FERMILAB-THESIS-2011-70

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

DIPARTIMENTO DI FISICA "G. GALILEI"

SCUOLA DI DOTTORATO DI RICERCA IN FISICA

CICLO XXIII

Search for a high-mass Higgs boson produced in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV with an hadronic τ in the final state

Direttore della Scuola:
Prof. Andrea Vitturi

Dottoranda:
Dott.ssa Maria d'Errico

Supervisore:
Prof.ssa Donatella Lucchesi

Operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy

Abstract

In this Thesis a search for Higgs boson production has been performed analysing data gathered by the Collider Detector at Fermilab (CDF) from proton-antiproton ($p\bar{p}$) collisions at the center of mass energy $\sqrt{s} = 1.96$ TeV.
Event selection has been optimized on Higgs decaying into W^+W^- bosons in order to reconstruct final states with two charged leptons formed by either an electron and a tau or a muon and a tau focusing on taus decaying into hadrons.
No evidence of Higgs production has been found analysing 4.8 fb^{-1} of data and an upper limit has been set on the production cross section as a function of the mass m_H in the range from 130 to 200 GeV/c^2 with a 95% Confidence Level. For a Higgs mass hypothesis of 160 GeV/c^2 the observed limit is 33.5 times the Standard Model predicted cross section σ_{SM} while the expected upper limit for the same mass is 19.6$\times\sigma_{SM}$. The high recorded luminosity along with the development of new analysis techniques are expected to improve the CDF sensitivity to the search for Higgs boson with hadronic taus reconstructed in the final state, in the mass range 130 to 200 GeV/c^2.

Sommario

In questa Tesi viene presentata la ricerca della particella di Higgs in un campione di dati raccolti dall'esperimento Collider Detector at Fermilab (CDF) e generati dalle collisioni protone-antiprotone all'energia del centro di massa di $\sqrt{s} = 1.96$ TeV.
La selezione degli eventi è stata ottimizzata per la ricerca del bosone di Higgs quando questo decade in due bosoni W^+W^- con stato finale costituito da un elettrone ed un tau o da un muone ed un tau, ricostruendo il tau tramite i suoi prodotti di decadimento adronici.
L'analisi, effettuata su 4.8 fb^{-1} di dati raccolti, non mostra alcuna evidenza della produzione dell'Higgs ma è stato possibile determinare il limite sulla sezione d'urto di produzione in funzione della massa m_H nell'intervallo tra 130 e 200 GeV/c^2 con il 95% di Livello di Confidenza. Per $m_H = 160$ GeV/c^2 il limite osservato è 33.5 volte la sezione d'urto predetta dal Modello Standard (σ_{SM}) mentre il limite atteso è di $19.6 \times \sigma_{SM}$.
L'utilizzo di nuove tecniche di analisi e l'aumento del campione di dati utilizzato potranno contribuire ad aumentare la sensibilità di CDF nella ricerca dell'Higgs con tau adronici nello stato finale, nella regione di massa 130 e 200 GeV/c^2.

Contents

Introduction

The Standard Model of elementary particles and electroweak interactions is the most successful theory of modern particle physics. It was extensively tested by recent experiments and no discrepancy from the theoretical predictions has been found yet. Particle masses are introduced in the Standard Model using the so called electroweak spontaneous symmetry breaking mechanism, which predicts the existence of a scalar particle, the Higgs boson, up to now, not experimentally observed.

The Tevatron, a proton-antiproton collider, is installed at the Fermi National Laboratory at Fermilab in Batavia (Illinois, USA). It provides proton-antiproton collisions at a center of mass energy of 1.96 TeV, collected and investigated at CDF and DØ detectors. The high recorded luminosity along with the great performance of the accelerator and of detectors, the good understanding of data and the development of new analysis techniques allow to search for tiny signal, such as the expected signature from a decay of a Higgs boson.
The search of the Higgs boson at Tevatron exploits different decay channels, according to the varying decay properties of the Higgs boson with its mass. One of the most relevant channel is the Higgs decay into two leptons through a WW intermediate state, used as a search channel in the Higgs mass range $135 \leq M_H \leq 200$ GeV/c^2. This Thesis concentrates on the Higgs search in the $H \to WW \to l\nu\tau\nu$ channel ($l = e,\ \mu$), with τ decaying into hadrons, based on 4.8 fb^{-1} of data.

A general introduction to the Higgs boson properties and to the limits on its mass, coming from theoretical arguments and from experimental searches, is given in Chapter 1. The chapter focuses on the Higgs boson searches at hadron colliders, describing the production and the decay properties as function of Higgs mass at the Tevatron and, for comparison, at the proton-proton Large Hadron Collider (LHC).
In Chapter 2 the Tevatron collider and the overall design of the CDF detector are described, together with the CDF trigger system, which plays a crucial role in the event reduction and selection.
Chapter 3 and 4 are dedicated to the event selection and to the description of the "objects" used in the event reconstruction, with the goal to maximize the total acceptance for the final state $(e - \tau)$ and $(\mu - \tau)$. The hadronic tau reconstruction algorithm and identification criteria in the CDF detector are studied in detail in Chapter 3. Jets misidentified as τ represent the main problem in tau reconstruction. Some variables able to select taus with a reduced fraction of fake taus coming from jets are taken into consideration. The selection of $(e-\tau)$ and $(\mu-\tau)$ events is discussed in Chapter 4. Key aspects of the analysis are represented by reconstruction and identification efficiency of leptons and by energy and direction resolutions. All these aspects are analyzed and discussed. Particular emphasis is given to the estimation of the processes with fake taus

entering the selected final sample. In this Thesis a method based on combined use of data and MC simulated event has been applied. However, a study about the feasibility of the fake background estimation from only data events has been carried out and reported in this chapter.
The expected sample composition of the selected data, having the required $H \to WW \to e/\mu + \tau$ signature, is discussed in Chapter 5, where 4.8 fb^{-1} of data collected by CDF have been analyzed.
The search for the Higgs boson is performed using multivariate techniques to improve the signal discrimination from backgrounds. This Thesis exploits Neural Network multivariate technique and results of the analysis based on 4.8 fb^{-1} of data are shown in Chapter 6 as a conclusion of the $H \to WW$ search in the $(e - \tau)$ and $(\mu - \tau)$ channels.

Chapter 1

Physics motivation: the Higgs boson in the Standard Model

Our current understanding of the subatomic world is summarized in the so called Standard Model of Particle Physics (SM), formulated in 1960's and 70's [1, 2, 3]. In this Chapter an overview of the SM is given, describing the elementary particles and their interactions [3]. The following Sections present an overview of the electro-weak sector in the SM and of the mechanism of spontaneous electro-weak symmetry breaking, which plays a central role by providing mass to particles [4, 5, 6]. The Higgs boson is predicted as a consequence of it, but its existence is still not experimentally confirmed. Indirect experimental constraints on the Higgs boson mass and the searches for the Higgs boson conducted at hadron colliders are also reviewed in this Chapter.

1.1 The Standard Model

The SM is a unified framework proposed to describe the electro-weak and strong interactions of elementary particles, based on the principles of the Quantum Field Theory [7], while the problem of incorporating gravity in the theory is still open [8]. Within SM, the fundamental constituents of matter are six leptons and six quarks, arranged into three families called *generations*. The SM incorporates the Glashow-Weinberg-Salam unified gauge theory [9, 10, 11] of Quantum Electro-Dynamics (QED) and of weak interactions between quarks and leptons, and the Quantum Chromo-Dynamics (QCD) gauge theory of strong interactions between quarks only [12], respectively based on the symmetry group SU(2)×U(1) and SU(3) [7].
In Quantum Field Theories the fundamental particles are described by *matter fields* and their interactions are invariant under space-time dependent phase transformations (local gauge invariance) of the fields. Indeed, in the SM all fundamental interactions derive from the requirement of local gauge invariance, which is guaranteed by introducing *gauge fields* representing each interaction. The fundamental interactions are then mediated by *gauge boson*, to which the gauge field is associated. The carrier of the electromagnetic force is the photon γ, which is massless and chargeless. Gluons are eight massless bosons of SU(3), mediators of the strong force. The weak force is mediated by three vector bosons: W^+, W^- and Z^0; having a large observed mass. The most up-to-date mass values for Z and W bosons, discovered at CERN SPS $p\bar{p}$ collider in 1983

[13, 14, 15, 16], are $M_W = (80.399 \pm 0.023)$ GeV/c^2 [3], and $M_Z = (91.1876 \pm 0.0021)$ GeV/c^2 [3].
Figure 1.1 shows the fundamental particles organized according to their mass and divided into three categories: gauge bosons, quarks and leptons.

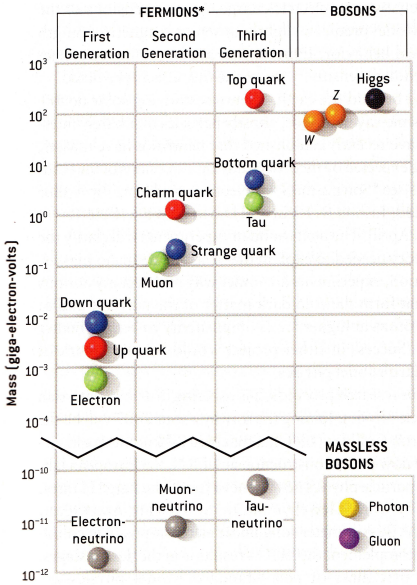


Figure 1.1: Elementary particles are organized according to their mass and divided in the three categories: gauge bosons, quarks and leptons.

1.1.1 The electro-weak sector and the SU(2)×U(1) gauge symmetry violation

The electro-weak theory is a unified gauge theory, invariant under SU(2)×U(1) transformations, according to the Glashow-Weinberg-Salam model formulated in 1961 [9, 10, 11]. This Section will show how, using the Lagrangian formalism to describe the electro-weak interactions, the prediction of massless fermions within the SM arises from the preservation of the SU(2)×U(1) invariance in the Lagrangian.
SU(2)$_L$ is the weak isospin non-Abelian group which acts on left-handed fermions. Weak interactions show common properties between doublets of fermions, therefore, left-handed fermions are in weak isodoublets, while the right-handed fermions are in weak isosinglets:

$$\psi_L = \begin{pmatrix} \nu_l \\ l \end{pmatrix}_L, \begin{pmatrix} u \\ d \end{pmatrix}_L \qquad \psi_R = e_R, u_R, d_R, \ldots \tag{1.1}$$

where "L" and "R" mean left-handed and right-handed respectively.
Electromagnetic interactions described by QED are invariant under U(1)$_Q$ Abelian gauge group, with the conservation of charge Q coming from local U(1) invariance. Table 1.1 summarizes the elementary particles with their electric charge Q, mass, third component of the weak isospin I^3 and hypercharge Y, defined as a function of I^3 and Q in unit of

the proton charge $+e$: $Y = 2(Q - I^3)$.

Generation	Quarks					
	Q $= +\frac{2}{3}$	$I^3 = +\frac{1}{2}$	$Y/2 = +\frac{1}{6}$	Q $= -\frac{1}{3}$	$I^3 = -\frac{1}{2}$	$Y/2 = +\frac{1}{6}$
1^{st}	up (u) $M_u \sim 1.5 \div 4.5$ MeV/c^2			down (d) $M_d \sim 4.0 \div 8.0$ MeV/c^2		
2^{nd}	charm (c) $M_c \sim 1.15 \div 1.35$ GeV/c^2			strange (s) $M_s \sim 80 \div 130$ MeV/c^2		
3^{rd}	top (t) $M_t = 173.1 \pm 1.3$ GeV/c^2			bottom (b) $M_b \sim 4.1 \div 4.9$ GeV/c^2		

Generation	Leptons					
	Q $= -1$	$I^3 = -\frac{1}{2}$	$Y/2 = -\frac{1}{2}$	Q $= 0$	$I^3 = +\frac{1}{2}$	$Y/2 = -\frac{1}{2}$
1^{st}	electron (e) $M_e = 0.511$ MeV/c^2			electron neutrino (ν_e) $0 < M_{\nu_e} < 3$ eV/c^2		
2^{nd}	muon (μ) $M_\mu = 106$ MeV/c^2			muon neutrino (ν_μ) $0 < M_{\nu_\mu} < 0.19$ MeV/c^2		
3^{rd}	tau (τ $M_\tau = 1.78$ GeV/c^2			tau neutrino (ν_τ) $0 < M_{\nu_\tau} < 18.2$ eV/c^2		

Table 1.1: Leptons and quarks: mass, electric charge Q, third component of the weak isospin T^3, hypercharge $Y/2$.

In the electro-weak sector four gauge fields can be defined. The field B_μ which corresponds to the Y, generator of the $U(1)_Y$ group, and three fields $W_\mu^{1,2,3}$ which correspond to the generators $T^a = \frac{\sigma^a}{2}$ (σ^a are Pauli matrices, $a = 1, 2, 3$) of the $SU(2)_L$ group. The commutation relations between generators are given by:

$$[T^a, T^b] = i\epsilon^{abc} T_c \qquad \text{and} \qquad [Y, Y] = 0 \tag{1.2}$$

where the ϵ^{abc} is the antisymmetric tensor.

In gauge theory the dynamics of particles is expressed by the Lagrangian density $\mathcal{L}$. An invariance of the $\mathcal{L}$ under a certain transformation of fields hides a symmetry for the system described by the Lagrangian, as the Noether's theorem states [7]. For the electro-weak interaction, the Lagrangian expression is a direct consequence of its invariance under $SU(2)_L \times U(1)_Y$ local gauge transformations.
The Lagrangian contains the kinetic terms for the gauge field:

$$\mathcal{L}_G = -\frac{1}{4} W^i_{\mu\nu} W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.3}$$

expressed through the Yang-Mills and Maxwell field strength tensors, $W^i_{\mu\nu}$ and $B_{\mu\nu}$, defined as follows in terms of electro-weak interaction eigenstates:

$$\begin{aligned} W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{ijk} W_{\mu j} W_{\nu k} \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \tag{1.4}$$

The coupling between the matter field and the gauge bosons is generated by $\mathcal{L}_f = \psi D_\mu \gamma^\mu \psi$. Since ψ_L and ψ_R, defined as (1.1), belong to different SU(2) representations, the covariant derivative D_μ is for convenience defined as:

$$\begin{aligned} D_\mu^L &= \partial_\mu - igT_a W_\mu^a - ig' \frac{Y}{2} B_\mu \\ D_\mu^R &= \partial_\mu - ig' Y B_\mu \end{aligned} \tag{1.5}$$

where g and g' are the gauge couplings for electro-weak interactions.
The resulting Lagrangian describing the electroweak interactions in the unified electroweak theory is thus:

$$\mathcal{L}_{EWK} = \mathcal{L}_f + \mathcal{L}_G \tag{1.6}$$

The mass terms in the form $m\bar{\psi}\psi$, $\frac{1}{2}M_Z^2 Z_\mu Z^\mu$ and $\frac{1}{2}M_W^2 W_\mu W^\mu$ are forbidden in (1.6), in order to preserve the SU(2)×U(1) gauge invariance. Therefore (1.6) can describe only massless gauge bosons and massless fermions, while symmetry breaking must occur to provide the right mass generation. The Standard Model Lagrangian for the electro-weak sector, invariant under gauge transformations, will thus be:

$$\mathcal{L} = \mathcal{L}_G + \mathcal{L}_f + \mathcal{L}_H \tag{1.7}$$

where the $\mathcal{L}_H$ accounts for the spontaneous symmetry breaking, needed in order for the SM to be compatible with the observed masses of fermions, W and Z bosons [3].

1.2 The Higgs mechanism

The mechanism of spontaneus electro-weak symmetry breaking proposed by Higgs-Brout-Englert-Guralnik-Hagen-Kibble [17, 18, 19], or Higgs mechanisms for short, was the first step towards the prediction of a new scalar massive particle in the SM, the Higgs boson, properly defined to be responsable for the spontaneous electro-weak symmetry breaking and the existence of massive $W^\pm$ and Z^0 gauge bosons.
The simplest example of mass generation is given considering the Lagrangian for a scalar complex field $\phi(x) = \frac{1}{\sqrt{2}}(\phi_1(x) + i\phi_2(x))$ with the potential V($\phi^*\phi$):

$$\mathcal{L}_H = \partial_\mu \phi^* \partial^\mu \phi - V(\phi^*\phi) = \partial_\mu \phi^* \partial^\mu \phi - \mu^2 \phi^* \phi - \frac{\lambda}{2}(\phi^*\phi)^2 \tag{1.8}$$

It is then simply the Lagrangian of a spin-zero particle of mass μ, if $\mu^2 > 0$. In turn, if $\mu^2 < 0$ the potential $V(\phi^*\phi)$ becomes the interacting potential V with the minimum:

$$< 0|\phi^*\phi|0 > \equiv |\phi_0|^2 = -\frac{\mu^2}{\lambda} \equiv v^2 \tag{1.9}$$

The quantity $\pm v \equiv < 0|\phi|0 >$ is called the vacuum expectation value (vev) of the scalar field ϕ. and λ is the positive coupling constant. In Figure 1.2 the $V(\phi^*\phi)$ potential is represented in the case $\mu^2 < 0$. Expanding the Lagrangian as a function of the minimum of the potential V, $\phi(x) = e^{i\theta(x)}[\rho(x) + v]$, a scalar massive field also appears:

$$\mathcal{L}_H = \underbrace{\partial_\mu \rho \partial^\mu \rho + 2\mu^2 \rho^2}_{m^2 = -2\mu^2} \underbrace{- \frac{\mu^2}{2} \partial_\mu \theta \partial^\mu \theta}_{m_\theta = 0} + O(3) \tag{1.10}$$

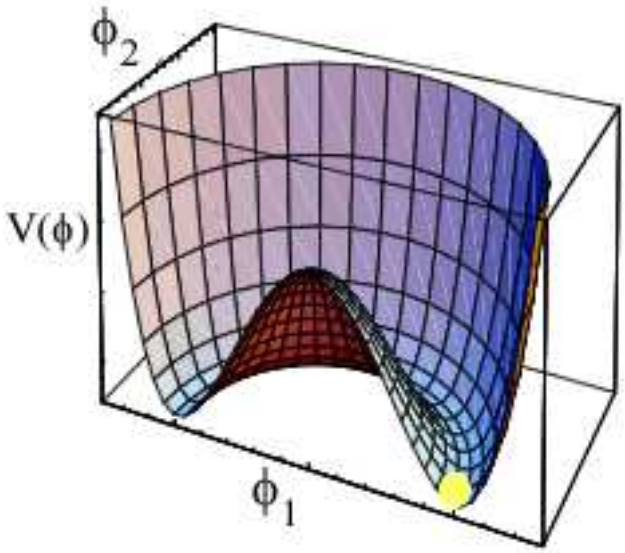


Figure 1.2: The potential V of the scalar complex field ϕ in the case $\mu^2 < 0$.

Thus, we have a massive boson ρ with $m = \sqrt{2}|\mu|$ and a massless boson θ, the Goldstone boson [1].
In the specific case of the local non-abelian SU(2)×U(1) gauge symmetry in SM, this mechanism should generate masses for the three gauge bosons $W^{\pm}$ and Z^0 but the photon should remain massless and QED must stay an exact symmetry [7, 20]. Therefore, the EWK theory described by $\mathcal{L}_{EWK}$ in (1.6) needs at least 3 degrees of freedom more (3 longitudinal polarizations, one for each new massive boson). Those can be introduced as 3 degrees of freedom for the scalar fields in the Higgs mechanism.
Therefore, the minimal choice in this case is four scalar fields ϕ in form of a complex SU(2) doublet [2]:

$$\phi = \begin{pmatrix} \phi^1 \\ \phi^2 \end{pmatrix} \qquad \begin{array}{ccccc} & I^3 & Q & Y \\ \phi^1 & \frac{1}{2} & 1 & 1 \\ \phi^2 & -\frac{1}{2} & 0 & 1 \end{array} \tag{1.11}$$

The field ϕ can be expressed as a function of new fields $\theta_{1,2,3}(\mathrm{x})$ and a physical gauge boson $H(\mathrm{x})$. When the new parameterization of the ϕ is inserted into the Lagrangian, one of the Higgs fields, $H(\mathrm{x})$, acquires mass, while the other three fields θ remain massless, hence the symmetry of the Lagrangian is broken. The invariant term relating to the scalar field, to be added to $\mathcal{L}_{EWK}$ in (1.6), becomes:

$$\mathcal{L}_H = (D^\mu \phi)^\dagger (D_\mu \phi) + \frac{m^2}{2}\phi^\dagger \phi - \frac{\lambda}{4}(\phi^\dagger \phi)^2 - \mathcal{L}_{YW} \tag{1.12}$$

The coupling of the scalar field to gauge bosons $(D^\mu \phi)^\dagger (D_\mu \phi)$ will show bilinear terms in the fields $W_\mu^{\pm}$, Z_μ, using the expression (1.4) for the field strength tensors and D_μ

[1] According to the Goldstone theorem [4]: for every spontaneously broken continuous symmetry, the theory contains massless scalar (spin-0) particles called Goldstone particles.

[2] The number of Goldstone bosons is equal to the number of broken generators. For a O(N) continuous symmetry there are $\frac{1}{2}N(N-1)$ generators. The residual unbroken symmetry O$(N-1)$ has $\frac{1}{2}(N-1)(N-2)$ generators and therefore, there are $N-1$ massless Goldstonee bosons.

defined in (1.5):

$$(D^\mu\phi)^\dagger(D_\mu\phi) = M_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2}M_Z^2 Z_\mu Z^\mu + \frac{1}{2}M_A^2 A_\mu A^\mu \tag{1.13}$$

$W_\mu^\pm$, Z_μ and A_μ are the mass eigenstates for the gauge bosons, expressed as a function of the electro-weak interaction eigenstates and the weak mixing angle θ_W [3]. This makes the W and Z bosons acquire mass, while the photon is still massless:

$$M_W = \frac{1}{2}vg, \quad M_Z = \frac{1}{2}v\sqrt{g^2+g'^2}, \quad \frac{M_W}{M_Z} = \cos\theta_W \qquad M_A = 0 \tag{1.14}$$

where $v = (-\mu^2/\lambda)^{1/2}$. The first hints of the W and Z bosons's mass from theory were:

$$M_W \approx (\frac{37.2}{\sin\theta_W}\text{GeV/c}^2) \sim 48 - 118 \text{ GeV/c}^2 \tag{1.15}$$

$$M_Z \approx (\frac{37.2}{\sin\theta_W \ \cos\theta_W}\text{GeV/c}^2) \sim 76 - 124 \text{ GeV/c}^2 \tag{1.16}$$

Those are in good agreement with experimental results [3].
Spontaneous symmetry breaking also generates lepton masses through the coupling to fermions described by the Yukawa invariant Lagrangian ($\mathcal{L}_{YW}$):

$$\mathcal{L}_{YW} = \sum_{e,\mu,\tau} G_l \frac{v}{\sqrt{2}}(\bar{l}_L l_R + \bar{l}_R l_L) + \frac{G_u v}{\sqrt{2}}\bar{u}_L u_R + \frac{G_d v}{\sqrt{2}}\bar{d}_L d_R \tag{1.17}$$

The parameter in front of the bracket can be interpreted as the mass of the fermions:

$$m_l = \frac{G_l v}{\sqrt{2}}, \quad m_u = \frac{G_u v}{\sqrt{2}}, \quad m_d = \frac{G_d v}{\sqrt{2}} \tag{1.18}$$

where G_l, G_d and G_u are the coupling constants of leptons and quarks to the Higgs field.
Thus, the electromagnetic $U(1)_Q$ and SU(3) color symmetry stay unbroken and the SM refers to SU(3)×SU(2)×U(1) gauge invariance when combined with the electroweak symmetry breaking mechanism. By spontaneously breaking the symmetry SU(2)×U(1), three Goldstone bosons have been absorbed by the $W^\pm$ and Z bosons to form longitudinal polarizations and acquire masses, while the remaining degree of freedom corresponds to a scalar particle, the Higgs boson.

In spite of predicting the existence of the Higgs boson, the theory does not predict its mass, which remains the only parameter in the SM still unknown. Further studies will be required to measure the properties of the Higgs boson, and to see if they match the expectations from the SM.
The Higgs boson couples to particles through spontaneous symmetry breaking of the SU(2)×U(1). As shown in Table 1.2, the Higgs boson couples to fermions proportionally to their mass and to bosons proportionally to the square of theirs. Therefore, it is most easily produced in association with heavy particles and its decays branching fractions are the largest for the heavy particles, such as top quarks, vector bosons and b-quarks, if kinematically allowed. The Higgs boson can also couple to $\gamma\gamma$ through a loop of charged particles, such as W bosons. In a similar way, the Higgs can couple to gluon via quark loop. The couplings of the Higgs boson largely define the strategies used in Higgs boson searches.

[3] The θ_W is the Weinberg angle defined by the coupling constants $g'/g = \tan\theta_W$

Coupling	Intensity
$Hf\bar{f}$	m_f/v
HW^+W^-	M_W^2/v
HZZ	M_Z^2/v

Table 1.2: The Higgs couplings to fermions and bosons.

1.3 Constraints on the Higgs mass

The Higgs boson's properties, such as production modes and corresponding cross sections, will be uniquely determined once its mass is fixed. Up to now, no direct indications of the production of the Standard Model Higgs boson have been found at the current operating hadron colliders, Tevatron and the Large Hadron Collider (LHC), or previously at the Large Electron Positron (LEP) e^+e^- collider at CERN. Therefore, only lower bounds and exclusion regions are allowed to be set using the data collected by the experiments.
Constraints from indirect experimental measurements and theoretical bounds on its mass value are available. Although the most definitive proof of the existence of the Higgs boson will come from the direct observation, indirect constraints on its mass help in determining the direction of the searches. However, the availability of both highly accurate measurements and theoretical predictions provides stringent tests of the SM.

1.3.1 Theoretical constraints

A lower bound can be put on a theoretical basis requiring the minimum of the Higgs potential to be an absolute minimum. Since the Higgs potential is affected by quantum loop corrections, which can generate instabilities in the potential, a stability bound is required. Theoretical considerations suggest that the Higgs boson mass must exceed about 4 GeV/c^2, according to the Linde-Weinberg lower bound [21].
At the same time, the running coupling constant λ in the Higgs potential (1.8) increases with the energy scale Λ according to the renormalization group equations [22]. If M_H is large, the value of λ becomes infinite and the request for λ to remain finite up to a scale corresponds to putting an upper bound on M_H:

$$M_H \leq M = (8\pi\sqrt{s}/3G_F)^{1/2} \sim 1 \text{ TeV}/c^2 \tag{1.19}$$

where G_F is the Fermi constant. If M_H exceeds the critical value M, the perturbation theory will cease to be a faithful representation of physics. A detailed discussion about the Higgs mass theoretical upper bound can be found in [22, 23]. The theoretical bounds on M_H as a function of Λ are shown in Figure 1.3. For Λ up to the Planck scale ($\Lambda = 10^{19}$ GeV), the Higgs mass must be in the range 130-200 GeV/c^2. Assuming a validity of the SM only up to 1 TeV, the Higgs mass can be up to 700 GeV/c^2. In any case, the main indication is that the Higgs boson should be searched in a range of masses below 1 TeV. More stringent limits are in fact coming from direct experimental Higgs searches.

1.3.2 Direct searches at LEP and Tevatron Run-I

Searches for the Standard Model Higgs boson have been performed at LEP in e^+e^- collisions at a center of mass energy $\sqrt{s}$ of about 133 GeV, in the LEP-I phase, and

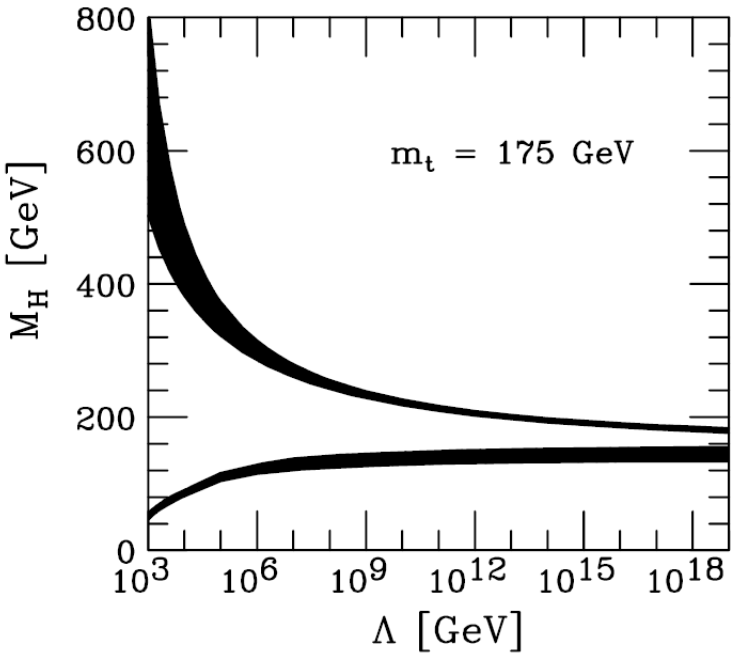


Figure 1.3: Summary of theoretical limits on the mass M_H of the Standard Model Higgs boson. The allowed region, as a function of the energy scale Λ at which the Standard Model breaks down, is between the two curves, obtained assuming the top mass $M_t = 175$ GeV/c^2.

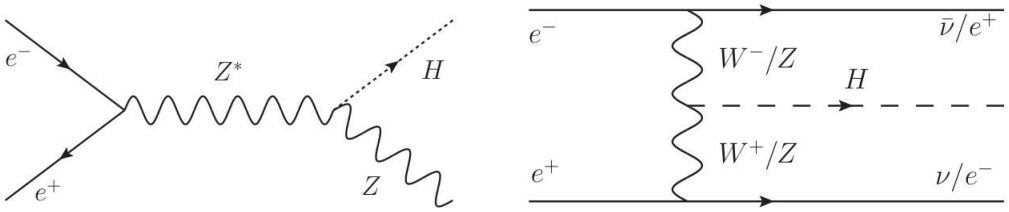


Figure 1.4: The SM dominant Leading Order Feynman diagrams for the production of the Higgs boson at the LEP e^+e^- collider: Higgs-strahlung from a virtual Z (left) and vector boson fusion (right).

$189 \leq \sqrt{s} \leq 209$ GeV at LEP-II. The main Higgs production at LEP is the Higgs-Strahlung process $e^+e^- \rightarrow Z^0 \rightarrow H^0Z^0$ [24]. Small additional contributions to the Higgs production add up via WW/ZZ fusion $e^+e^- \rightarrow \nu_e\bar{\nu}_eH/e^+e^-H$, while the direct production $e^+e^- \rightarrow H$ comes to be negligible because of the small coupling of the Higgs boson to electrons [24]. The Leading Order Feynman diagrams for the Higgs-Strahlung and WW/ZZ bosons fusion production are shown in Figure 1.4.
The accessible Higgs mass range continues to rise as the center of mass energy in the collisions increases at LEP-II. The Higgs-Strahlung process has a kinematic threshold at $M_H = \sqrt{s} - M_Z$, therefore this channel is sensitive to Higgs boson masses of about 115 GeV/c^2, with the final-state topology determined by the decay properties of the Higgs boson and by those of the associated Z boson. Combining the final results from all the experiments at LEP-II (ALEPH, L3, OPAL, DELPHI) [25], based on 2461 pb^{-1}

of data taken (1989 to 2000), the Higgs boson mass below 114.4 GeV/c^2 has been excluded at 95% Confidence Level (C.L.) [4].

Direct searches for the Higgs boson production are performed at $p\bar{p}$ Tevatron collider. During Run I (1992 to 1995), analyzing ~ 100 pb^{-1} of data collected at the center of mass energy of $\sqrt{s} = 1.8$ TeV, CDF and D0 experiments excluded a $M_H \sim 115$ GeV/c^2 Higgs boson at 95% C.L. [26].

1.3.3 Indirect experimental constraints

Each electroweak observable can be expressed as a perturbative series, through the so-called *radiative corrections*, in terms of a set of parameters of the SM directly measured in previous experiments. Predictions from radiative corrections can be either compared with direct observations of the electroweak observables, demonstrating the predictive power of quantum loop corrections and of electroweak unification in the SM theory, or used to set indirect experimental predictions for not yet discovered observables. Radiative corrections make the high-precision electroweak observables, especially radiative corrections of the W and Z properties, also sensitive to top quark M_t and Higgs mass M_H, showing contributions depending quadratically on M_t and logarithmically on M_H [5]. Since the Higgs particle contributes to the radiative corrections to the electroweak observables, the Higgs boson mass can be predicted via precision electroweak measurements [27].
The dependence of the SM observables on the Higgs mass is evaluated by the GFITTER Collaboration [28], using the latest experimental measurements of electroweak observables and the latest calculations of theoretical predictions of the SM [1]. A global fit to electroweak precision data is performed minimizing the test statistics χ^2 which accounts for the deviations between the measured value of the observables given as inputs to the fit and their SM prediction as a function of the M_H. Figure 1.5 shows the results for the M_H from the electroweak fit for separate measurements, with the M_t and M_W onces having the strongest impact on the fit. In Figure 1.6 is shown the $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ of the fit to electroweak precision measurements as a function of M_H, performed by the GFITTER Collaboration [6] and including all data and results from direct Higgs searches at LEP and Tevatron. The minimum χ^2 of the fit divided by the number of degrees of freedom is $\chi^2_{min}/\text{n.d.f} = 17.46/14 = 1.25$. The preferred Higgs mass from this fit corresponding to the minimum of the curve is:

$$M_H = 120.6^{+17.0}_{-5.2} \text{ GeV/c}^2 \tag{1.20}$$

and the 2σ interval $[114, 155]$ GeV/c^2 [27], where the error accounts for both experimental and theoretical uncertainties. The 95% one-side C.L. upper limit from the electroweak precision measurements becomes $M_H \lesssim 250$ GeV/c^2, thus cornering the Higgs mass in the range $114.4 \leq M_H \leq 250$ GeV/c^2. Therefore, indirect bounds from precision

[4] A benchmark for excluding a signal is set at 95% confidence level. This means that a 95% confidence level exclusion should be obtained no more than 5% of the time if a real signal is present

[5] The parameterized formula for the M_W, for example, is able to reproduce the full experimental result for M_W to better than 0.5 GeV/c^2 over the range 10 GeV/c^2 $< M_H <$ 1 TeV/c^2.

[6] Similar global fits used to be performed by the LEP Electroweak Working Group [29], using the ZFITTER [30] package. Results were found in excellent agreement.

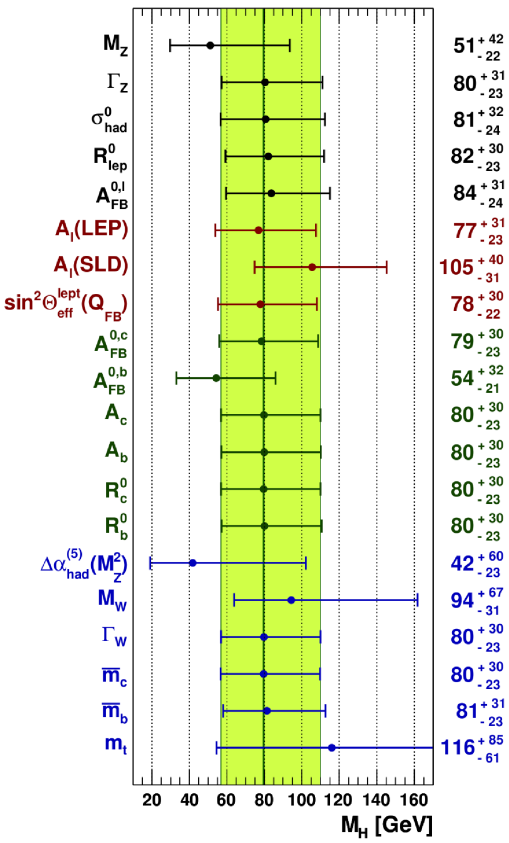


Figure 1.5: Results for M_H from the global fit for separate measurements. The fit is performed excluding the respective measured observable from the input set.

electroweak measurements show a strong preference for a low mass Higgs. This is not a proof that the SM Higgs boson exists, but it gives a guideline in what mass range to look for it.

1.4 Higgs production at hadron colliders

The four main Higgs production processes are :

$$\begin{aligned} \text{W/Z associated production :} &\quad q\bar{q} \to V + H \\ \text{vector boson fusion :} &\quad qq \to V^*V^* \to qq + H \\ \text{gluon-gluon fusion :} &\quad gg \to H \\ \text{heavy quarks associated production :} &\quad gg, q\bar{q} \to t\bar{t} + H \end{aligned}$$

The corresponding Leading Order Feynman diagrams are displayed in Figure 1.7. Their production cross section strongly depends on the center-of-mass energy of collisions and on the Higgs mass hypothesis. Therefore, the design of high-energy colliders is crucial in making the experiments sensitive to a large Higgs mass range and thus allowing as big cross section as possible. The $p\bar{p}$ collision energy at Tevatron RUN-II collider (2002

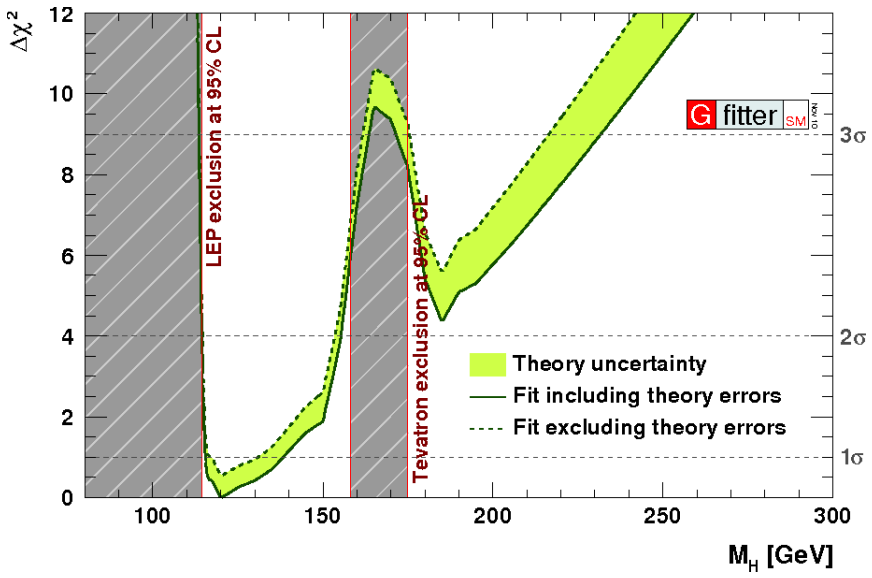


Figure 1.6: $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ of the fit to electroweak precision measurements performed by the GFITTER Collaboration, as a function of M_H. The solid line is the fit result, while the dashed curve is the result obtained ignoring the theoretical error due to missing higher order contribution. The vertical bands show the exclusion limit on the mass from direct searches at LEP-II (up to 114 GeV/c^2) and Tevatron (which will be discussed later in this Thesis).

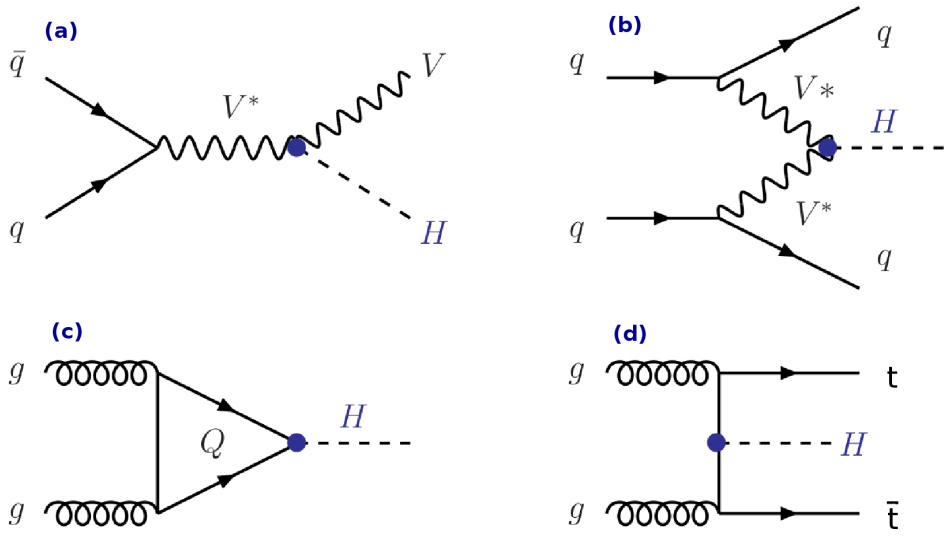


Figure 1.7: Leading Order Feynman diagrams of the dominant SM Higgs boson production mechanisms in hadron collisions: W/Z associated production (a), vector boson fusion (b), gluon-gluon fusion (c), heavy quarks associated production (d).

up to now) is of $\sqrt{s} = 1.96$ TeV in the center of mass while the LHC proton-proton collider is now operating at a world record center-of-mass collision energy of $\sqrt{s} = 7$ TeV. However, partons, as quarks and gluons in hadrons, are those really participating

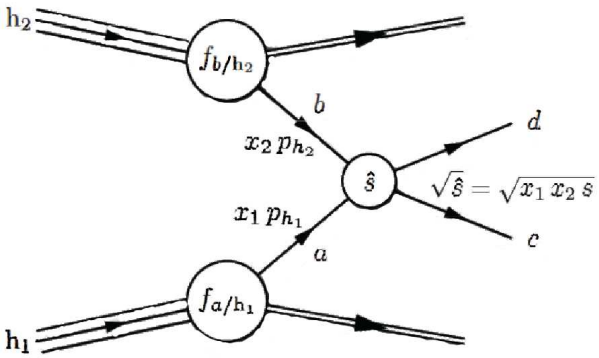


Figure 1.8: A schematic drawing of the hadrons collision. Partons really participate in the collision with a not fixed partonic center-of-mass energy $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$, function of the collision energy s and of the parton momentum fractions x_i.

in the hadron collision, as schematically shown in Figure 1.8. The center-of-mass energy of partons $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$ is not fixed and is a function of the collision energy s and of the parton momentum fractions x_i.
The study of the Higgs search sensitivity at hadron colliders relies on a correct evaluation of the Higgs production cross section [31]. Its calculation is based on the knowledge of the momentum fraction x distribution of partons, the Parton Distribution Functions (PDFs) $f_i(x, \hat{s})$, through the generic parton-model formula:

$$\sigma = \sum_{i,j} \int_0^1 dx_1 dx_2 f_i(x_1, \hat{s}) f_j(x_2, \hat{s}) \hat{\sigma_{ij}} \tag{1.21}$$

In Figure 1.9 the PDF luminosity curves $\frac{dL_{ij}}{d\hat{s}} \propto f_i(x_1, \hat{s}) f_j(x_2, \hat{s})$ at LHC are overlaid with the equivalent luminosity curves from the Tevatron [32]. At the $\sqrt{\hat{s}}$ values in order to produce an Higgs with a mass in the range 100 GeV/c^2 $\lesssim M_H \lesssim$ 1 TeV/c^2 allowed by the theory and by direct searches at LEP, the gg PDF luminosity is large with respect to other parton luminosities. Thus the $gg \to H$ gluon fusion is the Higgs production mechanism with the largest cross section at Tevatron and LHC [33, 34].

The gluon fusion cross section calculation is enhanced adding Higher-Order (HO) QCD corrections [35]. Usually the HO correction is expressed in the K-factor defined as the ratio of the HO cross section over the Leading-Order (LO) one. Considering only the Next-to-Leading-Order (NLO) correction the K-factor is:

$$K = \frac{\sigma_{NLO}}{\sigma_{LO}} \tag{1.22}$$

In Figure 1.10 the NLO corrections doubles the σ_{LO} cross section central value and the NNLO QCD correction gives an additional 40%. The updated central values for Higgs boson production cross sections are calculated folding the partonic cross section with the most recent MSTW2008 [36] set of PDFs of the (anti-)proton involved in the hadronic collision. The error bands in Figure 1.10 correspond to the systematic uncertainty on the cross section calculation, discussed in details in [35, 37].
The cross section calculated for the gluon-gluon channel, used in the Higgs searches at Tevatron are known up to Next-to-Next-to-Leading-Logarithm (NNLL) order [35],

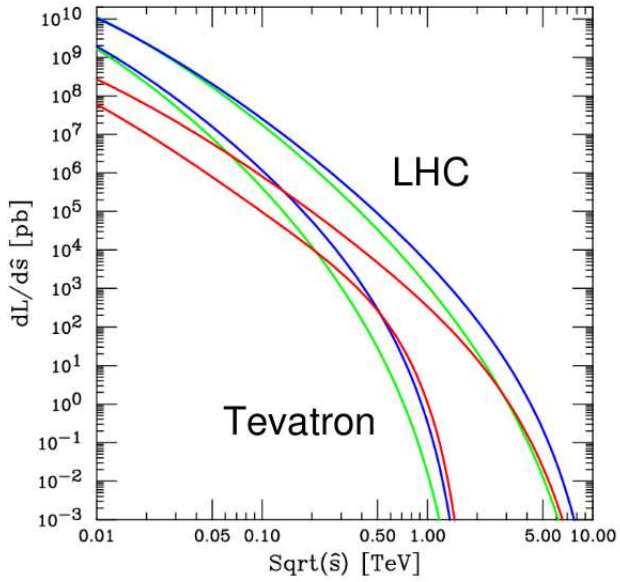


Figure 1.9: The parton-parton luminosity in picobarns. Green = gg, Blue = $\sum_i (gq_i + g\bar{q}_i + q_i g + \bar{q}_i g)$, Red = $\sum_i (q_i \bar{q}_i + \bar{q}_i q_i)$, where the sum runs over the five quark flavours $d,\ u,\ s,\ c,\ b$.

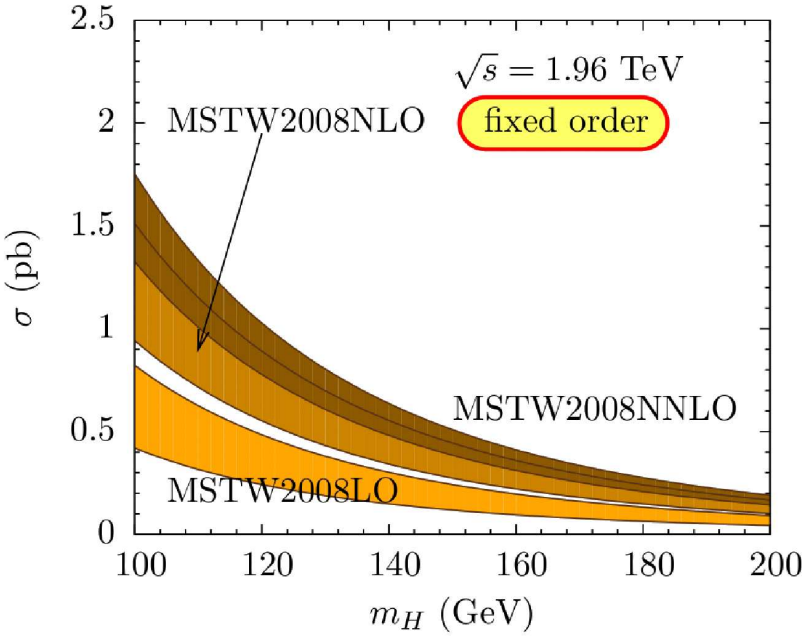


Figure 1.10: Cross section prediction $\sigma(gg \to H)$ at 1.96 TeV at different orders of QCD corrections. MSTW2008 [36] set of PDFs is used.

including the soft-gluon resummation contribution to the NNLO approximation. The NNLL calculation additionally increases the production rate by $\sim 10 - 15\%$.
Theoretical uncertainties on the Higgs production cross section are rather large for gluon-gluon channel. The production cross section estimated at NNLL is expected to suffer a $\sim$ 10% reduction due to the effect of the unknown Higher Order (HO) QCD con-

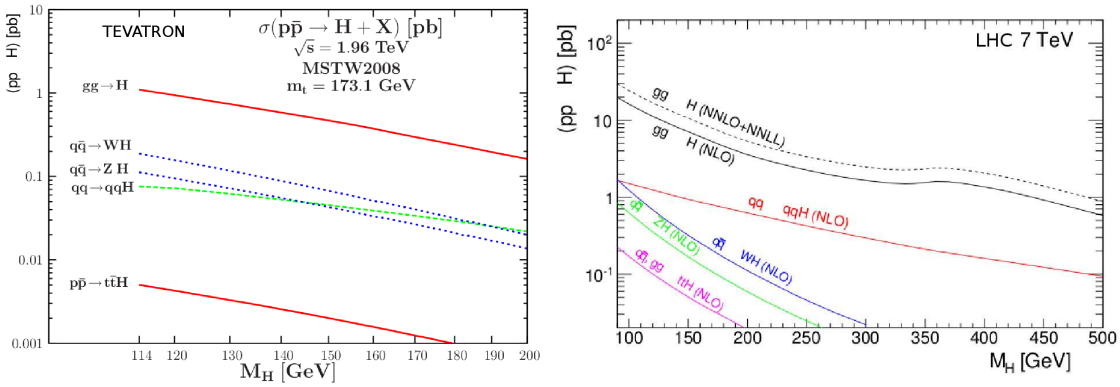


Figure 1.11: Higgs production cross sections at the $p\bar{p}$ collider Tevatron with center of mass energy of 1.96 TeV (left) and at the pp collider LHC with center of mass energy of 7 TeV (right). Cross sections for gluon-gluon fusion, associated production and vector boson fusion are approximated at the QCD Next-to-Next-to-Leading Order (NNLO). The $t\bar{t}H$ production cross section is at the NLO.

tributions [35]. Other contributions to the uncertainty come from the effective theory approach, assuming loop particle in the NNLO calculations to be very massive. It is a good approximation for top quark loop [38], however the same assumption for b quark loop leads to a $\sim 2-3\%$ uncertainty. The error due to imperfect parameterizations of the PDFs represents the second largest source of errors after the HO QCD correction. The total PDF uncertainty on the Higgs signal adopted by CDF/DØ analysis is of the order of $6-10\%$. An additional contribution is from the error on the α_s QCD coupling, obtained through the global fit to the available data. Those contributions to the uncertainty have been combined in [35] and [37], giving a total uncertainty of $\sim 10\%$ for the central value of the $gg \to H$ cross section, then assumed by the experiments at Tevatron. The associated production and vector boson fusion receive small contributions by HO corrections. The latest available calculations for associated production and vector boson fusion cross sections implement a full QCD Next-to-Next-to-Leading Order (NNLO) approximation [35, 37]. The $t\bar{t}H$ production cross section is only known at NLO [39, 40].

The cross sections of the various Higgs production mechanisms at the Tevatron and LHC are displayed in Figure 1.11 as a function of the Higgs mass. Over the entire mass range, the direct gluon-gluon is the dominant production mode at the two colliders.
At Tevatron the gluon fusion cross section is of the order of 0.1-1 pb in the range $100 < M_H < 200$ GeV/c^2. It contributes for the 78% to the inclusive Higgs production cross section and is 5 times larger than the associated production, which is the second highest contribution. Vector boson fusion has a behavior similar to the associated production and has a relative contribution of $\sim 7\%$.
At LHC the Higgs production cross sections are about two order of magnitude larger than at Tevatron. 88% of the inclusive cross section is the direct production, ranging between 75 pb and 11 pb for $100 < M_H < 300$ GeV/c^2. The associated production has smaller relative contribution to the total cross section, with respect to Tevatron, thus the second highest contribution at LHC is expected to come from the vector boson fusion, contributing $\sim 9\%$ to the Higgs production. The cross sections for WH, ZH and $t\bar{t}H$

m_H=160 Gev	Tevatron $\sqrt{s}$ = 1.96 GeV		LHC $\sqrt{s}$ = 7 TeV		LHC $\sqrt{s}$ = 14 TeV	
	σ[pb]	%	σ[pb]	%	σ[pb]	%
gg→H	0.439	78	8.49	87	32.6	88
WH	0.050	9	0.25	3	0.61	2
ZH	0.034	6	0.13	1	0.34	1
VBF	0.039	7	0.92	9	3.38	9
TOTAL	0.561	100	17.45	100	36.7	100
RATIO			17 X TEVATRON		66 X TEVATRON	

Figure 1.12: Higgs production cross sections at LHC and Tevatron hadron colliders, at 160 GeV/c^2 Higgs mass hypothesis. The gg production mode is dominant at the two colliders and its value and relative contribution to the inclusive production cross section is highlighted in the green box. At Tevatron the second highest contribution is the associated production, whose values are evidenced by the blue box.

productions are one to two orders of magnitude below the gluon and vector boson fusion cross sections.
The contribution of the four main production mechanisms at hadron colliders are summarized for the 160 GeV/c^2 Higgs mass hypothesis in Figure 1.12. The cross section values corresponding to the Tevatron energy in Figure 1.12 are taken from [35] and currently used by the CDF/DØ Collaborations.

1.5 Higgs decay modes and analysis strategies

To discuss the Higgs decays, it is useful to consider three distinct mass ranges:

- the *low mass range* 110 GeV/c^2 $< M_H <$ 135 GeV/c^2
- the *high mass range* 135 GeV/c^2 $< M_H <$ 200 GeV/c^2
- the *very high mass range* $M_H >$ 200 GeV/c^2

The branching rations of SM Higgs boson decays as a function of the Higgs mass are taken from [41] and shown in Figure 1.13.
This Thesis will focus on searches for the SM Higgs boson at CDF in the high mass range $135 < M_H < 200$ GeV/c^2. However, an overview of the SM Higgs analyses done at Tevatron RUN-II and of prospects for searches at LHC is presented in this Section, introducing the subject to which this Thesis will be devoted, the Higgs search in the channel $H \to WW \to l\nu\tau\nu$, with taus decaying hadronically.
The Higgs searches done at Tevatron discussed in the follow are based on 9 fb^{-1} of integrated luminosity collected by the experiments CDF and DØ, while LHC has delivered $\sim$ 50 pb^{-1} of integrated luminosity and prospects for searches based on at least 1 fb^{-1} will be discussed.

1.5.1 Low mass region: $M_H \leq 135$ GeV/c^2

For $M_H \lesssim 135$ GeV/c^2, the branching ratios of decays into two vector bosons is suppressed for the presence of at least one off-shell W or Z and the decay to a top pair is not

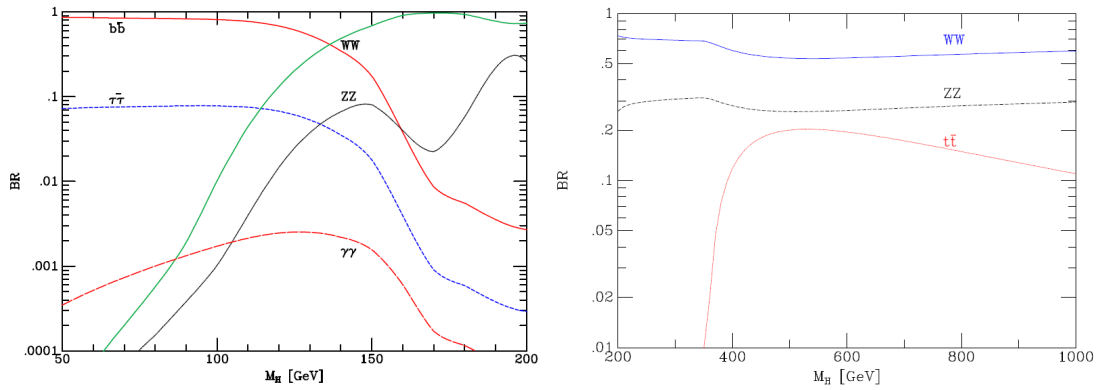


Figure 1.13: Branching ratios for the main decays of the SM Higgs boson for Higgs mass in the ranges $50 \leq M_H \leq 200$ GeV/c^2 (left) and $200 \leq M_H \leq 1000$ GeV/c^2 (right) separately. These results have been obtained with the program HDECAY [41].

kinematically allowed. Therefore, the $H \to b\bar{b}$ decay channel has the highest branching ratio, $\sim 75-50\%$ for $M_H = 115-130$ GeV/c^2. At Tevatron, as already seen in Section 1.4, the dominant production process for the Higgs boson is gluon fusion, with a cross section in the range 1.0-0.1 pb. However, the $gg \to H \to b\bar{b}$ channel is strongly overwhelmed by QCD production of non-resonant $b\bar{b}$ events, whose cross section is orders of magnitude larger than the Higgs production cross section. Associated production cross section is 0.2-0.5 pb in the low mass range, but the presence of an extra vector boson from the Higgs associated production helps to reduce backgrounds. Therefore, $H \to b\bar{b}$ searches at Tevatron exploit the associated production of the Higgs boson, with the W [42, 43] or Z [44, 45, 46, 47] boson decaying leptonically to increase the expected signal to background ratio. The three main final state, $l\nu b\bar{b}$, $\nu\bar{\nu}b\bar{b}$ and $l^+l^-b\bar{b}$ exhibit distinctive signatures: energetic leptons and/or undetected energy due to ν's. However, the $H \to b\bar{b}$ searches become not powerful for $M_H > 120$ GeV/c^2, due to the rapidly decreasing of the $H \to b\bar{b}$ branching ratio (see Figure 1.13).
Other Higgs signatures help improve the sensitivity of the Higgs search at the Tevatron. Higgs decays to leptons less often than to heavy quarks, given the lower mass. However the most favored leptonic decay channel is $H \to \tau^+\tau^-$ with a branching ratio of $\sim 7-5\%$ [48]. The Higgs can also decay to massless particles through loop-mediated processes, for example the $H \to \gamma\gamma$, which has a significant branching ratio for $M_H \lesssim 160$ GeV/c^2 (see Figure 1.13). However, the $H \to \gamma\gamma$ analysis suffers of low statistic and of a not good photon reconstruction at CDF or DØ detectors. Therefore the $H \to \gamma\gamma$ search channel [49, 50] is not competitive with $H \to b\bar{b}$ searches.

About prospects for low mass Higgs searches at LHC, the projections of the ATLAS and CMS Collaborations show that if the Higgs boson exists with a mass below 130 GeV/c^2, its discovery would be challenging with 1 fb^{-1} of integrated luminosity delivered [51, 52]. The preferred channel at the Tevatron, $q\bar{q} \to WH \to l\nu b\bar{b}$, is not a discovery mode at LHC due in part to the larger background cross section at $\sqrt{s} = 7$ TeV. A very high integrated luminosity of 300 fb^{-1} is necessary in order to have a reasonable sensitivity to this signal (with $S/\sqrt{B} \geq 5$), with the 5σ discovery limited to a Higgs mass of less than 123 GeV/c^2 [53].

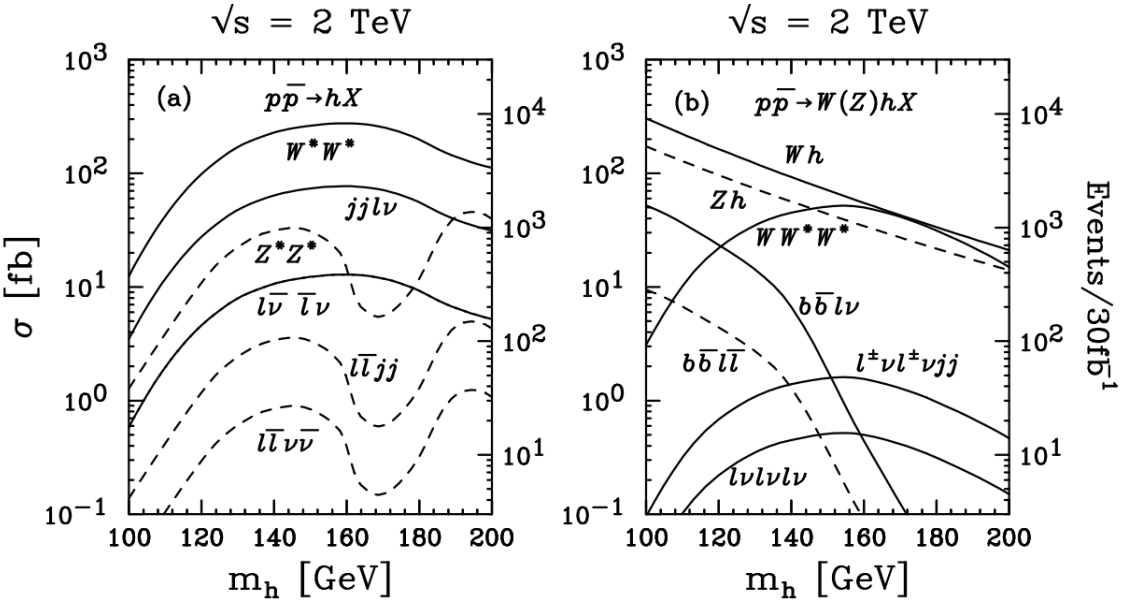


Figure 1.14: Standard Model Higgs boson production cross sections (in fb) and various subsequent decay modes versus M_H for (left) $gg \to H \to W^*W^*$ (solid curves) and Z^*Z^* (dashed curves). (right) $q\bar{q}' \to WH$ production with $H \to W^*W^*$ (solid curves) and ZH production (dashed curves). Also shown the $H \to b\bar{b}$ with leptonic W, Z decay channels.

However, at LHC in the $110 - 140$ GeV/c^2 mass range, the $H \to \gamma\gamma$ channel is expected to be one of the main channels for low mass Higgs searches at high luminosity, due to the small branching ratio. ATLAS and CMS preliminary results from simulations show an expected upper limit $\sim 5\times$ SM and $\sim 4\times$ SM cross section respectively, on 1 fb^{-1} of data using the $H \to \gamma\gamma$ decay channel. Predictions are significantly better than the current limits from each Tevatron experiments [49, 50, 54] in this channel ($\sim 20\times$ SM cross section). A significantly greater sensitivity can be achieved when considering more advanced analysis features and higher luminosity.

1.5.2 High mass region: $130 < M_H < 200$ GeV/c^2

The branching ratios for Higgs decays into vector bosons start to dominate for $M_H > 130$ GeV/c^2, in particular for $2M_W < M_H < 2M_Z$, where the W bosons are real, even if the ZZ decay channel is still partially suppressed. In the high mass region the Higgs decay into two massive gauge bosons W or Z represents the most favorite decay channel. $H \to WW$ and $H \to ZZ$ searches at Tevatron strongly depend on the decay mode of the bosons, since the background contribution can be very different. Figure 1.14(left) shows the cross section for $gg \to H \to WW^*/ZZ^*$ versus M_H at Tevatron. The WW^* and ZZ^* decay channels are separately shown usign solid and dashed curves:

$$H \to W^*W^* \to l\nu jj \quad \text{and} \quad l\bar{\nu}\bar{l}\nu \tag{1.23}$$

$$Z^*Z^* \to l\bar{l}jj \quad \text{and} \quad l\bar{l}\nu\bar{\nu} \tag{1.24}$$

where $l = e, \mu$ and j is a quark jet. Although the $l\nu jj$ mode has the larger production rate, the $WW^*/ZZ^* \to l\bar{\nu}\bar{l}\nu/l\bar{l}\nu\bar{\nu}$ mode is cleaner in terms of the SM background

contamination. Figure 1.14(right) shows the $H \to W^*W^*$ mode from the WH or ZH associated production. The two search channels with highest sensitivity are the dilepton plus jets ($WW^*W^* \to l^{\pm}\nu l^{\pm}\nu jj$) and trilepton ($WW^*W^* \to l^{\pm}\nu l'^{\pm}\nu l^{\mp}\nu$) events. The former signal rate is higher than the latter by about a factor of three due to the difference between the branching fractions of W decay into $l = e, \mu$ and into jets, while the signal rates for $WH \to b\bar{b}l\nu$ and $ZH \to b\bar{b}l\bar{l}$ channels via $H \to b\bar{b}$ decay (the main search channels in the *low mass region*, as discussed in the previous section) drop dramatically for higher M_H values. The trilepton channel has been explored at Tevatron, searching for events with three leptons, electrons or muons, plus missing energy [55, 56], while the $l\bar{l}jj$ search channel is still not explored at the Tevatron.
Requiring the $H \to WW$ leptonic decay, the final state involving dileptons from the vector boson fusion decay mode is given by:

$$qq \to qqH \to qqWW \to qql\nu l\nu \tag{1.25}$$

In particular the vector boson fusion has the distinctive signature of the two outgoing quarks which tend to have a small transverse momentum component and are very boosted in the direction of the initially colliding quarks, then have large rapidity.
The full hadronic final state is not explored at Tevatron due to the overwhelmed QCD background, although the high branching fraction of W decay in hadrons is $\sim 68\%$. .
The leptonic decay of the W bosons provides the best signature for the decay $H \to W^+W^- \to l^+\nu l^-\bar{\nu}$, which is the main Higgs discovery channel, for $2M_W < M_H < 2M_Z$, with the best sensitivity against background. Search analyses at CDF and DØ look for both Ws decaying leptonically, selecting events with two electrons or two muons or one electron and one muon for a total branching fraction of the WW pair of $\sim 6\%$, including the leptonic decay of the τ [55, 57].

At LHC the search channel $H \to W^+W^-$ is expected to bring the greatest sensitivity in the $130 < M_H < 200$ GeV/c^2 mass range, for low integrated luminosity. Dominant backgrounds for the $gg \to H \to WW/ZZ$ signal are the non-resonant WW and ZZ originating from the $q\bar{q}$ initial state. To compare the Tevatron and LHC sensitivity to this channel, it can be useful to study the expected $\sigma_{LHC}/\sigma_{TEV}$ ratio for signal and background processes, which is proportional to the ratio of the PDF luminosities $\frac{dL_{ij}}{d\hat{s}}$, defined in Section 1.4, at the LHC to the PDF at Tevatron, where i,j denote the initial partons. Figure 1.15 shows the parton-parton luminosity ratio at 7 TeV LHC and Tevatron for gg or $q\bar{q}q$ initial partons, as a function of $\sqrt{\hat{s}}$ [32]. The increase in PDF luminosity at the LHC with respect to Tevatron comes from gg initial states, followed by gq and then $q\bar{q}$ initial states. Searching a Higgs boson in the $135 \leq M_H \leq 200$ GeV/c^2 mass range, the Higgs signal production through gg annihilation proceeds with a factor of about 10 anhancement at the LHC compared to the Tevatron, while there is a commensurate increase in the rate for non-resonant background processes at LHC. Therefore, the $gg \to WW$ signal is favored at LHC with respect to Tevatron by the higher ratio of its cross section to the non-resonant $q\bar{q} \to WW$ production, $\sigma(gg \to WW)/\sigma(WW)$, with the best signature for $gg \to WW$ signal provided by the pure leptonic decay of the WW pair.
CMS and ATLAS Collaborations report sensitivity studies looking for the direct production of the Higgs boson decaying to a W pair. Projections at 1 fb^{-1} of integrated luminosity show an expected exclusion mass range between $145 < M_H < 183$ GeV/c^2

and $150 < M_H < 185$ GeV/c^2 at ATLAS [58] and CMS [51, 59] respectively. The 95% C.L. upper limit equal to the SM cross section for $M_H = 160$ GeV/c^2 is claimed by the experiments at LHC with just 250 pb^{-1}, then reaching the actual CDF and DØ sensitivity.

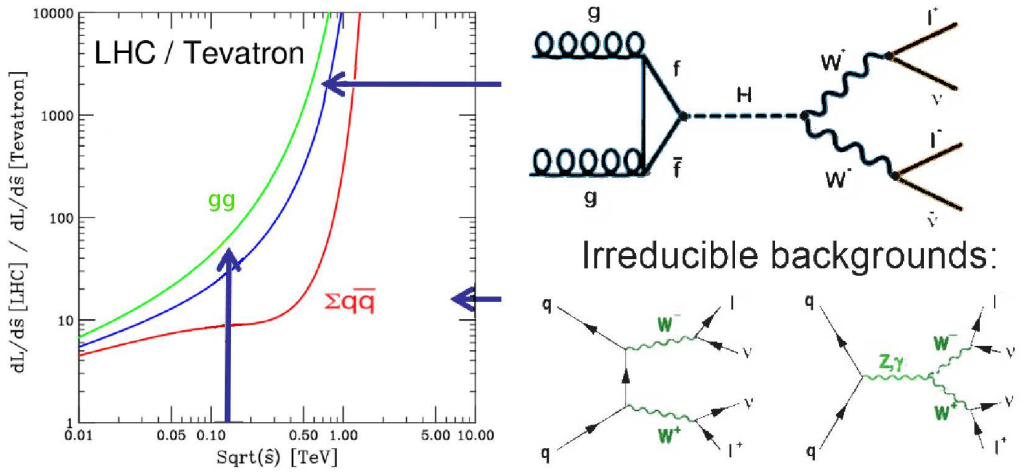


Figure 1.15: The ratio of parton-parton luminosity in picobarns at 7 TeV LHC and Tevatron [32]. Green = gg, Blue = $\sum_i (gq_i + g\bar{q}_i + q_i g + \bar{q}_i g)$, Red = $\sum_i (q_i\bar{q}_i + \bar{q}_i q_i)$, where the sum runs over the five quark flavours d, u, s, c, b. The gg Higgs production proceeds with a factor of about 10 enhancement at the LHC compared to the Tevatron, for $135 \leq M_H \leq 200$ GeV/c^2, while non-resonant backgrounds, originating from the $q\bar{q}$ initial state, have only a commensurate increase.

1.5.3 Very high mass region: $M_H > 200$ GeV/c^2

For $M_H > 2M_Z$, the $H \to ZZ$ is expected to be competitive with the WW decay channel, and for very high Higgs masses, $M_H > 2M_t$, also the decay into a top pair becomes important. There are not published or preliminary results at Tevatron, while LHC experiments are expected to be sensitive to the channels at larger Higgs mass values.
In this region the predominant decay channels are $H \to WW$ and $H \to ZZ$ with both vector boson *on-shell*. The Higgs boson decay to ZZ presents a range of experimental signatures depending upon the decays of the two Z bosons: $l^+l^-l'^+l'^-$, $l^+l^-\nu\bar{\nu}$ and $l^+l^-b\bar{b}$ signatures.
The $H \to ZZ^* \to l^+l^-l'^+l'^-$ decay channel has a smaller ZZ irreducible background than in the high mass region ($135 \leq M_H \leq 200$ GeV/c^2) and the leptonic signature offers excellent suppression of the QCD background. The inclusive Higgs signal cross section is larger than at the Tevatron and the ratio $\sigma(H \to ZZ)/\sigma(ZZ)$ is more favorable. Therefore, this channel is known as the golden Higgs decay mode in the very high mass region at the LHC collider, although it has a low branching fraction of $\sim 0.6\%$. With an integrated luminosity of 1fb^{-1} at 7 TeV the SM Higgs boson cannot be excluded for any mass value in the $H \to ZZ \to 4l$ channel alone. The analysis is most sensitive for SM Higgs mass of about 200 GeV/c^2 where an upper bound in cross section times the branching ratio of about $2.5 \times (\sigma_H \times BR_{H\to 4l})$ is expected at both ATLAS [52] and

CMS [51] experiments.
$H \to ZZ^* \to l^+l^-\nu\bar{\nu}$ and $H \to ZZ^* \to l^+l^-b\bar{b}$ decay modes are used to supplement $H \to ZZ \to 4l$ for very large masses, $M_H > 600$ GeV/c^2. The $H \to ZZ^* \to l^+l^-\nu\bar{\nu}$ decay channel has a branching ratio of 20%, thus the cross section times the branching ratio is greater than the $H \to ZZ^* \to l^+l^-l'^+l'^-$ by a factor of approximately 6, however the main difficulty is the unobserved Z decaying to neutrinos and the main background contributions are WW and WZ. The $H \to ZZ^* \to l^+l^-b\bar{b}$ channel has the highest branching ratio, since the BB($Z \to hadrons$) $\sim 70\%$, however it is largely contaminated by the $Z + jet$ background.

The combination of all results predicted for 1 fb^{-1} of data at 7 TeV is shown in Figure 1.16 for ATLAS and in Figure 1.17 for CMS. The $H \to \gamma\gamma$, $H \to WW$ and $H \to ZZ$ channels are important to achieve the best coverage for the full range of possible SM Higgs boson masses. When combined together, the expected exclusion limit for the SM Higgs at 7 TeV and with 1 fb^{-1} becomes $139 < M_H < 180$ GeV/c^2 and $145 < M_H < 190$ GeV/c^2, for ATLAS and CMS respectively.

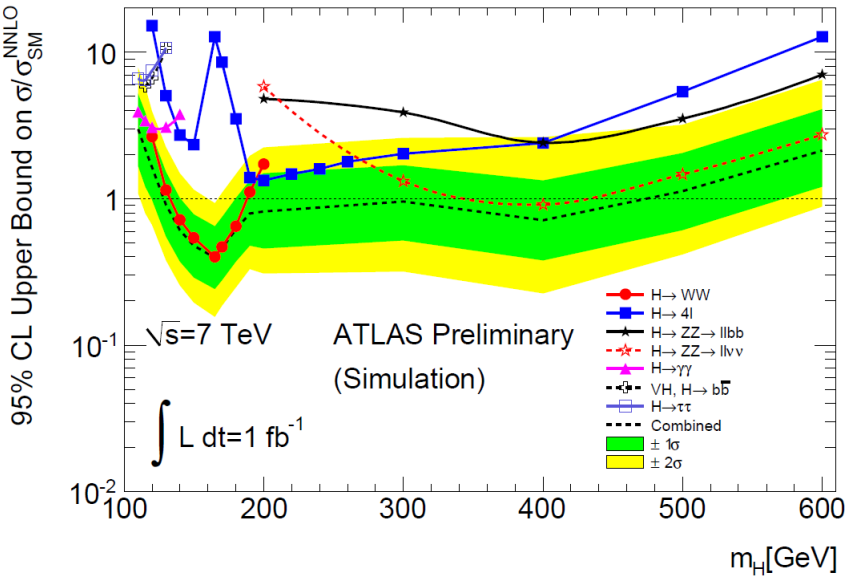


Figure 1.16: The projected exclusion limits for the combination of all Higgs search channels at ATLAS, using 1 fb^{-1} of data at 7 TeV. The green and yellow bands indicate the $\pm$ 1σ, 2σ intervals of the uncertainty on the expected limit [52].

1.6 $H \to WW \to l\nu\tau\nu$ search channel

The $H \to WW \to l\nu\tau\nu$ search channel in the *high mass range* $135 \leq M_H \leq 200$ GeV/c^2 is currently performed at CDF using 8.2 fb^{-1} of integrated luminosity [60]. The events can be classified by the flavor of the lepton pair and analyzed separately: one identified electron and one tau (e-τ) or one identified muon and one tau (μ-τ), with τ decaying into hadrons and missing transverse energy. The final states $(e - \tau)$ or $(\mu - \tau)$ from the $H \to WW$ decay process, with τ decaying into hadrons, correspond to an additional 3% of the

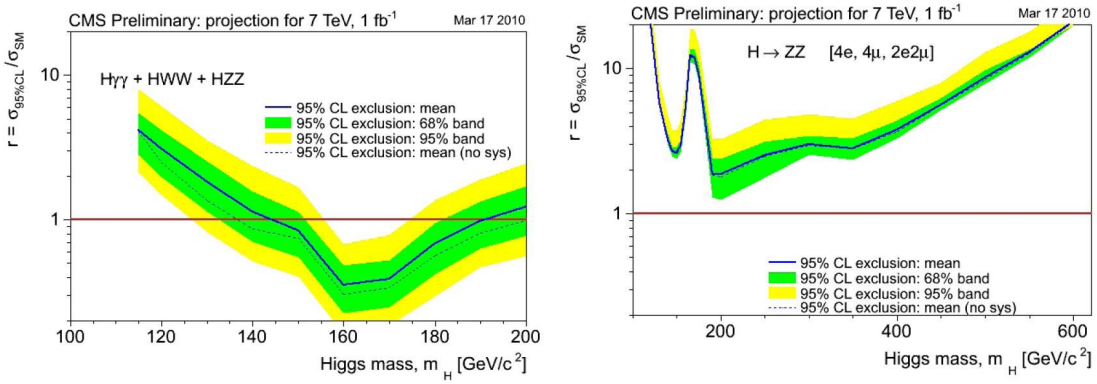


Figure 1.17: The projected exclusion limits for the combination of the $H \to WW$, $H \to ZZ$ and $H \to \gamma\gamma$ channels is presented (on the left). For Higgs mass above 200 GeV/c^2 the only contribution is from $H \to ZZ$ (on the right).

possible two-leptons final states from WW leptonic decay. However, τ hadronic decay reconstruction and the presence of an additional neutrino in the final state, make this analysis challenging. Events reconstruction involves leptons and quarks identification in the hard scattering final state, in addition to gluons and photons that can be radiated from either the incoming quarks (the Initial State Radiation, ISR) or from particles in the final state (Final State Radiation, FSR).
In this Thesis the Higgs search in the $H \to WW$ decay with $e-\tau$ and $\mu-\tau$ final states is performed, exploiting a different analysis technique with respect to [60]. The challenging τ identification will be deeply discussed in this work and one of the key parts of this analysis is the comprehension of the tau properties and of the CDF detector response in reconstructing the tau hadronic decay. Jets can be easily mis-reconstructed as tau and the ability to identify hadronic tau in the event reflects on the typology of physics processes entering the final selected sample of $(e-\tau)$ and $(\mu-\tau)$ events. This analysis has the major contribution of background coming from the processes with jets mis-identified as a τ, referred to in this Thesis as *fake background*, whose estimation represents another challenging aspect beside the tau reconstruction.

Chapter 2

Experimental apparatus

2.1 The Fermilab accelerator complex

The Tevatron is a proton-antiproton synchrotron collider located at Fermi National Accelerator Laboratory. First $p\bar{p}$ collisions were taken in 1985 at the center of mass energy of 1.8 TeV and the collider operation period can be divided into Run I, between 1988-1996, and Run II from 2000. Tevatron collider has been upgraded in the years 1996-2000 in order to increase the luminosity by about ten times and center of mass energy by about 10%. The on-going (2011) collisions are produced at center of mass energy of $\sqrt{s} = 1.96$ TeV.
The colliding proton and antiproton beams are the result of a complex acceleration apparatus which involve different stages. Once accelerated at the designed energy, they are transfered to the interaction points on an underground ring, where the CDF and DØ particle detectors are installed (see Figure 2.1).

Proton source

The acceleration process starts from the generation of protons used for collision, originally extracted from a very pure hydrogen gas, which is moved between two electrodes and ionized into electrons and H^+ ions. After the positive ions strike a cathode made of cesium, a gas of H^- is produced and sent to a Cockroft-Walton electrostatic preaccelerator.Here the H^- gas is accelerated to an energy of 750 KeV and steered to the linear accelerator, the Linac. The Linac [61, 62] accelerates the H^- ions to 400 MeV using radio frequency (RF) cavities. The oscillating electric field of the RF cavities groups the ions in bunches. The 400 MeV ions are then injected into the Booster [62, 63]. At the injection the electrons are stripped off the H^- ions, by passing the H^- through a carbon foil. Then, the resulting bare protons are collected into the Booster and accelerated to 8 GeV. The Booster is the first synchrotron in the Tevatron chain and consists of a series of magnets arranged around a 75-meter radius circle with 18 RF cavities interspersed.
The proton beam is extracted from the Booster and sent to the Main Injector. It will contain a maximum of 5×10^{12} protons divided among 84 bunches spaced by 18.9 ns, each consisting of 6×10^{10} protons.

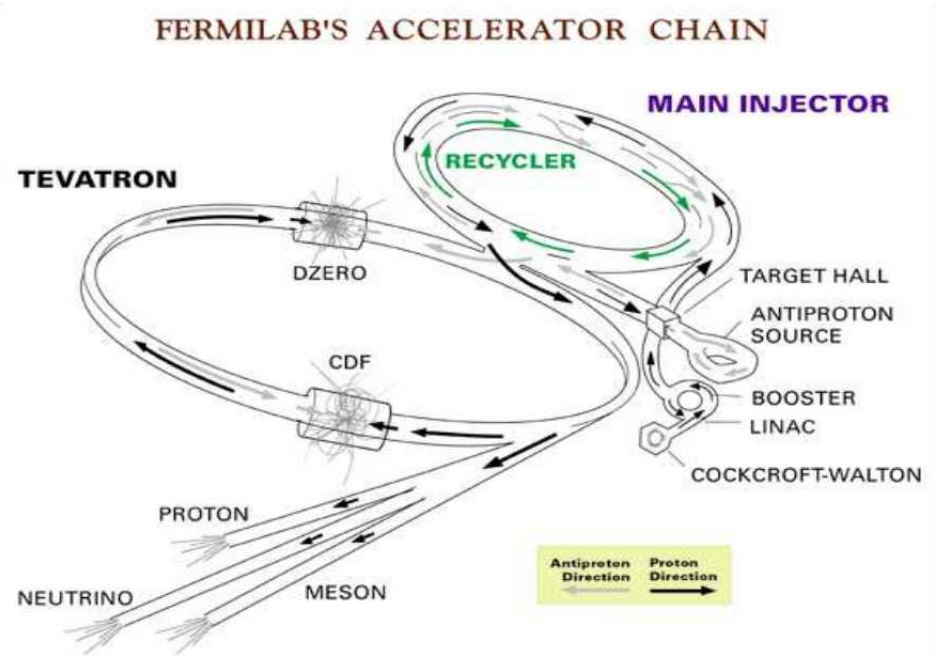


Figure 2.1: Layout of the Fermilab accelerator complex.

Main Injector

The Main Injector [62, 64] is a circular synchrotron seven times the circumference of the Booster. Depending on the Main Injector operation mode, it accelerates protons of 8 GeV energy up to 120 GeV for antiproton production only, while it is able to generate 150 GeV protons to be injected into the Tevatron if *Collider Mode* is set. Collider Mode is the most complex scenario that the Main Injector has to cope with: in addition to supplying 120 GeV protons for antiproton production, the Main Injector must also feed the Tevatron protons and antiprotons at 150 GeV. The protons and antiprotons need to be filled into super-bunches more intense than any individual bunch that can be accelerated by the Booster.

Debuncher and Recycler: Antiproton Production and Storage

For antiproton production a single pulse of 5×10^{12} protons at 120 GeV is extracted every $\sim$ 2.2 seconds from the Main Injector and directed to the nickel target. In the collision with the target, about 1 antiproton are produced for each 10^5 protons, with a mean energy of 8 GeV[62, 65]. The resulting particles spray contains some antiprotons with a broad momentum and wide-spread spatial distribution. The antiproton produced are collected by a lithium lens, separated from other particle species by a pulse magnet, focused by strong magnetic lenses and injected in the Debuncher Storage Ring. In the Debuncher ring, a rounded triangular-shaped synchrotron with mean radius of 90 meters, the batches of antiprotons are stochastically cooled [66] in order to reduce the spread of kinetic energy spectrum of the beam. The antiproton bunches are transferred to the Antiproton Accumulator, a 75 m mean radius storage ring of larger acceptance housed in the same tunnel as the Debuncher (see a sketch in Figure 2.2). In the Accumulator multiple beam pulses are stacked and $\bar{p}$ are further cooled to increase the antiproton

phase space density. It takes between 10 and 20 hours to build up a "stack"of antiprotons which is then used in collisions in the Tevatron.
This 8 GeV beam of stacked antiprotons (36 bunches of about 3×10^{10} antiprotons per bunch) is injected backwards to the Main Injector, accelerated to 150 GeV together with 36 bunches of roughly 3×10^{11} protons, and injected into the Tevatron.
When Accumulator reaches its maximum optimal capacity (10^{12} accumulated antiprotons), its antiprotons are passed into the Recycler Ring. The Recycler Ring [67] is a further improvement of the antiproton source, a post-accumulator storage ring of constant 8 GeV energy, located in the Main-Injector.
Figure 2.3 shows the antiproton production complex with the addition of the functionalities provided by the Recycler ring.

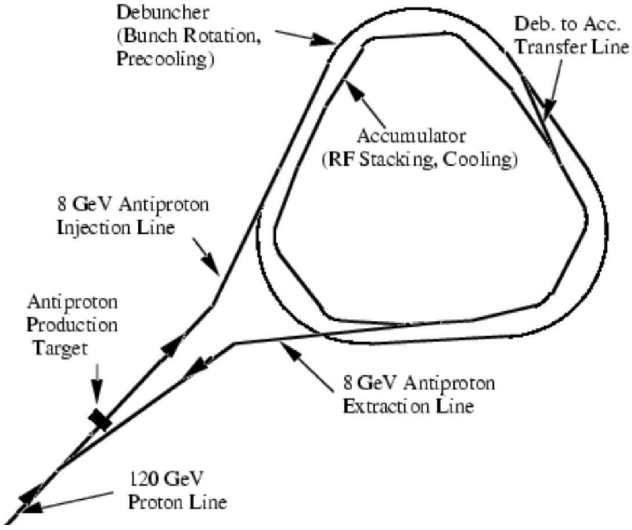


Figure 2.2: General layout of the antiproton source at Tevatron.

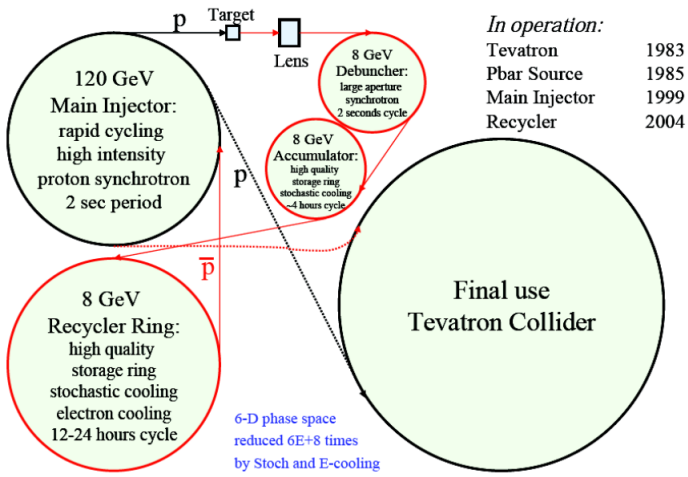


Figure 2.3: Antiproton production complex.

Tevatron

The Tevatron is a circular synchrotron with a 1 km radius. It is the last stage of the Fermilab's accelerator chain, able to accelerate the incoming 150 GeV proton and antiproton beams from Main Injector up to 980 GeV, providing a center of mass energy of 1.96 TeV . The antiprotons are injected after the protons have already been loaded. Protons and antiprotons circulate in opposite direction in the same vacuum pipe, since electrostatic separators reduce to a negligible amount the unwanted interactions, by keeping the beams on two non-intersecting helical closed orbits, except at the collision points. The beams are brought to collision at two collision points, BØ and DØ. The two collider detectors, the Collider Detector at Fermialb (CDF) and DØ, are built around the respective collision points.
The accelerator is composed of eight accelerating cavities, quadrupole and dipole focusing magnets. The accelerating cavities are made of superconducting magnets, requiring cryogenic cooling and consequently a large scale production and distribution of liquid helium. High-power focusing quadrupole magnets minimize the beam section at the interaction regions to maximize the collision rate. A tour of the Tevatron takes about 21 μs. About one minute is needed to reach the final beam energy.
Tevatron bunches are organized in three trains (see Figure 2.4). Within a train the inter-bunch time is 396 ns while inter-train time is 2.6 μs. The intra-train empty sectors allow enough time for fast kicker magnets to abort the beam into a dump before the arrival of the next train, in case of emergency. The Tevatron can then sustain both beams for hours at a time (called a *store*).

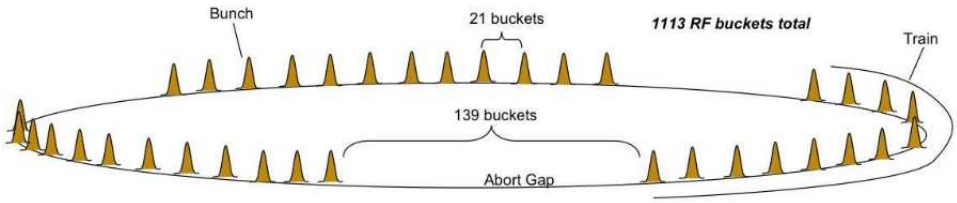


Figure 2.4: Bunch structure of the Tevatron beams in Run II.

The number of collisions per second is described by the *luminosity* L. The Tevatron collider operates, for the data taking period considered in this Thesis, with an instantaneous luminosity at the beginning of a store in the range between 0.1×10^{32} cm^{-2} sec^{-1} and 4×10^{32} cm^{-2} sec^{-1} (see Figure 2.5). The instantaneous luminosity in the Tevatron is approximatively given by the expression:

$$L = \frac{fBN_pN_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)} F(\sigma_l/\beta^*) \tag{2.1}$$

where f is the revolution frequency, B is the number of bunches in each beam, $N_p(N_{\bar{p}})$ is the number of protons (anti-protons) in a bunch. $\sigma_p(\sigma_{\bar{p}})$ is the effective width of the proton (anti-proton) beam size at the interaction point and F is a form factor that depends on the ratio of the bunch length, σ_l, to the beta function at the interaction point, β^*, which is a measure of the transverse beam width. Table 2.1 shows the RunII accelerator parameters [62].

Parameter	Run II
Number of bunches (N_B)	36
Protons/bunch (N_p)	2.7×10^{11}
Antiprotons/bunch ($N_{\bar{p}}$)	3.0×10^{10}
Total antiprotons	1.1×10^{12}
β^*[cm]	35
Bunch length [m]	0.37
Bunch spacing [ns]	396
Interactions/crossing	2.3
Energy [GeV/particle]	980

Table 2.1: Accelerator parameters for Run II configuration.

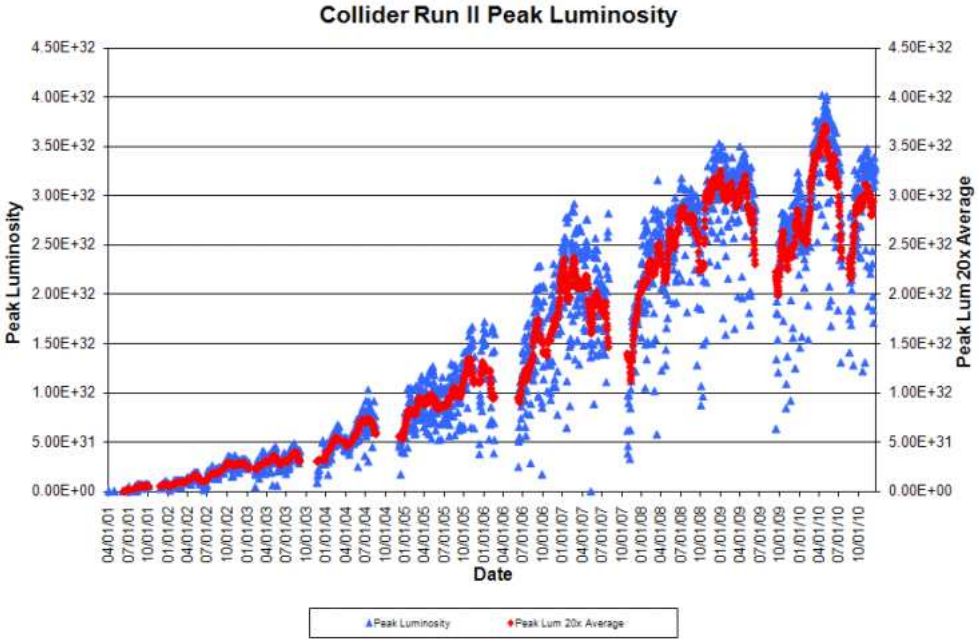


Figure 2.5: Tevatron initial instantaneous luminosity as a function of the time.

The luminosity in the Tevatron is proportional to the total antiproton intensity. The total number of antiprotons in the collider is determined by the product of the antiproton accumulation rate (25×10^9), the typical store duration (15-20 hours), and the transmission efficiency from the Antiproton Accumulator to storage (70%) in the Tevatron. Therefore, antiproton availability, their collection, cooling and stacking are the most limiting factors for the final Tevatron luminosity. The Recycler Ring provides a factor of ~ 2 in luminosity beyond that projected with the Main Injector alone, as well as providing a platform from which an additional increases in luminosity could be achieved. In particular, the Recycler was designed to stochastically cool enough antiprotons to obtain a luminosity of 2×10^{32}. Figure 2.5 shows the Tevatron peak luminosity as a function of the time. The blue squares show the peak luminosity at the beginning of each store. The red triangle displays a point representing the last 20 peak values averaged together. However, as the luminosity increases the number of interactions increases and adversely affects the particle detector performances and the ability of the experimenters

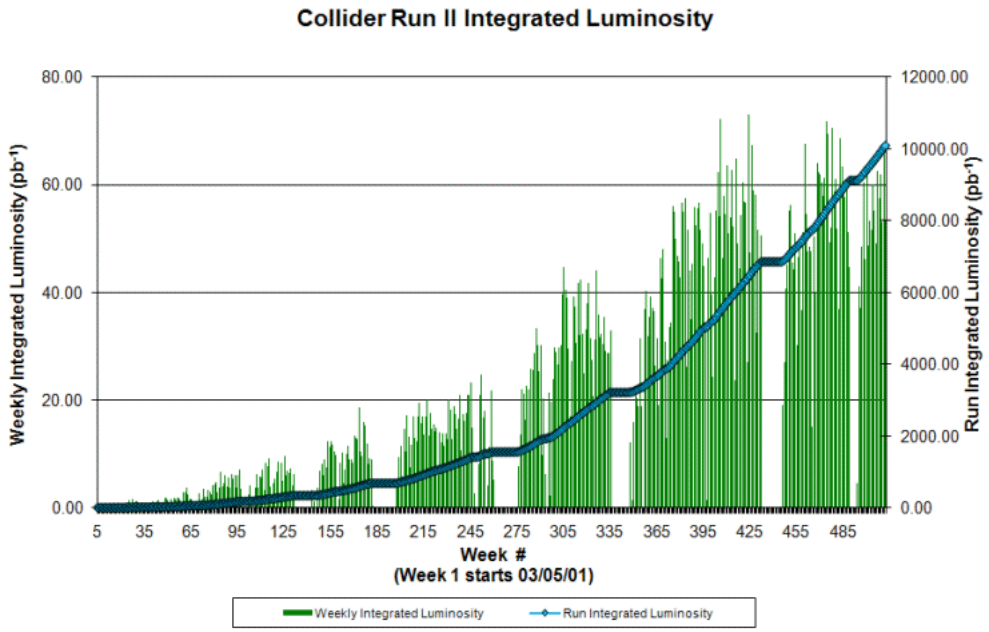


Figure 2.6: Weekly and total integrated luminosity delivered by the Tevatron collider to the CDF and DØ experiments.

to interpret the data. One way to reduce the number of interactions per crossing is to increase the number of bunches. Thus the Detector issues are driven by the luminosity, the number of bunches (36) and the time between crossings (396 ns).

Continuous improvements in the accelerator complex led to the rapid increase of the initial instantaneous luminosity and, more important, to the increase of the rate of integrated luminosity delivered to the experiments. The integrated luminosity as a function of time, by week and total to date, delivered by the Tevatron to the experiments are shown in Figure 2.6.

An average efficiency of $\sim$ 80% [68] and $\sim$ 92% [69] is reached by CDF and DØ experiments respectively in collecting the delivered luminosity. Inefficiency arises from beam conditions, from small detector problems or operational decisions to dedicate parts of a store.

2.2 The CDF Detector

The Collider Detector at Fermilab (CDF) is a general purpose solenoidal detector for the study of the Tevatron $p\bar{p}$ collisions. As illustrated in Figure 2.7, CDF uses a cartesian coordinate-system centered in the nominal point of interaction, with the z-axis coincident with the beamline and oriented parallel to the motion of the proton beam. The x-axis is in the horizontal plane of the accelerator ring, pointing radially outward, while the y-axis points vertically up.

For the symmetry of the detector, it is often convenient to work with cylindrical (z, r, ϕ) or polar (r , θ, ϕ) coordinate-system (see Figure 2.7). Polar θ and azimuthal ϕ angles relative to the z-axis are used to describe particles' trajectories: θ is measured from the

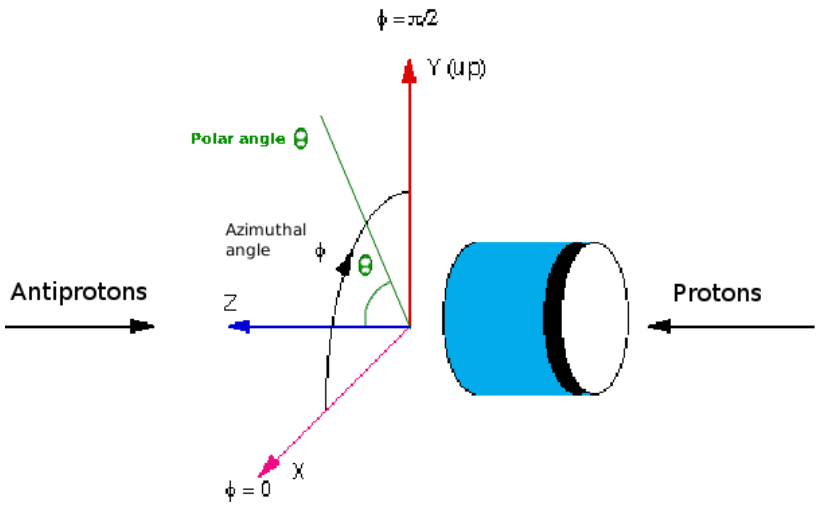


Figure 2.7: A schematic drawing of the CDF detector symmetry and of the polar (r, θ, ϕ) coordinate-system.

proton direction, ϕ is measured in the $x-y$ plane starting from the x axis, and it is defined positive in the anti-clockwise direction; the coordinate r defines the transverse distance from the z-axis. Another important coordinate that can be used instead of the polar angle θ, is called *pseudo-rapidity* [1]:

$$\eta = -ln \tan \frac{\theta}{2} \tag{2.2}$$

The pseudo-rapidity is usually preferred to θ at colliders, where events are boosted along the beamline, since it is invariant under Lorentz boosts. The direction of particles is usually measured in (η, ϕ). The quantity ΔR is often used to measure distances in $\eta - \phi$ space; it is defined as:

$$\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)} \tag{2.3}$$

The detector can be subdivided into two geometrical regions: the central part, covering regions with pseudorapidity $|\eta| \leq 1.0$, and the two plugs, covering the forward and backward regions with $1.0 \leq |\eta| \leq 3.6$.

CDF II combines precision charged particles tracking with fast projective calorimetry and fine grained muons detection. Figure 2.8 show the detector view. Tracking systems are contained in a superconducting solenoid, 1.5 m in radius and 4.8 m in length, which generates a uniform magnetic field $B = 1.4$ T parallel to the beam axis. Calorimetry and muon system are all outside the solenoid. The main feature of the detector systems are described below.
A detailed description of the upgraded detector can be found in [70].

[1] In CDF literature one distinguishes η_{det}, which is relative to the geometrical center of detector, and η, which is measured with respect to the interaction point z_0 where particles originated. Usually the former symbol is used for describing the detector geometry and the latter for outgoing particles. For simplicity the same symbol will be used in both cases.

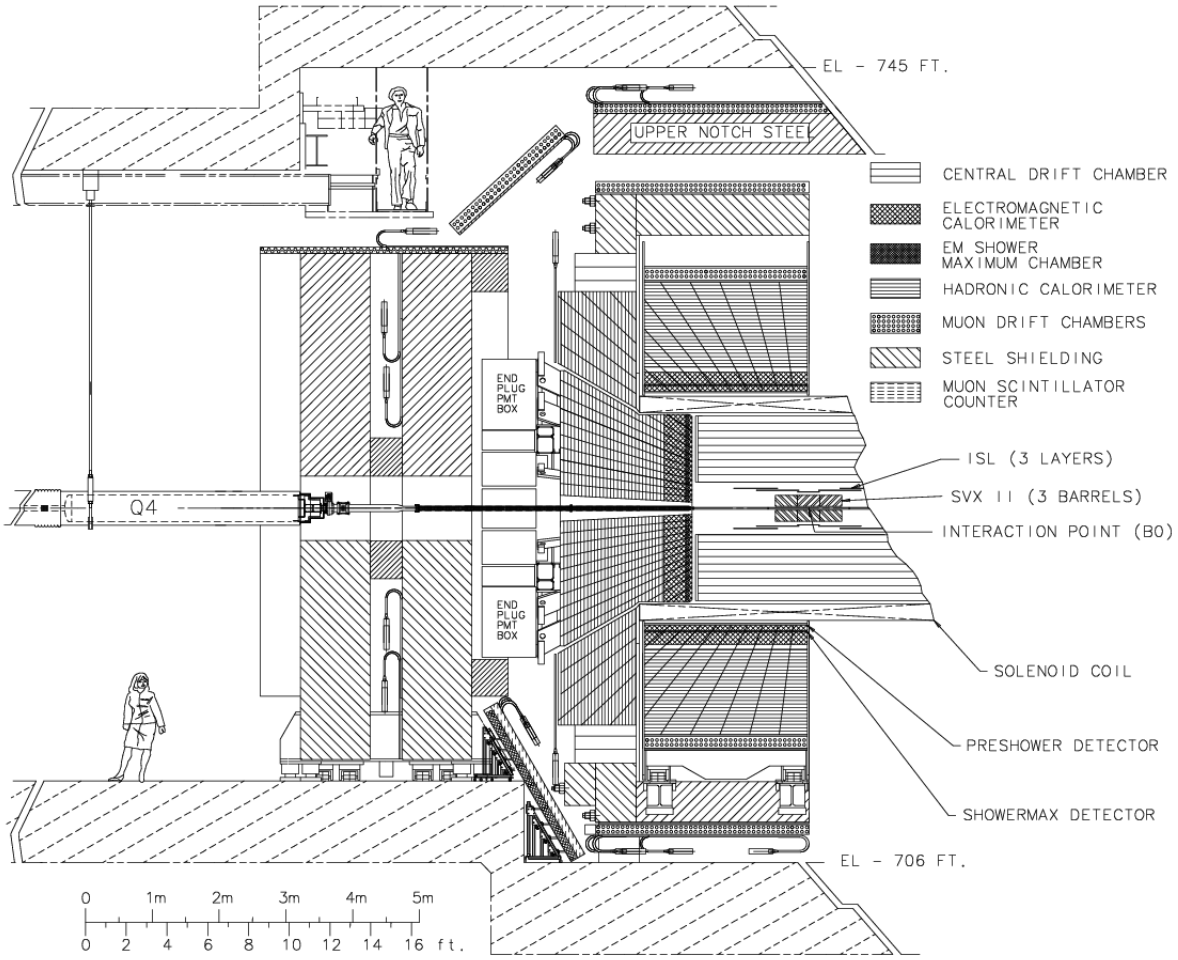


Figure 2.8: Elevation view of half of the CDF II detector.

2.2.1 Tracking system

Charged particles follow helical trajectories within the uniform magnetic field. Their trajectories are parameterized by five parameters:

- z_0 : the z coordinate of the closest point of track to the z-axis;
- $cotg(\theta)$: where θ is the track polar angle at z_0 ;
- ϕ_0 : azimuthal angle of the track direction at z_0 ;
- d_0 : impact parameter, the signed minimal distance from the z-axis;
- C : the helix curvature.

The impact parameter and the curvature are signed quantities defined by :

$$C = \frac{q}{2R} \tag{2.4}$$

$$d_0 = q(\sqrt{x_c^2 + y_x^2} - R) \tag{2.5}$$

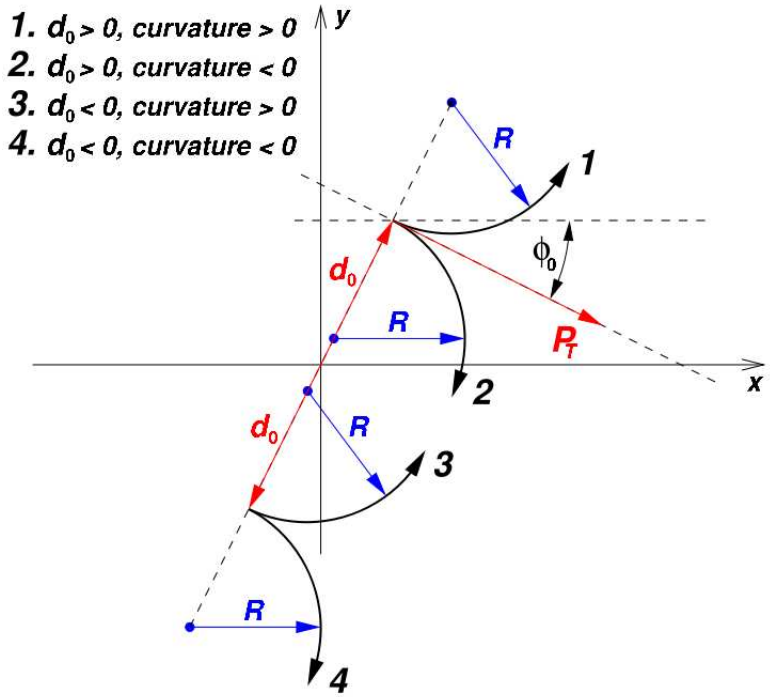


Figure 2.9: Illustration of helix track parametrization.

where q is the charge of the particle,$(x_c^2 + y_c^2)$ is the center of the helix as projected onto the $x-y$ plane and R is its radius. A graphical view of these variables together with the ϕ_0 is shown in Figure 2.9. Helix parameters are measured by the tracking system [70] in order to measure the charged particles' momentum. For a particle of momentum p and energy E, the transverse momentum p_T and the transverse energy E_T are defined as:

$$p_T = p \cdot \sin\theta \tag{2.6}$$

$$E_T = E \cdot \sin\theta \tag{2.7}$$

Particle transverse and longitudinal momenta can be derived as:

$$p_T = \frac{cB}{2|C|} \tag{2.8}$$

$$p_z = p_T \cot\theta \tag{2.9}$$

There are two types of inner tracking detectors used at CDF II. The innermost detector is a silicon microstrip based detector which provides very precise charged particle tracking. It consists of 3 major approximately cylindrical sub-systems coaxial with the beam-pipe: Layer 00 (L00) [70, 71, 72], the Silicon Vertex Detector (SVX II) [70], and the Intermediate Silicon Layers (ISL) [70, 73, 74]. Figure 2.10 zooms on the Inner Tracker from two different point of views. Outside of this is the Central Outer Tracker (COT) which is a wire drift chamber.
Some of the applications of the charged particle tracking system at the Tevatron collider includes:

- precision reconstruction of tracks' impact parameter and efficient reconstruction of

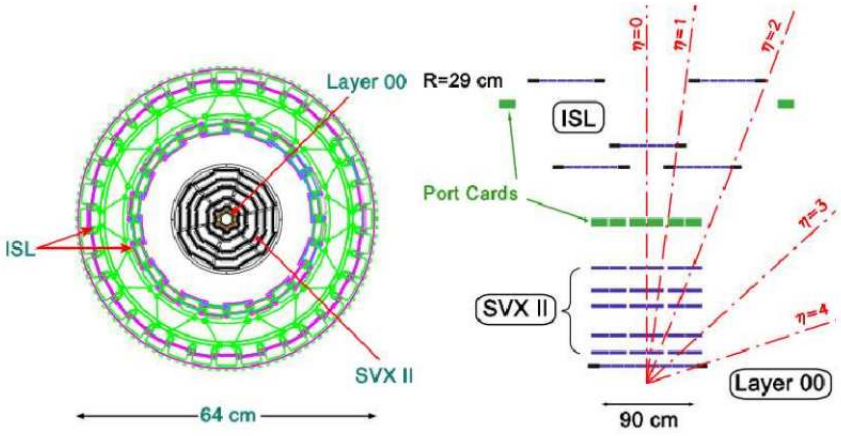


Figure 2.10: A sketch of the Inner Tracker silicon detector in a x/y projection (left) and a cutaway transverse to the beam of the three inner sub-systems (right): Layer 00 (L00), the Silicon Vertex Detector (SVX II) and the Intermediate Silicon Layers (ISL).

tracks in dense jets, which are the result of the recombination and hadronization of quarks or gluons leaving the interaction point;

- identification of jets which have a vertex slightly displaced from the initial interacting point (the displacement of these vertices is typically less than a millimeter, which is too small to be determined from the COT alone);
- precision reconstruction of the interaction vertex z coordinate to allow accounting and calibration of multiple interactions;
- the ability to combine tracks with information from calorimetry or muon chambers to provide efficient particles identification with excellent purity

Layer 00 (L00)

L00 detector consists of single-sided silicon micro-strip sensors assembled on the beam pipe. It is a 90 cm long radiation hardened cylindrical arrangement of silicon detectors at a radius of approximately 1.6 cm, providing a $|\eta| \leq 4.0$ coverage and a complete coverage in ϕ. It provides a precise measure of the track impact parameter with a resolution of $\sim 25/30$ μm [70, 71, 72] and additionally compensates for the degraded performance of the other silicon sub-detectors, due to radiation damages.

Silicon Vertex Detector (SVXII)

SVX II is a dual-sided silicon micro-strip detector, located outside L00, extends radially from 2.1 cm to 17.3 cm. It is segmented into three 29 cm barrels along the z-axis: this allows for a $|\eta| \leq 2.0$ coverage (Figure 2.11). It consists of 5 layers and its silicon micro-strips are aligned axially to the beam on one side, while in the other side are aligned with a small (1.2^o) angle stereo (layers 2 and 4) and with a 90^o stereo (layers 0, 1 and 3), providing information on ϕ as well as information in z from the stereo side. It provides high precision tracking and secondary vertex detection, exploiting 3D

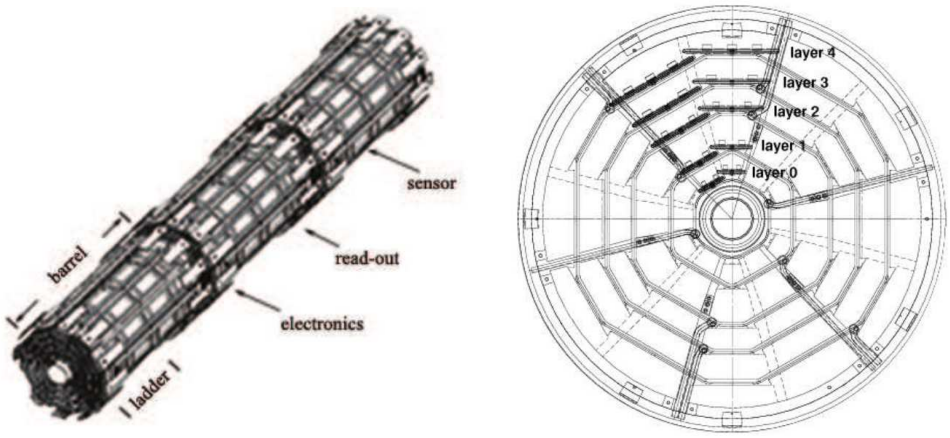


Figure 2.11: Isometric (left) and end-view (right) of the CDF Silicon Vertex Detector.

position measurements. This sub-detector has a 12 μm resolution on the single hit in the direction transverse to the beam.

Intermediate Silicon Layer (ISL)

The ISL, siting between the SVXII and COT, consists of double-sided silicon detectors (same as for SVXII), with axial strips on one side and small angle stereo strips on the other side. Four layers are at $1 \leq |\eta| \leq 2$ (at radii of 20 and 28 cm, as shown in Figure 2.10), one layer is at $|\eta| \leq 1.0$. The main ISL purpose is to compensate for incomplete coverage of other subdetectors in the region $|\eta| \geq 1.0$ by providing precision tracking at $1 \leq |\eta| \leq 2$.

Central Outer Tracker (COT)

The COT has a cylindrical shape and it is located outside SVX from $r = 40$ cm to $r = 137$ cm. COT provides full tracking in the central region ($|\eta| \leq 1$), even if its maximum geometrical acceptance reaches up to $|\eta| \leq 2$ (see figure 3.3), where the tracking performance reduces.
It is an open cell drift chamber operating with a gas mixture that is 50/50 argon-ethane with an admixture of 1.7% isopropanol. The COT is composed of 8 super-layers, each containing twelve layers of sense wires. The odd super-layers have wires parallel to the beam (axial super-layers) while the even ones have wires at a small (2.0^o) stereo angle (stereo super-layers). Figure 2.12 shows a portion of the COT endplate. Inside the solenoid magnetic field, the drifting electrons experience a Lorentz force which rotates their paths. The cells are tilted by 35^o with respect to the radial direction in order to have the electrons drifting perpendicularly to the radius for optimal momentum resolution[2]. The single hit position has been measured with an uncertainty of ~ 140 μm which translates into an overall transverse resolution $\sigma(p_T) = 0.15\% \cdot p_T$ [GeV/c][75]. Combined with the silicon detector the momentum resolution is $\sigma(p_T) = 0.07\% \cdot p_T$ [GeV/c] and for tracks with a transverse momentum p_T greater than 10 GeV/c the tracking efficiency of the COT is greater than 99% [76].

[2]For best momentum resolution, the optimal correlation between drift time and hit distance from wire is for electrons drifting in the direction transverse to the radius.

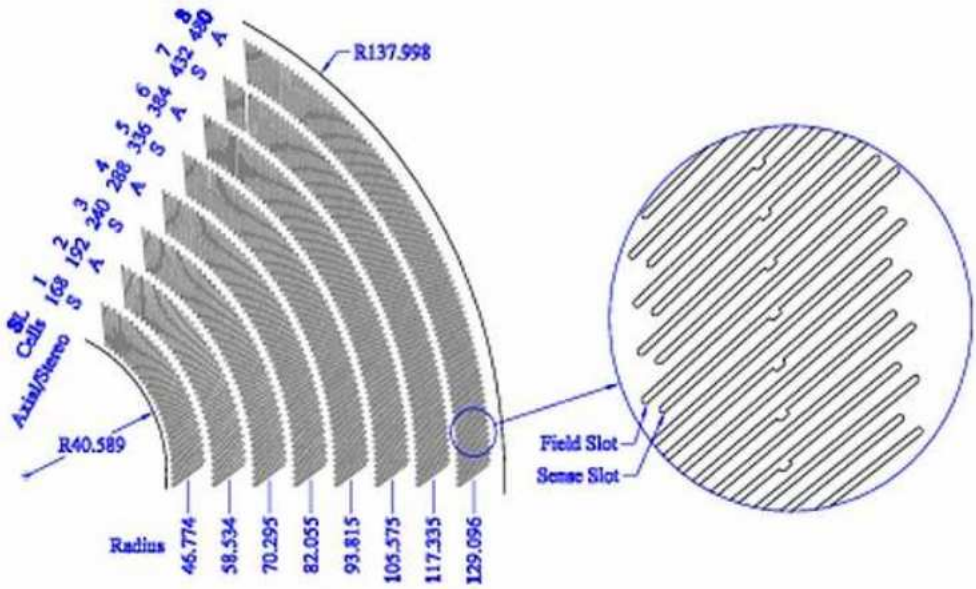


Figure 2.12: COT section: the eight superlayers (left) and the alternation of field plates and wire planes (right)

2.2.2 Calorimetric systems

The CDF calorimetry system [70] measures the particle energy and the direction of neutral and charged particles leaving the tracking region. Calorimeter information is also used to estimate the transverse energy of weakly interacting particles, such as neutrinos, by computing the imbalance in the total transverse energy. The calorimetry system consists of two different parts: an inner electromagnetic and an outer hadronic compartment, providing coverage up to $|\eta| \leq 3.6$. Therefore, particles hitting the calorimeter can be distinguished in electromagnetically interacting particles, such as electrons and photons, and hadronically interacting particles, such as mesons or baryons produced in hadronization processes.
Both calorimeters are scintillator sampling detectors segmented in $\eta - \phi$ sections, called *projective towers*, since they are projected towards the geometrical center of the detector, in order to supply information on particles' position. Each tower consists of alternating layers of passive absorber material (Pb in the electromagnetic and Fe in the hadronic compartment) and plastic scintillator for shower sampling. The light from the scintillator plates is read out by photo-multiplier tubes, through wavelength-shifting bars or plates and light guides.
High energy electrons and photons generate an electromagnetic shower which is mostly absorbed in the electromagnetic calorimeter compartment. For charged particles heavier than the electron, radiative energy losses are negligible to a first approximation. Hadrons interact with the detector matter mostly through inelastic collisions with nuclei of the absorbing medium. Particles produced in the nuclear interactions can loose their energy by ionization and secondary nuclear interactions. Thus, the energy of interacting particles is detected by measuring the deposited energy in hitted calorimetric towers. The energy E is assigned to a given tower as the sum of the energy deposited in that tower of the electromagnetic calorimeter and the corresponding one in the hadronic calorime-

ter: $E = E_{EM} + E_{HAD}$. Large localized energy deposits in both electromagnetic and hadronic calorimeters characterize *jets*.

The calorimetric system is subdivided into Central Electromagnetic Calorimeter (CEM) ($|\eta| \leq 1.1$), the Hadronic Calorimeter (CHA) in the central region ($|\eta| \leq 0.9$), the Endwall Hadronic Calorimeter (WHA) in the wall region ($0.9 < |\eta| \leq 1.3$) and the electromagnetic and hadronic plug calorimeters (PEM, PHA) in the plug region ($1.1 \leq |\eta| \leq 3.6$).

Central Calorimeter

The central calorimeters [77] CEM, CHA and WHA are azimuthally divided into 24 wedges. Each wedge in the CEM calorimeter is segmented in $\Delta\eta \times \Delta\phi = 0.11 \times 15^o$

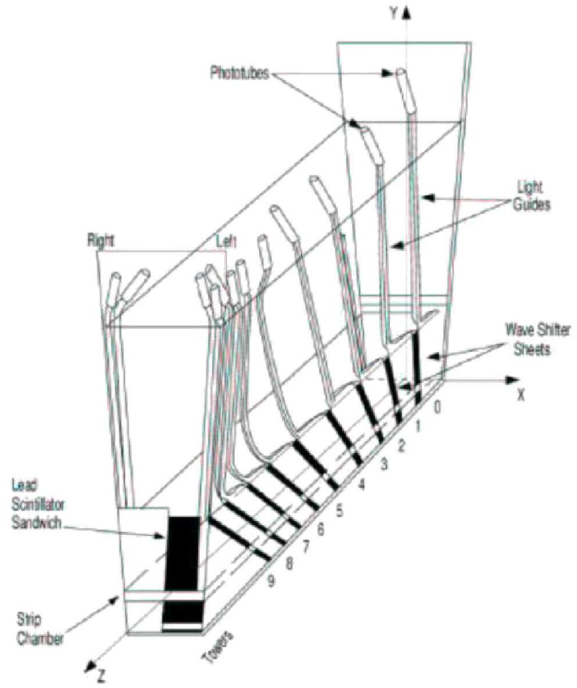


Figure 2.13: Perspective view of a CEM module: light-shifter plates connected to light guides and to photomultipliers of the front electromagnetic compartment of a central calorimeter wedge.

(see Figure 2.13), while the CHA and WHA geometric tower segmentation matches the CEM one in the central and wall regions respectively.
The CEM calorimeter is made of 31 alternate layers of 0.5 cm thick plastic scintillator plates and 0.32 cm thick lead absorbers: the total amount of material is $18 \cdot X_0$ (X_0 is the electron radiation length). The CEM energy resolution is:

$$\sigma_{E_T}/E_T = 13.5\%\sqrt{E_T[\text{GeV}]} \oplus 2\% \tag{2.10}$$

E_T being the energy of an electron or a photon hitting the calorimeter perpendicularly to its front.
The CHA and WHA calorimeters use instead iron layers as radiators. The total hadronic calorimeter thickness is $\sim 4.7\lambda_0$ (λ_0 is the charged pion absorption length) for both CHA

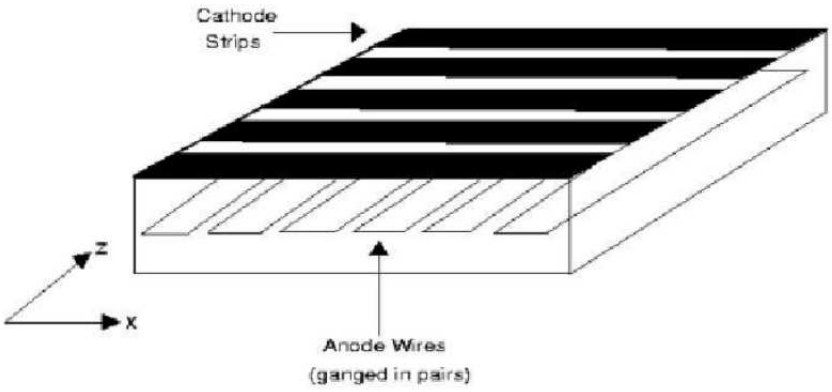


Figure 2.14: The CES detector in CEM. The cathode strips run in the x direction and the anode wires run in the z direction providing x and $(r - \phi)$ measurements.

and WHA. Resolutions of CHA and WHA for a perpendicular entering particle are:

$$\text{CHA}: \quad \sigma_{E_T}/E_T = 50\%\sqrt{E_T[\text{GeV}]} \oplus 3\% \tag{2.11}$$

$$\text{WHA}: \quad \sigma_{E_T}/E_T = 75\%\sqrt{E_T[\text{GeV}]} \oplus 4\% \tag{2.12}$$

The position on the reconstructed particles is determined by the ShowerMax detector. The central region is covered by the Central Electron Strip Chambers (CES), while Central Pre-Radiator (CPR) is located in front of the calorimeter. The two position detectors are embedded in each wedge of the CEM:

- Central Electron Strip Chambers (CES) : is a combined strip/wire gas proportional chamber embedded in CEM at about $6 \cdot X_0$ (184 cm) (See Figure 2.14). The CES purpose is to measure the position and the shape of electromagnetic showers in both transverse plane and longitudinal direction, by measuring the charge deposition of the electromagnetic showers.
  The CES detector identifies the hitting particle performing a strip/wire matching. Anode wires run parallel to the z-axis, and cathode strips orthogonally, thus the strip and wire clusters give independent measurements of the z and ϕ coordinates of the hit. The position of the hit comes from pairing strip and wire clusters, by examining the consistency between their energies. Typical deposited CES energy of an electron or photon at Tevatron is fully contained in a cluster of ~ 11 strip/wires (really wide 15-20 cm) [78], with the 95% of the energy of e/γ contained in clusters composed of 5 strips/wires. However, the choice of the cluster width in CES is a matter of compromise between containment of the deposited energy and the ability to resolve nearby clusters. CES has a finer azimuthal segmentation than the calorimeter tower: the resolution is about 1 cm in z and 1 mm in $r - \phi$. This results in an increased purity of electromagnetic object reconstruction.

- Central Pre-Radiator (CPR): is a set of multi-wire proportional chambers located in front of the calorimeter. It acts as a pre-shower detector, thus it helps distinguishing electrons from charged hadrons by gauging their probability of showering

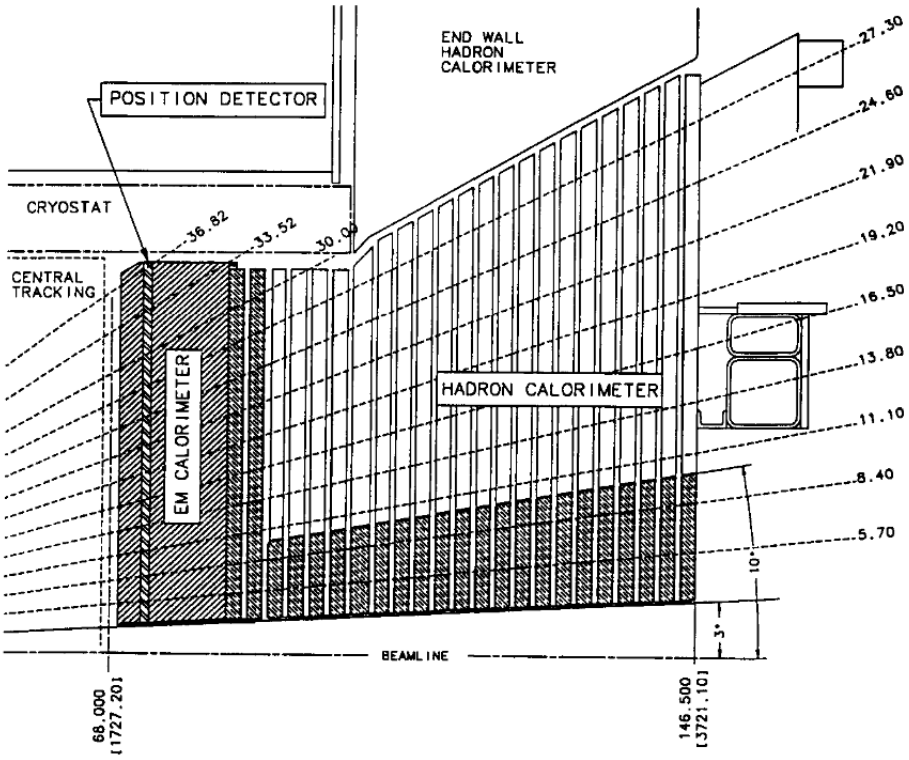


Figure 2.15: Plug Calorimeter (PEM and PHA) inserted in the Hadron End Wall calorimeter WHA and into the solenoid.

in the detector material prior to entering the calorimeter: the charged hadrons deposit a greater amount of energy in the chamber.

Plug calorimeter

The plug calorimeters (see Figure 2.15) cover the η region from 1.1 and 3.6. As in the central calorimeter, PEM and PHA consist of alternating layers of thick plastic scintillator plates and thick absorbers (lead and iron for electromagnetic and hadronic sectors respectively). Both electromagnetic and hadronic sectors are divided in 12 concentric η regions, with $\Delta\eta$ ranging from 0.10 to 0.64 according to increasing pseudorapidity, each segmented in 48 or 24 (for $|\eta| < 2.1$ or $|\eta| > 2.1$ respectively) projective towers. The segmentation is finer in ϕ (7.5 wide ϕ bins) than in the central calorimeter.
The transverse energy resolution in plug calorimeter is:

$$\text{PEM}: \quad \sigma_{E_T}/E_T = 16\%\sqrt{E_T[\text{GeV}]} \oplus 1\% \tag{2.13}$$

$$\text{PHA}: \quad \sigma_{E_T}/E_T = 80\%\sqrt{E_T[\text{GeV}]} \oplus 5\% \tag{2.14}$$

PEM is equipped with a shower maximum detector, PES. It consists of two layers of 200 scintillating bars each, oriented at crossed relative angles of 45^o ($\pm 22.5^o$ with respect to the radial direction). The position of a shower on the transverse plane is measured with a resolution of ~ 1 mm. More details are reported in [79].

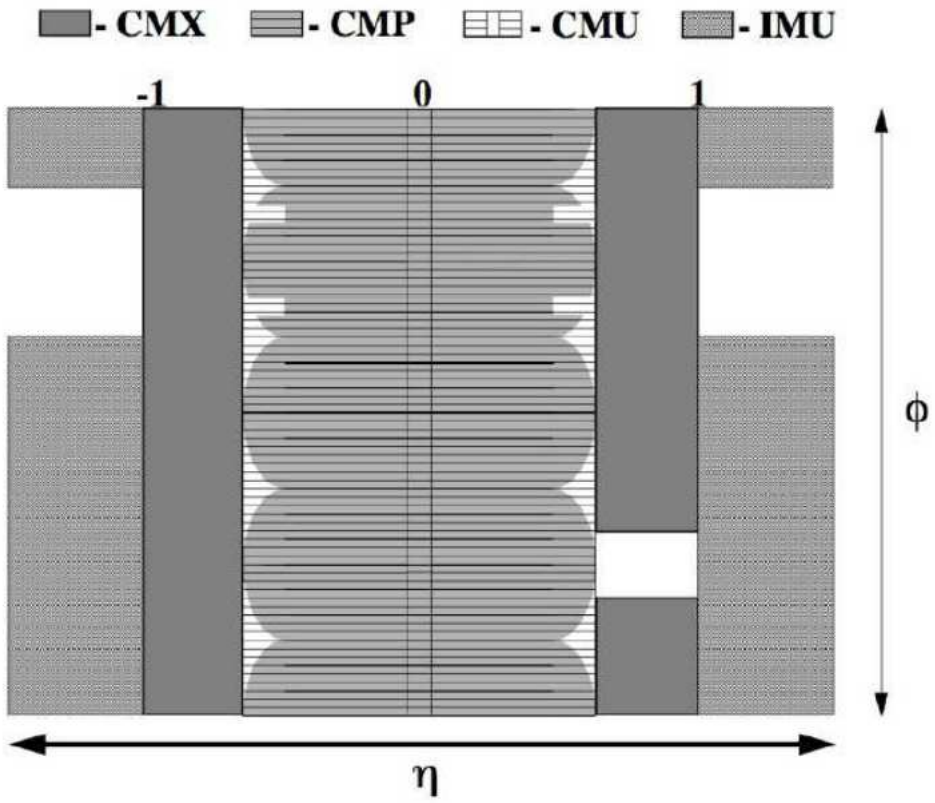


Figure 2.16: $\eta - \phi$ coverage of the muon detector system.

2.2.3 Muon detectors

Muon detectors are located in the outermost shell of the CDF detector. Since muon leptons interact as minimum ionizing particles (MIP) at the collision energy of the Tevatron, they are expected to pass through the calorimeters depositing only a small fraction of their energy, typically about 300 MeV in the electromagnetic compartment and about 2 GeV in the hadron compartment of the traversed calorimeter towers.

Four independent systems of drift chambers and scintillation counters placed at different radial distance from the beam axis are used to detect muons in the $|\eta| < 2$ region: the Central Muon Detectors (CMU), the Central Muon Upgrade Detectors (CMP), the Central Muon Extension (CMX), the Intermediate Muon Detectors (IMU). Muon detectors share common features [70, 80, 81] and their $\eta - \phi$ coverage is shown in Figure 2.16.
Chambers are coupled with scintillator counters in order to suppress backgrounds due to secondary interactions in the beam pipe material and to cosmic rays. Muons momenta are measured in the tracker, thus a muon candidate is reconstructed when a short track segment (stub) in the muon chambers corresponds to the extrapolation of a COT track. The momentum resolution of each muon sub-system is limited by multiple scattering within drift tubes.
The CMU detector is behind CHA at a radius of 347 cm from the beam axis and covers the $|\eta| < 0.68$ region. CMU consists of 4 drift tube layers sectioned by wedge matching the CHA towers. Each tube operates in proportional mode, with a maximum drift time of 0.8 μs. The multiple scattering resolution is 12/(p[GeV/c]) cm and the longitudinal

resolution is $\partial z \cong 10$ cm.
The CMP detector is located outside the CMU, it is arranged to enclose the $|\eta| < 0.68$ region in an approximately central box (see Figure 2.16). Unlike the other detectors in CDF, which are all (mostly) cylindrically symmetric around the beampipe, CMP is roughly box-shaped. This is because CMP uses the magnet return yoke steel as absorbing steel. The maximum drift time is 1.4 μs and it has a multiple scattering resolution of 15/(p[GeV/c]) cm. Scintillator layers (CSP) on the outermost side of the CMP chambers allow identifying bunch crossings. It provide timing information with a resolution of 1-2 ns. The CMU/CMP system is called CMUP.
The CMX detector extends the muon detector rapidity coverage to the $0.65 < |\eta| < 1$ region. It is located on either sides of the detector straddling the beamline, with a conical geometry of drift tubes with drift chambers. The CMX covers 360^o in ϕ, segmented in 15^o wedges in azimuthal angle. Its parts in $75^o \leq \phi \leq 105^o$ and $225^o \leq \phi \leq 315^o$ are called *miniskirt* and *keyston* detector respectively. CMX contains two folds of 4 layers of rectangular drift tubes. The multiple scattering resolution is 13/(p[GeV/c]) cm and the longitudinal position resolution is $\partial z \simeq 14$ cm. As for CMP, CMX is sandwiched to scintillators (CSX). One layer is mounted to the upper surface and the other to the lower surface of the CMX to provide timing information for the CMX.
The forward region of the muon system is the IMU detector ($1.0 < |\eta| < 1.5$). It contains four layers of proportional drift tubes (BMU), with a maximum drift time of 0.8 μs. The multiple scattering resolution is 13-25/(p[GeV/c]) cm and the longitudinal position resolution is $\partial z \backsimeq 16.5$ cm. Three layers of scintillators (BSU-F, BSU-R and TSU) are mounted outside the BMU to provide timing information.

2.2.4 Cherenkov Luminosity Counter: CLC

In Run II CDF measures the collider luminosity with a coincidence between two arrays of Cherenkov counters, the CLC, placed around the beam pipe on the two detector sides. The counters measure the number of hits for each bunch crossing then providing the average number μ of interactions per bunch crossing [82]. The luminosity is derived from the known average number of secondaries and inelastic cross section over the CLC angular coverage:

$$L_{inst} = \frac{\mu f_{bunch}}{\sigma_{p\bar{p}}} \tag{2.15}$$

In this expression, $\sigma_{p\bar{p}}$ is the value of the inelastic $p\bar{p}$ cross section at 1.96 TeV and f_{bunch} is the rate of bunch crossing. This method measures the luminosity with about a 6% systematic uncertainty.
Each CLC module contains 48 gas Cherenkov counters of conical shape projecting to the nominal interaction point, organized in concentric layers. It works on the principle that light produced by any particle originated at the collision point is collected with a much higher efficiency than for background stray particles. The CLC signal is thus approximately proportional to the number of traversing particles produced in the collision.

2.3 CDF Trigger

At Run II the bunch crossing rate is of 2.53 MHz (396 ns bunch spacing) and, considering the inelastic cross section for $p\bar{p}$ scattering of about 80.03 ± 2.24 mb [83] and an high luminosity of order $10^{32} cm^{-2} s^{-1}$, the interaction rate is much higher than any possible event recording rate (order of 100 Hz).
The identification of the most interesting events is accomplished by a three-level trigger system (see Figure 2.17). Events selected by the trigger system are saved permanently on a mass storage and subsequently fully reconstructed offline. Each level of the trigger system performs a slower but more accurate event reconstruction and applies a tighter filter with increasing trigger level according to a set of pre-defined conditions [70].

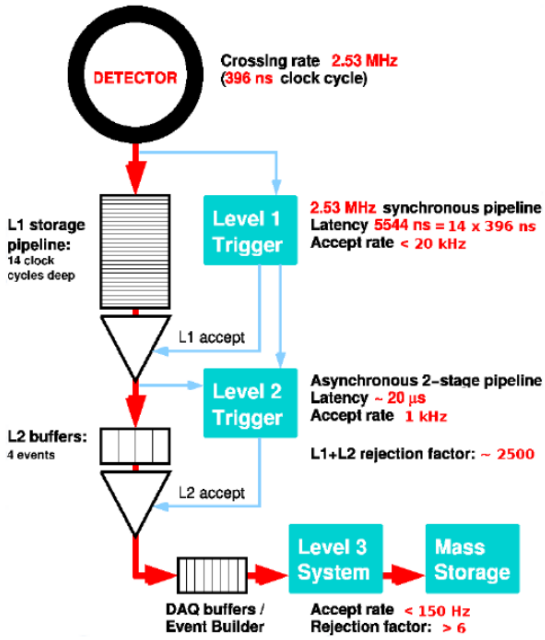


Figure 2.17: CDF trigger system [70]. The readout functional diagram: Dataflow for the CDF trigger design to accomodate bunch crossing of 396 ns; the front-end buffers data in 4μs pipelines while the Level 1 trigger processes the data and makes a trigger decision; for each Level 1 accept the data is stored in a Level 2 buffer while the Level 2 trigger works; A Level 2 accept initiates the readout of the event data into Level 3 processors farm where an accept causes the data to be written to disk/tape.

Level 1

The Level 1 (L1) task is to reduce the input crossing rate to 20 kHz. It is a synchronous pipeline system with an event read in every beam crossing and a decision (L1 accept/L1 reject) made every beam crossing. At L1 data must be buffered at least as deep as L1 decision time, expected to be approximately 4 μs (L1 Latency), to provide for near deadtimeless operations. The trigger and front-end electronics operate with short bunch spacing (396 ns). The L1 buffer can store 14 subsequent events, corresponding to a store time of about 5.5 μs, while the hardware is taking a decision. If an event is accepted, the information is stored in the Level 2 buffer memory.

To decide to retain an event for further processing, the L1 exploits most of the detector information before any physics object recognition starts. L1 typically rejects $\sim 97\%$ of the events.
The trigger objects available at L1 are:

- *L1 tracks.* An hardwired algorithm named eXtremely Fast Tracker (XFT) is a track finding tool which reconstructs approximate tracks, returning track p_T and ϕ by exploiting information from COT superlayers. These tracks are then extrapolated to the central calorimeter wedges and to the muon chambers (CMU and CMX), allowing the first electron or muon identification.

- *L1 trigger tower.* The energy released in the electromagnetic or hadronic towers are merged in pairs along η to define trigger towers, which are the basis for the definition of electron/photon and jet or global quantities such as $\sum E_T$ and the missing energy $\not\!\!E_T$.

- *L1 leptons.* L1 muon and electron triggers are obtained by matching a XFT track to a calorimetric trigger tower, for electron, or to clusters of hits in the muon chambers, for muons.

Level 2

The Level 2 (L2) trigger is an asynchronous system which processes events that have received a L1 accept. Buffers are used to store events until a decision is made. The L2 processing time is expected to be of about 20μs (L2 Latency) and L2 acceptance rate of ~ 1 kHz (rejection factor ~ 150). All the information used in the L1 decision is available to the L2 system, but with higher precision. In addition, data from the central calorimeter showermax detector allow improved identification of electrons and photons.

Trigger objects available at L2 are:

- *L2 calorimeter cluster.* An algorithm for cluster finding is implemented, clumping together neighbor calorimeter trigger towers inside a fixed cone centered at the *seed* tower, in order to improve the jet selection in the event. This L2CAL system [84] reduces the L2 trigger rate and provides at L2 improved measurements of jets energy and position. L2 clusters can be used to build trigger objects by applying a cut on their transverse energy and position, and global triggers by selecting on the number and $\sum E_T$ of clusters.

- *L2 SVT track.* At L2 the Silicon Vertex Tracker (SVT) [85, 86] is dedicated to charge tracks reconstruction in the $r - \phi$ plane transverse to the beam line. It uses information from the silicon vertex detector SVXII and the p_T and azimuthal angle ϕ of the tracks found by XFT, providing precise measurement of track impact parameter d_0, curvature and azimuthal angle.

- *L2 leptons.* L2 muons are optimized by a ϕ-matching (within 1.25^o) between XFT tracks and track segments (stubs) formed with hits in the muon chambers. For L2 electrons a finer ϕ-matching is performed exploiting the information from central and plug shower maximum detectors.

Level 3

Level 3 (L3) is a software trigger which almost fully reconstructs the entire event. L3 addresses event objects delivered by L2 to the Event Builder (EVB), which associates information from different detector parts to reconstruct the entire event. All physics objects in the event, such as tracks reconstructed in the whole detector, primary and secondary vertex, leptons, jets and missing energy, are reconstructed.
The final decision to accept an event is made on the basis of a list of observables indicating candidate events of physical interest (top production events, W/Z events, Drell-Yan events, etc.).
Accepted events exit L3 at a rate of up to 100 Hz and are permanently stored on tape for further offline analyses.

Many different *trigger paths* are defined to collect events for the different analyses performed at CDF. A trigger path identifies a unique combination of a L1, a L2, and a L3 trigger; data sets (or data streams) are then finally formed by merging the data samples collected via different trigger paths. During the full period of data taking the definition of trigger paths evolved: some triggers have been *pre-scaled* at L2 or *disabled* above some luminosities. Each trigger relays on different subdetectors therefore the total collected luminosity is different for each trigger path, while information about online status of the subdetectors is stored in the *good run list* [87].

2.4 Offline data processing

Row data from L3 triggers are grouped during real-time (*on-line*) acquisition in *run numbers* according to required trigger-paths and stored on fast-access disks as the data are collected. Physics objects, such as tracks, vertices, leptons and jets are generated after manipulation of raw data, referred as *offline* data processing. The algorithms used in the offline analysis are similar to those used for the reconstruction done at L3, except for the application of up-to-date detector calibrations, the best measured beamlines, etc. Several run numbers are grouped offline in *run periods*, each one with an integrated luminosity of the order of $\sim 100\text{pb}^{-1}$. The analysis described in this Thesis uses approximately 4.8 fb^{-1} of integrated luminosity, collected from the beginning of RunII and up to the and of March 2009. According to the CDF nomenclature, the collected luminosity used in this analysis corresponds to the data taken from *period 0* to *period 23* (runs 138425 to 274055).

2.5 Monte Carlo Simulation

Data analyses require deep knowledge of the physics processes and the detector response. To evaluate measured data, it is therefore useful to simulate all physics processes expected to contribute to the corresponding data sample and to emulate the detector response. Physics events that can be detected at the CDF detector can be simulated using Monte Carlo (MC) techniques, in order to predict the composition of the final state distributions and the contribution of not-yet observed processes.

The CDFSIM [88] simulation package provides the process simulation, reproducing the hard scattering and the generated particles, and the modeling of the various compo-

nents of the CDF detector, including the particle interaction through them. Standard Model processes are modeled using simulated events from several Monte Carlo generators which are then passed through the full GEANT [89] based CDF II detector simulation. Therefore the full simulation of the physics processes can be divided in the following components:

- Physics process simulation: simulation of the proton-proton interactions and of the produced particles;
- Parton showering and hadronization simulation: simulation of the jets formation;
- Detector simulation: simulation of the interactions of particle through the CDF detector and the effects of the online conditions at data taking;

Physics processes simulation

The hard scattering between partons in the $p\bar{p}$ collisions results in the production of up to hundreds of outgoing particles. A detailed description of the production of multiple particles is realized by Monte Carlo event generators which randomly produce collision data with assigned kinematic properties, according to the differential cross section of a given process.

Any theoretical model describing an elementary process starts from the knowledge of its cross section and must both contain a way to compute or to estimate the effects of higher-order in α_s perturbation theory and a way to describe hadronization effects. Event generators perform the finite higher-order corrections through the exact computation of a given number of perturbation orders, while the effects due to emissions at all orders in perturbation theory is estimated by the parton showering technique.

The signal and most of the background processes in this Thesis are generated using PYTHIA v6.4 [90], a general-purpose event generator. It contains showering routines as well as an event generator. The event generator in PYTHIA can handle simple Feynman diagrams, thus the hard scattering between partons in the $p\bar{p}$ collisions is generated at the Leading Order (LO). The CTEQ5L parton distribution functions [91] is used to describe the parton showering, accounting for the effects of initial- and final-state radiation and the hadronization of color-charged particles. More particles may be added in the final state of the event from effect of beam remnant or multiple interactions, therefore effects coming from the underlying event are also included in this simulation.
Processes involving γ are better simulated using the BAUR matrix element generator [92] for the hard scattering, providing a better description of the QED radiation.
Processes with an electroweak boson and radiated gluons are difficult to deal with because of the large amount of radiation they produce. ALPGEN Monte Carlo generator [93] is specifically designed for processes whose final state contains an electroweak boson and several radiated quarks and gluons, a major background in this analysis. The advantage of using ALPGEN is that it models accurately events with multiple jets, giving a more accurate modeling of the kinematics of the process than PYTHIA's showering approximation.

Parton showering and hadronization simulation

All events, regardless of how they were generated, are passed to PYTHIA for parton showering and hadronization. The showering procedure generates initial- and final-state

gluon radiation for each event and allows them to decay to quark pairs, increasing the number of particles in the final state of the event. This gives the final set of particles that are passed to the hadronization routine. Most of the particles resulting from hadronization are unstable, so PYTHIA causes them to decay into relatively stable particles (electrons and muons, protons and neutrons, pions and kaons) that can actually be detected. This step uses branching ratios and lifetimes measured in various experiments to calculate the final decay products.

Detector simulation

Once the particles have been generated, their interactions through the detector is simulated with the Geant3 package [89]. This requires a full detector simulation which simulates the response of the different subcomponents of the detector, including resolution effects, inefficiencies and the behavior of the particles as they pass through passive materials in the detector. The online beam position and the detector conditions at data taking are taken into account in the simulation. Simulations are available for the following subdetectors: Silicon detectors (SVX, ISL), COT, muon systems, Time-of-Flight system (ToF), calorimeters, Cherenkov Luminosity Counters (CLC) and forward detectors.
The geometry of the subdetectors has been implemented in great details. The charge deposition of traversing tracks on the silicon strips is calculated using a simple geometrical model based on the path length of the ionizing particle. The drift model for the COT simulation is based on the GARFIELD [94] package, a general drift chamber simulation program. The calorimeter simulation uses the GFLASH [95] shower developmen package. The EM and HAD showers simulation in GFLASH is initiated when particles undergo inelastic interactions inside the calorimeter volume, using the GEANT3 geometry description for the CDF calorimeters. The most challenging part in the muon simulation is the description of the complicated geometry of the muon systems. More details on the CDF II simulation can be found in the reference [88].

The simulation follows a run-dependent scheme reproducing a specific data taking condition. A sample used to reproduce a given period retrieves the status of the different systems and beam conditions of the runs recorded during that period, according to the instantaneous luminosity profile.
This implies that simulated events are corrected to represent data collected in a given time range, even if all of them will then be scaled to represent the full data set analyzed.
Once the detector data is obtained, either from Tevatron collisions or simulated MC events, it needs to be converted from the raw data in the detector to reconstructed physics quantities in order to be analyzed. First, information from subdetectors is combined to form high-level detector objects: tracks in the tracking detectors, stubs in the muon chambers, and clusters of energy towers in the calorimeters. Then these objects are analyzed to associate them with candidates of physical objects: electrons, muons, jets, or neutrinos. These can be finally used in a physics analysis.

Chapter 3

Reconstruction of physical objects

In this chapter the objects used in the event reconstruction are described with the goal of maximizing the total acceptance for the final states $e - \tau$ and $\mu - \tau$.
Lepton identification will be discussed in Section 3.2 and 3.3 respectively for electrons and muons. The τ hadronic decay reconstruction and its identification will be deeply discussed in Section 3.5 and 3.6.

3.1 Geometrical and kinematical selection

The total acceptance of $e\tau$ and $\mu\tau$ dileptons from $H \to WW$ decay is given by the combined geometrical and kinematical acceptances. The geometrical acceptance depends on the coverage of the relevant detector components (see Section 2.2). The kinematical acceptance is defined as the fraction of events for which both leptons l ($l = e$ or μ) and τ satisfy certain kinematical cuts.
The lepton identification requirements depend on the type of lepton, on its energy and on which part of the detector has reconstructed it. According to that, they can be classified as:

- Central lepton: $|\eta| < 1.0$
- Forward lepton: $1.0 < |\eta| < 2.0$

where η, defined as (2.2), refers to the pseudorapidity of reconstructed leptons.
Different (non-overlapping) electron and muon categories are defined according to which detector components have been used for the lepton reconstruction, since the subdetectors have a different segmentation resulting in different resolutions. Tau lepton reconstruction and identification instead depend on the decay mode. Taus decaying leptonically are identified following the standard electron or muon reconstruction, while the hadronic tau reconstruction is handled separately.
Other objects that can be reconstructed in the detector and enter the final state are photons and jets, generated by quarks and gluons produced directly from the $p\bar{p}$ hard scattering or by either ISR or FSR. Therefore, lepton identification aims at maximize the discrimination between real leptons and *fake leptons*, such as photon conversion or jet faking leptons. An explicit identification of photons is not required to the analysis described in this Thesis, and it will not be discussed, while algorithms developed for jet reconstruction are described in Section 3.7 .

3.2 Electron identification

Electrons can be identified in the detector searching for a charged and isolated track in the tracking system matched to a cluster in the electromagnetic calorimeter. The electron clusters have high electromagnetic fraction close to 1, since electrons, ideally, deposit all their energy in the electromagnetic calorimeter. Additionally, the ratio E/P of calorimeter energy to track momentum is expected to be close to 1.
The electron four-momentum (E, $\vec{p}$) is built using track information to determine the direction $\frac{\vec{p}}{|\vec{p}|}$ and energy of the matched calorimeter cluster for its magnitude $E \equiv |\vec{p}|$, in the massless approximation.
In this Thesis, the Likelihood-Based Electron (LBE) and Tight Central Electrons (TCE) categories are defined to identify central electrons, and Phoenix electrons (PHX) for electrons in the forward region.

Central electrons

The central electron categories relay on the reconstruction of energy deposited in the central region of the CDF detector, for $|\eta| \leq 1.1$, and are optimized for electrons with $p_T > 10$ GeV/c. The following quantities are used to define central electron categories:

- $\mathbf{E_{HAD}/E_{EM}}$: hadronic to electromagnetic energy ratio associated with the candidate cluster.

- $\mathbf{E/P}$: ratio of the electromagnetic cluster transverse energy to the COT track transverse momentum.

- $\mathbf{L_{shr}}$: the lateral shower profile in the plane transverse to the electron direction. It is defined as:
$$L_{shr} = 0.14 \frac{\sum_i (M_i - P_i)}{\sqrt{(0.14\sqrt{E_{EM}})^2 + \sum_i (\Delta P_i)^2}} \tag{3.1}$$
where the sums are over the towers in the electromagnetic cluster, M_i is the measured energy, P_i is the predicted energy in the i^{th} tower, E_{EM} is the total electromagnetic energy in the cluster, and ΔP_i is an estimate of the uncertainty in P_i [96]. L_{shr} is a measure of how well the measured lateral shower matches that expected for an electromagnetic shower from a single electron/photon, thus it can be properly viewed as an isolation energy cut.

- $\mathbf{Iso_{cal}}$: calorimetric isolation, defined as the energy E_T in a cone of radius $\Delta R = 0.4$ around the electron cluster without including it, divided by the energy in the electron cluster:
$$\mathrm{Iso_{cal}} = \frac{E_T^{\mathrm{cone}} - E_T^{\mathrm{electron}}}{E_T^{\mathrm{electron}}} \tag{3.2}$$

- $\mathbf{Iso_{trk}}$: track-based isolation, defined as $\mathrm{Iso_{cal}}$ but using all tracks with $p_T >$ 500 MeV/c and $|\eta| < 1.5$ in a cone $\Delta R = 0.4$ around the electron track.

- $\mathbf{Q \times \Delta x_{CES}}$: distance in $r-\phi$ plane between the extrapolated COT track and the nearest cluster in CES, times the charge Q of the track to account for asymmetric tails originated from bremsstrahlung radiation.

- **ΔZ_{CES}**: distance in $r-z$ plane between the extrapolated COT track and the best matching cluster in CES.
- **NCotHitsAx**: number of hits belonging to track associated to the candidate electron on COT axial super-layers.
- **NCotHitsSt**: number of hits belonging to track associated to the candidate electron on COT stereo super-layers .
- **χ^2_{COT}**: χ^2 of the track fit computed using only the COT hits belonging to the track.
- **NSvxHits** : number of SVX hits belonging to track associated to the candidate electron.

The Tight Central Electron (TCE) category is defined by means of a cut-based selection on the quantities described above and listed in Table 3.1. A *conversion removal* cut is

	TCE
Region	central
Track p_T	≥ 10 GeV/c (5 GeV/c if $E_T < 20$ GeV)
Track $\lvert z_0 \rvert$	≤ 60 cm
N. Ax SL (5hits)	≥ 3
N. St SL (5hits)	≥ 2
Conversion	false
E_{HAD}/E_{EM}	$\leq 0.055 + 0.00045 * E$
$\mathrm{Iso_{cal}}$	≤ 0.1
$\mathrm{L_{shr}}$	≤ 0.2
E/P	$< 2.5 + 0.015 * E_T$
$\Delta X \cdot Q$	$3 \leq q * \Delta X \leq 1.5$ cm
CES $\lvert \Delta Z \rvert$	< 3 cm
Track	Beam constrained

Table 3.1: Definition of cut-based Tight Central Electron (TCE).

applied to reduce the background due to electrons generating from photon conversion or trident events ($\gamma \to e^{\pm}e^{\mp}$ or $e^{\pm}\gamma \to e^{\pm}e^{\pm}e^{\mp}$). To reduce this background the removal algorithm requires two opposite sign tracks such that the separation in the $x-y$ plane is $|S_{xy}| < 0.2$ cm and the difference in the polar angle is $|\Delta \cot\theta| < 0.04$ [97]. The electron track is *beam constrained*: it is reconstructed constraining its origin to the beam position and the transverse momentum p_T has to be greater of 10 GeV/c ($p_T > 5$ GeV/c only if $E_T < 20$ GeV).

The Likelihood-Based Electron (LBE) category enhances the fake electron rejection with respect to TCE. A Likelihood function is created exploiting the combination of the detector quantities described above, to evaluate the probability for the candidate electron to be a real or fake lepton. The Likelihood $\mathcal{L}$ has the following form:

$$\mathcal{L}(\vec{x}) = \frac{\mathcal{L}_{sig}}{\mathcal{L}_{sig} + \mathcal{L}_{bkg}} = \frac{\prod_{i=0}^{N} P_i^{sig}(x_i)}{\prod_{i=0}^{N} P_i^{sig}(x_i) + \prod_{i=0}^{N} P_i^{bkg}(x_i)} \tag{3.3}$$

where N is the number of the used quantities. $P_i^{sig}(x_i)$ and $P_i^{bkg}(x_i)$ are the template functions which give the probability to have the value x_i for the i-th identification variable in real (signal) or fake (background) electron assumption respectively. By definition the $\mathcal{L}$ has values within the range [0.0, 1.0]. More details on the Likelihood method for lepton identification can be found in [98].
A cut on the Likelihood function is imposed out of mostly the same set of criteria defined for the TCE category. Therefore a full identified LBE electron satisfies the selection summarized in Table 3.2 .

	LBE
Region	central
E_{HAD}/E_{EM}	≤ 0.125
Track p_T	≥ 10 GeV/c (5 GeV/c if $E_T < 20$ GeV)
Track $\lvert Z_0 \rvert$	≤ 60 cm
Iso_{cal}	≤ 0.3
Conversion	false
Track	Beam constrained
Likelihood ($\mathcal{L}$)	$\mathcal{L} > 0.99$

Table 3.2: Identification criteria for LBE central electron.

Forward electrons

Tight Plug Phoenix Electrons (PHX) are forward electrons with a cluster in the PEM calorimeter and a silicon track reconstructed with the calorimetry-seeded "phoenix" tracking algorithm [99]. The following detector quantities are used for PHX identification:

- $\mathbf{Iso_{cal}}$: The total isolation energy (cone with $R = 0.4$) of a PEM cluster, it is defined as (3.2) and corrected for leakage [1].
- **PEM** $\boldsymbol{\chi^2_{3\times3}}$: An algorithm which compares the electromagnetic shower profile of a given PEM cluster to shapes derived from plug test beam data.
- **PES Profile Ratio** $\mathbf{5 \times 9}$ The ratio of the energy in the central five strips to the energy of the total nine strips of a PES cluster. Additional information can be found in [99].
- $\mathbf{\Delta R_{PES-PEM}}$: The matching distance in the $\eta - \phi$ plane between the PEM cluster and a PES 2D cluster.
- $\mathbf{N^{Si}_{hit}}$: number of layers of the silicon detector with hits from the reconstructed electron track, given the COT limited coverage in the plug region, it improves the cluster-matching track quality.

The forward electrons identification is discussed in more details in [99]. The cut values for PHX identification are shown in Table 3.3 .

[1] The electron's shower may not be fully contained in the 2×2 cluster of PEM towers and as such there may be some leakage to neighboring towers. The leakage is parameterized in terms of the position of the center of the electron's shower in the 2×2 cluster.

Variable	PHX Electrons
Region	PEM
E_T	≥ 20 GeV
PES 2D η	$1.2 \leq \lvert\eta\rvert \leq 2.0$
E_{HAD}/E_{EM}	≤ 0.05
PEM $\chi^2_{3\times3}$	≤ 10
PES 5×9	≥ 0.65
$\mathrm{Iso_{cal}}$	≤ 0.1
$\Delta R_{PES-PEM}$	≤ 3
Track $\lvert Z_0\rvert$	≤ 60 cm
$\mathrm{N^{Si}_{hit}}$	≥ 3

Table 3.3: Forward electron (PHX) definition cuts.

3.3 Muon identification

A muon signature is characterized by an isolated high-p_T track, which deposits only a small fraction of energy in either the electromagnetic or the hadronic calorimeter.
Muons are reconstructed in eight non-overlapping categories, depending on the instrumented region of the detector where the stub is reconstructed or the track is pointing to. Six of them require the muons to have hits in the muon detector subsystem, and are call *stubbed muons*. The other two are designed to recover muons that do not have hits in the muon detectors and are referred as *stubbless muons*.
The muon categories [100] are:

- central muons: CMUP, CMU, CMP, CMX and CMXMsKs [2] stubbed; CMIOCES stubbless.
- forward muons: BMU stubbed; CMIOPES subbless.

Kinematical and track quality criteria of the candidate muons are applied in each category to identify good muons. The cuts applied to all muon categories are summarized in Table 3.4 for central (stubbed and stubbless) muons. Tables 3.5, 3.6, 3.7 and 3.8

	For all central muons (stubbed and stubless muons)
Track χ^2/n.d.f	≤ 3 (≤ 4 if run ≤ 186598)
NCotHitsAx(5 hits)	≥ 3
NCotHitsSt(5 hits)	≥ 2
Track z_0	≤ 60 cm
Track d_0	≤ 0.2 cm (≤ 0.02 cm if $\mathrm{N^{Si}_{hit}} > 0$)
$\mathrm{Iso_{trk}}$	≤ 0.1
$\mathrm{Iso_{cal}}$	≤ 0.1
E_{EM}	$\leq 2 + \max(0, (p-100)\cdot 0.0115)$ GeV
E_{HAD}	$\leq 6 + \max(0, (p-100)\cdot 0.028)$ GeV

Table 3.4: Kinematical and track quality cuts for all central stubbed and stubbless muons.

[2] Pointing to either the miniskirt or keystone detectors (see Section 2.2.3).

show muon definitions specifically for the six muon stabbed categories. In Table 3.9 are reported CMIOCES and CMIOPES identification requirements.
Central muons are required to have a reconstructed track with a fit χ^2/n.d.f. < 3. This requirement is relaxed to χ^2/n.d.f. < 4 for data taken before run 186598. Requirements on the number of Axial and Stereo COT super-layers, track $|z_0|$ and impact parameter $|d_0|$ are applied (see Table 3.4).
For forward muons (BMU) there is a limited coverage from the COT drift chamber, so the cut on the number of hits is replaced requiring a number of hits in the chamber that is at least 60% of the expected one, based on the track direction, and at least three hits in the silicon detectors with a curvature significance $C/\sigma(C) > 12$ (see Table 3.8).
To reject fake muons an isolation requirement on the track is also applied to all the muon categories, $\mathrm{Iso_{trk}} < 0.1$, defined in the same way as for electrons in Section 3.1.1 .
The cut on the energy deposited in the electromagnetic and hadronic towers hit by the extrapolated track is required to be compatible with a minimum ionizing particle. An additional energy cut is applied to the calorimeter for muon candidates that have not been associated to a stub, CMIOCES and CMIOPEM.

	CMUP
Region	Central
ID cuts	See Table 3.4
ΔX_{CMU}	≤ 7 cm
ΔX_{CMP}	$\leq \max(6.0,\ 150/p_T[\mathrm{GeV/c}])$ cm

Table 3.5: Definition of CMUP central muons.

	CMU
Region	Central
ID cuts	See Table 3.4
ΔX_{CMU}	≤ 7 cm
	Not CMP or CMX

	CMP
Region	Central
ID cuts	See Table 3.4
ΔX_{CMP}	$\leq \max(6.0,\ 150/p_T[\mathrm{GeV/c}])$ cm
	Not CMU

Table 3.6: Definition of CMU and CMP central muons.

	CMX
Region	Central
ID cuts	See Table 3.4
ΔX_{CMX}	$< \max(6.0, 125.0/p_T)$ cm
	CMXMsKs
Region	Central
ID cuts	See Table 3.4
ΔX_{CMX}	$< \max(6.0, 125.0/p_T)$ cm

Table 3.7: Definition of CMX and CMXMsKs central muons.

	BMU
Region	Forward
COT hit fraction	≥ 0.6
$N_{\text{hit}}^{\text{Si}}$	≥ 3
Track z_0	≤ 60 cm
Track d_0	≤ 0.02 cm
C/σ(C)	> 12
Iso_{trk}	≤ 0.1
Iso_{cal}	≤ 0.1
E_{EM}	$\leq 2 + \max(0, (p-100)\cdot 0.0115)$ GeV
E_{HAD}	$\leq 6 + \max(0, (p-100)\cdot 0.028)$ GeV
$E_{EM} + E_{HAD}$	> 0.1 GeV

Table 3.8: Definition of BMU forward muons.

	CMIOCES
Region	Central
ID cuts	See Table 3.4
$E_{EM} + E_{HAD}$	> 0.1 GeV
Coverage	CES
Uniqueness	Not a CMUP, CMU, CMP, CMX, CMXMsKs

	CMIOPES
Region	Forward
ID cuts	See Table 3.4
$E_{EM} + E_{HAD}$	> 0.1 GeV
Coverage	PES
Uniqueness	Not a BMU

Table 3.9: Definition of stubless central (CMIOCES) and forward (CMIOPES) muons.

3.4 Electron and muon identification efficiency

Lepton efficiencies are measured using Drell-Yan events which have two electrons or two muons in the final state. Selected events have one identified lepton (the *tag*) and a second object (the *probe*) that passes the looser requirements, listed in Table 3.10. The tag and

probe leptons are required to have opposite charge and an invariant mass around the Z peak: $76 < M(ll) < 106$ GeV/c^2. The identification efficiency is measured for each lepton category l starting from the probe requirements and evaluating the probability to correctly identify the lepton. The lepton identification efficiency is:

$$\epsilon_{lep\ ID} = \frac{N_{TT}}{N_{TL}} \tag{3.4}$$

where N_{TT} is the number of events with two tag leptons and N_{TL} is the number of events with one probe and one tag. The lepton identification efficiencies used in this Thesis are taken from [101] and are around 80-90%.

Loose TCE
$E_T > 10$ GeV
$p_T > 5$ GeV/c
$\lvert z_0\rvert < 60$ cm

Loose PHX Trk
$E_{HAD}/E_{EM} \leq 0.05$
PEM $\chi^2_{3\times3} \leq 10$
PES $5 \times 9 \geq 0.65$
$\text{Iso}_{cal} \leq 0.1$
$\Delta R_{PES-PEM} \leq 3$

Loose PHX PEM
$E_T > 10$ GeV
$E_{HAD}/E_{EM} \leq 0.125$
$1.2 \leq \lvert\eta^{2D}_{PES}\rvert \leq 2.0$
Has a PHX track
$\text{N}^{\text{Si}}_{\text{hit}} \geq 3$
$\lvert z_0\rvert < 60$ cm

Loose CMUP/CMX
Stub: CMUP/CMX
$P_T > 10$ GeV/c
$\lvert z_0\rvert < 60$ cm

CMIOCES/CMIOPES
NCotHitsAx(5 hits) ≤ 3
NCotHitsSt(5 hits) ≤ 2
$P_T > 10$ GeV/c
$\lvert z_0\rvert < 60$ cm

Table 3.10: Definition of Loose requirements used to measure the identification efficiency for different lepton types.

3.5 Tau lepton: properties and reconstruction

Tau leptons have a very rich decay spectrum. In Figure 3.1 we list the dominant decay modes for tau leptons and the corresponding branching ratios. There are two groups of decay modes: leptonic and hadronic. Leptonically decaying taus appear as isolated electrons or muons accompanied by missing energy, while taus experiencing hadronic decay ($\tau \to X_h \nu$) leave a jet of particles (*tau-jet*) accompanied by a neutrino. Only charged leptons and the final state hadrons are visible. The visible decay products X_h can be $\pi^{\pm}/K^{\pm}$, or some short-lived intermediate resonances that decay directly to final states containing π, K. However, since the dimension of the detector and the shower sizes do not allow for particle-by-particle identification, in tau reconstruction all charged tracks are treated as $\pi^{\pm}$.

Tau hadronic decay channels can be further collected into two different decay modes: *one prong*, for τ decays into one charged hadron, and *three prong*, for decays into three charged hadrons. The total branching ratio for one prong τ decay is $(49.83 \pm 0.35)\%$ and $(14.38 \pm 0.24)\%$ for three prong decay, with a total of 64.21%. About 73% of all one-prong and 41% of all three-prong decays are associated with at least one π^0. Tau identification varies according to the analyses, but generally it consists of a set of rectangular cuts, which are applied to a selection of reconstructed tau candidates.

Decay Mode	Branching Ratio
"Leptonic Decay Modes"	
$\tau^- \to e\bar{\nu}_e\nu_\tau$	(18.01 ± 0.18) %
$\tau^- \to \mu\bar{\nu}_\mu\nu_\tau$	(17.65 ± 0.24) %
"One Prong Hadronic Decay Modes"	
0 neutrals:	
$\tau^- \to \pi^-\nu_\tau$	(11.7 ± 0.4) %
$\tau^- \to K^-\nu_\tau$	(0.67 ± 0.23) %
$\tau^- \to K^*(892)^-\nu_\tau$	(1.45 ± 0.18) %
Sum:	(13.82 ± 0.50) %
≥ 1 neutrals:	
$\tau^- \to \rho\nu_\tau \to \pi^-\pi^0\nu_\tau$	(25.2 ± 0.4) %
$\tau^- \to K^-\pi^0\nu_\tau$	(1.2 ± 0.5) %
$\tau^- \to a_1\nu_\tau \to \pi^-2\pi^0\nu_\tau$	(9.6 ± 0.4) %
$\tau^- \to \pi^-3\pi^0\nu_\tau$	(1.28 ± 0.24) %
Sum:	(37.28 ± 0.79) %
"Three Prong Hadronic Decay Modes"	
0 neutrals:	
$\tau^- \to a_1\nu_\tau \to \pi^-\pi^-\pi^+\nu_\tau$	(8.42 ± 0.31) %
≥ 1 neutrals:	
$\tau^- \to ... \to \pi^-\pi^-\pi^+ \geq 1\pi^0\nu_\tau$	(5.63 ± 0.22) %
$\tau^- \to (a_1\pi)^-\nu_\tau$	
$\tau^- \to \rho^0\pi^-\pi^0\nu_\tau$	
$\tau^- \to \rho^-\pi^-\pi^+\nu_\tau$	
$\tau^- \to \rho^+\pi^-\pi^-\nu_\tau$	
$\tau^- \to \omega\pi^-\nu_\tau$	(1.6 ± 0.4) %

Figure 3.1: List of dominant decay modes for tau leptons, with the corresponding branching ratios.

3.5.1 Reconstruction of hadronic taus

The *tau reconstruction* refers to the reconstruction of the visible decay products X_h of taus experiencing decays $\tau \to X_h\nu_\tau$. Tau particles can be reconstructed in the detector as narrow (due to the particle boost) jets, with low (1 or 3) track multiplicity plus an electromagnetic component from π^0s. At CDF both tracking and calorimeter information can be used for tau reconstruction. One can either look for isolated tracks pointing at hadronic energy in the calorimeter or search for a narrow cluster in the calorimeter matched to tracks.

In this analysis the tau reconstruction algorithm starts from a calorimeter cluster [102], requiring the calorimeter cluster to have at least one tower with total energy $E_T > 6$ GeV, and the highest E_T tower in the cluster is called a *seed tower*. All adjacent towers with $E_T > 1$ GeV are added to the seed tower to form the tau cluster. The size of the tau cluster is checked retaining for further reconstruction only clusters consisting of six

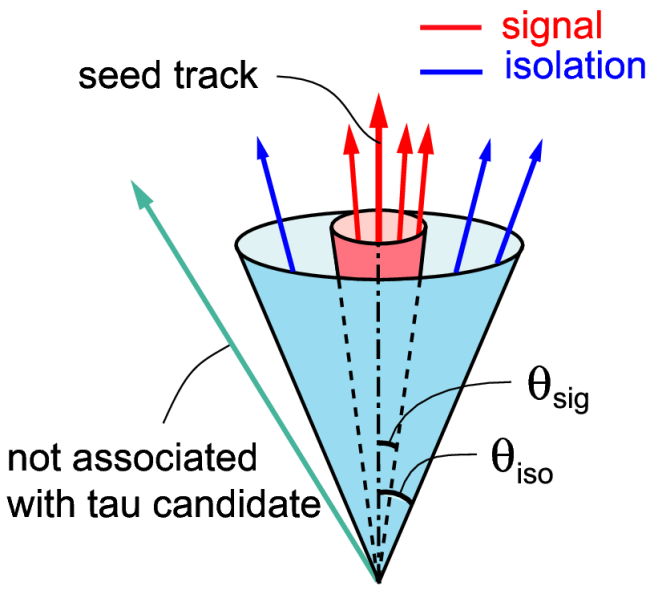


Figure 3.2: Tau-jet reconstruction scheme using track and π^0.

or fewer towers.
The candidate tau is reconstructed by matching the narrow calorimeter cluster with tracks. The one with the highest p_T is selected as the *seed track*, which is required to have $p_T > 4.5$ GeV/c. If there is no such track found the cluster is dropped. The seed track is used to define the signal cone, centered on it (see Figure 3.2 for a schematic drawing of the reconstructed tau cone) . The outer edge of the signal cone is defined by the opening angle θ_{sig}, function of the tau cluster energy E_{cal} associated to the candidate tau:

$$\theta_{sig} = \min(0.17, \max(\frac{5.0 \text{ rad/GeV}}{E_{cal}}, 0.05)) \tag{3.5}$$

All tracks in the signal cone with $p_T > 1$ GeV/c and z coordinate within 10 cm of the z-vertex of the tau jet, that also extrapolate to a tower in the cluster, are defined as *shoulder tracks* and associated to the candidate tau.
To a tau candidate is also associated an *isolation annulus*, with the inner radius equal to θ_{sig}, as defined above, and the outer radius $\theta_{iso} = 0.52$ radians. All tracks within the annulus are identified as *isolation tracks*. Requirements on the number and p_T of isolation tracks will specify the track isolation of the candidate tau, a powerful quantity in discriminating real tau-jet from fakes (see Section 3.6.2).

Similar to tracks, reconstructed π^0 with $E_T > 1$ GeV are associated to the candidate tau if inside a cone around the seed track with opening angle $\theta^{\pi^0}_{sig}$:

$$\theta^{\pi^0}_{sig} = \min(0.17, \max(\frac{5.0 \text{ rad/GeV}}{E_{cal}}, 0.1)) \tag{3.6}$$

Selected hadronic tau candidates are required to satisfy requirements on the energy of the seed calorimeter tower and on the seed track to ensure efficient reconstruction. The tau seed track is required to be fully contained in the volume covered by the COT,

and its trajectory extrapolation must point to a fully instrumented region of the CES detector in the z-direction, $9 < |Z_{CES}^{seed\ trk}| < 230$ cm, to ensure efficient π^0 reconstruction. The selection criteria used in this analysis for tau reconstruction are listed below:

- Transverse energy of the seed calorimeter tower, $E_T^{seed\ twr} > 6.0$ GeV
- Opening angle of the signal cone: $0.05 < \theta_{sig} < 0.17$
- Opening angle of the isolation cone: $\theta_{iso} = 0.52$ rad
- Seed track Z_0: $|Z_0^{seed\ trk}| < 60$ cm
- Impact parameter of the seed tracks: $|d_0^{seed\ trk}| < 0.2$ cm
- CES detector fiduciality [3]: $9 < |Z_{CES}^{seed\ trk}| < 230$ cm

Some requirements about the quality of the track are also specified:

- Seed track quality: ≥ 2 stereo and ≥ 3 axial COT segments with more than 6 hits.

The π^0 reconstruction plays an important role in the tau reconstruction. Since in some cases there is a non-negligible energy loss due to π^0 reconstruction inefficiency, a big effort has been spent to correct the tau four-momentum for missing reconstructed π^0 and to increase the acceptance for hadronic tau decays [102]. Therefore, tau reconstruction is affected by big uncertainties due to misidentified particles, such as neutrinos, and possible misreconstructed π^0. Corrections for π^0 mis-reconstruction can additionally improve the tau reconstruction acceptance, they are not applied in this analysis, however they will be discussed in Section 3.6.4 .

3.5.2 Reconstruction of π^0 from τ hadronic decay

π^0 particles from τ hadronic decay are identified exploiting the $\pi^0 \to \gamma\gamma$ decay, and their reconstruction involve the resolution of not-isolated π^0/γ, such as inside of the tau-jet where the decay products could be very close to each others.

π^0s are reconstructed at CDF in the central region, exploiting the informations from CES and CEM detectors to measure position and energy respectively. A collection of candidates π^0 is created by assigning each CES cluster the corresponding CEM tower energy. The π^0 is assigned the full electromagnetic calorimeter energy of the CEM tower corrected for the expected deposited energy from all charged tracks traversing this tower. The track energy deposition parameterization, in terms of the magnitude of the momentum, is obtained from data by studying the calorimeter response to isolated charged pions [103].
In the rest frame of the π^0 the 2 photons from $\pi^0 \to \gamma\gamma$ decay have an energy $m_{\pi^0}c^2/2 = 67.5$ MeV. The minimum separation in the $\eta - \phi$ plane is given by $S \sim R \cdot 1/\gamma_{boost} \sim R \cdot m_{\pi^0}/E_{\pi^0}$, where γ_{boost} is the boost of the π^0 and R is the range distance of the CES from the interaction point. For a π^0 of an energy of ~ 10 GeV the separation between the two photons is of the order of a couple of centimeters. As seen in Section 2.2.2, the

[3] The requirement to have the extrapolated seed track pointing to the CES detector (CES coverage is $9 < |Z_{CES}| < 216$ cm).

deposited CES energy of an electron or photon is fully contained in a cluster 15-20 cm wide, so the two photon showers from π^0 can have a significant overlap [78] and are almost always too close to be distinguished from single isolate photon showers. Only for π^0 with energy $E_{\pi^0} < 10$ GeV the photons can be resolved in some cases, except when the two photons have a very large energy asymmetry. When the two showers are close to each other they can get merged and reconstructed as a single shower, however this has little effect on the measured parameters of a reconstructed tau candidate. π^0 reconstruction [78] used in this Thesis refers to the CES showers as reconstructed π^0, either it is a true π^0 or a single photon coming from either a $\pi^0 \to \gamma\gamma$ decay or from any other source.
The π^0 reconstruction can limit the accuracy of the tau reconstruction, caused either by dead wires in the CES or by the overlap of a π^0 meson with a charged track. However the comparison between the reconstructed shower energy and tracks momenta can be exploited to correct tau momentum for π^0 mis-reconstruction (see Section 3.6.4).

3.6 Tau identification

Several tau related quantities are defined and a set of cuts has been chosen to be efficient for tau's identification, starting from a sample of τ candidates that have already passed the τ reconstruction selection defined in the previous section.
Tau reconstruction is a set of very loose requirements to the objects measured in the detector, reconstructed as tau candidates. Therefore, electrons, muons and jets can also get reconstructed as tau candidates, contributing as fake taus, as will be discussed in Section 3.6.2. Tau identification criteria are intended to be effective in rejecting background fakes and separate them from hadronic decay of real τ leptons.

3.6.1 Tau related variables

A sample of simulated taus in MC generated events has been used to set cuts for tau identification and study their efficiency, since, as was shown in [104] most of the tau identification (ID) variables are well modeled by MC simulation or corrected to account for differences between the data and MC distributions.

Track multiplicity: $N_{trks}^{sig\ cone}$ cut

The tau multiplicity is defined as the number of tracks (seed and shoulder tracks) in the signal cone. Since τs decay hadronically into one-prong or three-prongs, it follows that:

- $N_{trks}^{sig\ cone} = 1$ (1-prong) or 3 (3-prong)

Tau momenta reconstruction: visible p_T and mass cuts

Tracks and π^0s in the signal cone are used to construct the four-momentum of the hadronic system. Since the hadronic calorimeter has a poor energy resolution compared to that of the electromagnetic calorimeter or to the tracking p_T resolution (see Section 2.2.1 and 2.2.2), energy measurement based on reconstructed tracks and π^0s have been preferred to pure calorimetry. The four-vector momentum of a tau, p^τ , is defined as the

sum of four-vectors of tau tracks with $p_T > 1$ GeV/c and π^0s with $E_T > 1$ GeV (both are assumed massless) associated with the tau candidate:

$$p^\tau \equiv \sum_{\Delta\Theta<\theta_{sig}} p^{trks} + \sum_{\Delta\Theta<\theta^{\pi^0}_{sig}} p^{\pi^0} \tag{3.7}$$

where p^{trks} and p^{π^0} denote four-vector momenta of contributing tracks and π^0 candidates, and $\Delta\Theta$ is the angle between the seed track and a particular track or π^0 candidate. Such momentum is often called as *visible momentum* of tau candidate as it does not include neutrino contribution.
In Figure 3.3 is shown the p^τ_T distribution for 1-prong and 3-prong τ separately, where the p_T of the tau is defined as the scalar sum of the p_T of the tracks in the signal cone and the E_T(EM) of the π^0 candidates as measured in the CEM:

$$p^\tau_T = p_T(trks + \pi^0) = \sum_{trks} p^{trks}_T + \sum_{\pi^0 s} E_T(EM) \tag{3.8}$$

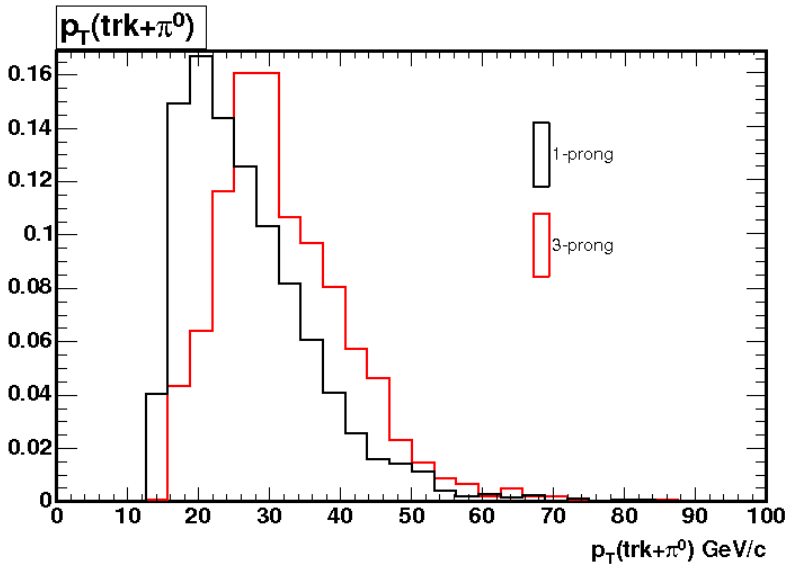


Figure 3.3: $p_T(trk+\pi^0)$ distribution for 1-prong and 3-prong τ decays. A sample of MC simulated events with hadronic taus has been used.

Due to the presence of neutrinos in the tau decay, the p_T measured using only tracks and π^0 is an underestimation of the real tau p_T. This results in a reduced kinematical acceptance for tau leptons compared with electrons and muons. In addition, the effect is even stronger for 1-prong decays compared to 3-prong, due to the π^0 reconstruction inefficiency and the large fraction of 1-prong decays accompanied by π^0's. To increase the acceptance for event selection with hadronic τs, a looser cut on tau p_T is applied to 1-prong reconstructed τ with respect to the cut on 3-prongs.
The following cuts on the tau visible p_T are applied for tau identification:

$$\begin{aligned} p_T(trks + \pi^0) &> 15 \text{ GeV/c} \quad \text{for} \quad N_{trk} = 1 \\ p_T(trks + \pi^0) &> 20 \text{ GeV/c} \quad \text{for} \quad N_{trk} = 3 \end{aligned} \tag{3.9}$$

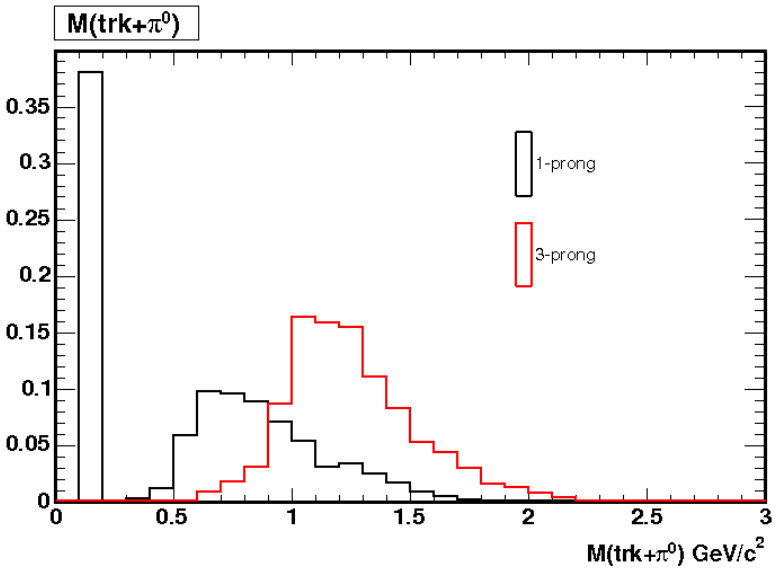


Figure 3.4: M(trks $+\pi^0$s) distribution for 1-prong and 3-prong τ decays. A sub-sample of MC simulated events with hadronic tau has been used.

The four-momentum is used for the determination of the invariant mass of the system, $M(trks\ +\pi^0\text{s})$. In Figure 3.4 in shown the $M(trks\ +\pi^0\text{s})$ distribution for 1-prong and 3-prong τs separately. The larger tails on the $M(trks+\pi^0)$ distribution for 3-prong with respect to 1-prong reconstructed taus is a consequence of the larger reconstructed p_T for 3-prong with respect to 1-prong taus. For 1-prong tau without any identified π^0 the mass is equal to the $\pi^\pm$ mass (140 MeV/c^2). Therefore the bin content corresponding to $\pi^\pm$ mass in the 1-prong mass distribution is not affected by π^0 reconstruction inefficiency. The standard tau identification applies the following cuts for the visible invariant mass, in order to maintain about the same cut efficiency for the 1-prong and 3-prong decay mode:

$$\begin{aligned} M(trks+\pi^0) &< 1.8\ \text{GeV/c}^2 \quad \text{for} \quad N_{trk}=1 \\ M(trks+\pi^0) &< 2.2\ \text{GeV/c}^2 \quad \text{for} \quad N_{trk}=3 \end{aligned} \tag{3.10}$$

A cut near the actual τ mass (1.78 GeV/c^2) has good efficiency for τ's: 97% and 90% cut efficiency for 1-prong and 3-prong respectively.

Tau calorimetric cluster: E_T^{cal} cut

In MC simulation the calorimeter energy deposit is not well modeled for τ due to a difference in lateral showering of calorimetric clusters between data and MC. This mis-modelling is more enhanced in hadronic showers: hadronic showers in MC simulation are significantly narrower than in the data, and the energy deposit in the calorimetric towers associated to the clusters is consequently higher, since the same amount of particle energy is distributed over a fewer number of calorimetric towers in MC simulation. This affects mostly tau clustering, since the clusters are made checking on the energy of the towers (only towers with $E_T > 1$ GeV are used), while all surrounding towers

contribute to the isolation energy. A correcting factor has been applied in this analysis to MC events, a 10% increase of the tau hadronic energy.

Using a sample of taus in MC simulated events, it has been studied the difference between the reconstructed tau cluster energy and the *visible energy* E^{gen} of the tau lepton, defined as its energy at generation in MC events subtracted by the energies of all neutrinos from its decay. Full tau reconstruction and identification cuts are applied to the candidate tau. In Figure 3.5 is shown the scatter plot between $(E^{cal}/E^{gen} - 1)$ and the E^{gen}. The black points correspond to the case with the 10% increase is not applied.

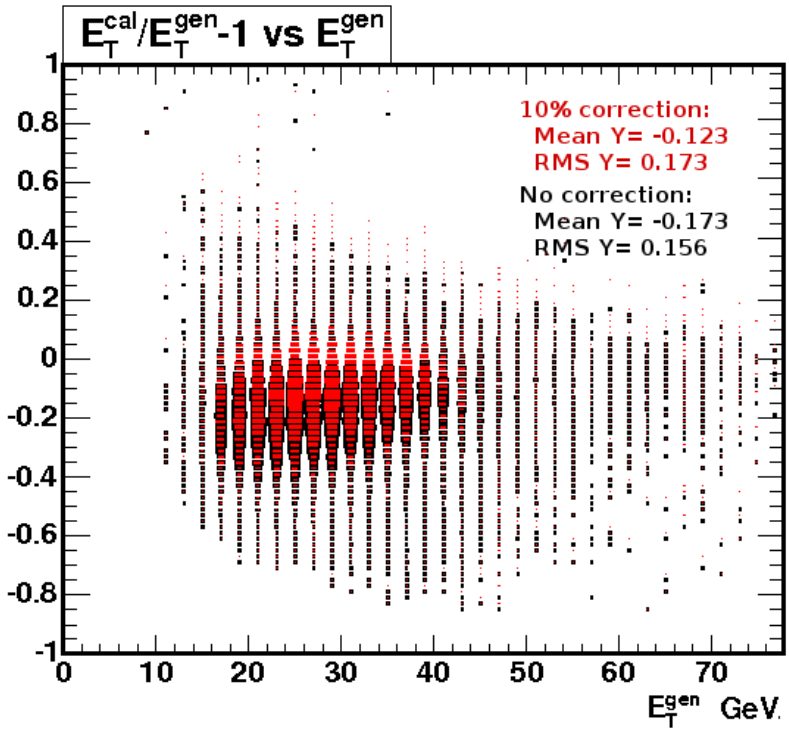


Figure 3.5: The scatter plot $(E_T^{cal}/E_T^{gen} - 1)$ VS E_T^{gen}. Red points correspond to the case with the 10% correction applied to the hadronic component of the cluster energy. The black points correspond to the case with no correction applied.

It can be seen that $(E_T^{cal}/E_T^{gen} - 1)$ distribution doesn't depend on the energy of the tau candidate. Therefore, the discrepancy between E_T^{cal} and E_T^{gen} can be accounted applying a scale factor. The 10% correction to the hadronic component of the cluster energy (red point in Figure 3.5) reduces the discrepancies, however, from Figure 3.5 the calorimetric energy still underestimates the real tau energy by about 12%. This can be explained as an effect of π^0 mis-reconstruction, as will be seen in Section 3.6.3. To better understand the effect of this correction, the 1-prong and 3-prong reconstructed tau's have been considered separately. In Figure 3.6 are shown the distribution in $(E_T^{cal} - E_T^{gen})/E_T^{gen}$ for 1-prong and 3-prong reconstructed taus. Again, the black distribution correspond to the case with no correction applied and the red one to the case with 10% shift in hadron energy. The mean value of the total calorimetric energy increases of about 32% for 1-prong tau's and 24% for 3-prong applying 10% shift. The residual underestimation on the total energy is of about 11% for 1-prong and of about 15% for 3-prong tau's.

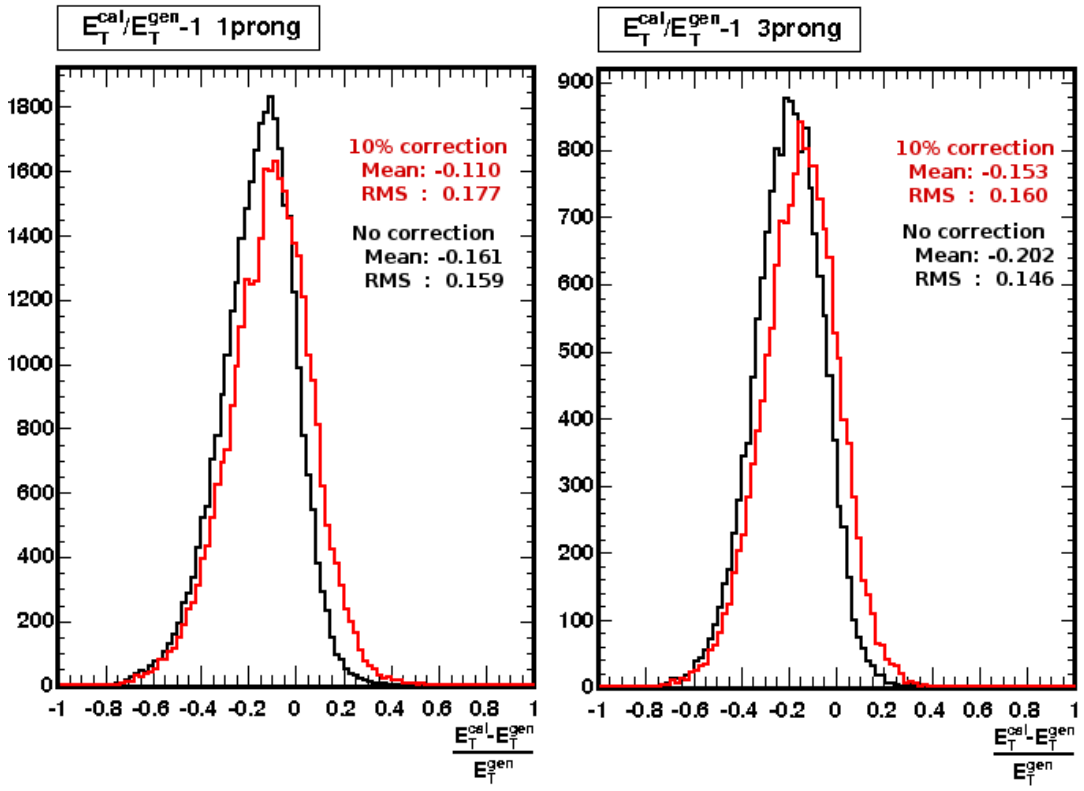


Figure 3.6: The distribution in the difference between the energy of the tau cluster E_T^{cal} of reconstructed tau and the E_T^{gen}, for 1-prong (left) and 3-prong tau's (right). The red curve correspond to the case with the 10% correction applied to the hadronic component of the cluster energy. The black curve correspond to the case with no correction applied.

In this analysis the 10% correction to the hadronic energy is applied for reconstructed tau in simulated events. The residual discrepancy in the cluster energy between data and MC will be taken into account in the total scale factor for the tau identification efficiency (Section 4.3.1).

3.6.2 Fake taus

Jets, electrons, and muons can pass tau identification cuts. Therefore, explicit vetoes for electrons and muons are included in the tau identification cuts while the isolation requirement helps in reducing fakes from jets, as it will be discussed in the follow.

Jet to tau fake: Isolation requirement against QCD jets

Isolation plays an important role in tau identification, and it is the most powerful cut against QCD jets. Isolation quantities can be defined using information from either calorimeters or tracks plus reconstructed π^0. However, tracking isolation has been found as the most powerful cut against jets faking taus. The track isolation is defined as the scalar sum of tracks around the seed track but outside the signal cone in $\eta - \phi$ space:

$$I_{trks} = \sum_{(\theta_{sig}<\Delta R<0.52)} p_T^{trks} \tag{3.11}$$

The tracks that define the track multiplicity associated with the tau cluster, the seed track and the shoulder tracks, are excluded from the sum.
In a similar way it is defined the tau candidate π^0 isolation as:

$$I_{\pi^0} = \sum_{(\theta^{\pi^0}_{sig}<\Delta R<0.52)} p_T^{\pi^0} \tag{3.12}$$

Figures 3.7 (top) and 3.8 (top) show the tracking and π^0 isolation (I_{trks} and I_{π^0}) respectively, both calculated for 1- and 3-prong reconstructed τs in a sample of MC simulated τs and in a sample of MC simulated QCD jets passing the τ identification requirements (fake taus). Jets are expected to be wider than tau-jets, indeed both I_{trks} and I_{π^0} distributions obtained in the jets (fake taus) sample show a bigger left tail than the distributions obtained from the real tau sample. For a given cut on I_{trks} or I_{π^0}, the *fake tau rejection* is defined as the number of reconstructed fake taus in the jet sample contributing with a larger I_{trks} or I_{π^0} value, thus failing the cut, while the *real tau acceptance* is defined as the number of real tau with a I_{trks} or I_{π^0} lower than the cut value. The distribution of real tau acceptance as a function of the cut and the distribution of fake tau rejection versus real tau acceptance are shown for I_{trks} and I_{π^0} cut in Figures 3.7(bottom) and 3.8(bottom) respectively. A compromise between fake tau rejection and real tau acceptance is reached setting the following cut for tracking and π^0 isolation:

$$\begin{aligned} I_{trk} &< 2.0 \text{ GeV/c} \\ I_{\pi^0} &< 1.0 \text{ GeV/c} \end{aligned} \tag{3.13}$$

The chosen cuts for both I_{trks} and I_{π^0} determine the same fake tau rejection for 1- and 3-prong reconstructed taus, since the distributions from fake taus do not show differences in the two decay modes. The $I_{trks} < 2$ GeV/c cut rejects about 65% of the fake taus, while the real tau acceptance is about 83% and 90% for 1-prong and 3-prong respectively. The $I_{\pi^0} < 1$ GeV/c cut, instead, rejects about 25% of the fake taus, while the real tau acceptance is about 92% and 94% for 1-prong and 3-prong respectively.

Electron removal cut

Electrons produce very narrow calorimetric clusters with one track pointing at them and can consequently be mistaken with τs hadronically decayed with one charged track. The best separation of electron faking taus from real taus relays on the hadronic energy associated with the taus.
Taus decaying hadronically are unlikely to have an electromagnetic fraction close to 1, and since π^0 are often present among the tau decaying products, the ratio E/P, where P is the tau momentum reconstructed only with charge tracks, is expected to be greater than 1. A tau cluster containing only charged pions, for example, has ideally an EM fraction equal to 0 and $E/P = 1$. Therefore, from the scatter plot of the electromagnetic energy fraction E_{EM} versus E/P electrons and tau leptons can be separated, and a cut in this plane can reject electrons efficiently. The variable ξ' is designed to remove electrons, and is defined as:

$$\xi' = \frac{E_{tot}}{\sum |\vec{p}|}(0.95 - \frac{E_{\text{EM}}}{E_{tot}}) \tag{3.14}$$

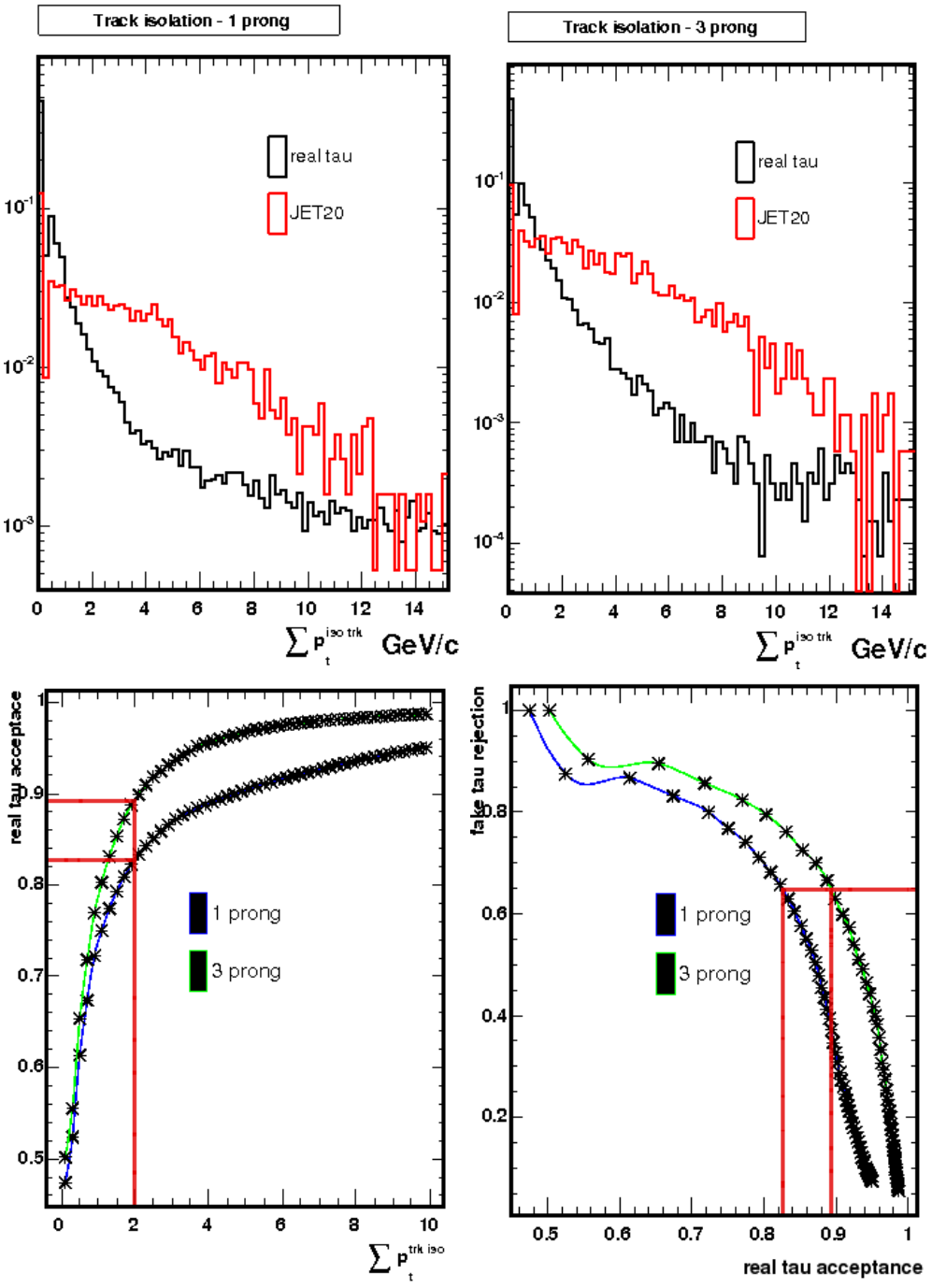


Figure 3.7: Tracking isolation I_{trks} for reconstructed 1-prong (top left) and 3-prong (top right) tau's. Normalized distributions for real taus, using a sample of reconstructed τ leptons in MC simulated $H \rightarrow WW$ signal events, and for fake taus in MC simulated QCD jet events are compared. Bottom left: Real tau acceptance as a function of the tracking isolation $\sum p_T^{iso\ trk}$ cut. The $I_{trk} < 2$ GeV/c applied cut corresponds to about 83% of 1-prong and about 90% of 3-prong real taus. Bottom right: Distribution of the fake tau rejection versus real tau acceptance. 65% of the fake taus are rejected. Each point corresponds to a cut on I_{trks}.

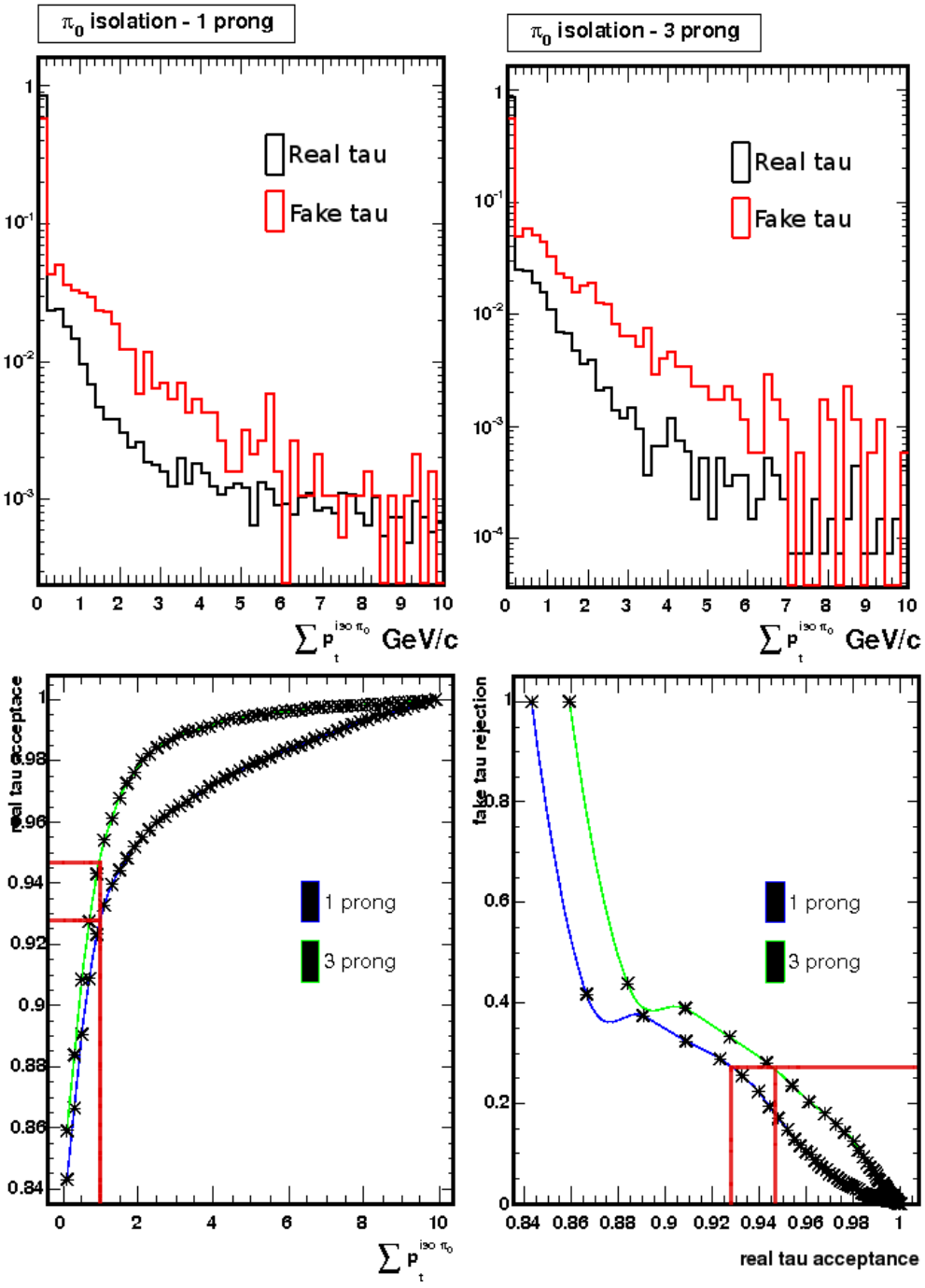


Figure 3.8: π^0 isolation I_{π^0} distributions for reconstructed 1-prong (bottom left) and 3-prong (bottom right) taus. Normalized distributions for real taus, using a sample of reconstructed τ leptons in MC simulated $H \to WW$ signal events, and for fake taus in MC simulated QCD jet events are compared. Bottom left: Real tau acceptance as a function of the tracking isolation $\sum p_T^{iso\ \pi^0}$ cut. The $I_{\pi^0} < 2$ GeV/c applied cut corresponds to about 92% of 1-prong and about 95% of 3-prong real taus. Bottom right: Distribution of the fake tau rejection versus real tau acceptance. 25% of the fake taus are rejected. Each point corresponds to a cut on I_{π^0}.

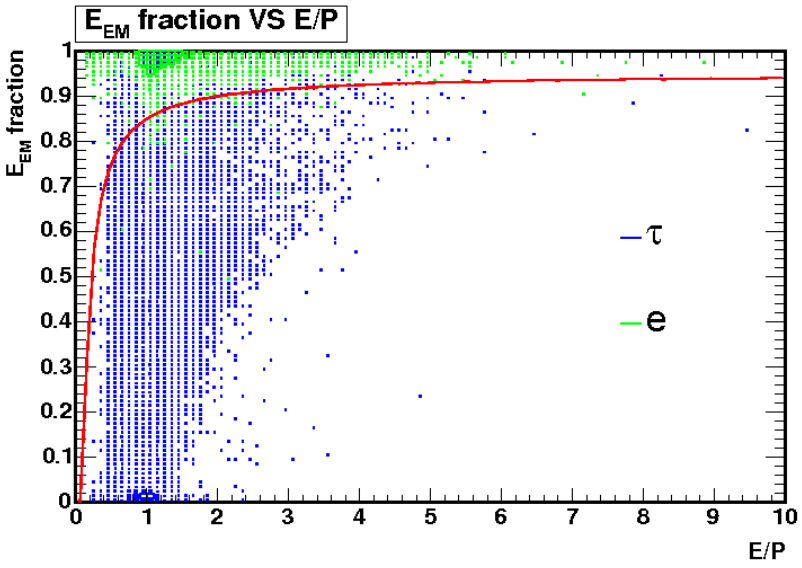


Figure 3.9: Electromagnetic energy fraction versus E/P for taus (blue) and electrons (green), from a sample of $Z \to \tau\tau$ and $Z \to ee$ MC events. Entries with $\xi' > 0.1$ are in the region below the solid red line.

Figure 3.9 shows the electromagnetic energy fraction vs E/P for electrons and hadronic taus. Fluctuations in the electromagnetic calorimeter energy measurement and strong Bremsstrahlung processes involving electrons can led to an electromagnetic fraction and E/P different from 1.0. If electrons emit high p_T photons, the electron track p_T has lower momentum and $E/P > 1.0$. In addition, high-energy electron showers may leak into the hadronic calorimeter, causing the EM fraction to be less than 1.0.
The electron removal cut is then defined as:

$$\xi' > 0.1 \tag{3.15}$$

The solid red line in Figure 3.9 corresponds to events with $\xi' = 0.1$ and only events below this curve are accepted after the cut.

Muon removal cut

The muons reconstructed as taus share many characteristics with the hadronically decaying taus: it is likely to be isolated and have a tendency to deposit its energy in the hadronic section. Since the requirement on the seed tower E_T for a tau cluster is only 6 GeV it is not unlikely that a reconstructed tau cluster is instead created by a muon, even if it is less likely than for electrons. However, tau clusters are rejected either if there is a track stub in a muon chamber matched to the tau clusters or if the energy deposition in the CEM and the CHA are consistent with that of a minimum ionizing particle.
In Figure 3.10 is shown the $E^{cal}/\sum p_T^{trks}$ distribution from a sample of simulated muons passing the taus identification requirements and being reconstructed as fake taus.
The muon removal cut is then defined as:

$$E^{cal}/\sum p_T^{trks} > 0.8 \tag{3.16}$$

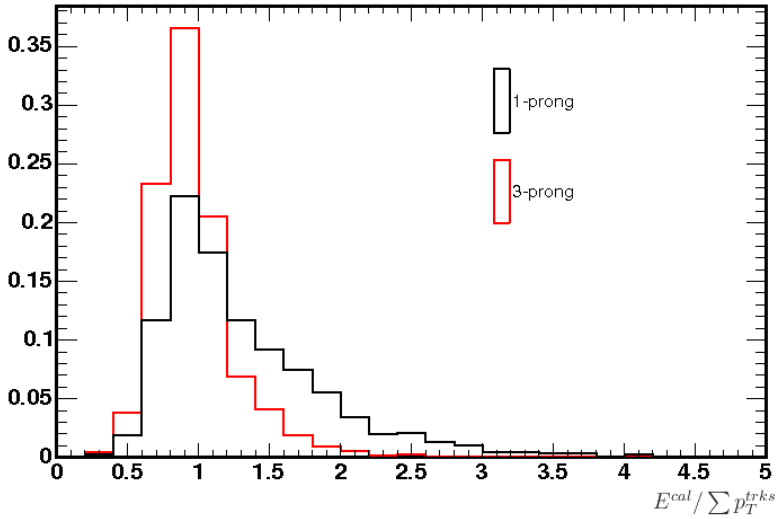


Figure 3.10: $E^{cal}/\sum p_T^{trks}$ distribution for MC taus distinguishing between 1-prong (black) and 3-prong (red) τs.

where p_T is reconstructed with only charged tracks.

3.6.3 Tau identification efficiency and optimization

The selection criteria implemented in this Thesis for tau identification are listed below:

- Transverse energy of the shoulder calorimeter towers $E_T^{sh\ twr} > 1.0$ GeV
- Number of towers in the τ cluster $N^{twr} \leq 6$
- Tau calorimetric cluster energy $E^{cal} > 10$ GeV
- Transverse momentum of the seed track $p_T^{seed\ trk} > 6.0$ GeV/c
- Transverse momentum of the shoulder tracks $p_T^{sh\ trks} > 1.0$ GeV/c
- Number of shoulder tracks $N_{trk}^{sig\ cone} = 1$ (1-prong) or 3 (3-prong)
- The charge of the tau candidate $|Q_\tau| = |\sum_{\tau\ tracks} Q^{trks}| = 1$
- Visible momentum $p_T(\text{trks}+\pi^0) > 15$ GeV/c (1-prong) or 20 GeV/c (3-prong)
- Visible mass $M(\text{trks}+\pi^0) < 1.8$ GeV/c^2 (1-prong) or 2.2 GeV/c^2 (3-prong)
- Track and π^0 isolation requirements $I_{trks} < 2$ GeV/c and $I_{\pi^0} < 1$ GeV/c
- Electron removal $\xi' > 0.1$
- Muon removal $E^{cal}/\sum p_T^{trks} > 0.8$

The τ identification efficiency used in this Thesis are documented in [105]. Some ID requirements are differently applied in the 1-prong and 3-prong cases, for the different distributions in the τ ID variables and in order to maintain the same cut efficiency. The efficiencies of all cuts for τ identification are reported in Table 3.11 and studied on τ leptons from $W \to \tau\nu$ MC events, by matching generated τs to the corresponding reconstructed τs. The following minimal kinematical requirements are applied to the hadronic tau: $|\eta| < 1.2$ and $p_T(\text{trks}+\pi^0) > 15$ GeV/c. The 10% correction on the hadronic component of the candidate tau energy has been applied.

	1-prong	3-prong
$N_{twr} \leq 6$	100%	100%
$E_T^{sho\ twr} > 1$ Gev	99.9%	99.9%
$E_T^{\tau\ clu} > 10$ GeV	99.2%	99.4%
$\sum Q_{sig}^{trk} = 1$	100%	100%
$M(trks + \pi^0) < 1.8$ for 1-prong (2.2 for 3-prong)	97.0%	90.0%
$\xi' > 0.1$	81.5%	87.6%
$\sum E_T^{iso\ \pi^0} < 1$ GeV	92.0%	94.0%
$\sum p_T^{iso\ trk} < 2$ GeV/c	83.0%	90.0%
$E^{cal}/\sum p_T^{trks} > 0.8$	99.0%	99.9%
Total efficiency	59%	66%

Table 3.11: Fraction of events passing the cut for events that pass the all τ reconstruction requirements. 1 and 3 prong tau decay modes are considered separately.

The major contributions to the tau identification inefficiency come from the isolation requirements and the invariant mass cut. The $N_{trks}^{\tau} = 1, 3$ requirement is mostly affected by the migration of events from 3-prong bin into the excluded 2-prong bin as the effect of the track reconstruction inefficiency. Another possible effect that is relevant for the $N_{trks} = 1, 3$ requirement is related to reconstructed tracks from underlying events and secondary interactions accidentally falling into the signal cone, causing a fraction of events to migrate from 1 and 3 prong bins into even prong bins. The π^0 isolation requirement is affected by the CES resolution, while the inefficiency for $M(trks + \pi^0)$ requirements is due to lost π^0s, thus not included in the tau visible mass calculation. However, the effect on the final result of the tau lepton identification will be taken into account as systematic uncertainty.

Tau momentum correction for π^0 mis-reconstruction

As already seen, π^0 reconstruction inefficiency may affect the tau reconstruction and in particular the tau four-momentum. Corrections to the measured tau energy from tracks and π^0's, can be determined detecting cases of excessive electromagnetic energy not assigned to any π^0. Those corrections are discussed in this section, although they are not applied in this analysis.
If the energy of the reconstructed shower is significantly larger than the track momentum, it is possibly the case of an overlapping track and π^0, in which the neutral particle does not get reconstructed. The π^0 correction algorithm makes attempts to exploit

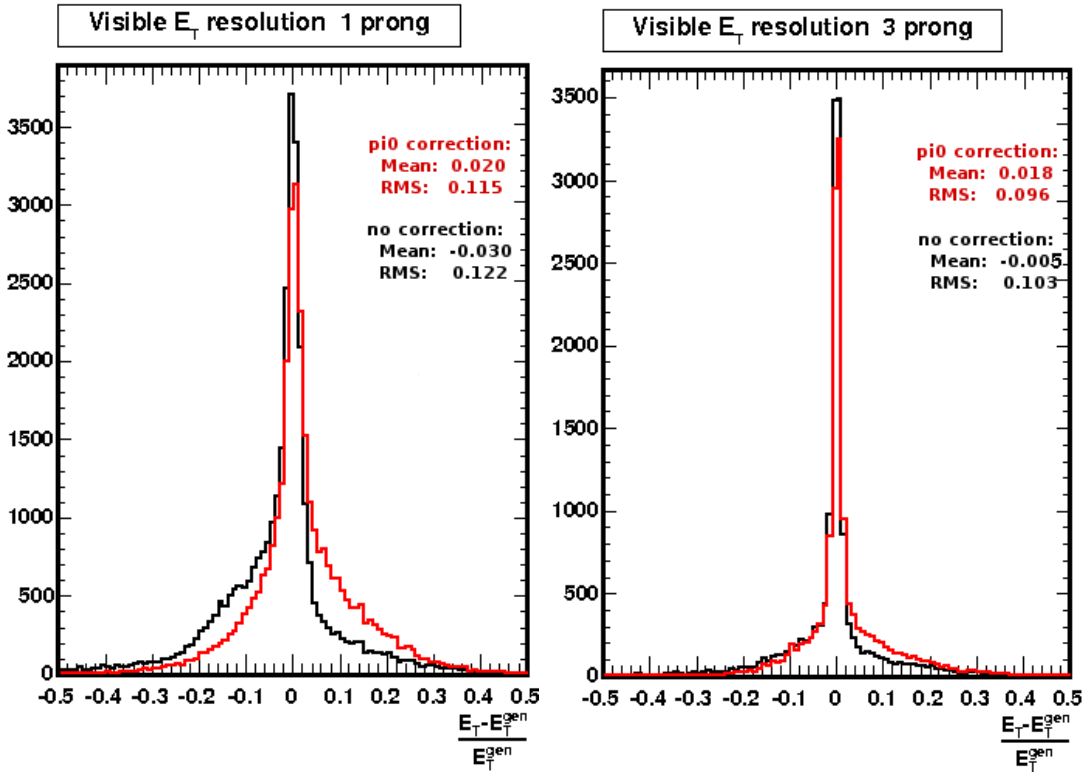


Figure 3.11: Energy resolution ($\frac{E_T - E_T^{gen}}{E_T^{gen}}$) plots for 1-prong taus (left) and 3-prong tau (right). The black line correspond to the resolution of the visible $E_T(\text{trks}+\pi^0)$, while the red line correspond to the reconstructed E_T^{rec} according to the π^0 correction algorithm output.

calorimetry/tracking information to estimate the energy of the mis-reconstructed π^0, determining which hypothesis would reliably yield the τ's energy, among $E(\text{trks}+\pi^0)$, E^{cal} or E^{corr}, defined as:

$$E^{corr} = E_{EM}^{cal} + \sum_{i=1}^{N_{trk}} (p_i - 0.32 \text{ GeV}) \tag{3.17}$$

E^{corr} is determined under the assumption of MIP-like $\pi^{\pm}$, when a $\pi^{\pm}$'s deposition in the electromagnetic calorimeter is taken to be a constant 0.32 GeV.
Corrections are applied when there is a substantial electromagnetic energy in the cluster ($E_{EM}/E_{tot} > 0.2$). First, the algorithm needs to decide wether a π^0 was mis-reconstructed, checking the eventually disparity between E^{corr} and $E(\text{trks}+\pi^0)$. If the π^0 is well reconstructed E^{corr} is equal to $E(\text{trks}+\pi^0)$, within detector resolutions, therefore $E(\text{trks}+\pi^0)$ remains the best estimation for the tau energy. If $E^{corr} \neq E(\text{trks}+\pi^0)$ the algorithm assigns E^{corr} when $\pi^{\pm}$ are MIPs, comparing E^{corr} to the E^{cal}, or E^{cal} when the the MIP-like hypothesis is not satisfied and thus for cases of potentially large hadronic energy contribution $E^{cal} > E(\text{trks}+\pi^0)$.

If a π^0 was mis-reconstructed, a new π^0 is assigned to the list of π^0's associated with the tau candidate, with direction coinciding with the hadronic tau candidate. In Figure 3.11 is shown the change in energy resolution $(E_T^{rec} - E_T^{gen})/E_T^{gen}$, using a sample

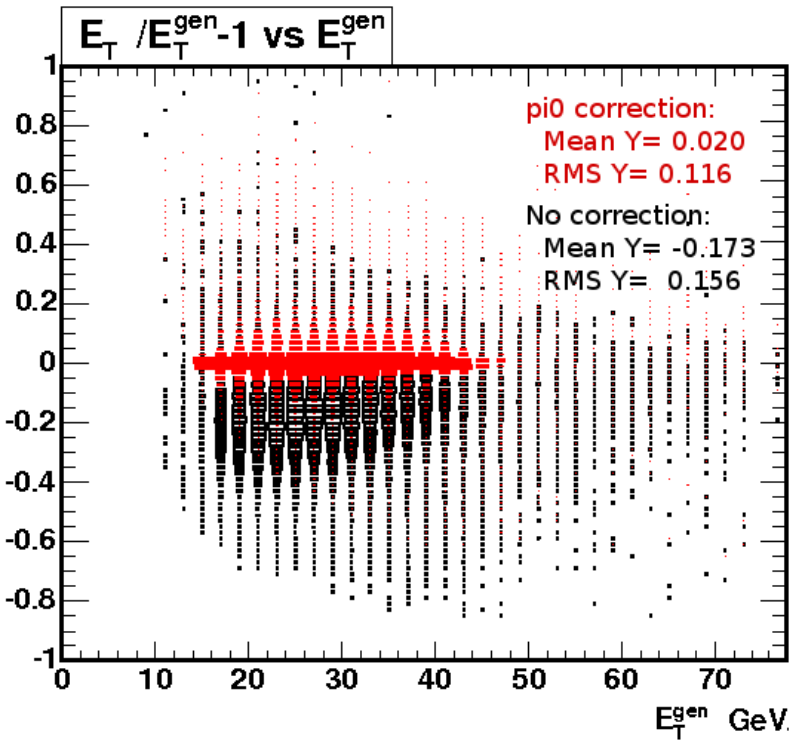


Figure 3.12: Scatter plot $(E_T - E_T^{gen})/E_T^{gen}$ VS E_T^{gen}. The black points correspond to the calorimetric energy resolution ($E_T = E_T^{cal}$), while the red points correspond to the E_T after the algorithm for π^0 correction is applied ($E_T = E_T^{rec}$).

.

of reconstructed taus in MC generated events, after the energy correction algorithm is applied. E_T^{rec} is the reconstructed tau energy, according to the algorithm output, and E_T^{gen} is the energy of the generated tau, to which the momenta of neutrinos from its decay are subtracted out. The set of standard identification cuts, listed above, are applied and the 1-prong and 3-prong cases are considered separately. The π^0 correction affects mostly the 1-prong tau, where π^0s carry a larger fraction of energy.
In some cases the calorimetric energy E^{cal} does not reproduce correctly the tau visible energy, however comparing the E_T^{rec} from the algorithm and the E_T^{cal} using only calorimetric information in Figure 3.12 the recovering of missing π^0 supplies a tau energy reconstruction with higher resolution with respect to the calorimetric energy.

3.7 Jet identification

The jet reconstruction algorithm used in this analysis, called JETCLU [106], relies only on calorimeter information. It relies on the assumption that particles close in space are originated from the same parton and particles which lie in a circle of given radius $\Delta R = 0.4$ in $\eta - \phi$ space are merged together. JETCLU starts considering all the calorimeter towers with $E_T = E_{EM} \sin\theta_{EM} + E_{HAD} \sin\theta_{HAD} > 1$ GeV, where E_{EM} and E_{HAD} are the energies of the electromagnetic and hadronic calorimeter towers respectively, and θ_{EM} , θ_{HAD} are the polar angles of the vectors pointing from the primary vertex

to the geometrical center of the electromagnetic (hadronic) towers. Starting from the highest E_T tower, adjacent towers within ΔR are clumped together to form a *pre-cluster*. Subsequently the energy-weighted center of the jet cone is calculated, and a new cluster is built around the new centroid. After the first iteration all towers with energy greater than 0.1 GeV are considered and added to each pre-cluster if within $\Delta R = 0.4$ from the new energy-weighted centroid of the pre-cluster. A new centroid for the cluster is recomputed each time that a new tower is added. This procedure is iterated until the stability of the jet configuration is reached.
Since jet candidates may share some amounts of transverse energy, they are merged if the shared E_T is more than 75% of the less energetic E_T jet. The jet centroid is then recalculated. Otherwise, shared E_T is assigned to the closest jet candidate. The four-momentum of the jet is then computed from its towers and final clusters are commonly referred to *raw jets*. Particles are specified by massless four-vectors with directions (η, ϕ), the jet four-vector is then defined by:

$$E_T^{jet} = \sum_p E_T^p \tag{3.18}$$

$$\eta^{jet} = \frac{1}{E_T^{jet}} \sum_p E_T^p \eta_p \tag{3.19}$$

$$\phi^{jet} = \frac{1}{E_T^{jet}} \sum_p E_T^p \phi_p \tag{3.20}$$

In this analysis a jet is defined as a calorimeter cluster of size $\Delta R < 0.4$ which has a total corrected transverse energy $E_T > 15$ GeV and is within pseudo-rapidity of $|\eta| < 2.5$. Jets are also required to be away (with $\Delta R \geq 0.4$) from identified leptons in the final state. Corrections applied to the raw detector energy are discussed in the following subsection.

Jet energy corrections

The 4-momentum assigned to a measured jet suffers from non-uniform and non-linear response of the calorimeter, from particles not involved in the hard interaction accidentally falling in the jet and the jet energy lost out of the jet cone. A set of corrections to the jet reconstructed energy (*raw energy*) have been developed to account for these limits and scale it back to the primary parton momenta.

- η-dependent correction. This corrects for non-uniformities in calorimeter response as a function of pseudo-rapidity. This dependence on η arises from the different response of the calorimeter in the central and the forward regions, and from the separation of calorimeter components at $\eta = 0$, where the two halves of the central calorimeter join, and at $\eta \sim 1.1$, where the plug and central calorimeters merge. The η-dependent corrections $f_\eta(R, p_T, \eta)$ are obtained by requiring that the two leading jets in dijet events are balanced in p_T in absence of hard QCD radiation, making the jet energy response uniform across the detector.

- Multiple $p\bar{p}$ Interaction corrections. This correction takes into account the extra energy entering the jet cone due to multiple interactions (M$p\bar{p}$I) in the event. At high luminosity more than one $p\bar{p}$ interaction occur in average in the same bunch

crossing. The jet energy would increase if final state hadrons of those extra interactions accidentally overlap with the jets. A good indicator of the number of additional interactions in the same bunch crossing is the number of the reconstructed primary vertices in the same event. The transverse energy in a random cone is measured in minimum-bias data and parameterized as a function of N_{vtx} in the event. This procedure allows to extract the average energy each extra vertex in the event is adding, and then to correct jet energies accordingly.

- Absolute Jet Energy Scale correction. It corrects the calorimeter energy for non-linearities and un-instrumented parts of the detector, which makes non linear the detector response to particles of different energy. The corrections $f_{jet}(R, p_T)$ are derived from Monte Carlo simulation, comparing the jets at calorimeter level and at particle level as a function of their transverse momentum p_T . After this correction the jet energy scale is independent of the CDF detector.

- The *Underlying event* correction. Energy can enter the jet cone as additional contribution in energy from ISR in the hard process or from soft interactions of spectator partons not involved in the hard interaction. These two contributions are indicated as Underlying Event (UE).

- Out-Of-Cone corrections. Those take into account the jet energy lost out of the jet cone due to large angle FSR or to jet particles exiting the cone in the fragmentation process.

The raw jet energy is corrected following the expression:

$$p_T(R, p_T, \eta) = [p_T^{\text{raw}}(R) \cdot f_\eta(R, p_T, \eta) - \text{M}p\bar{p}\text{I}(R)] \cdot f_{jet}(R, p_T) - \text{UE}(R) + \text{OOC}(R, p_T) \tag{3.21}$$

All the corrections applied in this analysis are described in detail in[107].

3.8 Missing transverse energy $\not\!\!E_T$

The presence of neutrinos in an event is inferred by an energy imbalance in the detector, since neutrinos weakly interact with the surrounding material, they escape the detector carrying away some amount of energy which cannot be directly measured. The z-component of the initially interacting partons is unknown on an event-basis, so the amount of missing energy cannot be determined. However, since particles escaping down the beam pipe carry little transverse momentum, the transverse momentum unbalance can be used to measure the transverse momentum carried away by the neutrinos. The missing transverse energy is defined as:

$$\vec{\not\!\!E}_T = \sum_i [-E_{T,i} \cos(\phi_i), -E_{T,i} \sin(\phi_i)] \tag{3.22}$$

where $E_{T,i}$ is the transverse energy of the calorimeter tower i and ϕ is its azimuthal angle. The sum is made over the energy of all calorimetric towers projected onto the plane perpendicular to the beam axis.

Missing transverse energy correction

Corrections are applied to the raw missing transverse energy, in three stages:

- corrections for the lepton (especially in muon case);
- adjusting $\not\!\!E_T$ due to jet energy corrections;
- tau related corrections.

The components of missing $\vec{\not\!\!E}_T$ vector in (3.22) are corrected for the energy deposited by minimum ionizing high-p_T muons by subtracting an amount of energy equal to the transverse momentum of the muon track and adding back the transverse energy in the calorimeter towers traversed by the muons.
Missing E_T is corrected also for the jet energy corrections (discussed in Section 3.7). The following trasformation is performed:

$$(\not\!\!E_T^{corr})_{x,y} = (\not\!\!E_T)_{x,y} - \sum_{\text{jets}} (p_T(\text{jet}))_{x,y}(f_{\text{jet}} - 1)$$

where p_T(jet) is uncorrected jet transverse momentum and f_{jet} is the correction factor for a jet.
The tau related correction to the missing energy takes into account the 10% scaling of the hadronic component of the tau energy, and it is applied to the missing energy in MC generated events as follows:

$$(\not\!\!E_T^{corr})_{x,y} = (\not\!\!E_T)_{x,y} + E_{x,y}^{\tau\ \text{HAD}} - 1.1 \cdot E_{x,y}^{\tau\ \text{HAD}}$$

Chapter 4

Selection and modeling of e-τ and μ-τ events

This Chapter describes the basic event selection for the $e - \tau$ and $\mu - \tau$ events, with exactly two opposite charged leptons and missing transverse energy. The background processes contributing to the final states $e - \tau$ and $\mu - \tau$ are also studied.

4.1 Base events selection

Events with the e-τ or μ-τ reconstructed final states are selected starting from a sample of events collected up to an integrated luminosity of ~ 4.8 fb^{-1} and triggered with the high-p_T lepton triggers, which select events with at least one high-p_T electron or muon reconstructed in the CDF detector. The specific trigger paths used in this analysis exploit information from CEM (see Section 1.2.2) for electrons, and from CMUP and CMX (see Section 1.2.3) for muons, as listed in the following:

- ELECTRON_CENTRAL_18: selects events with at least one electron in the CEM detector with $E_T > 18$ GeV.
- MET_PEM: selects events with one high-E$_T$ electron in the forward calorimeter region, such as an electron or photon with $E_T > 20$ GeV, and missing energy greater than 15 GeV.
- MUON_CMUP18: selects events with one muon with hits in the CMU and CMP detectors and $p_T > 18$ GeV/c.
- MUON_CMX18: selects events with high-p$_T$ muons with hits in the CMX detector.
- MUON_CMP18_PHI_GAP: events with at least one central muon with hits in the CMU detector only, pointing to region not covered by CMP chambers.

The detail requirements at trigger level for all the trigger paths used in this Thesis are described in [108].
The total integrated luminosity analyzed depends on the reconstructed leptons in the event and on the online status of the subdetectors during data taking. The luminosity for each trigger path listed above is summarized in Table 4.1.

Trigger	L [pb^{-1}]
Central electron	4829
Forward electron	4549
Central muon	4662
Forward muon	4395

Table 4.1: Integrated luminosity collected by different good run lists.

Leptons are reconstructed as described in Sections 3.1 and 3.2. In the events a central electron (TCE) or a muon identified as: CMUP, CMP, CMX, CMXMsKs, CMIOCES or CMIOPES are searched along with an additional object identified as a tau. Electrons and muons are required to have offline $E_T(p_T) \geq 20$ GeV(/c), while the reconstructed tau lepton $E_T^\tau \geq 10$ GeV. Trigger turn-on effects are avoided applying the 20 GeV cut on the electron or muon energy. The lower cut on the tau lepton E_T^τ takes into account the reduced kinematical acceptance for tau leptons compared with electrons and muons, due to the presence of neutrinos in the tau decay. However, the tau identification already requires the cluster E_T^τ cut at 10 GeV. A separation of $\Delta R(l, \tau) > 0.4$ is required for the two leptons passing the acceptance cuts, to avoid that the same lepton is reconstructed both as l and τ lepton type. Events with more than two reconstructed leptons are rejected. The selected event can thus be classified by the flavor of the lepton pair, (e-τ) or (μ-τ), and analyzed separately.
Very simple and minimal requirements are applied to selected events at this step of the analysis:

- collected by one of the high-p$_T$ lepton trigger paths;
- two opposite charged reconstructed leptons;
- a primary vertex in the axial direction with $|Z_0^{vtx}| \leq 60$ cm
- leptons coming from the same interaction point $|Z_0^1 - Z_0^2| \leq 5$ cm
- leptons close to the identified vertex $|Z_0^{1,2} - Z_0^{vtx}| \leq 5$ cm
- flavor requirement: $e - \tau$ or $\mu - \tau$
- invariant mass of the lepton pair M(l,τ)> 20 GeV/c^2
- ΔR(l,τ)> 0.4
- one trigger lepton with $E_T(p_T) \geq 20$ GeV(/c) and one hadronic tau with $E_T^\tau \geq 10$ GeV;

In a data sample of 4.8 fb^{-1}, 6722 events with $e - \tau$ signature and 3064 events with $\mu - \tau$ in the final state pass the minimal cut selection.

4.2 Sample composition

Many different SM processes can mimic the signal, satisfying the minimal selection defined above. They can be grouped into different classes of background, according to their contribution with real or fake taus:

- *physics background* (with only real tau):
  - $Z/\gamma^* \to \tau\tau$
  - WW, WZ, ZZ
  - $t\bar{t}$
- *single-fake background* (one fake lepton):
  - $W + jet$, where the jet is mis-identified as a τ
  - $W + \gamma$, where the photon is mis-identified as a τ
  - $Z/\gamma^* \to ee/\mu\mu$, where either an electron or a muon is mis-identified as a τ
  - $Z/\gamma^* \to ee/\mu\mu + jet$, where the jet is mis-identified as a τ
- *double-fake "QCD" backgrounds* (two fake leptons):
  - dijet, where one jet is mis-identified as a τ and the other jet as an electron or muon
  - $\gamma + jet$, where the γ or jet is mis-identified as a τ or l.

4.2.1 Physics background processes

The Drell-Yan $Z/\gamma^* \to \tau\tau$ process is the largest SM background giving 2 leptons in the final state and thus contributing to the e-τ or μ-τ reconstructed final state, when one of the two τ decays leptonically and the other one hadronically. The cross section for $p\bar{p} \to Z/\gamma^*$ production is 490 pb for $M_{ll} > 20$ GeV/c^2. In this process a quark pair annihilate to produce Z/γ^* which can decay into a lepton pair. The Z contribution is strongly peaked at around 90 GeV/c^2 while the γ^* contribution dominates the lower mass region. This process does not contain any neutrinos in the final state and thus ideally does not contributes with any *real* missing energy, but the required signature is satisfied when jet activity from initial/final state radiation or detector resolution effects cause an energy imbalance in the transverse plane. However, in $Z/\gamma^* \to \tau\tau$ events the missing energy comes also from neutrinos from the τ decay, although it is small.

The WW process looks very similar to the $H \to WW$ signature, but it has not the largest cross section among the backgrounds. The total $p\bar{p} \to WW$ cross section calculated at NLO is 12.4 ± 0.8 pb [109]. Considering the W boson decay to a lepton and a neutrino ($e\nu$, $\mu\nu$, or $\tau\nu$), the WW will yield a $e - \tau$ or $\mu - \tau$ final state the 3% of the time, with two high energy leptons and significant missing energy from neutrinos.

The WZ and ZZ cross sections are relatively small at Tevatron, predicted at NLO to be 3.7 ± 0.3 pb and 1.4 ± 0.1 pb [109] respectively. The WZ process contributes to the dilepton final state with the Z decaying leptonically and the W decaying into a lepton-neutrino pair ($WZ \to l\nu l^+l^-$) or two quarks ($WZ \to qq'l^+l^-$). In the former case one lepton must be missed and in the latter there is no neutrino present as a source for missing energy. The contribution to a dilepton sample from ZZ will mainly come from the case where the Z decays to a lepton pair and the other decays into neutrinos ($ZZ \to l^+l^-\nu\bar{\nu}$) or jets ($ZZ \to l^+l^-jj$). A much smaller contribution will come from the leptonic decay of both Z bosons with two leptons missed.

Top pair production can result from quark-antiquark annihilation or through gluon-gluon fusion. At the Tevatron the dominant production process is expected to be quark-antiquark annihilation and for a top mass of 175 GeV/c^2 the SM cross section at $\sqrt{s} =$

1.96 TeV is calculated to be $6.7^{+0.7}_{-0.9}$ pb [110]. The top quark decays to Wb with a branching ratio of 100%, therefore, it contributes to the final state with two real leptons and missing energy arising from the leptonic decay of the Ws.

4.2.2 Fake background processes

Some events that pass all event selection cuts, thus entering the final sample as background events, have jets, electrons or muons passing tau identification cuts, despite the isolation requirements and electron/muon removal cuts specifically applied in τ identification (see Section 3.6.2). Processes like $W + jets$, $W + \gamma$ and QCD backgrounds enter significantly in the final selected sample. They do not contain two leptons in the final state, but will appear in the dilepton sample as a result of mis-identification of either the photon (γ) or a jet originating from a quark or gluon. While, the $Z/\gamma^* \rightarrow ee/\mu\mu$ background events enter the $e - \tau$ or $\mu - \tau$ selected sample for e/μ passing tau identification criteria.
The ability of electrons, muons or jets to fake taus is expressed evaluating the lepton to tau fake rate $f_{(l\rightarrow\tau)}$ and jets to tau fake rate $f_{(jet\rightarrow\tau)}$.

$jet \rightarrow \tau$ fake rate

The $jet \rightarrow \tau$ fake rate is calculated as the ratio of the number of jets passing τ identification requirements (defined in Section 3.6.3) to the number of jets passing the looser criteria for τ reconstruction, therefore reconstructed as a candidate τ:

$$f_{(jet\rightarrow\tau)} = \frac{N(jet \rightarrow \tau^{ID})}{N(jet \rightarrow \tau^{reco})} \tag{4.1}$$

The fake rate is measured using a sample of jets in both MC and data events to verify that the rate in MC reproduces the rate in data. It is measured in data using four different jet samples, collected using different jet-trigger selections: JET_20, JET_50, JET_70 or JET_100. These trigger paths require at least one jet to be reconstructed in a cone $\Delta R = 0.7$ with different energy thresholds: 20, 50, 70 and 100 GeV. The corresponding MC samples have been used to measure the fake rate on simulated jets.

The topology of the event also affects the ability of a jet to be misidentified as a tau. Most of the QCD events have two transversely back-to-back jets with transverse energy close to each other, however in multi-jet events the jet has a lower probability to pass the isolation requirements in the tau identification, and thus to contribute as a fake tau. In Figure 4.1 is shown that the topology of events (expressed by N_{jet}), where a fake tau at a given energy is reconstructed, is different for events selected with different jet-triggers. Requiring a tau candidate with $E_T > 20$ GeV and regardless to which the isolation requirement is, we expect the fake rate to be higher in QCD events with two transversely back-to-back jets, mainly events collected with JET_20 trigger (Figure 4.1 (left)), than in multi-jet events, mainly events collected with triggers at higher E_T threshold (Figure 4.1 (right)).

The sample of fake taus is obtained requiring the reconstructed tau in the jet sample to be not the leading E_T jet in order to remove any trigger bias, since the leading E_T spectrum of the leading jet is limited by the trigger threshold energy (see Figure 4.2 (left)). Additionally, the invariant mass of the reconstructed tau and the leading jet

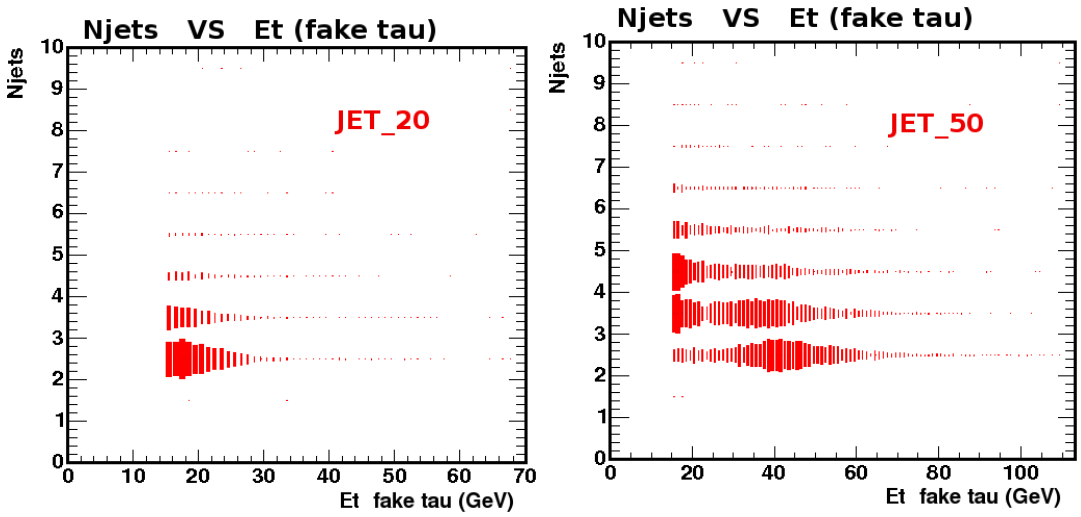


Figure 4.1: It is shown the topology of the events where a fake tau at a given energy is reconstructed, if the event is collected using JET_20 (left) or JET_50 (right) trigger.

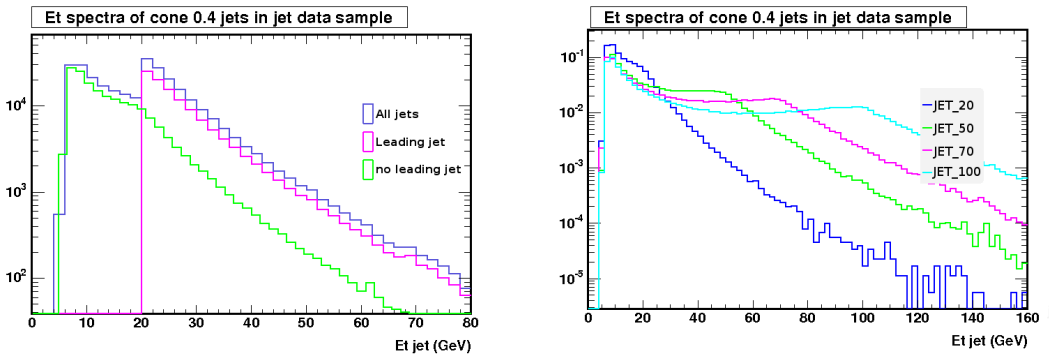


Figure 4.2: Left: E_T spectra for second leading and leading jets in events collected with the JET_20 trigger. The E_T spectra for the leading jets is limited by the trigger threshold energy, therefore leading E_T jets are not used, to avoid trigger biases. Right: E_T spectrum of jets in events where the leading jets have been removed.

system in the event is required to be out the Z peak region to remove Drell-Yan contribution: $M(\tau\text{-}jet) < 76$ GeV/c^2 or $M(\tau\text{-}jet) > 106$ GeV/c^2. In Figure 4.2 (right) is shown the E_T spectrum of jets in events where the leading jets have been removed, then used for fake rate calculation. The τ identification requirements are applied to select the sample of jet faking hadronic τs.

The fake rate is thus measured as a function of the calorimeter transverse energy of the fake tau. 1-prong and 3-prong tau candidates may exhibit different fake rates, therefore the scale factor is calculated for the two tau candidate types separately. In Figure 4.3 and 4.4 are shown the fake rates as a function of the calorimeter energy of the τ candidate in the JET_20, JET_50, JET_70 and JET_100 data and MC samples, for 1-prong and 3-prong tau respectively. The fake rate measured in different samples is the same within the statistical uncertainty and it is of the order of $5 - 10\%$ for transverse energies about the trigger threshold.

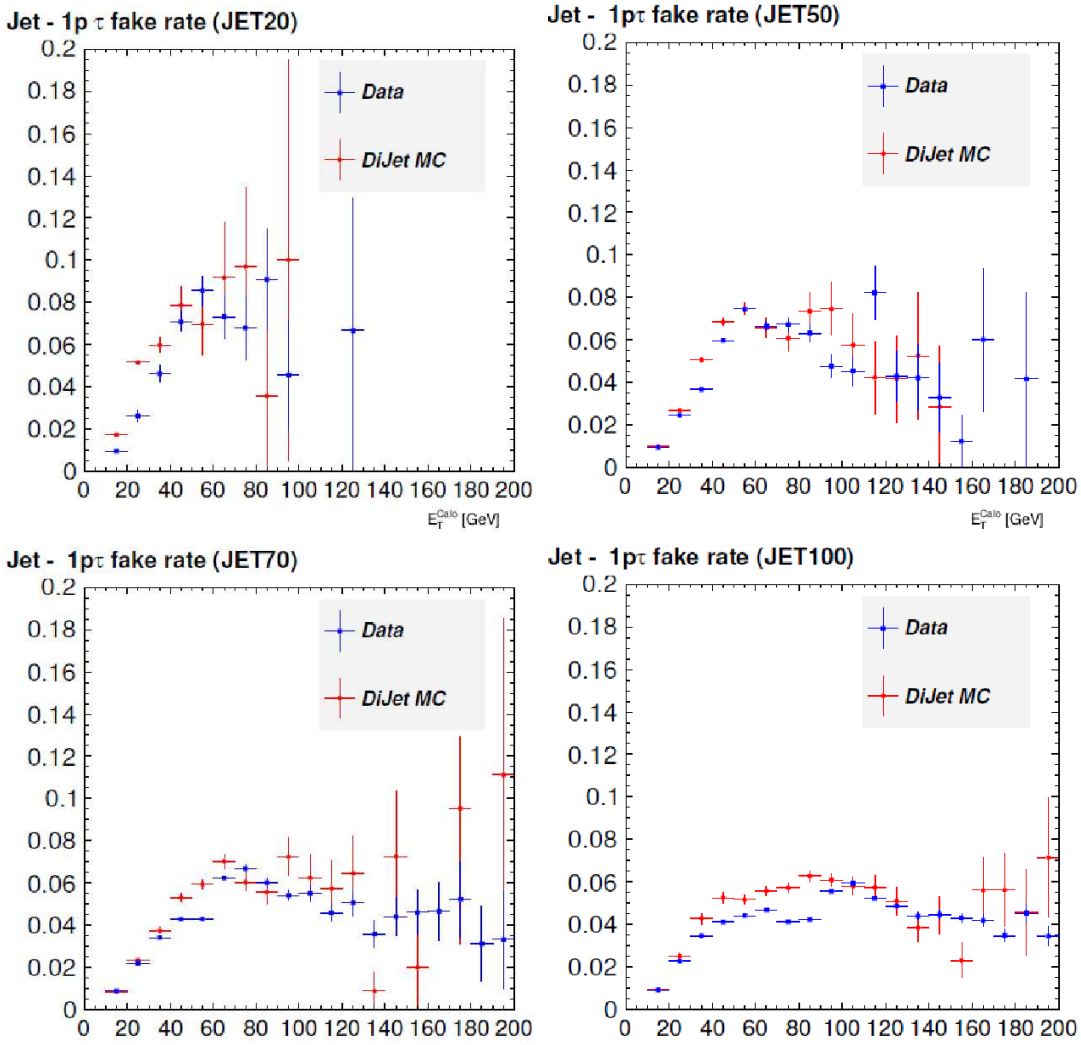


Figure 4.3: Jet into 1-prong tau fake rates as a function of the calorimeter energy of the τ candidate calculated in the JET_20, JET_50, JET_70 and JET_100 data samples and in the corresponding MC samples.

$e/\mu \to \tau$ fake rate

To measure the fake rate the $Z \to \mu\mu$ and $Z \to ee$ events are examined, where one of the leptons is identified as the primary electron or muon and the other fakes a tau. The fake rate is calculated as the ratio of the number of $Z \to ll$ events with one identified lepton and one identified τ over the number of $Z \to ll$ events with one identified lepton and one passing only the selection for τ reconstruction.

$$f_{l\to\tau} = \frac{N(l, \tau^{ID})}{N(l, \tau^{reco})} \tag{4.2}$$

The number of $Z \to ee$ and $Z \to \mu\mu$ events is estimated from the distribution in dilepton mass counting the number of events in the Z mass window [85, 95] GeV/c^2, for the e-τ and μ-τ channels respectively. The lepton into τ fake rate is calculated in data and MC events, selecting 1-prong taus with no π^0s, in order to untangle the contributions from

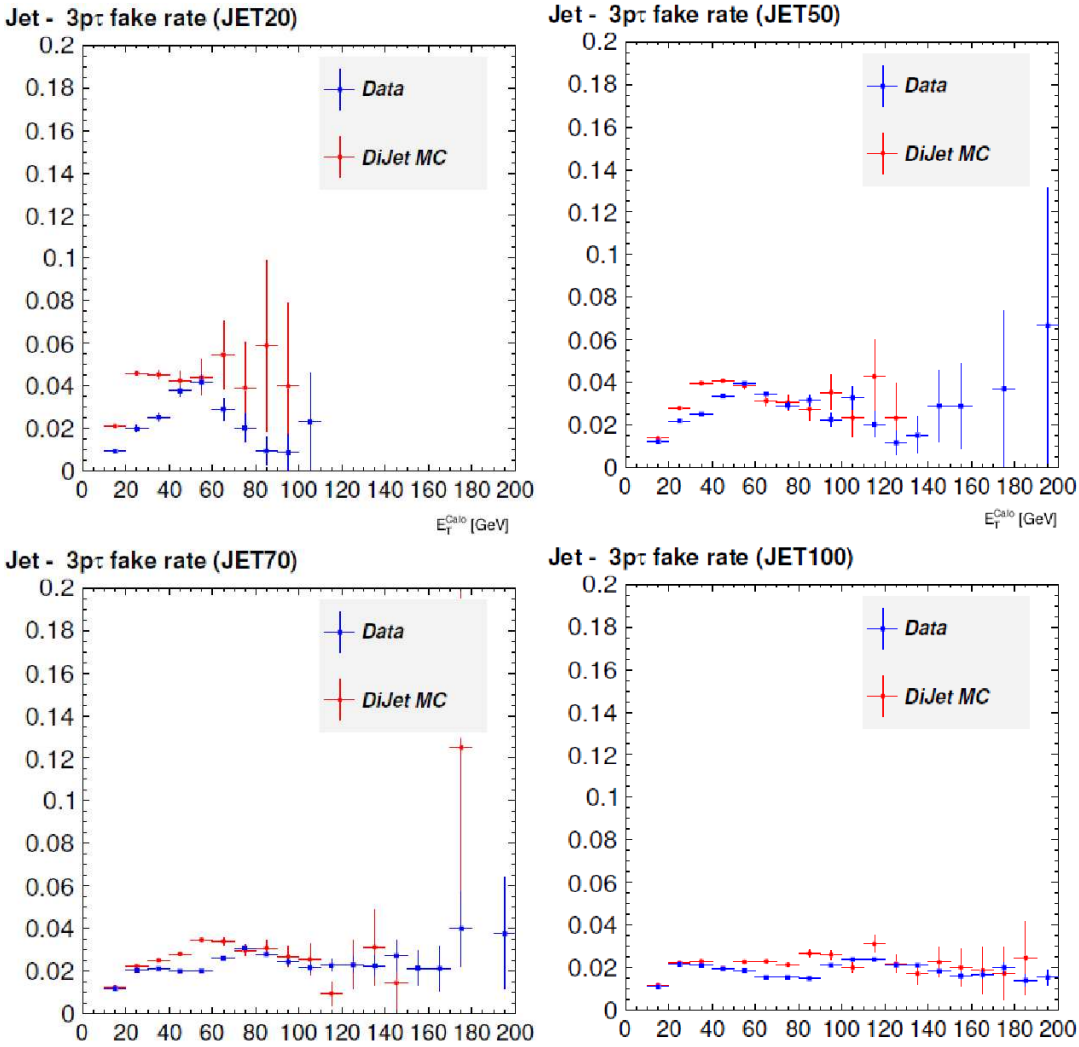


Figure 4.4: Jet into 3-prong tau fake rates as a function of the calorimeter energy of the τ candidate for the JET_20, JET_50, JET_70 and JET_100 data samples and corresponding MC samples.

leptons and leptons radiating a photon.
The distribution of the dilepton invariant mass in Drell-Yan events is given by a Breit-Wigner superimposed to the exponential distribution from $\gamma \rightarrow ee/\mu\mu$ events. The sample of $Z \rightarrow ee/\mu\mu$ events is selected in the Z mass window after the expected non-$Z \rightarrow ee/\mu\mu$ processes have been subtracted out. Fitting the dilepton invariant mass to an exponential in the tail of the mass distribution, used as side-band region, the fit result is then used to estimate the background in the Z mass window $85 \leq M(l\tau) \leq 95$ GeV/c^2.
In Figures 4.5 the fit results in $Z/\gamma^* \rightarrow ee/\mu\mu$ MC is shown for events with one identified electron/muon and one reconstructed 1-prong no-π^0 tau (on the left) and one identified electron/muon and one identified 1-prong no-π^0 tau (on the right).

For the fake rate estimation in data events, the non-$Z \rightarrow ee/\mu\mu$ processes also falling under the Z mass peak, reconstructed with data events, are extrapolated from their

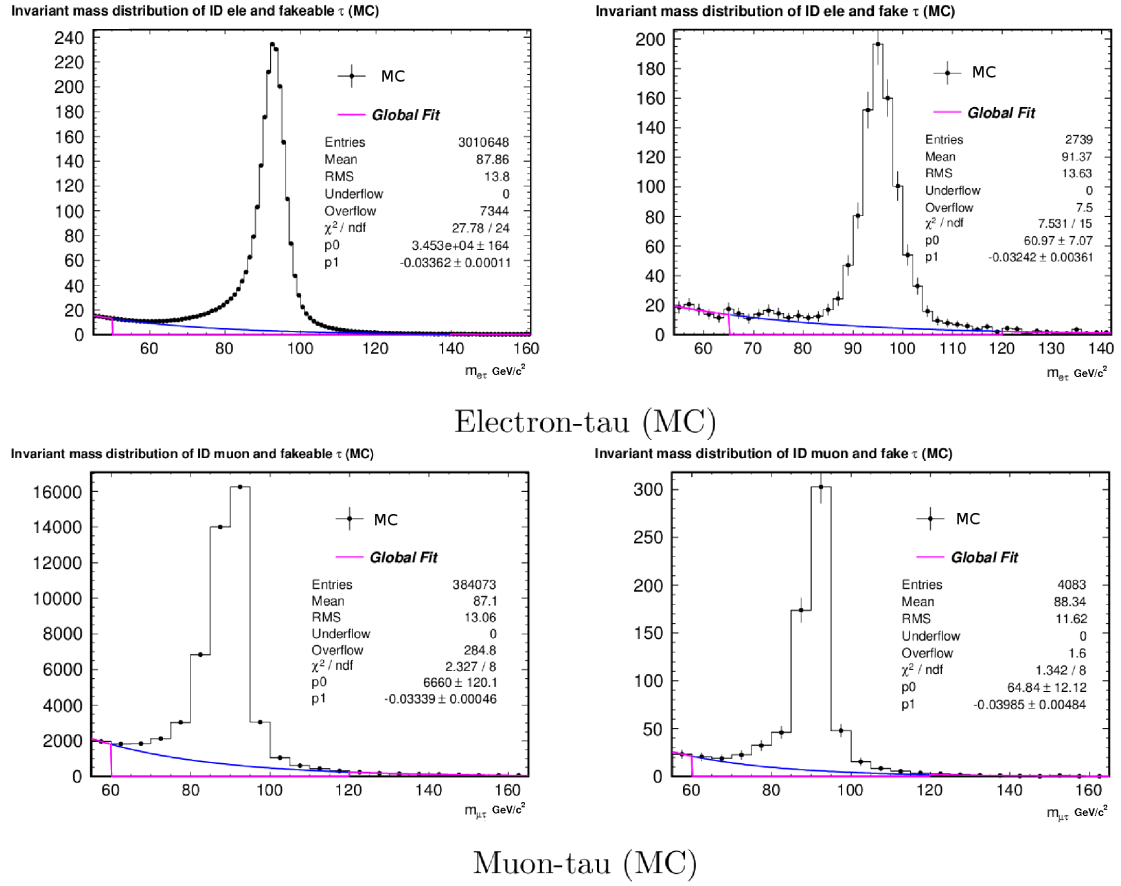


Figure 4.5: Invariant mass distribution of one identified electron or muon and one reconstructed/identified τ (left/right) from a subsample of $Z/\gamma^* \to ee/\mu\mu$ MC simulated events.

contribution in the tails of the mass distribution, used as side-band regions. QCD processes can fall under the Z peak, which contribution can be estimated selecting data events with leptons having the same charge sign, exploiting their symmetry in the charge of the leptons. The subtraction of data events with leptons having the same charge sign removes the QCD contribution and only a fraction of $W + jets$ events, since they are not symmetric in the charge of leptons. The reconstructed Z mass from $Z \to \tau\tau$ decays have an invariant mass distribution shifted to lower mass values with respect to the $Z \to ee/\mu\mu$ contribution, thus falling in the side band used as control region for the non-$Z \to ee/\mu\mu$ estimation. Since the reconstructed $Z \to \tau\tau$ mass has also a large resolution due to the presence of neutrinos in the final states, its mass distribution is expected to be not enough separated from the distribution of reconstructed $Z \to ee/\mu\mu$ decaying process. To avoid any bias in selecting the real $Z \to ee/\mu\mu$ process in data events, only events with missing transverse energy smaller than 30 GeV are considered. In Figure 4.6 is shown the invariant mass distribution in the $e - \tau$ and $\mu - \tau$ channels. Data distribution is compared to the estimated prediction for the most contributing processes $Z \to ll$, $Z \to \tau\tau$ and QCD. The fit results using data events are shown in Figures 4.7, for events with one identified electron/muon and one reconstructed 1-prong no-π^0 tau (on the left) and one identified electron/muon and one identified 1-prong no-π^0

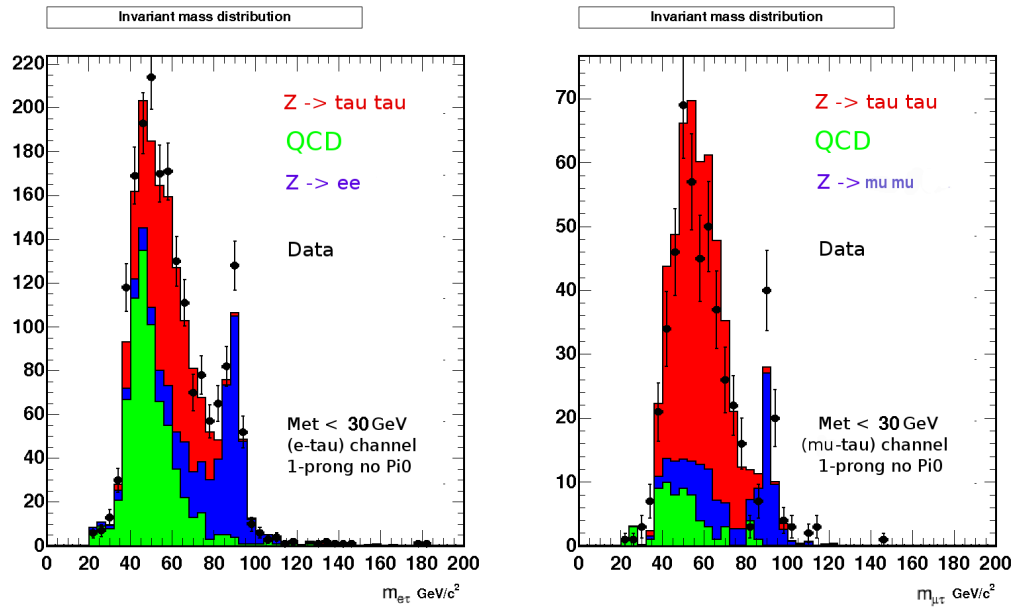


Figure 4.6: $M_{e\tau}$ (left) and $M_{\mu\tau}$ (right) distributions in data and MC. The data events with two leptons with opposite charge are shown as the plotted points while data events with same charge leptons used for QCD estimation are plotted as the green histogram. The mass distribution from $Z \to \tau\tau$ and $Z \to ll$ MC is stacked on top of the QCD distribution, normalized to the number of data events. The missing E_T less than 30 GeV is required to reduce the $Z \to \tau\tau$ contribution inside $85 \leq M(l\tau) \leq 95$ GeV/c^2 mass window. Only 1-prong with no-π^0 taus are considered.

tau (on the right). Those are used to estimate the background in the Z mass window $85 \leq M(l\tau) \leq 95$ GeV/c^2 .

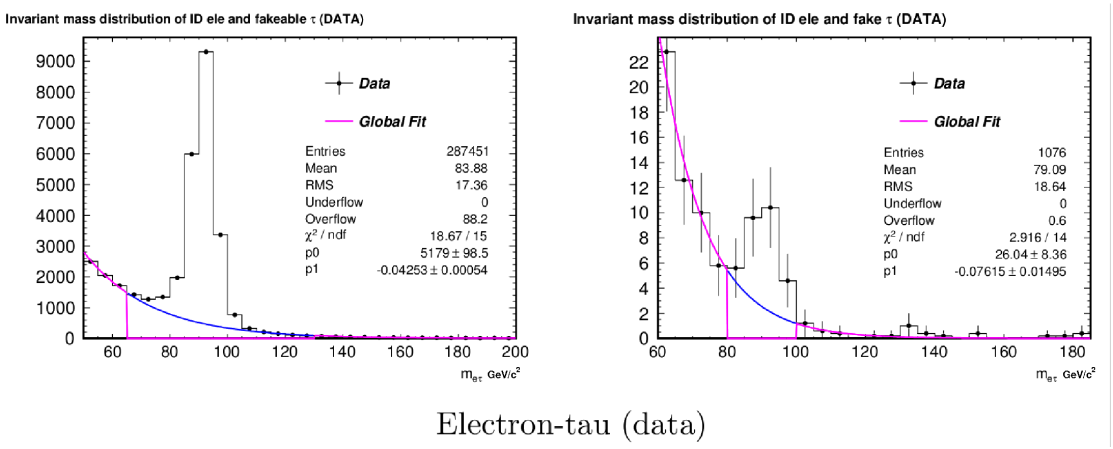
In Table 4.2 there are the $l \to \tau$ fake rates. The numerator and denominator in (4.2) are taken from the plots in Figures 4.5 and 4.7 for fake rate estimation in MC and in data respectively, counting the number of events in the Z mass window $85 \leq M(l\tau) \leq 95$ GeV/c^2, subtracted by the background contribution estimated from the fit. The lepton into tau fake rates are of the order of few percents.

	Electrons	Muons
Fake Rate Data	$(1.0 \pm 0.1) \times 10^{-3}$	$(3.1 \pm 0.2) \times 10^{-2}$
Fake Rate MC	$(0.67 \pm 0.01) \times 10^{-3}$	$(1.60 \pm 0.01) \times 10^{-2}$

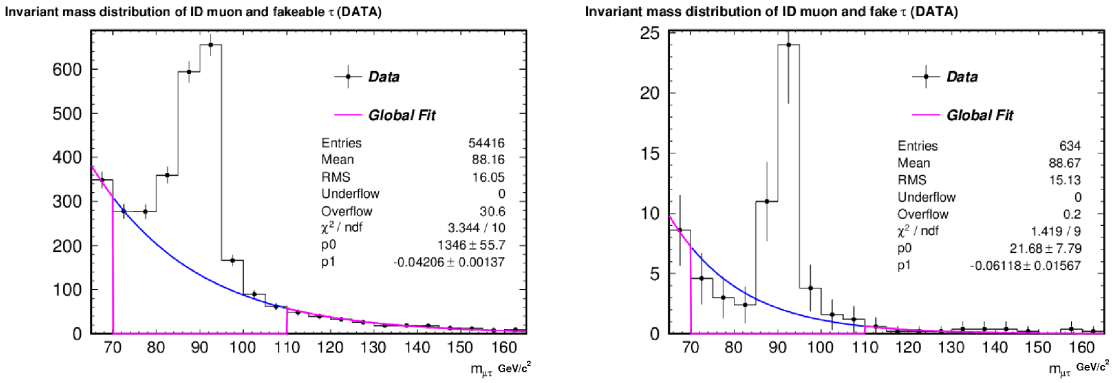
Table 4.2: Lepton into τ fake rates for 1-prong with no-π^0 taus.

4.3 Modeling of background

The knowledge about the physics processes and their contribution in the sample of selected events is required to analyze measured data. The most of background processes in this Thesis are simulated using MC generators, described in Section 2.5. Tables 4.3



Electron-tau (data)



Muon-tau (data)

Figure 4.7: Invariant mass distribution of one identified electron or muon and one reconstructed/identified τ (left/right) from data events.

and 4.4 gives a short summary of the MC simulated samples, with the process generator and the cross section used to normalize them.

channel	Generator	$\sigma \times B$[pb]	K-factor	Run Period
$Z/\gamma^* \to \tau\tau$	PYTHIA	355	1.38	0-23
WW	MC@NLO	11.66	1.0	0-7
WZ	PYTHIA	3.46	1.0	0-23
ZZ	PYTHIA	1.51	1.0	0-23
$t\bar{t}$	PYTHIA	7.04	1.0	0-23

Table 4.3: List of the MC simulated physics backgrounds. The event generator, the cross-section times the branching fraction used in the normalization and the run-period on which the simulation was tuned are also reported.

$p\bar{p} \to WZ, ZZ, t\bar{t}$ and Drell-Yan (DY) background processes are fully simulated using Pythia at the Next-to-Leading Order (NLO). Processes are scaled by the K-factor (1.22) from the generated cross sections to the most recent available cross section calculations at NLO accuracy.
Monte Carlo samples used to model DY processes have been produced under the con-

channel	Generator	$\sigma \times B$[pb]	K-factor	Run Period
$Z/\gamma^* \to ee$	PYTHIA	355	1.38	0-23
$Z/\gamma^* \to \mu\mu$	PYTHIA	355	1.38	0-23
$W\gamma \to e\nu\gamma$	BAUR	16.2	1.36	0-11
$W\gamma \to \mu\nu\gamma$	BAUR	16.2	1.36	0-11
$W\gamma \to \tau\nu\gamma$	BAUR	13.6	1.36	0-11

Table 4.4: List of the MC simulated $Z/\gamma^* \to ee/\mu\mu$ and $W\gamma$ backgrounds. The event generator, the cross-section times the branching fraction used in the normalization and the run-period on which the simulation was tuned are also reported.

straint that the two leptons are generated with invariant mass M($l\tau$) greater than 20 GeV/c^2. Due to unreconstructed neutrinos from the τ decay, and the consequently underestimation of the τ momenta, the invariant mass distribution from reconstructed leptons results shifted towards lower mass values with respect to the generated mass distribution (the two distributions are shown in Figure 4.8). The mass cut at the generator level in MC is not affecting the acceptance of the event selection, because due the shift of the reconstructed mass distribution if events were generated with invariant mass below 20 GeV/c^2 they would be expected not to pass the base selection which requires the invariant mass of reconstructed leptons to be greater than 20 GeV/c^2. Therefore, no corrections to the acceptance are needed.
The $Z/\gamma^* \to ee/\mu\mu$ contribution to background due to leptons faking taus is calculated using MC samples.

For the MC simulation of $W\gamma$ process, the BAUR matrix element generator is used

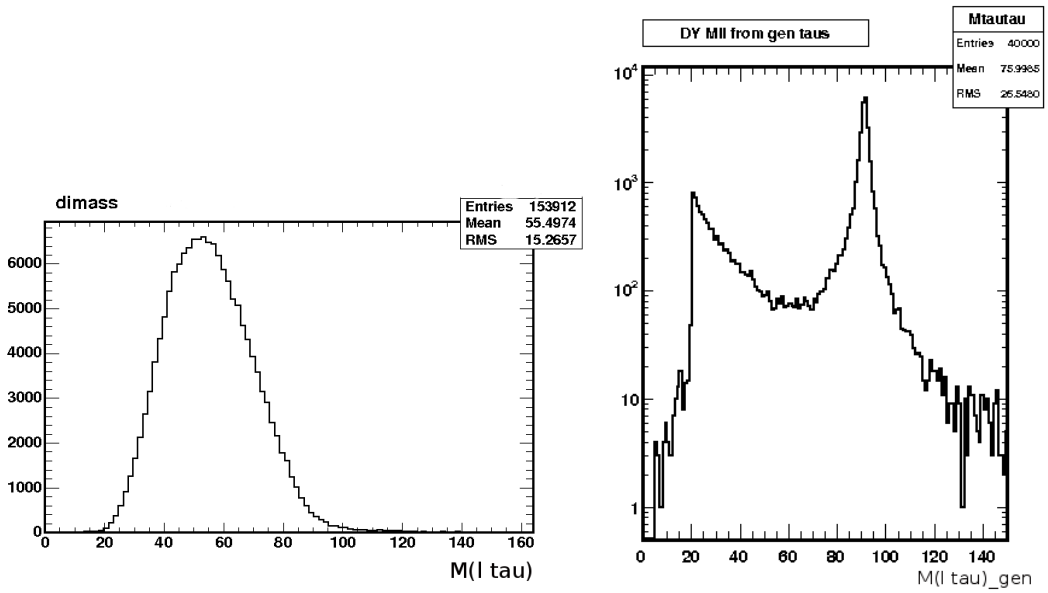


Figure 4.8: Left: Reconstructed invariant mass distribution using a sample of τ leptons in MC $Z \to \tau\tau$ events, with the M(l-τ) $>$ 20 GeV/c^2 cut applied at generator level. Right: invariant mass distribution from generated τ leptons, in a sample of MC $Z \to \tau\tau$ events.

channel	partons	Generator	$\sigma \times B$ [pb]	K-factor
W ($e\nu$)+ jet	0	ALPGEN	1800.0	1.35
	1	ALPGEN	225.0	1.35
	2	ALPGEN	35.3	1.35
	3	ALPGEN	5.59	1.35
	4	ALPGEN	1.03	1.35
W ($\mu\nu$)+ jet	0	ALPGEN	1800.0	1.35
	1	ALPGEN	225.0	1.35
	2	ALPGEN	35.3	1.35
	3	ALPGEN	5.59	1.35
	4	ALPGEN	1.03	1.35
W ($\tau\nu$)+ jet	0	ALPGEN	1800.0	1.35
	1	ALPGEN	225.0	1.35
	2	ALPGEN	35.3	1.35
	3	ALPGEN	5.59	1.35
	4	ALPGEN	1.03	1.35

Table 4.5: List of ALPGEN MC samples used to estimate the $W+jets$ background. The event generator and the cross-section times branching fraction used in the normalization are also reported.

and the WW process is generated using the NLO MC generator MC@NLO [111], since this process heavily contributes to final states where jets are reconstructed along with the two Ws.

Background processes with multi-jets, which contribute with double-fakes leptons, will be measured in data selecting events with leptons having the same electric charge. The number of di-jet and photon-jet events with opposite electric charge is expected to be the same as the number of events with fake leptons of same electric charge. The contamination of electro-weak processes with same charge leptons is properly taken into account and subtracted using predictions from MC samples.
The $jet \to \tau$ single-fake contribution has been estimated using a prediction from MC in simulated events. $W + jets$ and $Z/\gamma^* \to ee/\mu\mu + jet$ MC samples are used to estimate the corresponding contributions to the background. The single fake background from $W+jets$ is estimated using the ALPGEN generator. The $W+jets$ samples are generated with a specific number of partons (p) in the matrix element, including samples with $W + 0p$, $W + 1p$ and $W + 2p$. In order to get the total $W + jet$ contribution the three subsample are added up, removing double-counting events due to possible additional jets created during parton shower, according to the prescription in [93].
MC samples used to describe physics background processes also include jets coming from hard-scattering between partons. Therefore, they contribute also like single-fake background, since the reconstructed hadronic τ in the final state can be the real τ from W or Z decay or a fake tau when a jet from the radiative process is mis-reconstructed. To include events contributing with fake leptons, the reconstructed objects are not required to match their generator level parent particle.

In this Thesis most of the processes contribution with fake leptons are estimated from MC, therefore, MC corrections become fundamental also in correctly estimate the ability

of jets or other leptons in the event to mimic a tau. A study about the feasibility in this analysis of the estimation of the background due to $jet \to \tau_{had}$ misidentification using a *data-driven method* has been carried out and reported in Section 4.4. Big uncertainties arising from the data-driven method turned to the MC estimation advantage.

4.3.1 Correction to MC simulated events

For each simulated process a set of MC samples covering run periods is used. However, some samples have a restricted run range simulated. The MC samples for the WW , WZ, ZZ, and $W + jet$ are generated for a limited run range: WZ and ZZ cover periods 0-13, WW and $W + jet$ cover periods 0-17. Finally, the numbers from each run period is weighted by the appropriate luminosity to obtain the final acceptance numbers and possible difference in luminosity profiles will be taken into account as systematic uncertainty.

Thus, the per-event based weight applied to MC samples is:

$$w_i = \frac{s^i_{tot}\ \sigma_{prod}\ B\ L}{N_{gen}} \tag{4.3}$$

- σ_{prod} : is the production cross section used for the simulated process;
- B : is the branching ratio into the detected final state;
- L : is the luminosity of the good run list corresponding to the reconstructed leptons;
- N_{gen} : is the number of generated events in the MC samples;
- s^i_{tot} : is the global scale factor which accounts for differences between data and MC of trigger efficiencies, of the leptons reconstruction and identification efficiencies, and of the tau fake rates. It also accounts for filter and primary vertex requirements in MC generations: $(s^i_{tot} = \epsilon_{filter} \times \epsilon^{evt}_{trig} \times s_{lep\ ID} \times \epsilon_{vtx} \times s_{fake})$;

Trigger requirements introduce inefficiencies that have to be taken into account when comparing real data with simulation distributions. ϵ^{evt}_{trig} is the trigger efficiency for the event, defined as the probability for the event to be triggered, as a function of the reconstructed leptons firing the corresponding trigger path in the event. If only one lepton is found to fire one of the trigger paths, then the trigger efficiency for that lepton ϵ^{lep}_{trig} is also the trigger efficiency for the event. Trigger efficiencies ϵ^{lep}_{trig} will be calculated in the follows.

The correction factor ϵ_{vtx} account for the requirement on the generated z of the primary vertex to be within $\pm$60 cm from the center of the detector. It is measured in minimum bias events as function of the run period, which is on average $\epsilon_{vtx} = 0.9555 \pm 0.0031$ with less than 2% differences from one period to another.

In order to account for possible differences between simulated and collected data on the lepton identification (ID), and fake probability, scale factors have been applied. The lepton identification efficiencies $\epsilon_{lep\ ID}$ have been discussed in Section 3.4, for electrons and muons, and in Section 3.6.3 for taus, while lepton fake rates f_{fake} have been discussed in Section 4.2.2 for $jet \to \tau$ and $l \to \tau$ fake rates. The s_{fake} scale factors are given by the ratio of the fake rate calculated in data events to the fake rate calculated in MC

simulated events, using samples of leptons from $Z \to ll$ events, for the $l \to \tau$ fake rate scale factor ($s_{l\to\tau}$), and QCD jets samples, for the $jet \to \tau$ fake rate scale factor ($s_{jet\to\tau}$). Because of the relatively large sample size, the s_{fake}s are then applied to all MC processes, assuming their validity all over the background contribution and making the approximation that their calculation is not affected by the topology of the event. Scale factors for lepton identification $s_{lep\ ID}$ and for fake rates s_{fake} will be discussed in the follow.

Trigger efficiency ϵ_{trig}

For each lepton type a trigger efficiency is measured. Trigger efficiencies are taken from separate studies [112, 113, 114], with the efficiency plateau reached with lepton $p_T \geq 10$ GeV/c.

The trigger efficiency is estimated as the fraction of central leptons with $E_T > 20$ GeV firing the corresponding trigger path.

$$\epsilon_{trig} = \frac{N^{lep}_{\text{triggered}}}{N^{lep}_{\text{central}}} \tag{4.4}$$

Central and Forward electron trigger efficiencies are evaluated selecting $W \to e\nu$ events requiring an identified central electron with $E_T > 20$ GeV and a significant missing transverse energy. The central and forward electron triggers are found to be about 96% and 94% efficient respectively.
To measure muon trigger efficiencies $Z \to \mu\mu$ data events are selected, where one muon passes the trigger requirements, the second can be used to evaluate the trigger efficiency, requiring the invariant mass of the two identified muons to be $M(ll) \in [76, 106]$ GeV/c^2. The central muon efficiency is about 88%. The efficiency for forward muons is about 91%.

Trigger efficiencies are evaluated for each run period, to take into account differences in trigger requirements and detector performances. From *run period* 14 the trigger and detector conditions are stable, therefore the same values for the trigger efficiencies are used from this period on.

Lepton identification scale factors $s^{lep\ ID}$

Requirements for leptons identification have been studied on MC samples, however their consistency to data is evaluated through the lepton scale factor $s_{lep\ ID}$, defined as the ratio of lepton identification efficiencies measured in data events to the efficiency in MC. For electrons and muons with a minimum p_T of 10 GeV/c, the standard CDF data/MC scale factors for the medium and high-p_T ranges have been used [101].
For tau lepton identification, scale factor has been calculated in [105]. As seen in Section 3.6.3, the isolation and invariant mass requirements are the major contributions to the tau identification inefficiency, and they have been separately studied on data and MC events. The tau identification scale factors for the inefficiency due to the isolation

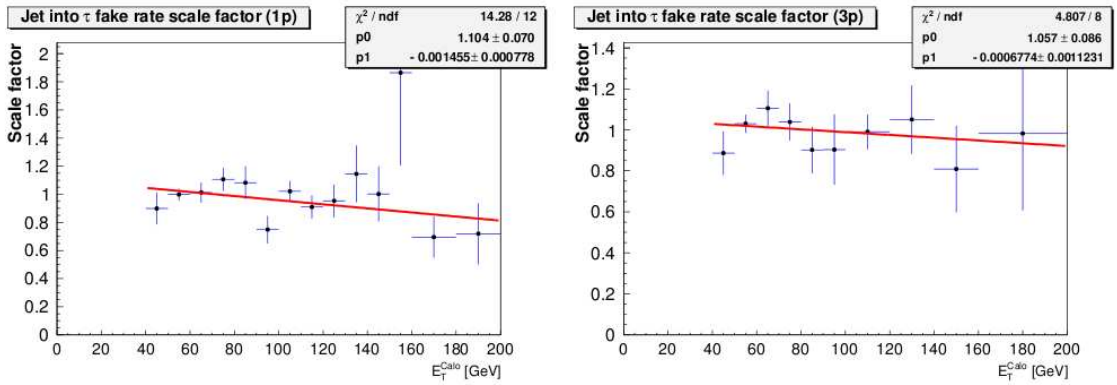


Figure 4.9: Fake rate scale factor for jet into 1-prong τ (left) and 3-prong τ as a function of the calorimeter E_T.

requirements are:

$$s_{lep\ ID}^{(\text{run}<209769)} = \frac{\epsilon_{data}^{iso}}{\epsilon_{MC}^{iso}} = 0.997 \pm 0.001 \quad (\text{runs} < 209769) \tag{4.5}$$

$$s_{lep\ ID}^{(\text{run}>209769)} = \frac{\epsilon_{data}^{iso}}{\epsilon_{MC}^{iso}} = 0.999 \pm 0.001 \quad (\text{runs} > 209769) \tag{4.6}$$

The run range distinction takes into account changes in trigger performances.
The scale factor due to the invariant mass cut inefficiency is evaluated for 1-prong and 3-prong taus separately, since two different requirements on the reconstructed mass is applied for the two decay modes. The resulting scale factors are:

$$s_{lep\ ID}^{(\text{1-prong})} = \frac{\epsilon_{data}^{m\ cut}}{\epsilon_{MC}^{M\ cut}} = 0.982 \pm 0.05 \quad (\text{1-prong}) \tag{4.7}$$

$$s_{lep\ ID}^{(\text{3-prong})} = \frac{\epsilon_{data}^{m\ cut}}{\epsilon_{MC}^{M\ cut}} = 0.965 \pm 0.005 \quad (\text{3-prong}) \tag{4.8}$$

Jet into tau fake rate scale factor $s_{jet\to\tau}$

The fake rate scale factor accounts for MC mis-modeling in reproducing the rate at which a jet, reconstructed as a τ, passes also tau identification cuts, in different phase space regions.
The fake rate scale factor is calculated in [60] using JET_20, JET_50, JET_70 and JET_100 data samples and the corresponding MC samples. It takes into account jet shape mis-modeling in MC samples for jets passing τ identification requirements, independently of the topology of the event. Thus, for convenience only τ candidates with E_T above the trigger threshold are considered. Figure 4.9 shows the scale factors for the 1-prong and 3-prong τ candidates as a function of the calorimeter energy. The energy dependence of the scale factor can be modeled with a first order polynomial. The results of the fit are also shown in Figure 4.9.

- $s_{jet\to\tau}^{1-prong}(E_T) = (1.104 \pm 0.07) - (0.0015 \pm 0.0008) \times E_T$
- $s_{jet\to\tau}^{3-prong}(E_T) = (1.06 \pm 0.09) - (0.0007 \pm 0.0011) \times E_T$

The lower cut-off at 40 GeV has been chosen to ensure to be on the plateau of the JET_20 trigger turn-on curve. The results of the fit are used to extrapolate the scale factor to lower τ energies. Since the coefficients of both cases are consistent within the statistical uncertainties, a single scale factor has been evaluated adding the two samples together. The result of the fit is shown in Figure 4.10.

- $s_{jet\to\tau}(E_T) = (1.104 \pm 0.050) - (0.0015 \pm 0.0006) \times E_T$

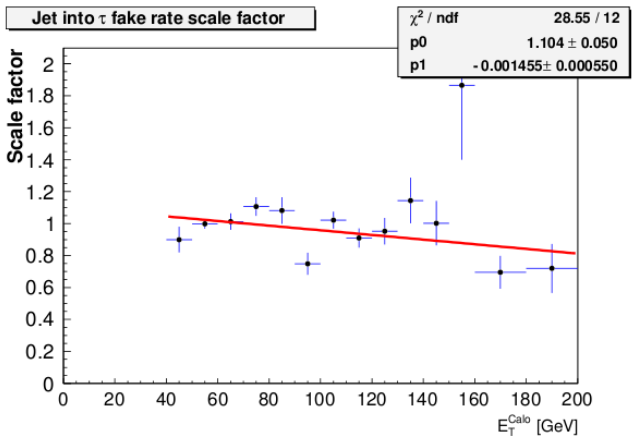


Figure 4.10: Combined 1-prong and 3-prong jet into τ fake rate scale factor as a function of the calorimeter E_T.

The main source of the uncertainty is due to the limited statistics in the samples used to derived the fit function.

Lepton into tau fake rate scale factor $s_{l\to\tau}$

The scale factor is calculated as the ratio of the lepton fake rate measured in data over the lepton fake rate as measured in MC and is reported in Table 4.6

	Electrons	Muons
Scale Factor	1.55 ± 0.21	1.95 ± 0.12

Table 4.6: Scale factor of the lepton into τ fake rate, for 1-prong with no-π^0 taus.

In the case of 1-prong taus with π^0s, data and MC agree within the limited statistics (see Figure 4.11) and we assume a scale factor of 1.

4.4 Data driven method for $jet \to \tau$ fake background estimation

The data-driven method statistically predicts the $jet \to \tau$ fake background contribution from a sample of jets, by applying the fake rate $f_{(jet\to\tau)}$, defined as the probability of a jet to pass tau identification criteria. Therefore, it assumes that the jets from the sources listed in Section 4.2 and classified as single-fake and double-fake backgrounds

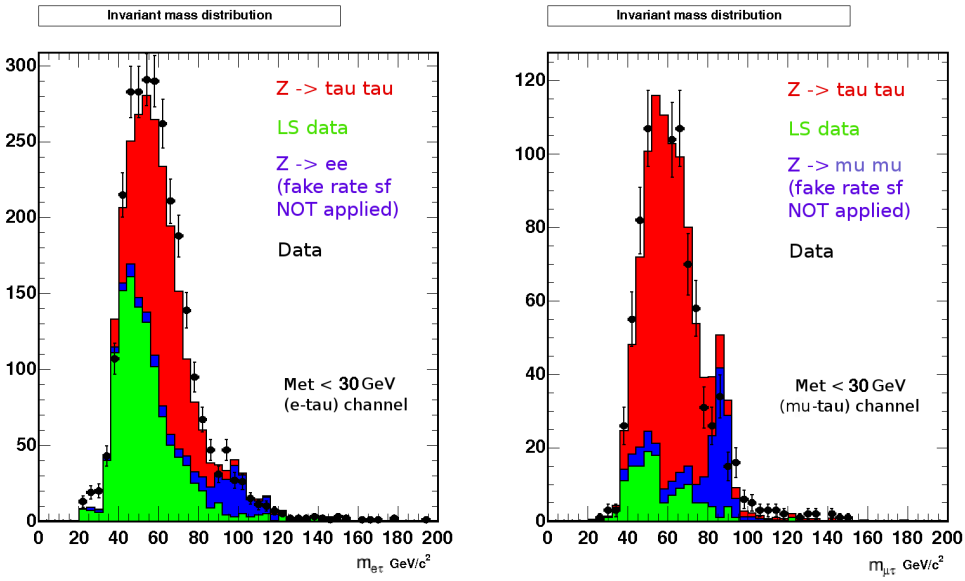


Figure 4.11: $M_{e\tau}$ (left) and $M_{\mu\tau}$ (right) distributions in data and MC, for 1-prong taus with π^0s. The data events with two leptons with opposite charge are shown as the plotted points while data events with same charged leptons used for QCD estimation are plotted as the green histogram. The mass distribution from $Z \to \tau\tau$ and $Z \to ll$ MC is stacked on top of the QCD distribution, normalized to the number of data events. The missing E_T less than 30 GeV is required to reduce the $z \to \tau\tau$ contribution in the $85 \leq M(l\ \tau) \leq 95$ GeV/c^2 mass windows. A scale factor of 1 has been considered.

have similar properties between each other.
The method is based on the selection of a sample dominated by jets where a low contribution from sources of real leptons is expected.
In Figure 4.12 is shown an example of a possible distribution of real taus and real jets in a sample of candidate τ, passing the *Loose* (L) tau requirements, defined in Section 3.5.2 as tau reconstruction. The sample (L-not T) of loose objects not passing the tau *Tighter* (T) cuts, defined in Section 3.6.3 as tau identification, can be considered as dominated by light quark jets, although a contribution from real τs is expected from τs non passing the identification criteria due to the tau identification inefficiency. The real τs contamination corresponds to the overlap between the two distributions in the (L-not T) region, as shown in Figure 4.12, whoose estimation depends on the identification efficiency ϵ. The $jet \to \tau$ fake background contribution to be estimated is the purple region in Figure 4.12 and depends on the $jet \to \tau$ fake rate $f_{(jet\to\tau)}$. Let $N_{(L-not\ T)}$ be the number of tau candidates not passing the tight tau criteria and $N_{(T)}$ the number of candidates passing them, the number of fakes in the (T) region is $N^{jets}_{(T)}$, while the real τs contamination in (L-not T) is $N^{real\ \tau}_{(L-not\ T)}$. The efficiency and the fake rate are

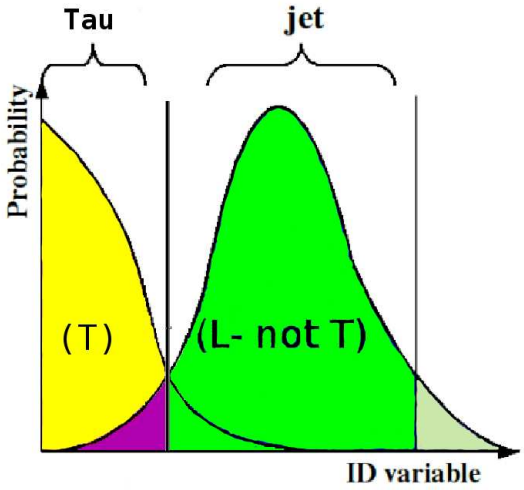


Figure 4.12: The distribution of real taus and jets in a sample of candidate τ. The purple region correspond to the $jet \rightarrow \tau$ fake background contribution to be estimated. The (L-not T) green region is chosen as the sample dominated by jets where a contribution from sources of real taus is expected.

re-defined for this study with respect to the (L-not T) sample, and are thus:

$$\epsilon = \frac{N^{real\ \tau}_{(T)}}{N^{real\ \tau}_{(L-not\ T)}} \tag{4.9}$$

$$f = \frac{N^{jets}_{(T)}}{N^{jets}_{(L-not\ T)}} \tag{4.10}$$

Therefore, the (L-not T) sample, dominated by jets, can be used to statistically predict the $jet \rightarrow \tau$ fake contribution to background by applying the fake rate defined in (4.10) (see Figure 4.13). The presence of a contamination with real taus determines an overestimation of the fake background in the (T) region as shown in Figure 4.14, thus, it needs to be accounted for when applying the fake rate to the (L-not T) sample.
The estimation of the number of fakes in the (T) region, $N^{jets}_{(T)}$, through the data-driven method can thus be summarized in the following steps:

- Selection of (L - not T) events in data sample
- Fake rate f calculation
- $N^{jets}_{(T)}$ calculation from total number $N_{(L-not\ T)}$ multiplied by f, with the assumption that $N^{real\ \tau}_{(L-not\ T)} \ll N^{jets}_{(L-not\ T)}$.

Therefore, the feasibility of the data-driven method for fakes estimation, relies on the estimation of the real tau contamination in the $N_{(L-not\ T)}$ samples and on the validity of the $N^{real\ \tau}_{(L-not\ T)} \ll N^{jets}_{(L-not\ T)}$ assumption.

In (T) and (L-not T), there are three sources that contribute to the observed events: real taus, leptons ($l = e,\ \mu$), and jets. The contribution from leptons ($l = e,\ \mu$) is small

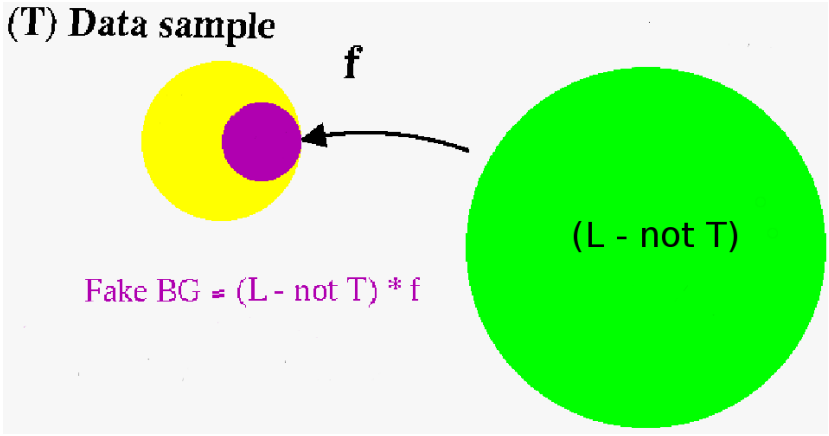


Figure 4.13: Schematical view of the fake background estimation in the (T) final sample of $e-\tau$ or $\mu-\tau$ selected events. The $jet \to \tau$ fake background contribution is statistically predicted from the (L-not T) sample through the $jet \to \tau$ fake rate f.

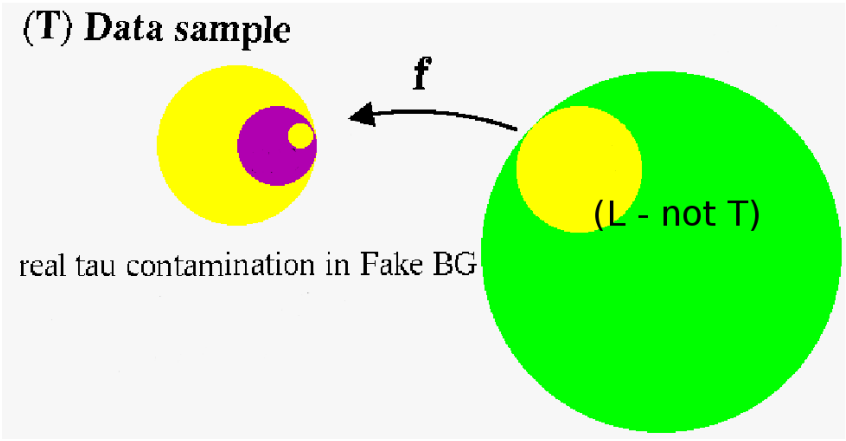


Figure 4.14: Schematical view of the effect on the fake background estimation due to the contamination from real taus in the (L-not T) sample $N^{real\ \tau}_{(L-not\ T)}$ (the yellow region on the right). It shows that the purple region calculated as $f \cdot N_{(L-not\ T)}$ overestimates the real fakes $jet \to \tau$ contribution $N^{jet}_{(T)}$, since the fake rate f multiplies also events in the contamination.

for the high electron and muon identification efficiency (about $80-90\%$, see Section 3.4), with respect to the tau identification efficiency, thus it has been neglected in this study:

$$N_{(L-not\ T)} = N^{real\ \tau}_{(L-not\ T)} + N^{jet}_{(L-not\ T)} + N^{l}_{(L-not\ T)} \approx N^{real\ \tau}_{(L-not\ T)} + N^{jet}_{(L-not\ T)} \quad (4.11)$$
$$N_{(T)} = N^{real\ \tau}_{(T)} + N^{jet}_{(T)} + N^{l}_{(T)} \approx N^{real\ \tau}_{(T)} + N^{jet}_{(T)}$$

If a contamination with real taus is present and correctly taken into accounts in the (L-not T) sample, the fake background $N^{jet}_{(T)}$ is expressed, using (4.11) and (4.10), as:

$$N^{jet}_{(T)} = \frac{f}{\epsilon - f}(\epsilon \cdot N_{(L-not\ T)} - N_{(T)}) \quad < \quad f \cdot N_{(L-not\ T)} \quad (4.12)$$

The uncertainty on the final Fakes contribution $N_{(T)}^{jet}$ due to the error on the fake rate f calculation has been compared to the uncertainty due to ignoring the real tau contamination in the (L-not T) sample, in order to evaluate how the contamination with real taus affects the final results, when the fake background is estimated using the relation $N_{(T)}^{jet} = f \cdot N_{(L-not\ T)}$.

Fake rate considerations for data-driven method

The fake rate as a function of the E_T of the identified tau lepton is shown in Figure 4.15 for all the jet triggered samples (JET_20, JET_50, JET_70 and JET_100) considering 1-prong and 3-prong taus together. The mean fake rate value is calculated as a weighted

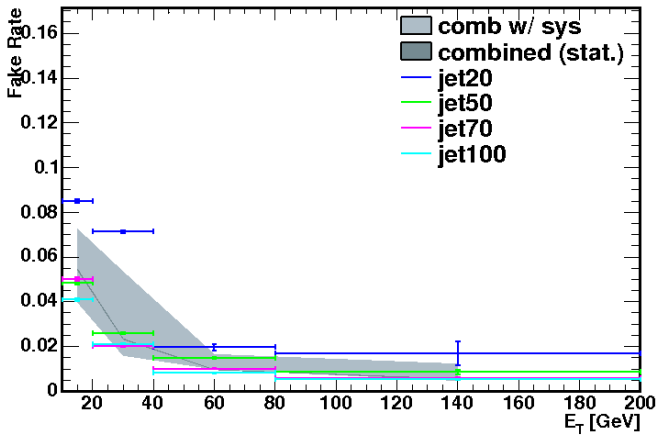


Figure 4.15: Fake rates for $jet \rightarrow \tau$, using samples of jets selected from data using JET_20, JET_50, JET_70 and JET_100 triggers.

mean of the f values from each jet triggered sample, and it shows an uncertainty of about 30-50% due to the spread among the fake rates on the various jet samples. The expected numbers of W and Z in the triggered samples are evaluated with the simulation and subtructed out, in order to correct for the fraction of real leptons in the jet data. In Table 4.7 the fake rate is parameterized in terms of the reconstructed tau energy, in four different ranges.

$E_T^{fake\ tau}$	10-20 GeV	20-40 GeV	40-80 GeV	80-200 GeV
$\bar{f}$	0.0555 ± 0.0004	0.0232 ± 0.0001	0.0097 ± 0.0001	0.0053 ± 0.0001

Table 4.7: The mean fake rate value in four different ranges in the E_T of the reconstructed fake tau. In each bin of energy, the fake rate value is calculated as a weighted mean of the f values from each jet sample. Statistical errors are reported.

For Fake background $N_{(T)}^{jet}$ estimation, the mean fake rate f is applied to jets in the $N_{(L-not\ T)}$ sample in the full spectrum of energy, therefore it strongly relays on how better the mean value approximates the correct one for the fake rate.
The fake rate estimation has been checked selecting events with $e - \tau$ opposite-sign

leptons in different ranges of energy. In Figures 4.16 the distribution of the dilepton invariant mass from data events of 4.8 fb^{-1}, after the base selection has been applied, is compared to the background prediction. The Fakes contribution is estimated from data applying the calculated mean fake rate. A cut in the missing energy $\not\!\!E_T < 25$ GeV is also applied to select a sample mostly dominate by QCD jets. Figures 4.16 clearly shows that the $jet \to \tau$ background is not properly estimated, mainly for events with low E_T taus. Therefore, the strong dependence of the fake rate on the jet-samples effects also the fakes background estimation in a not negligible way.

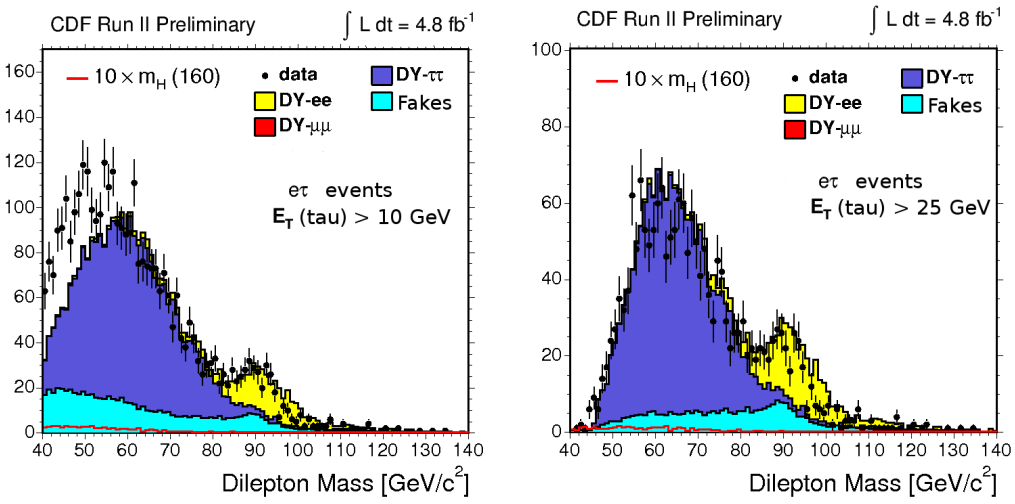


Figure 4.16: The invariant mass distribution of the $e-\tau$ leptons in events collected at 4.8 fb^{-1} integrated luminosity. The fakes contribution is estimated from data applying the estimated mean fake rate (see Figure 4.15). Two different ranges of energy have been considered: $E_T^\tau > 10$ GeV (left) and $E_T^\tau > 25$ GeV (right).

Different method in the fake rate calculation have been developed in previous analysis [115], modifying the fake rate parameterization. The jet sample dependence can be reduced calculating the fake rate as a function of additional quantities able to distinguish different jet event topologies. This new parameterization has not been developed in this Thesis, deferring it to a future improvement of the analysis using data-driven method for the $jet \to \tau$ background estimation.
The fake rate calculation parameterized in terms of the energy only is not a good estimation of the correct fake rate for energy below 15-20 GeV and it is a limiting factor to the applicability of the data-driven method in this analysis, where $jet \to \tau$ fakes background contribution is dominant.

Real tau contamination for data-driven method

In this section an estimate of the real τ contamination in the (L-not T) sample is given to verify the assumption that the real τ contamination in the (L-not T) sample is negligible and that the estimation of Fakes background contribution $N_{(T)}^{jet}$ from $N_{(L-not\ T)}$ can be feasible. A rough estimation of the real τ contamination is done in the $(e-\tau)$ and $(\mu-\tau)$ final states, with minimal requirements applied:

- invariant mass: $M(l\tau) \geq 20$ GeV/c^2

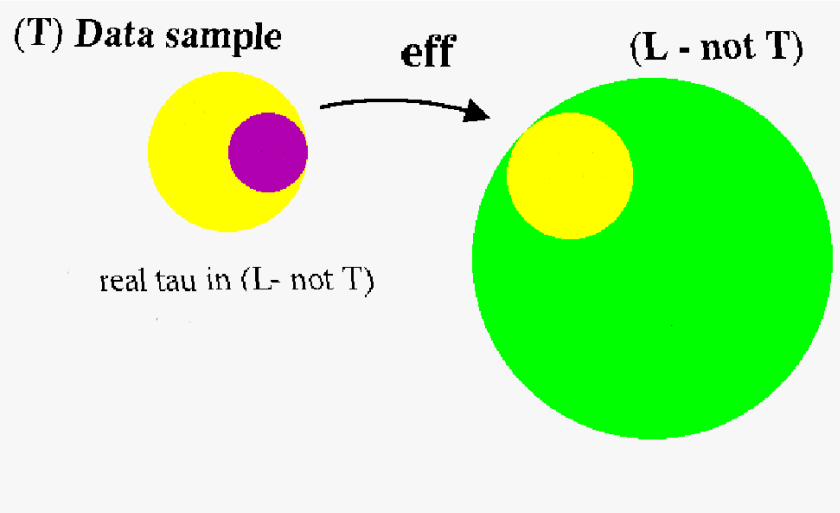


Figure 4.17: Schematical view of the method used to estimate the contamination from taus in the (L-not T) sample.

- separation between leptons in the $\eta - \phi$ plane: $\Delta R > 0.4$

The composition of the (T) sample, where the τ lepton passes all identification cuts, is shown in Table 4.8 for the two channel separately. Using the data-driven method for the prediction of the Fakes contribution, the predicted total background with respect to the number of events selected from 4.8 fb^{-1} of data results overestimated.

CDF RUN II Preliminary $\int L\ dt = 4.8$ fb^{-1}

Background contribution	$e - \tau$	$\mu - \tau$
$Z \to \tau\tau$	3424.2	2355.2
$Z \to ee$	463.0	0.0
$Z \to \mu\mu$	1	524.6
$t\bar{t}$	30.1	20.9
Diboson (WW, WZ, ZZ)	41.6	26.39
Fakes	5835.5	1460.3
Total BG	9795.4	4387.39
Data	8194	4414

Table 4.8: Sample composition in the $(e - \tau)$ and $(\mu - \tau)$ channel separately. The data-driven method for Fakes contribution has been used.

The real tau contamination in the (L-not T) sample is statistically predicted from the physics background contributions in the signal region (T), with the $Z \to \tau\tau$ contributing the most among them:

$$N^{real\ \tau}_{(L-notT)} = \frac{N^{real\ \tau}_{(T)}}{\epsilon} \tag{4.13}$$

In Figure 4.17 is shown the schematical view of the method used to estimate the real tau contamination, considering the efficiency ϵ of tau identification defined as in (4.10). The ϵ tau efficiency with respect to the (L-not t) region has been estimated using MC

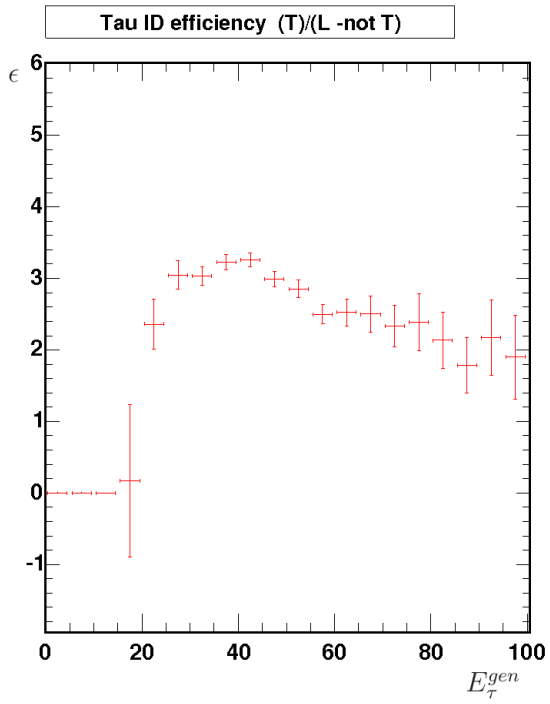


Figure 4.18: Efficiency ϵ (4.10) as a function of the energy E_τ^{gen} of the tau at generation, using a sample of simulated tau from MC sample. The efficiency is greater then one because it is defined with respect to the (L-not T) sample.

simulated taus as a function of the energy E_τ^{gen} of the tau generated at generation and shown in Figure 4.18. Given the energy dependence of the fake rate (discussed above) and of the tau efficiency ϵ, the real tau contamination has been studied in the (L-not T) sample dividing the events in four sets according to the reconstructed tau's energy: 10-20, 20-50, 50-70 and 70-200 GeV. In Table 4.9 is reported the tau efficiency in the (L-not T) sample for each range of tau's energy.

E_τ^{gen}	10-20 GeV	20-50 GeV	50-70 GeV	70-200 GeV
ϵ	0.08 ±0.09	3.1±0.1	2.9 ±0.1	2.0 ±0.2

Table 4.9: Tau efficiency estimated for different ranges in the tau's energy.

All contributions to the real tau contamination in the (L-not T) sample for the two channels are shown in Tables 4.10 and 4.11. Those are obtained weighting in each bin of tau energy the contribution from physics background in (T) (see Table 4.8) by the efficiency ϵ (4.10), according to the relation (4.13). The (L-not T) sample is estimated from 4.8 fb^{-1} of collected data, triggered by an high-p_T electron or muon and with a reconstructed tau lepton not passing all identification cuts. The fraction of the total contamination in each bin of energy expressed in percentage is also shown.

In the same range of energy, real τ $N^{real\ \tau}_{(L-not\ T)}$ and jet faking τ $N^{jets}_{(L-not\ T)}$ in the (L-not T) sample are weighted by the same fake rate value, from Table 4.7 to obtain the corresponding values in the (T) sample. Thus, the percentage of the contamination

(L-not T) ($\mu - \tau$ channel)	10-20 GeV	20-50 GeV	50-70 GeV	70-200 GeV
$Z \to \tau\tau$	104.4	583.2	119.4	0.8
$Z \to \mu\mu$	5.5	68.9	110.1	1.1
$Z \to ee$	0.0	0.0	0.0	0.0
$t\bar{t}$	1.0	4.6	2.1	0.2
$WW + WZ + ZZ$	1.6	6.1	1.6	0.1
Total $N^{real\ \tau}_{(L-not\ T)}$	483.7	662.8	233.2	2.2
$N_{(L-notT)}$	968	4201	2803	216
$N^{real\ \tau}_{(L-not\ T)}/N_{(L-not\ T)}$	50%	15.8%	8.3%	1%

Table 4.10: Real τ contamination in the (L-not T) sample statistically predicted from each physics background in the signal region (T) using (4.13). The contribution is parameterized in terms of the reconstructed τ $E_T(\text{trk}+\pi^0)$ energy, in four ranges of energy. The $N_{(L-not\ T)}$ number of events has been estimated from 4.8 fb^{-1} of collected data with the reconstructed tau not passing all identification cuts. The last raw shows the percentage of the total contamination in different ranges of energy.

(L-not T) ($e - \tau$ channel)	10-20 GeV	20-40 GeV	40-80 GeV	80-200 GeV
$Z \to \tau\tau$	156.1	846.3	170.1	1.2
$Z \to \mu\mu$	0.0	0.1	0.3	0.0
$Z \to ee$	10.6	65.3	83.2	1.72
$t\bar{t}$	1.6	6.4	3.1	0.2
$WW + WZ + ZZ$	2.7	9.7	2.2	0.1
Total $N^{real\ \tau}_{(L-not\ T)}$	171	927.8	258.8	3.2
$N_{(L-notT)}$	5407	17713	8796	722
$N^{real\ \tau}_{(L-not\ T)}/N_{(L-not\ T)}$	3.2%	5.3%	1.8%	0.4%

Table 4.11: Real τ contamination in the (L-not T) sample statistically predicted from each physics background in the signal region (T) using (4.13). The contribution is parameterized in terms of the reconstructed τ $E_T(\text{trk}+\pi^0)$ energy, in four ranges of energy. The $N_{(L-not\ T)}$ number of events has been estimated from 4.8 fb^{-1} of collected data with the reconstructed tau not passing all identification cuts. The last raw shows the percentage of the total contamination in different ranges of energy.

in $N^{jet}_{(T)}$ is the same as in (L-not T) from Tables 4.10 and 4.11. The uncertainty on the $N^{jet}_{(T)}$ estimation is dominated by the uncertainty on the fake rate $\sim 48\%$, however the contamination from real taus is not negligible and the validity of the data-driven method relays on the ability to discriminate jets from taus in the (L-notT) sample, in order to reduce the real tau contamination.

Although the data-driven method allows to account for all possible processes contributing with $jet \to \tau$ fakes to the background and to avoid problems with the MC mismodeling, the uncertainty on the fake background estimation results to be high. Therefore the fake background estimation exploiting MC simulated events has been chosen in this Thesis.

Chapter 5

$H \to e/\mu + \tau$ signal selection and background estimation

In this Chapter, the Higgs event rate and topology will be predicted. The sample composition have been mostly studied with MC simulated events, paying strong attention to the accurate Monte Carlo modeling of signal and background processes. Both measured and simulated objects are subject to the same event reconstruction algorithms for the $(e - \tau)$ or $(\mu - \tau)$ events selection.

5.1 Study of the sample composition

The four Higgs production mechanism considered, which may contribute to the signal, are the direct Higgs production via gluon fusion $gg \to H$, the associated production with the W and Z boson and vector boson fusion. Twelve Higgs signal samples have been used, which only differ by the Higgs mass hypothesis. To probe the $130 \leq M_H \leq 200$ GeV/c^2 range, the Higgs mass values have been chosen with a step of 10 GeV/c^2, and a step of 5 GeV/c^2 is used in the region around $M_H = 160$ GeV/c^2, where the larger number of signal events is expected.

In the samples of events passing the base selection, defined in Section 4.1, the highest

Process	Generator	$\sigma \times Br$ pb	K-factor	Run Period
$gg \to H \to WW$	PYTHIA	0.396	1.0	0-23
$qq' \to WH \to WWW$	PYTHIA	0.046	1.0	0-23
$q\bar{q} \to ZH \to WWZ$	PYTHIA	0.030	1.0	0-23
$q\bar{q}(') \to q\bar{q}(')H \to q\bar{q}(')WW$	PYTHIA	0.042	1.0	0-23

Table 5.1: List of MC simulated processes for signal contributions. The event generator, the cross-section times the branching fraction for $M_H = 160$ GeV/c^2 used in the normalization and the run-period on which the simulation was tuned are also reported. Other Higgs samples for different Higgs mass hypothesis only differ by the cross section and Branching ratio used in the normalization.

contribution to the signal is from direct production. In 4.8 fb^{-1} of collected data and Higgs mass hypothesis of 160 GeV/c^2, it is expected to be 0.846 and 0.612 respectively

in the $e - \tau$ and $\mu - \tau$ channels, while the expected background and observed yields are shown in Table 5.2. In order to devise the analysis strategy and improve the sensitivity to the Higgs signal, tightest cuts will be applied, exploiting differences in the kinematics of the $H \to WW \to l\nu\tau\nu$ process compared to backgrounds.

CDF RUN II Preliminary $\int L\ dt = 4.8$ fb^{-1}

Process	$e - \tau$	$\mu - \tau$
$Z \to \tau\tau$	2417	1701
$Z \to ee$	747	0
$Z \to \mu\mu$	0.89	420
WW	31	22
WZ	0.63	0.45
ZZ	1.50	0.98
$W\gamma$	5	3
$t\bar{t}$	29	20
$W + jet$	1543	1004
QCD	2091	181
Tot BG	6866	3352
Data	6722	3064

Table 5.2: Sample composition expected in 4.8 fb^{-1} after the base selection (defined in Section 4.1). The expected and observed number of events in the $e - \tau$ and $\mu - \tau$ final states are given separately.

5.1.1 Events selection for signal optimization

The background reduction exploits the spin correlation among the decay products of the Higgs in the dilepton final state. The Higgs is a particle with spin-0 which in this case decays into two spin-1 vector particles (WW pair) and their spins must be anti-aligned in the Higgs rest frame. The couplings of the leptons to Ws allows for only left-handed neutrinos and right-handed antineutrinos. As a result of these couplings the two neutrinos, as for other leptons (e, μ, τ), are preferentially emitted in the same direction and hence the separation angle $\Delta\phi(l, \tau)$ of the charged leptons will tend to be small in signal events, while background events tend to have the back-to-back charged leptons in the final state. Thus, the expected angular distribution of reconstructed leptons in $W + jets$ and non-resonant diboson events can be used to distinguish them from signal as can be seen from the distribution in Figure 5.1 of the expected and observed number of events as a function of the $\Delta\phi(l, \tau)$, before cutting on this variable.

After the base selection is applied, the data sample is dominated by Drell-Yan events (see Table 5.2). The event topology and the correlation between the missing energy and the lepton directions in $Z/\gamma \to \tau\tau$ events can be exploited to separate the $Z/\gamma^* \to \tau\tau$ Drell-Yan process from signal. In Figure 5.2 is shown the expected and observed distribution in $\Delta\phi(\not\!\!E_T, l - \tau)$, defined as the distance in the azimuthal plane between the missing energy and the system of the two reconstructed leptons $l - \tau$. All the background processes are stacked on top of each other, and the contribution of a 160 GeV/c^2 Higgs boson with yields scaled up by a factor of 10^3 is overlaid. In $Z/\gamma^* \to \tau\tau$ events the missing transverse energy $\not\!\!E_T$ likely comes from a mis-measurement of the

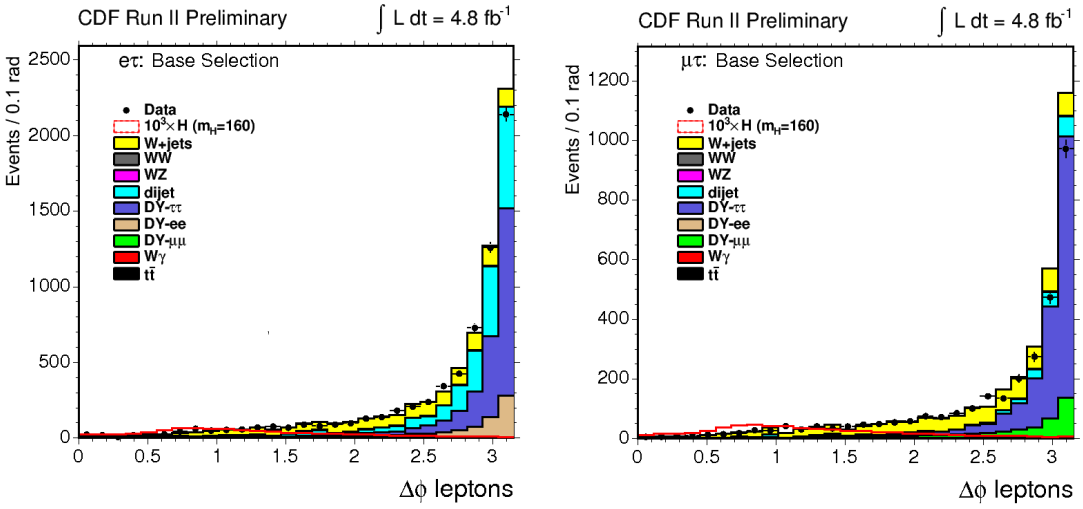


Figure 5.1: The expected and observed distribution of the number of events as a function of the $\Delta\phi(l,\tau)$ for the $e-\tau$ (left) and $\mu-\tau$ (right) channels passing the base selections and before any $\Delta\phi(l,\tau)$ requirement is applied. The sum of the direct production, the associated production and the vector boson fusion signal contributions is overlaid. The signal events are scaled up by a factor of 10^3 and a $M_H = 160$ GeV/c^2 is assumed.

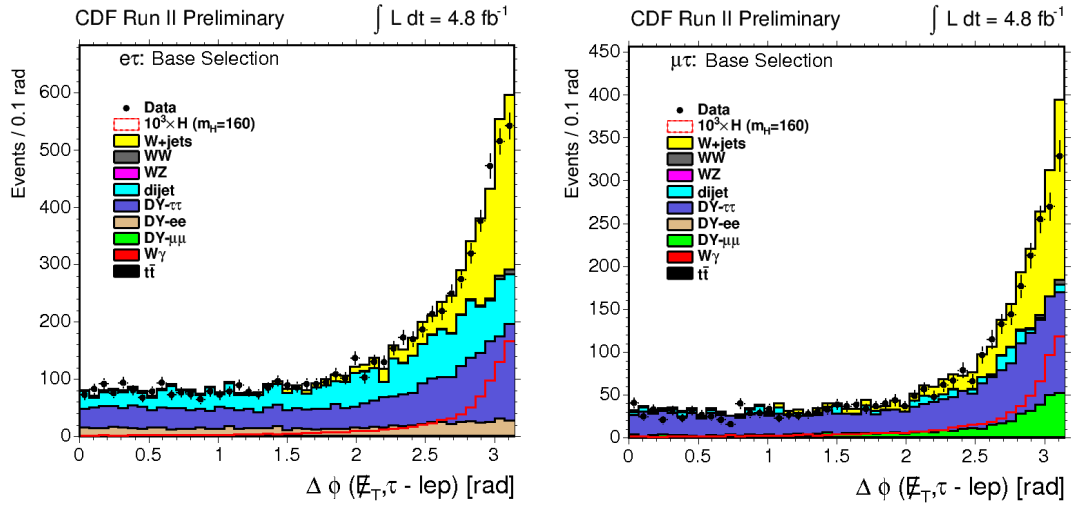


Figure 5.2: The expected and observed $\Delta\phi(\not\!\!E, l-\tau)$ distribution for the selected sample of $e-\tau$ (left) and $\mu-\tau$ (right) events, before the $\Delta\phi(\not\!\!E, l-\tau)$ requirement and in 4.8 fb^{-1} of integrated luminosity. The sum of the direct production, the associated production and the vector boson fusion signal contributions is overlaid. The signal events are scaled up by a factor of 10^3 and a $M_H = 160$ GeV/c^2 is assumed.

leptons, while in signal events the two neutrinos have a small angular separation $\Delta\phi$ giving an high transverse energy imbalance in signal events. This can be seen from the correlation between the missing transverse energy $\not\!\!E_T$ and the angle separation between $\not\!\!E_T$ and the $l-\tau$ system $\Delta\phi(\not\!\!E, l-\tau)$, shown in Figure 5.3 for the $\mu-\tau$ channel. Therefore, rectangular cuts on the missing transverse energy $\not\!\!E_T$ and the angle separation between

$\not\!\!E_T$ and the $l - \tau$ system in the event have been applied, defined in order to get the best separation from the signal and the $Z/\gamma^* \to \tau\tau$ background.

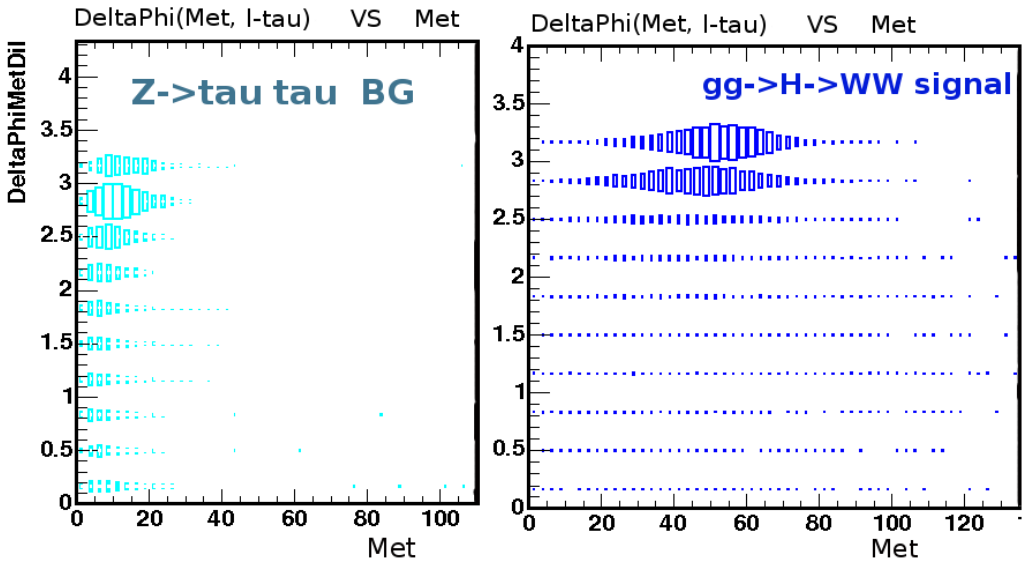


Figure 5.3: Correlation between the missing transverse energy $\not\!\!E_T$ and the angle separation between $\not\!\!E_T$ and the $l - \tau$ system for $Z/\gamma^* \to \tau\tau$ events (left) and signal $gg \to H \to WW$ events, with $M_H = 160$ GeV/c^2 (right). Here only $(\mu - \tau)$ events are selected, $(e - \tau)$ events show similar distributions.

In both $(e-\tau)$ and $(\mu-\tau)$ channels, the $\not\!\!E_T > 20$ GeV cut is expected to keep about 95% of the signal, reducing down to about 20% of the $Z/\gamma^* \to \tau\tau$ contribution. The signal and $Z/\gamma^* \to \tau\tau$ acceptance as a function of the cut on the missing energy in the event is shown in Figure 5.4 for the $(\mu - \tau)$ channel. The fraction of signal rejected by the requirement on $\Delta\phi(\not\!\!E, l - \tau) > 1.5$ rad alone is between 5% and 14% depending on its mass, to be compared with the $\sim 30\%$ rejection of $Z/\gamma^* \to \tau\tau$ events.

In the $r - \phi$ plane the two τ are expected to be nearly back-to-back with the missing energy along their same direction. Therefore, the $\Delta\phi_{l\not\!\!E} + \Delta\phi_{\tau\not\!\!E}$ is nearly equal to the $\Delta\phi(l\ \tau)$ (see Figure 5.5). This can be exploited to set an additional angular requirement to reject $Z \to \tau\tau$ events. A cut on the angular separation between leptons, $\Delta\phi(l\ \tau) < 1.5$ rad, results efficient in reducing the DY background. It rejects about 97% of the $Z/\gamma^* \to \tau\tau$ events, although between 10% and 20% of the signal is rejected depending on the Higgs mass hypothesis.

A cut on the missing transverse energy helps also in separating signal from QCD processes, while the correlation between the missing energy and $\Delta\phi(l, \tau)$ in $W + jets$ event, shown in Figure 5.6, has been exploited to separate signal from $W + jets$ event.

The set of cut optimizing the signal selection and background rejection defines the *signal region* as follow:

- $M_{\tau l} > 20$ GeV/c^2 , to suppress the low mass resonances;
- $\not\!\!E > 20$ GeV, to suppress the QCD background;

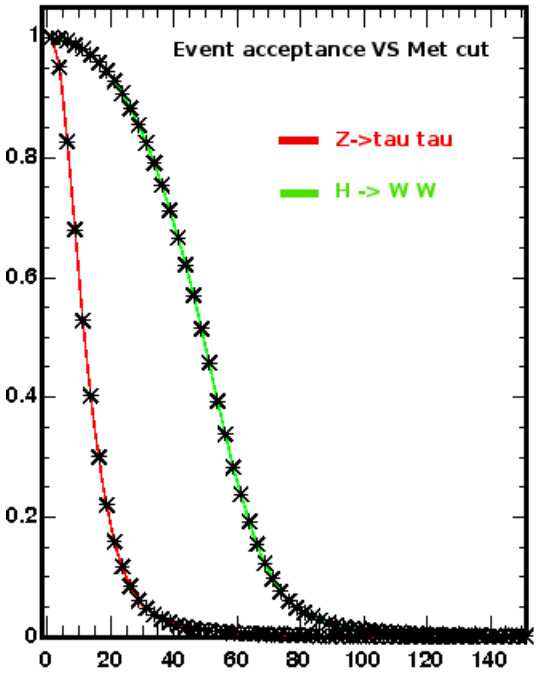


Figure 5.4: Signal $H \to WW$ (green) and the $Z/\gamma^* \to \tau\tau$ (red) acceptance as a function of the cut on the missing energy. From a sample of $H \to WW$ and $Z/\gamma^* \to \tau\tau$ MC simulated events. Here only $(\mu - \tau)$ events are selected, $(e - \tau)$ events show similar distributions.

.

- $\Delta\phi_{\tau l} < 1.5$ rad, to suppress the W+jet and WW components;
- $\Delta\phi_{\not{E},l-\tau} > 1.5$ rad, to suppress the $Z/\gamma^* \to \tau\tau$ contamination.

The expected number of signal events in the signal region is estimated at different Higgs mass hypotheses and listed in Table 5.3. The expected yields for all the processes contributing to the $e - \tau$ and $\mu - \tau$ channels in 4.8 fb^{-1} of integrated luminosity are reported in Table 5.4 .

Figures 5.7 and 5.8 show the invariant mass and the angular separation between leptons, in e-τ and μ-τ channels respectively. Additional distributions are reported in Appendix A. The background is stacked on top of each other, and for comparison the contribution of a 160 GeV/c^2 Higgs boson with yields multiplied by a factor of 100 is overplaied. All processes from MC simulation are corrected by the fake rate scale factor estimated in Sections 4.3.1. Observed data are in well agreement with the SM background expectations. The background modeling and normalization are tested in *control samples*, each of them targeting one specific background process.

5.2 Background control region and normalization

As a check that the data is being well modeled by the MC simulation it is desirable to have *control regions* which are independent of the signal region. Several distinct selection

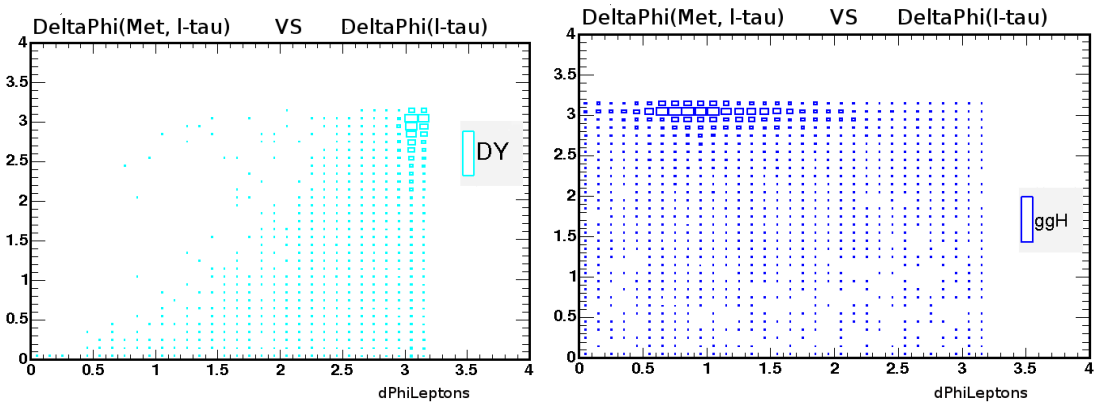


Figure 5.5: The expected and observed $\Delta\phi(l-\tau)$ distribution for the selected sample before the $\Delta\phi(l-\tau)$ requirement. From a sample of $H \to WW$ and $Z/\gamma^* \to \tau\tau$ MC simulated events. Here only $(\mu-\tau)$ events are selected, $(e-\tau)$ events show similar distributions.

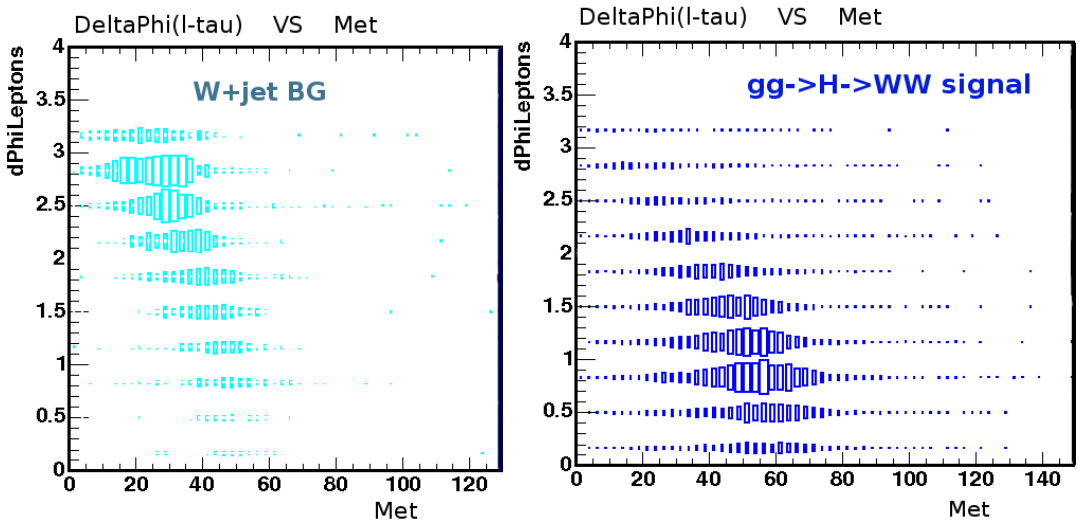


Figure 5.6: Correlation between the missing transverse energy and the angle separation between leptons for $W + jets$ events (left) and signal $gg \to H \to WW$ events, with $M_H = 160$ GeV/c^2 (right).

criteria (regions) are discussed below.
The control samples are selected exploiting the kinematics and topological properties of background against signal events:

- $W + jet$ region:
  $M_{l\tau} > 20$ GeV/c^2, $\not\!\!E_T > 20$ GeV, $\Delta\phi_{\tau l} > 2$ rad
- QCD region:
  $M_{l\tau} > 20$ GeV/c^2, $\not\!\!E_T < 20$ GeV
- $Z/\gamma^* \to \tau\tau$ region:
  $M_{l\tau} > 20$ GeV/c^2, $\not\!\!E_T < 20$ GeV, $\Delta\phi(\not\!\!E_T, l-\tau) < 0.5$ rad

M_H $(gg \to H)$ GeV/c^2	$e - \tau$	$\mu - \tau$
130	0.135 ± 0.002	0.095 ± 0.002
140	0.287 ± 0.004	0.206 ± 0.003
145	0.377 ± 0.005	0.260 ± 0.004
150	0.471 ± 0.006	0.335 ± 0.004
155	0.590 ± 0.006	0.413 ± 0.005
160	0.659 ± 0.007	0.470 ± 0.005
165	0.748 ± 0.007	0.541 ± 0.006
170	0.717 ± 0.007	0.509 ± 0.005
175	0.662 ± 0.006	0.471 ± 0.005
180	0.608 ± 0.006	0.432 ± 0.004
190	0.426 ± 0.004	0.292 ± 0.003
200	0.326 ± 0.003	0.302 ± 0.003

Table 5.3: Expected number of $gg \to H$ signal events for several Higgs mass hypotheses, for the $(e - \tau)$ and $(\mu - \tau)$ channels separately.

CDF RUN II Preliminary $\int L\ dt = 4.8$ fb^{-1}

Process	Yield (e-τ) Signal Region	Yield (μ-τ) Signal Region
$Z \to \tau\tau$	2.22 ± 0.71	1.7 ± 0.5
$Z \to ee$	9.73 ± 1.91	0.0
$Z \to \mu\mu$	0.33 ± 0.29	62.4 ± 9.1
WW	8.66 ± 1.27	6.12 ± 0.91
WZ	0.16 ± 0.02	0.14 ± 0.02
ZZ	0.20 ± 0.03	0.22 ± 0.03
$W\gamma$	1.7 ± 0.2	1.03 ± 0.14
$t\bar{t}$	7.9 ± 1.3	5.2 ± 0.8
$W + jets$	577 ± 96	383 ± 64
di-jet, γ+jet	36.5 ± 18.0	3.5 ± 16.1
Tot BG	645 ± 98	463 ± 67
$gg \to H$ $[M_H = 160$ GeV/c$^2]$	0.659 ± 0.105	0.470 ± 0.075
WH $[M_H = 160$ GeV/c$^2]$	0.169 ± 0.023	0.018 ± 0.003
ZH $[M_H = 160$ GeV/c$^2]$	0.099 ± 0.014	0.069 ± 0.010
VBF $[M_H = 160$ GeV/c$^2]$	0.052 ± 0.008	0.038 ± 0.006
Total signal	0.979 ± 0.109	0.595 ± 0.076
Data	636 ± 25	455 ± 21

Table 5.4: Expected signal and background events in the *signal region* for the opposite sign di-lepton $e - \tau$ and $\mu - \tau$ samples.

5.2.1 $W + jet$ control region

The $W + jet$ region:

- $M_{l\tau} > 20$ GeV/c^2, $\not\!\!E_T > 20$ GeV, $\Delta\phi_{\tau l} > 2$ rad

is defined in order to select a sample of events where the dominant contribution comes from the $W + jet$ process. The comparison between observed data and MC allows to

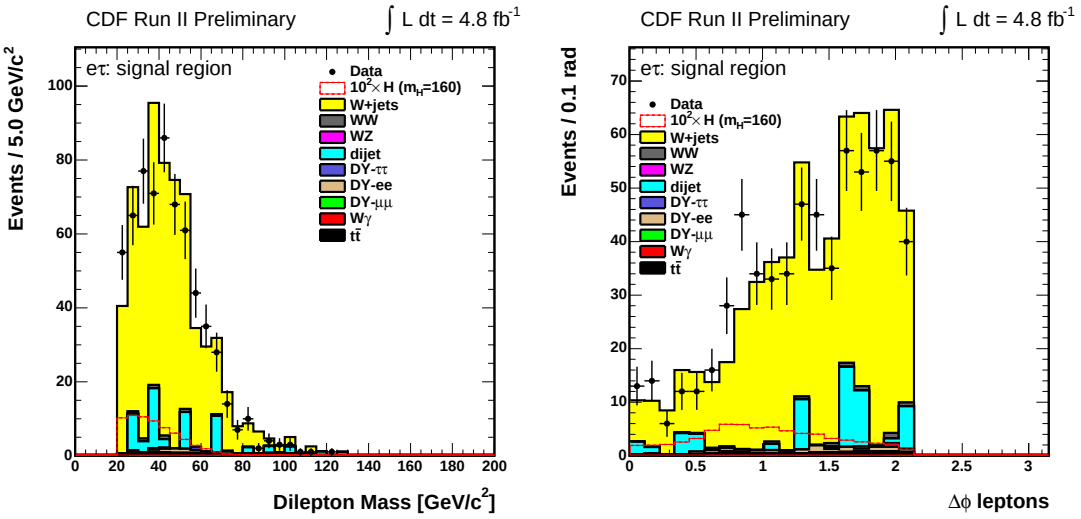


Figure 5.7: Distributions of the invariant mass (left) and $\Delta\phi(l-\tau)$ angle separation between the two leptons (right), in the opposite-sign signal region for events with the $e-\tau$ final state. All background processes are stacked on top of each other. The signal contribution (red line) is overlaid and multiplied by a factor 100. A $m_H = 160$ GeV/c^2 is assumed.

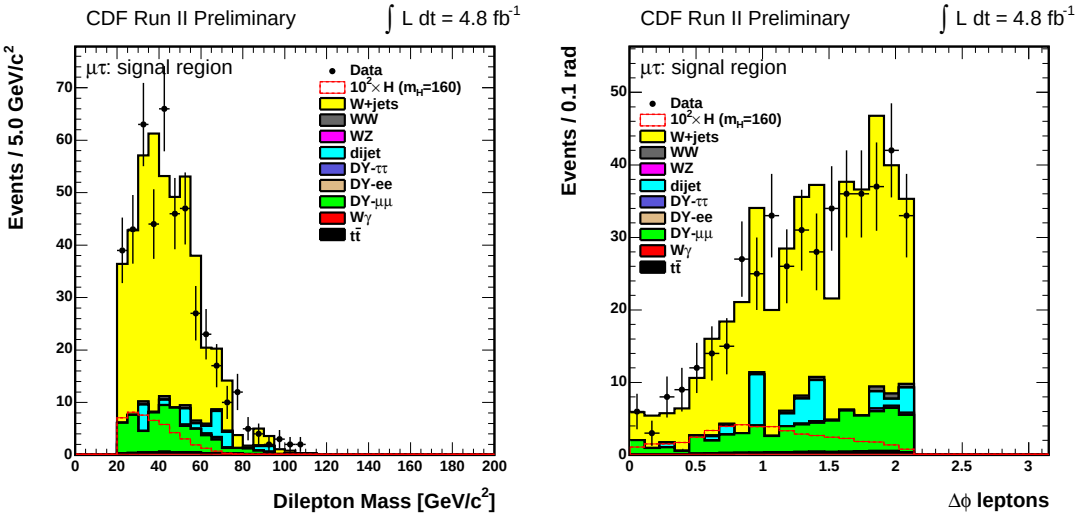


Figure 5.8: Distributions of the invariant mass (left) and $\Delta\phi(l-\tau)$ angle separation between the two leptons (right), in the opposite-sign signal region for events with the $\mu-\tau$ final state. All background processes are stacked on top of each other. The signal contribution (red line) is overlaid and multiplied by a factor 100. A $m_H = 160$ GeV/c^2 is assumed.

validate the $jet \rightarrow \tau$ fake rate scale factor applied to reconstructed τ in the $W + jets$ MC samples and verify the normalization of the $W + jet$ ALPGEN MC.
Both $e-\tau$ and $\mu-\tau$ channels are considered, and to achieve a more detailed investigation

the events in the control sample are split according to the τ decay type: 1 prong, 1 prong without π^0s and 3 prong. The sub-sample of MC events containing a reconstructed 1 prong τ with no π^0s among its decay products is fit to data by rescaling the $W + jet$ contribution by the 0.64 ± 0.07 scale factor. This correction to the normalization of the $W + jet$ ALPGEN MC is applied to all control regions and signal region, in both $e - \tau$ and $\mu - \tau$ channels in the 1 prong with no π^0 sub-sample. No correction needed for events with 3 prong or 1 prong with π^0 taus.
The comparison of expected and observed number of events in $W + jets$ control region is shown in Table 5.5. In Figure 5.9 and 5.10 are shown the distributions of two of the quantities used to define the $W + jets$ control region: invariant mass distribution and the angle separation among the reconstructed leptons in the $e - \tau$ and $\mu - \tau$ channels. Additional distributions can be found in Appendix B. Checks in this region demonstrate that estimates of the fake probabilities are well understood.

CDF RUN II Preliminary $\int L\ dt = 4.8$ fb^{-1}

Process	Yield (e-τ) $W + jets$ CR	Yield (μ-τ) $W + jets$ CR
$Z \to \tau\tau$	350.2 ± 46.7	235.4 ± 31.9
$Z \to ee$	33.1 ± 5.1	0.0
$Z \to \mu\mu$	0.2 ± 0.1	128 ± 18
WW	17.3 ± 2.5	11.5 ± 1.7
WZ	0.20 ± 0.03	0.174 ± 0.026
ZZ	0.40 ± 0.06	0.31 ± 0.05
$W\gamma$	2.4 ± 0.3	1.2 ± 0.2
$t\bar{t}$	15.1 ± 2.5	10.2 ± 1.7
$W + jets$	738 ± 123	446 ± 75
di-jet, γ+jet	54 ± 20	4 ± 17
Tot BG	1210 ± 133	837 ± 85
$gg \to H$ [$M_H = 160$ GeV/c^2]	0.091 ± 0.015	0.068 ± 0.011
WH [$M_H = 160$ GeV/c^2]	0.135 ± 0.018	0.094 ± 0.013
ZH [$M_H = 160$ GeV/c^2]	0.054 ± 0.007	0.040 ± 0.006
VBF [$M_H = 160$ GeV/c^2]	0.011 ± 0.002	0.008 ± 0.002
Total signal	0.291 ± 0.025	0.210 ± 0.018
Data	1159 ± 34	776 ± 28

Table 5.5: Expected signal and background events in the $W + jets$ control region for the opposite-sign dilepton $e - \tau$ and $\mu - \tau$ samples.

5.2.2 QCD control region

The dijet and γ-jet events are enhanced in low missing transverse energy events, in particular in the $e - \tau$ final state, requiring:

- $M_{l\tau} > 20$ GeV/c^2, $\not\!\!E_T < 20$ GeV

The comparison between observed data and the total background prediction allows to test the modeling of the QCD background, based on the selection of data events with same-sign leptons. The contribution of the electro-weak processes with same-sign leptons

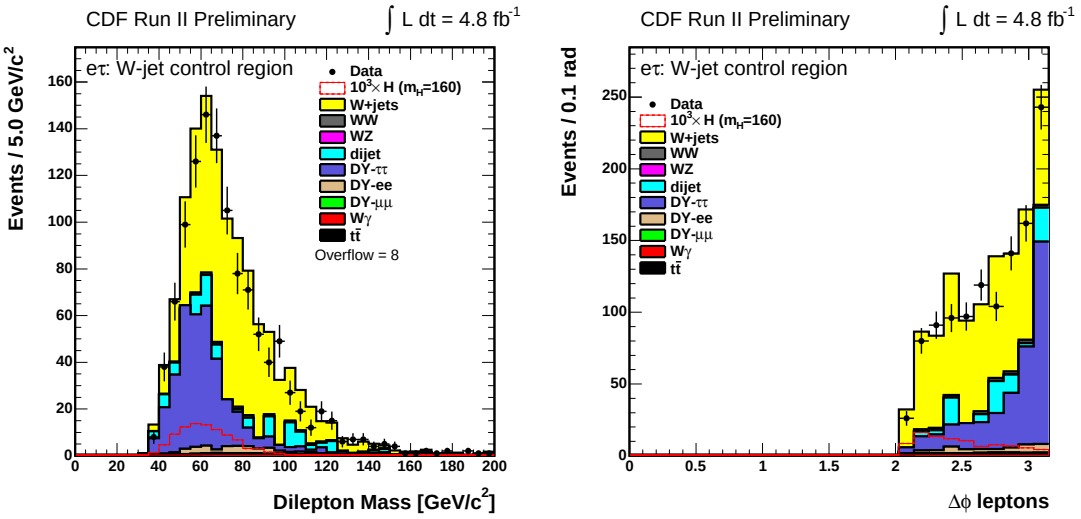


Figure 5.9: Invariant mass (left) and $\Delta\phi(l-\tau)$ angle separation between the two leptons (right) distributions in the opposite-sign $W + jets$ control region for events with the $e - \tau$ final state. All background processes are stacked on top of each other. The signal contribution (red line) is overlaid and multiplied by a factor 10^3. A $M_H = 160$ GeV/c^2 is assumed.

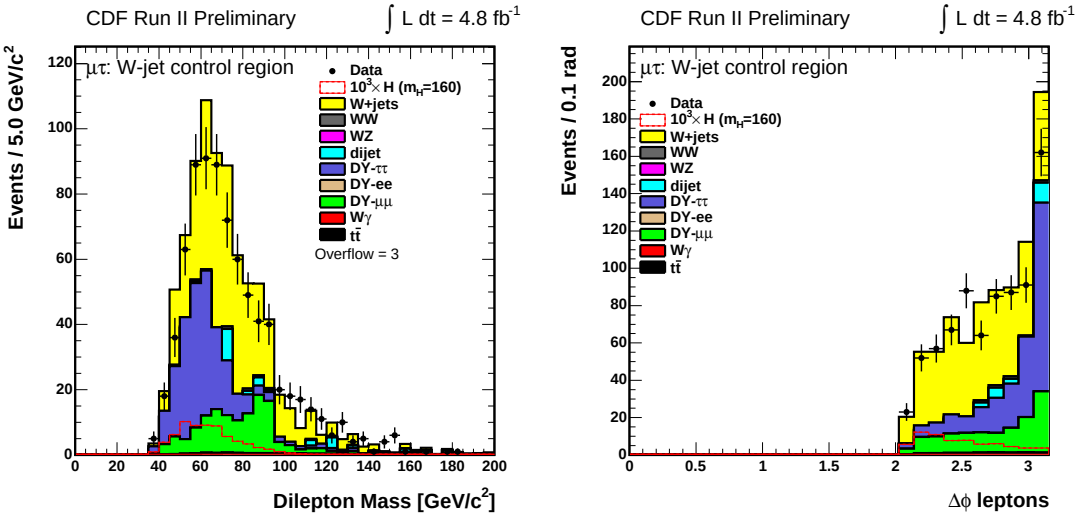


Figure 5.10: Invariant mass (left) and $\Delta\phi(l-\tau)$ angle separation between the two leptons (right) distributions in the opposite-sign $W + jets$ control region for events with the $\mu - \tau$ final state. All background processes are stacked on top of each other. The signal contribution (red line) is overlaid and multiplied by a factor 10^3. A $M_H = 160$ GeV/c^2 is assumed.

is subtracted using the same MC samples as used for estimating the opposite-charged yields. The background yields from MC and observed data in the QCD control region are reported in Table 5.6. In Figure 5.11 are shown the distributions for the quantities

used to define the QCD control region: invariant mass and angular separation between e and τ leptons. Additional distributions are shown in Appendix C. A good agreement in the observed and predicted number of events and kinematic distributions is found.

CDF RUN II Preliminary $\int L\ dt = 4.8\ \text{fb}^{-1}$

Process	Yield (e-τ) QCD CR
$Z \to \tau\tau$	2016.5 ± 265.0
$Z \to ee$	693.2 ± 95.8
$Z \to \mu\mu$	0.35 ± 0.17
WW	5.1 ± 0.8
WZ	0.16 ± 0.02
ZZ	0.7 ± 0.1
$W\gamma$	0.94 ± 0.14
$t\bar{t}$	2.2 ± 0.4
$W + jets$	216.4 ± 36.7
di-jet, γ+jet	1992.5 ± 48.7
Tot BG	4922 ± 288
$gg \to H$ [$M_H = 160$ GeV/c^2]	0.059 ± 0.010
WH [$M_H = 160$ GeV/c^2]	0.049 ± 0.007
ZH [$M_H = 160$ GeV/c^2]	0.032 ± 0.004
VBF [$M_H = 160$ GeV/c^2]	0.008 ± 0.002
Total signal	0.148 ± 0.013
Data	4823 ± 69

Table 5.6: Expected signal and background events in the QCD control region for the opposite-sign dilepton $e - \tau$ samples.

5.2.3 $Z/\gamma^* \to \tau\tau$ control region

Angular and low missing energy requirements:

- $M_{l\tau} > 20$ GeV/c^2, $\not\!\!E_T < 20$ GeV, $\Delta\phi(\not\!\!E_T, l - \tau) < 0.5$ rad

allow to test the $Z/\gamma^* \to \tau\tau$ MC prediction in a selected sample where the $Z/\gamma^* \to \tau\tau$ events are enhanced with respect to $W + jet$ and QCD dominant background, in particular when the $\mu - \tau$ final state is detected. In this control region the τ identification scale factor and the Drell-Yan normalization is verified.
The predicted yields along with the number of observed events from data in the $Z/\gamma^* \to \tau\tau$ control region is shown in Table 5.7. Distributions for the invariant mass and the angular separation between muon and tau, used to define the $Z/\gamma^* \to \tau\tau$ control region, are shown in Figure 5.12. Additional distributions are shown in Appendix D. In general, there is a good agreement between the predicted and observed rates and kinematic shapes.

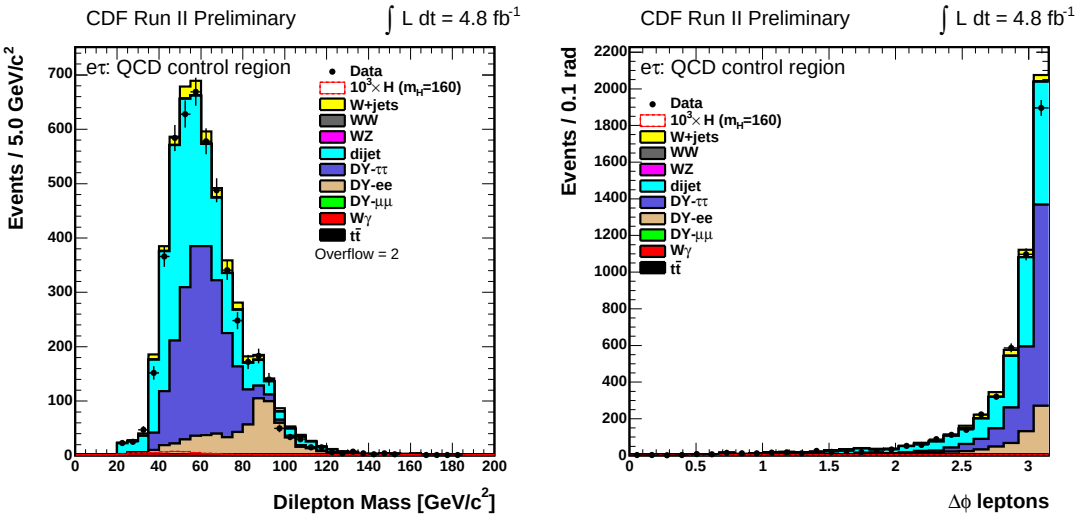


Figure 5.11: Invariant mass (left) and $\Delta\phi(l-\tau)$ angle separation between the two leptons (right) distributions in the opposite-sign QCD control region for events with the $e-\tau$ final state. All background processes are stacked on top of each other. The signal contribution (red line) is overlaid and multiplied by a factor 10^3. A $M_H = 160$ GeV/c^2 is assumed.

CDF RUN II Preliminary	$\int L\ dt = 4.8$ fb^{-1}
Process	Yield (μ-τ) $Z \to \tau\tau$ CR
$Z \to \tau\tau$	467.1 ± 61.8
$Z \to ee$	0.0
$Z \to \mu\mu$	20 ± 3
WW	0.26 ± 0.04
WZ	0.013 ± 0.002
ZZ	0.06 ± 0.01
$W\gamma$	0.02 ± 0.01
$t\bar{t}$	0.21 ± 0.04
$W + jets$	9.9 ± 2.4
di-jet, γ+jet	24.9 ± 6.2
Tot BG	$522.5 \pm 10.2_{stat}$
$gg \to H$ [$M_H = 160$ GeV/c^2]	0.0033 ± 0.0007
WH [$M_H = 160$ GeV/c^2]	0.0031 ± 0.0005
ZH [$M_H = 160$ GeV/c^2]	0.0018 ± 0.0003
VBF [$M_H = 160$ GeV/c^2]	0.0007 ± 0.0002
Total signal	0.009 ± 0.001
Data	465 ± 22

Table 5.7: Expected signal and background events in the $Z/\gamma^* \to \tau\tau$ control region for the opposite-sign dilepton $\mu - \tau$ samples.

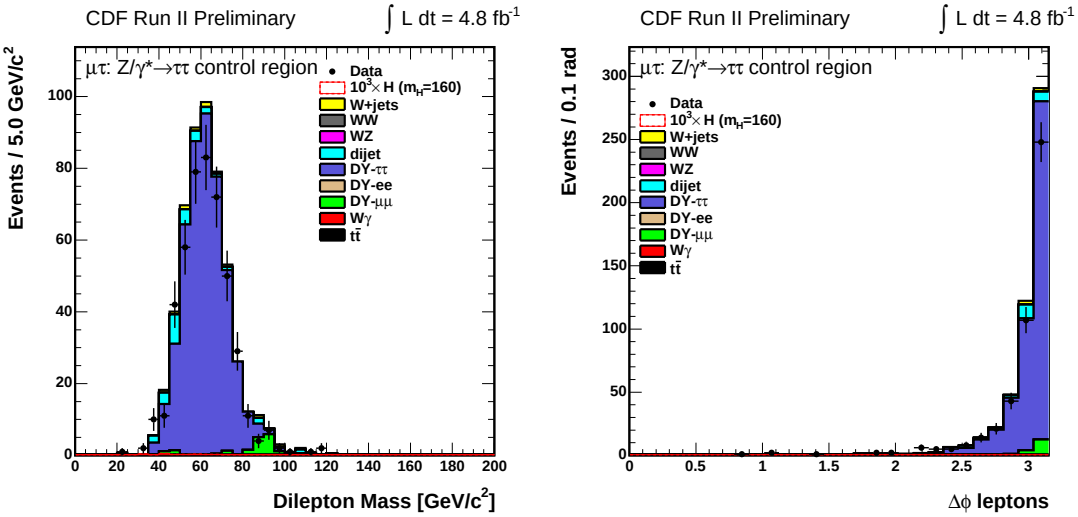


Figure 5.12: Invariant mass (left) and $\Delta\phi(l-\tau)$ angle separation between the two leptons (right) distributions in the opposite-sign $Z/\gamma^* \to \tau\tau$ control region for events with the $\mu - \tau$ final state. All background processes are stacked on top of each other. The signal contribution (red line) is overlaid and multiplied by a factor 10^3. A $M_H = 160$ GeV/c^2 is assumed.

Chapter 6

Higgs signal discrimination and limit results

In the signal region defined in Section 5.1.1. the fake backgrounds ($W + jets$ and $W\gamma$) contribute about 90% to the selected sample and Drell-Yan is important contribution too. From the expected signal (S) and total background (B) in Table 5.4, the signal to background $\frac{S}{\sqrt{B}}$ is less then 1%, therefore the sensitivity in both channels is not enough for a counting experiment. Neural Network advanced techniques exploiting multivariate analysis will be used and described in this Chapter.

6.1 Analysis strategy

In this analysis a Neural Network (NN) technique is used to exploit the differences in the kinematics of the $H \to WW \to l\nu\tau\nu$ process compared to backgrounds: the NN output will be used as the final signal-background discriminant for each $e-\tau$ and $\mu-\tau$ channel. The NN distribution from data events is compared to the MC background predictions, to determine the signal contribution or to quote an allowed upper limit on the SM Higgs boson production cross section.
To optimize the analysis technique and the significance of an experimental result, the expected limit is calculated in advance on MC events in the background-only hypothesis, as a measure of the expected sensitivity of the search, while details about how the limit is computed will be discussed later in Section 6.4. The sensitivity of the analysis depends on how better the NN output discriminates signal from background and, as it will be discussed in Section 6.2.1, this is a consequence of the choice of the NN inputs and NN *training*. Therefore, the NN used in this analysis has been built in order to obtain the minimum expected limit, setting the analysis strategy with the best expected sensitivity to the Higgs signal.

6.2 Neural Network

The problem of assigning to each given event the probability to be compatible with a signal or background event can be solved through a Neural Network technique. This probability is evaluated for the expected and the observed events. NNs are used to process a set of events, receiving as inputs kinematic variables which better describe the

Higgs signal. A *target* is defined, which represents the indicator of the expected true nature of the event: signal-like event (target value set to 1) or background-like event (target value set to -1).
The type of NN used in this analysis consists of 3 *layers* which are made up of *nodes*. An example is shown in Figure 6.1 with a single output node and one layer of *hidden nodes*, which provides flexibility in the event classification rule. The inputs to the first layer are the kinematic variables. The NN architecture have the following features:

- The hidden nodes are arranged in a series of layers
- With the inputs at the bottom and the outputs at the top, the only permissible connections are between nodes in consecutive layers and directed upwards. Consequently, the NN is called *feed-forward network*.

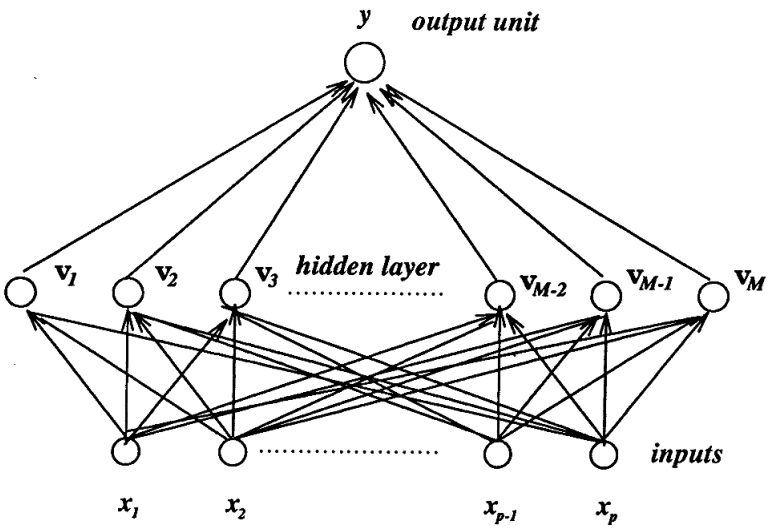


Figure 6.1: A single-output three-layered feed-forward neural network. Neural network general architecture consists of interlinked input, output and hidden nodes.

Each node receives inputs from each of a set of other nodes through internodal connections and transforms a linear combination of its inputs into an output value between -1 and 1. Weights assigned for the inputs to each node are *learned* in *training* the network. For a given set of input variables, the network is trained on a sample of simulated events, called *training sample*, in order to set weights and the network structure, optimized exploiting the network ability to learn the complex relationship between the input variables and the target. The two main aspects in neural network-based analysis are:

- optimize the choice of inputs variables
- specify the architecture of a suitable network

The NN package used in this analysis is the NeuroBayes Neural Network [116, 117]. It is a Bayesian NN, which provides for output a probability distribution of the target. Signal and background events are randomly divided in two sub-samples: the first one is used for training, and the second one is used to check the distribution of the NN output.

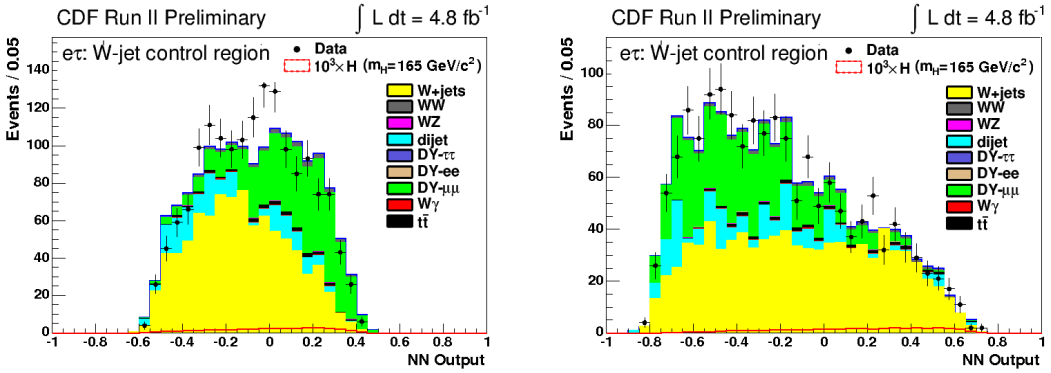


Figure 6.2: NN output in the $W+jets$ $e\tau$ control region for the Higgs mass 165 GeV/c^2. Only tau ID (left) or event (right) related input variables have used.

Once the NN is trained it can be used taking the same inputs from any data and gives output ranging from -1 and 1.
In this Thesis, different sets of input variables have been investigated, choosing a limited number of variables that statistically maximize the separation between signal and background events. One NN is trained at each of the 12 masses hypothesis investigated, for $e-\tau$ and $\mu-\tau$ channels separately, using a sample of $gg \to H$ signal against a mixture of background samples. All the background processes are used to train the network: DY, $W+jets$, Wγ, WW, WZ, ZZ and the data-driven QCD events. Therefore, the choice of the set of variables to be used as input to the NN aims at discriminating the Higgs signal from such a processes with a satisfactory efficiency by exploring the full set of available variables.
The Higgs signal discrimination from the main background processes $W+jets$, $Z \to \tau\tau$ and QCD have been studied in the correspondent control regions, where those backgrounds are dominant and the discriminating variables have been tested.

Figure 6.2(left) shows the NN output in the $W+jets$ control regions in the $e-\tau$ channel and for Higgs mass 165 GeV/c^2. Only tau related variables have been used as input to the NN, to exploit the difference between real taus and jets mis-identified as τ's. It can be seen that the $W+jets$ and QCD component tend to shift on the left, whereas the Drell-Yan component, mostly $Z/\gamma^* \to ee$, contributing with a lepton-faking τ, still accumulate at higher values of the NN score. However, the signal separation from the $Z/\gamma^* \to ee$ improves using variables related to the event topology and kinematics, as it is shown in Figure 6.2(right). However, training the NN using only variables related to the event topology and kinematics is not able to separate the signal from $W+jet$ events, which distribute almost omogeneusly between -1 and 1 (Figure 6.2(right)). The discriminating power is enhanced when event based observables are accompanied by observables specific to the τ identification.

In Figure 6.3 are shown the outputs of the NN trained using tau ID only or event only input variables, in the QCD control region, where contribution from dijet, $\gamma+jet$ and $Z \to ll$ are expected to be dominant. It can be seen that the signal discrimination from $Z \to ll$ and QCD events relays on variables exploiting the different topology of the events (as discussed in Section 5.1.1) rather then τ specific variables.

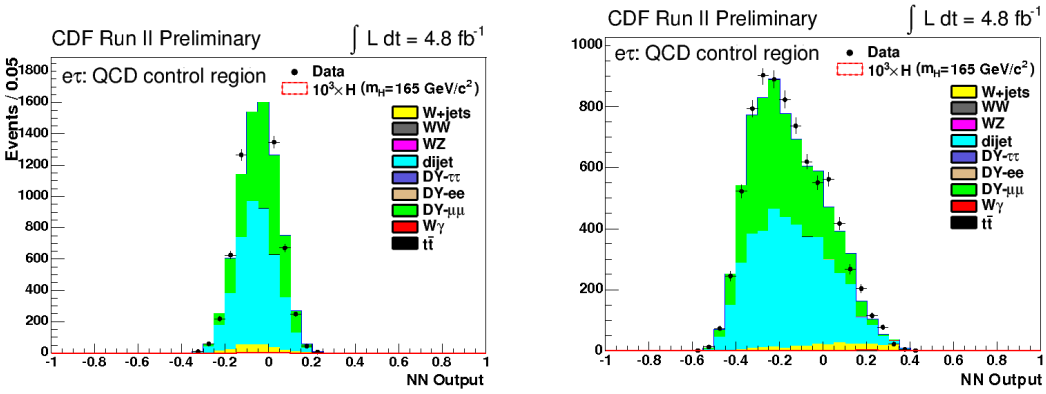


Figure 6.3: NN output in the QCD $e\tau$ control region for the Higgs mass 165 GeV/c^2. Only tau ID (left) or event (right) related input variables used.

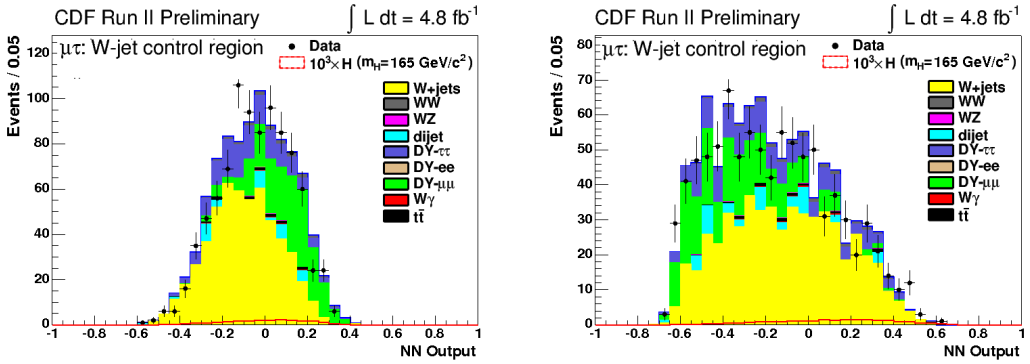


Figure 6.4: NN output in the $W + jet$ $\mu - \tau$ control region for the Higgs mass 165 GeV/c^2. Only tau ID (left) or event (right) related input variables used.

In the $\mu-\tau$ channel, separate studies on $W+jet$ control region have been performed. In Figure 6.4 is shown the output of the NN trained using tau ID only or event only input variables.

The variables used to train the NN are ordered according to the power in discriminating between signal and background, evaluated during the NN training.
Then, the set of input variables used to train the NN in the signal region has been chosen among the input variables found to be the most sensitive in discriminating signal from the dominant expected background.
The expected limit for each of these NN output is calculated and the set of variables which minimizes it has been chosen as input, in order to obtain the best expected sensitivity to the Higgs signal.
Finally, a set of 7 input variables have been chosen for both $e\tau$ and $\mu\tau$ channels.
Event and kinematic variables:

- transverse mass of the lepton and the missing energy, $M_T(l, \not\!\!E_T)$
- transverse mass of the tau and the missing energy, $M_T(\tau, \not\!\!E_T)$
- lepton and tau separation in $\Delta\phi$ and ΔR

Tau ID variables used:

- tau visible mass
- angle between the closest track and the tau track momentum ($\Delta\alpha^{closest\ trk}$)
- seed track $d_0^{seed\ trk}$
- scalar sum of the p_T's of all the tracks in the isolation cone which have not been used in the tau reconstruction ($\sum_{iso} p_T^{trk}$)

Since the distributions of input variables show different behaviors at different mass hypothesis, the discriminating power between signal and background of NN input variables depends on the Higgs mass. The expected distribution of some of the input variables for different mass hypothesis scaled up by a factor of 500 and superimposed as histograms to the background distribution are shown in Figure 6.5 and 6.6. In Tables 6.1 and 6.2 is reported the set of NN input variables chosen, ordered by their discriminating power for three mass points: $M_H = 130,\ 160,\ 190$ GeV/c^2.

Thus, the final NN architecture has one input layer with 8 nodes, one hidden layer

$e-\tau$	$M_H = 130$	$M_H = 160$	$M_H = 190$
1	$\Delta\alpha^{closest\ trk}$	$M_T(l, \not{E}_T)$	$M_T(l, \not{E}_T)$
2	τ visible mass	τ visible mass	τ visible mass
3	$M_T(l, \not{E}_T)$	$\Delta\alpha^{closest\ trk}$	$M_T(\tau, \not{E}_T)$
4	$\Delta\phi_{l\tau}$	$\Delta\phi_{l\tau}$	$\Delta\alpha^{closest\ trk}$
5	$d_0^{seed\ trk}$	$d_0^{seed\ trk}$	$d_0^{seed\ trk}$
6	$\sum_{iso} p_T^{trk}$	$M_T(\tau, \not{E}_T)$	$\sum_{iso} p_T^{trk}$
7	$M_T(\tau, \not{E}_T)$	$\sum_{iso} p_T^{trk}$	$\Delta\phi_{l\tau}$

Table 6.1: The set of NN input variables chosen in the $e\tau$ channel, ordered by their discriminating power for $M_H = 130,\ 160,\ 190$ GeV/c^2.

$\mu-\tau$	$M_H = 130$	$M_H = 160$	$M_H = 190$
1	τ visible mass	$M_T(l, \not{E}_T)$	$M_T(l, \not{E}_T)$
2	$\Delta\alpha^{closest\ trk}$	τ visible mass	τ visible mass
3	$\Delta R_{l,\tau}$	$\Delta\alpha^{closest\ trk}$	$\Delta\alpha^{closest\ trk}$
4	$M_T(l, \not{E}_T)$	$\Delta R_{l,\tau}$	$M_T(\tau, \not{E}_T)$
5	$d_0^{seed\ trk}$	$d_0^{seed\ trk}$	$d_0^{seed\ trk}$
6	$\sum_{iso} p_T^{trk}$	$M_T(\tau, \not{E}_T)$	$\sum_{iso} p_T^{trk}$
7	$M_T(\tau, \not{E}_T)$	$\sum_{iso} p_T^{trk}$	$\Delta R_{l,\tau}$

Table 6.2: The set of NN input variables chosen in the $\mu\tau$ channel, ordered by their discriminating power for $M_H = 130,\ 160,\ 190$ GeV/c^2.

with 9 nodes and one output layer constituted by just one node.

The control regions are used to test for proper training on the NN as well. The events of the $W + jet$ or QCD control region are given as input to the NN trained on the signal region. If the NN is trained properly, the control region processes are expected to output a very background-like NN. The NN output for these control regions

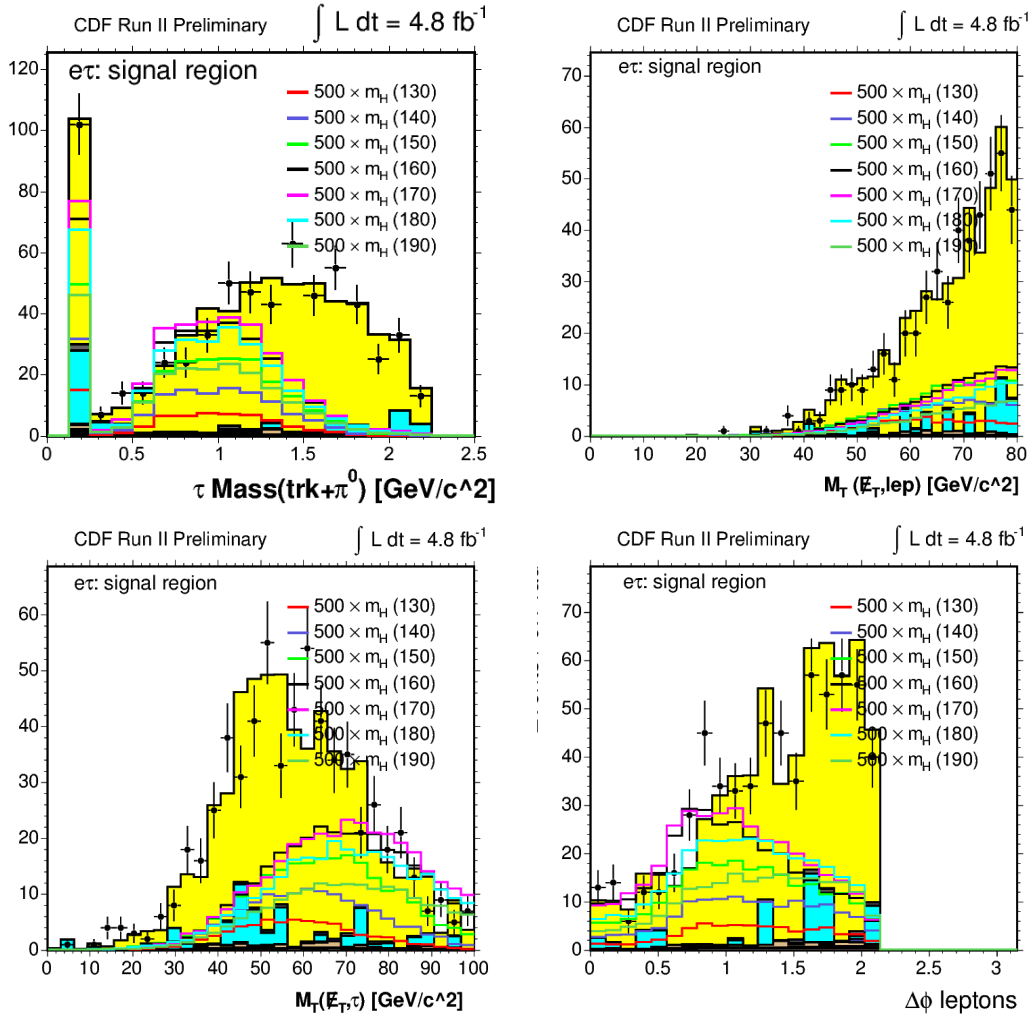


Figure 6.5: The expected distribution of some of the input variables for different mass hypothesis scaled up by a factor of 500 and superimposed as histograms to the background distribution in the $e - \tau$ channel.

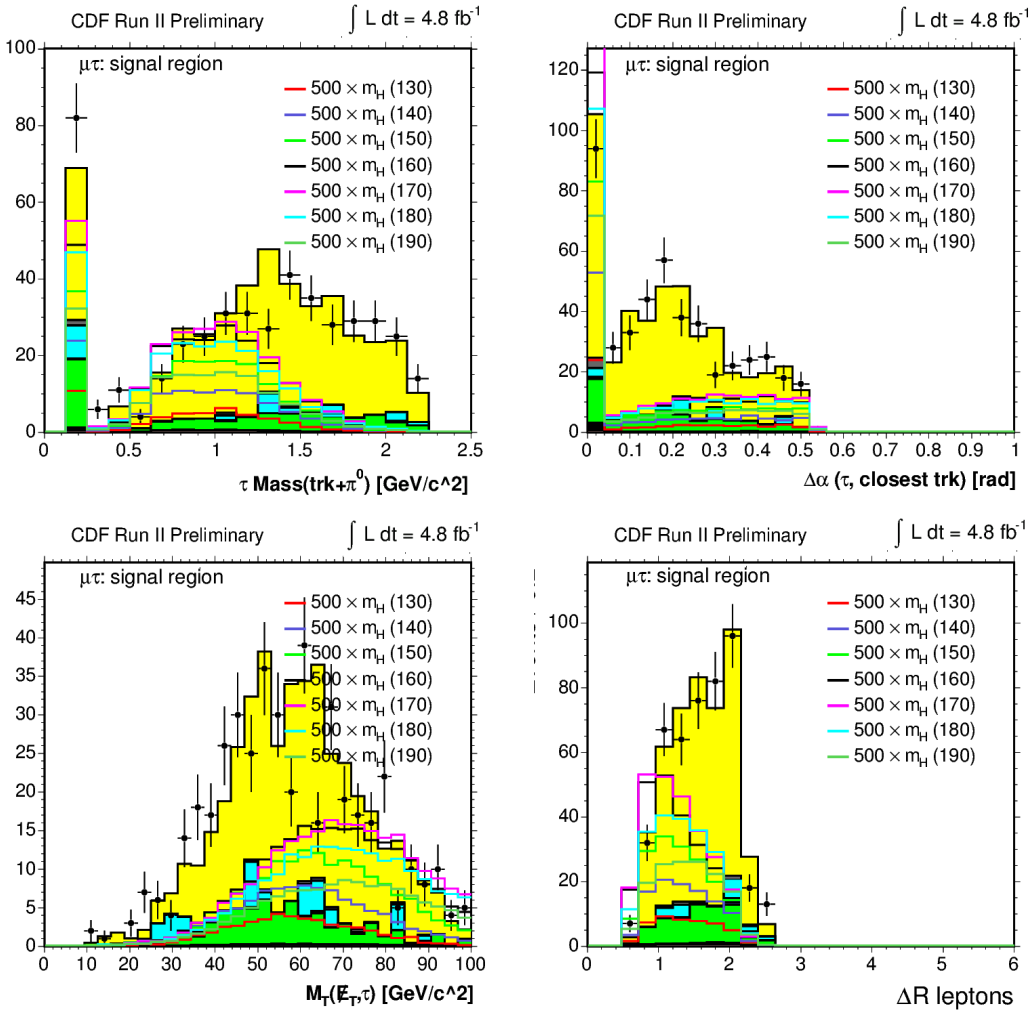


Figure 6.6: The expected distribution of some of the input variables for different mass hypothesis scaled up by a factor of 500 and superimposed as histograms to the background distribution in the $\mu - \tau$ channel.

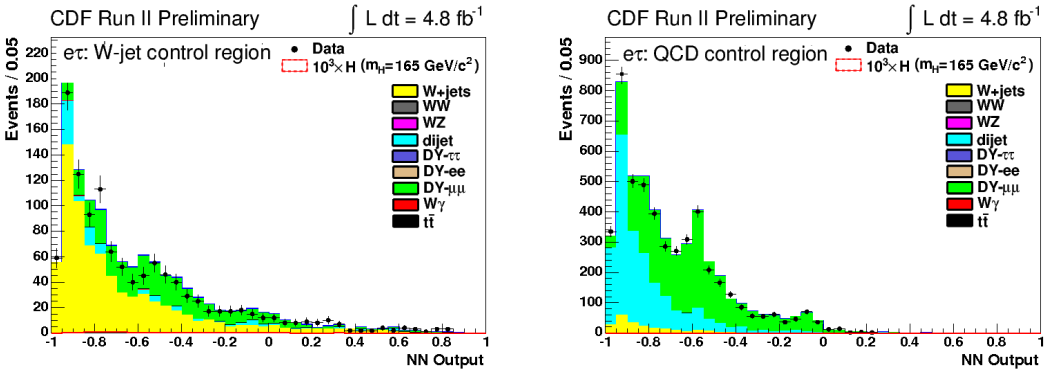


Figure 6.7: Outputs of NN trained on signal region events and applied to events in the $W + jet$ control region (left) and QCD control region (right), in the $e - \tau$ channel and mass $M_H = 165$ GeV/c^2.

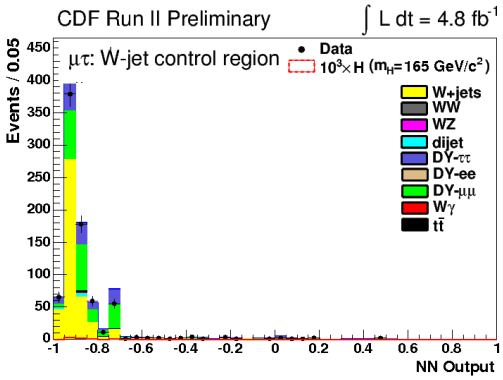


Figure 6.8: Output of NN trained on signal region events and applied to events in the $W + jet$ control region, in the $\mu - \tau$ channel and mass $M_H = 165$ GeV/c^2.

are illustrated in Figure 6.7 and 6.8, for a SM Higgs boson mass of $M_H = 165$ GeV/c^2. The NN outputs confirm the background-like nature of data found in the control regions.

As mentioned before, a different NN is trained for each of the 12 Higgs mass hypothesis considered. After the parameters used for the NN configuration have been optimized, the probability for each given event in the signal region to be compatible with signal or background processes is evaluated for expected and observed events. Templates are created for each NN and shown in Figures 6.9, 6.10, 6.11 and 6.12.

The histograms exhibit a stacked plot of the MC background prediction with the superimposed signal prediction, scaled up by a factor of 10^2. Signal events tend to have a score near the target value $+1$, while background events tend to be classified with score near -1. The score contribution of data events is also compared to the one of the expected sample composition from MC. Data are well described by the Standard Model mixture of the expected background processes, following the background-only prediction. Therefore, there is no evidence for signal and the NN output distributions are used in calculating a limit on the Higgs boson production cross section.

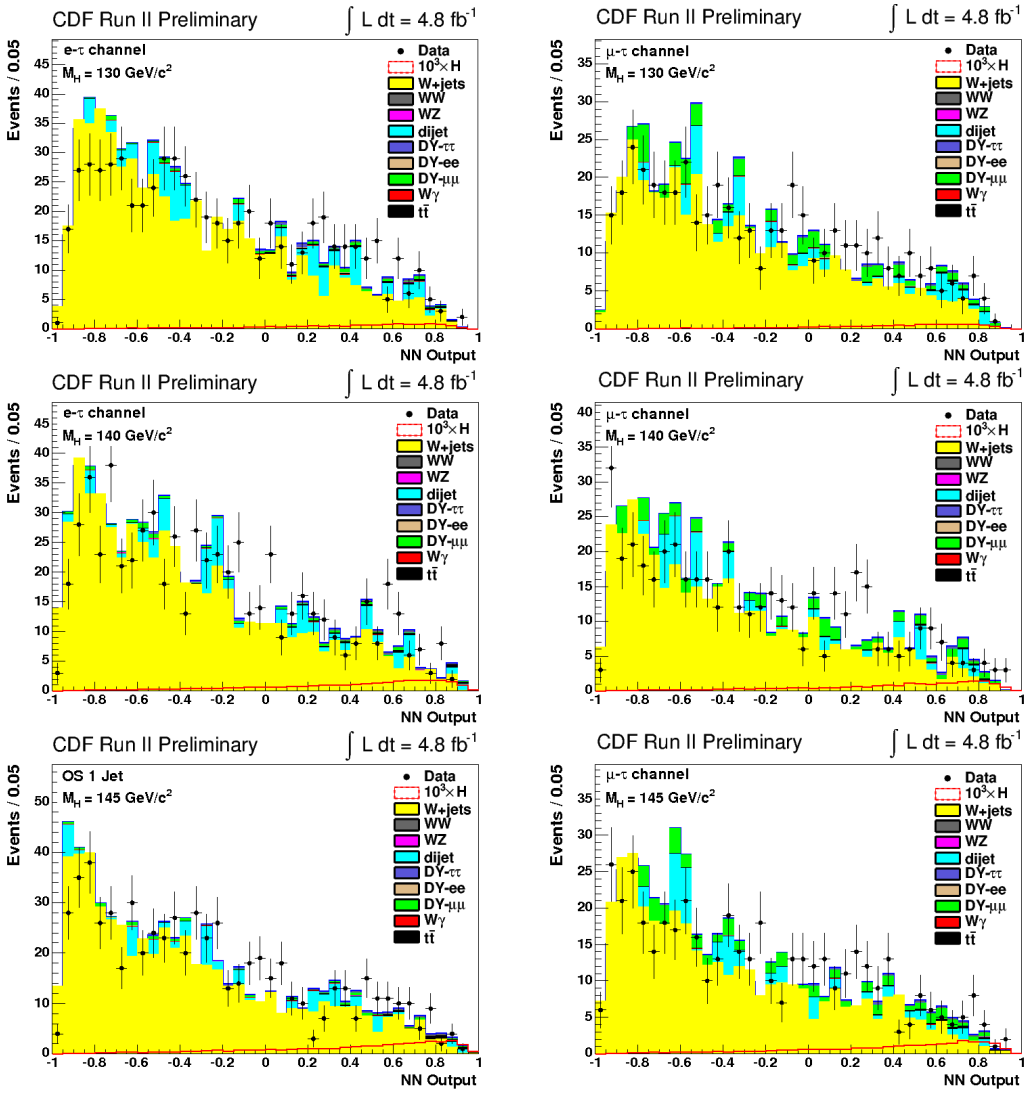


Figure 6.9: NN output templates for M_H from 130 and 145 GeV/c^2

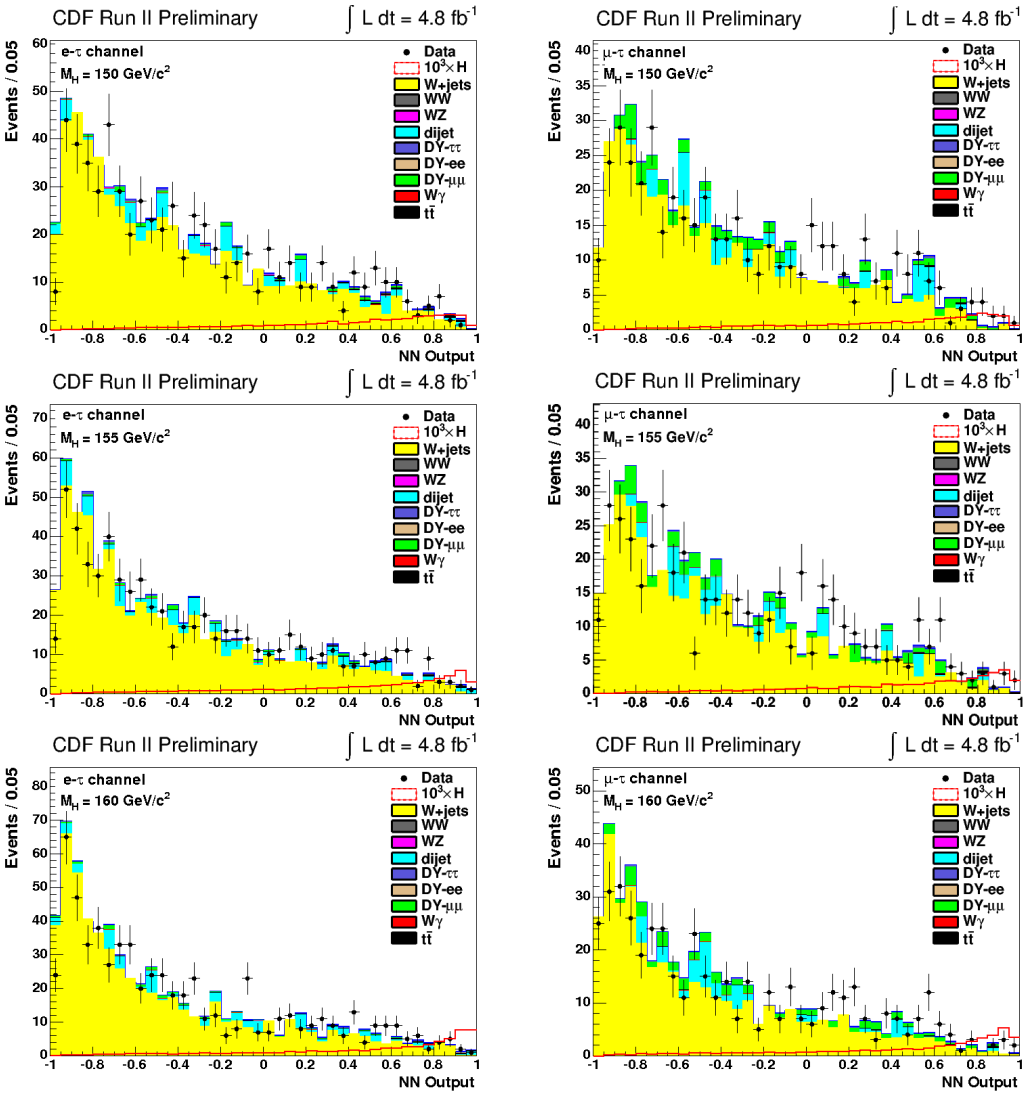


Figure 6.10: NN output templates for M_H from 150 and 160 GeV/c^2

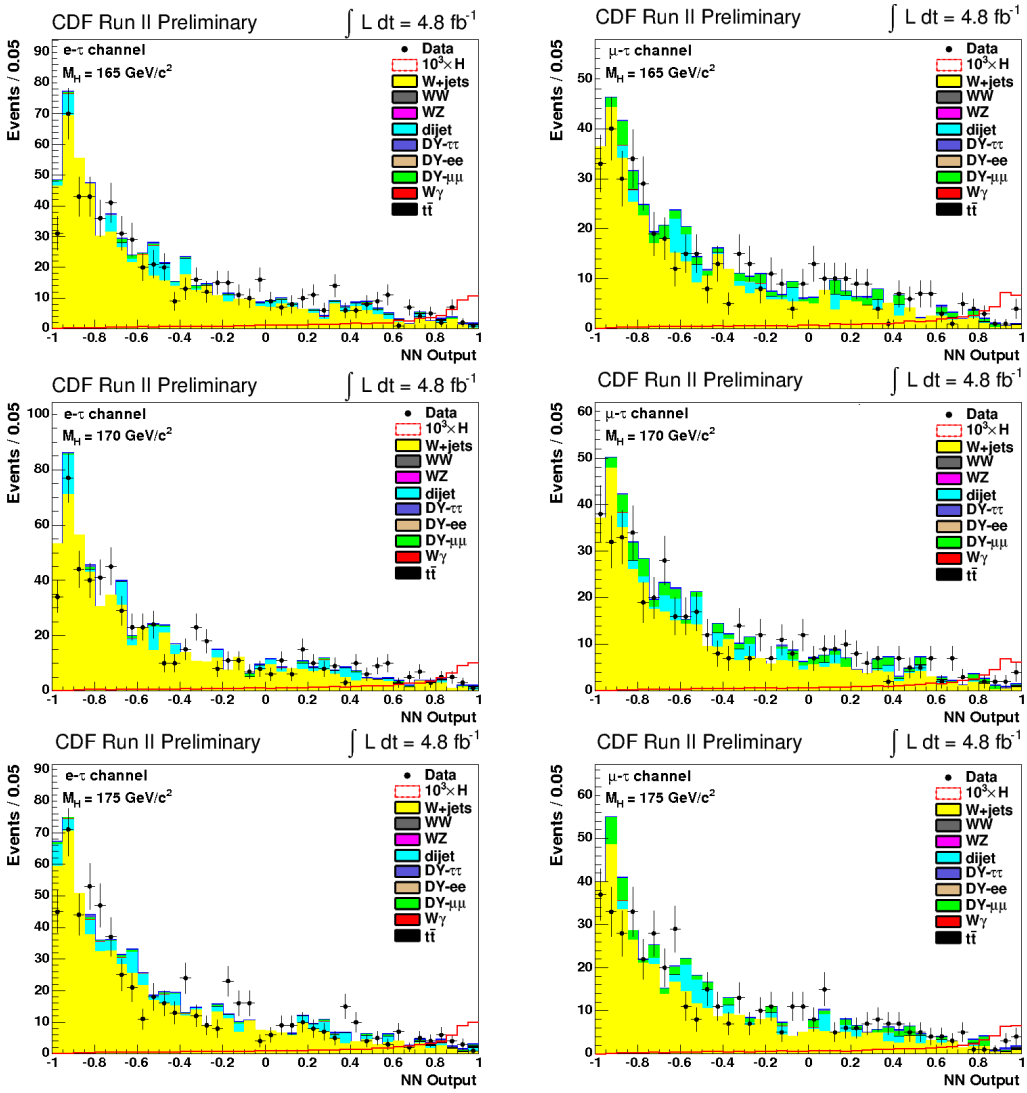


Figure 6.11: NN output templates for M_H from 165 and 175 GeV/c^2

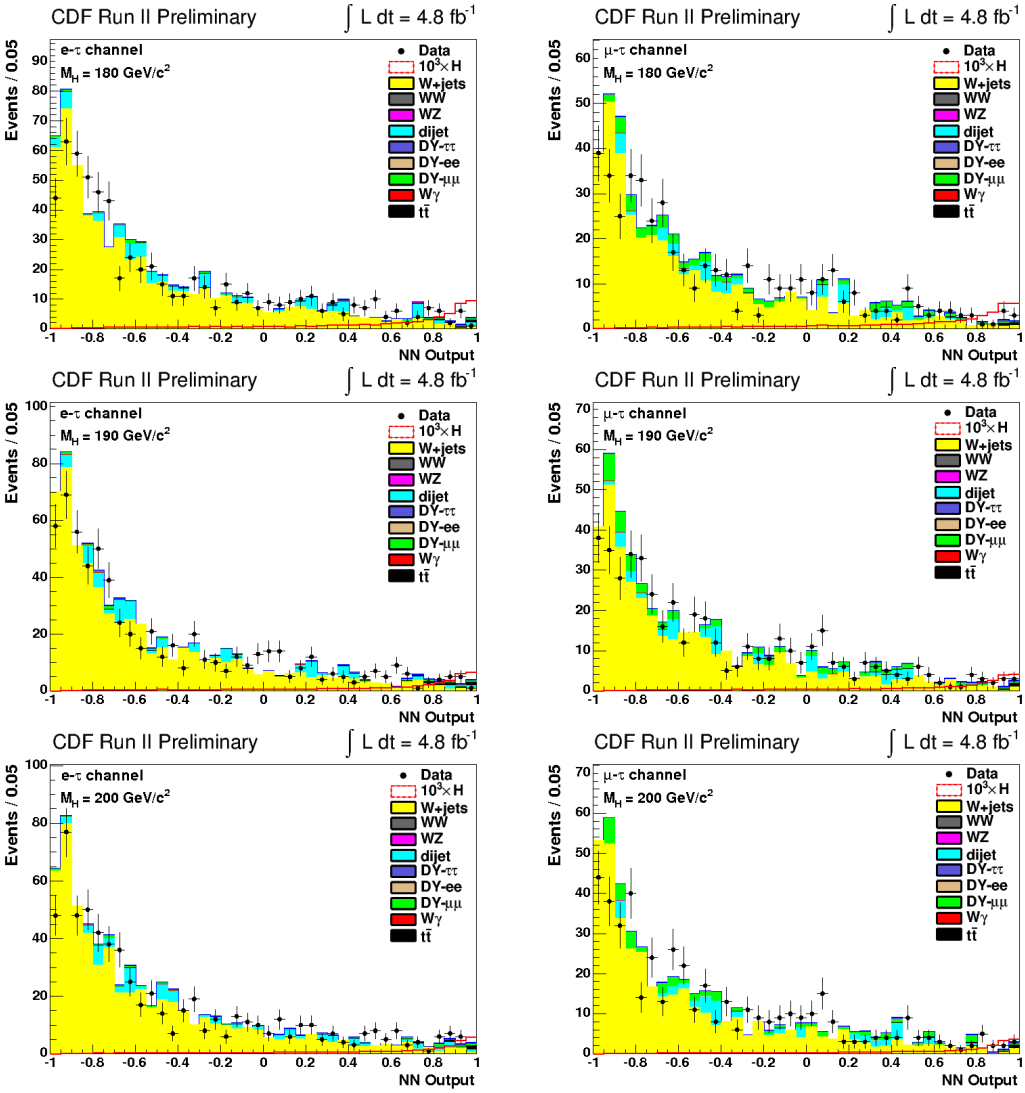


Figure 6.12: NN output templates for M_H from 180 and 200 GeV/c^2

6.3 Systematic uncertainties

The limit calculation may be affected by many sources of uncertainties. Systematic uncertainties affect the overall normalization (called *rate systematics*) of the background and signal cross section, thus affecting the expected number of events. Other sources of systematic error may change the shape of the kinematic distributions (*shape systematics*), and consequently of templates used in the limit calculation.
The effect of systematics affecting the shape of the NN output on the sensitivity of the measurement has been found to be negligible, therefore only rate systematics are considered. Tables 6.3, 6.4 and 6.5 report the rate systematic uncertainties estimated for this analysis.

Uncertainty source	WW	WZ	ZZ	$t\bar{t}$	$Z \to \tau\tau$	$Z \to ll$	$W + jet$	$W\gamma$
Cross section	6.0	6.0	6.0	10.0	5.0	5.0		
Measured W cross-section							12	
PDF Model	1.6	2.3	3.2	2.3	2.7	4.6	2.2	3.1
Higher order diagrams	10	10	10	10	10	10		10
Trigger Efficiency	0.5	0.6	0.6	0.6	0.7	0.5	0.6	0.6
Lepton ID Efficiency	0.4	0.5	0.5	0.4	0.4	0.4	0.5	0.4
τ ID Efficiency	1.0	1.3	1.9	1.3	2.1			0.3
Jet into τ Fake rate	5.8	4.8	2.0	5.1		0.1	8.8	
Lepton into τ Fake rate	0.2	0.1	0.6	0.2		2.3		2.1
$W + jet$ scale							1.6	
MC Run dependence	2.6	2.6	2.6				2.6	
Luminosity	5.9	5.9	5.9	5.9	5.9	5.9	5.9	5.9
Total	14.7	14.4	14.0	16.4	13.1	13.7	16.5	12.2

Table 6.3: Systematic uncertainties on the background for the $e\tau$ channel (expressed in %).

Uncertainty source	WW	WZ	ZZ	$t\bar{t}$	$Z \to \tau\tau$	$Z \to ll$	$W + jet$	$W\gamma$
Cross section	6.0	6.0	6.0	10.0	5.0	5.0		
Measured W cross-section							12	
PDF Model	1.5	2.1	2.9	2.1	2.5	4.3	2.0	2.9
Higher order diagrams	10	10	10	10	10	10		10
Trigger Efficiency	1.3	0.7	0.7	1.1	0.9	1.3	1.0	1.0
Lepton ID Efficiency	1.1	1.4	1.4	1.1	1.2	1.1	1.4	1.3
τ ID Efficiency	1.0	1.2	1.4	1.6	1.9			
Jet into τ Fake rate	5.8	5.0	4.4	4.4		0.2	8.8	
Lepton into τ Fake rate	0.06	0.05	0.09	0.04		1.9		1.2
$W + jet$ scale							1.4	
MC Run dependence	3.0	3.0	3.0				3.0	
Luminosity	5.9	5.9	5.9	5.9	5.9	5.9	5.9	5.9
Total	14.8	14.6	14.6	16.2	13.1	13.6	16.6	121

Table 6.4: Systematic uncertainties on the background for the $\mu\tau$ channel (expressed in %).

Uncertainty source	$gg \to WW$	WH	ZH	VBF		$gg \to WW$	WH	ZH	VBF
Cross section	14.3	5	5	10		14.3	5	5	10
PDF Model	2.5	2.0	1.9	1.8		2.6	2.2	2.0	2.2
Higher order diagrams		10	10	10			10	10	10
Trigger Efficiency	0.5	0.5	0.6	0.5		1.3	1.3	1.2	1.3
Lepton ID Efficiency	0.4	0.4	0.4	0.4		1.0	1.0	1.0	1.0
τ ID Efficiency	2.8	1.6	1.7	2.8		2.9	1.6	1.7	2.8
Jet into τ Fake rate		4.2	4.0	0.4			4.5	4.2	0.4
Lepton into τ Fake rate	0.15	0.06	0.15	0.11		0.04	0.02	0.02	0.04
Luminosity	5.9	5.9	5.9	5.9		5.9	5.9	5.9	5.9
Total	15.9	13.6	13.6	15.7		16.0	13.8	13.7	15.8

Table 6.5: Systematic uncertainties on the $e\tau$ and $\mu\tau$ signals (expressed in %).

Each row represents a source of systematic uncertainty, while the columns show the relative effects on the number of expected events for the given processes, expressed in %. The rate uncertainties can be divided into three categories:

- acceptance systematics: it includes all the sources of systematic uncertainties that vary the number of selected events either for the detector geometry or for the event selections applied. It include PDF modeling, high order diagrams approximation, lepton ID, tau ID and trigger efficiencies, tau fake rate scale factors, $W + jet$ rescaling, MC run dependence.
- cross section systematics: it accounts for the error on the theoretical cross section used to normalize signal and background simulations, affecting the number of expected events.
- luminosity systematics: it accounts for the error on the luminosity measurement which affect the number of events predicted by the simulation.

The calculation of all uncertainties is discussed below.

6.3.1 Acceptance systematics

PDF model

The effect on the $e\tau$ and $\mu\tau$ acceptances due to the uncertainties on PDF in the MC generation, due to the limited knowledge of the momentum distribution of the constituent partons in the colliding protons, is obtained by varying PDF parameters, renormalization and factorization scales. This uncertainty is calculated by looking at the maximum and minimum difference (CTEQ6M-CTEQ5L), where CTEQ5L is the default leading-order (LO) PDF and CTEQ6M [118] is the next-to-leading-order CTEQ PDF set with 1-loop α_s calculations. The method consists in reweighting the events of the MC samples, generated with the default CTEQ5L PDF set, to different sets of PDFs.

High order diagrams

The effect of the differences between LO and higher order calculations of background and signal processes on the acceptance are considered in the systematic uncertainty esti-

mation. It accounts for variations of processes generated with PYTHIA at LO (leading order + phenominological model of parton showering) from MC@NLO modeling.
The size of the higher order effects is determined by comparing a LO Pythia WW sample to the WW MC@NLO sample. It accounts for poor MC modeling beyond leading order. A difference of about 10% has been found and assigned as systematic error on the processes simulated by LO Pythia.

Lepton ID, tau ID and trigger efficiencies

A 3% uncertainty on the τ identification scale factor is applied as estimated in [102]. The electron and muon ID scale factors and trigger efficiencies are fluctuated up and down by 1σ. All lepton and trigger efficiencies are varied in the same direction simultaneously to determine their effect on the acceptance. Both uncertainties have a small effect, around 0.5% for electrons and 1.3% for muons, on the signal and background processes. The three contribution are indicated into the Tables 6.3, 6.4 and 6.5 separately.

Tau fake rate scale factors

The uncertainty on both $jet \to \tau$ and $lepton \to \tau$ fake rates has been considered. The effect of the uncertainty on the jet into τ scale factor is estimated by varying the values of the fit parameters in Figure 4.10, for the combined jet into τ fake rate scale factor, by $\pm 1\sigma$. Half of the maximum variation is taken as systematic uncertainty. The error on the lepton into τ scale factor is listed in Table 4.2. It accounts for the uncertainty on the data and on the expected background (see Section 5.2.1). The impact on background and signal estimation is determined by varying the central value of the lepton into τ scale factor by $\pm 1\sigma$.

$W + jet$ **rescaling**

The $W + jet$ background has an additional systematics uncertainty arising from the scaling applied to the single-prong no-π^0 $W + jet$ contribution, as discussed in Section 5.3.1. The uncertainties on the scale factor are assigned as systematic error to account for possible differences between data and MC. This systematic uncertainty is estimated by varying the central value of the scale factor used to rescale the $W + jets$ contribution by $\pm 1\sigma$, and it is called $W + jet\ scale$ in the systematic Tables.

MC run dependence

A systematic uncertainty is also assigned to MC samples which have a restricted run range simulated. The systematic uncertainty is assigned to account for a possible difference in luminosity profile between the MC samples and the data. This is determined by comparing a WW sample with partial run dependence (periods 0 7) with a fully run-dependent WW sample.

6.3.2 Cross section systematics

For the background processes, the theoretical uncertainties on the cross section are taken as systematic uncertainties. The $t\bar{t}$ cross section is calculated at NNLO in [119] and the estimated uncertainty is 10%. For di-boson processes MCFM [120] is used to perform a

NLO calculation with an uncertainty of 6%. PYTHIA MC generator is used to estimate the WW background. The cross section uncertainty on the WW production cross-section is taken from [109]. Drell-Yan is known at NNLO and the associated systematic is 5% [121]. The $W + jet$ cross section is normalized to the inclusive W production cross section measured by CDF [122].
Uncertainties on the theoretical cross sections vary for the different Higgs production mechanisms. Associated production cross sections are known to NNLO and it is less than 5% [123]. VBF production is known only to NLO, the residual theoretical uncertainty is higher and on the order of 10% [123]. The gluon fusion production cross section is larger than VH and VBF, even though it is known to NNLO [35]. For this studies, the HNNLO [124, 125] program is used to calculate the theoretical Higgs production cross section for $gg \to H$ production at Tevatron. The two main independent sources of theoretical uncertainty, as discussed in Section 1.4, are: uncalculated higher-order QCD radiative corrections, and knowledge of the parton distribution PDF. A $\pm 7\%$ variation on the renormalization and factorization scale of the cross section calculation is found. In [35] an additional expected PDF uncertainty of 7.7% is given.

6.3.3 Luminosity systematics

The integrated luminosity of the sample is defined in Section 2.1.3. The uncertainty on this measurement is estimated to be 5.9%. The two main sources contributing to the total systematic uncertainty are the inelastic $p\bar{p}$ cross section, which contributes with a 4% variation, and the geometry description of the detector in the simulation contributing with a 4.2%.

6.3.4 Systematic uncertainties on NN shape

Two possible sources of uncertainties that can modify the shape of the NN output distribution are the jet energy correction (see Section 3.7) and the initial state radiation (ISR). The systematic error due to the uncertainties in the jet energy correction can be evaluated applying a $+1\sigma$ and -1σ shifts to the jet momenta in signal and background events.
The effects of the ISR can be studied on Pythia $gg \to H$ samples, with the ISR either increased or decreased by $\pm 1\sigma$ from the default value.
In this analysis, there is a very little systematic change in the NN output shape, and the effect on the limit has been found to be negligible. Therefore, this contribution has been not taken into account.

6.4 Setting limits on the SM Higgs production cross section

The NN output created for each Higgs mass hypotheses is used to estimate Bayesian 95% CL upper limit on the Higgs production cross section. In this analysis an upper limit is set for 12 values of M_H hypothesis considered, which includes contributions from $gg \to H$, WH, ZH and VBF.

In the limit calculation, background and signal expectations in each bin of the NN output are treated according to the Poisson statistics. For each Higgs mass hypothesis a likelihood $\mathcal{L}$ is built as the product of the Poissonian probabilities of observing n_k data events in the kth bin of the NN output, given an expectation of μ_k:

$$\begin{aligned} \mathcal{L} &= \prod_{k=1}^{N} \frac{e^{-\mu_k}\ \mu_k^{n_k}}{n_k!} \cdot \prod_i e^{-\frac{\theta_i^2}{2}} \\ \mu_k &= R s_k + b_k \end{aligned} \tag{6.1}$$

s_k and b_k are the yields of signal and expected background in the kth bin, respectively, while R multiplies the number of expected signal events, that can also be written as $s_i = L{\cdot}\epsilon{\cdot}\sigma_H^{SM}$, where L is the integrated luminosity of the collected sample, ϵ the detector acceptance and analysis efficiency and σ_H^{SM} is the Standard Model Higgs cross section. Thus, the μ_k expectation, as a function of R, can differ depending on which hypothesis is used: the background-only hypothesis or the background-plus-signal hypothesis. A flat *prior* to R is assigned, since no experimental information on the Higgs production cross section is assumed. s_k and b_k carry systematic uncertainties, as discussed in Section 6.3, thus their predictions are systematically varied according the notation:

$$\begin{aligned} s_k(\theta) &= s_k \cdot \prod_j (1 - u_k^j \cdot \theta_j) \\ b_k(\theta) &= b_k \cdot \prod_j (1 - u_k^j \cdot \theta_j) \end{aligned} \tag{6.2}$$

where u_k^j is the scale associated with the systematic θ_j and the process k, so the Gaussian constraints on the θ_j in the likelihood (6.1) are all centered on zero with unit width. The product (6.2) runs over the index j, which accounts for different sources of systematic uncertainties. The likelihood is integrated over each systematic contribution and evaluated for the observed number of events. The only left dependence is on the parameter R, thus $\mathcal{L}(R)$ (6.1) is used to make inferences about the observation given a true value R.

6.4.1 Upper Limit Results

Upper limits for Higgs cross section are calculated as a ratio to the SM prediction. The confidence level chosen is 95%, a standard practice in high energy physics for setting limits. To investigate the confidence intervals, the p.d.f $P(R|\text{Observation})$ is integrated over a range of possible true R values, which can be written in terms of the likelihood (6.1) and the flat prior assigned to R, $P(R)$, applying the Bayes' Theorem:

$$\int_{R_1}^{R_2} P(R|\text{Observation})dR = \int_{R_1}^{R_2} \mathcal{L}(\text{Observation}|R)\frac{P(R)}{P(\text{Observation})}dR = \alpha \tag{6.3}$$

where α is the *confidence level* that the true value lies between R_1 and R_2 and $P(\text{Observation})$ is the normalization chosen to be:

$$P(\text{Observation}) = \int_0^{\infty} \mathcal{L}(\text{Observation}|R)dR \tag{6.4}$$

Since an upper limit is desired, R_1 is set to be 0, indicating that the preferred hypothesis is the one without any signal component, and R_2 is determined by choosing the value of $\alpha = 0.95$. The 95% upper limit on R is then calculated solving the equation:

$$\frac{\int_0^{R_2} \mathcal{L}(R)dR}{\int_0^{\infty} \mathcal{L}(R)dR} = 0.95 \tag{6.5}$$

The upper limit, R_2, is numerically determined in this Thesis using the MClimit [126] software package, written specifically for CDF analysis.

Upper limits on Higgs production in $(e-\tau)$ and $(\mu-\tau)$ channels

Upper limits are calculated on the SM Higgs production as a ratio to the SM prediction for 12 masses in the range $130 \leq M_H \leq 200$ GeV/c^2. The limits are calculated for the $e-\tau$ and $\mu-\tau$ channels separately, considering the different background composition in the two channels.
To estimate the sensitivity of the analysis before looking at data events, sets of pseudo-events n_k are randomly generated for each bin k taken from the Poisson distribution in the background only hypothesis ($\mu = b_k$). Pseudo-experiments are generated by Poisson-fluctuating the mean prediction μ for each process based on the systematic uncertainties associated to the process. This prediction is then selected from the NN template correspondent to the process according to the probability distribution of the template. The NN templates discussed in Section 6.2 have been used to calculate limits. The distribution of limits from pseudo-experiments gives an idea of the range of limits one should expect to find in the experiment, on which the optimization of the analysis is based.

At the Higgs mass hypothesis of 160 GeV/c^2 the expected limits are $25.14 \times \sigma_{SM}$ and $31.33 \times \sigma_{SM}$, and the observed limits are $30.87 \times \sigma_{SM}$ and $51.70 \times \sigma_{SM}$, for the $e-\tau$ and $\mu-\tau$ respectively. The expected and observed limits at all the masses investigated in the $e-\tau$ and $\mu-\tau$ channels are given in Tables 6.6 and 6.7 and plotted in Figure 6.13 and 6.14, respectively.
The observed limit is more than 1σ from the expected, however this excess is still well within the estimated uncertainty. As discussed in Section 5.1.1, the efficiency of selection cuts in discriminating signal from background depends on the Higgs mass hypothesis, being less efficient for light Higgs. Thus, for light Higgs boson the discriminating power is less causing the expected limit to become larger at low mass hypothesis.

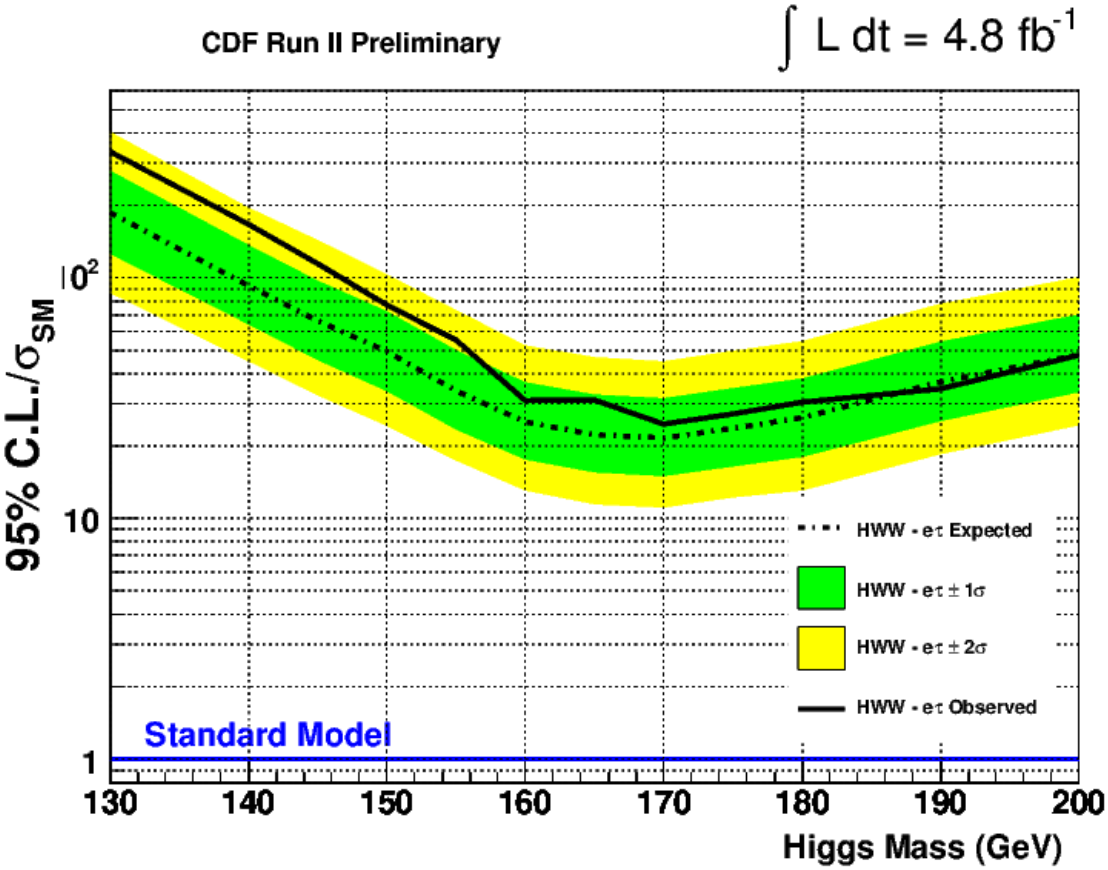


Figure 6.13: Observed and expected limits on the Higgs production as a ratio to the SM prediction for the $e - \tau$ channel.

Mass [GeV/c^2]	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Observed
130	86.47	125.75	186.53	277.40	404.20	333.26
140	45.01	63.94	92.63	136.33	194.86	166.25
145	32.46	45.43	65.95	96.92	142.81	114.35
150	24.23	33.79	49.48	72.48	103.03	76.89
155	17.39	23.48	33.85	49.78	73.21	54.99
160	13.06	17.61	25.14	36.76	52.24	30.87
165	11.45	15.54	22.24	32.63	46.64	30.97
170	11.11	15.01	21.58	31.57	44.71	24.64
175	12.24	16.47	23.64	34.70	49.63	27.10
180	13.09	18.04	26.18	38.12	54.30	30.33
190	18.54	25.47	36.91	54.17	77.67	34.58
200	24.35	33.43	48.01	70.29	100.50	47.74

Table 6.6: Expected and observed limits on the Higgs production as a ratio to the SM prediction for the $e - \tau$ channel.

$e - \tau$ and $\mu - \tau$ combined limits

The combination of the results for the two $e - \tau$ and $\mu - \tau$ channels in this Thesis is performed with the same method used to extract the individual limits, and described in

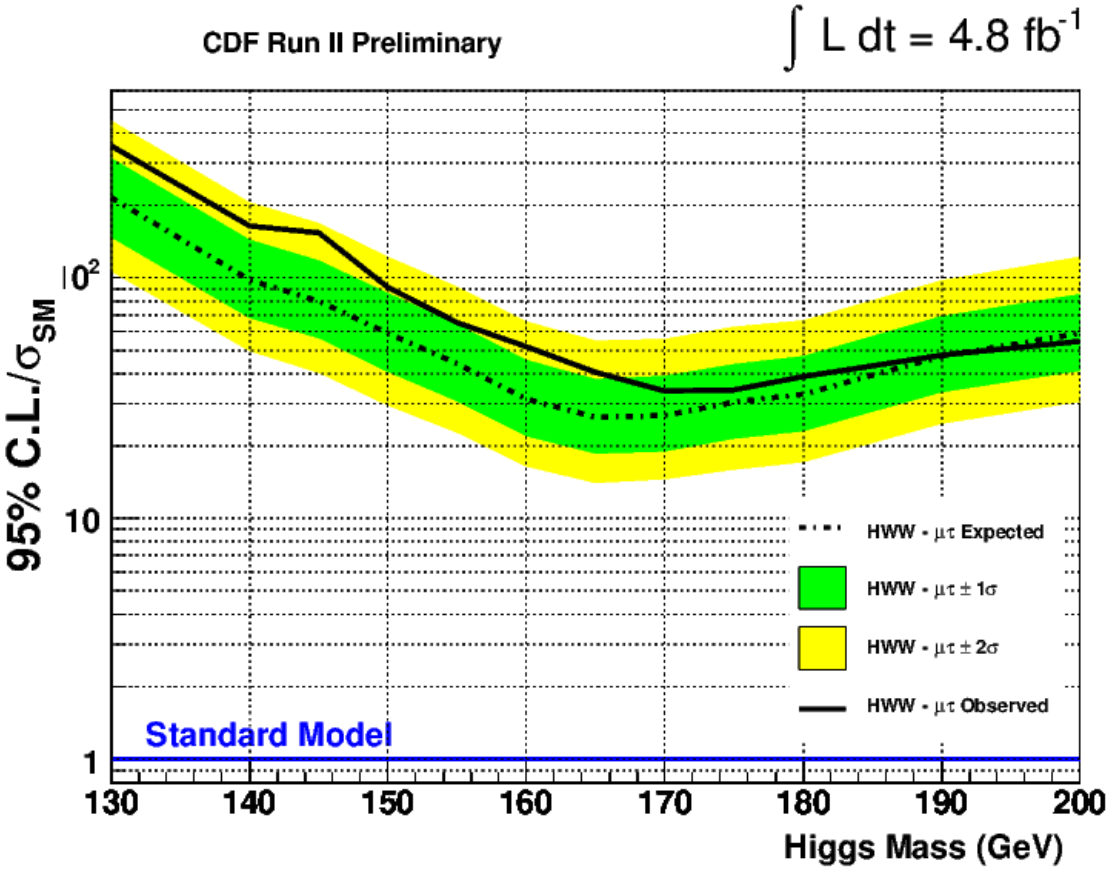


Figure 6.14: Observed and expected limits for the $\mu - \tau$ channel.

Mass [GeV/c^2]	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Observed
130	106.12	147.31	214.62	314.52	448.96	352.93
140	49.71	68.25	98.11	143.61	205.31	164.04
145	40.28	56.13	79.93	117.72	168.73	153.76
150	29.48	40.66	58.86	86.38	122.06	91.02
155	22.69	30.70	43.84	64.09	91.61	65.10
160	16.46	22.01	31.33	45.54	65.58	51.70
165	14.04	18.57	26.30	37.94	54.58	40.39
170	14.60	19.04	26.76	38.99	55.56	33.87
175	15.96	21.47	30.34	43.79	62.47	34.23
180	17.19	23.12	32.83	47.28	66.51	38.82
190	24.81	33.43	47.06	69.24	97.46	47.76
200	30.50	41.18	58.75	85.34	121.76	54.45

Table 6.7: Expected and observed limits in the $\mu - \tau$ channel.

Section 6.4.1.
The 95% upper limits are extracted using as input to the equation (6.1) the NN output

distributions of each channel for the corresponding mass hypothesis. The limits are evaluated for the twelve mass hypotheses investigated ranging from $130 \leq M_H \leq 200$ GeV/c^2. Systematics for the combination are taken from Tables 6.3, 6.4 and 6.5 for the two channels, however the systematic uncertainties that have the same names are assumed correlated.
The combined upper limits on the ratio of the SM Higgs production cross section to the Standard Model expectation are presented for each mass hypothesis in Table 6.8 and in Figure 6.15. The combination set a 95% C.L. limit of $33.54 \times \sigma_{SM}$ for the mass hypothesis $M_H = 160$ GeV/c^2, with an expected limit of $19.56 \times \sigma_{SM}$. The observed limit well agrees with the expectation, for all the mass hypotheses, although it is more than 1σ from the expected this excess is still within the estimated uncertainty. The combined result improves the sensitivity of the individual search channels on the direct search for the high mass Higgs boson.

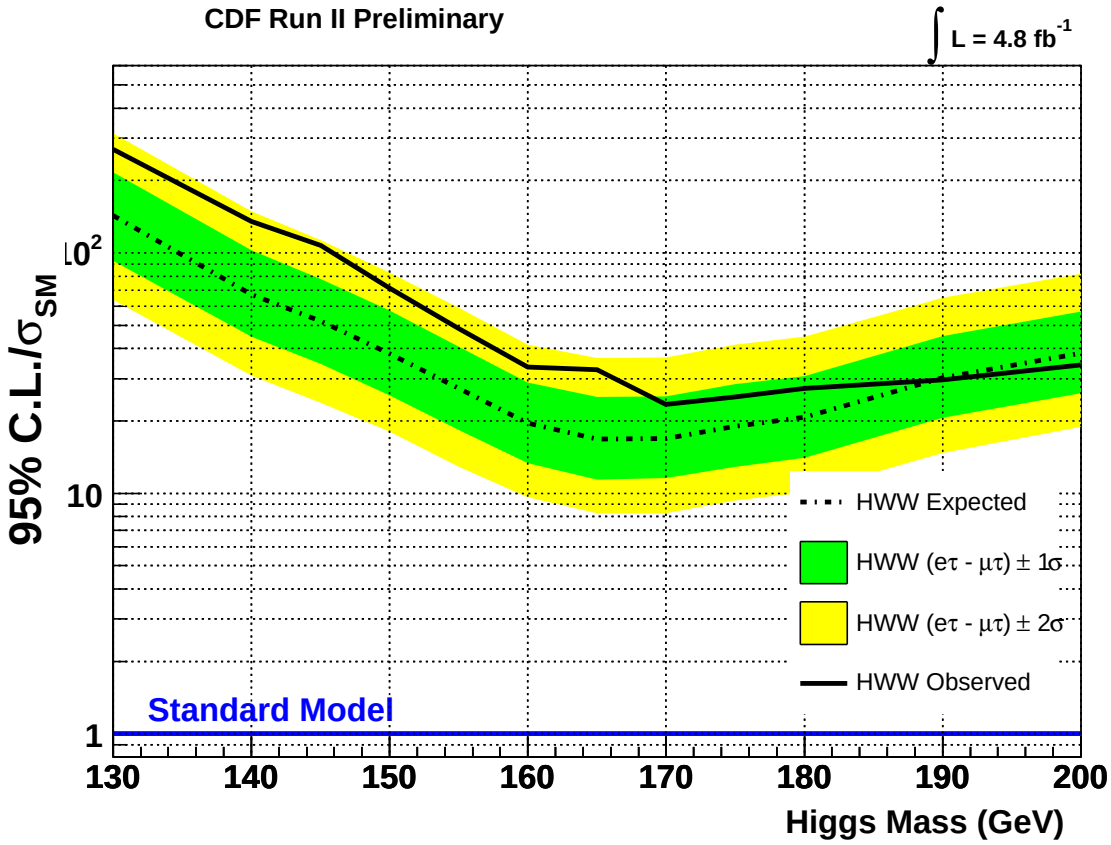


Figure 6.15: Combined observed and expected limits on the ratio of the SM Higgs production cross section for the $e - \tau$ and $\mu - \tau$ combined channels.

Mass [GeV/c^2]	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Observed
130	63.74	92.36	142.58	215.77	313.99	269.91
140	30.80	44.93	67.19	102.46	149.23	135.26
145	23.81	34.45	51.89	77.68	112.74	107.46
150	18.02	25.55	38.03	57.59	82.62	71.52
155	12.88	18.35	27.24	40.69	58.95	48.42
160	9.64	13.34	19.56	28.89	41.59	33.54
165	8.27	11.40	16.79	25.18	36.44	32.65
170	8.26	11.56	16.89	25.33	36.80	23.39
175	9.34	12.92	19.03	28.39	41.41	25.12
180	10.08	14.06	20.72	30.74	44.83	27.27
190	14.69	20.65	30.18	45.13	65.19	29.57
200	18.92	26.10	38.30	56.87	81.79	34.25

Table 6.8: Expected and observed limits for the $\mu - \tau$ and $e - \tau$ combined channel.

Chapter 7

Conclusions

In this Thesis a direct search for the Standard Model Higgs boson is presented. Hunting the Higgs boson at hadron colliders has acquired a strong interest and a lot of strengths have been invested to explore the expected scenarios involving the Higgs boson at different mass hypotheses made. Searches by previous experiments and constraints due to precision measurements of the Standard Model from theory indicate that the Higgs particle should have a mass in the range $114 \leq M_H \leq 250$ GeV/c^2. The available statistics collected by the experiments, CDF and DØ, and the improvements in the analysis techniques have allowed an impressive progress in the analysis of Tevatron data, as shown in Figure 7.1. Only the addition of data is not sufficient to provide enough sensitivity to probe Standard Model predictions. The orange band in Figure 7.1 shows the effect of improvements from 10% to 50% on the sensitivity.

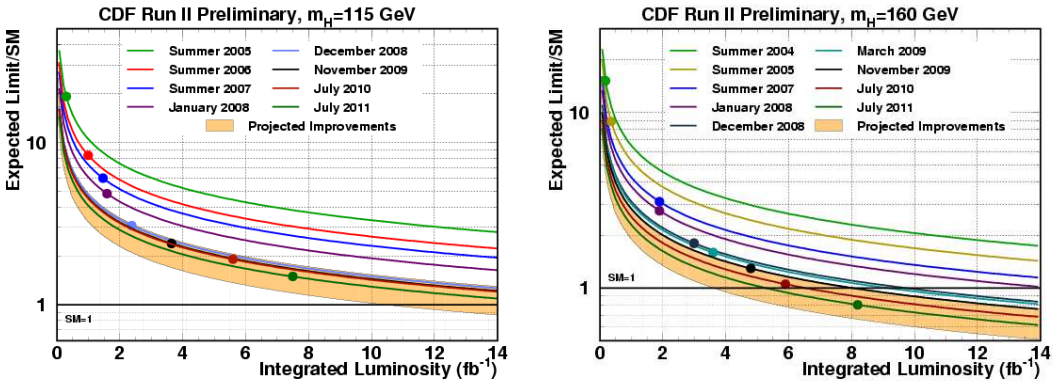


Figure 7.1: Achieved and projected median expected upper limits on the SM Higgs boson cross section, scaling the CDF performances to twice the luminosity. The solid lines are $1/\sqrt{L}$ projections, as function of integrated luminosity per experiment and each analysis update corresponds to a new point with a new curve. 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. (Left) Projections are shown for $M_H = 115$ GeV/c^2. The luminosity for the July 2011 point is 5.6 fb^{-1}. (Right) Projections are shown for $M_H = 160$ GeV/c^2. The luminosity for the July 2011 point is 5.9 fb^{-1}.

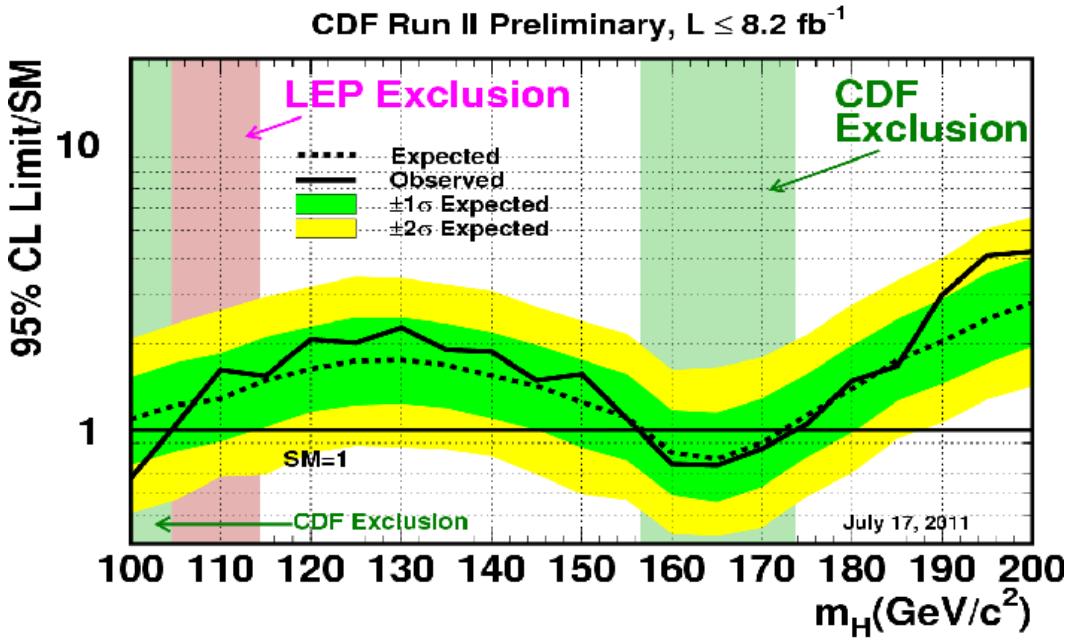


Figure 7.2: The expected and the observed limits on the Higgs production cross section obtained by combining CDF and DØ results as function of the Higgs mass. The limits are in units of expected SM cross section, for test masses (every 5 GeV/c^2) for which both experiments have performed dedicated searches in different channels.

This Thesis focuses on the Higgs search in the $H \to WW$ decay channel, with the two Ws decaying leptonically. It is the most promising channel for the Higgs mass hypothesis in the range $135 \leq M_H \leq 200$ GeV/c^2. Many analyses have been performed on statistically independent samples, covering all the expected final states given by the WW leptonic decays and the four Higgs production mechanisms.

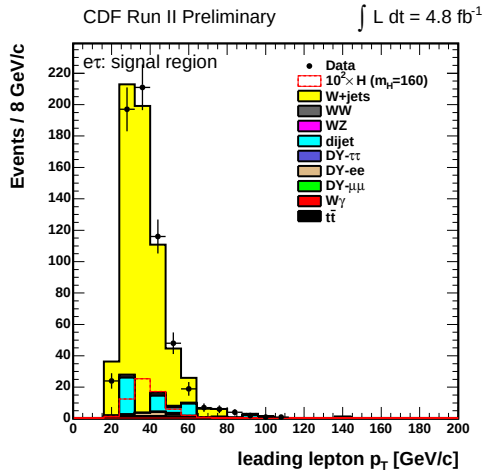
Four Higgs production mechanisms are considered giving a total expected signal yield, respectively for the $(e-\tau)$ or $(\mu-\tau)$ channels with τ decaying hadronically, of 0.93 ± 0.01 or 0.56 ± 0.01 events for a Higgs mass of 160 GeV/c^2 on top of an expected background of 644.7 ± 22.7 or 463.0 ± 19.6 events, in 4.8 fb^{-1} of data.
Advanced analysis techniques are used to further separate signal from background, setting an upper limit on standard model Higgs production for 12 Higgs mass hypotheses in the range $130 \leq M_H \leq 200$ GeV/c^2 using $H \to WW$ decay mode. Combining the two (e-τ) and (μ-τ) analyses, the observed upper limit con the SM Higgs cross section is $23.39 \times \sigma_{SM}$ for $M_H = 160$ GeV at 95% C.L. .
Although the increase of available data, the hadronic tau channel alone will not be sensitive enough to exclude the mechanism of Electroweak symmetry breaking [60]. However, it is included in the full combination of $H \to WW$ analyses, now setting an observed exclusion (CDF only) at 95% C.L. of $156 \leq M_H \leq 175$ GeV/c^2 [127] with 8.2 fb^{-1} of data.

The single CDF and DØ experiments are now sensitive to excluding Higgs bosons mass ranges, as more sophisticated data analysis techniques and more data from the Tevatron collider have increased experiments sensitivity to the Higgs boson. Combining

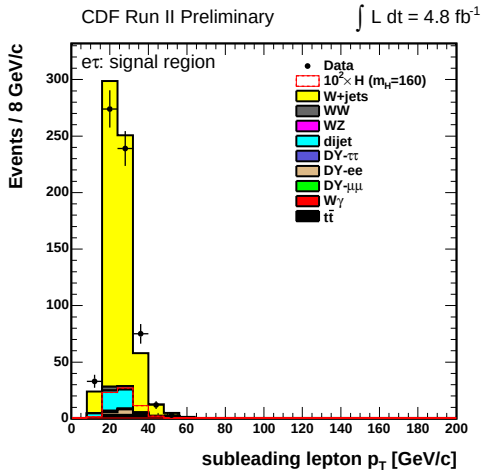
their independent Higgs analyses performed with 8.2 fb^{-1} of data, the Higgs boson mass ranges $100.0 \leq M_H \leq 104.5$ GeV/c^2 and $156.5 \leq M_H \leq 173.7$ GeV/c^2, have been excluded at 95% C.L. [128]. The ratios of the 95% C.L. expected and observed limit to the SM cross section are shown in Figure 7.2 for the combined CDF and DØ analyses. Therefore, the Tevatron has improved the SM sensitivity by combining analysis channels and combining the data from the two experiments. The result greatly improves the sensitivity of the individual searches and set the new world best limits on the Higgs boson production cross section by a single experiment, expected to grow substantially as more data are added and further improvements are made to the analysis techniques.

Appendix A

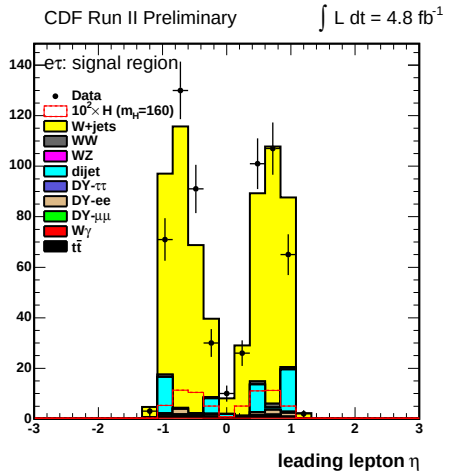
Signal region: $e\tau$ and $\mu\tau$ channels



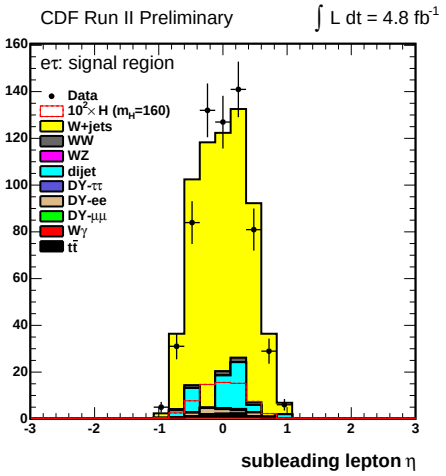
(a) p_T of the leading lepton.



(b) p_T of the subleading lepton.

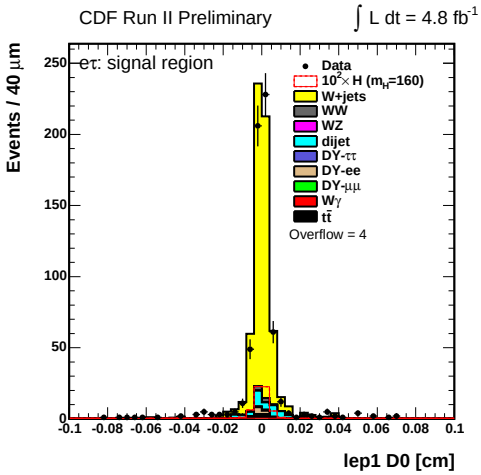


(c) Pseudorapidity of the leading lepton.

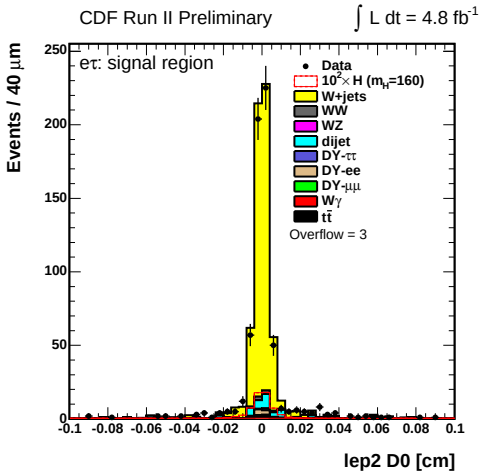


(d) Pseudorapidity of the subleading lepton.

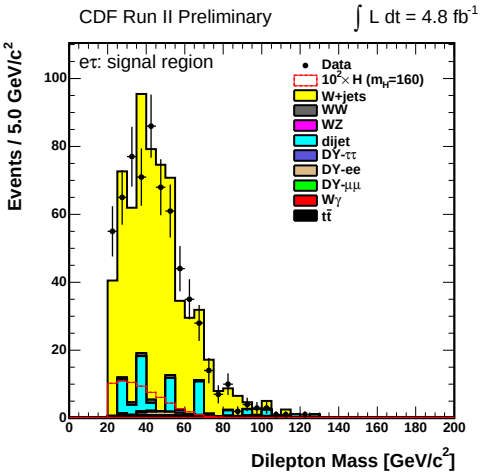
Figure A.1: Data-MC comparison in the $e\tau$ tight signal region.



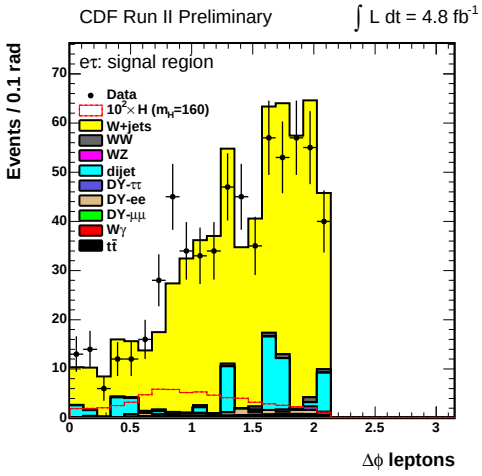
(a) Impact parameter of the leading lepton.



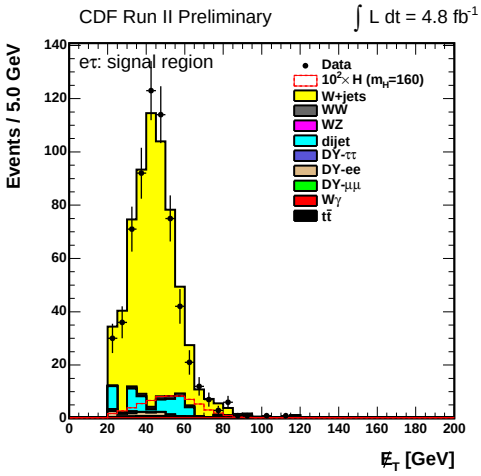
(b) Impact parameter of the subleading lepton.



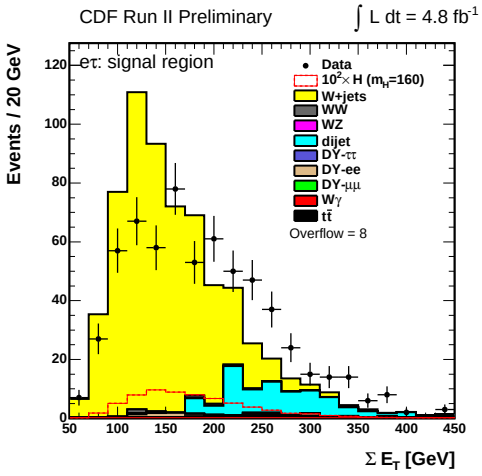
(c) Electron-tau invariant mass.



(d) $\Delta\varphi$ between the tau and the electron.

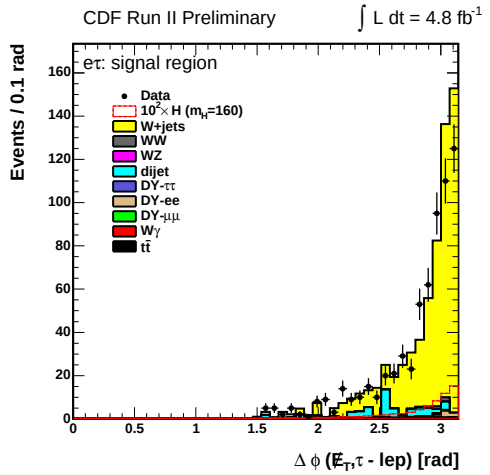


(e) Missing transverse energy.

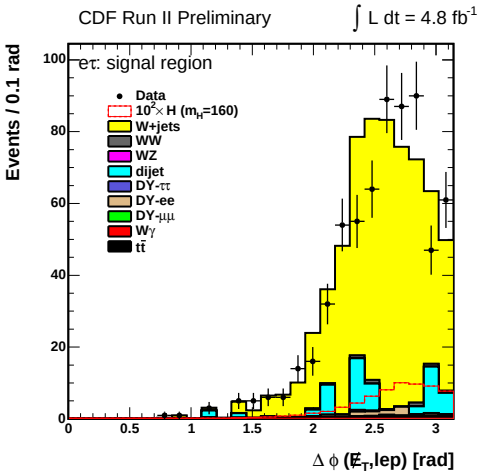


(f) Sum of the E_T of the reconstructed objects in the event

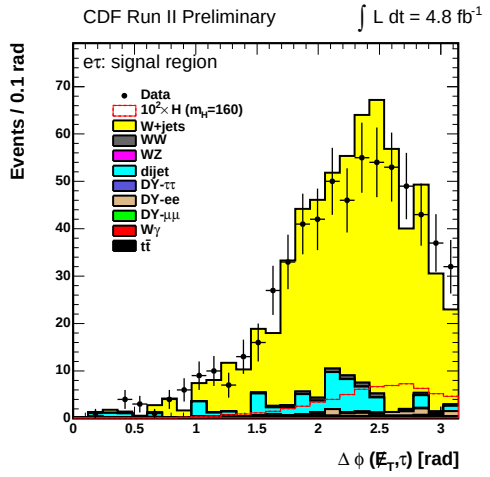
Figure A.2: Data-MC comparison in the $e\tau$ tight signal region.



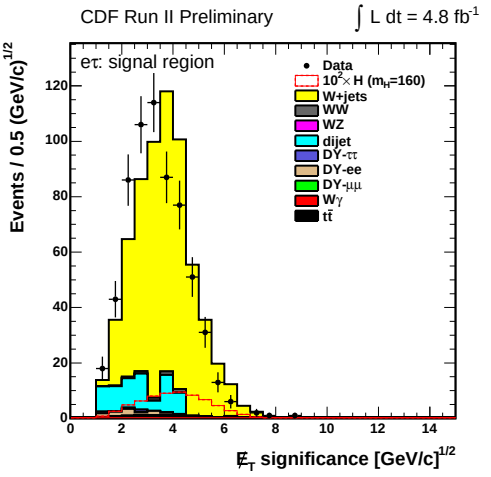
(a) $\Delta\varphi$ between the dilepton momentum and the $\not\!\!E_T$.



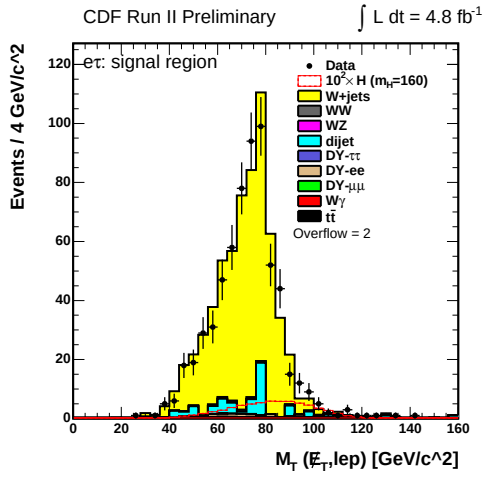
(b) $\Delta\varphi$ between the electron and the missing E_T.



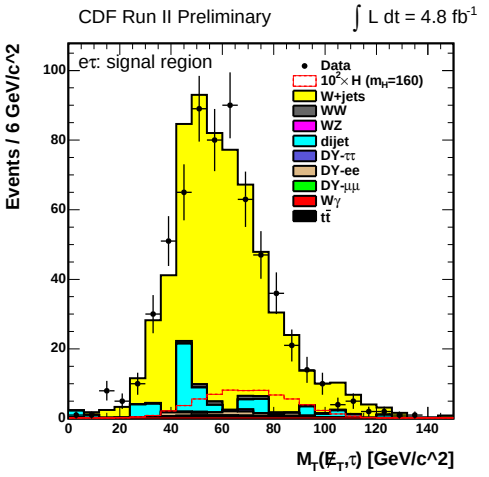
(c) $\Delta\varphi$ between the tau and the missing E_T.



(d) Significance of the missing transverse energy.

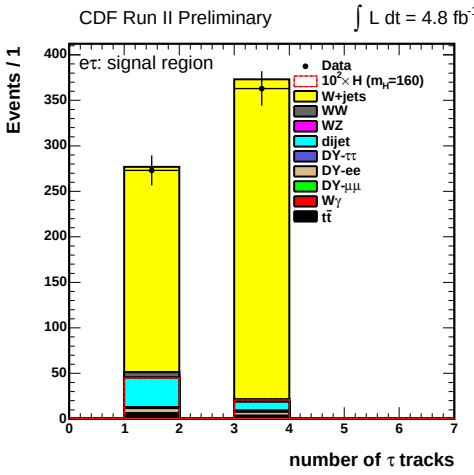


(e) τ-$\not\!\!E_T$ transverse mass.

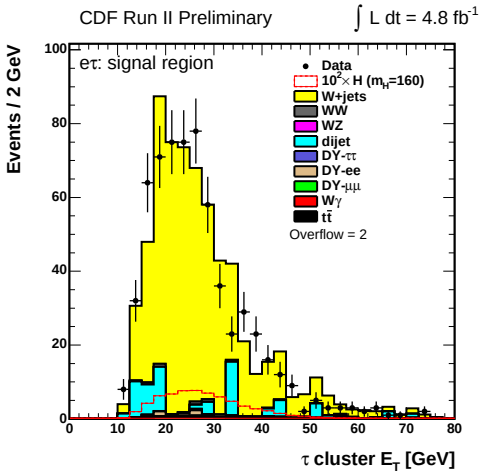


(f) e-$\not\!\!E_T$ transverse mass.

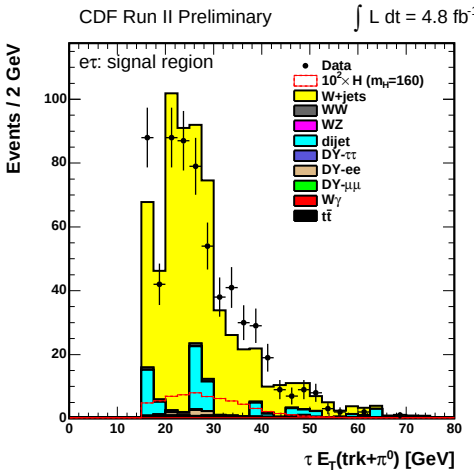
Figure A.3: Data-MC comparison in the $e\tau$ tight signal region.



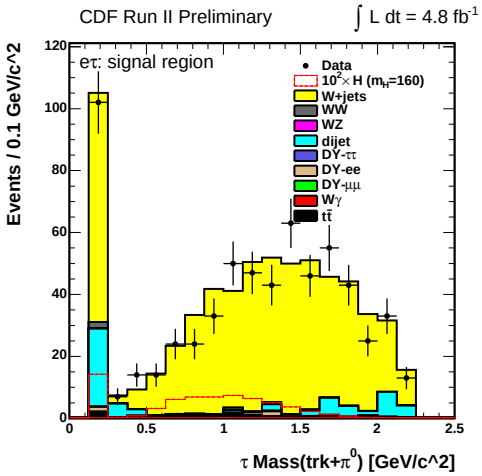
(a) Number of tracks of the reconstructed τ.



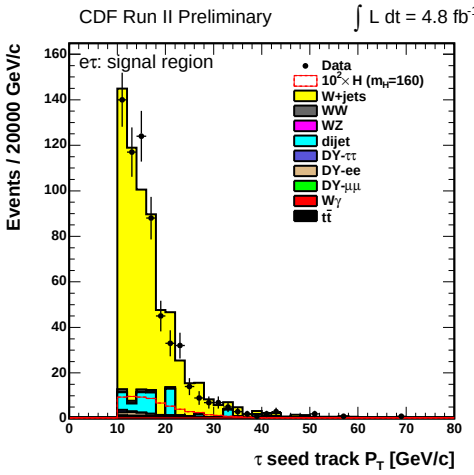
(b) Transverse energy of the τ calorimetric cluster.



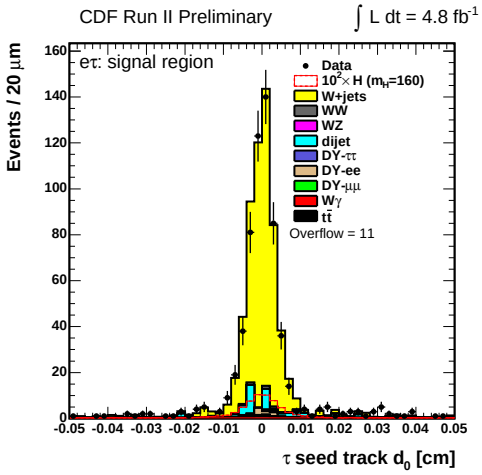
(c) Transverse energy of tracks and π^0's in the τ signal cone.



(d) Mass of tracks and π^0's in the τ signal cone.

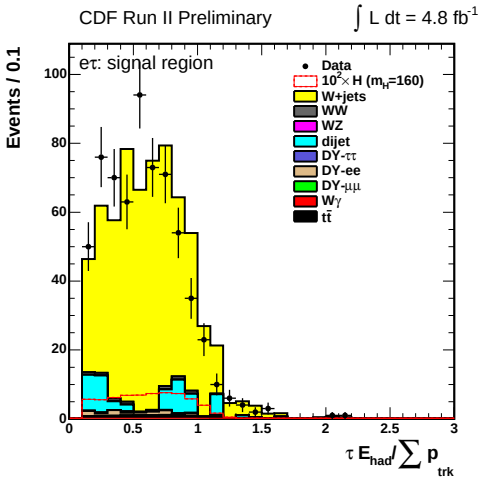


(e) Transverse momentum of the τ seed track.

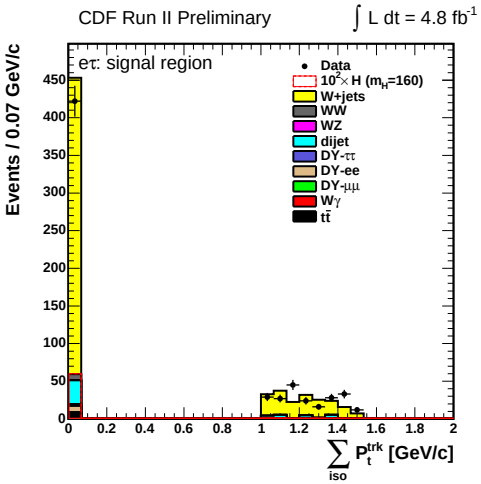


(f) Impact parameter of the τ seed track.

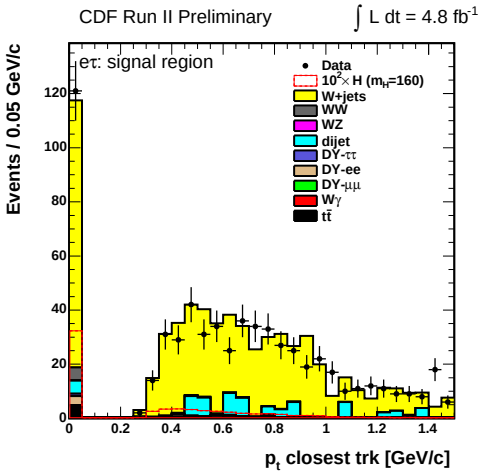
Figure A.4: Data-MC comparison in the $e\tau$ tight signal region.



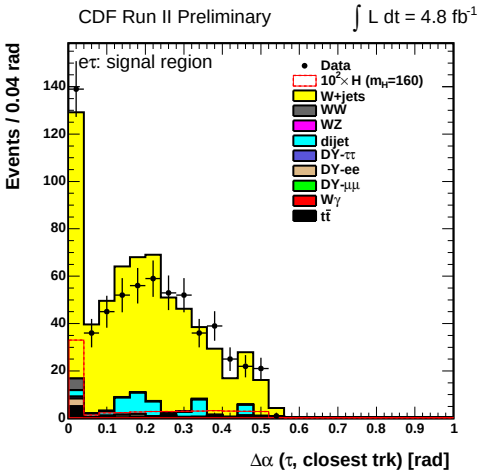
(a) Ratio of the τ cluster energy to the total momentum of the tracks associated to the τ.



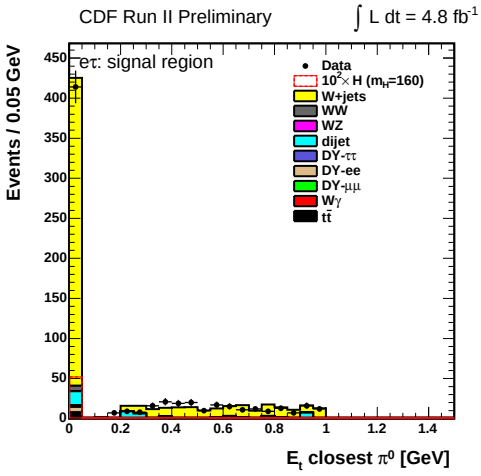
(b) τ track isolation.



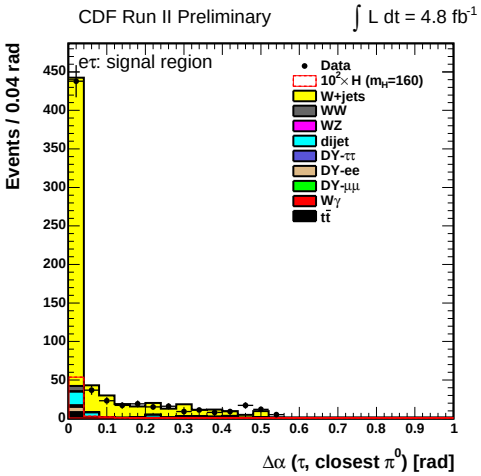
(c) p_T of the closest track to the τ.



(d) Angle between the τ and the closest track.

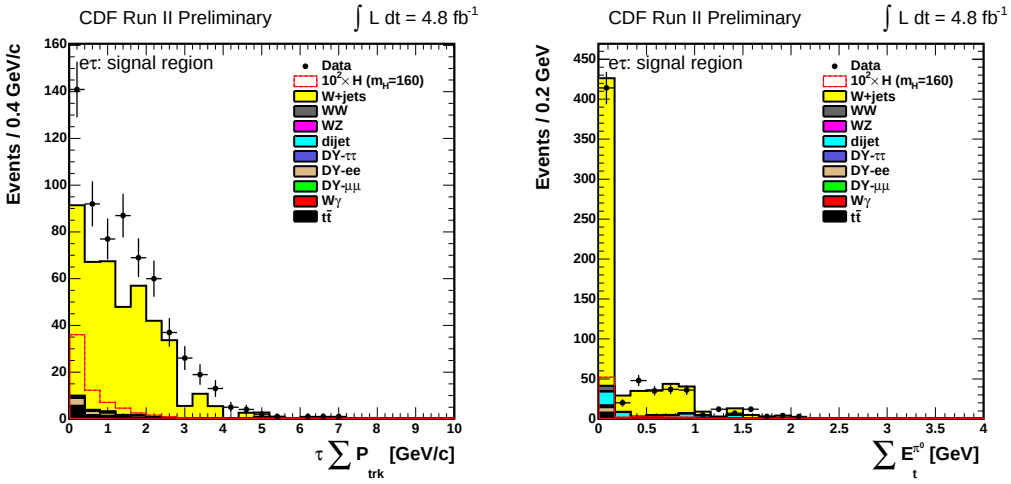


(e) E_T of the closest π^0 to the τ.



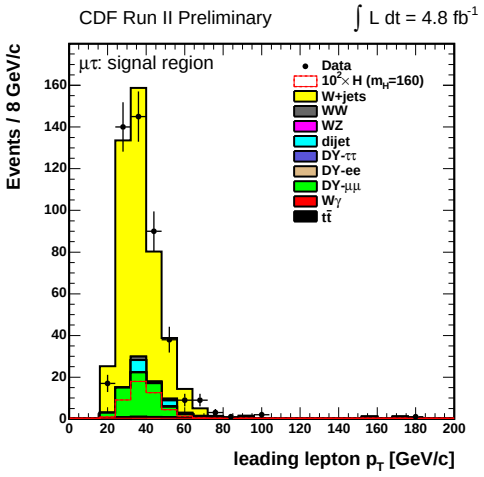
(f) Angle between the τ and the closest π^0.

Figure A.5: Data-MC comparison in the $e\tau$ tight signal region.

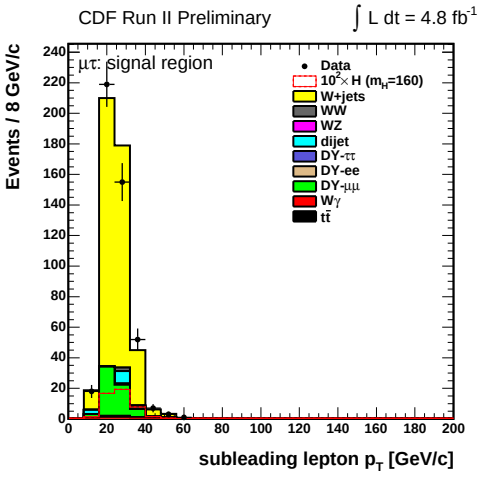


(a) Sum of p_T of all tracks in the τ isolation cone. (b) Sum of E_T of all π^0's in the τ isolation cone.

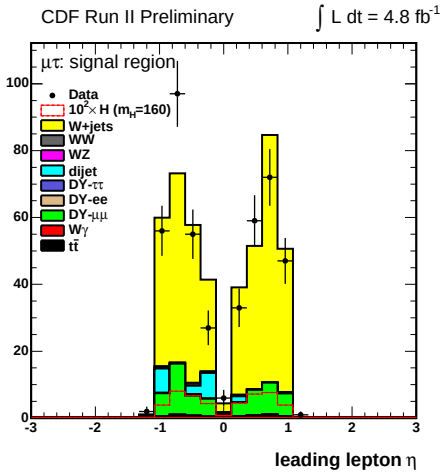
Figure A.6: Data-MC comparison in the $e\tau$ tight signal region.



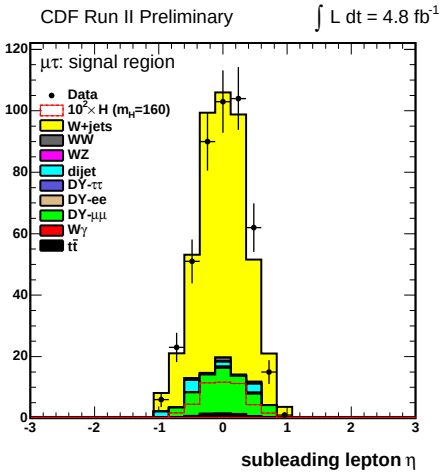
(a) p_T of the leading lepton.



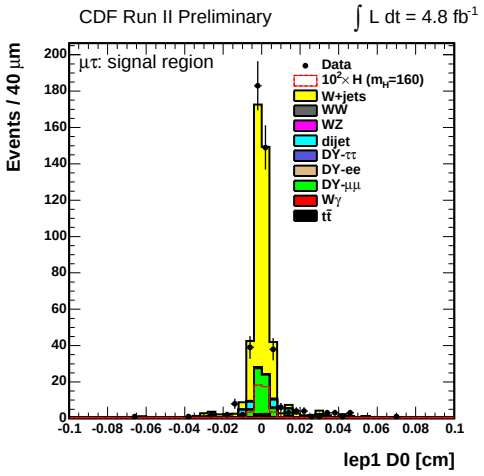
(b) p_T of the subleading lepton.



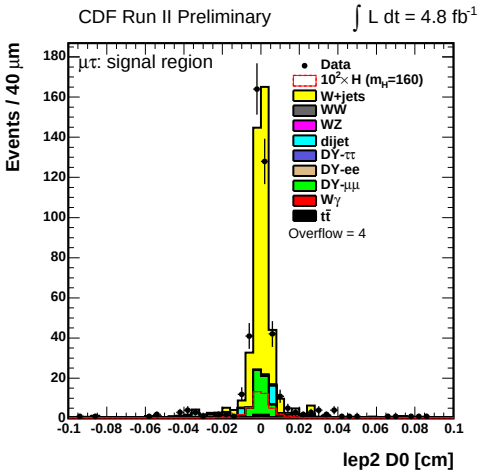
(c) Pseudorapidity of the leading lepton.



(d) Pseudorapidity of the subleading lepton.

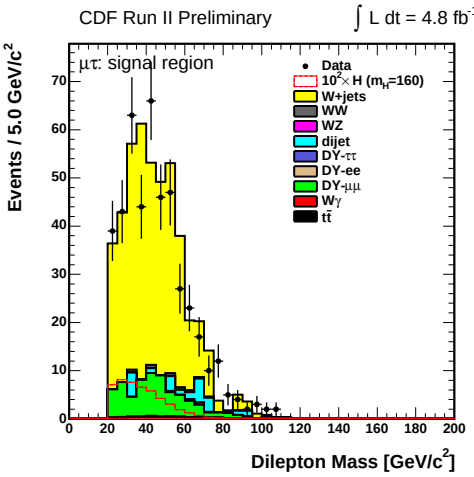


(e) Impact parameter of the leading lepton.

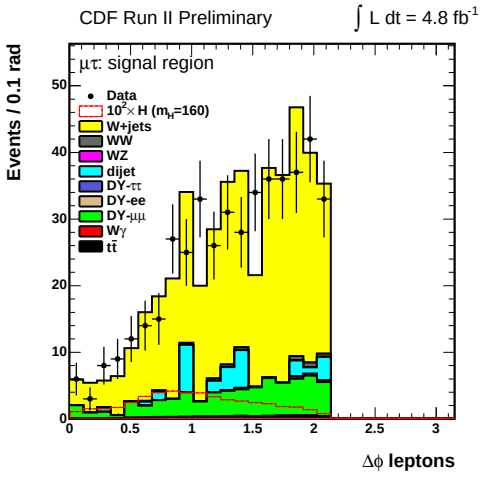


(f) Impact parameter of the subleading lepton.

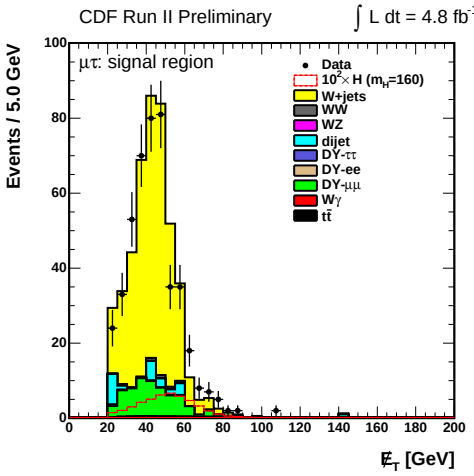
Figure A.7: Data-MC comparison in the $\mu\tau$ tight signal region.



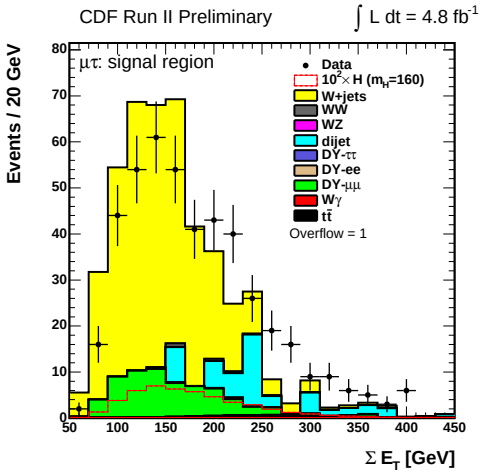
(a) Muon-tau invariant mass.



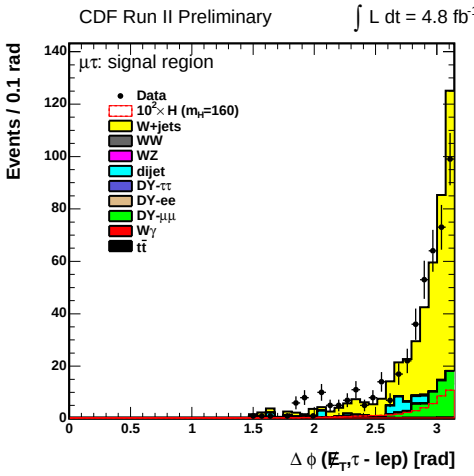
(b) $\Delta\varphi$ between the tau and the muon.



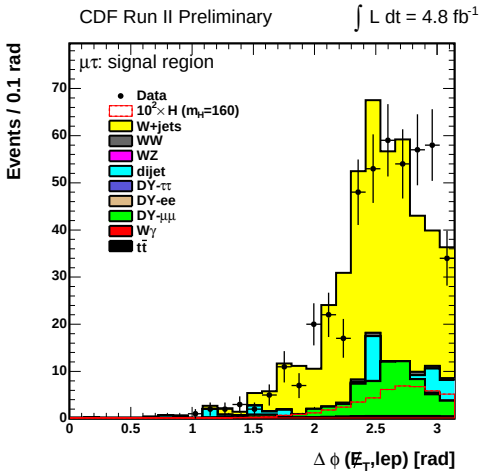
(c) Missing transverse energy.



(d) Sum of the E_T of the reconstructed objects in the event

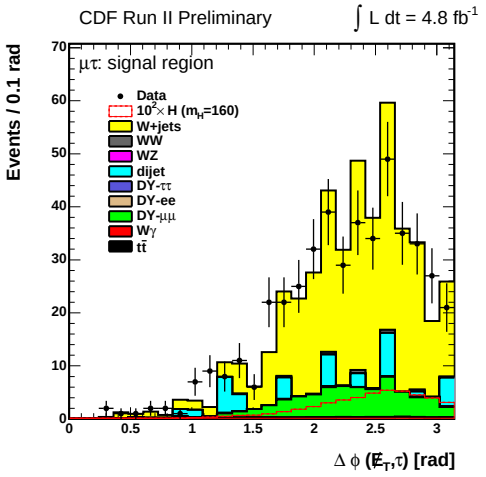


(e) $\Delta\varphi$ between the dilepton momentum and the $\not\!\!E_T$.

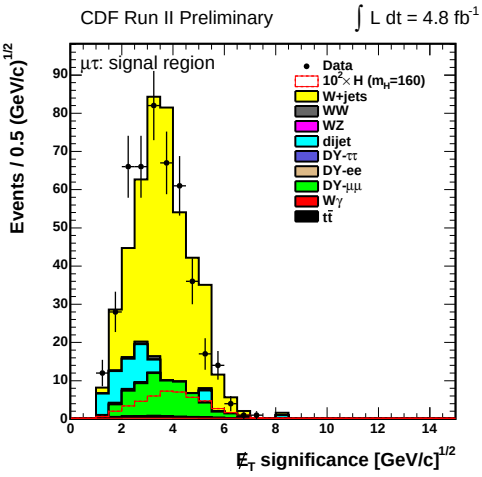


(f) $\Delta\varphi$ between the muon and the missing E_T.

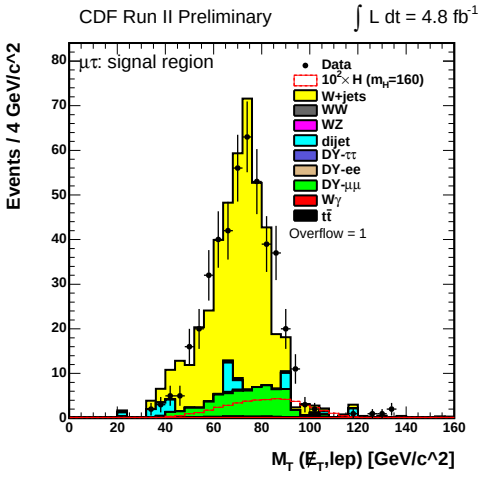
Figure A.8: Data-MC comparison in the $\mu\tau$ tight signal region.



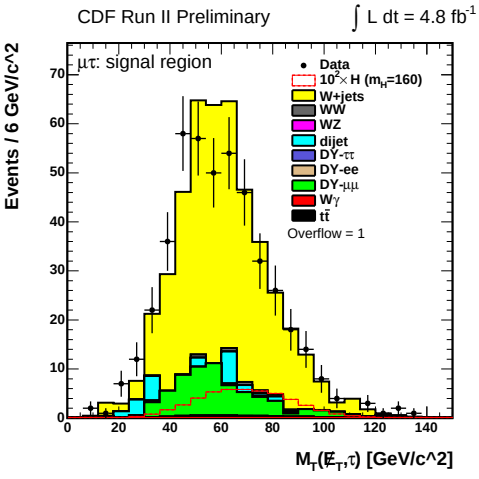
(a) $\Delta\varphi$ between the tau and the missing E_T.



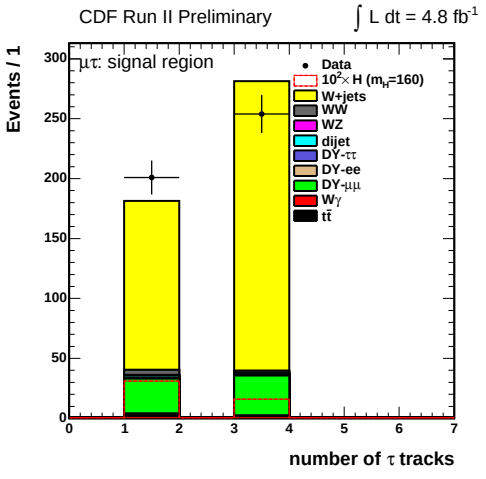
(b) Significance of the missing transverse energy.



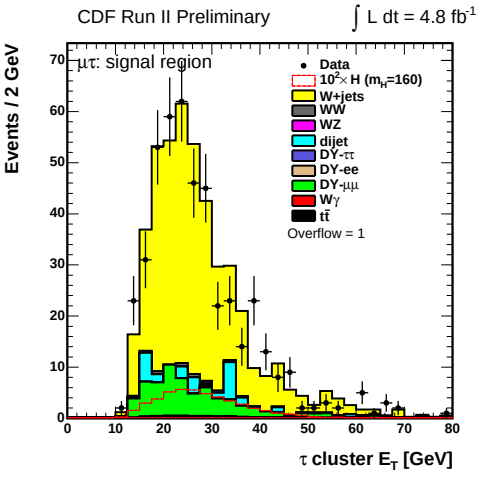
(c) τ-$\not\!\!E_T$ transverse mass.



(d) μ-$\not\!\!E_T$ transverse mass.

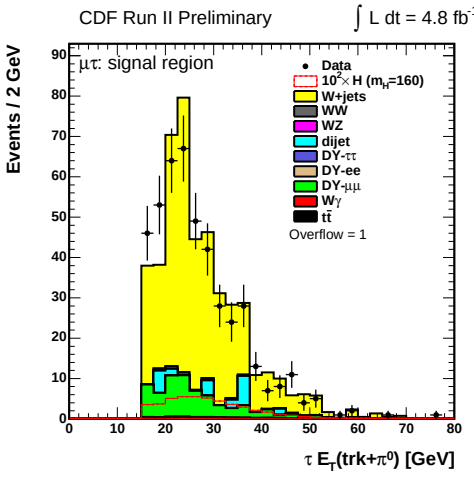


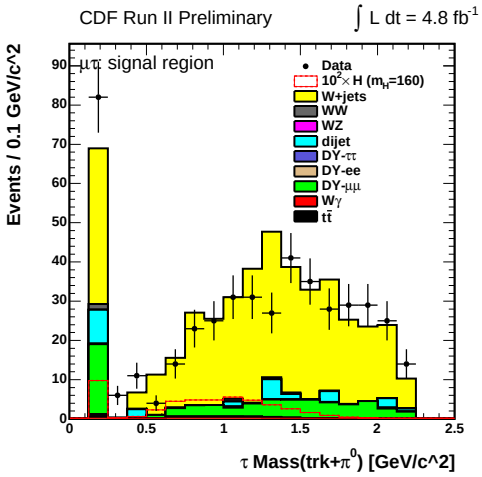
(e) Number of tracks of the reconstructed τ.



(f) Transverse energy of the τ calorimetric cluster.

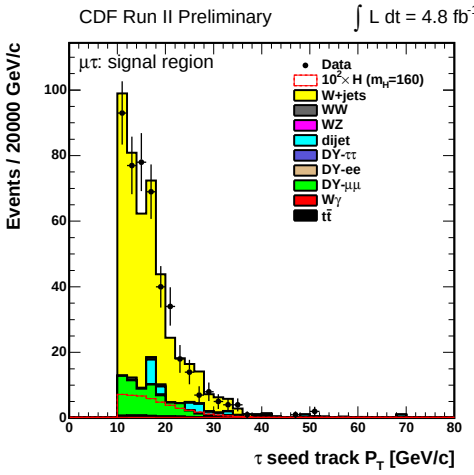
Figure A.9: Data-MC comparison in the $\mu\tau$ tight signal region.

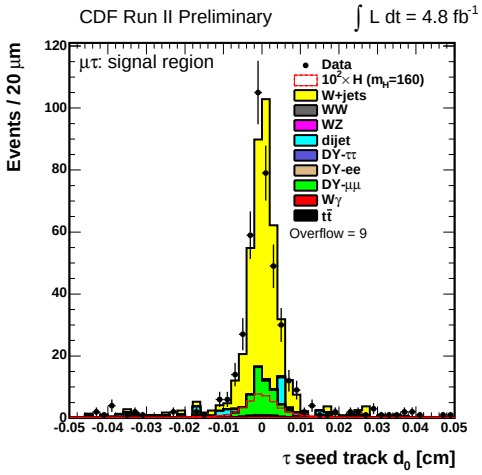




(a) Transverse energy of tracks and π^0's in the τ signal cone.

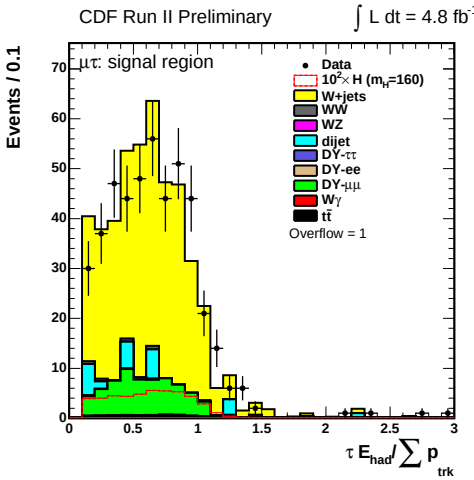
(b) Mass of tracks and π^0's in the τ signal cone.

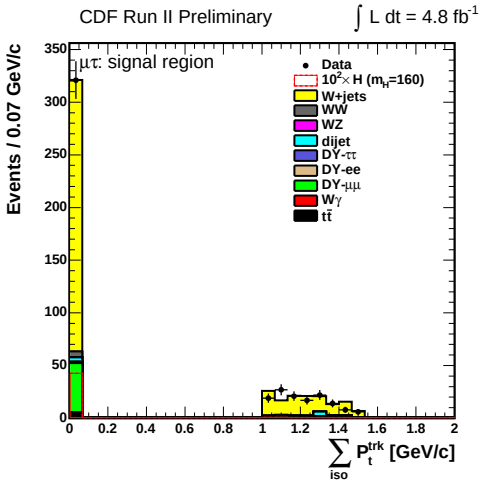




(c) Transverse momentum of the τ seed track.

(d) Impact parameter of the τ seed track.

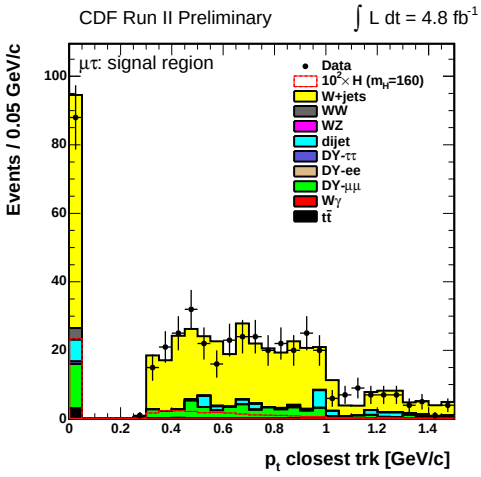




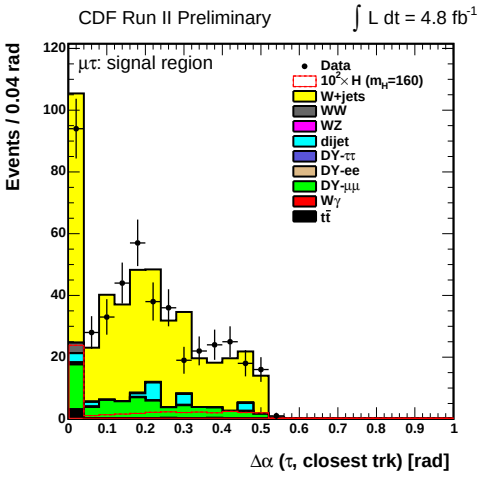
(e) Ratio of the τ cluster energy to the total momentum of the tracks associated to the τ.

(f) τ track isolation.

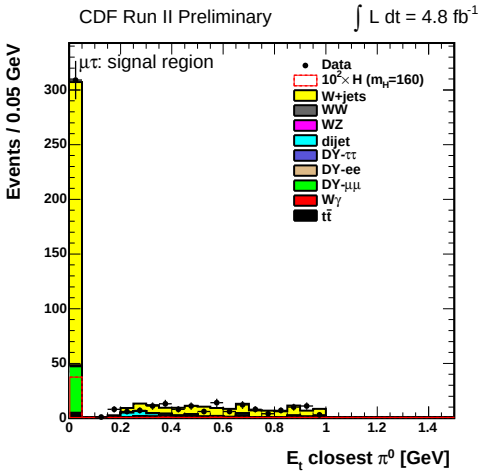
Figure A.10: Data-MC comparison in the $\mu\tau$ tight signal region.



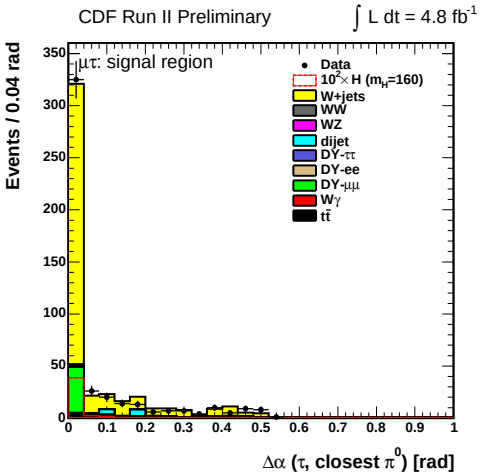
(a) p_T of the closest track to the τ.



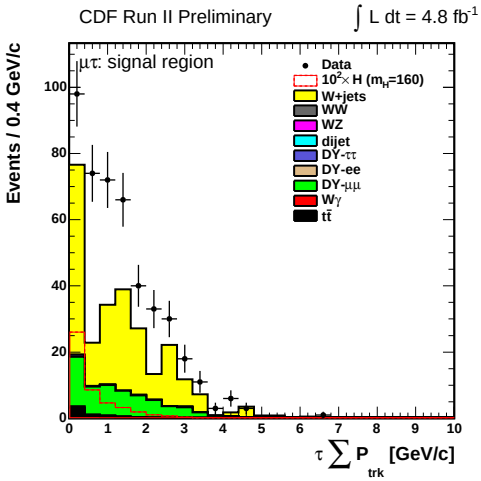
(b) Angle between the τ and the closest track.



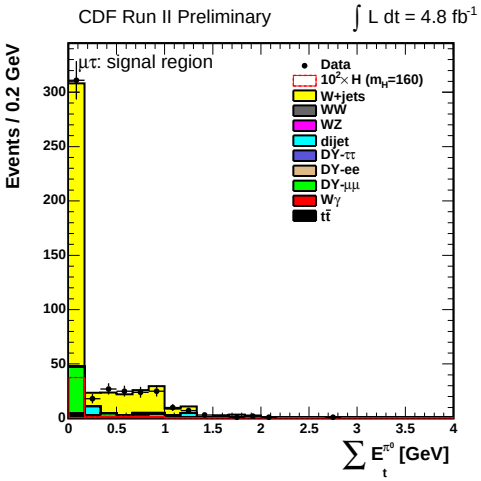
(c) E_T of the closest π^0 to the τ.



(d) Angle between the τ and the closest π^0.



(e) Sum of p_T of all tracks in the τ isolation cone.

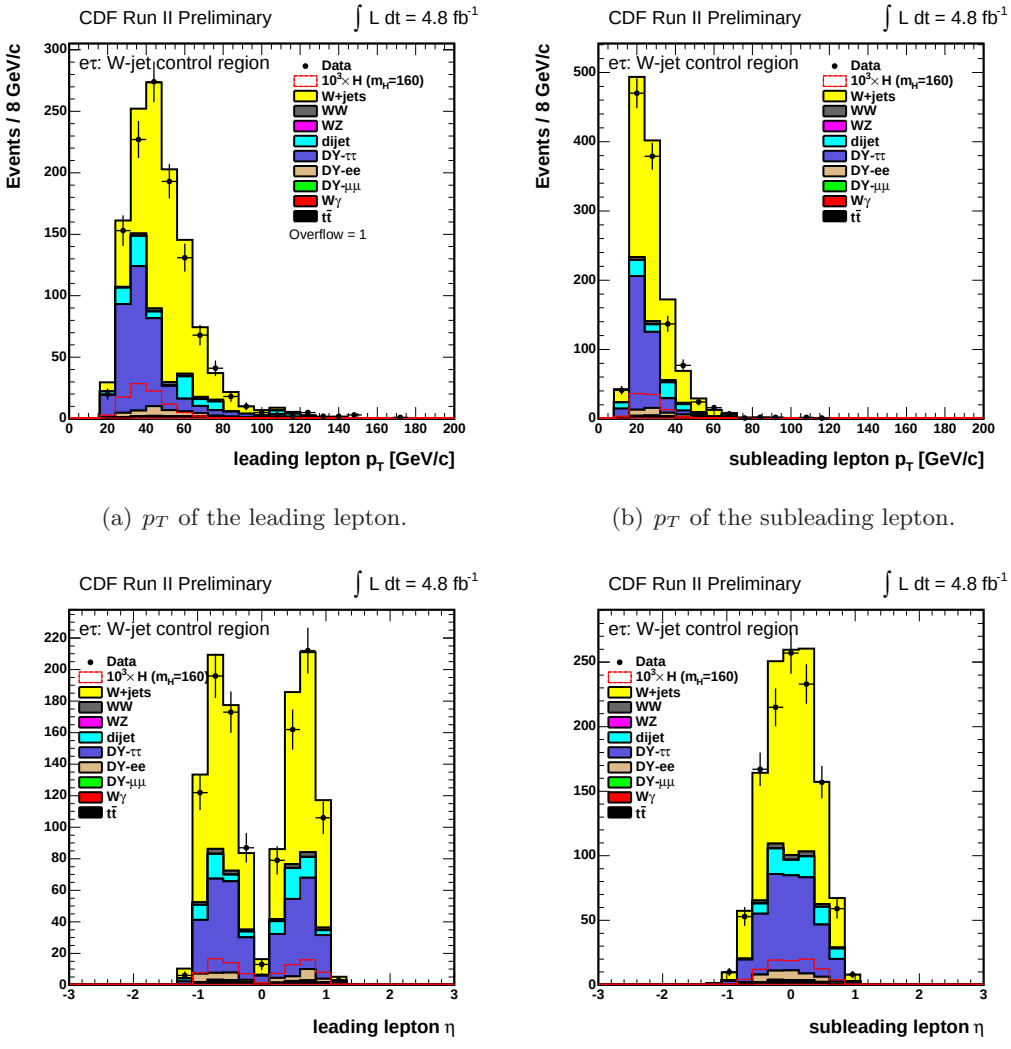


(f) Sum of E_T of all π^0's in the τ isolation cone.

Figure A.11: Data-MC comparison in the $\mu\tau$ tight signal region.

Appendix B

W+jet control region: $e\tau$ and $\mu\tau$ channels



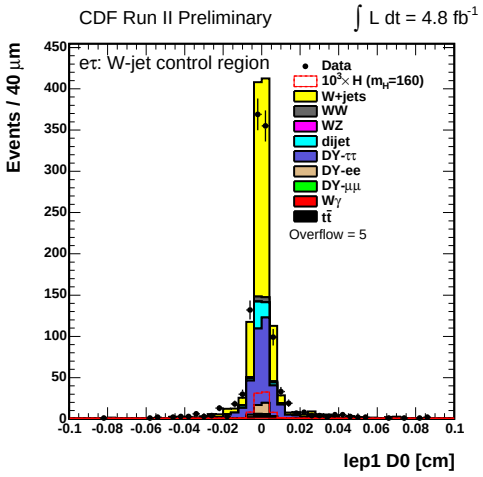
(a) p_T of the leading lepton.

(b) p_T of the subleading lepton.

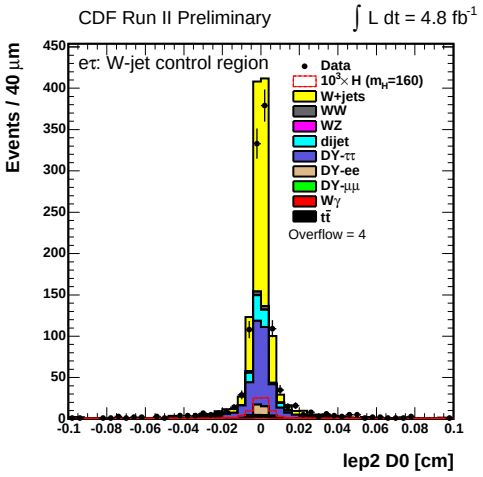
(c) Pseudorapidity of the leading lepton.

(d) Pseudorapidity of the subleading lepton.

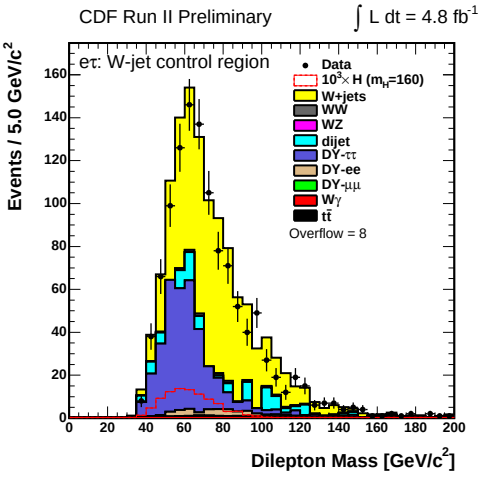
Figure B.1: Data-MC comparison in the $e\tau$ W+jet dominated region.



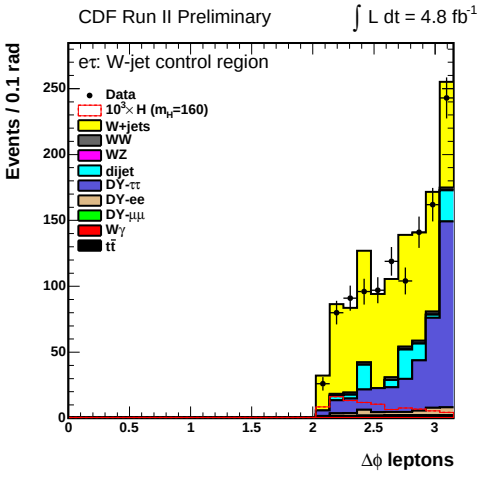
(a) Impact parameter of the leading lepton.



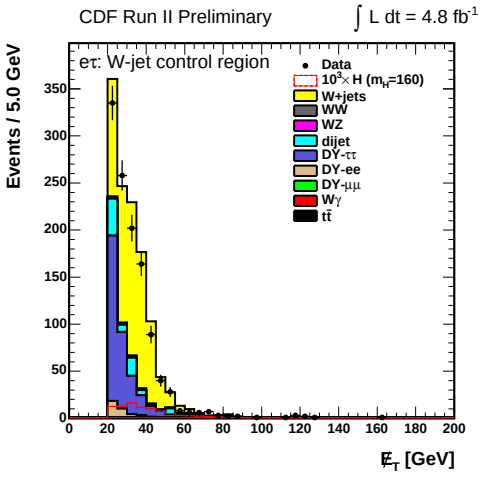
(b) Impact parameter of the subleading lepton.



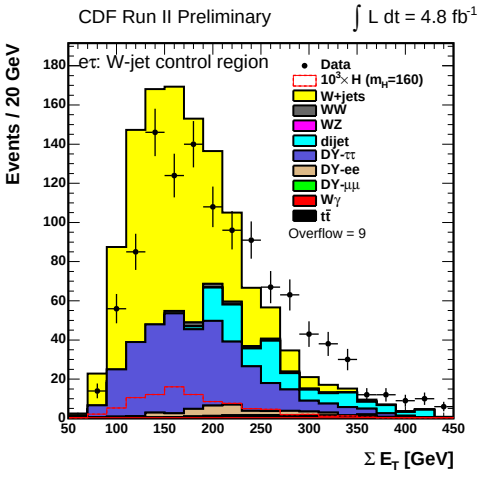
(c) Electron-tau invariant mass.



(d) $\Delta\varphi$ between the tau and the electron.

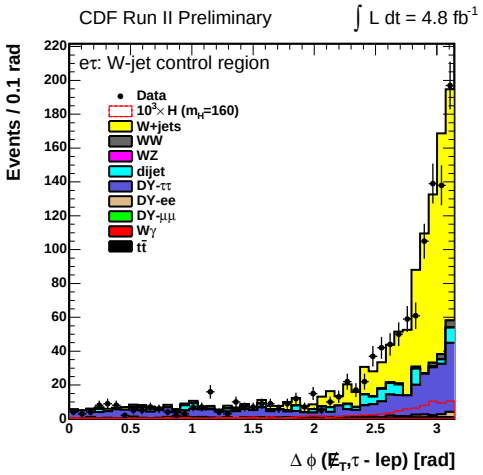


(e) Missing transverse energy.

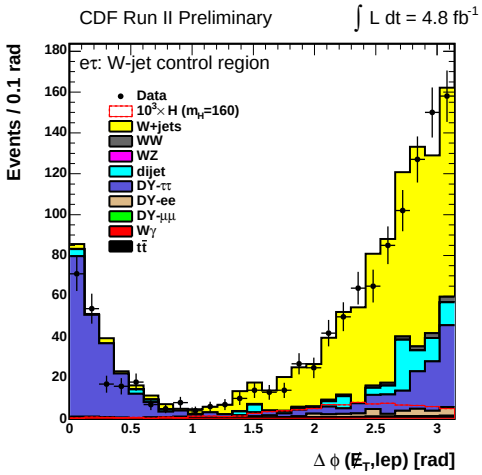


(f) Sum of the E_T of the reconstructed objects in the event

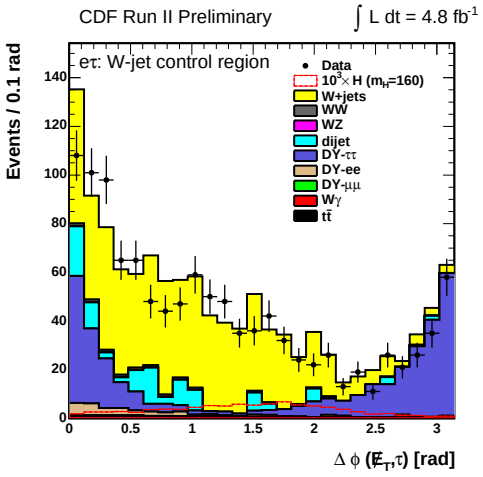
Figure B.2: Data-MC comparison in the $e\tau$ W+jet dominated region.



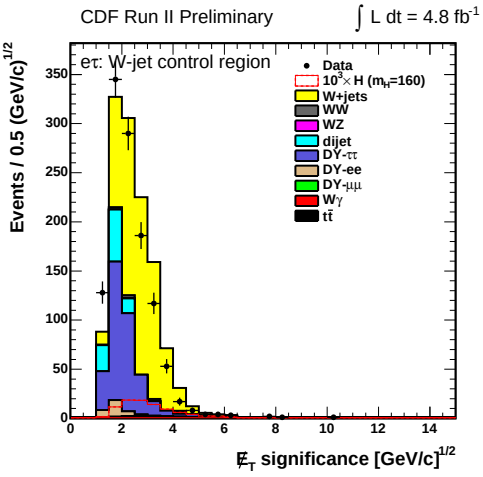
(a) $\Delta\varphi$ between the dilepton momentum and the $\not\!\!E_T$.



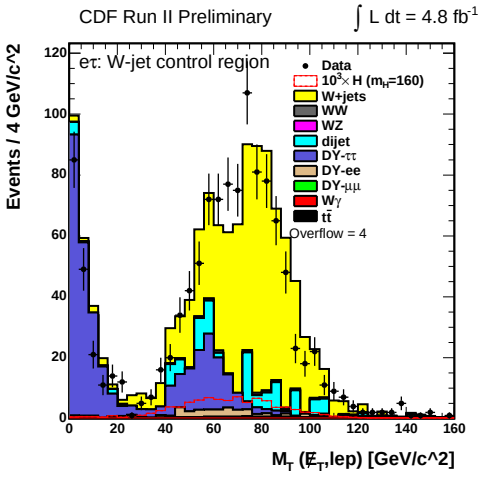
(b) $\Delta\varphi$ between the electron and the missing E_T.



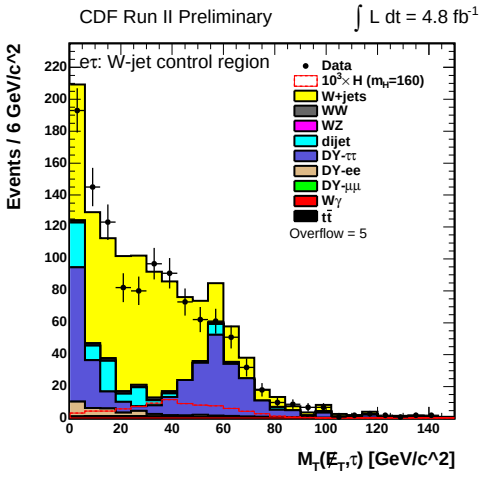
(c) $\Delta\varphi$ between the tau and the missing E_T.



(d) Significance of the missing transverse energy.

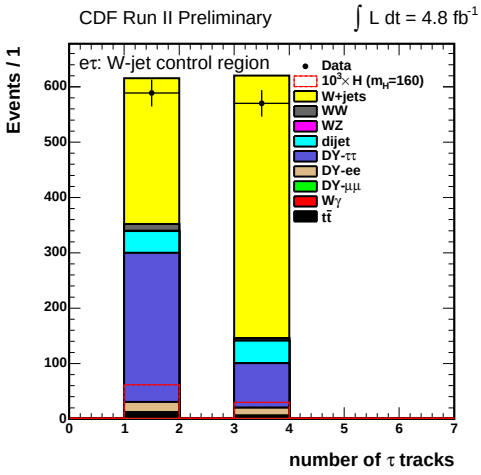


(e) τ-$\not\!\!E_T$ transverse mass.

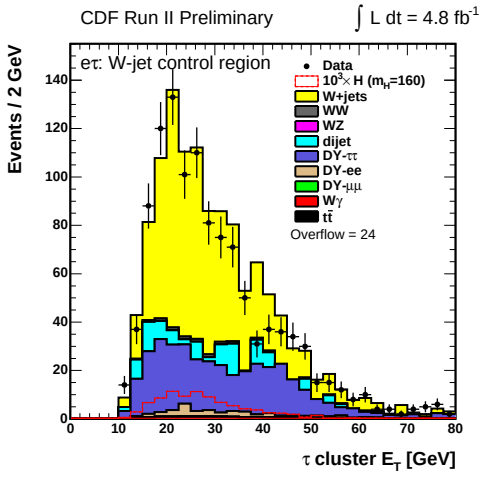


(f) e-$\not\!\!E_T$ transverse mass.

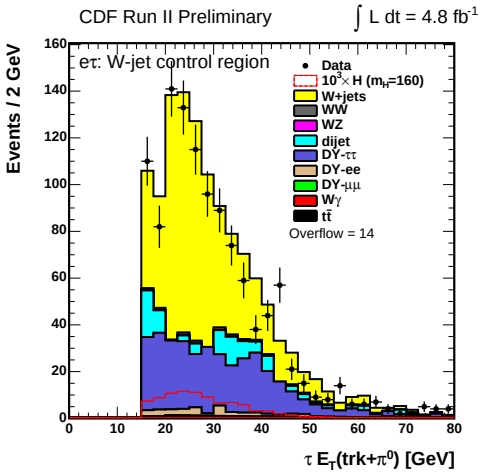
Figure B.3: Data-MC comparison in the $e\tau$ W+jet dominated region.



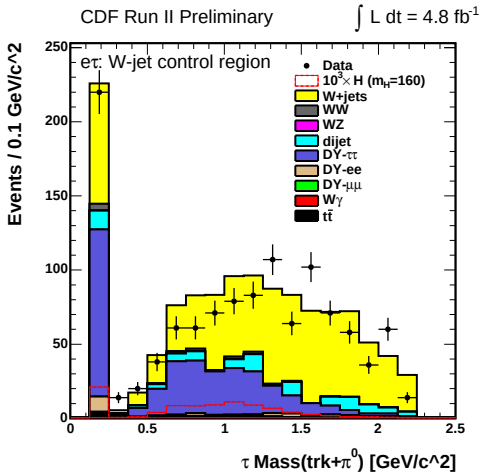
(a) Number of tracks of the reconstructed τ.



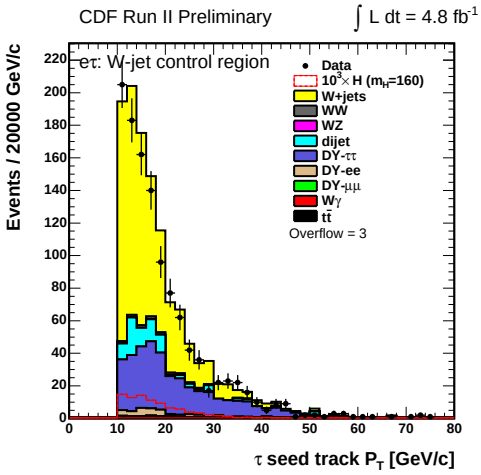
(b) Transverse energy of the τ calorimetric cluster.



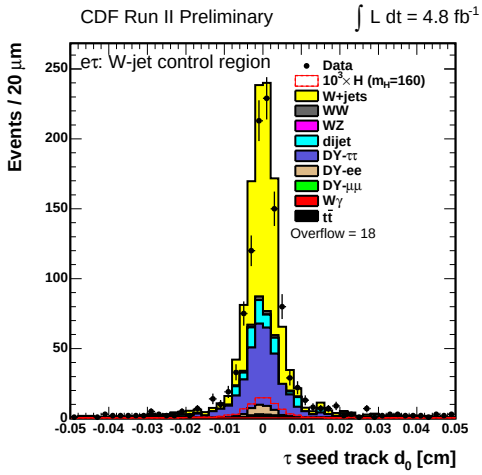
(c) Transverse energy of tracks and π^0's in the τ signal cone.



(d) Mass of tracks and π^0's in the τ signal cone.

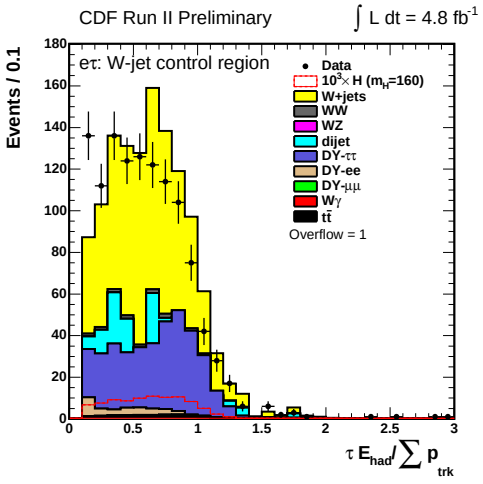


(e) Transverse momentum of the τ seed track.

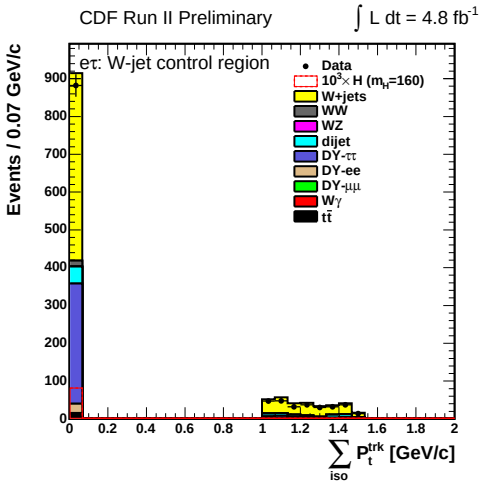


(f) Impact parameter of the τ seed track.

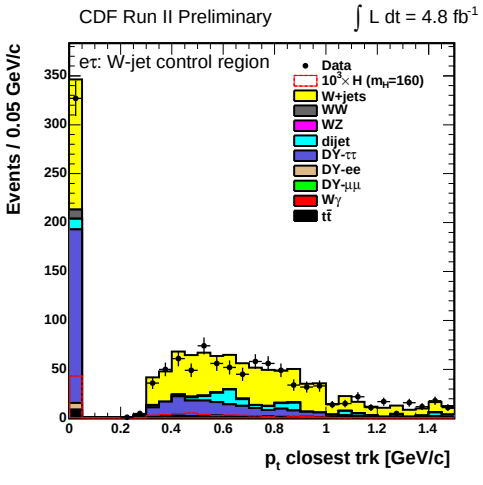
Figure B.4: Data-MC comparison in the $e\tau$ W+jet dominated region.



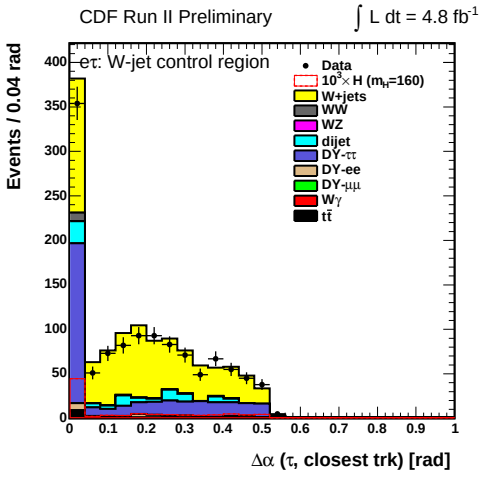
(a) Ratio of the τ cluster energy to the total momentum of the tracks associated to the τ.



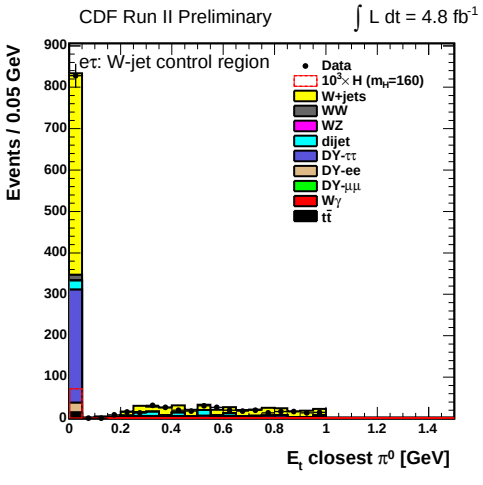
(b) τ track isolation.



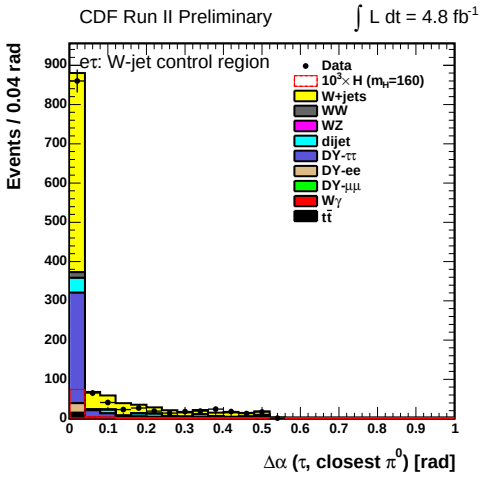
(c) p_T of the closest track to the τ.



(d) Angle between the τ and the closest track.

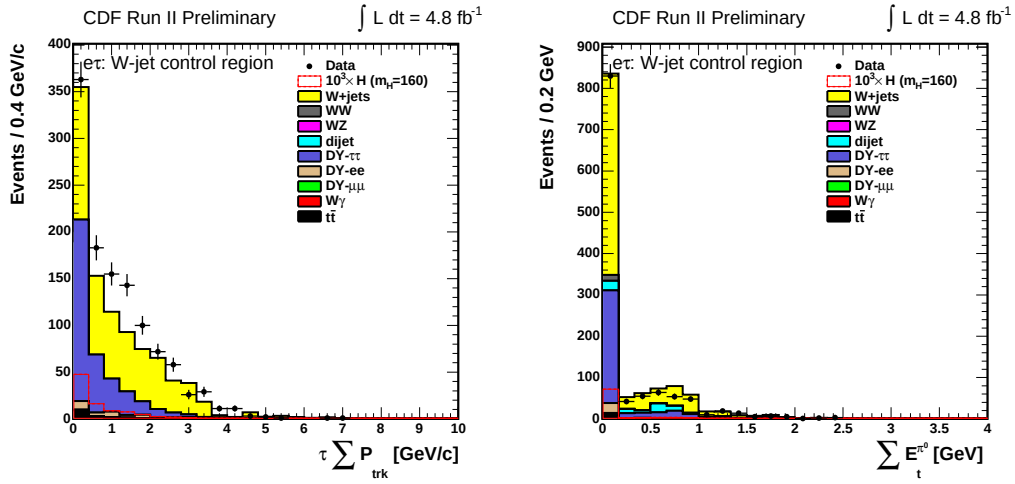


(e) E_T of the closest π^0 to the τ.



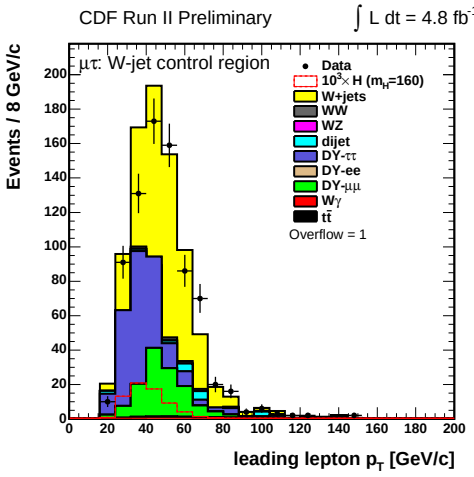
(f) Angle between the τ and the closest π^0.

Figure B.5: Data-MC comparison in the $e\tau$ W+jet dominated region.

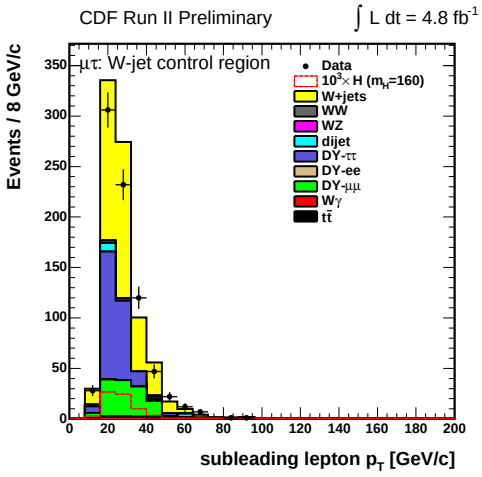


(a) Sum of p_T of all tracks in the τ isolation cone. (b) Sum of E_T of all π^0's in the τ isolation cone.

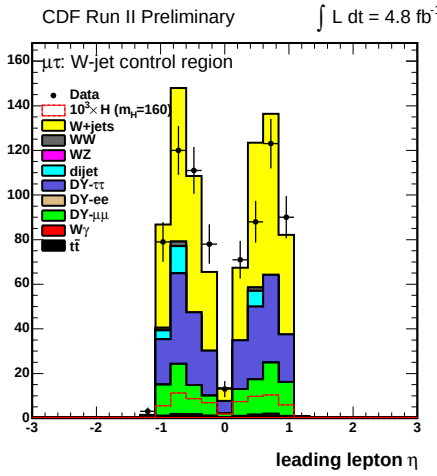
Figure B.6: Data-MC comparison in the $e\tau$ W+jet dominated region.



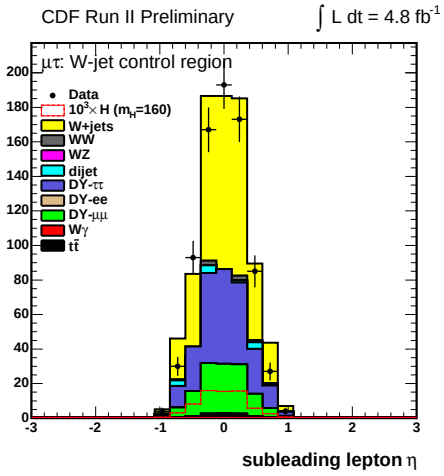
(a) p_T of the leading lepton.



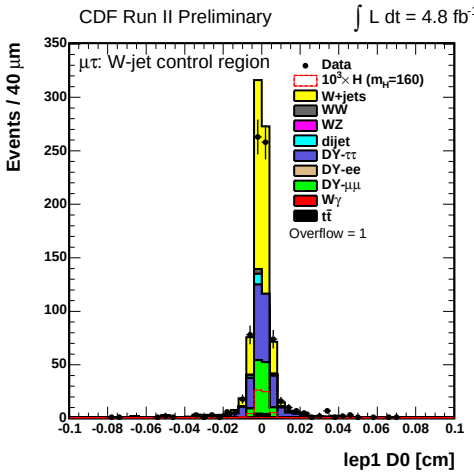
(b) p_T of the subleading lepton.



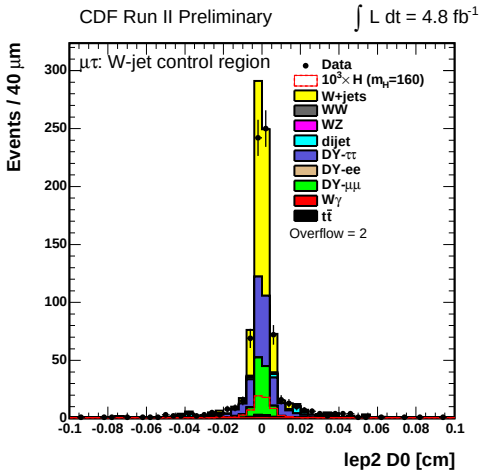
(c) Pseudorapidity of the leading lepton.



(d) Pseudorapidity of the subleading lepton.

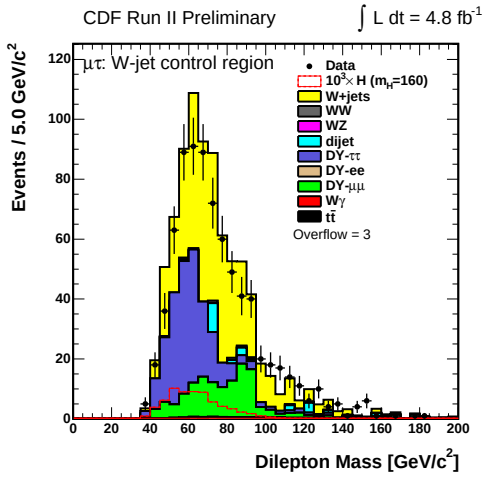


(e) Impact parameter of the leading lepton.

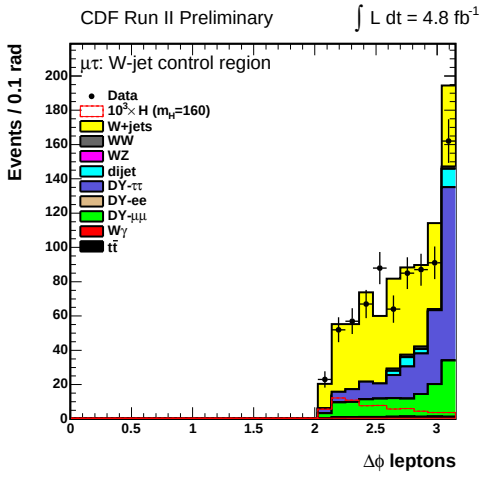


(f) Impact parameter of the subleading lepton.

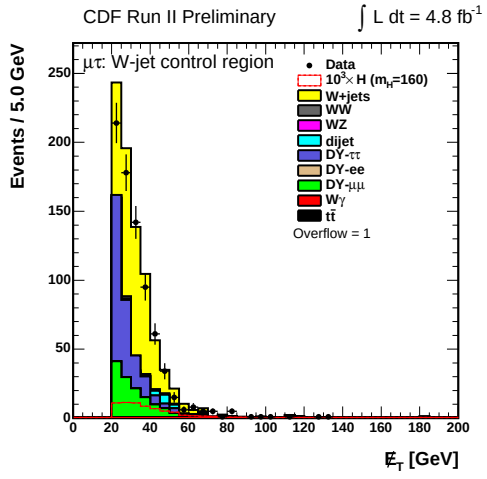
Figure B.7: Data-MC comparison in the $\mu\tau$ W+jet dominated region.



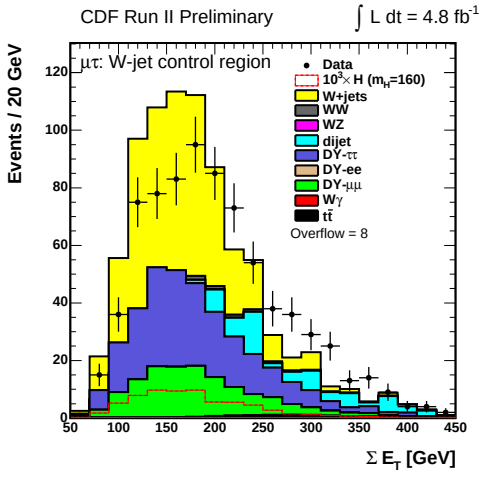
(a) Muon-tau invariant mass.



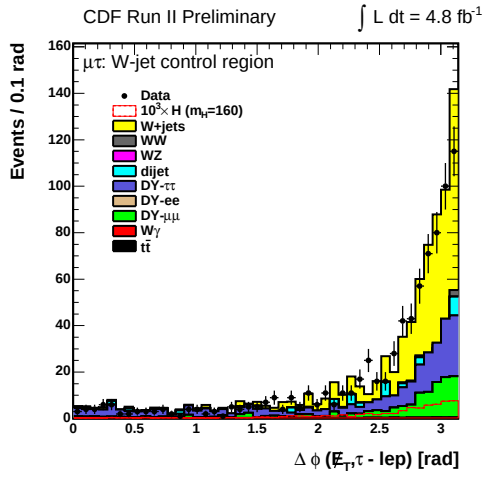
(b) $\Delta\varphi$ between the tau and the muon.



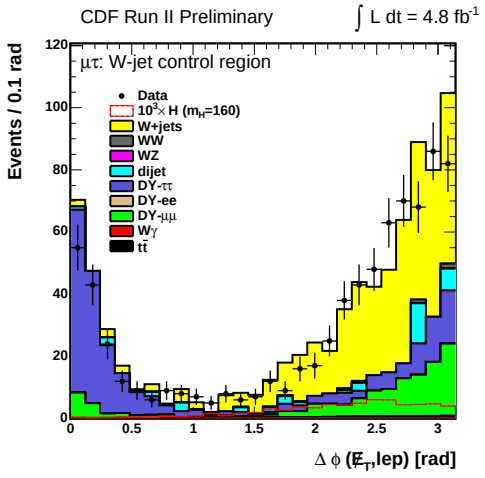
(c) Missing transverse energy.



(d) Sum of the E_T of the reconstructed objects in the event

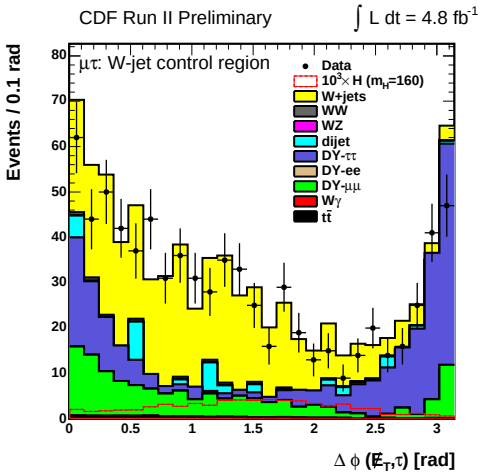


(e) $\Delta\varphi$ between the dilepton momentum and the $\not\!\!E_T$.

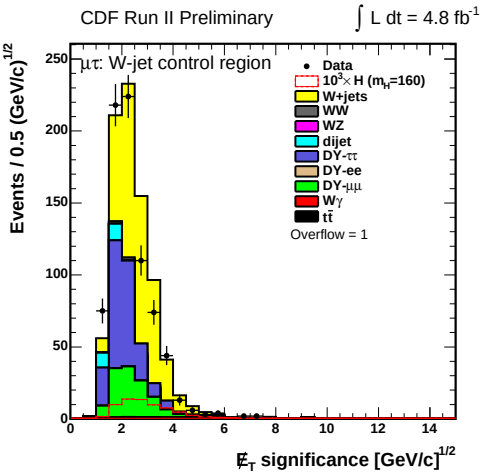


(f) $\Delta\varphi$ between the muon and the missing E_T.

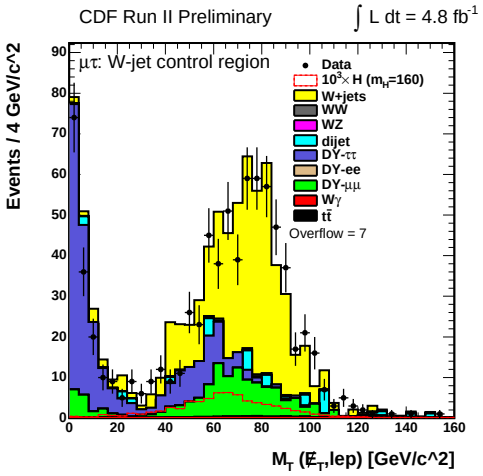
Figure B.8: Data-MC comparison in the $\mu\tau$ W+jet dominated region.



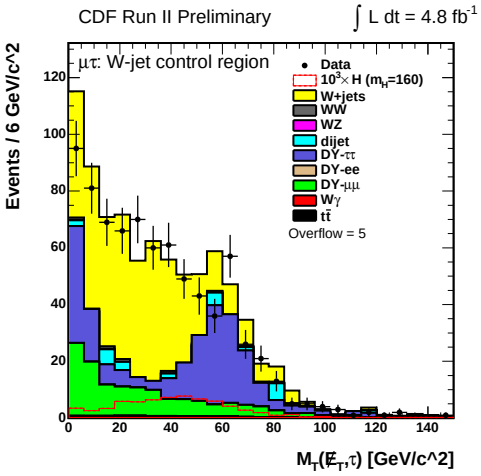
(a) $\Delta\varphi$ between the tau and the missing E_T.



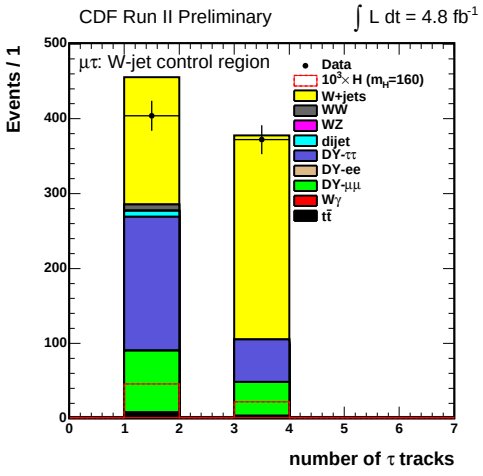
(b) Significance of the missing transverse energy.



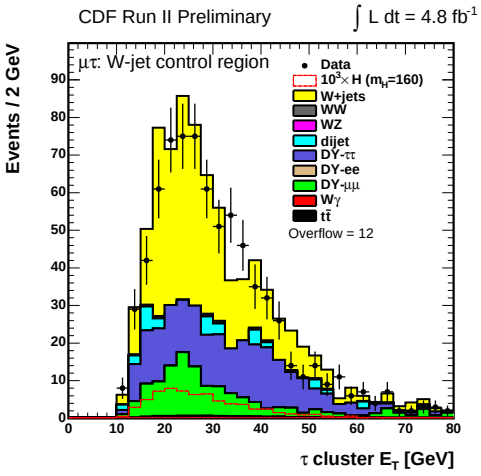
(c) τ-$\not\!\!E_T$ transverse mass.



(d) μ-$\not\!\!E_T$ transverse mass.

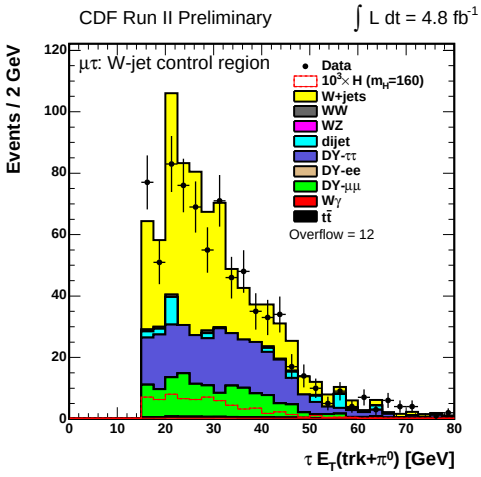


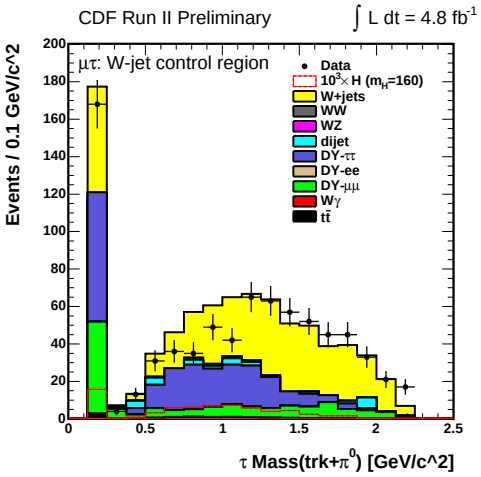
(e) Number of tracks of the reconstructed τ.



(f) Transverse energy of the τ calorimetric cluster.

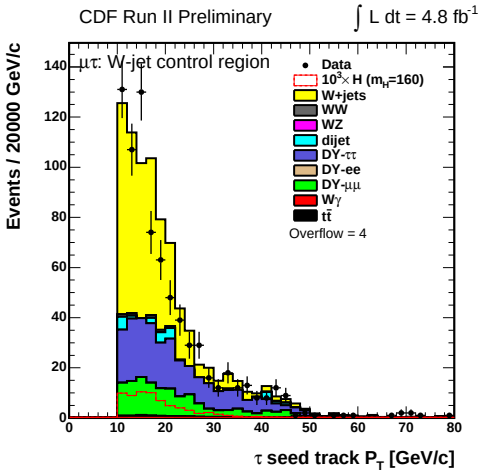
Figure B.9: Data-MC comparison in the $\mu\tau$ W+jet dominated region.

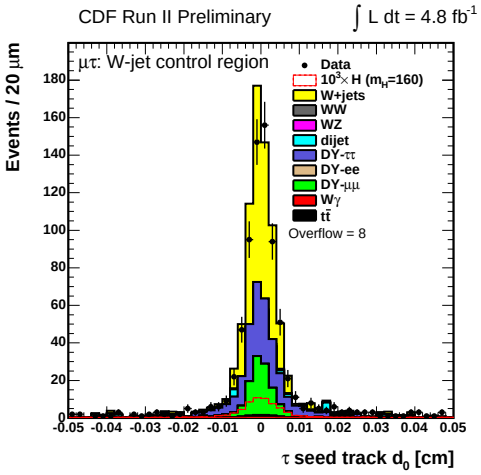




(a) Transverse energy of tracks and π^0's in the τ signal cone.

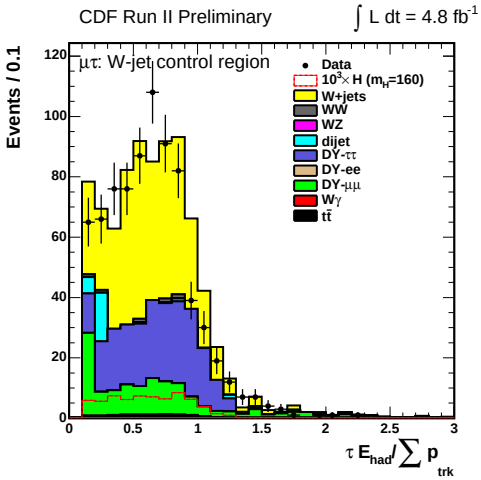
(b) Mass of tracks and π^0's in the τ signal cone.

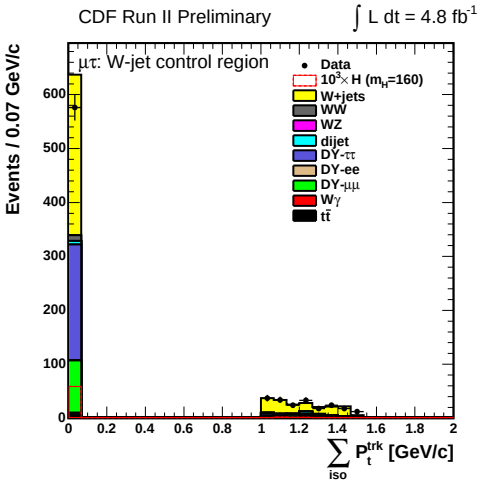




(c) Transverse momentum of the τ seed track.

(d) Impact parameter of the τ seed track.

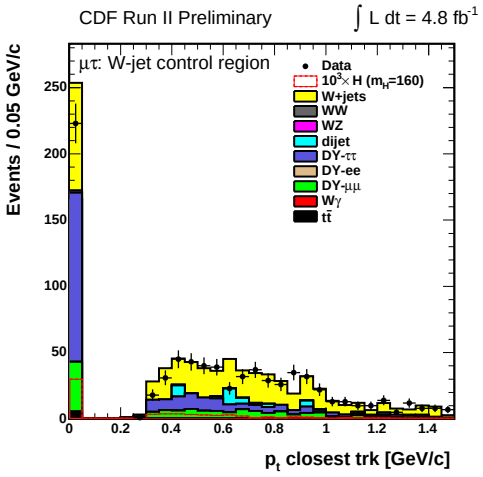




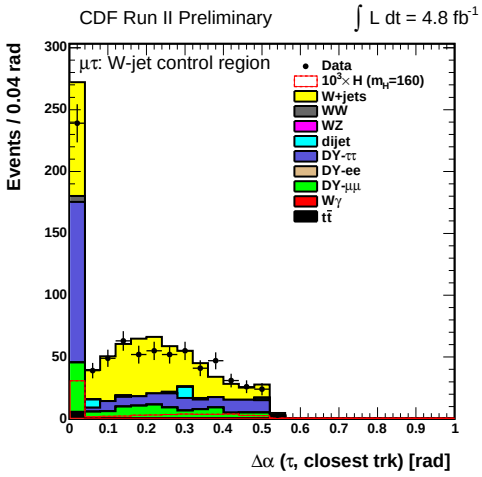
(e) Ratio of the τ cluster energy to the total momentum of the tracks associated to the τ.

(f) τ track isolation.

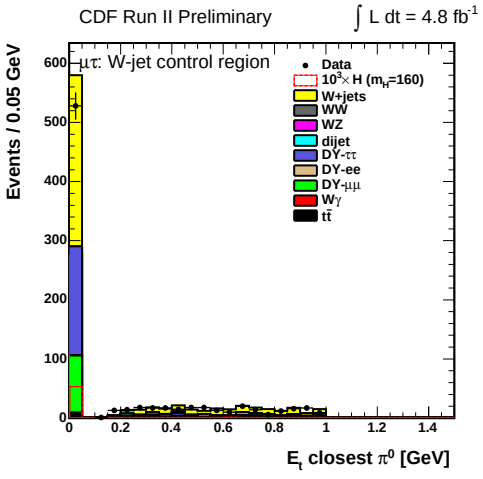
Figure B.10: Data-MC comparison in the $\mu\tau$ W+jet dominated region.



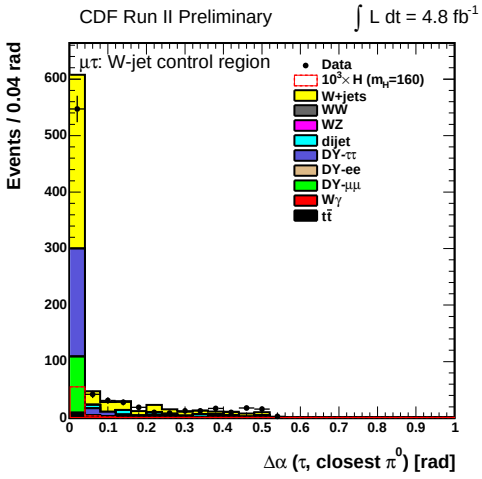
(a) p_T of the closest track to the τ.



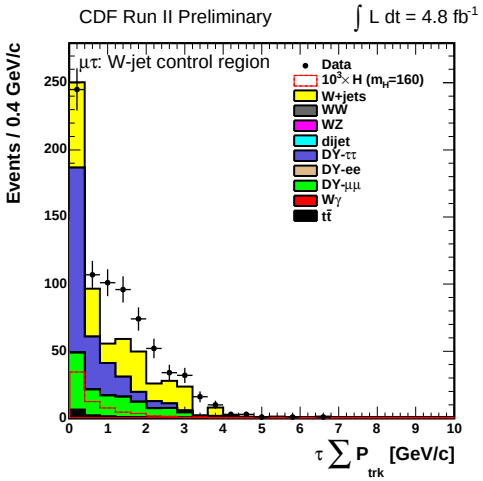
(b) Angle between the τ and the closest track.



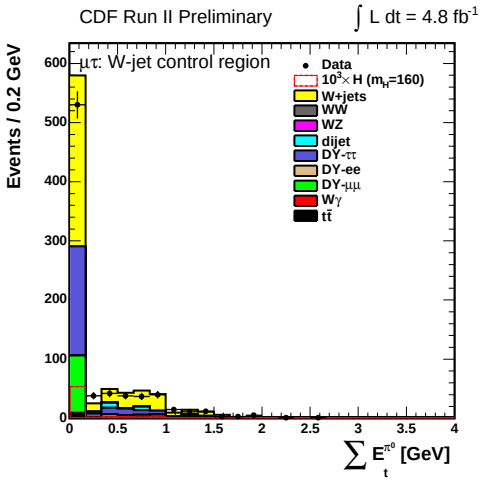
(c) E_T of the closest π^0 to the τ.



(d) Angle between the τ and the closest π^0.



(e) Sum of p_T of all tracks in the τ isolation cone.

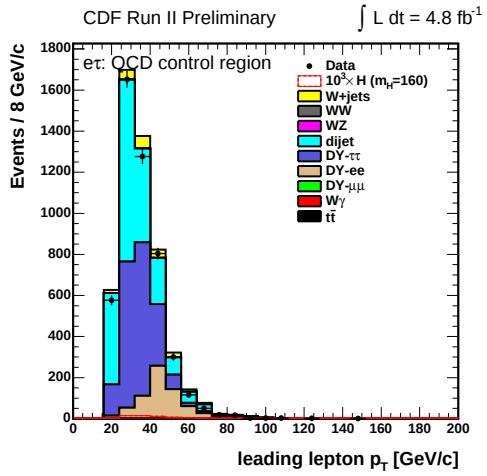


(f) Sum of E_T of all π^0's in the τ isolation cone.

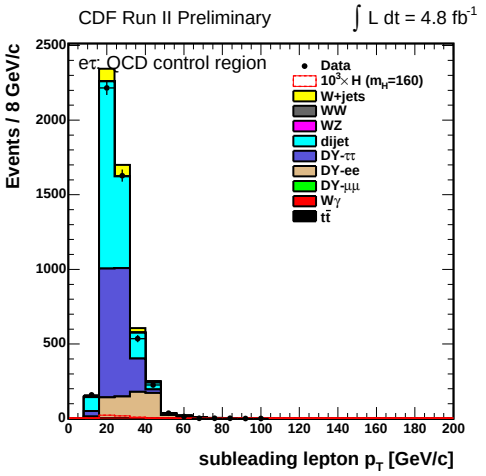
Figure B.11: Data-MC comparison in the $\mu\tau$ W+jet dominated region.

Appendix C

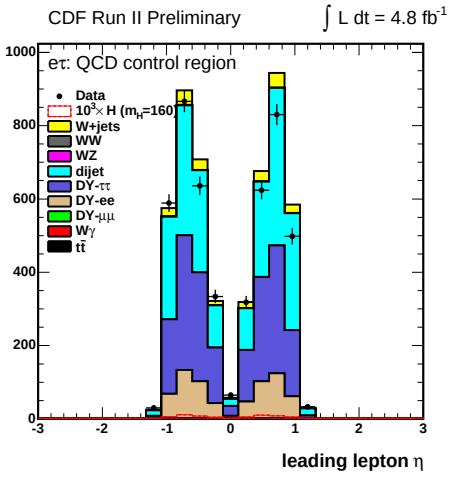
QCD control region: $e\tau$ channel



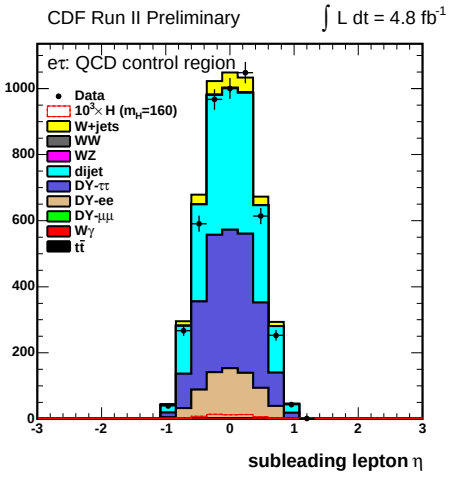
(a) p_T of the leading lepton.



(b) p_T of the subleading lepton.

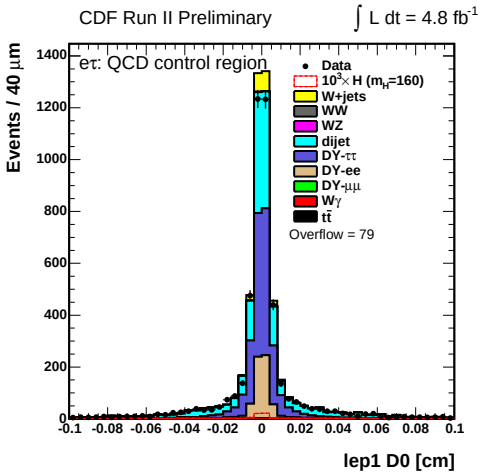


(c) Pseudorapidity of the leading lepton.

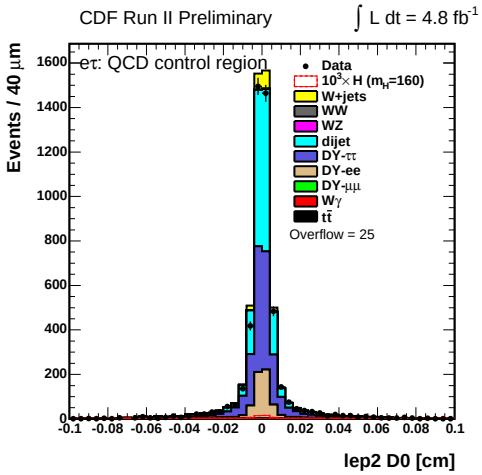


(d) Pseudorapidity of the subleading lepton.

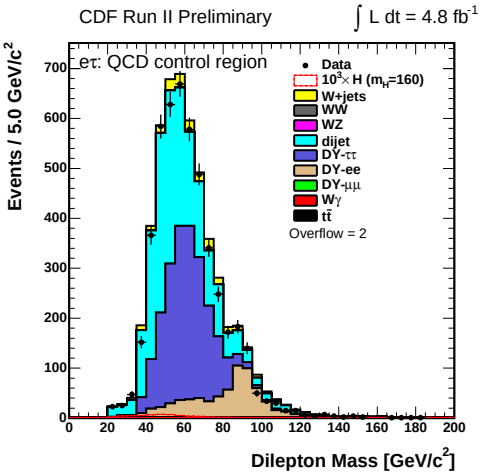
Figure C.1: Data-MC comparison in the $e\tau$ QCD enhanced region.



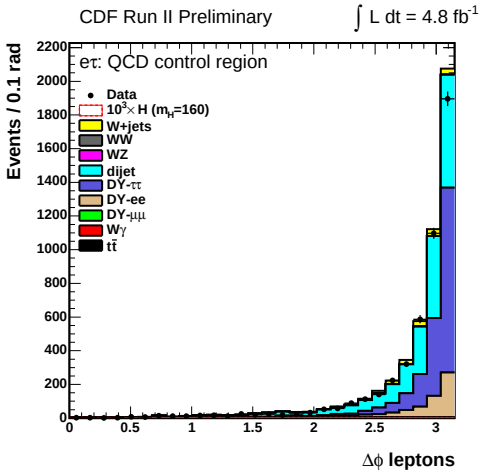
(a) Impact parameter of the leading lepton.



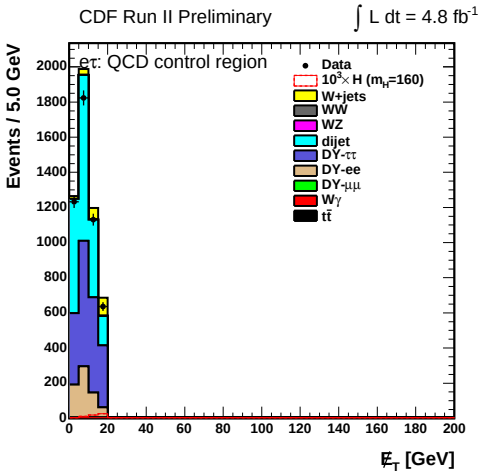
(b) Impact parameter of the subleading lepton.



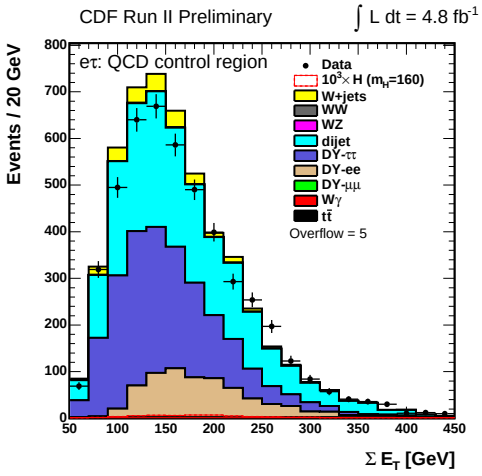
(c) Electron-tau invariant mass.



(d) $\Delta\varphi$ between the tau and the electron.

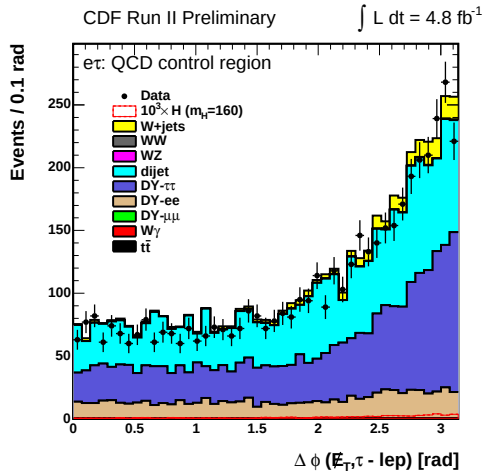


(e) Missing transverse energy.

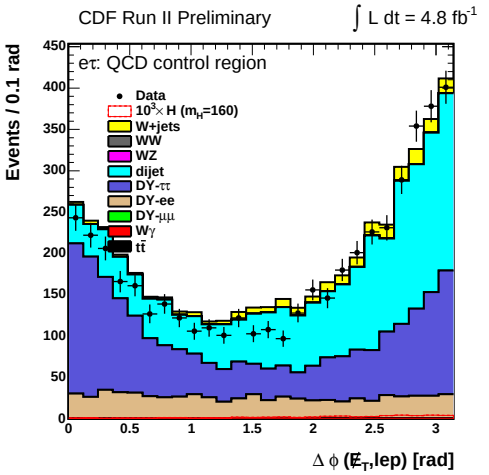


(f) Sum of the E_T of the reconstructed objects in the event

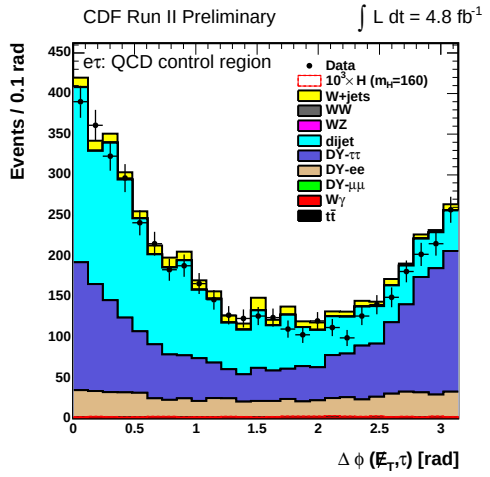
Figure C.2: Data-MC comparison in the $e\tau$ QCD enhanced region.



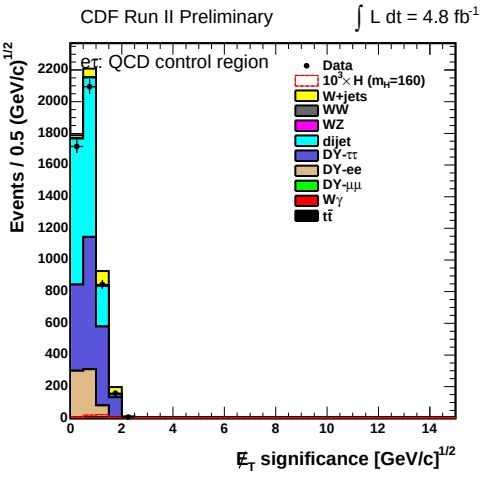
(a) $\Delta\varphi$ between the dilepton momentum and the $\not\!\!E_T$.



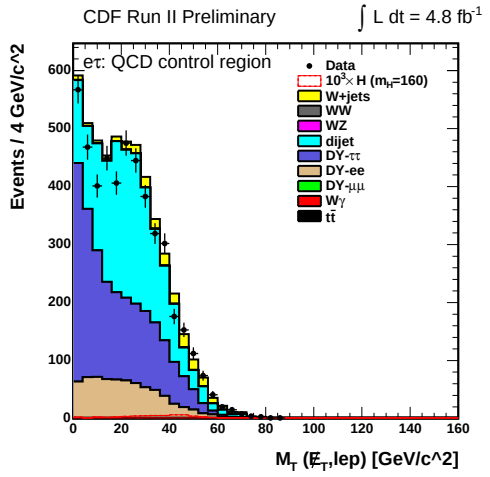
(b) $\Delta\varphi$ between the electron and the missing E_T.



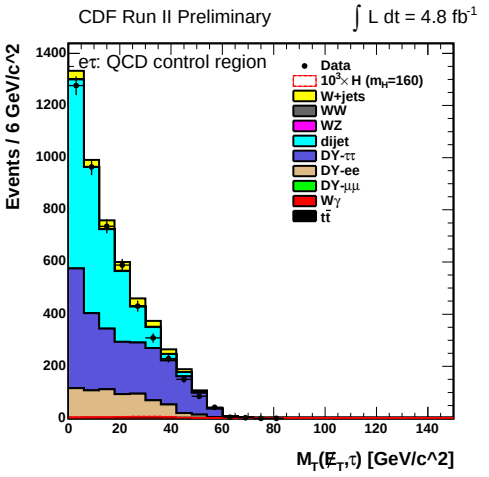
(c) $\Delta\varphi$ between the tau and the missing E_T.



(d) Significance of the missing transverse energy.

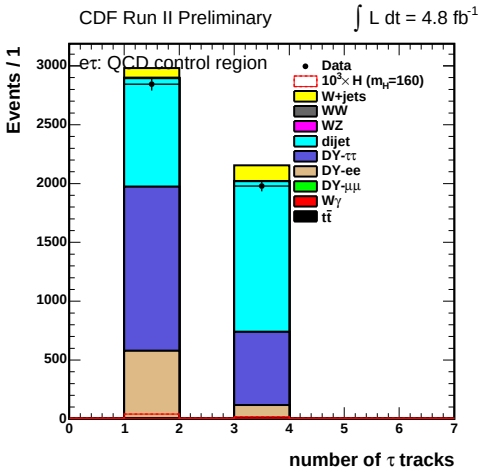


(e) τ-$\not\!\!E_T$ transverse mass.

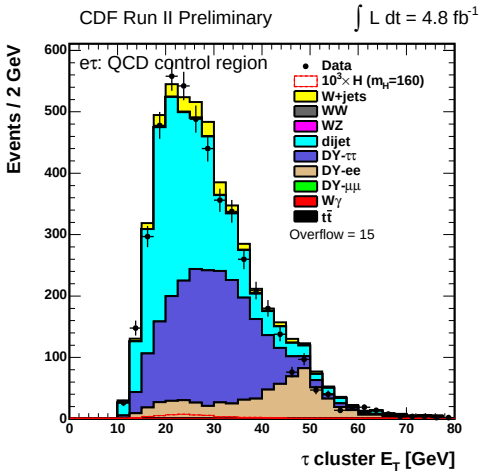


(f) e-$\not\!\!E_T$ transverse mass.

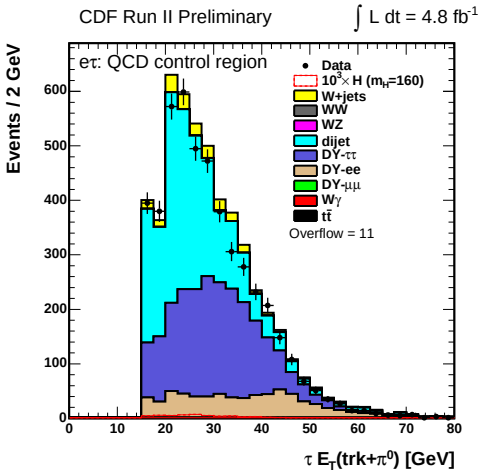
Figure C.3: Data-MC comparison in the $e\tau$ QCD enhanced region.



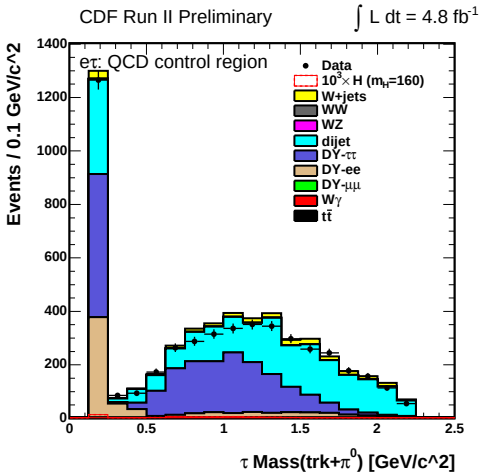
(a) Number of tracks of the reconstructed τ.



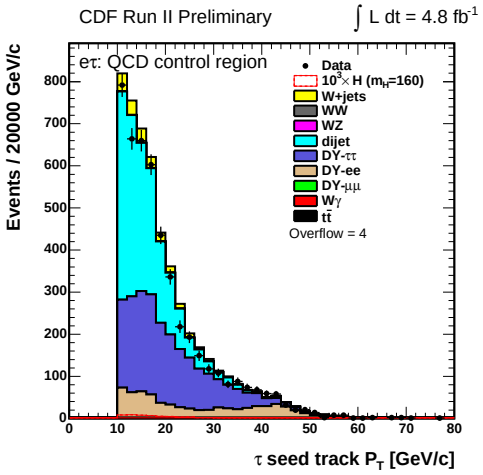
(b) Transverse energy of the τ calorimetric cluster.



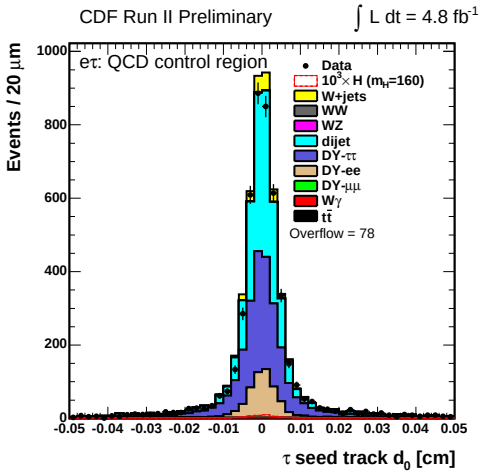
(c) Transverse energy of tracks and π^0's in the τ signal cone.



(d) Mass of tracks and π^0's in the τ signal cone.

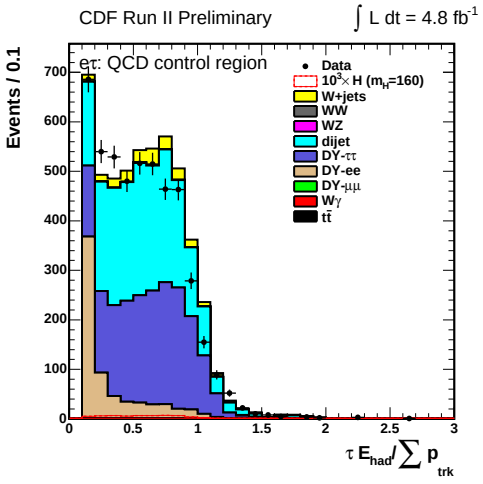


(e) Transverse momentum of the τ seed track.

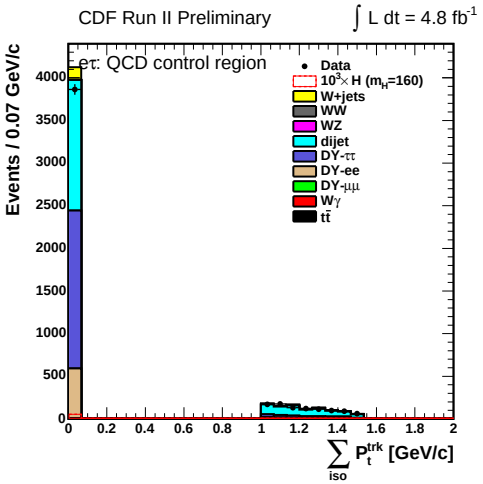


(f) Impact parameter of the τ seed track.

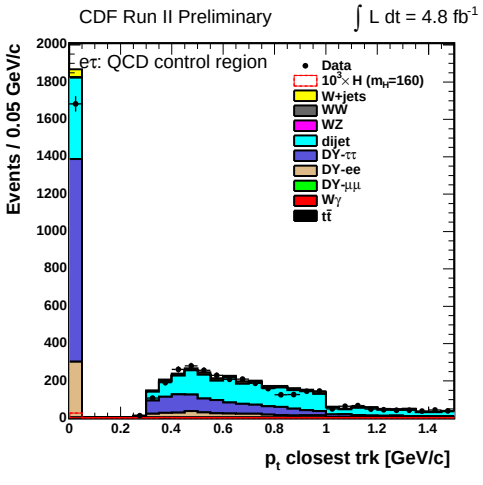
Figure C.4: Data-MC comparison in the $e\tau$ QCD enhanced region.



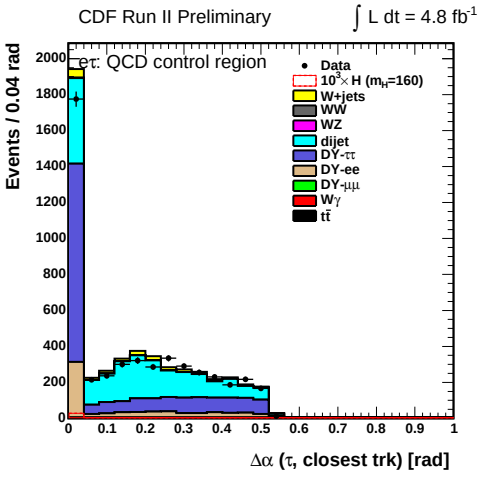
(a) Ratio of the τ cluster energy to the total momentum of the tracks associated to the τ.



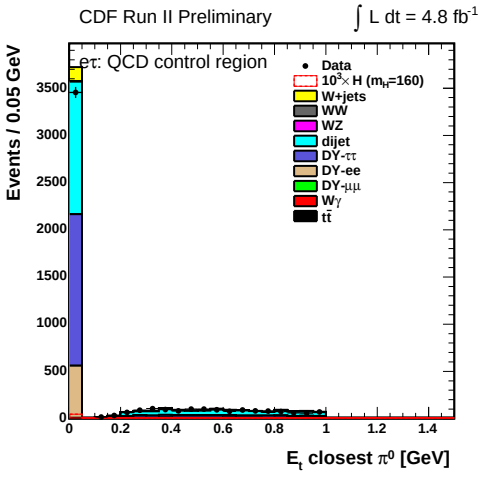
(b) τ track isolation.



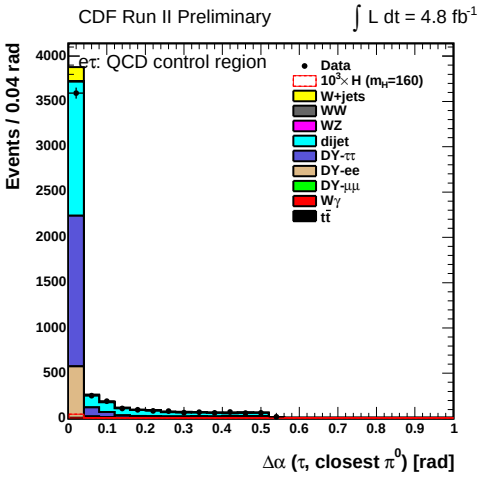
(c) p_T of the closest track to the τ.



(d) Angle between the τ and the closest track.

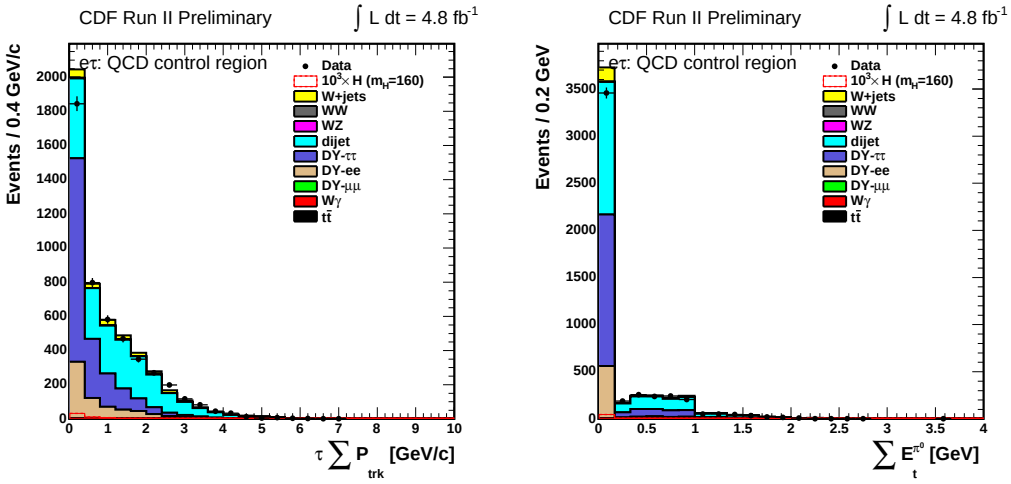


(e) E_T of the closest π^0 to the τ.



(f) Angle between the τ and the closest π^0.

Figure C.5: Data-MC comparison in the $e\tau$ QCD enhanced region.

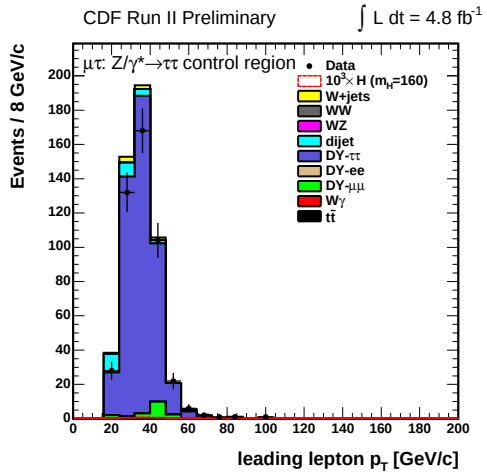


(a) Sum of p_T of all tracks in the τ isolation cone. (b) Sum of E_T of all π^0's in the τ isolation cone.

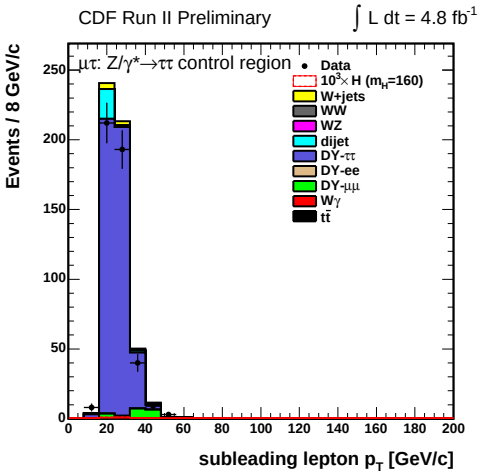
Figure C.6: Data-MC comparison in the $e\tau$ QCD enhanced region.

Appendix D

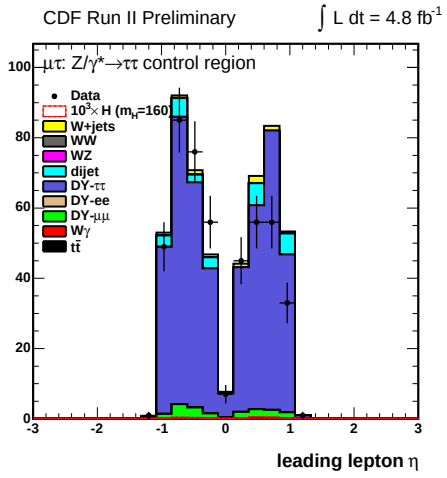
$Z \to \tau\tau$ control region: $\mu\tau$ channel



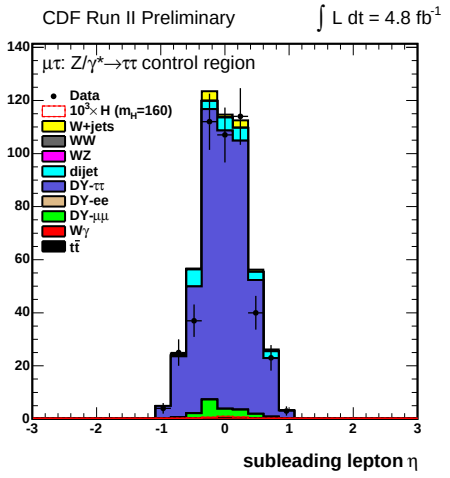
(a) p_T of the leading lepton.



(b) p_T of the subleading lepton.

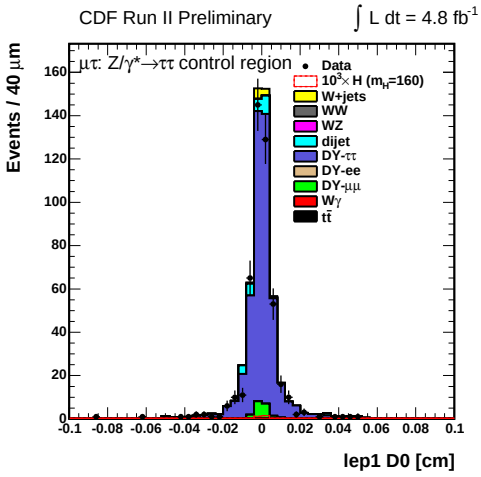


(c) Pseudorapidity of the leading lepton.

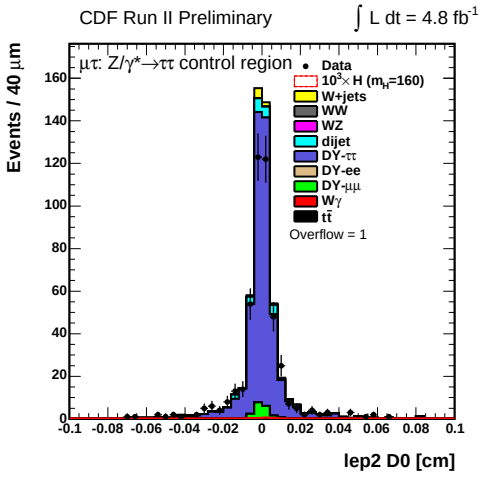


(d) Pseudorapidity of the subleading lepton.

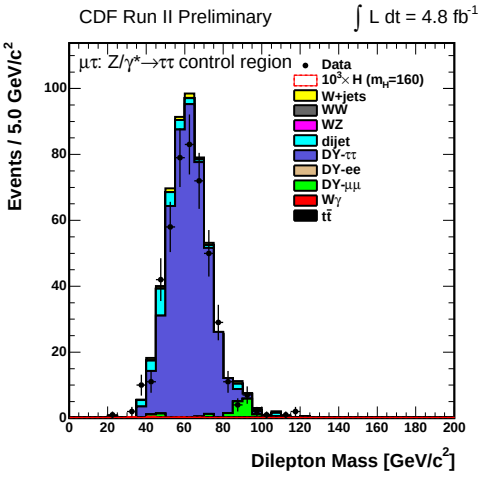
Figure D.1: Data-MC comparison in the $\mu\tau$ $Z \to \tau\tau$ region.



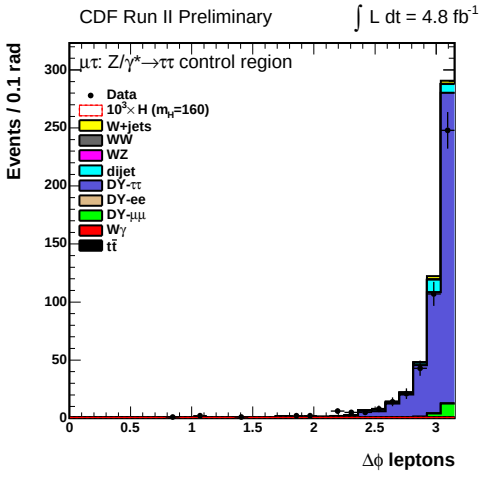
(a) Impact parameter of the leading lepton.



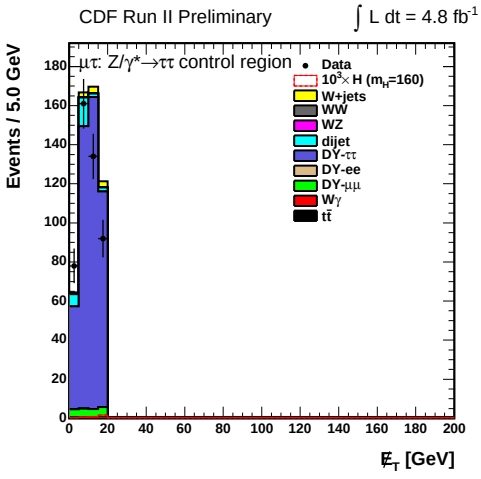
(b) Impact parameter of the subleading lepton.



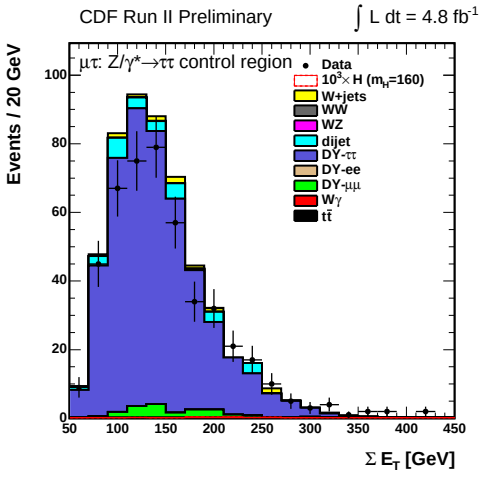
(c) Muon-tau invariant mass.



(d) $\Delta\varphi$ between the tau and the muon.

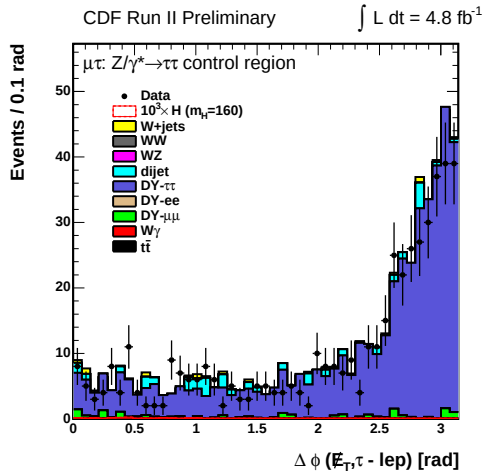


(e) Missing transverse energy.

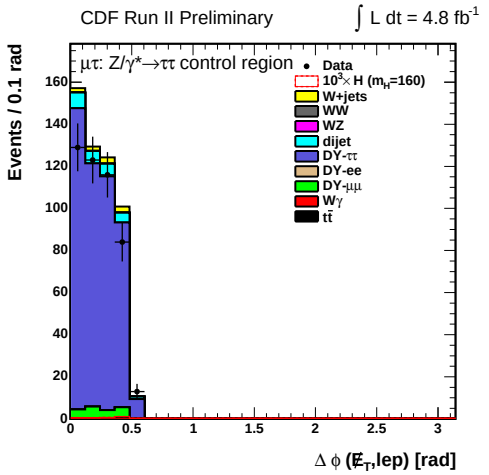


(f) Sum of the E_T of the reconstructed objects in the event

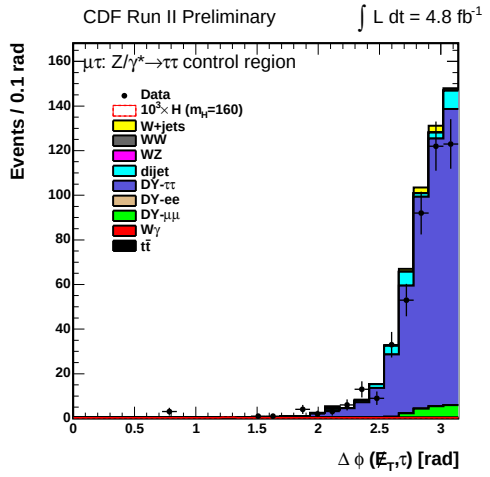
Figure D.2: Data-MC comparison in the $\mu\tau$ $Z \to \tau\tau$ region.



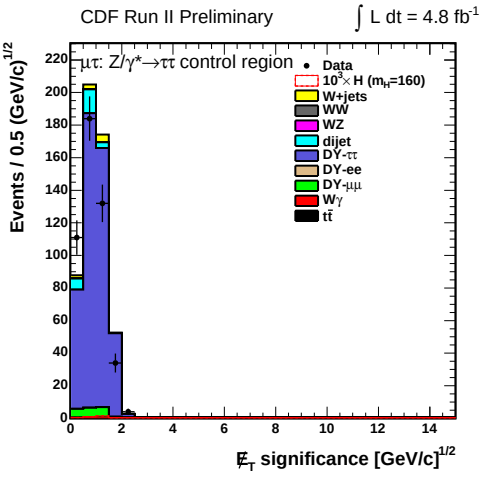
(a) $\Delta\varphi$ between the dilepton momentum and the $\not\!\!E_T$.



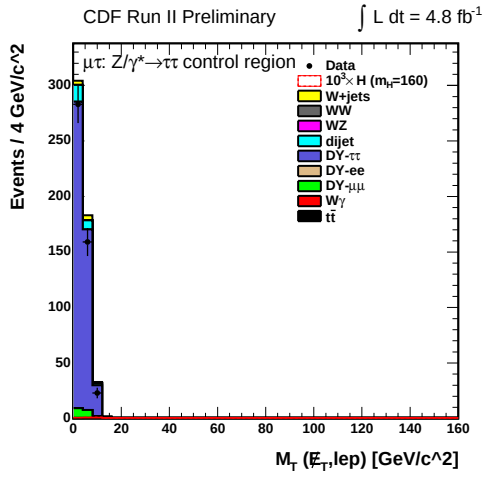
(b) $\Delta\varphi$ between the muon and the missing E_T.



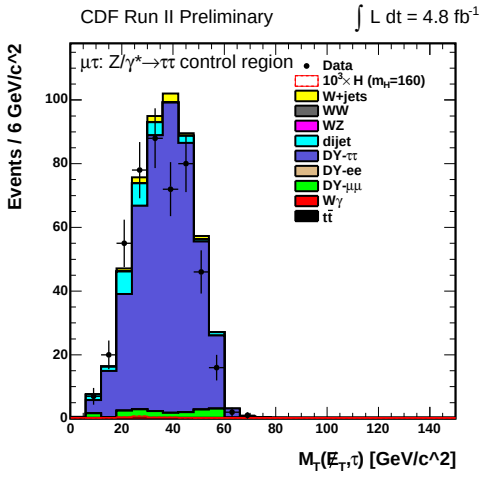
(c) $\Delta\varphi$ between the tau and the missing E_T.



(d) Significance of the missing transverse energy.

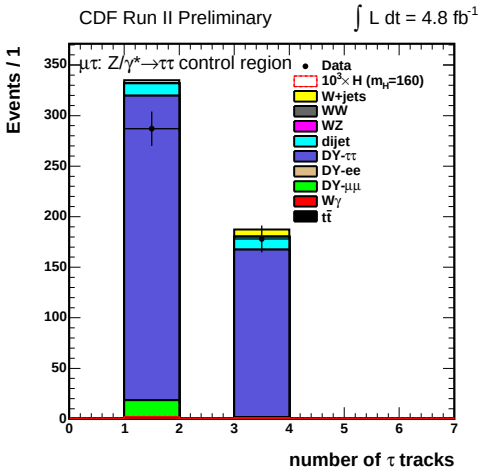


(e) τ-$\not\!\!E_T$ transverse mass.

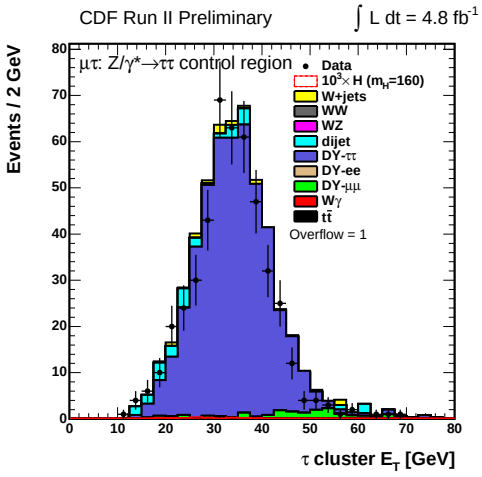


(f) μ-$\not\!\!E_T$ transverse mass.

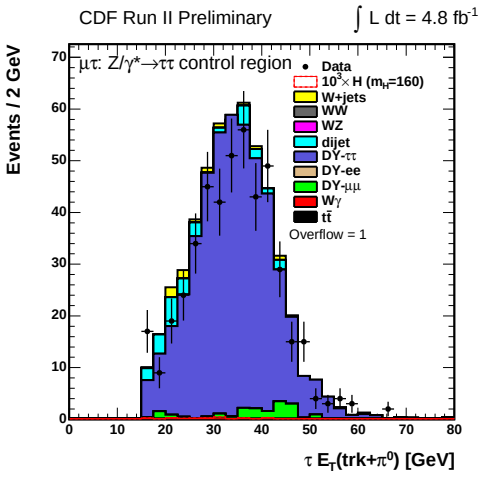
Figure D.3: Data-MC comparison in the $\mu\tau$ $Z \to \tau\tau$ region.



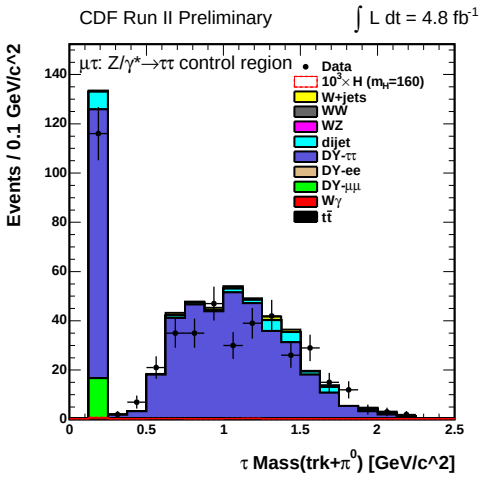
(a) Number of tracks of the reconstructed τ.



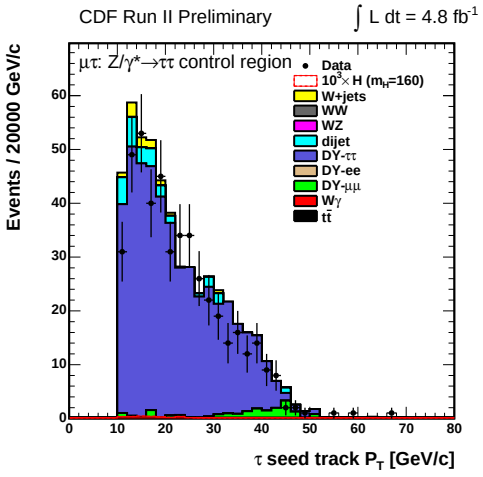
(b) Transverse energy of the τ calorimetric cluster.



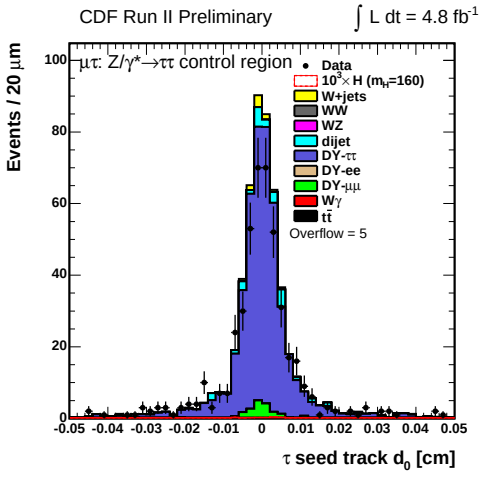
(c) Transverse energy of tracks and π^0's in the τ signal cone.



(d) Mass of tracks and π^0's in the τ signal cone.

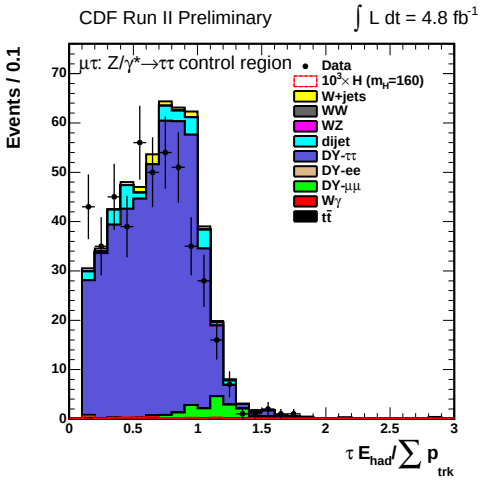


(e) Transverse momentum of the τ seed track.

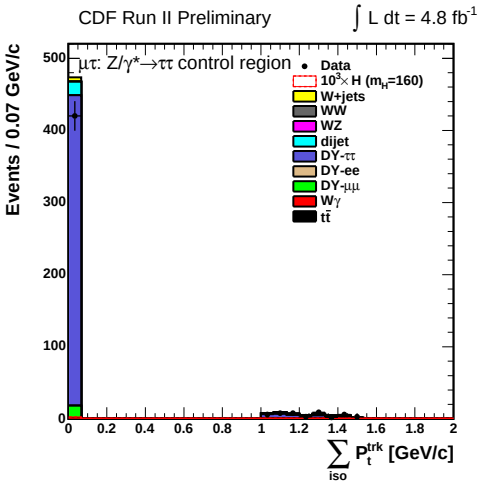


(f) Impact parameter of the τ seed track.

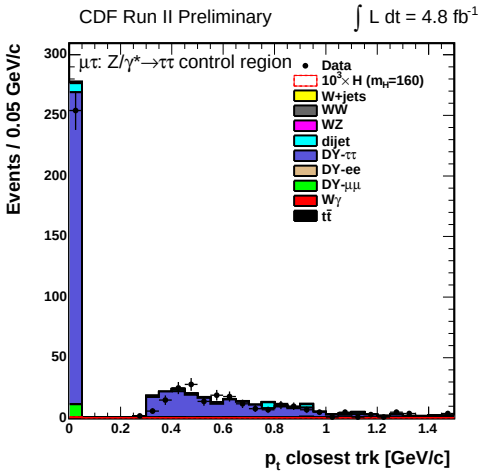
Figure D.4: Data-MC comparison in the $\mu\tau$ $Z \to \tau\tau$ region.



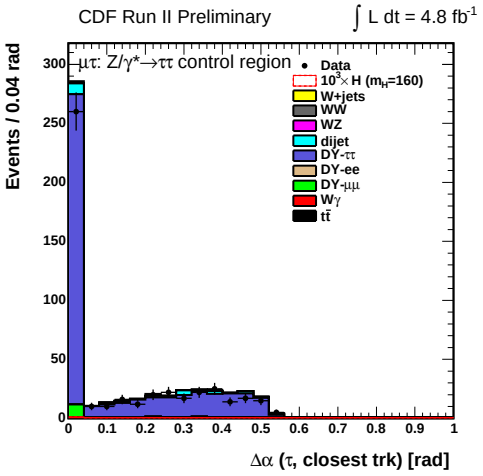
(a) Ratio of the τ cluster energy to the total momentum of the tracks associated to the τ.



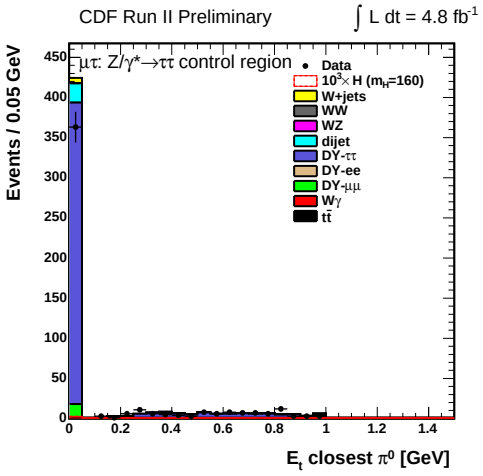
(b) τ track isolation.



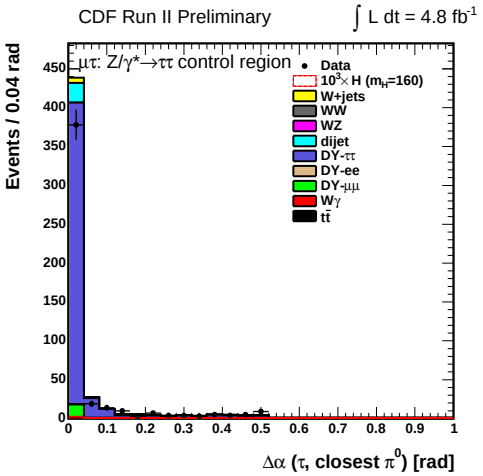
(c) p_T of the closest track to the τ.



(d) Angle between the τ and the closest track.

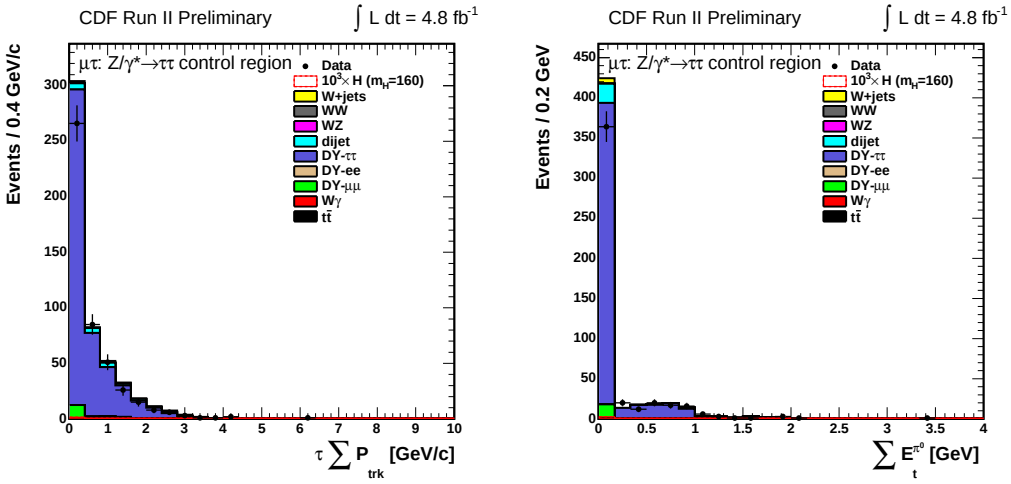


(e) E_T of the closest π^0 to the τ.



(f) Angle between the τ and the closest π^0.

Figure D.5: Data-MC comparison in the $\mu\tau$ $Z \to \tau\tau$ region.



(a) Sum of p_T of all tracks in the τ isolation cone. (b) Sum of E_T of all π^0's in the τ isolation cone.

Figure D.6: Data-MC comparison in the $\mu\tau$ $Z \to \tau\tau$ region.

List of Figures

List of Tables

Bibliography

[1] (LEP Collaborations: ALEPH, DELPHI, L3, OPAL, LEP Electroweak Working Group, SLD Electroweak, Heavy Flavour Groups), "A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model," Dec 2005, CERN-PH-EP/2004-069. arXiv:hep-ex/0412015v2.

[2] LEP Electroweak Working Group webpage.
URL: http:/lepewwg.web.cern.ch/LEPEWWG.

[3] K. Nakamura *et al.*, (Particle Data Group), "The review of particle physics," *J. Phys. G: Nucl. Part. Phys.*, vol. 37, p. 075021, 2010. doi:10.1088/0954-3899/37/7A/075021.

[4] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries," *Phys. Rev.*, vol. 127, pp. 965–970, Aug 1962, 3, American Physical Society. doi:10.1103/PhysRev.127.965.

[5] P. W. Higgs, "Broken symmetries, massless particles and gauge fields," *Phys. Lett.*, vol. 12, pp. 132–133, Sept 1964, 2, Elsevier B.V. doi:10.1016/0031-9163(64)91136-9.

[6] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964, 16, American Physical Society. doi:10.1103/PhysRevLett.13.508.

[7] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory.* Addison-Wesley Advanced Book Program, 1995.

[8] A. S. Goldhaber and M. M. Nieto, "Photon and Graviton Mass Limits," *Rev. Mod. Phys.*, vol. 82, pp. 939–979, 2010. arXiv:0809.1003v5.

[9] A. Salam, "Weak and Electromagnetic Interactions," pp. 367–377, 1968. Proceedings 8th Nobel Symposium (Stockholm), Ed. N. Svartholm 1968.

[10] S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967, 21, American Physical Society. doi:10.1103/PhysRevLett.19.1264.

[11] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear Physics*, vol. 22, pp. 579–588, Feb 1961, 4, Elsevier B.V. doi:10.1016/0029-5582(61)90469-2.

[12] H. Georgi and S. L. Glashow, "Unity of All Elementary-Particle Forces," *Phys. Rev. Lett.*, vol. 32, pp. 438–441, Feb 1974, 8, American Physical Society. doi:10.1103/PhysRevLett.32.438.

[13] G. Arnison *et al.*, (UA1 Collaboration), "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider," *Phys. Lett. B*, vol. 126, pp. 398–410, 1983.

[14] P. Bagnaia *et al.*, (UA2 Collaboration), "Evidence for $Z^0 \to e^+e^-$ at the CERN pp collider," *Phys. Lett. B*, vol. 129, pp. 130–140, 1983.

[15] G. Arnison *et al.*, (UA1 Collaboration), "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV," *Phys. Lett. B*, vol. 122, pp. 103–116, 1983.

[16] P. Banner *et al.*, (UA2 Collaboration), "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $p\bar{p}$ collider," *Phys. Lett. B*, vol. 122, p. 476, 1983.

[17] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964, 9, American Physical Society. doi:10.1103/PhysRevLett.13.321.

[18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles," *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964, 20, American Physical Society. doi:10.1103/PhysRevLett.13.585.

[19] T. W. B. Kibble, "Symmetry Breaking in Non-Abelian Gauge Theories," *Phys. Rev.*, vol. 155, pp. 1554–1561, Mar 1967, 5, American Physical Society. doi:10.1103/PhysRev.155.1554.

[20] A. Djouadi, "The Anatomy of Electro-Weak Symmetry Breaking. I: The Higgs boson in the Standard Model," *Phys.Rept.*, vol. 457, pp. 1–216, 2005. arXiv:hep-ex/0503172v2, doi:10.1016/j.physrep.2007.10.004.

[21] G. Altarelli and G. Isidori, "Lower limit on the Higgs mass in the standard model: An Update," *Phys. Lett. B*, vol. 337, pp. 141–144, 1994.

[22] B. W. Lee, C. Quigg, and H. B. Thacker, "Strength of Weak Interactions at Very High Energies and the Higgs Boson Mass," *Phys. Rev. Lett.*, vol. 38, pp. 883–885, Apr 1977, 16, American Physical Society. doi:10.1103/PhysRevLett.38.883.

[23] T. Hambye and K. Riesselmann, "Matching conditions and Higgs mass upper bounds revisited," *Phys. Rev. D*, vol. 55, pp. 7255–7262, 1997. arXiv:hep-ph/9610272.

[24] B. W. Lee, C. Quigg, and H. B. Thacker, "Weak interactions at very high energies: The role of the Higgs-boson mass," *Phys. Rev. D*, vol. 16, pp. 1519–1531, Sep 1977, 5, American Physical Society. doi:10.1103/PhysRevD.16.1519.

[25] G. Abbiendi *et al.*, (ALEPH, DELPHI, L3 and OPAL Collaborations, The LEP Working Group for Higgs Boson Searches), "Search for the Standard Model Higgs Boson at LEP," *Phys. Lett. B*, vol. 565, pp. 61–75, 2003. arXiv:hep-ex/0306033v1, doi:10.1016/S0370-2693(03)00614-2.

[26] M. T. P. Roco and P. D. Grannis, "Higgs boson discovery prospects at the tevatron," Jan 1999. Proceeding 13th Topical Conference on Hadron Collider Physics (HCP99), Mumbai, India.

[27] H. Flaecher, M. Goebel, J. Haller, A. Hoecker, K. Moenig, and J. Stelzer, (CERN, DESY and U. Hamburg), "Revisiting the Global Electroweak Fit of the Standard Model and Beyond with GFITTER," *Eur. Phys. J. C*, vol. 60, pp. 543–583, 2009. arXiv:0811.0009v3, doi:10.1140/epjc/s10052-009-0966-6.

[28] "GFITTER, A Generic Fitter Project for HEP Model Testing webpge." URL: http://gfitter.desy.de .

[29] (LEP Collaboration, LEP Electroweak Working Group, Tevatron Electroweak Working Group and SLD Electroweak and Heavy Flavour Groups), "Precision Electroweak Measurements and Constraints on the Standard Model," Nov 2009, CERN-PH-EP/2009-023. arXiv:0911.2604v2.

[30] A. B. Arbuzov *et al.*, "ZFITTER: A semi-analytical program for fermion pair production in e^+e^- annihilation, from version 6.21 to version 6.42," *Comput. Phys. Commun.*, vol. 174, pp. 728–758, 2006. arXiv:hep-ph/0507146.

[31] V. Ravindran, J. Smith, and W. L. van Neerven, "NNLO corrections to the total cross section for Higgs boson production in hadron-hadron collisions," *Nucl. Phys. B*, vol. 665, pp. 325–366, 2003. arXiv:hep-ph/0302135v2, doi:10.1016/S0550-3213(03)00457-7.

[32] J. M. Campbell, J. W. Huston, and W. J. Stirling, "Hard Interactions of Quarks and Gluons: a Primer for LHC Physics," *Rep. Prog. Phys.*, vol. 70, pp. 89–193, 2007. Online at: stacks.iop.org/RoPP/70/89. arXiv:hep-ph/0611148.

[33] H. M. Georgi, S. L. Glashow, M. E. Machacek, and D. V. Nanopoulos, "Higgs Bosons from Two-Gluon Annihilation in Proton-Proton Collisions," *Phys. Rev. Lett.*, vol. 40, pp. 692–694, Mar 1978, 11, American Physical Society. doi:10.1103/PhysRevLett.40.692.

[34] A. Djouadi, M. Spira, and P. M. Zerwas, "Production of Higgs bosons in proton colliders. QCD corrections," *Phys. Lett. B*, vol. 264, pp. 440–446, Aug 1991, Elsevier Science B.V. doi:10.1016/0370-2693(91)90375-Z.

[35] D. de Florian and M. Grazzini, "Higgs production through gluon fusion: updated cross sections at the Tevatron and the LHC," *Phys. Lett. B*, vol. 674, pp. 291–294, 2009. arXiv:0901.2427v2, doi:10.1016/j.physletb.2009.03.033.

[36] A. Martin, W. Stirling, R. Thorne, and G. Watt, "Heavy-quark mass dependence in global PDF analyses and 3- and 4-flavour parton distributions," *Eur. Phys. J. C*, vol. 70, pp. 51–72, 2010. hhtp://projects.hepforge.org/mstwpdf/. arXiv:1007.2624v2, doi:10.1140/epjc/s10052-010-1462-8.

[37] C. Anastasiou, R. Boughezal, and F. Petriello, "Mixed QCD-electroweak corrections to Higgs boson production in gluon fusion," *JHEP*, vol. 04, p. 003, 2009. arXiv:0811.3458v2, doi:10.1088/1126-6708/2009/04/003.

[38] R. V. Harlander, H. Mantler, S. Marzani, and K. J. Ozeren, "Higgs production in gluon fusion at next-to-next-to-leading order QCD for finite top mass," *Eur. Phys. J. C*, vol. 66, pp. 359–372, Dec 2009. arXiv:0912.2104v1, doi:10.1140/epjc/s10052-010-1258-x.

[39] L. Reina and S. Dawson, "Next-to-Leading Order Results for $t\bar{t}$ Production at the Tevatron," *Phys. Rev. Lett.*, vol. 87, p. 201804, Oct 2001, 20, American Physical Society. doi:10.1103/PhysRevLett.87.201804.

[40] Beenakker, W. and Dittmaier, S. and Kramer, M. and Plumper, B. and Spira, M. and Zerwas, P.M., "NLO QCD corrections to t anti-t H production in hadron collisions," *Nucl. Phys. B*, vol. 653, pp. 151–203, 2003. doi:10.1016/S0550-3213(03)00044-0.

[41] A. Djouadi, J. Kalinowski, and M. Spira, "HDECAY: a Program for Higgs Boson Decays in the Standard Model and its Supersymmetric Extension," *Comp. Phys. Commun.*, vol. 108, p. 56, 1998.

[42] T. Aaltonen *et al.*, "Search for a Higgs Boson Produced in Association with a W Boson in pbar-p Collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev. Lett*, vol. 103, p. 101802, Jun 2009. arXiv:0906.5613v1, doi:10.1103/PhysRevLett.103.101802.

[43] V. M. Abazov *et al.*, "Search for Associated W and Higgs Boson Production in pp Collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev. Lett*, vol. 102, p. 051803, 2009. arXiv:0808.1970.

[44] V. M. Abazov *et al.*, (D0 Collaboration), "Search for the Standard Model Higgs Boson in the $ZH \to \nu\nu b\bar{b}$ Channel in 5.2 fb^{-1} of $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev. Lett.*, vol. 104, p. 071801, Feb 2010, 7, American Physical Society. doi:10.1103/PhysRevLett.104.071801.

[45] T. Aaltonen *et al.*, "A search for the higgs boson using neural network in events with missing energy and b-quark jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," *Phys.Rev.Lett*, vol. 104, p. 141801, 2010. arXiv:0911.3935v4, doi:10.1103/PhysRevLett.104.141801.

[46] T. Aaltonen *et al.*, (CDF Collaboration), "Search for the higgs boson produced in association with $Z \to l^+l^-$ using the matrix element method at CDF II," *Phys. Rev. D*, vol. 80, p. 071101, Oct 2009, 7, American Physical Society. doi:10.1103/PhysRevD.80.071101.

[47] V. M. Abazov *et al.*, "Search for a Higgs boson produced in association with a Z boson in $p\bar{p}$ collisions," *Phys. Lett. B*, vol. 655, pp. 209–216, 2007. arXiv:0704.2000v1, doi:10.1016/j.physletb.2007.08.070.

[48] (CDF Collaboration), "Search for a low mass Standard Model Higgs boson in the di-tau decay channel with 2.3 fb^{-1} of CDF data," 2010. CDF/PUB/EXOTIC/PUBLIC/10133CDF. Public webpage for a low mass Standard Model Higgs boson in the $\tau\tau$ decay channel: http://www-cdf.fnal.gov/-physics/new/hdg/Results_files/results/smtautau_jul10/home.html.

[49] V. M. Abazov *et al.*, (D0 Collaboration), "Search for Resonant Diphoton Production with the D0 Detector," *Phys. Rev. Lett.*, vol. 102, p. 231801, 2009. arXiv:0901.1887.

[50] K. Bland, R. Culbertson, J. Dittmann, C. Group, and J. Ray, (CDF Collaboration), "Search for a SM Higgs Boson with the Diphoton Final State at CDF," 2010. CDF/PUB/EXOTIC/PUBLIC/10065.

[51] (CMS Collaboration), "The CMS physics reach for searches at 7 TeV," May 2010, CMS-NOTE-2010/008.

[52] (ATLAS Collaboration), "ATLAS Sensitivity Prospects for Higgs Boson Production at the LHC Running at 7, 8 or 9 TeV," Nov 2010, ATL-PHYS-PUB-2010-015.

[53] V. Drollinger, T. Muller, and D. Denegri, (CMS Collaboration), "Prospects for Higgs Boson Searches in the Channel $WH \to l\nu bb$," 2002. CMS-NOTE-2002/006. arXiv:hep-ph/0201249v1.

[54] (D0 Collaboration), "Search for the SM Higgs Boson in gamma-gamma final states at D0 with $L = 4.2$ fb^{-1} data," Feb 2009, D0 Note 5858-Conf.

[55] (CDF Collaboration), "Search for $H \to WW$ Production Using 8.2 fb^1," Jul 2011, CDF public Note 105611.

[56] (D0 Collaboration), "Search for associated Higgs boson production with like sign leptons in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," 2010. D0 Conference Note 6091.

[57] (D0 Collaboration), "Search for Higgs boson production in dilepton plus missing energy final states with 8.1 fb^1 of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," 2011. D0 Conference Note 6182.

[58] (ATLAS Collaboration), "ATLAS Sensitivity Prospects for Higgs Boson Production at the LHC Running at 7 TeV," Jul 2010, ATL-PHYS-PUB-2010-009.

[59] (CMS Collaboration), "Search strategy for a Standard Model Higgs boson decaying to two W bosons in the fully leptonic final state," Jul 2009, CMS PAS HIG-08-006.

[60] M. Bauce, S. Behari, D. Benjamin, P. Bussey, A. Canepa, B. Carls, S. Carron, M. Casarsa, R. St.Denis, M. d'Errico, P. M. Fernandez, M. Herndon, E. James, S. Jindariani, T. Junk, M. Kruse, A. Limosani, D. Lucchesi, R. Lysak, J. Nett, S. Oh, A. Robson, and G. Yu, (CDF Collaboration), "Search for H to WW Production in the e-tau and mu-tau Final States Using 8.2 fb-1," 2011. CDF/PUB/EXOTIC/PUBLIC/10597.

[61] B. Worthel, (Fermilab), "The Linac rookie book," 2006.
URL: http://www-bdnew.fnal.gov/operations/rookie_books/LINAC_v2.pdf.

[62] (Fermilab), "Run II Handbook," 2001. URL: http://www-ad.fnal.gov/runII/-index.html.

[63] B. W. Worthel, (Fermilab), "The Booster rookie book," 2009.
URL: http://www-bdnew.fnal.gov/operations/rookie_books/Booster_V4.1.pdf.

[64] T. Asher and S. Baginski, (Fermilab), "The Main Injector rookie book," 2006. URL: http://www-bdnew.fnal.gov/operations/rookie_books/-Main_Injector_v1.pdf.

[65] B. E. Drendel *et al.*, "The Antiproton Source rookie book," 2009. URL: http://www-bdnew.fnal.gov/operations/rookie_books/Pbar_v2.pdf.

[66] J. Marriner, (Fermilab), "Stochastic cooling overview," 2003. FERMILAB-CONF-03-158.

[67] P. Karns *et al.*, "The Recycler Injector rookie book," 2010. URL: http://www-bdnew.fnal.gov/operations/rookie books/Recycler_v1.41.pdf.

[68] CDF Summary of stores webpge:
URL: http://www-cdfonline.fnal.gov/ops/opshelp/stores/.

[69] D0 Summary of stores webpage:
URL: http://www-d0.fnal.gov/runcoor/RUN/run2_lumi.html.

[70] D. Acosta *et al.*, "The CDFII Detector Technical Design Report," *Phys. Rev. D*, vol. 71, p. 032001, 2005. Fermilab Report No. Fermilab-Pub-96/390-E.

[71] C. S. Hill, (CDF Collaboration), "Operational experience and performance of the CDFII silicon detector," *Nucl. Instrum A*, vol. 530, p. 1, 2004.

[72] T. K. Nelson, (CDF Collaboration), "The CDF Layer 00 Detector," 2001. FERMILAB-CONF-01-357-E.

[73] A. Affolder *et al.*, (CDF Collaboration), "The intermediate silicon layers detector at CDFII: Design and progress," *Nucl. Instrum. Meth. A*, vol. 435, p. 44, 1999.

[74] A. Affolder *et al.*, (CDF Collaboration), "The CDF intermediate silicon layers detector," *Nuovo Cim. A*, vol. 112, p. 1351, 1999.

[75] T. Affolder *et al.*, (CDF Collaboration), "CDF central outer tracker ," *Nucl. Instrum. Meth. A*, vol. 526, p. 249, 2004.

[76] C. Chen, I. Cho, C. Hays, M. Herndon, J. Kraus, J. Kroll, T. Miao, P. Murat, R. OldeMan, and J. C. Yun, (CDF Collaboration), "Measurement of the Track Reconstruction Efficiency in the COT Using a MC Track Embedding Technique," tech. rep., 2003. CDF/PHYS/BOTTOM/CDFR/6394.

[77] L. Balka *et al.*, (CDF Collaboration), "The CDF Central Electromagnatic Colorimeter," *Nucl. Instr. Electromagnetic A*, vol. 267, p. 272, 1988.

[78] F. Abe *et al.*, (CDF Collaboration), "Prompt photon cross section measurement in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV," *Phys. Rev. D*, vol. 48, pp. 2998–3025, Oct 1993, 7, American Physical Society. doi:10.1103/PhysRevD.48.2998.

[79] G. Apollinari *et al.*, (CDF Collaboration), "Shower maximum detector for the CDF plug upgrade calorimeter," *Nucl. Instr. Meth. A*, vol. 412, p. 515, 1998.

[80] C. M. Ginsburg, (CDF Collaboration), "CDF Run 2 muon system," *Eur. Phys. J. C*, vol. 33, p. 1002, 2004.

[81] T. Dorigo, (CDF Collaboration), "The muon system upgrade for the CDF II experiment," *Nucl. Instrum. Meth. A*, vol. 461, p. 560, 2001.

[82] Elias, J. and Klimenko, S. and Konigsberg, J. and Korytov, A. and Mitselmakher, G. and Nomerotski, A. and Northacker, D. and Safonov, A. and Vidal, R., (CDF Collaboration), "Luminosity monitor based on Cherenkov counters for P anti-P colliders," *Nucl. Instrum. Meth. A*, vol. 441, pp. 366–373, 2000.

[83] F. Abe *et al.*, (CDF Collaboration), "Measurement of the antiproton-proton total cross section at $\sqrt{s} = 546$ and 1800 GeV," *Phys. Rev. D*, vol. 50, pp. 5550–5561, Nov 1994, 9, American Physical Society. doi:10.1103/PhysRevD.50.5550.

[84] A. Bhatti *et al.*, (CDF Collaboration), "The CDF level 2 calorimetric trigger upgrade," *Nucl. Instrum. Meth.*, vol. A598, pp. 331–333, 2009.

[85] S. Belforte *et al.*, (CDF Collaboration), "Silicon Vertex Tracker TDR," 2009. CDF/DOC/TRIGGER/PUBLIC/3108 (SVT TECHNICAL DESIGN REPORT).

[86] A. Annovi *et al.*, (CDF Collaboration), "Upgrade for the cdf svt," 2004. CDF/-DOC/TRIGGER/CDFR/6947.

[87] W. Badgett, R. Frisch, Henry andSt. Denis, and T. Vaiciulis, (CDF Collaboration), "The Good Run List," 2001. CDF/MEMO/PRODUCTION/PUBLIC/5613.

[88] M. Paulini and E. Gerchtein, (CDF Collaboration), "CDF detector simulation framework and performance," march 2003. Talk from the Computing in High Energy and Nuclear Physics (CHEP03), La Jolla, CA, USA. arXiv:physics/0306031v1.

[89] R. Brun and F. Carminati, "Geant 3.21 detector description and simulation tool," 1993. CERN Programming Library Long Writeup W5013 edition, CERN, Geneva.

[90] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 05, p. 026, 2006.

[91] H. Lai *et al.*, (CTEQ Collaboration) *Eur. Phys. J*, vol. C12, p. 375, 2000.

[92] U. Baur and E. L. Berger, "Probing the weak boson sector in $Z\gamma$ production at hadron colliders," *Phys. Rev. D*, vol. 47, pp. 4889–4904, 1993.

[93] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions," *JHEP*, vol. 0307, p. 001, 2003. arXiv:hep-ph/0206293.

[94] R. Veenhof, "GARFIELD, recent developments," *Nucl. Instrum. Methods A*, vol. 419, pp. 726–730, 1998.

[95] S. Peters, G. Grindhammer, and M. Rudowitz, "The Fast Simulation of Electromagnetic and Hadronic Showers," *Nucl. Instrum. Methods A*, vol. 290, p. 469, 1990.

[96] R. G. Wagner, (CDF Collaboration), "Electron Identification for Run II: Understanding and Using Lshr," tech. rep., 2003. CDF/DOC/ELECTRON/-CDFR/6249.

[97] D. Acosta *et al.*, " Direct photon cross section with conversions at CDF," *Phys.Rev.D*, vol. 70, p. 074008, 2004. Fermilab-Pub-04/044E. arXiv:hep-ex/0404022v2.

[98] S. Levy, (CDF Collaboration), "Updated Central Electromagnetic Identification," 2003. CDF/PUB/ELECTRON/PUBLIC/6721.

[99] P. de Barbaro, A. Bodek, H. S. Budd, J. Liu, W. Sakumoto, and Y. Fukui, (CDF Collaboration), "Measurement of the Shape of the Signal from the Plug Upgrade Calorimeter for Electrons and Hadrons," 1998. CDF/MEMO/PLUG_UPGR/PUBLIC/4573.

[100] S. C. Hsu, E. Lipeles, K. McCarthy, M. Neubauer, and F. Wurthwein, (CDF Collaboration), "Lepton ID and Fake Rates for the WW and WZ Analyses," Tech. Rep. CDF/ANAL/ELECTROWEAK/CDFR/8093, 2006.

[101] (CDF Collaboration), "Search for $H \to WW$ Production at CDF using 4.8 fb^{-1} of Data," 2009. CDF/PUB/EXOTIC/PUBLIC/9887.

[102] A. Anastassov, J. Conway, D. Jang, A. Lath, F. Ratnikov, and Z. R. U. Wan, (CDF Collaboration), "Tau Reconstruction Efficiency and QCD Fake Rate for Run 2," 2003. CDF/ANAL/EXOTIC/PUBLIC/6308.

[103] J. Insler, S. Demers, and K. McFarland, (CDF Collaboration), "Isolated, Low Momentum Hadron Response in the Central Calorimeter," 2001. CDF/ANAL/CALORIMETRY/PUBLIC/5703.

[104] A. Anastassov, J. Conway, C. Cuenca, D. Jang, and A. Lath, (CDF Collaboration), "Search for Neutral MSSM Higgs Bosons Decaying to Tau Pairs with 1.8 fb^{-1} of Data," 2007. CDF/PHYS/EXOTIC/PUBLIC/9071.

[105] T. Aaltonen *et al.*, CDF Collaboration, "Search for Higgs Bosons Predicted in Two-Higgs-Doublet Models via Decays to Tau Lepton Pairs in 1.96 TeV $p\bar{p}$ Collisions," *Phys. Rev. Lett.*, vol. 103, p. 201801, Nov 2009, 20, American Physical Society. doi:10.1103/PhysRevLett.103.201801.

[106] G. C. Blazey *et al.*, "Run II jet physics," 2000. arXiv:hep-ex/0005012.

[107] A. Bhatti *et al.*, "Determination of the Jet Energy Scale at the Collider Detector at Fermilab," *Nucl. Instrum. Meth. A*, vol. 566, 2006.

[108] "CDF Run2 triggers and filters report group triggers and bits on cdf online production report physics table physics 5 04 v-3. Website: http://www-cdfonline.fnal.gov/," 2010.

[109] J. M. Campbell and R. K. Ellis, "Update on vector boson pair production at hadron colliders," *Phys. Rev. D*, vol. 60, p. 113006, Nov 1999, 11, American Physical Society. doi:10.1103/PhysRevD.60.113006.

[110] M. Cacciari, S. Frixione, G. Ridolfi, M. L. Mangano, and P. Nason, "The $t\bar{t}$ cross-section at 1.8 and 1.96 TeV: a study of the systematics due to parton densities and scale dependence," *JHEP*, pp. 068–074, 2004. doi:10.1088/1126-6708/2004/04/068.

[111] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," *JHEP*, vol. 06, p. 029, 2002. arXiv:hep-ph/0204244.

[112] V. Boisvert, (CDF Collaboration), "Trigger Efficiencies for the High E_T Central Electrons in Gen6," 2006. CDF/DOC/ELECTRON/CDFR/7939.

[113] L. L. Simona Rolli, (CDF Collaboration), "Muon ID, REconstruction and Trigger Efficiencies and Scale Factors for Perdiod 9-12 data," 2007. CDF/ANAL/MUON/CDFR/9085.

[114] D. Bortoletto, A. Canepa, O. Gonzalez, and E. Lytken, (CDF Collaboration), "Trigger Efficiency for medium p_T Dimuon Triggers in Run II," 2004. CDF/DOC/EXOTIC/CDFR/7196.

[115] S. Tourneur, *Search for top pair production in the dilepton decay channel with tau lepton at the CDF experiment.* PhD thesis, l' Université Paris 6, 2006.

[116] M. Feindta and U. Kerzel, "The NeuroBayes Neural Network package," *Nucl. Instr. and Meth. A*, vol. 559, pp. 190–194, 2006.

[117] M. Feindt, "A Neural Bayesian Estimator for Conditional Probability Densities," 2004. arXiv:physics/0402093v1.

[118] J. Pumplin *et al.*, "New generation of parton distributions with uncertainties from global QCD analysis," *JHEP*, vol. 07, p. 012, 2002. arXiv:hep-ph/0201195.

[119] S. Moch and P. Uwer, "Heavy-quark pair production at two loops in QCD," *Nucl. Phys. Proc. Suppl.*, vol. 183, pp. 75–80, 2008. arXiv:0807.2794.

[120] J. M. Campbell and R. K. Ellis, " Radiative corrections to Z b anti-b production," *Phys. Rev., vol. D*, vol. 62, p. 114012, 2000. arXiv:hep-ph/0006304.

[121] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini, "Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO," *Phys. Rev. Lett.*, vol. 103, p. 082001, 2009. arXiv:0903.2120.

[122] A. Bellavance, (CDF Collaboration), "W/Z Production Cross Sections and Asymmetries at $E_{CM} = 2$ TeV," 2005. FERMILAB-CONF-05-235-E. arXiv:hep-ex/0506025v1.

[123] The TeV4LHC Working Group:
http://maltoni.home.cern.ch/maltoni/TeV4LHC/SM.html.

[124] S. Catani and M. Grazzini, "An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC," *Phys. Rev. Lett.*, vol. 98, p. 222002, 2007. arXiv:hep-ph/0703012v2.

[125] HNNLO program is publically available at:
http://theory.fi.infn.it/grazzini/codes.html.

[126] T. Junk, (CDF Collaboration), "Sensitivity, exclusion and discovery with small signal, large background, and large systematics," tech. rep., 2007. CDF/DOC/STATISTICS/PUBLIC/8128.

[127] M. Bauce, S. Behari, D. Benjamin, P. Bussey, A. Canepa, B. Carls, S. Carron, M. Casarsa, R. St.Denis, M. d'Errico, P. M. Fernandez, M. Herndon, E. James, S. Jindariani, T. Junk, M. Kruse, A. Limosani, D. Lucchesi, R. Lysak, J. Nett, S. Oh, A. Robson, and G. Yu, (CDF Collaboration), "Search for $H \to WW^*$ Production at CDF Using $8.2 fb^{-1}$ of Data," 2011. CDF/PUB/EXOTIC/PUBLIC/10599.

[128] (CDF and D0 Collaborations, the TEVNPHWG Working Group), "Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb^{-1} of Data," 2011. Public note for summer 2011 conferences, CDF Note 10606.