet for $SU(2)_L$.

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

In Table 1.2 are listed their properties.

Flavour	Charge (Q)	Mass [GeV/c^2]	Interaction
e	-1	$0.511\cdot 10^{-3}$	EM, Weak
ν_e	0	$< 225\cdot 10^{-9}$	Weak
μ	-1	$105.6\cdot 10^{-3}$	EM, Weak
ν_μ	0	$< 0.17\cdot 10^{-3}\cdot 10^{-3}$	Weak
τ	-1	1.776	EM, Weak
ν_τ	0	$< 15.5\cdot 10^{-3}$	Weak

Table 1.2: Observed properties of the leptons of the SM [2]

1.1.2 Gauge Bosons

SM interactions are mediated by spin-1 bosons, which are the generators of the symmetry group. The 8 colored gluons g are the massless mediators of the strong force, i.e. they are the generators of the group $SU(3)_C$. The photon γ, $W^{\pm}$ and Z are the force carriers of the electroweak interactions, i.e. they are the generators of the group $SU(2)_L \otimes U(1)_Y$.
The gauge bosons and their properties are summarized in Table 1.3. The basic electroweak unification formulae are recalled in Appendix A.

Flavour	Charge (Q)	Mass [GeV/c^2]	Width [GeV/c^2]	Interaction
g	0	0		Strong
γ	0	0		EM
$W^{\pm}$	± 1	80.399 ± 0.0023	2.141 ± 0.041	Weak
Z	0	91.188 ± 0.002	2.4952 ± 0.0023	Weak

Table 1.3: Observed properties of the force carriers [2]

1.2 Quantum Chromodynamics

Quantum chromodynamics (QCD) is the theory of the strong interactions between quarks and gluons. Quarks carry a single color charge while gluons, which are mediator of color flow are bicolored (i.e. carrying color and anti-color). $SU(3)_C$ has 8 gauge bosons, which can be foreseen since the three colors and three anti-colors give nine total color states for the gluon: a color octet and a color singlet. The singlet has not net color and so has to be excluded as interaction carrier.
Quarks carry color charge, but colored particles have never been observed in nature so all terms of the lagrangian must be singlets of $SU(3)_C$, i.e. quarks cannot be observed as free particles. Colored partons[2] will be confined in objects which are as a whole colorless. We can only observe color singlet quark-antiquarks bound states (*mesons*) or of three quarks or three antiquarks (*baryons*). Mesons and baryons are collectively denoted as hadrons and, being composed of quarks, are subject to strong interactions.
At small distances QCD is well described in terms of weakly interacting quarks and gluons, while at large distances, of the order of 1 fm, the theory becomes non-perturbative and quarks are confined. The strong interactions generate dynamically a characteristic energy scale $\Lambda_{QCD} \sim 1/\text{fm} \sim 200$ MeV, which is the energy scale where SM should be considered valid.
The crucial property of QCD is *asymptotic freedom*. The coupling constant of QCD (α_S) is a running function of the momentum transferred in the interaction, q^{μ}, which is approximately given by

$$\alpha_S(q^2) = \frac{12\pi}{(33 - 2n_f)\log(q^2/\Lambda^2_{QCD})} \tag{1.2}$$

[2] Quarks as well as gluons.

where n_f is the number of quark flavors whose mass is greater than the q^2 of interest [11]. At very large q^2 (corresponding to very short approach distances) α_S becomes increasingly small. This phenomena is known as asymptotic freedom. This property allows, for high-q^2 interaction, perturbative expansion of QCD processes which remain finite. This is often the case of the collisions at Tevatron, where it is possible to calculate interaction cross sections as perturbative expansions.
When highly energetic quarks or gluons are produced in high energy physics experiment a process called *hadronization* or *showering* takes place: after a parton is produced in an interaction, the potential with the parton system of origin, tries to keep it bound until the strength reaches a breaking point where a $q\bar{q}$ pairs are created. The new partons are approximately collinear with the original parton and combine into meson or baryons in such a way that a spray of colorless particles is observed which move close to the same direction. The final state in which we observe the parton generated in the interaction is a collimated *jet*.
Most of the fragmentation process is non-perturbative and not completely theoretically under control, since during the fragmentation process the particle energies become successively smaller and perturbative QCD is no longer applicable. Phenomenological models are usually applied in order to describe completely jet features [12].

1.3 The Higgs Mechanism

An explicit mass term in the interaction Lagrangian cannot be accepted because it would violate gauge symmetry. The gauge invariance of $SU(2)_L \otimes U(1)_Y$ implies massless fermions and weak bosons, in total contradiction with reality. Non-zero mass can be given, while preserving a local gauge symmetry, by *spontaneous symmetry breaking*.
Spontaneous breaking of symmetry is based on the possibility, in systems with infinite degrees of freedom, to have a Lagrangian invariant under a group G of transformation that produces non symmetric states.
The Higgs mechanism was proposed by P. Higgs in 1964 [13] and has been fully incorporated into the Standard Model by Weinberg and Salam [15, 16] in the $SU(2)_L \otimes U(1)_Y$ theory.
This mechanism gives rise to particle masses, at the price of introducing a new spin-0 scalar gauge field, called the Higgs boson, whose corresponding Lagrangian density is

$$\mathscr{L}_\Phi = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) \tag{1.3}$$

$\mathscr{L}_\Phi$ in eq. (1.3) maintains invariance under the local gauge transformation. The potential is defined as

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$$

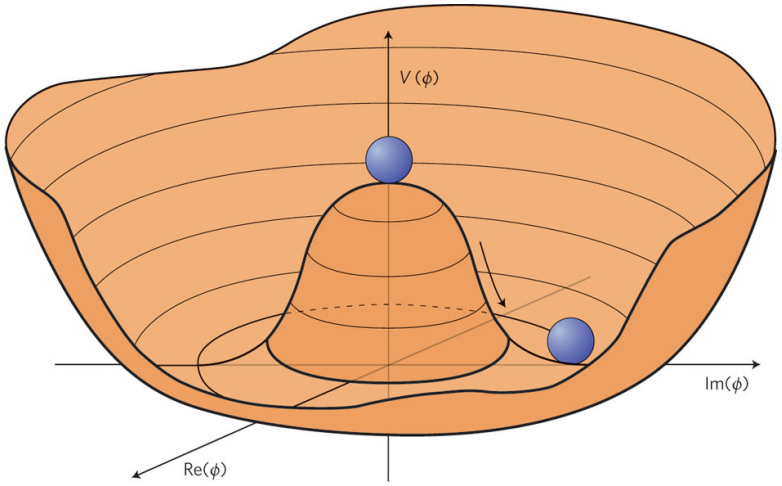

Figure 1.1: Higgs potential V(Φ) for μ^2 <0

If μ^2 is positive then the potential V(Φ) is symmetric about its minimum which is 0. However, in the case where μ^2 <0 the potential has a minimum at $\Phi^\dagger\Phi = \mu^2/2\lambda$ as shown in Fig. 1.1. Thus, the field Φ has a non-zero *vacuum expectation value* (VEV) and choosing one of a set of degenerate states of minimum energy, the gauge symmetry is broken.

The **Goldstone theorem** states that fields that acquire a VEV will have an associated massless boson which will disappear, transformed into the longitudinal component of a massive gauge boson.

Since the photon is known to be massless, the symmetry is chosen to be broken so that only the fields with zero electric charge (the ones that cannot couple to the electromagnetic interaction) acquire a VEV.

Expanding around the true minimum of the theory, the physical field can be written in terms of the vacuum expectation value and two real fields with zero vacuum expectation value ξ and H as

$$\Phi(x) = e^{i\sigma_j \xi_j(x)/\mathrm{v}} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \mathrm{v} + H(x) \end{pmatrix}$$

H is the Higgs field, σ_j are the Pauli matrices and ξ_j are non-physical Goldstone bosons.

Since the SM theory must be invariant under local gauge transformations, with an appropriate unitary gauge transformation it is possible to cancel the dependence on the ξ fields. When we introduce this specific representation of Φ in the SM Lagrangian, the Goldstone bosons vanish while the gauge bosons acquire terms which can be identified as mass terms. From the broken Lagrangian one finds the following relations between the masses of the gauge bosons

$$M_W = M_Z \cos\theta_W \qquad M_H = \sqrt{-2\mu^2}$$

Experimentally one can measure both M_W and M_Z. Given the relationship between g and g'[3] one finds that v =246 GeV [17]. Then the only undetermined parameter is μ which implies that the mass of the Higgs is undetermined and needs to be measured experimentally. So far, the Higgs boson has still not been observed, only experimental limits from both LEP and the Tevatron exist [18, 20].

1.4 WZ Associated Production

The work reported in this thesis is related to the measurement of the cross section for WZ in which W boson decays leptonically ($W \to \ell\nu$, $\ell = e$ or μ) while the Z boson decays hadronically ($Z \to q\bar{q}$). The leading order (LO) Feynman diagrams for these processes are shown in Fig. 1.2.

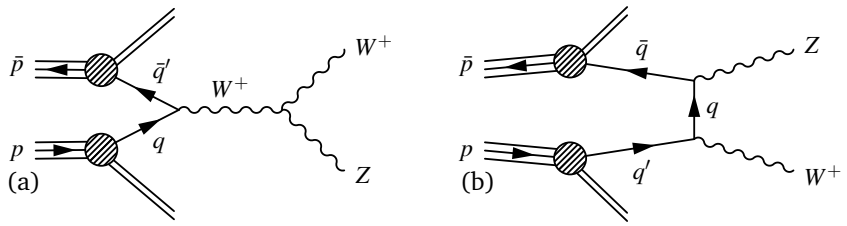

Figure 1.2: Tree-level s-channel (a) and t-channel (b) Feynman diagram for WZ diboson production at the Tevatron.

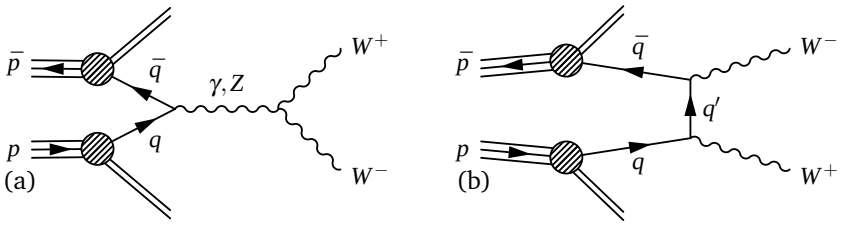

Figure 1.3: Tree-level s-channel (a) and t-channel (b) Feynman diagram for WW diboson production at the Tevatron.

Next-to-leading order (NLO) calculations for the diboson production at $\sqrt{s} = 1.96$ TeV predict the cross sections [1],

$$\sigma(p\bar{p} \to WW) = 11.34^{+0.56}_{-0.49} \text{ pb}$$
$$\sigma(p\bar{p} \to WZ) = 3.22^{+0.20}_{-0.17} \text{ pb}$$
$$\sigma(p\bar{p} \to ZZ) = 1.20^{+0.05}_{-0.04} \text{ pb}$$

[3] $\cos\theta_W$, g and g' are defined in Appendix A

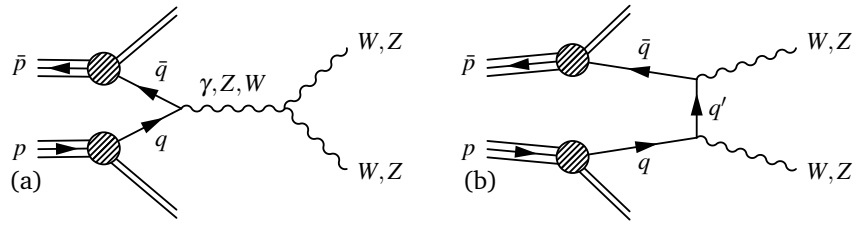

Figure 1.4: Tree-level s-channel (a) and t-channel (b) Feynman diagram for ZZ diboson production at the Tevatron.

Process	Decay Mode	Branching Ratio
WZ	$\ell\nu\ell\ell$	3.3
	$\ell\nu\nu\nu$	6.5
	$\ell\nu q\bar{q}$	22.6
	$q\bar{q}'\ell\ell$	6.8
	$q\bar{q}'\nu\nu$	13.5
	all hadronic	47.2
ZZ	$\ell\ell\ell\ell$	1.2
	$\ell\ell\nu\nu$	4.0
	$\nu\nu\nu\nu$	4.0
	$\ell\ell q\bar{q}$	14.1
	$\nu\nu q\bar{q}$	27.9
	all hadronic	48.8
WW	$\ell\nu\ell'\nu'$	10.5
	$\ell\nu q\bar{q}'$	43.8
	all hadronic	45.7

Table 1.4: Branching ratios for WW, WZ and ZZ processes [2].

Each boson can decay either leptonically or hadronically, which yields three classes of final state: fully leptonic, fully hadronic and semileptonic.

The semileptonic final state studied in this thesis has a higher rate than the fully leptonic mode but suffers from larger backgrounds due to W+jets and Z+jets production, which have relatively larger production cross section.

A primary motivation for studying diboson physics comes from the fact that their production and interactions provide a test of the Electroweak sector of the SM, in particular of the very characteristic prediction of trilinear gauge boson couplings [21, 22, 23]. The virtual W propagator diagram in WZ production in particular is sensitive to the WWZ coupling.

The deviation of a TGC or production cross section from the values predicted by the SM would be an hint of new physics beyond the SM and could give some clues about the mechanism responsible for Electroweak Symmetry Breaking.

Furthermore, the SM Higgs boson search is often performed in the diboson channel. In fact, one of the most promising channels for discovering a low mass Higgs boson ($M_H \lesssim 135$ GeV/c^2) is in associated production with a W boson and H decay into two beauty jets.

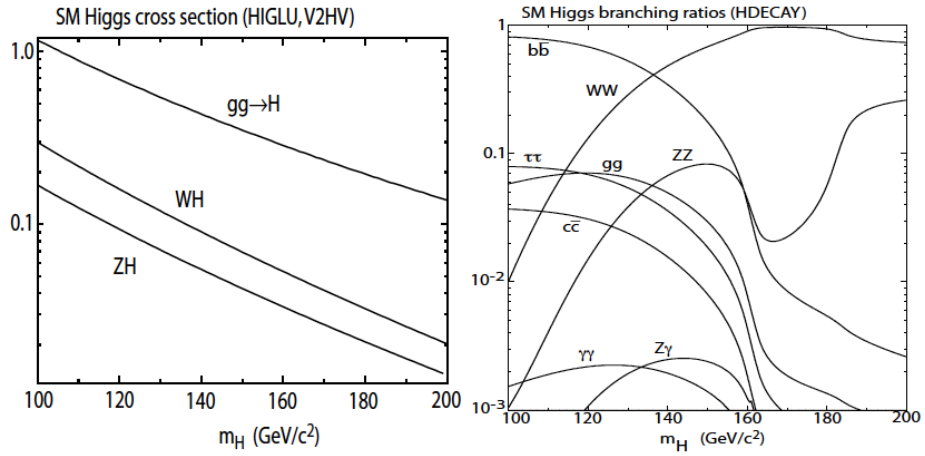

Figure 1.5: SM Higgs boson production cross section at the Tevatron and the BR of Higgs.

At the Tevatron, the process $WH \to Wb\bar{b}$ has an expected $\sigma \cdot BR$ [4] about five times lower than $WZ \to Wb\bar{b}$ for $m_H \simeq 120$ GeV/c^2, as shown in Table 1.5 and in Fig. 1.5.

	BR
$W \to \ell\nu$ $(\ell = e, \mu)$	21.6%
$Z \to b\bar{b}$	15.1%
$H \to b\bar{b}$	64.9%
	$\sigma \times$ **BR** [pb]
$p\bar{p} \to WZ$	3.22
$p\bar{p} \to WH$	0.15
$WZ \to \ell\nu b\bar{b}$ $(\ell = e, \mu)$	0.105
$WH \to \ell\nu b\bar{b}$ $(\ell = e, \mu)$	0.021

Table 1.5: BR and expected cross sections for WZ/WH [2, 19].

The Feynman diagrams for $W^{\pm}H \to \ell\nu b\bar{b}$ are notably similar to those for $W^+W^-/W^{\pm}Z \to \ell\nu q\bar{q}$ shown in Fig. 1.2-1.3. 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 boson or a weak gauge boson (W or Z). Therefore, the $W^{\pm}Z \to Wb\bar{b}$ is an irreducible background in the search for low mass Higgs. Measuring directly this diboson production supplies a in situ measurement of the size of this background.
Detecting WZ in this channel represents in a number of ways a benchmark for Higgs searches with similar final states. The Higgs boson searches and this analysis share the same challenge of separating a small signal with kinematical features similar to WZ from a large background. Of course, this analysis has the advantage that the signal is in many times larger and that the Z mass is known.

[4] BR being the Branching Ratios of $Z/H \to b\bar{b}$

Discovering the Higgs boson (as well as measuring $W^+W^-/W^\pm Z \to \ell\nu q\bar{q}$) depends largely on how precisely a resonance in the dijet invariant mass is reconstructed. A measurement of $W^+W^-/W^\pm Z \to \ell\nu q\bar{q}$ would determine the actual resolution of the measurement of a dijet resonance.

$W^\pm Z$ production has been measured only in the fully leptonic finale state. The reconstruction of $W^\pm Z$ events when the Z decays hadronically represents a challenge regarding the separation of signal from the dominant background, a challenge shared by Higgs boson searches. One can therefore state that the measurement of $W^\pm Z \to \ell\nu q\bar{q}$ produciton provides a *standard candle* with which to calibrate and optimize many techniques used in the SM light Higgs searches.

Weak diboson production is also a significant background for higher mass SM Higgs searches ($M_H > 135$ 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 inclusive diboson production impact the power of the search of a H resonance.

The event selection for this search shares most of the trigger, MonteCarlo (MC) simulation and normalization methods of the H$\to b\bar{b}$ search. The multivariate event classification schemes 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.

1.5 Diboson Physics at the Tevatron

The WW and ZZ productions have been extensively studied at LEP [24]. The WW production cross section measured at LEP agrees well with the SM predictions. WZ production was not allowed at LEP by charge conservation.

WW, WZ and ZZ production have been observed in their fully leptonic decay modes by CDF and DØ and their measured cross sections are in agreement with the theoretical predictions [26, 27, 28, 29].

At Tevatron **WW** production cross section at next-to-leading order is expected to be ~ 11.34 pb. The first observation has been reported by DØ [26] consistent with SM prediction. CDF most recent analysis, on approximately 3.6 fb^{-1} of data, based on a likelihood ratio method calculated a cross section [27], again consistent with SM predictions.

WZ production cross section was first observed by CDF collaboration with a data sample of approximately 1.1 fb^{-1} in the $WZ \to \ell\nu\ell'\ell'$ channel ($\ell = e, \mu$) [25]. More recent measurements have been done both by CDF and DØ in the $WZ \to \ell\nu\ell'\ell'$channel [29, 28].

CDF also reported the evidence for combined WW and WZ production with W/Z decaying to jet-jet $WW/WZ\to \ell\nu jj$[5]. The branching ratio for this process is higher than for the leptonic decay channel but it has not been measured before at hadron colliders because of the

[5]It's not possible with actual jets resolution to distinguish between a W jet pair and a Z jet pair so the analysis has been done combining the two processes.

Process	Experiment	$\mathscr{L}$ (fb^{-1})	Measured σ (pb)	Theory σ (pb)
$WW \to \ell\nu\ell\nu$	CDF [27]	3.6	$12.1\pm0.9(\text{stat.})^{+1.0}_{1.4}(\text{syst.})$	$11.34^{+0.56}_{-0.49}$
	DØ [26]	1	11.5±2.2(stat.+syst.)	
$ZZ \to \ell\ell\ell\ell$	CDF [29]	6	$1.7^{+1.2}_{-0.7}(\text{stat.})\pm0.2(\text{syst.})$	$1.20^{+0.05}_{-0.04}$
$ZZ \to \ell\ell\nu\nu$	CDF [30]	5.9	$1.45^{+0.6}_{-0.5}$(stat.+syst)	$1.20^{+0.05}_{-0.04}$
	DØ [32]	1.7	$1.60\pm0.63(\text{stat})^{\pm0.16}_{0.17}$(syst)	
$WZ \to \ell\ell\ell\nu$	CDF [29]	6	4.1±0.6(stat.)±0.4(syst.)	$3.22^{+0.20}_{-0.17}$
	DØ [28]	1	$2.7^{+1.7}_{-1.3}$(stat.+syst.)	
$WW/WZ/ZZ \to \nu\nu$+jets	CDF [33]	3.5	18.6±2.8(stat.)±02.6(syst.)	16.8±0.5
$WW/WZ \to \ell\nu$+jets	CDF[36]	4.6	$16.5^{+3.3}_{-3.0}$(stat.+syst.)	15.9±0.9
	DØ [34]	1	20.2±4.5(stat.+syst.)	

Table 1.6: Measurements of the diboson production cross section in leptonic and semileptonic final state at the Tevatron.

large background due to W+jets events. The first observation was achieved in a final state with two jets and large invisible energy [33]. In the channel with an identified lepton and jets, DØ had evidence of diboson production [34] while CDF published an observation [35].
ZZ production at Tevatron has been studied in $ZZ\to \ell\ell\ell\ell$ and $ZZ\to \ell\ell\nu\nu$ decay channel in which sensitivity is best. The former has a very low background but also a small branching ratio, while the latter has a greater branching ratio but a higher background due mainly to Z+jets and WW events.
Tab. 1.6 summarizes the current status of measurements for the two Tevatron experiments in the leptonic and semileptonic final states.

Chapter 2

Experimental Apparatus

The measurement described in this thesis is based on a data sample collected by the CDF II detector during Run II operations at the Fermilab's Tevatron Collider, where the WZ and WH searches are in progress. This chapter provides a general description of the experimental apparatus, both collider and detector, focusing on the tracking system, calorimeters and trigger systems, which are the more relevant elements for this analysis. A more complete description of the entire detector can be found in [37].

2.1 The Tevatron Collider

The Tevatron [38] located at the Fermi National Accelerator Laboratory (*Fermilab*) in Batavia (Illinois, USA) is a proton-antiproton ($p\bar{p}$) collider with a center-of-mass energy of 1.96 TeV. The Tevatron started operating in 1985 at a center of mass energy of 1.8 TeV. Since the year 2002 its c.m.s. energy was increased to 1.96 TeV, and it operates only in the collider mode. The upgraded machine collides 36×36 bunches every 396 ns. As shown in Fig. 2.1, the Tevatron complex has five major accelerators and storage rings used in successive steps, to produce, store and accelerate the particles up to 980 GeV . The main components are:

- the proton source, including a linear accelerator and a *Booster* synchrotron;
- *Main Injector*, feeding the antiproton source and the Tevatron collider;
- antiproton source, including a *Debuncher*, an *Accumulator* and a *Recycler* ring;
- Tevatron;
- the extraction lines to fixed target experiment.

The so-called Run II is the data acquisition period which started in 2002 and is expected to end the 30^{th} of September 2011. The present analysis deals with the data collected in Run II.

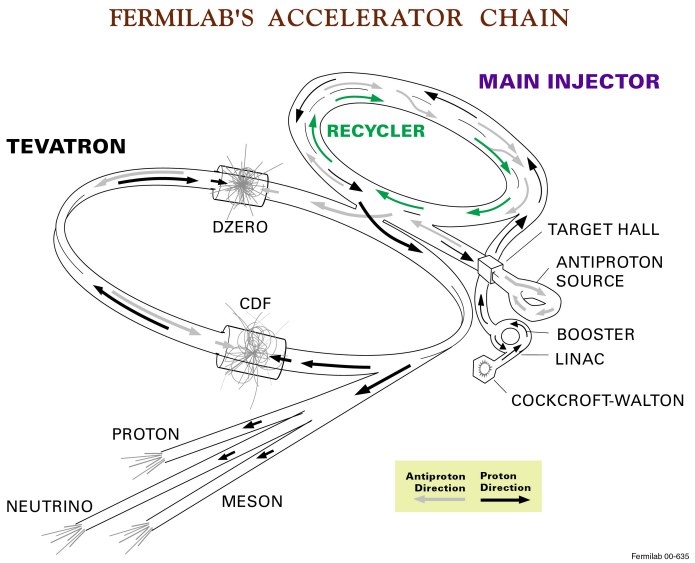

Figure 2.1: The Tevatron Collider Chain at Fermilab.

2.1.1 The Proton Source

The acceleration cycle starts with the production of protons from negatively ionized hydrogen atoms, H^-, which are accelerated to 750 KeV of kinetic energy by a Cockroft-Walton electrostatic accelerator. Pre-accelerated H^- ions are then injected into the LINAC where they are accelerated up to 400 MeV by passing through a 150 m long chain of radio-frequency (RF) accelerator cavities grouping the ions into bunches. To obtain protons, the H^- ions are passed through a carbon foil which strips their electrons off.
Inside the *Booster* the protons are merged into bunches and accelerated from 400 MeV to 8 GeV by a number of RF cavities, prior to enter into the *Main Injector*.
The protons are then transfered to a synchrotron, called the *Main Injector*[1], which brings their energy to 150 GeV. This is the last step before protons are injected into the Tevatron.

2.1.2 Antiproton Production

The production of the antiproton beam is significantly more complicated. The cycle starts with extracting a 120 GeV proton beam from the *Main Injector* onto a stainless steel target. This process produces a variety of different particles, among which are $\bar{p}$[2]. The particles come off the target at many different angles and they are focused into a beam line with a lithium magnetic focusing lens. In order to select only the antiprotons, the beam is sent through a pulsed magnet which acts as a charge-mass spectrometer.

[1] Completed in 1999 for Run II, it is located in a 3 Km circumference tunnel (which houses also the antiproton *Recycler*) and is approximately tangent to the Tevatron.
[2] The production rate, for 8 GeV antiprotons, is about $18\bar{p}/10^6 p$

The emerging antiprotons, having a bunch structure similar to that of the incident protons and a large energy spread, are stored in a *Debuncher*, a storage ring where their momentum spread is reduced via stochastic cooling stations[3]. The bunch structure is destroyed resulting in a continuous beam of 8 GeV antiprotons which are transferred to the next ring before the start of the next cycle.
At the end of the debunching process the antiprotons are stored in the *Accumulator*, housed in the same tunnel of the *Debuncher*, which is a triangle-shaped storage ring where they are further cooled and stored until all cycles of the *Debuncher* are completed.
After the accumulator has collected an amount of antiprotons saturating its acceptance ($\sim 6 \times 10^{11}$), they are transferred to the *Recycler*[4] which is a larger acceptance 8 GeV fixed energy storage ring, placed in the *Main Injector* enclosure, made of permanent magnets and further cooled using stochastic cooling and accumulated.
When a current sufficient to create 36 bunches with the required density is available, the $\bar{p}$ are injected into the Main Injector where they are accelerated to 150 GeV.

2.1.3 Tevatron

The Tevatron is a large synchrotron, 1 Km in radius, that accelerates particles from 150 GeV to 980 GeV. It keeps both protons and antiprotons in the same beam pipe, revolving in opposite directions. Electrostatic separators produce a strong electric field that keeps the two beams away form each other except at the collision point. The beam is steered by 774 superconducting dipole magnets and 240 quadrupole magnets with a maximum magnetic field of 4.2 Tesla. They are cooled by liquid helium to 4.2 K, at which point the niobium-titanium alloy of the magnet coils becomes superconducting.
The process of injecting particles into the machine, accelerating them, and initiating collisions, referred to as a “shot", starts with injection of protons, one bunch at a time, at 150 GeV from the *Main Injector*. The antiprotons are injected three bunches at a time. RF cavities accelerate the beams to 980 GeV, and then some electrostatic separators switch polarity to cause the beams to collide at two points.
Each interaction point lies at the center of a particle detector: one named DØ (according to its location in the Tevatron optics) and the other named the Collider Detector at Fermilab (CDF) located at BØ .
Stable running conditions and data-taking by the experiments are reached after beams are scraped with remotely-operated collimators to remove the beam halo. A continuous period of collider operation using the same collection of protons and antiprotons is called a “store".

[3] Stochastic cooling is a technique used to reduce the transverse momentum and energy spread of a beam with a very small beam loss.

[4] Antiproton availability is the most limiting factor at Tevatron for attaining high luminosities: keeping a large stash of antiprotons inside the *Recycler* has been one of the most significant engineering challenges and the excellent performance of the *Recycler* is an achievement of prime importance for the good operation of the accelerator.

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

Table 2.1: Accelerator parameters for Run II configuration.

2.2 Luminosity and Tevatron Performance

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, $\mathscr{L}$. The former defines the accessible phase-space for the production of final states particles. The latter is defined as the interaction rate per unit cross section of the colliding beams (collisions/(cm^{2}s)). In the absence of a crossing angle or position offset, the luminosity at the CDF or DØ is given by the expression:

$$\mathscr{L} = \frac{f_{bc} N_b N_p N_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)} F\left(\frac{\sigma_l}{\beta^*}\right), \tag{2.1}$$

where f_{bc} is the revolution frequency, N_b is the number of bunches, $N_{p(\bar{p})}$ is the number of protons (antiprotons) per bunch, and $\sigma_{p(\bar{p})}$ is the transverse and longitudinal rms proton (antiproton) beam size at the interaction point. F is a form factor with a complicated dependence on the low beta value, β^*, and the bunch length, σ_l[5]. Table 2.1 shows the design Run II accelerator parameters [38].

Figs. 2.2 (a)-(b) show, respectively, the evolution in the integrated luminosity, defined as $L = \int \mathscr{L} dt$, and the instantaneous luminosity at the start of Tevatron stores, since the start of Run II up to September 2011. The progressive increase in the integrated luminosity and the continuous records in the instantaneous luminosity [6] prove the good performance of the accelerator.

At the end of September 2011, physics quality data corresponding to about 12 fb^{-1} are expected to be collected by both CDF and DØ .

[5] The beta function is a measure of the beam width, along the accelerator and it is proportional to the beam's x and y extent in phase space. β^* is the value of this function in the collision point

[6] In May 2011, the record in the instantaneous luminosity was $4.31 \times 10^{32} cm^{-2} s^{-1}$.

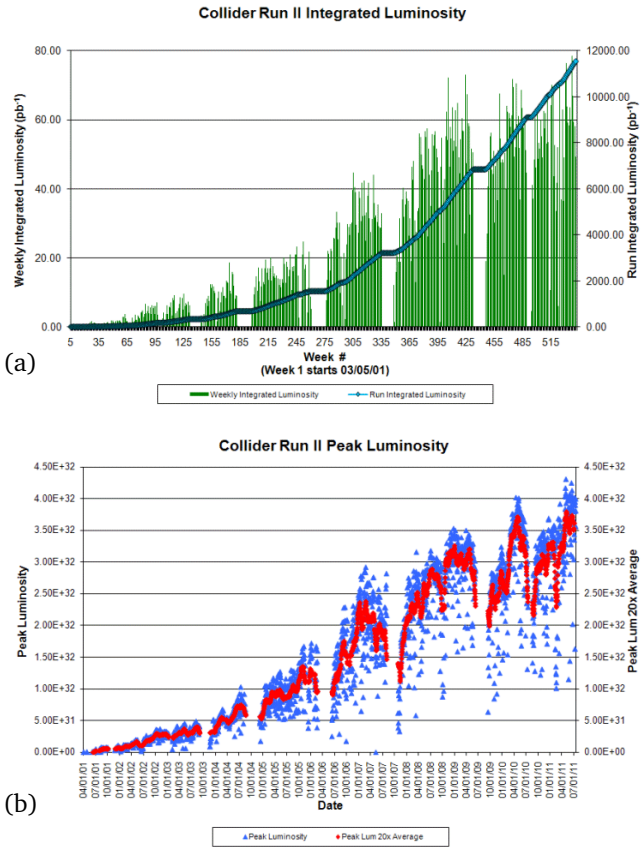

Figure 2.2: (a) Integrated luminosity as a function of Run II weeks. Empty periods are due to Tevatron shut down. (b) Run II Peak luminosity in the stores and its average, as function of calendar date (up to September 2011).

2.3 CDF Run II detector

The CDF II detector [37], in operation since 2001, is an azimuthally and forward-backward symmetric apparatus designed to study $p\bar{p}$ collisions at the BØ interaction point of the Tevatron. It is a general purpose, cylindrical-shaped detector (Fig. 2.3), which consists of:

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 constant 1.4 T magnetic field parallel to the beam. 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.

Time-of-Flight system, radially outside the COT for particle identification up to momenta below 2 GeV.

Calorimeter system, with the purpose of measuring the energy of charged and neutral particles produced in the interaction.

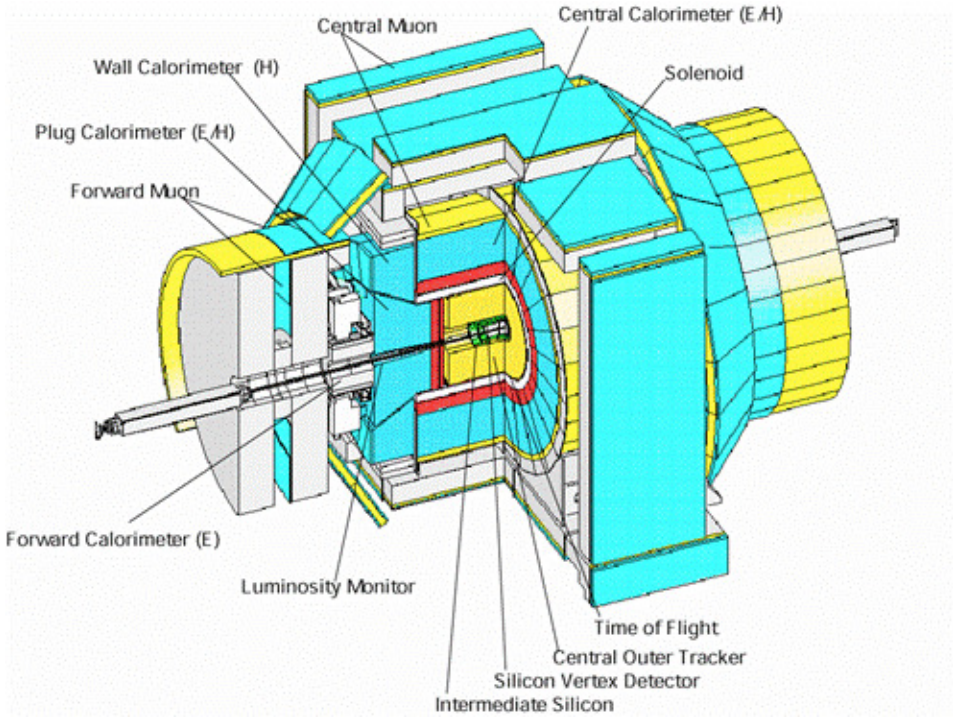

Figure 2.3: Isometric view of the CDF II detector.

Muon chambers and scintillators, used to track and identify muons that pass through the calorimeters interacting as minimum-ionizing-particles (MIP).

Luminosity monitors, for the instantaneous luminosity measurement, necessary to predict event yields. Monitoring the instantaneous luminosity is also critically important to safeguard delicate detector components.

2.3.1 Coordinate System and Useful Variables

The CDF detector is approximately cylindrically symmetric around the beam axis. Its geometry can be described in cartesian as well as in cylindrical coordinates.

The left-handed cartesian system is centered on the nominal interaction point with the z axis in the direction of the proton beam and the x axis on the Tevatron plane pointing radially outside.

Since the colliding beams of the Tevatron are unpolarized, the resulting physical observations are invariant under rotations around the beam line axis.

A cylindrical coordinate system is particularly convenient to describe the detector geometry. The cylindrical coordinates are the azimuthal angle, ϕ, and the polar angle, θ,

$$\phi = \tan^{-1}\frac{y}{x} \qquad \theta = \tan^{-1}\frac{\sqrt{x^2+y^2}}{z}$$

A momentum-dependent particle coordinate named *rapidity* is also commonly used. The rapidity is defined as

$$Y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z} \tag{2.2}$$

where E is the energy and p_z is the z component of the momentum of the particle[7]. It is used instead of the polar angle θ because it is Lorentz invariant under $\hat{z}$ boost. In the ultra relativistic limit, or in general away from $\theta = 0$ when the mass of the particle is ignored, rapidity becomes dependent only upon the production angle of the particle with respect to the beam axis. This approximation is called *pseudorapidity* η and is defined by

$$Y \xrightarrow{p\gg m} \eta = -\ln\left(\tan\frac{\theta}{2}\right)$$

A value of $\theta = 90^\circ$ would be perpendicular to the beam axis and correlates to $\eta = 0$. The pseudorapidity is commonly used to identify different detector regions according to their position respect to the beamline and interaction vertex position, as shown in Fig. 2.4 (a).
As the event-by-event longitudinal position of the interaction is distributed around the nominal interaction point with 30 cm r.m.s. width, it is useful to distinguish **detector pseudo-rapidity**, η_{det}, measured with respect to the (0, 0, 0) nominal interaction point, from **event pseudo-rapidity**, η, which is measured with respect to the z_0 position of the event vertex where the particle originated.[8]
Since both η and $\Delta\eta$ between particles are Lorentz invariant, a useful Lorentz invariant separation, ΔR, can be defined between two particles as

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

Often, ΔR is used to define a Lorentz invariant cone around a single particle or detector position, in order to study nearby detector activity.
Certain other quantities are useful for studying $p\bar{p}$ interactions, among these the transverse momentum, the transverse energy and the five parameter of helices describing the trajectory of a charged particle in a constant magnetic field.
The transverse momentum, p_T, of a particle is defined as $p \cdot \sin\theta$ that is $p_z/\sinh\eta$.
Charged particles moving through a homogeneous solenoidal magnetic field along the z direction follow helical trajectories. Knowing that the projection of the helix on the $x-y$ plane is a circle, to uniquely parameterize a helix in three dimensions, five parameters are needed:

[7] The rapidity can be derived from the Lorentz-invariant cross-section: $E\frac{d^3\sigma}{(dp)^3} = E\frac{d^2\sigma}{2\pi p_T dp_T dp_z}$. Observing that only E and p_z change under z boosts, we can replace them by a variable Y such as $E\frac{dY}{dp_z} = 1$. Solving for Y we get (2.2).

[8] An idea of the difference is given by considering that $|\eta_{\rm det} - \eta_{part}| \approx 0.2$ if the particle is produced at $z = 60$ cm from the nominal interaction point.

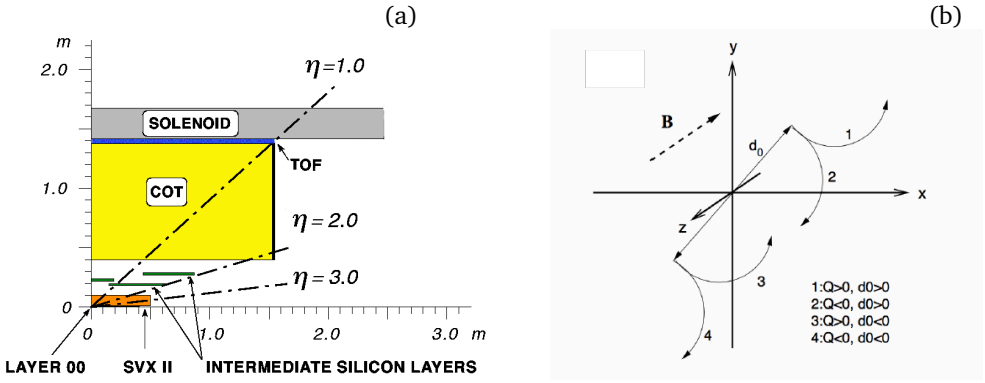

Figure 2.4: (a) The CDF II tracker layout showing the different subdetector systems. (b) Schematic drawing illustrating the impact parameter d_0.

$\boldsymbol{C}$ – signed helix (half)-curvature, defined as $C = q/2R$, where R is the radius of the helix and q is the charge of the track. This is directly related to the transverse momentum. When the magnetic field (B) is measured in Tesla, C in m^{-1} and p_T in GeV: $p_T = 0.15\ qB/|C|$;

$\boldsymbol{\phi_0}$ – ϕ angle of the particle at the point of closest approach to the z-axis;

$\boldsymbol{d_0}$ – signed impact parameter[9], i.e. the radial distance of the closest approach in the transverse plane between the helix and the beam line (z-axis), defined as $d_0 = q(\sqrt{x_0^2 + y_0^2} - R)$, where (x_0, y_0) are the coordinates of the track, in the transverse plane, in the point of closest approach to the beam line. This is schematically drawn in Fig. 2.4 (b);

$\boldsymbol{\lambda}$ – the helix pitch, i.e. $\cot(\theta)$, where θ is the polar angle of the particle at the point of its closest approach to the z-axis. This is directly related to the longitudinal component of the momentum: $p_z = p_T \cot\theta$;

$\boldsymbol{z_0}$ – the z coordinate of the point of closest approach, defined as the interception between the track and the z-axis in the transverse plane.

2.3.2 Tracking

Three-dimensional charged particle tracking is achieved through an integrated system consisting of three silicon inner sub-detectors and a large outer drift-chamber, all contained in the solenoid field.

The silicon detectors provide excellent impact parameter, azimuthal angle and z resolution. The COT provides excellent resolution in C, ϕ and η. Together they provide a very accurate measurements of the helical paths of charged particles.

We will describe this system starting from the devices closest to the beam and moving outwards (see Fig. 2.4 (a)).

[9] With the impact parameter significance, defined as $|d_0/\sigma_{d_0}|$ is possible to estimate if the particle comes from primary vertex, $d_0 \sim 0$, or from a secondary one.

Silicon Tracker

The full CDF Silicon Detector [39] is composed of three approximately cilindrical coaxial subsystems (see Fig. 2.6): the **L00** (Layer 00), the **SVX** (Silicon VerteX), the **ISL** (Intermediate Silicon Layer).
The silicon wafers operate as reverse-biased $p-n$ junctions. When a charged particle passes through the detector material, it causes ionization. In the case of a semiconductor material, it means that electron-hole pairs will be produced. The amount of charge gathered at the anode and the cathode is proportional to the path length traversed in the material by the charged particle.
By segmenting the p or n side of the junctions into “strips" and reading out the charge deposition separately on every strip we can measure the position of the charged particle. At CDF the typical distance between two strips is about 60 μm.

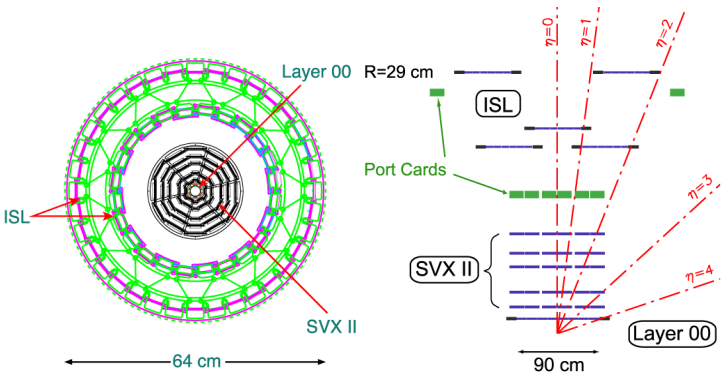

Figure 2.5: Schematic representation of the Silicon Vertex Detector at CDF showing the different layers and parts of the detector in the $r-\phi$ (left) and in the plane $r-z$ planes (right).

- **L00** is a 90 cm long, radiation hard, assembly of single sided silicon detectors, structured in longitudinal strips. It is mounted directly on the beam pipe at $1.35-1.62$ cm from the beam axis. The detector support structure is in carbon fiber with integrated cooling system. Being so close to the beam, L00 allows to reach a resolution of $\sim 25/30$ μm on the impact parameter of tracks of moderate p_T, providing a powerful help to identify long-lived hadrons containing a b quark.

- **SVX** is composed of three 29 cm long cylindrical barrels [40], radially organized in five layers of double-sided silicon wafers extending from 2.5 cm to 10.7 cm. Each barrel is segmented into 12 (each covering $\sim 30^o$ in ϕ) wedges. The double-side structure of the wafers allows a three dimensional position measurements: one side of the wafer has axial strips (parallel to the beam), the other one has either 90^o strips (perpendicular to the beam) or 1.2^o stereo strips (at small angle with respect to the beam).
  This detector provides position information with a 12 μm resolution on the single hit and some dE/dx ionization information.

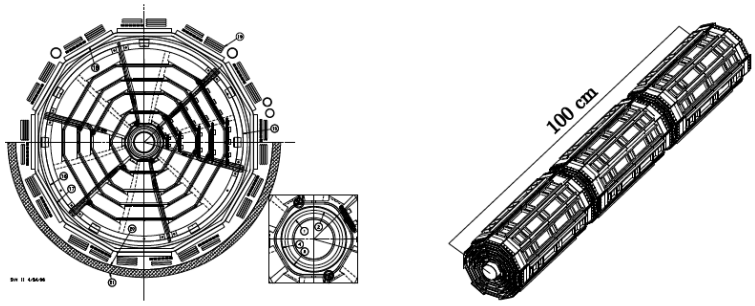

Figure 2.6: Cross section view (left) and isometric view (right) of the SVX II detector.

- **ISL** consists of 5 layers of double sided silicon wafers similar to those of SVX II: four are assembled in twofold telescopes with planes at 22 cm and 29 cm radial distance from the beamline covering $1 < |\eta| < 2$. One single central layer is located at $r = 22$ cm, covering $|\eta| < 1$. The two ISL layers are important to help tracking in a region where the COT coverage is incomplete and also help matching between SVX II and COT track.

The combined resolution of the CDF inner trackers for high momentum tracks is $\sim$40 μm in impact parameter and $\sim$70 μm along z direction [41]. All silicon detectors are used in the offline track reconstruction algorithms, while SVX plays a crucial role both for the online reconstruction and for triggering on B hadrons.
CDF employs an innovative processor SVT [43, 42] for online track reconstruction in the silicon detector. The SVT was upgraded [43] to cope with the higher Tevatron luminosity. The SVT reconstruction is precise enough for identifying at the trigger level secondary vertexes of B hadron decays (displaced from the primary interaction point).

COT

Surrounding the silicon detector is the Central Outer Tracker (COT) [44]. It is a 3.1 m long cylindrical drift chamber cohaxial with the beam that covers the radial range from 40 to 137 cm ($|\eta| < 1$). The COT contains 96 sense wire layers, which are radially grouped into 8 *superlayers*, as inferred from the end plate slot structure shown in Fig. 2.7. Each superlayer is divided in ϕ into *supercells*, and each supercell has 12 alternated sense and field shaping wires. So within the supercell width the trajectory of a particle is sampled 12 times. The maximum drift distance is approximately the same for all superlayers. Therefore, the number of supercells in a given superlayer scales approximately with the radius of the superlayer. The entire COT contains 30,240 sense wires. Approximately half the wires run along the z direction (*axial*), the other half are strung at a small stereo angle (2°) with respect to the z direction (*stereo*). The combination of the axial and stereo information allows us to measure the z positions. Particles originated from the interaction point, which have $|\eta| < 1$, pass through all the 8 COT superlayers.

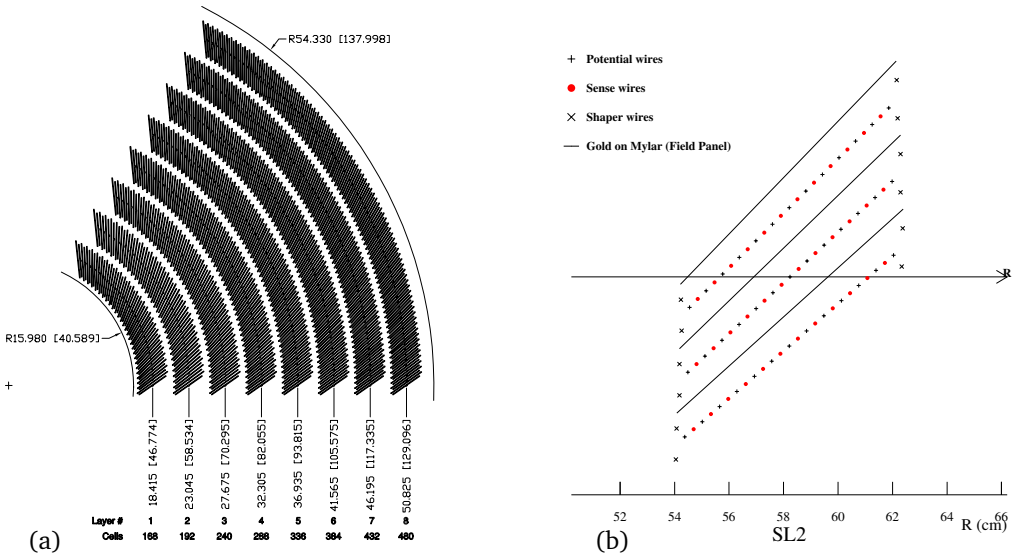

Figure 2.7: (a) Layout of wire planes on a COT endplate. (b) Layout of wires in a COT supercell.

The COT is filled with an Argon-Ethane gas mixture and Isopropyl alcohol (49.5:49.5:1). The mixture is chosen to have a constant drift velocity, approximately 50 μm/ns, across the cell width and the small content of isopropyl alcohol is intended to reduce the wire aging and carbon build up on the wires. When a charged particle passes through, the gas is ionized. Electrons are attracted toward the sense wires. However, due to the magnetic field that the COT is immersed in, electrons drift at a Lorentz angle of 35°. The supercell is tilted by 35° with respect to the radial direction to compensate for this effect and make the drift path perpendicular to the radial direction. The hit resolution in $r-\phi$ is about 140 μm and the transverse momentum resolution of the tracks in the COT chamber depends on the p_T and is measured to be $\sigma(p_T)/p_T^2 \approx 0.15\%$ (GeV/c)$^{-1}$ for tracks with $p_T > 2$ GeV/c [45].
In addition to the measurement of the charged particle momenta, the COT is used to identify particles based on dE/dx measurements.

2.3.3 Time of Flight

Just outside the tracking system, supported on the inside of the solenoid, see Fig. 2.4 (a), CDF II has a Time of Flight (TOF) detector [46]. It is a barrel of scintillator bars almost 3 m long located at 140 cm from the beam line with a total of 216 bars, each covering 1.7° in ϕ and the pseudorapidity range $|\eta| < 1$.
Light is collected by photomultipliers at both ends of each scintillator bar. Single hit position in the TOF is determined by the comparison of the signal times of the photomultipliers. Particle identification is achieved by measuring the time of arrival of a charged particle at the scintillators with respect to the collision time. Thus, combining the measured time-of-flight, the momentum and the path length the mass of the particle can be estimated.

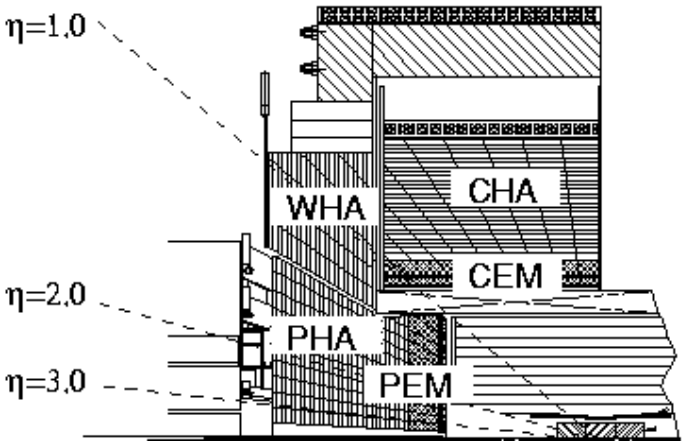

Figure 2.8: Elevation view of 1/4 of the CDF detector showering the components of the CDF calorimeter: CEM, CHA, WHA, PEM and PHA.

$$m = \frac{p}{c}\sqrt{\frac{1}{\beta^2} - 1} = \frac{p}{c}\sqrt{\left(\frac{cT}{L}\right)^2 - 1}$$

where the momentum p and the path length L are precisely measured by the tracking system. For the TOF measurement the collision time must be known and this is found with a 50 ps uncertainty by a best-fit process over all tracks in the event.
The resolution in the time-of-flight measurement is $\approx$ 120 ps and provides at least two standard deviation separation between $K^{\pm}$ and $\pi^{\pm}$ for momenta $p < 1.6$ GeV/c.

2.3.4 Calorimeter System

Surrounding the CDF tracking volume, there is the calorimeter system. The CDF calorimeter measures the particle energy by absorbing their total energy and providing a signal in principle proportional to it. The calorimeters are scintillator-based detectors segmented in projective towers (or wedges), in $\eta \times \phi$ space, that point to the interaction region. The total coverage of the system is 2π in ϕ and about $|\eta| < 3.64$ units in pseudorapidity. The calorimeter system is divided in two regions: central and plug. The **central** calorimeter covers the region $|\eta| < 1.1$ and is split into two halves at $|\eta| = 0$. The forward **plug** calorimeters cover the angular range corresponding to $1.1 < |\eta| < 3.64$, as it is shown in Fig. 2.8. Due to this structure two "gap" uninstrumented regions are found at $|\eta| = 0$ and $|\eta| \sim 1.1$. The two outer towers in one wedge are missing to allow accessing the solenoid liquid helium input gasket for checks and repairs if needed.

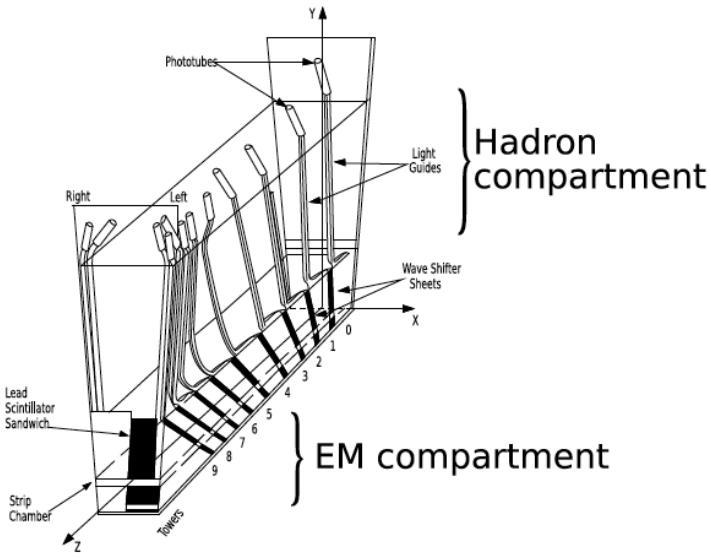

Figure 2.9: Structure of a wedge in the central calorimeters. The front compartment (lower section in figure) is the electromagnetic calorimeter CEM, while the rear one (upper section) is the hadron calorimeter CHA.

Central Calorimeters

The central calorimeters consist of 478 towers, each one is 15° in azimuth by about 0.11 in pseudorapidity. Each wedge consists of an electromagnetic component backed by a hadronic section. The light from each tower is read out through wavelength-shifting (WLS) bars or plates with light guides leading to two photo-multiplier tubes (PMT) per tower (see Fig. 2.9).

In the central electromagnetic calorimeter (**CEM**) [47], there are 31 layers of polystyrene scintillator interleaved with layers of lead. The total material has a depth of 19 X_0 [10]. CEM also includes two additional specialized detector: the Central Electron Strip Chambers (**CES**) and the Central Preshower (**CPR**), called also *shower maximum* and *preshower* detectors.

CES is a combined strip/wire gas proportional chamber embedded in CEM at about 6 X_0, where the maximum longitudinal development of the electromagnetic shower is expected. The CES purpose is to measure the position and the shape of electromagnetic showers in both transverse plane and longitudinal direction, which is used to distinguish electrons and photons from hadrons. CES resolution is about 1 cm in z and 1 mm in $r-\phi$.

CPR is a set of scintillator tiles located in front of the calorimeter wedges which helps distinguishing electrons from charged hadrons by gauging their probability of showering in the detector material prior to entering the calorimeter.

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

The central hadronic calorimeter (**CHA**) [48] surrounds the CEM covering the region $|\eta| < 0.9$ and consists of 32 steel layers sampled each 2.5 cm by 1.0 cm-thick acrylic 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 $0.7 < |\eta| < 1.3$, the wall hadronic calorimeter (**WHA**), which continues the tower structure of the CHA but with reduced sampling (each 5.0 cm). The total thickness of the hadronic section is approximately constant and corresponds to 4.5 interaction lengths (λ_0) [11].
The energy resolution for each section was measured in the test beam and, for a perpendicular incident beam, it can be parameterized as

$$\frac{\sigma}{E} = \frac{\sigma_1}{\sqrt{E}} \oplus \sigma_2,$$

where the first term comes from sampling fluctuations and the photostatistics of PMTs, and the second term comes from the non-uniform response of the calorimeter.
In the **CEM**, the energy resolution for high energy electrons and photons is

$$\frac{\sigma(E)}{E} = \frac{13.5\%}{\sqrt{E_T}} \oplus 1.5\%,$$

where $E_T = E \sin\theta$ being θ the beam incident angle.
Charge pions were used to obtain the energy resolution in the **CHA** and **WHA** detectors at the cell center, that are

$$\frac{\sigma(E)}{E} = \frac{50\%}{\sqrt{E_T}} \oplus 3\% \quad \text{and} \quad \frac{\sigma(E)}{E} = \frac{75\%}{\sqrt{E_T}} \oplus 4\%$$

respectively.

Plug Calorimeters

One of the major components upgraded for the Run II was the plug calorimeter [49]. The new plug calorimeters are built with the same technology as the central components and replace the Run I gas calorimeters in the forward region. The $\eta \times \phi$ segmentation depends on the tower pseudorapidity coverage. For towers in the region $|\eta| < 2.1$, the segmentation is 7.5^o in ϕ and from 0.1 to 0.16 in the pseudorapidity direction. For more forward wedges, the segmentation changes to 15^o in ϕ and about 0.2 to 0.6 in η.
As in the central calorimeters, each wedge consists of an electromagnetic (**PEM**) and a hadronic section (**PHA**). The PEM, with 23 layers composed of lead and scintillator, has a total thickness of about 21 X_0.
As for CEM, PEM is equipped with a *shower maximum* detector (**PES**), positioned at about

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

6 X_0, and with a *preshower* detector, the Plug PreRadiator (**PPR**) located at the inner face of the calorimeters. More details can be found in [50].
The PHA is a steel/scintillator device with a depth of about 7 λ_0. In both sections the scintillator tiles are read out by WLS fibers embedded in the scintillator. The WLS fibers carry the light out to PMTs tubes located on the back plane of each end plug. Unlike the central calorimeters, each tower is only read out by one PMT.
The **PEM** energy resolution for high energy electrons and photons is:

$$\frac{\sigma}{E} = \frac{16\%}{\sqrt{E_T}} \oplus 1\%.$$

The **PHA** energy resolution, for charged pions that do not interact in the EM component, is:

$$\frac{\sigma}{E} = \frac{80\%}{\sqrt{E_T}} \oplus 5\%.$$

2.3.5 Muons System

Muons can be separated from other particles because their typical radiation length is $4{\cdot}10^4$ times greater than for electrons. High p_T muons at CDF interact in calorimeters as minimum ionizing particles (MIP) that loose only modest quantities of energy when passing through large amounts of material. To exploit the penetration power of muons, the CDF calorimeters are surrounded by muon detectors which signal crossing charged particles which were not absorbed in the calorimeters.
Four indipendent system are used to detect muons [51, 52]: the Central Muon Detector, **CMU**, the Central Muon Upgrade Detector, **CMP**, the Central Muon Extension Detector, **CMX** and the Intermediate Muon Detector, **IMU**. The coverage of the muon systems is almost complete in ϕ, except for some gaps, and spans in polar angle up to $|\eta| \sim 1.5$. The $\eta - \phi$ coverage of the Run II muon detectors is shown in Fig. 2.10.
Muon detectors share common features. They consists of stacks of rectangular drift chambers modules [12] composed of single-wire cells. Stacks are four layers deep with laterally displaced cells from layer to layer to compensate for cell edge inefficiencies. z and ϕ coordinates of the muon candidate are provided by the chambers. The difference in drift-electron arrival-times between neighbor cells provides a typical resolution of 250 μm for the hit position in the transverse plane. Charge division at the wire ends measures the z coordinate with a 1.2 mm resolution. Muon momenta are measured in the tracker.
Chambers are coupled with scintillator counters in order to suppress backgrounds due to secondary interactions in the beam pipe material and to cosmic rays.
A muon candidate is reconstructed when a short track segment in the muon chambers corresponds to the extrapolation of a COT track.

[12] chambers are filled with a mixture of argon and ethane (50% each)

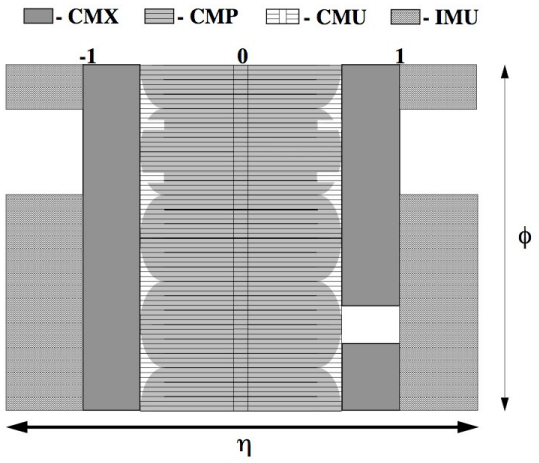

Figure 2.10: The η/ϕ coverage of the muon system. The shape is irregular because of the obstruction by systems such as cryo pipes or structural elements.

- **CMU** : The Central Muon Chambers (CMU) [37] is a set of four layered drift chamber sandwiches, housed on the back of wedges inside the central calorimeter shells covering the region $|\eta| < 0.6$. It is approximately cylindrical in geometry with a radius of 350 cm, arranged into 12.6° wedges. Each wedge contains three modules (stacks) with four layers of four rectangular drift cells. The cell are 266 cm x 2.68 cm x 6.35 cm wide and have 50 μm sense wire at the center of the cell, parallel to the z direction. The system is filled with Argone-Ethane gas mixture and alcohol like the COT.

- **CMP** : The Central Muon uPgrade (CMP) consists of a 4-layer sandwich of wire chambers operated in proportional mode covering most of the $|\eta| < 0.6$ region where it overlaps with CMU (see Fig. 2.10). It is located outside an additional layer of 60 cm thick steel partially used for the magnetic field return, providing the needed shielding to absorb particles, other than muons, leaking on the back of the calorimeter. Unlike mostly of the CDF components, this subdetector is not cylinrically-shaped but box-like, because CMP uses the magnet return yoke steel as an absorber, along with some additional pieces of steel to fill gaps in the yoke. On the outer surface of CMP a scintillator layer, the Central Scintillator Upgrade (CSP), measures the muon traversal time. The system CMU/CMP, which is called CMUP, detects muons having a minimum energy of ~ 1.4 GeV.

- **CMX** : The muon extension CMX is a large system of drift chambers-scintillator sandwiches arranged in two truncated conical arches detached from the main CDF detector to cover the region $0.6 < |\eta| < 1.0$. Due to main detector frame structure, some regions of this subdetector are characterized by their peculiar geometry. Two layers of scintillators are mounted on the internal and external sides of the CMX, the so-called central muon extension scintillator, CSX.

- **IMU** : Muons in a more forward region at $1.0 < |\eta| < 1.5$ are detected by the Intermediate Muon Extensions (IMU) on the back of the Plug Calorimeters. The IMU incorporates two scintillator systems: the barrel scintillator upgrade, BSU, and the Toroid Scintillator Upgrade, TSU.

2.3.6 CLC and Measurement of the Luminosity

Absolute luminosity measurements by the machine based on measurements of beam parameters have uncertainties of the order of 15-20%.
For this reason in CDF, the beam luminosity is determined using gas Cherenkov counters (CLC) [53] located in the pseudorapidity region $3.7 < |\eta| < 4.7$, which measure the average number of inelastic interaction per bunch crossing.
Each module consists of 48 thin, gas-filled, Cherenkov counters. The counters are arranged around the beam pipe in three concentric layers, with 16 counters each pointing to the center of the interaction region. The cones in the two outer layers are about 180 cm long and the inner layer counters, closer to the beam pipe, have a length of 110 cm.
The Cherenkov light is detected with photomultiplier tubes and the momentum threshold for light emission is 9.3 MeV/c for electrons and 2.6 GeV/c for pions.
The number of $p\bar{p}$ interactions in a bunch crossing follows a Poisson distribution with mean μ, where the probability of empty crossing is given by

$$P_0(\mu) = e^{-\mu},$$

which is correct if the acceptance of the detector and its efficiency were 100%. In practice, there are some selection criteria, α, to define an "interaction". An "interaction" is defined as a $p\bar{p}$ crossing with hits above a fixed threshold on both sides of the CLC detector. Therefore, an empty crossing is a $p\bar{p}$ crossing with no interactions. Given these selection criteria, the experimental quantity P_0, called $P_0^{exp}\{\alpha\}$, is related to μ as:

$$P_0^{exp}\{\mu;\alpha\} = (e^{\epsilon_\omega \cdot \mu} + e^{-\epsilon_e \cdot \mu} - 1) \cdot e^{-(1-\epsilon_0)\cdot\mu},$$

where the acceptances ϵ_0 and $\epsilon_{\omega/e}$ are, respectively, the probability to have no hits in the combined east and west CLC modules and the probability to have at least one hit exclusively in west/east CLC module. The evaluation of these parameters is based on Monte Carlo simulations, and typical values are ϵ_0=0.07 and $\epsilon_{\omega/e} = 0.12$.
From the measurement of μ we can extract the luminosity. Since the CLC is not sensitive at all to the elastic component of the $p\bar{p}$ scattering, the rate of inelastic $p\bar{p}$ interactions is given by:

$$\mu \cdot f_{bc} = \sigma_{in} \cdot \mathscr{L},$$

where f_{bc} is the bunch crossings frequency at Tevatron and σ_{in} is the inelastic $p\bar{p}$ cross section. $\sigma_{in} = 60.7 \pm 2.0$ mb, is obtained by extrapolating the combined results for the inelastic $p\bar{p}$ cross section of CDF at $\sqrt{s} = 1.8$ TeV and E811 measurements at $\sqrt{s} = 1.96$ TeV [54].
Different sources of uncertainties have been taken into account to evaluate the systematic uncertainties on the luminosity measurement [55]. The dominated contributions are related to the detector simulation and the event generator used, and have been evaluated to be about 3%. The total uncertainty in the CLC luminosity measurements is 5.8%, which includes uncertainties on the measurement (4.2%) and on the inelastic cross section value (4%).

2.4 Silicon Detector: Performances and Aging Studies

Performances of the Silicon Detectors has been vital to the success of the CDF's physics program, which includes a wide variety of precision electroweak measurements, bottom and charm physics, studies of the strong interaction as well as searches for new physics. These physics goals benefit from a high-resolution tracking and vertexing. For example, efficient identification of jets originating from b-quarks, by detection of displaced secondary vertices, which is crucial for studies of the top quark and searches for the Higgs boson, would be impossible without information provided by the finely-segmented silicon detector surrounding the interaction region.
The CDF silicon vertex detector was designed to withstand 2-3 fb^{-1} of accumulated $p\bar{p}$ collisions, with an upgrade planned thereafter [41]. The cancellation of the replacements for the Run IIb upgrade forced the experiment to keep the current detector much longer. Later delays in the start-up of the LHC experiments has led to unexpected extension of data taking, with an expected total delivered integrated luminosity of 12 fb^{-1} or more by September 2011.
The silicon detector at the CDF Experiment in Run II has been collecting data steadily in a hard radiation environment for the last 10 years. This has led to the presence of some aging effects, which have been monitored in order to preserve the detector performance and to predict the useful lifetime of the sensors.
Several preventive measures were taken to keep the original detector operational, the aging under control and the detectors fully functioning as long as possible.
Issues arising from radiation damage of the sensors, aging infrastructure and electronics are continually being addressed. In addition, there are basic challenges posed by the inaccessibility of the detector volume and large number of readout channels (about 750 thousand).
We will describe the response of the detector to accumulated radiation doses and performance of the silicon detector, introducing some parameters used for monitoring radiation exposure effects.

2.4.1 Effects of radiation exposure

The CDF Run II Silicon Detector is located is exposed to a radiation dose of 300 ± 60 kRad/fb^{-1} at a radius of 3 cm [56]. The effects caused by radiation are generally classified in

- **surface damage** caused by the Ionizing Energy Loss that is characterized by an accumulation of charge in the SiO_2 and Si/SiO_2 interface, which increases inter-strip capacitance of silicon sensor due to charge spreading.
- **crystal damage** which arises in displacement damage and defects in the lattice structure of the sensor, caused by Non-Ionizing Energy Loss. This generates an increase of noise, changes the effective doping concentration of the bulk and increases the charge carrier trapping. It represents the main concern for detector longevity.

From the operational point of view, these effects translate into performance degradation due to the increase of noise from the sensors and a decrease of charge collection efficiency.
Bias current is expected to increase with received radiation dose because radiation creates defects in the silicon lattice. An increase in bias current causes an increase in noise and reduces the signal-to-noise ratio. Thus, measuring the bias current is an important method for estimating the lifetime of a silicon detector.
Change in leakage current has been shown [57] to increase linearly with the absorbed dose,

$$\Delta I_{leak} = \alpha \kappa \Phi$$

where α is the damage factor, Φ is the particle fluence and κ is a constant depending on the source of non-ionizing radiation.
Measurement of the radiation dose in various points of the detector was done early in Run II using more than 1000 thermal luminescent dosimeters (TLDs)[13] spread over the entire detector volume. From these measurements the radiation field at any point is extracted and compared the result to bias current data, so the value of the damage constant α was extrapolated by the best fit.
Using a 95 pb^{-1} data sample collected in 2004 a damage factor of 1.65 ± 0.12 was extracted from bias current data [58].

2.4.2 Studies of the full depletion voltage

Depletion voltage is the bias voltage required to get rid of free carriers in the bulk of the detector. The expected evolution depends on the dose, precisely the full depletion voltage increases linearly with the integrated dose of radiation (data are well modeled by the phenomenological *Hamburg Model*, see [59] for more details).

[13] The Thermo Luminescent Dosimeter measures ionizing radiation exposure by measuring the amount of visible light emitted from a crystal in the detector when the crystal is heated. The amount of light emitted is dependent upon the radiation exposure.

To keep the detectors fully depleted is fundamental in order to reduce the noise of the sensor to a minimum, while enhancing the collection efficiency of charge induced by ionization.
As radiation damage increases, effects induced on the sensor lead to the *type-inversion process*, which changes the effective doping concentration of the bulk and mutates behavior from n-type of doped bulk to an effective opposite p-type. This change occurs in a continuous way as the sensor is exposed to radiation and can be detected by a constant monitoring the depletion voltage, which decreases until the type inversion is complete.
The measurement of the full depletion voltage has been performed at several times for each sensor through the *signal-to-bias* scan technique, which consists of measuring the charge collected for traversing particles as a function of the bias voltage.
Tracks traversing the silicon layer under study are extrapolated from the COT to the silicon sensors. The charge collected by the sensor increases by increasing bias voltage up to a saturation point.
Taken a set of bias voltage, for each of them a distribution of charge is measured. Through the interpolation with a Landau-Gaussian function it is possible to extract the most probable value of the fitted curve. That value is considered as measurement of the collected charge.
The charge collection as a function of bias voltage is well parametrized with a sigmoid function, and from the interpolation of the points one can extract the depletion voltage: it is defined as the bias voltage which correspond to a charge collection equal to 95% of the asymptotic value at the efficiency plateau, as shown in Fig. 2.11.
The evolution of the depletion voltage with integrated luminosity facilitates quantitative studies on the aging of specific sensors. The measured values at different times are fitted with a third-degree polynomial to extract the inversion point and with a straight line to obtain the behavior after inversion and the expectations for the future.

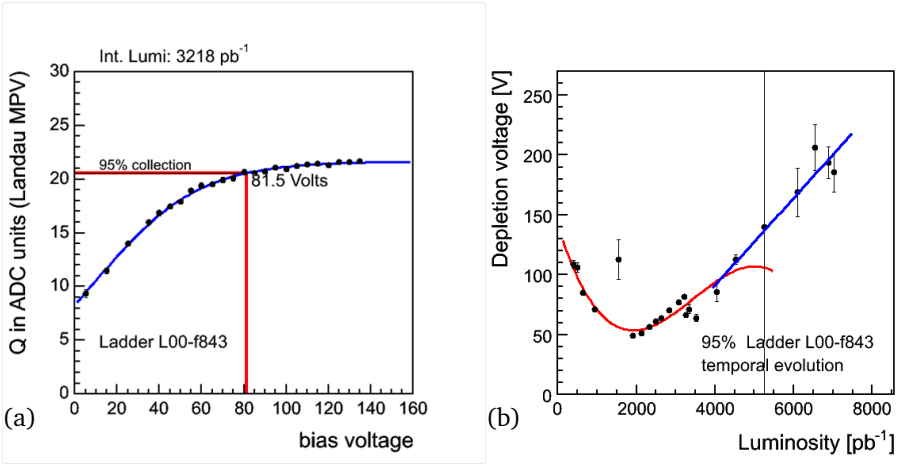

Figure 2.11: Example of determination of the full depletion voltage in a particular sensor (a). Evolution of the full depletion voltage of a sensor in CDF Layer-00 detector as a function of the integrated delivered luminosity (b) [41].

Fig. 2.12 (a) shows the prediction based on the linear fits of the depletion voltage for sensors in L00, while Fig. 2.12 (b) for sensors of SVX-L0. As regards L00 the predicted depletion voltages lie below the high voltage power supply limit of 500 V and the sensor breakdown region. The majority of the sensors is expected to be working properly for delivered luminosity over 12 fb^{-1}.

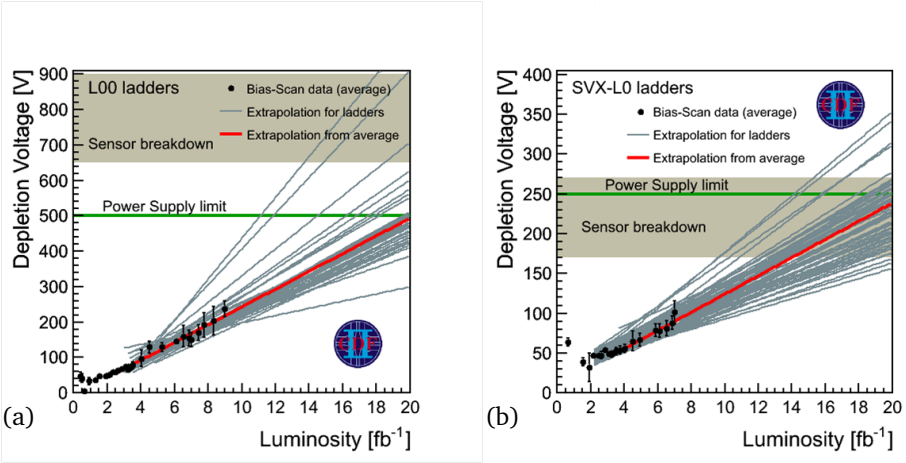

Figure 2.12: Summary of the extrapolation fits of L00 (a) and SVX-L0 (b) depletion voltage using data until 6.9 fb^{-1} [41]

As regards SVX-L0, while the depletion voltages for all sensors will remain below the breakdown region through 10 fb^{-1} of delivered luminosity, many of them will likely not be able to be fully depleted for higher values of integrated luminosity. This may eventually limit our ability to obtain useful data from this layer.
If we consider outer layers of the silicon detector the increase of the depletion voltage due to the radiation dose does not represent a limit. We expect to be able to fully deplete all sensors for delivered luminosity over 12 fb^{-1}.

2.4.3 Signal-to-Noise ratio

Another variable of interest for monitoring the CDF silicon detector performances is the study of the evolution of signal over noise ratio. The monitoring of sensor aging does not provide direct information on how the aging of the sensors is impacting the physics analysis for track reconstruction. For evaluating it, the most common variable is the ratio of the most probable value of the charge collected in a particular sensor over the noise. These studies provide estimates of usability of the detector in charged particle tracking and in turn for physics analyses.
Signal-to-noise ratio expected to decrease over time.
The signal is determined from $J/\Psi \rightarrow \mu^+\mu^-$ tracks, which provide a clean track sample, while noise is estimated from special calibrations runs with non interacting beams.

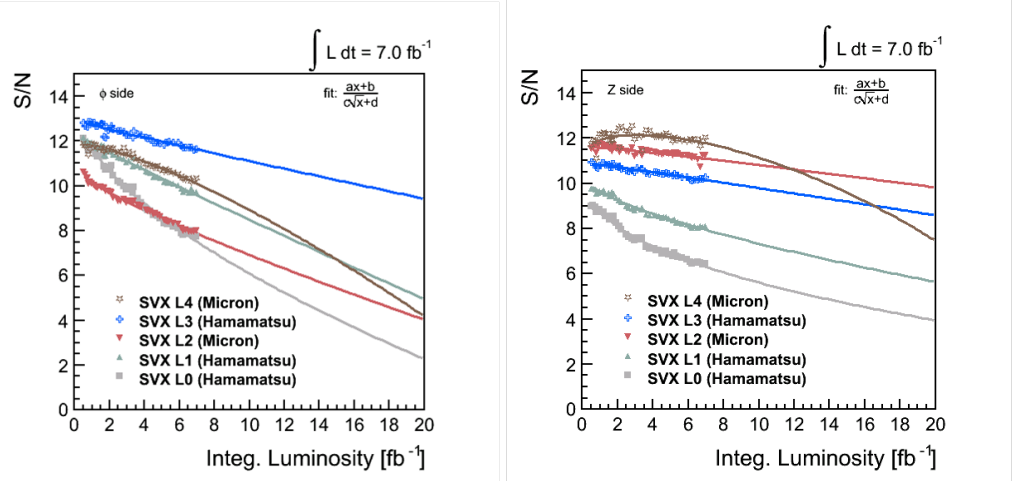

Figure 2.13: Evolution of the S/N of all the layers of the SVX detector of CDF as a function of the integrated luminosity [41].

Plots in Fig. 2.13 and Fig. 2.14 shows on average the evolution of this variable for all the sensors belonging to the layers of SVX detector in relation with the integrated delivered luminosity. The lines interpolating the points in Fig. 2.13 show the expected values for S/N in the future. It has to be taken in account that the predicted values are valid in a condition of fully depleted sensor, otherwise the signal value is expected to decrease. The evolution related to L00 is showing a knee that can be explained by the sensor being slightly underdepleted in the earlier period.

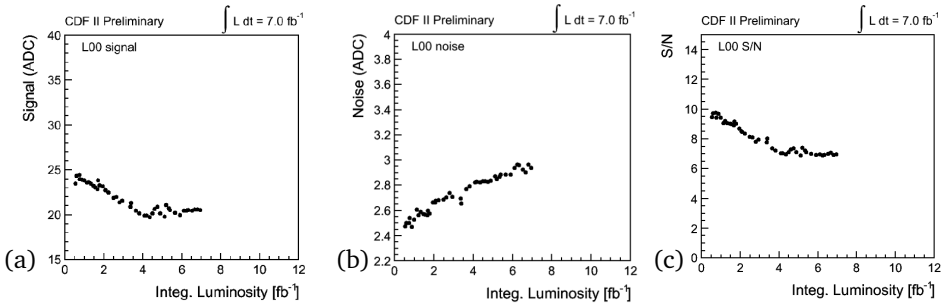

Figure 2.14: Evolution of the Signal (a), Noise (b) and S/N (c) of the sensors in L00 component as function of the integrated luminosity [41].

2.4.4 Final Comments on Silicon Radiation Aging

The CDF Run II silicon detector have been exposed to large radiation dose since 2001, much beyond what was originally planned. For this reason, the radiation damage is continuously monitored in order to keep the detector working as well as possible.
As a matter of fact, the CDF silicon detector has provided at all times precise measurements of the trajectories of charged particles important to identify and measure heavy flavored hadrons which are crucial to CDF's physics program, including B hadron, top quark and Higgs boson physics.
In particular, the detector is still showing good performance even if the inner layers have since long passed through effective doping inversion. Most of the sensors in L00 and SVX-L0 are expected to operate with high efficiency until the delivered luminosity will approach approximately 12 fb^{-1}.
Even after DOE announced that Tevatron will be closed in September 2011, aging studies preserve their attractive in the interest of designing new experiments, since these studies on radiation damage are performed in a very unique situation, in which the sensors are strongly irradiated during a long period of time. LHC experiments will eventually face the aging process of their silicon detectors and Tevatron experiments experience can provide an useful indication for designing their upgrades. In fact all the CMS and ATLAS strip silicon sensors were fabricated with the same technology as L00.
Simulations of radiation effects in silicon detectors can predict the lifetime of detectors not yet built and also suggest new fabrication methods to improve resistance to radiation damage. One can easily predict that the LHC experiments will make extensive use of CDF data to tune their simulations.

2.5 Trigger and Data Acquisition

For equally spaced 36 on 36 bunches the average interaction rate at the Tevatron would be 1.7 MHz. The actual interaction rate is higher because the beams circulate in three trains of 12 bunches spaced 396 ns, which leads to a crossing rate of 2.53 MHz. This interaction rate is orders of magnitude higher than the maximum rate that the CDF data acquisition system can handle. Furthermore, most collisions are of no interest. This imposes implementing a trigger system that preselects events online and decides if the information is written to tape or discarded. The identification of the interesting events is accomplished by dedicated fast online electronics, called the Trigger System, which evaluates the information from the detector in real time. The CDF trigger system is organized in three trigger levels (see Fig. 2.15). The first two levels are hardware based while the third one is software filter run on a processor farm. The decisions taken by the system are based on increasingly more complex event information. Each level of electronics performs a slower but more accurate event reconstruction and filter with increasing selectivity according to a set of predefined conditions.

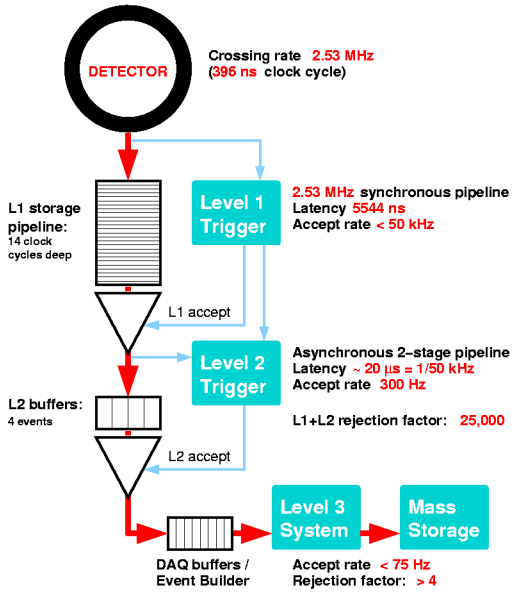

Figure 2.15: Block diagram showing the global trigger and DAQ systems at CDF II.

LEVEL 1 is a synchronous system with event reading and a decision made every beam crossing. In a synchronous pipeline up to 42 subsequent events can be stored for ∼5.5 μs while the hardware is taking a decision. If no acceptance decision is made within that time the event is lost. L1 decision are made in average in about 4 μs: no dead time is expected from this level.
Level 1 rejects 97% of the events, it reduces the event rates from 2.53 MHz to less than 50 kHz. The L1 decision is generated by:

- XFT (extremely fast tracker), which reconstructs approximate tracks (p_T >1.5 GeV) in the transverse plane by exploiting information from COT superlayers. These tracks are extrapolated to the other detector parts to contribute to all trigger levels.
- the calorimeter trigger, which indicates large energy releases in single electromagnetic or hadronic cells (these can be seed for electrons or jets identification).
- the muon trigger, which matches XFT tracks segments in the muon chambers.

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

After the Level 2, the event rate is reduced to about 300 Hz (rejection factor ∼150). L2 purposes are:

- to add the energy deposited in the towers in small regions around L1 seeds, as an approximate measure of an electron or jet energy.
- to use calorimeter and CES chamber information to improve separation of $e^{\pm}$ from γ
- to reconstruct a full COT track and associate it to an outer muon stub in order to improve muon signature.
- to feed startup information to the Silicon Vertex Trigger (SVT) [42, 43] which generates triggers on secondary vertexes from decay of long-lived heavy flavour hadrons.

LEVEL 3 is a software trigger. L3 addresses event objects delivered by L2 to the Event Builder (EVB), which reconstructs the entire event with the same accuracy as in the offline analysis.
The final decision to accept an event is made on the basis of a list of observables indicating candidate events of physical interest (top quark production events, W/Z events, Drell-Yan events, etc.).
Events that satisfy the Level 3 trigger requirements are then transferred onward to the Consumer Server/Data Logger (CSL) system for storage first on disk and later on tape. The average processing time per event in Level 3 is on the order of one second. The Level 3 leads to a further reduction in the output rate, of about 50 Hz in average, with an accepted maximum of about 120 Hz.

A set of requirements that an event has to fulfill at Level 1, Level 2 and Level 3 constitutes a *trigger path*. The CDF II trigger system implements about 150 trigger paths, which are periodically adjusted depending on machine luminosity and physics needs. A typical list of the datasets at CDF Run II can be found in [61].
The trigger system described above exploits the information of all detector subsystem. Combining the measurements of the various subsystems it is possible to efficiently record, at the same time, events characterized by different signatures.
Triggers which occupy most of the band width can be *dynamically prescaled* (DPS). A trigger path is said to be prescaled by a factor N if it is configured to accept only one event each N accepted events. Prescaling is dynamically implemented by luminosity-dependent factors during data taking. This is important in order to ensure that no trigger path reaches rates so high as to create unacceptable dead time to triggers on rare events of primary importance.
During data taking the luminosity decreases with time, and consequently a number of prescale factors can be relaxed. The prescale factors decrease proportionally to the rate of triggered events, so as the number of recorded events is constant. Using dynamic prescaling ensures that optimal use is made for physics of the available luminosity.

2.6 Consumer Server/Logger

Once an event is accepted by the trigger, it must be recorded. This is managed by a software package named *Consumer Server and Logger* (CSL), which categorizes events by the triggers they fired and writes them to hard disk, while using a small set for online monitoring. The data on disks are then copied to tapes to be processed by offline reconstruction algorithms and made usable for physics analyses.

Chapter 3

Objects Identification

Outgoing particles from $p\bar{p}$ interactions are identified using the information provided by the CDF sub-detectors described in the previous chapter. The raw outputs form the CDF detector are electronic signal recorded by the hardware components which must be converted into physical information.
From the raw data, high level objects (such as tracks, vertices, calorimeter clusters) are reconstructed and combined to identify physical objects (electrons, muons, neutrinos and jets) of interest for the analysis.
The ability to detect and reconstruct charged particle trajectories is essential for particle identification and momentum measurement.
Precise, high efficiency tracking plays a central role for charged particle identification and separation. Photons give electron-like signals in the electromagnetic calorimeter but have no associated track, and can be identified by this signature.
In addition, as mentioned in the previous chapter, track reconstruction in the silicon tracker allows precise measurement of track impact parameter and thus the identification of heavy-flavor hadrons in jets, which is a fundamental tool for the light Higgs search.

3.1 Track Reconstruction

In a uniform axial magnetic field, neglecting energy loss by ionization in the detector material, the trajectory of a charged particle is an helix (see Fig. 3.1). The helix is completely described by five parameters (see subsection 2.3.1), i.e. every point along the trajectory satisfies the following equations:
The equations describing the helix are:

$$x = r\sin\phi - (r - d_0)\sin\phi$$
$$y = r\cos\phi - (r - d_0)\cos\phi$$
$$z = z_0 + s\lambda$$

where s is the length projected along the track, and $\phi = 2Cs + \phi_0$.

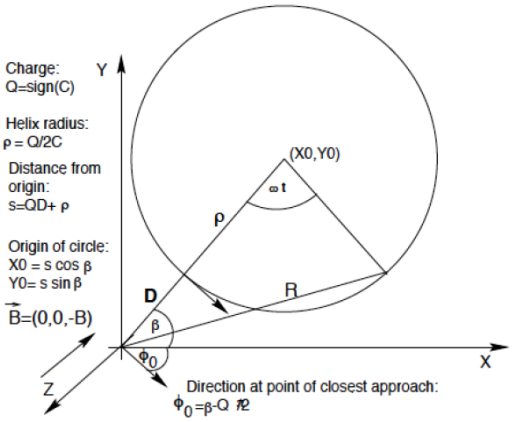

Figure 3.1: The schematic trajectory of a positive charged particle.

3.1.1 Tracking Algorithms

Charged particles leave small charge depositions as they pass though the layers of the tracking system (COT and silicon detectors). Using these depositions, pattern recognition algorithms can reconstruct the particle original trajectory measuring the five parameters of helix that best match the hits.

A track pattern recognition algorithm searches among the hits in the tracking system for those that can be associated to the same track. Then a track fitting algorithm use those hits to reconstruct the track and its parameters. The helical fit takes into account field non-uniformities and scattering in the detector material.

The experiment exploits several tracking algorithms [63], each optimized for the information available in different detector regions. The most used tracking algorithms are three: *Stand-alone* (SiSA), *Inside-out* (IO) and *Outside-in* (OI). A special algorithm named *Phoenix* is used for tracking electrons at small angles.

Outside-In Algorithm

The OI algorithm is the default CDF tracking algorithm. COT tracks are used as input and the available SVX hits are linked. This algorithm allows tracking in $|\eta| < 1$ region.

Track pattern recognition starts in the COT outer layers, where the hit density is smaller and proceeds through four steps.

The *first step* is the segment finding: each of the 8 COT superlayers is searched for groups of three aligned hits, which are fitted to a straight line with the least squares method. Then, tracks are reconstructed (*second step*) from the information of the axial superlayers only, which are linked by the standard CDF linking algorithms[1].

[1] At CDF two algorithms have been implemented: “segment linking" and “histogram linking" [62]

During the *third step* in COT track reconstruction stereo hits are added by the linking algorithms and searches for the track point of origin are performed. The *fourth* and last step of COT tracking is a global refit of the track taking into account all known corrections for the non-uniformity of the magnetic field and for the modeling of the electron drift.
The track found in the COT [64] is propagated inwards into the silicon system. A road around a track is defined using the uncertainties on track parameters and silicon hits are added if they lie inside this predefined road. When a hit is added, the track parameters are recalculated and the hit search is repeated. The accepted track has the largest number of hits. If this criterion is satisfied by more than one track, the one with the lowest χ^2 is selected. The impact parameter resolution of COT + SVX tracks is found to be $\sigma_{d_0} \sim 20$ μm.

Silicon-Stand-Alone Algorithm

SiSA algorithm uses only SVX information. The hits in silicon subdetectors not used by OI tracking are available [65]. The region $|\eta| < 2$ is covered with some residual efficiency up to $|\eta| \sim 2.8$.
The SiSA algorithm starts from collecting at least four hits in the SVXII detector in the $r-\phi$ plane and fits the C, d_0 and λ parameters to obtain a projection of the helix on the transverse plane to the beams. Then the algorithm creates a 3-D seed track adding small angle hits and event primary vertex information.
At this point the 90 stereo silicon hits are added and a global refit is performed. SiSA tracks reconstructed only with SVXII have a poor resolution for high pT tracks so hits are searched in L00 and ISL with the SVXII track as seed. The track is refit if information from these layers is added. Since, SiSA performances on momentum and impact parameter resolution are limited, SiSA tracks are not used for secondary vertexing.

Inside-Out Algorithm

The IO Algorithm, [66], tries to recover efficiency and p_T resolution in the region $1.2 < |\eta| < 1.8$ where the COT coverage is poor. SiSA tracks used as seeds are extrapolated to the COT inner cylinder. Matching hits in the available COT layers are added and the track is refitted. Possible duplicates are removed.

Phoenix Forward Electron Tracking

Phoenix Forward Electron Tracking [67] combines calorimetric information and standalone silicon tracks to track electrons in the forward region, $(1 < |\eta| < 3.6)$ which is not covered by the COT. The Phoenix algorithm reconstructs the track of an electron by using:

1. the reconstructed 3D point of the $p\bar{p}$ interaction
2. the 3D position of the EM shower estimated by PES.
3. the transverse energy deposited in the calorimeter.

The position of the EM shower in PES and the primary event vertex are used as seed points of the track helix and the helix curvature is varied to match the calorimeter E_T. If a match is found, available hits from SVXII are added to allow for a better fit. The track is called *Phoenix electron* (**PHX**).

3.2 Primary Vertex Identification

Precise determination of the primary vertex is important to signal displaced secondary vertices and to properly correct jets energy. The location of the primary $p\bar{p}$ vertex is required to compute transverse energies in the calorimeter towers and to build objects such as jets and missing energy.
The position of the interaction point, the so-called “primary vertex" of the event, is reconstructed using track information by two finding algorithms: *PrimVtx* and *ZVertex*.
The primary vertex location for a given event is found by fitting high quality tracks to a common point of origin. At high luminosities, multiple collisions frequently occur in the same bunch crossing. For example, with a luminosity of 10^{32} cm^{-2}s^{-1}, there are in average of 2.3 interactions per bunch crossing. Typically, since the luminous region is sufficiently long (with $\sigma_z = 29$ cm), the primary vertices of each collision are separated in z enough to allow isolating the vertex of the large p_T collision of interest.

PrimVtx uses a seed vertex calculated as the average event z position of all tracks passing pre-defined quality requirements, which are: $|z_{trk} - z_{vtx}|$ <1 cm, $|d_0| < 1$ cm and $|d_0/\sigma_{d_0}| < 3$. Then they are ordered in decreasing P_T and fitted to a common vertex. The tracks with $\chi^2 > 10$ after the fit are removed and the procedure is iterated until all accepted tracks have $\chi^2 < 10$.

ZVertex selects vertices from tracks passing minimal quality requirements and removes those without a minimal associated number of tracks with P_T >300 MeV[2]. The z position of each vertex is the mean z_0 of its associated tracks, weighted on its σ_{z_0}.

3.3 Electron/Photon Identification

High momentum electrons, photons, and jets deposit energy in small contiguous groups of calorimeter towers which can be identified as energy clusters.
The physical space corresponding to the calorimeter towers is mapped in an $\eta - \phi$ plane, the CDF EM algorithm [67] creates two lists of towers ordered by decreasing energy revealed on them: the *usable* list (towers with E > 100 MeV) and the *seed* list (towers with E > 2 GeV). After, the algorithm takes the first seed tower and create a cluster by adding the adjacent

[2]A track is associated to a vertex if its distance from the vertex is less than 1 cm (silicon vertex) or 5 cm (COT vertex).

towers, with energy greater than a lower threshold, to form a 2×2 or 3×3 $\eta - \phi$ area.
The position of the cluster is defined by the energy-weighted position mean of the towers in the cluster. Usually, 3×3 clustering is used in the CEM region while 2×2 clusters are used in the PEM region, because this reduces the probability to overlap the clusters of two different electrons. Also, the energy measured in the shower max (PES) and pre-shower (PPR) detectors is added to the final reconstructed energy.
Then, the used towers are removed from the lists and the algorithm selects the next seed tower and iterates the process until all the seed towers have been used.
Several corrections are applied to reconstruct the true energy of the EM object. Clusters are corrected for lateral leakage and a calibration is applied according to the response curve drawn by the test beam data.
After the electromagnetic cluster reconstruction, the algorithm searches for a track to be associate to the cluster in order to obtain the complete electron reconstruction. From all the tracks reconstructed in the COT which point to an instrumented region of the CEM detector, the one with the highest p_T that best matches the cluster is chosen as the geometrical center of the main cluster .
The track's momentum is used to define the direction and the transverse momentum of the electron while its energy is total energy of the calorimeter cluster.
CDF uses standard requirements optimized to efficiently select high p_T electrons [68], briefly described in the following. In Tab. 3.1 are given the specific quantitative values for the parameter employed to select the *Tight Central Electrons* (**TCE**), which are used in this analysis.

- The ratios between the hadronic and the electromagnetic cluster energies ($\mathbf{E_{EM}/E_{HAD}}$) and between the total cluster energy and the track momentum ($E = p$) are required to be consistent with a typical electron's energy deposition in the calorimeters as measured in the test beam.
  The $E = p$ distribution is peaked at 1 but it has large tails because the electrons can radiate collinear photons in the passage through matter in the tracking volume. The EM energy measurement is weakly influenced by that (the photon usually deposits energy in the same EM cluster), but the momentum measurement decreases.
- The **shape** of the calorimeter cluster is required to be consistent with that of a typical electromagnetic shower. The variable used for this comparison is

$$L_{shr} = 0.14 \cdot \sum_i \frac{E_i - E_i^{EXP}}{\sqrt{(0.14)^2 E_i + (\Delta E_i^{exp})^2}}$$

  where E_i is the measured energy in the i-th tower, E_i^{exp} is the expected value for the energy in the i-th tower, according to a parametrization based on test- beam data, and the index i runs over all the towers adjacent to the seed one. L_{shr} is required to be less than 0.2.

- **Isolation** is a measurement of the calorimetric activity around the candidate electron. The leptons from decays of W are often isolated from hadronic jets. The isolation requirement is meant to reject leptons from semileptonic decays of heavy flavor hadrons and hadrons faking leptons. In these processes the lepton candidate is produced in a jet environment and tends to be less isolated then a lepton produced by a W boson decay. The isolation variable for electrons is defined in Appendix B.
- The track associated to the calorimeter cluster is required to have a large number of hits attached. At least three **axial** and two **stereo** superlayers must provide segments formed by at least five hits.
- The track is required to match a CES cluster in both axial ($|\Delta z| < 3$ cm) and azimuthal (-3< $Q \cdot \Delta X$ <1.5 where Q is the charge of the the candidate electron and X is the distance in the $r - \phi$ plane) directions, in order to remove possible hadron contamination.

All leptons in this analysis are required to be tight, i.e. to pass the isolation criterion; however, loose leptons are still used to model the QCD background.

Parameter	Requirement
Region	$\|\eta\| < 1.1$
Fiduciality	tracks fiducial to CES
Track p_T [3]	> 10 GeV/c
Track $\|z_0\|$	≤ 60 cm
# Axial SL	≤ 3
# Stereo SL	≤ 2
Iso/E_T	≤ 0.1
E/p	$\leq 2.5 + 0.015 \cdot E_T$
E_{HAD}/E_{EM}	$< 0.055 + 0.00045 \cdot E_T$
L_{shr}	< 0.2
ΔX	$-3 \leq Q \cdot \Delta X \leq 1.5$
$\|\Delta z\|$	< 3 cm

Table 3.1: TCE identification cuts [71].

A photon conversion veto is also implemented: an electron can be misidentified as a photon if the electron track is not reconstructed. Alternatively, a photon can be misidentified as an electron if the photon converts to an electron-positron pairs as it transverses the tracker, or if a track is mistakenly associated to the EM cluster [69]. Electrons from photon conversion are identified and rejected if another opposite-sign COT track is found close in space.

Parameter	Requirement
Track p_T [4]	> 10 GeV/c
Track $\lvert z_0 \rvert$	≤ 60 cm
Track d_0 [5]	≤ 0.2 cm
Iso/p_T	≤ 0.1
# Axial SL	≤ 3
# Stereo SL	≤ 2
E_{EM}	$\leq 2+\max\left[0,(p-100)\cdot 0.0115\right]$
E_{HAD}	$\leq 6+\max\left[0,(p-100)\cdot 0.028\right]$
CMU Fiduciality	$x < x_{fid}, z < z_{fid}$
CMP Fiduciality	$x < x_{fid}, z < z_{fid} - 3$ cm
CMX Fiduciality	$x < x_{fid}, z < z_{fid} - 3$ cm

Table 3.2: CMX and CMUP muons identification cuts[71].

3.4 Muons Identification

While electrons and hadrons lose all their energy and stop in the calorimeter, most muons deposit only a small fraction of their energy through ionization, interacting as *minimum ionizing particle* (m.i.p.). So, they reach the outer part of the detector, where muon chambers are located.
An algorithm fit the hits produced by the muon in the muon chambers, to a track segment which is called *stub*. The complete identification of a muon is composed by three requirements:

1. a charged-particle track reconstructed by the tracking algorithm
2. the track must point to a stub in a muon chamber
3. a m.i.p. deposit in the traversed calorimeters towers.

Real muons that are backgrounds for this analysis are cosmic rays and muons from decays in flight of kaons and charged pions or from heavy flavor semi-leptonic decays.
A veto, based on timing information from the muon chamber and the COT, is used to reject cosmic rays [70]. Muons can also be mimicked by hadrons that shower unusually late or not at all in the calorimeter ("punch-throughs").
In this analysis we define two types of muons according to the detector used to identify them: **CMX** muons, with a stub in the CMX system that covers the region $0.7 < |\eta| < 1$, and **CMUP** muons, with a stub in both the CMU and CMP systems with $|\eta| < 0.7$.
The extrapolation (x, z) of the track to the relevant muon detector is required to be fiducial to the detector. In the case of CMP and CMX muons it must also be within 3 cm in z_{fid} from the edge of the detector.
The observables and the requirements used to select the muons are summarized in Tab. 3.2, the muon isolation parameter is defined in Appendix B.

3.5 Jets Reconstruction

To describe parton confinement, a potential is introduced which diverges with increasing inter-parton distance (see section 1.2). The strong potential can be roughly parametrized as a function of the inter parton distance as

$$V(r) = -\frac{a}{r} + b \cdot r \qquad a,\ b > 0$$

With increasing r increases, since the production of new partons (eventually hadrons) is energetically less demanding than stretching the parton bound, partons undergo a cascade of gluon emissions and production of additional parton pairs. This process is named *hadronization*. Partons from colorless hadrons combine into stable particles which reach the detector material. The properties of a hadron jet are measured primarily by calorimeters[6]. The carried information on the energy-momentum of the primary parton can be degraded on three levels, for non-accounted physical reasons and for instrumental features of the detectors, as pictured in Fig. 3.2:

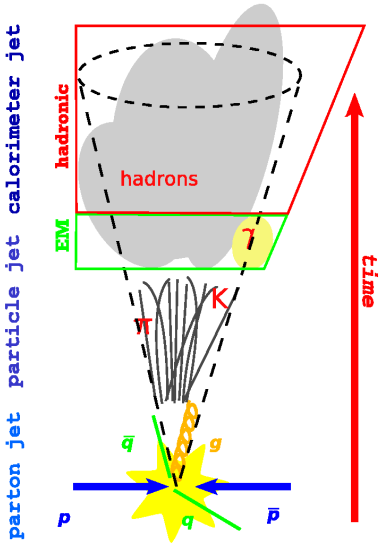

Figure 3.2: A parton originating from a hard scattering hadronizes and generates a jet.

1. at the **parton level**: whether the jet is generated by a quark or by a gluon is not known. This introduces an uncertainty because the fragmentation process may be parton-dependent;
2. the particles produced in the hadronization process define the **particle level** of the jet: the details of particle production in the fragmentation process are unknown;
3. the signals produced in the calorimeters correspond to the **calorimeter level**: the

[6]The jet energy is typically shared by 60% in EM and 40% in the hadron calorimeter.

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

From the experimental point of view, the resulting shower of particles appears as a large energy deposit in a localized area of the detector (see Fig. 3.2). The challenge of a physics analysis is to recover from the detector information the initial energy, momentum and, possibly, the nature of the parton produced in the original interaction. A *jet algorithm* is a tool to reconstruct such information. Several different jet reconstruction algorithms have been developed by various groups [72]. A number of key features that any algorithm must possess in order to be theoretically interpreted are mentioned in Appendix C.

3.5.1 Jet Cone Algorithm

CDF uses several algorithms, none of them really satisfying all the above requirements. The most common one is `JETCLU` [73], an iterative fixed cone jet reconstruction algorithm based only on calorimetric information.

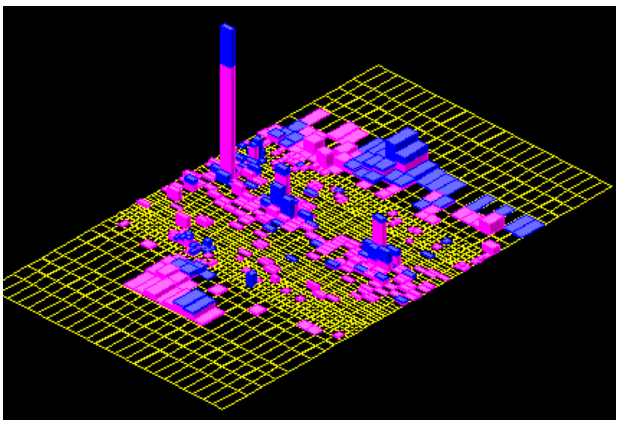

Figure 3.3: Transverse plane projection of an event with a reconstructed jets: calorimetric deposits are shown in red (EM) and in blue (HAD); the height of the signals is proportional to the deposited energy.

The algorithm begins by creating the list of seed towers with EM+HAD transverse energy above 1 GeV. Starting with the highest-E_T seed tower, a precluster is formed by combining together all adjacent seed towers within a cone of given radius R[7]. This procedure is repeated, starting with the next unused seed tower, until the list is exhausted. The E_T-weighted centroid is then formed from the towers in the precluster and a new cone of radius R is formed around this centroid.
All towers with energy above 100 MeV within this new cone are added to the cluster. Then, a new centroid is calculated from the set of towers within the cluster and a new cone drawn.

[7] CDF reconstructs jets using R= 0.4, 0.7 and 1.0.

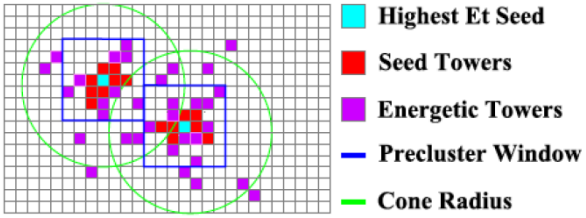

Figure 3.4: Jet reconstruction by the JETCLU algorithm.

This process is iterated until the centroid of the energy deposition within the cone is consistently aligned with geometric axis of the cone (stable solution).
Since each tower may belong to only one jet, in case of jet overlap two clusters are merged if the total energy of the overlapping towers is greater than 75% of the energy of the smaller cluster. If the shared energy is below this cut, the shared towers are assigned to the cluster that is closer in $\eta - \phi$ space.
This process is iterated again until the list of clusters remains fixed.
Massless four-vector momenta are assigned to the towers in the clusters for EM and HAD components with a magnitude equal to the energy deposited in the tower and the direction defined by a unit vector pointing from the event vertex to the centre of the calorimeter tower at depth that corresponds to the shower maximum[8].
A cluster four-vector is then defined summing over the towers in the cluster:

$$E = \sum_{i=1}^{N} \left(E_i^{EM} + E_i^{HAD} \right) \tag{3.1}$$

$$p_x = \sum_{i=1}^{N} \left(E_i^{EM} \sin\theta_i^{EM} \cos\phi_i^{EM} + E_i^{HAD} \sin\theta_i^{HAD} \cos\phi_i^{HAD} \right) \tag{3.2}$$

$$p_y = \sum_{i=1}^{N} \left(E_i^{EM} \sin\theta_i^{EM} \sin\phi_i^{EM} + E_i^{HAD} \sin\theta_i^{HAD} \sin\phi_i^{HAD} \right) \tag{3.3}$$

$$p_z = \sum_{i=1}^{N} \left(E_i^{EM} \cos\theta_i^{EM} + E_i^{HAD} \cos\theta_i^{HAD} \right) \tag{3.4}$$

where the index i runs over the towers in the cluster. Other variables are added to the recorded jet-object to be used in the analysis: E_T, η and ϕ (calculated from the jet vertex with an energy weighted average over the calorimeter towers associated with the cluster) or other useful information like the number of tracks reconstructed inside the jet cone, the vertex quality or the energy deposited in the HAD or EM calorimeter.

[8] Although CDF uses projective towers, the jets do not always originate in the geometrical center of the detector. As a consequence, the direction of the vector pointing to a tower may depend on the chosen depth along the tower.

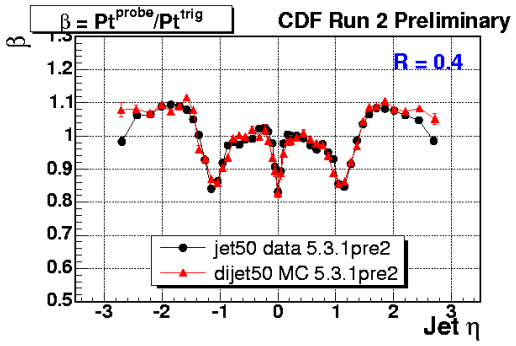

Figure 3.5: η-dependent energy scale correction factor for JETCLU with radius 0.4; a sample of events with at least one trigger tower above 50 GeV is used.

3.5.2 Jet Energy Corrections

The four-momentum assigned to a jet must be corrected to account for detector defects and reconstruction algorithm imperfections. In order to convert the measured transverse jet energy into the transverse energy of the partons, a set of corrections to the measured jet energy ("raw energy") have been developed.

The corrections, developed using data and simulation of the CDF detector, address the response inhomogeneity in η, the contributions from multiple interactions, the non-linearity of the calorimeter response, the contribution by the underlying event and the jet energy flow out of the jet cone.

Each of those corrections has a fractional uncertainty, $\sigma_{JES}(p_T)$ which can be parameterized as a function of the corrected transverse momentum of the jet p_T.

They are applied in a sequence of levels (of "L-levels"') in order to correct for each bias independently [74].

The correction can be parameterized as follows

$$p_T^{\text{parton}} = (p_T^{jet} \cdot C_\eta - C_{MI}) \cdot C_{Abs} - C_{UE} + C_{OOC} = p_T^{\text{particle}} - C_{UE} + C_{OOC},$$

where the terms are described below.

C_η: pseudorapidity-dependent correction (L1)

The L1 corrects for non-uniformities in calorimeter response along η. It is obtained by studying the p_T balancing in dijet events, which are selected in order to have one jet ("trigger jet") in the $0.2 < |\eta| < 0.6$ region (far away from detector cracks). The other jet, called "probe jet", is free to span over the entire $|\eta| < 3$ region.

Since in a perfect detector the two jets must be balanced in p_T, a balancing fraction is formed

$$f_b = \frac{\Delta p_T}{p_T^{ave}} = 2 \cdot \frac{p_T^{probe} - p_T^{trigger}}{p_T^{probe} + p_T^{trigger}}$$

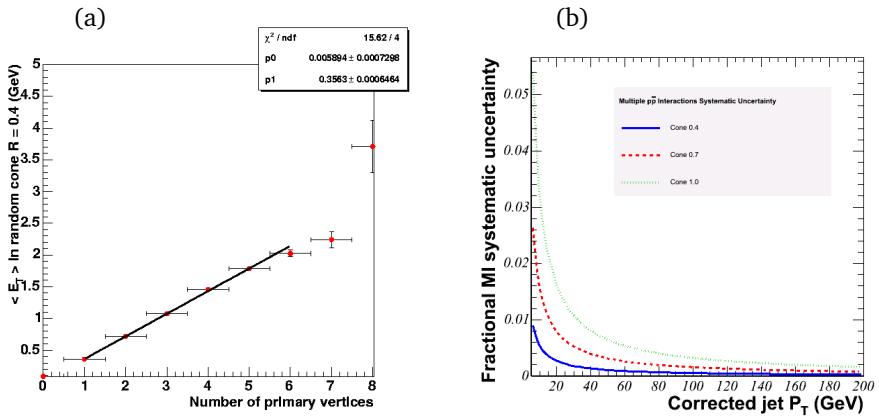

Figure 3.6: (a) E_T in R=0.4 cone as a function of the number of reconstructed primary vertexes in minimum bias events. (b) Fractional systematic uncertainties due to multiple interactions for different cone sizes as a function of jet transverse momentum.

The average of f_b in the analyzed η bin is used to define the β factor[9] (see Fig. 3.5)

$$\beta = \frac{2 + \langle f_b \rangle}{2 - \langle f_b \rangle} \tag{3.5}$$

The final L1 correction is defined as $f_{L1}(\eta, E_T^{raw}, R) = 1/\beta$ and the uncertainty associated with this correction is estimated to be of the order of 1% for central jets and 7.5% for forward jets.

C_{MI}: multiple interactions correction (L4)

The number of interactions occurring during beam bunch crossings follows a Poisson distribution whose mean increases with instantaneous luminosity.
These additional interactions, dominantly soft *minimum bias* events, cause extra unwanted energy to be deposited in the calorimeter.
The number of reconstructed vertices is used to estimate the effect. The average energy flow in minimum bias events, which are triggered by the luminosity monitor CLC, is measured in the best-performing region ($0.2 < |\eta| < 0.6$) of the calorimeter as a function of the number of reconstructed vertices. The resulting plot is fit to a straight line (see Fig. 3.6). This is used to correct the energy of the jets. Because of the finite reconstruction efficiency of the vertices, this linear approximation works well for events with less than seven vertices.
This is not a serious limitation because in practice events with so many vertices are very rare. The uncertainty on this correction is estimated to be of the order of 15%.

[9] The definition of 3.5 is in average equal to $p_T^{\text{probe}}/p_T^{\text{trigger}}$ but it reduces the sensitivity to the presence of non-Gaussian tails which affect the usual ratio.

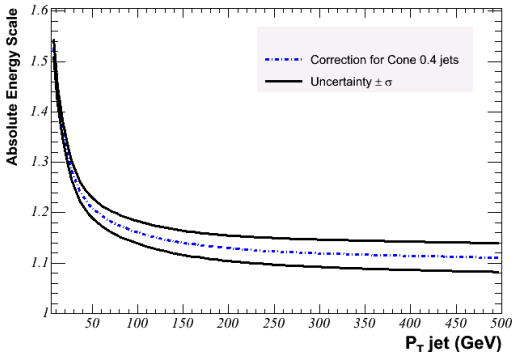

Figure 3.7: Absolute energy corrections for jets with cone size $\Delta R = 0.4$ as a function of jet p_T with uncertainty.

C_{Abs}: absolute energy scale corrections (L5)

While L1 and L4 are corrections at calorimeter level, L5 steps back to particle level[10]. The procedure used to estimate the L5 correction factor is described accurately in [74] and uses a MC sample of inclusive dijet events simulated with PYTHIA [75].
The correction is derived comparing particle jets at generator level (before they are passed through the detector simulation), with calorimeter jets, as obtained from the detector simulation. These are required to be within 0.1 of each other in the $\eta - \phi$ plane to ensure that they are the same object. The probability at the distribution maximum of measuring a value of p_T^{jet} given p_T^{particle} is taken as a correction factor (see Fig. 3.7). The uncertainty on this corrections is estimated to be of the order of 3.5% (15% near the edge of the calorimeter).

C_{UE} and C_{OOC}: underlying event (L6) and out-of-cone (L7) corrections

Reconstructed jet energies in hard $p\bar{p}$ interactions may contain contributions by particles created by spectator interactions or by gluons from initial state radiation in the hard interaction. These contributions are called *underlying event*. On the other hand a fraction of the parton energy may be lost outside the jet cone because of final state gluon radiation, fragmentation at large angles relative to the jet axis or low p_T particles bending in the magnet field. This energy is modeled imperfectly in MC events, so a systematic uncertainty is assigned by examining photon + jets events in data and MC. A ring around the jet with a radius between 0.4 and 1.3 in the $\eta - \phi$ plane is examined, and the energy in this region is compared between data and MC simulation. The largest difference between MC events and data is taken as a systematic uncertainty.
Fig. 3.8 shows the individual fractional systematic uncertainties as a function of jet p_T in the central region, $0.2 < |\eta| < 0.6$, of the calorimeter. They are independent from each other and thus are added in quadrature to derive the total uncertainty.

[10] L2 and L3 have survived in the CDF jargon but are not used anymore.

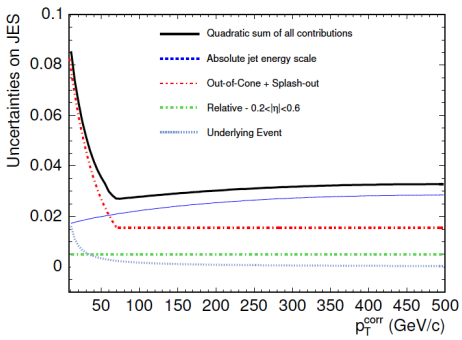

Figure 3.8: The fractional systematic uncertainty of the JES corrections as a function of the jet transverse momentum. The total uncertainty is taken as the sum in quadrature of all individual contributions.

For $p_T > 60$ GeV/c the largest uncertainty is contributed by the absolute energy scale, which is caused by the uncertainty of the calorimeter response to charged hadrons. For an analysis that relies on jet quantities as the subject of this thesis, the uncertainty on the jet energy scale results is the dominant source of uncertainty in the final result.

3.6 Neutrino Reconstruction

Neutrinos produced in the final state escape detection causing an energy imbalance in the observed event. Momentum conservation is the only way available handle to reveal the presence of neutrinos. Since, the z-component of the momentum the colliding partons is not known, one cannot determine the net “missing" energy caused by neutrinos. However, the total momentum in the transverse plane is zero to a very good approximation and therefore one can define the transverse missing energy (which is actually missing transverse momentum) as the vector sum of the transverse energies of all the calorimeter towers:

$$\vec{\not{E}}_T = -\sum_i \vec{E}_T^i$$

where i indexes each tower. The sum involves all the towers with total energy above 0.1 GeV in the region $|\eta| < 3.6$. This vector gives the energy and the direction in the transverse plane of a (massless) undetected particle.
At offline level, the algorithm corrects for the position of the reconstructed event vertex and for any reconstructed muon [76].
The largest correction is due to muons which have only ionization energy losses in the calorimeter and hence produce an apparent missing energy in the detector. The $\vec{\not{E}}_T$ is corrected in events where a muon is identified by adding back its transverse momentum, as measured in the tracker, to the total transverse energy deposited in the towers and subtracting any small amount of transverse energy which it may have deposited in the calorimeters.

In this case

$$\vec{\not\!\!E}_T = \vec{\not\!\!E}_T^{raw} - p_T^{\mu} \cdot \left(1 - \frac{E_{cal}^{\mu}}{p^{\mu}}\right)$$

where $\vec{\not\!\!E}_T^{raw}$ is the "raw" uncorrected missing energy, $E_{cal}^{\mu} \sim 2$ GeV is the energy deposited by the muon in the calorimeter and p^{μ} is its momentum.
Another correction is needed due to jets whose raw measured energy within the jet-cone is systematically shifted from the hadron's one. For the jet corrections in the $\not\!\!E_T$, for each jet the uncorrected jet energy is subtracted by the total transverse energy deposited in the towers and the corrected transverse jet energy is added:

$$\vec{\not\!\!E}_T = \vec{\not\!\!E}_T^{raw} - \Delta\vec{E}_T^{jet}$$

where $\Delta\vec{E}_T^{jet}$ is the difference between the corrected and the raw jet energy.
The source of real $\vec{\not\!\!E}_T$ in this analysis is from neutrinos created in electroweak interactions. There are also several sources of false $\vec{\not\!\!E}_T$ which are often difficult to control. These sources include the mis-measurement of jet and lepton energies as well as apparent missing energy seen when an event prong enters a crack in the detector.

3.7 Bottom Jets Identification

The identification of B-hadrons in jets was fundamental for the discovery of the top quark in 1995 and is one of the crucial features of the searches for SM light Higgs boson at the Tevatron collider. Jets carrying b-flavour (*b-jets*) are produced from b-quark hadronization in top or Higgs decays, while mostly light-quark jets are produced from the main background processes (e.g. W+jets), which contaminate the candidate event data sample.
Several methods to identify b-jets exist at CDF. We will describe the official CDF b-tagger, *SecVtx*, and the one used in this analysis, *b*ness.

3.7.1 SecVtx

SecVtx takes advantage of the long proper lifetime of the B-hadrons[11], which, together with a large relativistic boost, can fly several millimeters before decaying. Therefore, the vertex where the decay products originate may be enough displaced to be distinguishable from the primary one. For this reason this algorithm is called "Secondary vertex algorithm", *SecVtx*.
In order to identify secondary vertices, displaced tracks are reconstructed by the inner tracker with sufficient precision (about 50 μm) to be distinguished from the prompt tracks[12].
SecVtx selects tracks with a large impact parameter and reconstructs a secondary vertex out

[11] The B^0 has a proper life time of 1.5 ps ($c\tau = 476$ μm). The average transverse momentum of a B-hadron coming from a WZ event is about 40 GeV/c. A B^0 meson has typically $\beta\gamma = 7.6$ and the average decay length is 3.5 mm

[12] those produced at the primary interaction

of them. It doesn't account of cases in which decay products are neutral and/or may include charmed hadrons, which produce tertiary vertices by decaying.
By this procedure C hadrons cannot be distinguished by B hadrons. Although C hadrons decay with a lower track multiplicity and in a shorter average time than b-hadrons[13], they are tagged with a relativity high efficiency as well. Therefore *SecVtx* is actually a "heavy-flavour tagger".
Although this is not desired for directly produced charm hadrons, it increases the b-tagging efficiency by tagging secondary charm hadrons produced int b-decays. Details on how *SecVtx* operates are given in Appendix D.

3.7.2 Jet-$b$ness

In the analysis described in this thesis the b-tagger employed is the $b$ness, which is a multivariate, neural network (NN) based tagger. It provides an output value which serves as a figure of merit to indicate how b-like a jet appears to be.
What distinguishes this tagger from other multivariate taggers used at CDF is its emphasis on studying individual tracks for characteristics indicating they may have come from a B-hadron decay. This is done by implementing a *track-by-track* NN, whose output is fed into the final *jet-by-jet* NN.

track-by-track NN

A NN output is assigned to each track within a jet as a figure of merit to indicate how likely the track appears to be associated to a B-hadron decay product. Only tracks with $p_T > 0.4$ GeV are considered. Within a pool of investigated variables in simulations, there those with the largest discrimination power were selected.
The track observables used to train this NN are

- p_T
- P_{perp}, momentum perpendicular to the jet axis
- rapidity with respect to the jet axis
- d_0 and σ_{d_0}
- z_0 and σ_{z_0}
- parent jet E_T

Note that the choice of the observables in the track NN reflect the fact that tracks from B-hadron decay products will tend to have higher momentum and be more collinear with the jet than other tracks, and also have an impact parameter farther from the primary vertex.
The parent jet E_T is used as an input as distributions of B vs. non-B track observables may change depending on the jet's E_T. However, when the training was performed a weighting was applied such that the parent jet E_T distributions were the same both for B and non-B

[13] The D^0 proper life time is about 0.4 ps.

tracks: this was to ensure that application of the tagger wouldn't introduce a kinematic bias. The same weighting is applied to the jet-by-jet NN. In Fig. 3.9 the input observables for the track NN are shown.

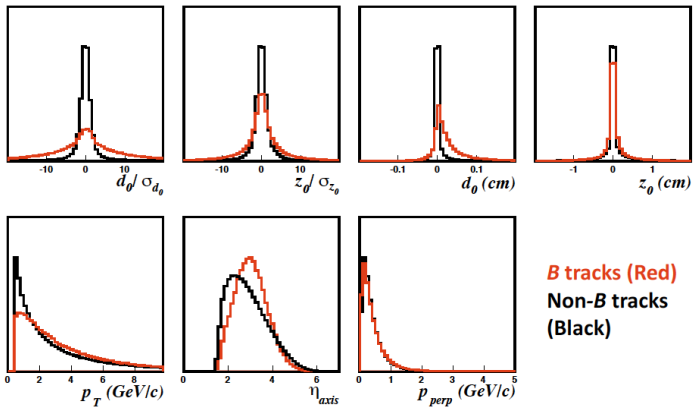

Figure 3.9: Comparisons of track observables input into the track NN, for tracks from B-hadron decays vs. tracks not from B-hadron decays.

jet-by-jet NN

The NN trained to identify b-jets uses as input the following variables:

- top five highest track's $b$ness
- number of tracks with $b$ness> 0
- invariant mass of tracks with $b$ness> 0
- $\sigma_{L_{xy}}$
- number of K_S candidate
- muon likelihood

Here, not only individual track $b$ness, but also the multiplicity and invariant mass of high-$b$ness tracks are used as inputs into the NN to evaluate the jet $b$ness.
Furthermore L_{xy}, the displaced vertex significance, is calculated by taking all tracks with a $b$ness> -0.5, and removing the lowest-b-ness tracks in the set until the fit for a secondary vertex is returned with a $\chi^2 < 50$.
Two other input variables are: the number of candidate K_S or Λ particles in the jet and the highest-likelihood muon candidate in the jet [79].
In Fig. 3.10 the input variables of the jet $b$ness for jets matched to b-hadrons[14] and the other jets are compared.
Finally, in Fig. 3.11 we see the jet $b$ness output for jets matched to b quarks versus the other jets. The difference between the two distributions is evident.

[14] a dR<0.4 is required in order to consider a B-hadron matched to a jet.

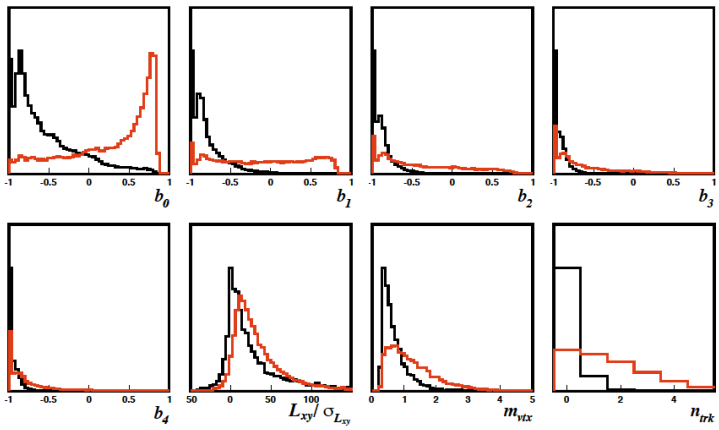

Figure 3.10: Comparisons of jet observables input into the jet NN, for jets including B-hadron decays (red) vs. jets without B-hadron decays (black).

$b$ness-Tagger Validation

In order to verify that jet-$b$ness response is similar in data and MC data (collected with high-p_T lepton triggers) and MC by using $t\bar{t} \rightarrow$ leptons + jets and $Z \rightarrow \ell\ell$+1 jet selections are compared. The former offers a comparison of jets in a b-enhanced region (b-jets are about 50% of the total), while the latter compares jets that largely do not originate from bottom quarks. These studies are described in [80].

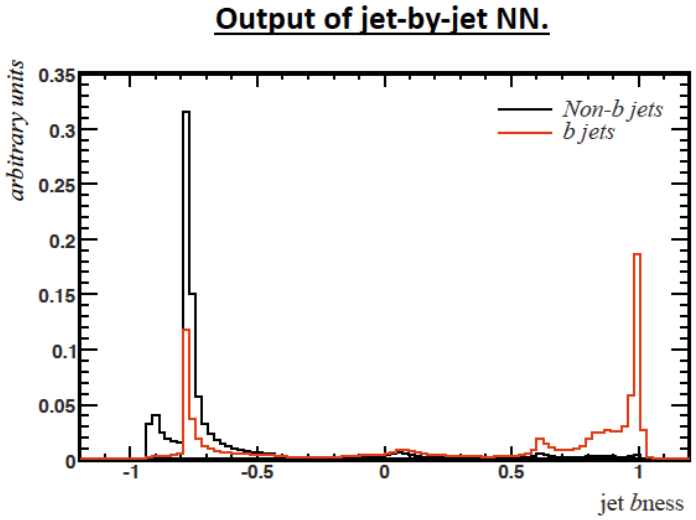

Figure 3.11: Jet $b$ness distribution from our electroweak MC, for jets matched to b quarks (red) vs. jets not matched to b quarks (black).

Chapter 4

Composition of the selected data sample

This chapter reports on how the sample of experimental data used for our studies is selected. The events of interest in this analysis have the experimental signature of two or three energetic jets, high p_T isolated leptons (electrons or muons) and large missing transverse energy.
The details of the offline selection, operating on the physical objects discussed in the previous chapter, will be described.
The criteria are tuned in order to efficiently select our signal events and reduce the backgrounds that share the same signature.

4.1 Event Selection

Our final goal will be to observe the associated production of W and Z bosons in the $WZ \to \ell\nu jj$ final state. Two samples corresponding to an integrated luminosity of 6.6 fb^{-1} are studied separately. One, the *tag* sample in which we require to have two b-jets in the final state. The second, the *notag* sample is the *pretag* subsample[1] where the *tag* contribution has been removed in order to be able to combine the results in the two samples. The first is the golden channel for the light SM Higgs boson search at Tevatron ($WH \to Wb\bar{b}$).
The invariant mass distributions in the *notag* and *tag* samples will be fitted separately and the results will combined in order to extract the signal.
The offline event selection cuts are summarized in Table 4.1.

[1] Pretag sample is the one where no constrain on jets flavor are applied

Leptons

First, high p_T electrons and muons are identified. The trigger requires an electron or a muon with $E_T > 18$ GeV. The basic lepton selection cuts applied at trigger level still accept a large number of fake leptons. Offline *good* leptons satisfying some quality criteria are identified in order to purify the sample, making it easier to be modeled and to estimate the background contribution.
Finally, the triggered lepton is required to be a *tight central electron* (the TCE selection cuts are listed in section 3.3) with $E_T > 20$ GeV and $|\eta| < 1.1$, or a CMUP or CMX muon with $P_T >$20 GeV/c and $|\eta| < 1$ (the CMUP and CMX selection cuts are listed in section 3.4).
We allow for more than one *good* lepton per event and we consider also ZZ as our signal. This increases the number of events by about 5% for the electron sample and by 18% for the muon sample, albeit at the expense of some degradation in resolution, see Fig. 4.1-4.2

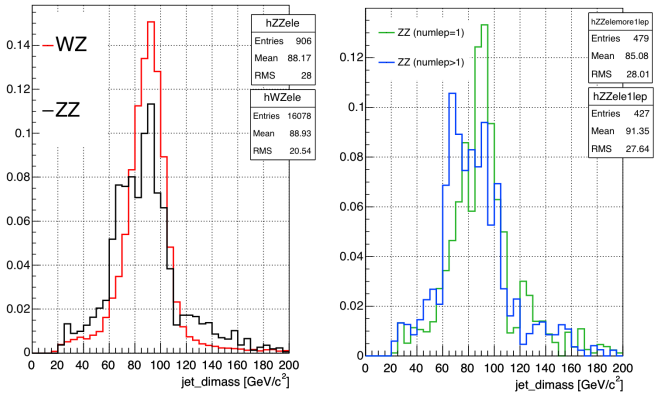

Figure 4.1: Left, dijet mass distribution built with the two E_T leading jets in WZ (red) and ZZ (black) MC samples for TCE electrons. Right, in a ZZ MC sample, dijet mass distributions in events where exactly one (green) or more than one electrons (blue) are reconstructed.

In the sample in which we allow for more than one *good* lepton the $\not\!\!E_T$ comes completely from a jet energy mismeasurement. Since we have required $\not\!\!E_T > 20$ GeV we accept only events in which the mismeasurement is large. This is the reason why the resolution is spoiled so much. The dijet mass resolution is more degraded in the electron sample than in the muon one. This is mostly due to the requirement on M_T^W which is more restrictive for electrons.

Missing Transverse Energy

In order to compute the transverse mass of the W we need also the missing transverse energy to account for the non interacting neutrino.

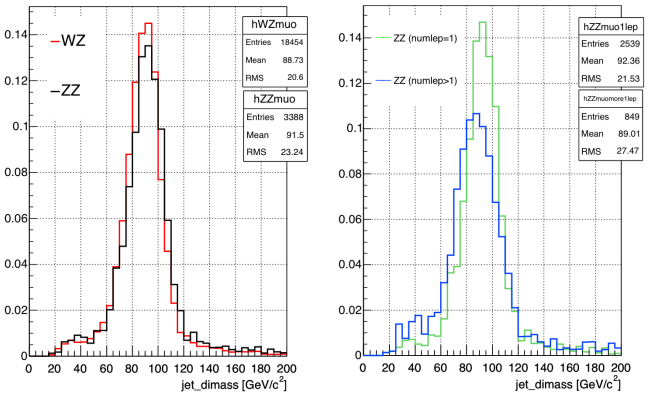

Figure 4.2: Left, dijet mass distribution built with the two E_T leading jets in WZ (red) and ZZ (black) MC samples for CMUP or CMX muons. Right, in a ZZ MC sample, dijet mass distributions in events where exactly (green) one or more than one muons (blue) are reconstructed.

We require the missing transverse energy to be greater than 20 GeV. With this cut the multijet background is reduced by 75% while the signal only by 15%.

Transverse Mass

From the $\not\!\!E_T$ and the lepton we reconstruct the transverse mass of the W candidate[2].

$$M_T^W = \sqrt{2 \cdot E_T^\ell \cdot \not\!\!E_T \cdot (1 - \cos\Delta\phi_{\ell\nu})} \tag{4.1}$$

In Fig. 4.3 are shown M_T^W distributions for electrons and muons, after the offline lepton selection. We observe that while the distributions peak approximately at the W mass as expected for W events, there is a large tail at low mass due mainly to QCD background where the $\not\!\!E_T$ is a fake. The tail at low M_T^W values is more enhanced for electrons, corresponding to a larger fraction of QCD events since it is easier for a jet to fake an electron than a muon. For these reasons, the transverse mass of the W is required to satisfy $M_T^W \geq 30$ GeV/c^2 for electrons and ≥ 10 GeV/c^2 for muons.

[2]The event is not longitudinally balanced, because the momenta of the partons inside the colliding protons are unknown.

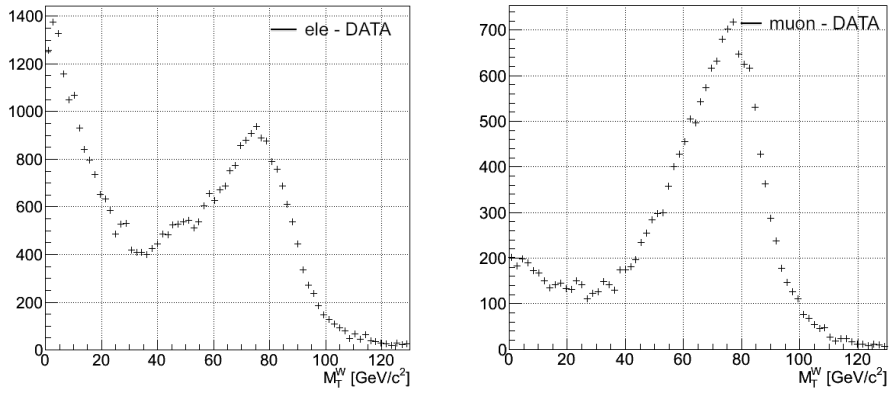

Figure 4.3: M_T^W distributions for electrons (left) and muons (right).

metsig

In order to reduce the QCD background in the electron data sample, also a cut in metsig[3] has been optimized in terms of the signal over background S/B ratio. A cut on this variable allows us to reduce by about 50% the QCD fraction in the sample while the signal is reduced only by 7%[4].

Jets

The offline event selection identifies jets using the JETCLU cone algorithm with radius 0.4, corrected for detector effects as described in section 3.5.2.
Currently in similar CDF analisyses two exclusive highest E_T jets are required and the WZ/ZZ signal is extracted using a fit to their invariant mass.
Here we are including also the three exclusive jets sample which is the scope of this thesis. The final fit will use separately the *two* and *three jets regions* in both *notag* and *tag* samples.

We look for two jets with $E_T > 15$ GeV and $|\eta| < 2$. In the two jets study we exclude events in which a third jet with $E_T > 15$ GeV is found in this range.
For the three jets studies, we allow for a third jet with $E_T > 10$ GeV and $|\eta| < 3.6$, and we accept events in which only three jets are found. The requirements chosen for jet E_T are shown in Table 4.1. They have been optimized for jets produced from a heavy boson decay which have higher transverse energies than background QCD jets.
In order to select the *tag* sample we require the two leading jets to have $b$ness > 0.75, -0.2 respectively. These cuts have been optimized against the sensitivity of the measurement, and are the same for the *two jets region* and *three jets region* studies [6]. Consequently we require

[3] `metsig` $= (-\log_{10}(\mathrm{P}(\not\!\!E_T^{fluct} > \not\!\!E_T)))$, where P is the probability and $\not\!\!E_T^{fluct}$ is the expected missing transverse energy arisen from fluctuations in the energy measurements [82].
[4] This estimation has been done after cuts on $\not\!\!E_T$ and M_T^W were applied.

Variable	Requirement
E_T (TCE, CMUP, CMX)	$\geq$ 20 GeV
$\not\!\!E_T$	$\geq$ 20 GeV
M_T^W if $\ell = \mu$	$\geq$ 10 GeV
M_T^W if $\ell = e$	$\geq$ 30 GeV
metsig if $\ell = e$	$\geq$ 1.8
number of jets	$= 2$
E_T J1	$\geq$ 25 GeV
E_T J2	$\geq$ 25 GeV
$\|\eta\|$ J1	≤ 2
$\|\eta\|$ J2	≤ 2
number of jets	$= 3$
E_T J1	$\geq$ 25 GeV
E_T J2	$\geq$ 15 GeV
E_T J3	$\geq$ 15 GeV
$\|\eta\|$ J1	≤ 2
$\|\eta\|$ J2	≤ 2
$\|\eta\|$ J3	≤ 3.6

Table 4.1: Summary of the cuts defining the candidate events samples.

in *notag* sample the two leading jets to not satisfy both the conditions for the *tag* sample, i.e. $b$ness > 0.75, -0.2 respectively, in order to separate from each other *tag* and *notag* samples. Later in this thesis we will refer to the two leading jets as the

- E_T-highest jets in the *notag* sample
- $b$ness-highest jets in the *tag* sample. In this case the third selected jet is the one with the highest E_T among the others.

4.2 Data sample

This analysis uses data collected by CDF II detector between February 2002 and July 2010, corresponding to an integrated luminosity of 6.6 fb^{-1}.
At the trigger level, one attempts to select signal events with high efficiency while keeping the accept rate low. The data used in our $WZ/ZZ \to \ell\nu jj$ search has been collected exploiting only one signature, i.e. the presence of a high p_T central lepton. A high p_T electron trigger and two high p_T muon triggers are used, which feed the CDF standard `BHEL` and `BHMU` datasets, respectively. All data events are required to fire one of the following *trigger paths*:

```
ELECTRON_CENTRAL18
MUON_CMUP18
MUON_CMX18
```

which apply specific requirements at each trigger level.

The ELECTRON_CENTRAL18 path is used to select centrally produced high-p_T electrons, reconstructed in the CEM calorimeter.
The MUON_CMUP18 and MUON_CMX18 paths, used for the central muon selection, look for muons in CMU and CMP chambers and CMX chambers respectively. In Fig. 4.4 the coverage in $\eta - \phi$ of muons and electrons is shown. An extensive overview of the trigger selection is given in [61].

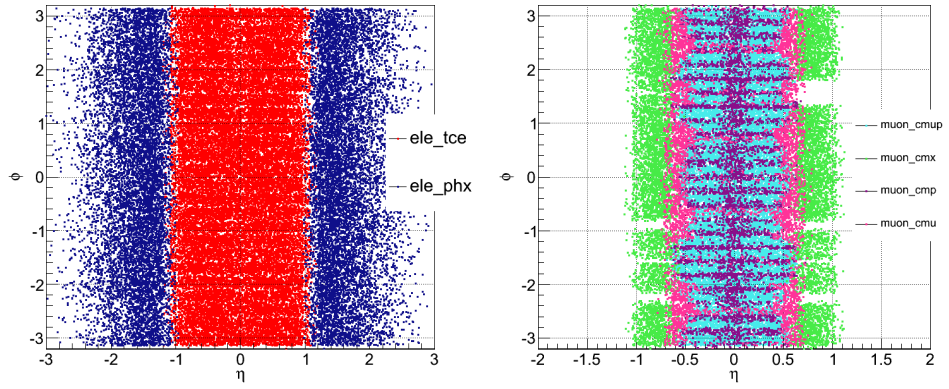

Figure 4.4: Coverage in the $\eta - \phi$ plane of TCE trigger electrons (left) and CMUP and CMX trigger muons (right). Also Phoenix electrons (PHX), CMP and CMU muons are shown, even if they are not used in this analysis.

4.3 Signal and Background

The study presented in this thesis is a contribution to a search for the associated production of $W(Z)$ and Z bosons, where the primary $W(Z)$ boson decays leptonically and the associated Z boson decays into two quarks that hadronize into jets.
Also the ZZ contribution is included in our signal, since even if Z boson decays into two charged leptons, a lepton can escape detection and fake large $\not\!\!E_T$ as in W decay.
Several other processes can simulate the same signature, primarily $W(Z)$ events with associated hadron jets, or jet events where a jet fakes a lepton and jet mismeasurements fake large missing E_T. Our data sample has a large contamination of different backgrounds (see Table 4.2 for some important backgrounds). Online and offline selections are tuned in order to improve the signal over background ratio.
W+jets, i.e. the production of a W boson in association with multiple high p_T jets, is the main background in the *notag* sample.
The production of multiple jets in association with a Z boson (**Z+jets**) is second in importance. In Fig. 4.5 are shown some lowest order Feynman diagrams for these processes.
Top quarks are produced mostly in pairs at the Tevatron (Fig. 4.6), though single top quark production is also possible with a factor of ∼3 lower cross section. Single top quarks are

Process	σ (pb)
$p\bar{p} \to W^{\pm} \to \ell^{\pm}\nu_{\ell}$ + jets	$\sim 2.1 \cdot 10^3$
$p\bar{p} \to Z \to \ell^{+}\ell^{-}$ + jets	~ 787
$p\bar{p} \to t\bar{t}$	~ 7.0
$p\bar{p} \to t\bar{b}$	~ 2.9

Table 4.2: Estimated cross sections of some important background processes [81, 19].

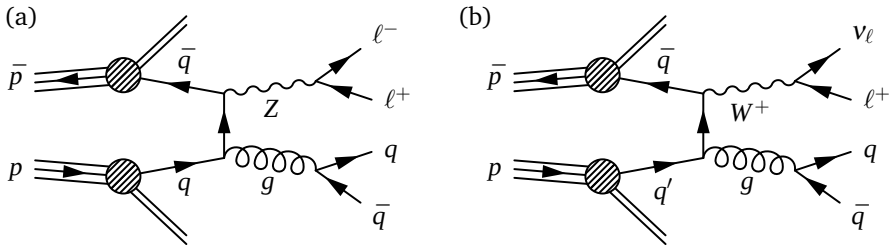

Figure 4.5: Lowest order Feynman diagrams for the production of $Z+$ jets (a) and $W+$ jets (b).

produced through two electroweak processes: a charged current interaction generates a top quark and a bottom quark (s-channel) or a virtual W interacts with a b-quark in the proton sea to produce a top quark (t-channel), Fig. 4.7.
Top production processes may have final states very similar to the signal, with a leptonically decaying W and two detected jets. Since top production is characterized by b-jets in the final state we expect it to give a more significant background in the *tag* sample.

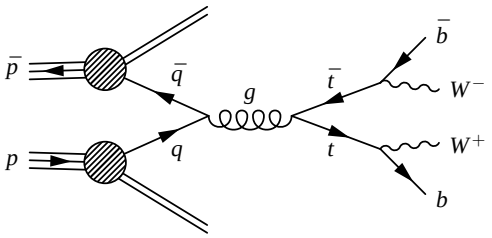

Figure 4.6: Feynman diagram for $t\bar{t}$ production.

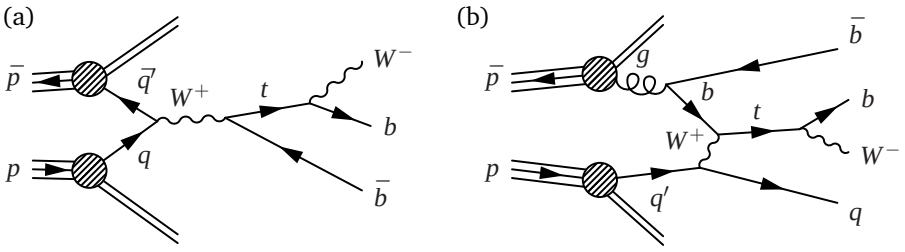

Figure 4.7: Feynman diagrams for single top production: $s-$channel (a) and $t-$channel (b).

Because of the extremely large cross section of **QCD** multijet events even rare mismeasured events can contaminate the sample significantly (Fig. 4.8). Signal signature can be faked by a jet passing all lepton cuts and simultaneously the energies are so poorly measured that a large missing transverse energy is found.
The low probability for a jet to fake a lepton would require to generate a huge MC sample in order to simulate the effect. Moreover, one would have to be very confident in details of the detector simulation. Since it would be highly unreliable to simulate this background with MC studies, a data driven technique has been used to compute it, as explained in the next chapter.

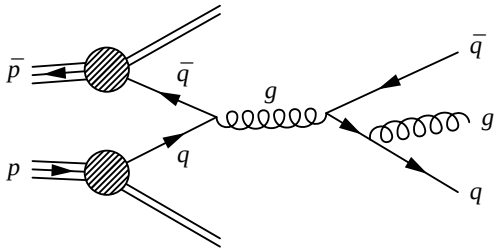

Figure 4.8: Example of one Feynman diagram for QCD multijets production.

Chapter 5

Modeling the selected data sample

In this chapter the methods used to model the data and to estimate each contributing process will be presented.
This analysis uses a fit to the invariant mass distribution of the jets associated to the hadronic decay of Z boson to extract the diboson signal in the selected data sample (see chapter 4). Thus, modeling plays a fundamental role, since the sensitivity to the signal and its significance depend on an accurate modeling of all components of the invariant mass distribution.

5.1 Estimate of Signal and Background in the sample

The expected composition of the selected sample is estimated in a sequence of steps by using a method which is very similar to the *Method II* used in many analyses of CDF data with leptons plus jets signature [83].
The method assumes that the following process contribute to the selected data sample:

- **Electroweak and top** processes characterized by accurately predicted cross section
- **QCD** multi-jets production faking the lepton and $\not\!\!E_T$
- **W+jets** production of a $W-$boson associated to hadron jets
- **Z+jets** production, which is characterized by a well measured cross section

The method consists of two steps. First of all, the contribution of processes whose cross sections are well known is estimated. Then, the contribution of QCD fakes and of W+jets is estimated with a data driven method. In this way we avoid to use the imprecise theoretical predictions for the production cross section of W boson with associated jets [84, 85].

5.1.1 MC based background estimates

Processes whose contributions can be accurately computed based on a known cross section are WW, WZ, ZZ, Z+jets, $t\bar{t}$, single-top (see Table 5.1). Each of these processes can produce a signature consistent with one or more leptons, large $\not\!\!E_T$, and more than two jets.

Process	σ (pb)
WZ	3.22 ± 0.06
WW	11.34 ± 0.35
ZZ	1.20 ± 0.07
$t\bar{t}$	7.04 ± 0.8
single-top	
t-channel	2.10 ± 0.07
s-channel	1.04 ± 0.09
Z+jets	787.4 ± 50

Table 5.1: NLO theoretical cross sections for EW and top processes, assuming a top quark mass $m_t = 173.0 \pm 1.2$ GeV/c^2 [19]. Measured cross section for Z+jets production [86].

The yields in our sample are estimated using the cross section, the luminosity for the sample, the trigger efficiency and an overall selection efficiency derived from MC simulation. The contamination of each of the EW processes to the *notag* and *tag* samples is determined as,

$$N_X^{notag} = \sigma_{p\bar{p}\to X} \cdot \epsilon_{evt} \cdot (1 - \epsilon_{tag}) \cdot \int \mathscr{L} \mathrm{d}t$$
$$N_X^{tag} = \sigma_{p\bar{p}\to X} \cdot \epsilon_{evt} \cdot \epsilon_{tag} \cdot \int \mathscr{L} \mathrm{d}t$$

where:

- X = WW, WZ, ZZ, Z+jets, $t\bar{t}$, single top

- $\sigma_{p\bar{p}\to X}$ is the theoretical/measured cross-section for $p\bar{p} \to X$, (Table 5.1).

- ϵ_{evt} is the MC-derived acceptance in the *pretag* sample. It is estimated by applying the selection criteria. Calibration factors accounting for discrepancies in lepton identification efficiencies and trigger efficiencies between data and MC are applied[1].

- ϵ_{tag} is the event-by-event tagging efficiency, defined as the probability for the two leading jets to have $b$ness>0.75, -0.2 respectively. Such a probability is estimated in the *pretag* sample. Rather than applying a renormalization factor based on the calculated different MC versus data scale factors for the jet-by-jet efficiency and mistag rates, we locate the equivalent $b$ness cuts in the Monte Carlo that matches the measured efficiencies (for processes that have real b jets, like top backgrounds and our signal) and mistag rates (for processes without real b jets, like WW) in the data [80].

[1] The event detection efficiency can be decomposed in

- event detection efficiency obtained from our samples of MC events, before tag requirements are applied
- BR factor, which takes into account that in some MC samples used W boson is only allowed to decay into leptons
- trigger efficiency, which takes into account the fact that in the MC the trigger requirements are not applied
- the lepton ID scale factor, which is the ratio of the lepton reconstruction efficiency in MC to that in data

5.1.2 Non-W and W+jets background estimate

The background from **QCD** multijets production arises when a jet fakes a lepton and mismeasurements of the jet energies lead to large missing transverse energy. This effect is most relevant in the case of electrons, which can be easily faked by jets of low track multiplicity and large electromagnetic fraction. A particular problem is multijets production where some of the jets are originated by B-hadrons. In this case the semileptonic decays of B-hadrons produce electrons or muons and neutrinos, altogether faking the signature of W's. This background is small relative to the dominant W+jets background, but larger than the signal. A data driven method is chosen to model it.

Two different models for QCD events are used, one for electrons and one for muons, since different lepton types may induce different rates and shapes of multijet background. Both are based on the principle that this process must contain a jet that is misidentified as a lepton. It is modeled from a control data sample with jets that come close to passing leptons cuts. This sample is built requiring the same selection requirements as the ones described before except that some of lepton identification criteria are reversed. The criteria to be failed are different according to the type of the lepton which fires the trigger.

For **muons** we require the isolation cut to fail. The isolation cut for the muons used in this analysis is defined as $Iso < 0.1$[2]. So, we require $Iso > 0.1$ to model the QCD in muon sample. Since the isolation is not a kinematical variable, we expect the non-isolated muon sample to simulate well the real QCD multijets kinematical behavior in the muon sample. Inverting the isolation cut of muon allows to have enough statistics for modeling the QCD background.

For **electrons** we use an *antielectrons* sample which is built with events that fail at least two of the non-kinematic cuts (listed in section 3.3) but pass all the others. These cuts are based on variables such as E_{HAD}/E_{EM}, χ^2_{CES}, L_{shr}, $Q \cdot \Delta X_{CES}$ and $|\Delta z_{CES}|$, that are designed primarily to reject fake electrons but do not much affect the kinematic properties of the event. This fake electron is chosen as the candidate electron, and on the rest of the event the standard selection cuts are applied. In order to estimate the multijet background and W+jets normalization, we perform a fit of the $\not\!\!E_T$ distribution in the *notag* and *tag* electron data samples. Removing the $\not\!\!E_T$ cut creates samples with a large QCD component, since QCD events have no real neutrino and dominate the region with small $\not\!\!E_T$. For muons, the fit is performed in the *notag* and *tag* data samples using the M_T^W distribution which has best sensitivity to the QCD contribution.

Since the backgrounds contributing to the invariant mass distribution are not simply parameterizable, we use directly their dijet mass *templates* (histograms normalized to unit area), taken either from MC simulation or from data driven procedures. With those we perform a binned χ^2 fit to the data.

[2] Iso is the ratio between the transverse energy surrounding (within a ΔR=0.4 cone) the muon and the transverse energy of the muon.

In order to verify our modeling, a fit is performed in terms of 3 known $\not\!\!E_T$ (M_T^W) *templates*[3]:

1. **EW** and **Z+jets** MC-based template (ALPGEN + PYTHIA [87] and PYTHIA MC [88] after the simulation of the CDF detector) built by using EW and Z+jets events selected according to the requirements described before. We sum each contributes and we obtain a template (**EWK**) whose normalizations is constrained to the expected value of each contributing process.
2. **W+jets** MC-based template (ALPGEN + PYTHIA) built with W+jets selected events. The normalization of this template is a free parameter of the fit and is obtained from data.
3. **QCD**

   a - *notag* data-based template, whose normalization is allowed to float in the fit.

   b - *tag* we measure in data the tag rate in bins of $\not\!\!E_T$ (M_T^W for muons) and we use this rate to weight the *notag* QCD template. The resulting distribution is used as the QCD template in the tag sample[4].

The results of the fit are shown in Fig. 5.1-5.2 and in Table 5.2 separately for electrons and muons in *notag* and *tag* samples. We reported in Table 5.2 the uncertainties on W+jets returned by the fit. However, in the final fit to estimate the WZ/ZZ cross-section, we allow the W+jets normalization to float (sec. 7.1).
The QCD fraction can fluctuate substantially when fitting different ($\not\!\!E_T$, M_T^W) observables or with a different choice of fit ranges. A systematic uncertainty of 50% covers all observed effects. This number is the largest difference obtained by choosing different fit ranges and different observables.

Process	Electrons	Muons
NOTAG		
W+jets	0.76 ± 0.01	0.79 ± 0.01
QCD	0.129±0.064	0.037 ± 0.018
TAG		
W+jets	0.42 ± 0.08	0.47 ± 0.07
QCD	0.089±0.044	0.023±0.011

Table 5.2: The resulting QCD and W+jets fractions returned by the fit.

[3] We perform the fit in terms of a sum of the background templates. The templates are built separately in the electron and muon samples and then they are added together before performing the fit.

[4] The requirements on the $b-$flavor content of the jets drastically reduce the number of events of the data-based QCD template, thus making it impossible to build a significant QCD template directly in the tag sample.

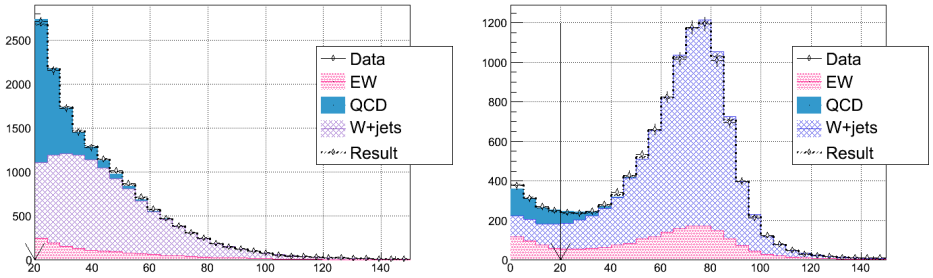

Figure 5.1: Notag Sample. Left, fit to the $\not{E}_T$ distribution for electrons. Right, fit to the M_T^W distribution for muons.

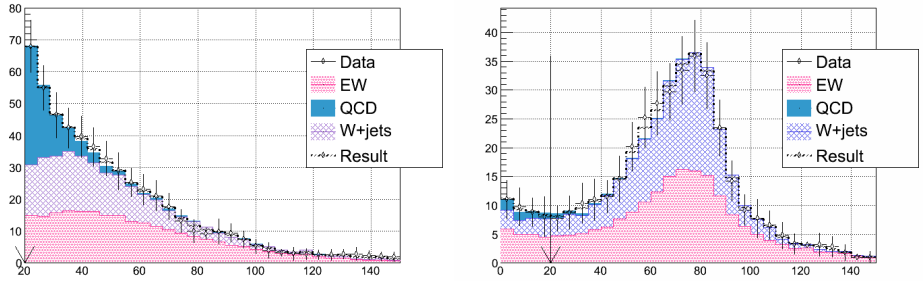

Figure 5.2: Tag Sample. Left, fit to the $\not{E}_T$ distribution for electrons. Right, fit to the M_T^W distribution for muons.

5.2 Modeling

Once the background levels are predicted, we want to investigate the agreement between data and MC of various kinematic distributions. Before doing that, we need to apply another correction to the MC to take into account the different luminosity profiles with respect to the data.

A quantity directly related to the instantaneous luminosity is the number of reconstructed vertices. We deal with the observed disagreement between data and MC prediction by re-weighting the MC events.

In Table 5.3-5.4 we show the estimated number of events for each process contributing to both the *notag* and *tag* sample data sample for the MJ1J2 distribution.

5.2.1 Notag sample

In this section we compare a number of observables (see Figs. 5.3-5.18) in data and simulated events in order to validate the MC, the QCD templates, and the expected rates in the notag sample:

- `jet1_Et`, `jet2_Et`, `jet3_Et`: the transverse energies of the first/second/third E_T-leading jet.
- `jet1_Eta`, `jet2_Eta`, `jet3_Eta`: η of the first/second/third E_T-leading jet.
- `jet1_Phi`, `jet2_Phi`, `jet3_Phi`: ϕ of the first/second/third E_T-leading jet.
- `jet1_bness`, `jet2_bness`, `jet3_bness`
- `lep_et`: the transverse energy of the highest-E_T lepton
- $\not\!\!E_T$
- M_T^W: invariant mass of the highest-E_T lepton and $\not\!\!E_T$ system in the transverse plane
- `metsig`.

The agreement is satisfactory in all distributions.

Process	Rate (Electrons)	Rate (Muons)
Signal	66.2 ± 0.9	69.5 ± 0.9
WW	386.2 ± 3.0	311.1 ± 3.1
$t\bar{t}$	333.0 ± 1.4	288.5 ± 1.2
single-top	68.9 ± 0.4	57.8 ± 0.3
Z+jets	350.0 ± 3.2	1167.8 ± 4.5
W+jets	10304.2 ± 29.6	8275 ± 22.8
QCD	1600.4 ± 60.0	352.3 ± 5.4
Total Observed	13109.0 ± 114.5	10522.0 ± 102.6

Table 5.3: Predicted and observed number of events of the *notag* three jets sample. The expected rates are separated for the triggered lepton type. We also require the invariant mass of the two E_T-leading jets to be within [0,300] GeV/c^2. By construction the overall expected rates are the same as the observed ones in each region. The uncertainty shown is statistical.

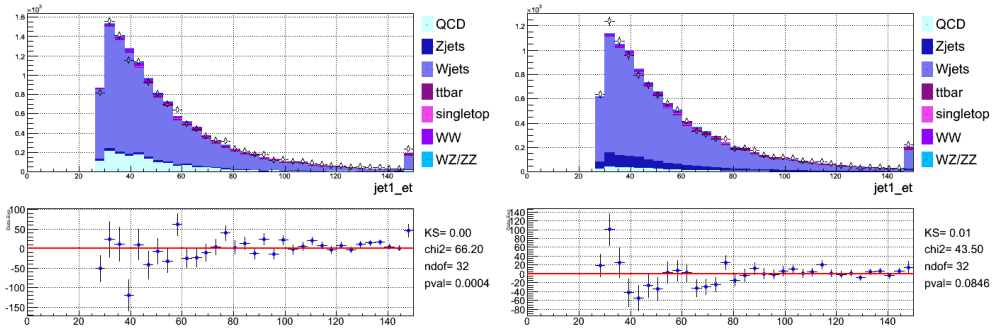

Figure 5.3: Left, `jet1_Et` (GeV) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

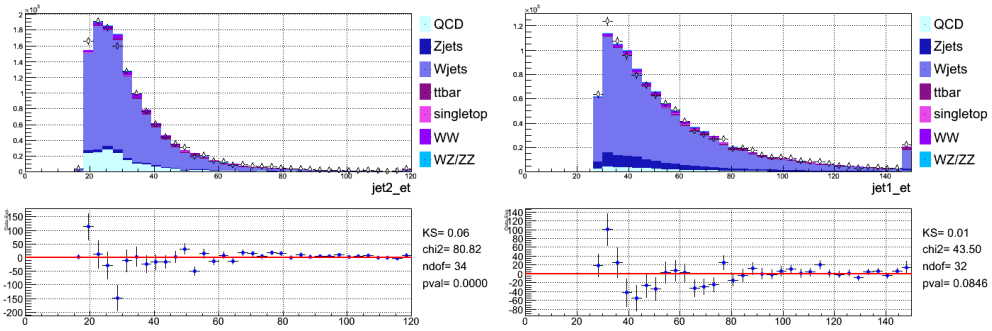

Figure 5.4: Left, `jet2_Et` (GeV) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

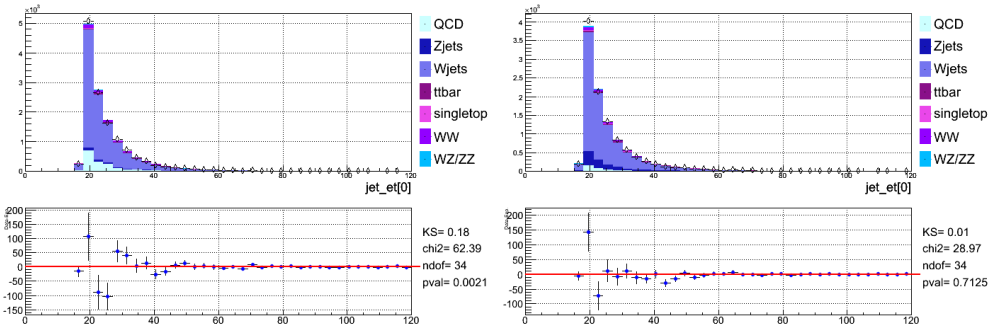

Figure 5.5: Left, `jet3_Et` (GeV) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

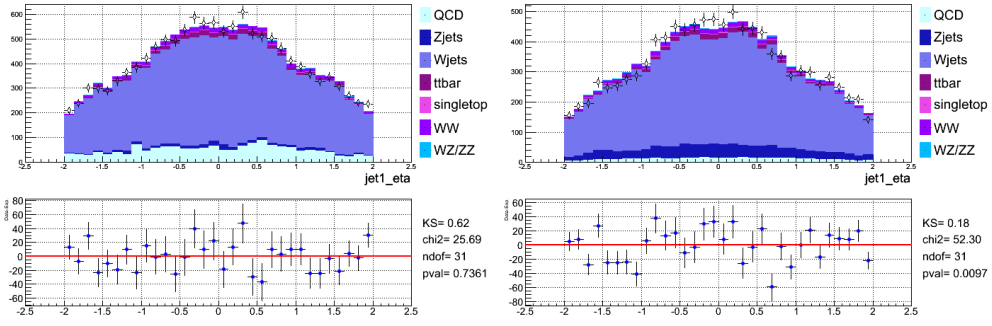

Figure 5.6: Left, `jet1_Eta` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

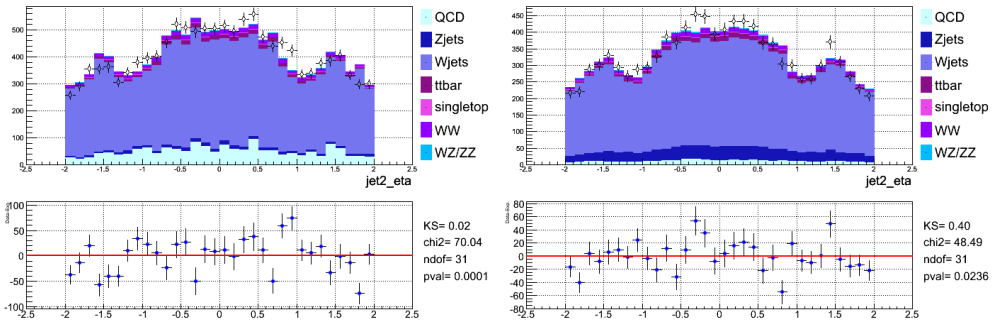

Figure 5.7: Left, `jet2_Eta` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

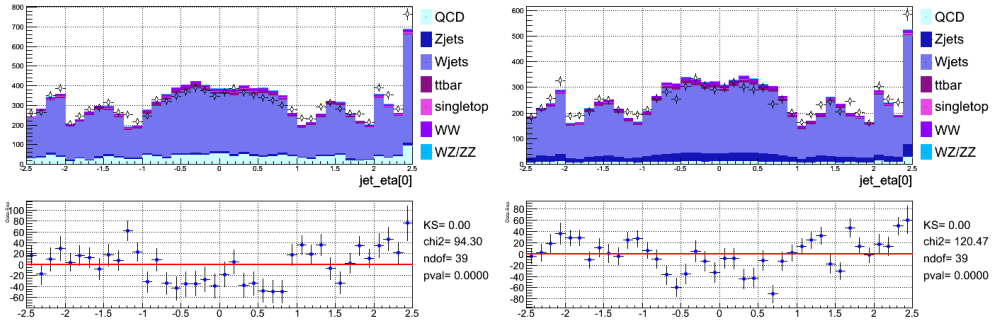

Figure 5.8: Left, `jet3_Eta` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

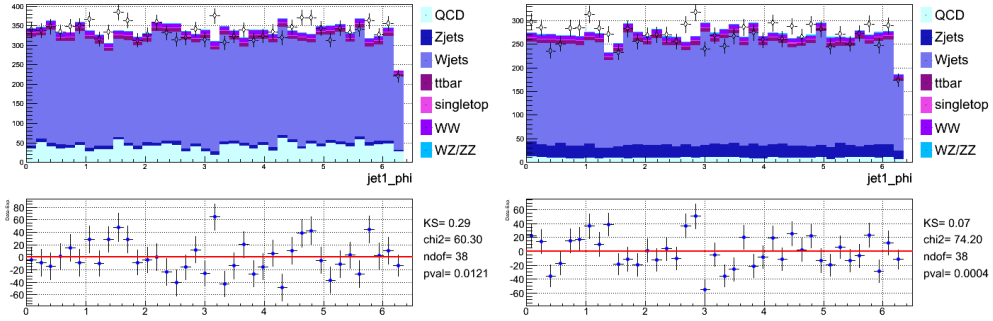

Figure 5.9: Left, `jet1_Phi` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

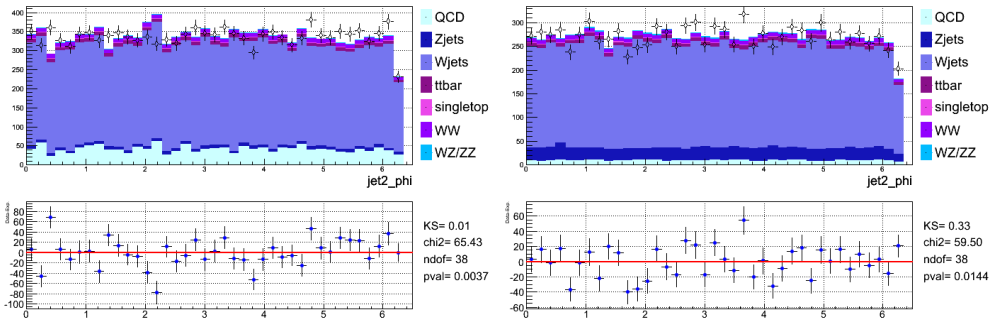

Figure 5.10: Left, `jet2_Phi` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

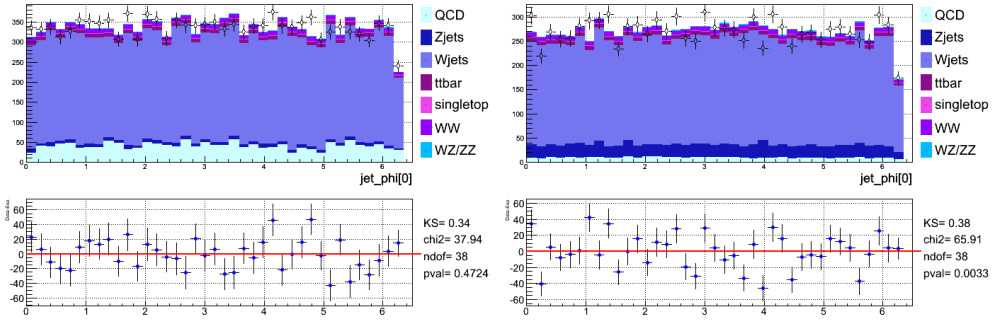

Figure 5.11: Left, `jet3_Phi` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

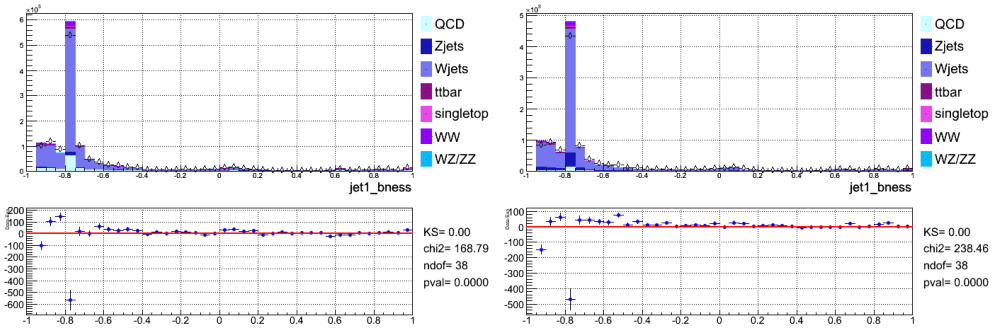

Figure 5.12: Left, `jet1_bness` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

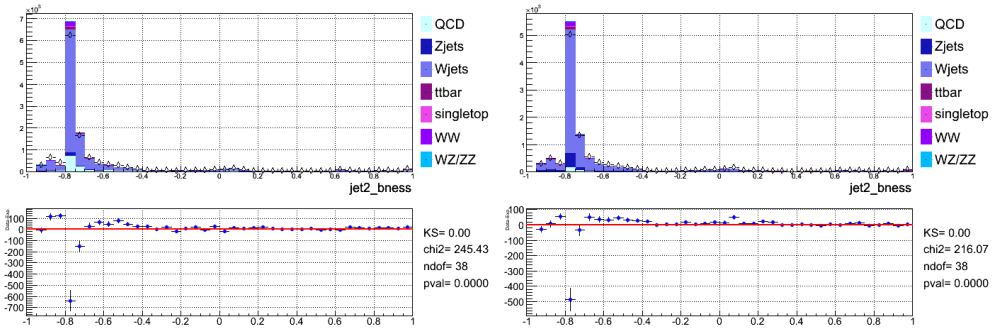

Figure 5.13: Left, `jet2_bness` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

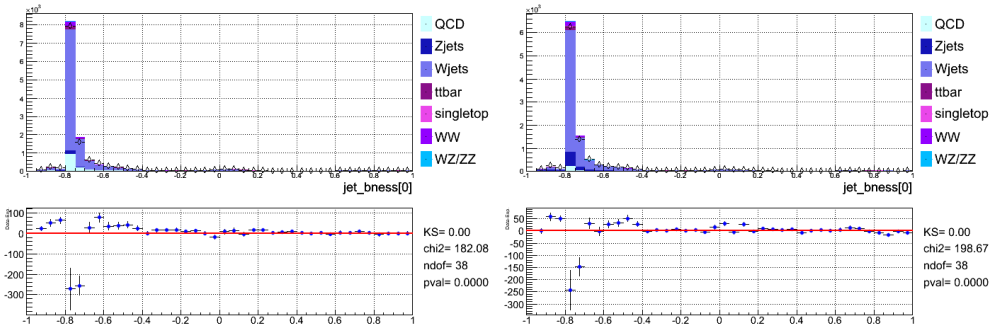

Figure 5.14: Left, `jet3_bness` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

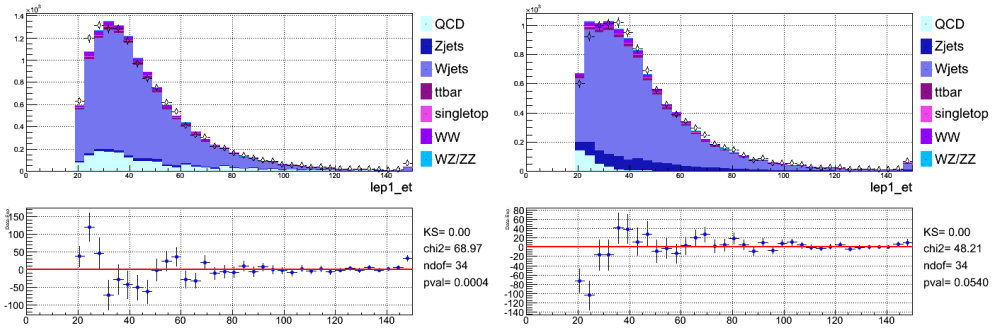

Figure 5.15: Left, `lep_et` (GeV) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

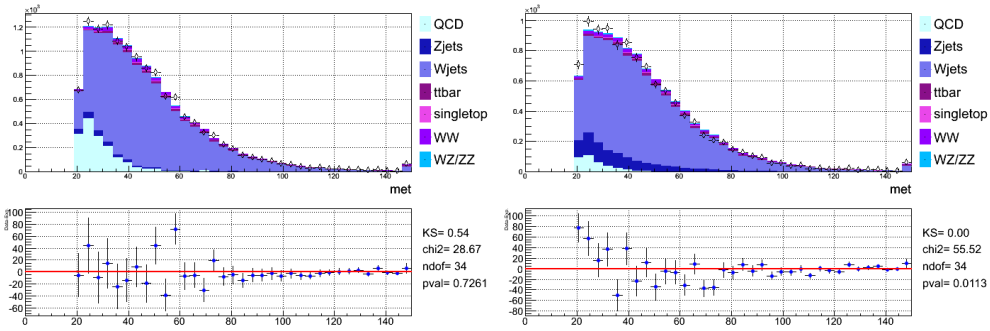

Figure 5.16: Left, $\not\!\!E_T$ (GeV) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

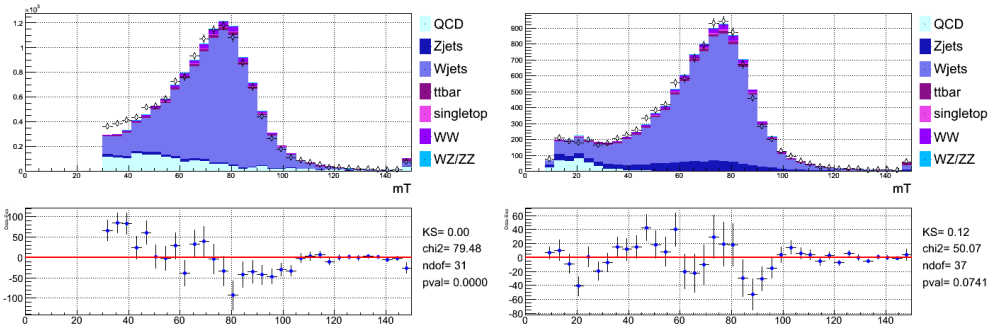

Figure 5.17: Left, M_T^W (GeV/c^2) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

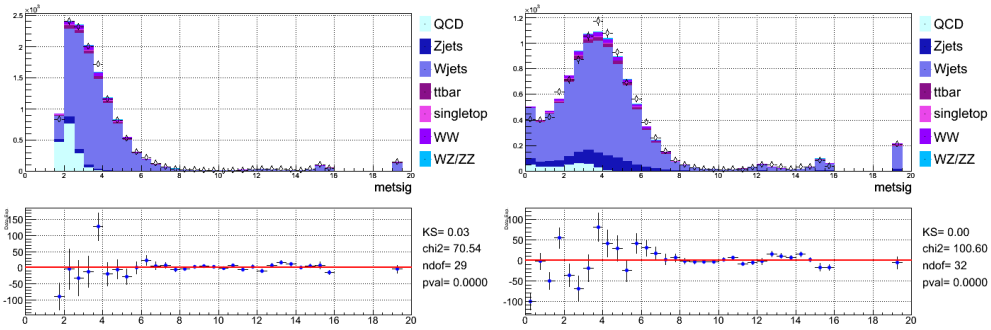

Figure 5.18: Left, `metsig` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

5.2.2 Tag sample

In this section we plot the same variables listed for the *notag* sample in order to validate the MC, the QCD templates and the expected rate in the *tag* sample, (see Figs. 5.19-5.31).

- `jet1_Et_et`, `jet2_Et`, `jet3_Et`: the transverse energies of the first/second/third *b*ness-leading jet.
- `jet1_Eta`, `jet2_Eta`, `jet3_Eta`: η of the first/second/third *b*ness-leading jet.
- `jet1_Phi`, `jet2_Phi`, `jet3_Phi`: ϕ of the first/second/third *b*ness-leading jet.
- `lep_et`: the transverse energy of the highest-E_T lepton
- $\not\!\!E_{\mathrm{T}}$
- $\mathrm{M}_{\mathrm{T}}^{\mathrm{W}}$: invariant mass of the highest-E_T lepton and $\not\!\!E_{\mathrm{T}}$ system in the transverse plane
- `metsig`.

The agreement is satisfactory in all distributions.

Process	Rate (Electrons)	Rate (Muons)
Signal	3.5 ± 0.2	3.6 ± 0.2
WW	6.2 ± 0.4	4.7 ± 0.3
$t\bar{t}$	146.4 ± 0.9	127.9 ± 0.8
single-top	22.5 ± 0.2	18.7 ± 0.2
Z+jets	8.0 ± 0.4	23.6 ± 0.6
W+jets	212.0 ± 3.9	189.9 ± 3.2
QCD	32.5 ± 0.3	5.7 ± 0.0
Total Observed	431 ± 20.8	374.0 ± 19.3

Table 5.4: Predicted and observed number of events of the *tag* three jets sample. The expected rates are separated for the triggered lepton type. We also require the invariant mass of the two E_T-leading jets to be within [0,300] GeV/c^2. By construction the overall expected rates are the same as the observed ones in each region. The uncertainty shown is statistical.

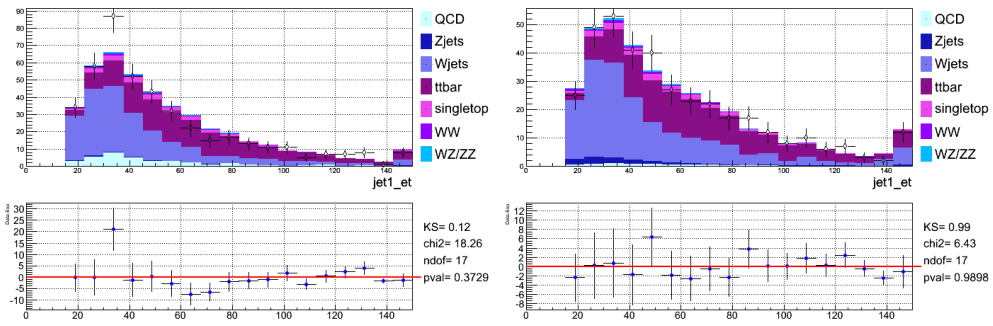

Figure 5.19: Left, `jet1_Et` (GeV) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

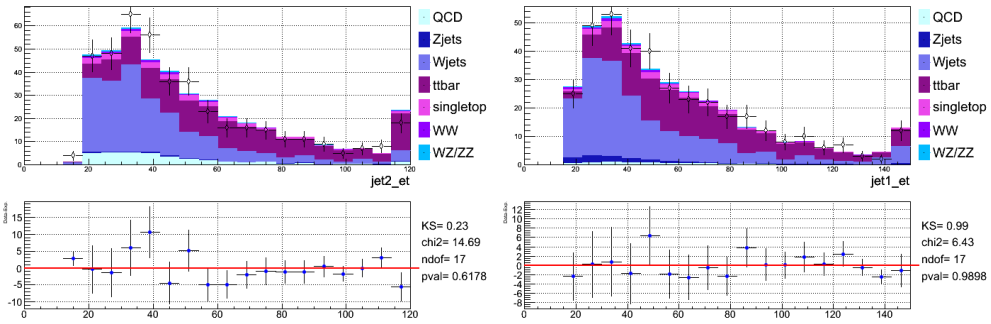

Figure 5.20: Left, `jet2_Et` (GeV) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

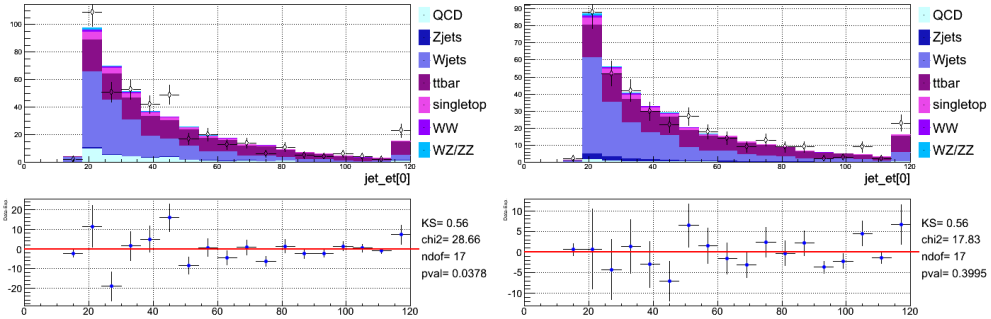

Figure 5.21: Left, `jet3_Et` (GeV) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

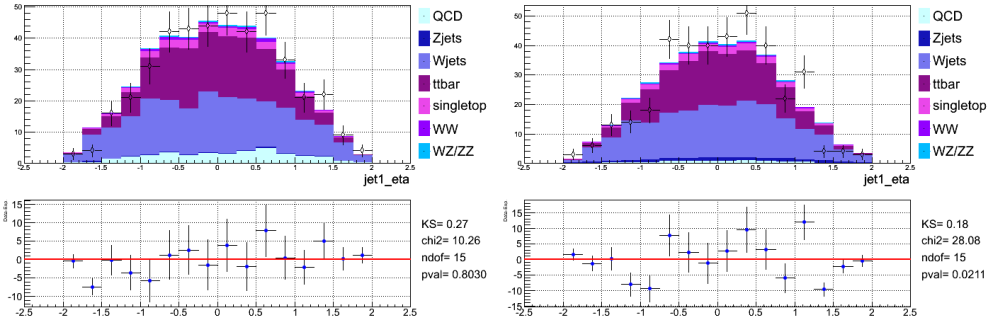

Figure 5.22: Left, `jet1_Eta` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

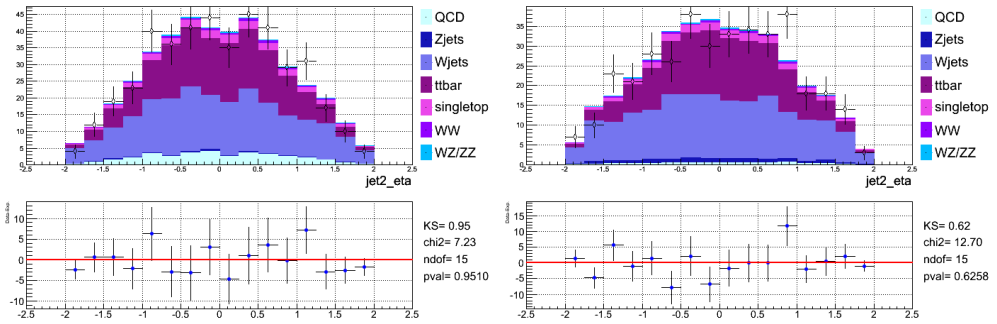

Figure 5.23: Left, `jet2_Eta` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

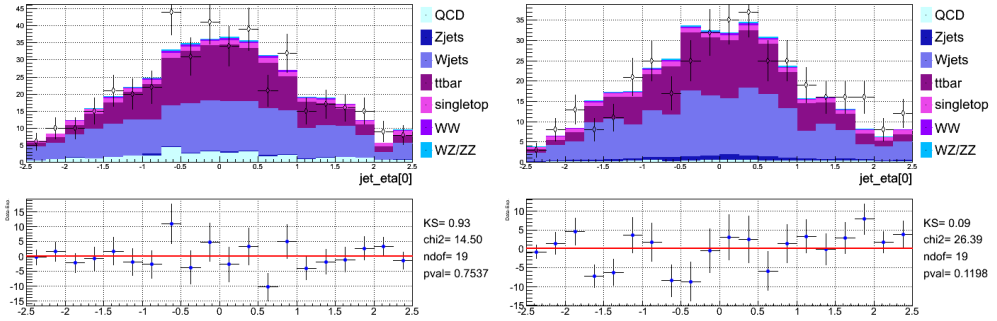

Figure 5.24: Left, `jet3_Eta` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

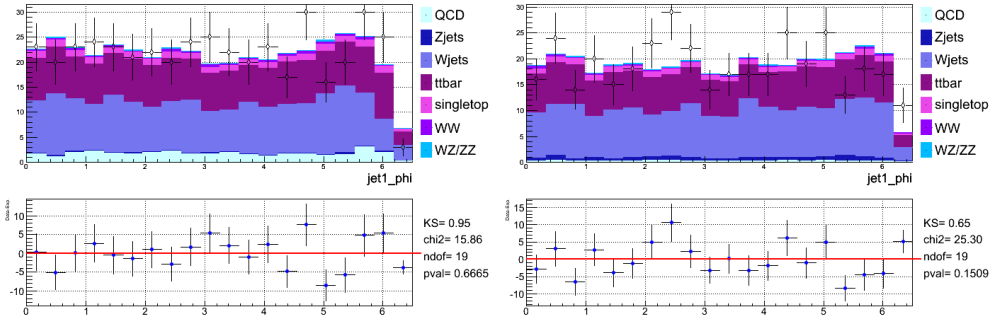

Figure 5.25: Left, `jet1_Phi` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

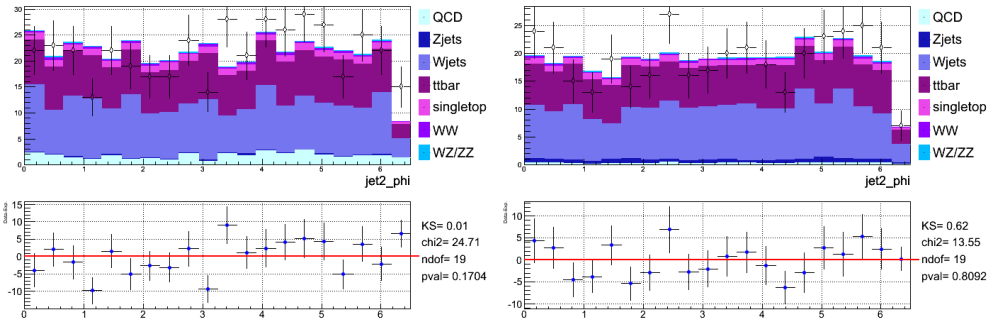

Figure 5.26: Left, `jet2_Phi` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

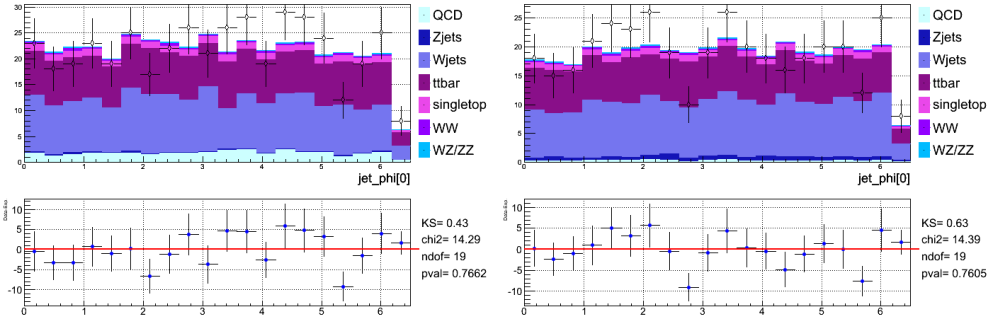

Figure 5.27: Left, `jet3_Phi` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

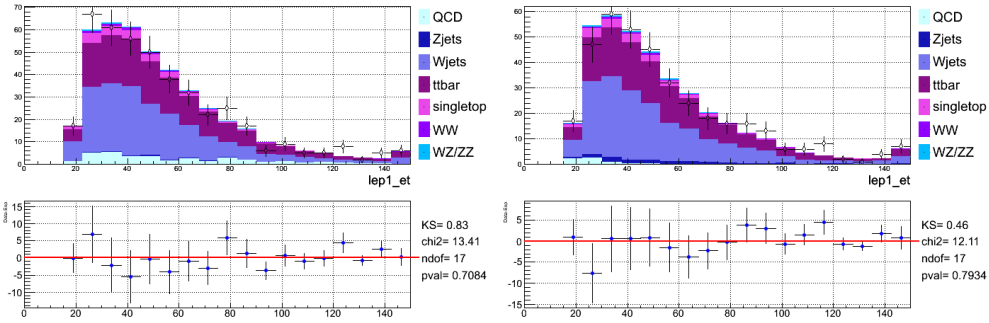

Figure 5.28: Left, `lep_et` (GeV) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

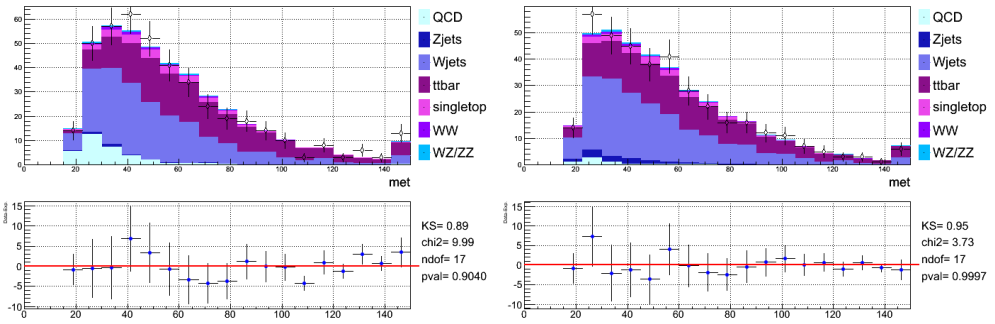

Figure 5.29: Left, $\not\!\!E_T$ (GeV) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

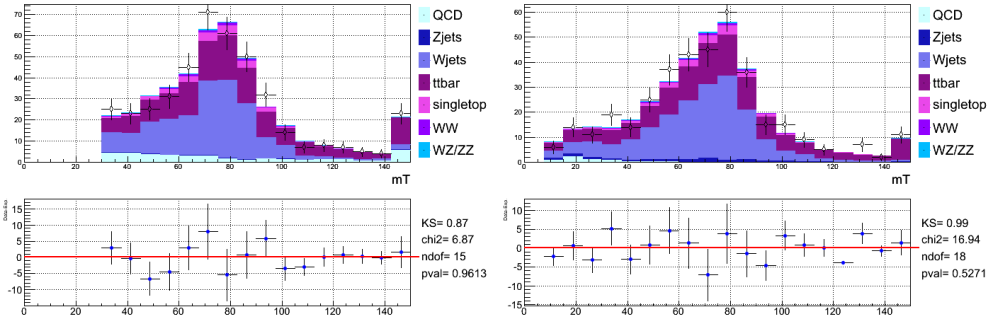

Figure 5.30: Left, M_T^W (GeV/c^2) distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

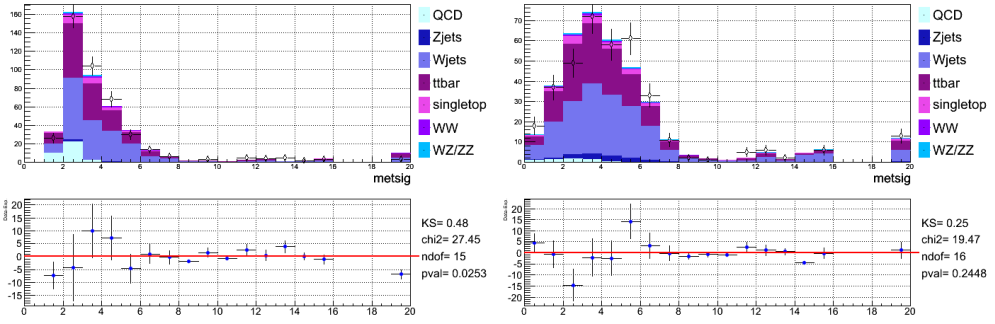

Figure 5.31: Left, `metsig` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

Chapter 6

Three jets sample Strategy

In this chapter the "novel technique" with which the Z-mass is reconstructed in the three jets sample is described.
In order to increase the acceptance of the WZ/ZZ analysis we are interested in events in which three high energetic jets are found. Our goal is to resolve combinatoric problem for building the Z mass and consequently improve the resolution of the invariant mass distribution in this sample. This would be fundamental since a fit of this distribution will be used for extracting the WZ/ZZ signal.

6.1 Motivations

A standard kinematical cut requests exactly two high energy jets (i.e. $E_T > 20$ GeV) in the candidate sample. Simulations show that if a third high energy jet is allowed the signal acceptance is increased by $1/3$. Therefore, it would be extremely important to be able to detect the Z signal also in events with more than two high energy jets.
A serious difficulty is that the signal to background ratio is very poor, due primarily to the contribution of associated production of W and incoherent jets. Optimal mass resolution of jet systems is of utmost importance for discriminating this background, since a fit to the invariant mass distribution of the jets associated to the hadronic decay of Z is used to disentangle the diboson signal from the backgrounds.

6.2 Origin of the extra Jet

In this thesis we investigate at generator level a sample of simulated CDF WZ events for finding a means to determine the origin of the extra jet and the right jet combination to be chosen for the best reconstruction of the Z mass. This is attempted for the first time in CDF [3, 4, 5]. In the case of WZ events, additional jets may be initiated by gluon(s) radiated from the interacting partons (Initial State Radiation, ISR) or from the Z-decay products (Final State Radiation, FSR).

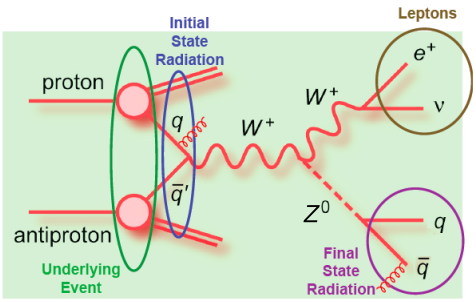

Figure 6.1: Feynman diagram in which both ISR and FSR are shown.

Extra-activity produced by spectator partons or by pile-up of events is negligible in our studies. Details on how we recognize ISR and FSR jets in simulation are given in Appendix E.

6.3 Three jets Region

In diboson analyses the standard kinematical cut at CDF requires the number of jets to be exclusively two, *two jets region*. Precisely, events with a third jet with $E_T > 15$ GeV are rejected.

The studies presented in this thesis are performed on the sample in which we allow for a third jet with $E_T > 15$ GeV[1], the *three jets region*.

About 33% diboson events lye in the three jets region, but jets due to initial or final state radiation confuse the choice of the jet system to be attributed to Z decay.

In Fig. 6.2 the invariant mass built using the two E_T leading jets for WZ MC events in the *two jets region* is compared with the same distribution built in the *three jets region*. In the sample

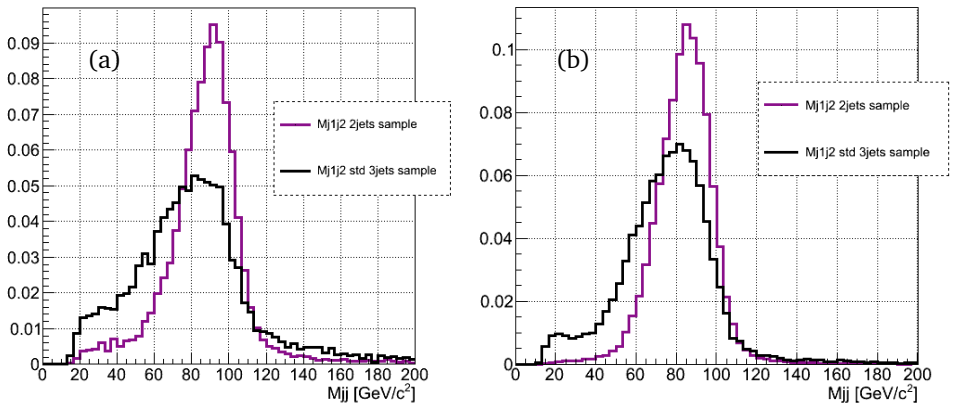

Figure 6.2: The black distribution is dijet mass built with the two leading jets in the *three jets region*, while the violet one is the invariant mass in the *two jets region*, in *notag* (a) and *tag* (b) sample.

[1]We exclude events with a fourth jet with $E_T > 10$ GeV.

where three jets are found MJ1J2 has a degraded resolution, high mass and low mass tails due to wrong combinations are present.
Choosing the correct jet combination coming from Z for building the Z mass would improve the resolution.

6.3.1 The importance of knowing the correct jet system

We started from studying the three jets sample in WZ MC. Jets are ordered in decreasing E_T in *notag* sample and in decreasing $b$ness in *tag* sample[2].
We investigate at generator level the origin of the not-matched jet (NMJ) in order to understand the Right Jet Combination (RJC). Jets are matched in direction to quarks from Z decay, requiring $\Delta R < 0.4$ between partons from Z and jets.
When a number of matches different from 2 is found the event is not considered. In $\sim 30\%$ of cases two or more jets are not matched. We are presently studying how to rescue these events by implementing a more efficient matching algorithm. This algorithm will search for hadrons rather than quarks in the jet cone and will trace back the origin of the hadrons in order to understand if they were produced by a Z-decay.
In Fig. 6.3 the invariant mass distribution is shown in the sample in which we require both jets to be matched to quarks from Z, in the whole sample and in the subsample in which there is no double-match. The MJ1J2 distribution in the sample which is rejected for our studies, i.e. the one with no double-match, has the distribution peak shifted to a mass region lower than expected $Z-$mass value. We expect that improving the matching algorithm will improve our NNs efficiency, since it will allow train NNs with a larger set of correctly assigned events (the blue distribution in Fig. 6.3).

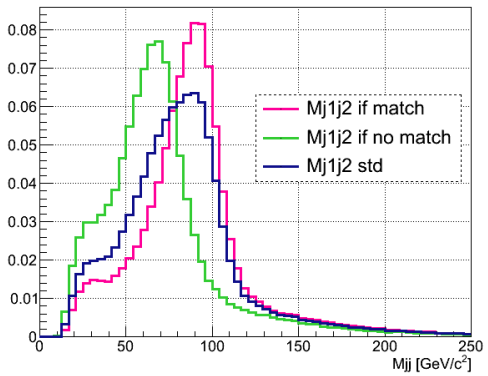

Figure 6.3: Invariant mass distribution in the 3 jets region built with the two E_T-leading jets in the whole *notag* sample, in the subsample in which the double-match requirement is satisfied and in the complementary one.

[2] J1, J2 would be the two with highest $b$ness value, J3 the one with highest E_T among the others.

Notag

In terms of the frequency of RJC the pretag sample is composed as follows (see Fig. 6.6):

1. NMJ = J3 is from ISR ↦ RJC = J1J2 - 35% of events
2. NMJ = J2 is from ISR ↦ RJC = J1J3 - 21% of events
3. NMJ = J1 is from ISR ↦ RJC = J2J3 - 10% of events
4. NMJ is from FSR ↦ RJC = J1J2J3 - 19% of events

The best resolution we can get in this sample is shown in Fig. 6.4-6.5, where we compare the invariant mass built using the proper RJC for each event with the distribution built with the two E_T leading jets (Fig. 6.4) and with the dijet mass in the *two jets region* (Fig. 6.5). The low and high mass tails affecting the MJ1J2 distribution are drastically reduced by choosing the correct combination. Moreover, choosing the correct combination would allow us to have in the three jets region the same resolution as in the two jets region.

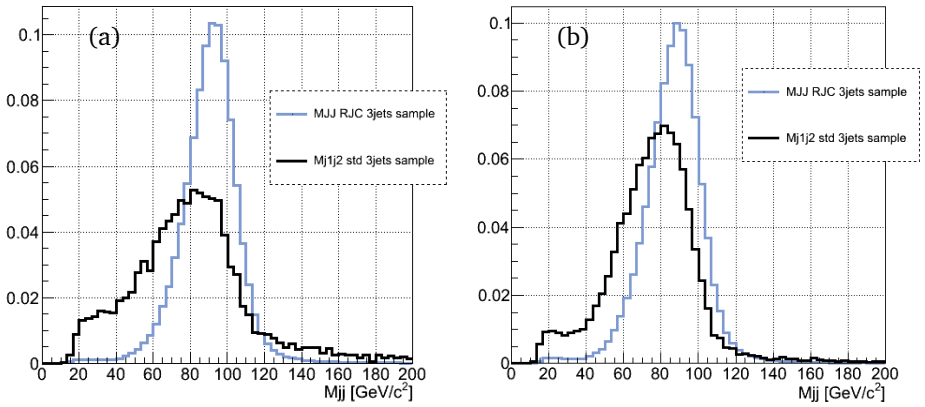

Figure 6.4: The blue invariant mass is built using the two 'Z-jets' when the 3^{rd} jet is from ISR and combining the 3 jets if FSR. The black distribution is dijet mass built with the two leading jets, in *notag* (a) and *tag* (b) sample.

Tag

In terms of the frequency of RJC the tag sample is composed as follows (see Fig. 6.6):

1. NMJ = J3 is from ISR ↦ RJC = J1J2 - 44% of events
2. NMJ = J2 is from ISR ↦ RJC = J1J3 - 17% of events
3. NMJ = J1 is from ISR ↦ RJC = J2J3 - 7% of events
4. NMJ is from FSR ↦ RJC = J1J2J3 - 19% of events

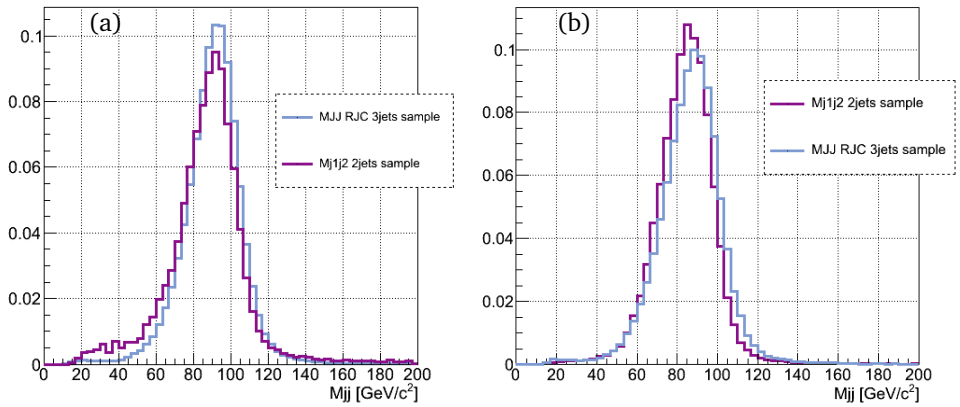

Figure 6.5: The blue invariant mass is built using the two 'Z-jets' when the 3^{rd} jet is from ISR and combining the 3 jets if FSR; the violet distribution is the invariant mass in the tight dijet sample, in *notag* (a) and *tag* (b) sample.

Notice that in *tag* sample MJ1J2 is the RJC in the $\sim$ 44% of cases, since jets are ordered in *b*ness and we require the two *b*ness leading jets to satisfy some criteria, as specified in chapter 4. The greater contribution of MJ1J2 in the whole sample is the reason why in the *tag* sample the resolution is already good for the distribution built with the two jets with highest *b*ness. As a consequence, we expect that there is less to improve in this sample.

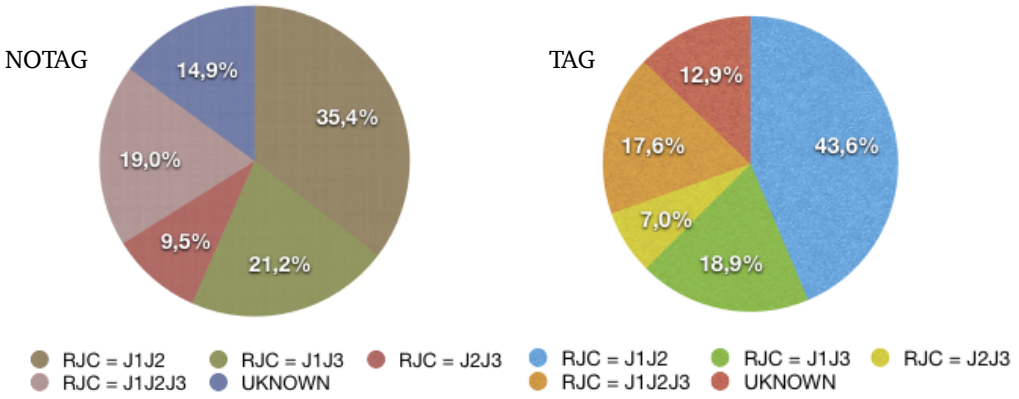

Figure 6.6: Notag and tag sample composition in terms of RJC.

The best resolution we can get in this sample is shown in Fig. 6.4-6.5, where we compare the invariant mass built using the proper RJC for each event with the distribution built with the two E_T leading jets (Fig. 6.4) and with the dijet mass in the *two jets region* (Fig. 6.5).

6.3.2 Adopted strategy: Neural Networks

Four different Neural Networks (NNs) are trained in MC signal events to isolate each of the above cases: MJ1J2, MJ1J3, MJ2J3 and MJ1J2J3.
These NNs combine kinematical information and some tools developed by CDF collaboration for isolating gluon-like and b-like jets from light-flavored jets. Using NNs outputs, we have developed a set of criteria to decide event by event which is the most likely correct jet combination to be used for building the Z-mass.

Neural Networks are more and more used in various fields for data analysis and classification, both by research and by commercial institutions. In particle physics they are used mainly for classification tasks, i.e. signal over background discrimination.
NNs are non-linear statistical data modeling tools, used to model complex relationships between inputs and outputs or to find patterns in data [91].
A Neural Network is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information, using a connectionist approach to computation. The key element is the novel structure of the information processing system, which consists of an interconnected group of artificial *neurons* (this is why they are named networks), working in unison to solve specific problems.
NNs, like people, learn by example. The examples must be selected carefully otherwise time may be wasted or even worse the network might be functioning incorrectly. The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable.
So our role has been to chose representative data and then exploit training algorithms for helping the NN to learn the structure of the data.

MultiLayer Perceptrons

More than 50% of neural networks are **MultiLayer Perceptrons** (MLP). The MLP is a simple feed-forward network: signals flow from inputs, forwards through any hidden units, eventually reaching the output units. Such a structure has stable behavior.
An example system has three layers, see Fig. 6.7. The first layer has input neurons, which send data via synapses to the second layer of neurons, and then via more synapses to the third layer of output neurons. The synapses store parameters called *weights* that manipulate the data in the calculations.
Each neuron in any subsequent layer first computes a linear combination of the outputs of the previous layer. The output of the neuron is then function of that combination.
More details about activation function and MLP methods are given in [89].

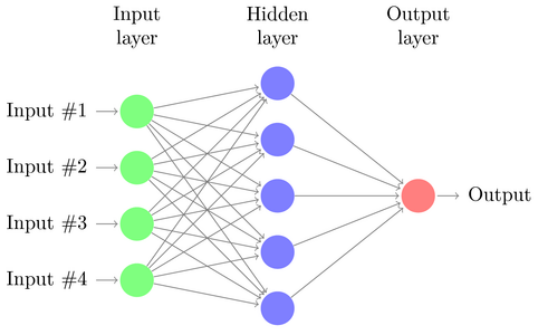

Figure 6.7: Multilayer perceptron with one hidden layer.

6.4 The “Novel technique” in notag sample

In the following sections are described the four NNs used in the *notag* sample, while in Appendix G are shown their input variables. In Appendix J.1 NNs outputs and MJJ distributions in data and simulated events and QCD events are shown.

6.4.1 Exploring MJ1J2: NN_{12}

In order to isolate events when RJC = J1J2 we analyze differences of some variables in two subsamples:

- RJC = J1J2
- Other jet combinations (RJC = J1J3, J2J3, J1J2J3, UNKNOWN) which we name "OJC"

Below is the list of the variables used:

1. $m_{jj'}/m_{j_1j_2j_3}$[3]
2. $\gamma_{jj'} = (E_j + E_{j'})/m_{jj'}$
3. $\mathrm{d}\eta_{jj'}$
4. $dR_{jj'} = \sqrt{d\eta^2_{jj'} + d\phi^2_{jj'}}$
5. $dR_{j_1j_2,j_3}$, dR between the third jet and vectorial sum of the two leading jets.
6. $dR_{j_1j_2j_3,j_3}$, dR between the third jet and vectorial sum of the three jets.
7. Quark Gluon Discriminator for J1, J2 used for discriminating a quark from a gluon jet. For further details see [90].

[3] jj’ refers to the three possible combinations: J1J2, J1J3 and J2J3.

In order to avoid background (W+jets, $t\bar{t}$, etc...) being sculpted later, the input variables are weighted. Weights applied are calculated such that the MJ1J2 distribution in RJC = J1J2 and OJC samples become approximately the same. By doing this we will decorrelate the NN_{12} output from the numerical value of MJ1J2 and make it sensitive only to the difference in kinematical distributions of the involved variables. In Fig. 6.8 the weights used and MJ1J2 distributions before and after normalization in both subsamples are shown.

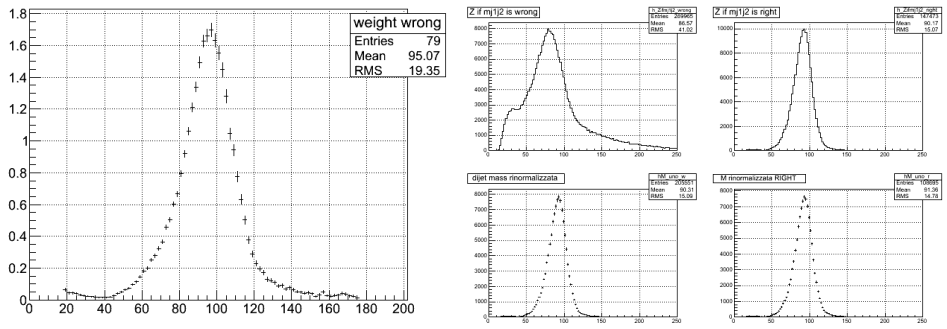

Figure 6.8: MJ1J2 distribution in the two subsamples before and after normalization (Right). Weights used for normalizing (Left). The horizontal scale is in GeV/c^2. Above 110 GeV/c^2 and below 80 GeV/c^2 weights are lower than 1 in order to reduce the tails.

The variables described above are weighted accordingly and are used for training a Neural Network, employing the MLP method.
The NN_{12} response is shown in Fig. 6.9[4].

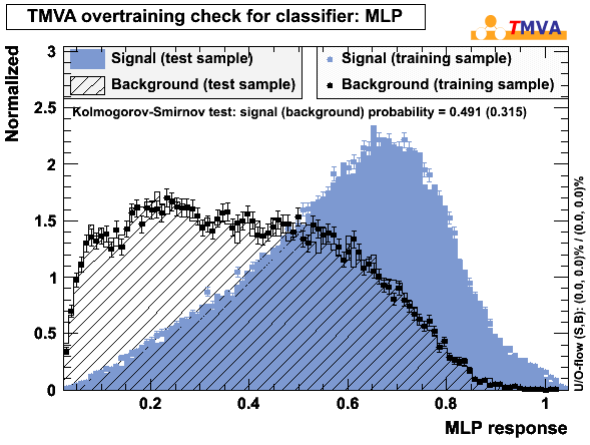

Figure 6.9: NN_{12} MLP response.

[4]The training data contains examples of signal inputs together with background inputs, and the network learns to infer the relationship between the two. Test data are used for checking the progress of the algorithm.

6.4.2 Exploring MJ1J3: NN_{13}

As described before for NN_{12}, in order to isolate events when RJC = J1J3 we analyze differences of some variables in two subsamples, which now are defined as:

- RJC = J1J3
- Other jet combinations (RJC = J1J2, J2J3, J1J2J3, UNKNOWN) which we name "OJC"

Below is the list of the variables used:

1. $m_{jj'}/m_{j_1j_2j_3}$[5]
2. $\gamma_{jj'} = (E_j + E_{j'})/m_{jj'}$
3. $\mathrm{d}\eta_{j_1j_2}$, $\mathrm{d}\eta_{j_2j_3}$
4. $dR_{j_1j_2}$, $dR_{j_2j_3}$
5. $\eta(j_1 + j_3)/\eta(j_2)$
6. “pt-imbalance” : $p_{Tj_1} + p_{Tj_3} - p_{T\ell}-$MET
7. EMfr for j_2 which is the ratio between EM and total energy.
8. $dR_{j_1j_2,j_3}$, $dR_{j_2j_3,j_1}$
9. $dR_{j_1j_2j_3,j_2}$
10. $dR_{j_2,\ell}$
11. Quark Gluon Discriminator for J2

Even in this case we decorrelate the NN_{13} output from the numerical value of MJ1J3 and make it sensitive only to the kinematical distributions of the involved variables (see Fig. 6.10). The NN_{13} response is shown in Fig. 6.11.

6.4.3 Exploring MJ2J3: NN_{23}

In this case the two subsamples are defined as:

- RJC = J2J3
- Other jet combinations (RJC = J1J2, J1J3, J1J2J3, UNKNOWN) which we name "OJC"

Input variables in this case are:

1. $m_{jj'}/m_{j_1j_2j_3}$[6]
2. $\gamma_{jj'} = (E_j + E_{j'})/m_{jj'}$

[5] jj' refers to the three possible combinations: J1J2, J1J3 and J2J3.
[6] jj' refers to the three possible combinations: J1J2, J1J3 and J2J3.

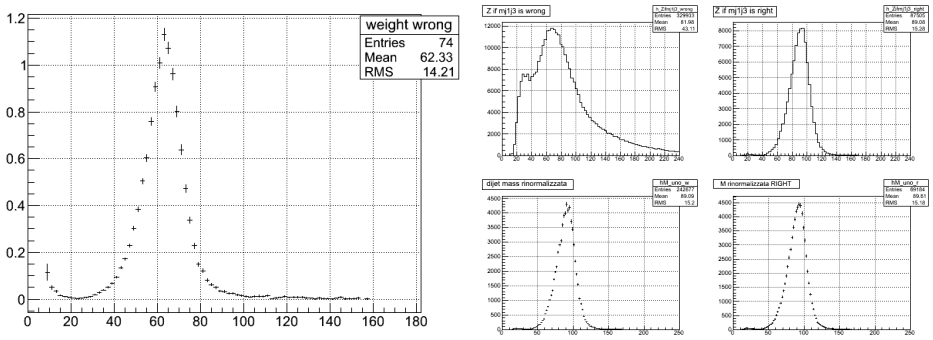

Figure 6.10: MJ1J3 distribution in the two subsamples before and after normalization (Right). Weights used for normalizing (Left). The horizontal scale is in GeV/c^2.

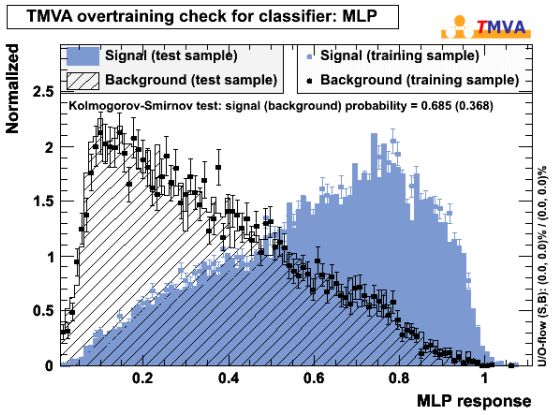

Figure 6.11: NN_{13} MLP response.

3. $d\eta_{j_1 j_2}$, $d\eta_{j_1 j_3}$
4. $p_T(j_2 + j_3)/p_T(j_1)$
5. $dR_{j_1 j_2}$, $dR_{j_1 j_3}$
6. $dR_{j_1 j_2, j_3}$, $dR_{j_1 j_3, j_2}$
7. $dR_{j_1 j_2 j_3, j_1}$
8. $dR_{j_1, \ell}$
9. Quark Gluon Discriminator for J1

Even in this case we decorrelate the NN_{23} output from the numerical value of MJ2J3 in order to make it sensitive only to the kinematical distributions of the involved variables (see Fig. 6.12). The NN_{23} response is shown in Fig. 6.13.

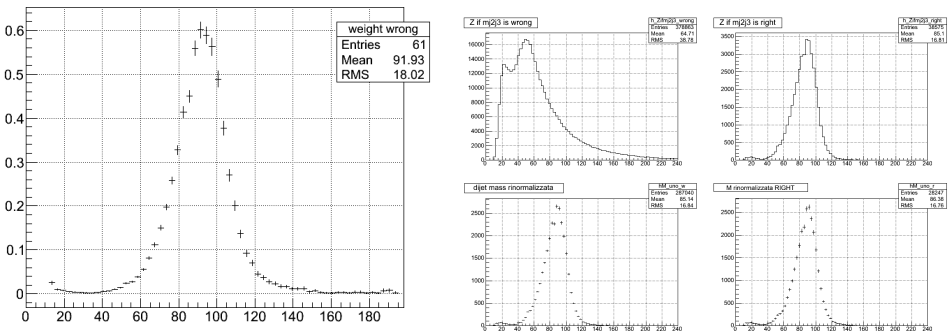

Figure 6.12: MJ2J3 distribution in the two subsamples before and after normalization (Right). Weights used for normalizing (Left). The horizontal scale is in GeV/c^2.

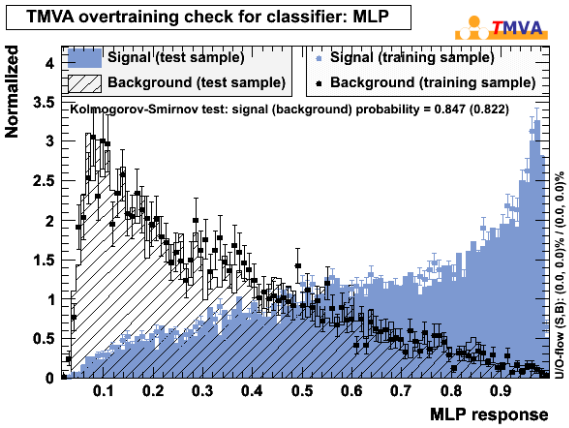

Figure 6.13: NN$_{23}$ MLP response.

6.4.4 Exploring MJ1J2J3: NN$_{123}$

NN$_{123}$ aims at isolating the events in which FSR has occurred and for which Z-mass should be reconstructed using all the three jets. The subsamples in this case are defined as:

- RJC = J1J2J3
- Other jet combinations (RJC = J1J2, J1J3, J2J3, UNKNOWN) which we name "OJC"

The input variables are:

1. $\gamma_{jj'} = (E_j + E_{j'})/m_{jj'}$ [7]
2. $\gamma = (E_{j_1} + E_{j_2} + E_{j_3})/MJ1J2J3$
3. “pt-imbalance” : $p_{T\,j_1} + p_{T\,j_2} + p_{T\,j_3} - p_{T\,\ell}$−MET

[7] jj’ refers to the three possible combinations: J1J2, J1J3 and J2J3.

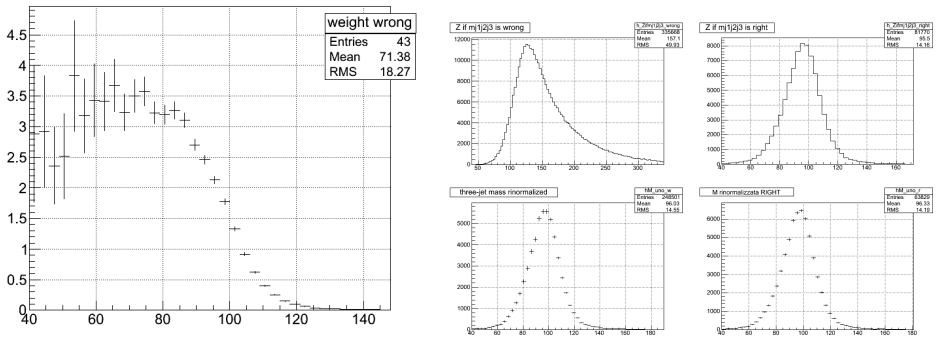

Figure 6.14: MJ1J2J3 distribution in the two subsamples before and after normalization (Right). Weights used for normalizing (Left). The horizontal scale is in GeV/c^2.

4. $\mathrm{d}\eta_{j_1 j_3}$, $\mathrm{d}\eta_{j_2 j_3}$
5. $dR_{j_1 j_3}$, $dR_{j_2 j_3}$
6. $dR_{j_1 j_3, j_2}$, $dR_{j_2 j_3, j_1}$
7. $dR_{j_1 j_2 j_3, j_3}$
8. $dR_{j_3 \ell}$
9. Quark Gluon Discriminator for J2, J3

In Fig. 6.14 are shown the weights applied to MJ1J2J3 distribution, calculated with the same criteria described for the other NNs. In Fig. 6.15 is shown the NN_{123} output.

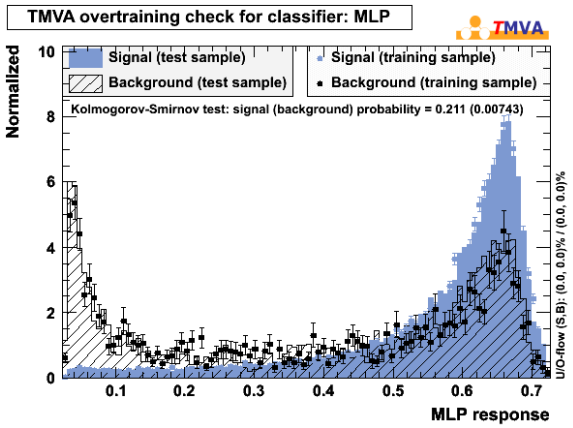

Figure 6.15: NN_{123} MLP response.

NN	MJJ_{COMB}
$NN_{12} > 0.43$	MJ1J2
$NN_{123} > 0.57$	MJ1J2J3
$NN_{13} > 0.72$	MJ1J3
$NN_{23} > 0.9$	MJ2J3

Table 6.1: Criteria used for building MJJ_{COMB} in the *notag* sample.

6.4.5 Criteria for the notag sample

For combining the information provided by the outputs of the four NNs, a criterion for building the invariant mass has been developed. We started with a requirement on NN_{12} and so we select the 35% of the sample where MJ1J2 is chosen for reconstructing the Z. Next we apply a requirement on NN_{123}, in order to select the subsample where MJ1J2J3 would be used. After we apply a cut on the NN_{13} and NN_{23} outputs.
We apply these cuts sequentially since we notice that NN's outputs are decoupled, see Appendix I. The above criterion (see Table 6.1) allows us to build a MJJ_{COMB} to be used in the *three jets region*, which is shown in Fig. 6.16 compared with MJ1J2 distribution in the *two jets region*. It is seen that an improvement in resolution is obtained. In order to understand

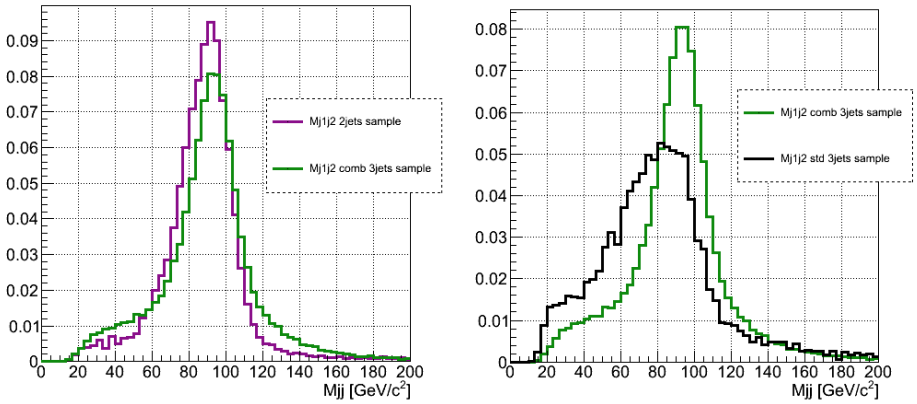

Figure 6.16: MJ1J2 in the *two jets region* (violet) and MJ1J2 in the *three jets region* (black) are compared with MJJ_{COMB} in the *three jets region*.

the impact of this method on the sensitivity of the measurement we apply the method also to the major sources of background (W+jets, Z+jets, $t\bar{t}$ and single top) and compare the result to WZ events. In Fig. 6.17 MJ1J2 and MJJ_{COMB} distributions are shown in signal and background events. The signal is multiplied by 100 in order to facilitate a visual comparison.

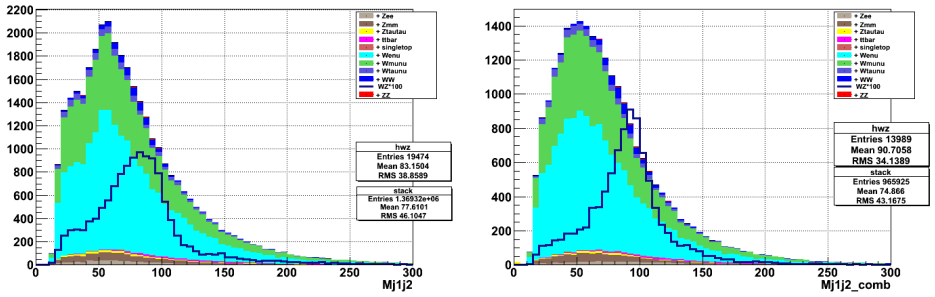

Figure 6.17: Simulation of signal+background. Left, MJ1J2. Right, MJJ_{COMB} built with the criteria described in the text. The horizontal scale is in GeV/c^2.

In Table 6.2 the acceptance, the *purity* p defined in 6.1 and the σ over μ ratio[8] are given.

$$Acc = \frac{Evt^{sel}}{Evt^{tot}} \quad p = \frac{MJJ^{RIGHT}}{Evt^{sel}} \tag{6.1}$$

	std	if **criteria**
Acc	100%	72%
p	35%	64%
σ/μ	0.27	0.13

Table 6.2: Performance of MJJ_{COMB} in the *notag* sample.

We note that as a consequence of the above mentioned decorrelation procedure, the invariant mass distribution in the background events doesn't get sculpted.

6.5 The "Novel technique" in tag sample

In the tag sample we used a very similar technique. Differences from the criteria developed in the *notag* sample are mainly in the variables used for training each NN.
Since we expect two b-jets in this sample we also use $b$ness information in our NNs. A detailed description of the NNs used in the tag sample is in Appendix H. In Appendix J.2 NNs outputs and MJJ distributions in data and simulated events and QCD events are shown.
Here we present only the criteria and the obtained results.

[8] σ and μ are estimated by a Gaussian fit in the mass window [70,110] GeV/c^2, see also Appendix F.1 for more details.

NN	MJJ_{COMB}
$NN_{12} > 0.57$	MJ1J2
$NN_{123} > 0.64$	MJ1J2J3
$NN_{13} > 0.7$	MJ1J3
$NN_{23} > 0.9$	MJ2J3

Table 6.3: Criteria used for building MJJ_{COMB} in the *tag* sample.

6.5.1 Criteria for the tag sample

The developed criteria (see Table 6.3) allow building a MJJ_{COMB} for the *three jets region*, as appropriate for the *tag* sample.
The obtained mass distribution is compared with MJ1J2 in the *two jets region* in Fig. 6.18. An improvement in resolution is obtained, although in this sample the resolution of MJ1J2 is already good. The persistence of the low mass tail in MJJ_{COMB}, presumably due to NN_{13} and NN_{23} inefficiency, will be studied in future work.

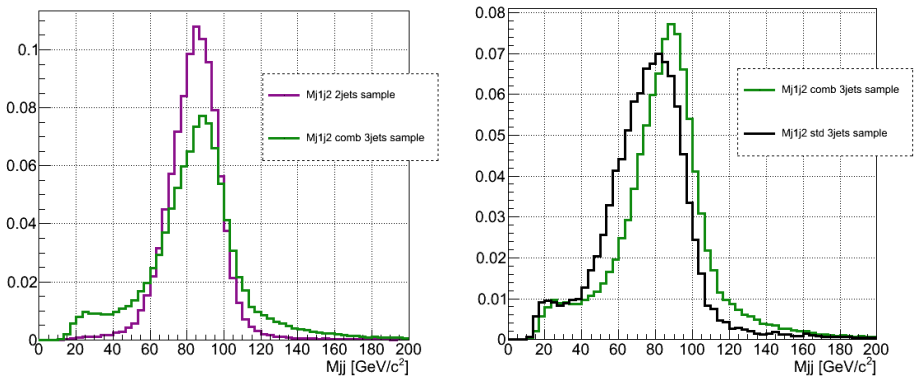

Figure 6.18: MJ1J2 in the *two jets region* (violet, Left) and in the *three jets region* (black, Right) are compared with MJJ_{COMB} in the *three jets region*.

In order to understand the impact of this method on the sensitivity of the measurement we build MJJ_{COMB} in the main sources of background (W+jets, Z+jets, $t\bar{t}$ and single top) and compare it to WZ events. In Fig. 6.19 MJ1J2 and MJJ_{COMB} distributions are shown in signal and background events. The signal is multiplied by 60 in order to facilitate a visual comparison.

	std	if **criteria**
Acc	100%	75%
p	39%	64%
σ/μ	0.21%	0.15%

Table 6.4: Performance of MJJ_{COMB} in the *tag* sample: acceptance, purity and resolution parameters.[9]

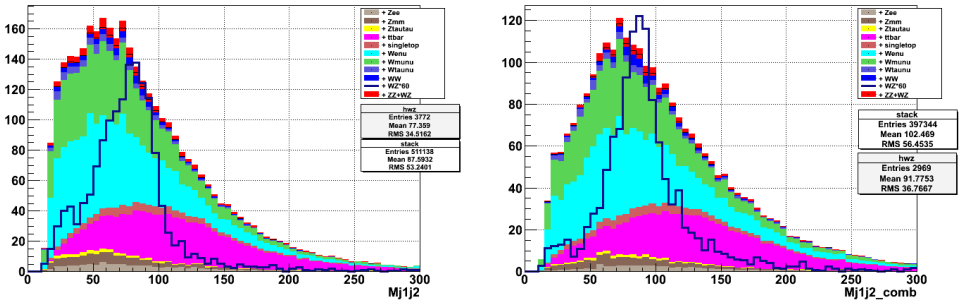

Figure 6.19: Simulation of signal+background. Left, MJ1J2. Right, MJJ_{COMB} built with the criteria described in the text. The horizontal scale is in GeV/c^2.

We note that as a consequence of the decorrelation procedure, the invariant mass distribution in the background events doesn't get sculpted under the signal.

Figures 6.17 together with Figs. 6.19 illustrate in a clear summary the progress obtained with our MJJ_{COMB} method.

Chapter 7

Results

The technique used to extract the diboson contribution is a fit of the dijet mass distribution to a combination of signal and backgrounds, whose modeling has been described in the previous chapter. In chapter 6 our technique to compute the invariant mass in the three jets region was described. This chapter presents the results on data. The systematic uncertainties included in the fit are discussed. At the end the significance of the measurement is given.

7.1 Signal Extraction from the 3-jets sample

We extract the signal from the background by fitting the invariant mass distribution in data events collected with the high-P_T triggers. The fit is performed simultaneously in a region with no b-tags (*notag*) and with two b-tags (*double-tag* or *tag*).
Fits to data (or pseudo-data) are performed using the `mclimit_csm.C` code [93]. Histograms for background and signal are inputed, as well as various systematics like rate and/or shape uncertainties.
We adopt two different fit methods in the *three jets* region:

1. we treat $WW/WZ/ZZ$ as our signal in the *pretag*[1] sample
2. we treat WZ/ZZ as our signal in events with no b-tags and events with two b-tags

The first method is just a test of our technique. Since WW/WZ signal has been observed in CDF [35, 36], it represents a standard candle to understand if using only the 3-jets sample a $WW/WZ/ZZ$[2] signal could be extracted. The second method aims at sensing a WZ/ZZ signal. We decide to treat separately the *notag* and *tag* regions and then combine the results in order to reach a greater sensitivity.

[1] No requirement on jet *b*ness: jets of all flavors are accepted.
[2] we expect the ZZ contribution to be negligible due to the requirement on $\not\!\!E_T$

The *templates* used are :

- W+jets, whose normalization is unconstrained in the fit and is independent in the two channels.
- $t\bar{t}$, single top and Z+jets (EWK), Gaussian-constrained to the theoretical/measured cross section with uncertainties of 6%.
- QCD, data driven estimate, Gaussian-constrained with a systematic uncertainty of 50% in the no-tag channel and its statistical uncertainty in the two-tag channel.
- When treated as a source of background, WW is scaled to the NLO cross section, Gaussian-constrained with an uncertainty of 6%.
- WZ/ZZ signal, with normalization allowed to float unconstrained in the fit, but unlike the W+jets background, with rates constrained to the expected ratio in the two channels.

7.2 The Test Statistic

In order to separate signal from background, typical CDF analyses use histograms of a number of observables in data events (discriminants). Various signal hypotheses may be tested, parameterized by mass, cross section, BR. Signals and background depend on a variety of parameters which may not be of primary interest but which are of use in the discriminant. Their values are useful in order to extract measurements of, or limits on, the parameters of interest, and uncertainty in their values usually results in reduced sensitivity in the measurement of the parameters of interest. The systematic errors on observables are parameterized in terms of these *nuisance parameters*.
Data must be compared to a model. The predictions of the model are uncertain because of the lack of precise knowledge of the nuisance parameters. If two models are being compared, a null hypothesis (background only) and a test hypothesis (signal + background), the uncertainties in the nuisance parameters can reduce the ability of the analysis to reject one of the hypotheses [92].
In addition, if a background histogram is allowed to float freely in a fit to match data in which a signal is actually present, the sensitivity for detecting the signal may be greatly reduced. This problem in encountered in our case in particular when handling the W+jets background, whose normalization cannot reliably be obtained from theory.
In order to evaluate the significance of an experimental result, the data must be compared to the expectations of the two hypothesis being tested, by means of a *test statistic*.
A test statistic is a single real number which is a function of the experimental outcome, and is chosen to maximize the separation between experimental outcomes expected when a signal is present and those expected when only the background contributes [93].

The likelihood ratio Q is defined as a ratio of probabilities P in the fit of the data sample:

$$Q = \frac{P(s+b)}{P(b)} \tag{7.1}$$

where s and b are the number of signal and background events respectively. The goal is to compute Q, the ratio of probability of observing the data in the test hypothesis to that in the null hypothesis. The accuracy in predicting Q is reduced if the background estimates in each hypothesis are poorly known.
A `MINUIT` [94] fit is usually applied to the observed spectrum, assuming either the signal+ background hypothesis or the background-only hypothesis and the consistency of the data with the fits is compared to see how well the two hypothesis agree with the data.
A χ^2 function is useful in comparing Poisson distributed data with a sum of contributions, each of which may have correlated and uncorrelated systematic uncertainties on rate and shape.
A general χ^2 function is computed, which includes all known uncertainties of the contributions. Once the number of degrees of freedom is known one can estimate the goodness of the fit using the value of χ^2 [95].
We use the χ^2 function to select one of the two hypotheses by taking the difference of the two χ^2 computed under the two hypotheses.
Computing a difference of χ^2 values for the two hypotheses H_1 and H_0 is equivalent to taking the logarithm of the likelihood ratio,

$$\Delta\chi^2 = \chi^2(H_1) - \chi^2(H_0) = -2\ln Q \tag{7.2}$$

7.3 Systematic uncertainties

Since this analysis relies so heavily on MC simulation, a large number of systematic uncertainties must be considered when checking if data and MC are consistent. Two classes of systematics are considered, those affecting the signal extraction (i.e. the number of signal events estimated in our data samples) and those affecting the acceptance and therefore the signal cross-section.
Table 7.1 shows the summary of the systematic uncertainties considered in the analysis.

7.3.1 Signal Extraction

We estimate the systematics on the signal extraction by generating pseudo-experiments (PE) using two additional models, corresponding to upward or downward fluctuations of the nuisance parameter for each systematic source. The pseudo-experiments are then fitted using the templates of the main fit on data. For each PE nuisance parameters are varied to get the best agreement between pseudo-data and fitted model. The difference between the

central value of the fit on data and the mean of the estimator of the signal content on the two corresponding pseudo-experiments is taken as systematic uncertainty on the corresponding source. For the signal extraction we consider the following sources of systematic uncertainty:

Jet Energy Scale (JES). The corrections used to set the JES are described in sec. 3.5.2. The effect of the JES uncertainty on the measurement is estimated by varying the energy of all jets in MC samples by $\pm 1\sigma$, where σ is the p_T-dependent uncertainty curve shown in Fig. 3.8. This procedure is applied at the same time to all the MC based processes. The evaluated systematic is given in Table 7.1.

b-tagger efficiency/mistag rate. Rather than applying scale factors to the tagging efficiency and mistag rate in order to make simulations to agree with data, we locate the cut in the MC computed efficiency parameter that matches the measured efficiency and mistag rates in the data. In the same manner, we determine the cuts on $b$ness in the MC that match the $\pm 1\sigma$ uncertainty values in the mistag rate and efficiency [80].

Renormalization and Factorization Scales in the W+jets MC (Q^2): the ALPGEN event generator used for W+jets events requires renormalization and factorization scales, Q^2, to be set to account for the finite order perturbative calculations of cross sections and for the factorization approximation of structure functions and cross sections.
Since the Q^2 values are not known, and indeed not physically measurable since they are an artifact of the theoretical approximation, an uncertainty is assigned to cover a variety of different possibilities. As a default, the renormalization and factorization scales are set to be the same to $Q^2 = {M_W}^2 + \sum p_T^2$, where M_W is the W boson mass and p_T^2 is the parton transverse momentum squared and the sum extends over all the final state partons [87]. This parameter is doubled and halved to create two samples which are used to determine the shape uncertainty on the W+jets template[3].

7.3.2 Cross Section Evaluation

In the cross section estimation, we assumed that the luminosity, the lepton trigger efficiency and the MC acceptance were exact. All of these assumptions are later considered as sources of systematics that contribute to the uncertainty of the cross section measurement, in addition to the signal fraction extraction systematics.

Jet Energy Resolution (JER): the modeling of the jet energy resolution (JER) can be a source of systematic uncertainty. It can affect the signal acceptance: if the dijet resonance is wider, more signal events may fall below the jet E_T thresholds or outside the dijet mass window used in the fit. The uncertainty in the jet energy resolution is found to vary as $\Delta(\sigma/p_T)$=(0.03±1.7)/p_T [GeV/c] [96]. Smearing the dijet mass due to energy resolution results in a 0.7% uncertainty in the measured cross section.

[3] Since the theoretical cross section of W+jets is only known to the lowest order in QCD and suffers from large uncertainties, its normalization is derived from data

Acceptance		Signal	EWK	W+jets	QCD	WW
JER		0.7%	0.7%	0.7%	no	0.7%
lumi		6%	6%	6%	no	6%
ISR/FSR		2.5%	2.5%	2.5%	no	2.5%
PDF		2%	2%	2%	no	2%
trigger		2.2%	2.2%	2.2%	no	2.2%
Systematic	**channel**	**Signal**	**EWK**	**W+jets**	**QCD**	**WW**
JES shape/rate	notag	yes/$\pm^{1\%}_{18\%}$	yes/$\pm^{3\%}_{13\%}$	yes	no	yes/$\pm^{2\%}_{17\%}$
	tag	yes/$\pm^{9\%}_{16\%}$	yes/$\pm^{13\%}_{18\%}$	yes	no	yes/$\pm^{10\%}_{17\%}$
$b$ness cuts rate	notag	$\mp^{11\%}_{12\%}$	$\mp^{12\%}_{13\%}$	$\pm^{17\%}_{19\%}$	no	$\mp^{6\%}_{18\%}$
	tag	$\pm^{0.7\%}_{0.7\%}$	$\pm^{2\%}_{2\%}$	$\pm^{0.3\%}_{0.3\%}$	no	$\pm^{0.0\%}_{0.3\%}$
Q^2	tag	no	no	yes	no	no
	notag	no	no	yes	no	no

Table 7.1: Summary of the cuts defining the candidate events samples.

ISR/FSR: this systematics affects the acceptance of the MC events. The PYTHIA showering algorithm is found to describe satisfactorily the spectrum of ISR/FSR, but the amount of ISR/FSR depends on input parameters, including the momentum scale of the interaction and the value of Λ_{QCD}, the energy scale at which perturbative quantum chromodynamics becomes not reliable. A systematic uncertainty is applied to predicted yields to account for an incomplete knowledge of these parameters. The impact of this uncertainty is evaluated by generating new simulated samples for WW/WZ signal varying the parameters within a plausible range of values. We determine a 2.5% systematic uncertainty due to more/less ISR and FSR.

PDF, the parton distribution functions (PDFs) used in generating the MC models, are determined by fitting a large number of experimental results which are subject to errors. We determine the change in the signal acceptance due to the PDF uncertainty to be 2% [97].

Luminosity and trigger efficiency. A 6% uncertainty on the cross section is assigned due to the uncertainty on the luminosity, as determined by the CDF luminosity counters. Uncertainties due to trigger efficiencies are calculated by varying the trigger scale factors[4] within their uncertainties and then applying the shifted scale factors in MC weight. A variation of 2.2% in the number of the expected events is found.

[4]Trigger scale factors are chosen run-by-run and during data-taking, and are used to weight MC events according to the corresponding trigger efficiency.

7.4 Sensitivity and Optimization

The `mclimit_csm.C` code provides a means of obtaining an estimate for the probability of a 2σ and 3σ measurement by generating pseudo-experiments and constructing $\Delta\chi^2$ distributions for test and null hypotheses. To reach acceptable accuracy we generated about 100,000 pseudo-experiments.
P2σ and P3σ are the probability of a 2σ (3σ) evidence assuming test hypothesis is true. They are computed as follows:

1. we generate 100,000 PEs in the null hypothesis and we compute the value of $\Delta\chi^2_{2\sigma}$ ($\Delta\chi^2_{3\sigma}$) such that the probability to have $\Delta\chi^2 < \Delta\chi^2_{2\sigma}$ ($\Delta\chi^2_{3\sigma}$) is 2.3%, (0.13%) [5]
2. we generate 100,000 PEs in the test hypothesis and we compute the integral from $-\infty$ to $\Delta\chi^2_{2\sigma}$ ($\Delta\chi^2_{3\sigma}$), corresponding to P2σ (P3σ)

In order to quantify the significance of the signal, i.e. the level of agreement between the data and the diboson hypothesis the corresponding *p-value* is computed.
The *p*-value represents the probability that the background fluctuate to a number of events equal or greater than the number of expected signal events.
We define as test statistic the $\Delta\chi^2$ defined in 7.2. Then:

1. we generate 100,000 PEs in the test hypothesis and we compute the median of $\Delta\chi^2$ distribution as an estimator of the most probable outcome in real data.
2. we generate 100,000 PEs in the null hypothesis and we compute the integral from $-\infty$ to the median of $\Delta\chi^2$ distribution.

The integral of the resulting distribution that exceeds the median is used to obtain the *p*-value.

7.4.1 WZ/ZZ/WW pretag in the 3-jets region

We estimate the probability at two and three standard deviations level to extract an inclusive diboson signal in the 3-jets sample alone (P2σ, P3σ). Systematic uncertainties are not yet included for generating PEs and for the fits of the pseudo-data. After our procedure for building the Z mass is applied, P3σ is about 4 times greater than when building the Z mass "by default" with the two E_T leading jets, see Table 7.2.
In Fig. 7.3 the $\Delta\chi^2$ distributions are shown for MJ1J2 and MJJ$_{COMB}$. The vertical lines show the value of $\Delta\chi^2_{2\sigma}$ ($\Delta\chi^2_{3\sigma}$) such that the probability to have $\Delta\chi^2 < \Delta\chi^2_{2\sigma}$ ($\Delta\chi^2_{3\sigma}$) is 2.3% (0.13%).

[5] 2.3% and 0.13% are the integrals in [0,2] and [0,3] of a normal distribution

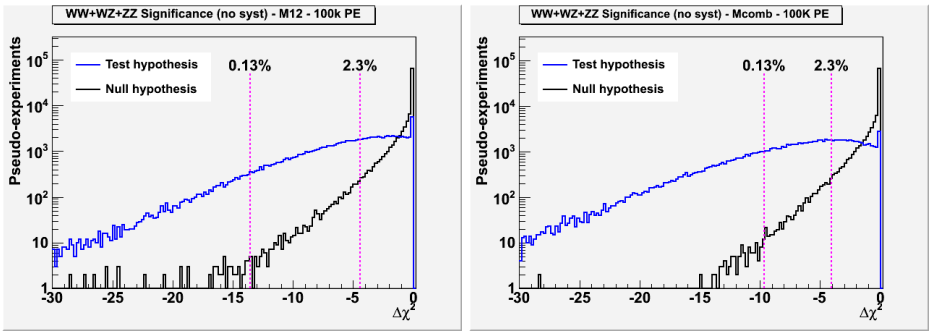

Figure 7.1: The $\Delta\chi^2$ distributions for null and test hypotheses for the fit to the $WZ/ZZ/WW$ signal, using MJ1J2 (left) and MJJ_{COMB} (right) as the invariant mass in the 3-jets region.

7.4.2 WZ/ZZ combined double tag+notag

As a first step of work in progress, we estimate the expected p-value to extract the WZ/ZZ signal in the 3-jets sample combining the information of the *notag* and *tag* channels. The results are shown in Table 7.2.

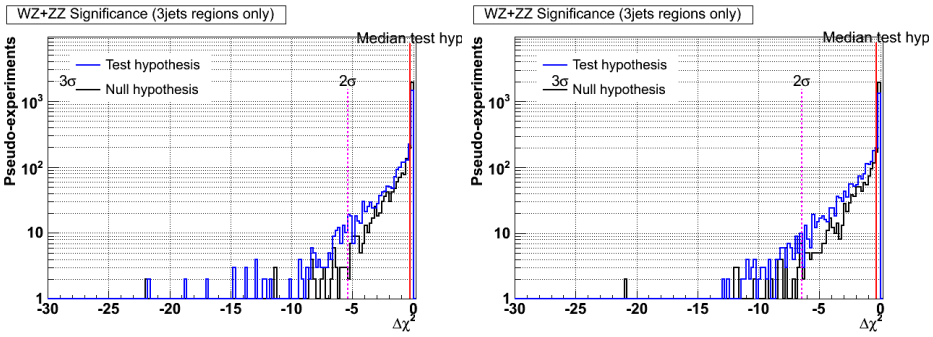

Figure 7.2: The $\Delta\chi^2$ distributions for null and test hypotheses for the fit to the WZ/ZZ signal, using MJ1J2 (left) and MJJ_{COMB} (right) as the invariant mass in the 3-jets *notag* and *tag* regions.

After applying our technique to build Z mass in the three jets region, the sensitivity increases although only a modest improvement is reached.
We expect this sample to be typo in the global CDF WZ/ZZ analyses in order to increase the acceptance. Then we estimate the expected p-value to extract a WZ/ZZ signal combining the information of the *notag* and *tag* channels in the 2-jets region including also the 3-jets sample. The expected p-value estimated considering only the *two jets region* is 0.75 σ [6]. We estimate a p-value $= 1.05\ \sigma$ including the *three jets region*.

Fit Method	P2σ	P3σ
$WZ/ZZ/WW$ pretag		
- MJ1J2	51.2%	6.4%
- MJJ_{COMB}	66.7%	25.9%
	p-value	
WZ/ZZ notag+tag		
- MJ1J2	0.35 σ	
- MJJ_{COMB}	0.45 σ	

Table 7.2: Sensitivity for the different performed fits considering only the 3-jets region.

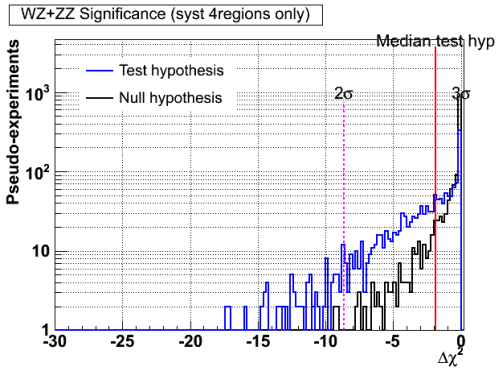

Figure 7.3: The $\Delta\chi^2$ distributions for null and test hypotheses for the fit to the WZ/ZZ considering both *notag* and *tag* sample combining the 2-jets and 3-jets samples.

7.5 Concluding Comments

To qualify the potential of the method we have studied an experimental data sample accepting events with a leptonically decaying W and 3 large transverse momentum jets, as in the studies of the simulated WZ sample. The selection cuts accept jets of all flavors (*pretag* sample), and all diboson events including WW besides WZ, ZZ may pass the cuts. We estimate the probability at three standard deviations level to extract an inclusive diboson signal in the 3-jets sample alone (P3σ). After our procedure for building the Z mass is applied, P3σ is about 4 times greater than when building the Z mass "by default" with the two E_T leading jets. Even if the systematic errors were not yet accounted for in this estimate, this appears as a significant progress.

However, in order to discriminate against the WW contribution we apply our technique considering WZ/ZZ as the signal. By investigating the 3-jets *notag* and *tag* regions the sensitivity increases although only a modest improvement is reached. Still, our technique allows including the three jets sample in the WZ/ZZ analyses in order to increase the acceptance and the sensitivity in the search for the hadronically decaying Z-boson.

Appendix A

Electroweak Unification

The electroweak theory unified the electromagnetic and the weak interactions as the so called electroweak force [14]. Both interactions are combined into one single gauge group, $SU(2)_L \otimes U(1)_Y$, in which they would appear as two manifestations of the same fundamental force.
Weak isospin (T) and hypercharge (Y) are the respective generators of symmetry transformations, related by the Gell-Mann-Nishijima (1.1). The electroweak Lagrangian is given by

$$\mathscr{L} = -\frac{1}{4} W_{\mu\nu} W^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{\Psi} i \gamma^{\mu} D_{\mu} \Psi \tag{A.1}$$

where the covariant derivative is defined as

$$D_{\mu} = \partial_{\mu} + i g W_{\mu} T + i \frac{g'}{2} B_{\mu} Y \tag{A.2}$$

T is the weak isospin operator, g and g′ are the two electroweak coupling constants and $B_{\mu\nu}$ is defined as

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

where B_{μ} is the massless gauge field representing the singlet of $U(1)_Y$. W_{μ} are the gauge fields of $SU(2)_L$ and $W_{\mu\nu}$ the field tensor which is given by

$$W_{\mu\nu} = \partial_{\mu} W_{\nu} - \partial_{\nu} W_{\mu} - g W_{\mu} \times W_{\nu}$$

The electroweak gauge fields W_{μ}^1, W_{μ}^2, W_{μ}^3 and B_{μ} used to write the electroweak lagrangian have basically a theoretical meaning. Instead, in particle physics we can express $\mathscr{L}$ as function of four physical fields: A_{μ}, Z_{μ}, W_{μ}^+ and W_{μ}^-.
A_{μ}, is the (neutral) electromagnetic field, Z_{μ} is the field corresponding to weak neutral current while $W_{\mu}^{\pm}$ correspond to the weak charged currents. By requiring the electromagnetic and weak forces to be unified and to describe the gauge bosons observed experimentally it

can be inferred that there be two neutral and two charged bosons.
Thus, the electromagnetic field A_μ and neutral current Z_μ must be some linear combination of the unified electroweak fields. This can be written in terms of the electroweak mixing angle θ_W as

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} = \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}$$

From which it can be shown that the parameters g and g′, satisfy the relation $g' = g\tan\theta_W$, and are also related to the charge of the electron e by the relation $e = g\sin\theta_W = g'\cos\theta_W$. The real fields are then

$$W_\mu^\pm = \sqrt{\frac{1}{2}}(W_\mu^1 \mp iW_\mu^2) \tag{A.3}$$

$$A_\mu = \frac{gB_\mu + g'W_\mu^3}{\sqrt{g^2 + g'^2}} \tag{A.4}$$

$$Z_\mu = \frac{-g'B_\mu + gW_\mu^3}{\sqrt{g^2 + g'^2}} \tag{A.5}$$

$$\tag{A.6}$$

When we introduce the physical fields in the lagrangian, from the first term we get up to quartic interaction vertices between charged bosons or charged and neutral bosons, while the second term produces vertices with no more than two neutral bosons. Triple gauge couplings (TGC or quartic interaction vertices) of only neutral bosons such as ZZZ, ZZγ, Z$\gamma\gamma$, are then absent in the SM.

Appendix B

Definition of Isolation

The isolation variable for **electrons** is defined as

$$Isol = \frac{E_T^{CONE} - E_T^{CLUS}}{E_T^{CLUS}}$$

where E_T^{CONE} is the sum of the transverse energy of the towers inside a cone of radius $\Delta R = 0.4$ around the cluster axis, and E_T^{CLUS} is the transverse energy of the cluster associated to the candidate electron. A small isolation indicates that there is little extra activity in the calorimeter near the lepton. If this quantity is less than 0.1, the lepton is said to be isolated or *tight*. Otherwise, it is non-isolated or *loose*.
The isolation variable for **muons** is defined as

$$Isol = \frac{1}{p_T}\left(\sum_i E_T^i - E_T^\mu\right)$$

where E_T^i is the transverse energy in the i-th tower, E_T^μ is the M.I.P. energy deposited in the tower crossed by the track and p_T is the track transverse momentum. The sum runs over all the towers inside a cone of radius $\Delta R = 0.4$ around the track direction.

Appendix C

Jet Algorithm Requirement

An acceptable jet algorithm must satisfy at best the following requirements [72]:

Infrared safety: the presence of soft radiation between two jets may cause a merging of the two jets. The solutions found by the algorithm should be insensitive to soft radiation in the event.

Collinear safety: the jet reconstruction should be independent of any collinear radiation in the event, i.e. different energy distribution of particles inside calorimetric towers.

Invariance under boost: the algorithm should find the same solutions independent of boosts in the longitudinal direction.

Boundary stability: kinematic variables should be independent of the details of the final state.

Order independence: the same reconstructed quantities should be defined at parton, particle and detector levels.

Straightforward implementation: the algorithm should be easy to implement in theoretical perturbative calculations.

Appendix D

SecVtx

D.1 Track Selection

SecVtx tries to merge large impact parameter tracks into a common vertex. The study is performed for all jets with $|\eta| < 2.4$ in a event, and only tracks within the jet cone are examined.
Tracks that are displaced from the interaction point can be:

- mis-reconstructed tracks (mostly due to multiple scattering in the material)
- secondary particles produced by nuclear interactions in the detector material
- decay products of long lived strange particles (mostly K_S and Λ)
- decay products of long lived charm and beauty hadrons

As we mentioned before, we are interested in the last category. The following track selection cuts are used to enhance the fraction of tracks belonging to the last category over the rest. Additional cuts on the reduced χ^2 returned by the track algorithm and on the number of good silicon hits are applied [77]. Moreover track pairs, whose invariant mass is within appropriate K_S or Λ mass windows, are removed.

D.2 Secondary Vertex Identification

Given the selected tracks, *SecVtx* uses a two pass approach to find vertices:

Pass 1 At least three tracks with $P_T^{min} = 0.5$ GeV/c and $S_{d_0}^{min} = 2$ are used to fit for the secondary vertex. At least one of the tracks used in the fit must have a $P_T > 1$ GeV/c.

Pass 2 if a vertex is not found by the first pass, only two displaced tracks are required to reconstruct the vertex. Compared to *Pass 1*, these tracks have to pass tighter quality requirements: the first (second) track must have $P_T^{min} = 1$ GeV/c ($= 1.5$ GeV/c) and $S_{d_0}^{min} = 3 (= 3)$.

Parameter	Requirement
$\Delta R = \sqrt{(\eta_{trk} - \eta_{jet})^2 + (\phi_{trk} - \phi_{jet})^2}$	< 0.4
P_T	$> P_T^{min}$
d_0	< 0.3 cm
$\lvert z - z_{PrimVtx} \rvert$	< 5 cm
$S_{d_0} = d_0/\sigma_{d_0}$	$> S_{d_0}^{min}$

Table D.1: Requirements on tracks.

Once the vertex is found in a jet, the two dimensional decay length L_{xy} is calculated as the projection in the transverse plane of the vector pointing from the primary to the secondary vertex. The sign of L_{xy} is defined by the angle α between the jet axis and the $SecVtx$ vector, see Fig. D.1.

A number of errors or fluctuations can generate fake vertices. The rate of negative L_{xy} is a sensor of these effects. Secondary vertices corresponding to the decay of heavy-flavor hadrons are expected to have large (positive) L_{xy}, while secondary vertices from mis-measured tracks are expected to be less displaced and to feature with the same probability positive and negative d_0.

Therefore *SecVtx* requires $L_{xy}/\sigma_{L_{xy}} > 3$, where $\sigma_{L_{xy}}$ is the estimated uncertainty on L_{xy}. Other cuts to reject secondary vertices due to material interaction are described in [78].

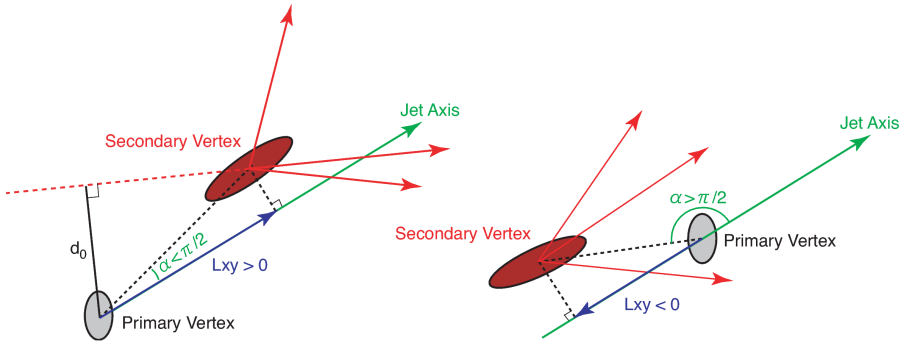

Figure D.1: Left, true reconstructed secondary vertex. Right, negative *SecVtx* tag (falsely reconstructed secondary vertex).

Appendix E

Investigating ISR/FSR

In order to reproduce the final state, we use the ALPGEN generator which is interfaced to the generator PYTHIA to include jet fragmentation. We investigate at generator level the origin of the not-matched jet (NMJ) in order to understand the Right Jet Combination (RJC).
Jets are matched in direction to quarks from Z decay, requiring $\Delta R < 0.4$ between partons from Z and jets.
When a number of matches different from 2 is found the event is not considered.
We focus on the sample in which two jets are matched to the $q\bar{q}$ pair from Z, which is less than the approximately 70% of the whole sample. Then we investigate the origin of the NMJ. We identify all the gluons which have $\Delta R < 0.4$ from the NMJ and we separate two cases:

1. gluon is radiated from the interacting partons, i.e. ISR
2. gluon is radiated from the $q\bar{q}$ pair from Z, i.e. FSR

Then the sum of the energy contribute to the jet energy coming form each case is computed. We state that a jet comes from ISR if the energy contribute coming from FSR is 0 and viceversa.
In the about 6% of events we have both contributions to jet energy. These events are rejected.
In the about 10% of cases PYTHIA can't state the origin of the gluon matched to the jet, also this event are rejected.

Appendix F

Resolution Parameters

F.1 Notag Sample

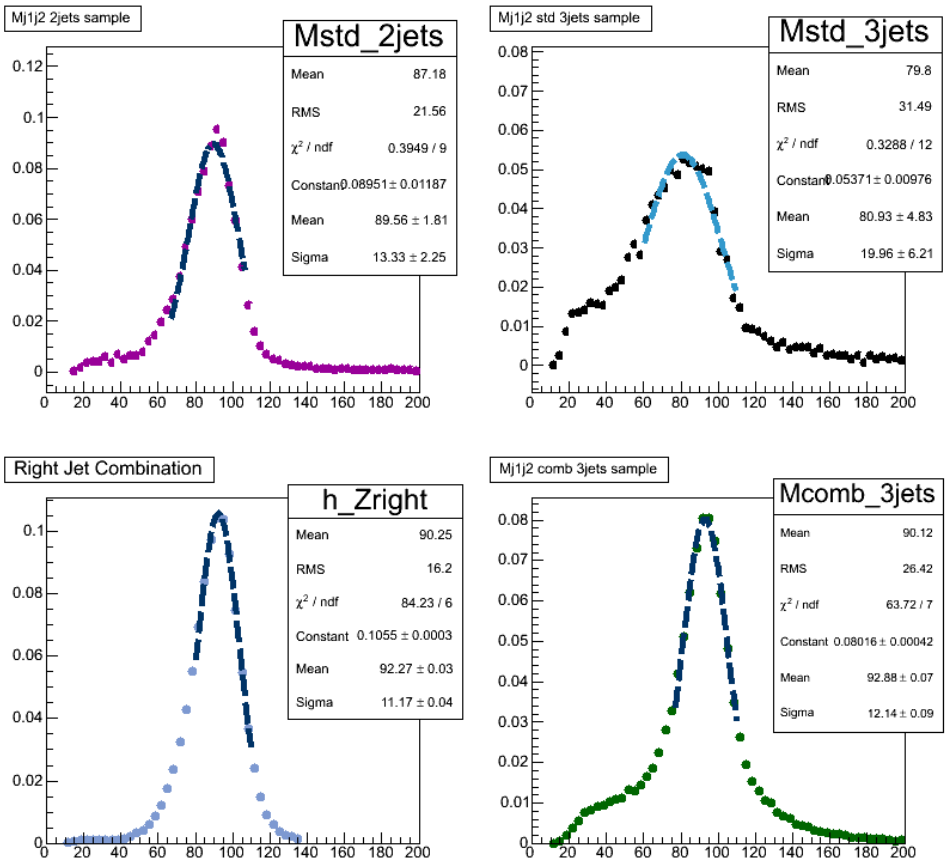

Figure F.1: Invariant mass distributions in different subsamples in the *notag* region.

F.2 Tag Sample

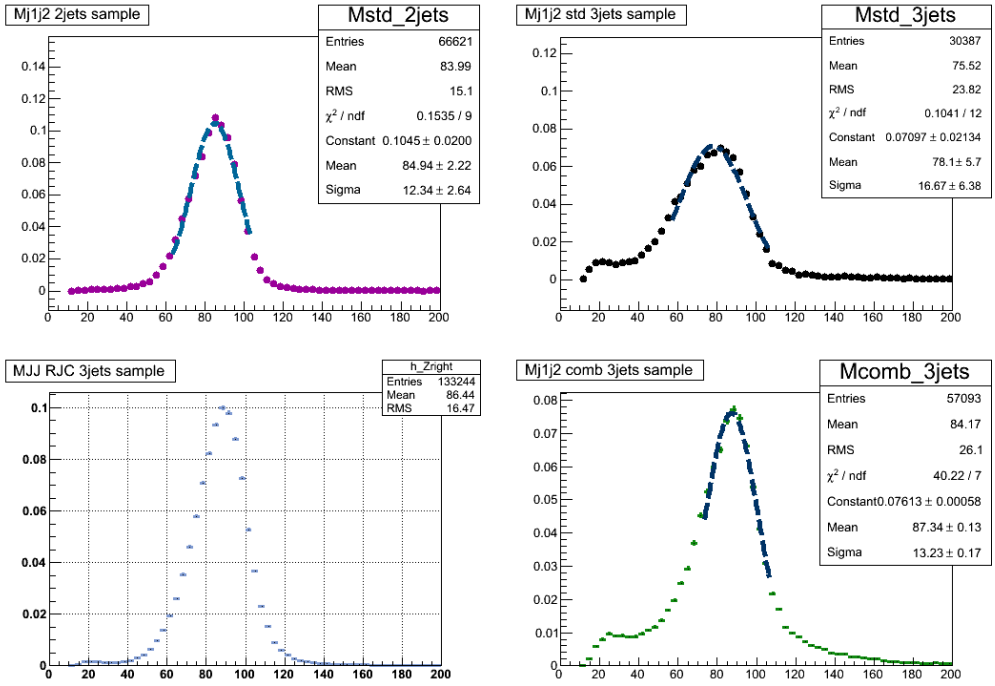

Figure F.2: Invariant mass distributions in different subsamples in the *tag* region.

Appendix G

NNs input in notag sample

G.1 NN_{12}

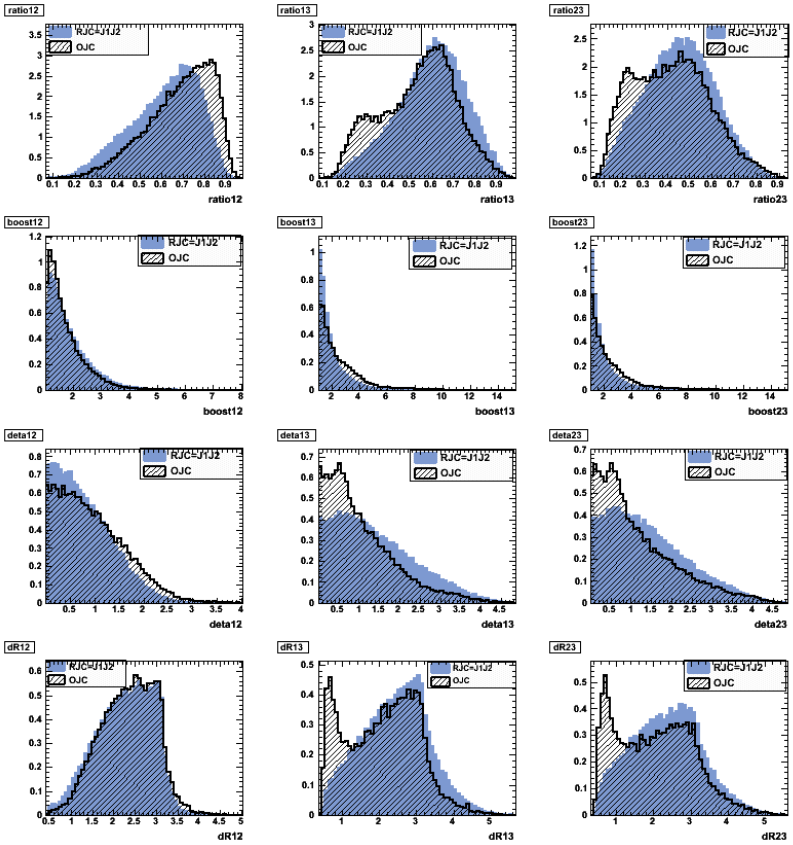

Figure G.1: NN_{12} input variables after weighting.

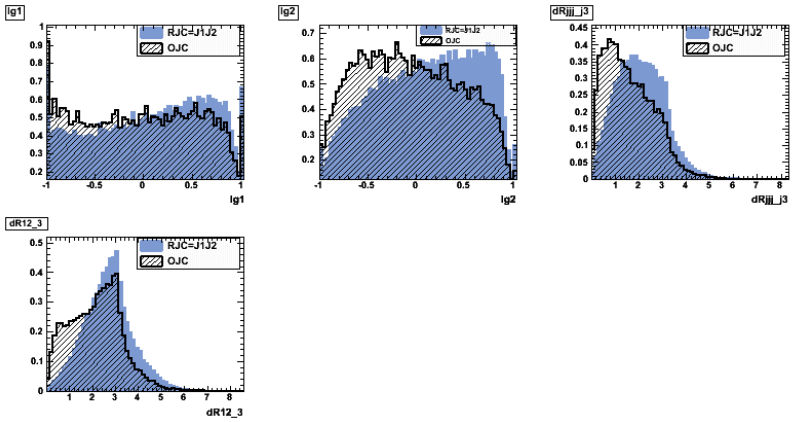

Figure G.2: NN_{12} input variables after weighting.

G.2 NN_{13}

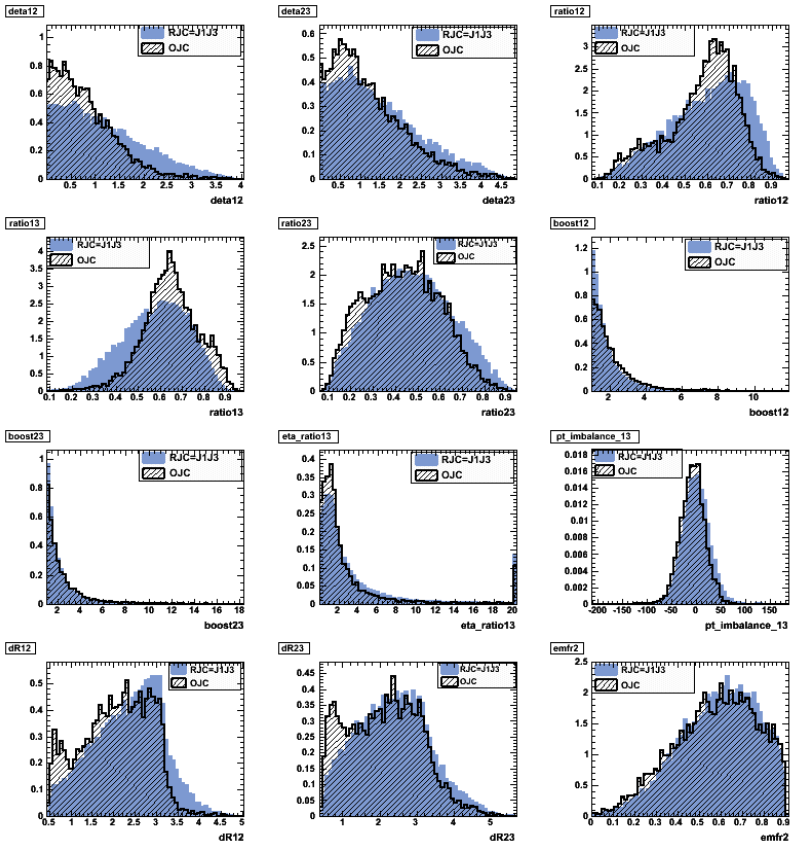

Figure G.3: NN_{13} input variables after weighting.

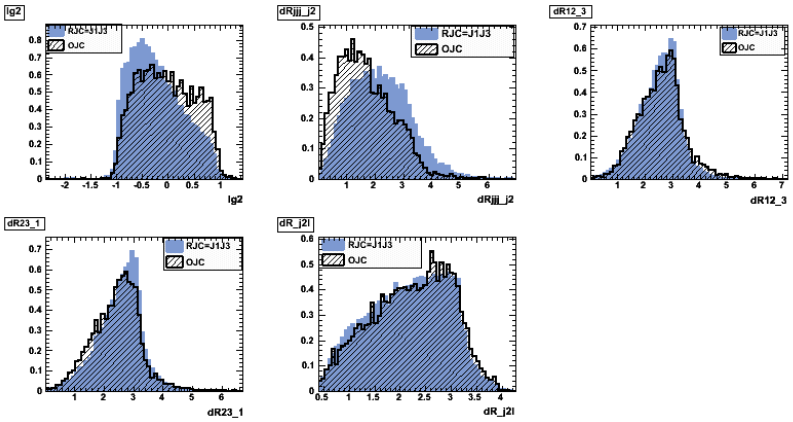

Figure G.4: NN_{13} input variables after weighting.

G.3 NN_{23}

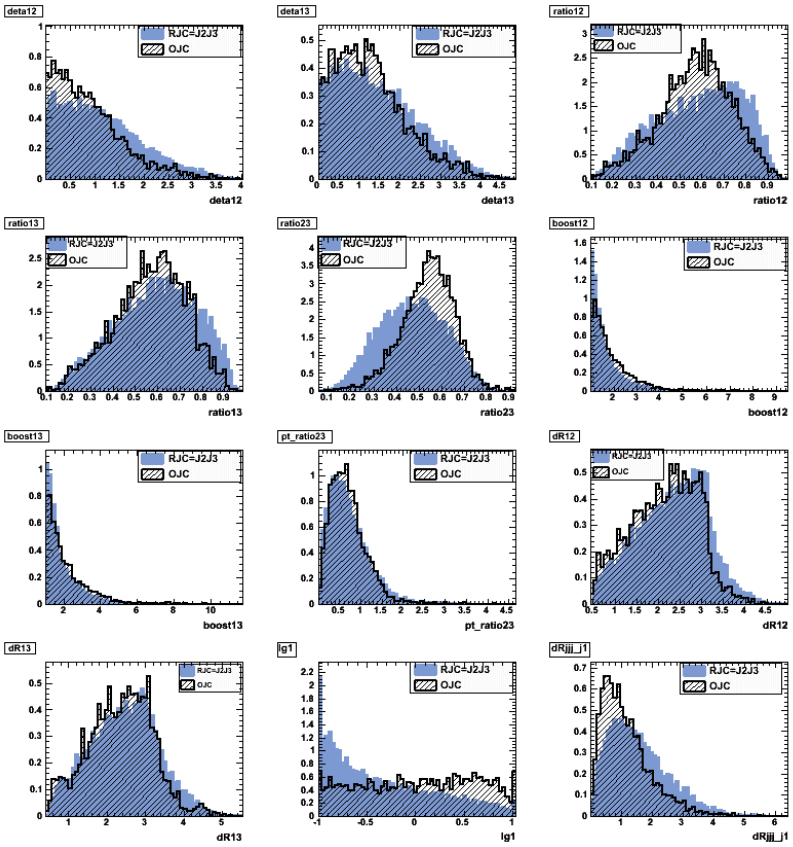

Figure G.5: NN_{23} input variables after weighting.

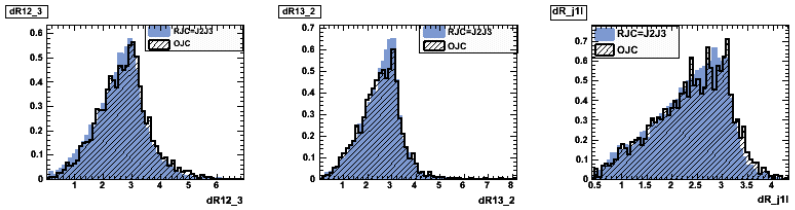

Figure G.6: NN_{23} input variables after weighting.

G.4 NN_{123}

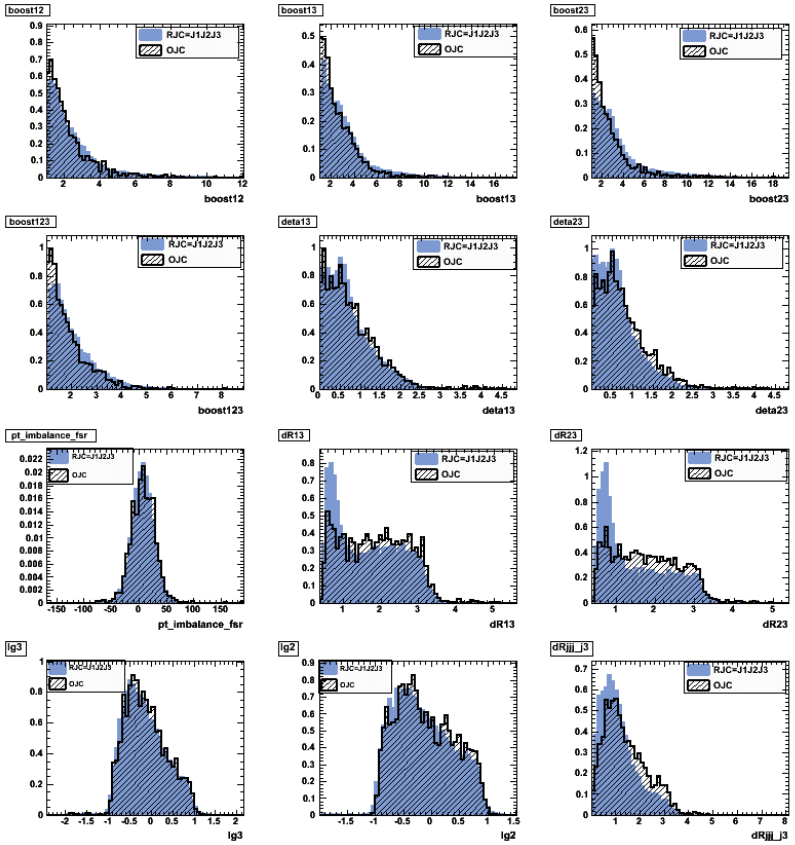

Figure G.7: NN_{123} input variables after weighting.

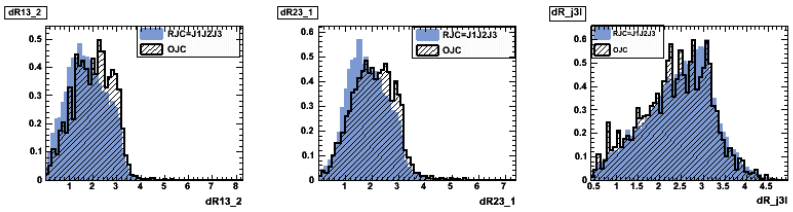

Figure G.8: NN_{123} input variables after weighting.

Appendix H

NNs input in tag sample

H.1 Exploit MJ1J2: NN_{12}

H.1.1 Input

Below is the list of the variables used:

1. $m_{jj'}/m_{j_1j_2j_3}$ [1]
2. $\gamma_{jj'} = (E_j + E_{j'})/m_{jj'}$
3. $\mathrm{d}\eta_{j_1j_3}$, $\mathrm{d}\eta_{j_2j_3}$
4. $\eta(j_1 + j_2)/\eta(j_3)$
5. $dR_{j_1j_3}$, $dR_{j_2j_3}$
6. $dR_{j_1j_2,j_3}$, $dR_{j_2j_3,j_1}$
7. $dR_{j_1j_2j_3,j_3}$
8. $dR_{j_3,\ell}$
9. $b$ness for J1, J2
10. Quark Gluon Discriminator for J3

H.1.2 Output

The variables described above are weighted accordingly and are used for training a Neural Network, employing MLP method.
The NN_{12} response is shown in Fig. H.2.

[1] jj' refers to the three possible combinations: J1J2, J1J3 and J2J3.

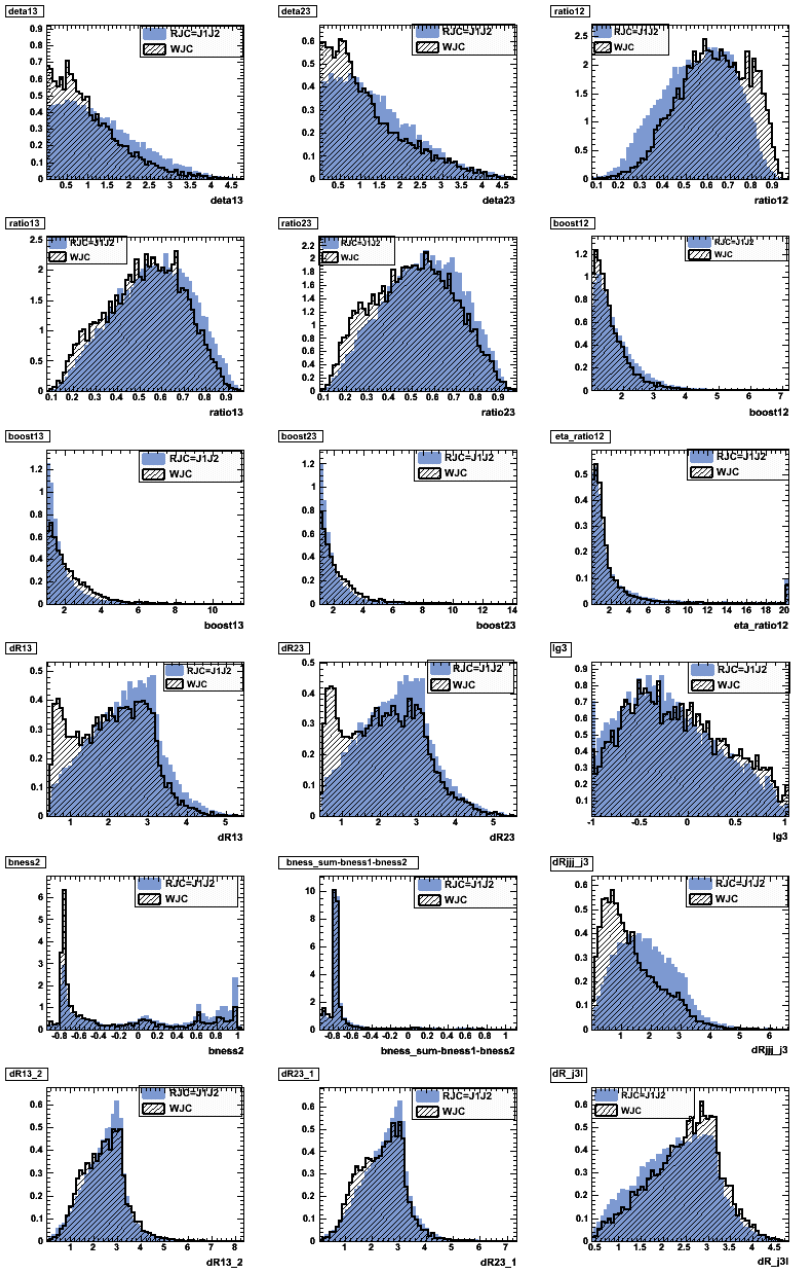

Figure H.1: NN_{12} input variables after weighting.

H.2 Exploit MJ1J3: NN_{13}

H.2.1 Input

Below is the list of the variables used:

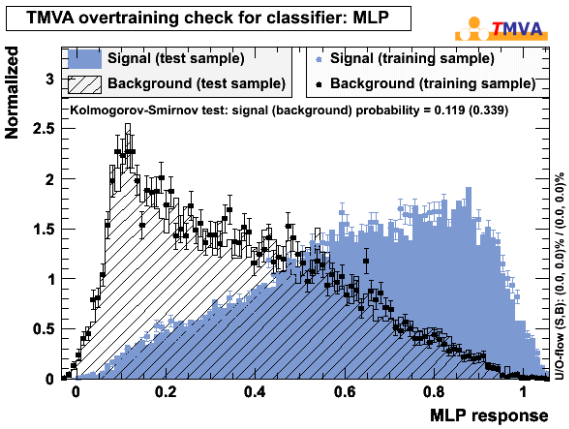

Figure H.2: NN_{12} MLP response

1. $m_{jj'}/m_{j_1j_2j_3}$
2. $\gamma_{jj'} = (E_j + E_{j'})/m_{jj'}$
3. $\mathrm{d}\eta_{j_1j_2}$, $\mathrm{d}\eta_{j_2j_3}$
4. $dR_{j_1j_2}$, $dR_{j_2j_3}$
5. "pt-imbalance" : $p_{Tj_1} + p_{Tj_3} - p_{T\ell}-$MET
6. EMfr for J2 which is the ratio between EM and total energy.
7. $dR_{j_1j_2,j_3}$, $dR_{j_2j_3,j_1}$
8. $dR_{j_1j_2j_3,j_2}$, $dR_{j_1j_2j_3,j_3}$
9. $dR_{j_2,\ell}$, $dR_{j_3,\ell}$
10. Quark Gluon Discriminator for J2
11. $b$ness for J2

H.2.2 Output

Even in this case we decorrelate the NN_{13} output from the numerical value of MJ1J3 and make it sensitive only to the kinematical distributions of the involved variables. The NN_{13} response is shown in Fig. H.4.

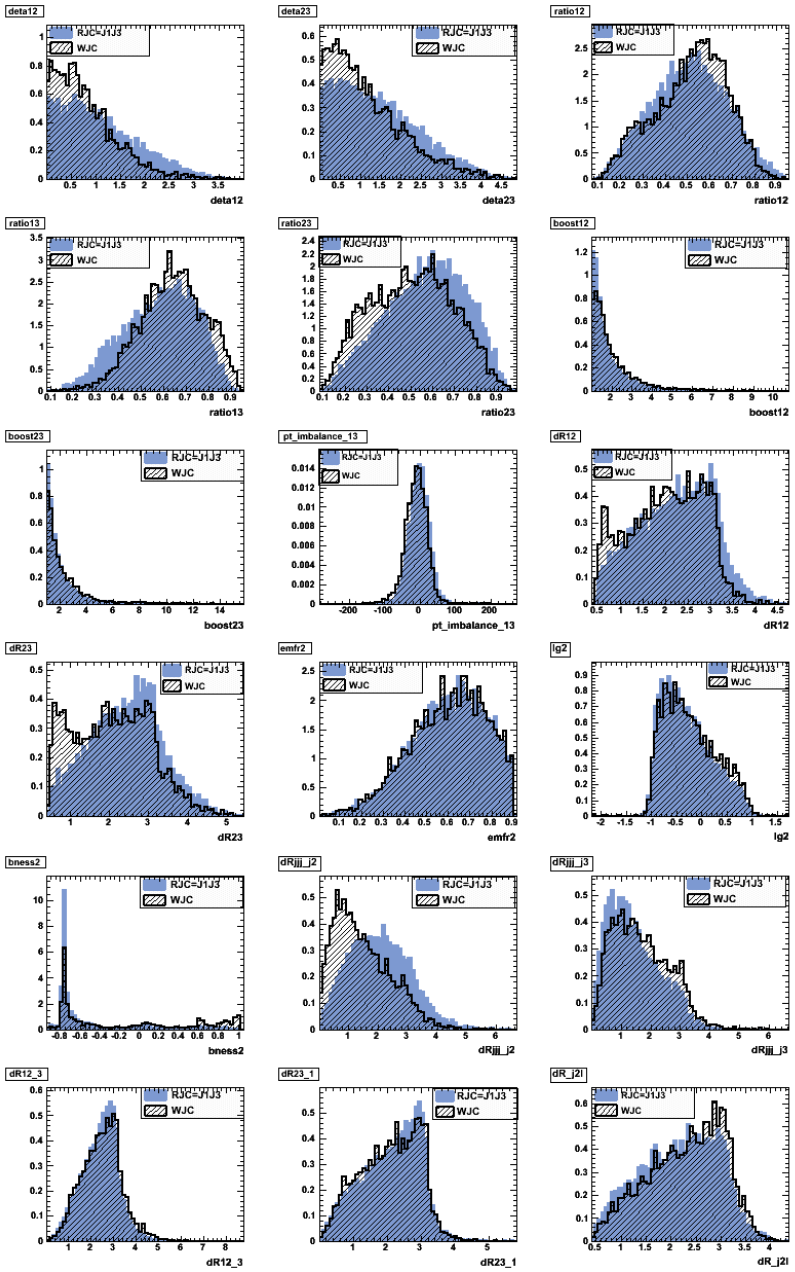

Figure H.3: NN_{13} input variables after weighting.

H.3 Exploit MJ2J3: NN_{23}

H.3.1 Input

Input variables in this case are:

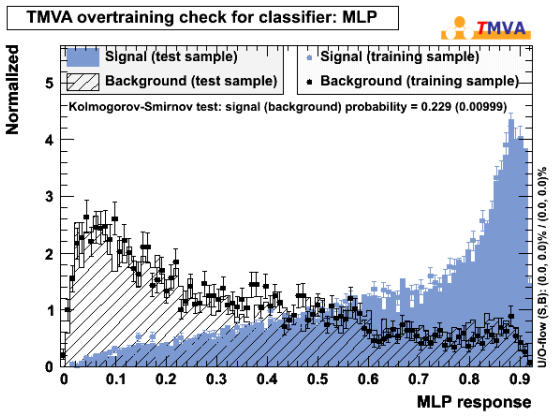

Figure H.4: NN_{13} MLP response

1. $\gamma_{jj'} = (E_j + E_{j'})/m_{jj'}$
2. $\mathrm{d}\eta_{j_1 j_2}$, $\mathrm{d}\eta_{j_1 j_3}$
3. $p_T(j_2 + j_3)/p_T(j_1)$
4. "pt-imbalance" : $p_{T\,j_2} + p_{T\,j_3} - p_{T\,\ell}$−MET
5. $dR_{j_1 j_2}$, $dR_{j_1 j_3}$
6. $dR_{j_1 j_2, j_3}$, $dR_{j_1 j_3, j_2}$
7. $dR_{j_1 j_2 j_3, j_1}$
8. $dR_{j_1, \ell}$
9. $b$ness for J1, J3
10. Quark Gluon Discriminator for J1, J3

H.3.2 Output

Even in this case we decorrelate the NN_{23} output from the numerical value of MJ2J3 and make it sensitive only to the kinematical distributions of the involved variables.The NN_{23} response is shown in Fig. H.6.

H.4 Exploit MJ1J2J3: NN_{123}

H.4.1 Input

The input variables are:

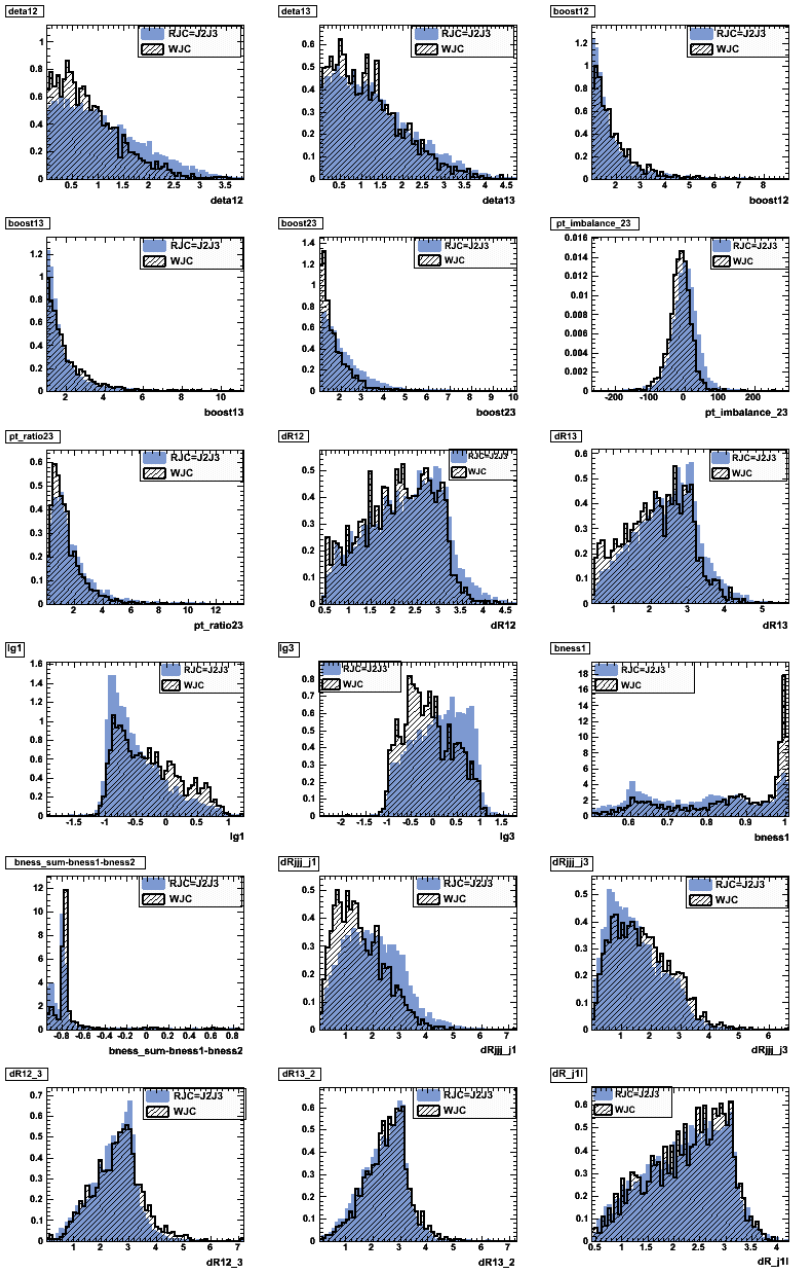

Figure H.5: NN_{23} input variables after weighting.

1. $\gamma_{jj'} = (E_j + E_{j'})/m_{jj'}$ [2]
2. $\gamma = (E_{j_1} + E_{j_2} + E_{j_3})/MJ1J2J3$
3. "pt-imbalance" : $p_{T\,j_1} + p_{T\,j_2} + p_{T\,j_3} - p_{T\,\ell}$−MET

[2] jj' refers to the three possible combinations: J1J2, J1J3 and J2J3.

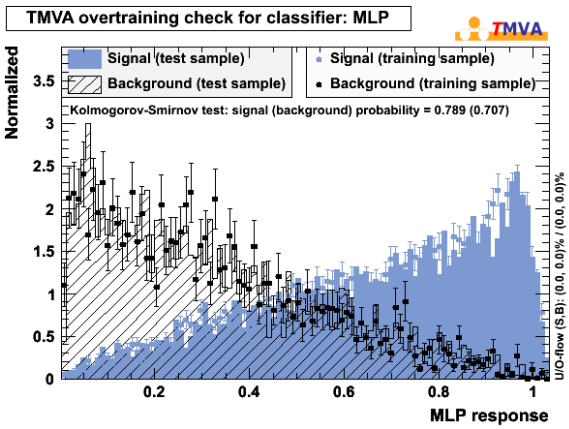

Figure H.6: NN_{23} MLP response

4. $\mathrm{d}\eta_{jj'}$
5. $dR_{j_1 j_3}$, $dR_{j_2 j_3}$
6. $dR_{j_1 j_3, j_2}$, $dR_{j_2 j_3, j_1}$
7. $dR_{j_1 j_2 j_3, j_3}$, $dR_{j_1 j_2 j_3, j_2}$
8. $dR_{j_2 \ell}$, $dR_{j_3 \ell}$
9. EMfr for J2, J3
10. $b$ness J2, J3
11. Quark Gluon Discriminator for J2, J3

H.4.2 Output

In Fig. H.9 is shown the NN_{123} output.

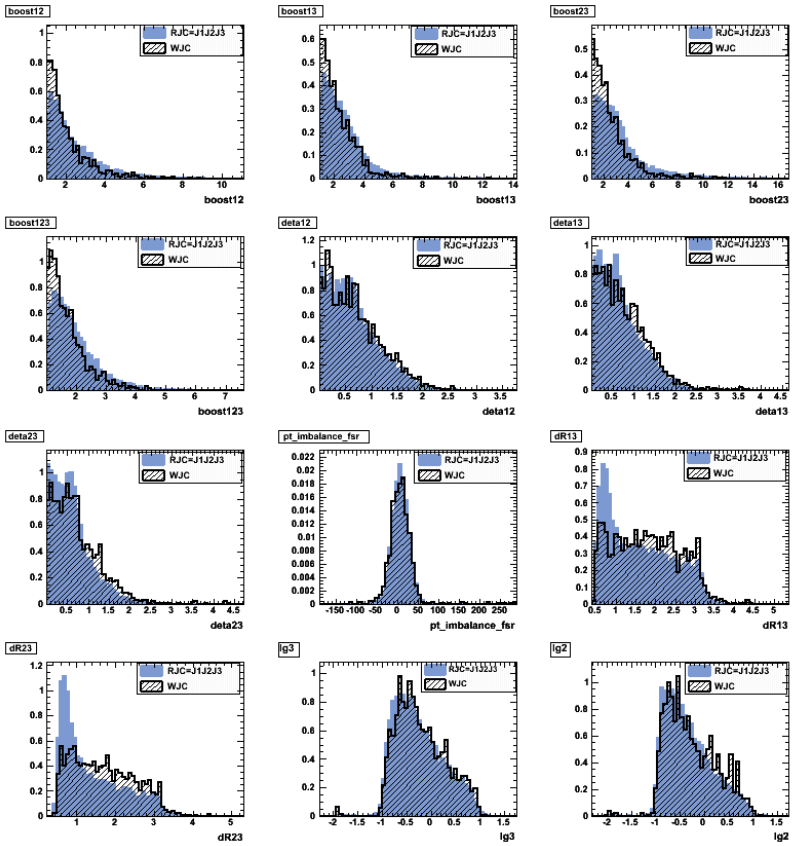

Figure H.7: NN_{123} input variables after weighting.

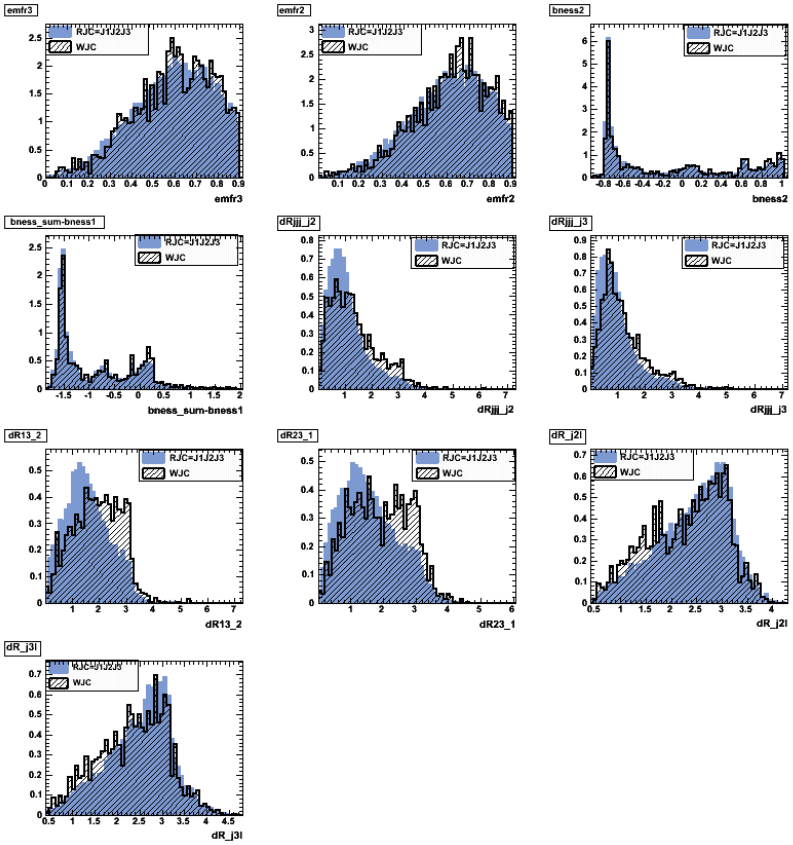

Figure H.8: NN_{123} input variables after weighting.

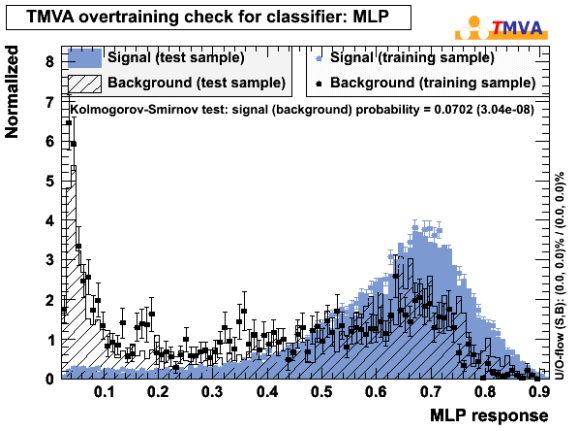

Figure H.9: NN_{123} MLP response

Appendix I

NNs Correlations

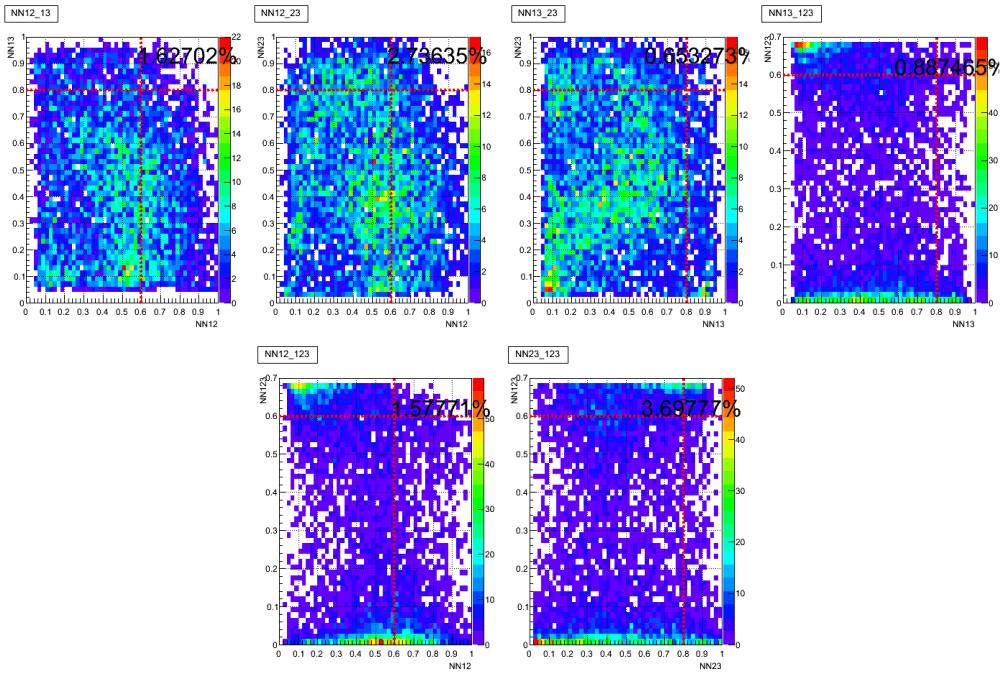

Figure I.1: Correlations between NNs output

Appendix J

NNs Modeling

J.1 Notag sample

In this section we compare the output of NNs and MJJ distributions (see Figs. J.1-J.8) in data and simulated events and QCD events in order to validate the MC, the QCD templates, and the expected rates in the *notag* sample:

- `NN12`, `NN13`, `NN23`, `NN123`: the outputs of NNs.
- `MJ1J2`, `MJ1J3`, `MJ2J3`, `MJ1J2J3`.

The agreement is satisfactory in all distributions.

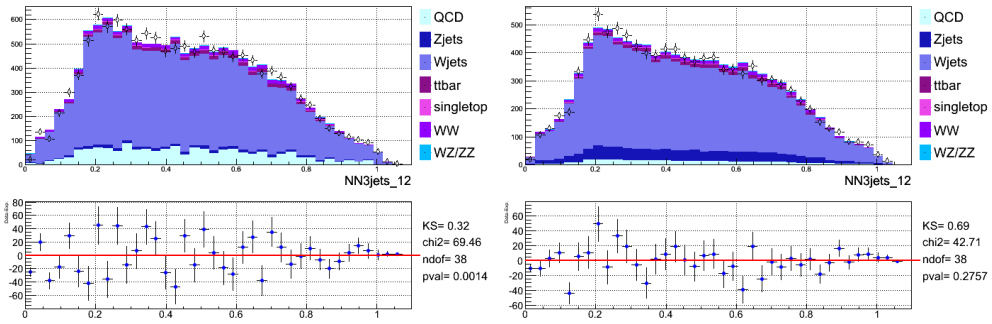

Figure J.1: Left, `NN12` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

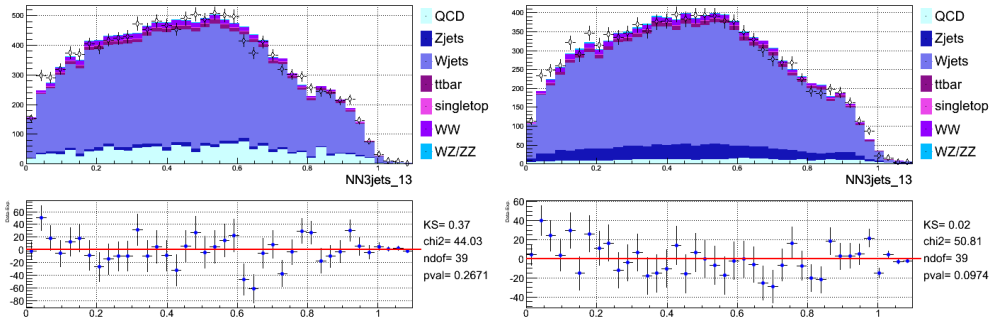

Figure J.2: Left, `NN13` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

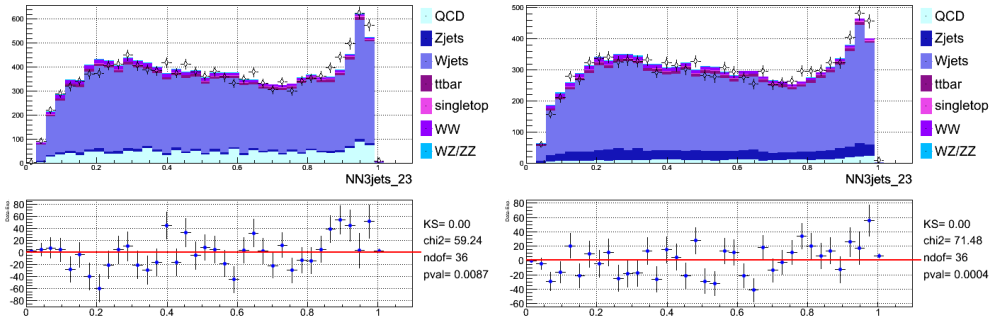

Figure J.3: Left, `NN23` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

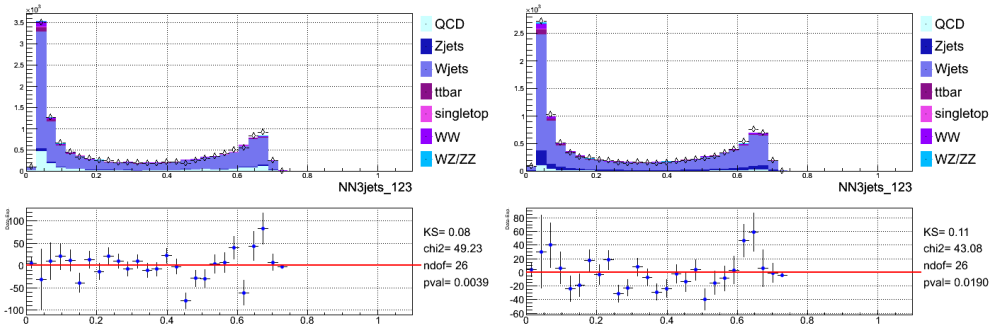

Figure J.4: Left, `NN123` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

J.2 Tag sample

In this section we plot the same variables listed for the *pretag* sample in order to validate the MC, the QCD templates and the expected rate in the *tag* sample, (see Figs. J.9-J.16).

- `NN12`, `NN13`, `NN23`, `NN123`: the outputs of NNs.
- `MJ1J2`, `MJ1J3`, `MJ2J3`, `MJ1J2J3`.

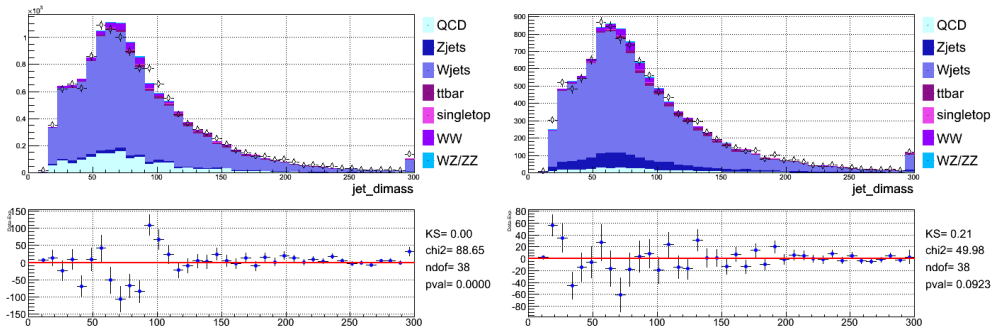

Figure J.5: Left, `MJ1J2` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation. The horizontal scale is in GeV/c^2.

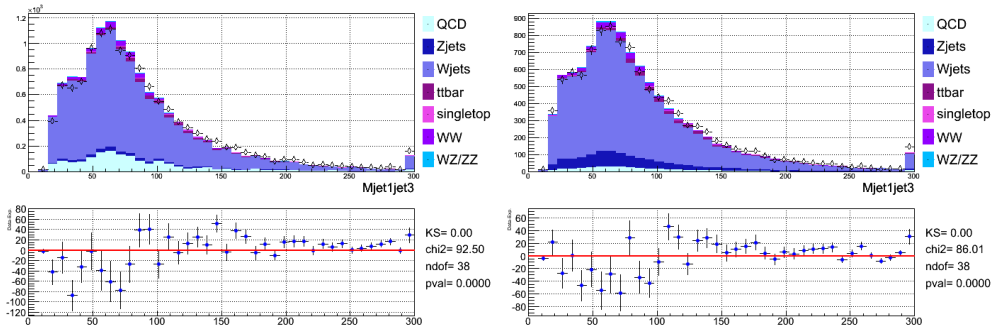

Figure J.6: Left, `MJ1J3` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation. The horizontal scale is in GeV/c^2.

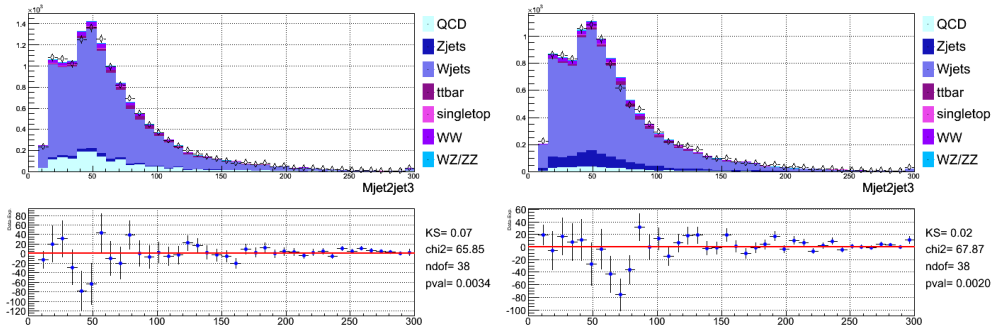

Figure J.7: Left, `MJ2J3` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation. The horizontal scale is in GeV/c^2.

The agreement is good in all the distributions.

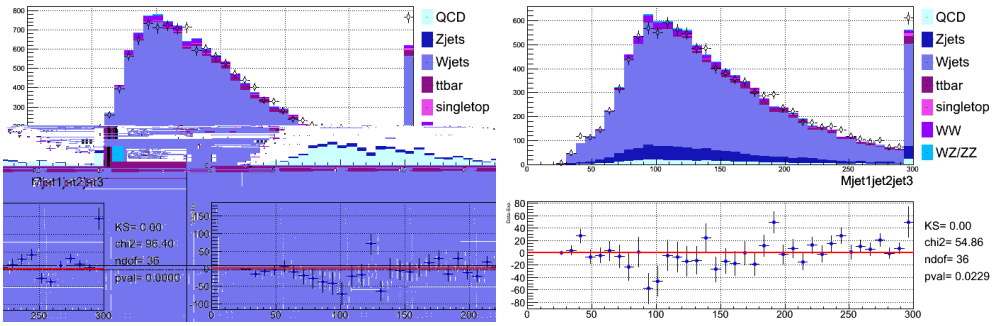

Figure J.8: Left, `MJ1J2J3` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation. The horizontal scale is in GeV/c^2.

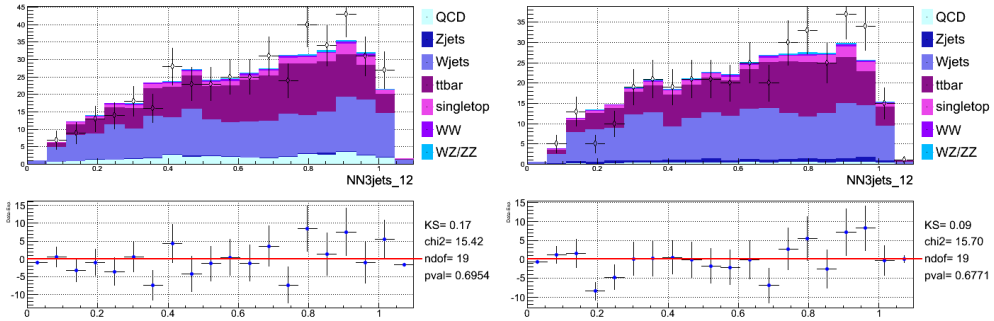

Figure J.9: Left, `NN12` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

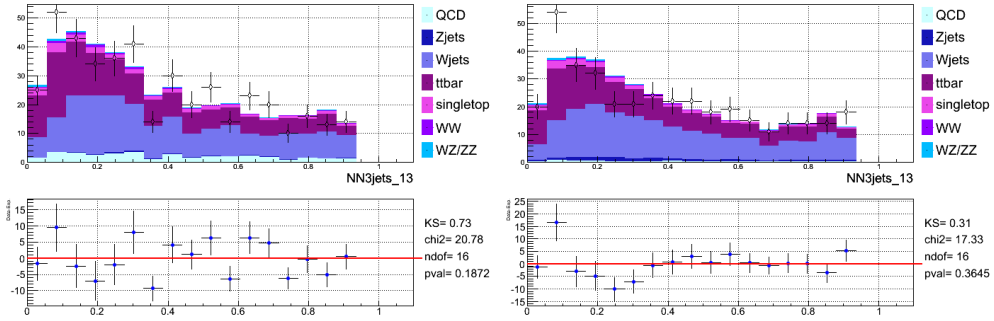

Figure J.10: Left, `NN13` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

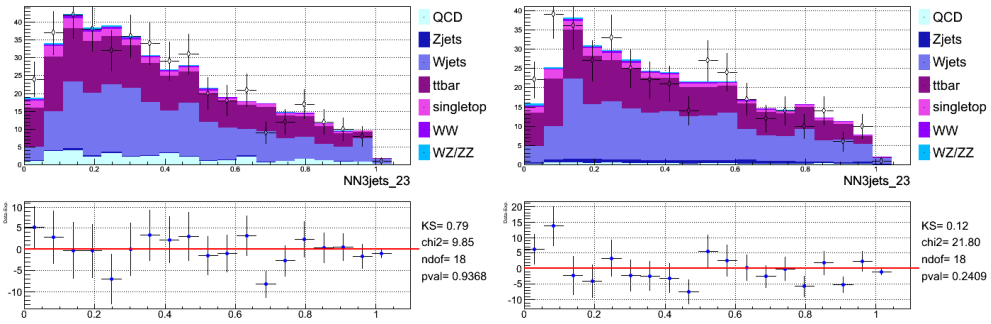

Figure J.11: Left, `NN23` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

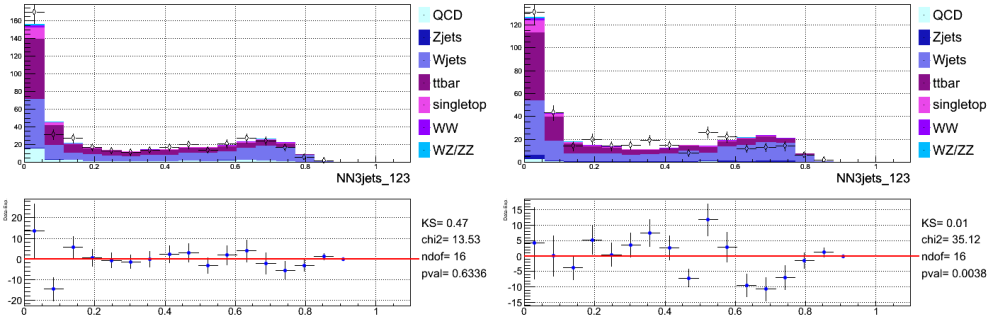

Figure J.12: Left, `NN123` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation.

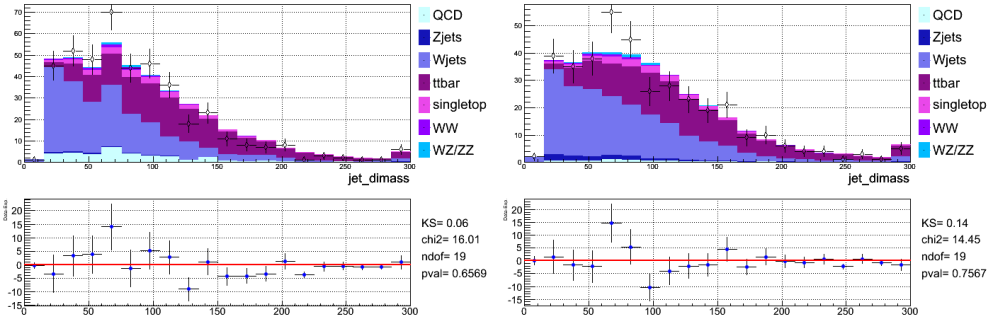

Figure J.13: Left, `MJ1J2` distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation. The horizontal scale is in GeV/c^2.

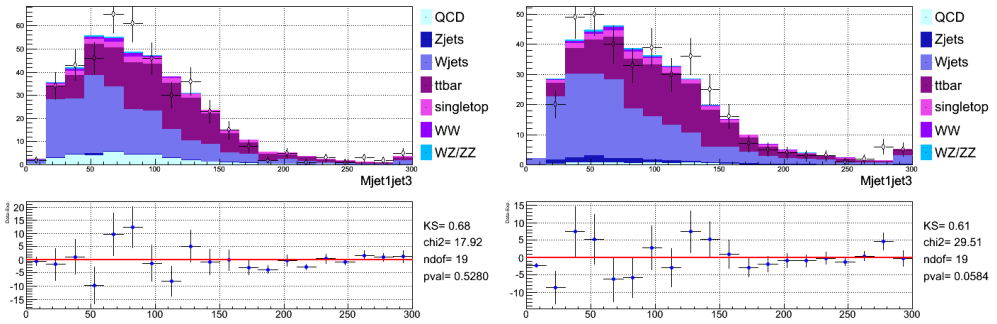

Figure J.14: Left, MJ1J3 distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation. The horizontal scale is in GeV/c^2.

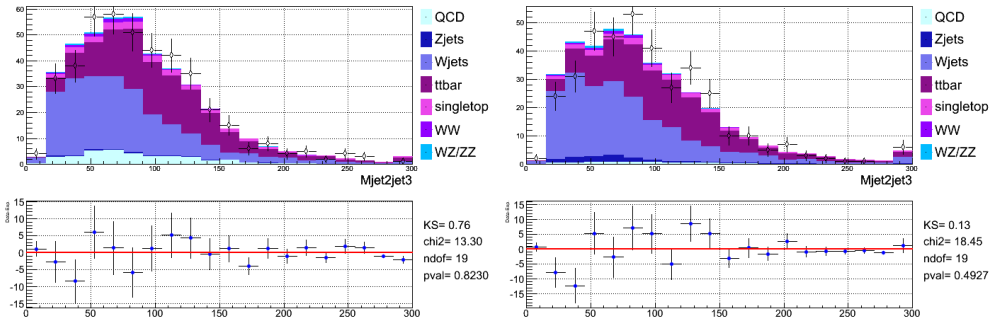

Figure J.15: Left, MJ2J3 distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation. The horizontal scale is in GeV/c^2.

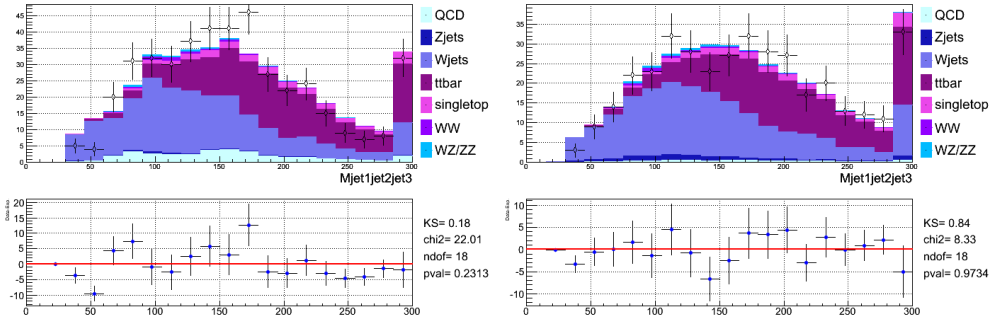

Figure J.16: Left, MJ1J2J3 distribution for electrons. Right, for muons. The lower distributions show the difference data-simulation. The horizontal scale is in GeV/c^2.

Bibliography

[1] J. M. Campbell and R. K. Ellis *Update on Vector Boson Pair Production at Hadron Colliders* Phys. Rev. D **65** (2002) 113007.

[2] K. Nakamura et al. (Particle Data Group) J. Phys. G **37** (2010) 075021.

[3] G. Bellettini, G. Latino, S. Leo, M. Trovato, C. Vernieri, V. Rusu, G. Velev, *Jet studies on the WZ* $\rightarrow \ell\nu(\geq 3)j$ *sample*, CDF Note **9944** (2010)

[4] M. Trovato, C. Vernieri *Search for diboson production in lepton+MET+bb channel* CDF Note **10551** (2011)

[5] G. Bellettini, G. Latino,V. Rusu, M. Trovato, G. Velev, C. Vernieri *A Novel Technique to reconstruct the Z mass in WZ/ZZ events with lepton(s), MET and Three Jets* CDF Note **10642** (2011)

[6] G. Bellettini, G. Latino,V. Rusu, M. Trovato, G. Velev, C. Vernieri *Search for WZ/ZZ production in events with lepton(s) plus jets plus missing transverse energy* CDF Note **10630** (2011)

[7] Y. Fukuda et al. *Evidence for oscillation of atmospheric neutrinos* Phys. Rev. Lett. **81** (1998) 1562 1567.

[8] M. Maggiore *A Modern Introduction to Quantum Field Theory* Oxford University Press 2005.

[9] M. Peskin and D. Schroeder *An Introduction to Quantum Field Theory* Westview Press 1995.

[10] M. Goldhaber, L. Grodzins and A. W. Sunyar *Helicity of Neutrinos*, Phys. Rev. Lett. **109** (1958) 1015.

[11] I. J. R. Aitchison and A. J. G. Hey. *Gauge Theories in Particle Physics* 2 Volume Set. Taylor & Francis, 3 edition, January 2004.

[12] R. Ellis, W.J. Stirling, and B.R. Webber *QCD and Collider Physics* Cambridge University Press 2003.

[13] P. W. Higgs *Broken Symmetries, Massless Particles And Gauge Fields* Phys. Rev. Lett. **12** (1964) 132; *Broken Symmetries And The Masses Of Gauge Bosons* Phys. Rev. Lett. **13** (1964) 508.

[14] S. L. Glashow *Partial Symmetries of Weak Interactions* Nucl. Phys. A **22** (1961) 579.

[15] S. Weinberg *A Model of Leptons* Phys. Rev. Lett. **19** (1967) 1264

[16] A. Salam, *Elementary Particle Theory* ed. by N. Svartholm (Almqvist and Wiksells, Stockholm, 1968).

[17] Donald H. Perkins *Introduction to High Energy Physics* Cambridge University Press 2000.

[18] LEP Working Group for Higgs Boson Searches (ALEPH, DELPHI, L3 and OPAL) *Search for the Standard Model Higgs Boson at LEP* Phys. Rev. B **565** (2003) 61

[19] Tom Junk et al. *Combination of CDF's Searches for the Standard Model Higgs boson with up to 8.2 fb^{-1} of Data.* CDF Note **10609** (2011)

[20] CDF and (DØ Collaboration) *Combination of Tevatron Searches for the Standard Model Higgs Boson in the WW Decay Mode* Phys. Rev. Lett. **104** (2010) 061802.

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

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

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

[24] The LEP Collaborations *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model* [arXiv:hep-ex/0511027].

[25] T. Aaltonen et al. *Observation of WZ production* Phys. Rev. Lett. **98** (2007)161801

[26] V. Abazov (DØ Collaboration), *Measurement of the WW Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV* Phys. Rev. Lett. **94** (2005) 151801.

[27] CDF Collaboration, *Measurement of the WW Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using 3.6 fb^{-1} of CDF Run II Data* Phys. Rev. Lett. **104** (2010) 201801.

[28] V. Abazov et al. (DØ Collaboration) *Measurement of the $p\bar{p} \rightarrow WZ+X$ Cross Section at $\sqrt{s} = 1.96$ TeV and Limits on WWZ Trilinear Gauge Couplings* Phys. Rev. D **76** (2007) 111104.

[29] T. Aaltonen et al. (CDF Collaboration) *Measurement of the WZ and ZZ Production Cross Sections Using WZ→ $3\ell + \not{E}_T$ and ZZ → 4ℓ Channels* CDF Note **10238** (2010).

[30] T. Aaltonen et al. (CDF Collaboration) *ZZ cross section measurement in $\ell\ell\nu\nu$ decay channel at CDF* CDF NOTE **10358** (2010).

[31] V. M. Abazov et al. (DØ Collaboration)*$ZZ \to \ell\ell\nu\nu$ production in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96 TeV* Phys. Rev. D **78** (2008) 072002.

[32] V. M. Abazov et al. (DØ Collaboration)*Observation of ZZ production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV* Phys. Rev. Lett. **101** (2008) 171801.

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

[34] V. M. Abazov et al. (DØ Collaboration) *Evidence of WW +WZ Production with Lepton+Jets Final States in Proton-Antiproton Collisions at $\sqrt{s} = 1.96$ TeV* Phys. Rev. Lett. **102** (2009) 161801.

[35] T. Aaltonen et al. (CDF Collaboration) *Measurement of the WW + WZ Production Cross Section Using the Lepton + Jets Final State at CDF II* Phys. Rev. Lett. **104** (2010) 101801.

[36] T. Aaltonen et al. (CDF Collaboration) *Invariant Mass Distribution of Jet Pairs Produced in Association with a W Boson in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV* Phys. Rev. Lett. **106** (2011) 171801.

[37] CDF Collaboration *The CDF II Detector-Technical Design Report* 1996 FERMILAB-Pub-96/390-E CDF.

[38] Fermilab Beam Division *Run II Handbook* http://www-bd.fnal.gov.

[39] C.S.Hill et al. *Initial Experience with the CDF Layer 00 Silicon Detector* Nucl. Inst. and Meth. **A511** (2002) 118.

[40] A. Still et al. (CDF Collaboration) *SVX-II: CDF Run II Silicon Tracking Projects* Nucl. Instrum. Meth. **A447** (2000) 1.

[41] C. S. Hill et al. (CDF Collaboration) *Operational Experience and Performance of the CDF II Silicon Detector* Nucl. Instrum. Meth. **A530** (2004) 1.

[42] I. Vila for the CDF Collaboration *Performance and First Physics Results of the SVT Trigger at CDF II* ECONF 2003 [arXiv:hep-ph/0307165].

[43] B. Ashmanskas et al. (CDF Collaboration) *The CDF Silicon Vertex Trigger* Nucl. Instrum. Meth. **A518** (2004) 532 [arXiv:physics/0306169]

[44] T. Aolder et al. (CDF Collaboration) *COT Central Outer Tracker* Nucl. Instrum. Meth. **A526** (2004) 249.

[45] See CDF Run II Detector Parameters, *COT Simulation Validation Plots*.

[46] D. Acosta et al. (CDF Collaboration) *A Time-Of-Flight Detector in CDF II* Nucl. Instrum. Meth. **A518** (2004) 605.

[47] L. Balka et al. (CDF Collaboration) *The CDF Central Electromagnetic Calorimeter* Nucl. Instrum. Meth. **A267** (1988) 272.

[48] S. Bertolucci et al. (CDF Collaboration) *The CDF Central and Endwall Hadron Calorimeter* Nucl. Instrum. Meth. **A267** (1988) 301.

[49] M. Albrow et al. (CDF Collaboration) *The CDF Plug Upgrade Electromagnetic Calorimeter: Test Beam Results* Nucl. Instrum. Meth. **A480** (2002) 524.

[50] G. Apollinari et al. (CDF Collaboration) *Shower maximum detector for the CDF plug upgrade calorimeter* Nucl. Instrum. Meth. **A412** (1998) 515.

[51] G. Ascoli et al. (CDF Collaboration) *CDF central muon detector* Nucl. Instrum. Meth. **A268** (1988) 33.

[52] K. Byrum et al. (CDF Collaboration) *The CDF forward muon system* Nucl. Instrum. Meth. **A268** (1988) 46.

[53] D. Acosta et al. (CDF Collaboration) *The CDF Luminosity Monitor* Nucl. Instrum. Meth. **A461** (2001) 540.

[54] S. Klimenko J. Konigsberg and T. Liss *Averaging of the inelastic cross sections measured by the CDF and the E811 experiments* Fermilab-FN-0741 (2003).

[55] S. Jindariani et al. *Luminosity Uncertainty for Run 2 up until August 2004* CDF Note **7446**.

[56] R.J. Tesarek et al. *A Measurement of the Radiation Environment in the CDF Tracking* Nucl. Inst. and Meth. **A514** (2003) 188.

[57] M. Moll et al. *Relation between Microscopic Defects and Macroscopic Changes in Silicon Detector Properties after Hadron Irradiation* Nucl. Inst. and Meth. **B186** (2002) 110.

[58] P. Dong et al. *An Analysis of Bias Currents in SVX II* CDF Note **8219** 2006.

[59] G. Lindström, M. Moll and E. Fretwurst *Radiation Damage in Silicon Detectors* Nucl. Instrum. Meth. **A426** (1999) 1.

[60] A. Canepa et al. *Level-3 Calorimetric Resolution available for the Level-1 and Level-2 CDF Triggers* [arXiv:0810.3738].

[61] The Trigger and Datasets Working Group *Run II Trigger Table and Datasets Plan* CDF Note **4718**.

[62] F. D. Snider (CDF Collaboration) *Tracking at CDF: Algorithms and experience from Run I and Run II* Nucl. Instrum. Meth. **A566** (2006) 133.

[63] P. Azzi at al. *Histogram Tracking in the COT* CDF Note **5562**.

[64] K. Bloom, W. M. Yao *Outside-In Silicon Tracking at CDF* CDF Note **5991**.

[65] Byeong Rok Ko et al. *SVXII Stand-alone Tracking* CDF Note **6440**.

[66] C. P. Hays et al. *Inside-out tracking at CDF* Nucl. Instrum. Meth. **A538** (2005) 249.

[67] J. Goldstein at al. *Silicon Tracking for Plug Electrons* CDF Note **5970**.

[68] A. Abulencia et al. (CDF Collaboration) *Measurements of Inclusive W and Z Cross Sections in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV* J. Phys. **G** Nucl. Part. Phys. **34** (2007) 2457.

[69] D. Acosta et al. (CDF Collaboration) *Direct Photon Cross Section with Conversions at CDF* Phys. Rev. **D** 70 (2004) 074008.

[70] A. Taffard *Run II Cosmic Ray Tagger Performances* CDF Note **6255**.

[71] E. Lipeles et al. *Lepton ID for Multilepton Diboson Analyses* CDF Note **8538**.

[72] G.C. Blazey et al. *Run II Jet Physics* CDF Note **5293**.

[73] F. Abe et al. (CDF Collaboration) *The Topology of Three Jet Events in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV* Phys. Rev. **D** 45 (1992) 1448.

[74] A. Bhatti et al. (CDF Collaboration) *Determination of the Jet Energy Scale at the Collider Detector at Fermilab* Nucl. Instrum. Meth.**A566** (2006) 375.

[75] T. Sjostrand et al. *High-Energy-Physics Event Generation with PYTHIA 6.1* Comput. Phys. Commun. **135** (2001) 238.

[76] B. Cooper and A. Messina, *Estimation of the Background to W→ e+n jets Events* CDF Note **7760**.

[77] Henry Bachacou *A Measurement of the Production Cross Section of tt-bar Pairs Using Secondary Vertex b-tagging* CDF Note **7632**.

[78] W. Yao et al. *A seed vertexing b-Tag Algorithm for Top* CDF Note **2716**.

[79] L. Cerrito, A. Taffard *A Soft Muon Tagger for Run II* CDF Note **7122**

[80] J. Freeman, W. Ketchum, S. Poprocki, S. Pronko, V. Rusu, and P. Wittich *Search for Diboson Production in MET+$b\bar{b}$ channel* CDF Note **10204** [arXiv:1108:2060].

[81] M. L. Mangano et al. *ALPGEN, a Generator for Hard Multiparton Processes in Hadronic Collisions* J. High Energy Phys. **07** (2003) 001

[82] R. Culbertson et al *Search for Anomalous Production of di-photon+MET Events in 2 fb^{-1} of Data* CDF **9184**

[83] J. Adelman et al. *Method II for you* CDF Note **9185** (2008)

[84] Frankiln et al. *Calibration of Heavy-Flavor Production in QCD Data* CDF Note **8768** 2007

[85] Frankiln et al. *Heavy-Flavor Content of the W+Jets Sample* CDF Note **8765** 2007

[86] T. Aaltonen et al (CDF Collaboration) *Measurement of Inclusive Jet Cross Sections in $Z/\gamma^*(\rightarrow e^+e^-)$+ jets Production in $p\bar{p}$ Collisions at $\sqrt{s}1.96$ TeV* Phys. Rev. Lett. **100** (2008) 102001

[87] M. L. Mangano et al *ALPGEN, a generator for hard multiparton processes in hadronic collisions J. High Energy Phys.* **0307** (2003) 001

[88] T. Sjöstrand et al *Computer Phys. Commun.* **135** (2001) 238

[89] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss *TMVA - Toolkit for Multivariate Data Analysis* (2007) arXiv:physics/0703039

[90] W. Ketchum, V. Rusu, Y.K. Kim *Search for WZ/ZZ Production in leptons+jets channel* CDF Note **10581** (2011)

[91] G. Cowan *Statistical Data Analysis* Oxford Science Publications (1998)

[92] Tom Junk *Frequentist Exclusion and Discovery of Signals in Combined Searches with Large Systematic Uncertainties on the Background* CDF Note **6525** (2003)

[93] Tom Junk *Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematics* CDF Note **8128**

[94] F. James and M. Roos *MINUIT: A system For Function Minimization And Analysis Of The Parameter Errors and Correlations* Comput. Phys. Commun. **10** (1975) 343

[95] Tom Junk *Building a More General χ^2* CDF Note **7904** (2006)

[96] F. Canelli et al. *Jet Energy Resolution* CDF Note **7856**

[97] H. L. Lai et al. *Global QCD Analysis of Parton Structure of the Nucleon: CTEQ5 Parton Distributions* Eur. Phys. J. C **12** (2000) 375 [arXiv:hep-ph/9903282v3].

Acknowledgements

Usually I don't say thank you and I let facts to demonstrate how much I am grateful. But at this point I feel duty bound to thank all the people that in different ways have supported me in the last 5 years.

Thanks to my parents, who allowed me to study in Pisa, for giving me the freedom to follow my ambitions and to do my choices.

Thanks to Marco Trovato for helping me in resolving codes and conceptual issues, for reproaching me for my laziness in the optimization of codes, but especially for supporting me everyday in my experience in CDF.
Thanks to Prof. Giorgio Bellettini who let me work independently, but always following my steps, for his patience in correcting this thesis.
Thanks to Giuseppe Latino, Gueorgui Velev and Vadim Rusu for the illuminating suggestions.
Thanks to Matteo Corbo for his precious help for silicon stuff.
Thanks to Giorgio Chiarelli for the patience in helping me in organizing the periods I spent in Fermilab.

During the period I worked at this thesis I spent some months in Fermilab. I want to especially thank some of the guys I met there for the funny weekends and thoughtless breaks. So, thanks to Mirco, Lucia, Sabatino and Patrizia.
Thanks to Viviana and Angelo for being so thoughtful towards me and also for the precious help in writing this thesis and the suggestions about my future plans.
Thanks to Matteo for spending so much time on chatting with me, giving "almost always" an useful suggestion.

This thesis is just the final part of a very long way.
Thanks to Matteo, *Jon* for saying always what I don't want to admit.
Thanks to Daria, who has always the right words even when she uses only two fingers on the keyboard as during the periods I spent in Chicago.
Thanks to Elia encouraging me since the first months in Pisa, for always being there for me.
I want to say thank to Valeria, my friend and roommate in Pisa, who has been my family in the last years.

A special thanks to Peo, with whom I shared the most part of these years at the University. I could do it even without you but it wouldn't be so beautiful.