Università degli Studi di Trieste

Sede Amministrativa del Dottorato di Ricerca

Scuola di Dottorato di Ricerca in Fisica
(XXII ciclo del Dottorato di Ricerca)

Search for a Standard Model Higgs boson in the $\tau\tau$ decay channel produced in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV at Tevatron

Settore scientifico-disciplinare FIS/01

Dottorando:
Pierluigi Totaro

Coordinatore del Collegio dei Docenti:
Ch.mo Prof. Paolo Camerini
(Università degli Studi di Trieste)

Supervisore:
Dott. Giuseppe Della Ricca
(Università degli Studi di Trieste)

Relatore:
Dott.sa Anna Maria Zanetti
(INFN, Sezione di Trieste)

Anno Accademico 2009 – 2010

Alla mia amata Fabietta
e al Pincifruttolo
che già tanta gioia ci sta donando

Abstract

This thesis describes the search for the Standard Model Higgs boson decaying to tau lepton pairs, in the Tevatron proton-antiproton collisions at a center of mass energy $\sqrt{s}$ = 1.96 TeV. The search is based on approximately 2.3 fb^{-1} of CDF Run II data and is performed by considering the following signal processes: WH($\rightarrow \tau\tau$), ZH($\rightarrow \tau\tau$), qHq'$\rightarrow$q$\tau\tau$q' and gg$\rightarrow$H$\rightarrow \tau\tau$.

Events are selected by requiring an hadronic tau and one isolated electron or muon, coming from the leptonic decay of one of the two taus. In addition, at least one calorimeter jet must be present in the final state.

We expect 921.8$\pm$48.9 background events in the 1 jet channel and 159.4$\pm$11.6 in the $\geq$ 2 jets channel, while in data we observe 965 and 166 events, respectively.

In order to improve the search sensitivity we employ a multivariate technique, based on a set of Boosted Decision Trees trained to get the best separation between signal and the dominant sources of background. We observe no evidence for a Higgs boson signal and therefore we set a 95% confidence level (C.L.) upper limit on the cross section relative to the SM predictions ($\sigma/\sigma_{\rm SM}$).

Results are presented for the Higgs boson mass varying from M_H = 100 GeV/c^2 to M_H = 150 GeV/c^2. For the mass hypothesis of 120 GeV/c^2 the observed limit is 27.2, while the corresponding expected value is $23.4^{+9.8}_{-6.4}$.

Sommario

In questa tesi viene descritta la ricerca del bosone di Higgs del Modello Standard nel canale di decadimento in coppie di leptoni τ, condotta al Tevatron in collisioni protone-antiprotone ad un'energia nel centro di massa $\sqrt{s} = 1.96$ TeV. L'analisi si basa su circa 2.3 fb^{-1} di dati raccolti dall'esperimento Collider Detector at Fermilab (CDF) ed è stata effettuata considerando i seguenti processi di produzione di segnale: WH($\rightarrow \tau\tau$), ZH($\rightarrow \tau\tau$), qHq'$\rightarrow$q$\tau\tau$q' and gg$\rightarrow$H$\rightarrow \tau\tau$.

Gli eventi sono selezionati richiedendo un τ adronico ed un elettrone (o muone) isolato, quest'ultimo proveniente dal decadimento leptonico di uno dei due τ originari. In aggiunta, almeno un getto calorimetrico deve essere presente nello stato finale.

Il numero di eventi di fondo attesi è 921.8$\pm$48.9 nel canale con un unico getto e 159.4$\pm$11.6 nel canale con due o più getti calorimetrici, mentre gli eventi osservati sono rispettivamente 965 e 166.

Per aumentare la sensitività dell'analisi si è impiegato di un metodo multivariato, basato su un set di *Boosted Decision Trees* allenati per separare il segnale dalle principali sorgenti di fondo. I dati non mostrano alcuna evidenza di produzione di bosoni di Higgs, pertanto sono stati calcolati i limiti al 95% di Livello di Confidenza sulla sezione d'urto di produzione, relativamente ai valori previsti dal Modello Standard.

I risultati sono presentati per diversi possibili della massa del bosone di Higgs, nell'intervallo comopreso tra 100 GeV/c^2 e 150 GeV/c^2. Nell'ipotesi $M_H = 120$ GeV/c^2 il limite osservato è 27.2 volte superiore alla sezione d'urto predetta, mentre il corrispondente valore atteso è $23.4^{+9.8}_{-6.4}$.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Standard Model (SM) of Particle Physics is the current theory that provides the best description of the fundamental particles, quarks and leptons, and their interactions which result from the exchange of vector gauge bosons.

The theory was developed in the early and middle 20th century and the current formulation, finalized in the 1970's, incorporates three of the four known forces of Nature: the electromagnetic, the weak and the strong one.

The SM was able to successfully explain over the past few decades an incredible wide range of phenomena and on the experimental side many precision measurements have been performed providing us with stringent confirmations of the SM predictions with increasing accuracy over time. Figure 1.1 shows how well several SM processes, which differ in the production cross section at the Tevatron for several orders of magnitude, are described by the CDF experiment.

Despite its many successes, the SM in its original form does not account for one of the most relevant properties of the observed particles, the mass. The simplest solution to incorporate in the theory the description of masses, in a consistent and gauge invariant way, is by means of the *Higgs mechanism*: a new scalar field is postulated and a non-zero mass is acquired by the gauge bosons through the spontaneous symmetry breaking process.

This mechanism implies the existence of a new scalar particle, the Higgs boson, which has not been experimentally observed yet. The validity of the Higgs model as the way to explain the origin of mass can be confirmed only by the experimental evidence of the existence of the Higgs boson. However, the small predicted production cross section, compared to the other SM processes, makes this search extremely challenging.

Extensive searches were performed first by the experiments of the e^+e^- LEP collider, which reached the operational center of mass energy of 209 GeV in 2000: a constraint to the Higgs boson mass $m_H > 114.4$ GeV/c^2 at 95% C.L. was set. Searches are now being

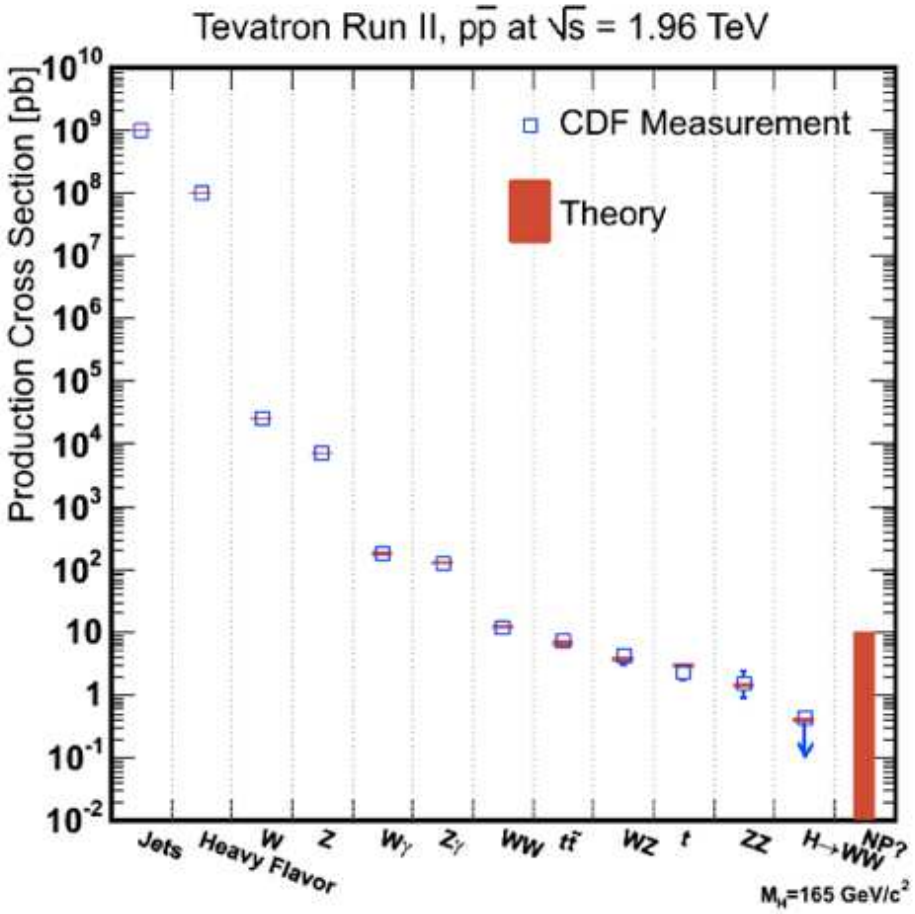

Figure 1.1: SM Production cross sections at Tevatron: the blue squares correspond to the CDF measurements, while the theoretical predictions are reported in red.

carried on at the Tevatron $p\bar{p}$ collider, where CDF and DØ have continuously improved the sensitivity by including more data in the analyses and by developing advanced techniques to separate small signal contributions from huge background samples. The most recent Tevatron combined results allowed us to exclude the Higgs boson at 95% C.L. in the $100 < m_H < 109$ GeV/c and $^2158 < m_H < 175$ GeV/c^2 mass ranges.

In this thesis we present the search for the Higgs boson in the $\tau\tau$ decay mode using approximately 2.3 fb^{-1} of CDF Run II data, collected at the Tevatron $p\bar{p}$ collider at the center of mass energy of 1.96 TeV.

An overview of the key concepts of the SM theory is given in chapter 2, focusing on the electroweak sector and on the spontaneous symmetry breaking provided by the Higgs mechanism.

The main properties of the Higgs boson, the theoretical and experimental constraints to the mass and its relations to the production cross sections and decay branching ratios are discussed in chapter 3.

Chapter 4 introduces the experimental apparatus used in this thesis: the Fermilab accelerator complex and the CDF detector.

The Higgs search strategies, the most recent Tevatron results and an overview of the Higgs channels studied in this thesis are introduced in chapter 5, while a description of the data sample and of the $H \to \tau\tau$ event selection is provided in chapter 6.

The modeling of the signal and background contributions is discussed in chapter 7. Results are finally presented in chapter 8: a multivariate technique, adopted to improve the separation of signal from background is described; systematic uncertainties related to the analysis are also discussed, along with the bayesian method adopted to place an upper limit to the production cross section.

In chapter 9 we conclude the thesis by summarizing the results obtained in the analysis, by describing the most recent combined measurement of the CDF and DØ collaborations and by discussing the future prospects for the $H \to \tau\tau$ analysis and the low mass Higgs search in general at the Tevatron and LHC experiments.

Chapter 2

The Standard Model and the Higgs boson

"La mathematica è l'alfabeto in cui Dio à scritto l'universo."
Galileo Galilei

The Standard Model represents our best knowledge to explain nature at the subatomic level, by providing a description of the fundamental constituents of matter and of the forces which act between them. The SM is a quantum field theory which unifies the electromagnetic, weak and strong interactions, giving predictions for a wide variety of phenomena, which have been tested in several experiments over the last decades, with an unprecedented level of accuracy.

The SM does not incorporate in its framework the fourth force which characterizes the existence of the universe and our everyday life, the gravitation: however, at the considered energy scale the effect of gravity is extremely small if compared to the other forces, and can thus be neglected.

An exhaustive description of the theory can be found elsewhere in the scientific literature [14–16]; in this chapter we present a short overview of the key concepts, focusing on the electroweak sector and on the Higgs mechanism, which provides in an elegant mathematical formalism a simple way to assign mass to particles, via spontaneous symmetry breaking.

A consequence of this mechanism is the prediction of a new heavy neutral particle, the Higgs Boson, whose existence has not yet been experimentally confirmed. The search of this ultimate fundamental particle is currently the most challenging quest for the high energy physics community and it represents the main topic of this thesis.

2.1 Fundamental particles

The SM describes the universe in terms of interacting particle fields, which represent the building blocks of nature and are divided into two categories: spin $\frac{1}{2}$ fermions are the basic constituent of observable matter, while spin 1 bosons are responsible for carrying the interactions among fermions.

Fermions are classified into lepton and quarks, which are both further organized into three parallel families, called *generations*, that describe the mass hierarchy and the flavor structure among particles, with the lightest ones belonging to the first generation. There are six types (*flavors*) of quarks, plus their antiparticle partners. Quarks interact via both strong and electroweak forces, have a fractional electric charge and combine together to form hadrons, bound states of 2 or 3 quarks, called mesons and baryons respectively. Up and down quarks, members of the first generation, form neutrons and protons, which are the baryons with the smallest mass and represent the familiar constituents of common matter, inside nuclei.

In analogy with quarks, there are six kinds of leptons with their corresponding antiparticles, which are classified into three doublets, each consisting of a neutral and a charged lepton. They are subject to the electroweak interaction and not to the strong force.

Quarks and leptons properties of electric charge and mass, for the three generations, are summarized in table 2.1.

Fundamental fermions (spin 1/2)					
Quarks			Leptons		
Flavor	Charge	Mass[GeV/c^2]	Flavor	Charge	Mass[MeV/c^2]
up (u)	+2/3	0.0017-0.0033	el.neutrino (ν_e)	0	<0.003
down (d)	-1/3	0.0042-0.0058	electron (e)	-1	0.511
charm (c)	+2/3	1.18-1.34	muon neutrino (ν_μ)	0	<0.19
strange (s)	-1/3	0.080-0.130	muon (μ)	-1	105.6
top (t)	+2/3	172.0±0.9±1.3	tau neutrino (ν_τ)	0	<18.2
bottom (b)	-1/3	$4.19^{+0.18}_{-0.06}$	tau (τ)	-1	1776.82±0.16

Table 2.1: Fermions (quarks and leptons) in the Standard Model. Mass values are taken from [1].

Vector bosons are the particles that mediate the forces: the photon carries the electromagnetic force, gluons carry the strong one, while the charged $W^\pm$ and the neutral Z^0 bosons are responsible for the weak interaction. Table 2.2 lists these gauge bosons, with their corresponding electric charge and mass.

Fundamental bosons (spin 1)			
Name	Interaction	Charge	Mass[GeV/c^2]
photon (γ)	Electromagnetic	0	0
W bosons (W$^{\pm}$)	Weak	±1	80.399$\pm$0.023
Z boson (Z^0)	Weak	0	91.1876$\pm$0.0021
gluons	Strong	0	0

Table 2.2: The gauge bosons in the Standard Model and the corresponding interaction [1]

2.2 Gauge theory overview

The Standard Model mathematical formulation is based on the gauge invariance of a relativistic quantum field theory [17]. This theoretical approach was first adopted with great success in the late 1940s and early 1950s for the development of Quantum ElectroDynamics (QED), in the attempt to describe the electromagnetic interactions in a quantized relativistic invariant form.

The basic idea is to require that the physical laws, which are described by the Lagrangian of a system, remain unchanged under a certain symmetry transformation. A transformation that is space-time dependent is said to be *local*, and a physical system which exhibits symmetry to it is called *gauge invariant*.

In the case of QED, let us consider the Lagrangian density of a massive fermion, represented by the Dirac field ψ

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{2.1}$$

where γ^μ are the Dirac matrices and $\bar{\psi}$ is defined as $\psi^\dagger\gamma^0$.

Under the local U(1) transformation[1]

$$\psi \rightarrow \psi' = e^{iQ\theta(x)}\psi \tag{2.2}$$

where Q is the fermion charge and $\theta(x)$ is a change of phase which is a function of the space-time coordinates. The Lagrangian is not invariant because of the non vanishing extra term coming from the partial derivative of $\theta(x)$.

The problem can be overcome by introducing an additional vector field $A_\mu(x)$, called gauge field, which transforms as

$$A_\mu(x) \rightarrow A'_\mu(x) = A_\mu(x) + \frac{1}{e}\partial_\mu\theta(x) \tag{2.3}$$

where e is the usual charge of the electron. By redefining the derivative of the particle field

[1]In QED, the global symmetry under U(1) transformation, where θ is a fixed parameter in space-time, is responsible for the conservation of the electric charge.

$$\partial_\mu \rightarrow D_\mu = \partial_\mu - ieQA_\mu(x) \tag{2.4}$$

the Lagrangian of the system becomes

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi + eQA_\mu(\bar{\psi}\gamma^\mu\psi) \tag{2.5}$$

where the invariance is elegantly restored and the interaction of the particle field ψ with A_μ, which represents in this formulation the electromagnetic (EM) field, appears in a natural way in the last term. The Lagrangian is finally completed by adding the kinetic term, which comes from the Maxwell's equations:

$$\mathcal{L}_K = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.6}$$

where the field strength tensor is defined as $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$.

A mass term like $\frac{1}{2}m^2 A_\mu A^\nu$ is not allowed by gauge invariance, therefore the quantum of the EM field (the photon) has to be massless, as it is largely confirmed by the experimental observations.

In analogy with QED, the procedure followed in the construction of the SM is defined by building an overall gauge symmetry group:

$$\mathrm{SU(3)_C} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y} \tag{2.7}$$

where C is a reminder that SU(3) represents the symmetry group which describes the *colored* strong force of quantum chromodynamics (QCD) [18]. The L indicates that the SU(2) group is related to the left-handed weak isospin doublets, while the Y refers to the right-handed weak hypercharge singlets contained in the U(1) group.

The $\mathrm{SU(2)}_L \times \mathrm{U(1)}_Y$ groups provide a unified formulation of the electromagnetic and the weak interactions. A more detailed overview of this sector of the SM is given in the next section.

2.3 The Electroweak theory

The electroweak theory, which results from the unification of QED and weak interactions, was formulated in the middle of the 1960s, with the purpose of describing the two forces as different manifestations of the same fundamental interaction [19, 20].

The theory, based on the $\mathrm{SU(2)}_L \times \mathrm{U(1)}_Y$ symmetry group, is found to be *chiral*, which means that the right-handed and the left-handed components of the spinors, given by

$$\psi_L = \frac{1-\gamma_5}{2}\psi \tag{2.8}$$

$$\psi_R = \frac{1+\gamma_5}{2}\psi \tag{2.9}$$

transform in different ways under the local gauge transformation for the group,

$$\psi'_L(x) = U\psi_L(x) \tag{2.10}$$

$$\psi'_R(x) = U\psi_R(x) \tag{2.11}$$

U is defined as

$$U = e^{i\boldsymbol{\alpha}(x)\cdot\mathbf{T_i}+i\beta(x)Y} \tag{2.12}$$

where $T_i = \frac{1}{2}\sigma_i$ (with i=1,2,3) are the three generators for the $SU(2)_L$ group, the hypercharge Y is the generator of $U(1)_Y$, with $\boldsymbol{\alpha}(x)$ and $\beta(x)$ representing real and continuous functions which parametrize the space-time dependent phase change of the transformation. σ_i are the Pauli spin matrices. The commutation relations of the group are

$$[T_i, T_j] = i\epsilon_{ijk}T_k \tag{2.13}$$

$$[T_i, Y] = 0; \quad i, j, k = 1, 2, 3 \tag{2.14}$$

$$\tag{2.15}$$

$SU(2)_L$ represents the weak isospin group, which acts only on the left-handed fermions, while $U(1)_Y$ is the weak hypercharge group. As a consequence, left-handed fermions form isospin doublets, while right handed components transform as isospin singlets:

$$f_l^i = \begin{pmatrix} \nu_L^i \\ l_L^i \end{pmatrix}, \begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix} \tag{2.16}$$

$$f_R^i = l_r^i, u_R^i, d_R^i \tag{2.17}$$

where i = 1,2,3 corresponds to the family index. The weak hypercharge Y is related to electric charge Q and the third component T_3 of the weak isospin T by the following formula:

$$Q = T_3 + \frac{Y}{2} \tag{2.18}$$

Quantum numbers for the first generation of fermions are shown in table 2.3.

When requiring the Lagrangian describing a massless fermion field

$$\mathcal{L}_f = i\bar{\psi}\gamma^\mu\partial_\mu\psi \tag{2.19}$$

to be invariant under the local gauge transformation described above, the following covariant derivative and vector boson fields need to be introduced:

Fermions	T	T_3	Q	Y
ν_L	1/2	1/2	0	-1
e_L	1/2	-1/2	-1	-1
e_R	0	0	-1	-2
u_L	1/2	1/2	2/3	1/3
d_L	1/2	-1/2	-1/3	1/3
u_R	0	0	2/3	4/3
d_R	0	0	1/3	-2/3

Table 2.3: First generation fermion electroweak quantum numbers.

$$D_\mu = \partial_\mu + ig\mathrm{T}\cdot\mathrm{W}_\mu(x) + ig'\frac{Y}{2}B_\mu(x) \tag{2.20}$$

$$\mathrm{W}_\mu(x) \to (W)'_\mu = \mathrm{W}_\mu(x) - \frac{1}{g}\partial_\mu\boldsymbol{\alpha}(x) - \boldsymbol{\alpha}(x)\times\mathrm{W}_\mu(\ x) \tag{2.21}$$

$$B_\mu(x) \to B'_\mu = B_\mu - \frac{1}{g'}\partial_\mu\beta(x) \tag{2.22}$$

where W_μ and B_μ are the gauge fields that correspond to $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ respectively, and g and g' are the coupling constants for the corresponding interaction.

Similarly to QED, the kinetic terms for the gauge fields need to be added to the Lagrangian:

$$\mathcal{L}_G = -\frac{1}{4}\mathrm{W}_{\mu\nu}\mathrm{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{2.23}$$

with the field strength tensors defined as

$$\mathrm{W}_{\mu\nu} = \partial_\mu\mathrm{W}_\nu - \partial_\nu\mathrm{W}_\mu - gW_\mu\times W_\nu \tag{2.24}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.25}$$

The physical experimentally observed gauge bosons, the two charged $\mathrm{W}^\pm_\mu$ and the neutral Z_μ and A_μ are obtained as linear combinations of the electroweak eigenstates, shown in the following expressions:

$$W^\pm_\mu = \frac{W^1_\mu \mp iW^2_\mu}{\sqrt{2}} \tag{2.26}$$

$$Z_\mu = W^3_\mu\cos\theta - B_\mu\sin\theta \tag{2.27}$$

$$A_\mu = W^3_\mu\sin\theta + B_\mu\cos\theta \tag{2.28}$$

where a rotation angle θ, called Weinberg angle (or weak mixing angle), is introduced. It comes out that the coupling constants g and g' and the electron charge have the following simple relations:

$$g' = g \tan\theta \tag{2.29}$$

$$e = g \sin\theta \tag{2.30}$$

Again the gauge invariance is granted by ensuring that the gauge bosons and the fermions involved in the interactions are massless: terms such as $m^2 W_\mu W^\nu$ or $m^2 B_\mu B^\nu$ for bosons and $m^2 \bar{f} f$ for fermions are thus not allowed in the Lagrangian. However, this requirement contradicts the experimental observations of massive particles.

In the SM the Higgs mechanism of the Spontaneous Symmetry Breaking (SSB) is introduced to provide us with the generation of masses without violating the gauge invariance.

2.4 The Higgs Mechanism

In a paper published in 1964 [21], Peter Higgs described a mechanism by which gauge bosons may acquire mass via spontaneous symmetry breaking of a transformation group. The SSB occurs when the Lagrangian which describes the dynamics of a physical system exhibits a symmetry which is no longer preserved by the configuration of the field in its ground state.

The Higgs mechanism has been incorporated into the SM and it is considered responsible for the generation of gauge boson and fermion masses. In addition, the model predicts the existence of a new spin-0 scalar, not yet experimentally observed, whose mass and self-interaction are not determined by the theory.

Generally speaking, when the SSB is introduced in a given gauge theory based on a local invariance with respect to a symmetry group G, with dimension N, and the vacuum state keeps a symmetry H of dimension M, N-M massless scalar Goldstone bosons, corresponding to the generators of the broken symmetry, will be absorbed into the longitudinal components of N-M massive vector bosons.

In the specific case of the electroweak theory, the symmetry group $\mathrm{SU(2)}_L$ X $\mathrm{U(1)}_Y$ (dim(G)=4) is spontaneously broken down to the $\mathrm{U(1)}_{em}$ symmetry of the electromagnetism ((dim(H)=1), related to the electric charge, which need to be conserved in the vacuum:

$$\mathrm{SU(3)}_C \times \mathrm{SU(2)}_L \times \mathrm{U(1)}_Y \rightarrow \mathrm{SU(3)}_C \times U(1)_{em} \tag{2.31}$$

As a consequence, three vector Goldstone bosons appear and are absorbed by three of the four gauge bosons of the electroweak theory, leaving the last one (the photon) massless. Due to this process, gauge bosons naturally acquire the proper mass in a gauge invariant way.

The mechanism requires the introduction of a new field Φ (the Higgs field), with at least 3 degrees of freedom. The simplest choice is a complex SU(2) doublet of electroweak scalar fields, such as

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.32}$$

With hypercharge $Y(\Phi) = 1$. The additional term in the Lagrangian is:

$$\mathcal{L}_{SSB} = (D_\mu \Phi)^\dagger D^\mu \Phi - V(\Phi) \tag{2.33}$$

where the simplest normalizable potential $V(\Phi)$ is

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2. \tag{2.34}$$

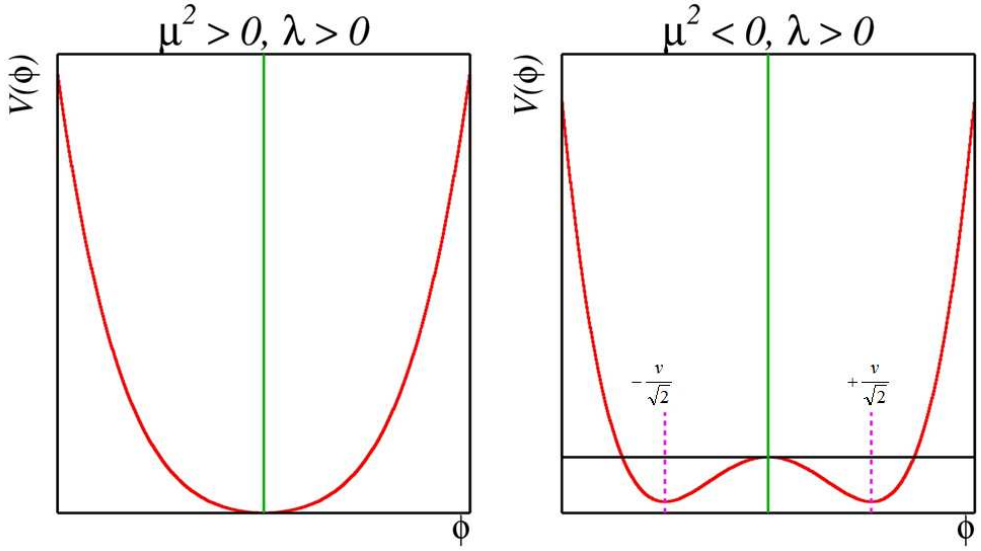

Figure 2.1: The Higgs potential V(Φ), with μ^2 >0 and λ >0 (left) and μ^2 <0 and λ >0 (right)

If $\mu^2 < 0$ and $\lambda > 0$ the potential shows a minimum for $\sqrt{\frac{-\mu^2}{2\lambda}}$, i.e. the field has a non zero vacuum expectation value (VEV):

$$\langle \Phi \rangle_0 = \langle 0 | \Phi | 0 \rangle = \frac{v}{\sqrt{2}} \tag{2.35}$$

with $v = \sqrt{\frac{-\mu^2}{\lambda}}$. This means that in the lowest energy state the system no longer reflects the symmetry of the original potential $V(\Phi)$. In order to preserve the electromagnetic symmetry, the ground state is chosen in such a way that only the component of the field with no electric charge acquires a VEV:

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.36}$$

By expanding the field around the minimum,

$$\Phi(x) = \frac{1}{\sqrt{2}} e^{\frac{i\xi(\vec{x})\vec{\sigma}}{v}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.37}$$

and by applying a gauge transformation in order to eliminate the fields $\vec{\xi}(x)$ which represent the Goldstone bosons, we obtain the tree level masses for the gauge bosons from the term

$$(D_\mu \Phi)^\dagger D^\mu \Phi = \frac{g^2 v^2}{4} W_\mu^+ W^{-\mu} + \frac{(g^2 + g'^2)v^2}{8} Z_\mu Z^\mu + \dots \tag{2.38}$$

while the Higgs boson itself, coming from the survived field H(x), gets the mass from the potential term $V(\Phi)$, which becomes

$$V(\Phi) = \frac{1}{2}(2\mu^2)H^2 + \dots \tag{2.39}$$

The explicit mass values are:

$$M_{W^\pm} = \frac{vg}{2} \tag{2.40}$$

$$M_Z = \frac{v\sqrt{g^2 + g'^2}}{2} \tag{2.41}$$

$$M_H = \sqrt{2\lambda} v \tag{2.42}$$

$$M_\gamma = 0 \tag{2.43}$$

where v, which represents the energy scale of the spontaneous symmetry breaking, is determined experimentally from the muon decay measurement at low energies, by exploiting the relation with the Fermi constant G_F:

$$\frac{G_F}{2} = \frac{g^2}{8M_W^2} = \frac{1}{2v^2} \tag{2.44}$$

. It turns out that $v \approx$246 GeV. The theory does not provide any direct information for the Higgs mass, because λ is a free parameter.

2.4.1 Fermion masses

Not only gauge bosons, but also fermions acquire mass from the Higgs doublet previously defined, by introducing in the Lagrangian a term which describes the Yukawa interaction between the Higgs field Φ and the left handed isospin doublets and the right-handed singlet spinors:

$$\mathcal{L}_{YW} = -g_f(\bar{\Psi}_L \Phi \Psi_R + \bar{\Psi}_R \bar{\Phi} \Psi_L) \tag{2.45}$$

the coupling constant is not fixed by the theory. By expliciting the expression in terms of the VEV, we obtain the following expression for the fermion masses:

$$m_f = \frac{g_f v}{2} \tag{2.46}$$

We now have all the elements to build the complete Lagrangian for the electroweak sector of the SM, by collecting all the terms described throughout the last sections:

$$\mathcal{L}_{SM} = \mathcal{L}_f + \mathcal{L}_G + \mathcal{L}_{SSB} + \mathcal{L}_{YW} \tag{2.47}$$

Chapter 3

Properties of the Higgs boson

"The God Particle: If the Universe Is the Answer, What Is the Question?"
Leon M. Lederman

The Higgs mechanism, by introducing a new scalar field in the Standard Model theory, describes in an elegant and consistent way how fundamental particles may acquire mass. This mechanism also predicts the existence of a new neutral scalar particle, whose properties have been extensively studied in the last decades.

Even if the mass of the SM Higgs boson is directly connected to the energy scale at which the electroweak symmetry is broken to give mass to the W and Z bosons, the value of the mass itself m_H is a free parameter of the model.

No experimental evidence for the existence of this particle has been observed yet; however, several theoretical and experimental constraints, which are discussed in this chapter, limit the possible mass value to a restricted range, where all the current searches are focused.

For a given mass value, Higgs production modes cross sections and decay branching ratios (B.R.'s) can be easily predicted by the theory. In the SM the Higgs boson couples preferentially to heavy particles: the dominant mechanisms involve therefore the $W^{\pm}$, Z^0 bosons and the third generation of quarks and leptons. At higher order, when loop processes are considered, Higgs may couple also to massless particles, like gluons and photons.

The main properties of the Higgs boson at the Tevatron, where proton-antiproton collisions take place at the center of mass energy of 1.96 TeV, are here described.

3.1 The mass

3.1.1 Theoretical constraints

A first theoretical constraint to the Higgs mass is provided by the *unitarity* condition [22]. At high energy, the elastic scattering amplitude of longitudinal vector boson pairs, WW→WW and ZZ→ZZ, increase indefinitely with the energy of the incoming particles. The unitarity can be restored by considering the exchange of the Higgs scalar particle and by imposing the s-wave unitarity condition, from which we get an upper bound on the Higgs mass:

$$m_H < \sqrt{\frac{8\pi\sqrt{2}}{3G_F}} \approx 1 \text{ TeV} \tag{3.1}$$

A second upper limit comes from the condition of *triviality* of the Higgs potential [23]. The coupling constant λ has a functional dependence on the energy scale Λ, from the renormalization group equation [24]. A solution of the equation can be written by choosing the electroweak symmetry breaking (EWSB) scale Λ_0=v:

$$\lambda(\Lambda^2) = \lambda(v^2)\left[1 - \frac{3}{4\pi^2}\lambda(v^2)\log\left(\frac{\Lambda^2}{v^2}\right)\right]^{-1} \tag{3.2}$$

According to this solution, if the energy is much smaller than the EWSB scale, $Q^2 \ll v^2$, $\lambda(\Lambda^2)$ becomes small and the theory is trivial as the coupling approaches zero. On the other side, if $Q^2 \gg v^2$, λ eventually becomes infinite at the so called Landau pole energy Λ_C. The relation between this energy Λ_C and the Higgs boson mass is

$$m_H^2 = \frac{8\pi^2 v^2}{3\log(\Lambda_C^2/v^2)} \tag{3.3}$$

The Higgs mass has to be smaller than m_H in order to avoid the Landau pole. For $\Lambda = 10^{19}$ GeV(the Planck scale) it is found that $m_H < 180\pm6$ GeV [23].

A lower bound is obtained by the *stability* condition, which requires that the minimum of the Higgs potential $\langle\Phi\rangle_0 = \langle 0|\Phi|0\rangle = \frac{v}{\sqrt{2}}$ is absolute, preventing from instabilities generated by quantum loop corrections. This condition is equivalent to the request that the quartic self coupling constant of the Higgs potential remain positive at large energy scales Λ, $\lambda(\Lambda) > 0$. Such a study has been performed at the two-loop level [25], where at $\Lambda = 10^{19}$ GeV it results that $m_H > 135$ GeV.

The triviality upper bound and the stability lower bound are shown as a function of the energy scale Λ in figure 3.1.

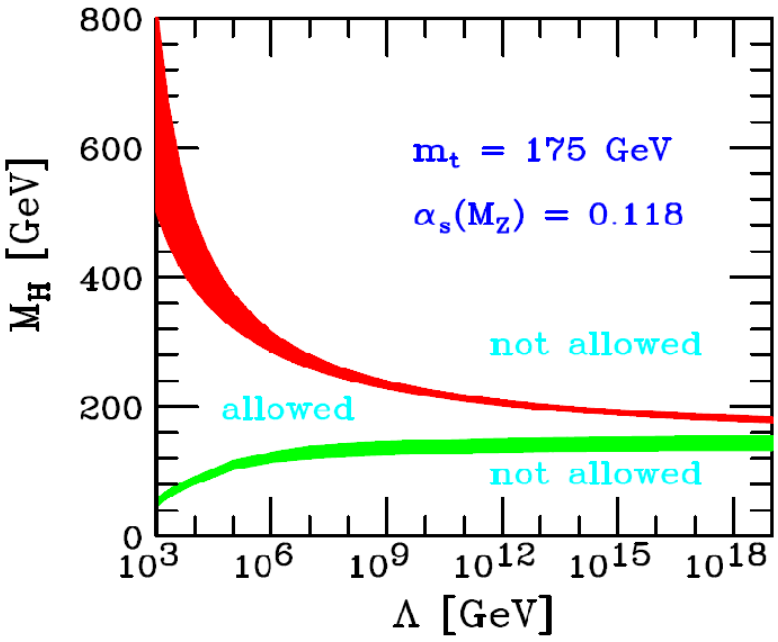

Figure 3.1: Summary of the theoretical bounds on the Higgs mass: triviality (red) and vacuum stability (green) bounds as a function of the cut-off scale Λ.

3.1.2 Experimental constraints

Direct searches at LEP II

No experimental evidence of the Higgs boson existence has been found yet. The most stringent limit on the Higgs mass comes from the direct searches performed at the Large Electron Positron Collider (LEP) by the experiments ALEPH, DELPHI, L3 and OPAL, which analyzed e^+e^- collisions at a center of mass energy $\sqrt{S}$ between 189 and 209 GeV [7].

These searches, covering a total of 2461 pb^{-1} of integrated luminosity collected by the four detectors, considered any possible signature provided by Higgs bosons produced in association of a Z or through vector boson fusion.

Figure 3.2 shows the leading order Feynman diagrams for these Higgs production processes, while figure 3.3 presents the reconstructed candidate Higgs mass distributions for two different event selections.

Since no significant excess of signal was found, the LEP searches were combined to provide a lower limit on the Higgs Mass, $m_H > 114.4$ GeV/c^2, at 95% confidence level (C.L.).

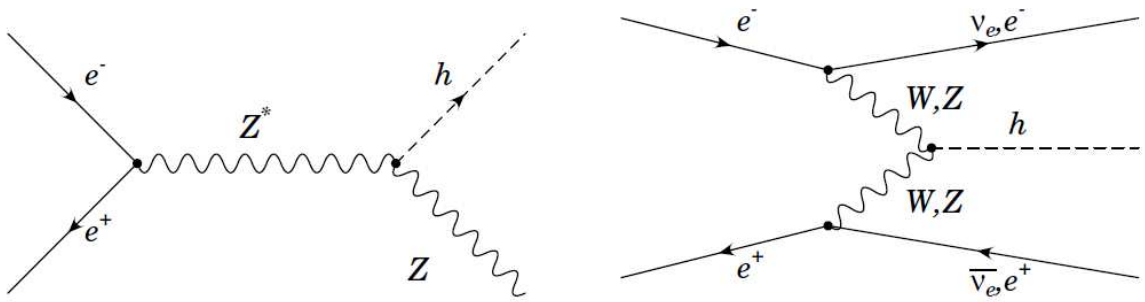

Figure 3.2: The SM leading order diagrams for the Higgs boson production at the LEP electron-positron collider. Higgsstrahlung from a virtual Z boson (left) and W/Z vector boson fusion(right).

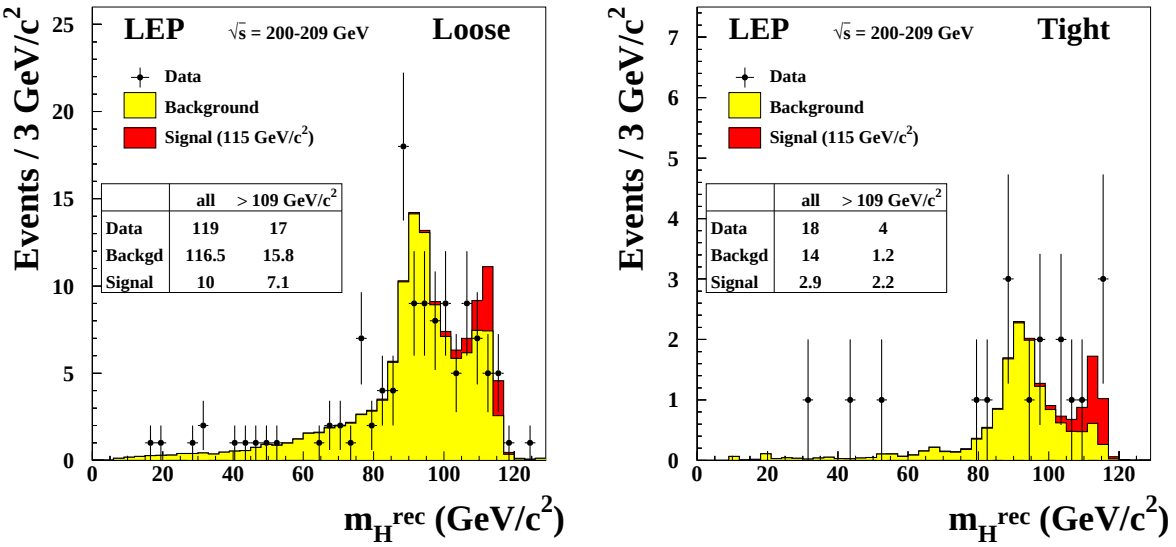

Figure 3.3: Distributions of the reconstructed Higgs boson mass, obtained from two selections with different expected signal purities, at LEP II experiments. The Monte Carlo predictions for background (yellow) and signal (red) are shown, with an hypothetical Higgs mass of 115 GeV/c^2 [7].

Direct searches at Tevatron

CDF and DØ experiments have recently reached the sensitivity to probe the SM Higgs production by analyzing $p\bar{p}$ collisions provided by the Tevatron accelerator complex at the center of mass energy $\sqrt{S} = 1.96$ TeV. At the moment of this writing (Autumn 2010) the combination of both experiments' searches have allowed to exclude the existence of the Higgs boson in the 158-175 GeV/c^2 interval at 95% C.L., by looking

at the H→WW* decay process, with a mean integrated luminosity per experiment corresponding to 5.9 fb^{-1} [26].

More details about the most promising Tevatron search channels and the different analysis strategies are provided in chapter 5.

3.1.3 Indirect constraints

In addition to the constraints provided by direct searches, precision measurements of electroweak parameters of the SM give an indirect indication of the preferred Higgs mass range.

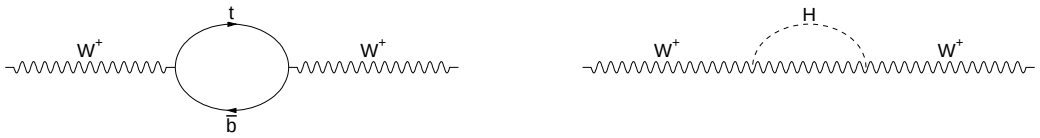

Figure 3.4: One loop Feynman diagrams affecting the mass of the W boson.

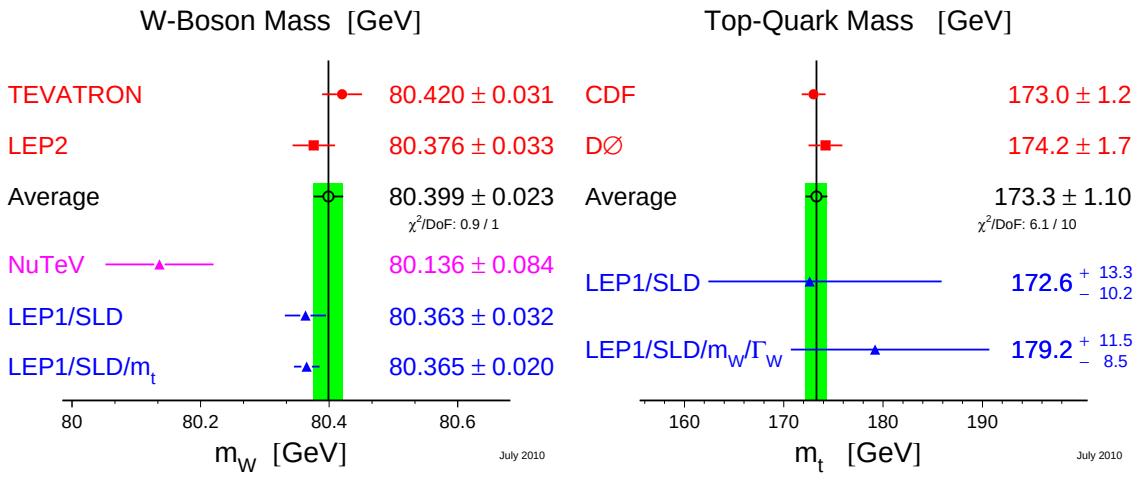

Figure 3.5: Left: W mass direct measurements performed by LEP II and Tevatron experiments; LEP1, SLD and NuTeV indirect constraints within the Standard Model. Right: top mass direct measurements performed by CDF and DØ; LEP1 and SLD indirect constraints [8].

The relative strength ρ of the charged and neutral currents, which relates the W and Z boson masses and the Weinberg θ_W angle,

$$\rho = \frac{m_W^2}{m_Z^2 \cos\theta_W{}^2} = \frac{1}{1-\Delta r} \tag{3.4}$$

is different from unity if loop corrections, involving the top quark and the Higgs boson (figure 3.4), are considered. These corrections, included in the Δr term, have a quadratic dependence from the top mass m_t and the W mass m_W, while the Higgs contribution has a logarithmic dependence, $\ln(m_H^2)$; as a consequence, precision measurements of m_W and m_t can be used to constraint the Higgs mass. The latest results for m_W and m_t are reported in figure 3.5, while the relationship between m_H, m_W and m_t, expressed in 68% C.L. regions, is shown in figure 3.6.

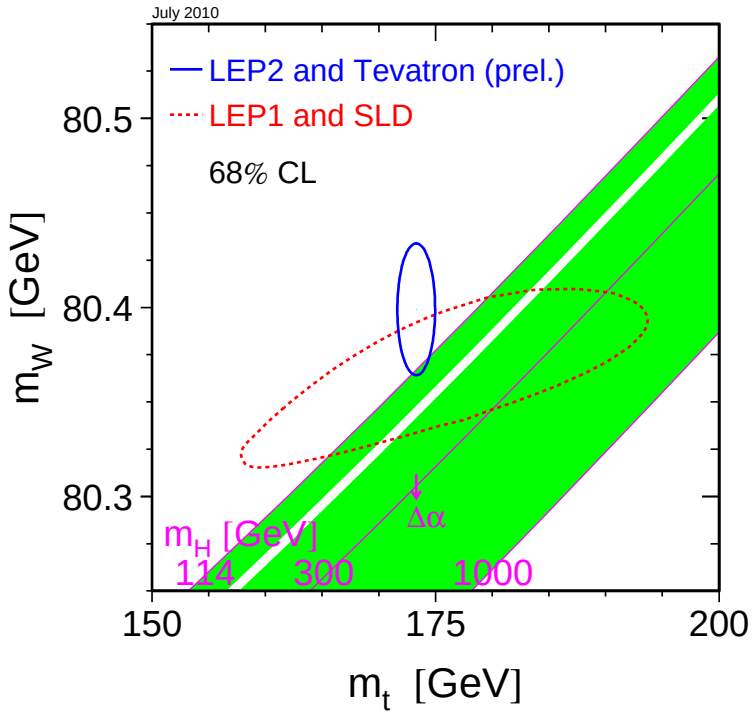

Figure 3.6: Measured W mass versus top mass 68% confidence intervals. The diagonal green bands show the SM allowed regions for a Higgs mass constrained by the current lower LEP limit and the 158-175 GeV/c^2 Tevatron excluded interval (Winter 2010) [8].

Figure 3.7 presents the $\Delta\chi^2=\chi^2-\chi^2_{min}$ curve, as a function of m_H, of a global fit performed by the LEP ElectroWeak Working Group (LEP-EWWG) [8], by combining the precision measurements on electroweak observables from the four LEP experiments, CDF, DØ, NuTeV and SLD [9]. The fit was evaluated by employing the ZFITTER package [27]. The minimum χ^2 divided by the number of degrees of freedom is 17.3/13 = 1.33, which corresponds to a p-value of 19% and a preferred value for the Higgs mass given by $m_H = 89^{+35}_{-26}$ GeV/c^2. The 95% C.L. excluded regions, the one derived from the LEP experiments and the one coming from the CDF and DØ combined searches are also reported.

The 95% C.L. upper limit is $m_H < 158$ GeV/c^2, which becomes $m_H < 185$ GeV/c^2 if the LEP direct search results are included in the fit.

A similar study has been performed by the GFITTER group [10], which used a Bayesian approach [28] instead of the frequentist one applied in [27]. Results are in agreement: the preferred value for the Higgs mass is given in this case by $m_H = 95.7^{+30.6}_{-24.2}$ GeV/c^2. The corresponding $\Delta\chi^2$ curve is reported in figure 3.8.

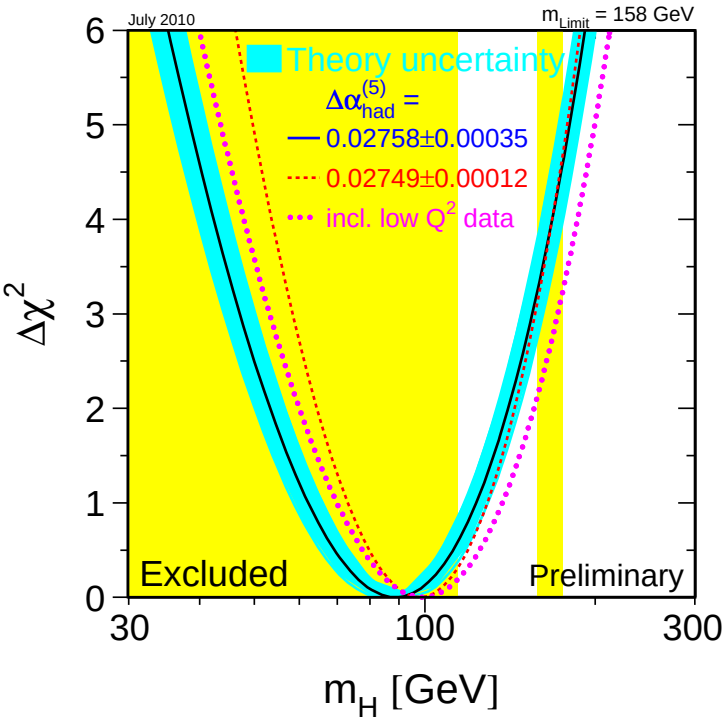

Figure 3.7: $\Delta\chi^2$=χ^2-χ^2_{min} of the global fit to the electroweak precision measurement performed by the LEP-EWWG [8]. The black line is the fit result, while the blue band represents the theoretical uncertainty due to the missing higher order corrections. The dashed lines are the fit results obtained by using different estimation of $\Delta\alpha_{had}$, the hadronic contribution to the running of the QED fine structure constant. The dotted curve represents the results obtained by including low-Q^2 data in the fit. Vertical yellow bands show the 95% C.L. excluded limits on the Higgs mass from the LEP direct searches and from the preliminary Tevatron combined results [9].

3.2 Higgs boson production cross section

Figure 3.9 shows the Higgs boson production cross section at the Tevatron energy, as a function of the Higgs mass. The Leading Order Feynman diagrams for the four most

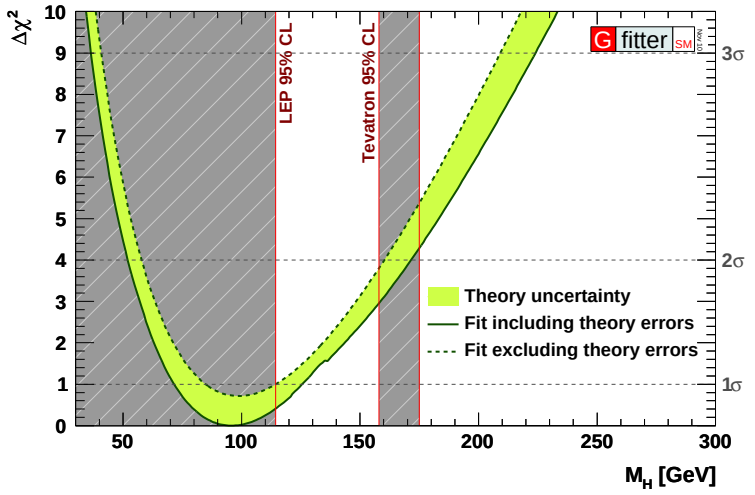

Figure 3.8: $\Delta\chi^2$=χ^2-χ^2_{min} of the global fit to the electroweak precision measurement performed by the GFITTER group [10]. The solid and dashed lines give the results when including or ignoring the theoretical errors.

relevant processes are reported in figure 3.10 and 3.11, while the corresponding cross section values are summarized in table 3.1 for several masses.

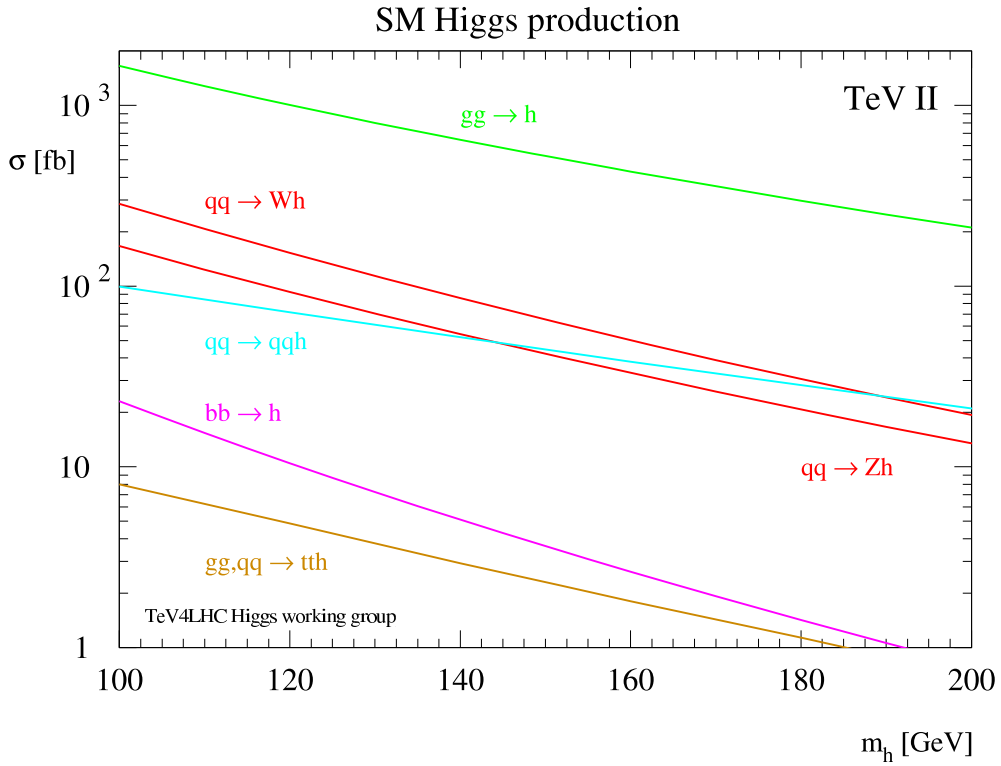

Figure 3.9: Standard Model Higgs boson production cross sections for $p\bar{p}$ collisions, at the Tevatron energy scale $\sqrt{S}$ =1.96 TeV, as a function of mass.

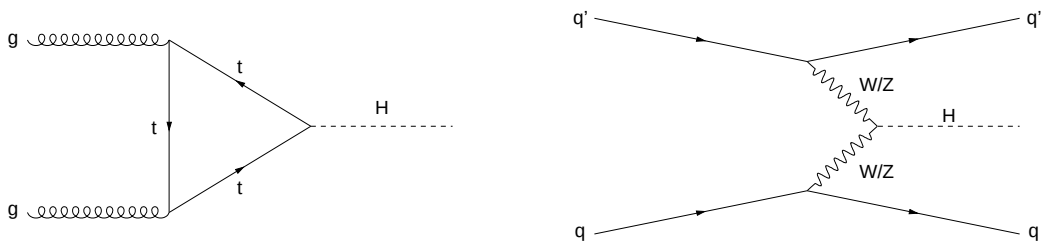

Figure 3.10: Higgs production processes at hadron colliders: gluon-gluon fusion (left) and vector boson fusion (right).

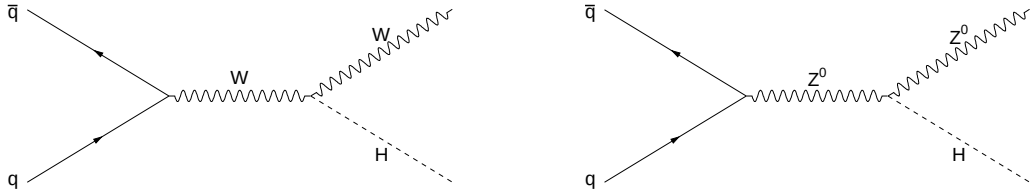

Figure 3.11: Higgs production processes at hadron colliders: associated production with a W boson (left) and associated production with a Z boson(right).

The dominant production mode is represented by the gluon-gluon fusion, gg→H, which proceeds mainly via a virtual top quark loop. A relevant contribution to the cross section is provided by the QCD radiative corrections, which are evaluated at the Next-to-Next-to-Leading Order (NNLO) approximation [29, 30]. Soft gluon resummation [31], two-loop electroweak corrections [32] and NLO effects of the b-quark in the loop [33] are also taken into account. The resulting inclusive theoretical cross sections, which is known with an uncertainty of about 10%, ranges from 1.8 to 0.2 pb for the 100 GeV/c^2 $< M_H <$ 200 GeV/c^2 interval. Larger theoretical uncertainties due to the renormalization scale and the gluon Parton Distribution Function (PDF) variations are evaluated and assigned separately to each jet multiplicity [34].

The associated production is an electroweak process which is characterized by a quark-antiquark annihilation into a virtual vector boson, that subsequently undergoes Higgs radiation (Higgsstrahlung), $q\bar{q} \rightarrow$V$^* \rightarrow$VH. Radiative corrections are calculated at the NNLO [35], with a total accuracy of less than 5%. Resulting cross sections are approximately one order of magnitude smaller than the gluon fusion one: it ranges from 0.3 to 0.02 pb for $q\bar{q} \rightarrow$WH and from 0.2 to 0.01 pb for $q\bar{q} \rightarrow$ZH in the 100 GeV/c^2 $< M_H <$ 200 GeV/c^2 interval.

The vector boson fusion takes place when the incoming quark and antiquark radiate vector bosons which annihilate into a Higgs: $q\bar{q} \rightarrow$ H$q\bar{q}$. This is a pure electroweak process and the QCD corrections are computed at the Next-To-Leading Order (NLO) with a 10% uncertainty, and the cross section evaluated with MCFM [36] ranges from 0.1 to 0.02 pb.

At the Large Hadron Collider (LHC), the current highest-energy particle accelerator in the world, the Higgs production cross section at the nominal center of mass energy $\sqrt{S}$= 14 TeV[1] is about two order of magnitude larger than at the Tevatron. However, relative contributions from the four main production mechanisms previously described change significantly: as shown in figure 3.12, direct production through gluon fusion is still the dominant process, while the production in association with a W or a Z boson becomes less relevant than at the Tevatron, being its contribution to the total cross section much smaller than that provided by the vector boson fusion process.

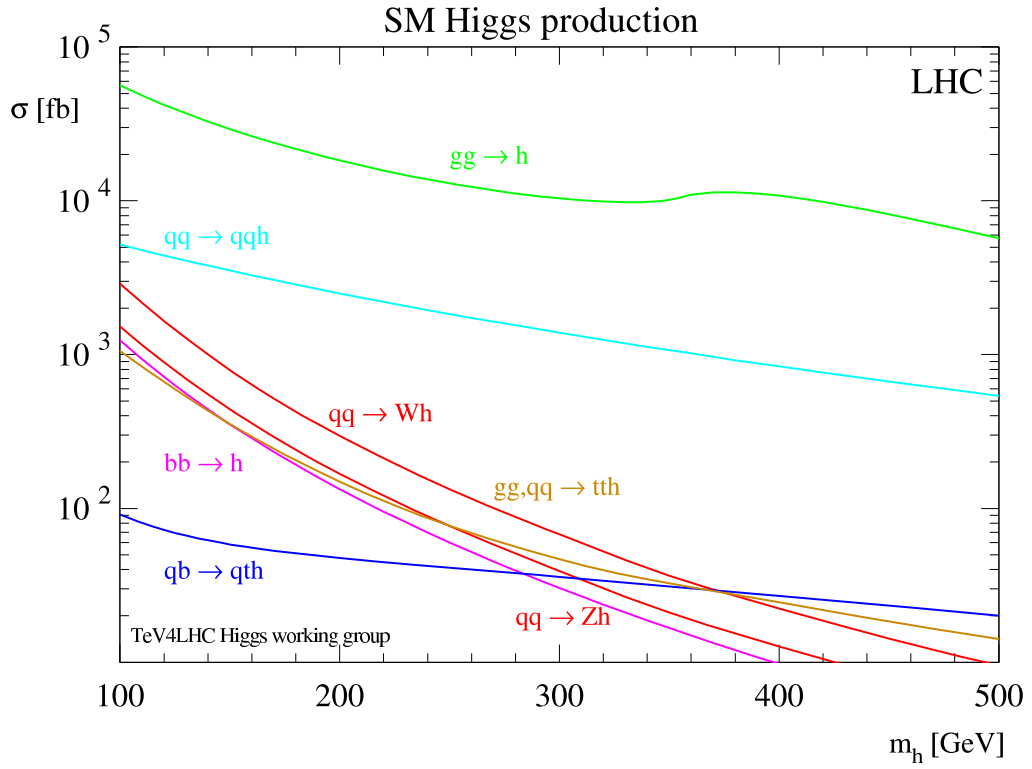

Figure 3.12: Standard Model Higgs boson production cross sections for *pp* collisions, at the LHC energy scale $\sqrt{S}$ =14 TeV, as a function of mass.

3.3 Higgs boson decay

Because the Higgs boson couples strongly to heavy particles, the main decay modes involve preferentially vector bosons and b-quarks; the probabilities of different possible decay channels are a strong function of the Higgs mass, as can be seen in figure 3.13 and 3.14, where the partial widths and the corresponding branching ratios for the most relevant decays processes, calculated at the NLO by using version 3.53 of HDECAY program [2], are shown.

[1] As of this writing, Autumn 2010, LHC beam energies are set to 3.5 TeV, thus providing $\sqrt{S}$= 7 TeV.

Higgs mass (GeV/c^2)	$\sigma_{gg\to H}$ (fb)	σ_{WH} (fb)	σ_{ZH} (fb)	σ_{VBF} (fb)
100	1861	291.9	169.8	99.5
105	2618	248.4	145.9	93.3
110	1413	212.0	125.7	87.1
115	1240	181.9	108.9	79.07
120	1093	156.4	94.4	71.65
125	967	135.1	82.3	67.37
130	858	116.9	71.9	62.5
135	764	101.5	63.0	57.65
140	682	88.3	55.3	52.59
145	611	77.0	48.7	49.15
150	548	67.3	42.9	45.67
155	492	58.9	37.9	42.19
160	439	50.8	33.1	38.59
165	389	44.6	30.0	36.09
170	349	40.2	26.6	33.58
175	314	35.6	23.7	31.11
180	283	31.4	21.1	28.57
185	255	28.2	18.9	26.81
190	231	25.1	17.0	24.88
195	210	22.4	15.3	23.00
200	192	20.0	13.7	21.19

Table 3.1: Summary of the Higgs production cross sections for the four most relevant processes at the Tevatron, in the mass range explored by CDF and DØ experiments: gluon fusion ($\sigma_{gg\to H}$), associated production with a W boson (σ_{WH}) or a Z boson (σ_{ZH}) and vector boson fusion (σ_{VBF}).

The two largest decay channels in the mass range explored by the Tevatron experiments are H$\to b\bar{b}$, which dominates for M_H <135 GeV/c^2, and H$\to W^+W^-$, whose contribution is prevalent for higher mass values. The branching fraction to W boson pairs reaches its peak at around $2m_W \approx 160$ GeV/c^2, when both bosons can be produced on-shell; consequently the total width, which was below 10 MeV in the low mass region, rapidly increases up to 1 GeV. In the mass range above $2m_Z$ also the decay into ZZ starts to be relevant, while the decay into top pairs is important only at very high Higgs masses, $m_H > 2m_{top}$, when this process becomes kinematically accessible (outside the mass range accessible at the Tevatron).

The Higgs decay into leptons and light quarks is suppressed by their low mass. However, the decay channel into a $\tau\tau$ pair represents a not negligible contribution in the low mass region, with a branching fraction of 5-8% in the 100 GeV/c^2 < M_H <

130 GeV/c^2 range.

Higgs mass (GeV/c^2)	B(H$\rightarrow b\bar{b}$) (%)	B(H$\rightarrow \tau^+\tau^-$) (%)	B(H$\rightarrow W^+W^-$) (%)
100	80.33	7.920	1.052
105	78.57	7.821	2.307
110	75.90	7.622	4.585
115	71.95	7.288	8.268
120	66.49	6.789	13.64
125	59.48	6.120	20.78
130	51.18	5.305	29.43
135	42.15	4.400	39.10
140	33.04	3.472	49.16
145	24.45	2.585	59.15
150	16.71	1.778	68.91
155	9.88	1.057	78.92
160	3.74	0.403	90.48
165	1.29	0.140	95.91
170	0.854	0.093	96.39
175	0.663	0.073	95.81
180	0.535	0.059	93.25
185	0.415	0.046	84.50
190	0.340	0.038	78.70
195	0.292	0.033	75.88
200	0.257	0.029	74.26

Table 3.2: Summary of the Higgs boson decay branching ratios for the three most relevant channels explored by the CDF and DØ experiments [2].

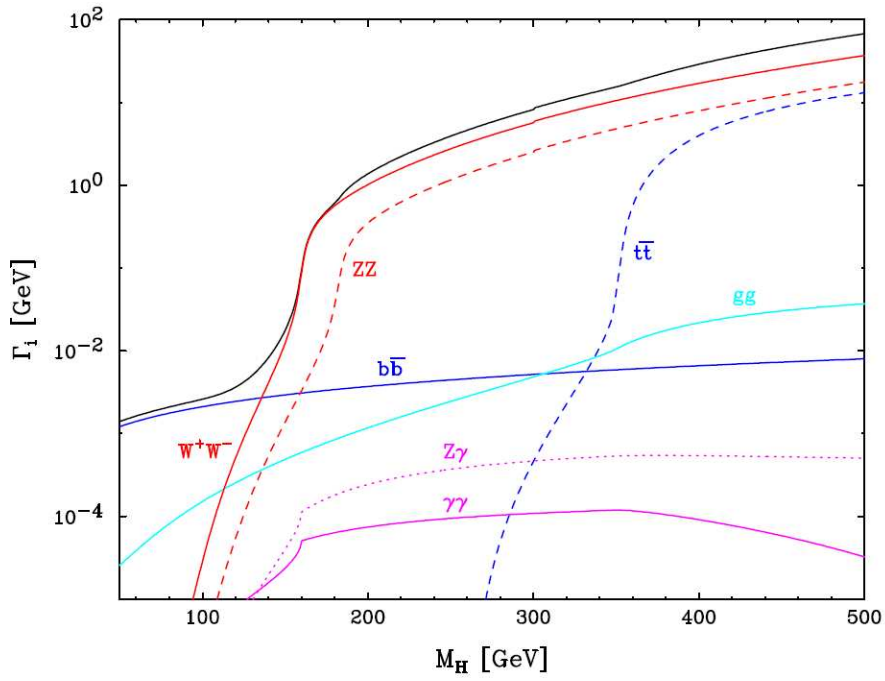

Figure 3.13: Partial widths of the Higgs boson decay channels in the SM, as a function of mass.

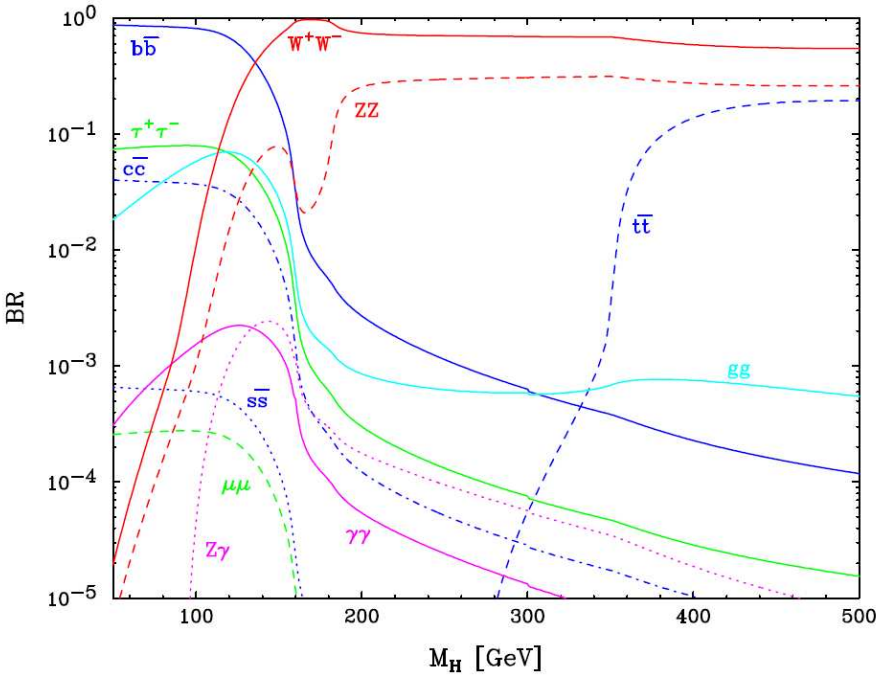

Figure 3.14: Decay branching ratios of the Higgs boson in the SM, as a function of mass.

Chapter 4

Experimental apparatus

"No amount of experimentation can ever prove me right;
a single experiment can prove me wrong"
Albert Einstein

In this work we investigate the data collected by the upgraded Collider Detector at Fermilab (CDF II) in head-on collisions between protons and antiprotons, at the center of mass energy of 1.96 TeV.

The CDF II detector is located at one of the two collision points on the Tevatron $p\bar{p}$ accelerator ring, hosted at the Fermi National Accelerator Laboratory (FNAL) in Batavia, about 50 Km west from Chicago (Illinois, USA).

A brief description of the accelerator complex and the detector is provided in this chapter.

4.1 The Fermilab accelerator complex

Fermilab's Tevatron [37] is a high energy hadron collider, which represents the last stage of a complex chain of particle (proton and antiproton)production, transfer, storage and acceleration up to the final energy of 980 GeV. Resulting final collisions, at the center of mass energy of 1.96 TeV, take place in two dedicated interaction points, where the CDF and DØ detectors are installed.

A schematic diagram of the Fermilab accelerator complex is shown in figure 4.1.

The Tevatron is an approximately circular synchrotron, 1 Km in radius, which employs eight radio-frequency (RF) accelerating cavities and a set of 772 dipole, 2 half-dipole and 204 quadrupole superconducting magnets, kept at a temperature of 4.3 K by a large cryogenic cooling system.

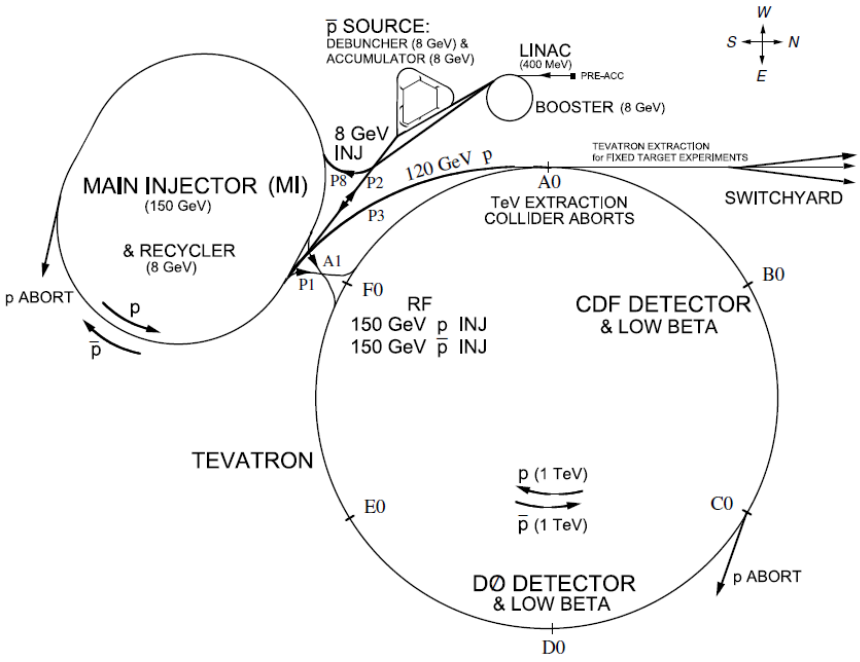

Figure 4.1: Schematic illustration of the Fermilab's accelerator complex.

In collider operation mode, 36 bunches of protons circulate clockwise, spaced by 396 ns each other, and collide against a similar beam of antiprotons which travel in the opposite direction in the same ring.

The performance of the Tevatron is evaluated in terms of two key parameters: the available center of mass energy, $\sqrt{S}$, which defines the accessible phase-space for the production of new particles in the collisions, and the instantaneous luminosity $\mathcal{L}$, which quantifies the number of interactions per unit time and thus represents the coefficient of proportionality between the rate of a specific process and its cross section.

$$rate[events\ s^{-1}] = \mathcal{L}[cm^{-2}s^{-1}] \times \sigma[cm^2] \tag{4.1}$$

The time integral of the previous formula is therefore a measure of the expected number of events n, for a given process of cross section σ, in a finite time T:

$$n(T) = \int_0^T \mathcal{L}\sigma dt \tag{4.2}$$

while the time integral of the luminosity, usually referred to as integrated luminosity, is:

$$\mathcal{L}_{int} = \int_0^T \mathcal{L} dt \tag{4.3}$$

Assuming an ideal head-on $p\bar{p}$ collision with no crossing angle between the beams, $\mathcal{L}$ is approximately given by the following expression:

$$\mathcal{L} = \frac{f N_B N_p N_{\bar{p}}}{2\pi(\sigma_{p^2} + \sigma_{\bar{p}^2})} F\left(\frac{\sigma_l}{\beta^*}\right) \tag{4.4}$$

where f is the revolution frequency, N_B is the number of bunches, $N_{p(\bar{p})}$ is the number of (anti)protons per bunch, $\sigma_{p(\bar{p})}$ is the r.m.s. (anti)proton beam width at the interaction point and F is a form factor which corrects for the beam shape and depends on the ratio of the longitudinal r.m.s. width σ_l of the bunch and the beta function β^*. The latter is a measure of the beam width and is proportional to the beam's x and y extent in phase space. The dominant limiting factor of the luminosity is the availability of monochromatic antiprotons that need to be efficiently produced and transferred through the accelerator chain for final collisions.

Parameter	Run II
Circumference[Km]	6.12
Collided particles	$p\bar{p}$
Injection beam energy[TeV]	0.15
Maximum beam energy[TeV]	0.980
Number of bunches N_B per species	36
Protons/bunch N_p	3×10^{11}
Antiprotons/bunch $N_{\bar{p}}$	10^{11}
Total antiprotons	3.5×10^{12}
beam width β^*[cm]	35
bunch length σ_l [cm]	57
bunch spacing [ns]	396
revolution frequency f [kHz]	47.713

Table 4.1: Run II configuration of accelerator's parameters.

Given the current parameters of the accelerator, as summarized in table B.1, Tevatron typically achieves initial instantaneous luminosities which usually exceeds 3.5×10^{32} cm^{-2} s^{-1}, with a current record of 4.024×10^{32} cm^{-2} s^{-1}, registered on April the 16^{th}, 2010.

Figure 4.2 shows the Tevatron peak luminosity for each store[1]: the blue markers represent the initial instantaneous luminosity, while the red points show the peak average value among each set of contiguous 20 stores. Figure 4.3 reports the weakly and total integrated luminosity delivered by Tevatron to date, as a function of time, since the beginning of Run II. As of December 2010, physics quality data corresponding to about 8.5 fb^{-1} have been stored.

Tevatron delivered integrated luminosity is also shown in figure 4.4, along with the value of good quality data recorded by the CDF detector. The average data taking

[1]Store: a continuous period of collider operation using the same collection of protons and antiprotons.

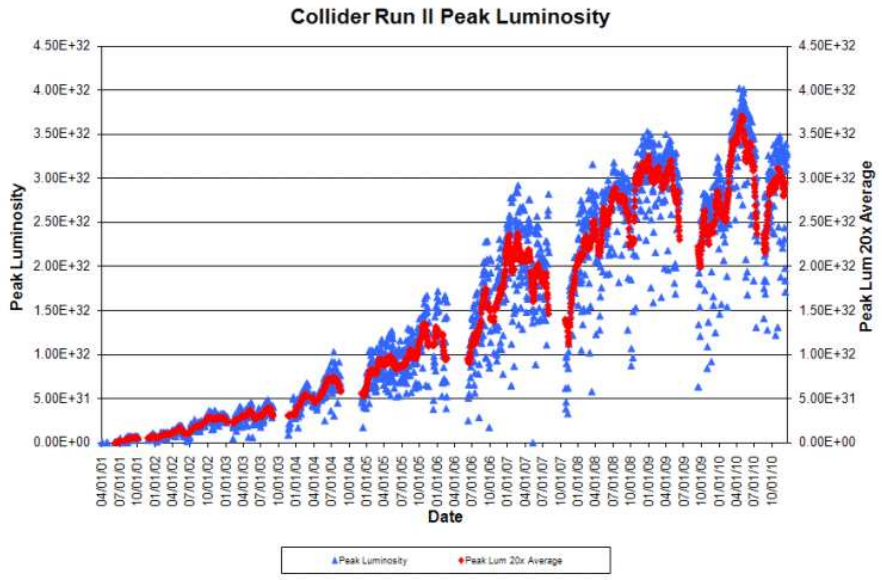

Figure 4.2: Tevatron Collider Run II peak luminosity, up to December 2010. Blue markers show the luminosity at the beginning of each store, while red markers represent the peak average value among a set of 20 stores.

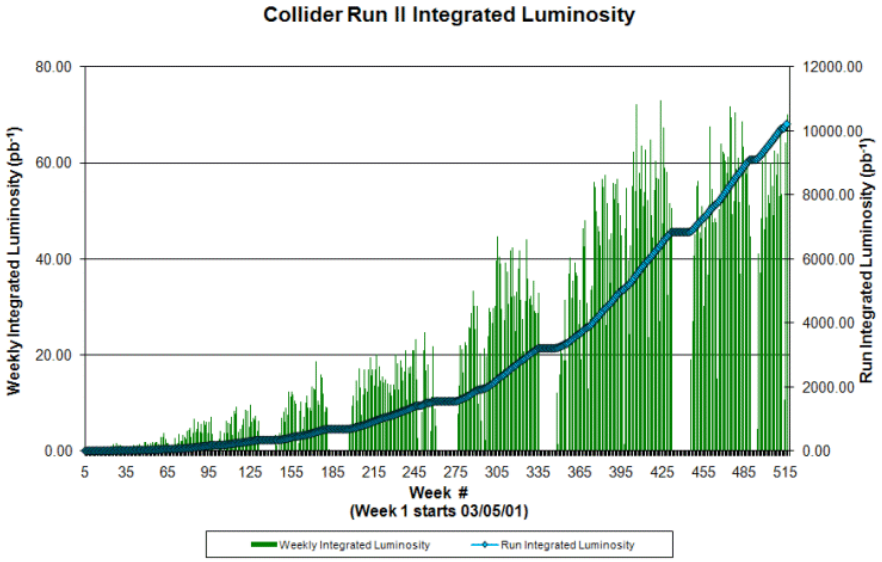

Figure 4.3: Tevatron Collider Run II weakly and total integrated luminosity, up to December 2010.

efficiency of the experiment is of about 85%, as can be seen in figure 4.5: a 5% of inefficiency arises at the beginning of the store, when the detector is not powered while

waiting for stable beam conditions, an additional 5% is due to trigger dead time, while the last 5% comes from unexpected detector and data acquisition problems.

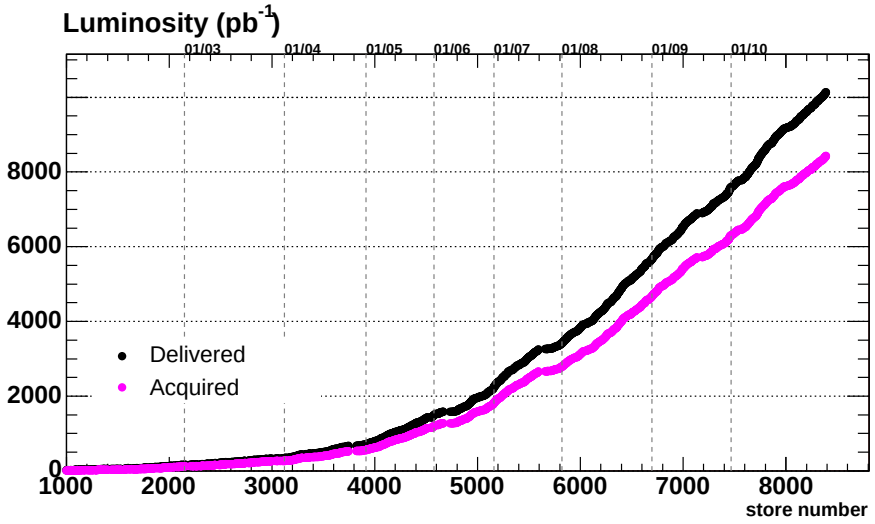

Figure 4.4: Total integrated luminosity delivered by the Tevatron collider and acquired by the CDF detector, as a function of the store number.

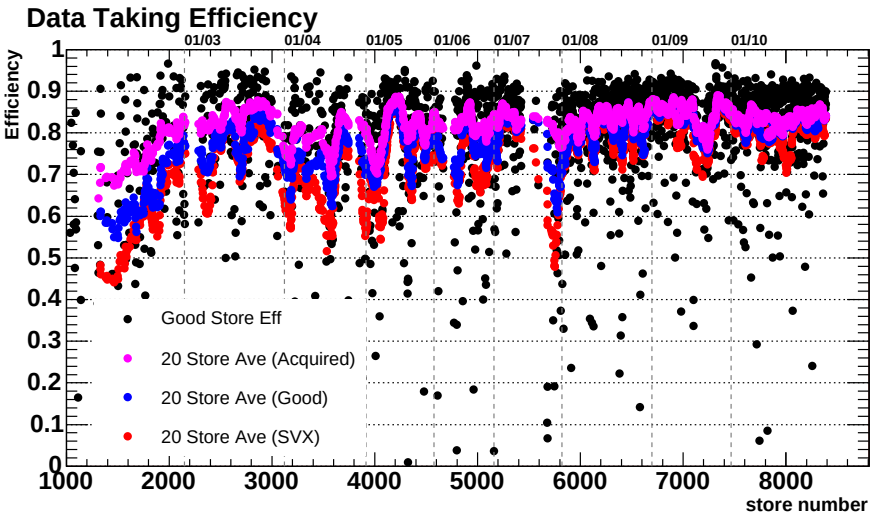

Figure 4.5: CDF Data taking efficiency, as a function of the store number.

In the following sections we describe the different steps of the particle production and acceleration chain at Fermilab, along with the procedure for obtaining proton-antiproton collisions in the Tevatron ring.

4.1.1 Proton production

The acceleration cycle starts with the production, in a Cockroft-Walton chamber, of hydrogen ions (H^-), which are accelerated by a positive voltage up to an energy of 750 KeV. The ions are then transported to a linear accelerator, the *Linac* [38], where they reach an energy of 400 MeV by passing through a 150 m long chain of radio-frequency (RF) cavities. The oscillating electric fields gradually accelerate the beam and group the ions into bunches.

The Booster [39] is an alternating gradient synchrotron (orbit radius of 75.5 m, 18 RF cavities) which takes the 400 MeV negative hydrogen ions and strips the electrons off by means of a carbon foil. A *batch* of bare protons is then accelerated up to an energy of 8 GeV and subsequently transferred into the Main Injector [40], divided into 84 bunches spaced by 18.9 ns, each of them consisting of about 6×10^{10} protons.

4.1.2 The Main Injector

The Main Injector [40] is a 3 Km circular synchrotron, with 18 accelerating RF cavities and conventional magnet, which is capable to accelerate particles to either 120 GeV or 150 GeV, depending on the operation status of the machine.

When the Main Injector's task is to provide beam to the antiproton production source (*accumulation mode*) one batch of 8 GeV protons is extracted from the Booster, accelerated up to 120 GeV and then focused, by means of quadrupole magnets, to a rotating 7 cm thick nickel target, which yields 8 GeV antiprotons.

When the Main Injector is used to inject the beams into the Tevatron for the last stage of acceleration (*collider mode*), bunches of alternatively protons and antiprotons are accelerated up to 150 GeV.

4.1.3 Antiproton production and accumulation

The choice of providing collisions of protons against antiprotons, instead of collisions between proton beams (as it is in the Large Hadron Collider at CERN), has several advantages: because of the opposite charges, antiprotons can be accelerated in the same ring of the protons, thus reducing the costs of a possible second ring and an additional set of magnets; what is more, the production cross section for a wide variety of physics processes is higher in $p\bar{p}$ collisions, for $\sqrt{s}$ up to 3 TeV, because of the dominating quark-antiquark interaction.

However, the production of antiprotons represents an extremely complex, expensive and time consuming task, and the number of antiprotons available for collisions is one of the most relevant factors affecting the luminosity, the quality and duration of the physics runs of the Tevatron [41].

Figure 4.6 shows the schematic layout of the antiproton production chain at the Tevatron. When a 120 GeV proton beam coming from the Main Injector strikes the nickel target, a particle spray with a broad momentum distribution is produced in the interaction. The beam of secondary particles is then collected and focused with

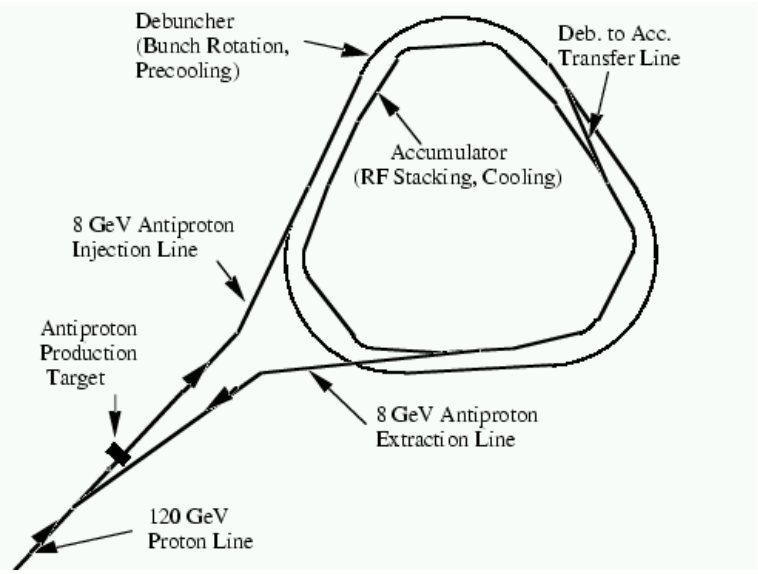

Figure 4.6: Layout of the antiproton production chain.

a cylindrical lithium lens into a beam line, where 8 GeV/c negative charged particles are selected by a 1.5 T pulsed dipole magnet. Particles with the wrong charge-to-mass ratio are removed and dumped to a graphite target.

The produced antiprotons, approximately 10-20 for each 10^6 protons on the target[2], are delivered to the Debuncher storage ring [41],a rounded triangular-shaped synchrotron with 90 m of mean radius, where the spread in the momentum distribution is reduced using the stochastic cooling [41] and the bunch rotation techniques.

From the Debuncher, antiprotons are transferred into the Accumulator [41], another triangular-shaped synchrotron, with 75 m of mean radius, located in the same tunnel of the Debuncher, where they are stacked and cooled up to the maximum intensity achievable.

From the Accumulator, 8 GeV antiprotons can be transferred either to the Main Injector or to the Recycler ring [42]. The latter is a 3.3 Km long ring, located in the Main Injector enclosure, which is designed to gather antiprotons as a final storage before the injection into the Tevatron, thus allowing the Accumulator to operate at its optimal efficiency.

It usually takes about 10 to 20 hours to collect a stack of 3.5×10^{12} antiprotons, to be used in Tevatron collisions.

4.1.4 Injections and collisions

When the antiproton storage reaches the desired amount, the machine starts the procedure for the beam injections into the Tevatron and the subsequent collisions.

[2]The resulting $\bar{p}$ production rate is expressed at Tevatron in mA per hour (typically 10-20 mA/h); a mA is a current measurement corresponding to 1×10^{10} antiprotons.

A set of 7 proton bunches is extracted from the Booster, transferred to the Main Injector, accelerated up to 150 GeV, coalesced[3] into a single bunch of $\approx 300 \times 10^9$ particles and then injected into the Tevatron. This process is repeated until 36 proton bunches are loaded. A set of electrostatic separators (about 30 pairs of metal plates) are then activated in order to create a pair of non interacting helicoidal closed orbits, separated by approximately 5 mm, in preparation for the antiproton injection.

Four set of antiproton bunches are then extracted from the Accumulator (or the Recycler ring), transferred to the Main Injector, accelerated, coalesced into 4 bunches of $\approx 90 \times 10^9$ particles and injected in the opposite direction, with respect to the protons, in the Tevatron. The antiproton load is repeated 4 times, until 36 bunches circulate in the ring.

The beams are finally accelerated from 150 GeV up to 980 GeV and bought into collisions at the two interaction points, where the DØ and CDF detectors are located.

Particle acceleration takes place through the RF buckets. A bucket is one interval of the longitudinal restoring force provided by the RF cavities that results in a stable phase-space where a bunch may be captured and accelerated. There are 1113 buckets along the Tevatron ring, and the 36 bunches of protons(antiprotons) are distributed among them in three equispaced "trains" of 12 bunches each. The inter-bunch spacing within each train is 396 ns, while an "abort" gap of 2.6 μs, corresponding to 139 buckets, is kept between the trains. The abort gap allows antiproton injection without perturbing the orbits of the already circulating protons. Furthermore, when beam abort is needed, it allows to ramping-up the deflecting magnets without interfering with the beam.

As a consequence of this not regular bunch distribution along the ring, the average bunch-crossing rate is 1.7 MHz, resulting from a 2.53 MHz when the proton and antiproton trains are crossing, and zero rate in correspondence of the abort gaps.

Near the collision points, the beams are focused by special quadrupole magnets (*low-β squeezers*) installed on the beam pipe just out of the detectors, in order to reduce the transverse spatial spread of the particles and then increase the luminosity. The transverse profile is additionally shaped by a set of retractable collimators (iron plates) which remove the halo coming from the tails of p and $\bar{p}$ distributions which interact with the beam pipe. When the beam conditions are stable and safe, the detectors are powered and start the data taking.

During collisions, the instantaneous luminosity decreases as a function of the time, because of the interaction of the beam with the residual molecules of gas that escaped the vacuum of the beam pipe, beam-halo interactions and $\bar{p}$ depletion due to collisions.

Meanwhile a new antiproton accumulation cycle has started and when antiproton stack is sufficiently large and the instantaneous luminosity becomes too small, the detectors are switched off, the store is dumped and the Tevatron is soon ready for a new beam injection.

It usually takes a couple of hours between the end of a store and the beginning

[3]Coalescing is the process of compacting several particle bunches into one bigger individual bunch.

of another one, during which time subdetector calibrations and cosmic rays tests are usually performed by both experiments.

4.2 The CDF II detector

The CDF detector [43] is a large multipurpose cylindrically symmetric spectrometer, surrounded by 4π fast projective calorimeter towers and fine-grained muon detectors. It is installed at the BØinteraction point of the Tevatron and it is designed to study 1.96 TeV $p\bar{p}$ collisions. The original facility, commissioned in 1985, has been subject to several upgrades during the time; the most extensive one started in 1995 and led to the current detector configuration whose operation, begun in 2001, is generally referred to as CDF II.

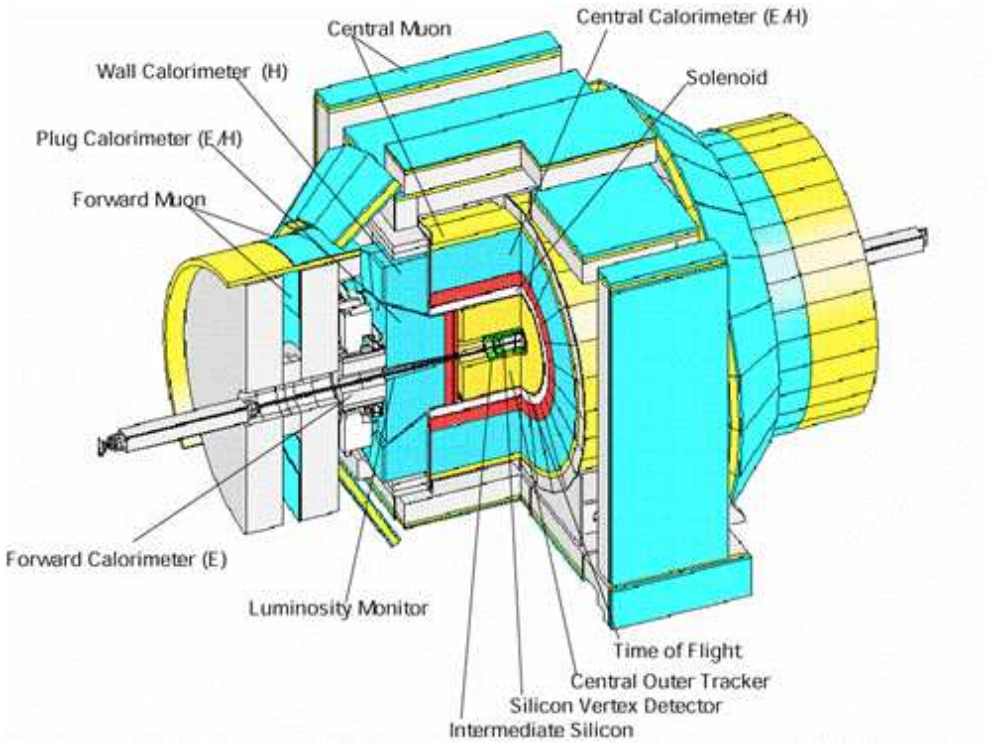

Figure 4.7: Isometric view of the CDF Run II detector.

CDF II detector, shown in a schematic view in figure 4.7, is composed of several specialized subsystems arranged in concentric layers, each one aimed at performing a specific task. Starting from the interaction point, particles generated from the $p\bar{p}$ collisions encounter in sequence:

- the thin beryllium wall of the beam vacuum pipe;
- a high-precision tracking system, that provides charged-particle trajectory reconstruction; it is composed by a inner set of silicon microstrip detectors and an

outer drift chamber, ;

- a time of flight detector, for the particle identification;
- a solenoidal magnet and its return steel joke;
- a set of sampling calorimeters segmented with a projective tower geometry, with the purpose of measuring the energy of charged and neutral particles;
- a set of drift chambers and scintillators counters, used to detect muons.

In this section we summarize the main features of each subdetector, by providing more details for the components relevant for the analysis presented in this thesis.

4.2.1 Coordinate system

CDF II employs a right-handed coordinate system, with the origin in the BØinteraction point, assumed coincident with the geometric center of the detector. The positive z-axis is defined along the direction of the proton beam. The (x,y) plane is then perpendicular to either beams, with y positive direction pointing vertically upward and positive x pointing radially outward with reference to the center of the Tevatron ring.

Cylindrical (r,ϕ,z) or polar(r,ϕ,θ) coordinate systems are also introduced, where r and ϕ define the transverse plane, with r indicating the radial distance from the origin and the azimuthal angle ϕ starting from the x axis and defined positive in the counter-clockwise direction. The polar angle θ is the angle relative to the z axis.

In high-energy collisions, where particles produced in the interaction are boosted along the beam line, it is useful to define a new variable, which is relativistically invariant under z boosts and can be used instead of the polar angle θ: the rapidity Y. It is defined as

$$Y = \frac{1}{2} \ln \left(\frac{E + p \cos\theta}{E - p \cos\theta} \right) \tag{4.5}$$

In the ultra-relativistic limit, when p$\gg$m, Y can be replaced with its approximated expression η, called pseudorapidity,

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \tag{4.6}$$

which is only function of the polar angle θ. It is particularly convenient to map solid angles in the detector in terms of (pseudo)rapidity and azimuthal angle, because the density of final-state particles, in high-energy collisions, is approximately flat in the (Y,ϕ) space. Other useful variables are the transverse component with respect to the beam axis of energy (E_T) and momentum (p_T) and the approximately Lorentz-invariant angular distance ΔR, defined as

$$E_T \equiv E \sin\theta \qquad p_T \equiv p \sin\theta \qquad \Delta R \equiv \sqrt{\eta^2 + \phi^2} \tag{4.7}$$

4.2.2 Tracking system

An integrated cylindrical system consisting of three silicon inner subdetector and an outer drift chamber, immersed in a B = 1.4 T solenoidal magnetic field parallel to the beam axis, provides three-dimensional charged particle tracking.

Within the magnetic field, particle trajectories are described by an helix, whose arc can be parametrized by using three transverse and two longitudinal parameters:

- **C** - defined as $C \equiv \frac{q}{2R}$, where R is the radius of the helix, it is the signed helix curvature, which is directly related to the transverse momentum of the particle, from the relation: $p_T = \frac{cB}{2C}$;
- $\boldsymbol{Z_0}$ - the z coordinate of closest approach to the z-axis;
- $\boldsymbol{\phi_0}$ - the ϕ direction of the particle at the point of closest approach to the z-axis;
- $\boldsymbol{d_0}$ - the signed impact parameter, defined as the distance of closes approach to the z-axis: $d_0 \equiv q\sqrt{x_c^2 + y_c^2} - R$, where ($x_c$,$y_c$) is the center of the helix, as seen from the (x,y) plane;
- $\boldsymbol{\lambda}$ - the helix pitch, $\lambda \equiv \cot\theta$, where θ is measured at the point of closest approach to the z-axis. It is directly related to the longitudinal component of the particle momentum $p_Z = p_T \cos\theta$.

The reconstruction of the trajectories and the determination of the parameters described above, is done by dedicated algorithms which perform helical fits of a set of spatial measurements coming from the tracking detectors. Magnetic field non-uniformities and effects of scattering in the material are opportunely taken into account.

A diagram of the CDF tracking volume in the (r,z) plane is shown in figure 4.8. Here we describe in more details each subsystem, starting from the one closer to the interaction point.

Layer 00 (L00)

L00 [44] is a single sided silicon microstrip detector directly mounted on the beam pipe, at an alternating radius of 1.35 cm or 1.62 cm. It is 87 cm long, provides a full coverage in ϕ and up to 4.0 in $|\eta|$. The strips are parallel to the beam axis, thus providing position measurements in (r,ϕ) plane, with a resolution up to 11 μm.

Silicon Vertex Detector (SVXII)

Located outside L00, SVXII [45] consists of five layers of double sided silicon microstrip detectors, extending from a radius of 2.45 cm to 10.6 cm, full covering up to $|\eta| \leq 2.0$. The system has a cylindrical geometry, coaxial with the beam, and it is segmented into three 32 cm *barrels*, each of them divided into twelve 30°azimuthal *wedges*, with a small overlap between them (see figure 4.9).

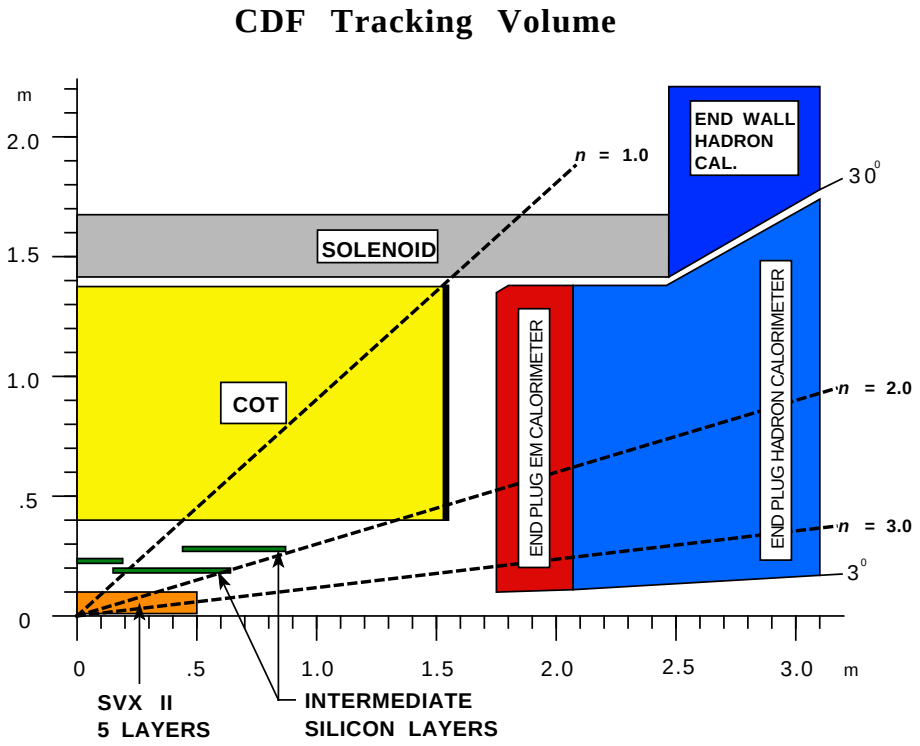

Figure 4.8: Elevation view of one quadrant of the inner portion of the CDF detector; the tracking volume inside the solenoid and the forward calorimeters are shown.

Strips are axially aligned to the beam on one side, while on the reverse side they have an orientation of 1.2°in layers 2 and 4, and of 90°in layers 0, 1 and 3. A precise reconstruction of r, ϕ and z coordinates is provided, with a position resolution of 9 μm.

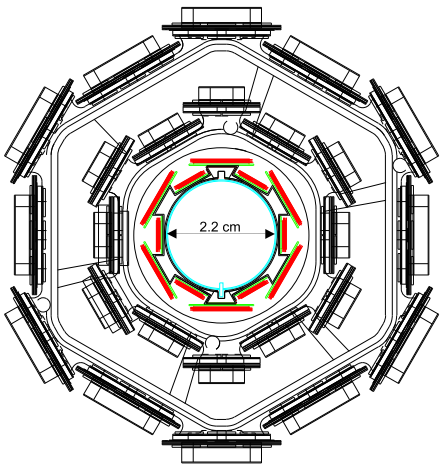

Figure 4.9: Schematic illustration of LOO and the first two layers of SVXII.

Intermediate Silicon Layer (ISL)

The ISL [46] is a silicon tracker made of three layers, positioned outside the SVXII at different radii and covering specific regions in η : one central layer ($|\eta| < 1.0$) is placed at r=22 cm, two forward and backward layers ($1< |\eta| <2$) are located at 20 cm and 28 cm respectively. Each of them is azimuthally divided into 30°sections, matching the SVXII structure, and consists of a double sided silicon microstrip detector, with axial strips on one side and stereo strips (with 1.2°angle) on the other one.

The impact parameter and z_0 resolutions of the SVX+ISL system are 40 μm and 70 μm, respectively.

Relative positions of the three silicon detectors is shown in details in figure 4.10.

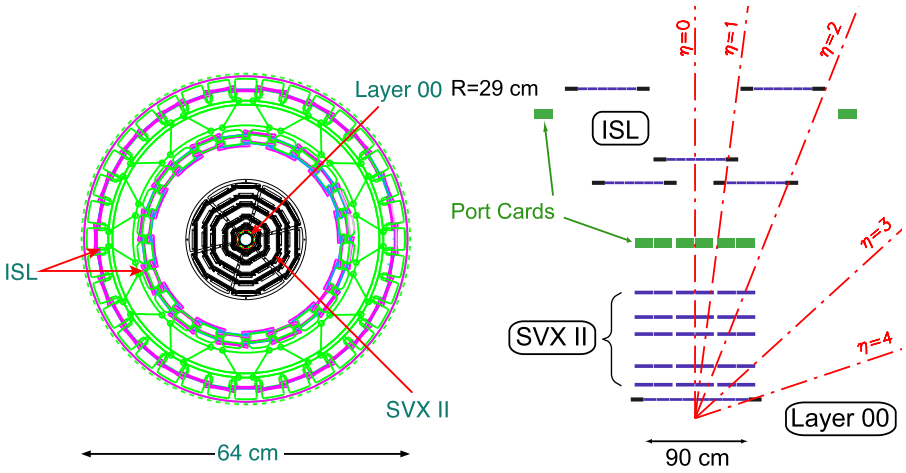

Figure 4.10: A radial view of the three silicon detector subsystems (left) and their coverage in the r-z plane (right).

Central Outer Tracker (COT)

It is a multi-wire, open cell drift chamber, which provides charged particle tracking in the central volume ($|\eta|$ <1.0), occupying the radial region from 44 to 132 cm, with $|z|$ <155 cm [47]. It consists of 96 sense wire layers, radially arranged in eight *super-layers*, each of them subdivided in ϕ into *supercells*, made of 12 sense wires plus a set of potential wires, shaper wires and field panels. Both sense and potential wires are 40 μm diameter gold plated tungsten. The field panel is 6.5 μm thick Mylar with a vapor deposited gold on both sides. A schematic view of one COT sector is shown in figure 4.11.

The entire COT contains 30240 sense wires; four superlayers have axial wire planes, while the remaining four superlayers have wires displaced with ±2°angles with respect to the beam line.

The COT is filled with an Argon-Ethane mixture, with a small contamination of Isopropyl alcohol (49.5 : 49.5 : 1.0), chosen to have a constant drift velocity. Hit posi-

tion resolution is $\approx 140\ \mu m$, while the momentum resolution of the reconstructed tracks is $0.15\% \times p_T[GeV/c]^{-1}$ which reduces to $0.07\% \times p_T[GeV/c]^{-1}$ if the COT information is combined to those coming from silicon detectors.

Tracking efficiency is greater than 98-99% for tracks with p_T >1 GeV/c.

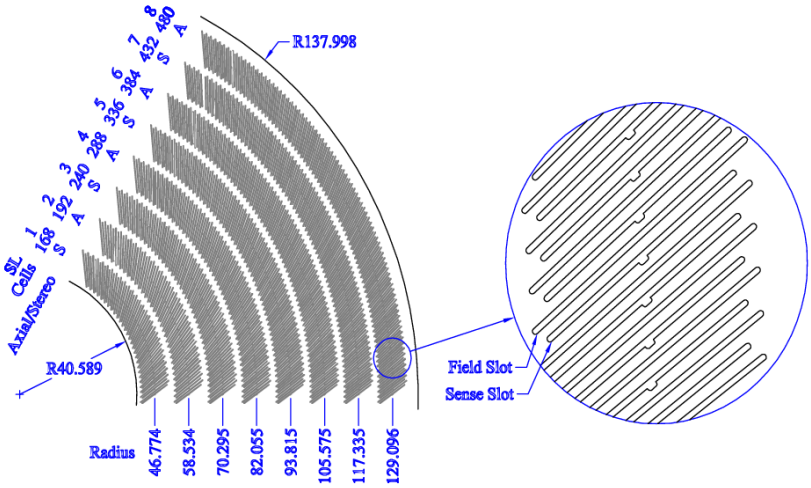

Figure 4.11: A COT section, showing the eight superlayers and a detail of the alternating field and wire planes.

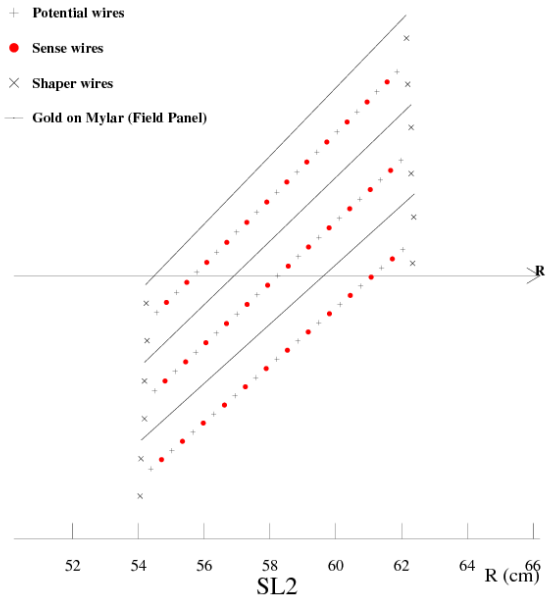

Figure 4.12: Three supercells in superlayer 2 looking along the beam z direction.

4.2.3 Time of flight detector

CDF II is equipped with a Time of Flight detector [48]: it is a cylindrical array of 216 scintillator bars, 280 cm long, positioned around the COT at an average radial distance of 138 cm. Each bar covers $|\eta|$ <1 in pseudorapidity and 1.7°in azimuthal angle.

Light pulses, collected by a set of PMTs displaced at each end of the bars, allow to measure the time of arrival of charged particles with respect to the bunch crossing time, with a resolution of about 100 ps. By combining the time-of-flight information with the momentum and path length measurement provided by the tracking system, it is possible to determine the mass of the particles. Pions and kaons can be discriminated with 2σ of separation for momenta p_T <1.6 GeV.

4.2.4 Calorimeters

Outside the solenoid coil, sampling calorimeters, consisting of alternated layers of passive material and plastic scintillators, cover the region $|\eta|$ <3.6 and are designed to measure the energy depositions of neutral and charged particles which escape from the tracking volume.

These calorimeters are organized in projective towers with a truncated pyramidal geometry, where the vertexes point to the center of the detector and the base is a rectangular cell in the (η,ϕ) space. Each tower, finely segmented in solid angle around the collision point and coarsely segmented radially outward, is divided into two independent compartments: the inner one is devoted to the detection of the electromagnetic component of showers, the outer one measures energy losses of hadrons.

The read-out system is composed of a set of wavelength shifting fibers and light guides, which collect the signal pulses generated in the scintillators and carry them to a couple of photomultiplier tubes (PMT) for each tower. PMT's current amplitudes are converted into measurement of energy, which for a given tower consists in the sum of the energy deposited in the electromagnetic and in the hadronic portions.

According to the geometry, the CDF calorimetric system is composed of two regions: central calorimeter covers the pseudorapidity range $|\eta|$ <1.1, is split in two halves at η=0 and is divided in Central ElectroMagnetic (CEM) and HAdronic (CHA) sectors. Plug calorimeter is placed in the 1.1< $|\eta|$ <3.6 region and consists of a Plug ElectroMagnetic (PEM) and a HAdronic (PHA) compartment. A supplementary endWall HAdronic calorimeter (WHA) is located behind the CEM/CHA system and above the plugs, as shown in figure 4.13, providing additional coverage in the region 0.9< $|\eta|$ <1.3.

The most relevant characteristics of each part of the calorimeter system are summarized in table 4.2.

The central calorimeter

The central calorimeter [49–51] has a radial extension ranging from 173 cm to 347 cm. Each of the three subsystems (CEM, CHA, WHA) is divided into four azimuthal arches (NE,NW,SE,SW) subtending 180°and organized in twelve 15°modules, called *wedges*.

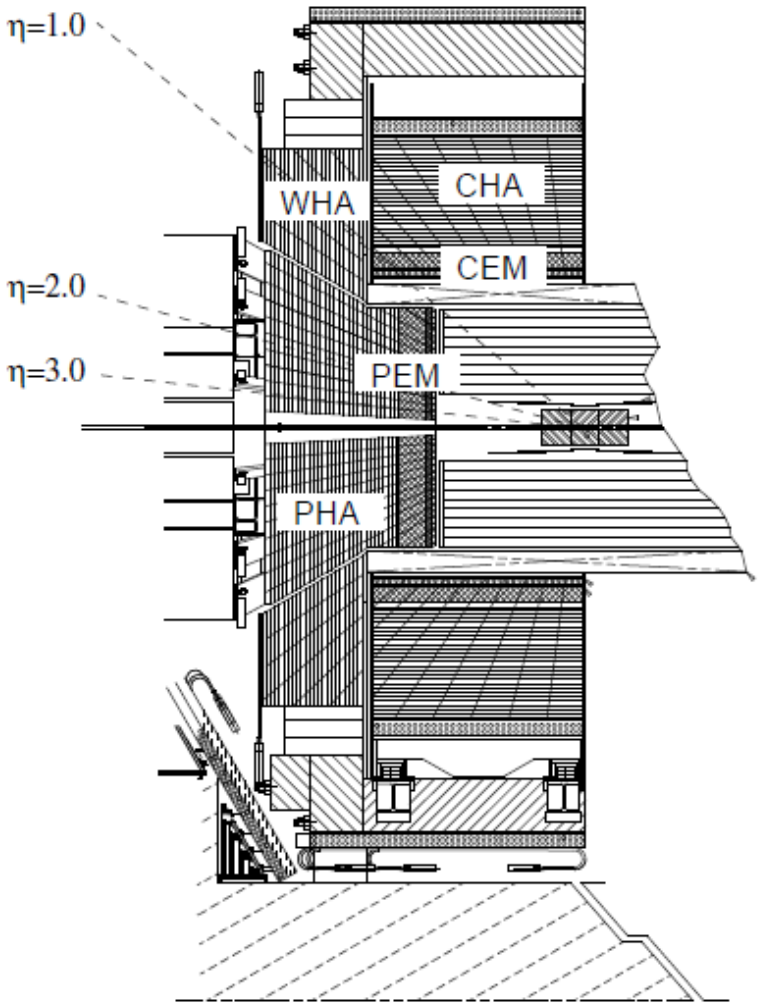

Figure 4.13: Elevation view of one half of CDF II detector, showing the layout of the different components of the calorimeter: CEM, CHA, WHA, PEM and PHA.

A CEM wedge consists of 31 layers of 5 mm thick polystyrene scintillator, alternated with 3.2 mm thick lead sheets, and is subdivided into 10 towers along pseudorapidity, with a resulting segmentation of 0.11 × 15°in the (η,ϕ) space. Two towers (*chimneys*) in one wedge are missing to allow solenoid access, so the total number of instrumented towers is 478.

The material thickness of the electromagnetic sector corresponds to 19 radiation lengths X_0[4] with an energy resolution:

[4]The radiation length represents the mean distance, expressed in units of g/cm^2, over which a high-energy electron loses all but 1/e of its energy by bremsstrahlung. It corresponds to about 7/9 of the mean free path for a high-energy photon before producing a e^+e^- pair.

$$\frac{\sigma_E}{E} = \frac{13.5\%}{\sqrt{E_T[GeV]}} \oplus 2\% \quad [CEM] \tag{4.8}$$

CHA and WHA detectors follow the same segmentation of CEM. CHA wedges are divided into 9 η towers matching in size and position the CEM towers; each of them is made of 32 layers of 2.5 cm thick steel absorber and 1.0 cm thick scintillator. WHA consists of 6 towers, 3 of them matching CHA; wedges are made of 15 layers of 5.1 cm thick steel absorbing material and 1 cm thick scintillator.

Both CHA and WHA have a total thickness of approximately 4.5 interaction lengths λ_{int} [5], with an energy resolution which is given respectively by

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E_T[GeV]}} \oplus 3\% \quad [CHA] \tag{4.9}$$

and

$$\frac{\sigma_E}{E} = \frac{75\%}{\sqrt{E_T[GeV]}} \oplus 4\% \quad [WHA] \tag{4.10}$$

Two additional position detectors are located within each CEM wedge.

At a radial depth of 5.9 X_0 (r≈184 cm), approximately corresponding to the region were the maximum energy deposition is expected for a high-p_T electron [49], a CEntral Strip multi-wire proportional chamber (CES) provides two dimensional information of shower position and profile, with a finer segmentation then the calorimeter [52, 53].

This results in an increased purity of electromagnetic objects reconstruction and in the capability of discriminating a single photon shower from that one generated by, for example, a neutral meson which decay into two photons. For the analysis described in this thesis, this feature helps in the reconstruction of taus decaying into hadronic systems which contain π^0s.

The Central Pre-Radiator (CPR) [54] is a further set of multi-wire chambers, placed immediately in front of the calorimeter, in the gap between the outer surface of the solenoid and the CEM. It monitors shower energy depositions started before the first layer of the calorimeter, helping in the discrimination of pions from electron and photons.

The plug calorimeter

Based on the same technology of the central components, the plug calorimeter [55–57], shown in picture 4.14, is composed of two identical devices, installed symmetrically on both sides of the detector, providing coverage to the forward regions ($1< |\eta| <3.6$).

The Plug ElectroMagnetic calorimeter (PEM) is composed of 23 lead plates, 2.77 m in outer diameter, each 45 mm thick, built with a "doughnut" structure around a

[5] The interaction length λ_{int} is the average distance a particle travels before interacting with a nucleus, expressed in units of g/cm^2. Here we refer to the interaction length of a charged pion.

central hole, where the beam pipe is located. Scintillator tiles 4 mm thick are placed between these layers of absorbing material, organized in 24 15°wedges for $|\eta|$ >2.1 and 48 7.5°wedges for $|\eta|$ <2.1.

Unlike the central calorimeter, the read-out system of each plug tower is made of a single PMT.

As in the central calorimeter, a preshower detector (Plug Pre-Radiator, PPR) consisting of a thicker (10 mm) scintillator, is installed in the first layer of the PEM, while a shower profile detector (the Plug showEr Maximum, PES) made of two tilted layers of scintillator strips, is located at a depth corresponding to 6 X_0. The total thickness of material corresponds to 21 X_0, with an energy resolution of

$$\frac{\sigma_E}{E} = \frac{16\%}{\sqrt{E_T[GeV]}} \oplus 1\% \quad [PEM] \tag{4.11}$$

The Plug HAdronic calorimeter (PHA) is divided into 12 wedges in ϕ, each subtending 30°and consisting of 23 layers of 5.08 cm thick iron absorbing material, alternated with 6 mm scintillator tiles. The total thickness corresponds to 7 λ_{int}, with an energy resolution given by

$$\frac{\sigma_E}{E} = \frac{74\%}{\sqrt{E_T[GeV]}} \oplus 4\% \quad [PHA] \tag{4.12}$$

	CEM	CHA	WHA	PEM	PHA
η coverage	<1.1	<0.9	0.7< $\|\eta\|$ < 1.3	1.3< $\|\eta\|$ <3.6	1.3< $\|\eta\|$ <3.6
Absorber	Pb	Fe	Fe	Pb	Fe
Thickness	$19X_0(1\lambda_{int})$	$4.5\lambda_{int}$	$4.5\lambda_{int}$	$21X_0(1\lambda_{int})$	$7\lambda_{int}$
Energy res.	$\frac{13.5\%}{\sqrt{E_T}} \oplus 2\%$	$\frac{50\%}{\sqrt{E_T}} \oplus 3\%$	$\frac{75\%}{\sqrt{E_T}} \oplus 4\%$	$\frac{16\%}{\sqrt{E_T}} \oplus 1\%$	$\frac{74\%}{\sqrt{E_T}} \oplus 4\%$

Table 4.2: CDF Run II calorimeter system summary table.

4.2.5 Muon detectors

At the GeV energy scale, muons generated in the $p\bar{p}$ collisions are minimum ionizing particles (MIP), which pass through the entire tracking system and calorimeter by depositing only a small fraction of their energy.

Four independent devices, designed to detect such muons which escape from the inner volume of the detector, are placed radially outside the calorimeters [58, 59].

They consist of a set of drift chambers and scintillating counters which cover specific (η,ϕ) regions up to $|\eta|$ <1.5, as shown in figure 4.15. Each chamber contains an array of drift tubes operating in proportional modem, with an Argon-Ethane mixture gas and a single sense wire each: absolute differences of drift arrival time between two tubes

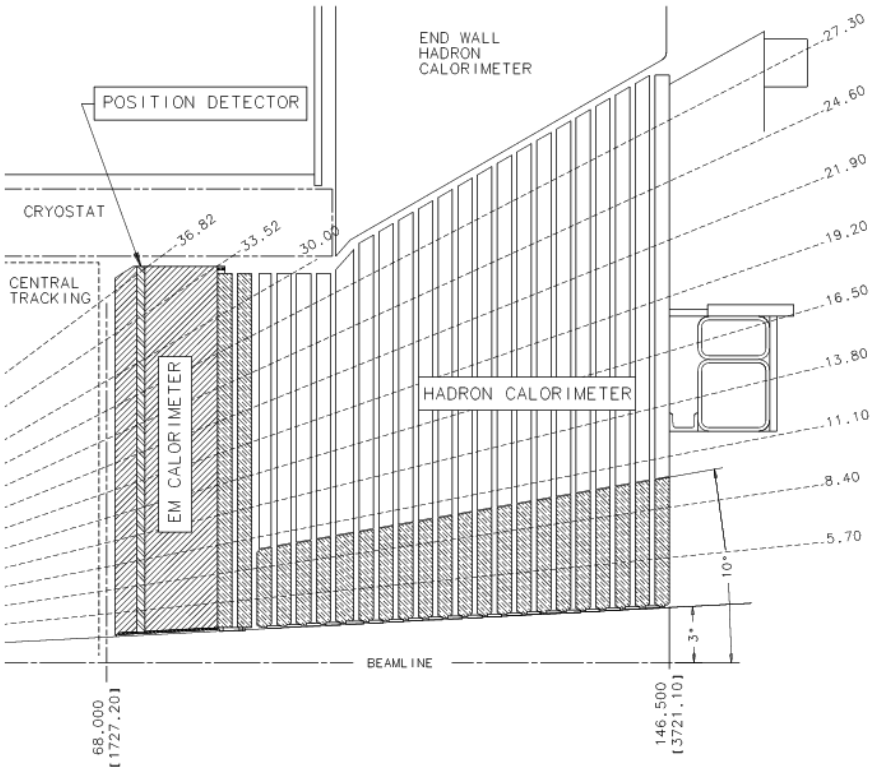

Figure 4.14: Elevation view of one quarter of the plug calorimeter.

provide a measurement of the azimuthal coordinate, while the charge division at each end of a wire can be used to determine the z coordinate.

Penetrating muons are identified by reconstructing small segments of their path (*stubs*) within the chamber, which are then matched to tracks extrapolated from the COT.

In the $|\eta|$ <0.6 region, outside the CHA at a radial distance of about 3.5 m, is located the Central MUon detector (CMU), consisting of four drift tube layers segmented in ϕ as the CHA (15°wedges). CMU is separated by $5.4\lambda_{int}$ of material from the interaction point; as a consequence, about 0.5% of high-energy hadrons will pass through the calorimeter and reach the muon detectors, thus creating a fake background. This contribution can be reduced by discarding CMU muon candidates which are not confirmed by hits in an outer detector, the Central Muon UPgrade, which lies in the same pseudorapidity region, is made of 4 layers of drift tubes and is separated from the CMU by 60 cm of absorber steel (more than 2 additional interaction lengths).

Timing information, with a resolution of 1-2 ns, is provided by a layer of scintillators (CSP) mounted on the outer surface of the CMP.

The muon coverage in the 0.6< $|\eta|$ <1.0 volume is granted by the Central Muon eXtension (CMX) conical section chambers, placed at a radius of about 3.5 m, divided in 15°wedges, each containing 8 layers of drift chambers. The total thickness from the

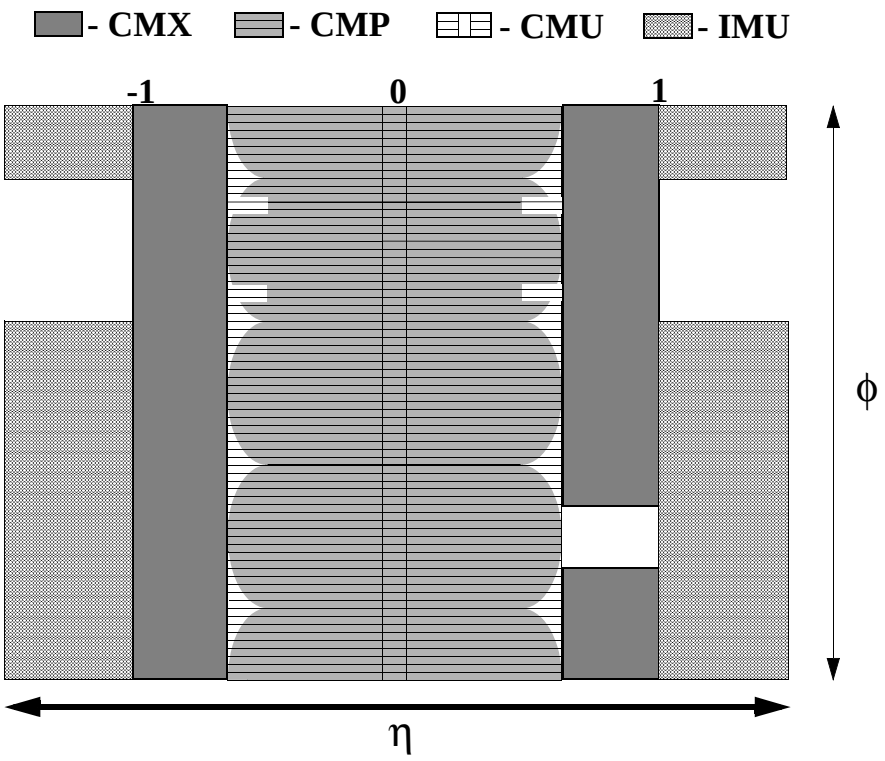

Figure 4.15: Coverage of the muon detector system in the η-ϕ plane.

interaction point corresponds to about $6.2\lambda_{int}$.

Two layers of scintillator (CSX) are mounted on the upper and on the lower surfaces of the CMX system.

The Intermediate MUon detector (IMU) consists of two barrels made of a set of 4 layers of proportional drift tubes (BMU) geometrically covering up to $|\eta|$ <1.5. A scintillator system, composed of two different devices (the Barrel Scintillator Upgrade, BSU, and the Toroidal Scintillator Upgrade, TSU) is mounted outside the BMU, providing timing information.

4.2.6 Luminosity measurement: CLC detector

The Tevatron instantaneous beam luminosity can be inferred from the average number $\bar{N}$ of inelastic interactions per bunch crossing, according to the formula

$$\bar{N} \times f_{b.c.} = \sigma_{p\bar{p}} \times \varepsilon \times \mathcal{L} \tag{4.13}$$

where $\sigma_{p\bar{p}}$=59.3±2.4 mb is the total $p\bar{p}$ inelastic cross section resulting from the averaged CDF and E811 luminosity measurements at $\sqrt{S}$=1.8 TeV, extrapolated to $\sqrt{S}$=1.96 TeV [60] $f_{b.c.}$ is the bunch crossing frequency known from the Tevatron RF and ε is the efficiency in the inelastic scattering detection.

The average number of interaction per bunch crossing is measured with a couple of Cerenkov Luminosity Counters (CLC) [61], installed on both sides of the detector along the beam pipe, inside the end plug calorimeter, as it is shown in figure 4.16.

They cover the 3.7< $|\eta|$ <4.7 regions and consist of 48 thin 110-180 cm long, conical, isobuthane filled Cerenkov modules (figure 4.17), arranged in 3 concentric layers and pointing toward the interaction point.

The base of each cone contains a mirror that collects the light into a PMT. Given a Cerenkov angle θ_C of 3.4°, momentum thresholds for light emission are 9.3 MeV/c for electrons and 2.6 GeV/c for charged pions.

$\bar{N}$ is measured with an uncertainty of 4.4%, which is combined to the 4% uncertainty on $\sigma_{p\bar{p}}$, thus leading to an instantaneous luminosity relative uncertainty of 5.9%.

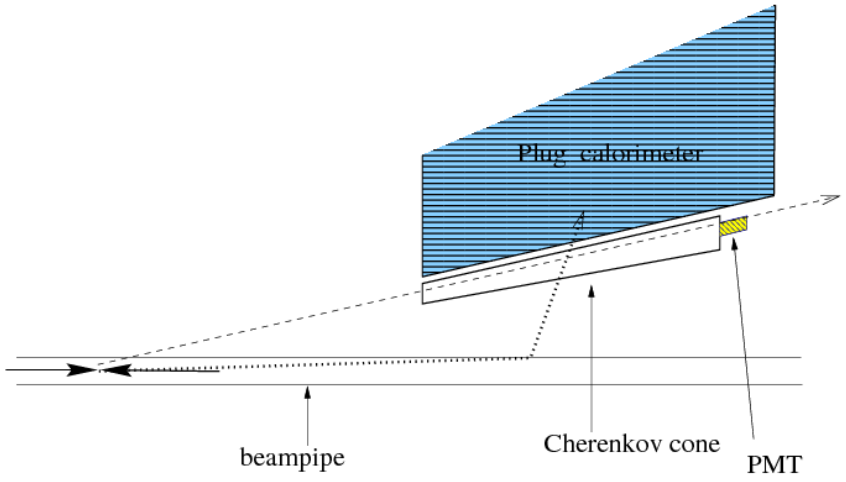

Figure 4.16: A view of one CLC cone module installed inside a quadrant of CDF.

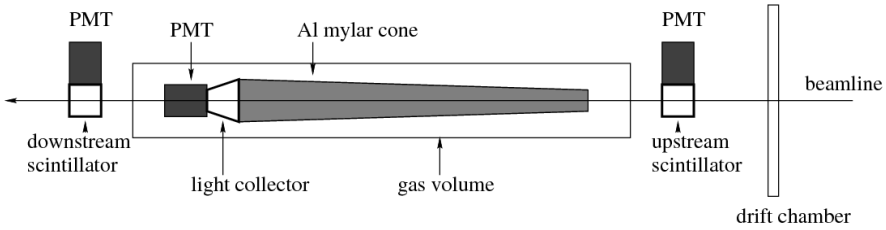

Figure 4.17: A schematic drawing of the CLC cone module.

4.2.7 The trigger system

With a predicted inelastic cross section for $p\bar{p}$ scattering of $\sigma_{p\bar{p}}$=59.3±2.4 mb, we obtain that at a typical Tevatron instantaneous luminosity of 10^{32} cm^{-2} s^{-1}, approximately 6 millions collisions per second occur, while the maximum rate that the CDF data acquisition system can handle is of about 100 Hz.

Furthermore, because the average size of the information associated to each event is 140 Kb, even in case of a detector capable of acquire and record all the events, an appropriate throughput and storage rate of 840 Gb/s would be needed, largely beyond any possible current available technology.

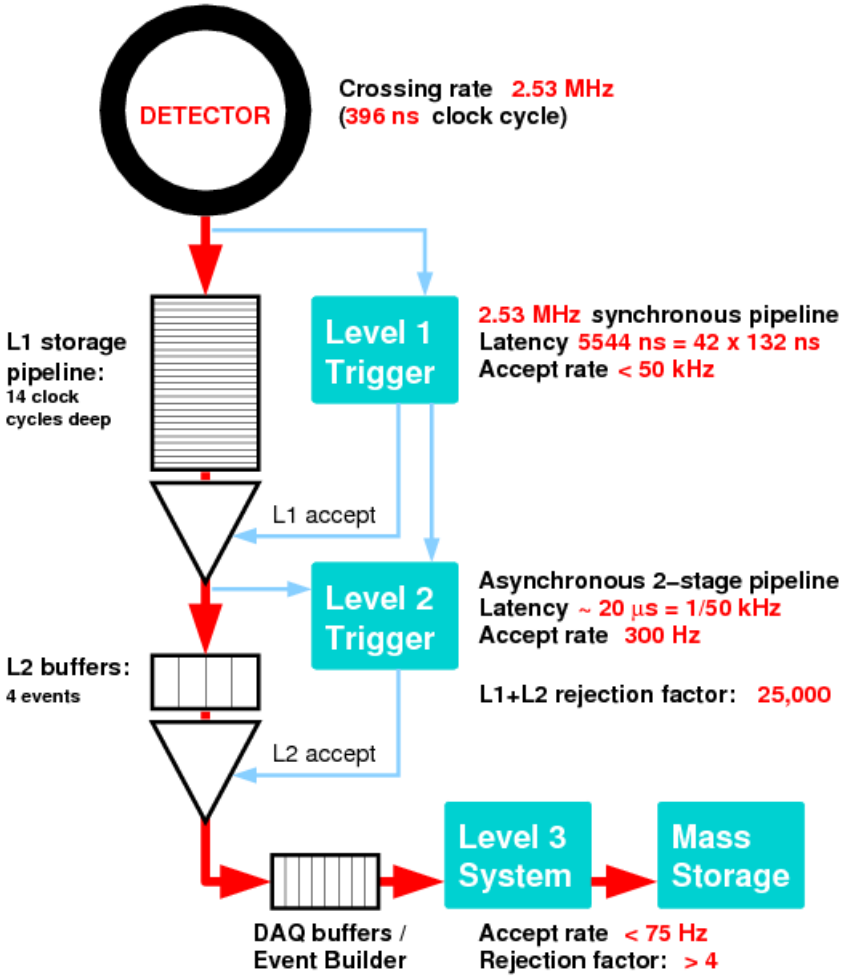

Figure 4.18: Diagram of the CDF trigger system data flow.

The trigger system allows to overcome both of these issues, by performing the task of selecting only the most interesting physics processes, which are permanently written to tape and subsequently made available for a fully offline reconstruction. This

online filtering process, which reduces the data flow of several orders of magnitude, is performed by evaluating the partial information provided by the different subdetectors.

The CDF trigger system is organized with a three-stage architecture, with a progressively greater sophistication of event reconstruction. At each level events are selectively accepted or discarded, with a reduction rate sufficient to allow data to be processed by the subsequent level with minimal dead time. This means that each event may be kept in the trigger memory for the decision without inhibiting the acquisition of the following event. A schematic view of the thee level trigger system data flow at CDF is shown in picture 4.18.

Level 1 is based on hardware electronics, L2 consists of a combination of custom hardware and commodity processors, while L3 is entirely implemented as a software in a PC farm with more than 500 dedicated CPU's, organized in a modular and parallelized structure.

The final decision is taken by applying a logical "or" among a collection of several programmable selection criteria, which are stored in a "trigger table": a specific combination of L1, L2 and L3 requirements univocally identifies a "trigger path".

Level 1

The front-end electronics of each subdetector, packaged in Versa Module Eurocard (VME) modules, hosted in about 120 crates, is interfaced to a 42 cell deep pipeline, synchronized with the Tevatron clock cycle (132 ns)[6]. Since the interbunch time is three times larger, the pipeline collects data corresponding to a maximum of 14 bunch crossings; each event is kept up to 132 ns$\times$42 $\approx$ 5.5μs for the decision, before the content of the buffer is definitely deleted. The L1 maximum accept rate is of about 30 kHz and is limited by the level 2 execution time.

The L1 hardware system consists of three parallel processing streams, which collect the basic raw information provided by three subdetectors (the calorimeters, the COT and the muon chambers), build low resolution physics objects, called "primitives", and feed inputs to the Global Level 1 Decision Unit, where the L1 trigger decision are taken.

- The four axial superlayers of the COT are used by the eXtremely Fast Tracker (XFT) [62], a custom processor that identifies two dimensional tracks in the (r,ϕ) plane. A pattern-matching algorithm searches for correspondences between the observed combination of hits in each superlayers and a set of predefined patterns, which represent a specific segment. If a coincidence between the segments crossing the four superlayers is found, the segments are linked to reconstruct a track. the algorithm returns the p_T and ϕ, by comparing each track to a set of about 2400 predetermined patterns, corresponding to all tracks with p_T >1.5 GeV/c originating from the beam line, the algorithm returns the p_T and ϕ.

[6]The CDF Run II detector trigger system was originally designed to operate with a Tevatron bunch spacing of 132 ns

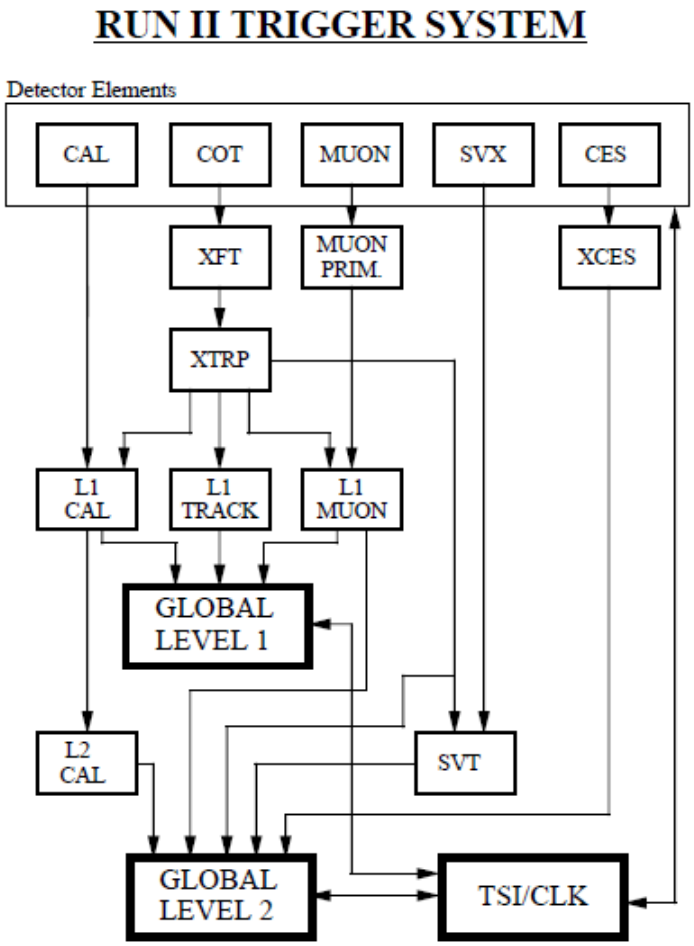

Figure 4.19: Block diagram of the CDF II Level 1 and Level 2 trigger system.

A recent upgrade to the XFT system [63], which was necessary to keep trigger rates under control at the high luminosity runs of the latest periods, allows to reject axial tracks which are not confirmed by the outer three COT stereo segments (*SLAM* [7] confirmation): fake tracks are thus reduced of about a factor of 7.

The resulting track list is then sent to the eXTRaPolation module (XTRP), a digital system which extrapolates each of them to the central calorimeter wedges and to the muon chambers (CMU and CMX), thus allowing to define electron and muon objects primitives.

- Calorimetric primitives are built by merging towers in pairs along η, dividing the calorimeter into 24×24 trigger towers in the (η,ϕ) space, with a transverse energy for each tower which is measured with a granularity of 500 MeV.

[7] The SLAM acronym refers to the Stereo Linker Association Modules, introduced in the XFT upgrade to replace the Linker Output Modules.

The information coming from each trigger tower is used to define two classes of primitives: *global primitives* are the sum of all transverse energy deposits above the threshold of 1 GeV in all trigger towers (ΣE_T) and the missing transverse energy $\not\!\!E_T$[8]; *object primitives* are electrons, photons and jets, defined by evaluating the hadronic and electromagnetic transverse energy relative contributions of the trigger towers and by exploiting the information coming from the extrapolated XFT track (i.e. for the discrimination of electrons from photons).

- Muon primitives are generated by matching the information coming from the muon detectors (single or coincidence hits in the scintillators, pattern of hits in the wire chambers) with the XFT tracks.

Level 2

The Level 2 is an asynchronous system with a latency time of 20 μs and a maximum accept rate of about 1 kHz.

Events accepted by Level 1 are transferred to a 4-cell level buffer, which is integrated in the front-end electronics of each subdetector, and queued for the Level 2 decision.

The buffering is organized as a two stage pipeline: in the first step signals from a particular section of the detector are analyzed; in the second stage the outcomes of step one are collected for the trigger decision. While data in the buffer are being processed, they cannot be overridden by a new event: if a Level 1 accept occurs when all four buffers are occupied, the system is subject to trigger dead time.

Additional information coming from the shower maximum strips chambers in the central calorimeter and the axial hits in the SVX II is combined with Level 1 primitives, to produce Level 2 primitives.

- A raw energy clustering is performed by an iterative algorithm which merges the energy of adjacent towers which exceed a predefined threshold, typically a few GeV. L2 clusters can be used to build objects to trigger on, by applying cuts on their total transverse energy, their position in the (η,ϕ) plane or their multiplicity.

  CDF has recently upgraded the L2 calorimeter trigger [64]: the new system increases the energy resolution up to a granularity of 125 MeV, by exploiting the full 10 bit trigger tower information[9]. It also implements a new fixed cone clustering algorithm, which turns out to be less affected by fake cluster formation. A cluster is defined by adding the energy of all towers which lie within a radius R = $\sqrt{\Delta\eta^2 + \Delta\phi^2}$= 0.7 around a seed tower.

- The XCES boards process the energy measurement registered by the shower maximum detectors, which provide a much better spacial resolution than the

[8] See section 6.4.5 for a full definition of the *met* variable

[9] Due to hardware limitations, the old system based on the clustering algorithm developed for Run I used only 8 bit energy information

calorimeter towers. By matching CES clusters with L1 tracks, electron and photon primitives are obtained, with a position known with an accuracy of 2°.

- The Silicon Vertex Tracker (SVT) combines the high precision information provided by the SVXII detector with the L1 XFT primitives, to form two-dimensional tracks with a reconstruction efficiency and a resolution close to the offline ones. In addition to the ϕ and p_T, also a measurement of the track impact parameter d_0 is performed.

  The SVT architecture is based on the pattern recognition technique: the algorithm receives in input the axial COT tracks found by XFT and the digitized pulse-heights coming from the $r \times \phi$ sides of the SVXII layers. As a first step, the Hit Finder collects the silicon hits and stores them into a buffer. Adjacent SVX detector channels are grouped into "superbins": candidate tracks, called "roads", are created by combining the XFT tracks with the superbins which contain hits from at least four different silicon layers. These roads are then compared to a set of pre-calculated combinations which are recorded in special designed memories, called Associative Memories (AM): when a matching is found, the hits belonging to that road are retrieved from the buffer and sent to the Track Fitter, where the tracks parameter are computed.

  The impact parameter is measured by SVT with a r.m.s. width $\sigma_{d_0} \approx 35$ μm, thus allowing to efficiently trigger on displaced tracks with respect to the beam spot.

- L2 muon primitives are characterized by an improved ϕ matching up to 1.25°, between the XFT tracks and the hits segments in the muon chambers.

  With the upgraded XFT system [65], a 3D track reconstruction can now be performed, with a good resolution on $\cot\theta$ ($\sigma_{\cot\theta} = 0.12$) and z_0 ($\sigma_{z_0} = 11$ cm).

Level 3

When an event is accepted by Level 2, all the raw detector output is sent via optical fibers to the Level 3, where it is fragmented and transferred to a Linux PC farm for a full reconstruction, performed by C++ based algorithms, similar to those used at offline level for the Physics analyses. These algorithms define high level objects and benefit from a full detector information and an improved resolution, including 3D track reconstruction and a tighter matching between tracks, calorimeter clusters and muon stubs.

Information related to the different subdetectors are then collected and re-assembled by the Event Builder into an "event record", a block of data which corresponds univocally to a bunch crossing.

Events satisfying Level 3 trigger requirements are finally transferred to the Consumer Server/Data Logger (CSL) system [66]. The CSL buffers the data into separate streams and store them on disk, records file meta information into a database and sends

on disk. The average processing time per event in L3 is of the order of the second and the admitted accept rate is of about 100 Hz. When A fraction of the output is monitored in real time to search for detector malfunctions, to derive calibrations constants and to graphically display events.

Trigger prescale

In high instantaneous luminosity conditions, the output rate of some trigger paths can easily exceed the maximum allowed values, thus leading to the occurrence of unwanted dead time; this effect has gained increasing relevance with the Tevatron performances improvements over the course of Run II.

The problem has been solved, without introducing tighter selection cuts which could create biases and inconsistencies in the data samples collected over different time periods, by applying a prescale to such trigger paths: this means that only a fraction of the events that satisfy the requirements at a specific trigger level are actually accepted and passed to the following level.

This feature is implemented in the first two stages of the CDF trigger system, in two different ways: a *static* prescale characterized by a fixed predefined prescale factor; a *dynamic* prescale, where the prescale factor is allowed to change during the data taking, according to the available bandwidth at each trigger level, thus achieving a real-time optimization of the online trigger rates.

4.2.8 Data processing and acquisition

Each time that at least one of the trigger paths fires, an *event* is labeled with a progressive number. Events are then grouped into *runs*, which represent periods of continuous data taking with the same configuration of the active subdetectors and of the trigger table. Different parameters, like the set of calibration values, the beam-line position and slope, are stored in the database.

All data manipulations which take place after the data are written to tapes are referred to as *off-line* operations. The most important of these operations is the *production* analysis, which consists in the unpacking of low-level raw data and in a complete reconstruction of the collected event information: high-level objects suitable for analysis, such as tracks, vertexes, leptons and jets are generated, in a similar way as it is done at Level 3, but with more precise information about the detector and beam-line conditions, and by applying more sophisticated reconstruction algorithms [67].

The output of the production is then organized into datasets, which are made available for the physics analyses.

In the production stage, several run numbers are grouped into *run periods*, each one corresponding to an integrated luminosity of a few hundreds of pb^{-1}.

Chapter 5

Higgs boson search strategies at the Tevatron

Given the direct and indirect constraints to the Higgs mass discussed in chapter 3, the Tevatron and the LHC hadron colliders currently represent the only accelerators in the world where the Higgs boson, as predicted by the Standard Model, can be produced.

The different center of mass energies at which the interactions take place (1.96 TeV at Tevatron, 14 TeV at LHC), as well as the kind of particles which participate to the collisions ($p\bar{p}$ at Tevatron, pp at LHC), lead to significant differences in the search strategies adopted in the two laboratories: according to the relative contributions of each production mechanism and to the specific background environments, some search channels can be preferred to others, because of their higher potential discovery sensitivity.

In this chapter we present an overview of the most promising channels which are studied in the Tevatron experiments, CDF and DØ, covering the entire mass range allowed by the theoretical and experimental constraints. In particular, we focus on the analyses carried out in the low mass hypothesis ($m_H \leq 135$ GeV/c^2) and we introduce in more details the H$\rightarrow \tau\tau$ search, which represents the main topic of this thesis.

5.1 High and low mass region

The search strategies for the Higgs boson are driven at the Tevatron by the decay modes, which are shown in figure 3.14, as a function of the mass. In addition, because the W and Z vector bosons are largely involved in many Higgs mechanisms and background processes, a good understanding of their decay channels and production cross sections is crucial. We report in table 5.1 the branching ratios for the dominant W and Z decay channels.

W^+ decay	B.R.(%)	Z^0 decay	B.R.(%)
$e^+\nu_e$	10.75	e^+e^-	3.36
$\mu^+\nu_\mu$	10.57	$\mu^+\mu^-$	3.37
$\tau^+\nu_\tau$	11.25	$\tau^+\tau^-$	3.37
hadrons	67.70	hadrons	69.91
		$\nu_l\bar{\nu_l}$	20.00

Table 5.1: Branching ratios for the W and Z bosons decays [1].

It is particularly useful to split the considered Higgs mass range into two different intervals, commonly denoted as *low mass* and *high mass* regions.

For $m_H \leq 135$ GeV/c^2, the dominant decay is into a $b\bar{b}$ pair. The process gg$\rightarrow$H$\rightarrow$ $b\bar{b}$, although the most abundant, is experimentally prohibitive due to the overwhelming QCD multijet background, which is estimated to be more than 10^6 times larger than the predicted signal yield. The cross section for the Higgs production in association with a vector boson is approximately one order of magnitude smaller, but the experimental signature is cleaner because of the leptonic decay of the W and Z, which provides a good handle to reduce the background contribution.

For $m_H \geq 135$ GeV/c^2, the decay process H$\rightarrow WW^*$ represents the primary mode for the Higgs boson search and both the direct and the associated production modes, as well as the vector boson fusion, are explored.

However, due to the extremely low signal to background ratio, no single channel analysis at the Tevatron can reach the sensitivity to the SM Higgs production cross section. The best performances can only be achieved by optimizing each search in every channel separately and by combining the results of both CDF and DØ experiments into one single measurement.

5.1.1 High mass Higgs boson searches

Even if the W boson decays into hadrons with a branching fraction of about 68%, the H$\rightarrow WW^*$ final state involving only hadronic jets is not considered, because of the large multijet background. Candidate events are thus selected by requiring that at least one of the two W's decays leptonically. Several categories are then considered and analyzed individually, according to the number of identified leptons, their charge, the jet multiplicity and the dilepton reconstructed invariant mass.

- H$\rightarrow WW^* \rightarrow l^+\nu l^-\bar{\nu}$: the fully leptonic channel represents the most promising for the Higgs discovery, given the clean signature provided by the two reconstructed high p_T leptons. The search is organized by splitting the final state into five independent categories: opposite charged leptons with either zero, one or more calorimeter jets; events with low two lepton invariant mass; three lepton events with no same-flavor opposite charged leptons inside the Z mass peak[1].

Even if the invariant mass of the Higgs boson cannot be reconstructed because of the presence of the undetected neutrinos, the sensitivity of this channels is nevertheless significant. The property of spin correlation among the decay products of the Higgs is exploited in the analysis, by looking at the angular distribution of the observed leptons in the final states. The Higgs is a spin-0 scalar particle, which decays into two spin-1 vector bosons, whose spins have to be anti-aligned in the Higgs rest frame: as a result of the couplings to the W's, leptons are emitted preferentially in the same direction. The resulting small separation angle differs significantly from that of the dominant background, represented by the non-resonant electroweak WW production.

- H$\rightarrow WW^* \rightarrow l\nu q\bar{q}'$: when only one of the two W decays leptonically, the final state is characterized by one high p_T lepton, two jets and a significant amount of missing energy due to the undetected neutrino[1].

  This channel is studied by looking at the associated production, where one or two additional leptons coming from the W or Z decay are identified:

$$WH \rightarrow WWW^* \rightarrow l\nu l\nu q\bar{q}' \tag{5.1}$$

$$ZH \rightarrow ZWW^* \rightarrow lll\nu q\bar{q}' \tag{5.2}$$

  When only two leptons are selected, the background is heavily suppressed by requiring that the charge sign of the two leptons is the same. In this case the major contributions to the background originate from events with either an erroneously reconstructed lepton charge or a jet faking the lepton signature.

  A second category which is considered is when three leptons are identified and two of them have the same flavor (ee or $\mu\mu$), opposite charge signs and their invariant mass is consistent with that of the Z boson. The signal contribution arises predominantly from the ZH production channel, while the dominant background is given by the WZ electroweak production.

The latest published combined results of the CDF and DØ analyses on the high mass region cover about 4.8 fb^{-1} and 5.4 fb^{-1} of data, respectively [68][2].

.
.

5.1.2 Low mass Higgs boson searches

At masses below 135 GeV/c^2 the search for the SM Higgs production is performed by considering the following different channels:

[1]For the detailed definition of calorimeter jets, missing energy and lepton variables, see section 6.4

[2]More recent results, covering higher integrated luminosities, have not been published yet, but they enter in the Tevatron combination which will be discussed in chapter 9

- WH$\rightarrow l\nu b\bar{b}$: the W decays leptonically, while the Higgs goes into a $b\bar{b}$ pair. The signal events are characterized by a high p_T electron or muon (the extension to the tau category is under study), large $\not\!\!E_T$ and two high E_T calorimeter jets identified in the central region of the detector.

  Background reduction is provided by the implementation of advanced algorithms and multivariate techniques, such as matrix element methods or artificial neural networks, which are used to separate the signal from the principal sources of background and to discriminate b-quark jets from jets originated from gluons or light quarks.

  The published CDF and DØ results on this channel are based on 2.7 fb^{-1} [69] and 1.0 fb^{-1} [70] of data, respectively.

- ZH$\rightarrow l^+l^-b\bar{b}$: the Z boson is identified through the decay into pairs of opposite charged electrons or muons and the reconstruction of the dilepton invariant mass, which has to be consistent with the Z mass. Events are then selected by requiring two additional central energetic jets.

  Neural networks (CDF, 1.0 fb^{-1} [71]) and event probability distributions based on matrix elements (CDF, 2.7 fb^{-1} [72]) are implemented to discriminate the signal events from the main background, represented by Z+jets, top pairs and ZZ. DØ's measurement refers to the first 4.2 fb^{-1} of data [73].

- ZH$\rightarrow \nu\bar{\nu}b\bar{b}$: due to the Z$\rightarrow \nu\nu$ decay, the final state is characterized by the presence of a large missing energy and two b-jets.

  The main source of background, represented by multijet events, is reduced by applying a neural network technique. The sensitivity of this channel is further increased by including also events from the WH production channel, in which the lepton coming from the W decay fails the identification requirements.

  The most recent published results cover 2.1 fb^{-1} and 5.2 fb^{-1} of data, collected by the CDF [74] and DØ [75] experiments.

- WH$\rightarrow qq'b\bar{b}$/ZH$\rightarrow q\bar{q}b\bar{b}$: when the vector bosons are searched in the hadronic decay, the final state is characterized by four jets final state, two of which are required to be generated from b-quarks. Signal events are separated from the background by defining a log-likelihood ratio function, which is based on the matrix element method.

  CDF published this search covering 2.0 fb^{-1} of data [76].

- H$\rightarrow \gamma\gamma$: the branching ratio of this decay mode is extremely small but the sensitivity is good, due to the much better energy resolution of the identified photons, with respect to the b-jets. The narrow invariant mass peak $M_{\gamma\gamma}$ can be easily exploited to reduce background, mainly represented by SM diphoton or photon+jet events. Event if the results in this decay channel are not competitive

with H$\rightarrow b\bar{b}$ searches, the analysis gives a non negligible contribution to the combined sensitivity. DØ's published analysis refers to 2.7 fb^{-1} of collected data [77].

- H$\rightarrow \tau\tau$: this decay mode allows us to simultaneously explore all the four different Higgs production channels. The main source of background, represented by QCD multijet events, is reduced by looking for the leptonic decay of one of the two taus, while the request of jets in the final state maximizes the sensitivity to the vector boson fusion and to the associated production. This decay channel is the subject of this thesis so a more detailed description is given in the next section.

5.1.3 Higgs boson decaying to two τ leptons

The Tevatron experiments have recently reached the remarkable result of the SM Higgs boson exclusion at 95% C.L. in the high mass region and are intensively working to extend the sensitivity to lower masses as well(more details in chapter 9): both CDF and DØ have focused their efforts to the Higgs decaying to b quarks, but many difficulties on this search come from the low signal yield, because the direct production is not accessible, as discussed in the previous sections. From the experimental point of view, the low mass search is much harder because of the not trivial estimation and modeling of $b\bar{b}$ background processes and the poor resolution of the jet energy measurements compared to that of the leptons.

For this reasons, any search aimed at including other Higgs decay channels may help to improve the overall sensitivity. In this context, the Higgs going to two τ leptons can play a crucial rule: though the branching ratio is small (<10%), the direct production from gluon and vector boson fusion become feasible, when the leptonic decay of one of the two τ is considered. As a matter of fact, given the clean and efficient detection of electrons and muons provided by the CDF and DØ detectors, the multijet background is reasonably kept under control. Furthermore, in the associated production channels a significant fraction of the signal acceptance can be recovered, combined to the other searches, by considering also the hadronic decay of the W and Z bosons.

The additional request of at least one calorimeter jet increases the signal over background ration and potentially improves the search sensitivity. For the vector boson fusion, which Leading Order (LO) Feynman diagram is shown in figure 5.2, two jets arise naturally from the hadronization of the outgoing quarks; in the case of the associated production, one or more jets can directly come from the hadronic decays of the W and Z bosons or from gluons emitted by the incoming quarks, as shown in figures 5.3 and 5.4; in the gluon fusion production mechanism no jets are expected at the tree level, however in a sizable fraction of events they can be generated from initial state radiation, as can be seen in the Feynman diagram reported in figure 5.1.

To summarize, the following signal processes are considered, where the two reconstructed leptons in the final state are explicitly required to be primary or secondary decay products of the Higgs boson:

- gg→H→ $\tau^+\tau^-$
- qq'→Hqq'→ $\tau^+\tau^-$qq'
- qq'→WH→qq'$\tau^+\tau^-$
- qq'→WH→ l$\nu\tau^+\tau^-$
- qq'→ZH→qq'$\tau^+\tau^-$
- qq'→ZH→ l$^+$l$^-\tau^+\tau^-$
- qq'→WH→ $\nu\nu\tau^+\tau^-$

In this thesis, events are selected by requiring that one of the two taus decays leptonically, while the other one is reconstructed by its hadronic decay products. In order to have an independent event selection with respect to the other CDF Higgs analyses, a veto to any additional identified lepton is applied. In this way no signal overlap is introduced between our $H \rightarrow \tau\tau$ search and that one optimized to the tri-lepton final state. The effect of this veto is that the W or a Z decaying leptonically are admitted only when these leptons are lost in the detector, fail the identification requirements or are misreconstructed as jets.

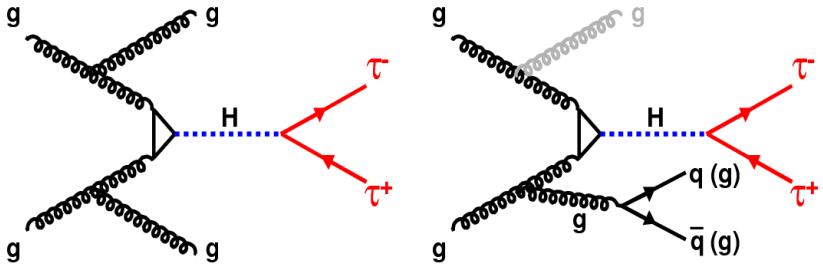

Figure 5.1: Tree Level Feynman diagram of Higgs production through gluon fusion with the Higgs subsequently decaying into two τ leptons.

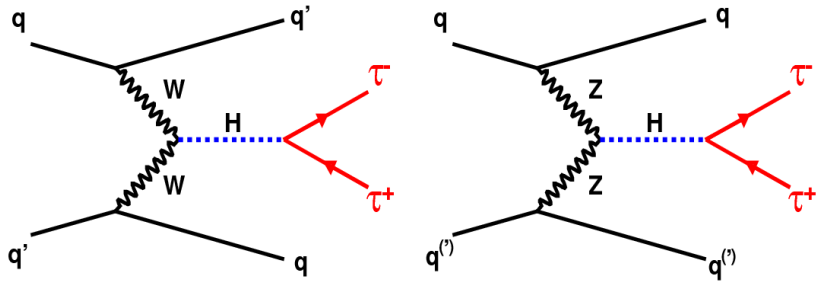

Figure 5.2: Tree Level Feynman diagram of Higgs production through vector boson fusion with the Higgs subsequently decaying into two τ leptons.

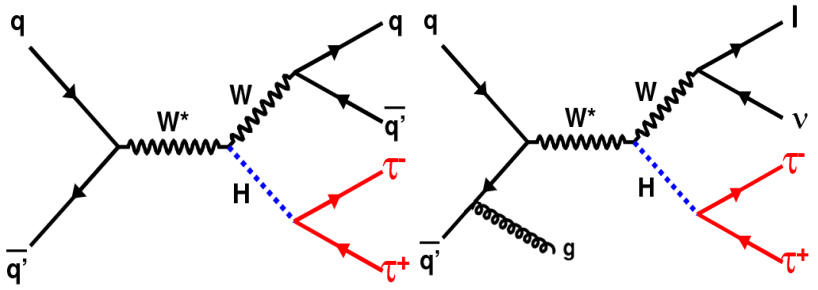

Figure 5.3: Tree Level Feynman diagrams for the Higgs produced in association with a W boson with the Higgs subsequently decaying into two τ leptons.

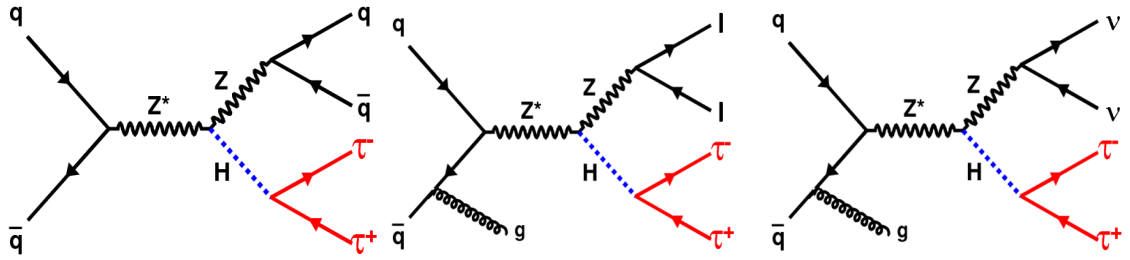

Figure 5.4: Tree Level Feynman diagrams for the Higgs produced in association with a Z boson with the Higgs subsequently decaying into two τ leptons.

Future prospects at the LHC

The interest on the H$\rightarrow \tau\tau$ decay mode increases in the perspective of the LHC experiments. As shown in figure 3.12, at 14 TeV the VBF production cross section becomes larger than the associated production: in the low mass region, the qHq$\rightarrow$q$\tau\tau$q channel is expected to be the most sensitive [78], because of the difficulties to probe the gluon-gluon fusion production channel followed by the Higgs decay into b quark pairs, overwhelmed by the large multijet background.

Additional studies have been performed by CMS [12] and ATLAS [13] collaborations, to extrapolate the sensitivity projections to lower LHC center of mass energies (7, 8 and 9 TeV): even though the importance of the H$\rightarrow \tau\tau$ decay mode is in this cases slightly reduced, this channel still play an important rule to extend the discovery potential down to 115 GeV/c^2, as can be seen in figure 5.5. The plot reports the CMS projected exclusion limits, as a function of the mass, for a SM Higgs search at 8 TeV, with an integrated luminosity of 5 fb^{-1}. The contributions of the individual channels used in the overall combination are also shown.

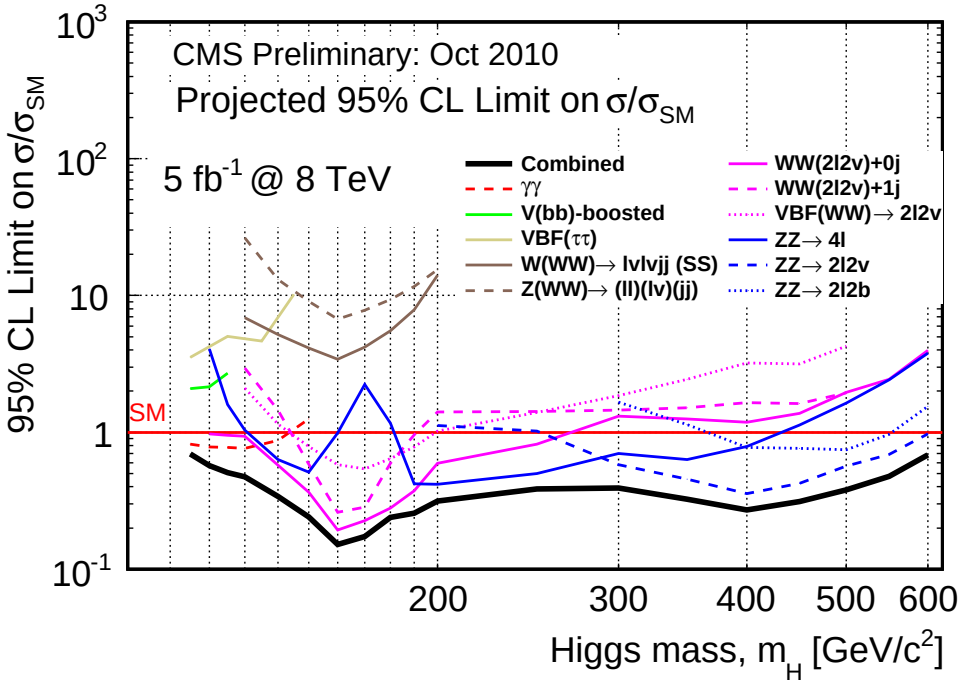

Figure 5.5: CMS expected 95% C.L. upper limits on the Higgs production cross section, as a function of m_H, for an integrated luminosity of 5 fb^{-1} at 8 TeV. Both the combination and the individual contributions of the different channels are shown.

Chapter 6

Data samples and event selection

6.1 Two τ Higgs decay

The analysis described in this thesis studies the SM Higgs decay mode into tau leptons, by exploring the most relevant Higgs production mechanisms in the hypothetical mass region between 100 and 150 GeV/c^2. Tau leptons are very short lived particles ($\tau = 290.6 \times 10^{-15}$ s [1]), which can be detected by CDF only through their visible decay products. Table 6.1 summarizes the most relevant decay modes and the corresponding branching ratios.

τ decay mode		B.R.(%)
leptonic	$e^-\bar{\nu}_e\nu_\tau$	17.85
	$\mu^-\bar{\nu}_\mu\nu_\tau$	17.36
hadronic	$h^-\pi^0\nu_\tau$	25.94
	$h^-\nu_\tau$	11.61
	$h^-h^-h^+\nu_\tau$	9.80
	$h^-\pi^0\pi^0\nu_\tau$	9.51
	$h^-h^-h^+\pi^0\nu_\tau$	4.75
	$h^-\pi^0\pi^0\pi^0\nu_\tau$	1.18

Table 6.1: Main final states for the hadronic and leptonic τ^- decays, with the corresponding branching ratios [1]; $h^\pm$ stands for $\pi^\pm$ or $K^\pm$. Final states for positive taus are represented by the charge conjugates of the modes listed here.

Hadronic decays are of the form

$$\tau \rightarrow X_h\nu_\tau \tag{6.1}$$

where X_h is a system of hadrons consisting of charged pions or kaons, or some short-lived intermediate resonances which decay directly to final states containing $\pi^\pm$, π^0,

$K^{\pm}$ and K^0. Their branching ratio is about 65% and their signature is characterized by a very narrow jet in the calorimeter, matching reconstructed tracks and π^0's.

The remaining channels (B.R.$\sim$35%) are represented by the leptonic decays,

$$\tau \rightarrow l\nu_l\nu_\tau \tag{6.2}$$

which appear in the detector as isolated electrons or muons. From now on in this writing we will denote the τ detected in a particular decay channel as τ_h, τ_μ and τ_e in the hadronic, muonic, electronic decay mode respectively.

Depending on the combinations of these modes, several possible final states are defined for the H$\rightarrow \tau\tau$ search, with different branching ratios and background contributions.

When looking at both τ's decaying hadronically, the search suffers from the extremely large QCD multijet background due to the high probability of erroneously identifying a jet as a τ, which is of the order of a few percent. On the other side, electrons and muons are well understood and characterized by a clean signature in the CDF detector, with an extremely small mis-identification probability($< 10^{-4}$).

For this reason, the most promising channels (the only ones considered in this thesis) consist in the $\tau_h\tau_\mu$ and $\tau_h\tau_e$ combinations, which represent a good compromise between the high branching ratio of the hadronic decay of one of the two taus and the large multijet background suppression provided by the request of an electron or a muon in the final state: the total fraction of the two τ decays which are then taken into account corresponds to about 46%.

di-tau decay mode	fraction(%)
$\tau_h\tau_h$	42.0
$\tau_h\tau_e$	23.2
$\tau_h\tau_\mu$	22.4
$\tau_e\tau_\mu$	6.2
$\tau_e\tau_e$	3.2
$\tau_\mu\tau_\mu$	3.1

Table 6.2: H$\rightarrow \tau\tau$ decay mode combinatorics.

6.2 Event topology

The search presented in this thesis is optimized for the detection of the SM Higgs boson in the $\tau_h\tau_\mu$ and $\tau_h\tau_e$ decay modes. The four production mechanisms under consideration (see section 5.1.3) provide a very characteristic signature: in both the vector boson fusion process (qq'$\rightarrow$q$\tau\tau$q') and in the associated production (qq'$\rightarrow$WH; qq'$\rightarrow$ZH) with a hadronically decaying W or Z, the Higgs boson is accompanied by two

quarks in the final state, which can be identified by the presence of hadronic calorimeter jets; in the gluon fusion case, even if no jets are expected at the tree level, they can be generated as well, when energetic gluons are emitted by the incoming partons.

Therefore, the optimized baseline selection of this analysis is the following:

- One isolated electron or muon with E_T (p_T)$\geq$ 10 GeV(/c);
- One hadronically decaying τ with $p_T \geq 15$ GeV/c;
- One or more calorimeter jets with $E_T \geq 20$ GeV.

In addition, the two leptons must have opposite charges, since they represent the decay products of the SM neutral Higgs boson.

A detailed description of the trigger selection, of the physical objects employed in the analysis (electrons, muons, hadronic τ's, calorimeter jets and missing transverse energy) and of the list of specific event cuts is provided in the following sections.

In table 6.3 the fractional contribution of the Higgs signal processes in the different jet multiplicity channels is reported.

process	0 jet	1 jet	$\geq$2 jets
ggH	74.0%	21.2%	4.9%
WH	23.1%	28.7%	48.2%
ZH	27.7%	24.3%	48.0%
VBF	8.3%	38.0%	53.7%

Table 6.3: Signal fractional contribution in the different jet multiplicity channels, calculated for M_H= 120 GeV/c^2, after the baseline selection is applied.

6.3 Trigger selection

This analysis covers the first 2.3 fb^{-1} of CDF data, collected by exploiting the three level trigger system described in section 4.2.7. Depending on the lepton type associated to the hadronic tau, the following trigger paths are used:

- TAU_ELECTRON8_TRACK5_ISO
- TAU_CMUP8_TRACK5_ISO
- TAU_CMX8_TRACK5_ISO

All these triggers, generally denoted as *lepton plus track*, require one electron or muon with $p_T \geq 8$ GeV/c and one additional isolated track with $p_T \geq 5$ GeV/c, which is used as a starting point for the hadronic τ reconstruction.

In the latest periods of data taking, additional *luminosity enabled* and *dynamically prescaled* versions of the trigger path involving CMX muons have been introduced.

- TAU_CMX8_TRACK5_ISO_LUMI_200: trigger disabled at instantaneous luminosities above 200×10^{30} cm^{-2} s^{-1}.

- TAU_CMX8_TRACK5_ISO_LUMI_250: trigger disabled at instantaneous luminosities above 250×10^{30} cm^{-2} s^{-1}.

- TAU_CMX8_TRACK5_ISO_DPS: Dynamic PreScale (DPS) enabled at level 2.

dataset	period	run range	lum [pb^{-1}]	cumulative lum[pb^{-1}]	dates
etlpad	0	138425-186598	520	520	04 Feb 02 - 22 Aug 04
etlpbh	1	190697-195408	130	650	07 Dec 04 - 18 Mar 05
	2	195409-198379	130	780	19 Mar 05 - 20 May 05
	3	198380-201349	100	880	21 May 05 - 19 Jul 05
	4	201350-203799	100	980	20 Jul 05 - 04 Sep 05
etlpai	5	203819-206989	150	1130	05 Sep 05 - 09 Nov 05
	6	206990-210011	120	1250	10 Nov 05 - 14 Jan 06
	7	210012-212133	50	1300	14 Jan 06 - 22 Feb 06
	8	217990-222426	210	1510	09 Jun 06 - 01 Sep 06
	9	222529-228596	200	1710	01 Sep 06 - 22 Nov 06
	10	228664-233111	290	2000	24 Nov 06 - 30 Jan 07
etlpaj	11	233133-237795	260	2260	30 Jan 07 - 31 Mar 07
	12	237845-241664	180	2440	01 Apr 07 - 13 May 07
	13	241665-246231	320	2760	13 May 07 - 04 Aug 07

Table 6.4: Data acquired by the CDF detector in the run periods considered in the H$\rightarrow \tau\tau$ search presented in this work: for each period the run number range, the starting and ending dates of data collection and the integrated luminosity are reported.

More details of the requirements applied at each level for all the trigger paths listed above are provided in appendix A.

The datasets which have been analyzed in this search, cover the run range from 141544 to 246231, corresponding to the first 14 periods of CDF Run II data taking. In table 6.4 we report for each run period the specific run range and the approximated luminosity.

At analysis level a good run selection is applied to the datasets, in order to keep only the events which have been collected with a good operational condition of the CDF detector subsystems used in this analysis: COT, calorimeters, CMX and CMU+CMP muon chamber, with no specific requirements on the condition of the silicon trackers. After considering this good run selection, the total integrated luminosity reduces from about 2.8 fb^{-1} to 2.3^{-1}.

6.3.1 Trigger efficiency

Analysis objects (electrons, muons and hadronic τ's in our case) are generally defined at offline level with tighter requirements than those applied at trigger level. However, only a fraction of these leptons actually satisfy the corresponding trigger selection, because of the different, worse resolution of the online variables with respect to the offline ones.

This effect, denoted as *trigger efficiency*, needs to be carefully evaluated and then applied to Monte Carlo (MC) samples in order to properly estimate the fraction of events which satisfy a specific set of trigger requirements.

The efficiencies of the lepton+track trigger paths used in this analysis have been calculated separately for the lepton and the isolated track. The overall efficiency is then given by the product of the efficiencies of the two objects, for the specific final state considered. Changes in the trigger selections and in the detector performances among different run periods are taken into account.

For the electron and muon legs we quote to the values commonly used by CDF for the high-p_T triggers (those requiring one electron(muon) with $E_T(p_T)$ greater than 18 GeV(/c), see appendix A), since it has been demonstrated [79, 80] that the efficiency plateau (90-95%) is already reached for a lepton with $p_T \geq 10$ GeV, with values consistent with those calculated for $p_T \geq 20$ GeV. For the tau leg, we refer to the studies performed in a previous search [79].

The procedure to compute the trigger efficiency is based on the definition of an offline *triggerable object*, starting from the categories used at the analysis level. The number of triggerable objects which have been identified in an unbiased dataset, i.e. a set of events collected with a different and uncorrelated trigger, is the denominator of the efficiency measurement. The number of triggerable objects which actually fired the corresponding trigger path represents the numerator.

6.4 High p_T objects identification

The physics objects which are of interest in the definition of the final state selections of our H$\rightarrow \tau\tau$ search are:

- electrons and muons, which come from the leptonic τ decays;
- the visible products of the hadronically decaying τ;
- Jets from quarks and gluons produced in the hard scattering processes or coming from initial (ISR) and final state radiation (FSR);
- neutrinos, which appear in all τ decays.

The selection criteria for electrons and muons are those commonly recommended by the CDF collaboration for high (p_T >20 GeV/c) and medium (p_T >10 GeV/c) transverse momentum leptons, belonging to the CEM, CMU+CMP and CMX categories. A

complete description of all the specific variables and the identification cuts is provided in sections 6.4.1 and 6.4.2.

Hadronic τ candidates are selected by applying a dedicated two-cone reconstruction algorithm, implemented at off-line level and described in section 6.4.3. In order to increase the jet$\rightarrow \tau_h$ fake rejection and keep a high efficiency on real τ's, the usual CDF set of rectangular identification cuts has been substituted by a new multivariate selection based on the Boosted Decision Tree method [81], which combines the discriminating power of the different identification variables into one single more powerful variable. The details of this new τ identification algorithm are provided in section 6.4.3.

The hadronic jets produced by quarks and gluons that undergo the hadronization and fragmentation processes in the detector, are measured in the calorimeters: the jet reconstruction algorithm and the energy correction procedure are described in section 6.4.4.

Finally, in section 6.4.5 a definition of the missing transverse energy, which is the indirect sign of the presence of neutrinos in the final state, is given.

6.4.1 Electrons

Central electrons are characterized by narrow showers in the central electromagnetic (EM) calorimeter ($|\eta_{\rm det}| <1.1$), matched to a track reconstructed either in the COT or in the silicon detector, originating from the interaction vertex. An electromagnetic cluster is defined by looking for a seed tower with $E_T >2$ GeV and by collecting the energy depositions of the adjacent towers in the η direction and within the same ϕ wedge, having $E_T >100$ MeV. E_T is the transverse component of the energy, calculated assuming that the lepton has originated from the center of the detector.

Electron candidates are selected by applying a set of quality and fiducial cuts, which are summarized in table 6.5. The complete list of the identification variables which are used, along with their description, is provided below. For more details, refer to the CDF documentation [82]

- $E_T^{\rm corr}$: the transverse component of the electron energy, defined as $E_T^{\rm corr} = E\times \sin\theta$ where θ is the polar angle calculated with respect to interaction vertex from which the electron has originated; E is the energy deposited in the electromagnetic calorimeter cluster.

- $E_{\rm HAD}/E_{\rm EM}$: the ratio of the energy deposited by the electron in the hadronic to the electromagnetic compartments of the calorimeter. A real electron deposits most of its energy in the EM calorimeters: by requiring $E_{\rm HAD}/E_{\rm EM}$ to be smaller than a selected value, it is possible to distinguish them from hadronic particles, which release a large fraction of their energy in the HAD calorimeters.

- E/P: the ratio of the energy of the EM cluster associated to the electron and the momentum of the track measured in the COT. For high energies, real electrons

can be considered as massless and their energy should equal to the magnitude of the momentum, thus E/P is expected to be $\widetilde{1}$.

- L_{shr}: a variables which accounts for the shape of the lateral shower profile of the electron calorimeter cluster. It compares the energy distribution of the EM towers adjacent to the seed tower, to the distribution which is expected from measurements performed with electron test-beam data.

- E_{rel}^{iso}: the calorimetric isolation, defined as the fraction of the transverse energy contained in a cone of radius $\Delta R = 0.4$ (after subtracting the energy of the electron itself) and the transverse energy associated to the electron candidate.

$$E_{\rm rel}^{\rm iso} \equiv \frac{E_T^{\Delta R=0.4} - E_T}{E_T} \tag{6.3}$$

- p_T: the transverse component of the momentum of the track matched to the electron object.

- Z_0: the z coordinate of the track extrapolated to the beam line.

- $|\Delta Z_{CES}|$: the distance in the (r,z) plane between the coordinates of the track, extrapolated to the plane of the CES detector, and the position of the nearest CES cluster.

- $Q \times \Delta x_{CES}$: the distance in the (r,ϕ) plane between the coordinates of the extrapolated track and the CES cluster, multiplied by the charge of the track. It takes into account for asymmetries in the shapes of the calorimeter clusters, originated from electrons subject to bremsstrahlung.

- CES χ^2_{strip}: a χ^2 comparison of the CES shower profile in the (r,z) plane to the expected one as measured in the electron test-beam.

- $N_{COT}(hits)$: number of hits for each COT superlayer, belonging to the track associated to the electron candidate.

- $N_{COT}(axSL)$: number of axial COT superlayers belonging to the electron track.

- $N_{COT}(stSL)$: number of stereo COT superlayers belonging to the electron track.

6.4.2 Muons

At the GeV energy scale, muons are Minimum Ionizing Particles(MIP): they thus traverse the entire detector with very little energy depositions in the tracking systems and in the calorimeters. Muon candidates are selected by matching COT tracks to segments left in the outer drift chambers.

Electron ID	
Variable	Cut
$\lvert\eta_{det}\rvert$	≤ 1.1
E_T	≥ 10 GeV
p_T	≥ 8 GeV/c
E_{had}/E_{em}	$\leq 0.055+0.00045\times E$
L_{shr}	≤ 0.2
E/P	≤ 2.0 unless $p_T \geq 50$GeV/c
$\lvert Z_0\rvert$	≤ 60 cm
$\lvert\Delta Z_{CES}\rvert$	≤ 3 cm
$Q \times \Delta x_{CES}$	$\geq$ -3.0 cm, ≤ 1.5 cm
CES χ^2_{strip}	≤ 10
E^{iso}_{rel}	≤ 0.1
N_{COT}(axSL)	≥ 3
N_{COT}(stSL)	≥ 2
N_{COT}(hits)	≥ 5
fiduciality	YES

Table 6.5: Baseline ID cuts for the central tight electrons (CEM).

The four momentum is defined by measuring the $\vec{p}$ of the track and by assuming a massless particle, i.e. E$\equiv |\vec{p}|$.

Two different categories are used in the analysis, consisting in central muons reconstructed in the CMU+CMP and in the CMX detectors, respectively. All the identification variables, except those already described for the electron case, are described below. The quality cuts applied to muon candidates are summarized in table 6.4.2.

- d_0: the track impact parameter, corrected for the measured position of the beam line. Different cuts are applied if the track contains silicon hits or if it doesn't.

- ρ_{COT}: it is the radius at which the track appears to leave the end plates of the COT:
$$\rho_{COT} \equiv \frac{\eta}{|\eta|} \times \frac{Z_{COT} - Z_0}{tan\lambda} \tag{6.4}$$
where
$$\lambda = \frac{\pi}{2} - \theta, \qquad and \qquad \theta = 2 \times tan^{-1}(e^{\eta}) \tag{6.5}$$
and $Z_{COT} = 155$ cm is the radial dimension of the COT. The lower cut on this variables ensures that the reconstructed muon passes through all the eight COT superlayers.

- E_{em}: energy deposited in the central electromagnetic calorimeter.

- E_{had}: energy deposited in the central hadronic calorimeter.
- E_{rel}^{iso}: in analogy to the electron identification (see equation 6.3), the calorimetric isolation is defined as

$$E_{rel}^{iso} \equiv \frac{E_T^{\Delta R=0.4}}{p_T} \tag{6.6}$$

- Δx_{CMU}: distance along the x coordinate of the CMU detector, between the extrapolated track and the stub in the muon chamber.
- Δx_{CMP}: distance along the x coordinate of the CMP detector, between the extrapolated track and the stub in the muon chamber.
- Δx_{CMX}: distance along the x coordinate of the CMX detector, between the extrapolated track and the stub in the muon chamber.

The fiduciality refers to some additional cuts on the fiducial distances of the tracks from the muon chambers, plus the removal of muons from specific regions of the subdetectors for run periods where these subdetectors were not operating or triggering properly. More details and definitions can be found in the specific documentation [83].

6.4.3 Taus

In this section we discuss the reconstruction of the visible decay products of the τ which undergo hadronic decay. Two separate categories, defined by counting the number of charged hadrons in the final state, are considered in this thesis:

1. *one-prong* τ's: B.R.$\sim$50%;
2. *three-prong* τ's: B.R.$\sim$15%;

The five-prong case is extremely rare (B.R.$\sim 10^{-3}$) and thus it is neglected.

Reconstruction algorithm

A reconstruction algorithm implemented at offline level, the *TauFinder*, creates a collection of objects which are identified as τ candidates, the so called *CdfTaus* [84].

First the algorithm builds a calorimeter cluster around a seed tower, by adding adjacent (*shoulder*) towers up to a predefined maximum number, which is kept low due to the narrowness of τ jets. Then tracks pointing to the cluster are searched and the one with the highest p_T is labeled as seed track. If other tracks have sufficient p_T and small separation from the z intercept of the seed track, they are associated to it.

Neutral pions, if present, are selected by exploiting the information provided by the central electromagnetic calorimeter and the CES detector [85]: CES clusters of 5

Muon ID	
Variable	Cut
p_T	$\geq$ 10 GeV/c
$\lvert Z_0 \rvert$	$\leq$ 60 cm
E^{iso}_{rel}	$\leq$ 0.1
$\lvert d_0 \rvert$	$\leq$ 0.2 cm if no Si hits
	$\leq$ 0.02 cm if Si hits
N_{COT}(axSL)	$\geq$ 3
N_{COT}(stSL)	$\geq$ 2
N_{COT}(hits)	$\geq$ 5
ρ_{COT}	$\geq$ 140 cm for CMX
fiduciality	YES
	if $p_T \leq$ 20 GeV/c
E_{em}	$\leq$ 2 GeV
E_{had}	$\leq$ 3.5 + p_T/8 GeV
Δx_{CMU}	$\leq$ 7 cm or $\chi^2_{CMU} \leq$ 9 for CMUP
Δx_{CMP}	$\leq$ 5 cm or $\chi^2_{CMP} \leq$ 9 for CMUP
Δx_{CMX}	$\leq$ 6 cm or $\chi^2_{CMX} \leq$ 9 for CMX
	if $p_T \geq$ 20 GeV/c
E_{em}	$\leq$ 2 + max(0,0.0115 $\times$ (p-100)) GeV
E_{had}	$\leq$ 6 + max(0,0.028 $\times$ (p-100)) GeV
CMU Δx_{CMU}	$\leq$ 7 cm for CMUP
CMP Δx_{CMP}	$\leq$ 5 cm for CMUP
CMX Δx_{CMX}	$\leq$ 6 cm for CMX

Table 6.6: Baseline ID cuts for CMUP and CMX muons.

wires/strips, not associated to charged tracks, determine the z and ϕ coordinates of π^0 candidates, which are assigned the energy of the corresponding CEM towers.

Reconstructed tracks and π^0's lying within a signal cone ($\theta \leq \theta_{\rm sig}$) defined around the seed track, as shown in figure 6.1, are considered τ decay products and used to build the *visible* four-momentum of the hadronic system. On the other hand, objects which fall in what is called the isolation region ($\theta_{\rm sig} \leq \theta \leq \theta_{\rm iso}$) are used to veto τ candidates because these more likely have originated from hadronic jets of QCD background events.

TauFinder provides two different definitions for the 3-D angle $\theta_{\rm sig}$: it may be kept to the constant value of 0.175 radiants (10°), or dependent on the cluster energy ($E^{\rm clus}$), as described in the following formula:

$$\theta_{\rm sig} = min(0.175, \frac{5.0GeV}{E^{\rm clus}})rad \tag{6.7}$$

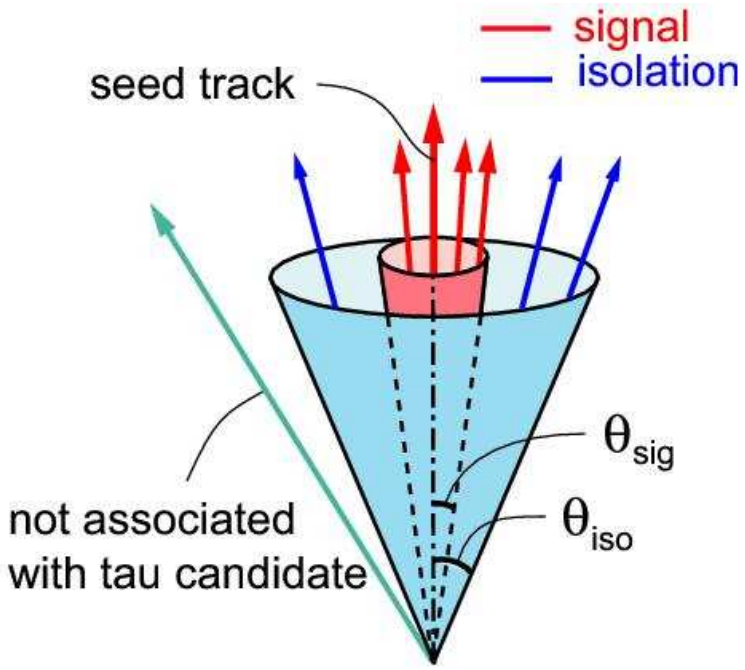

Figure 6.1: Signal cone and isolation region for tracks and π^0's of the τ candidate.

In addition, it is required that $\theta_{\rm sig}$ cannot be smaller than 0.05(0.1) rad, when adding tracks (π^0's) to the signal cone.

This *shrinking* signal cone is found to be more efficient in the discrimination of τ's against jets originated from quarks or gluons, as can be seen in figure 6.2; because of the relativistic boost, the decay products of real τ's are more collimated as the τ energy increases.

Due to te Cabibbo angle and the phase-space suppression, the fraction of $K^{\pm}$ in the final state is small compared to $\pi^{\pm}$: for this reason, in the reconstruction procedure all tracks are assumed to be charged pions and are assigned an invariant mass of 139 MeV/c^2.

A detailed description of all the τ variables implemented in the *TauFinder* algorithm is provided below.

- $N_{\rm twr}$: number of towers in the calorimeter cluster;
- $|\eta_{\rm det}|$: detector eta of the cluster;
- $\mathrm{E}_T^{\rm seed\ twr}$: transverse energy of the seed tower;
- $\mathrm{E}_T^{\rm sh\ twr}$: transverse energy of the shoulder towers;
- $\mathrm{E}_T^{\rm cal}$: total transverse energy of the calorimeter cluster;
- $\mathrm{E}_T^{\rm em}$: electromagnetic transverse energy in the calorimeter cluster;

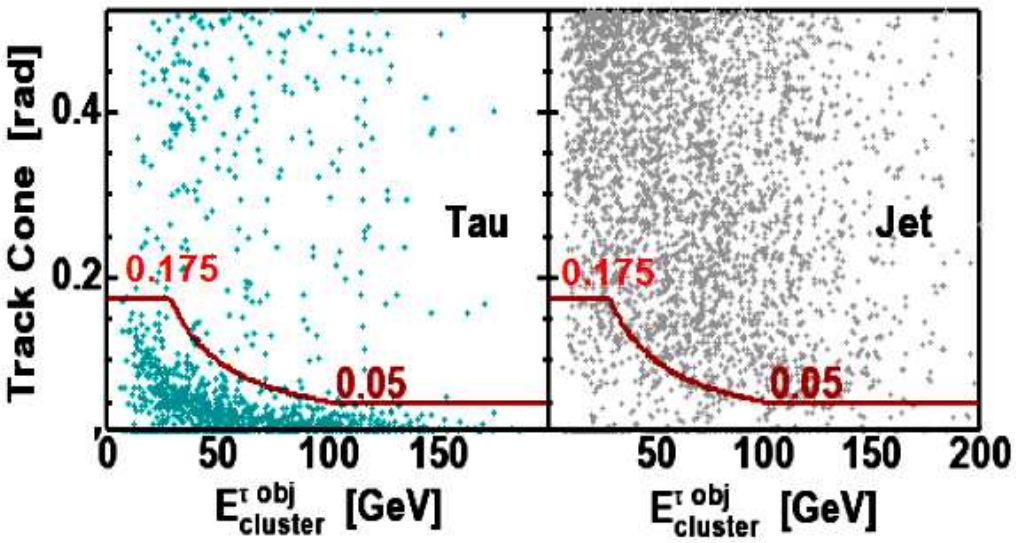

Figure 6.2: Signal cone as a function of the cluster energy.

- E_T^{had}: hadronic transverse energy in the calorimeter cluster;
- E_{em}^{cal}/E^{cal}: the fraction of electromagnetic energy in the calorimeter cluster;
- E_{rel}^{iso}: in analogy to the electron identification (see equation 6.3), the calorimetric isolation is defined as the fraction of the transverse energy contained in a cone of radius $\Delta R = 0.4$ around the seed track (after subtracting the cluster energy) and the cluster transverse energy.

$$E_{\rm rel}^{\rm iso} \equiv \frac{E_T^{\Delta R=0.4} - E_T^{\rm cal}}{E_T^{\rm cal}} \tag{6.8}$$

- $p_T^{\text{seed trk}}$: transverse momentum of the seed track;
- $p_T^{\text{sh trk}}$: transverse momentum of the shoulder tracks;
- $\Delta z^{\text{sh trk}}$: shoulder track distance from the seed track along the z coordinate;
- $\theta_{\rm sig}$: 3-D angle with respect to the seed track direction, defining the τ signal cone;
- $\theta_{\rm iso}$: 3-D angle with respect to the seed track direction, defining the isolation annulus $\theta_{\rm sig} < \theta < \theta_{\ iso}$.

- $d_0^{\text{seed trk}}$: the seed track impact parameter, corrected for the measured position of the beam line;

- Z_0: the z coordinate of the track extrapolated to the beam line.

- $|z_{\text{CES}}^{\text{seed trk}}|$: z coordinate of the extrapolation of the seed track to the CES detector radius

- $N_{\text{sig}}^{\text{trk}}$: number of tracks with p_T >1.0 GeV/c in the signal cone;

- $N_{\text{sig}}^{\pi^0}$: number of neutral pions with E_T >0.5 GeV in the signal cone;

- $|\sum_{i=1}^{N_{\text{trk}}^{\text{sig}}} Q_i|$: sum of tracks's charge for tracks with p_T >1.0 GeV/c in the signal cone;

- $\sum_{i=1}^{N_{\text{trk}}^{\text{sig}}} p_{Ti}$: sum of transverse momenta for tracks with p_T >1.0 GeV/c in the signal cone;

- $N_{\text{trk}}^{\text{iso}}$: number of tracks with p_T >1.0 GeV/c in the isolation region;

- $N_{\pi^0}^{\text{iso}}$: number of neutral pions with E_T >0.5 GeV in the isolation region;

- $\sum_{i=1}^{N_{trk}^{\text{iso}}} p_{Ti}$: sum of transverse momenta for all tracks in the isolation region(no minimum track p_T is applied);

- $\sum_{i=i}^{N_{\pi^0}^{\text{iso}}} E_{Ti}$: sum of transverse energy for all neutral pions in the isolation region (no minimum π^0 E_T is applied);

- $M_{trk+\pi^0}$: invariant mass of the visible hadronic system defined by tracks and neutral pions in the signal cone;

- p_T: τ visible transverse momentum. It is estimated by applying the procedure described in 6.4.3.

- ξ: this variable is designed to suppress electrons and muons depositing a large amount of energy in the electromagnetic calorimeter. It is defined as

$$\xi = \frac{E^{\text{cal}}}{\sum|\vec{p}|}\left(1 - \frac{E_{\text{em}}^{\text{cal}}}{E^{\text{cal}}}\right) \tag{6.9}$$

  where $|\vec{p}|$ is the scalar momenta of tracks in the signal cone.

A slightly different definition of this variable, which has not been adopted in this search, is the following:

$$\xi' = \frac{E^{\rm cal}}{\sum|\vec{p}|}\left(0.95 - \frac{E^{\rm cal}_{\rm em}}{E^{\rm cal}}\right) \tag{6.10}$$

The discriminating power of the ξ and ξ' variables is shown in figures 6.3 and 6.4 [11].

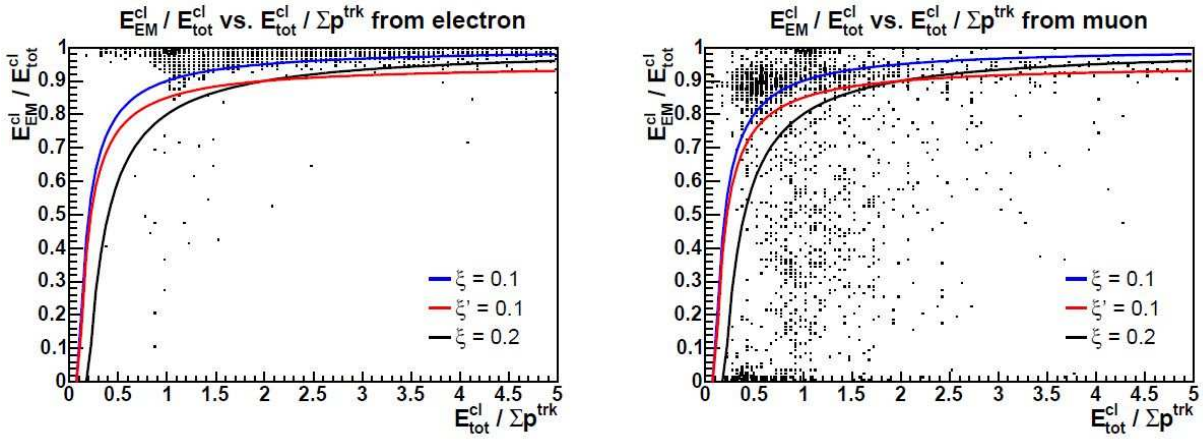

Figure 6.3: $E^{\rm cal}_{\rm em}/E^{\rm cal}$ vs $E^{\rm cal}/\sum|\vec{p}|$ for electrons and muons. The different cuts reject τ candidates which lay above the corresponding solid lines [11].

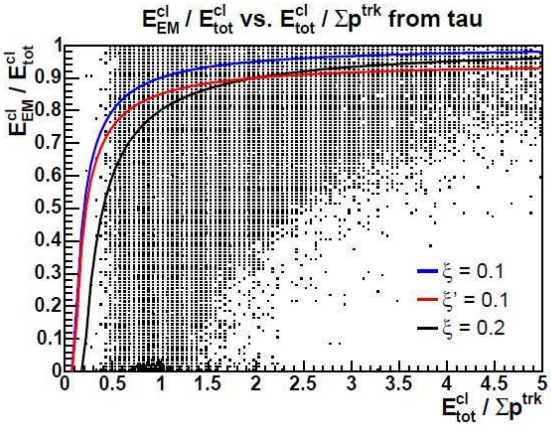

Figure 6.4: $E^{\rm cal}_{\rm em}/E^{\rm cal}$ vs $E^{\rm cal}/\sum|\vec{p}|$ for hadronic τ. The different cuts reject τ candidates which lay above the corresponding solid lines [11].

In this work we define a pre-selection of τ candidates, starting from the *CdfTau* category and applying a set of additional quality requests on the seed track, which must also satisfy the replication at offline level of the trigger isolation, as reported in

table 6.7. These cuts, summarized in table 6.8, are much looser than those used in any CDF analysis but they represent the starting point for the determination of the final identification selection.

	Seed track isolation	
Variable	Cut (run≤209770)	Cut (run>209770)
isolation region	$0.175 \leq \Delta R \leq 0.524$	$10^\circ \leq \Delta\theta \leq 30^\circ$
$p_T^{\text{sh trk}}$	$\geq$ 1.5 GeV/c	$\geq$ 1.5 GeV/c
$\Delta z^{\text{sh trk}}$	≤15 cm	≤ 5 cm

Table 6.7: Offline seed track isolation, representing a replication of the L3 trigger cuts: τ candidates with additional tracks satisfying these requirements are rejected.

Tau preselection	
Variable	Cut
N_{towers}	≤ 6
$\lvert\eta_{\text{det}}\rvert$	≤ 1.0
$E_T^{\text{seed twr}}$	$\geq$ 4.5 GeV
$E_T^{\text{sh twr}}$	$\geq$ 1.0 GeV
$p_T^{\text{seed trk}}$	≥6.0 GeV/c
$p_T^{\text{sh trk}}$	$\geq$ 1.0 GeV/c
$\Delta z^{\text{sh trk}}$	≤ 5.0 cm
θ_{sig}	see formula 6.7
θ_{iso}	= 0.52 rad
$\text{N}_{\text{COT}}(\text{axSL})^{\text{seed trk}}$	≥ 3
$\text{N}_{\text{COT}}(\text{stSL})^{\text{seed trk}}$	≥ 2
$\text{N}_{\text{COT}}(hits)^{\text{seed trk}}$	≥ 5
$\lvert z_{\text{CES}}^{\text{seed trk}}\rvert$	≥9 cm, ≤230 cm
seedtrack trigger isolation	YES
p_T	$\geq$ 15 GeV/c
$M_{trk+\pi^0}$	≤ 5.0 GeV/c^2

Table 6.8: Baseline ID cuts for the τ preselection.

Visible energy correction

The τ energy defined using visible decay products within the signal cone, $E_{\text{trk}+\pi^0}$, can be affected by π^0 reconstruction inefficiencies: it happens mostly when a π^0 hits the CES detector near the edges, or when its coordinates are too close to a reconstructed

track. Energy losses can thus be non-negligible, especially in the 1-prong case, where π^0's carry a larger fraction of τ momentum.

We apply an energy correction algorithm, described in [86], based on the choice among the following three different energy measurements:

1. $E_{\mathrm{trk}+\pi^0} = \sum_{i=1}^{\mathrm{N_{trk}}} p_i + \sum_{i=1}^{N_{\pi^0}} E_{\pi_i^0}$
2. $E_{\mathrm{cal}} = E_{\mathrm{cal}}^{\mathrm{EM}} + E_{\mathrm{cal}}^{\mathrm{had}}$
3. $E_{\mathrm{corr}} = E_{\mathrm{cal}}^{\mathrm{EM}} + \sum_{i=1}^{\mathrm{N_{trk}}} p_i - \sum_{i=1}^{\mathrm{N_{trk}}} E_{\pi^{\pm}}^{\mathrm{EM}} = E_{\mathrm{cal}}^{\mathrm{EM}} + \sum_{i=1}^{\mathrm{Ntrk}} (p_i - 0.32GeV)$

The latter is desirable when a π^0 reconstruction is suspected ($E_{\mathrm{corr}} > E_{\mathrm{trk}+\pi^0}$), provided that charged pions are actually MIP[1]. This is verified by checking the consistency of track momenta with the hadronic calorimeter measurement, within its resolution ($|p_{\mathrm{trk}} - E_{\mathrm{cal}}^{\mathrm{had}}| < 3\sigma_{\mathrm{had}}$) and by asking that $|E_{\mathrm{corr}} - E_{\mathrm{cal}}| < |E_{\mathrm{trk}+\pi^0} - E_{\mathrm{cal}}|$. These requirements prevent us from overestimating the τ energy in the cases where there are no real π^0's in the final state and there is a large amount of electromagnetic release due to the $\pi^{\pm}$.

When the conditions for this correction are not met, it has been demonstrated that E_{cal} gives a better estimate of the τ energy with respect to $E_{\mathrm{trk}+\pi^0}$.

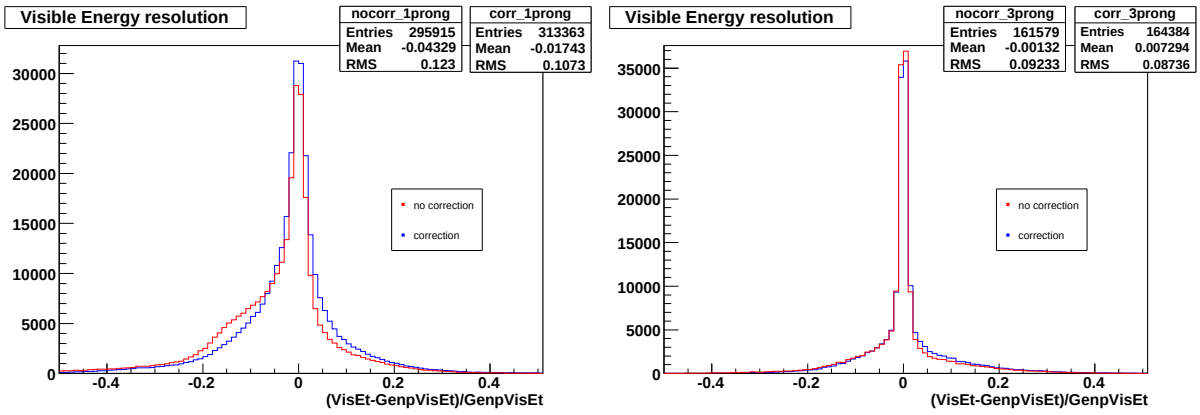

Figure 6.5: Energy resolution ($\frac{p_T^{rec} - p_T^{MC}}{p_T^{MC}}$) for 1-prong τ's and 3-prong τ's, before and after the application of the correction algorithm.

In figure 6.5 the energy resolution before (red) and after (blue) applying the correction is shown, for 1-prongs and 3-prong τ's selected in a MC sample, using the set of standard ID cuts reported in table 6.9. Resolution is defined as the relative difference

[1] $\pi^{\pm}$ test beam measurements have shown that the mean energy release in the CDF electromagnetic calorimeter is 0.32 GeV at the MIP peak

between the assigned transverse visible momentum ($p_T^{\rm rec}$) and the value reconstructed at the MC generator level, without the neutrino ($p_T^{\rm MC}$).

In our analysis we decided to apply the correction algorithm only in the 1-prong case, where the resolution improvement is more appreciable: the fraction of events with the energy lying within 10% of the real value, increases from 69% to 75%, the mean value is closer to 0, the Root Mean Square (RMS) is reduced. In addition, nearly an additional a 6% of the low energy τ's satisfy the selection criteria by passing the 15 GeV/c cut.

Standard ID and BDT-based selections

The final τ identification criteria usually differ slightly to the different CDF analyses, but they generally consist of a set of rectangular cuts applied to some discriminating variables. These cuts are optimized to distinguish τ's from electrons and quark or gluon jets and generally represent a compromise between the need to keep a high signal efficiency and to reduce the fake rate.

The distributions with the largest separation power are those related to the objects reconstructed in the isolation annulus: a tight selection is defined by putting a veto to any additional track or π^0 in this region; a looser selection may admit some small charged or neutral pion contamination, by setting an upper limit to the corresponding sum of transverse momenta and energy.

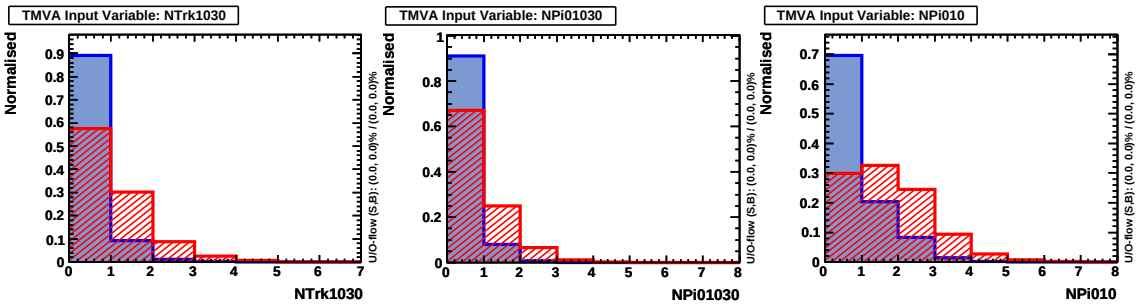

Figure 6.6: 3-prong tau candidates, passing the preselection summarized in table 6.8; distributions for real τ's extracted from a W$\rightarrow \tau\nu$ MC (blue) and from a jet background data sample (red) are shown. From the left to the right: number of tracks with $p_T > 1.0$ GeV/c in the isolation region, number of neutral pions with $E_T > 0.5$ GeV in the isolation region, number of neutral pions with $E_T > 1.0$ GeV in the signal region.

Another variable which efficiently separates hadronic τ's from fakes is the visible invariant mass. Because of the lost neutrino, the reconstructed value is almost always smaller than the real τ mass ($m_\tau = 1776.82 \pm 0.16$ MeV/c^2 [1]), while for quark or gluon jets this variable has a wider distribution.

Figures 6.6 and 6.7 show a comparison between 3-prong τ's extracted from MC datasets and τ-like jets selected in data, satisfying the preselection summarized in table 6.8.

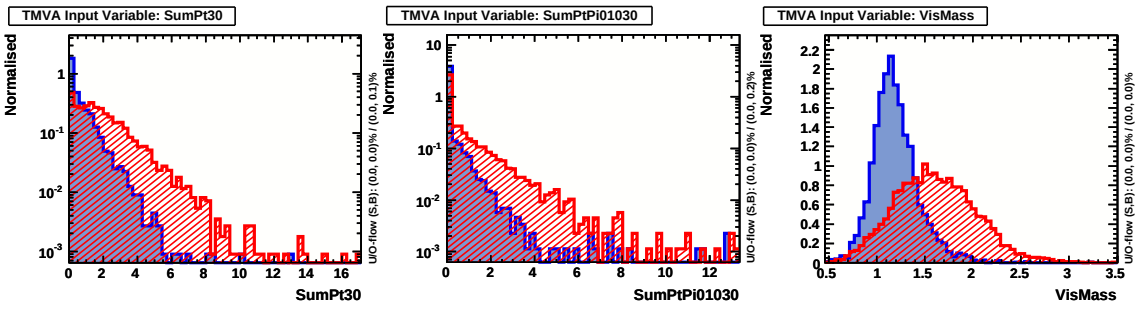

Figure 6.7: 3-prong τ candidates, as in figure 6.6. From the left to the right: sum of transverse momenta for all tracks in the isolation region, sum of transverse energy for all neutral pions in the isolation region, reconstructed visible mass.

Standard tau ID	
Variable	Cut
pass preselection	YES
$N_{\rm trk}^{\rm sig}$	= 1-3
$\left\|\sum_{i=1}^{N_{\rm trk}^{\rm sig}} Q_i\right\|$	= 1
$N_{\rm trk}^{\rm iso}$	= 0
$N_{\pi^0}^{\rm iso}$	= 0
$M_{trk+\pi^0}$	$\leq$ 1.8 GeV/c^2
$E_{\rm rel}^{\rm iso}$	$\leq$ 0.1
$\|Z_0\|$	$\leq$ 60 cm
ξ	$\geq$ 0.2

Table 6.9: Baseline ID cuts for standard τ's at CDF [3].

The standard ID presented in table 6.9 was adopted in a previous search [3] and has been chosen here as a benchmark for our studies aimed at improving the identification performances. The idea is to exploit all the information available in the τ candidate objects, by replacing this cut-based selection with a more sophisticated algorithm based on the *Boosted Decision Tree* (BDT) method. The basic concepts of this multivariate event classifier, which has been proved to be particularly robust and reliable especially for particle identification purposes [81, 87] are provided in appendix B; a complete description of the method can also be found in [6].

The starting point of the new identification algorithm is the building of the training samples: τ candidates are defined by applying the preselection summarized in table 6.8 and are extracted from a W$\rightarrow \tau_h\nu_\tau$ MC and from a data sample containing events with at least one jet with E$_T$ $\geq$20 GeV (JET_20 trigger path [2]).

[2]See appendix A for more details about the JET_20 trigger path

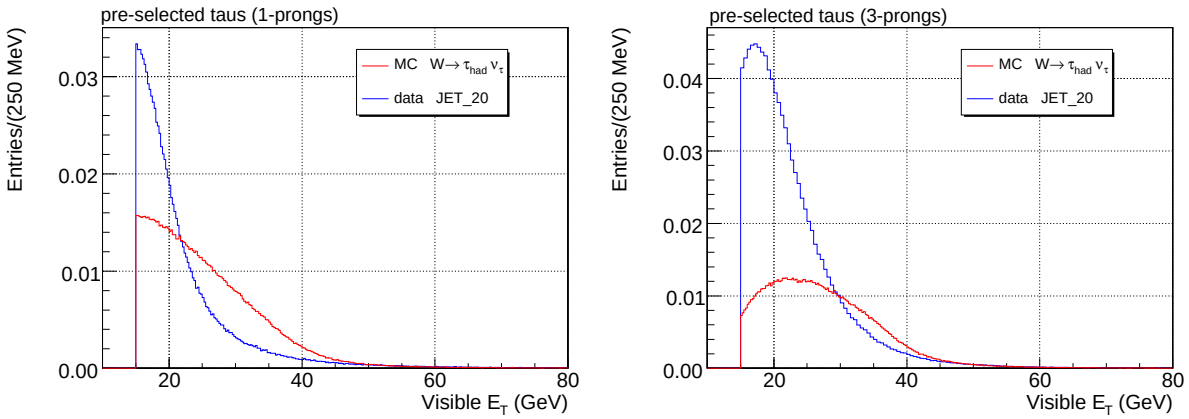

Figure 6.8: Normalized visible E_T spectra for the preselected τ candidates (1-prong on the left, 3-prongs on the right) extracted from a $W \rightarrow \tau_h \nu_{tau}$ MC and from a jet data sample.

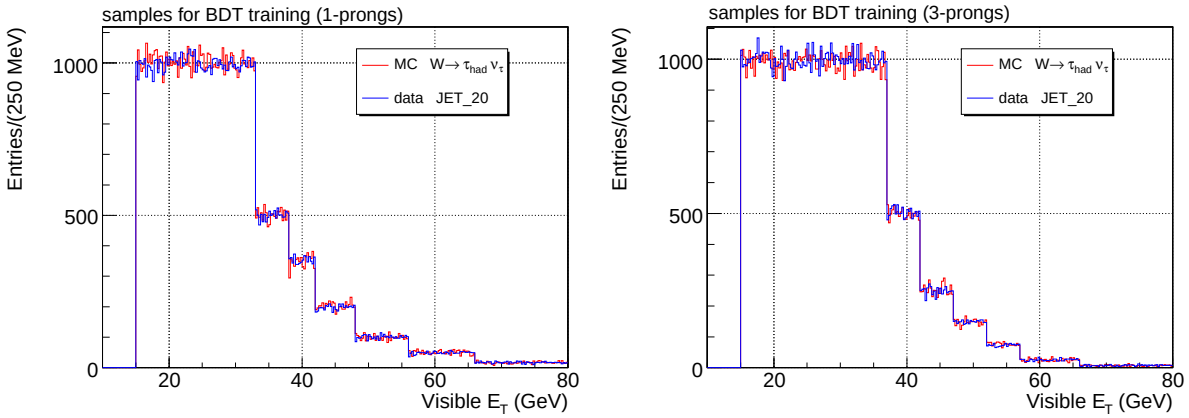

Figure 6.9: Flat visible E_T spectra of the selected for training samples (1-prong on the left, 3-prongs on the right): the number of events in each energy subrange has been chosen according to the available statistics.

The visible energy spectra of the preselected τ's are reported in figure 6.8, separately for the 1-prong and 3-prong categories. Given the strong dependence of the identification performances from the τ energy, it is important that the training procedure is not affected by the particular energy population of the datasets used to extract the training samples. We overcome this problem by randomly retaining only a fraction of the events, in such a way to have a flat distributions within a specific interval of visible energy. The normalized flat distributions for signal (real taus) and background (jets) candidates are shown, superimposed, in figure 6.9. Energy subrange widths and number of events have been chosen accordingly to the statistics available in the original samples.

The second step is the definition of several τ subcategories, according to the number of tracks in the signal cone (either 1 or 3) and to the visible transverse energy. A specific BDT is then implemented for each category and trained to distinguish between real τ's and calorimeter jets. The complete list of input training variables is reported in table 6.11.

At the end, a total of 16 and 17 BDTs have been trained for the 1-prong and 3-prong τ categories, in the transverse energy range between 15 and 80 GeV, as summarized in table 6.10. In order to maximize the discriminating power of the training variables, we applied a decorrelation procedure by diagonalizing the symmetric covariance matrices, separately for the signal and background samples, as described in [6] and in appendix B.

1-prong τ's		3-prong τ's	
Energy range	BDT cut value	Energy range	BDT cut value
15 - 16 GeV	0.445	15 - 16 GeV	0.62
16 - 17 GeV	0.32	16 - 17 GeV	0.62
17 - 18 GeV	0.40	17 - 18 GeV	0.56
18 - 19 GeV	0.365	18 - 19 GeV	0.605
19 - 20 GeV	0.27	19 - 20 GeV	0.60
20 - 22 GeV	0.32	20 - 22 GeV	0.57
22 - 24 GeV	0.25	22 - 24 GeV	0.53
24 - 26 GeV	0.315	24 - 26 GeV	0.435
26 - 29 GeV	0.215	26 - 28 GeV	0.42
29 - 33 GeV	0.11	28 - 32 GeV	0.395
33 - 38 GeV	0.26	32 - 37 GeV	0.45
38 - 42 GeV	0.24	37 - 42 GeV	0.4
42 - 48 GeV	0.275	42 - 47 GeV	0.395
48 - 56 GeV	0.305	47 - 52 GeV	0.42
56 - 66 GeV	0.4	52 - 57 GeV	0.42
66 - 80 GeV	0.295	57 - 66 GeV	0.395
		66 - 80 GeV	0.31

Table 6.10: BDT output cuts applied to the different τ subcategories.

Once a given event is analyzed by the trained BDT, it is assigned an output value, called "score", which can range between -1 and +1 depending on whether the τ candidate is more compatible with a quark or gluon jet or a real hadronically decaying τ. Figure 6.10 shows the BDT output distributions for two of the predefined τ categories. The training samples are compared with statistically independent samples, and Kolmogorov-Smirnov tests are performed in order to check the consistency of the two shapes and to determine the occurrence of *overtraining* [3].

[3] The overtraining is a seeming increase in the classification performance with respect to the real one.

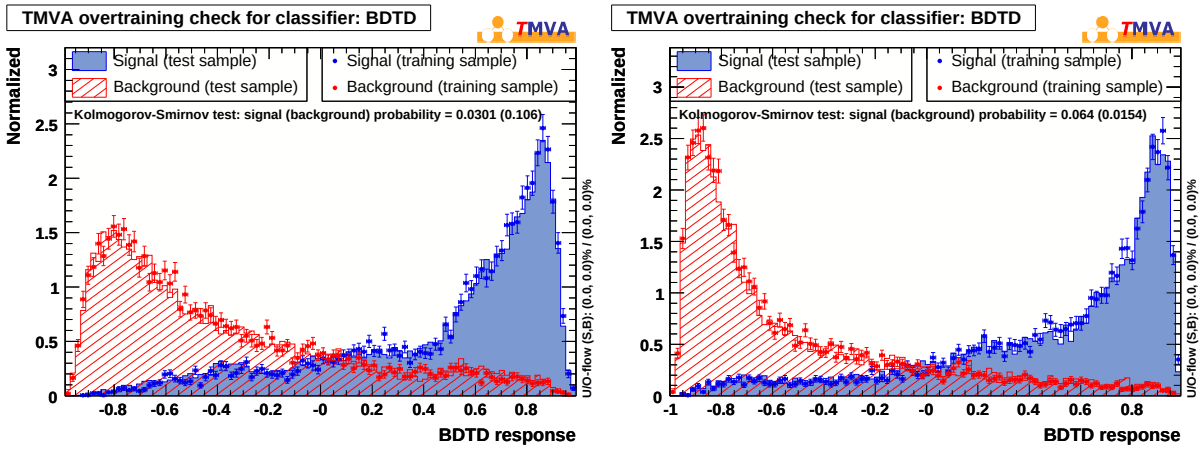

Figure 6.10: BDT output distributions for signal and background samples: the 1-prong and 3-prong cases, in the 29-33 GeV and 28-32 GeV visible energy ranges are shown. Training and test samples are superimposed.

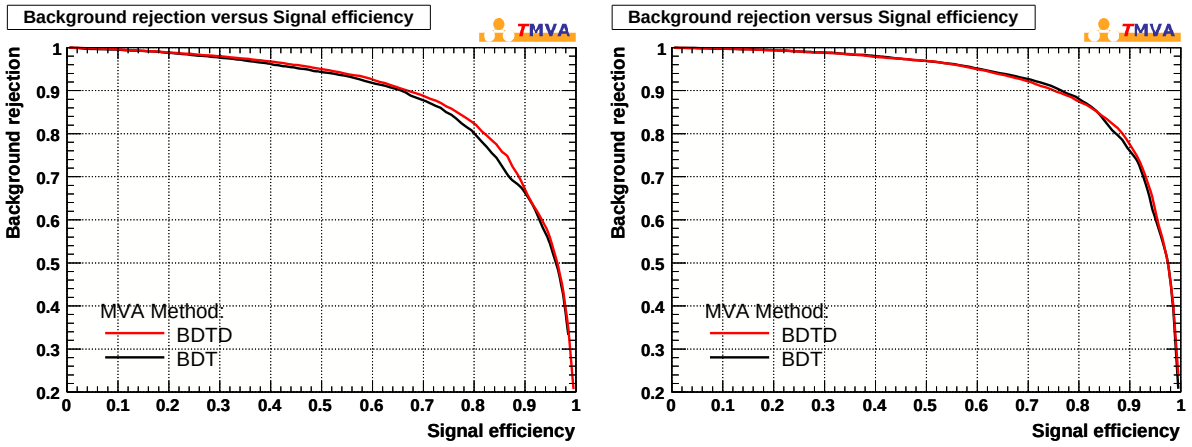

Figure 6.11: Background rejection rate versus signal efficiency, obtained by cutting on the BDT output distributions for the 1-prong and 3-prong cases shown in figure 6.10. The BDTD curve represents the performances obtained after the decorrelation procedure on the training variables is applied.

By placing a lower limit cut on the classifier response, most of the real τ candidates are retained, along with a small fraction of jet$\rightarrow \tau_h$ fakes background. The performances of the method can be studied by looking at the relationship between the background rejection rate and the τ identification efficiency, as reported in figure 6.11: each point in the plot corresponds to a specific lower cut on the classifier output distribution.

A failure of the K-S test, (i.e. a probability smaller than a few percent that the output distributions for the training and test samples are statistically consistent), indicates that overtraining is probably affecting the results of the algorithm. See appendix B for additional details.

BDT input variable	description
Visible p_T	visible transverse momentum
$M_{trk+\pi^0}$	invariant mass
$N^{\rm iso}_{\rm trk}$	# of tracks with p_T >1.0 GeV/c in the isolation region
$N^{\rm iso}_{\pi^0}$	# of π^0 with E_T >0.5 GeV in the isolation region
$N^{\pi^0}_{\rm sig}$	# of π^0 with E_T >0.5 GeV in the signal cone
$\sum_{i=1}^{N^{\rm sig}_{\rm trk}} p_{Ti}$	sum of transverse momenta of tracks in the signal cone
$\sum_{i=1}^{N^{\rm iso}_{trk}} p_{Ti}$	sum of transverse momenta of tracks in the isolation region
$\sum_{i=i}^{N^{\rm iso}_{\pi^0}} E_{Ti}$	sum of transverse energies of π^0s in the isolation region
$d_0^{\rm seed\ trk}$	seed track impact parameter
$\rm E^{cal}_{EM}/E^{cal}$	fraction of electromagnetic energy
$\rm E^{em}_T$	electromagnetic transverse energy
$\rm E^{had}_T$	hadronic transverse energy

Table 6.11: Input variables used to train the tau ID BDT.

It turns out that such a τ selection can be particularly flexible, because the cuts on the different BDT output distributions can be easily changed to increase signal efficiency or alternatively reduce the fake rates, and this optimization procedure can be done separately for each τ category: in this search we decided to define the cuts in order to keep a similar identification efficiency with respect to the set of cuts summarized in table 6.9, with the purpose of maximizing the fake rejection.

We show here the final performance plots, separately for the 1-prong and 3-prong categories: the relative efficiency and fake rates, as a function of the visible corrected transverse energy, are shown in figures 6.12 (W$\rightarrow \tau_h\nu_\tau$ MC) and 6.13 (JET_20 data) for the BDT-based and the cut-based standard ID. The relative efficiency is defined as the fraction of preselected τ's, matched with the generated ones within ΔR <0.2, which are selected by the ID algorithm. Similarly the relative fake rate is provided by estimating the fraction of preselected τ's, matched with a calorimeter jet (ΔR <0.4), which pass the identification requirements.

As can be seen, the efficiency is consistent within a few percent with the standard selection, while the corresponding fake rate is much smaller, in particular for low energy tau candidates. In the analysis presented in this thesis, a 25% reduction of jet$\rightarrow \tau_h$ fakes is achieved by implementing the new τ ID algorithm.

Loose ID selection

It is useful to define a looser and completely independent τ ID selection, defined by vetoing candidates which pass the BDT based ID and by applying the cuts summarized in table 6.12.

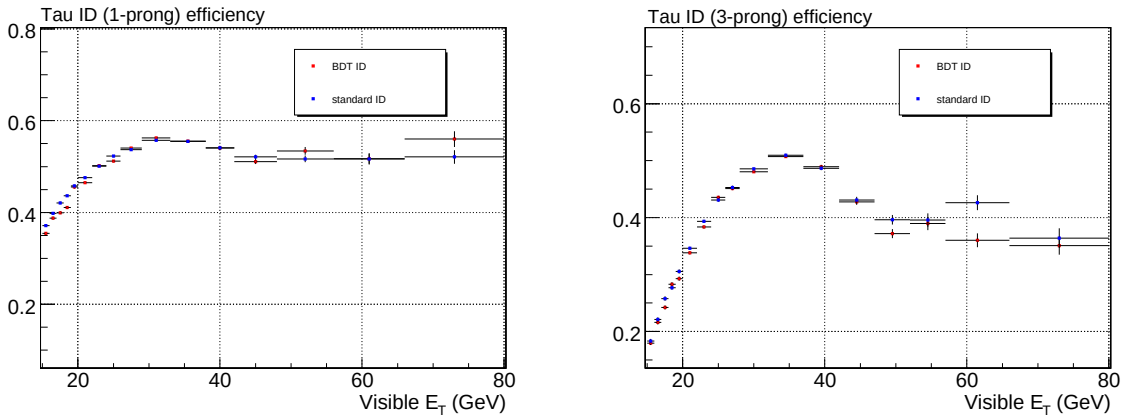

Figure 6.12: Standard and BDT reconstruction efficiency for hadronic τ's (1-prong on the left, 3-prong on the right) extracted from a W$\rightarrow \tau_h \nu_\tau$ MC sample, relative to the preselection level described in table 6.8.

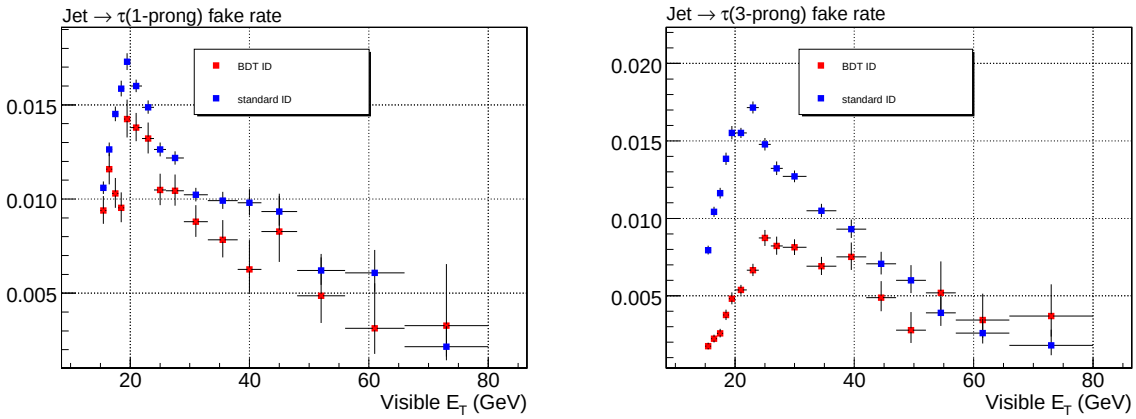

Figure 6.13: Standard and BDT jet$\rightarrow \tau_h$ fake rate (1-prong on the left, 3-prong on the right) for jets extracted from a JET_20 data sample, relative to the preselection level described in table 6.8.

6.4.4 Calorimeter jets

Several different types of clustering algorithms have been developed at CDF and implemented at offline level, for the reconstruction of calorimeter jets.

For this study, we use a fixed cone algorithm, JetClu [88], in which the center of a jet is defined as (η^{jet},ϕ^{jet}) and the size of the cone is R = $\sqrt{(\eta^{twr} - \eta^{jet})^2 + (\phi^{twr} - \phi^{jet})^2} \leq$ 0.4, meaning that only the calorimeter towers whose coordinates in the (η,ϕ) space satisfy this relation are included in the jet.

The clustering procedure is the following. First, a list of seed towers with E_{Ti} >1 GeV is created, where $E_{Ti} = E_i \sin\theta_i$ represents the transverse energy measured in the electromagnetic and hadronic calorimeter sectors. For each seed tower, a cluster is

Loose tau ID	
Variable	Cut
pass preselection	YES
pass BDT ID	NO
$N_{\rm trk}^{\rm sig}$	= 1-3
$\lvert\sum_{i=1}^{N_{\rm trk}^{\rm sig}} Q_i\rvert$	= 1
$E_T^{\rm seed\ twr}$	$\geq$ 6.0 GeV
$E_T^{\rm cal}$	$\geq$ 9.0 GeV
$D_0^{\rm seed\ trk}$	$\leq$ 0.2 cm
$\sum_{i=1}^{N_{trk}^{\rm iso}} p_{Ti}$	$\leq$ 2 GeV/c
$\sum_{i=i}^{N_{\pi^0}^{\rm iso}} E_{Ti}$	$\leq$ 1 GeV
ξ	$\geq$ 0.2

Table 6.12: Loose tau ID cuts.

built by collecting the adjacent towers within a radius of size R.

The cluster location and the transverse energy are calculated with the following definitions:

$$E_T^{jet} = \sum_{i=0}^{N_{tow}} E_{T_i} \tag{6.11}$$

$$\phi^{jet} = \sum_{i=0}^{N_{tow}} \frac{E_{Ti}\phi_i}{E_T^{jet}} \tag{6.12}$$

$$\eta^{jet} = \sum_{i=0}^{N_{tow}} \frac{E_{Ti}\eta_i}{E_T^{jet}} \tag{6.13}$$

where N_{tow} is the number of towers inside the radius R. The procedure is repeated, as a new list of towers around the new E_T-weighted center of the cluster is determined; iteration ends when the assigned list of towers becomes stable.

Overlapping clusters are then merged into one single jet if the overlap is greater than 50%, otherwise each tower in the common region is assigned to the nearest cluster.

Jet energy corrections

The direct measurement provided by the calorimeter, commonly referred as *raw energy*, requires several corrections before giving a reliable estimate of the energy of the final state at the particle level, that is the sum of hadrons, leptons and photons momenta

within the clustering cone. Furthermore, additional corrections need to be applied to associate the particle jet energy to the original parent parton.

The transition from the *calorimeter jet* to the *parton jet*, passing through the *particle jet*, is summarized in the diagram reported in figure 6.14.

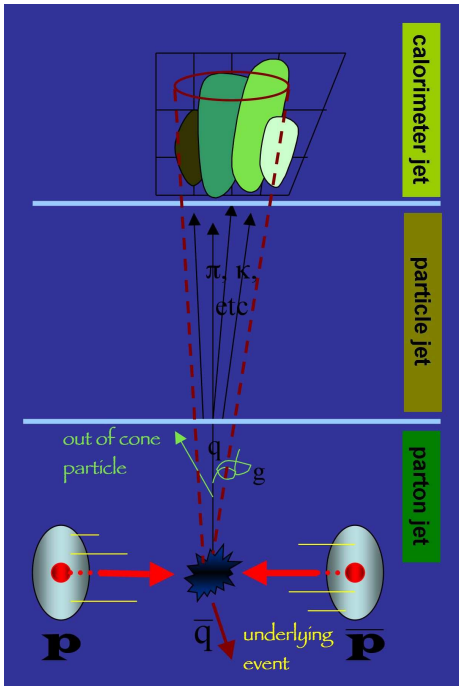

Figure 6.14: A diagram showing the transition from the measured calorimeter jet in the detector to the parton level jet.

The correction procedure [88] is organized into different levels, each of them considering a specific effect which could distort the energy measurement, like the response of the calorimeter to different particles, non linear response to different momenta of the same particle, un-instrumented or non-uniform regions of the detector, spectator interactions and energy radiated outside the clustering cone.

The original parton transverse energy (p_T^{parton}) is determined by using the following formula

$$p_T^{\text{parton}} = (p_T^{\text{jet}} \times C_\eta - C_{\text{MI}}) \times C_{\text{ABS}} - C_{\text{UE}} + C_{\text{OOC}} = p_T^{\text{particle}} - C_{\text{UE}} + C_{\text{OOC}} \quad (6.14)$$

which incorporates all the correction levels listed in table 6.13 and described below:

- **Level 0**: p_T^{jet} is calculated by setting the calorimeter energy scale. Electromagnetic compartments are calibrated by imposing that the reconstructed mass of the Z boson in the $Z \rightarrow e^+e^-$ decay mode is consistent with the mass measured at LEP [89]. Resolution is further improved for each tower by applying an additional calibration factor given by the ratio of the calorimeter energy and the

Level	Parameter	Description
0		Online/Offline calibration
1	C_η	η-dependent
2		Not in use
3		Not in use
4	$C_{\rm MI}$	Multiple interactions
5	$C_{\rm ABS}$	Absolute
6	$C_{\rm UE}$	Underlying event
7	$C_{\rm OOC}$	Out-of-cone

Table 6.13: CDF jet energy corrections.

track momenta of the electron candidates. On the other side, the energy scale of the hadronic towers is derived from test beam data, by evaluating the response to 50 GeV/c charged pions.

The calorimeter stability is monitored online by means of various calibration methods: laser systems, radioactive Co^{60} sources, muons from $J/\Psi \rightarrow \mu^+\mu^-$ decays and minimum bias data. The response decrease in time, due to the aging of scintillators and photo-multipliers, is then taken into account by periodically updating the energy scales.

- **Level 1 (C_η)**: the CDF calorimeter is not uniform in pseudorapidity. A lower response arises in the poorly instrumented regions due to the physical separation between the different calorimeter components: at η=0, where the two halves of the central calorimeter join, and at $|\eta| \approx$1.1, on the border between the central and the plug calorimeters.

  The η-dependent correction is obtained by using the so called *dijet balancing method*, based on the assumption that the two leading jets in dijet events are equal in p_T in absence of hard QCD radiation. Corrections are determined separately for data and Monte Carlo and for different p_T bins.

- **Level 4 ($C_{\rm MI}$)**[4]: the energy coming from different $p\bar{p}$ interactions, which take place in the same bunch crossing, is adequately subtracted when falling within the cone of a cluster. This correction is derived from minimum bias data, by measuring the transverse energy contained in a random cone of radius R = 0.4 and by parametrizing the value as a function of the number of primary vertexes of the events.

  The number of $p\bar{p}$ interactions per bunch crossing follows a Poisson distribution and the mean <N> increases linearly with the instantaneous luminosity: it is one

[4]Level 2 and level 3 corrections are obsolete: they were applied only at the beginning of Run II and the original numbering has been kept for consistency.

for $\mathcal{L} = 0.4 \times 10^{32}$ cm^{-2}s^{-1} and reaches the value of 8 for $\mathcal{L} = 3 \times 10^{32}$ cm^{-2}s^{-1}. Consequently this kind of correction has increased its relevance over the time of CDF data taking.

- **Level 5** (C_{ABS}): the absolute correction aims to transform the measured jet energy into the energy corresponding to the underlying particle jet. The procedure is based on MC simulations and its accuracy depends on how well the response of the calorimeter to the single particle is modeled (calorimeter simulation) and on how well the multiplicity and the p_T spectra of the particles inside a jet are reproduced (fragmentation simulation).

- **Level 6** (C_{UE}) and **level 7** (C_{OOC}): the reconstruction of the original parton momentum starting from the particle jet requires to consider the following additional effects. The underlying event, i.e. spectator quarks or particles from initial state radiation, can contribute to the total amount of cluster energy. On the other hand, a fraction of the parton energy can escape from the jet cone because of the presence of fragmentation products or final state gluon radiation with a large exit angle, or low p_T particles widely bent by the magnetic field. The effect of these processes is estimated using MC simulations.

Systematic uncertainties

The total systematic uncertainty assigned to the energy of the reconstructed jets used in the analysis is obtained by adding in quadrature each specific source of systematic, related to the different levels of correction.

The total and relative contributions are reported, as a function of the corrected jet p_T, in figure 6.15.

Jet definition

Calorimeter jets suitable for the analysis are defined by applying the following selection cuts:

- E_T (corrected at Level 5)$>$ 20 GeV
- $|\eta| \leq 2.5$
- $E_{em}/E < 0.9$

Jets are also required to be not matched with any reconstructed lepton in the final state ($\Delta R_{lep-jet} > 0.4$).

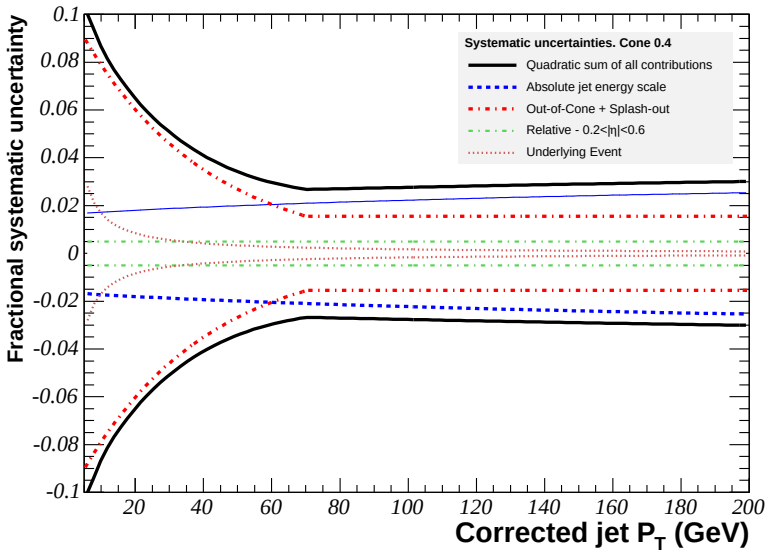

Figure 6.15: Fractional systematic uncertainties, coming from each energy correction level, for ΔR = 0.4 cone size jets, as a function of the corrected p_T.

6.4.5 Missing transverse energy

Neutrinos do not interact with the detector, but they can be indirectly identified by the presence of unbalanced energy in the event. Since the longitudinal component of the colliding partons momentum is unknown, only the missing energy in the transverse plane, $\not\!\!E_T$, is calculated: apart from instrumental mismeasurements, it represents the transverse component of the sum of the neutrinos momenta.

The raw $\not\!\!E_T$ is defined as:

$$\vec{E_T^{raw}} = -\sum_i E_T^i \hat{n_i} \tag{6.15}$$

where E_T^i is the transverse energy of the i-th calorimeter tower and $\hat{n_i}$ is the transverse unit vector pointing to the center of each tower.

Several corrections are applied to the raw value: first the transverse energy components are calculated with respect to the event interaction vertex instead of the center of the detector (z=0); then $\not\!\!E_T$ is adjusted for the possible presence of identified muons in the final state, by replacing the transverse energy left by muons in the calorimeter towers with the track p_T; finally, the same corrections used for the jet energy of all the jets with $E_T^{raw} \geq 10$ GeV and $|\eta| \leq 2.5$ are applied.

6.5 Event cuts

The general strategy of this analysis is to apply very simple and minimal requirements to keep Higgs signal acceptance as high as possible and then to rely on a multivariate

technique to exploit all the kinematical and topological information to discriminate signal against the main sources of background.

The starting point is represented by the events collected the trigger paths listed in section 6.3. Leptons are then reconstructed accordingly to the algorithms and the sets of identification cuts described in previous sections.

The final state selection consists of exactly one electron or muon with $E_T(p_T)$ >10 GeV(/c) and one hadronic τ with visible p_T > 15 GeV/c, spatially well separated one from the other (ΔR >0.4).

Lepton charges must have opposite signs, given the fact that we are trying to isolate the decay products of the SM Higgs boson, which is a neutral particle. Events with additional identified leptons in the final state are then rejected.

By cutting on the distance between leptons in the extrapolated Z_0 coordinate, $|Z_0^1 - Z_0^2| < 5$ cm, we ensure that they have originated from the same $p\bar{p}$ interaction. A good quality primary vertex must also be detected along the axial direction z, in the fiducial central region $|Z_0^{vtx}| < 60$ cm and close to both leptons, $|Z_0^{1,2} - Z_0^{vtx}| < 5$ cm.

An additional set of requirements described in the following sections is applied in order to further clean-up the event selection, by removing muons from cosmic rays, electrons originated from photon conversion, $\mu \rightarrow \tau_h$ fakes and Z→ee background events.

Finally, each event is classified by counting the number of calorimeter jets identified in the final state and not matched to any other reconstructed lepton.

The complete list of event cuts is reported in table 6.14.

Event Selection Summary	
Variable	Cut
pass lep+track trigger	YES
goodrun list	YES
electron/muon	$E_T(p_T)$> 10 GeV(/c)
hadronic τ	p_T >15 GeV/c
3^{rd} lepton	VETO
$Q_1 \times Q_2$	= -1
ΔR_{12}	> 0.4
$\lvert Z_0^1 - Z_0^2 \rvert$	<5 cm
$\lvert Z_0^{vtx} \rvert$	<60 cm
$\lvert Z_0^{1,2} - Z_0^{vtx} \rvert$	< 5 cm
cosmics	VETO
conversions	VETO
$\mu \rightarrow \tau_h$ removal	YES
Z→e^+e^- removal	YES

Table 6.14: Event selection cuts.

6.5.1 Cosmic rays and conversions removal

Cosmic rays

Cosmic rays consisting in high-p_T muons which interact with the detector simultaneously with a bunch crossing, can easily fire the muon triggers and contaminate the data sample in the τ_h+τ_μ channel.

A cosmic tagger has been implemented within the CDF offline code [90], by exploiting the muon chambers, COT, TOF, calorimeter energy and timing information; the algorithm provides a nearly 100% rejection of cosmic ray events, with a negligible loss of signal efficiency.

Conversions

Electron originating from photon conversions, $\gamma \rightarrow e^+e^-$, represent a significant background for the $\tau_h\tau_e$ samples, which is reduced by applying the standard CDF conversion tagger [4].

The algorithm looks for two opposite sign tracks (one of them belonging to the identified electron) and requires that $\Delta\cot(\theta)$ and the distance at the closest approach between them, D_{xy}, are less than the specific values reported in table 6.15. The efficiency of the conversion veto is also shown: about 65% of the electrons originating from a conversion is actually rejected.

Photon Conversion tagger		
$\Delta\cot(\theta)$	D_{xy}	ε
$\leq$0.02	$\leq$0.1 cm	64.6$\pm$0.4%

Table 6.15: Conversion cuts and removal efficiency in data [4].

6.5.2 $\mu \rightarrow \tau_h$ fake removal

Muons can mimic the τ signature when they release a not negligible amount of energy in the calorimeter volume: in this case the muon track appears to be associated to a small calorimeter cluster and can be easily mis-identified as a 1-prong τ.

We can consider two different situations: in some cases a muon, before hitting the external drift chambers and exiting the detector, may interact with the calorimeter and then deposit a certain amount of energy in the hadronic compartments, thus failing the muon ID calorimeter cuts; in other cases a muon can be subject to bremsstrahlung and radiate photons, which release energy in the adjacent electromagnetic towers, causing the muon to fail the EM and the isolation requirements.

If a τ candidate is found to be close to such partially identified muons, ΔR <0.4, the event is rejected.

6.5.3 Z→e$^+$e$^-$ veto

Most of the events containing electrons which could be misidentified as hadronic τ's are rejected by the tight τ preselection requirement $\xi \geq 0.2$, which is defined in section 6.4.3.

We apply an additional cut to remove most of the residual Z→e$^+$e$^-$ background contamination affecting the $\tau_e\tau_h$ final state: events containing 1-prong τ's with a low hadronic energy fraction and a lepton+τ invariant mass consistent with that of a Z boson (see table 6.16), are vetoed. This selection removes about 70% of Z→e$^+$e$^-$ events, with an efficiency of about 95% on the Higgs signal.

Z→ee veto	
event category	$\tau_e\tau_h$
τ category	1-prong
ξ	≤ 0.4
$M_{e-\tau_h}$	≥ 80, $\leq$110 GeV/c^2

Table 6.16: Z→e$^+$e$^-$ veto cuts.

6.6 Signal and control regions

When applying the complete baseline selection summarized in table 6.14, we observe a total of 8451 events in 2.3 fb^{-1} of integrated luminosity.

As a preliminary step, we classify each event by counting the number of identified jets in the final state, and roughly estimate for each category the expected background(B) and signal(S) yields, as shown in table 6.17.

It is useful to report the signal over background ratio S/B and the significance s, defined as $S/\sqrt{B}$. This variable is related to the probability that the expected amount of signal events is greater than the statistical fluctuation of the background.

Unlike S/B, the significance is dependent on the amount of data considered. For example, a doubling of the integrated luminosity corresponds to an increase of significance of $\sqrt{2}$.

Given the significance estimated for each jet multiplicity value, we decided to focus the Higgs search on the *1 jet* and *≥2 jets* channels, and exploit the events with no jets in the final state to test the reliability of the background modeling: for this purpose a set of control regions has been defined, as it will described in details in the next chapter.

Jet mult.	0	1	$\geq$2
observed events	7320	965	166
expected S	$\sim$2	$\sim$0.75	$\sim$0.5
expected B	$\sim$7500	$\sim$950	$\sim$150
S/B	0.0003	0.0008	0.003
$S/\sqrt{B}$	0.023	0.024	0.041

Table 6.17: Number of observed events for 2.3 fb^{-1} of integrated luminosity and the corresponding estimated H$\rightarrow \tau\tau$ signal (at m$_H$=120 GeV/c^2) and background contribution.

Analysis channels		
0 jets	1 jet	$\geq$2 jets
control region	signal region	signal region

Table 6.18: Signal and control regions, according to the event jet multiplicity.

Chapter 7

Signal and background modeling

"Le savant doit ordonner; on fait la science avec des faits comme une maison avec der pierres; mais une accumulation de faits n'est plus science qu'un tas de pierres n'est une maison."
Jules Henri Poincarè

In this chapter we give an overview of all the physics processes which satisfy the event selection and contribute to the signal and background yields of our Higgs search. In particular, we describe the techniques, based on simulations and data-driven methods, adopted to model the the different physics contributions.

Events with no reconstructed jets in the final state are used to build three different orthogonal control samples, where the agreement between data and expectation is verified, thus providing a test of the goodness of the background modeling procedure.

7.1 Monte Carlo Simulation

Monte Carlo (MC) simulations are widely used in particle physics to estimate the fraction of events of a certain physics process which lie within the detector acceptance and satisfy the selection requirements, and to evaluate object reconstruction and identification efficiencies or algorithms' performances.

For a typical analysis the production and validation of several millions of simulated events is needed, and this process usually requires big human efforts and a huge amount of CPU time.

The full simulation procedure can be divided into three separate steps:

1. generation of the physics events of interest in the proton-antiproton interaction;
2. simulation of the parton showering and the hadronization process;

3. simulation of the detector geometry, material and response to the interacting particles passing through it;

The output of the detector simulation is stored in files with the same format of real data and thus it can be studied with the same analysis code, by applying the usual reconstruction algorithms, the event selection, etc.

7.1.1 Event generator

Different generators are employed at CDF. The Higgs signal and several electroweak backgrounds processes (WW,WZ,ZZ and $t\bar{t}$) are modeled with PYTHIA [91], which uses LO matrix elements for the description of the hard parton scattering inside the colliding protons and antiprotons.

Background arising from W and Z bosons, produced in association with jets, are estimated using ALPGEN generator [92], which is specifically designed and finalized to give a reliable modeling of multiparton events in hadron collisions. This is the case of the final states containing radiated quarks and gluons. For this specific purpose, this generator is organized in such a way that several samples, with a different number of partons accompanying the hard scattering process[1], are produced separately. The generated independent sets of events, one for each Z(W)+n partons subprocess, are then added together by applying the proper specific cross sections.

Both PYTHIA and ALPGEN use the CTEQ5L setting for the parton distribution functions(PDFs) [93][2]

7.1.2 Parton showering and hadronization

PYTHIA is not only an event generator, but it is also used to describe the parton showering, gluon radiation and hadronization of color-charged partons. The showering procedure generates initial and final state radiation; gluons are allowed to decay to quark pairs, thus increasing the number of final state particles. The effect coming from the pile-up of multiple $p\bar{p}$ interactions in the same bunch crossing and from the interaction of spectator quarks (beam-beam remnants, BBR) are also included at this stage of the simulation.

It is important to note that τ leptons are treated as stable particles by PYTHIA, and their decay is simulated by another package, named TAUOLA [94].

Table 7.1 summarizes the MC models which have been used to generate all the signal and background processes employed in the analysis.

[1] When building the different n-parton subprocesses, a cut on the generated parton p_T is applied: in CDF Monte Carlo, this cut is set to $p_T^{\rm gen} \geq$15 GeV/c. This means, for example, that in a Z+3p sample there are exactly three partons with p_T above 15 GeV/c. This configuration is appropriate and does not introduce any bias in the analysis, since the transverse energy cut applied in the jet reconstruction is well above the threshold of 15 GeV, as described in section 6.4.4.

[2] A parton distribution function is defined as the probability density to find a parton with a certain longitudinal momentum fraction x at the momentum scale Q^2.

MC samples	
Process	Generator
$gg \rightarrow$H	PYTHIA
WH	PYTHIA
ZH	PYTHIA
qqH	PYTHIA
$Z/\gamma^* \rightarrow \tau\tau$ + np	ALPGEN
$Z/\gamma^* \rightarrow ee$ + np	ALPGEN
$Z/\gamma^* \rightarrow \mu\mu$ + np	ALPGEN
W$\rightarrow e\nu$ + np	ALPGEN
W$\rightarrow \mu\nu$ + np	ALPGEN
W$\rightarrow \tau\nu$ + np	ALPGEN
$t\bar{t}$	PYTHIA
WW	PYTHIA
WZ	PYTHIA
ZZ	PYTHIA

Table 7.1: MC sample generators used in the analysis for the different processes. *np* refers to the number of partons accompanying the hard scattering process.

7.1.3 Detector simulation

In the standard CDF II simulation, the detector geometry and material are modeled using the version 3 of the GEANT package [95]. GEANT receives as input from the event generator the position coordinates, the four-momenta and the identities of all particles produced in the simulated collisions, which have enough lifetime to exit the beam pipe and interact with the detector volume. All kind of interactions, like bremsstrahlung, multiple scattering, nuclear interactions, photon conversions are fully simulated.

Specific packages substitute GEANT for some sub-detectors: GFLASH [96] simulates the calorimeter response, while the drift time within the COT is reproduced using the GARFIELD package [97], both tuned on data.

Beam position, instantaneous luminosity profile and detector conditions are taken into account by generating run-dependent MC samples.

7.2 Sample composition

The Higgs processes discussed in this thesis are those described in section 5.1.3. Several MC samples for each process are generated for the hypothetical masses ranging from 100 GeV/c^2 to 150 GeV/c^2, in steps of 5 GeV/c^2.

The different sources of background which pass the requirements of the analysis

selection are divided into two types: the main physical irreducible processes (Drell-Yan, diboson, $t\bar{t}$ production), which are described by means of MC simulations, and the background due to calorimeter jets erroneously identified as leptons, which are modeled by a data-driven procedure.

The expected number of events for each signal and background contribution derived from MC samples is given by the following formula:

$$N^i = \sigma^i \times A^i \times \varepsilon_{\text{trig}} \times \varepsilon_{\text{ID}} \times \varepsilon_{\text{vtx}} \times \int Ldt \tag{7.1}$$

where the index i refers to the specific process under investigation. The variables which enter in the expression are:

- σ: the cross section of the simulated physics process, given from theoretical calculations or experimental measurements;

- A: the acceptance, both due to the detector geometry and the event selection efficiency. It is calculated in the MC samples by using the formula:

$$A^i = \frac{N^i_{\text{sel}}}{N^i_{\text{gen}}(|Z_0^{P.V.}| \leq 60\,\text{cm})} \tag{7.2}$$

  where N^i_{gen} is the number of generated events passing the good run list and with a primary vertex within 60 cm from the center of the detector, while N^i_{sel} is the number of events which satisfy all the selection requirements.

- $\varepsilon_{\text{trig}}$: the effective trigger efficiency, given by the product of the track (corresponding to the τ leg) and the other lepton (electron or muon) trigger efficiency; typical values, for each leg, are of the order of 90-95%: see section 6.3.1 for additional details about the estimation procedure.

- ε_{ID}: the resulting identification scale factor for the two identified leptons, $\varepsilon_{\text{ID}} = \varepsilon_{\text{IDlep}} \times \varepsilon_{\text{IDtau}}$, which takes into account the possible differences in the lepton identification performances between the simulated and the collected data. For the evaluation of $\varepsilon_{\text{IDlep}}$ we refer to [82, 83, 98, 99] and to the following updated studies performed by the CDF collaboration, separately for the medium-p_T and the high-p_T leptons; for $\varepsilon_{\text{IDtau}}$, we apply the procedure described in 7.6.2. This scale factors generally range between 0.9 and 1.0.

- ε_{vtx}: the run-dependent correction factor (around 0.95) for the z vertex position requirement $|Z_0^{vtx}|$ <60 cm [100];

- $\int Ldt$: the integrated luminosity of the data samples, with the appropriate good run list applied.

It is possible that some MC events pass the final state selection because of the presence of generated quark/gluon jets which are misidentified as leptons. Given that the jet$\rightarrow \tau_h$ and jet$\rightarrow \mu$/e fake background is already estimated through data samples, as explained in section 7.5, we have to avoid double counting by considering only the MC contributions of real leptons.

This is done by requiring that each reconstructed lepton is geometrically close to the right generated particle, $\Delta R_{l^{\rm rec}\_l^{\rm gen}} \leq 0.4$. In the case of the hadronic τ, the matching is done between the direction of the reconstructed seed track and the direction of the generated visible τ decay products, that is the vector difference between the original τ and the emitted neutrino directions.

7.3 Signal processes

$\mathbf{M_H}$	Acceptance % ($\mathbf{N_{exp}}$)			
(GeV/c²)	**0 jet**	**1 jet**	**≥2 jets**	**TOT**
100	1.10(2.996)	0.29(0.792)	0.06(0.167)	1.45 (3.955)
105	1.17(2.723)	0.31(0.732)	0.07(0.168)	1.55 (3.623)
110	1.23(2.439)	0.33(0.656)	0.08(0.152)	1.63 (3.247)
115	1.30(2.165)	0.36(0.610)	0.08(0.138)	1.74 (2.912)
120	1.36(1.866)	0.39(0.535)	0.09(0.129)	1.85 (2.530)
125	1.41(1.549)	0.42(0.456)	0.10(0.111)	1.93 (2.115)
130	1.43(1.207)	0.45(0.374)	0.10(0.087)	1.98 (1.669)
135	1.49(0.931)	0.47(0.295)	0.11(0.068)	2.08 (1.294)
140	1.54(0.675)	0.49(0.215)	0.12(0.054)	2.15 (0.944)
145	1.54(0.452)	0.50(0.147)	0.12(0.036)	2.17 (0.635)
150	1.61(0.292)	0.52(0.094)	0.13(0.024)	2.26 (0.409)

Table 7.2: MC samples used to estimate the gluon fusion process, along with the acceptance and the corresponding number of expected events, for the different Higgs masses and number of jets in the final state.

The Higgs production cross sections and decay branching ratios used in this analysis are summarized in tables 3.1 and 3.2. The acceptance for each process and the expected number of events in 2.3 fb^{-1} ($N_{\rm exp}$), obtained by applying all the baseline selection cuts and the proper correction factors, trigger and identification efficiencies, are shown in tables 7.2, 7.3, 7.4 and 7.5, separately for the different Higgs masses and the number of reconstructed jets in the final state.

As can be seen, the overall acceptance is between 1.30% and 2.50% for all the signal processes, with different relative contributions according to the jet multiplicity. We note that a very small fraction of gluon fusion events fall in the ≥2 jet channels, because in this case jets do not arise from the leading order interaction, but from initial

M_H (GeV/c^2)	Acceptance % (N_{exp}) 0 jet	1 jet	≥2 jets	TOT
100	0.31(0.133)	0.39(0.164)	0.62(0.265)	1.31 (0.563)
105	0.33(0.117)	0.40(0.145)	0.66(0.237)	1.39 (0.499)
110	0.35(0.104)	0.42(0.126)	0.70(0.208)	1.46 (0.438)
115	0.35(0.087)	0.44(0.108)	0.74(0.181)	1.53 (0.377)
120	0.37(0.072)	0.47(0.091)	0.77(0.150)	1.60 (0.314)
125	0.38(0.058)	0.47(0.072)	0.79(0.121)	1.64 (0.252)
130	0.39(0.045)	0.50(0.057)	0.83(0.096)	1.72 (0.198)
135	0.41(0.034)	0.51(0.042)	0.86(0.071)	1.78 (0.148)
140	0.41(0.023)	0.51(0.029)	0.88(0.050)	1.80 (0.102)
145	0.43(0.016)	0.54(0.020)	0.93(0.034)	1.91 (0.070)
150	0.42(0.009)	0.54(0.012)	0.96(0.021)	1.92 (0.047)

Table 7.3: MC samples used to estimate the WH process, along with the acceptance and the corresponding number of expected events, for the different Higgs masses and number of jets in the final state.

M_H (GeV/c^2)	Acceptance % (N_{exp}) 0 jet	1 jet	≥2 jets	TOT
100	0.40(0.098)	0.37(0.091)	0.67(0.167)	1.43 (0.356)
105	0.42(0.088)	0.37(0.079)	0.74(0.156)	1.53 (0.322)
110	0.44(0.079)	0.39(0.069)	0.77(0.137)	1.61 (0.285)
115	0.45(0.066)	0.42(0.061)	0.80(0.117)	1.66 (0.244)
120	0.48(0.056)	0.42(0.050)	0.83(0.099)	1.73 (0.205)
125	0.48(0.045)	0.43(0.041)	0.86(0.081)	1.78 (0.166)
130	0.50(0.036)	0.46(0.033)	0.91(0.064)	1.88 (0.133)
135	0.51(0.027)	0.46(0.024)	0.95(0.049)	1.93 (0.099)
140	0.53(0.019)	0.48(0.017)	1.00(0.036)	2.01 (0.071)
145	0.55(0.013)	0.48(0.011)	1.01(0.023)	2.04 (0.048)
150	0.55(0.008)	0.50(0.007)	1.05(0.015)	2.10 (0.030)

Table 7.4: MC samples used to estimate the ZH process, along with the acceptance and the corresponding number of expected events, for the different Higgs masses and number of jets in the final state.

state radiation. However, due to the higher production cross section, the event yield is comparable to the other processes.

At the reference hypothetical Higgs mass of 120 GeV/c^2 the expected numbers of signal events for 2.3 fb^{-1} are summarized in table 7.6.

M_H (GeV/c^2)	Acceptance % (N_{exp}) 0 jet	1 jet	≥2 jets	TOT
100	0.15(0.021)	0.64(0.093)	0.89(0.130)	1.68 (0.245)
105	0.15(0.020)	0.68(0.091)	0.93(0.125)	1.76 (0.237)
110	0.16(0.020)	0.70(0.086)	1.02(0.125)	1.88 (0.231)
115	0.17(0.018)	0.75(0.080)	1.05(0.111)	1.97 (0.210)
120	0.17(0.016)	0.78(0.070)	1.10(0.099)	2.05 (0.184)
125	0.20(0.015)	0.79(0.061)	1.13(0.086)	2.13 (0.162)
130	0.19(0.012)	0.83(0.051)	1.16(0.071)	2.18 (0.134)
135	0.21(0.010)	0.89(0.042)	1.22(0.057)	2.32 (0.109)
140	0.22(0.008)	0.90(0.031)	1.24(0.042)	2.37 (0.080)
145	0.23(0.005)	0.94(0.022)	1.26(0.030)	2.43 (0.057)
150	0.22(0.003)	0.96(0.014)	1.31(0.020)	2.49 (0.037)

Table 7.5: MC samples used to estimate the VBF process, along with the acceptance and the corresponding number of expected events, for the different Higgs masses and number of jets in the final state.

M_H=120 GeV/c^2	
N_{jets}	N_{sig}
0	2.010
1	0.746
≥2	0.477
TOT	3.233

Table 7.6: H→ $\tau\tau$ signal yield at M_H = 120 GeV/c^2, after the event selection is applied, for 2.3 fb^{-1}.

7.4 Background from MC simulations

Table 7.7 lists all the cross section values which have been used to calculate the number of estimated events for each background modeled by MC simulations.

For the Z+jets process, we refer to the inclusive Z→ll CDF measurement [101][3].

Cross sections for WW, WZ and ZZ production [102] are computed at the NLO with the MCFM package [36]. NLO calculations are performed also to predict $t\bar{t}$ events, using m_t = 173±1.2 GeV/c^2 [103]. In both cases the MSTW2008 PDF [104] setting have been used.

[3]The cross section value shown in table 7.7 refers to the reconstructed dilepton invariant mass in the 66< M <116 GeV/c^2 range.

process	cross section(pb)		
Z/γ^*+jets	254.9	$\pm$	5.6
WW	11.34	$\pm$	0.68
WZ	3.22	$\pm$	0.19
ZZ	1.20	$\pm$	0.07
$t\bar{t}$	7.04	$\pm$	0.70

Table 7.7: Cross section for background contributions estimated with MC samples.

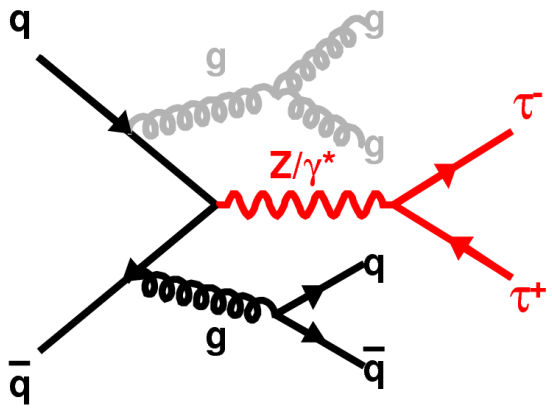

Figure 7.1: Feynman diagram for the principal source of MC-derived background which pass the event selection, Z$\rightarrow \tau\tau$.

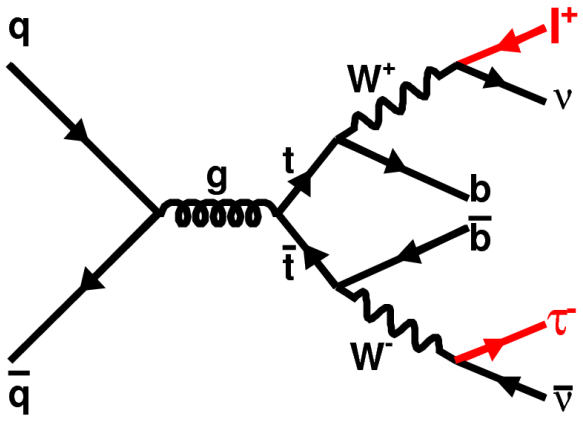

Figure 7.2: Feynman diagram for the $t\bar{t}$ background, which is relevant in the final state containing or more two reconstructed jets.

7.4.1 Drell Yan normalization

The Drell-Yan process represents the main irreducible background which enters in our data samples. The choice of ALPGEN as a MC generator is driven by the need to have

a good understanding of the jet multiplicity. This is crucial in our analysis, where data are divided into different subsamples by counting the number of reconstructed jets in the final state.

However, even if the parton multiplicity is correctly included in the MC, some discrepancies between the expected and the observed event rates may still appear, e.g. because of a non perfect correspondence between the jet definition at offline level and the clustering procedure applied in the MC after the PYTHIA showering.

As a matter of fact, when the ALPGEN n-parton events are generated, each parton is also required to be geometrically matched to a showered particle cluster ($\Delta R \leq 0.4$), with a lower E_T threshold which is set to 15 GeV: this value is generally different from the offline jet energy measurement, where there is the effect of the particles interaction with the detector as well. Therefore, a non perfect modeling of the jet energy spectrum around the offline energy threshold, may arise, thus leading to event migrations between the different jet-multiplicity bins.

In order to check the level of agreement between ALPGEN MC and data, we reconstruct Z events at the mass peak, by looking at the di-electron and di-muon final states.

We start by selecting events with the following high p_T lepton trigger paths:

- ELECTRON_CENTRAL_18: it requires events with at least one central electron with $E_T \geq 18$ GeV;
- MUON_CMUP18: it selects events with at least one muon candidate with hits in both the CMU and CMP chambers and $p_T \geq 18$ GeV;
- MUON_CMX18: it selects events with at least one muon candidate with hits in the CMX chamber and $p_T \geq 18$ GeV.

The detailed description of the cuts applied at each level of these trigger paths is reported in appendix A.

We then require two reconstructed opposite charged electrons or muons with a resulting dilepton invariant mass consistent with that of the Z boson ($70 \leq M_Z \leq 110$ GeV/c^2); fake events (with jet$\rightarrow$e or jet$\rightarrow \mu$ misidentification) and other physics processes have a negligible contribution and we can thus directly compare data to Z$\rightarrow$ee/$\mu\mu$ simulations.

Before looking at the jet multiplicity, we first compared ALPGEN expectations with PYTHIA in the inclusive dataset: we found that PYTHIA well reproduced the observed event yield in both the ee/$\mu\mu$ cases, while ALPGEN inclusive acceptance is smaller of about 4.5% than PYTHIA, thus resulting in an underestimation of data. We correct for this effect by applying an overall scale factor of 1.045 to the acceptance of all Drell-Yan ALPGEN processes.

The invariant mass distributions for the different jet multiplicity channels, obtained after applying the acceptance correction, are shown in figures 7.3 and 7.4. Results are also reported in table 7.8.

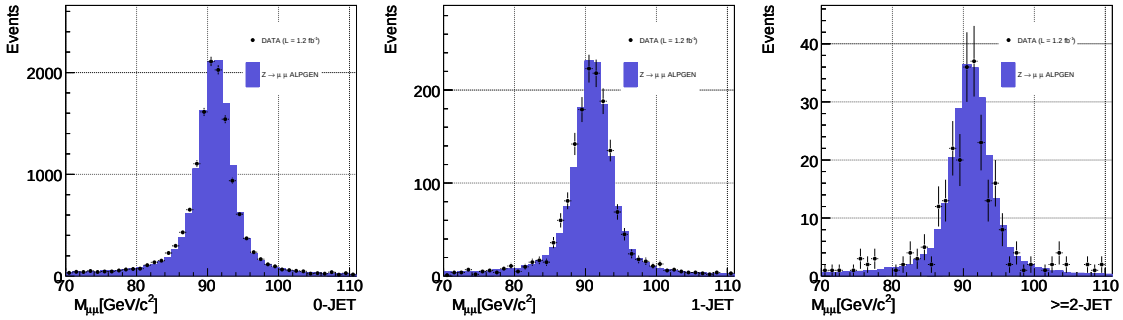

Figure 7.3: Di-muon invariant mass in the Z mass peak, for 0 jet, 1 jet and $\geq$2 jets in the final state.

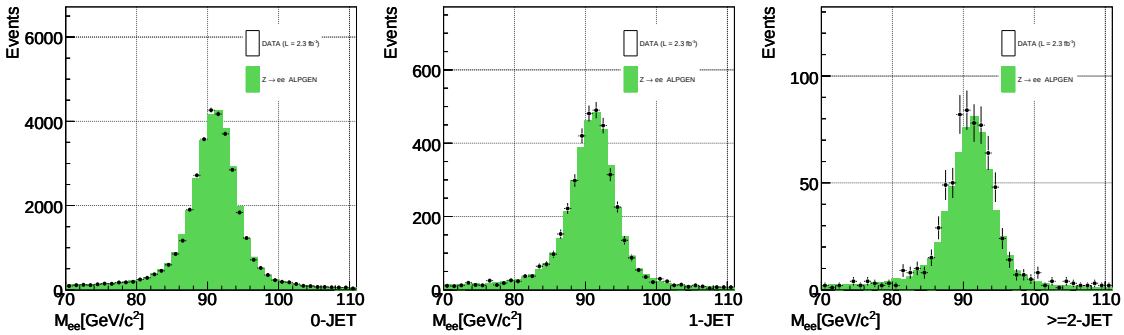

Figure 7.4: Dielectron invariant mass in the Z mass peak, for 0 jet, 1 jet and $\geq$2 jets in the final state.

The agreement between data and MC is in general good: we verified that the residual discrepancy of 10% in the "$\geq$2 jet" bin of the di-electron selection is fully covered by the systematic uncertainty related to the jet energy scale correction.

7.5 Background from misidentified leptons

Three different processes can contribute to the background coming from misidentified leptons:

- γ + **jet**: it is relevant for the $\tau_e\tau_h$ channel, when the jet fakes a hadronic τ and an electron from photon conversion is not removed by the conversion tagger.
- **QCD multijet**: when one jet fakes a τ and another one is reconstructed as an electron or a muon as shown in figure 7.5; it represents the most significant source of background coming from misidentified leptons.
- **W**$\rightarrow l\nu$**+jets**: when one of the jets in the final state is misidentified as a hadronically decaying τ.

Category	Jets	MC	data	diff (%)
$\mu\mu$	0	13814.4	13877	-0.4%
	1	1568.8	1607	-2.4%
	$\geq$2	258.3	252	+2.5%
	TOT	15641.6	15736	-0.6%
ee	0	34269.4	34387	+0.7%
	1	3925.2	3996	-1.8%
	$\geq$2	645.5	724	-10.8%
	TOT	39200.1	39107	+0.2%

Table 7.8: Expected and observed event counts in the dielectron (period 0-13) and dimuon final states (period 0-8).

In order to estimate these contributions we apply a method based on a combination of same sign (SS) data events ($Q_1 \times Q_2 = 1$) and W+jets MC samples.

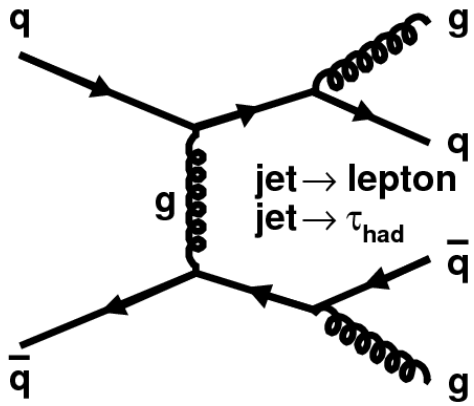

Figure 7.5: Feynman diagram for the QCD multijet process.

7.5.1 Same sign data method

We model the first two processes listed in the previous section, by exploiting the very little or no correlation which is expected between the charges of the two reconstructed lepton candidates. As a consequence of this assumption, the number of opposite sign(OS) events should approximately be equal to the number of same-sign(SS) events, as it has been widely verified in previous similar searches [3], by looking at QCD-enriched data samples defined in the lepton calorimeter isolation sidebands.

Since the analysis selection requires two opposite charged leptons in the final state, we recover the remaining SS data events and use them to model these two categories

of jet$\rightarrow \tau_h$ fake background.

Rates and event kinematic shapes are evaluated separately:

1. the background rate is predicted by simply counting the number of data events which pass the same criteria of the baseline analysis selection summarized in table 6.14, except for the $Q_1 \times Q_2$ cut, which is reversed (*tight SS* control region);

2. the kinematic shapes are estimated by using the events which fall in a *loose SS* control region, obtained by replacing the BDT-based τ ID with the loose selection described in table 6.12: provided that the different τ identification does not change the event topology, as shown in figures 7.6 and 7.7, the higher statistics of this sample allows us to better reproduce the tails of the distributions, in particular for the high-jet multiplicity channels.

This method has the great advantage to be entirely based on data and not on MC simulations, which are known to be unreliable for QCD processes.

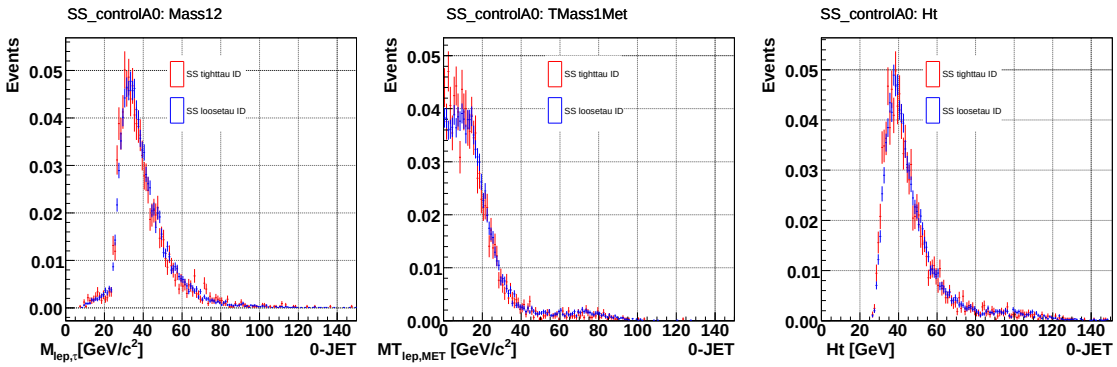

Figure 7.6: From the left to the right: normalized distributions for the lepton+τ invariant mass,lepton+$\not\!\!E_T$ transverse mass and $\sum E_T$, for events with no jets in the final state. A comparison between loose and tight SS data is shown.

W+jets background correction

The hypothesis of symmetry between SS and OS events is true for the QCD multijet and γ+jet background, but it is not valid in the case of W+jets, where a correlation may appear between the charge of the lepton coming from the W boson and the charge of the outgoing quark, which generates the jet faking the hadronic τ.

This feature can be understood by looking at one typical Feynman diagram of this process, shown in picture 7.8, from which we observe that it is more likely that the jet originated from the hadronization of the u quark has an opposite charge sign with respect to the accompanying W. The 1-prong or 3-prong τ reconstructed on this jet will keep memory of this charge correlation, thus leading to a greater number of OS events, $N_{\rm OS}$, compared to SS events, $N_{\rm SS}$.

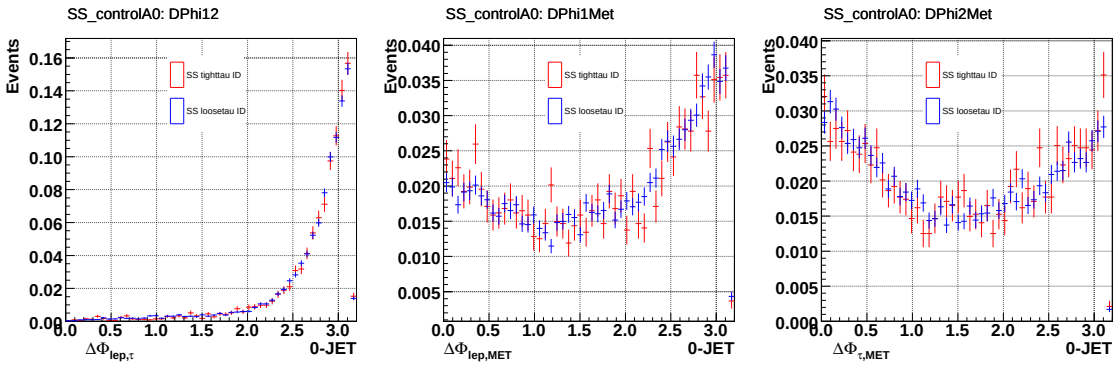

Figure 7.7: From the left to the right: normalized distributions for angular separation in φ between lepton and τ, lepton and $\not\!\!E_T$, τ and $\not\!\!E_T$ for events with no jets in the final state. A comparison between loose and tight SS data is shown.

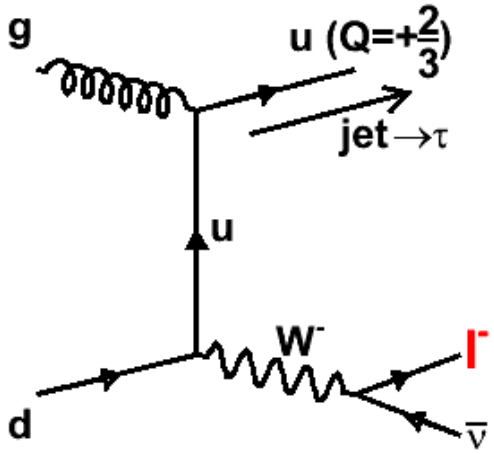

Figure 7.8: A Feynman diagram for the W+jets production, showing the charge correlation between the lepton coming from the W and the outgoing quark which originates a jet.

A fraction of W+jets contribution is already included within the SS data sample; we just need to estimate the extra contribution coming from $N^{W+jets}_{add-on} = N^{W+jets}_{OS}$ - N^{W+jets}_{SS} and add it to the total background prediction. In order to perform this estimate, we apply the procedure described in what follows.

We use the W + n partons ALPGEN MC samples and we compare the event yield estimation to data in a control region defined by the cuts described in table 7.9: The looser τ identification increases the jet$\rightarrow \tau_h$ fake contamination, while the latter two cuts kill almost all QCD background and maximize the W+jets contribution, respectively.

Within this control region, we scale first MC to the LO cross section, following formula 7.1, and we then evaluate the appropriate NLO correction factors by simply taking the ratio between data and W+jets MC predictions, after subtracting the resid-

W+jets enriched selection	
Variable	Cut
loose hadronic τ	see section 6.4.3
$\not\!\!E_T$	$\geq$ 25 GeV if 0 jets
	$\geq$ 30 GeV if 1 jet
	$\geq$ 35 GeV if $\geq$2 jets
$M_T(lep, \not\!\!E_T)$	$\geq$ 40 GeV/c^2;

Table 7.9: Event cuts for a W+jets enriched control region. A further division in a OS and SS selection is applied.

ual physics contributions (mainly Drell-Yan), which are of the order of 10%. This procedure is applied separately for the OS and SS categories and for each jet-multiplicity channel. The correction factor also includes any possible secondary effect due to mismodelings of the OS/SS asymmetry and of the jet$\rightarrow \tau_h$ fake rate and finally provides a good estimation of the W+jets event rate; as can be seen in figures 7.9 to 7.14, after applying the correction factor we find a very fine agreement in the shapes of a large set of kinematical and τ variable distributions, both for the OS and for the SS channels.

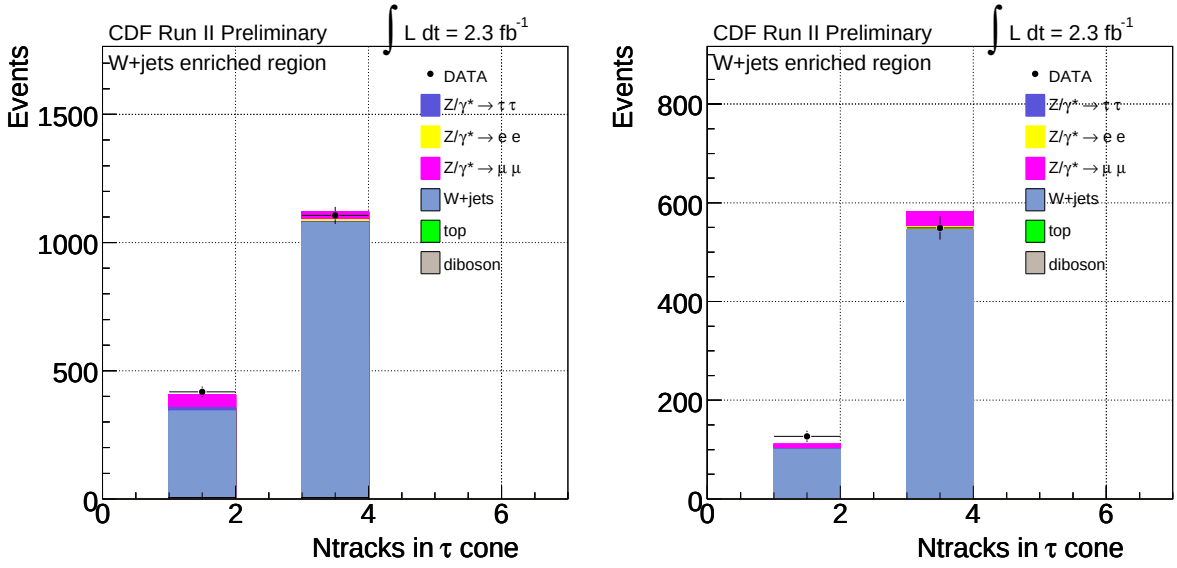

Figure 7.9: Number of tracks in τ signal cone, for the OS(left) and SS(right) W+jets enriched control region defined in the text.

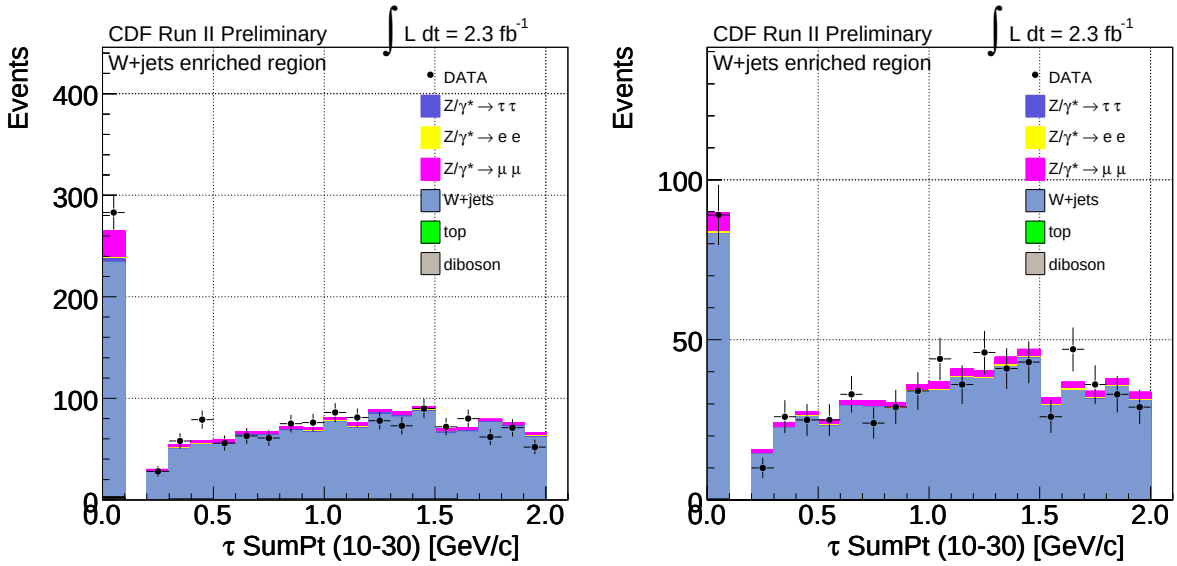

Figure 7.10: Sum of p_T of all tracks in the τ isolation annulus, for the OS(left) and SS(right) W+jets enriched control region defined in the text.

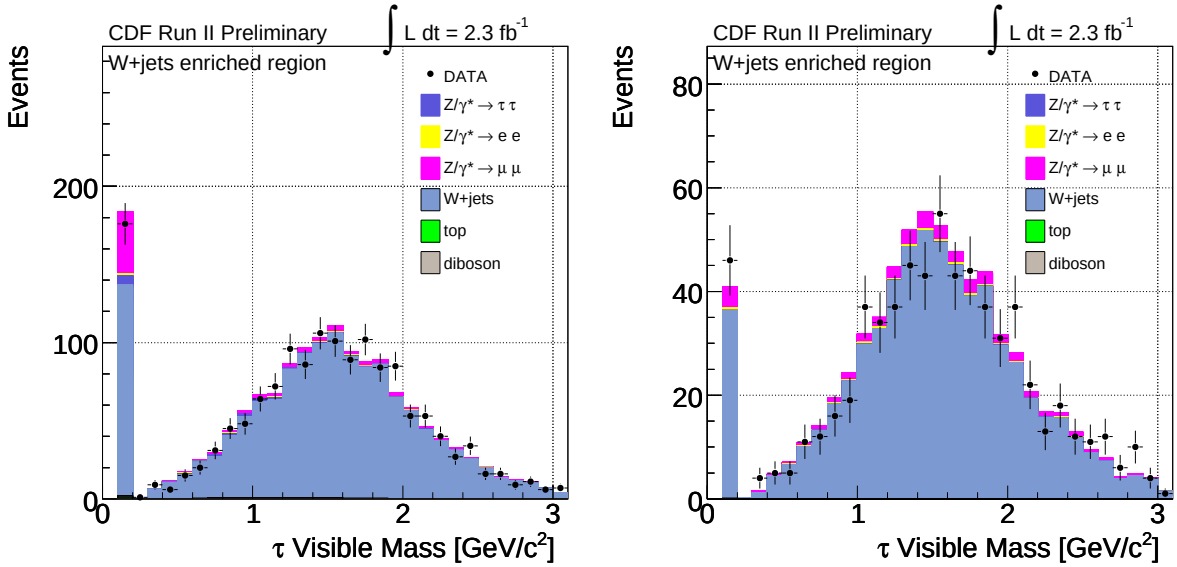

Figure 7.11: τ visible mass, for the OS(left) and SS(right) W+jets enriched control region defined in the text.

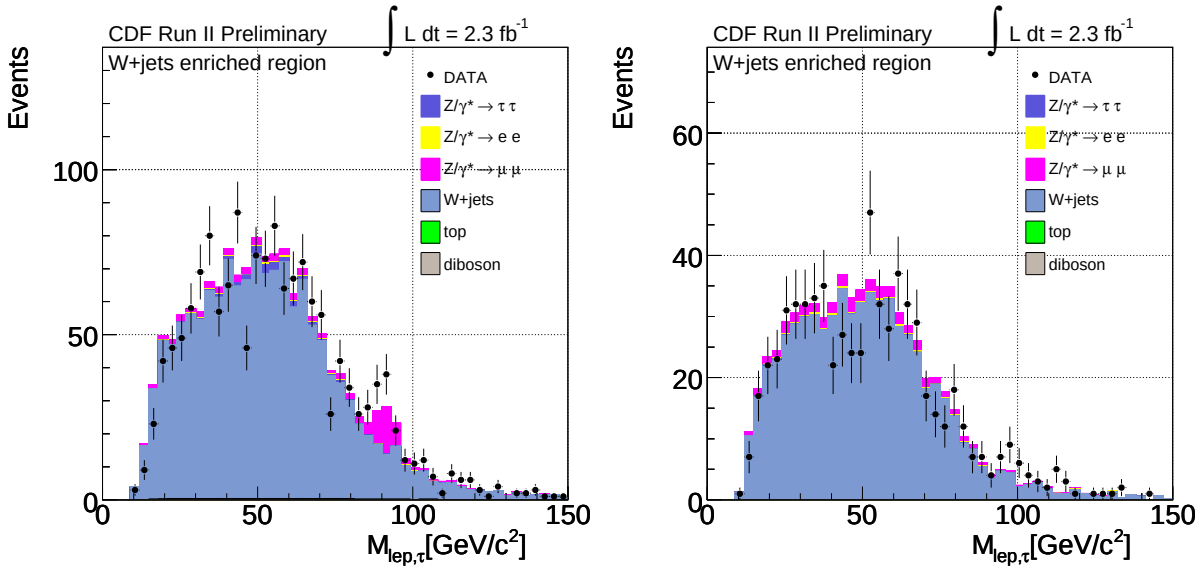

Figure 7.12: Lepton+τ invariant mass, for the OS(left) and SS(right) W+jets enriched control region defined in the text.

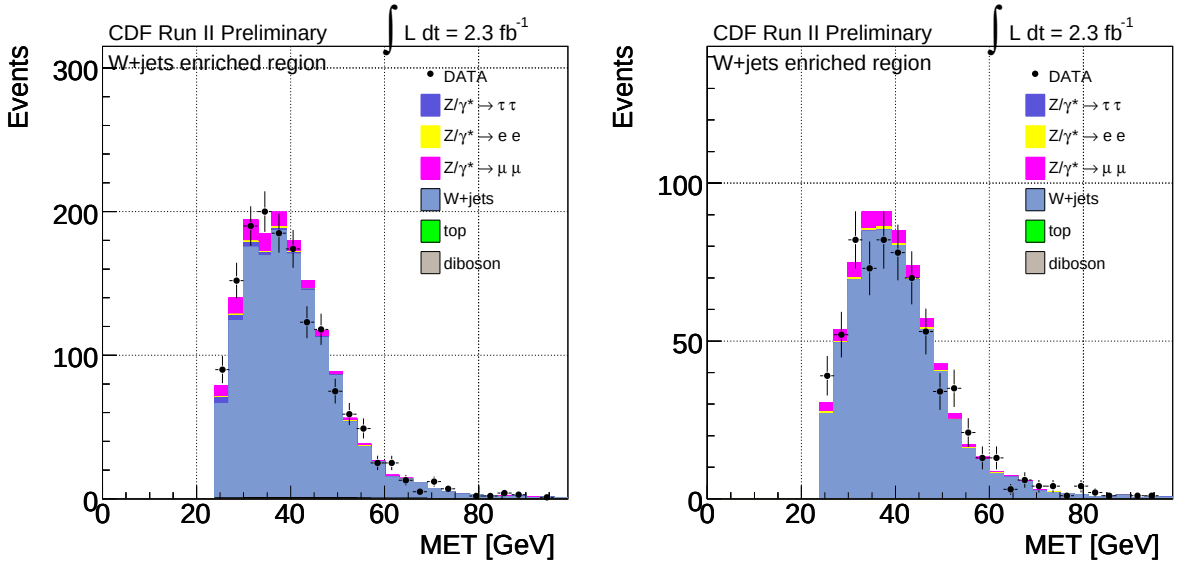

Figure 7.13: Missing transverse energy, for the OS(left) and SS(right) W+jets enriched control region defined in the text.

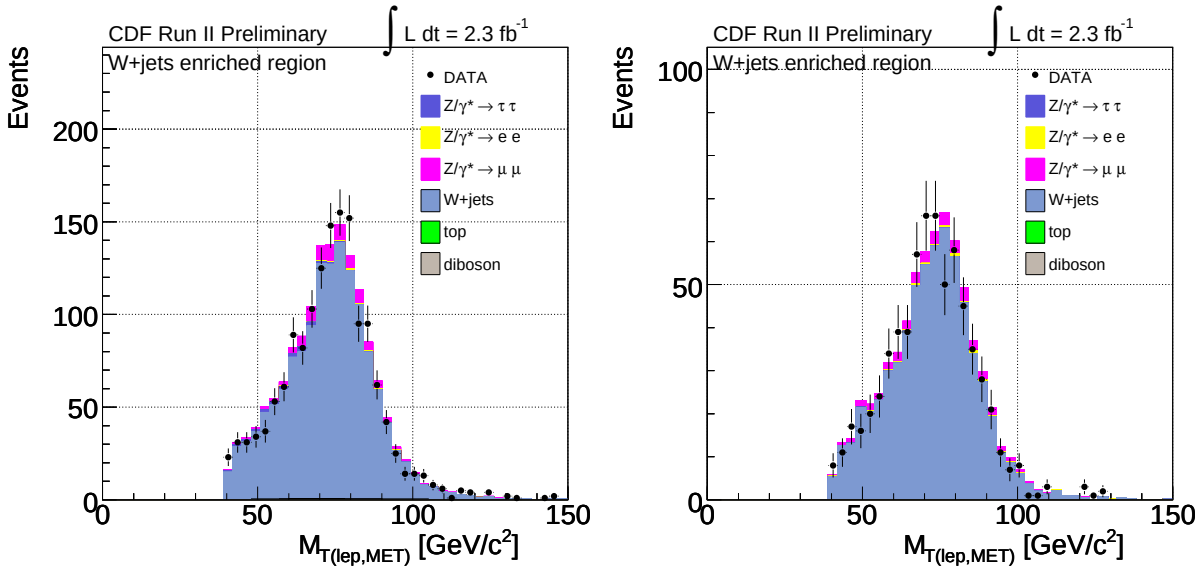

Figure 7.14: Transverse mass of the lepton+$\not\!\!E_T$ system, for the OS(left) and SS(right) W+jets enriched control region defined in the text.

We can now evaluate, directly from MC, the W+jets asymmetry coefficient A_{asym}, defined as N_{OS}/N_{SS} and apply it to get the additional N_{OS}^{W+jets} - N_{SS}^{W+jets} W+jets contribution in the analysis event selection, by means of the following formula:

$$N_{add-on}^{W+jets} = N_{OS}^{W+jets} - N_{SS}^{W+jets} = N_{OS}^{W+jets} - N_{OS}^{W+jets}/A_{asym} = N_{OS}^{W+jets}(1-1/A_{asym}) \quad (7.3)$$

where N_{OS}^{W+jets} is the number of OS events obtained by applying the W+jets MC in the analysis selection.

The application of this procedure is valid only under the assumption that the QCD multijet and γ+jet event yields are actually negligible in the control region where the OS/SS asymmetry coefficient is evaluated. This is verified by removing the $\not\!\!E_T$ cut from the selection, in order to highlight the incoming QCD/γ+jet contribution. This check is shown in figures 7.15, 7.16 and 7.17, where a clear excess in data is visible at low values of $\not\!\!E_T$ and $M_T(lep, \not\!\!E_T)$ and at high values of $\Delta\varphi_{12}$: this is the unambiguous topological signature of di-jet and γ+jet events, where the fake leptons have opposite angular directions.

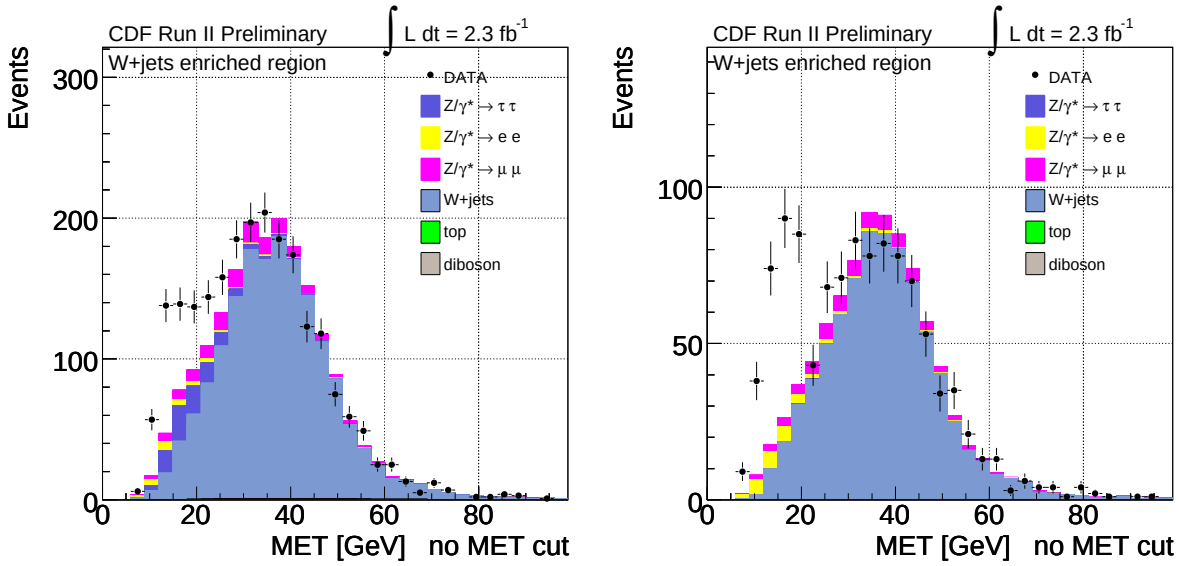

Figure 7.15: Missing transverse energy, for the OS(left) and SS(right) W+jets enriched control region without the $\not\!\!E_T$ cut.

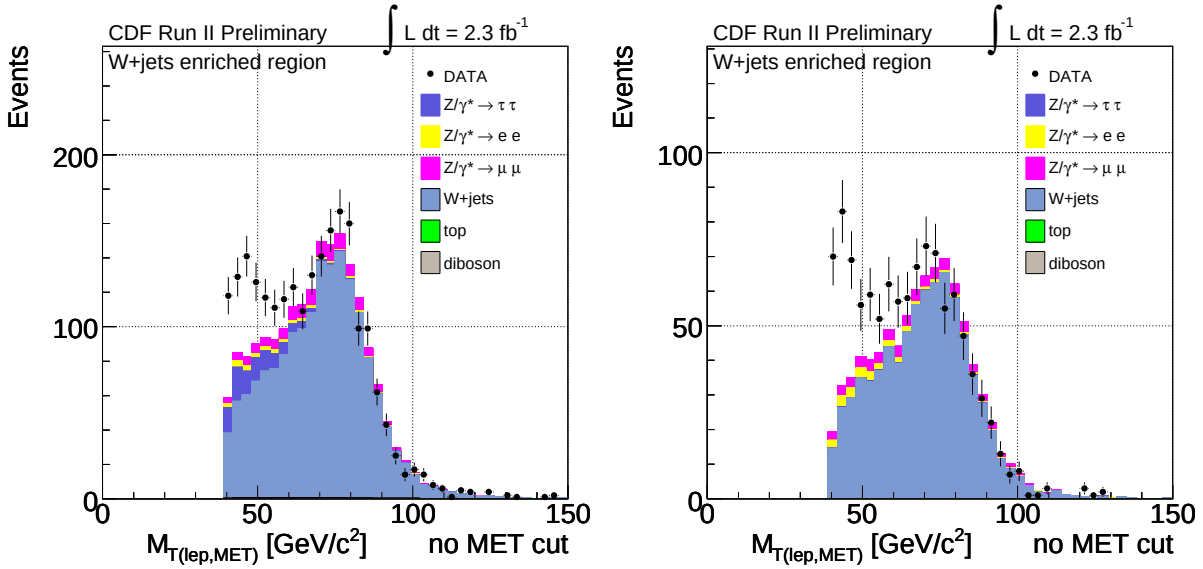

Figure 7.16: Transverse mass of the lepton+$\not\!\!E_T$ system, for the OS(left) and SS(right) W+jets enriched control region without the $\not\!\!E_T$ cut.

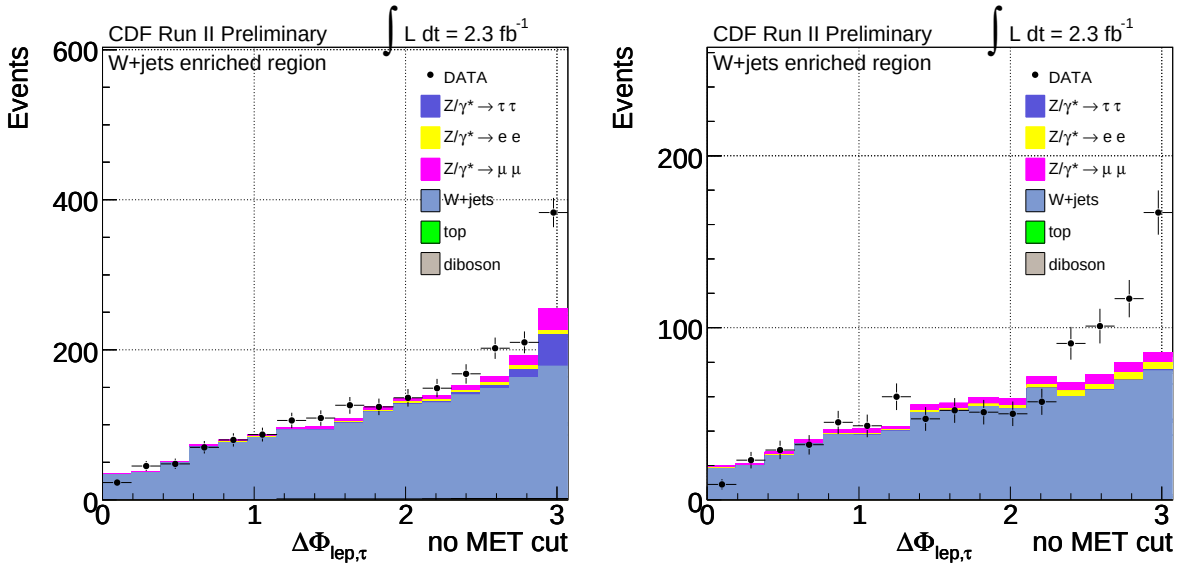

Figure 7.17: Angular separation in φ between lepton and τ, for the OS(left) and SS(right) W+jets enriched control region without the $\not\!\!E_T$ cut.

7.6 Control samples

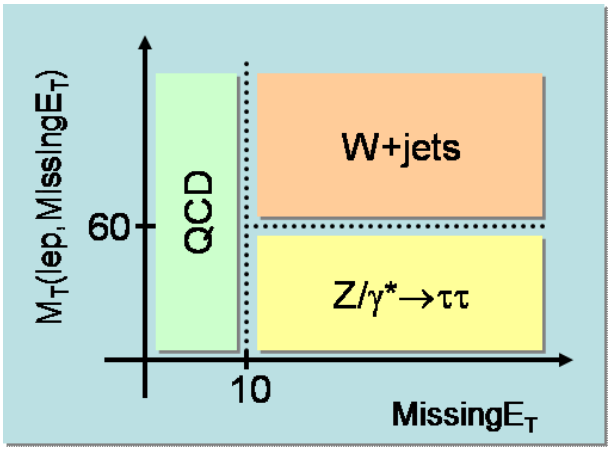

Figure 7.18: Schematic representation of the three control samples obtained by cutting on $\not\!\!E_T$ and M_T(lepton,$\not\!\!E_T$)

As reported in section 7.3, events with no jets in the final state are used to build several control samples, useful to test the physics processes modeling and the event yield normalization procedure. These control samples are orthogonal one from each other and are defined by the following requirements, chosen to maximize the different background contributions:

- QCD region: $\not\!\!E_T \leq 10$ GeV
- Z/$\gamma^* \rightarrow \tau\tau$ region: $\not\!\!E_T \geq 10$ GeV & $M_T(lep, \not\!\!E_T) \leq 60$ GeV/c^2
- W+jets region: $\not\!\!E_T \geq 10$ GeV & $M_T(lep, \not\!\!E_T) \geq 60$ GeV/c^2

A schematic representation of the three control regions in the ($\not\!\!E_T$, $M_T(lep, \not\!\!E_T)$) plane is shown in figure 7.18.

7.6.1 QCD region

In this subset of 0-jet events the QCD contribution is enhanced by requiring the missing transverse energy to be smaller than 10 GeV: we can test here the jet $\rightarrow \tau$ fake modeling based on SS data.

In figure 7.19, 7.20 and 7.21 some significant kinematic and τ distributions are reported. Background and data rates are also summarized in table 7.10, with the systematics uncertainties shown as well (a complete description of systematic errors is given in section 8.5).

We find an acceptable agreement between the overall estimations and the observed number of events. The small residual discrepancy, of the order od 10%, is taken into account by assigning a systematic uncertainty to this background contribution.

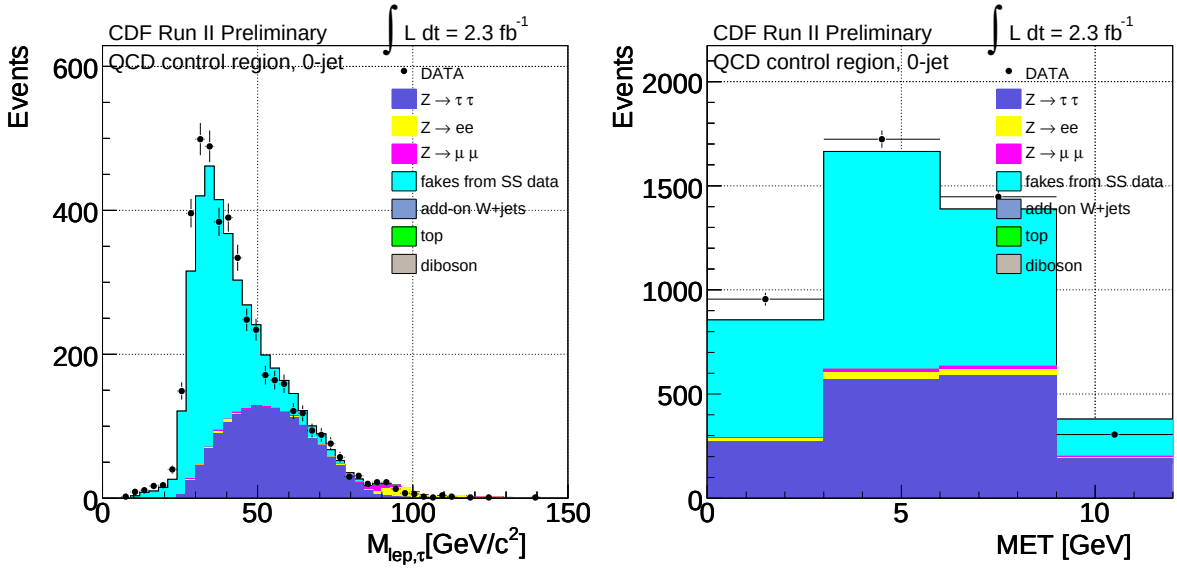

Figure 7.19: Lepton+τ invariant mass (left) and missing transverse energy (right) for the QCD control region.

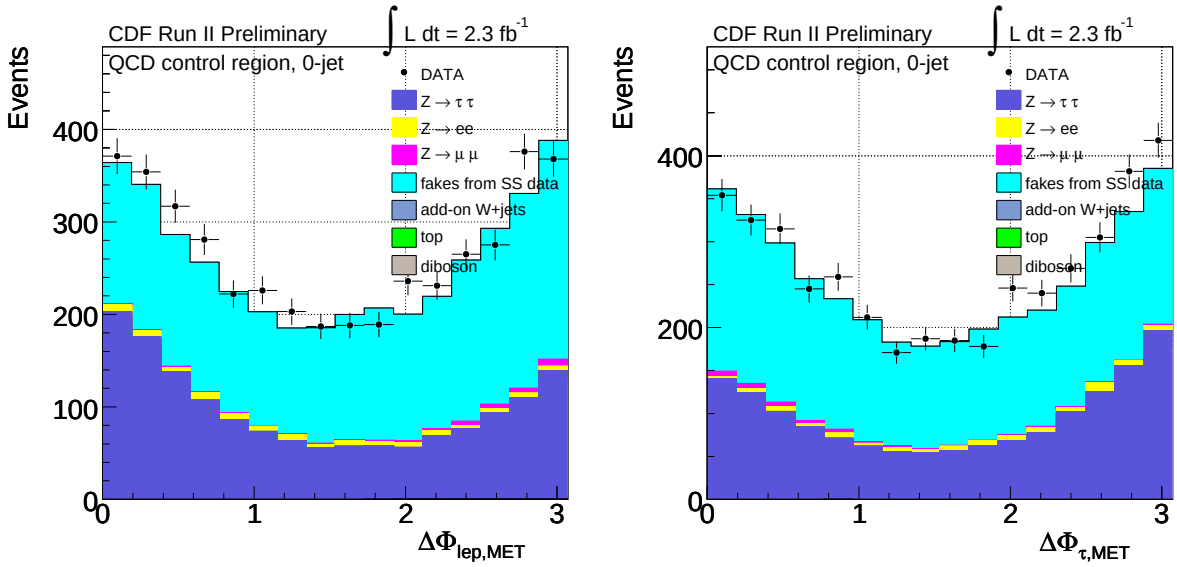

Figure 7.20: Angular separation in φ between lepton and $\not\!\!E_T$ (left) and τ and $\not\!\!E_T$ (right), for the QCD control region.

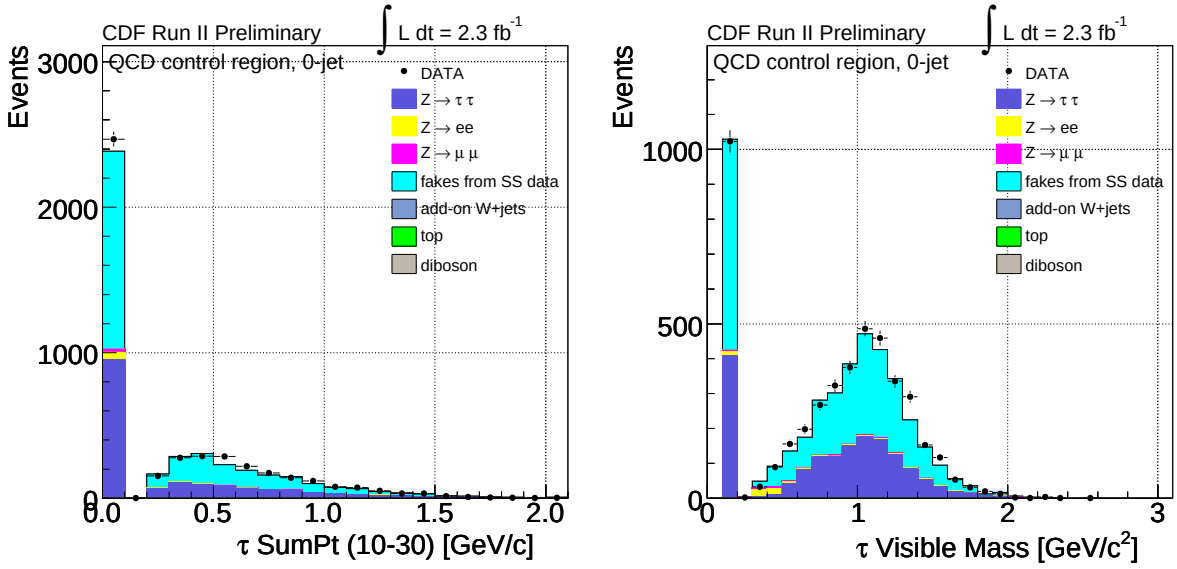

Figure 7.21: Sum of transverse momenta of all tracks in the τ isolation annulus (left) and τ visible mass (right), for the QCD control region.

CDF Run II Preliminary	$\int L = 2.3\ fb^{-1}$		
QCD region			
$Z/\gamma^* \rightarrow \tau\tau$	1610.4	$\pm$	72.2
$Z/\gamma^* \rightarrow ee/\mu\mu$	120.6	$\pm$	5.4
WW/WZ/ZZ	1.2	$\pm$	0.1
$t\bar{t}$	0.011	$\pm$	0.002
fakes from SS data	2542.0	$\pm$	254.2
add-on W+jets	13.3	$\pm$	0.7
Total Background	4287.8	$\pm$	203.9
Data	4433		
Total Signal (M_H = 120 GeV/c^2)	0.63	$\pm$	0.11

Table 7.10: Event yield in the 0-jet QCD region, with 2.3 fb^{-1} of CDF data. Systematic errors discussed in section 8.5 are included.

7.6.2 Z$\rightarrow \tau\tau$ region

The cuts on $\not\!\!E_T \geq 10$ GeV and $M_T(lep, \not\!\!E_T) \leq 60$ GeV/c^2 define a Drell-Yan enriched sample, useful to check the validity of the new hadronic τ identification and to evaluate the correction factor, if any differences appear in the identification performances between data and MC.

Figures 7.22, 7.23 and 7.24 present the resulting distributions for some relevant kinematical and τ variables. Table 7.11 summarizes the corresponding background and event yields.

We describe here the procedure followed to determine the scale factor for the new τ ID BDT algorithm by comparing data and background expectations in this control sample. The starting point is the formula for the Z$\rightarrow \tau\tau$ cross section:

$$\sigma(Z \rightarrow \tau\tau) = \frac{N_{\text{obs}} - N_{\text{bkg}}}{A \times \varepsilon_{\text{trig}} \times \varepsilon_{\text{IDlep}} \times \varepsilon_{\text{IDtau}} \times \varepsilon_{\text{vtx}} \times \int L dt} \tag{7.4}$$

where

- $\sigma(Z \rightarrow \tau\tau)$ is the cross section for the $Z/\gamma^* \rightarrow \tau\tau$ process (in the 66$\leq M_Z \leq$ 116 GeV/c^2 mass window [101]);
- A is the corresponding acceptance, evaluated from MC samples;
- N_{obs} is the number of observed events in the control region;
- N_{bkg} is the estimated amount of background contributions (except $Z/\gamma^* \rightarrow \tau\tau$ in the 66$\leq M_Z \leq$ 116 GeV/c^2 mass range);

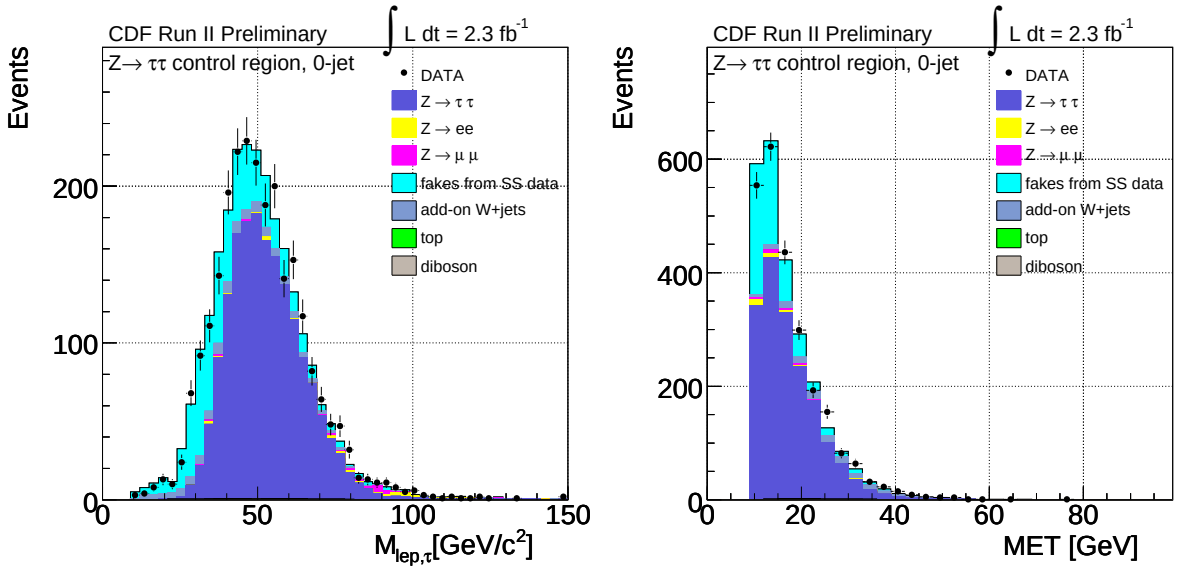

Figure 7.22: Lepton+τ invariant mass (left) and missing transverse energy (right) for the $Z/\gamma^* \rightarrow \tau\tau$ control region.

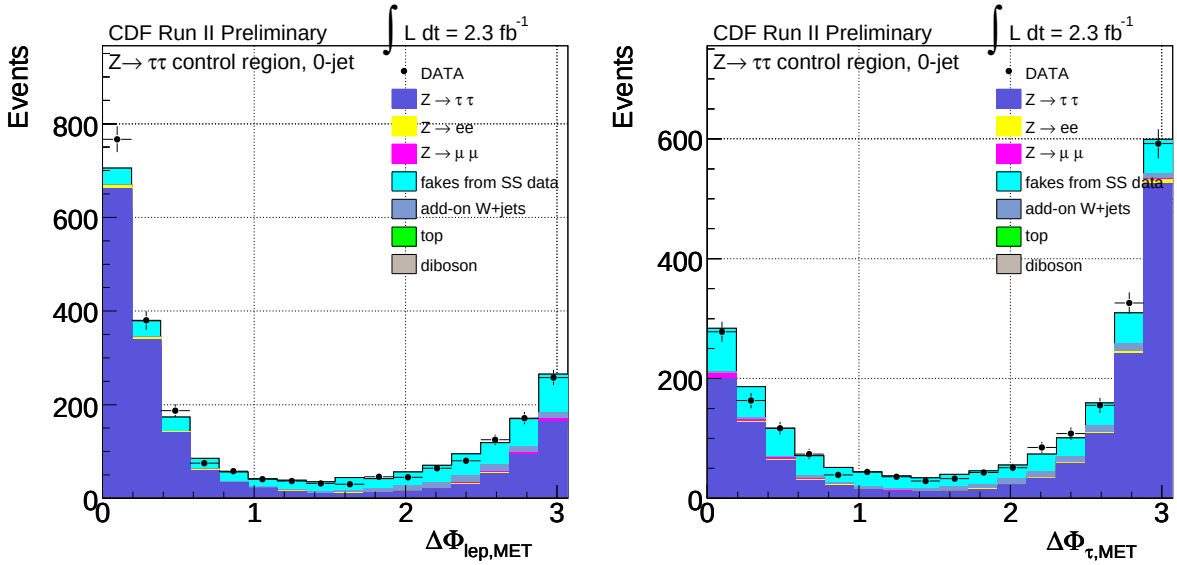

Figure 7.23: Angular separation in φ between lepton and $\not\!\!E_T$ (left) and τ and $\not\!\!E_T$ (right), for the $Z/\gamma^* \rightarrow \tau\tau$ control region.

- $\varepsilon_{\text{trig}}$ is the trigger efficiency;
- $\varepsilon_{\text{IDlep}}$ is the electron/muon ID scale factor;
- ε_{vtx} is the efficiency of the z-vertex position requirement.

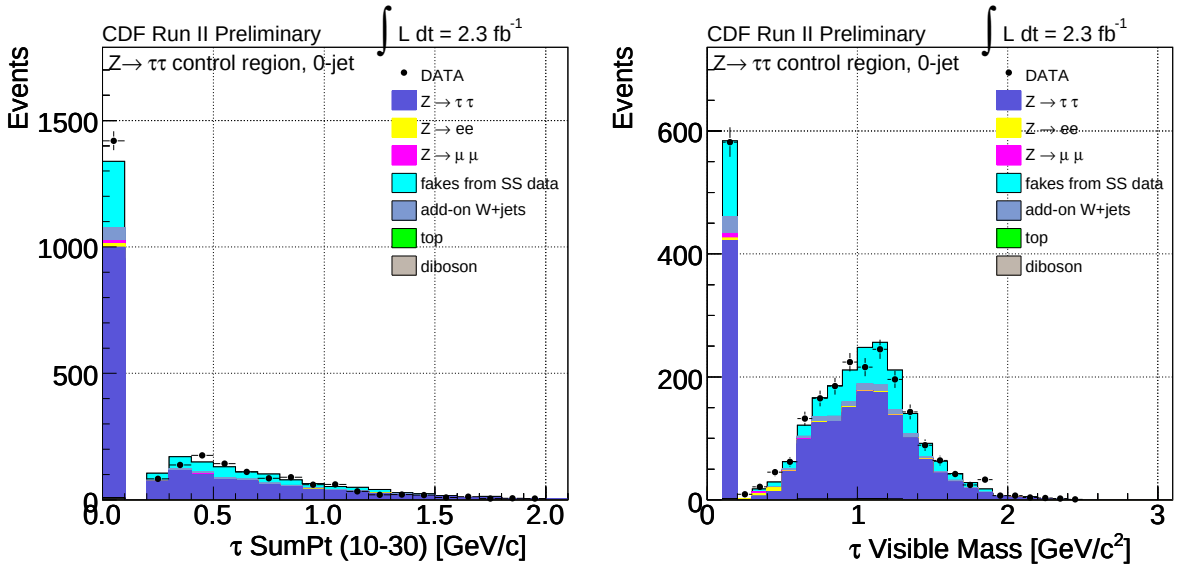

Figure 7.24: Sum of transverse momenta of all tracks in the τ isolation annulus (left) and τ visible mass (right), for the $Z/\gamma^* \to \tau\tau$ control region.

- $\varepsilon_{\mathrm{IDtau}}$ represents the quantity to be determined.

In the numerator we can split N_{bkg} into the different contributions, by taking into account that multijet, γ+jet and W+jets processes are estimated from data and thus are not affected by the τ ID scale factor. Therefore:

$$N_{\mathrm{bkg}} = N_{\mathrm{SSdata}} + N^{add-on}_{\mathrm{W+jets}} + N^{\mathrm{MC}}_{\mathrm{bkg}} \tag{7.5}$$

where

$$N^{\mathrm{MC}}_{\mathrm{bkg}} = \sum_{\mathrm{i_{bkg}}=1}^{N} \sigma^i \times A^i \times \varepsilon_{\mathrm{trig}} \times \varepsilon_{\mathrm{IDlep}} \times \varepsilon_{\mathrm{IDtau}} \times \varepsilon_{\mathrm{vtx}} \times \int Ldt \tag{7.6}$$

The formula of the $Z \to \tau\tau$ cross section can finally be reversed and used to provide an estimate of $\varepsilon_{\mathrm{IDtau}}$:

$$\varepsilon_{\mathrm{IDtau}} = \frac{1}{\varepsilon_{\mathrm{trig}} \times \varepsilon_{\mathrm{IDlep}} \times \varepsilon_{\mathrm{vtx}} \times \int Ldt} \times \frac{N_{\mathrm{obs}} - N^{\mathrm{QCD}}_{\mathrm{SSdata}} - N_{\mathrm{W+jets}}}{A \times \sigma(Z \to \tau\tau) + \sum^{N}_{\mathrm{i_{bkg}}=1}(\sigma^i \times A^i)} \tag{7.7}$$

The obtained value is consistent with 1, while the systematic uncertainties related to this measurement will be discussed in section 8.5. From now on, all the MC-based background and signal rates will be derived by applying formula 7.1, where $\varepsilon_{\mathrm{IDtau}}$=1.00.

CDF Run II Preliminary	$\int L = 2.3\ fb^{-1}$		
$Z/\gamma^* \to \tau\tau$ region			
$Z/\gamma^* \to \tau\tau$	1735.5	$\pm$	73.0
$Z/\gamma^* \to ee/\mu\mu$	44.0	$\pm$	2.0
WW/WZ/ZZ	9.7	$\pm$	0.8
$t\bar{t}$	0.10	$\pm$	0.02
fakes from SS data	594.8	$\pm$	59.5
add-on W+jets	112.7	$\pm$	5.6
Total Background	2496.3	$\pm$	54.9
Data	2501		
Total Signal ($M_H = 120$ GeV/c^2)	1.253	$\pm$	0.203

Table 7.11: Event yield in the 0-jet $Z/\gamma^* \to \tau\tau$ region, with 2.3 fb^{-1} of CDF data. Systematic errors discussed in section 8.5 are included.

7.6.3 W+jets region

We check the validity of the normalization of the W+jets ALPGEN MC. Background and data number of events are reported in table 7.12, while the distributions of some kinematic and τ variables are shown in figures 7.25, 7.26 and 7.27.

CDF Run II Preliminary	$\int L = 2.3\ fb^{-1}$		
W+jets region			
$Z/\gamma^* \to \tau\tau$	42.0	$\pm$	1.9
$Z/\gamma^* \to ee/\mu\mu$	6.9	$\pm$	0.3
WW/WZ/ZZ	14.9	$\pm$	1.2
$t\bar{t}$	0.24	$\pm$	0.05
fakes from SS data	137.0	$\pm$	13.7
add-on W+jets	180.1	$\pm$	9.0
Total Background	381.4	$\pm$	14.9
Data	386		
Total Signal ($M_H = 120$ GeV/c^2)	0.126	$\pm$	0.016

Table 7.12: Event yield in the 0-jet W+jets region, with 2.3 fb^{-1} of CDF data. Systematic errors discussed in section 8.5 are included.

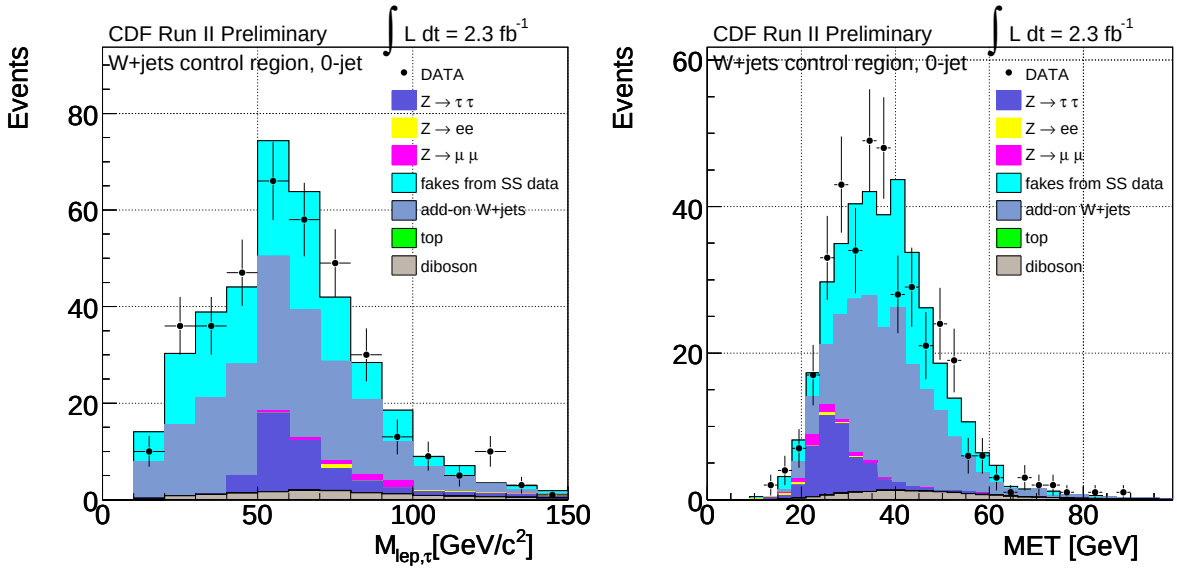

Figure 7.25: Lepton+τ invariant mass (left) and missing transverse energy (right) for the W+jets control region.

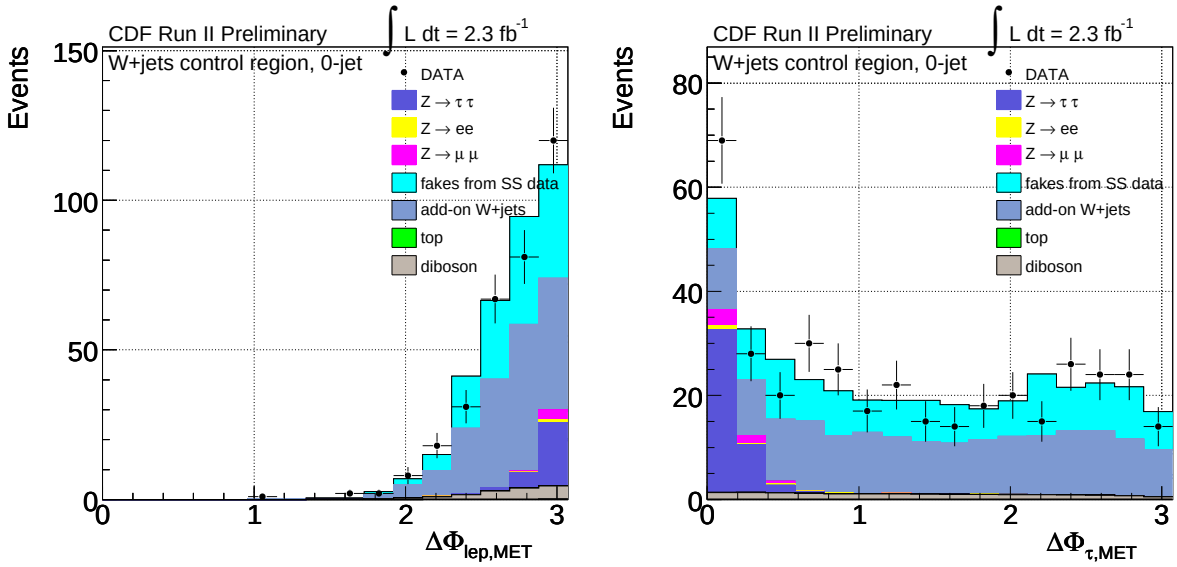

Figure 7.26: Angular separation in φ between lepton and $\not\!\!E_T$ (left) and τ and $\not\!\!E_T$ (right), for the W+jets control region.

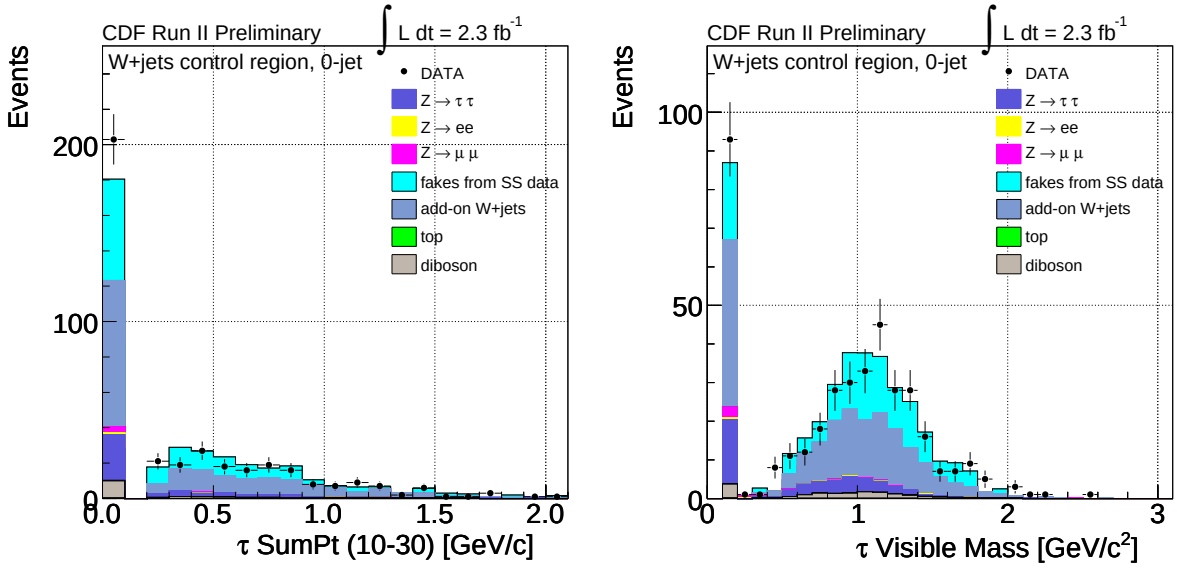

Figure 7.27: Sum of transverse momenta of all tracks in the τ isolation annulus (left) and τ visible mass (right), for the W+jets control region.

Chapter 8

Results

"An experiment is a question which science poses to Nature,
and a measurement is the recording of Nature's answer"
Max Planck

This chapter presents the analysis of the signal datasets. We show a comparison between the expected background and the observed data in several kinematical and topological variables, followed by the discussion of the different systematic uncertainties which affect the measurement. A description of the multivariate technique adopted to improve the separation of the Higgs contribution from the principal sources of background is then provided.

Given no evidence for a signal excess in any distribution, a Bayesian method [105] is employed to set an upper limit to the Higgs production cross section in the $100 \leq M_H \leq 150$ GeV/c^2 mass range.

8.1 Event yield

The expected background and the observed event rates for the signal channels, defined by applying the selection cuts listed in table 6.14, are reported in table 8.1 separately for events containing exactly one or at least two reconstructed jet with transverse energy greater than 20 GeV.

The table also shows the expected signal yield for the different Higgs production processes, in the mass hypothesis of 120 GeV/c^2. The complete summary of signal predictions for the other Higgs masses is provided in section 7.3.

CDF Run II Preliminary						$\int L = 2.3\ fb^{-1}$
Signal channels $M_H = 120$ GeV/c^2						
Background source	1 JET			≥ 2 JETS		
$Z/\gamma^* \rightarrow \tau\tau$	357.9	$\pm$	33.1	59.3	$\pm$	8.8
$Z/\gamma^* \rightarrow ee/\mu\mu$	26.4	$\pm$	2.0	4.8	$\pm$	0.7
WW/WZ/ZZ	3.9	$\pm$	0.4	0.9	$\pm$	0.1
$t\bar{t}$	4.6	$\pm$	0.6	16.3	$\pm$	1.9
fakes from SS data	483.0	$\pm$	48.3	64.0	$\pm$	6.4
add-on W+jets	45.8	$\pm$	8.2	14.1	$\pm$	4.2
Total Background	921.7	$\pm$	48.9	159.4	$\pm$	11.6
Data	965			166		
ggH	0.535	$\pm$	0.154	0.129	$\pm$	0.092
WH	0.091	$\pm$	0.010	0.150	$\pm$	0.014
ZH	0.050	$\pm$	0.005	0.099	$\pm$	0.009
VBF	0.070	$\pm$	0.009	0.099	$\pm$	0.013
Total Signal	0.746	$\pm$	0.163	0.477	$\pm$	0.121

Table 8.1: Event yield in the 1 jet and $\geq$2 jets channels, with 2.3 fb^{-1} of CDF data. Systematic errors discussed in section 8.5 are included.

In the next sections some of the most relevant τ, kinematical and topological distributions are shown, separately for the two signal channels. In each plot the data are represented by the black points, with the statistical uncertainties; the colored regions are the different background contributions estimated following the procedure explained in chapter 7. The Higgs signal is also reported, with the rate incremented by a factor of 100.

The variables related to the hadronic tau identification have been defined in section 6.4.3, while the variables which characterize the event kinematics and topology are described in section 8.4.1.

As can be seen, a satisfying agreement between the observed data and the expected background is achieved in a very large set of distributions.

8.2 1 jet signal channel

8.2.1 τ variables

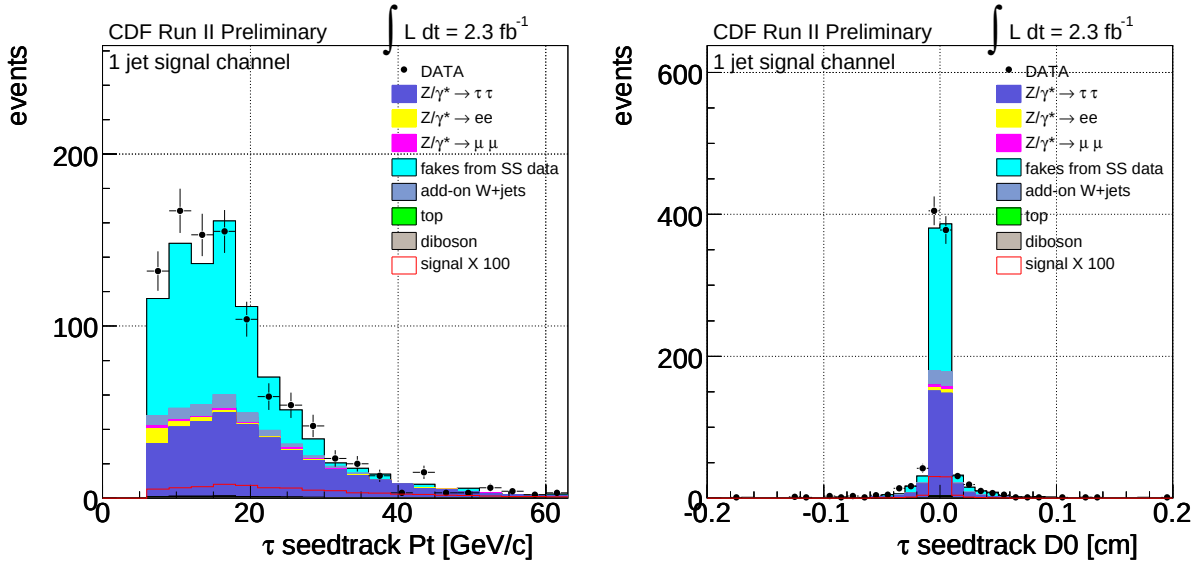

Figure 8.1: τ seedtrack p_T (left) and seedtrack d_0 (right) for the 1 jet channel.

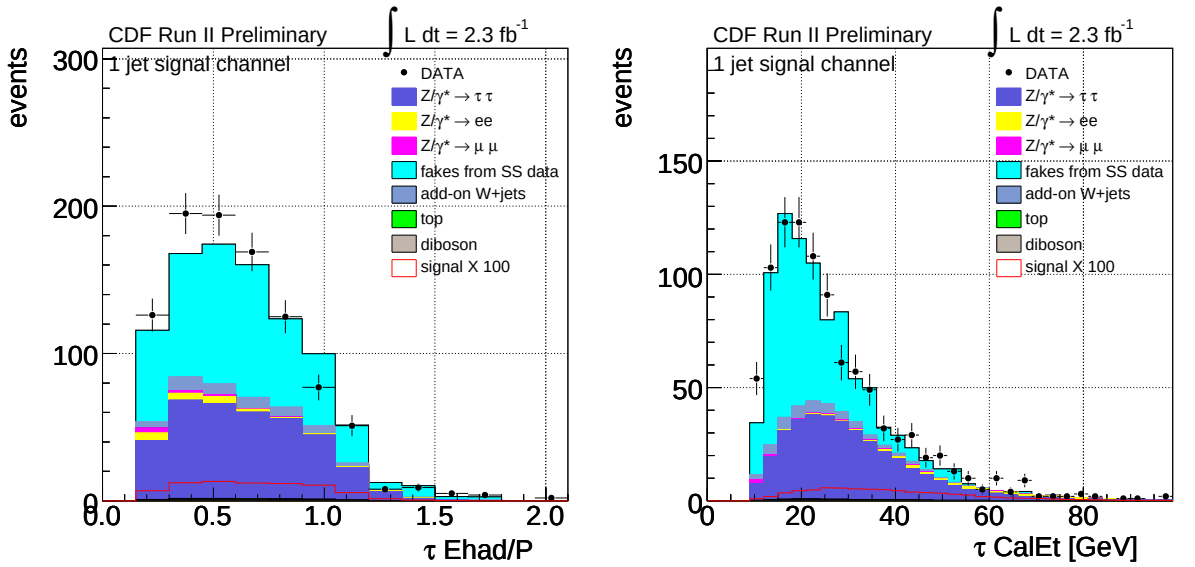

Figure 8.2: Ratio between the τ hadronic energy and the sum of track momenta in the signal cone (left) and tau cluster E_T number of tracks in the τ signal cone (right) for the 1 jet channel.

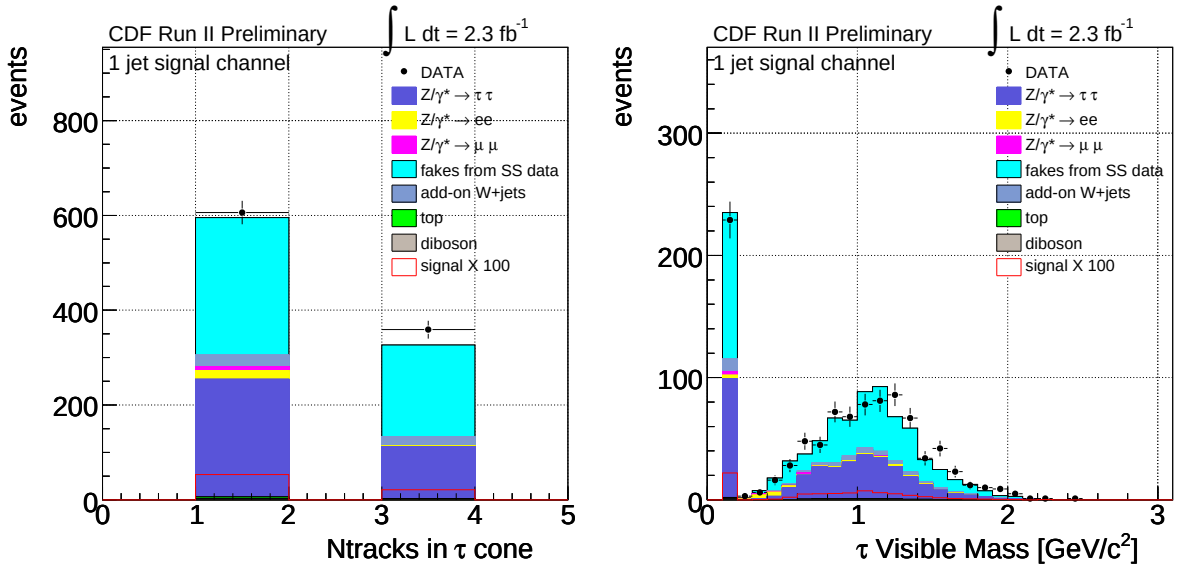

Figure 8.3: Number of tracks in the τ signal cone (left) and τ visible mass (right) for the 1 jet channel.

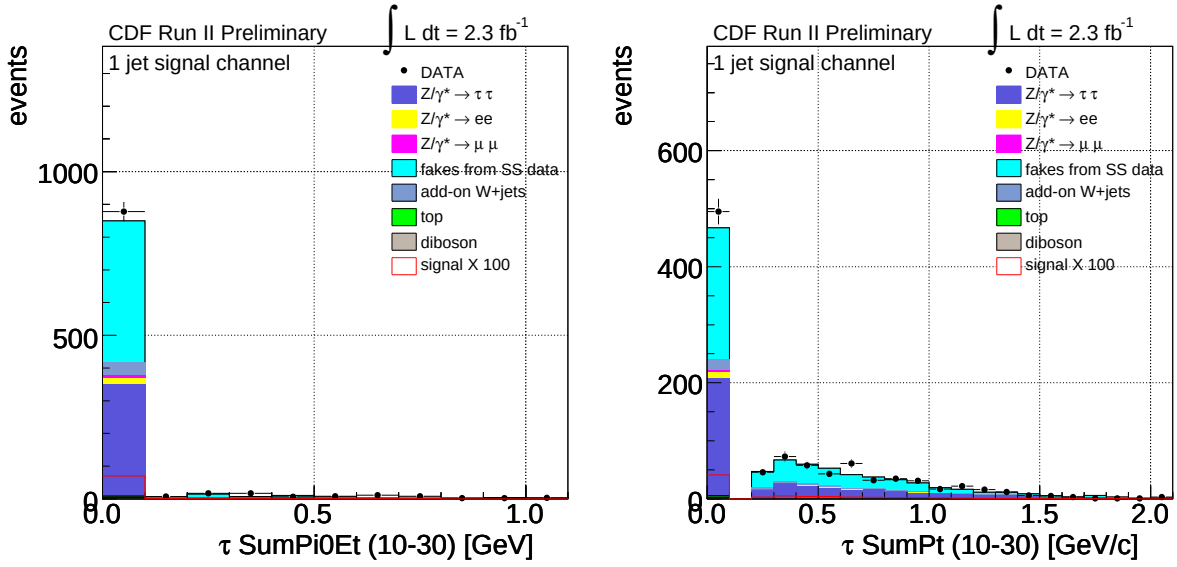

Figure 8.4: Sum of p_T of all tracks (right) and sum of E_T of all the π^0's in the isolation annulus for the 1 jet channel.

8.2.2 Event kinematics and topology

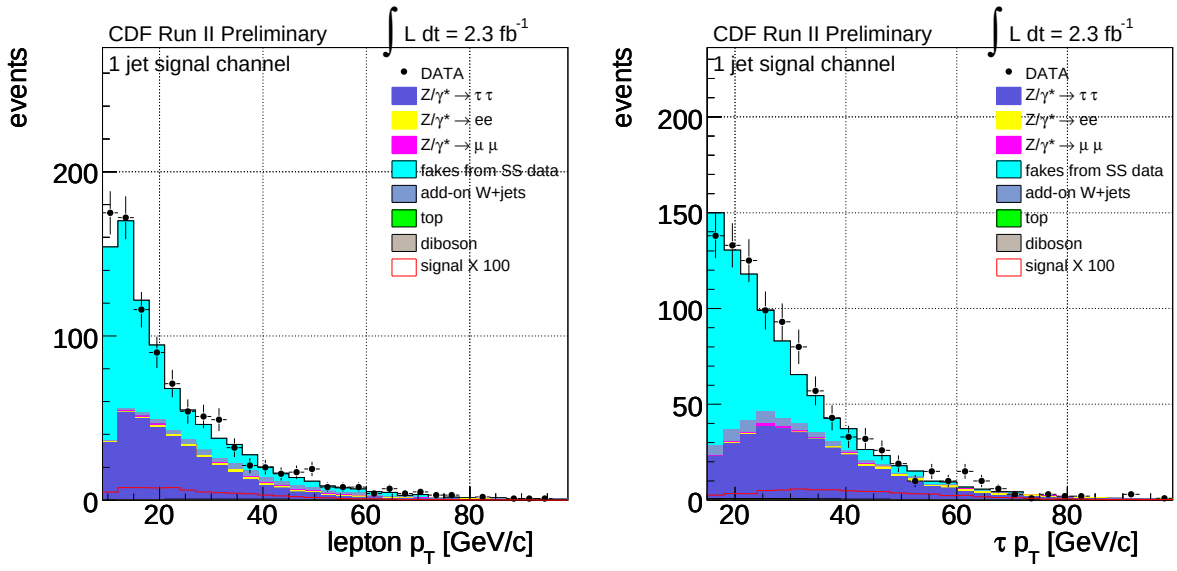

Figure 8.5: Lepton p_T (left) and τ visible p_T (right) for the 1 jet channel.

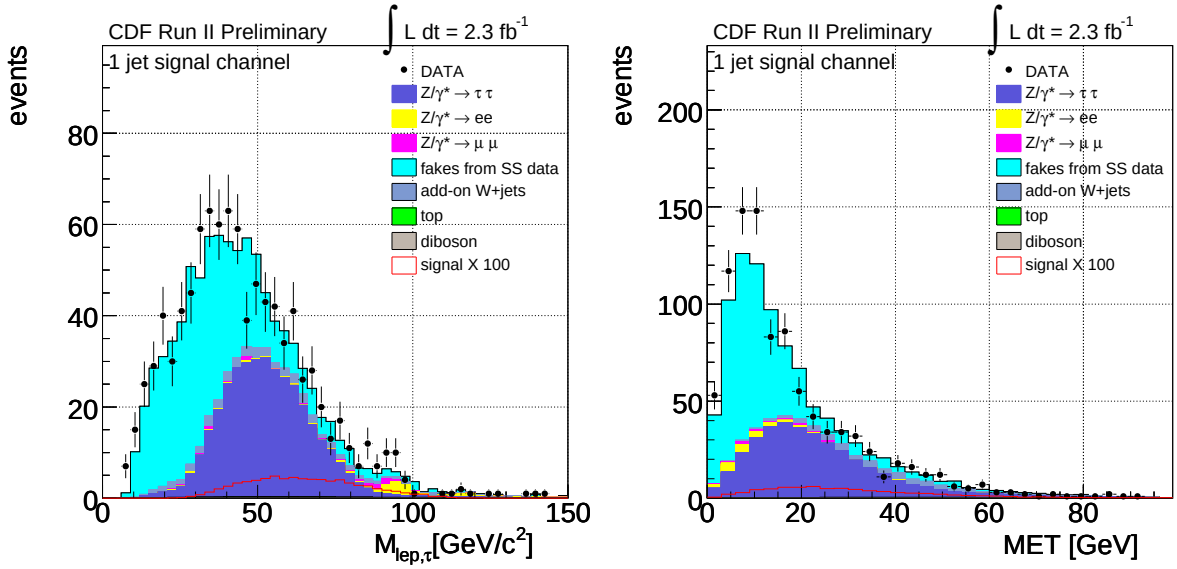

Figure 8.6: Invariant mass of the lepton+τ system (left) and $\not\!\!E_T$ (right) for the 1 jet channel.

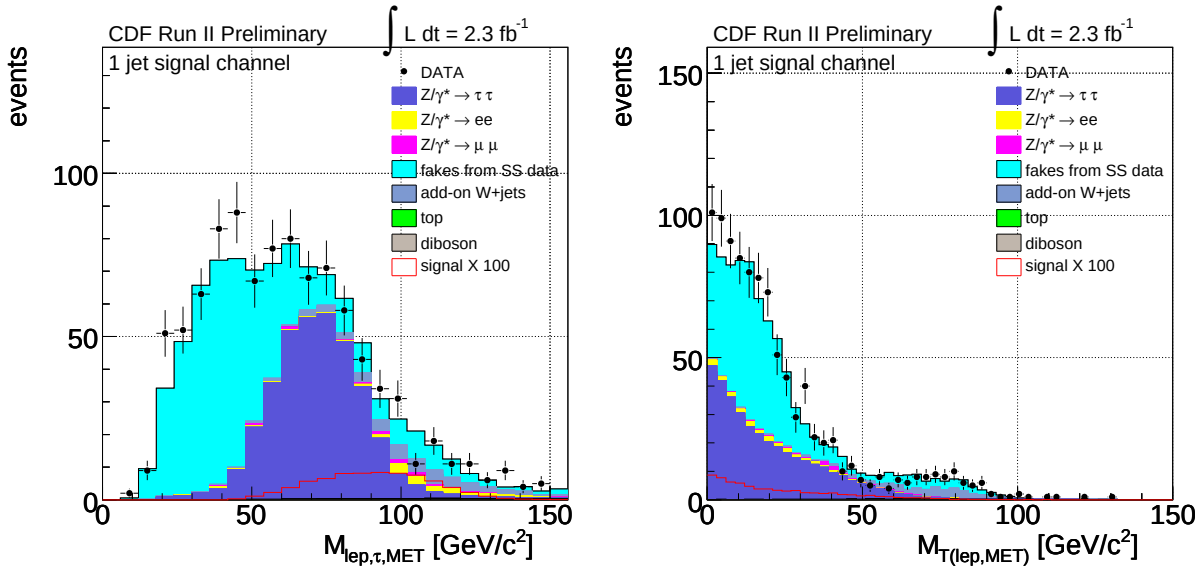

Figure 8.7: Invariant mass of the lepton+τ+$\not\!\!E_T$ system (left) and transverse mass of the lepton+$\not\!\!E_T$ system (right) for the 1 jet channel.

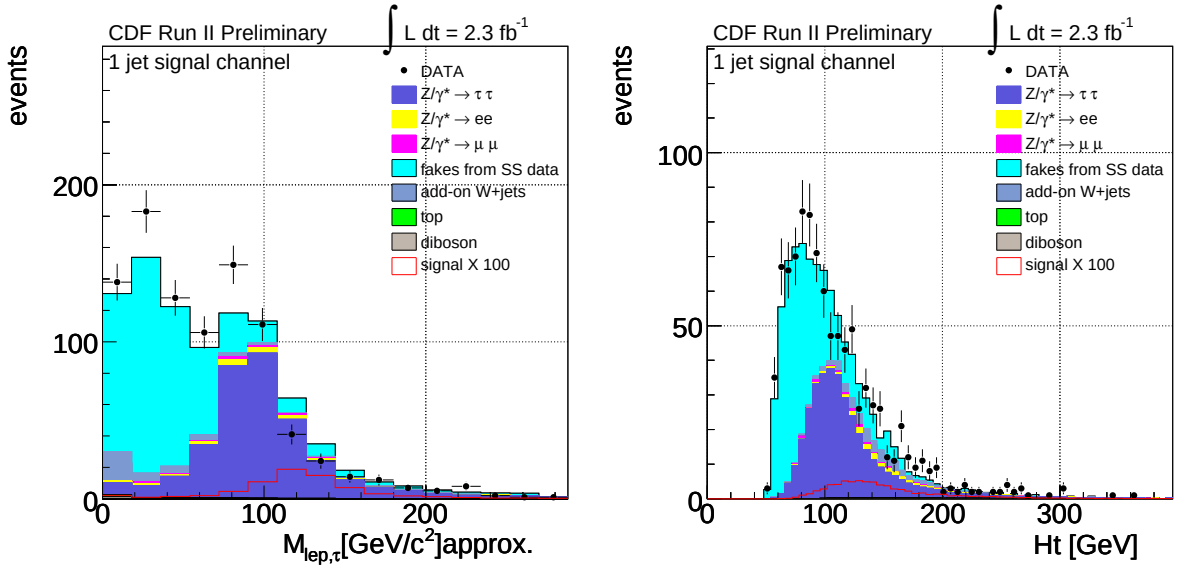

Figure 8.8: Invariant mass of the lepton+τ system in the collinear approximation for neutrinos' energy (left) and sum of the transverse energies for all the reconstructed objects in the event (right) for the 1 jet channel.

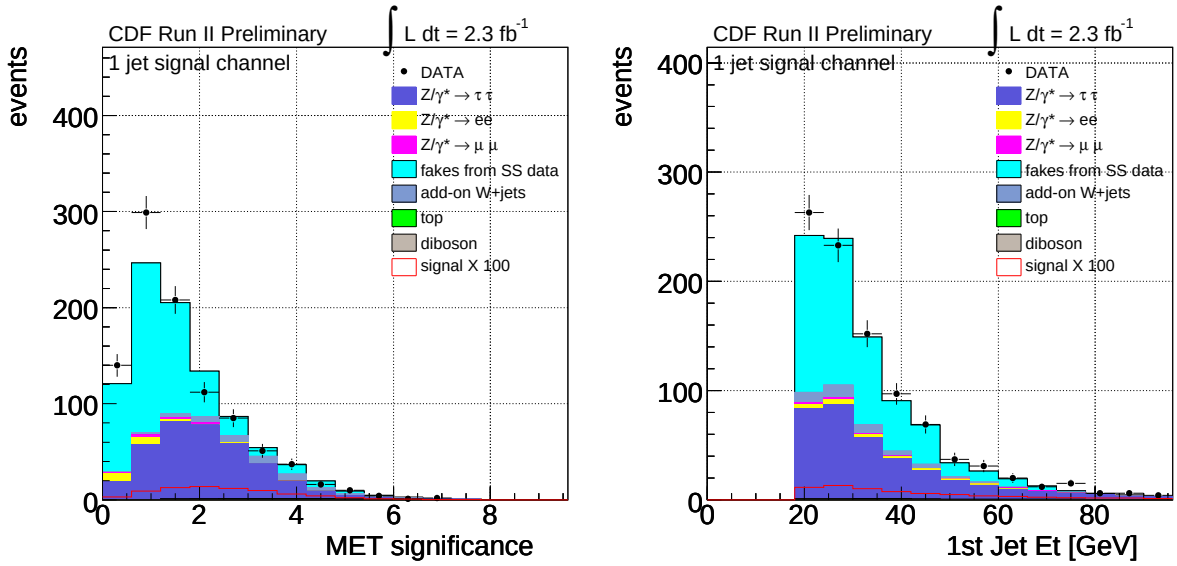

Figure 8.9: $\not\!\!E_T$ significance (left) and corrected E_T of the most energetic jet (right) for the 1 jet channel.

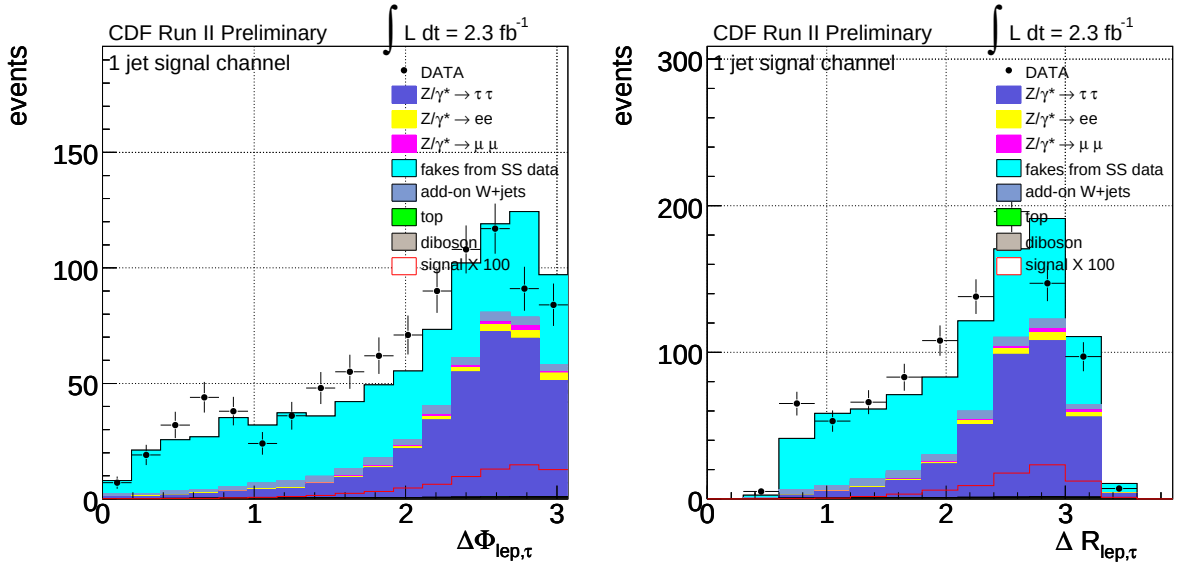

Figure 8.10: Angular separation in φ between the lepton and the τ (lef) and separation in R between lepton and τ (right) for the 1 jet channel.

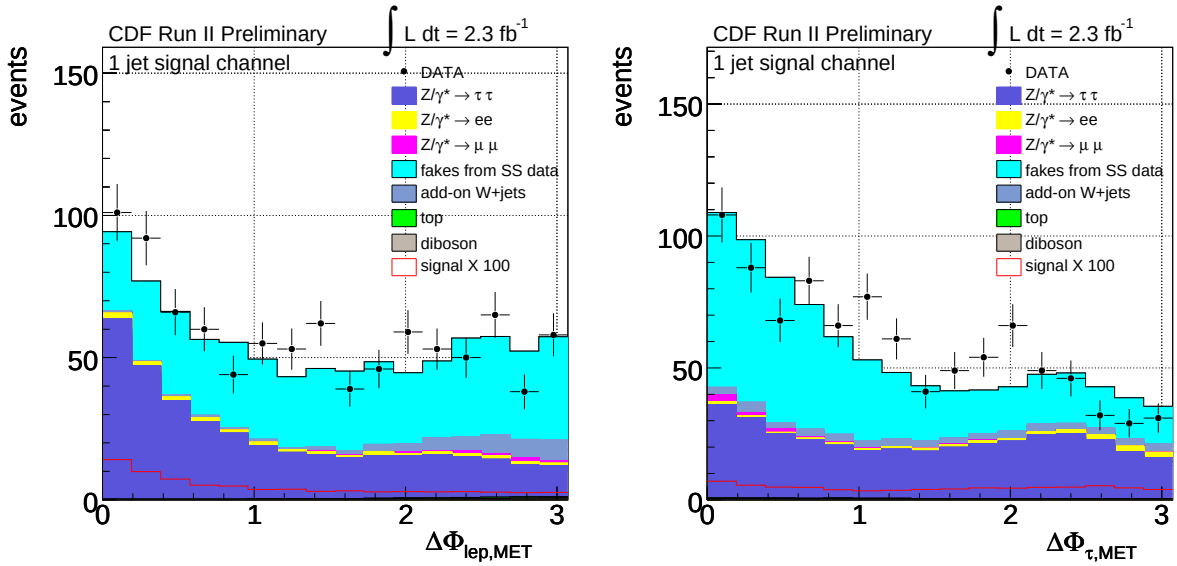

Figure 8.11: Angular separation in φ between the lepton and $\not\!\!E_T$ (left) and between the τ and $\not\!\!E_T$ (right) for the 1 jet channel.

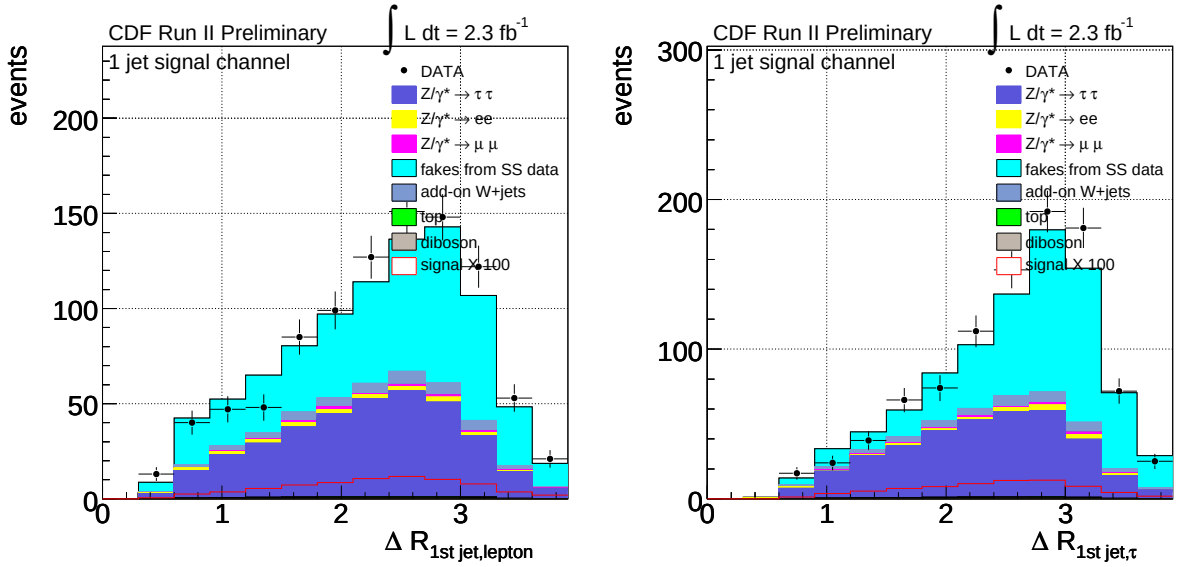

Figure 8.12: Separation in R between the leading jet and the lepton (left) and between the leading jet and the τ (right) for the 1 jet channel.

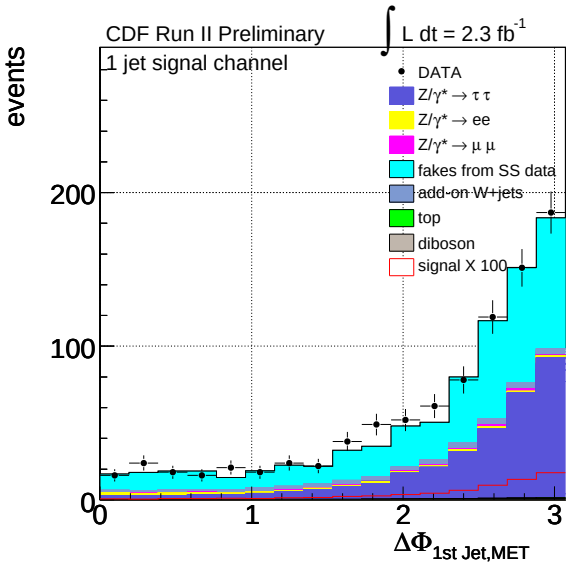

Figure 8.13: Angular separation in φ between the leading jet and $\not\!\!E_T$ for the 1 jet channel.

8.3 ≥2 jets signal channel

8.3.1 τ variables

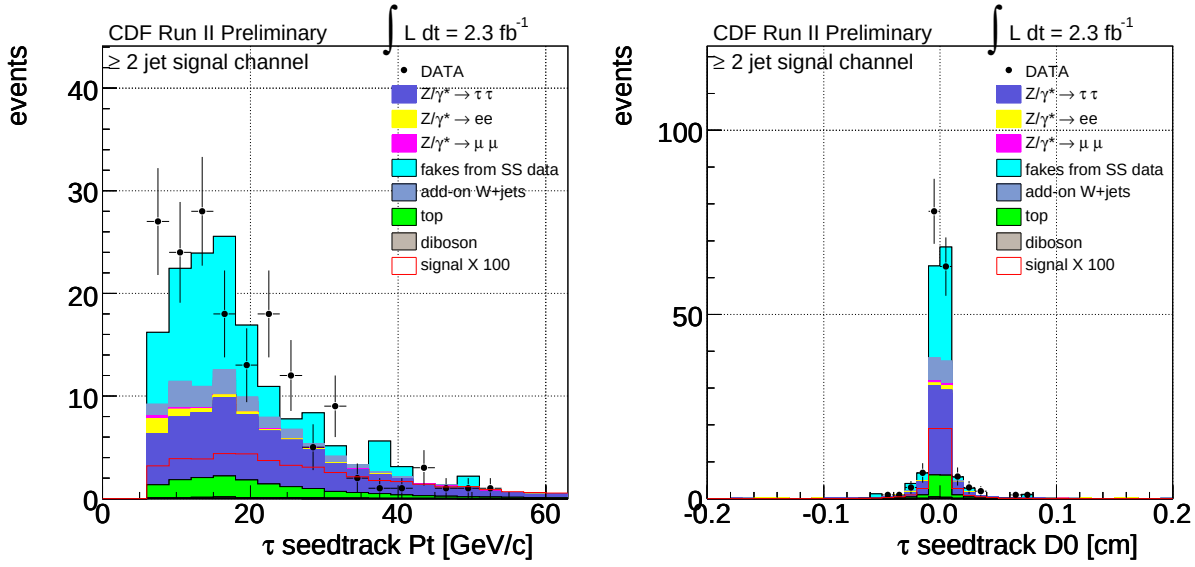

Figure 8.14: τ seedtrack p_T (left) and seedtrack d_0 (right) for the ≥2 jets channel.

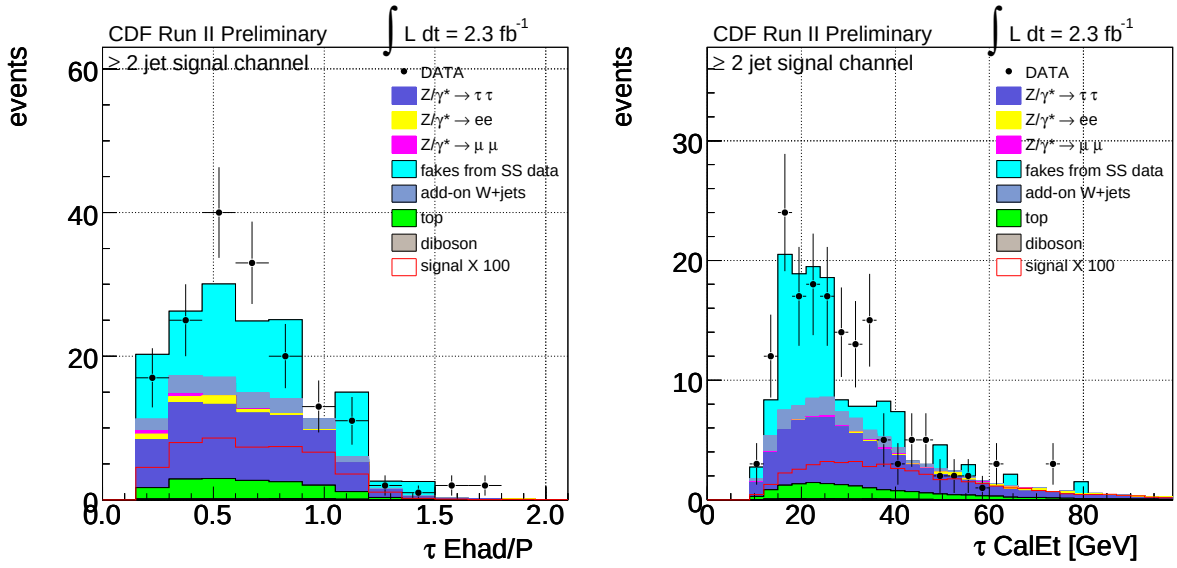

Figure 8.15: Ratio between the τ hadronic energy and the sum of track momenta in the signal cone (left) and tau cluster E_T number of tracks in th e τ signal cone (right) for the $\geq$2 jets channel.

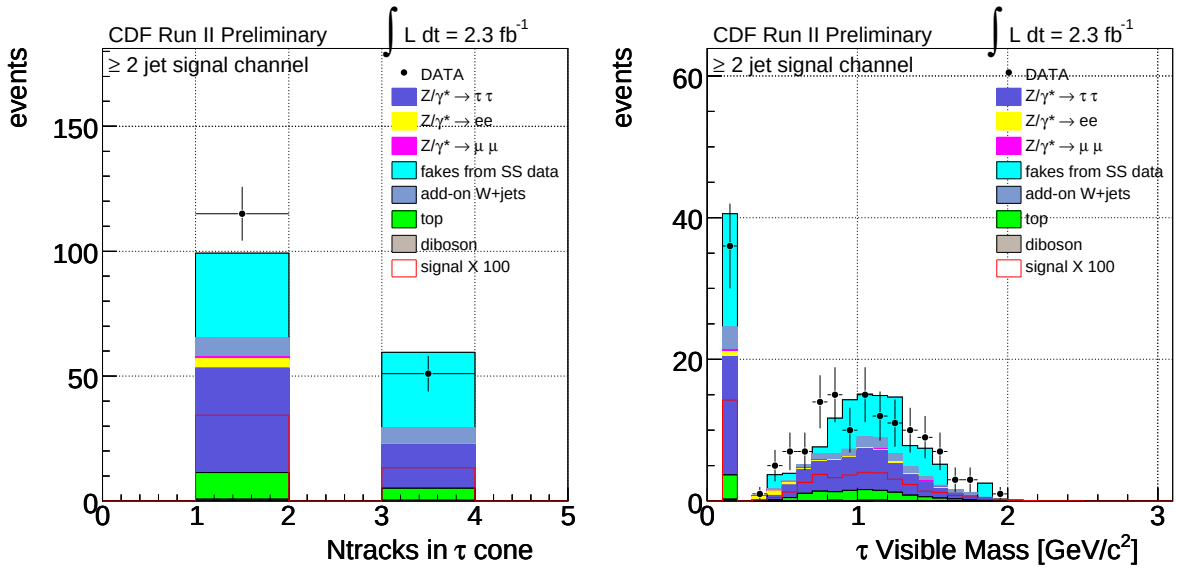

Figure 8.16: Number of tracks in the τ signal cone (left) and τ visible mass (right) for the $\geq$2 jets channel.

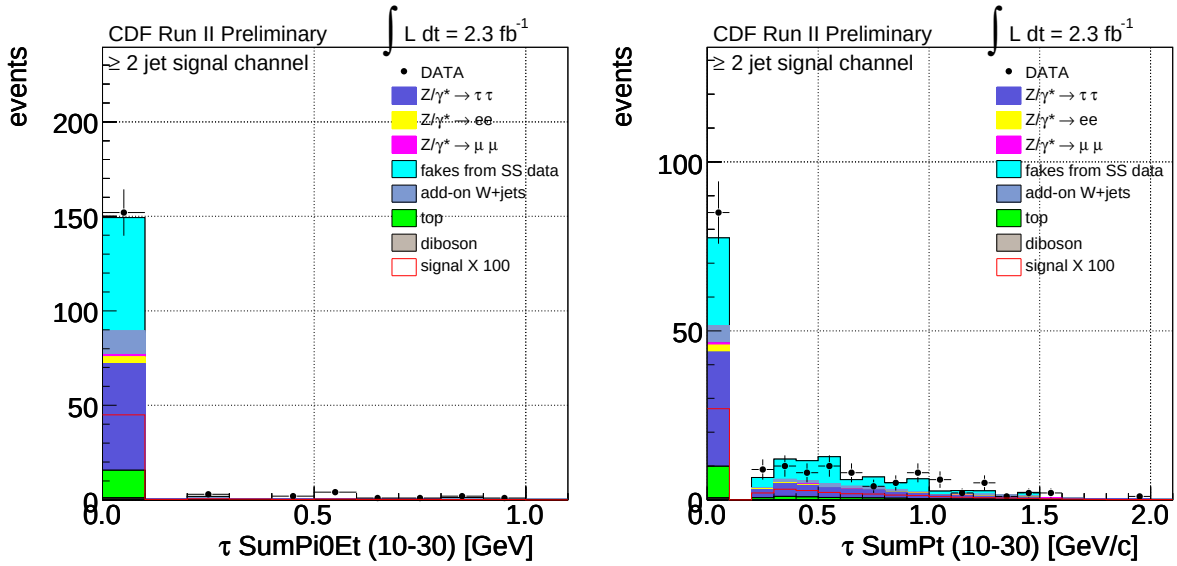

Figure 8.17: Sum of p_T of all tracks (right) and sum of E_T of all the π^0's in the isolation annulus for the $\geq$2 jets channel.

8.3.2 Event kinematics and topology

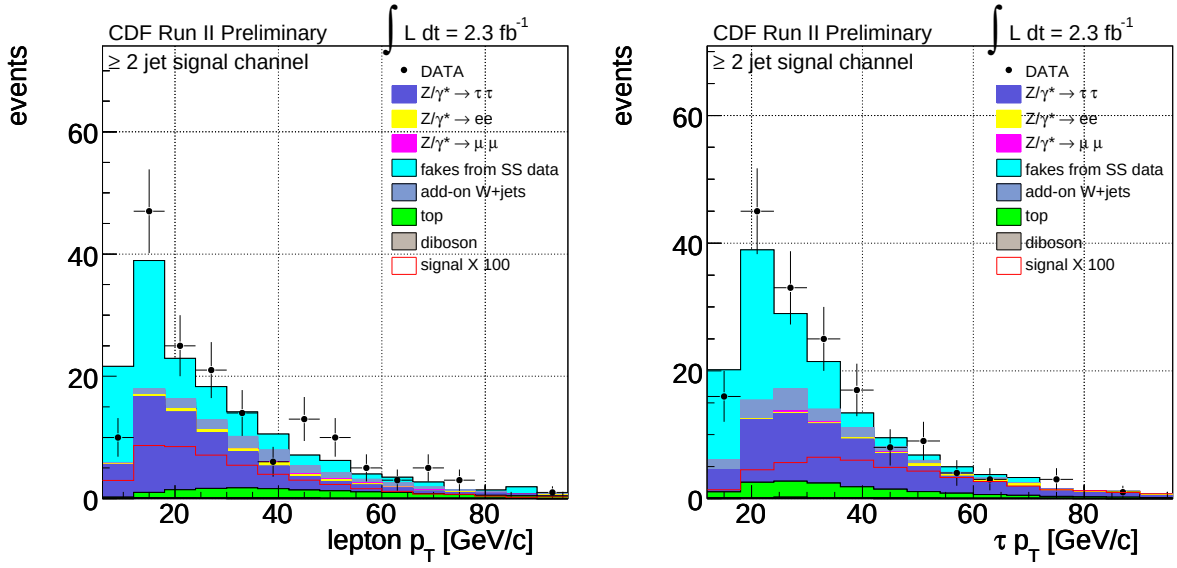

Figure 8.18: Lepton p_T (left) and τ visible p_T (right) for the $\geq$2 jets channel.

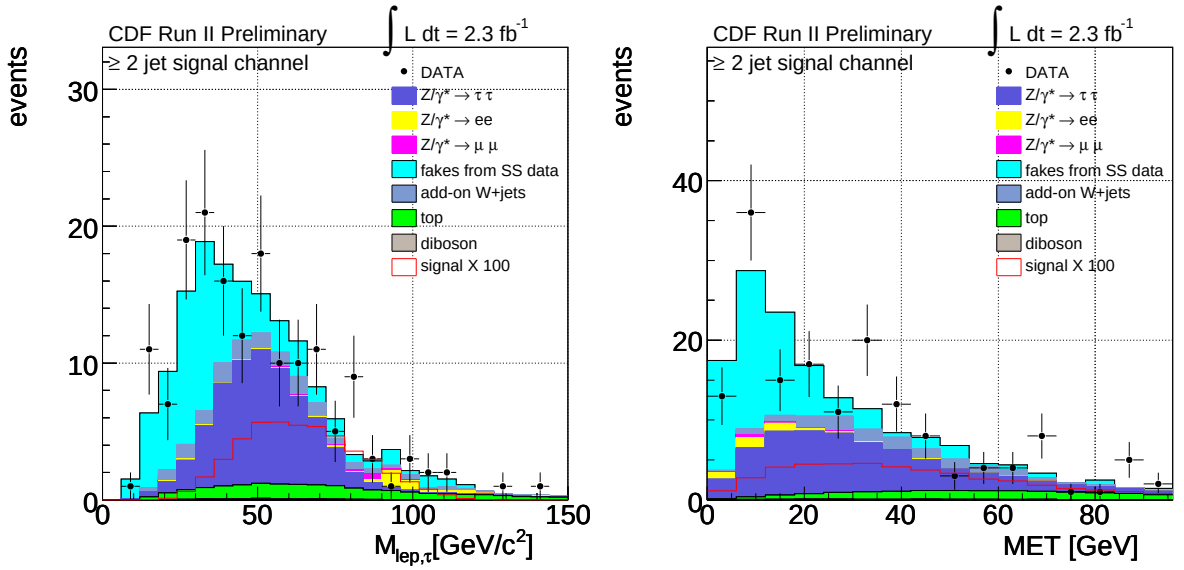

Figure 8.19: Invariant mass of the lepton+τ system (left) and $\not\!\!E_T$ (right) for the $\geq$2 jets channel.

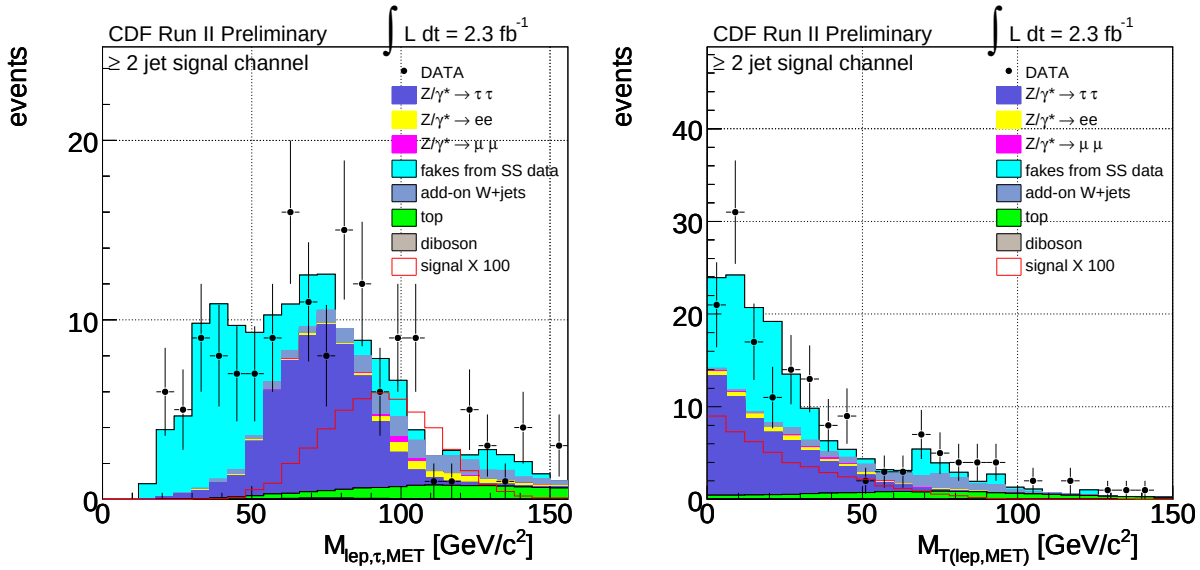

Figure 8.20: Invariant mass of the lepton+τ+$\not\!\!E_T$ system (left) and transverse mass of the lepton+$\not\!\!E_T$ system (right) for the $\geq$2 jets channel.

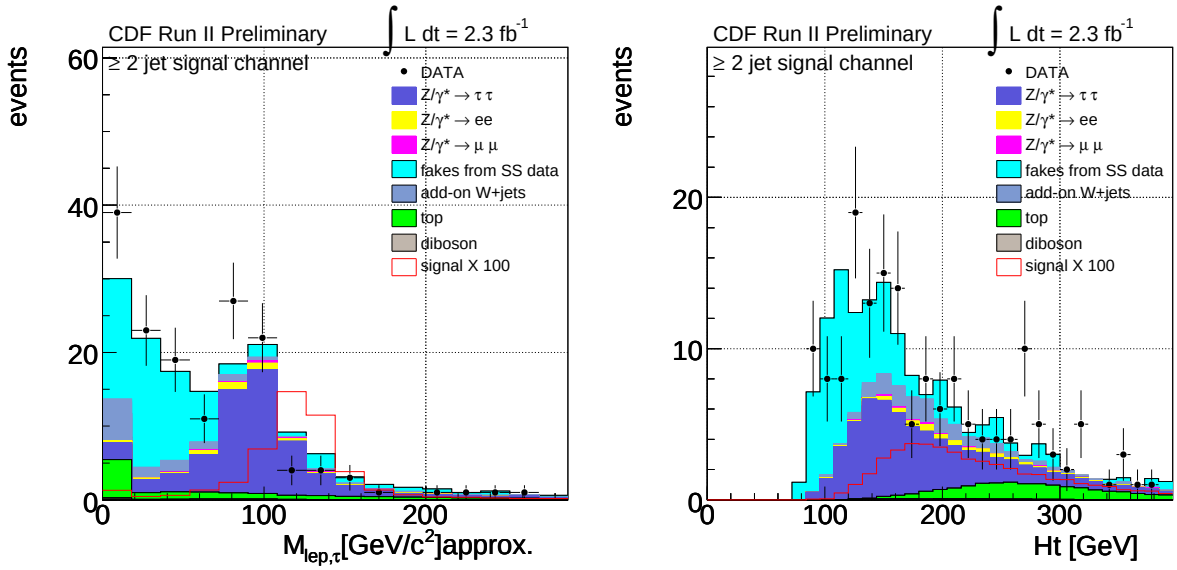

Figure 8.21: Invariant mass of the lepton+τ system in the collinear approximation for neutrinos' energy (left) and sum of the transverse energies for all the reconstructed objects in the event (right) for the $\geq$2 jets channel.

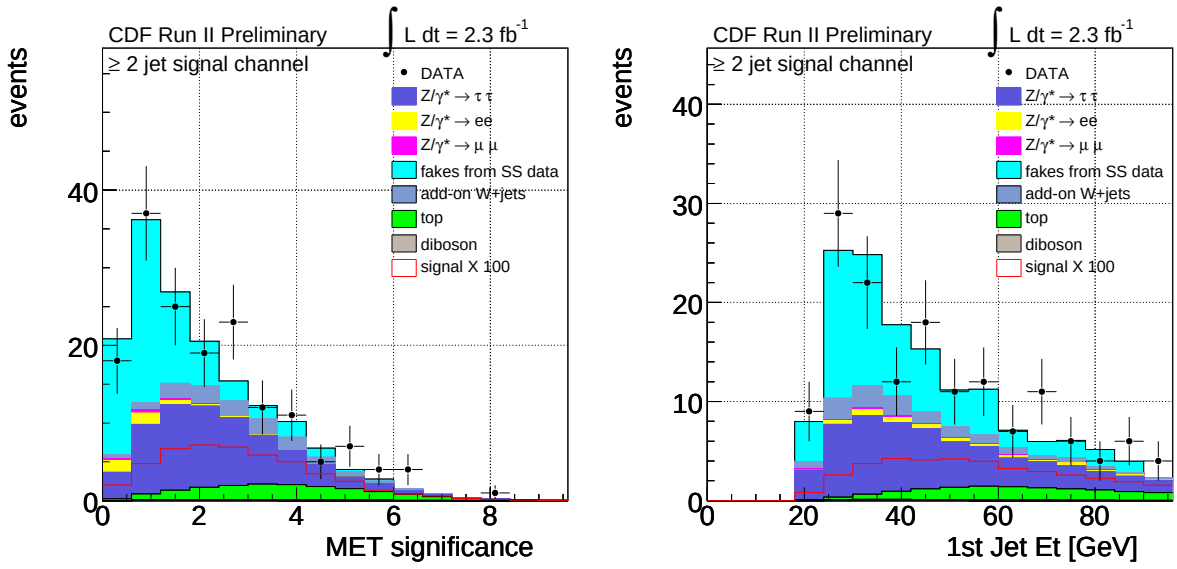

Figure 8.22: $\not\!\!E_T$ significance (left) and corrected E_T of the most energetic jet (right) for the $\geq$2 jets channel.

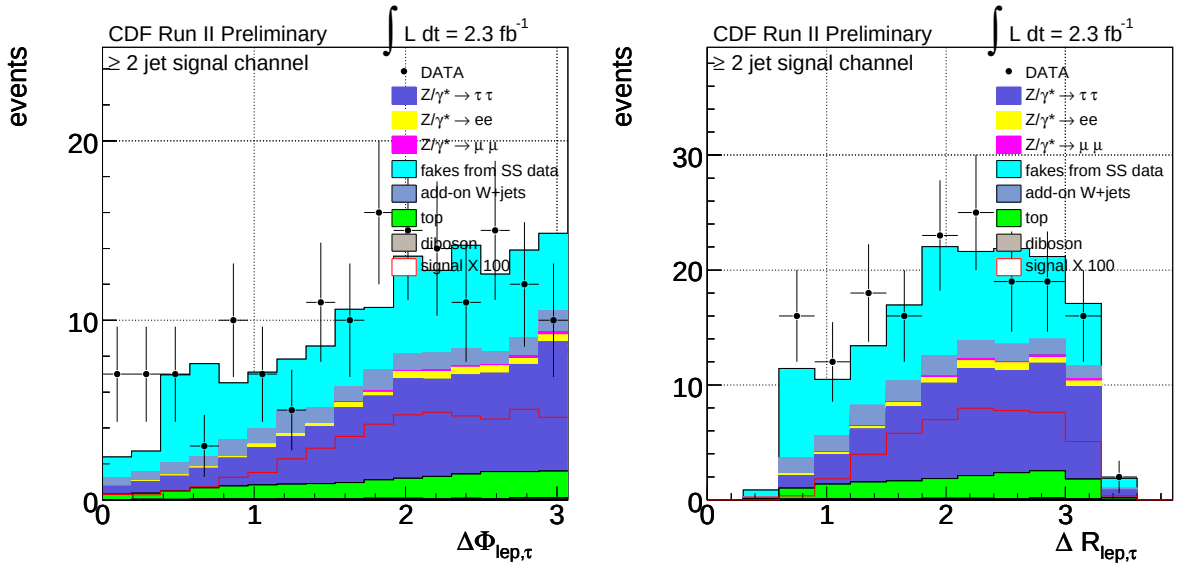

Figure 8.23: Angular separation in φ between the lepton and the τ (lef) and separation in R between lepton and τ (right) for the $\geq$2 jets channel.

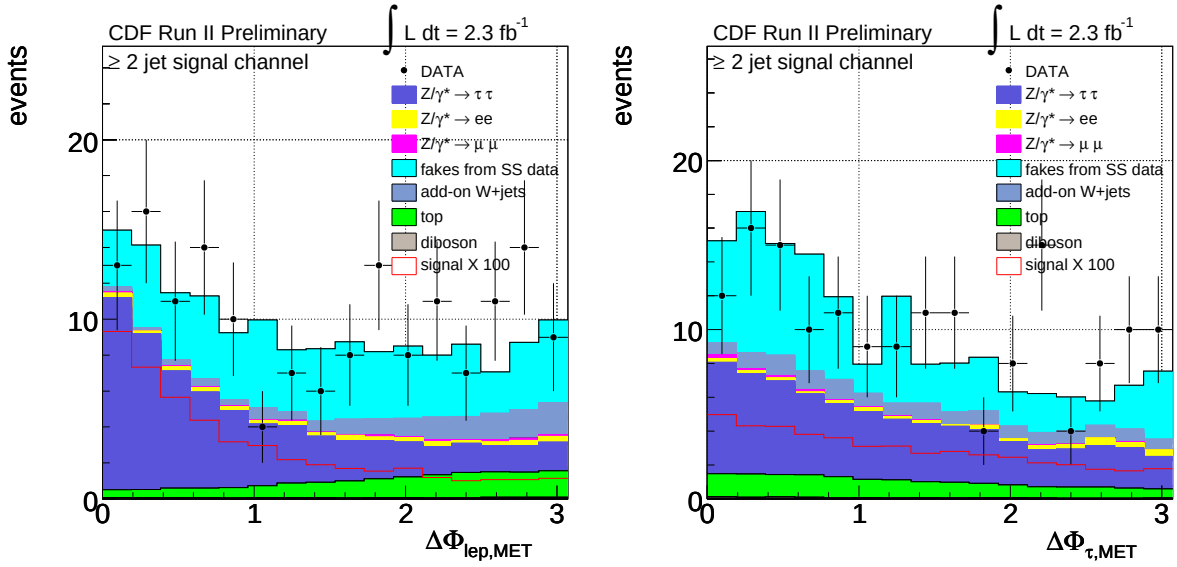

Figure 8.24: Angular separation in φ between the lepton and $\not\!\!E_T$ (left) and between the τ and $\not\!\!E_T$ (right) for the $\geq$2 jets channel.

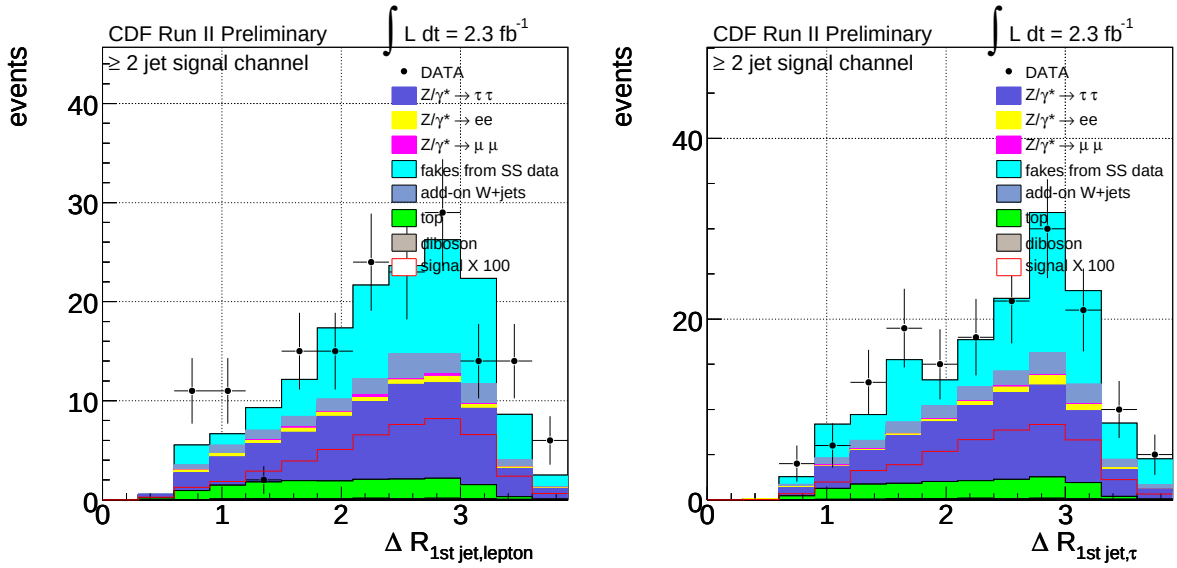

Figure 8.25: Separation in R between the leading jet and the lepton (left) and between the leading jet and the τ (right) for the $\geq$2 jets channel.

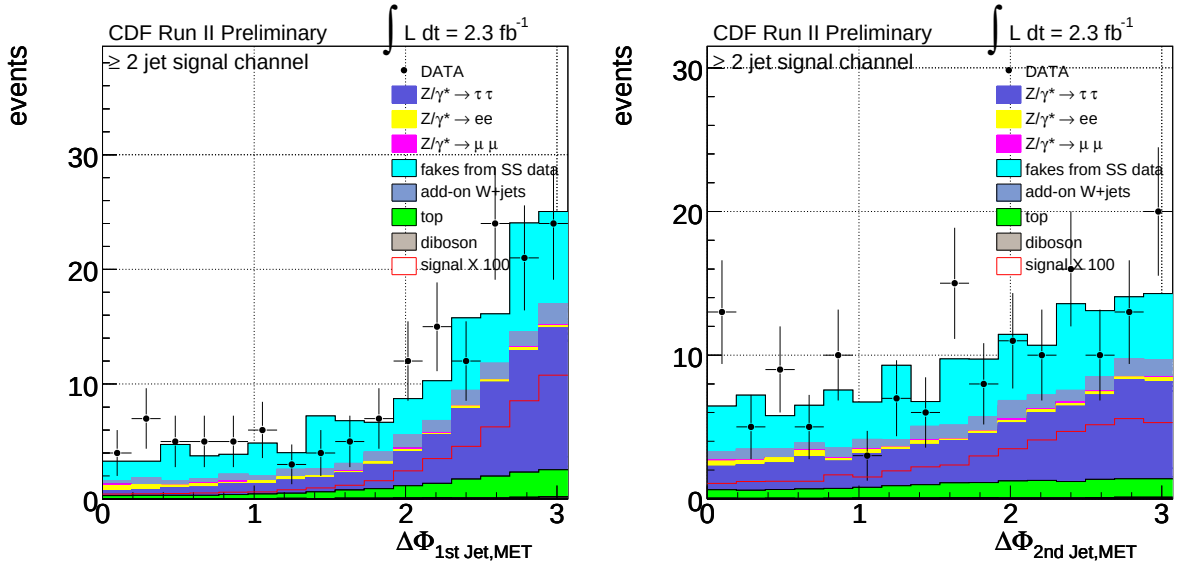

Figure 8.26: Angular separation in φ between the leading jet and $\not\!\!E_T$(left) and the second jet and $\not\!\!E_T$ (right) for the $\geq$2 jets channel.

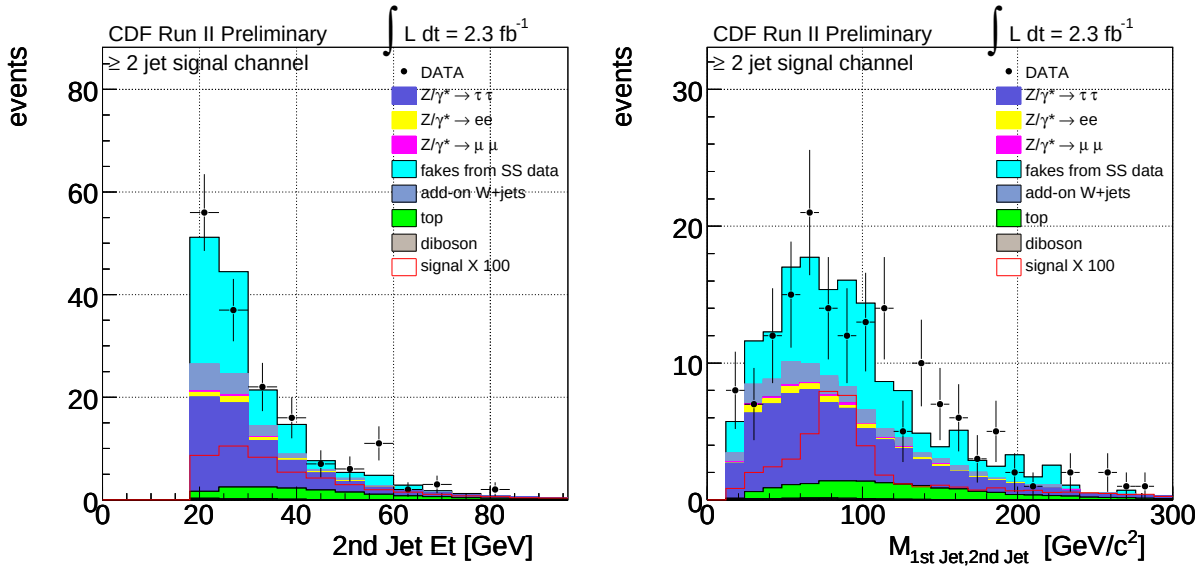

Figure 8.27: Corrected E_T of the second jet (left) and invariant mass of the 2-jet system (right) for the $\geq$2 jets channel.

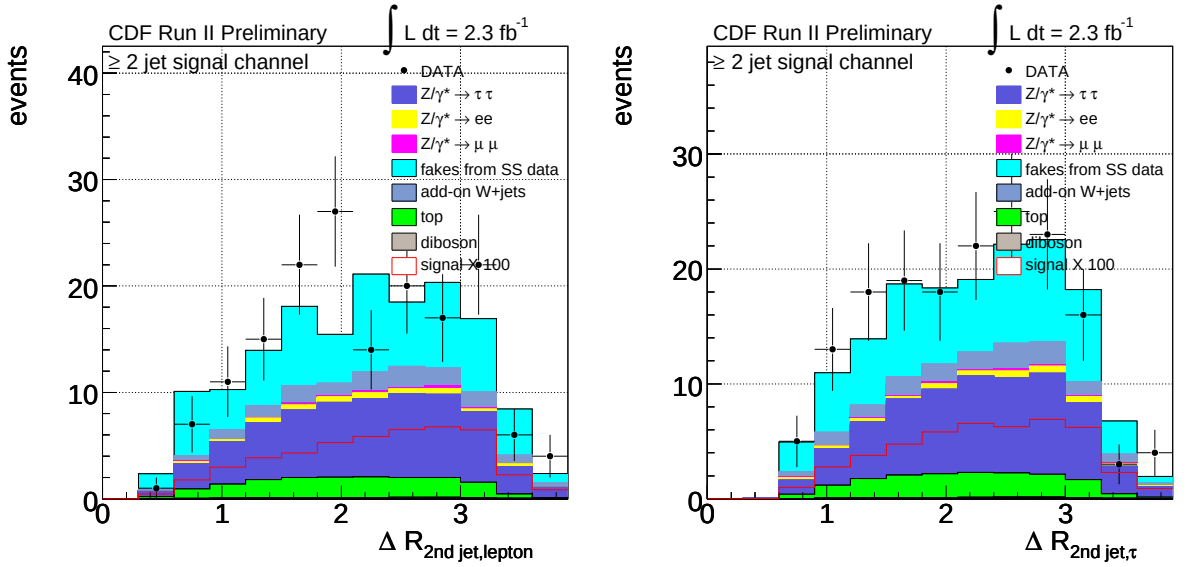

Figure 8.28: Separation in R between the second jet and the lepton (left) and between the second jet and the τ (right) for the $\geq$2 jets channel.

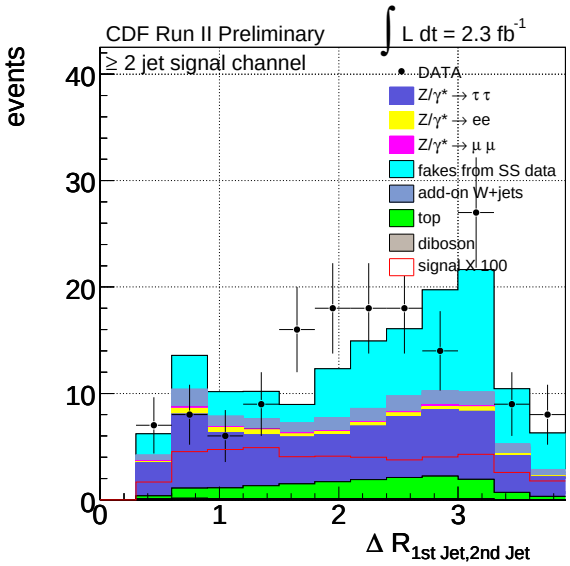

Figure 8.29: Separation in R between the two most energetic jets for the ≥2 jets channel.

8.4 Multivariate analysis: the BDT method

Given the large background contamination in the analysis channels, a simple counting experiment cannot provide any significant result to the Higgs search. In this context, we consider advanced statistical techniques to exploit all the information available in the event observables and take advantage of all the differences between the kinematic properties of the signal events with respect to the background ones: such techniques collect and combine the discriminating power of several distributions into one single variable, in such a way that the separation between signal and background is maximized.

We decided to use a multivariate method, the Boosted Decision Tree (see appendix B for a detailed description), the same adopted for the hadronic τ identification algorithm implemented in this analysis. However, in this case instead of simply applying a lower cut on the output score assigned to each signal and background event, we retain the entire final distributions of several BDTs and we exploit them to define a new discriminant variable which will then be used for the limit calculation. The procedure is described in details in the next sections.

8.4.1 Input variables

We examine the two signal channels separately and we train several independent BDTs, each of them optimized to get the best separation between the signal and one of the dominant sources of background which appear in the event selection, as shown in table 8.2: the training procedure is repeated for different Higgs mass hypotheses, ranging

from 110 GeV/c^2 to 150 GeV/c^2, in steps of 10 GeV/c^2. The training signal samples are defined by taking a weighted admixture of all the four Higgs production processes considered in this search.

	Channel	
	1 JET	$\geq$ 2 JETS
BDT 1	$Z/\gamma^* \rightarrow \tau\tau$	$Z/\gamma^* \rightarrow \tau\tau$
BDT 2	QCD	QCD
BDT 3	–	top-antitop

Table 8.2: Background processes for which each BDT is trained for the two signal channels.

The kinematical and topological variables we considered in the BDT training optimization procedure are:

- Lepton p_T: transverse momenta of the reconstructed electron or muon;
- τ p_T: transverse momentum of the reconstructed hadronically decaying τ;
- $\Delta R_{lep,\tau}$: separation in R between lepton and τ;
- $M_{lep,\tau}$: invariant mass of the τ-lepton system;
- $M^{approx}_{lep,\tau}$: invariant mass of the τ-lepton system in the collinear approximation for neutrinos' energy[1];
- $\not\!\!E_T$: missing transverse energy;
- $M_{T(lep,\not\!\!E_T)}$: transverse mass of the lepton and the $\not\!\!E_T$;
- $M_{lep,\tau,\not\!\!E_T}$ invariant mass of the lepton,τ and $\not\!\!E_T$ system;
- $H_T = \sum E_T$, the sum of the transverse energies of all the reconstructed objects in the event: $\not\!\!E_T$, lepton p_T, tau p_T, jet E_T;
- $\not\!\!E_T/\sqrt{\sum E_T}$: $\not\!\!E_T$ significance;
- $\Delta\varphi_{lep,\not\!\!E_T}$, angular separation in φ between the lepton and the $\not\!\!E_T$;
- $\Delta\varphi_{\tau,\not\!\!E_T}$, angular separation in φ between the τ and the $\not\!\!E_T$;
- $\Delta\varphi_{lep-\tau,\not\!\!E_T}$, angular separation in φ between the lepton-τ system and the $\not\!\!E_T$;
- E_T(Jet1), corrected(L5) transverse energy of the most energetic jet;
- E_T(Jet2), corrected(L5) transverse energy of the second most energetic jet[2];

- $\Delta R_{jet1,lep}$, separation in R between the most energetic jet and the lepton;
- $\Delta R_{jet1,\tau}$, separation in R between the most energetic jet and the τ;
- $\Delta\varphi_{jet1,\not\!\!E_T}$, angular separation in φ between the most energetic jet and the $\not\!\!E_T$;
- $\Delta R_{jet2,lep}$, separation in R between the the second jet and the lepton[2];
- $\Delta R_{jet2,\tau}$, separation in R between the second jet and the τ[2];
- $\Delta\varphi_{jet2,\not\!\!E_T}$, separation in R between the second jet and the $\not\!\!E_T$[2];
- $\Delta R_{jet1,jet2}$, separation in R between the two most energetic jets[2];
- $M_{jet1,jet2}$, invariant mass of the two most energetic jets, with energy corrected up to L7[2];

In the $\geq$2 jets channel, we calculate the dijet invariant mass after correcting the energies up to Level 7 (see section 6.4.4), since we want to go back to the original partons and reconstruct the peak corresponding to the hadronic decay of the W and Z bosons in the Higgs associated production.

The choice of BDT internal parameters (see table B.1 in appendix B) and the number of effective variables used for each BDT have been tuned and defined in order to get the best results in terms of signal efficiency, background rejection power and algorithm robustness and reliability. The latter criterion has been constantly verified by checking the consistency between the performances provided by the samples used for the training and the samples used to test the method. In some cases we increased the pruning strength[3] in order to reduce the occurrence of overtraining..

8.4.2 BDT performance distributions

The performance plots of the different trained BDTs, in the hypothesis of a Higgs boson mass of 120 GeV/c^2 are presented in this section. Pictures on the left show the output distributions for the Higgs signal and the specific background samples used in the training stage. Pictures on the right show the corresponding background rejection factor, as a function of the signal efficiency: each point of the curve corresponds to a specific lower cut applied on the BDT output.

As stated before, in this analysis search the BDT output distributions are combined to build templates, which are then used in the limit calculation. Therefore, performance plots are presented here just to provide us with an estimate of the separation power between the signal process and the source of background under consideration.

[1]The $\tau\tau$ invariant mass in the collinear approximation for neutrinos is defined and discussed in appendix C.

[2]Training variable used only for the $\geq$2 jets channel.

[3]Pruning: see definition in appendix B.

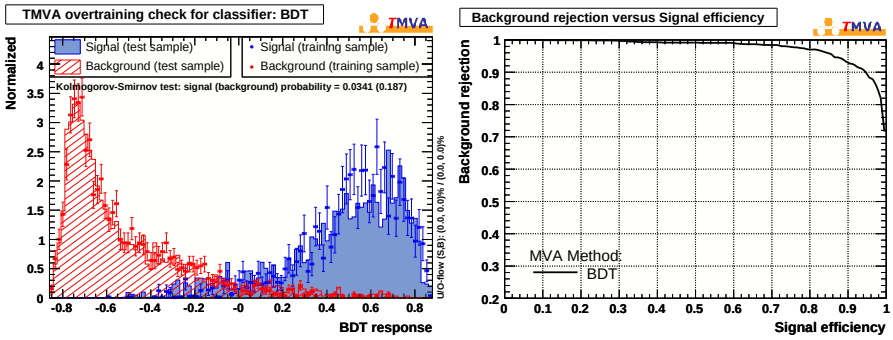

Figure 8.30: BDT output distribution (left) and background rejection vs. signal efficiency plot (right). BDT trained to separate Higgs signal events (M_H = 120 GeV/c^2) from QCD events in the 1 jet channel.

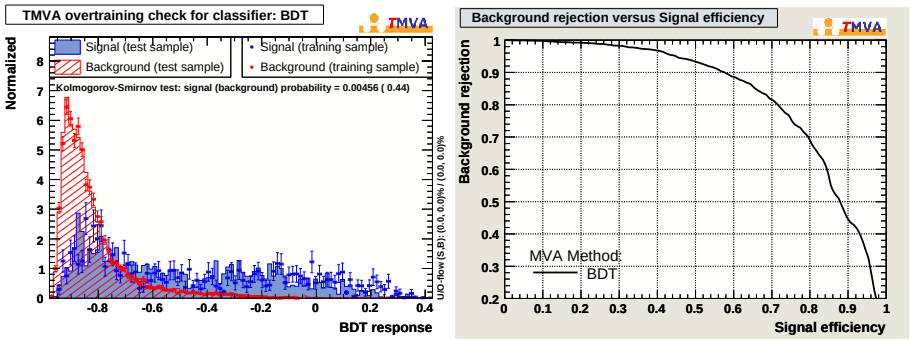

Figure 8.31: BDT output distribution (left) and background rejection vs. signal efficiency plot (right). BDT trained to separate Higgs signal events (M_H = 120 GeV/c^2) from Z$\rightarrow \tau\tau$ events in the 1 jet channel.

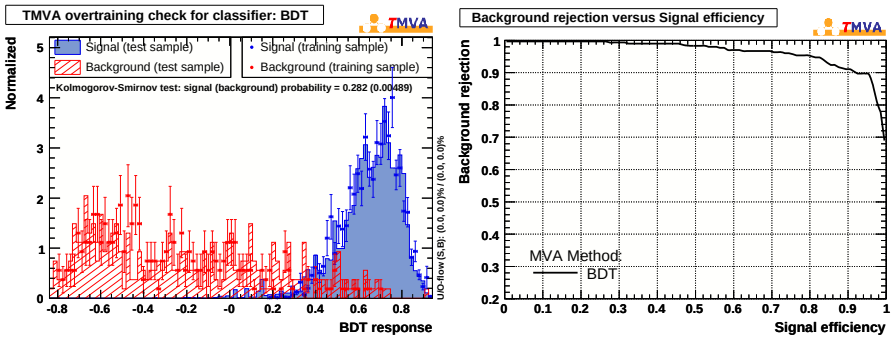

Figure 8.32: BDT output distribution(left) and background rejection vs. signal efficiency plot (right). BDT trained to separate Higgs signal events (M_H = 120 GeV/c^2) from QCD events in the $\geq$2 jets channel.

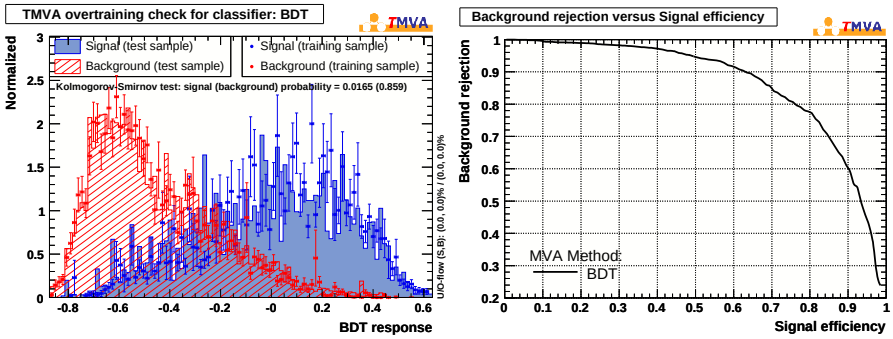

Figure 8.33: BDT output distribution (left) and background rejection vs. signal efficiency plot (right). BDT trained to separate Higgs signal events ($M_H = 120$ GeV/c^2) from Z$\rightarrow \tau\tau$ events in the $\geq$2 jets channel.

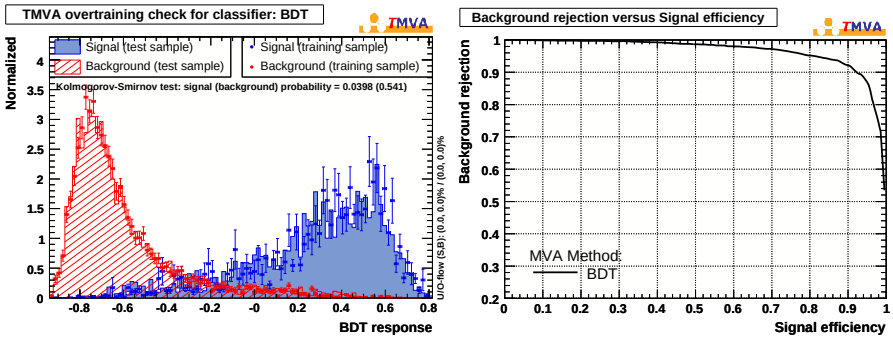

Figure 8.34: BDT output distribution (left) and background rejection vs. signal efficiency plot (right). BDT trained to separate Higgs signal events ($M_H = 120$ GeV/c^2) from $t\bar{t}$ events in the $\geq$2 jets channel.

8.4.3 Final discriminant templates

In this section the distributions of the BDT outputs, obtained by applying the trained BDTs (in the mass hypothesis of 120 GeV/c^2), to the different background, signal and data samples, are shown in figure 8.35 and 8.36. A multiplicative factor of 250 and 100 is applied to the expected signal rate, respectively in the 1 jet and $\geq$2 jet channel plot.

Starting from the set of BDT distributions previously described, several solutions have been studied to build the final discriminant templates (one for the 1 jet and another one for the $\geq$2 jets channels) used to calculate the limit to the Higgs cross section. It turned out that the choice which provides the best results in terms of search sensitivity is that one first adopted in a previous search [3, 106]: each event is simultaneously analyzed by the different BDTs and the minimum output score is retained. In this way, we ensure that a non-Higgs event, classified as background-like at least once, will fall in the left region of the BDT output admitted range; on the other side, a Higgs event will be on average recognized as signal-like by each BDT and

the minimum score will still lay in the right region of the output range. The result is an increase of the separation between signal and background and then an enhancement of the sensitivity.

Final plots obtained by adopting this procedure are shown in figure 8.37 (linear scale) and 8.38 (logarithmic scale). Distributions for the other Higgs boson masses are reported in appendix D. Because of computational limitations, the BDT training procedure was performed by considering only five hypothetical Higgs masses, while the final sensitivity have been evaluated in 21 different mass values, separated by 5 GeV/c^2 in the 100$\leq$M$_H$ $\leq$150 GeV/c^2 range: for the intermediate mass points the discriminant templates have been built by taking the set of BDTs trained with the closest greater hypothetical Higgs mass [1].

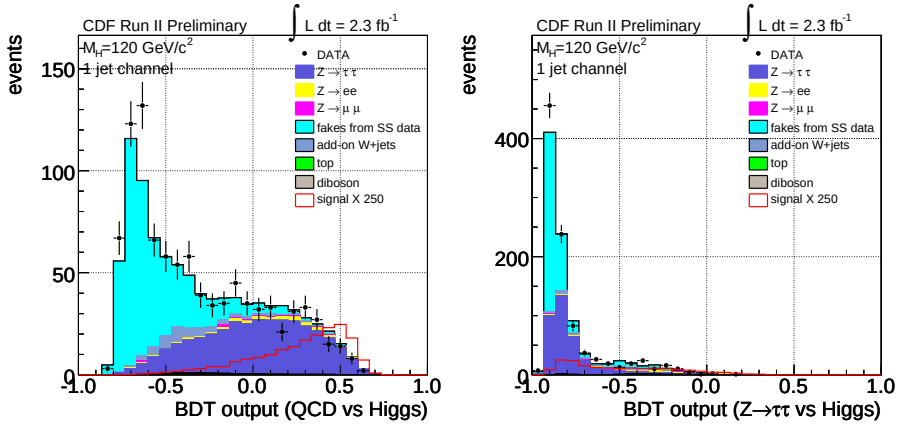

Figure 8.35: BDT templates for the Higgs mass hypothesis of 120 GeV/c^2, in the 1 jet channel. Left:signal against QCD; right: signal against Z$\rightarrow \tau\tau$.

[1]For example, the template for the 145 GeV/c^2 case is defined by using the BDTs trained to discriminate the background from a 150 GeV/c^2 Higgs signal.

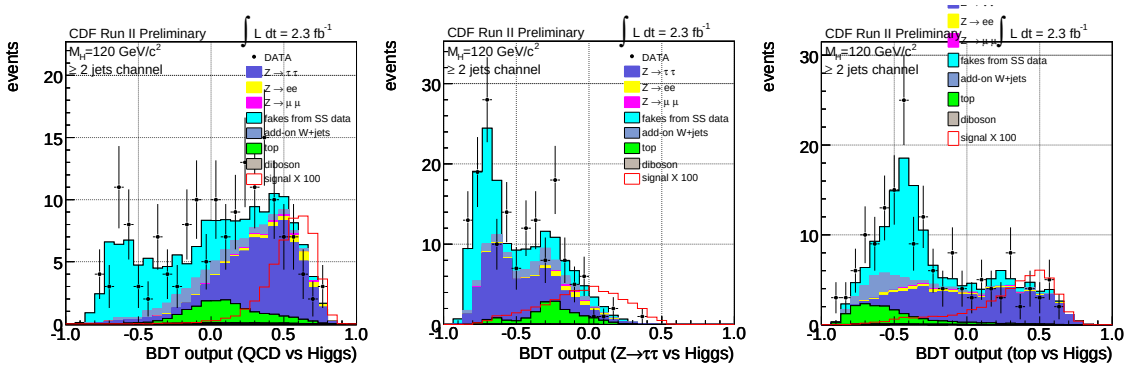

Figure 8.36: BDT templates for the Higgs mass hypothesis of 120 GeV/c^2, in the ≥2 jets channel. From the left to the right: signal against QCD; signal against Z→ $\tau\tau$, signal against $t\bar{t}$.

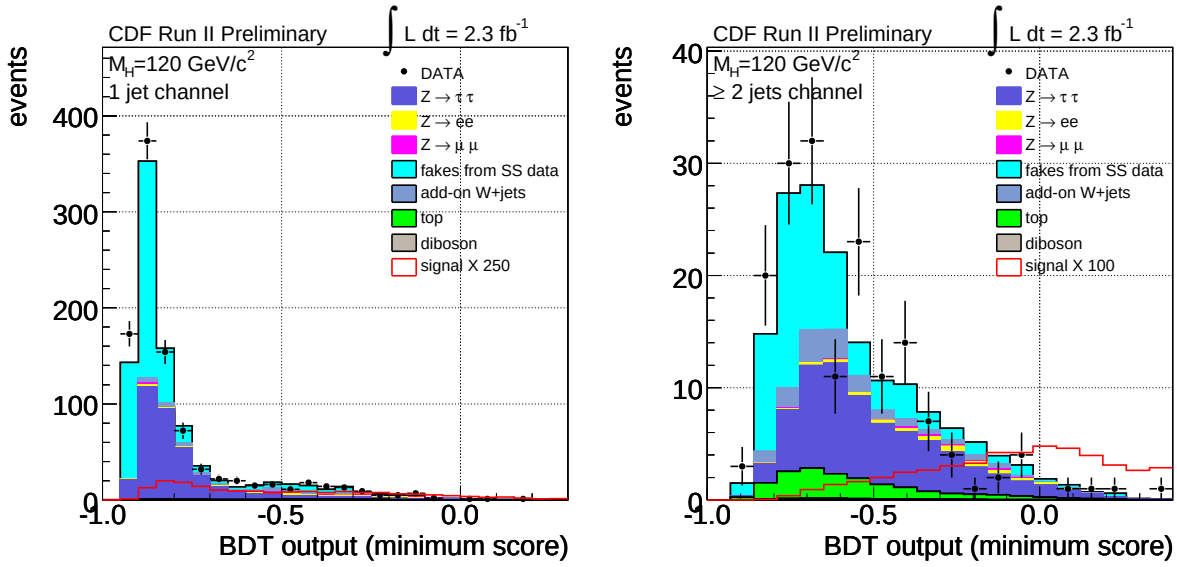

Figure 8.37: Final discriminant for the Higgs mass hypothesis of 120 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: ≥2 jets channel.

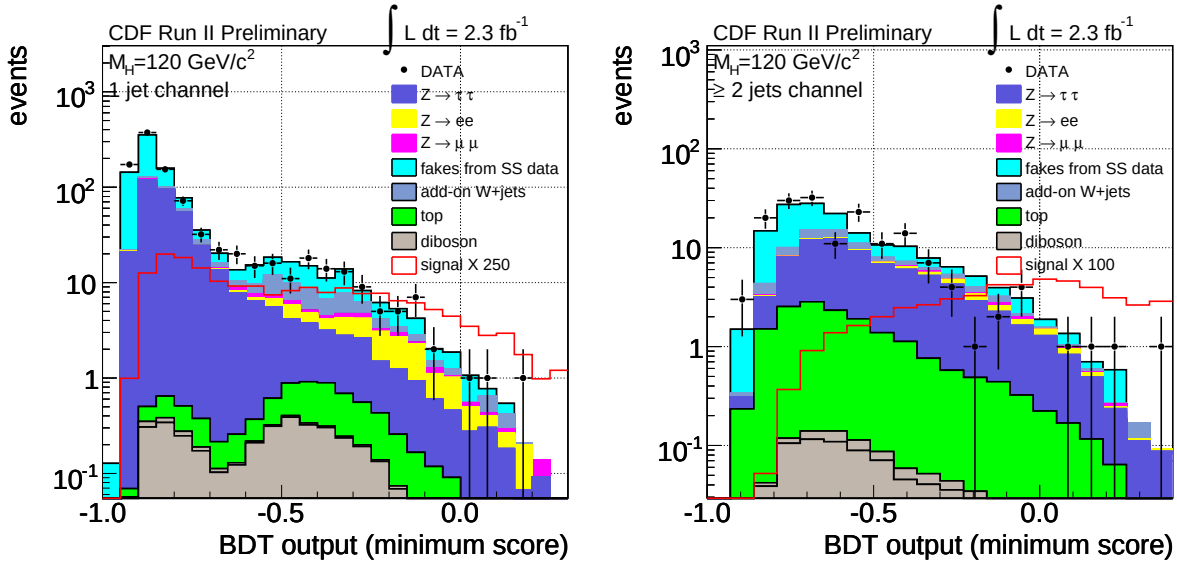

Figure 8.38: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 120 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

8.5 Systematic uncertainties

Systematic uncertainties affect the analysis search mainly in two different ways: they influence the normalization (rate) of background and signal expectations; they change the shape of the kinematic distributions and consequently that of the final discriminant used to set the upper limit to the Higgs cross section.

Since the event selection of this Higgs search is extremely sensitive to the number of reconstructed jets in the final state, the dominant source of systematics is represented by the jet energy scale, for which both rate and shape effects have been carefully estimated and taken into account throughout the analysis.

The complete list of systematic uncertainties which have been considered in this work is the following:

1. **Jet Energy Scale (JES)**: it refers to the set of corrections which are applied to the raw calorimeter measurements to estimate the momenta of particles which lie within the jet cone (level 5) or the momentum of the parent parton (level 7), as described in section 6.4.4. Following the instructions provided in [88], the effect of this source of uncertainty is estimated by applying $\pm 1\sigma$ shifts to the energy scale: due to the threshold of 20 GeV in the jet definition, event migrations between the different jet multiplicity channels are observed, thus affecting the kinematic acceptances of each signal and MC-derived background process.

2. **Cross Sections**: a rate systematic error due to the cross section uncertainty is assigned to each physics process estimated using MC samples. $t\bar{t}$ theoretical cross section is calculated at NNLO in [103], with an uncertainty of 10%. MCFM [36] has been used to compute a NLO calculation for the diboson processes (WW, WZ and ZZ) [102], with an uncertainty of 6%. The experimental uncertainty of a recent CDF measurement [101] is assigned to Drell-Yan events. For the Higgs signal processes, theoretical cross section uncertainties have been already discussed in section 3.2.

3. **PDF functions**: the uncertainty associated to the specific choice of the parton distribution functions used in the MC generation (CTEQ5L), is estimated by applying a re-weight procedure to the MC samples. A first step consists in re-weighting each event to several alternative sets of PDFs (MRST72, MRST75 and 20 orthogonal CTEQ6M sets); then the acceptance differences between pairs of CTEQ6M eigennvectors and the reference PDF set, between MRST72 and CTEQ5L and between MRST72 and MRST75 are evaluated; the total PDF uncertainty is finally calculated by summing the latter difference in quadrature to the larger between the first two. The effect is of the order of a few percent. In this analysis we quote the values which have been calculated in [106].

4. **ISR/FSR**: dedicated MC samples for Higgs signal events at the reference mass of 120 GeV/c^2 are generated, with an increased and decreased strength of the ISR

and FSR in the parton showering process, by $\pm 1\sigma$, with respect to the default value. The ISR/FSR uncertainty is evaluated by taking half of the acceptance difference between the two cases.

5. **Jet$\rightarrow \tau_h$ fakes**: an uncertainty of 10% is applied to the background due to jet$\rightarrow \tau_h$ fakes derived from the same sign data, in order to take into account of the discrepancies observed in the QCD control region (see section 7.6.1).

6. **W+jets normalization**: the limited data statistics affects the normalization procedure of the W+jets contribution which comes from the asymmetry between opposite sign and same sign events, as described in section 7.5.1: an uncertainty of 5%, 18% and 30% is computer for the 0 jet, 1 jet and $\geq$ 2 jets channels, respectively.

7. **Drell-Yan MC acceptance**: an additional uncertainty to the rate of Drell-Yan processes arises from the difference observed between ALPGEN and PHYTIA generators. Half of the correction factor applied to ALPGEN acceptance (see section 7.4.1) is taken as systematic error.

8. **Tau ID scale factor**: as outlined in section 7.6.2, the performances of the new τ identification algorithm have been tested by comparing the observed data with theoretical predictions, in a control sample specifically designed to enhance the $Z/\gamma^* \rightarrow \tau\tau$ contribution. When applying the τ ID efficiency scale factor (formula 7.7) to the expression for the estimation of the MC-based physics processes yield(formula 7.1), we obtain

$$N^i = \sigma^i \times A^i \times \frac{N_{\rm obs} - N^{\rm QCD}_{\rm SSdata} - N_{\rm W+jets}}{A \times \sigma(Z \rightarrow \tau\tau) + \sum_{\rm jbkg=1}^{N} (\sigma^j \times A^j)} \tag{8.1}$$

and we observe that some sources of systematics errors that enter in other CDF analyses cancel out. They are the data luminosity, the Z vertex cut, the trigger and electron/muon identification efficiencies, which areincorporated within the uncertainty of the τ ID efficiency.

The quantities which affect the τ ID scale factor and thus propagate their uncertainty to the MC event rate N^i are reported in table 8.5.

In tables 8.4 and 8.5 we summarize the different contribution to the systematic uncertainties for each background and signal process. Values are expressed in %; a negative sign means that a positive one sigma variation of a specific uncertainty source produces a rate reduction to the considered physics process. As can be seen, Drell-Yan backgrounds have a double dependence on the cross section and acceptance systematics: the first dependence is the standard one, the second arises from the way τ ID scale factor is calculated in the $Z/\gamma^* \rightarrow \tau\tau$ control sample.

Uncertainty sources for the τ ID efficiency	
$N_{\rm obs}$	# of observed event in the $Z/\gamma^* \to \tau\tau$ C.R.
$N^{\rm QCD}_{\rm SSdata}$:	# of SS events in the $Z/\gamma^* \to \tau\tau$ C.R.
$N_{\rm W+jets}$	W+jets contribution in the $Z/\gamma^* \to \tau\tau$ C.R.
$\sigma(Z\to \tau\tau)$	DY cross section
$A(Z\to \tau\tau)$	acceptance in the $Z/\gamma^* \to \tau\tau$ C.R.

Table 8.3: Systematic uncertainties which affect the estimation of the τ ID efficiency scale factor. C.R. stands for Control Region.

Source		$Z/\gamma^* \to ll$	$t\bar{t}$	WW/WZ/ZZ	fakes from SS	W+jets
JES	(0 jet)	−0.6	−19.0	−0.9		
	(1 jet)	6.2	−7.7	7.1		
	(≥2 jets)	14.2	3.2	11.7		
Cross section		2.2	10.0	6.0		
PDF		1.0	1.0	1.0		
SS data					10.0	
W+jets scale	(0 jet)					5.0
	(1 jet)					18.0
	(≥2 jets)					30.0
$A(Z\to \tau\tau)$		2.3				
tau ID SF:						
$N_{\rm obs}$		2.8	2.8	2.8		
$N^{\rm QCD}_{\rm SSdata}$		−3.3	−3.3	−3.3		
$N_{\rm W+jets}$		−0.3	−0.3	−0.3		
$\sigma(Z\to \tau\tau)$		−2.1	−2.1	−2.1		
$A(Z\to \tau\tau)$		−2.2	−2.2	−2.2		

Table 8.4: Systematic uncertainties on the background processes, expressed in %.

8.6 Limit calculation

As we have seen in the figures reported in sections 8.2 and 8.3, the observed data is reasonably consistent with the SM predictions; no Higgs component and no significant excess is observed in any kinematical and topological distribution. An interpretation of these results is given by estimating the upper limit on the possible Higgs signal contribution at a specific confidence level (C.L.).

This section provides a description of the method implemented to determine such upper limit at 95% C.L. on the Higgs production cross section. The method, which represents the standard choice for the Higgs analyses performed by the CDF collabo-

Source		ggH	WH	ZH	VBF
JES	(1 jet)	5.1	−4.8	−5.3	−3.7
	(≥2 jets)	13.2	5.4	4.8	5.2
cross section	(1 jet)	23.5	5.0	5.0	10.0
	(≥2 jets)	67.5	5.0	5.0	10.0
PDF		4.9	1.2	0.9	2.2
ISR	(1 jet)	13.0	−6.1	−1.7	−2.9
	(≥2 jets)	15.5	−1.5	0.1	−2.7
FSR	(1 jet)	−5.0	4.3	1.0	1.7
	(≥2 jets)	−5.2	−2.1	0.4	−1.1
tau ID SF:					
N_{obs}		2.8	2.8	2.8	2.8
$N^{\text{QCD}}_{\text{SSdata}}$		−3.3	−3.3	−3.3	−3.3
$N_{\text{W+jets}}$		−0.3	−0.3	−0.3	−0.3
$\sigma(\text{Z}\rightarrow \tau\tau)$		−2.1	−2.1	−2.1	−2.1
$A(\text{Z}\rightarrow \tau\tau)$		−2.2	−2.2	−2.2	−2.2

Table 8.5: Systematic uncertainties on the signal, expressed in %. For JES we quote the average values calculated among the different Higgs masses.

ration [105], is based on a pure bayesian approach; it allows us to easily combine the results of several searches performed on statistically independent samples and it naturally accommodates the treatment of systematic uncertainties (and their correlation among different channels) in the limit calculation.

The starting point is the definition, for a given Higgs boson mass, of the combined likelihood function $\mathcal{L}$, as the product of Poisson probabilities of observing n_{ij} events in the j-th bin of the i-th input histogram, given an expectation value μ_{ij}:

$$\mathcal{L}(R\,\vec{s},\vec{b}|\vec{n},\vec{\theta}) = \prod_{i=1}^{N_c}\prod_{j=1}^{N_b}\frac{e^{-\mu_{ij}}\mu_{ij}^{n_{ij}}}{n_{ij}!} \tag{8.2}$$

where the first product is over the total number of channels, N_c, and the second one is over the N_b histogram bins; one input binned histograms is generally employed for each search channel. For this analysis we use the final discriminant distributions of figure 8.37.

The expected bin contents are parametrized as the sum of signal $s_{ij}(\vec{\theta})$ and background $b_{ij}(\vec{\theta})$ expectation

$$\mu_{ij} = R \times s_{ij}(\vec{\theta}) + b_{ij}(\vec{\theta}) \tag{8.3}$$

where R is an artificial scaling factor applied to the signal, used to test the sensitivity level of the search. Given no experimental information on the Higgs production cross

section, a flat prior distribution is assigned to R. Each of the signal and background contribution depend on the so called *nuisance* parameters θ_k, which are introduced to incorporate in the likelihood the systematic uncertainties associated to each production rate, as well as the uncertainties related to the distribution shapes:

$$s_{ij}(\vec{\theta}) = s_{ij}^{\text{central}} \prod_k (1 + u_{ij}^k \theta_k) \tag{8.4}$$

$$b_{ij}(\vec{\theta}) = b_{ij}^{\text{central}} \prod_k (1 + u_{ij}^k \theta_k) \tag{8.5}$$

where s^{central} and b^{central} are the central values for the expected event yields of each signal and background contribution, while u_{ij}^k is the relative uncertainty due to the k-th nuisance parameter θ_k.

Nuisance parameter are independent one from each other, but most of them affect all the signal and background predictions in any of the channels: their values are constrained to Gaussian distributions with zero average and unit width[2], by adding to the likelihood the appropriate multiplicative terms

$$\mathcal{L}(R\,\vec{s},\vec{b}|\vec{n},\vec{\theta}) = \prod_{i=1}^{N_c}\prod_{j=1}^{N_b} \frac{e^{-\mu_{ij}}\mu_{ij}^{n_{ij}}}{n_{ij}!} \prod_{k=1}^{N_{np}} e^{-\theta_k^2/2} \tag{8.6}$$

where N_{np} represents the total number of nuisance parameters.

Correlations among different analyses provided by the systematic uncertainties are in such a way naturally included in the limit calculation.

The likelihood is then integrated over the values of each nuisance parameter and evaluated for the observed number of events n_{ij}. This procedure is called *marginalization* and the result, normalized to unit area, is a posterior density function where the only left dependence is on the parameter R. The 95% C.L. upper limit is then estimated by calculating the value of R that subtend the 95% of the area of the the resulting distribution:

$$\int_0^{R_{95}} \mathcal{L}(R)dR = 0.95 \tag{8.7}$$

R_{95} corresponds to the limit on the production cross section of the Higgs boson, normalized to the SM prediction.

8.6.1 Estimated and observed limits

Before computing at the observed limits it is useful to estimate the sensitivity of the analysis by generating simulated data in a background only hypothesis. Pseudo experiments (PE) are randomly generated by fluctuating each background source with a

[2]A lower truncation is applied to the Gaussian constraints, in order to prevent from negative predictions of s_{ij} and b_{ij}, which could occur when the nuisance parameters are allowed to fluctuate.

Poisson function, with the mean given by the expected rate. Systematic uncertainties are also included.

Each PE is then used to evaluate the 95% C.L. limit, as it is done for real data, following the procedure described in the previous section. All the results obtained for a specific hypothetical Higgs mass are then collected in a single histogram, and the expected limit is estimated by taking the median of the distribution, while $\pm 1\sigma$ and $\pm 2\sigma$ variations are given by defining the intervals containing 68% and 95% of the total area. Examples of the resulting distributions of the estimated limits are shown in figure 8.39 for the 120 GeV/c^2 and 140 GeV/c^2 masses; for each of the eleven masses under consideration, a set of 10000 PE have been generated and the complete set of histograms is reported in appendix E.

Expected and observed limits are calculated in this search by implementing in the analysis code the MCLIMIT package, whose description is provided in [107].

The results for the expected and the observed limits are shown in linear and logarithmic scale in figures 8.40, as a function of the Higgs mass: values are expressed in units of the SM expectation. A summary is also reported in table 8.6 for the eleven Higgs boson masses varying from $M_H = 100$ GeV/c^2 to $M_H = 150$ GeV/c^2, in steps of 5 GeV/c^2. As can be seen, the observed limits are well in agreement with the expected ones, being within 1σ in the entire explored mass range. The maximum sensitivity is reached at the reference mass of 120 GeV/c^2, where the expected limit is $23.4^{+9.8}_{-6.4}$, while the corresponding observed value is 27.2. The sensitivity decreases as the Higgs mass becomes bigger, mainly because of the reduced H$\rightarrow \tau\tau$ branching ratio and production cross section.

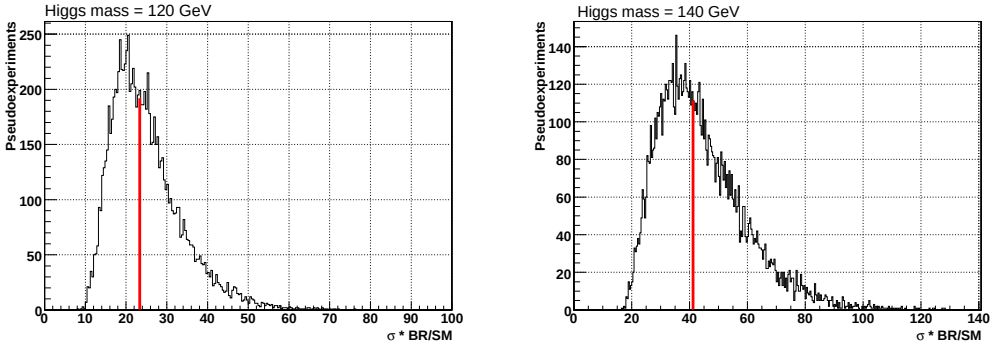

Figure 8.39: Expected distribution of 95% C.L. limits in the background-only hypothesis. Left: M_H=120 GeV/c^2; Right: M_H=140 GeV/c^2. The red line shows the median of the distribution; $\pm 1\sigma$ and $\pm 2\sigma$ intervals around the median are obtained by taking 68% and 95% of the total area, respectively.

Higgs Mass	Expected limit/σ(SM)					Observed
GeV/c^2	-2 σ	-1 σ	median	+ 1 σ	+ 2 σ	limit/σ(SM)
100	15.2	20.1	28.2	40.4	56.6	37.6
105	13.4	17.8	25.2	35.2	48.0	34.5
110	13.1	17.2	23.9	34.2	48.2	34.8
115	13.8	17.7	24.5	35.4	50.2	27.9
120	13.1	17.0	23.4	33.2	46.8	27.2
125	14.5	18.8	26.5	37.8	52.6	25.3
130	15.5	20.3	28.1	40.3	56.9	30.0
135	18.4	23.8	33.7	48.2	65.6	30.3
140	22.6	29.6	41.2	58.6	81.2	38.2
145	31.2	41.0	57.4	81.9	114.8	46.3
150	45.1	59.1	82.6	118.0	166.5	67.0

Table 8.6: Summary table of the expected and observed 95% C.L. limit for the H$\rightarrow \tau\tau$ search, covering 2.3 fb^{-1} of CDF data. Values are all expressed in units of the theoretical SM cross section. The expected limit and the $\pm1\sigma$ and $\pm2\sigma$ intervals are obtained by generating 10000 pseudo experiments in the background-only hypothesis.

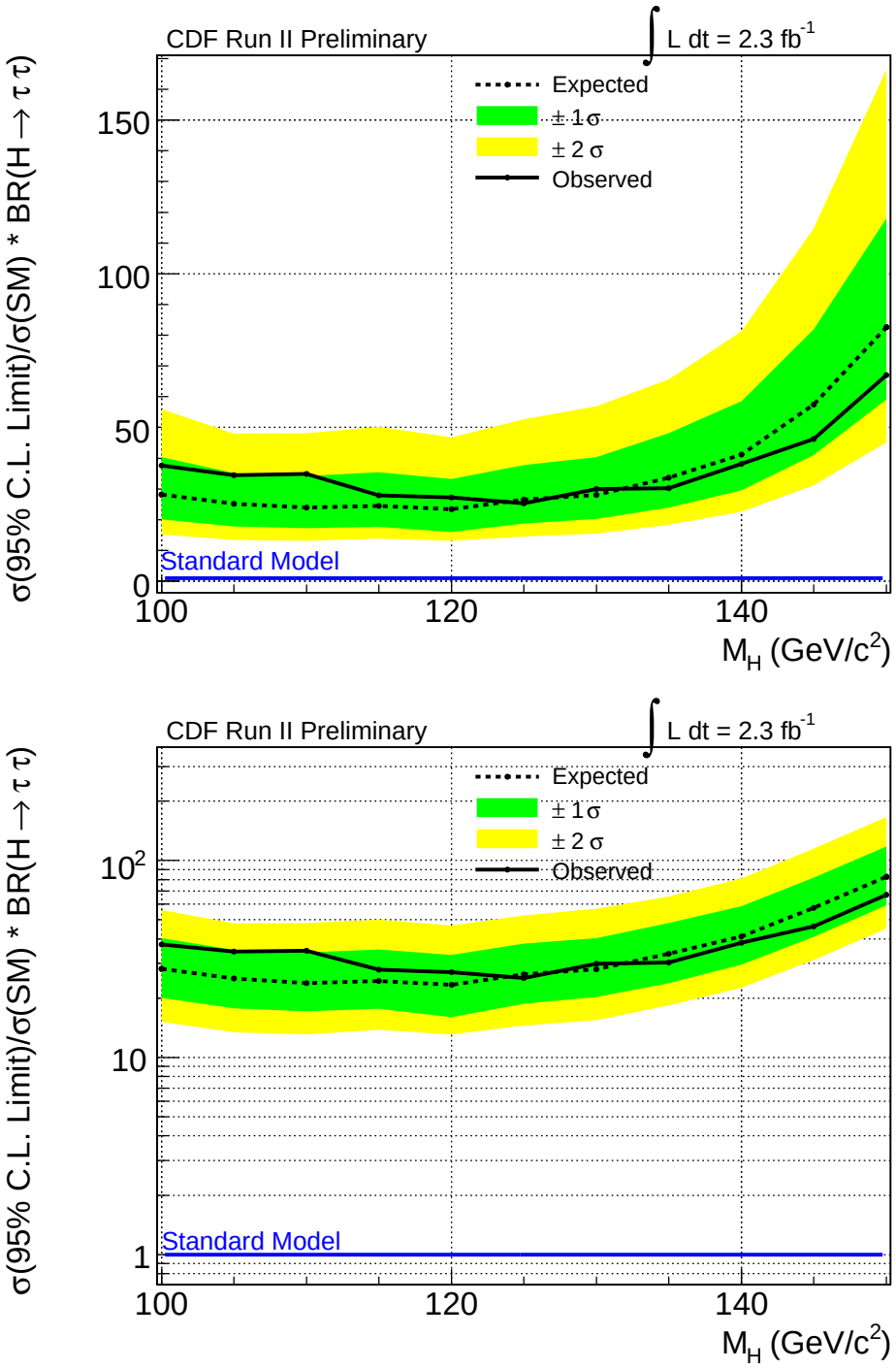

Figure 8.40: The observed and expected 95% C.L. upper limits on the Higgs production cross section, as a function of m_H, for the H$\rightarrow \tau\tau$ search, covering 2.3 fb^{-1} of CDF data. Values are expressed in units of the theoretical SM cross section. The solid and dashed black curves represent the observed and the median expected values respectively; the colored bands show the 68% (green) and 95% (yellow) probability regions. The top plot is in linear scale, while the bottom one is in logarithmic scale.

Chapter 9

Conclusions

9.1 Analysis summary

We have reported in this thesis the direct search for the Standard Model Higgs boson in the two τ lepton final state, performed at the Tevatron proton-antiproton collider. The presented results are based on 2.3 fb^{-1} of data collected at the center of mass energy $\sqrt{s} = 1.96$ TeV by the CDF detector and involve several signal processes, which have been studied simultaneously: gg$\rightarrow$H$\rightarrow \tau\tau$, WH($\rightarrow \tau\tau$), ZH($\rightarrow \tau\tau$) and qHq'$\rightarrow$q$\tau\tau$q'.

As described in chapter 6, the analysis has been optimized for the detection of one τ decaying hadronically and the other one leptonically. The selection of the $\tau_h\tau_e$ and $\tau_h\tau_\mu$ modes covers about 46% of the possible di-τ decay combinations, as summarized in table 6.1. This choice represents the best compromise between the large background suppression provided by the very pure and efficient electron and muon identification, and the high B.R. of the hadronic τ decay.

The request of at least one additional jet in the final state maximizes the sensitivity to the vector boson fusion and the associated production signal channels, where jets arise from the hadronization of the outgoing quarks in the calorimeters. Even though no jets are expected at the tree level for the direct production mechanism, Higgs events from gluon fusion are nevertheless selected, when the process is accompanied by initial state radiation: acceptance for this channel is therefore small, but signal yield is comparable to the other processes because of the much higher production cross section.

This analysis improves the results of a previous search performed by the CDF collaboration [106]: the major changes consist in the increase of the data statistics, in the extension of the signal acceptance (obtained by including the 1 jet channel) and in the reduction of jet$\rightarrow \tau_h$ background provided by the implementation of a new τ identification algorithm. In addition, a new statistical advanced method has been adopted, to further increase the Higgs search sensitivity.

In chapter 7 we provided a description of the techniques adopted to model the different SM background sources which contribute to the event selection. These processes are categorized into two classes, according to the origin of the reconstructed leptons: the physical irreducible backgrounds, yielding two real leptons, are represented by

$Z \rightarrow \tau\tau$, diboson and top pair processes and have been described by MC simulations; events containing one or two jets which are misidentified as leptons have been modeled by applying data driven procedures.

Events with no jet activity have been used to build different control samples and test the agreement between the observed data and the expected background rates.

We then implemented a multivariate method, based on sets of Boosted Decision Trees, to increase the separation between signal and background in the two signal channels, defined as the events with one or more than one jet in the final state (chapter 8). Unfortunately, no evidence for the Higgs boson has been found: the expected background yields are 921.7±48.9 in the 1 jet channel and 159.4±11.6 in the ≥ 2 jets channel, while the observed event rates are 965 and 166, respectively.

We finally applied a bayesian technique to compute a 95% C.L. upper limit to the Higgs production cross section times the branching ratio of the decay to τ pairs, for each hypothetical Higgs mass varying from 100 to 150 GeV/c^2, as reported in table 8.6 and shown in figure 8.40.

The analysis described in this thesis helps to increase the Tevatron Higgs search sensitivity in the low mass region, where the branching fraction of the decay process into τ leptons, ranging from about 8% at $M_H = 100$ GeV/c^2 to 2% at $M_H = 150$ GeV/c^2, actually gives a non negligible contribution.

9.2 The CDF Higgs analyses combination

Several Higgs searches performed at CDF contribute to improve the overall sensitivity in the entire Higgs mass range allowed by the theoretical and experimental constraints [105]. The efforts have recently increased in particular in the low mass region, both by improving the analysis techniques and by adding new channels previously unexplored. The limits on the Higgs cross section are now getting close to the SM expectations, thanks to the contribution of a large set of analyses, which cover almost all the Higgs production modes and decay channels.

The results obtained in the different searches are combined by applying the bayesian method described in section 8.6, where all the systematic uncertainties and their correlations among the specific analyses are also taken into account. The CDF combined 95% C.L. upper limit on the Higgs production cross section is shown in figure 9.1, as a function of the Higgs mass. The corresponding values of the expected and observed limits, expressed in units of the SM cross section, are reported in table 9.1 for each mass point. The results refer to the searches performed up to July 2010 and cover up to 5.9 fb^{-1} of collected Run II data [105]. Figure 9.2 shows the individual contributions of the different analyses which enter in the combination.

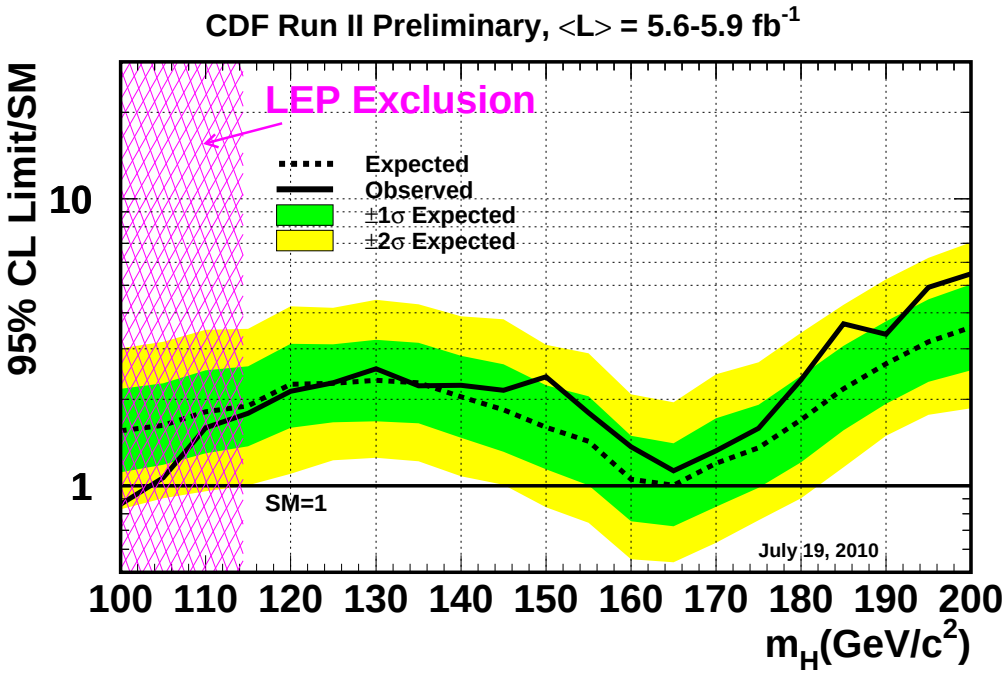

Figure 9.1: The observed and expected 95% C.L. upper limit on the Higgs production cross section, as a function of m_H, obtained by combining all CDF's SM Higgs search channels. Values are expressed in units of the SM expectation. The solid and dashed black curves represent the observed and the median expected values, respectively; the colored bands show the 68% (green) and 95% (yellow) probability intervals. The statistical method employed to extract the limits is described in section 8.6.

9.3 The Tevatron combination

The Tevatron sensitivity to the SM Higgs boson is obtained by combining the search channels of both the CDF and DØ experiments. DØ has performed several analyses, with a sligthly larger integrated luminosity, up to 6.7 fb^{-1} [108].

Two types of combination techniques have been independently applied, in order to verify that the final results don't depend on the details of the statistical method: the first technique uses the bayesian formulation described in section 8.6, the second one is based on a modified frequentist approach [26, 109, 110]. Both methods exploit the same final discriminant distributions of each analysis channel and combine the different sources of systematic uncertainties by considering the appropriate correlations among the two experiments. The methods provide similar results, within 10%, with the modified frequentist approach being a little bit more optimistic on average.

At the moment, the latest results on the combined 95% C.L. upper limits [26] allow to exclude the presence of the Higgs boson at 95% C.L. in the mass ranges

Higgs Mass	Expected limit/σ(SM)					Observed
GeV/c^2	-2 σ	-1 σ	median	+ 1 σ	+ 2 σ	limit/σ(SM)
100	0.83	1.11	1.55	2.18	3.02	0.86
105	0.90	1.18	1.63	2.28	3.17	1.07
110	0.96	1.30	1.81	2.53	3.50	1.59
115	1.00	1.37	1.90	2.61	3.52	1.79
120	1.10	1.59	2.26	3.12	4.22	2.13
125	1.23	1.66	2.28	3.11	4.18	2.28
130	1.25	1.68	2.33	3.23	4.44	2.56
135	1.22	1.65	2.28	3.15	4.30	2.23
140	1.08	1.47	2.04	2.84	3.90	2.24
145	1.01	1.31	1.84	2.65	3.80	2.16
150	0.84	1.14	1.60	2.24	3.10	2.40
155	0.74	1.01	1.43	2.05	2.90	1.79
160	0.55	0.75	1.05	1.49	2.08	1.37
165	0.54	0.72	1.00	1.41	1.96	1.13
170	0.63	0.84	1.20	1.72	2.45	1.32
175	0.76	0.98	1.36	1.92	2.69	1.59
180	0.91	1.21	1.70	2.42	3.41	2.34
185	1.16	1.56	2.18	3.07	4.27	3.66
190	1.49	1.93	2.66	3.75	5.25	3.37
195	1.77	2.30	3.17	4.46	6.24	4.92
200	1.87	2.53	3.56	5.04	7.06	5.47

Table 9.1: Summary table of the expected and observed 95% C.L. limits for all the CDF Higgs boson search channels combined, between 100 and 200 GeV/c^2. The limits are all given in units of the expected SM cross section.

158<m$_H$ <175 GeV/c^2 and 100<m$_H$ <109 GeV/c^2, as shown in figure 9.3 and summarized in table 9.2.

The impact of the combined Tevatron results on the Higgs boson searches has been evaluated by the GFITTER group [10], which performed a global fit to the Higgs mass, similar to the one described in section 3.1.3; in this case the constraints coming from the Higgs direct searches are added as input to the other SM observables. The minimum χ^2 is 17.46 for 14 degrees of freedom, which corresponds to a p-value of 21.0%. This means that, given the results provided by the searches performed at the LEP and Tevatron, the SM is still compatible with the presence of the Higgs boson, even if it is excluded in the high mass region. The $\Delta\chi^2$ curve of the global fit, as a function of m$_H$, is shown in figure 9.4. The preferred value for the Higgs boson mass is m$_H$ = $120.6^{+17.9}_{-5.2}$ GeV/c^2.

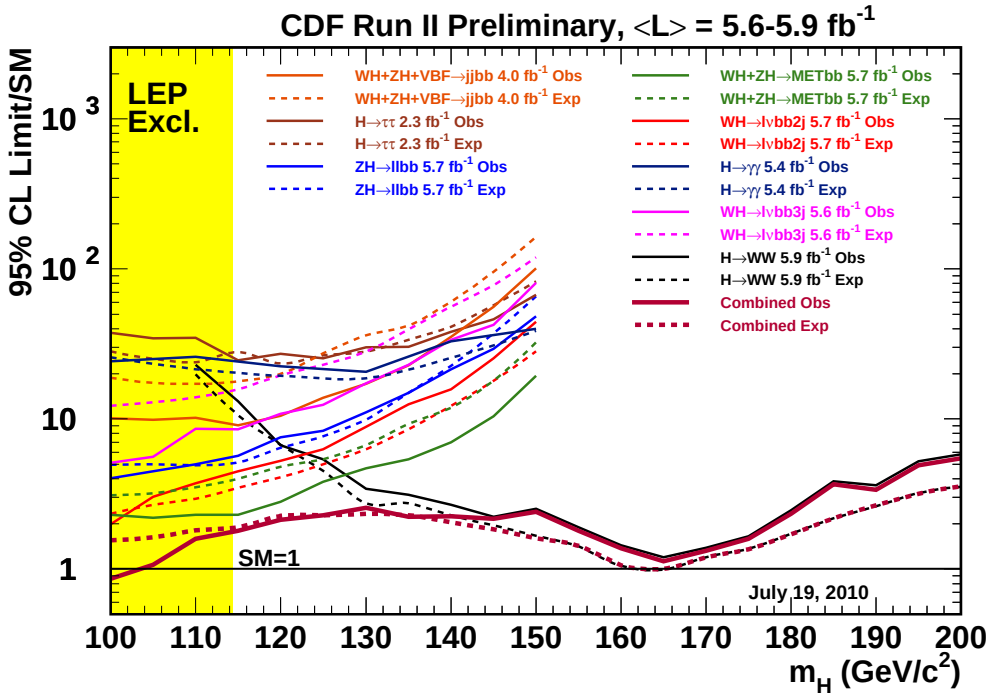

Figure 9.2: The observed and expected 95% C.L. upper limits on the Higgs production cross section, as a function of m_H, shown separately for each CDF analysis and for the combination (thick red line). Values are expressed in units of the SM expectation. The solid and dashed lines indicate the observed and the median expected limits, respectively.

9.4 Future prospects

The Tevatron collider has delivered more than 10 fb^{-1} of integrated luminosity to both CDF and DØ detectors from 2001 to the end of 2010. The steady improvements occurred over the years to each stage of the accelerator chain led to an increase of the Tevatron performances in terms of initial instantaneous luminosity, beam stability and stores quality. These improvements resulted ultimately in a continuous increase of the rate of the integrated luminosity made available to the two experiments. This can be clearly seen in figure 9.5, where the accumulated luminosity is reported, as a function of time, separately for the different years since the beginning of Run II.

The Tevatron is scheduled to provide proton-antiproton collisions at least up to September 2011 and we can reasonably assume that in this time period about 2.5 fb^{-1} of data will be delivered.

As described in chapter 4, the data taking efficiency of the CDF detector is steadily around 85%. This value reduces by about 10% when quality requirements are applied on the functionality of the different subdetectors. This means that CDF could real-

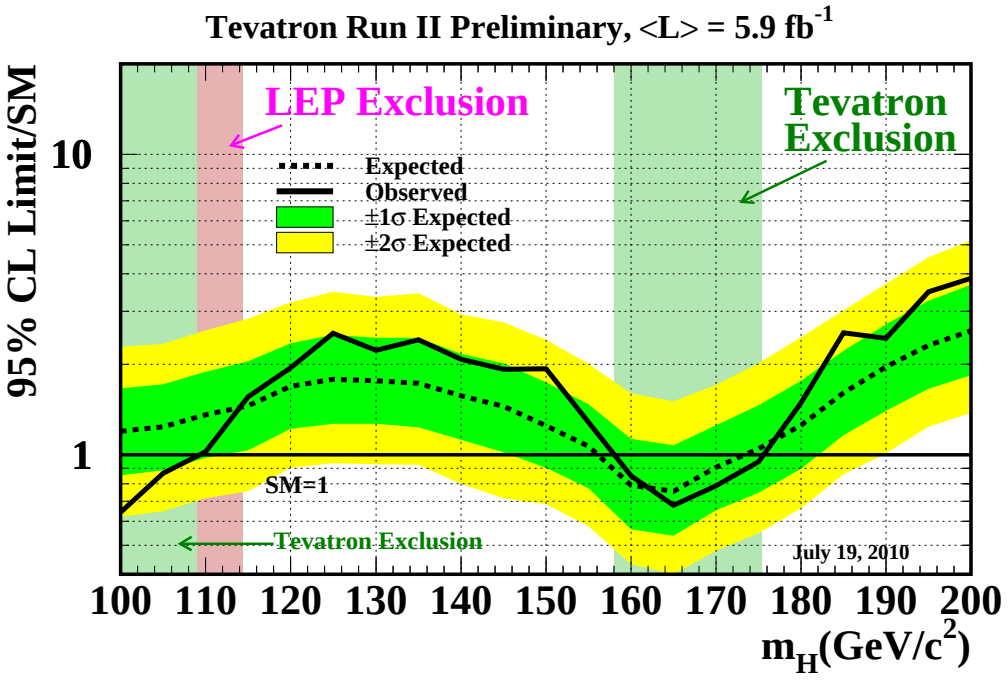

Figure 9.3: The observed and expected 95% C.L. upper limits on the Higgs production cross section, as a function of m_H, obtained by combining all CDF and DØ analyses. Values are expressed in units of the SM cross section and are obtained with the bayesian calculation described in section 8.6. The solid and dashed black curves represent the observed and the expected values respectively; the colored bands show the 68% (green) and 95% (yellow) probability regions.

istically expect about 10 fb^{-1} of data available to be analyzed by the end of 2011. An estimate of the possible improvements in the Higgs search sensitivity which can be reached by adding more data can be done by simply scaling the 95% C.L. limits with the luminosity increase:

$$L^{exp} = L_0^{exp}\sqrt{\frac{\mathcal{L}_0}{\mathcal{L}}} \tag{9.1}$$

where L_0^{exp} is the expected limit obtained, for a specific Higgs mass value, with the current integrated luminosity $\mathcal{L}_0$, while L^{exp} is the extrapolated value which corresponds to an integrated luminosity $\mathcal{L}$. These projections are shown in figures 9.6 and 9.7 for two reference masses, 115 GeV/c^2 and 160 GeV/c^2. The solid lines are the $1/\sqrt{\mathcal{L}}$ extrapolations of the expected limits on the SM Higgs cross section, obtained starting from different time periods (represented by the circular markers) over the last five years. As can be seen, the sensitivity increased over the time much more than by

Higgs Mass(GeV/c^2)	100	105	110	115	120	125	130
Exp. limit/σ(SM)	1.20	1.24	1.36	1.45	1.69	1.78	1.76
Obs. limit/σ(SM)	0.64	0.87	1.02	1.56	1.95	2.54	2.23
Higgs Mass(GeV/c^2)	135	140	145	150	155	160	165
Exp. limit/σ(SM)	1.73	1.57	1.45	1.25	1.07	0.79	0.76
Obs. limit/σ(SM)	2.41	2.07	1.92	1.93	1.28	0.85	0.68
Higgs Mass(GeV/c^2)	170	175	180	185	190	195	200
Exp. limit/σ(SM)	0.91	1.04	1.25	1.61	1.96	2.31	2.58
Obs. limit/σ(SM)	0.79	0.95	1.49	2.55	2.44	3.49	3.87

Table 9.2: The expected and observed 95% C.L. limits on the Higgs production cross section for the combined CDF and DØ analyses, for 21 Higgs mass hypothesis. Values are expressed in units of the SM cross section and are obtained with the bayesian method described in section 8.6.

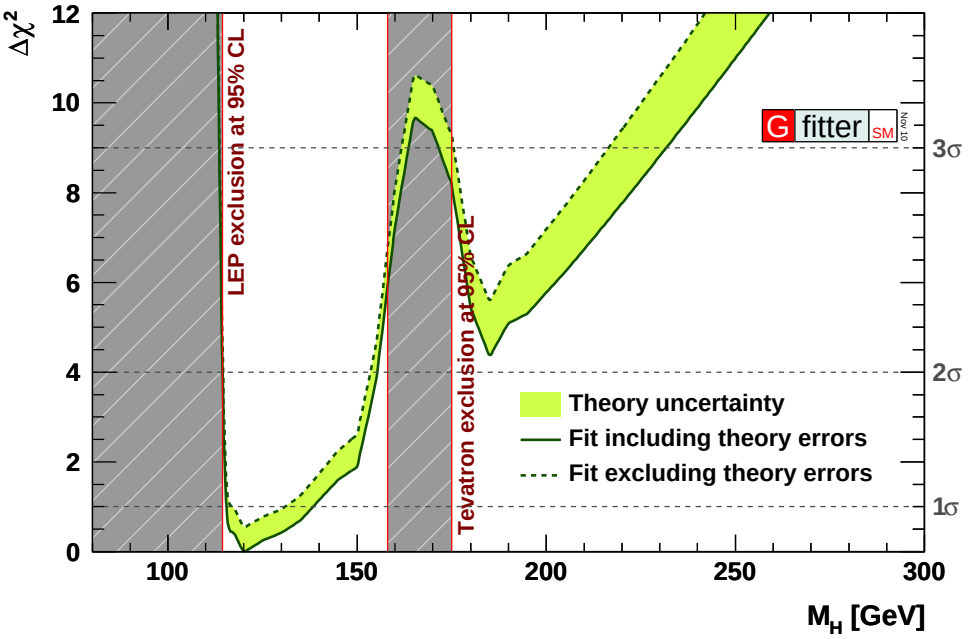

Figure 9.4: $\Delta\chi^2$=χ^2-χ^2_{min} of the global fit to the electroweak precision measurement performed by the GFITTER group [10]. The fit includes the constraints provided by the direct searches of the Higgs boson performed at LEP and Tevatron. The solid and dashed lines give the results when including or ignoring the theoretical errors, respectively.

simply adding data statistics: this has been possible by exploiting new final states and by optimizing the analyses techniques. It is believed that there is still the possibility of further improvements in the near future and this is quantified by the orange bands

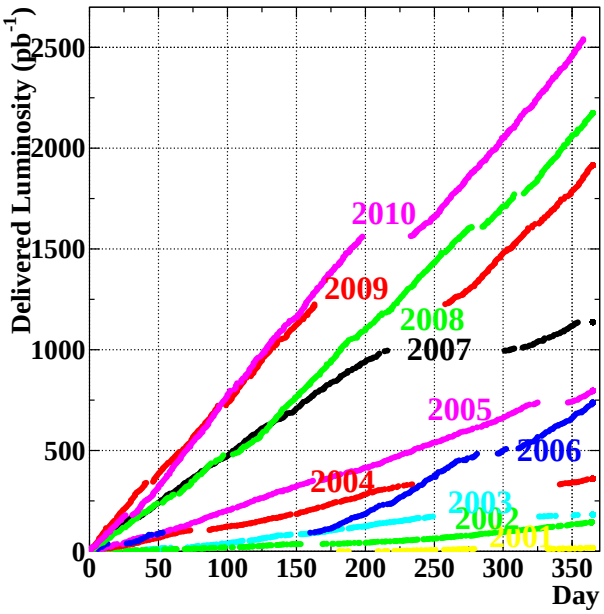

Figure 9.5: Tevatron Run II delivered luminosity as a function of time, for the different years of operations.

in the figures.

Regarding the H$\rightarrow \tau\tau$ analysis, several improvements are in progress, which may soon lead to interesting results and significantly help to increase the sensitivity in the low mass regime. The results presented in this thesis cover only the first 2.3 fb^{-1} of data; by extending the search to 10 fb^{-1} (four times the current luminosity) the expected limit would reduce by a factor of 2 and any improvements in the analysis techniques could further lower this value. Studies are already ongoing to exploit the potential gain coming from loosening the hadronic τ identification requirements and from extending the search to additional di-τ decay channels, like the $\tau_e\tau_\mu$ mode. The expected additional improvements could be of the order of 10%.

9.4.1 Higgs observation probabilities

In addition to the extrapolation of the 95% C.L. exclusion limits, the probabilities of observing the Higgs boson at the Tevatron have been calculated as well. This has been done by scaling the CDF's estimations by $\sqrt{2}$, assuming that the contribution of the DØ analyses being approximately identical to those performed by CDF. Results are reported in figures 9.8 and 9.9, where the probabilities of having at least a 2σ excess and a 3σ evidence above the expected background are shown. Dashed lines refer to the probabilities obtained by assuming additional improvements of 50% compared with the results obtained by only the luminosity scaling, with respect to Winter 2009 results.

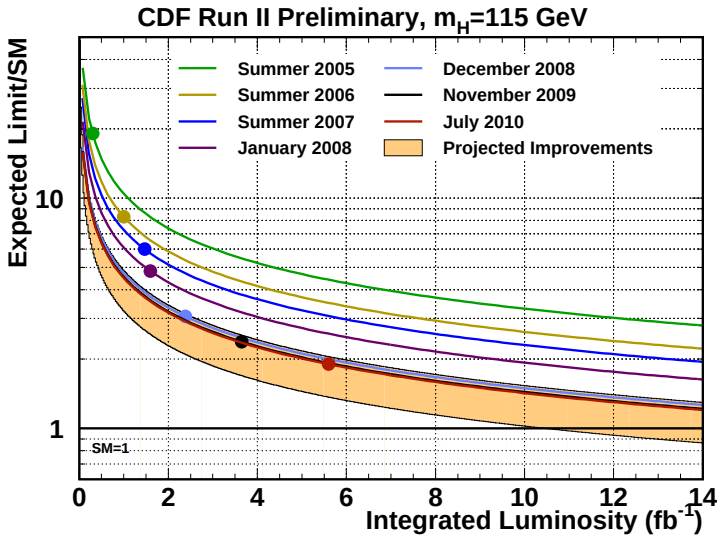

Figure 9.6: Achieved and projected median expected upper limits on the SM Higgs boson cross section, for the Higgs mass hypothesis m_H=115 GeV/c^2. The solid lines are $1/\sqrt{\mathcal{L}}$ projections. The top of the orange band corresponds to the Summer 2007 performance expected limit divided by 1.5, and the bottom of the orange band corresponds to the Summer 2007 performance expected limit divided by 2.25.

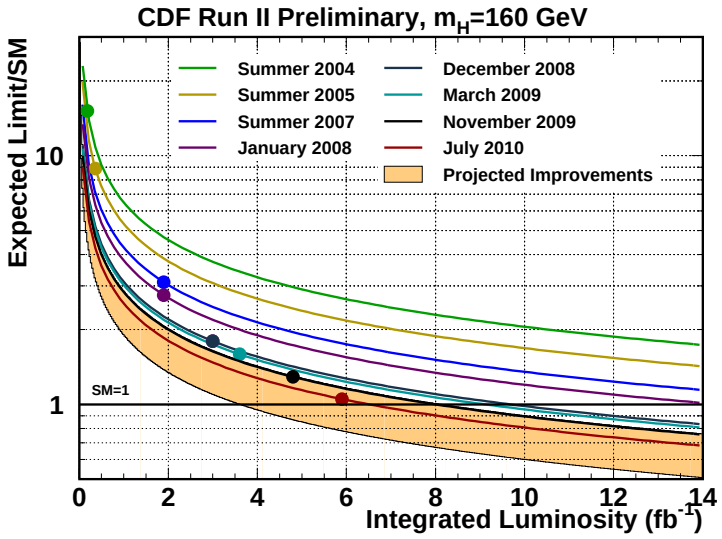

Figure 9.7: Achieved and projected median expected upper limits on the SM Higgs boson cross section, for the Higgs mass hypothesis m_H=160 GeV/c^2. The solid lines are $1/\sqrt{\mathcal{L}}$ projections. The top of the orange band corresponds to the Summer 2007 performance expected limit divided by 1.5, and the bottom of the orange band corresponds to the Summer 2007 performance expected limit divided by 2.25.

In this case, the analysis of 10 fb^{-1} per experiment would give a probability for a 3σ evidence of a Higgs signal greater than 30% in the entire mass range between 114 and 186 GeV/c^2. The probability would further increase up to more than 60% in the case of a 2σ excess.

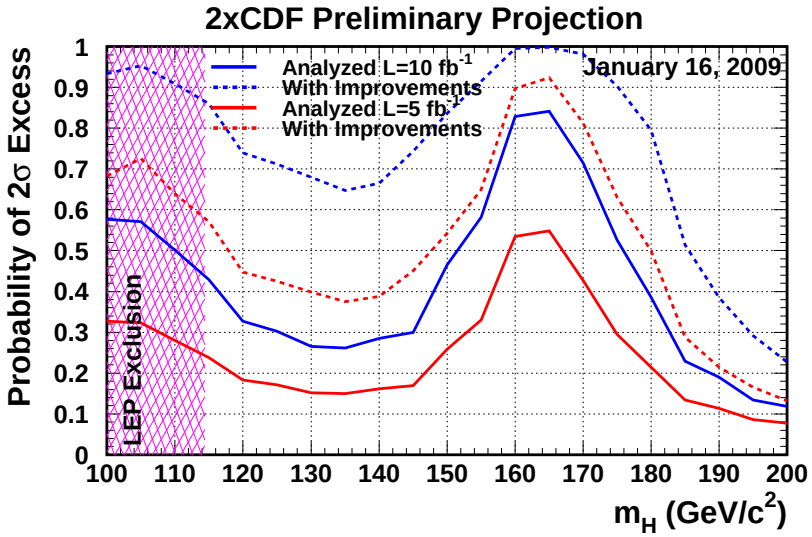

Figure 9.8: Probability of seeing a 2σ excess as a function of the Higgs mass, for 5 fb^{-1} (red) and 10 fb^{-1} (blue) of data analyzed per experiment, assuming that the CDF and D0 performances are the same. Two scenarios are shown, in which the analyses have the same performance as for the Winter 2009 combination (solid lines), and for the case with another factor of 1.5 in the sensitivity (dashed lines).

9.4.2 Tevatron Run III hypothesis and LHC prospects

The possibility of a three years extension of the Tevatron operations, from the end of 2011 through 2014, is currently under discussion [5]. The major goal is to provide us with a very good change of obtaining at least a 3σ evidence of the SM Higgs boson if its mass is below the WW threshold, by almost doubling the existing datasets, up to a total analyzable integrated luminosity of 16 fb^{-1} per experiment. The projected sensitivity for different luminosities is shown, as a function of the mass, in figure 9.10. As can be seen in table 9.3, 15 fb^{-1} per experiment should be enough to probe the entire low mass region with enough sensitivity to claim a 3σ excess. Furthermore, the observation of the H$\rightarrow b\bar{b}$ decay mode is essential for a complete understanding of the electroweak symmetry breaking mechanism. This channel is actively searched for at the Tevatron, while at the LHC it will be inaccessible in the first years of operations.

As of this writing, the LHC will probably not be able to overcome the Tevatron in the low mass Higgs searches in the next few years. LHC plans are to reach 1 fb^{-1} of integrated luminosity at the center of mass energy of 7 TeV by the end of 2011, followed by a long shut down which will last up to the middle of 2013. Projections

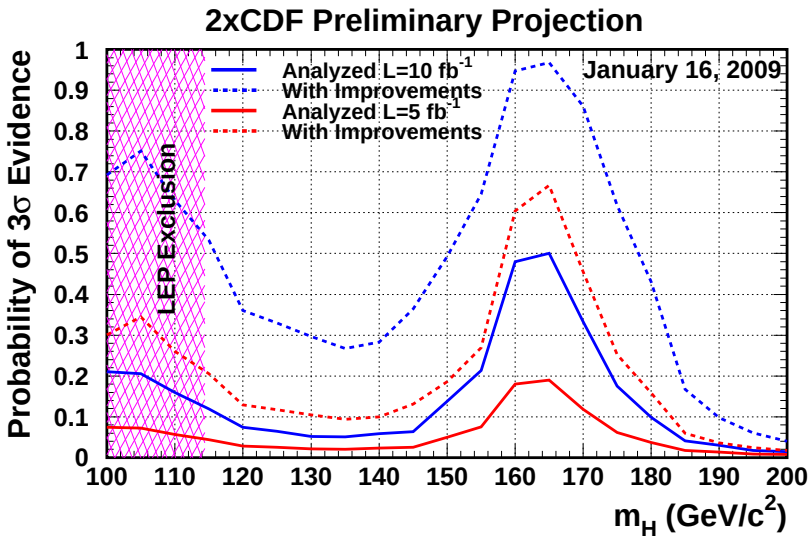

Figure 9.9: Probability of seeing a 3σ observation as a function of the Higgs mass, for 5 fb^{-1} (red) and 10 fb^{-1} (blue) of data analyzed per experiment, assuming that the CDF and D0 performances are the same. Two scenarios are shown, in which the analyses have the same performance as for the Winter 2009 combination (solid lines), and for the case with another factor of 1.5 in the sensitivity (dashed lines).

Lum/exp	115 GeV/c^2	115 GeV/c^2	115 GeV/c^2
5 fb^{-1}	2.2 σ	1.7 σ	1.9σ
10 fb^{-1}	3.1 σ	2.5 σ	2.7σ
15 fb^{-1}	3.8 σ	3.0 σ	3.2σ
20 fb^{-1}	4.4 σ	3.5 σ	3.7σ

Table 9.3: Tevatron Sensitivity to the low mass SM Higgs boson for different integrated luminosities per experiment [5].

from CMS and ATLAS experiments show that the available data from this initial run will not be enough to be as sensitive as CDF and DØ in the low mass regime, because of the lower cross sections in the modes initiated by quark-antiquark annihilation (the associated production, WH/ZH) and the analysis difficulties for the H$\rightarrow b\bar{b}$ channels, overwhelmed by the QCD dijet background [12, 13]. In addition, the cross section for the Higgs production from gluon fusion is about 3.5 times smaller at 7 TeV than at 14 TeV. In this context, the ATLAS and CMS Higgs searches with 1 fb^{-1} of data would yield a 5σ discovery only in a limited mass range, while the 95% C.L. exclusion limits would be at most comparable to those achievable at the Tevatron, as can be seen in figures 9.11, 9.12, 9.13 and 9.14.

At the operating center of mass energy of 14 TeV, it is expected that the LHC will need 1.5 fb^{-1} per experiment to set a 2σ exclusion on the full mass range down

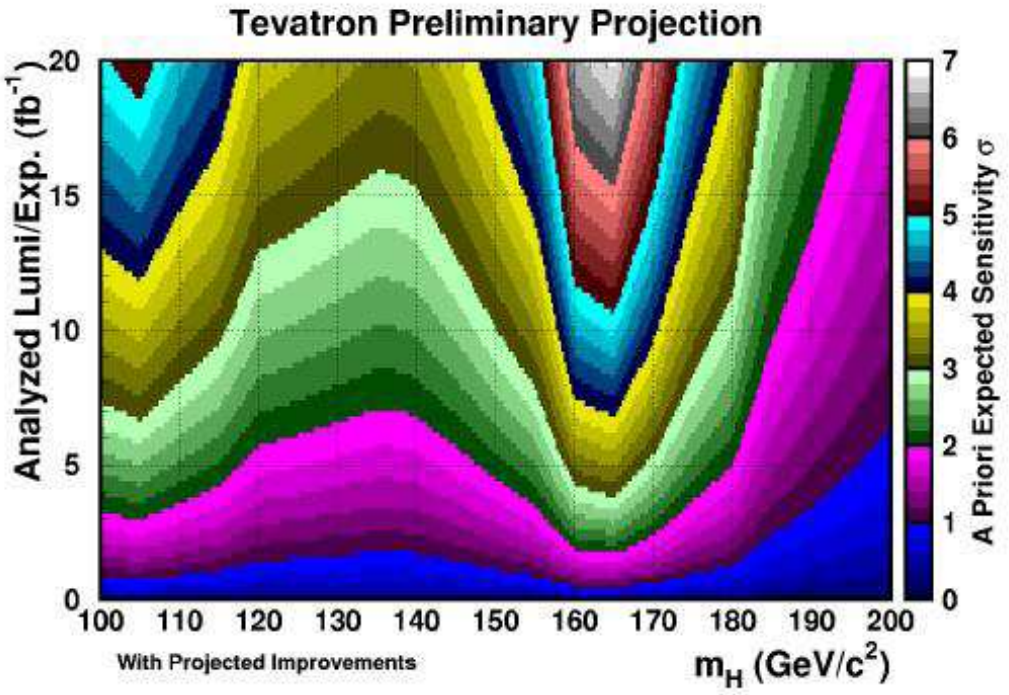

Figure 9.10: Higgs boson expected sensitivity projections in the hypothesis of a Tevatron Run III [5]. The colored bands represent the levels of sensitivity which are expected, as a function of M_H, for a specific amount of integrated luminosity per experiment.

to $M_H = 115$ GeV/c^2, while about 10 fb^{-1} will be required to provide a discovery at $M_H = 115$ GeV/c^2. Furthermore, the low mass Higgs boson sensitivity at the LHC is dominated by the search in the $\gamma\gamma$ decay channel, while the H$\rightarrow$ will be feasible only with at least 30 fb^{-1} of data per experiment [5, 111].

From this perspective, Tevatron experiments could continue to play a crucial rule and be competitive well beyond 2011, providing us with very interesting and stringent results in the exciting quest for the Higgs boson.

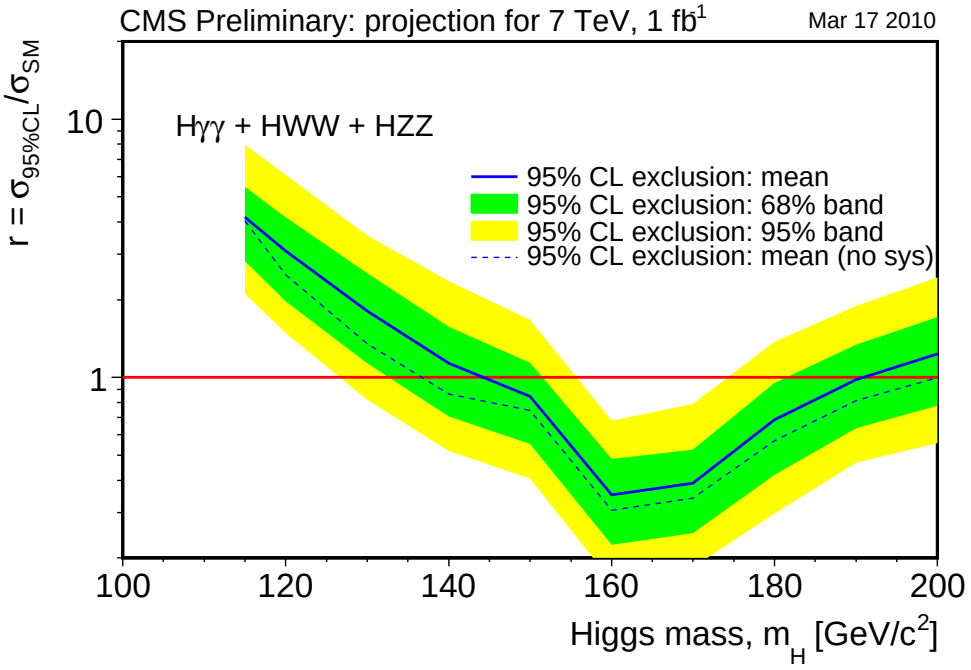

Figure 9.11: CMS expected 95% C.L. upper limits on the Higgs production cross section, as a function of m_H, obtained by combining the projections for an integrated luminosity of 1 fb^{-1} at 7 TeV of three decay modes: WW, ZZ and $\gamma\gamma$ [12].

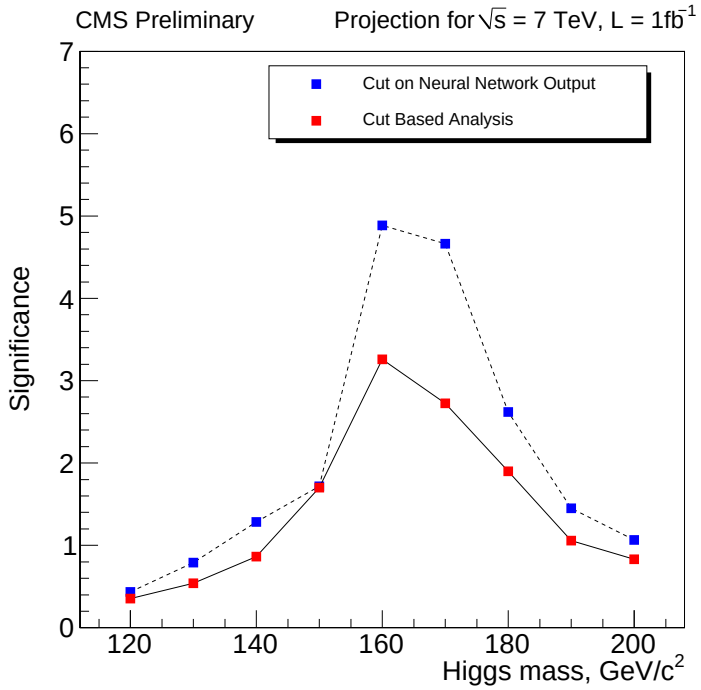

Figure 9.12: CMS sensitivity projections, as a function of m_H, in a search for H→WW→ $ll\nu\nu$, with an integrated luminosity of 1 fb^{-1} at 7 TeV [12]. A 5σ discovery excess is expected for the mass range $160 \leq M_H \leq 170$ GeV/c^2

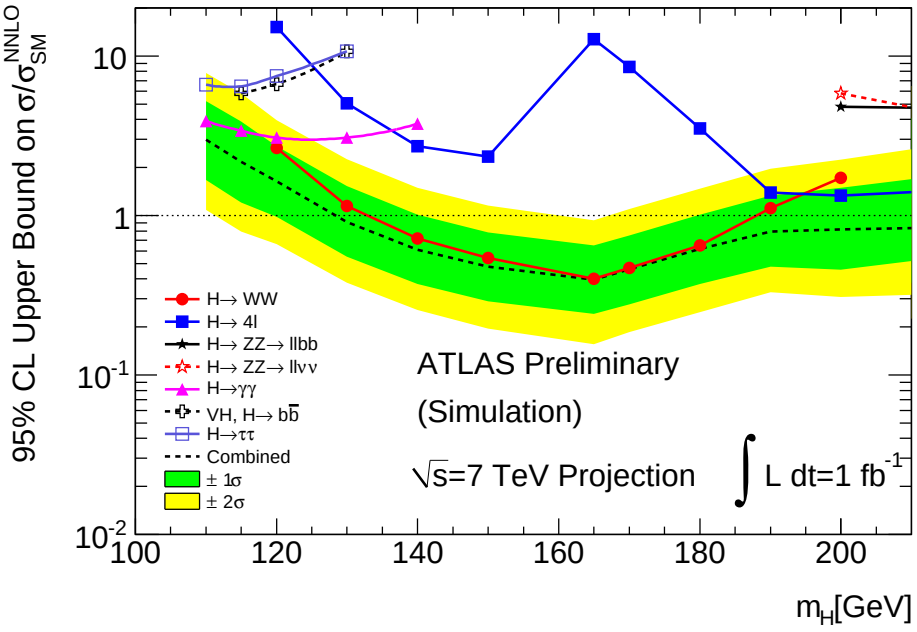

Figure 9.13: ATLAS expected 95% C.L. upper limits on the Higgs production cross section, as a function of m_H, obtained by combining the projections for an integrated luminosity of 1 fb^{-1} at 7 TeV of different channels: WW, ZZ, $\gamma\gamma$, VH($\rightarrow b\bar{b}$) and $\tau\tau$ [13].

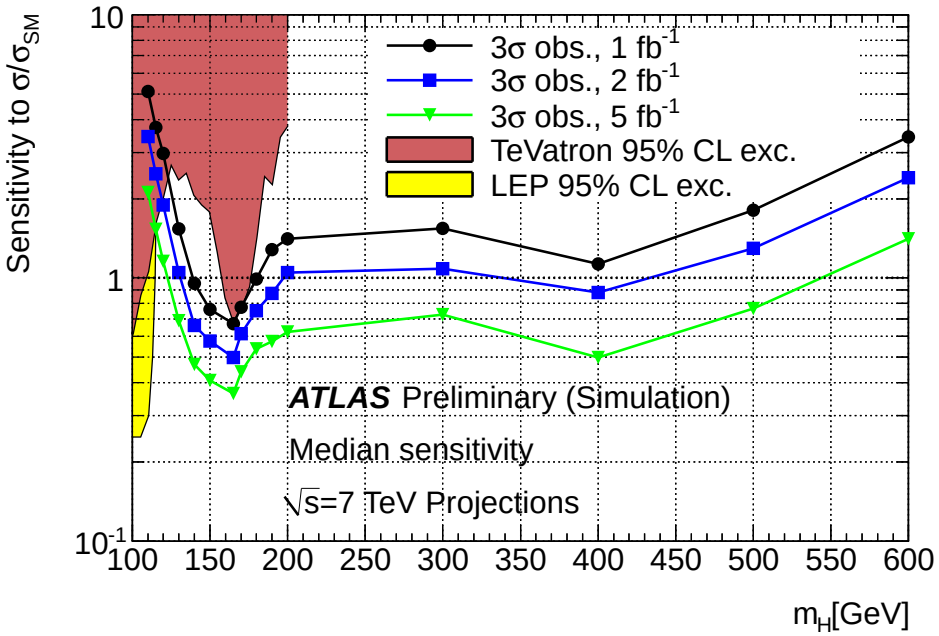

Figure 9.14: ATLAS combined sensitivity projections, as a function of m_H, for different integrated luminosity scenarios at 7 TeV [13]. With 1 fb^{-1}, a 3σ excess is expected for the mass range 139$\leq$M$_H$ $\leq$180 GeV/c^2

Appendix A

Trigger requirements

In this appendix we give a detailed description of all the trigger paths employed at several levels of the analysis.

As outlined in section 6.3, the final state selection is based on data collected by the *lepton plus track* triggers; the procedure for the development of the new tau identification algorithm (see section 6.4.3) required a jet-enriched sample; finally, for the additional studies performed to derive specific background contributions (section 7.2), we used high-p_T muon and high-p_T electron triggers.

A.1 Lepton plus track triggers

Trigger family	Trigger path considered	Tags
Electron plus track	TAU_ELECTRON8_TRACK5_ISO	1-12
CMUP plus track	TAU_CMUP8_TRACK5_ISO	1-11
CMX plus track	TAU_CMX8_TRACK5_ISO	1-8
	TAU_CMX8_TRACK5_ISO_LUMI_200	1
	TAU_CMX8_TRACK5_ISO_LUMI_250	1
	TAU_CMX8_TRACK5_ISO_DPS	1-4

Table A.1: Lepton plus track trigger paths for the run range (141544-246231) used in this analysis.

Several path families refer to this trigger category, according to the different lepton type (CEM electron, CMU+CMP and CMX muon) which is associated to the isolated track. These triggers underwent many modifications during the time and each version is univocally identified by a specific tag number, as reported in table A.1.

Some changes had a very small impact on the trigger behavior or refer to path versions used only for a few test runs, not included in the analysis; other modifications, adopted to accommodate the XFT and L2 cluster upgrades occurred between 2006 and

2007, presented more relevant effects both on the lepton and the track leg, as described in the following sections. Another important change consisted in the definition of the L3 track isolation, which was modified after run 209770, as described in section 6.4.3.

A.1.1 TAU_ELECTRON8_TRACK5_ISO

TAU_ELECTRON8_TRACK5_ISO			
Level	Path name	Object	Requirements
1	L1_CEM8_PT8	CAL cluster	$E_T \geq 8$ GeV
			$E_{had}/E_{em} \leq 0.125$
		XFT track	COT layers ≥ 4
			$p_T \geq 8.34$ GeV/c
2	L2_TAU4_PT5_CEM8_PT8_CES3	CES cluster	$E_T \geq 3$ GeV
		CAL cluster 1	$\lvert\eta\rvert \leq 1.1$
			$E_T \geq 8$ GeV
			$E_{had}/E_{em} \leq 0.125$
		XFT track 1	$p_T \geq 8.34$ GeV/c
		CAL cluster 2	$N_{twr} \leq 4$
			$E_T \geq 4$ GeV
		XFT track 2	$p_T \geq 5.18$ GeV/c
			$\Delta\phi^{SL6}_{track1-track2} \geq 10^\circ$
3	L3_CEM8_TRACK5_ISO	electron	central
			$\chi^2_{strip} \leq 20$
			$L_{shr} \leq 0.4$
			$\lvert\Delta Z_{CES}\rvert \leq 8$ cm
			$E_T \geq 8$ GeV
			$p_T \geq 8$ GeV/c
		track	$\lvert\eta\rvert \leq 1.5$
			$p_T \geq 5$ GeV/c
			isolation
			(table 6.7)
			$\Delta R_{e-\tau} \geq 0.175$
			$\lvert\Delta Z_{e-track}\rvert \leq 15$ cm

Table A.2: Trigger requirements for the TAU_ELECTRON8_TRACK5_ISO path, tag 12.

This trigger requires an electron with $E_T \geq 8$ GeV, which is defined by matching an XFT track to a cluster in the central calorimeter, plus an additional isolated track with $p_T \geq 5$ GeV/c.

The selections applied at each level of the trigger are summarized in table A.2, specifically for the last version used in the analysis, corresponding to tag 12.

The main modifications which occurred from the first version to the last one are the following:

- tag 4: a second L2 XFT track is required;
- tag 8: the second L2 XFT track has to match a calorimetric cluster with $E_T \geq$ 4 GeV and $N_{\rm twr} \leq$5; a new L3 track isolation definition is implemented;
- tag 10: the minimum number of COT layers for the the electron L1 XFT track is increased from 3 to 4;
- tag 12: the upgraded L2 clustering algorithms are included, both for the electron and the tau objects. The maximum number of towers for the tau cluster is reduced to 4.

A.1.2 TAU_CMUP8_TRACK5_ISO

This trigger is aimed at collecting events characterized by central muons, which have a reconstructed XFT track with $p_T \geq 8$ GeV/c, matched to stubs in both the CMU and CMP chambers; in addition, a track with $p_T \geq 5$ GeV/c, satisfying the isolation requirements, has to be found.

The detailed list of trigger selections applied at each level is reported in table A.3 for tag 11, which corresponds to the last version used in the analysis.

The most relevant modifications occurred from the first to the last version of this trigger are the following:

- tag 7: a L2 XFT track with $p_T \geq 8.34$ GeV/c matched to a muon stub is required; previous versions had a L2 *autoaccept*[1];
- tag 8: a second L2 XFT track with $p_T \geq 5.19$ GeV/c is required;
- tag 9: the new L3 track isolation definition is implemented;
- tag 10: Stereo confirmation[2] for the muon and the isolated track;

A.1.3 TAU_CMX8_TRACK5_ISO (DPS/LUMI_200/LUMI_250)

These triggers are aimed at collecting central muons which are detected through a reconstructed XFT track with $p_T \geq 8$ GeV/c, matched to stubs in the CMX Arches or, for specific run periods, CMX Miniskirt and Keystone detectors. Similarly to the other *lepton plus track* triggers, an additional isolated track with $p_T \geq 5$ GeV is also required.

The original version underwent the following relevant transitions:

[1] *Autoaccept* means that no trigger requirements are applied and each event coming from the previous level is passed to the following one

[2] See section 4.2.7 for the description of the L1 XFT upgrade and the SLAM stereo confirmation

TAU_CMUP8_TRACK5_ISO			
Level	Path name	Object	Requirements
1	L1_CMUP6_PT4	CMU stub	$p_T \geq 6$ GeV/c CMP hits = yes
		XFT track	$p_T \geq 4.09$ GeV/c
2	L2_CMUP6_PT8_TRK5_3D	muon	type=CMUP
		XFT track 1	$p_T \geq 8.34$ GeV/c stereo confirmed
		XFT track 2	$p_T \geq 5.19$ GeV/c stereo confirmed
3	L3_CMUP8_TRACK5_ISO	muon	type=CMUP $p_T \geq 8$ GeV/c $\Delta x_{CMP} \leq 20$ cm $\Delta x_{CMU} \leq 15$ cm
		track	$\|\eta\| \leq 1.5$ $p_T \geq 5$ GeV/c isolation (table 6.7) $\Delta R_{\mu-\tau} \geq 0.175$ $\|\Delta Z_{\mu-track}\| \leq 15$ cm

Table A.3: Trigger requirements for the TAU_CMUP8_TRACK5_ISO path, tag 11.

- tag 7: added the requirement of two L2 XFT tracks, the first one ($p_T \geq 8.34$ GeV/c) matched to a CMX stub; previous versions had a L2 *autoaccept*;
- tag 8: implemented the new L3 track isolation;

As already mentioned in section 6.3, several dynamically prescaled or luminosity enabled versions have been implemented and used in the last periods of data taking, in order to keep the trigger rate under control at high instantaneous luminosity.

The trigger which is enabled at luminosities below 250 $\times 10^{30}$ cm^{-2}s^{-1} requires that the two tracks are stereo confirmed, as well as the DPS version, tag 4, whose selections are summarized in table A.4.

TAU_CMX8_TRACK5_ISO_DPS			
Level	Path name	Object	Requirements
1	L1_CMX6_PT8_CSX	CMX stub	$p_T \geq 6$ GeV/c CSX hits = yes
		XFT track	COT layers ≥ 4 $p_T \geq 8.34$ GeV/c
2	L2_CMX6_PT8_&_TRK5_3D_DPS	muon	type=CMX
		XFT track 1	$p_T \geq 8.34$ GeV/c stereo confirmed
		XFT track 2	$p_T \geq 5.19$ GeV/c stereo confirmed
			MAX_PRESCALE = 20 MIN_PRESCALE = 1
3	L3_CMX8_TRACK5_ISO	muon	type=CMX $p_T \geq 8$ GeV/c $\Delta x_{CMX} \leq 30$ cm
		track	$\|\eta\| \leq 1.5$ $p_T \geq 5$ GeV/c isolation (table 6.7) $\Delta R_{\mu-\tau} \geq 0.175$ $\|\Delta Z_{\mu-track}\| \leq 15$ cm

Table A.4: Trigger requirements for the TAU_CMX8_TRACK5_ISO_DPS (tag 4) path.

A.2 Jet triggers

Jet-based triggers select events with at least one reconstructed calorimeter jet above a specific transverse energy threshold. Given the high cross section for QCD multi-jet processes in $p\bar{p}$ collisions, most of these triggers have to be prescaled to keep rates at a reasonable level.

A.2.1 JET_20

The requirements of this trigger path are summarized for each level in table A.5.

JET_20			
Level	Path name	Object	Requirements
1	L1_JET5_PS50	cal tower	$E_T \geq 5$ GeV PRESCALE = 50
2	L2_JET15_PS25	cal cluster	$E_T \geq 15$ GeV $\|\eta\| \leq 3.6$ PRESCALE = 25
3	L3_JET_20	jet(fixed cone ΔR = 0.7)	$E_T > 20$ GeV

Table A.5: Trigger requirements for the JET_20 path.

A.3 High p_T lepton triggers

A.3.1 ELECTRON_CENTRAL_18

In this trigger, energetic electron candidates are selected by looking for an XFT track with $p_T \geq$9 GeV/c matched to an $E_T \geq$18 GeV central calorimeter cluster, satisfying some quality requirements in the L_{shr}, $|\Delta Z_{CES}|$ and E_{had}/E_{em} variables.

The complete list of selections applied at each level is reported in table A.6, for tag version 13.

ELECTRON_CENTRAL_18			
Level	Path name	Object	Requirements
1	L1_CEM8_PT8	CAL cluster	$E_T \geq 8$ GeV $E_{had}/E_{em} \leq 0.125$
		XFT track	COT layers ≥ 4 $p_T \geq 8.34$ GeV/c
2	L2_CEM18_PT8	CAL cluster	$|\eta| \leq$1.317 $E_T \geq 18$ GeV $E_{had}/E_{em} \leq 0.125$
			COT layers ≥ 4 $p_T \geq 8$ GeV/c
3	L3_ELECTRON_CENTRAL_18	electron	central $L_{shr} \leq 0.4$ $\|\Delta Z_{CES}\| \leq 8$ cm $E_T \geq 18$ GeV $E_{had}/E_{em} \leq 0.125$ $p_T \geq 9$ GeV/c

Table A.6: Trigger requirements for the ELECTRON_CENTRAL_18 path, tag 13.

This path experienced very few changes in time, the most important ones are the following:

- tag 6: the minimum number of COT layers for the L1 XFT track is reduced from 4 to 3;
- tag 8: cuts to L_{shr} and $|\Delta Z_{CES}|$ are added at L3;
- tag 9: the minimum number of COT layers for the XFT track is increased from 3 to 4 both at L1 and L2;
- tag 13: a new clustering algorithm, which exploits the L2 calorimeter upgrade, is implemented.

A.3.2 MUON_CMUP18

This trigger selects high p_T muon candidates with a reconstructed XFT track with p_T $\geq$18GeV/c, matched to stubs in the CMU and CMP chambers. The detailed requirements are shown in table A.7 for the last version used in this work. The main change occurred to this trigger path is between tag 7 and 8 and is related to the XFT upgrades.

- tag 3: added the requirement of a L2 XFT track with $p_T \geq 8.34$ GeV/c; previous versions had a L2 *autoaccept*;
- tag 7: the L2 track has to match stubs in the CMU and CMP chambers;
- tag 8: the p_T cut for the L2 track is increased to 14.77 GeV/c; stereo confirmation is also required.

A.3.3 MUON_CMX18

This trigger collects high p_T muon candidates with a reconstructed track with p_T $\geq$18 GeV/c matched to stubs in the CMX Arches or, for specific run periods, CMX Miniskirt and Keystone detectors.

Similarly to the MUON_CMUP18 path, the most relevant changes were done to exploit the upgrade of the XFT system. The L2 p_T increase up to 14.77 GeV/c and the stereo confirmation requirements were implemented at the same time of the introduction of a luminosity enabling at 250 $\times 10^{30}$ $cm^{-2}s^{-1}$, quickly replaced by a dynamic prescale as reported in table A.8.

The other changes which affected the original not prescaled version of this trigger are the following:

- tag 3: added the requirement of hits in the CSX scintillator;
- tag 6: a four-layers L2 XFT tracks, with $p_T \geq 10.1$ GeV/c, matching a CMX stub, has to be found; previous versions had a L2 *autoaccept*;

MUON_CMUP18			
Level	Path name	Object	Requirements
1	L1_CMUP6_PT4	CMU stub	$p_T \geq 6$ GeV/c CMP hits = yes
		XFT track	$p_T \geq 4.09$ GeV/c
2	L2_CMUP6_PT15 _3D	muon	type=CMUP
		XFT track 1	$p_T \geq 14.77$ GeV/c COT layers ≥ 4 stereo confirmed
3	L3_CMUP8_TRACK5_ISO	muon	type=CMUP $p_T \geq 18$ GeV/c $\Delta x_{CMP} \leq$20 cm $\Delta x_{CMU} \leq$10 cm

Table A.7: Trigger requirements for the MUON_CMUP18 path, tag 10.

MUON_CMX18_DPS			
Level	Path name	Object	Requirements
1	L1_CMX6_PT8_CSX	CMX stub	$p_T \geq 6$ GeV/c CSX hits = yes
		XFT track	COT layers ≥ 4 $p_T \geq 8.34$ GeV/c
2	L2_CMX6_PT15_3D_DPS	muon	type=CMX
		XFT track 1	$p_T \geq 14.77$ GeV/c COT layers ≥ 4 stereo confirmed MAX_PRESCALE = 10 MIN_PRESCALE = 1
3	L3_MUON_CMX18	muon	type=CMX $p_T \geq 18$ GeV/c $\Delta x_{CMX} \leq$10 cm

Table A.8: Trigger requirements for the MUON_CMX18_DPS path, tag 3.

Appendix B

Multivariate techniques: the Boosted Decision Tree method

The Boosted Decision Tree is an advanced method of event classification.

A *decision tree* is a sequence of rooted binary splits, performed using a set of discriminating variables. A sketch of the underlying logic is presented in Figure B.1. Given a training sample made of known signal and background events, repeated decisions are performed: at each *node* the variable and the split value which give the best separation are selected and two classes (or child nodes) are created. When a predefined criterion is met, splitting stops and the terminal nodes are called *leafs* and tagged as signal (S) or background (B) according to their purity P, defined as the weighted signal fraction of the training sample:

$$P = \frac{\Sigma_s W_s}{\Sigma_s W_s + \Sigma_b W_b} \tag{B.1}$$

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

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

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

Overtraining affects a multivariate method when too many parameters need to be adjusted to too few data points. BDT is particular sensitive to this effect, when

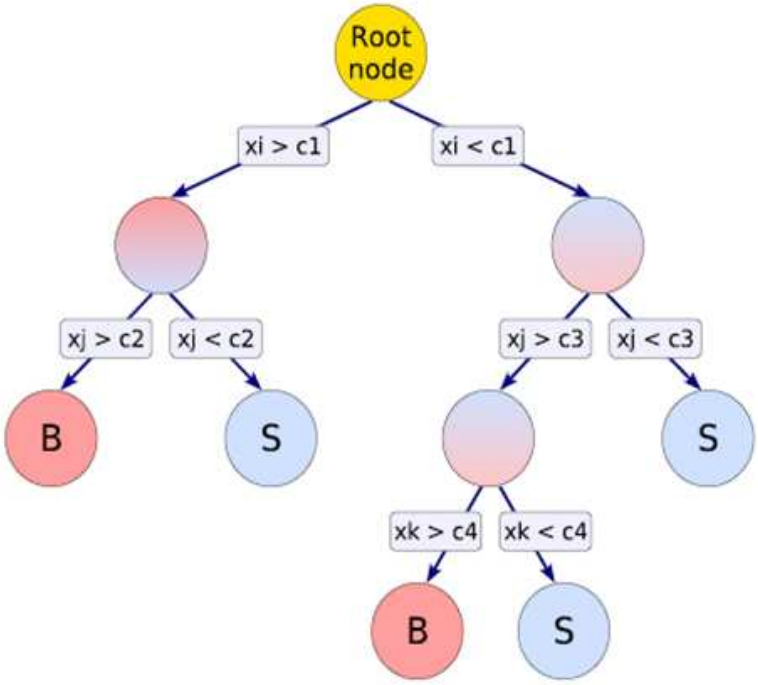

Figure B.1: An example picture of a decision tree. At each node, the most discriminating variable x_i is chosen and the cut which gives the best separation between signal and background is applied. When the predefined criterion is reached, splitting stops and each leaf is labeled as S or B.

splitting is extended to nodes with very poor statistics. In this case a decision tree must be *pruned*. Pruning is the process of cutting back a tree from the bottom, by removing final statistically insignificant nodes. This procedure is performed after the growth of the tree to its maximum size, because it has been demonstrated that some splits may appear useless at one node but may lead to good splits in subsequent steps.

B.1 TMVA framework

In this study we use the BDT method implemented, along with several other classifiers, in the Toolkit for MultiVariate Analysis (TMVA), available in the ROOT package.

In the TMVA framework, an analysis procedure is typically organized in two independent stages, the *training* and the *application*.

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

parameter	value	description
`nTrees`	`400`	Number of trees in the forest
`BoostType`	`Adaboost`	Boosting type
`SeparationType`	`GiniIndex`	Separation criterion for the node splitting
`nEventsMin`	`20`	Minimum number of events in a node
`nCuts`	`20`	Number of steps in the scan for cut optimization at a node
`PruneMethod`	`CostComplexity`	Pruning method
`PruneStrength`		Amount of pruning

Table B.1: Configuration options for the BDT classifier adopted in this analysis; the *PruneStrength* parameter has been tuned for each specific BDT to keep overtraining under control. For more details, refer to [6].

In the *application* stage the user defines a *Reader* object, where the results of the training are retrieved and applied to a dataset of unknown composition: each event is thus analyzed and given a score.

In the *weight* files a large variety of control and performance plots are stored, as well and they can be displayed via a graphical user interface (GUI), by running a set of ROOT macros.

More details about TMVA structure and all configuration parameters can be found in the User Guide [6].

B.2 Variable decorrelation

It has been proved that BDTs, like other multivariate event classifiers, underperform in presence of strong correlations among discriminating variables.

This could be the case for the tau identification algorithm, because several tau variables described in section 6.4.3 partially share the same physical information coming from the detector: for example, the sums of track momenta and π^0 energy in the isolation region are related to the number of identified objects; π^0's are reconstructed using electromagnetic cluster energy; visible mass and visible transverse energy collect information from tracks and π^0's, etc.

Additional information can be recovered if a decorrelation procedure is applied to data before presenting it to the classifier: this is done by TMVA by diagonalizing the symmetric covariance matrices (separately for signal and background samples), as described in [6].

In the analysis discussed in this thesis we trained the BDTs for tau identification, both with and without decorrelation and we actually found a slight performance improvement when applying this data pre-processing procedure. A Boosted Decision Tree trained with decorrelated variables is typically called BDTD in TMVA framework.

Appendix C

Final $\tau\tau$ mass reconstruction using the missing transverse energy

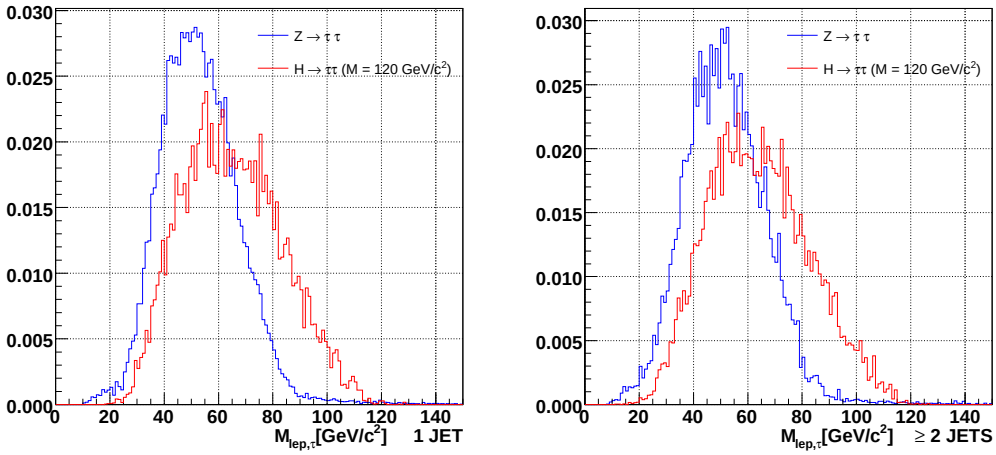

Figure C.1: Invariant mass of the visible τ decay products in the $\tau_h\tau_e$ and $\tau_h\tau_\mu$ final state, for events containing 1(left) and $\geq$2 calorimeter jets (right). The distributions for the Z$\rightarrow \tau\tau$ and the H$\rightarrow \tau\tau$ processes are shown, in the hypothesis of M_H=120 GeV/c^2.

As can be seen in figure C.1, where the Z$\rightarrow \tau\tau$ and the H$\rightarrow \tau\tau$ (in the 120 GeV/c^2 mass hypothesis) processes are compared in the $\tau_h\tau_e$ and $\tau_h\tau_\mu$ final states, the invariant mass calculated with the visible τ decay products is shifted to low values and suffers from a very poor resolution, since a sizeable fraction of the τ energy is carried by the undetected neutrinos.

The separation between the two distributions can be sligthly increased if the information provided by the missing transverse energy (see section 6.4.5), which represents a measure of the transverse component of the sum of the neutrino momenta, is exploited. A first attempt can be performed by defining the invariant mass of the $\tau_l+\tau_h+\not{E}_T$ system. Although some improvements are achieved, the results, shown in figure C.2, are

still unsatisfying.

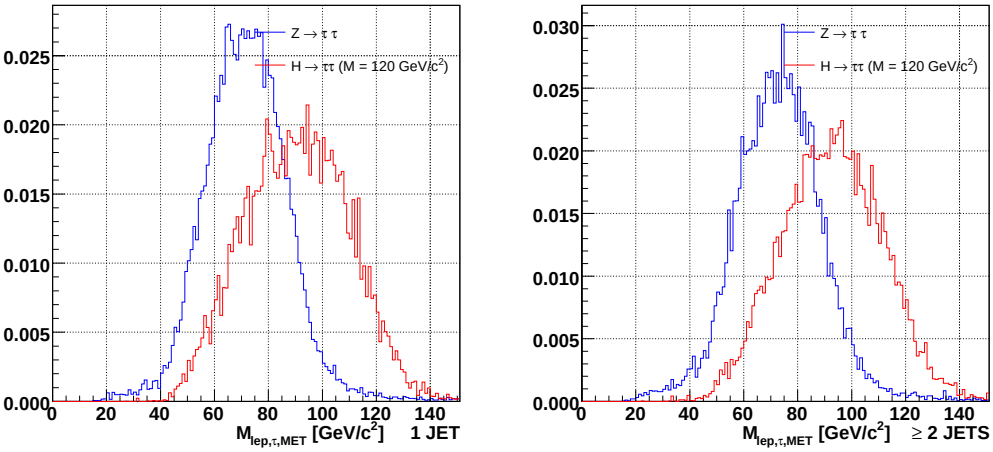

Figure C.2: Invariant mass of the lepton+τ_h+$\not\!\!E_T$ system in the $\tau_h\tau_e$ and $\tau_h\tau_\mu$ final state, for events containing 1(left) and $\geq$2 calorimeter jets (right). The distributions for the Z$\rightarrow \tau\tau$ and the H$\rightarrow \tau\tau$ processes are shown, in the hypothesis of M_H=120 GeV/c^2.

When the particle which decays into a couple of τ leptons is boosted in the transverse plane and therefore the τ decay products are not emitted back to back [1], it is possible to fully reconstruct the invariant mass, by means of the so called *collinear approximation*: as shown in figure C.4, the neutrino directions are assumed to be the same of the other visible decay products and the missing transverse components are assigned to the reconstructed τ's by solving the following set of equations:

$$\not\!\!E_{Tx} = \alpha p_x^{\tau_1} + \beta p_x^{\tau_2} \tag{C.1}$$
$$\not\!\!E_{Ty} = \alpha p_y^{\tau_1} + \beta p_y^{\tau_2} \tag{C.2}$$

where α and β are defined in such a way that the full τ four-momenta are given by

$$P_{full}^{\tau_1} = (1+\alpha)P_{vis}^{\tau_1} \tag{C.3}$$
$$P_{full}^{\tau_2} = (1+\beta)P_{vis}^{\tau_2} \tag{C.4}$$

This method is very promising and allows us to reconstruct the right original invariant mass peak in events with at least one reconstructed jet, as can be seen in figure C.3, thus considerably improving the discrimination between events originated from the Higgs boson and those coming from the Z decay.

[1] This condition is easily met when considering events with jet activity, as it is in the Higgs event topology considered in this thesis.

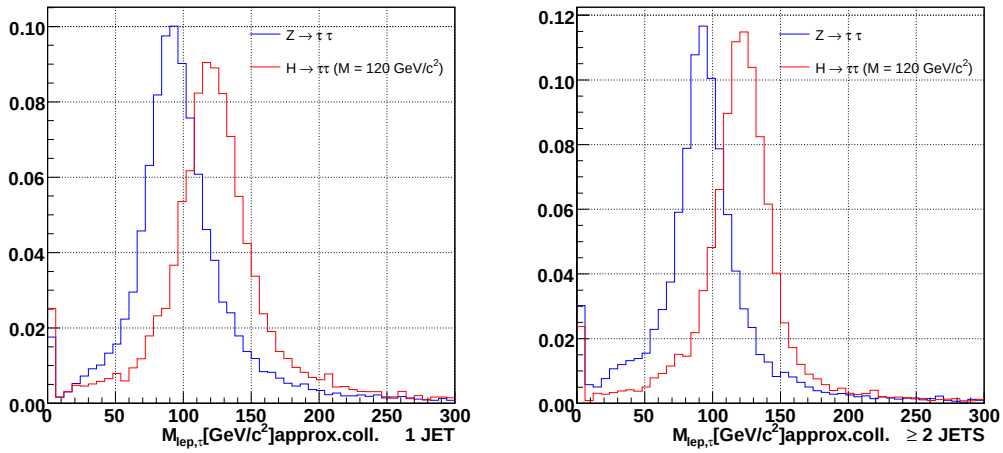

Figure C.3: Invariant mass of the $\tau_l\tau_h$ system, in the collinear approximation for neutrinos' energy, for events containing 1(left) and $\geq$2 calorimeter jets (right). The distributions for the Z$\rightarrow \tau\tau$ and the H$\rightarrow \tau\tau$ processes are shown, in the hypothesis of M_H=120 GeV/c^2.

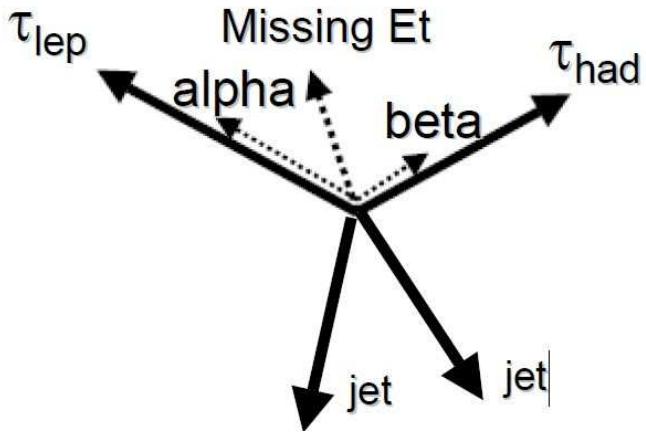

Figure C.4: Schematic view of the assignment of missing transverse energy to τ's, in the collinear approximation for neutrino's direction.

Appendix D

Final discriminant templates

D.1 $M_H = 100$ GeV/c^2

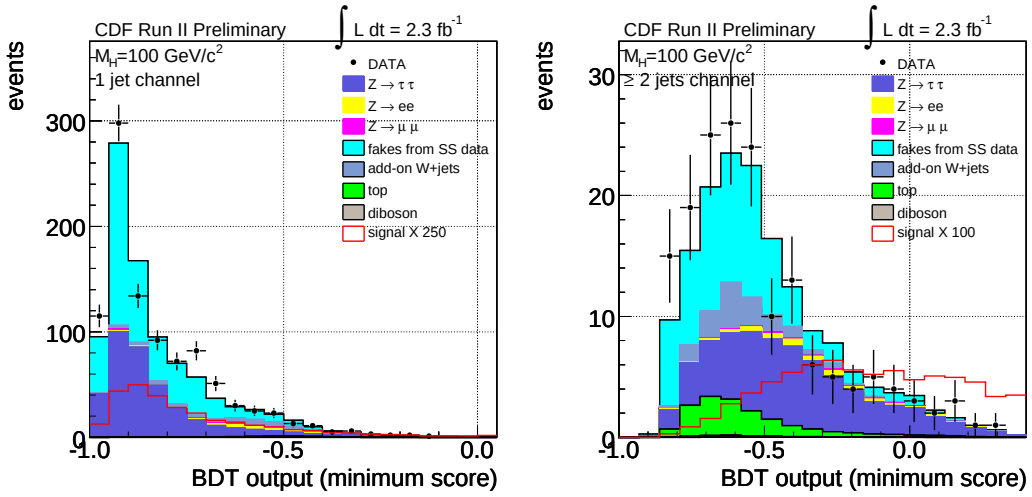

Figure D.1: Final discriminant for the Higgs mass hypothesis of 100 GeV/c^2. Left: 1 jet channel; right: $\geq$2 jets channel.

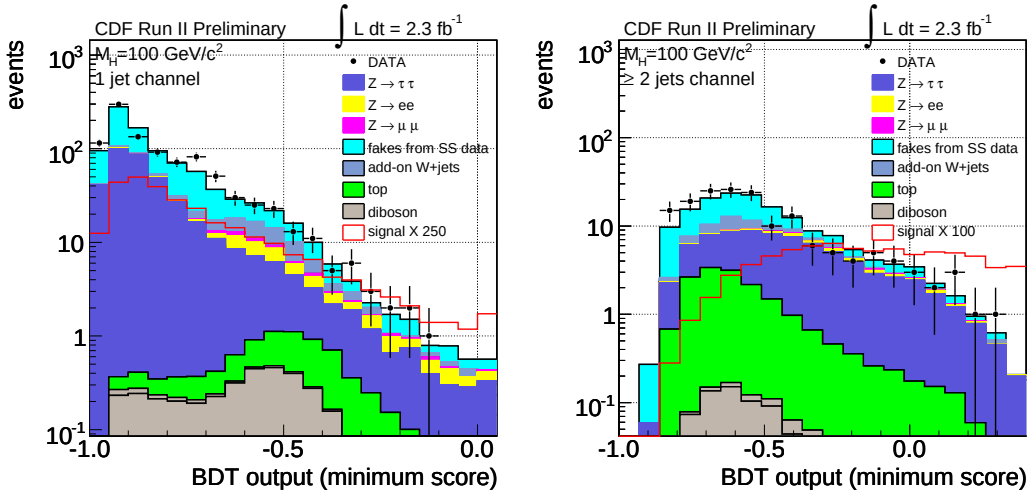

Figure D.2: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 100 GeV/c^2. Left: 1 jet channel; right: $\geq$2 jets channel.

D.2 $M_H = 105$ GeV/c^2

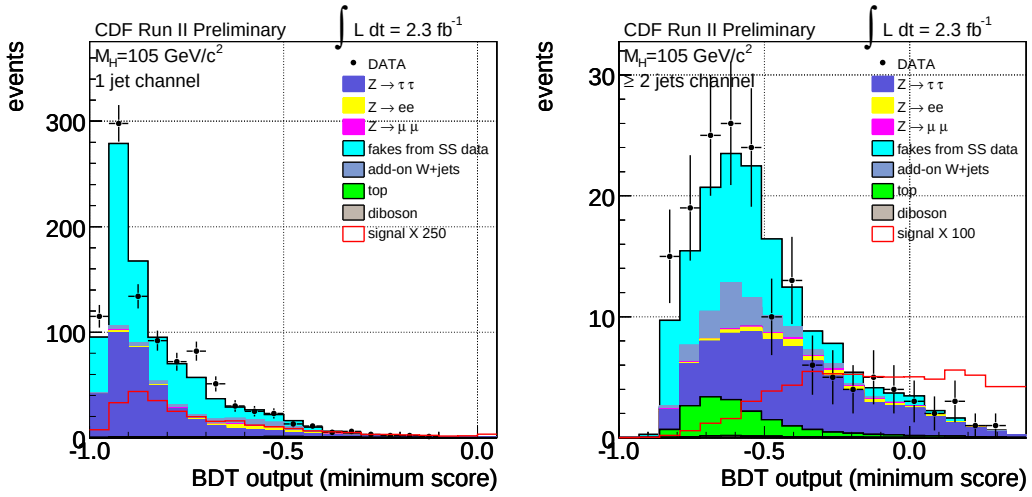

Figure D.3: Final discriminant for the Higgs mass hypothesis of 105 GeV/c^2. Left: 1 jet channel; right: $\geq$2 jets channel.

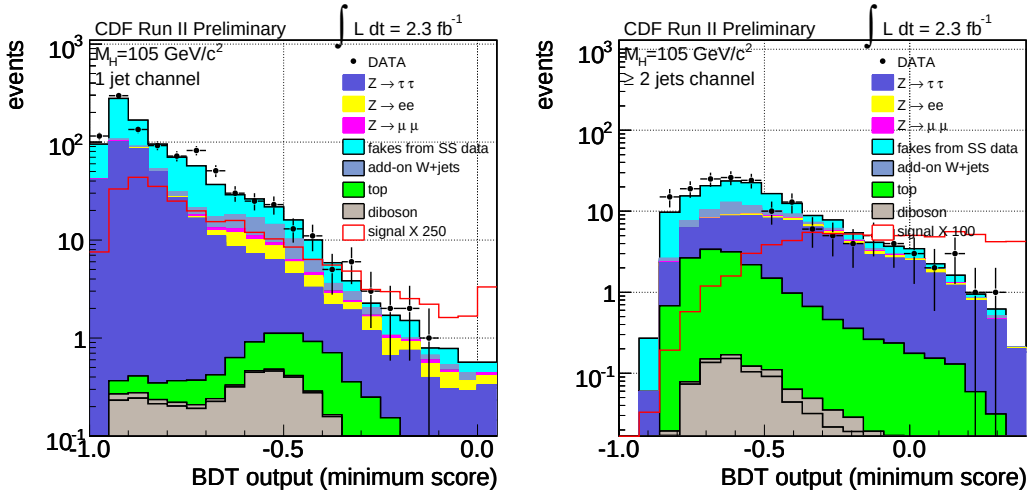

Figure D.4: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 105 GeV/c^2. Left: 1 jet channel; right: $\geq$2 jets channel.

D.3 $M_H = 110$ GeV/c^2

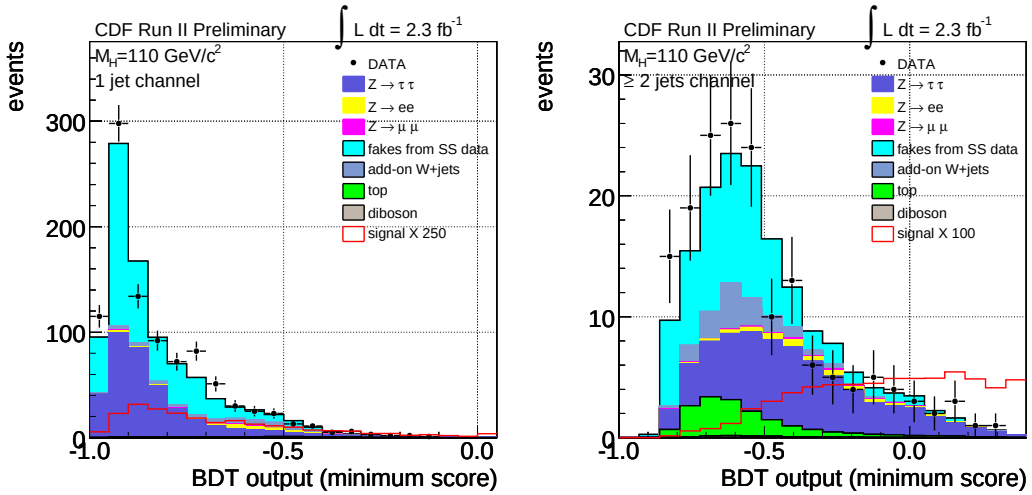

Figure D.5: Final discriminant for the Higgs mass hypothesis of 110 GeV/c^2. Left: 1 jet channel; right: $\geq$2 jets channel.

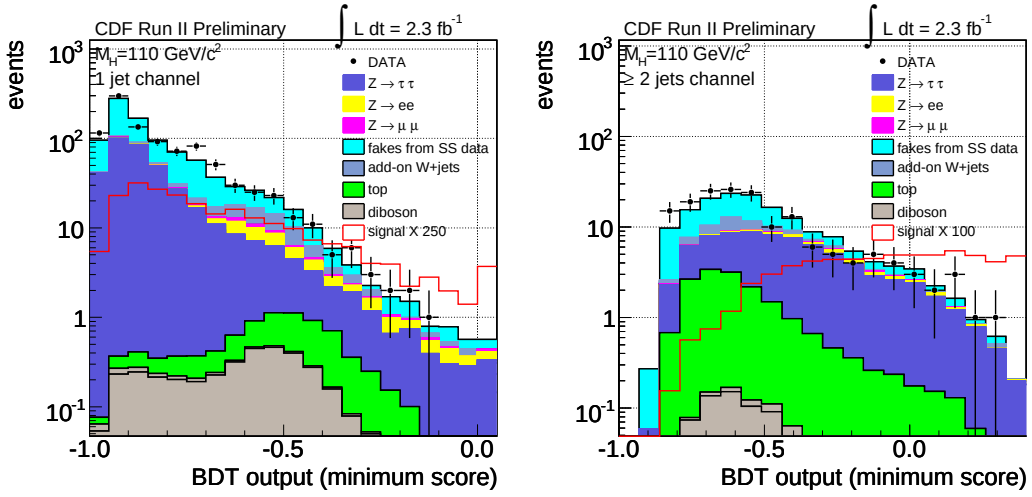

Figure D.6: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 110 GeV/c^2. Left: 1 jet channel; right: $\geq$2 jets channel.

D.4 $M_H = 115$ GeV/c^2

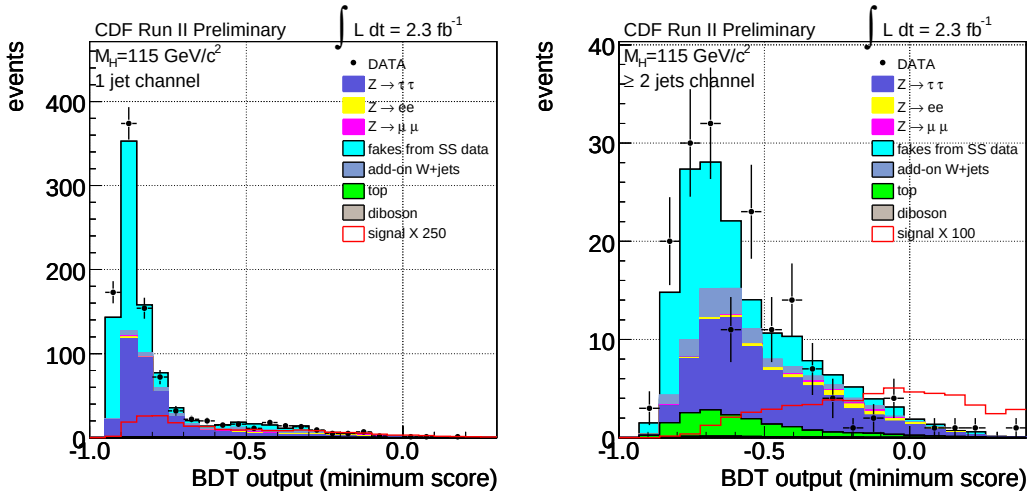

Figure D.7: Final discriminant for the Higgs mass hypothesis of 115 GeV/c^2. Left: 1 jet channel; right: $\geq$2 jets channel.

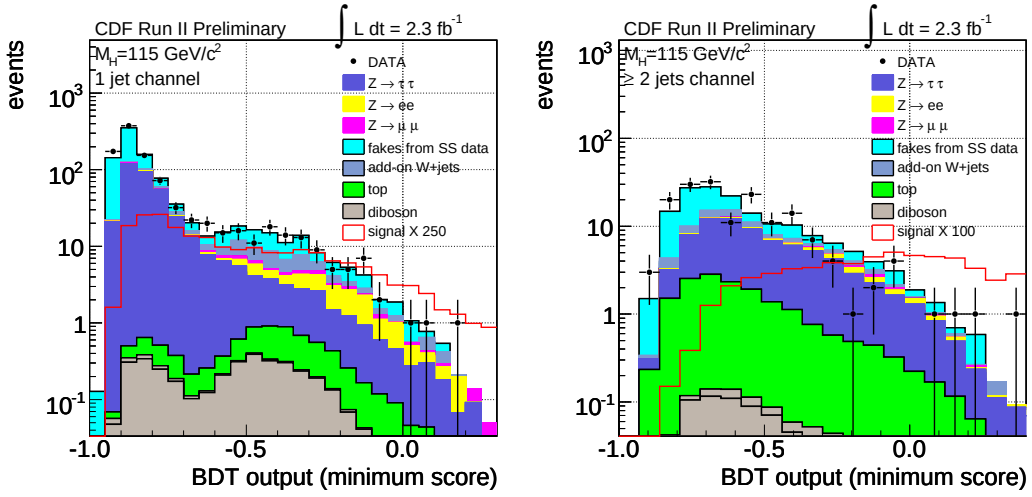

Figure D.8: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 115 GeV/c^2. Left: 1 jet channel; right: $\geq$2 jets channel.

D.5 $M_H = 125$ GeV/c^2

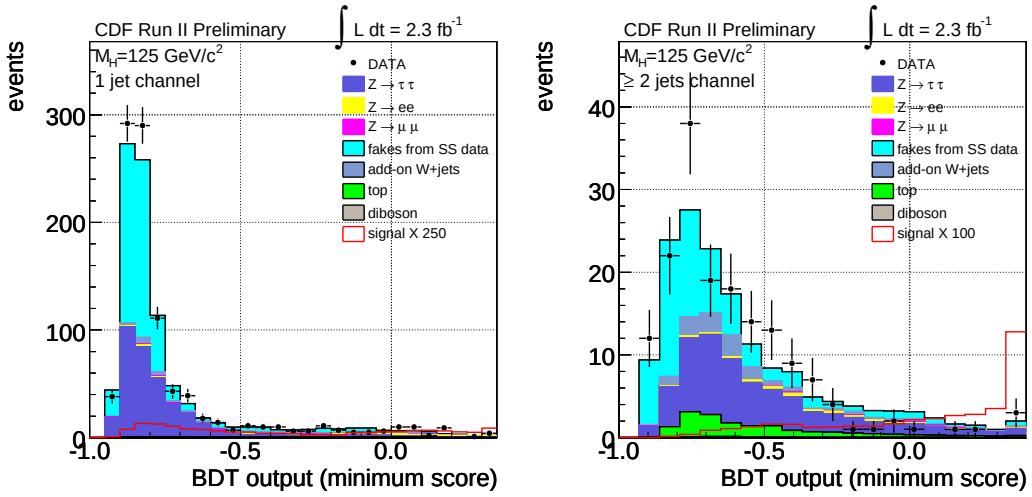

Figure D.9: Final discriminant for the Higgs mass hypothesis of 125 GeV/c^2. Left: 1 jet channel; right: $\geq$2 jets channel.

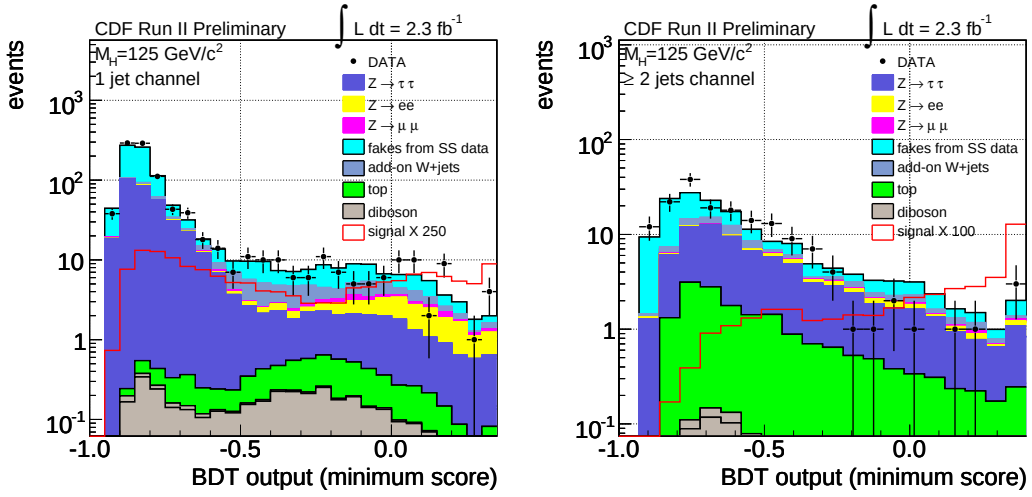

Figure D.10: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 125 GeV/c^2. Left: 1 jet channel; right: $\geq$2 jets channel.

D.6 $M_H = 130$ GeV/c^2

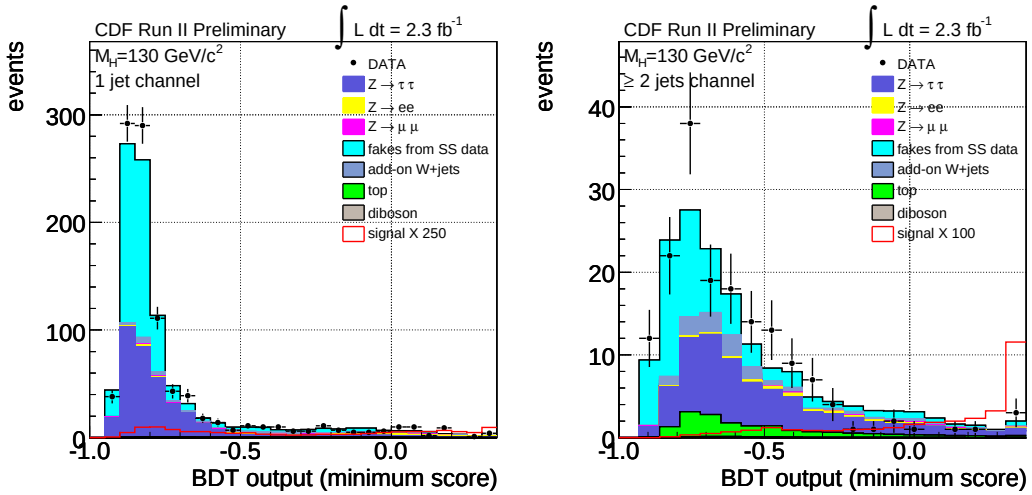

Figure D.11: Final discriminant for the Higgs mass hypothesis of 130 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

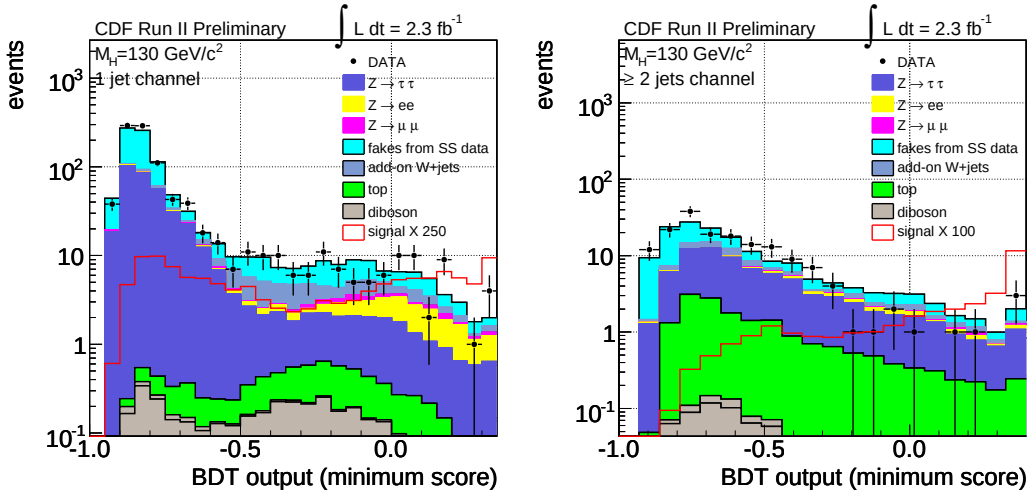

Figure D.12: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 130 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

D.7 $M_H = 135$ GeV/c^2

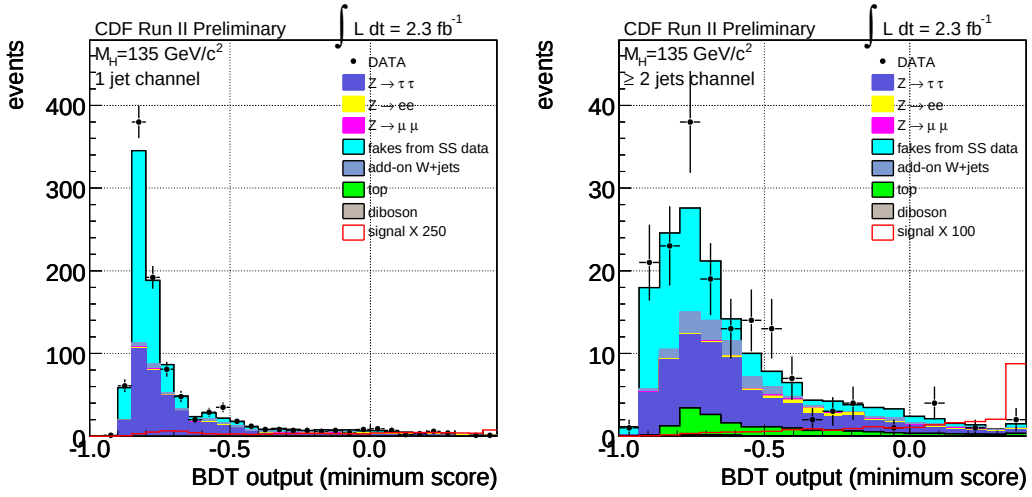

Figure D.13: Final discriminant for the Higgs mass hypothesis of 135 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

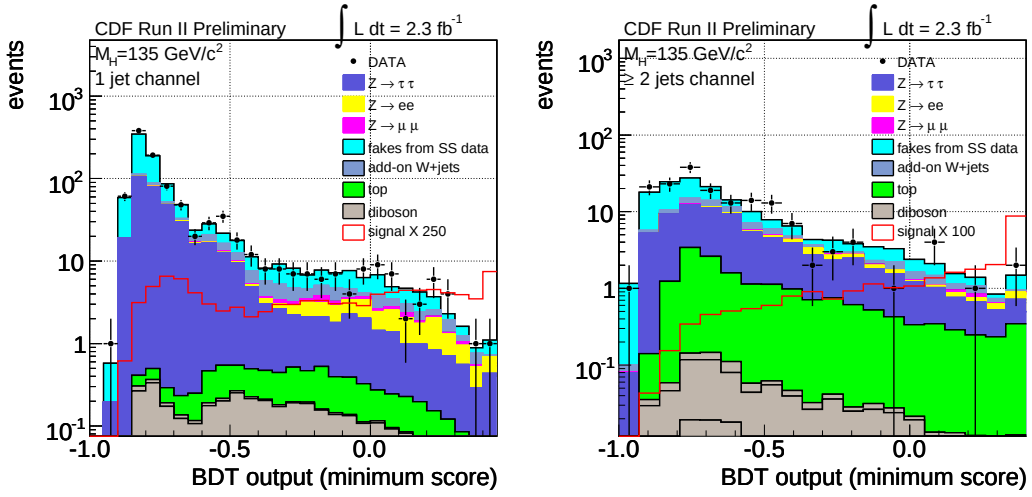

Figure D.14: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 135 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

D.8 $M_H = 140$ GeV/c^2

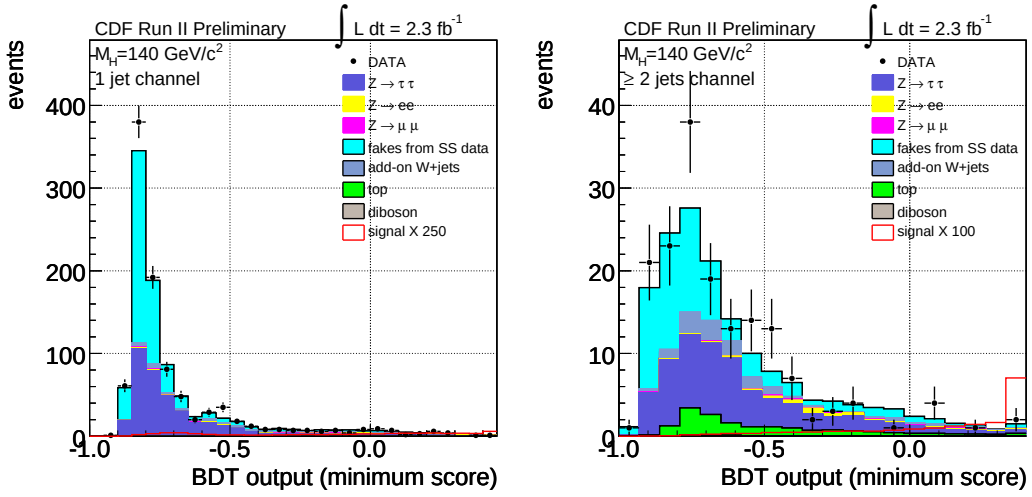

Figure D.15: Final discriminant for the Higgs mass hypothesis of 140 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

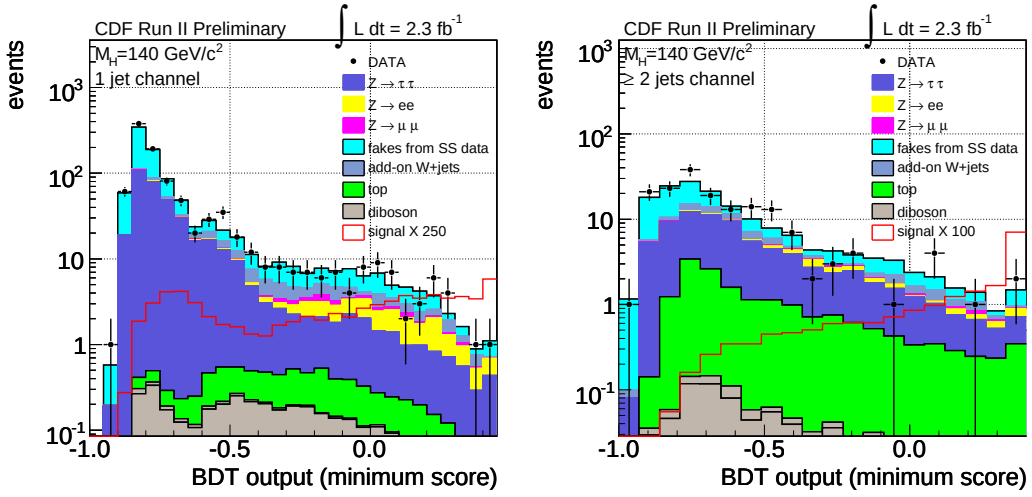

Figure D.16: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 140 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

D.9 $M_H = 145$ GeV/c^2

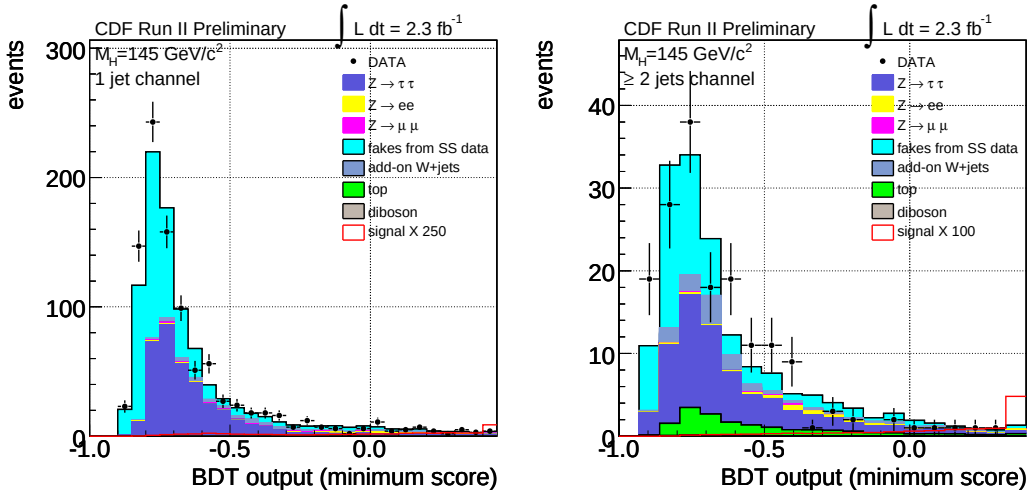

Figure D.17: Final discriminant for the Higgs mass hypothesis of 145 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

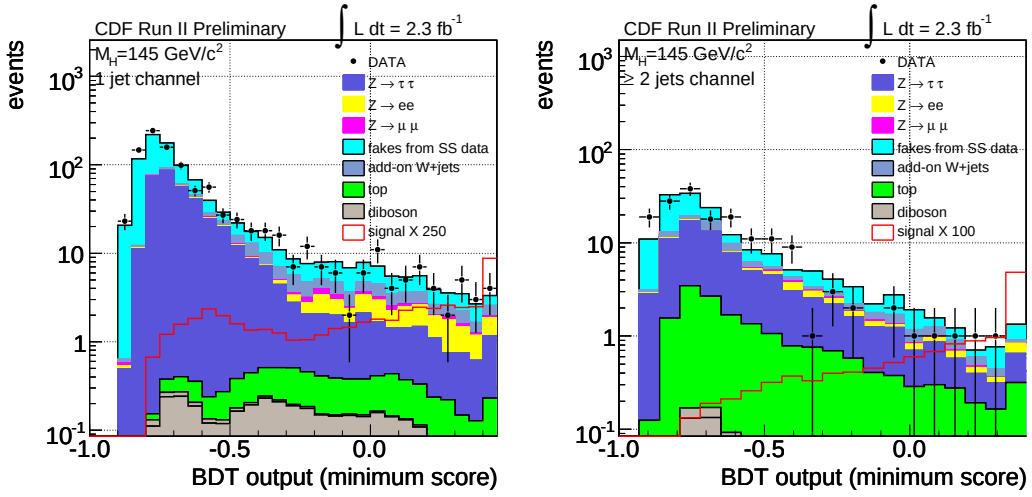

Figure D.18: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 145 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

D.10 $M_H = 150$ GeV/c^2

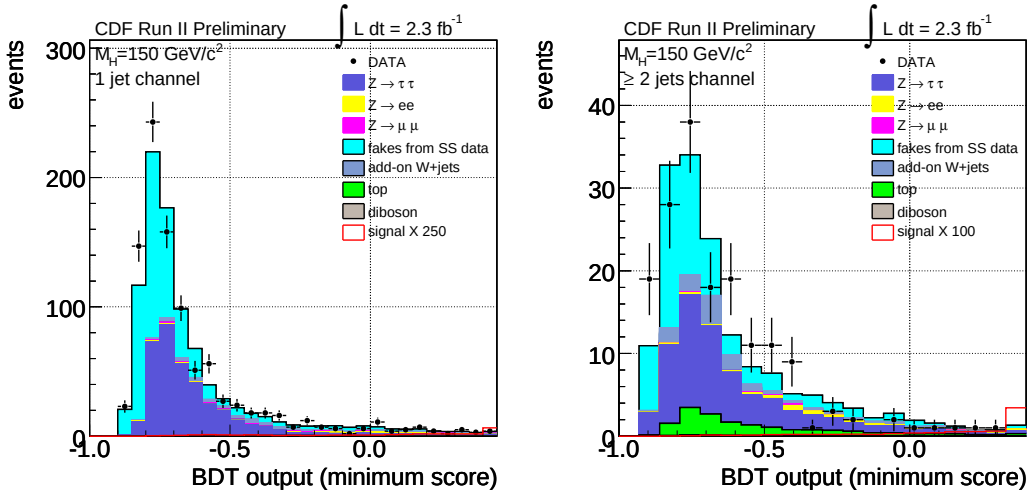

Figure D.19: Final discriminant for the Higgs mass hypothesis of 150 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

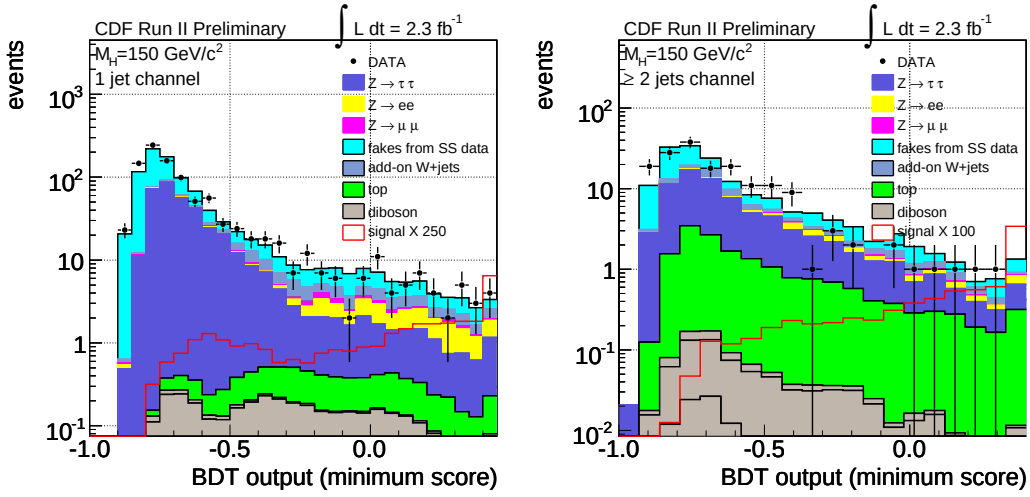

Figure D.20: Final discriminant in logarithmic scale for the Higgs mass hypothesis of 150 GeV/c^2, obtained by taking for each event the minimum score of each trained BDT. Left: 1 jet channel; right: $\geq$2 jets channel.

Appendix E

Distributions for the expected limit

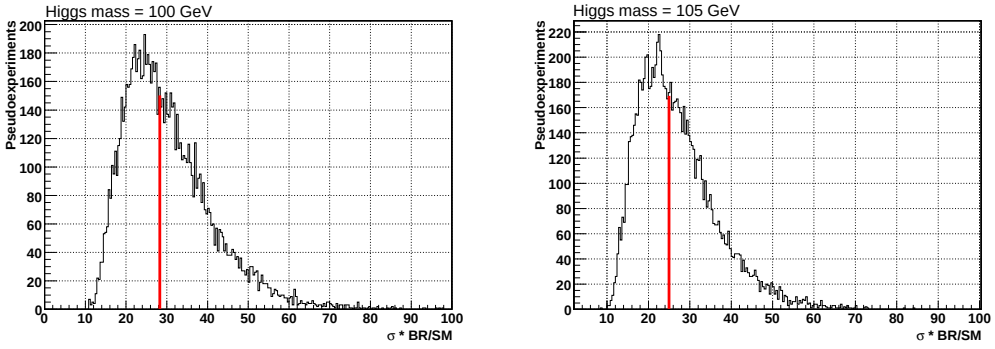

Figure E.1: Expected distribution of 95% C.L. limits in the background-only hypothesis. Left: M_H=100 GeV/c^2; Right: M_H=105 GeV/c^2. The red line shows the median of the distribution; $\pm 1\sigma$ and $\pm 2\sigma$ intervals around the median are obtained by taking 68% and 95% of the total area, respectively.

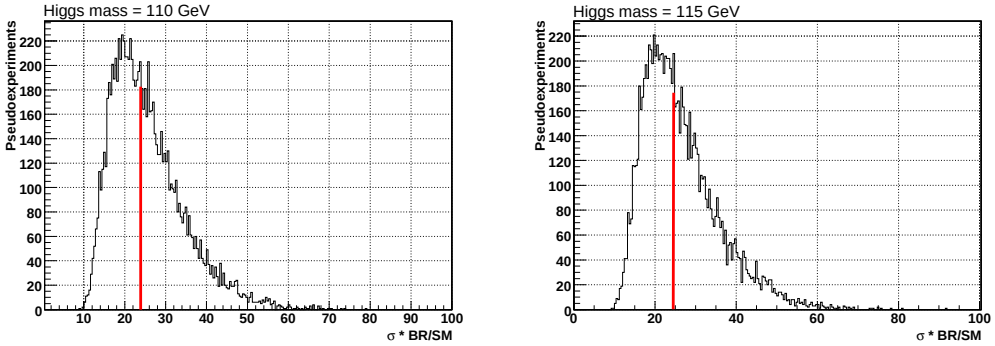

Figure E.2: Expected distribution of 95% C.L. limits in the background-only hypothesis. Left: M_H=110 GeV/c^2; Right: M_H=115 GeV/c^2. The red line shows the median of the distribution; $\pm 1\sigma$ and $\pm 2\sigma$ intervals around the median are obtained by taking 68% and 95% of the total area, respectively.

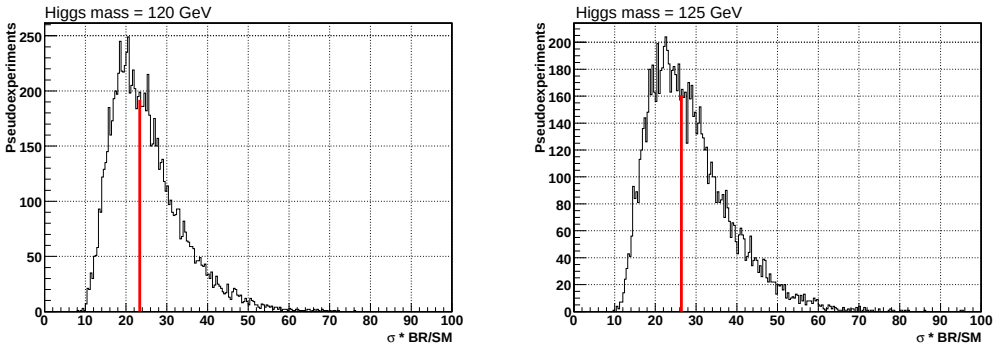

Figure E.3: Expected distribution of 95% C.L. limits in the background-only hypothesis. Left: M_H=120 GeV/c^2; Right: M_H=125 GeV/c^2. The red line shows the median of the distribution; $\pm 1\sigma$ and $\pm 2\sigma$ intervals around the median are obtained by taking 68% and 95% of the total area, respectively.

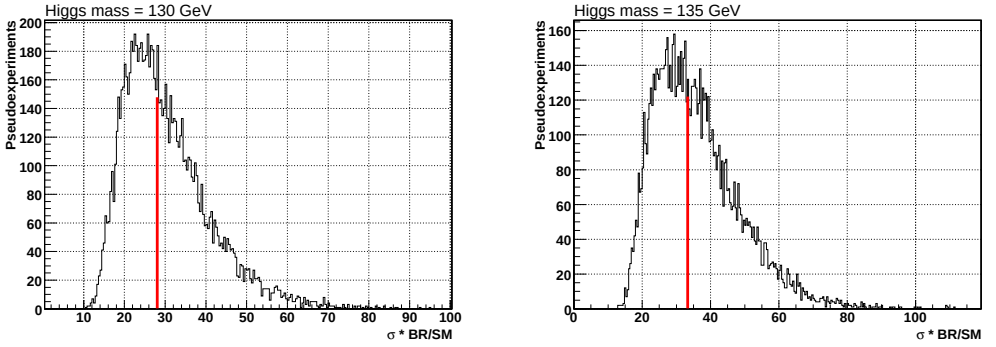

Figure E.4: Expected distribution of 95% C.L. limits in the background-only hypothesis. Left: M_H=130 GeV/c^2; Right: M_H=135 GeV/c^2. The red line shows the median of the distribution; $\pm 1\sigma$ and $\pm 2\sigma$ intervals around the median are obtained by taking 68% and 95% of the total area, respectively.

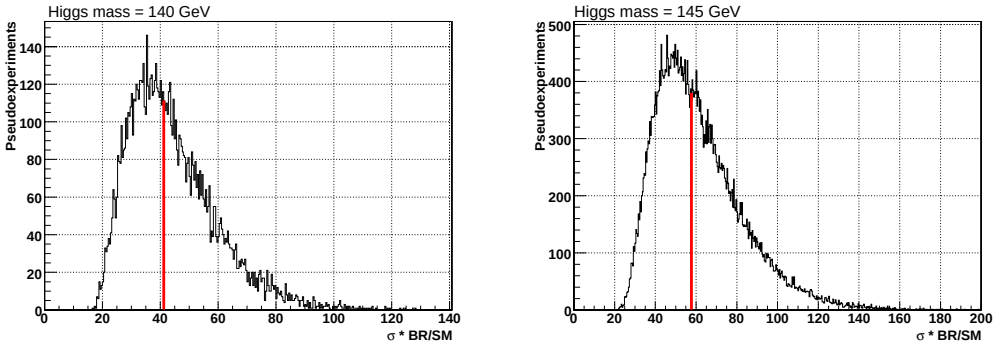

Figure E.5: Expected distribution of 95% C.L. limits in the background-only hypothesis. Left: M_H=110 GeV/c^2; Right: M_H=115 GeV/c^2. The red line shows the median of the distribution; $\pm 1\sigma$ and $\pm 2\sigma$ intervals around the median are obtained by taking 68% and 95% of the total area, respectively.

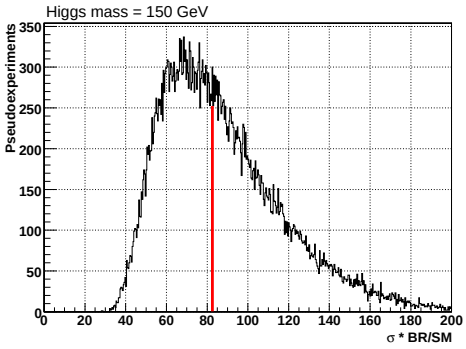

Figure E.6: Expected distribution of 95% C.L. limits in the background-only hypothesis, for M_H=150 GeV/c^2. The red line shows the median of the distribution; $\pm 1\sigma$ and $\pm 2\sigma$ intervals around the median are obtained by taking 68% and 95% of the total area, respectively.

Bibliography

[1] K. Nakamura et al. (Particle Data Group). Review of Particle Physics. *J.Phys. G*, 37, 2010.

[2] J.Kalinowski A.Djouadi and M.Spira. HDECAY: a Program for Higgs Boson Decays in the Standard Model and its Supersymmetric Etension. *Comput.Phys.Commun.*, 108:56, 1998. arXiv:9704448.

[3] Y-K.Kim K.Yorita. Search for the Standard Model Higgs Boson using tau, 2008. CDF/ANAL/EXOTIC/CDFR/9179.

[4] A.Canepa A.Attal. Photon Conversion Removal Efficiency, 2006. CDF/ANAL/ELECTRON/CDFR/8073.

[5] M.Carena et al. Run III: Continues Running of the Tevatron Collider Beyond 2011, 2010. `http://beamdocs.fnal.gov/AD-public/DocDB/ShowDocument?docid=3617`.

[6] A.Hocker et al. Toolkit for multivariate data analysis with root, 2007. `http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf`.

[7] R.Barate et al. Search for the Standard Model Higgs Boson at LEP. *Phys.Lett.B*, 565:61–75, 2003.

[8] The Lep Electro-Weak Working Group, 2010. `http://lepewwg.web.cern.ch/LEPEWWG`.

[9] ALEPH Collaboration et al. Precision Electroweak Measurements and Constraints on the Standard Model, 2010. arXiv:1012.2367.

[10] GFITTER: A Generic Fitter Project for HEP Model Testing, 2010. `http://gfitter.desy.de`.

[11] A.Anastassov et al. Search for Neutral MSSM Higgs boson(s) in the $\tau\tau$Decay Channel, 2005. CDF/PHYS/EXOTIC/CDFR/7622.

[12] The CMS Collaboration. The CMS physics reach for searches at 7 tev, 2010. CMS NOTE 2010/008.

[13] The ATLAS Collaboration. ATLAS Sensitivity Prospects for Higgs Boson Production at the LHC Running at 7, 8 or 9 TeV, 2010. ATL-PHYS-PUB-2010-015.

[14] D.Perkins. *Introduction to High Energy Physics*. Cambridge University Press, 2000.

[15] D.Griffiths. *Introduction to Elementary Particles*. Wiley, 1987.

[16] F.Halzen A.D.Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1984.

[17] M.E.Peskin and D.V.Schroeder. *An Introduction To Quantum Field Theory*. Addison-Wesley, 1995.

[18] B.R.Webber R.K.Ellis, W.J.Stirling. *QCD and Collider Physics*. Cambridge Monogr.Part.Phys.Nucl.Phys.Cosmol., 1996.

[19] S.Weinberg. A Model of Leptons. *Phys.Rev.Lett.*, 19:1264, 1967.

[20] A.Salam. Weak and Electromagnetic Interactions. *Elementary Particle Theory, Proc. 8th Nobel Symp.*, pages 367–377, 1968.

[21] P.W.Higgs. *Phys.Lett.*, 12:132, 1964.

[22] C.Quigg B.W.Lee and H.B.Thacker. The Strength of Weak Interactions at Very High Energies and the Higgs Boson Mass. *Phys.Rev.Lett.*, 38:883–885, 1977.

[23] T.Hambye and K.Riesselmann. Matching Conditions and Higgs Mass Upper Bounds Revisited. *Phys.Rev.D*, 55:7255–7262, 1997.

[24] A.Djouadi. The Anatomy of Electro-Weak Symmetry Breaking. I: The Higgs boson in the Standard Model. *Phys. Rept.*, 457:1–216, 2008.

[25] G.Altarelli and G.Isidori. Lower Limit on the Higgs Mass in the Standard Model: an Update. *Phys.Lett.B*, 337:141–144, 1994.

[26] The Tevatron New Physics Higgs Working Group The CDF Collaboration, the D0 Collaboration. Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 6.7 fb^{-}1 of Data, 2010. arXiv:1007.4587.

[27] Zeuthen ZFITTER new page at DESY, 2010. `http://www-zeuthen.desy.de/theory/research/zfitter/`.

[28] GSM: A GFitter Package for the Global Electroweak Fit, 2010. `http://gfitter.desy.de/GSM/`.

[29] D.de Florian and M.Grazzini. Higgs production through gluon fusion: updated cross sections at the Tevatron and the LHC. *Phys.Lett.B*, 674:291, 2009.

[30] R.Boughezal C.Anastasiou and F.Petriello. Mixed QCD-electroweak corrections to Higgs boson production in gluon fusion. *JHEP*, 0904:003, 2009.

[31] M.Grazzini S.Catani, D.de Florian and P.Nason. Soft-gluon resummation for Higgs boson production at hadron colliders. *JHEP*, 0307:028, 2003.

[32] U.Aglietti et al. Two-loop electroweak corrections to Higgs production in proton-proton collisions, 2006. arXiv:hep-ph/0610033v1.

[33] C.Sturm S.Actis, G.Passarino and S.Uccirati. NLO Electroweak Corrections to Higgs Boson Production at Hadron Colliders. *Phys.Lett.B*, 670:12, 2008.

[34] C.Anastasiou et al. Perturbative QCD effects and the search for a h→ww→ $l\nu l\nu$ signal at the Tevatron, 2009. arXiv:0905.3529v3.

[35] J.Baglio and A.Djounadi. arXiv:1003.4266.

[36] J.Campbell and K.Ellis. Monte Carlo for FeMtobarn processes. `http://mcfm.fnal.gov`.

[37] Fermilab. Run II Handbook, 2009. http://ww-ad.fnal.gov/runII/index.html.

[38] F.B. Division. The Linac Rookie Book, 2006. `http://www-bdnew.fnal.gov/operations/rookie_books/LINAC_v2.pdf`.

[39] F.B. Division. The Booster Rookie Book, 2009. `http://www-bdnew.fnal.gov/operations/rookie_books/Booster_v4.1.pdf`.

[40] F.B. Division. The Main Injector Rookie Book, 2010. `http://www-bdnew.fnal.gov/operations/rookie_books/Main_Injector_v1.1.pdf`.

[41] F.B. Division. The Antiproton Source Rookie Book, 2009. `http://www-bdnew.fnal.gov/operations/rookie_books/Pbar_v2.pdf`.

[42] F.B. Division. The Recycler Rookie Book, 2010. `http://www-bdnew.fnal.gov/operations/rookie_books/Recycler_v1.141.pdf`.

[43] R.Blair et al. *The CDF-II detector: Technical design report*. FERMILAB-PUB-96-930-E, 1996.

[44] N.Timothy. The CDF L00 Detector, 2001. CDF/PUB/SEC_VTX/PUBLIC/5780.

[45] A.Sill. CDF Run II Silicon Tracking Projects. *Nucl.Instrum. Methods A*, 447, 2000.

[46] A.Affolder et al. Status Report of the Intermediate Silicon Layers Detector at CDF II. *Nucl.Instrum. Methods A*, 485, 2002.

[47] A.Affolder et al. CDF Central Outer Tracker. *Nucl.Instrum. Methods A*, 526, 2004.

[48] C.Grozie et al. A Time-of-Flight Detector for CDF. *Int.J.Mod.Phys.*, A16S1C:1119–1121, 2001.

[49] L.Balka et al. The CDF Central Electromagnetic Calorimeter. *Nucl.Instrum. Methods A*, 267:272, 1988.

[50] S.R.Hahn et al. Calibration Systems for the CDF Central Electromagnetic Calorimeter. *Nucl.Instrum. Methods A*, 267, 1988.

[51] S.Bertolucci et al. The CDF Central and Endwall Hadron Calorimeter. *Nucl.Instrum. Methods A*, 267:301, 1988.

[52] L.Nodulman. Hybrid Shower Counter for CDF. *Nucl.Instrum. Methods A*, 176:345, 1980.

[53] L.Nodulman. Central Shower Counter Prototype for the Fermilab Collider Detector Facility. *Nucl.Instrum. Methods A*, 204:351, 1983.

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

[55] M.Albrow et al. The CDF Plug Upgrade Electromagnetic Calorimeter: Test Beam Results. *Nucl.Instrum. Methods A*, 480, 2002.

[56] M.Albrow et al. A Preshower Detector for the CDF Plug Upgrade: Test Beam Results. *Nucl.Instrum. Methods A*, 431, 1999.

[57] G.Apollinari et al. Shower Maximum Detector for the cdf Plug Upgrade Calorimeter. *Nucl.Instrum. Methods A*, 412, 1998.

[58] G.Ascoli et al. CDF Central Muon Detector. *Nucl.Instrum. Methods A*, 268, 1988.

[59] C.M.Ginsburg et al. CDF Run II Muon System. *Eur.Phys.J.*, 33(S1002), 2004.

[60] C.Avila et al. A Measurement of the Proton-Antiproton Total Cross Section at $\sqrt{s} = 1.8$ tev. *Phys.Lett.B*, 445:419, 2002.

[61] D.Acosta et al. The Performance of the CDF Luminosity Monitor. *Nucl.Instrum.Methods*, 494, 2002.

[62] E.J.Thomson et al. Online Track Processor for the CDF Upgrade. *IEEE Trans.Nucl.Sci.*, 49:1063–1070, 2002.

[63] A.Abulencia et al. The CDF II Level 1 Track Trigger Upgrade. *IEEE Trans.Nucl.Sci.*, 55:126–132, 2008.

[64] A.Bhatti et al. Level-2 Calorimeter Trigger Upgrade at CDF. *IEEE Trans.Nucl.Sci.*, 56:1685–1689, 2009.

[65] A.Abulencia et al. The CDF II 3D-Track Level 2 Trigger Upgrade. In *Real-Time Conference, 2007 15th IEEE-NPSS*, pages 1–5, 2007.

[66] M.Shimojima et al. Consumer-server/logger system for the CDF experiment. *IEEE Trans.Nucl.Sci.*, 47:236–239, 2000.

[67] J.Antos et al. Data Processing Model for the CDF Experiment. *IEEE Trans.Nucl.Sci.*, 53:2897–2906, 2006.

[68] V.M. Abazov et al. (D0 Collaboration) T. Aaltonen et al. (CDF Collaboration). Combination of Tevatron Searches for the Standard Model Higgs Boson in the W^+W^- Decay Mode. *Phys.Rev.Lett.*, 104(6):061802, 2010.

[69] T.Aaltonen et al. (CDF Collaboration). Search for a Standard Model Higgs Boson in wh$\rightarrow lvb\bar{b}$ in $p\bar{p}$ Collisions at $\sqrt{s}$= 1.96 tev. *Phys. Rev. Lett.*, 103(10):101802, 2009.

[70] V.M. Abazov et al. (D0 Collaboration). Search for Associated w and Higgs Boson Production in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96 tev. *Phys. Rev. Lett.*, 102(5):051803, 2009.

[71] T.Aaltonen et al. (CDF Collaboration). Search for the Higgs Boson Produced in Association with z$\rightarrow \ell^+\ell^-$ in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96 tev. *Phys. Rev. Lett.*, 101(25):251803, 2008.

[72] T.Aaltonen et al. (CDF Collaboration). Search for the Higgs boson produced in association with z$\rightarrow \ell^+\ell^-$ using the matrix element method at CDF II. *Phys. Rev. D*, 80(7):071101, 2009.

[73] V.M.Abazov et al. (D0 Collaboration). Search for $zh \rightarrow l^+l^-b\bar{b}$ Production in 4.2 fb^{-1} of $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96 tev. *Phys. Rev. Lett.*, 105(25):251801, 2010.

[74] T.Aaltonen et al. (CDF Collaboration). Search for the Higgs Boson Using Neural Networks in Events with Missing Energy and b-Quark Jets in $p\bar{p}$ Collisions at $\sqrt{s}$= 1.96 tev. *Phys. Rev. Lett.*, 104(14):141801, 2010.

[75] V.M.Abazov et al. (D0 Collaboration). Search for the Standard Model Higgs Boson in the zh$\rightarrow \nu\nu b\bar{b}$ Channel in 5.2 fb^{-1} of $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96 tev. *Phys. Rev. Lett.*, 104(7):071801, 2010.

[76] T.Aaltonen et al. (CDF Collaboration). Search for the Associated Production of the Standard-Model Higgs Boson in the All-Hadronic Channel. *Phys. Rev. Lett.*, 103(22):221801, 2009.

[77] V.M.Abazov et al. (D0 Collaboration). Search for Resonant Diphoton Production with the DØ Detector. *Phys. Rev. Lett.*, 102(23):231801, 2009.

[78] The ATLAS Collaboration. Expected performance of the ATLAS experiment : detector, trigger and physics, 2009. arXiv:0901.0512.

[79] Cristobal Cuenca Almenar. *Search for the neutral MSSM Higgs bosons in the tau-tau decay channels at CDF Run II.* PhD thesis, Unversitat de Valencia, 2008.

[80] A.Anastassov et al. Search for Neutral MSSM Higgs boson(s) in the $\tau\tau$Decay Channel, 2007. CDF/PHYS/EXOTIC/CDFR/8972.

[81] B.P.Roe et al. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nucl.Instrum. Methods A*, 543:577–584, 2005.

[82] S. Rolli D. Hare, E. Halkiadakis. Electron id efficiencies and scale factors for periods 9 to 12, 2007. CDF/DOC/ELECTRON/CDFR/9148.

[83] X.Zhang U.Grundler, A.Taffard. High-pT muons reccommended cuts and efficiencies for summer 2006, 2006. CDF/ANAL/TOP/CDFR/8262.

[84] A.Anastassov et al. Tau Reconstruction Efficiency and QCD Fake Rate for Run 2, 2003. CDF/ANAL/EXOTIC/CDF/6308.

[85] A.Anastassov. Non-isolated π^0/γ Reconstruction, 2003. CDF/ANAL/EXOTIC/CDF/6688.

[86] S.Baroiant et al. Energy Measurement for Hadronic Taus, 2003. CDF/ANAL/EXOTIC/CDF/6654.

[87] H.-J.Yang et al. Studies of stability and robustness for artificial neural networks and boosted decision trees. *Nucl.Instrum. Methods A*, 574:342–349, 2007.

[88] A.Bhatti et al. Determination of the jet energy scale at the collider detector at Fermilab. *Nucl.Instrum. Methods A*, 566:375–412, 2006.

[89] S.Eidelman et al. (Particle Data Group). Review of Particle Physics. *Phys.Lett.B*, 592:1, 2004.

[90] A.Taffard. Run II Cosmic Ray Tagger Performances, 2003. CDF/PUB/MUON/PUBLIC/6255.

[91] P.Skands T.Sjosrand, S.Mrenna. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[92] M.Mangano et al. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003.

[93] CTEQ Collaboration H.L.Lai et al. Global QCD Analysis of Parton Structure of the Nucleon: CTEQ5 Parton Distributions. *Eur.Phys.J.*, C12:375–392, 2000.

[94] S.Jadach et al. The tau decay library TAUOLA: Version 2.4. *Comput.Phys.Commun.*, 76:361–380, 1993.

[95] R.Brun et al. GEANT: Simulation Program for Particle Physics Experiments. User Guide and Reference Manual, 1978. CERN-DD-78-2-REV.

[96] S.Peters G.Grindhammer, M.Rudowicz. The Fast Simulation of Electromagnetic and Hadronic Showers. *Nucl.Instrum. Methods A*, 290:469–488, 1990.

[97] R.Veenhof. GARFIELD, recent developments. *Nucl.Instrum. Methods A*, 419:726–730, 1998.

[98] S.Dube et al. Medium E_t Electron identification efficiency and scale factors, 2006. CDF/DOC/EXOTIC/CDFR/8321.

[99] E.Lytken et al. Intermediate p_t muon ID efficiencies with the 1 fb^{-1} dataset, 2006. CDF/DOC/EXOTIC/CDFR/8336.

[100] W.K.Sakumoto. Event —Zvtx—¡60 Cut Acceptance for Run II, 2006. CDF/ANAL/ELECTROWEAK/CDFR/8318.

[101] D.Acosta et al. First Measurements of Inclusive W and Z Cross Sections from Run II of the Fermilab Tevatron Collider. *Phys.Rev.Lett.*, 94(9):091803, 2005.

[102] J.Campbell and K.Ellis. Update on vector boson pair production at hadron colliders. *Phys. Rev. D*, 60(11):9, 1999.

[103] P.Uwer U.Langenfeld, S.Moch. Measuring the running top quark mass. *Phys.Rev.D*, 80(5):054009, 2009.

[104] A.D.Martin et al. Parton distributions for the LHC. *Eur.Phys.J.*, 63(2):189–285, 2009.

[105] The CDF Collaboration. Combined Upper Limit on Standard Model Higgs Boson Production in up to 5.9 fb^{-1} of Data, 2010. CDF Public Note 10223.

[106] The CDF Collaboration. Search for the Standard Model Higgs Boson in H$\rightarrow \tau\tau$ Channel at CDF Run II. Simultaneous Search for WH/ZH/VBF/H Processes, 2008. CDF Public Note 9248.

[107] T.Junk. Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds and Large Systematic Uncertainties, 2007. CDF/DOC/STATISTICS/PUBLIC/8128.

[108] The DØ Collaboration. Combined Upper Limits on standard model Higgs boson production from the DØ experiment in up to 6.7 fb^{-1} of Data, 2010. DØ Conference Note 6094.

[109] T.Junk. Confidence Level Computation for Combining Searches with Small Statistics. *Nucl.Instrum. Methods A*, 434:435–443, 1999.

[110] A.L.Read. Presentation of search results: the CLs technique. *J.Phys. G*, 28:2693, 2002.

[111] H.E.Haber. Present status and future prospects for a Higgs boson discovery at the Tevatron and LHC, 2010. arXiv:1011.1038.