



Università degli Studi di Pisa  
Corso di Laurea Specialistica in Scienze Fisiche  
Anno Accademico 2006/2007

Tesi di Laurea Specialistica

## Improving acceptance for Higgs events at CDF

*Tesi di Federico Sforza*

*Relatore: Prof. Giorgio Chiarelli*

*To my father and my family*

# Contents

<b>Introduction</b>	<b>vii</b>
<b>1 Standard Model of Elementary Particles and Higgs Physics</b>	<b>1</b>
1.1 Standard Model of Elementary Particles . . . . .	1
1.1.1 Gauge Theory . . . . .	2
1.1.2 Standard Model Theory . . . . .	3
1.2 Higgs Mechanism . . . . .	6
1.2.1 Spontaneous Symmetry Breaking in SM . . . . .	7
1.3 Experimental Limits on Higgs Boson Mass . . . . .	9
1.4 Higgs Production and Search at the Tevatron . . . . .	11
1.4.1 Higgs Production and Decay . . . . .	12
1.4.2 $WH$ Search at the Tevatron . . . . .	15
<b>2 The Tevatron Collider and the CDF II experiment</b>	<b>19</b>
2.1 The Tevatron . . . . .	19
2.1.1 Proton and Antiproton Production . . . . .	19
2.1.2 Collision and Performance . . . . .	21
2.2 The CDF Detector . . . . .	23
2.2.1 Overview and Coordinate system . . . . .	24
2.2.2 Tracking System . . . . .	26
2.2.3 Central Outer Tracker . . . . .	29
2.2.4 Calorimeter System . . . . .	30
2.2.5 Other Detectors . . . . .	35
2.3 Trigger and Data Handling . . . . .	37
2.3.1 CDF Software Framework . . . . .	40
2.4 Monte Carlo Simulation . . . . .	41

---

<b>3</b>	<b>Physical Objects Reconstruction</b>	<b>43</b>
3.1	Tracks Reconstruction . . . . .	44
3.1.1	Tracking Algorithms . . . . .	45
3.2	Calorimeter Electron Identification . . . . .	47
3.3	Muon Identification . . . . .	49
3.4	Primary Vertex Identification . . . . .	50
3.5	Neutrino Identification . . . . .	50
3.6	Jet Identification . . . . .	51
3.6.1	CDF Cone Algorithm . . . . .	52
3.6.2	Jet Corrections . . . . .	54
3.7	Secondary Vertex Tagging . . . . .	57
3.7.1	The SecVtx Algorithm . . . . .	59
3.7.2	Tagging Performances and Scale Factors . . . . .	61
<b>4</b>	<b>Event Selection and Signal Acceptance</b>	<b>65</b>
4.1	Data Sample and Run Interval . . . . .	65
4.2	Event Selection Requirements . . . . .	66
4.3	Forward Tracking Efficiency Study . . . . .	70
4.4	WH Signal Acceptance . . . . .	71
<b>5</b>	<b>Background Analysis</b>	<b>75</b>
5.1	Overview of the Background Composition . . . . .	75
5.1.1	Background Estimate . . . . .	76
5.2	Non- $W$ QCD Background . . . . .	77
5.2.1	Tagged non- $W$ Background . . . . .	79
5.3	Heavy Flavor Background . . . . .	80
5.4	Light Flavors/Mistags . . . . .	84
5.5	Electroweak and Top Background . . . . .	88
5.6	Background Summary . . . . .	89
<b>6</b>	<b>Comparison of Kinematic Quantities</b>	<b>95</b>
6.1	Kinematic Comparison . . . . .	95
<b>7</b>	<b>Conclusions</b>	<b>105</b>
7.1	Shapes Comparison . . . . .	106
7.2	Signal Improvements . . . . .	106
7.3	Future Prospect . . . . .	107
<b>A</b>	<b>Trigger Efficiency Studies</b>	<b>109</b>
A.1	MET_PEM Trigger Path . . . . .	109
<b>B</b>	<b>Track Matching Optimization</b>	<b>113</b>

## CONTENTS

v

---

C Tight Electrons and Muons

115

D Kolmogorov Smirnov Test

117



# Introduction

The Standard Model of elementary particles predicts the existence of the Higgs boson as the responsible of the electroweak symmetry breaking, the process by which fermions and vector bosons acquire mass. The Higgs existence is one of the most important questions in the present high energy physics research.

This work concerns the search of  $WH$  associate production at the CDF II experiment (Collider Detector at Fermilab). Even if  $WH$  production is one of the favoured search channels, the expected cross section is very tiny:

$$0.1 \sim 0.2 \text{ pb for } m_H < 140 \text{ GeV}/c^2, \quad (1)$$

therefore it is of fundamental importance to exploit the maximum capability of the detector.

This analysis searches for  $WH$  events in the decay channel:

$$W^\pm H \rightarrow e^\pm \nu b \bar{b}. \quad (2)$$

Appropriate cuts are applied to select candidate events: one electron is reconstructed through a calorimetric cluster in the forward region of the detector ( $1.2 < |\eta| < 2.8$ ) with a track matched to it, the neutrino is revealed as missing energy and at least one jet compatible with  $b$ -hadrons decay must be identified.

This kind of selection improves the CDF acceptance in a twofold way: first of all signal and background are studied in the forward region of the detector, part not fully exploited up to now because most of the analyses are based on the central part of the detector. Second a new set of track reconstruction algorithms is tested, this provides an increased tracking efficiency maintaining background under control.





# Chapter 1

## Standard Model of Elementary Particles and Higgs Physics

The present physics theories of Nature identify four kinds of fundamental interactions: gravitational, electromagnetic, weak and strong interaction. The *Standard Model of elementary particle* (SM) unifies and describes in an excellent way the last three interactions, leaving out gravitational force, that is, however, negligible at atomic and subatomic scale.

The theory is verified with very good accuracy and predicted the existence of new particles (like the  $W$  and  $Z$  bosons or the *top* quark) that were later discovered. The SM also predicts the existence of a not yet discovered particle, the *Higgs boson*, an essential element to introduce particle masses in the equation of motion[1, 2, 4].

Indirect limits can be set on the expected Higgs mass and production cross section on the basis of SM assumptions and exploiting electroweak measurements. At the moment the direct experimental limits are above the SM expectation but the high-energy physics community is pushing forth in the search. The Tevatron, being the most powerful hadron collider currently in operation, plays a fundamental role in this search.

### 1.1 Standard Model of Elementary Particles

High energy particle physics inquires nature at fundamental level, its constituents and the basic interactions. In theoretical physics language particles are quantum local fields interacting via the exchange of force-mediator vector bosons. A free field is completely described only by spin and mass, the SM introduces interactions through gauge symmetries, with new quantum numbers that classify type and strength of forces.

The fundamental building blocks of matter observed up to now are the spin 1/2 fields (*fermionic* fields) called *quarks* and *leptons* and the spin 1 vector boson fields called *gauge bosons*.

The leptons are divided into three *generation* or *families* and are grouped in a left weak isospin doublet<sup>1</sup> and a right weak isospin singlet. Also quarks are divided into three *flavor families* but weak isospin classification mixes quark doublets of different families, besides there is the *color* quantum number to take into account strong interaction.

The force mediators are  $W^\pm$ ,  $Z^0$ ,  $\gamma$ , that carry electroweak force, and  $g$  (*gluons*), which mediate strong interaction. A short summary of the SM fundamental particles is reported in Table (1.1).

Generation			Proprieties		
1st	2nd	3rd	Spin ( $\hbar$ )	Charge ( $e$ )	Interaction
$\begin{pmatrix} u \\ d \end{pmatrix}$	$\begin{pmatrix} c \\ s \end{pmatrix}$	$\begin{pmatrix} t \\ b \end{pmatrix}$	1/2	$\begin{pmatrix} +2/3 \\ -1/3 \end{pmatrix}$	EM,Weak,Strong
$\begin{pmatrix} \nu_e \\ e \end{pmatrix}$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$	1/2	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	EM,Weak
Gauge Boson	Mass (GeV/c <sup>2</sup> )	Spin ( $\hbar$ )	Charge ( $e$ )	Interaction	
$\gamma$	0	1	0	EM	
$W$	80.4	1	$\pm 1$	EM,Weak	
$Z$	91.2	1	0	EM,Weak	
$g$	0	1	0	Strong	

Table 1.1: Quarks, leptons and gauge bosons in Standard Model and some of their characteristics, for each particle the corresponding antiparticle exists.

### 1.1.1 Gauge Theory

The SM is a local quantum field theory based on a *local gauge symmetry* and on the *least-action principle* to constrain the equation of motion.

The importance of gauge invariance comes directly from free Dirac Lagrangian, the equation that describes free fermionic fields:

$$\mathcal{L}(x) = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x), \quad (1.1)$$

where  $\psi$  is the Dirac field of mass  $m$  and  $\gamma^\mu$  are the Dirac's matrices. Eq. 1.1 satisfies the *global*  $U(1)$  symmetry transformation:

$$\psi(x) \rightarrow e^{iQ\alpha}\psi(x), \quad (1.2)$$

---

<sup>1</sup>See section 1.1.2 for the explanation of the weak isospin quantum number.

with the electric charge  $Q$  and the space independent parameter  $\alpha$  ( $x$  is a space-time 4-vector). The *Noether theorem*[4] states that when a symmetry appears in a Lagrangian there is a corresponding conserved current. In the case of the Dirac field:

$$\partial_\mu j^\mu = 0, \quad (1.3)$$

with the current 4-vector  $j^\mu = -Q\bar{\psi}\gamma^\mu\psi$ . This leads to the conservation of the charge, i.e the time component of  $j^\mu$  (integrated over the space).

An elegant way to introduce interaction in the free Lagrangian is to shift from the global, i.e. space independent,  $U(1)$  transformation to a *local*  $U(1)$  transformation, i.e. with a space dependent parameter  $\alpha(x)$ :

$$\psi(x) \rightarrow e^{iQ\alpha(x)}\psi(x), \quad (1.4)$$

to maintain the gauge invariance condition in the Lagrangian 1.1, a covariant derivative  $D_\mu$  is introduced:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iQA_\mu, \quad D_\mu\psi(x) \rightarrow e^{iQ\alpha(x)}D_\mu\psi(x), \quad (1.5)$$

where it was defined a new vector field  $A_\mu$  transforming in the following manner:

$$A_\mu \rightarrow A_\mu - \frac{1}{Q}\partial_\mu\alpha(x). \quad (1.6)$$

The final result is the QED Lagrangian:

$$\mathcal{L}_{QED} = \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (1.7)$$

where  $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$  is the covariant kinetic term of  $A_\mu$ . If the Euler-Lagrange equation[4] is applied, we obtain the Dirac equation of motion for a field  $\psi$  undergoing electromagnetic interaction:

$$(i\gamma^\mu\partial_\mu - m)\psi(x) = Q\gamma^\mu A_\mu\psi(x), \quad (1.8)$$

the force is mediated by the massless vector field  $A_\mu$ . A mass term in the form  $\frac{1}{2}m^2 A_\mu A^\mu$  would break apart gauge invariance of Eq. 1.7, indeed this is consistent with zero mass of the photon.

### 1.1.2 Standard Model Theory

The SM is based on a gauge group  $SU(2) \otimes U(1)^2$ .  $SU(2)$  is the non-Abelian group of the spin algebra (the so-called *weak isospin*) and it is characterized

---

<sup>2</sup>Only electroweak interaction in the leptonic sector is considered.

by three generators linked to three gauge vector fields, beyond the vector field produced by  $U(1)$  group generator.

Electroweak interaction can be explained with a simplified model containing only two spin  $1/2$ , elementary, massless, fermions,  $f$  and  $f'$ , such that  $Q_f = Q_{f'} + 1$  ( $Q$  is the electric charge).

Weak interaction is built from V-A currents. Left and right components are defined and they are collected into a left doublet field and into two right singlet fields:

$$\psi_1 \equiv \begin{pmatrix} f_L(x) \\ f'_L(x) \end{pmatrix}, \quad \psi_2 \equiv f_R(x) \quad \psi_3 \equiv f'_R(x), \quad (1.9)$$

with:

$$f_{L,R}(x) = \frac{1}{2}(1 \pm \gamma_5)f(x), \quad \bar{f}_{L,R}(x) = \frac{1}{2}\bar{f}(x)(1 \pm \gamma_5), \quad (1.10)$$

$$f'_{L,R}(x) = \frac{1}{2}(1 \pm \gamma_5)f'(x), \quad \bar{f}'_{L,R}(x) = \frac{1}{2}\bar{f}'(x)(1 \pm \gamma_5). \quad (1.11)$$

The leptonic sector of the SM can be explained by such pattern: we define  $T_3$  as the third component of weak isospin and  $Y$  (the *hypercharge*) as  $U(1)$  part of the interaction, the Gell-Mann-Nishijima relation binds  $Q$ ,  $T_3$  and  $Y$ :

$$Q = T_3 + \frac{Y}{2}. \quad (1.12)$$

The left doublet with  $T_3 = \pm 1/2$ ,  $Y = 1$  is the charged lepton  $f$  plus the corresponding neutrino  $f'$ , while the right singlet with  $T_3 = 0$ ,  $Y = -2$  is only the charged lepton.

The electroweak interaction is introduced through  $SU(2) \otimes U(1)$  gauge transformation:

$$\psi_j(x) \rightarrow \psi'_j(x) = e^{i\frac{\tau}{2} \cdot \vec{\alpha}(x) + iY_j\beta(x)} \psi_j(x), \quad (1.13)$$

in the free field Lagrangian:

$$\mathcal{L}(x) = \sum_{j=1}^3 i\bar{\psi}_j(x) \gamma^\mu \partial_\mu \psi_j(x), \quad (1.14)$$

and a covariant derivative is introduced in Eq. 1.14 to maintain gauge invariance:

$$\mathcal{L}_I(x) = \sum_{j=1}^3 i\bar{\psi}_j(x) D_\mu^j \partial_\mu \psi_j(x), \quad (1.15)$$

$$\text{with} \quad D_\mu^j = \partial_\mu - ig\frac{\tau}{2} \cdot \vec{W}_\mu(x) - ig'Y_j B_\mu(x). \quad (1.16)$$

Eq. 1.16 contains three vector bosons ( $\vec{W}_\mu$ ) from  $SU(2)$  generators, one ( $B_\mu$ ) from  $U(1)$  generator and four coupling constants:  $g, g'Y_j$  ( $j = 1, 2, 3$ ). After some algebra the Lagrangian 1.15 can be written in the form:

$$\mathcal{L}_I(x) = \mathcal{L}_{CC}(x) + \mathcal{L}_{NC}(x), \quad (1.17)$$

with a “charged current” contribution ( $\mathcal{L}_{CC}$ ) and a “neutral current contribution” ( $\mathcal{L}_{NC}$ ). The charged current contribution is seen only by left doublet fields:

$$\mathcal{L}_{CC}(x) = \frac{g}{2\sqrt{2}} \left\{ \bar{f}(x) \gamma^\mu (1 - \gamma_5) f'(x) \frac{1}{\sqrt{2}} W_\mu^+(x) + h.c. \right\}, \quad (1.18)$$

with  $W_\mu^+(x)$  a linear combination of  $W_\mu^1(x)$  and  $W_\mu^2(x)$ . Eq. 1.18 is the correct Lagrangian for charged current behavior mediated by the  $W$  boson.

The fermion coupling to  $Z^0$  and photon is produced in a similar way: an appropriate orthogonal linear combination of neutral vector fields  $B_\mu(x)$  and  $W_\mu^3(x)$  produce the correct fermion coupling:

$$\mathcal{L}_{NC}(x) = \mathcal{L}_{NC}^A(x) + \mathcal{L}_{NC}^Z(x) : \quad (1.19)$$

$$\mathcal{L}_{NC}^A(x) = \sum_{j=1}^3 \bar{\psi}_j(x) \gamma^\mu \left[ g \frac{\tau_3}{2} \sin \theta_W + g' Y_j \cos \theta_W \right] \psi_j(x) A_\mu(x), \quad (1.20)$$

$$\mathcal{L}_{NC}^Z(x) = \sum_{j=1}^3 \bar{\psi}_j(x) \gamma^\mu \left[ g \frac{\tau_3}{2} \cos \theta_W + g' Y_j \sin \theta_W \right] \psi_j(x) Z_\mu(x), \quad (1.21)$$

$\theta_W$  is the Weinberg angle and the generic four coupling constants have now a physical meaning:

$$g \sin \theta_W = e, \quad (1.22)$$

$$g' \cos \theta_W Y_1 = e(Q_f - 1/2), \quad g' \cos \theta_W Y_2 = eQ_f, \quad (1.23)$$

$$g' \cos \theta_W Y_3 = eQ_{f'} \quad (1.24)$$

Preceding equations are the core of the Standard Model. However the masses of the field do not appear: the spontaneous symmetry breaking and Higgs mechanism can generate the mass without breaking the gauge invariance.

## 1.2 Higgs Mechanism

Spontaneous breaking of symmetry is based on the possibility, in systems with infinite degrees of freedom, to have a Lagrangian invariant under a group  $G$  of transformation that produces non symmetric states.

A toy Higgs mechanism[5] can be realized in Lagrangian density of scalar electrodynamics:

$$\begin{aligned} \mathcal{L}(x) = & -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \\ & [(\partial_\mu + ieA_\mu(x))\phi^\dagger(x)] \cdot [(\partial^\mu - ieA^\mu(x))\phi(x)] \\ & -\mu^2\phi^\dagger(x)\phi(x) - h[\phi^\dagger(x)\phi(x)]^2 \end{aligned} \quad (1.25)$$

where  $h > 0$ ,  $\mu^2 < 0$ ,  $\phi(x)$  is the scalar field undergoing electromagnetic interaction via the Abelian gauge field  $A^\mu(x)$ , the last part of the Lagrangian is the Higgs potential. Eq. 1.25 maintains invariance under the local gauge transformation:

$$\phi(x) \rightarrow \phi'(x) = e^{i\alpha(x)}\phi(x) \quad A_\mu(x) \rightarrow A'_\mu(x) = A_\mu(x) - \partial_\mu\alpha(x) \quad (1.26)$$

The solution of the equation of motion corresponds to the minimal energy solution, in other words the vacuum expectation values of the fields in lowest order perturbation theory. Because of  $\mu^2 < 0$  there is not only the trivial solution  $\phi(x) = 0$ , but there exist a set of degenerate solutions with  $|\phi|^2 = \frac{-\mu^2}{h} = \frac{\lambda^2}{2}$ . This reflects the underlying gauge symmetry:  $\phi(x) = \frac{\lambda}{\sqrt{2}}e^{i\alpha(x)}$  (see Fig. 1.1). Gauge freedom allows to choose  $\alpha(x)$  such that  $\phi'(x)$  is real and

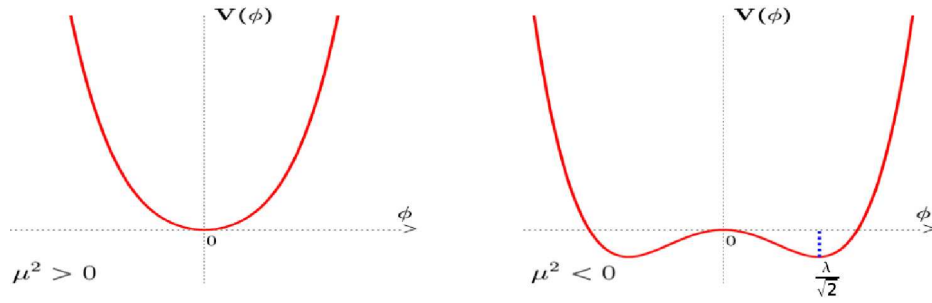


Figure 1.1: Symmetry breaking depending from  $\mu^2$  parameter:  $\mu^2 > 0$  on the left,  $\mu^2 < 0$  on the right.

the lowest state is  $\phi(x) = \frac{\lambda}{\sqrt{2}}$ . To first order we can write:

$$\phi'(x) = \frac{1}{\sqrt{2}}[\lambda + \phi_1(x)], \quad \phi_2(x) = 0, \quad A'_\mu(x) = B_\mu(x). \quad (1.27)$$

Replacing Eq. 1.27 in Eq. 1.25 and writing the Lagrangian in powers of  $\phi_1(x)$  one obtains:

$$\begin{aligned}
 \mathcal{L}(x) = & -\frac{1}{4}B_{\mu\nu}(x)B^{\mu\nu}(x) + \frac{1}{2}e^2\lambda^2 B_\mu(x)B^\mu(x) \\
 & + e^2\lambda B_\mu(x)B^\mu\phi_1(x) + \frac{1}{2}e^2\lambda B_\mu(x)B^\mu\phi_1^2(x) \\
 & + \frac{1}{2}[\partial^\mu\phi_1(x)\partial_\mu\phi_1(x) + 2\mu^2\phi_1^2(x)] \\
 & + \frac{\mu^2}{\lambda}\phi_1^3(x) + \frac{\mu^2}{4\lambda^2}\phi_1^4(x) - \frac{1}{4}\lambda^2\mu^2.
 \end{aligned} \tag{1.28}$$

Each line in Eq. 1.28 has a physical meaning:

- the first line describes a *massive* vector field with mass  $|e\lambda|$  instead of the original massless gauge field;
- the second line is the interaction of the vector field with the neutral scalar field with coupling strength  $e^2\lambda$  and  $\frac{1}{2}e^2$ ;
- the third line is the free scalar Lagrangian of a particle, called *Higgs*, with mass  $M_H = \sqrt{-2\mu^2}$ ;
- the last line is the self interaction of the scalar field.

Spontaneous symmetry breaking takes place in Eq. 1.25, the initial complex scalar field (two degrees of freedom) and the massless vector field (other two degrees of freedom for the helicity states) turns into a scalar real (neutral) particle (one degree of freedom) and a massive vector boson (three degrees of freedom).

### 1.2.1 Spontaneous Symmetry Breaking in SM

Spontaneous symmetry breaking can be also applied to Eq. 1.9 to give mass to  $W^\pm$  and  $Z^0$  bosons. Two complex scalar fields are introduced to adapt Higgs mechanism to gauge groups  $SU(2) \otimes U(1)$ . They form an isodoublet with respect to  $SU(2)$  group:

$$\phi(x) \equiv \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix}, \tag{1.29}$$

where the field  $\phi^+(x)$  is the charged component of the doublet and  $\phi^0(x)$  is neutral. SM Lagrangian with the added Higgs potential gives:

$$V_H(x) \equiv -\mu^2\phi^\dagger(x)\phi(x) - h[\phi^\dagger(x)\phi(x)]^2, \tag{1.30}$$

with  $h > 0$  and  $\mu^2 < 0$ . Eq. 1.27 states that the neutral scalar field  $\phi^0(x)$  has a vacuum expectation value of  $\frac{\lambda}{\sqrt{2}}$ , so that (at first order) the field 1.29 is:

$$\phi(x) = e^{\frac{i}{\lambda} \vec{\tau} \cdot \vec{\theta}(\vec{x})} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(\lambda + \chi(x)) \end{pmatrix}, \quad (1.31)$$

where the  $SU(2)$  gauge freedom is explicit; it permits to *gauge away* three of the four components of field  $\phi(x)$ , only one real scalar field remains:  $\phi^0(x) = \frac{1}{\sqrt{2}}(\lambda + \chi(x))$ . Last section of the SM Lagrangian comes out composing all the previous equations:

$$\begin{aligned} \mathcal{L}(x) = & \frac{1}{4}g^2\lambda^2 W_\mu^\dagger(x)W^\mu(x) + \frac{1}{1}(g^2 + g'^2)\lambda^2 Z_\mu(x)Z^\mu \\ & + \frac{1}{2}g^2\lambda W_\mu^\dagger(x)W^\mu(x)\chi(x) + \frac{1}{4}g^2 W_\mu^\dagger W^\mu \chi^2(x) \\ & + \frac{1}{4}(g^2 + g'^2)\lambda Z_\mu(x)Z^\mu(x)\chi(x) + \frac{1}{8}g^2 Z_\mu(x)Z^\mu(x)\chi^2(x) \\ & + \frac{1}{2}[\partial^\mu \chi(x)\partial_\mu \chi(x) + 2\mu^2 \chi^2(x)] \\ & + \frac{\mu^2}{\lambda}\chi^3(x) + \frac{\mu^2}{4\lambda^2}\chi^4(x) - \frac{1}{4}\lambda^2\mu^2, \end{aligned} \quad (1.32)$$

Eq. 1.32 has to be added to SM Lagrangian. We conclude that now the  $Z^0$  and  $W^\pm$  bosons have acquired mass:

$$M_W = \frac{1}{2}\lambda g, \quad (1.33)$$

$$M_Z = \frac{1}{2}\lambda\sqrt{g^2 + g'^2} = \frac{1}{2}\frac{\lambda g}{\cos\theta_w}. \quad (1.34)$$

Some parameters are now constrained, for example:

$$M_Z = \frac{M_W}{\cos\theta_w} \geq M_W, \quad (1.35)$$

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2}, \quad (1.36)$$

however Higgs mass,  $M_\chi = \sqrt{-2\mu^2}$  (sometimes  $M_H$  is used), remains a free parameter to be measured by experiments. Higgs mechanism generates also fermion masses if a Yukawa coupling is added:

$$\begin{aligned} \mathcal{L}(x) = & c_{f'} \left[ (\bar{f}(x), \bar{f}'(x))_L \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix} \right] f'_R(x) \\ & + c_f \left[ (\bar{f}(x), \bar{f}'(x))_L \begin{pmatrix} -\bar{\phi}^0(x) \\ \phi^-(x) \end{pmatrix} \right] f_R(x) + h.c., \end{aligned} \quad (1.37)$$



therefore, after symmetry breaking, fermion masses have the form:

$$m_f = -c_f \frac{\lambda}{\sqrt{2}}, \quad m_{f'} = -c_{f'} \frac{\lambda}{\sqrt{2}}, \quad (1.38)$$

where the constants  $c_f$  and  $c_{f'}$  can be derived by measurements of the fermion masses.

### 1.3 Experimental Limits on Higgs Boson Mass

Limits on the Higgs mass come both from direct searches or from accurate electroweak measurements that indirectly constrain SM parameters.

**Direct Experimental Limits:** the most important direct limit on the Higgs mass comes from LEP experiments[7]. The experiment performed a direct Higgs search using  $2461 \text{ pb}^{-1}$  of data at a center of mass energy between 189 and 209 GeV. Channels used were  $e^+e^- \rightarrow Z^0 H$ , with  $Z^0$  decaying into all possible modes and  $H \rightarrow b\bar{b}$ , and the channel with  $H \rightarrow \tau^+\tau^-$  and  $Z^0 \rightarrow q\bar{q}$ .

Figure 1.2 shows reconstructed Higgs mass distribution. No significant mass peak was found, so a 95% confidence level lower mass limit was established:

$$m_H > 114.4 \text{ GeV}/c^2 \quad (1.39)$$

However the ALEPH experiment claimed some inconsistency of observed data with expected background.

**Indirect Experimental Limits:** Indirect Higgs mass estimates are done in a model dependent way, assuming the correctness of SM with the Higgs mechanism included. Accurate mass measurements of the heavier SM particles, like  $W^\pm$ ,  $Z^0$  and top quark, pose theoretical limits on the allowed Higgs mass. The mass of those particles is increased by loop diagram corrections as shown in Figure 1.3. The contribution of Higgs mass to gauge boson masses has the following form[8]:

$$\rho = \frac{M_W^2}{M_Z^2(1 - \sin^2 \theta_W)} = 1 + \Delta\rho, \quad (1.40)$$

$$\Delta\rho \equiv \frac{3G_F}{8\pi^2\sqrt{2}}M_t^2 + \frac{\sqrt{2}G_F}{16\pi^2}M_W^2 \left[ \frac{11}{3} \ln\left(\frac{M_H^2}{M_W^2}\right) + \dots \right] \quad (1.41)$$

where  $G_F$  is the Fermi coupling constant,  $\theta_W$  is Weinberg angle,  $M_t$ ,  $M_W$ ,  $M_Z$  and  $M_H$  are, respectively, the masses of top quark,  $W$  and  $Z$  bosons

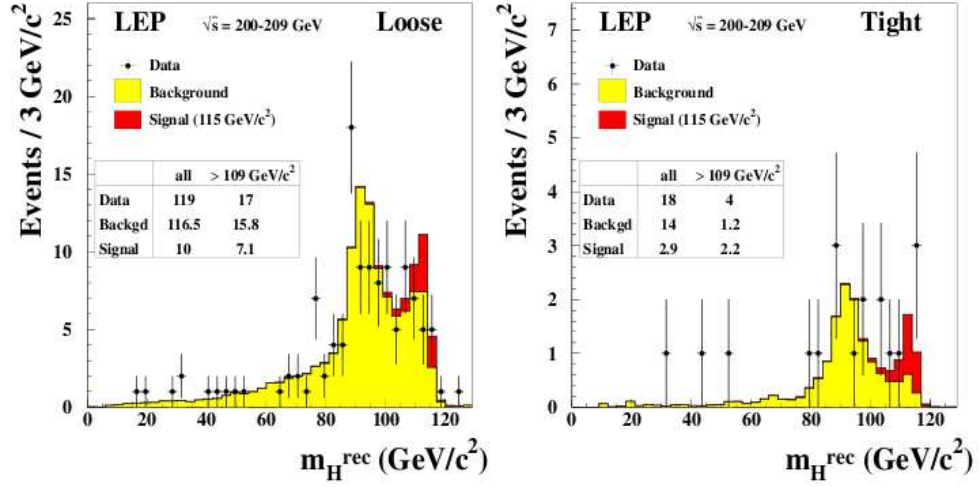


Figure 1.2: Reconstructed Higgs boson mass obtained from two selection at LEP. Monte Carlo predicted background (yellow) and Standard Model Higgs boson signal (red) for a mass of  $115 \text{ GeV}/c^2$  is shown together with data.

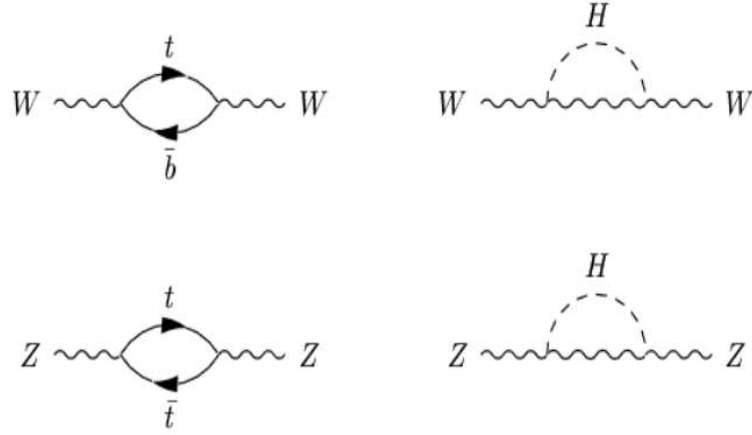


Figure 1.3: Radiative loop contribution to masses of electroweak objects. Precision measurements of the gauge bosons and of the top quark masses can provide a limit on the SM Higgs boson mass.

and Higgs boson. Figure 1.4 shows the limits on  $M_H$ , derived by  $M_W$  and  $M_{top}$  measurements. Contour curves are obtained varying experimental mass values of  $M_t$ ,  $M_W$  of  $\pm 1\sigma$  [9, 10] (68% confidence level).

Fitting all electroweak measurements performed at LEP, SLD, CDF and

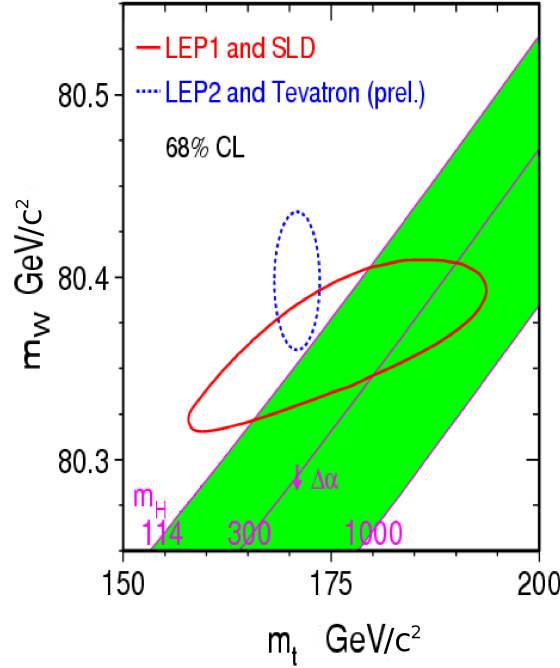


Figure 1.4: SM relationship between  $M_t$ ,  $M_W$  and  $M_H$ . Contour curves are obtained varying experimental mass values of  $\pm\sigma$ [9, 10]. The arrow labeled as  $\Delta\alpha$  shows the global variation if  $\alpha(M_Z)$  is changed by one standard deviation.

DØ with Higgs mass as a free parameter in SM, we derive the  $\Delta\chi^2$  curve in Figure 1.5. The preferred value corresponds to the minimum of the curve and gives  $M_H = 87^{+37}_{-27}$  GeV/ $c^2$  at 68% CL. If also LEP-2 direct search limit is included (yellow region in Fig. 1.5), it gives:

$$114.4 < M_H < 190 \text{ GeV}/c^2, \quad (1.42)$$

95% CL constraint derived both from direct and indirect searches.

## 1.4 Higgs Production and Search at the Tevatron

The Tevatron  $p\bar{p}$  collider, with  $\sqrt{s} = 1.96$  TeV center-of-mass energy, is the only place where operative experiments can explore the Higgs existence in a mass range  $100 - 200$  GeV/ $c^2$ . Currently the CDF II and DØ experiments are collecting quality data and research groups are performing analyses in

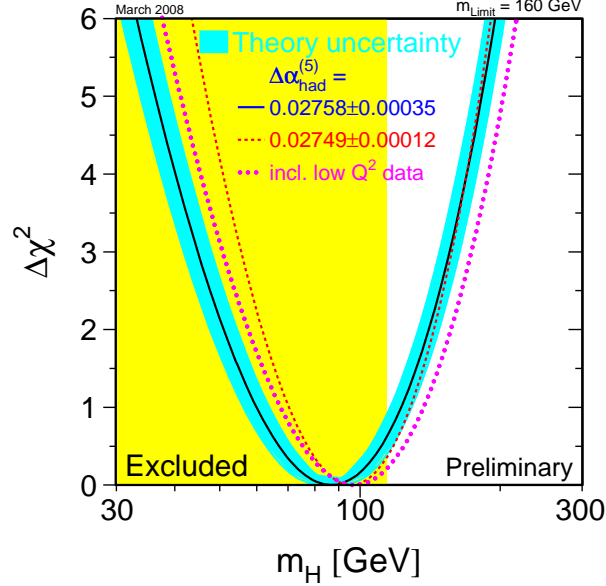


Figure 1.5:  $\Delta\chi^2$  distribution as a function of  $M_H$  from a global fit of electroweak parameters measured at LEP, SLD, CDF, DØ and NuTeV.

many different channels. During 2003 the “CDF and DØ Working Group”[11] estimated how much data should be collected for a  $5\sigma$  discovery,  $3\sigma$  evidence or to exclude Higgs of given mass with 95% CL (see Fig. 1.6). Such study has not been updated recently, and therefore can be considered no more than an indication of the physics reach of the Tevatron.

#### 1.4.1 Higgs Production and Decay

The Higgs production processes accessible at the Tevatron are gluon fusion and vector boson associate production (see Fig. 1.7). Their cross sections have been calculated taking into account QCD radiative corrections[6]. Results are plotted in Figure 1.8 for a wide range of masses.

Higgs mass value also sets the branching ratios and the allowed decay channels (see Fig. 1.9).

On the basis of the Higgs Working Group sensitivity study, the searches for the Higgs are divided in two categories: “high Higgs mass” searches, with  $M_H > 140 \text{ GeV}/c^2$ , and “low Higgs mass” searches, with  $M_H < 140 \text{ GeV}/c^2$ .

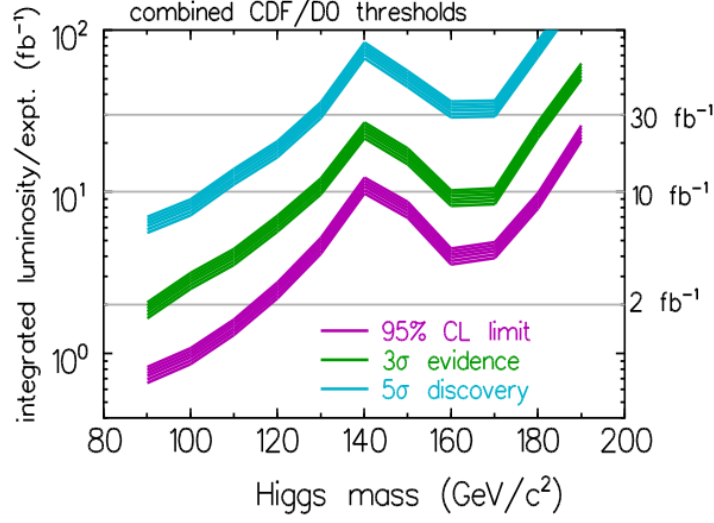


Figure 1.6: Prospect of sensitivity and integrated luminosity for the SM Higgs boson search as a function of  $M_H$  at the Tevatron.

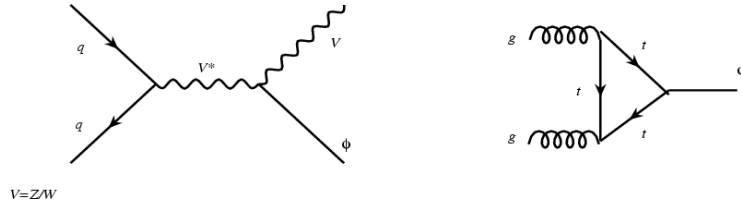


Figure 1.7: Tree diagrams for Higgs ( $\phi$ ) production at the Tevatron. Left: Higgs production associated with a vector boson ( $V$ ). Right: Higgs production through gluon fusion.

In the high mass region, the main Higgs decay channels are  $W^+W^-$  and  $Z^0Z^0$ . The processes investigated are:

$$p\bar{p} \rightarrow H \rightarrow WW^* \rightarrow l\nu jj \text{ and } l\nu\bar{l}\bar{\nu}, \quad (1.43)$$

$$p\bar{p} \rightarrow H \rightarrow ZZ^* \rightarrow lljj \text{ and } l\bar{l}\nu\bar{\nu}, \quad (1.44)$$

$$p\bar{p} \rightarrow W^\pm H \rightarrow l^\pm \nu WW^* \rightarrow l\nu l\nu l\nu, \quad (1.45)$$

$$p\bar{p} \rightarrow W^\pm H \rightarrow l^\pm \nu WW^* \rightarrow l\nu l\nu jj. \quad (1.46)$$

The clean signature of two vector bosons decaying to leptons is an advantage for background rejection but in many cases the vector bosons decay into

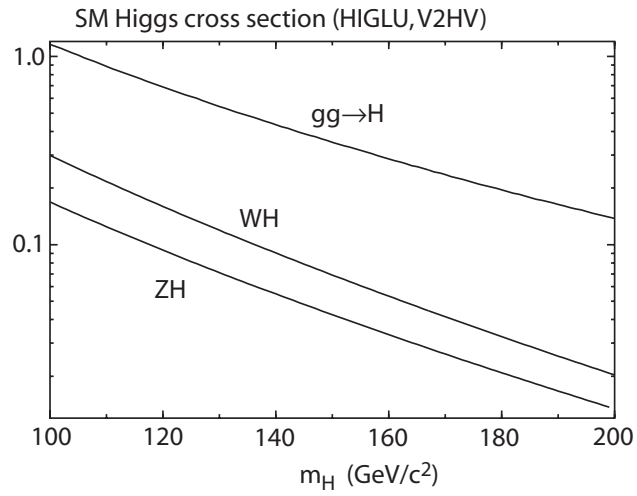


Figure 1.8: Higgs production cross sections as a function of  $m_H$  at the Tevatron center of mass energy  $\sqrt{s} = 1.96 \text{ TeV}$ [6].

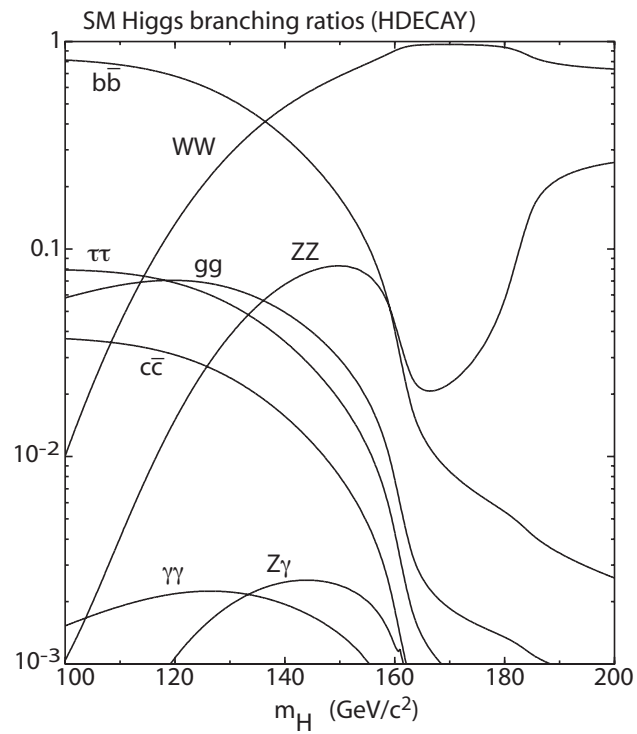


Figure 1.9: Higgs branching ratios as a function of  $m_H$ [6].

high-multiplicity final states therefore it becomes difficult to collect all the information.

Below the  $W^+W^-$  threshold, fermionic branching ratio is favored. Being is proportional to the fermion mass squared:

$$\Gamma(H \rightarrow f\bar{f}) = \mathcal{N}_c \frac{G_F m_f^2}{4\sqrt{2}\pi} m_H \beta^3, \quad (1.47)$$

the favored decay channel is  $H \rightarrow b\bar{b}$ .

In the low mass region the search is focused on  $VH$  production (where  $V$  is either a  $W$  or a  $Z$  boson). Even though these are not the primary production modes, they are the easier to detect. The main Higgs production mode is by gluon fusion but, due to the hadronic environment, this channel has a very large background from QCD processes. The channels currently investigated in the low mass region are:

$$p\bar{p} \rightarrow WH \rightarrow l\nu b\bar{b}, \quad (1.48)$$

$$p\bar{p} \rightarrow ZH \rightarrow l^+l^- b\bar{b}, \quad (1.49)$$

$$p\bar{p} \rightarrow ZH \rightarrow \nu\bar{\nu} b\bar{b}. \quad (1.50)$$

The branching fraction of  $W^\pm$  and  $Z^0$  to leptons further reduce the number of expected events but the clean signature of an isolated lepton and a neutrino provides a simple signature for triggering and background rejection.

### 1.4.2 $WH$ Search at the Tevatron

The channel  $WH \rightarrow l\nu b\bar{b}$  is exploited by both CDF II and DØ collaborations and is one of the most promising for the low Higgs mass region.  $WH$  analyses are based on the lepton plus neutrino selection and on the successive identification of one or more  $b$ -jets (i.e. jets produced by  $b$  quarks).

The usual approach is to optimize the signal to background significance in different  $b$ -jet tagging categories (see section 3.7). Various studies are focused on the optimization of analysis tools, such as neural-network discriminants, jet reconstruction algorithms or  $b$ -taggers. Other works (and this thesis is among them) aim to increase the signal acceptance. Figure 1.10 shows the acceptance for  $WH$  events in the last CDF analysis performed at CDF[13] with an integrated luminosity of  $1.9 \text{ fb}^{-1}$  collected in the  $WH \rightarrow e\nu b\bar{b}$ ,  $WH \rightarrow \mu\nu b\bar{b}$  channels in the central part of the detector.

CDF II and DØ experiments are getting near to the SM limit on the Higgs production but further refinement and more integrated luminosity is necessary. Figure 1.11 shows the experimental limits found in  $WH$  analysis

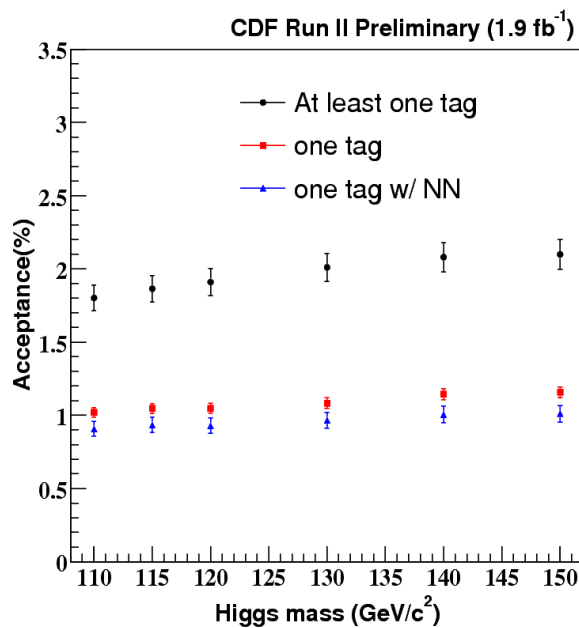


Figure 1.10: Acceptance for  $WH \rightarrow l\nu b\bar{b}$  events with  $l = e, \mu$  in the central region of CDF detector for different  $b$ -tagging categories[13].

on the  $\sigma(p\bar{p} \rightarrow WH) \times BR(H \rightarrow b\bar{b})$ , both at DØ experiment, with an integrated luminosity of  $1.7 \text{ fb}^{-1}$ [12], and at CDF II experiment, with an integrated luminosity of  $1.9 \text{ fb}^{-1}$ [14].

Figure 1.12 displays the upper limit at 95% CL on Higgs production after the combination of all the low mass and high mass channels of the Higgs search both for CDF II and DØ.



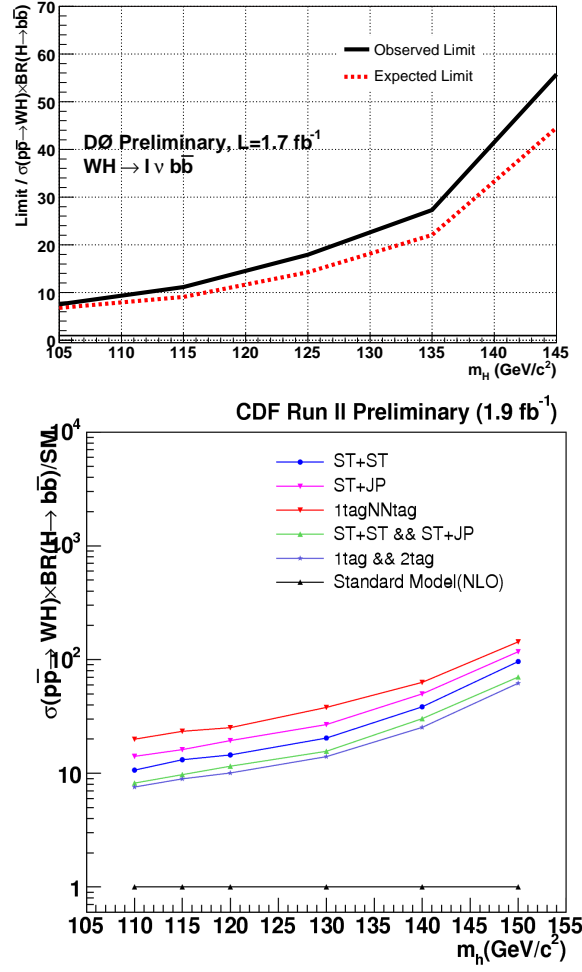


Figure 1.11: Experimental limits on  $\sigma(pp \rightarrow WH) \times BR(H \rightarrow b\bar{b}) / SM_{expect}$  for the DØ experiment with  $1.7 \text{ fb}^{-1}$  (top) and for the CDF II experiment with  $1.9 \text{ fb}^{-1}$  (bottom). CDF II limits (bottom) are quoted for different  $b$ -tagging categories.

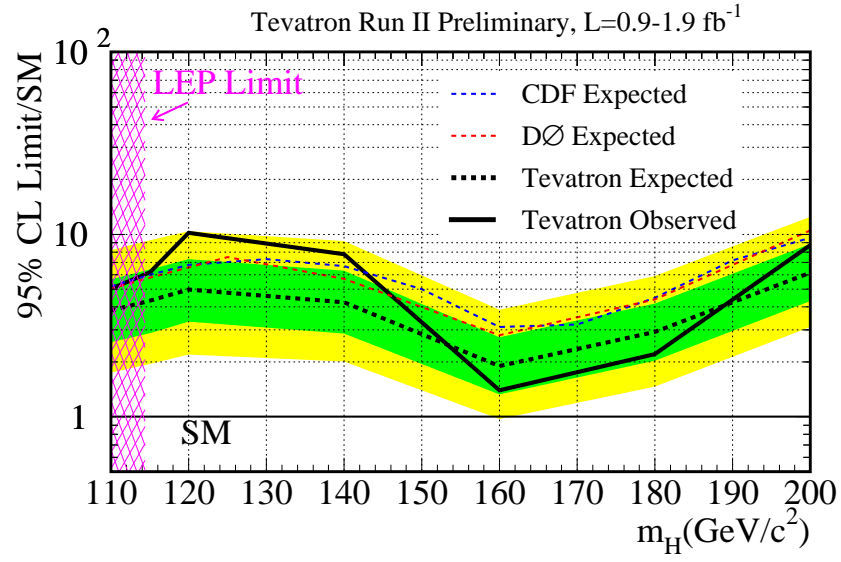


Figure 1.12: 95% CL upper limit on Higgs production at the Tevatron. The limit is obtained combining all the search channels of CDF II and DØ experiments[14].

## Chapter 2

# The Tevatron Collider and the CDF II experiment

The structure of CDF II (Collider Detector at Fermilab for RUN 2 ) experiment is described in this chapter. The three main sections describe: proton-antiproton beams production and acceleration, the CDF detector composition and main characteristics, the trigger and data acquisition system.

### 2.1 The Tevatron

The Tevatron collider is a proton-antiproton storage ring, circular accelerator located at the Fermi National Accelerator Laboratory (FNAL or Fermilab), 50 Km west from Chicago (Illinois, U.S.A.). With a center-of-mass energy of  $\sqrt{s} = 1.96$  TeV, it is the world highest energy accelerator[15] in operation.

Proton-antiproton production and acceleration is a technological challenge and involves the simultaneous operation of several accelerator machines. Figure 2.1 shows a view of Tevatron complex and of its various sections[16].

#### 2.1.1 Proton and Antiproton Production

The first stage, proton extraction and initial acceleration, is done by the PreAccelerator (*PreAc*). Hot hydrogen gas molecules ( $H_2$ ) are splitted by an intense local electrostatic field and charged with two electrons;  $H^-$  ions are accelerated up to 750 KeV by a Cockcroft-Walton accelerator every 66 ms.

PreAc ion source<sup>1</sup> constantly produces beams at 15 Hz rate and send them to the *Linac*: a linear accelerator that increases the ions energy 750 KeV to

---

<sup>1</sup>Actually there are two of them, named H- and I-, working alternatively.

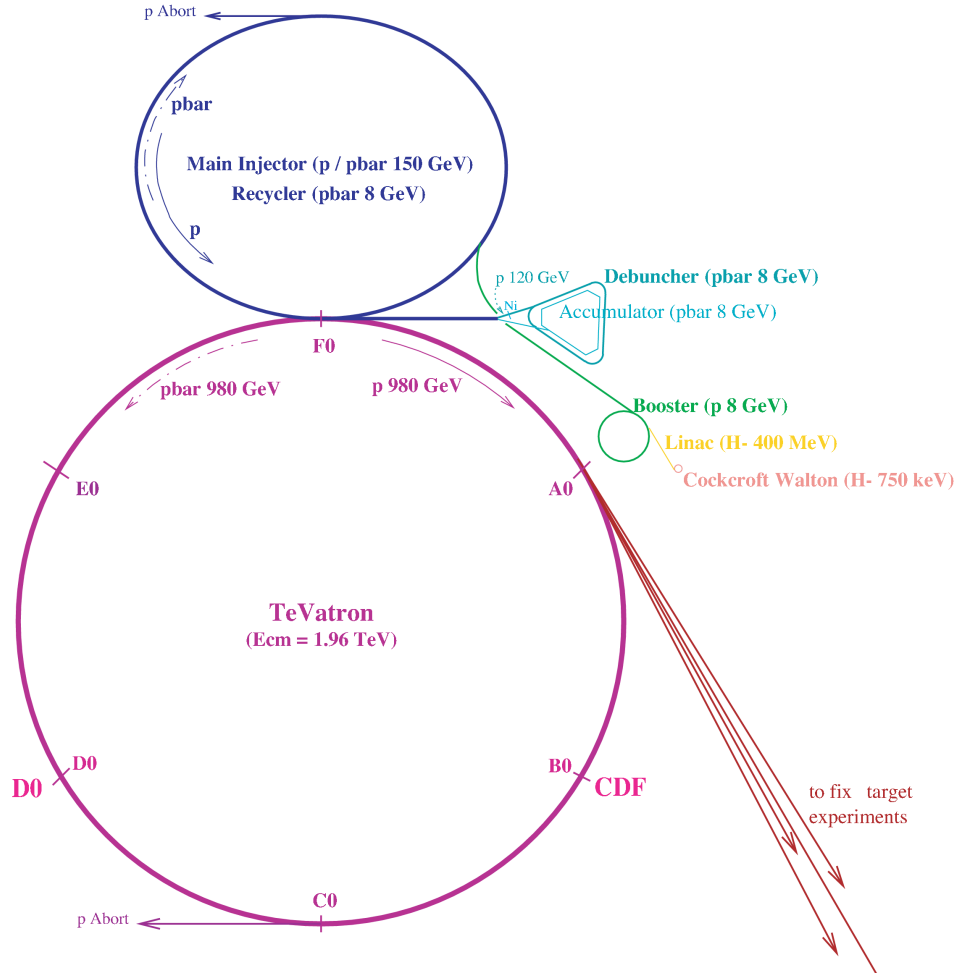


Figure 2.1: Schematic view of Tevatron accelerator complex at Fermilab, different colors mark different accelerator sections.

400 MeV. It is made of two sections: a low energy drift tube and a high energy coupled cavity at the end.

The next acceleration stage is performed by a circular accelerator (synchrotron) of 75 m radius called *Booster*. The insertion of a thin carbon foil strips off electrons from the 400 MeV ions and a sweep, from  $\sim 38$  to  $\sim 53$  MHz in radio-frequency (RF), carries resulting protons to an energy of 8 GeV. The use of negative ions permits injection of more particles from the Linac, otherwise the magnetic field needed to catch the protons would also kick away protons already inside the Booster. Bunches are extracted when about  $8 \cdot 10^{12}$  protons are collected.

Resulting bunches are injected into another larger synchrotron, the *Main Injector*: it has a radius of more than 0.5 Km, 18 accelerating cavities and conventional magnets. It can accelerate protons up to 120 or 150 GeV, depending of the use: 120 GeV protons are used to stack antiprotons, while 150 GeV protons are used to continue the acceleration chain into Tevatron main ring.

The *Antiproton Source* is not an accelerator but it is essential to obtain and store antiprotons. This machine is composed by several parts (see Fig. 2.1): first there is a target station where the 120 GeV protons, extracted from Main Injector, collide with a nickel target and antiprotons of 8 GeV are selected from all the resulting particles. Typically, 21 antiprotons are collected for each  $10^6$  protons on target.

After production, antiprotons have a large spatial and momentum spread while acceleration into Main Injector requires narrow packets. The *Debuncher* is a triangular shape synchrotron, with a mean radius of 90 m, where *stochastic cooling* and *bunch rotation*[17] is applied. Practically the bunch signal is picked up and analyzed from one side of the ring and then it is corrected on the other side.

In the last stage of antiproton production the antiproton beam is sent to the *Accumulator*, a smaller synchrotron (with a mean radius of 75 m) inside Debuncher ring, where antiprotons are stored and some more cooling methods are applied until desired antiproton intensity is reached and antiprotons can be sent to the *Recycler*.

The Recycler is an antiproton accumulator of increased acceptance located in the same tunnel of the Main Injector. It can store many more antiprotons than the Accumulator. Originally it was built to recover antiprotons at the end of a Tevatron run but, at the moment, has the only (very important) function to store antiprotons before the injection in the last accelerator machine. This important new feature boosted Tevatron performance in the last years.

## 2.1.2 Collision and Performance

Last acceleration stage takes place into the *Tevatron*. With a radius of one kilometer this is the largest of the Fermilab accelerators, and, thanks to superconducting magnets, it can store and accelerate  $p\bar{p}$  beams from an energy of 150 GeV (Main Injector result) to 980 GeV.

When beams production and acceleration is complete, a Tevatron *store* is started: 36 protons and 36 antiprotons bunches, containing respectively

$\sim 10^{12}$  and  $\sim 10^{11}$  particles, are injected into the Main Ring at location<sup>2</sup> “F0”. Proton and antiproton bunches are kept  $\sim 0.5$  mm thin and share the same beam pipe, the magnets and vacuum system, separated by 5 mm in two non intersecting orbits. Beam control is obtained through nearly 1000 superconducting magnets 6 m long, cooled to 4.3 K and capable of 4.2 T fields.

Beside energy, the other fundamental parameter of an accelerator is the instantaneous luminosity ( $\mathcal{L}$ ), as the rate of a physical process with cross section *sigma* is:

$$\frac{dN}{dt} = \mathcal{L}\sigma. \quad (2.1)$$

High energy permits an insight to incredibly small scale physics but only very high instantaneous luminosity and very large integrated (in time) luminosity allow to see rare events. Figure 2.2 shows the production cross section of different physical processes. Due to the tiny cross sections we deal with in most of this work, we will be using *picobarns* (pb) where  $1 \text{ pb} = 10^{-36} \text{ cm}^2$ . The instantaneous luminosity of an accelerator, usually measured in  $\text{cm}^{-2}\text{s}^{-1}$ , is given by:

$$\mathcal{L} = \frac{N_p N_{\bar{p}} B f}{4\pi\sigma_x\sigma_y}, \quad (2.2)$$

where  $N_p$  and  $N_{\bar{p}}$  are the number of protons and antiprotons per bunch,  $B$  is the number of bunches inside accelerator,  $f$  is the bunch crossing frequency,  $\sigma_x$  and  $\sigma_y$  are the beam dimensions in the plane transverse to the interaction point. Inside the Tevatron the 36  $p\bar{p}$  bunches have a crossing frequency of 396 ns and, at interaction points, quadrupole magnets focus beams in  $\approx 30 \mu\text{m}$  spots.

While several parameters are (almost) fixed being linked to the accelerators lattice, higher luminosities have been achieved in Run II thanks to the increased antiproton stack rate. At this moment the record instantaneous luminosity is  $\mathcal{L} = 2.85 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}$  obtained on 18 February 2007.

CDF writes on tape about 80% of the total integrated luminosity delivered by the Tevatron (see Fig. 2.3), inefficiency are due to detector calibration or monitoring stores not used to collect physics quality data. On this side, another important record is the  $\approx 40 \text{ pb}^{-1}$  per week integrated luminosity on tape[19].

---

<sup>2</sup>The Tevatron is divided into six sections (see Fig. 2.1) and each junction zone, named from “A” to “F”, has a different function: most important areas are B0, D0 and F0, the first two are experimental areas where CDF and DØ detectors are placed, while F0 contains RF cavity for beam acceleration and switch areas to connect Main Injector and the Tevatron.

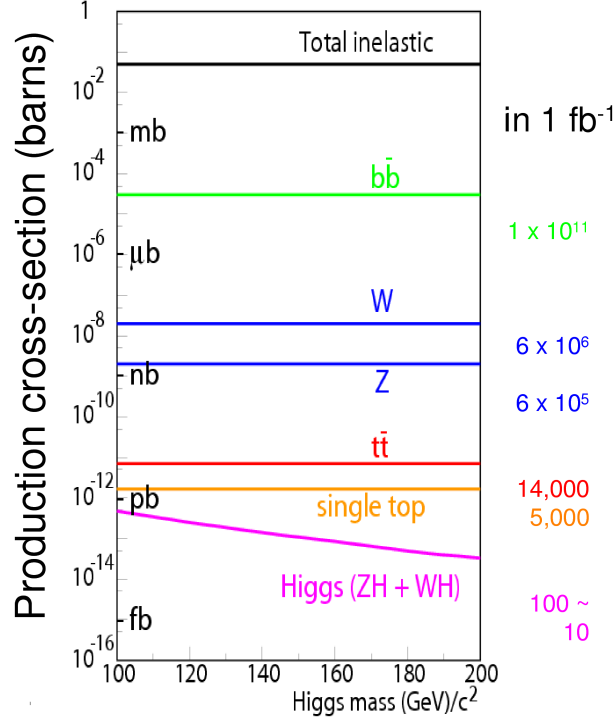


Figure 2.2: Production cross section of physical processes at CDF. The number of expected events in  $1 \text{ fb}^{-1}$  of integrated luminosity is reported on the right  $y$ -axis. Higgs cross section varies with the Higgs mass.

This analysis use data collected up to the end of 2006 that correspond to about  $1.4 \text{ fb}^{-1}$ , a half of the total now on tape.

## 2.2 The CDF Detector

CDF II is a multi-purpose solenoidal detector situated at the BØ interaction point along the Tevatron main accelerator ring. Thanks to charged particle tracking, fast projective calorimetry and fine grained muon detection, the CDF II detector can measure energy, momentum and charge of most particles resulting from  $1.96 \text{ TeV } p\bar{p}$  collisions.

The first original design go back to 1981 but CDF underwent many upgrades during the past twenty years. The last and most extensive upgrade began in 1996 and ended in 2001 when Tevatron *RUN II* started. At the present the CDF II experiment is operated by an international collaboration that embraces more than 60 institutions from 13 different countries, for a

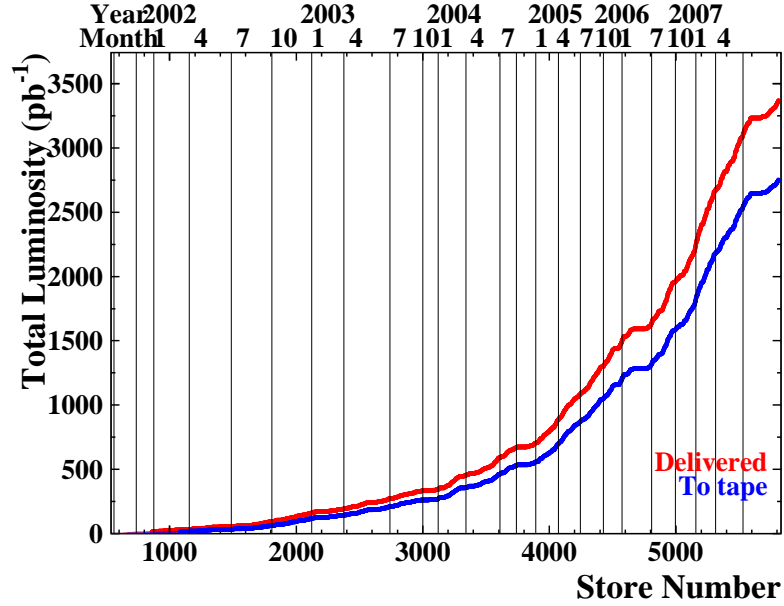


Figure 2.3: Integrated luminosity delivered by the Tevatron (red) and stored by CDF (blue)[18].

total of about 600 researchers.

### 2.2.1 Overview and Coordinate system

The detector is composed by many subsections (subdetectors) for a total of about 5000 tons of metal and electronics, a length of  $\sim 16\text{m}$  and a diameter of  $\sim 12\text{m}$ . The detector is approximately of cylindrical shape with axial and forward-backward symmetry about the  $B\bar{O}$  interaction point. Before going further we describe the coordinate system used at CDF and through this thesis.

$B\bar{O}$  is taken as the origin of CDF right-handed coordinate system:  $x$ -axis is horizontal pointing North<sup>3</sup>,  $y$ -axis is vertical pointing upward and  $z$ -axis is along beam line pointing along proton direction, it identifies “forward” and “backward” regions, respectively at  $z > 0$ , East, and  $z < 0$ , West. Sometimes it is convenient to work in cylindrical  $(r, z, \phi)$  coordinates where the azimuthal angle  $\phi$  is on the  $xy$ -plane and is measured from the  $x$ -axis. The  $xy$ -plane is called *transverse*, quantities projected on it are noted with a T subscript. Two useful variables are the transverse momentum,  $p_T$ , and

<sup>3</sup>Outward with respect to the center of Tevatron.



energy,  $E_T$ , of a particle:

$$\vec{p}_T \equiv p \sin(\theta), \quad E_T \equiv E \sin(\theta). \quad (2.3)$$

Besides Cartesian coordinates, another system is commonly used in collider physics, in it the polar angle  $\theta$  is replaced by *pseudorapidity*:

$$\eta \equiv -\ln[\tan(\theta/2)]. \quad (2.4)$$

If  $(E, \vec{p})$  is the 4-momentum of a particle, the pseudorapidity is the high energy approximation ( $p \gg m$ ) of the *rapidity*:

$$y = \frac{1}{2} \ln \frac{E + p \cos(\theta)}{E - p \cos(\theta)} \xrightarrow{p \gg m} \frac{1}{2} \ln \frac{p + p \cos(\theta)}{p - p \cos(\theta)} = -\ln[\tan(\theta/2)] = \eta. \quad (2.5)$$

A Lorentz boost along the  $\hat{z}$  direction adds a constant  $\ln(\gamma + \gamma\beta)$  to  $y$ , therefore rapidity differences are invariant. In hadronic colliders interactions take place between the (anti)proton constituents which carry only a fraction of the energy of the nucleon so resulting particles usually have momentum imbalances along  $\hat{z}$ . As a result statistical distribution of final state particles is roughly flat in  $y$ , this makes  $\eta$  a good parameter for detector segmentation.

Figure 2.4 shows a 3D cross-section of the CDF detector and of its various subdetectors. The part inside the 1.4T superconducting solenoid contains the integrated tracking system: three silicon subdetectors (the *LayerØØ*, the *Silicon Vertex detector II* and the *Intermediate Silicon Layers*) are the inner core of CDF II. The high resolution capability of silicon microstrips is necessary to have good track resolution near the interaction point, where particle density is higher. Afterward an open cell drift chamber (the *Central Outer Tracker*) covers until  $r \simeq 130\text{cm}$ , in the region  $|\eta| < 1.0$ , the extended lever arm provides very good momentum measurement ( $\Delta P_T/P_T^2 \simeq 10^{-3} \text{ GeV}/c^{-1}$ ).

Calorimeter systems are located outside the superconducting solenoid. They are based on *shower sampling calorimeters* made of sequential layers of high-Z passive absorbers and active signal generator plastic scintillators. The system is composed by towers with  $\eta - \phi$  segmentation, each one divided in electromagnetic and hadronic part, they cover the region up to  $|\eta| \simeq 3.6$  ( $\theta \simeq 3^\circ$ ) and are organized in two main sections: the *Central Calorimeter* covering the region  $|\eta| \lesssim 1.1$  and the *Plug Calorimeter* extending the coverage up to  $|\eta| \simeq 3.6$ . While the central calorimeter is unchanged since 1985, the “Plug” calorimeter active part was completely rebuilt for Run II, replacing gas chambers with plastic scintillator tiles to better cope with the higher luminosity.

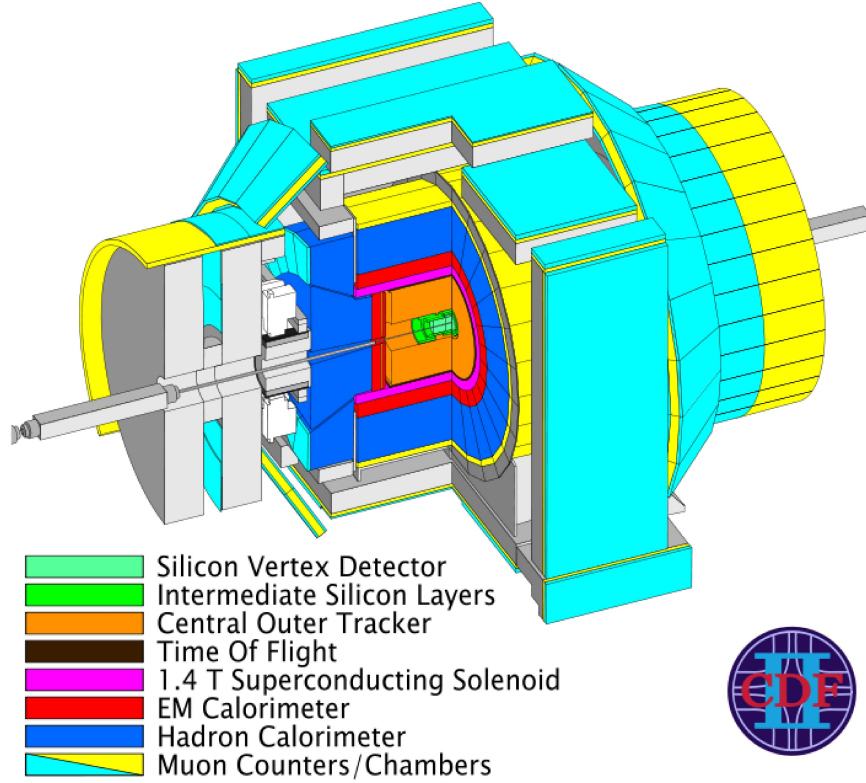


Figure 2.4: 3D cross section of CDF II detector, various subdetectors have different colors.

The outermost part of CDF detector, outside calorimeters, is occupied by the muon detectors. They are multiple layers of drift chambers arranged in various subsections which cover the region  $|\eta| \lesssim 1.5$ . Only high penetrating charged particles, such as muons, can go across the entire detector.

Other detectors are used for a better particle identification, calibration or monitoring. However a detailed description of the entire CDF detector is far from the scope of this work. The next paragraphs will focus on tracking and calorimeter systems which play a significant role in the analysis. A complete description of CDF II detector can be found in[20, 21].

## 2.2.2 Tracking System

The trajectory of a charged particle with non-zero momentum in a uniform magnetic field in vacuum is a helix. A tracking detector identifies (possibly

with minimal perturbation) some points, *hits*, along particle path so that it is possible to obtain momentum measurements by reconstructing the helix parameters<sup>4</sup>. A schematic view of CDF tracking volume can be seen in Figure 2.5: the three main components are the superconducting magnet, the silicon subdetectors and the central drift chamber.

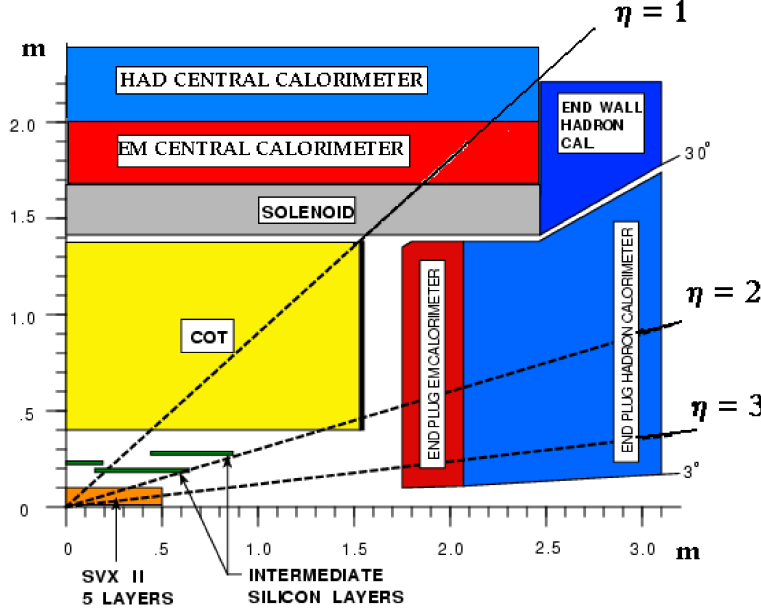


Figure 2.5: View of CDF tracking volume and calorimeter location.

The solenoidal magnet, made by NbTi/Cu superconducting coils, maintains a bending magnetic field with a central value of 1.4116 Tesla, oriented along the positive  $\hat{z}$  direction and nearly uniform in all the tracking volume ( $r \lesssim 150$  cm and  $|z| \lesssim 250$  cm). The momentum threshold for a particle to radially escape the magnetic field is  $p_T \gtrsim 0.3$  GeV/ $c$  and the radial thickness of the coil is 0.85 radiation lengths ( $X_0$ ).

## Silicon System

The silicon system is the first tracking subdetector encountered by particles exiting from the primary interaction vertex. Semiconductor detectors offer excellent spatial resolution and fast response time. Therefore it permits the

<sup>4</sup>See Section 3.1 for track reconstruction details.

reconstruction of secondary vertices displaced from the primary, produced in the decay of long lived  $b$ -hadrons<sup>5</sup>.

CDF employs  $\sim 7 \text{ m}^2$  silicon active-surface for a total of 722,432 different channels read by about 5500 integrated custom chips. The complete silicon tracking detector is displayed in Figure 2.6. Of the three subsystems composing the core of CDF, the *Layer00*[22] (L00) is the innermost. It consists of a single layer of single-sided silicon sensors directly mounted on the beam pipe at radii, alternating in  $\phi$ , of 1.35 cm or 1.62 cm, covering the region  $|z| \lesssim 47 \text{ cm}$ . During the construction of the SVXII microvertex (see below) CDF realized that the multiple scattering due to the presence of read-out electronics and cooling systems installed inside tracking volume was going to degrade the impact parameter resolution. L00 was designed to recover it thanks to its proximity to the beam. Furthermore, being made of state-of-the-art radiation-tolerant sensors, it will ensure a longer operating lifetime to the entire system.

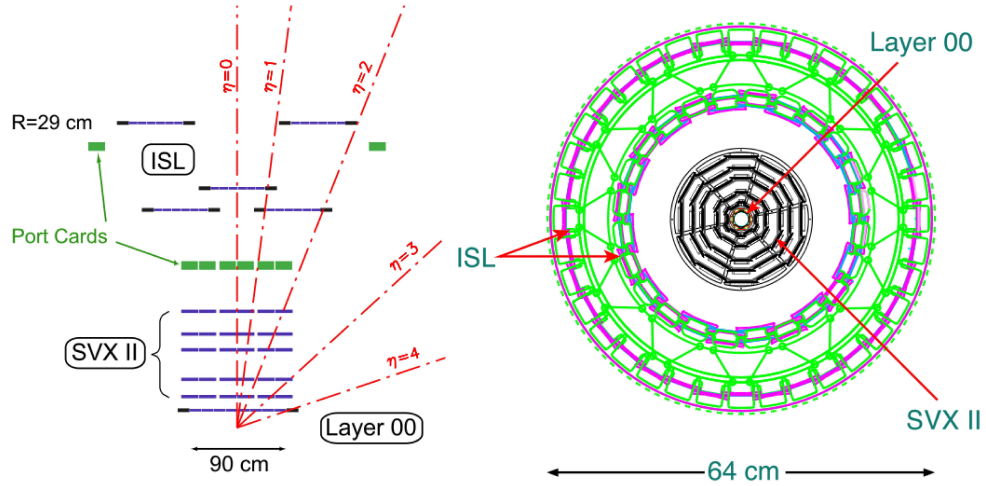


Figure 2.6: Side and front view of silicon tracking system at CDF.

The main component of the silicon system is SVX II[23], the *Silicon Vertex detector* is made of three cylindrical barrels for a total length of about 96 cm along  $z$ , covering the luminosity region until  $\simeq 2.5 \sigma_z$ , and with a pseudo-rapidity range  $|\eta| \lesssim 2$ . Each barrel is divided in twelve identical wedges in  $\phi$ , arranged in five concentric layers between radii 2.4 cm and

<sup>5</sup>Correct identification of  $b$ -hadrons is fundamental in many analyses e.g.  $b$ -hadrons are one of the decay products of  $top$  quark and also Higgs boson has a high branching ratio to  $b$  quark.

10.7 cm. Each layer is divided into independent longitudinal read-out units, called “ladders”. Each ladder consists of a low-mass support for a double-sided silicon microstrip detector. Three out of five layers combine an  $r - \phi$  measurement on one side with  $90^\circ$  stereo measurement on the other, the remaining two layers combine an  $r - \phi$  measure with a small angle  $r - z$  stereo measurement (with tilt angle of  $1.2^\circ$ ). The highly parallel fiber based data acquisition system reads out the entire subdetector in approximately  $10 \mu s$ .

ISL[24] (*Intermediate Silicon Layers*) is the outermost of the three silicon subdetectors, radially located between SVX II and the drift chamber covering the region  $|\eta| \lesssim 2$ . It is divided in three barrels segmented into  $\phi$  wedges. The central barrel ( $|\eta| \lesssim 1$ ) is made of one layer of silicon sensors at radius of 22 cm, instead the two outer barrels ( $1 \lesssim |\eta| \lesssim 2$ ) are made of two layers at radii of 20 cm and 28 cm. Its purpose is to strengthen the CDF tracking in the central region and to add precision hits in a region not fully covered by the drift chamber. This improves the tracking at large  $\eta$  and also allows silicon stand alone track reconstruction in the whole region  $|\eta| < 2$ .

The complete silicon subdetector (LØØ, SVXII and ISL) has an asymptotic resolution of  $40 \mu m$  in impact parameter and of  $70 \mu m$  along  $z$  direction. The total amount of material varies roughly as:

$$\frac{0.1X_0}{\sin(\theta)} \quad (2.6)$$

in the central region and doubles in the forward region because of the presence of read-out electronics, cooling system and support frames[25].

### 2.2.3 Central Outer Tracker

The *Central Outer Tracker*[26] (COT) is an open-cell drift chamber used for charged particles tracking at large radii. It has an hollow-cylindrical geometry and covers  $43.3 < r < 132.3$  cm,  $|z| \lesssim 155$  cm. Figure 2.5 shows that COT fully covers the central region ( $|\eta| \lesssim 1$ ) with some residual capability up to  $|\eta| \approx 2$ .

The COT (see Fig. 2.7) is structured into eight “super-layers” each divided into  $\phi$  cells; each cell contains twelve sampling wires, spaced 0.583 cm, to collect the ions produced by passing charged particles. The arrangement of the cells has a  $\chi = 35^\circ$  tilt with respect to the chamber radius to partially compensate the Lorentz angle of the electrons drifting in the magnetic field and obtain the best resolution<sup>6</sup>.

---

<sup>6</sup>Electrons drifting in a gas within an electromagnetic field ( $\vec{E}, \vec{B}$ ) move with an angle

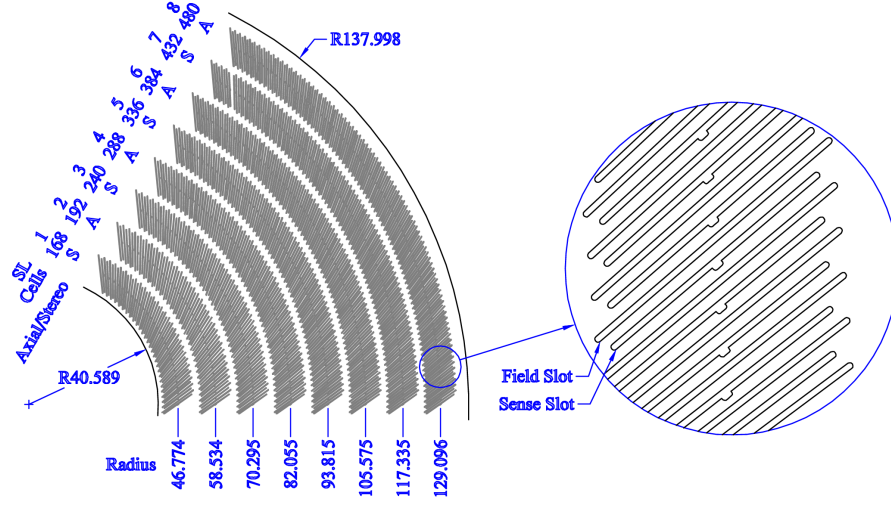


Figure 2.7: A 1/6 section of the COT end-plate with the eight super-layers structure and the location of cell slots.

The final structure has 8x12 sampling planes alternated with planes of potential wires (see Fig. 2.8), 96 hits are measured for a particle crossing the entire COT ( $|\eta| < 1$ ). Four super-layers employ sense-wires parallel to the beam axis for measurements in  $r - \phi$  plane, the other four interspersed super-layers are called “stereo super-layers” because their wires are alternately canted at angles of  $+2^\circ$  and  $-2^\circ$  with respect to the beam line and are used to measure  $r - z$  coordinates. The applied electric drift field (see Fig. 2.8) is 1.9 kV/cm. A 50 : 50 gas admixture of argon and ethane bubbled through isopropyl alcohol (1.7%) constantly flows in the chamber volume. The obtained drift velocity is about 100  $\mu\text{m}/\text{cm}$  for a maximum drift space of 0.88 cm. The material of the COT is about  $0.017X_0$ , mostly concentrated in the inner and outer shell.

## 2.2.4 Calorimeter System

Located immediately outside the solenoid, the calorimeter system at CDF covers a solid angle of nearly  $4\pi$  around  $p\bar{p}$  interaction point and has the fundamental role to measure energies of electrons, photons, particle clusters (*jets*) and the imbalance in transverse energy flow (signature of *neutrinos*).

---

$\chi \simeq \arctan\left(\frac{v(E, B=0)B}{kE}\right)$ , where  $k$  is empirical parameter of gas and electric field and  $v(E, B = 0)$  is the velocity without the magnetic field. The angle  $\chi$  is also known as Lorentz angle.

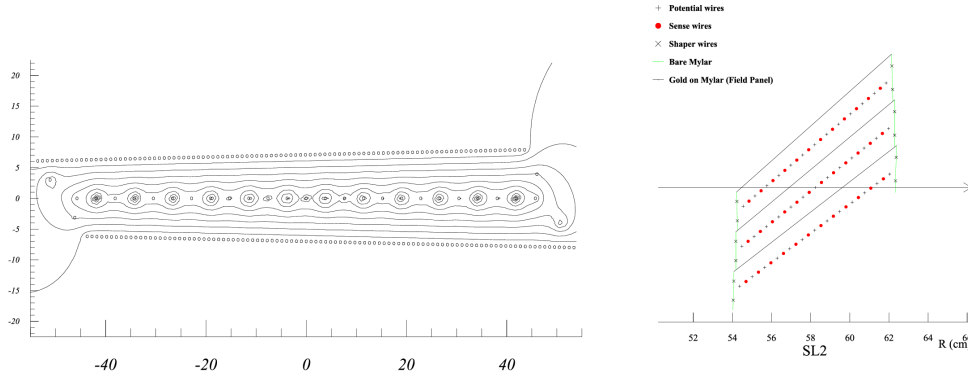


Figure 2.8: Left: equipotential line inside one of the COT super-layer cell. Right: layout of sense-wires, field-wires and shaper-wires inside one COT cell.

	CEM	CHA, WHA	PEM	PHA
Energy Resolution	$14\%/\sqrt{E}$	$75\%/\sqrt{E}$	$16\%/\sqrt{E}$	$80\%/\sqrt{E}$
Angular Coverage	$ \eta  < 1.1$	$ \eta  < 1.3$	$1.1 <  \eta  < 3.6$	$1.3 <  \eta  < 3.6$
Absorber	lead	iron	lead	iron
Longitudinal Depth	$19X_0, 1\lambda$	$4.5\lambda$	$21X_0, 1\lambda$	$7\lambda$

Table 2.1: Main characteristics of CDF calorimeter system.

The location of calorimeter sections is visible in Figure 2.5. Both “Plug” and “Central” are sampling calorimeters divided in an electromagnetic section (lead/scintillator), optimized to collect all the energy of electrons and photons, and a subsequent hadronic section (iron/scintillator), thicker and optimized for hadron energy measurement. Calorimeters have an in-depth segmentation, finer near the collision point and coarser outward. The  $\eta - \phi$  plane is mapped in rectangular cells, each one corresponding to the independent read-out of a projective electromagnetic or hadronic tower. Thanks to the fast response of scintillators, signals from calorimeters are quickly processed and used at various trigger levels. Following paragraphs explains in more detail the composition of the different subsections and Table 2.1 summarizes their main characteristics.

## Central Calorimeter

The central region of the detector is covered by the *Central Electromagnetic* (CEM) and the *Central HAdronic* (CHA) calorimeters[27], corresponding to the pseudo-rapidity region  $|\eta| < 1.1$  and  $|\eta| < 0.9$  respectively.

The CEM is a hollow cylinder located at  $173 < r < 208$  cm, divided in four  $180^\circ$  arches each composed by 12 azimuthal sections ( $\Delta\phi = 15^\circ$ ) and 10 pseudo-rapidity sections ( $\Delta\eta \simeq 0.11$ ) for a total of 478 instrumented towers<sup>7</sup>.

The CHA covers region  $|\eta| < 0.9$  and it is divided into  $9 \times 12$   $\eta - \phi$  towers corresponding to CEM segmentation for a total of 384 towers. Central hadronic calorimeter covering is extended up to  $|\eta| \simeq 1.3$  thanks to the *end-Wall HAdron Calorimeter*[28] (WHA). It has same  $\phi$  segmentation and six additional  $\eta$  towers: the first three overlap CHA and the last three extend  $\eta$  coverage.

Figure 2.9 shows a wedge of the central calorimeter system. Each CEM sector is a sampling device made of 31 layers of polystyrene scintillator (5 mm thick) radially alternated with layers of aluminum-clad lead (3.18 mm thick). Some of the 30 lead layers are replaced by acrylic (Plexiglas) as a function of  $\theta$  to maintain a uniform thickness in  $X_0$ . As particles loose energy into

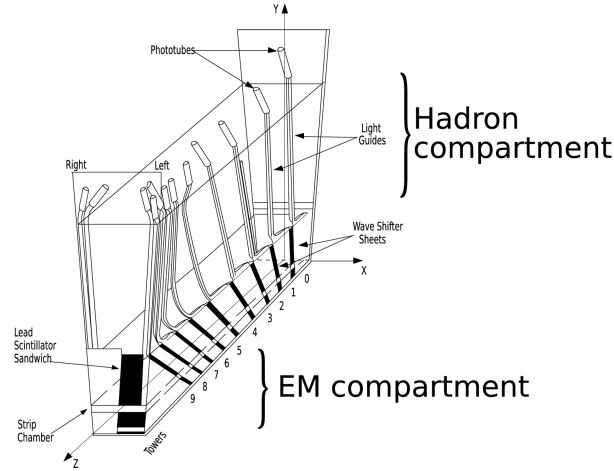


Figure 2.9: Structure of a wedge of CDF central calorimeter.

the absorber medium, the blue light emitted by active scintillator medium is collected by thin bars of blue-to-green wave-length shifter acrylic material placed on the sides of each tower that bring the light to two photomultiplier tubes (PMT) outside CHA. CEM contains also the *Central Electron Strip*

<sup>7</sup>Two towers are missing to permit access to the solenoid, the so-called “chimney”.



*chambers* (CES) and the *Central Preshower detector* (CPR). CES is a multi-wire proportional chamber placed at a radial depth of  $\sim 6X_0$  and is used to determine shower position and transverse shower development with an accuracy of  $\sim 0.2$  cm. CPR is a layer of finely segmented scintillators located immediately outside the solenoid and is used to monitor photon conversion started in the tracking region.

The structure of hadronic calorimeters is similar to electromagnetic ones but absorber materials are 32 steel, 2.5 cm-thick, layers in CHA and 15 steel, 5.1 cm thick, layers in WHA both alternated with acrylic scintillator, 1.0 cm thick.

The total thickness of electromagnetic section is nearly uniform and corresponds to 19 radiation lengths ( $X_0$ ) or 1 interaction length ( $\lambda_{int}$ ). Based on test beam data, the CEM energy resolution for an electron going through the center of a tower is found to be:

$$\frac{\sigma_E}{E} = \frac{13.5\%}{\sqrt{E(\text{GeV})}} \oplus 2\%. \quad (2.7)$$

The total thickness of hadronic section is  $\sim 4.5\lambda_{int}$  and the energy resolution is:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E(\text{GeV})}} \oplus 3\%, \quad \frac{\sigma_E}{E} = \frac{75\%}{\sqrt{E(\text{GeV})}} \oplus 4\%. \quad (2.8)$$

respectively for CHA and WHA.

## Forward Calorimeter

*Plug calorimeters*[29] are two identical structures, East and West, covering region  $1.1 \lesssim |\eta| \lesssim 3.6$ . Figure 2.10 shows the structure of plug calorimeters, in a way similar to the central device: there is a *Plug ElectroMagnetic calorimeter* section (PEM), a *Plug PReshower* (PPR) detector before the calorimeter, a *Plug Electromagnetic Shower-maximum* detector (PES) embedded (at  $6X_0$ ) and a subsequent *Plug HAdronic calorimeter* section (PHA).

Electromagnetic section is  $21X_0$  thick and is composed by 23 annular plates, of 2.77 m outer diameter and an inner hole for the beam pipe made of 4.5 mm thick lead absorber. Towers have a segmentation with varying  $\Delta\eta$  and  $\Delta\phi$  as Table 2.2 shows, with an azimuthal-angle covering of  $7.5^\circ$  down to  $\eta = 2.11$  and of  $15^\circ$  further. Active elements are 4 mm thick scintillator tiles read-out by embedded wavelength shifters connected to PMT. All is assembled in triangular shape *pizza-pans* which enclose a slice of a  $\Delta\phi = 15^\circ$  sector. Two layers are different: first scintillator layer is 10 mm thick and is used as a preshower detector, PES layer is made by two strips of scintillators that provide shower maximum position measurement with  $\sim 1$  mm accuracy.

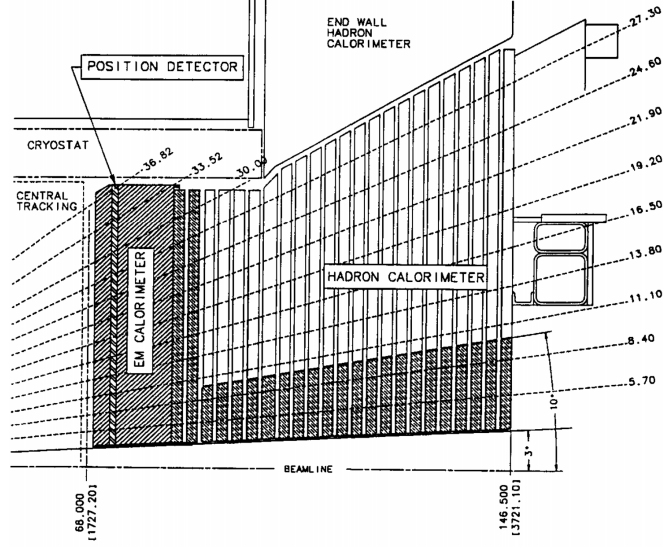


Figure 2.10: Elevation view of one quarter of the CDF Plug calorimeter.

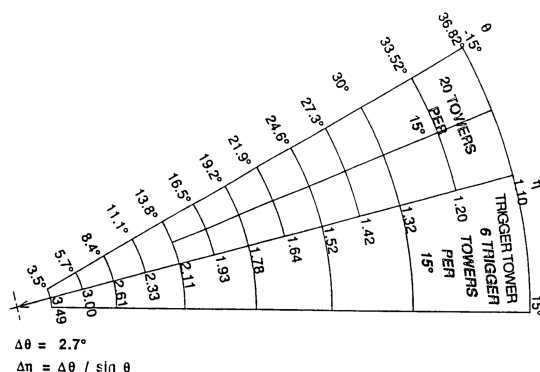
$ \eta $ Range	$\Delta\phi$	$\Delta\eta$
0. – 1.1(1.2H)	$15^\circ$	$\sim 0.1$
1.1(1.2H) – 1.8	$7.5^\circ$	$\sim 0.1$
1.8 – 2.1	$7.5^\circ$	$\sim 0.16$
2.1 – 3.64	$15^\circ$	0.2 – 0.6

Table 2.2: CDF II calorimeter segmentation,  $H$  stands for the hadronic section.

Hadronic section is about  $7\lambda_{int}$  thick and segmented in  $\Delta\phi = 30^\circ$  for a total of 12 sections of 23 iron 5 cm-thick layers alternated with 6 mm scintillator active material layers. The characteristic “plug” shape is due to the growing radii of the layers far from interaction point to match WHA coverage. Energy resolution is:

$$\frac{\sigma_E}{E} = \frac{16\%}{\sqrt{E(\text{GeV})}} \oplus 1\%, \quad \frac{\sigma_E}{E} = \frac{74\%}{\sqrt{E(\text{GeV})}} \oplus 4\%. \quad (2.9)$$

respectively for PEM and PHA. Figure 2.11 shows the segmentation of a  $\Delta\phi = 15^\circ$  sector and describes the distribution of trigger towers.



### 2.2.5 Other Detectors

This analysis is based on reconstructed jets and electrons which mainly use calorimeter and tracking capability of the CDF detector, however other sub-detectors data are used indirectly or play a key role in other important researches made at CDF. The following paragraphs make an overview of the main aspects of the remaining components.

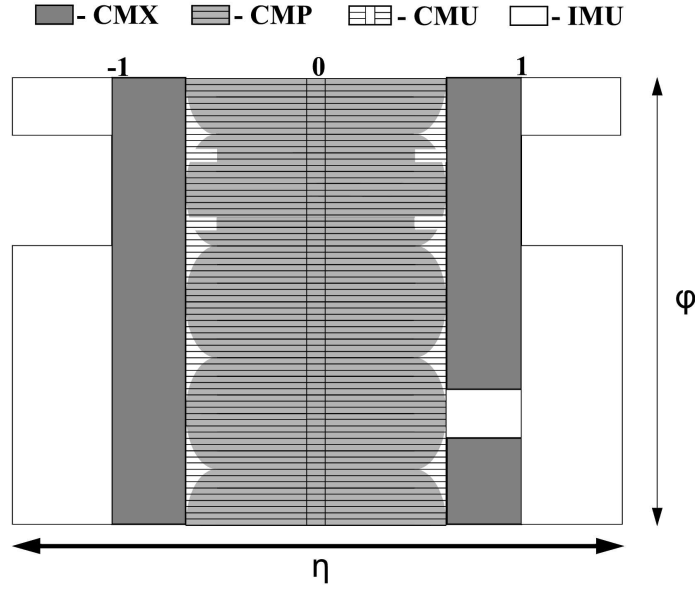
## Muon Detectors

Muons lose very little of their energy into the tracking volume and in the calorimeter system<sup>8</sup> so only they, except neutrinos and particles going into the “cracks” of the detector, pass throughout.

Muon subsystems[30] are organized in four subdetectors: the *Central MUon detector* (CMU) covers region  $|\eta| \lesssim 0.6$ , the *Central Muon Upgrade* (CMP) is placed over CMU and over a  $\sim 2.4\lambda_{int}$  steel shielding to reduce hadrons escaping from CHA, the *Central Muon eXtention* (CMX) covers region  $0.6 < |\eta| < 1.0$  and the *Intermediate MUon* system (IMU) reaches  $|\eta| \simeq 1.5$ . The  $\eta - \phi$  coverage and subdetectors disposition is summarized in Figure 2.12.

Each muon subsystem is composed by arrays of azimuthal, single wire, rectangular, drift chambers overlaid in different patterns (to cancel hits position ambiguities) and coupled with scintillators (to grant fast timing measurements). Single hit resolution is of  $\sim 0.25$  mm in  $\eta - \phi$  plane, thanks

<sup>8</sup>Muons from  $Z^0$  decays, for instance, deposit on average about 0.4 GeV in the electromagnetic portion of the calorimeter and 4 GeV in the hadronic one.

Figure 2.12:  $\eta - \phi$  coverage of the CDF muon system.

to the measurements of the difference of drift electron arrival times between neighboring cells, and of  $\sim 1.2$  mm along  $z$  thanks to the measurements of the different charge collected at the end of each wire. When a track-segment results from three matching radial layers (a *stub*) and a corresponding COT track point outward to it, a muon candidate is identified.

## Cherenkov Luminosity Counters

The *Cerenkov Luminosity Counters*[31] (CLC) are two symmetrical detector modules designed to measure the instantaneous luminosity through the rate of inelastic processes of known cross section in the forward region.

Each counter is made of 48 conical, gaseous, Cerenkov counters posed inside each Plug Calorimeter in a high pseudo-rapidity region ( $3.7 < |\eta| < 4.7$ ) and pointing to the nominal interaction region. Cones are disposed in concentric manner with the smaller counters at the center (length 110 cm, initial diameter 2 cm) and the bigger outward (length 180 cm, initial diameter 6 cm). The narrow shape and orientation is optimal to collect particles outgoing from the interaction point that produce an important Cerenkov light yield. On the other hand particles from beam halo or from secondary interactions have larger crossing angle, hence they produce a much smaller signal. Time coincidence between two modules (East and West) is allowed by the

good time resolution (less than 100 ps) and is an additional tool to remove background interactions. Figure 2.13 shows the time distribution of hits on the two modules.

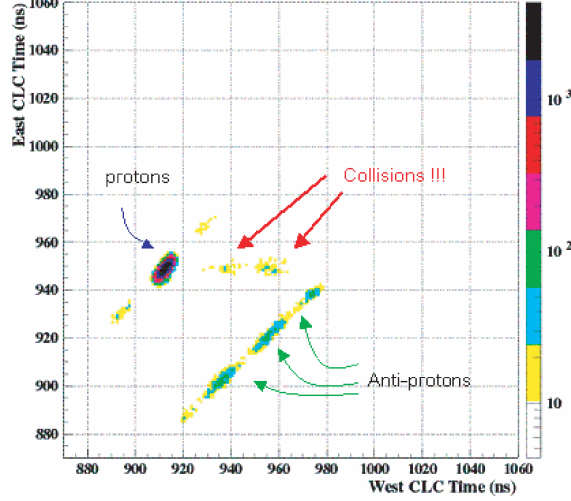


Figure 2.13: Time distribution of East and West CLC modules signals. A  $p\bar{p}$  collision deposits a coincidence signal in the two modules.

Different signal shapes on CLC can measure the average number of inelastic interactions per bunch crossing ( $\bar{N}$ ) with an uncertainty of  $\sim 4\%$ [32], the instantaneous luminosity  $\mathcal{L}$  is inferred from the relation:

$$\bar{N}f_{b.c.} = \mathcal{L}\sigma_{in}\epsilon, \quad (2.10)$$

where the bunch crossing frequency ( $f_{b.c.}$ ) is precisely known from the Tevatron RF,  $\epsilon$  is the CLC acceptance for inelastic scattering and  $\sigma_{in}$  is the inelastic  $p\bar{p}$  cross section resulting from the averaged CDF and E811 luminosity independent measurements at  $\sqrt{s} = 1.80$  TeV[33], and extrapolated to  $\sqrt{s} = 1.96$  TeV:

$$\sigma_{in}(1.80\text{TeV}) = 60.4 \pm 2.3\text{mb} \rightarrow \sigma_{in}(1.96\text{TeV}) = 61.7 \pm 2.4\text{mb}. \quad (2.11)$$

The combined uncertainty of different factors of Eq. 2.10 results in a total  $\simeq 6\%$  error on instantaneous luminosity values.

## 2.3 Trigger and Data Handling

Trigger purpose is the on-line selection of highly interesting physics events from the overwhelming background.

At B $\bar{O}$  the bunch crossing frequency is 2.5 MHz and the inelastic  $p\bar{p}$  cross section is  $\sigma_{in} \simeq 60$  mb. There are about 1–2 inelastic collision in each bunch crossing with an instantaneous luminosity of  $\mathcal{L} \simeq 10^{32} \text{cm}^{-2} \text{s}^{-1}$ . Storing the entire detector information for each event would be impossible, because maximum recording rate is  $50 \div 100$  Hz, and useless, because interesting processes have much smaller cross section than the inelastic  $p\bar{p}$  cross section (Higgs production cross section is  $\mathcal{O}(10^{-9} \text{mb})$ , see Fig. 1.8).

CDF trigger is composed by three levels, (L1, L2 and L3, see Fig. 2.14), each one provides rate reduction sufficient to allow the processing in the next level with virtually no dead-time<sup>9</sup>.

Each level bases its decision on a set of programmable conditions, step by step more complex as detector read-out and data elaboration becomes available. The so-called *trigger path* is the logic combination of criteria from different levels. All trigger paths are arranged in a *trigger table* optimized to get best acceptance for interesting events and a maximum acquisition dead time of 5%.

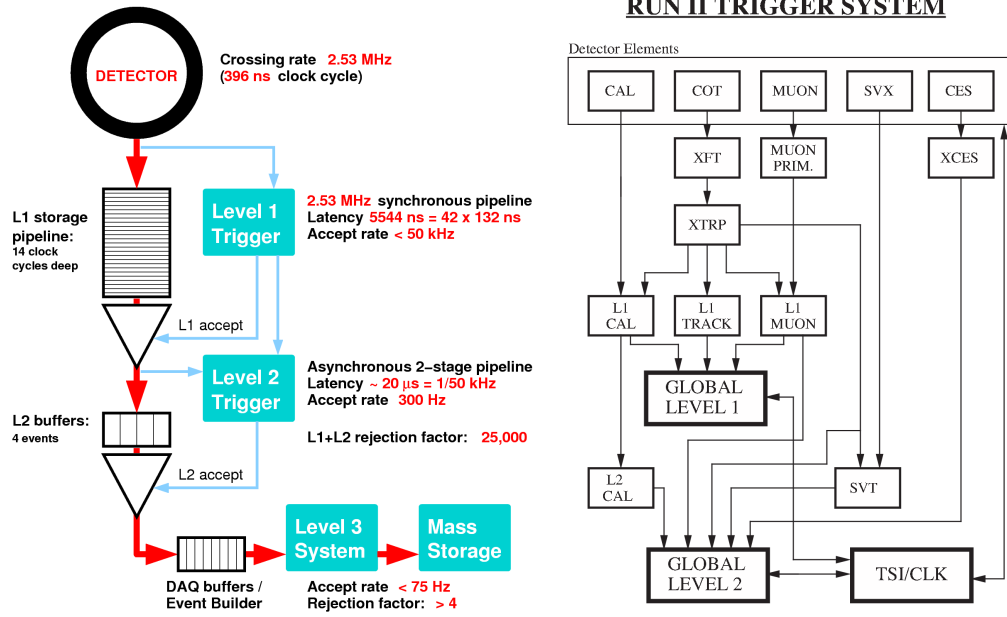


Figure 2.14: The CDF II trigger system. Left: block diagram of three-levels trigger and DAQ system. Right: L1 and L2 trigger data streams.

<sup>9</sup>I.e. *dead-time* occurs when events must be rejected because trigger system is occupied processing a preceding event.

**Level-1** is a custom designed hardware system composed by three parallel streams which feed inputs to Global Level-1 decision unit (see Fig. 2.14). One stream collects prompt calorimetric response: it is divided into “object triggers”, i.e. single hadronic or electromagnetic deposits above threshold, and “global triggers”, i.e. total event transverse energy ( $\sum E_T$ ) or the presence of raw missing transverse energy (MET or  $\cancel{E}_T$ )<sup>10</sup>. A second stream collects information from muon chambers thus identifying muon stubs. The last stream comes from the eXtreme Fast Tracker (XFT), a powerful parallel pattern recognition algorithm used to feed COT raw tracking information to L1 trigger. This raw and fast information permits to reconstruct candidate physical objects: a track plus matched EM deposit is an electron candidate, a track plus a matched stub is a muon candidate, MET may stand for a neutrino, a hadronic cluster in the calorimeter can be a jet and so on. L1 has 42 pipelined buffers synchronized with Tevatron clock cycles (132 ns) and used to store events while trigger takes the decision. L1 is where the first drastic rate reduction in accepted events takes place: it has  $132 \cdot 42 \simeq 5.5 \mu\text{s}$  latency time and feeds events to next level with a rate  $< 50 \text{ KHz}$ .

**Level-2** can use information collected by shower-maximum detector to ensure a better electron recognition with azimuthal information. Furthermore it can use crude reclustering algorithms to merge more towers together and calculate better jet information. Furthermore silicon SVX is read-out and so can reconstruct tracks with enough precision (thanks to the SVT<sup>11</sup>) that the impact parameter resolution is similar to what obtained off-line. L2 has only four memory buffers so the total latency time is  $5.5 \cdot 4 \simeq 21 \mu\text{s}$ , the accept rate for Level-3 is 300 Hz.

**Level-3** is last stage of the trigger system, it is made of a farm of several hundreds commercial computers running LINUX and organized in a modular and parallel structure. Whole detector information is available as well as the raw objects reconstructed by previous levels. The *EVent Builder*[34] (EVB) assembles the different fragments into a block of data corresponding to a unique bunch crossing, the *event record*. Level 3 farm uses selection criteria based on a simplified version of the offline algorithms. Therefore a precision close to the offline analysis is obtained. While events accepted are transferred for mass storage, a fraction of the output is used as a real time

---

<sup>10</sup>Missing transverse energy is defined as  $\vec{\cancel{E}}_T \equiv -\sum_i \vec{E}_T^i$ , see Section 3.5.

<sup>11</sup>I.e it reveals displaced secondary vertex, a very important tool to select events containing beauty hadrons.

monitoring.

A collection of events of similar provenience or common physical interest forms a “dataset”, for instance the dataset `bhmu` collects all events coming from the high  $p_T$  central muon triggers, or the dataset `bpe1` contains events with candidate high  $p_T$  electron in the plug region. Two more characters are added at the end of the dataset name, they point out the data taking period, for instance period `0d` corresponds to data collected from 2002 to 2004. Many parameters can vary between different data periods because extensive maintenance of the detector or machinery upgrades can occur.

### 2.3.1 CDF Software Framework

The CDF experiment uses a specifically developed object-oriented software to manage data acquisition, offline event reconstruction and most of the analysis. The C++ program language is used to define *objects*, i.e self-consisting structures containing properties and definition common to all entries of the same kind. Usually an object is a physics interesting quantity. It can be simple, as a track reconstructed with a particular algorithm or the energy deposit in the calorimeters, or more complex, like a “jet object”, that will contain links to tracks, calorimeter energy deposits, vertex information and much more.

All the objects characterizing a reconstructed event are stored into large arrays (“n-tuples”) suitable for high level analysis. Usually it is performed with “ROOT”[35], an object oriented data analysis framework.

The different objects are collected in AC++ *modules* that make easier to do parallel studies and upgrades, however all research starts from a common stable framework. When a large set of new features is available, a new stable “Generation” (or “Gen”) of the CDF software is released. Actually CDF collaboration is using *Gen6* but *Gen7* is near to be fully tested.

The release of a new Generation improves several aspects of CDF analysis because, as time goes, new algorithms are implemented and a better knowledge of the detector is obtained. Anyway all the data acquired until the date of the new release as well as all the Monte Carlos need to be reprocessed with the new software and some older analysis may need adjustments because of backward compatibility problems. This is called “production” and is a time and CPU consuming task.



## 2.4 Monte Carlo Simulation

An accurate simulation of physics events and of the expected detector response are of fundamental importance in many analyses. The Monte Carlo samples of physics processes used along this work are produced with a three step approach.

**Parton level:** the process is calculated at parton level, the matrix element is derived with numerical integration of leading-order or next-to-leading order equation. `ALPGEN`<sup>12</sup>[36] is used for many physics processes in this thesis. In order to reproduce the final state, it is interfaced to `PYHTIA`<sup>13</sup>[37] (see below). `PYHTIA` is also used as generator for some physics processes ( $t\bar{t}$  production etc). For this studies `PYHTIA` was used to produce the signal sample ( $WH$ ,  $W \rightarrow e\nu$ ,  $H \rightarrow b\bar{b}$ ). At CDF it is also possible to use `HERWIG`[38] to produce physics processes.

**Particles level:** higher order or non-perturbative processes are simulated on the basis of analytical results and QCD models. Parton shower programs like `PYHTIA` and `HERWIG` can be used to simulate quark hadronization, soft gluon emission or minimum-bias events inside a process, i.e. secondary soft interactions or spectator quark interactions.

**Detector level:** the last step is the detector simulation. The `GEANT3`<sup>14</sup>[39] simulator is an AC++ module composed by C++ objects that simulate each subdetector. The simulator receives in input the positions, the four-momenta, and the identities of all particles produced by the simulated collisions that have long enough lifetimes to exit the beam pipe. The detector simulation is tuned to test-beam data, and, most important, it has access to the run database, therefore run-dependent conditions like dead detector channels, beam alignment, beam luminosity etc can be simulated in a run-dependent way. Therefore for each physics process a definite number of events/pb<sup>-1</sup> is produced and passed through simulation.

---

<sup>12</sup>The version used is 2.1

<sup>13</sup>The version used is 6.326

<sup>14</sup>The version used is V.3.15.



## Chapter 3

# Physical Objects Reconstruction

This analysis is centered on the search of the *Higgs* scalar boson decaying into two *b*-jets and produced in association with a *W* vector boson decaying in the channel  $W^\pm \rightarrow e^\pm \nu_e$  with the electron identified in the forward region of the CDF detector. Therefore the final state is composed of an electron, a neutrino and two *b*-jets. Each of the four final state objects is reconstructed combining data of several subdetectors, in the followings we explain some of the methodologies used.

Electron candidates are reconstructed as an energy cluster in the electromagnetic calorimeter matched with a reconstructed track. In this work we use the standard CDF clustering procedure while the track reconstruction algorithms contain many features used here for the first time. Indeed most of this work is devoted to the comparison between track reconstruction in *Generation 6* and *Generation 7*<sup>1</sup>, pointing out the increased acceptance in the forward tracking.

Jets are the experimental signature of high momentum quark and gluon production that hadronize in a collimated shower of particles. They appear as a (clustered) deposit of energy in both electromagnetic and hadronic calorimeters.

Heavy flavor quarks (particularly *b*-quarks) hadronize to meta-stable particles that can travel a distance away from the primary interaction vertex. This characteristic is used by the Secondary Vertex (**SecVtx**) tagging algorithm to mark the difference between light and heavy flavor jets.

Neutrinos are found by looking for transverse energy imbalance (missing energy,  $\cancel{E}_T$ ) as measured by the whole CDF calorimeter system.

---

<sup>1</sup>Gen6 is actually the official CDF software release while Gen7 release used here is the development version “6.1.4int6”, available since spring 2007.

### 3.1 Tracks Reconstruction

The ability to detect and reconstruct charged particle trajectories is essential for particle identification and momentum reconstruction. Precise, high efficiency tracking plays a central role for lepton and photon identification, the latter, being neutral, has no track associated. Moreover track reconstruction allows precise measurement of the track impact parameter, and thus the identification of heavy-flavor quarks in jets.

At CDF the following five parameters are used to describe the helix trajectory that is drawn by a charged particle traveling inside a uniform magnetic field  $\vec{B}$  (see Fig. 3.1):

- $C$ : the half-curvature of the trajectory,  $C \equiv 1/2qr$  with  $r$  equal to the helix radius and  $q$  the measured charge of the particle. It has the same sign of the particle charge and it is related to the transverse momentum of the track:

$$p_T = \frac{cB}{2|C|} \quad (3.1)$$

- $d_0$ : the impact parameter, i.e. the distance of closest approach in the transverse plane between the helix and the origin. It has the opposite sign of the angular momentum of the track with respect to the  $z$ -axis and it is defined as (see Fig 3.1):

$$d_0 = q \left( \sqrt{x_0^2 + y_0^2} - r \right), \quad (3.2)$$

where  $x_0$  and  $y_0$  are the coordinates of the center obtained by the projection of the helix on the transverse plane and  $r = q/2qC$ . The quality of  $d_0$  measurement is often parametrized by the “impact parameter significance” defined as  $|d_0/\sigma_{d_0}|$ .

- $\lambda$ : the helix pitch, i.e. the cotangent of the polar angle between the track and the  $z$ -axis ( $\cot \theta_0$ ). The longitudinal component of the momentum is given by:

$$p_z = p_T \cot \theta_0. \quad (3.3)$$

- $z_0$ : the  $z$  position of the track vertex.
- $\phi_0$ : the azimuthal angle of the track at its vertex.

The helix is completely described by these five parameters, indeed every point along the trajectory satisfies the following equations[42]:

$$x = r \sin \phi - (r - d_0) \sin \phi_0, \quad (3.4)$$

$$y = -r \cos \phi + (r - d_0) \cos \phi_0, \quad (3.5)$$

$$z = z_0 + s\lambda, \quad (3.6)$$

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

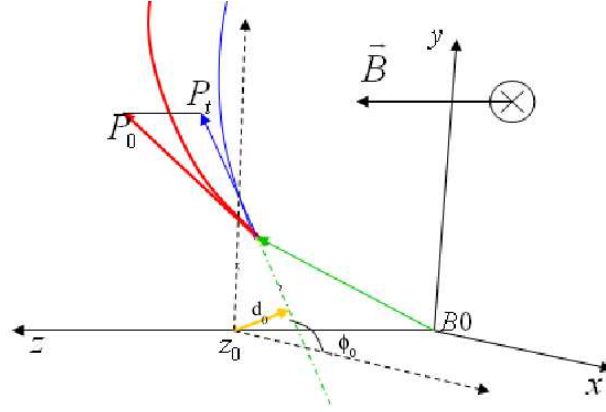


Figure 3.1: CDF track parameters and coordinate system.

### 3.1.1 Tracking Algorithms

A track pattern recognition algorithm is able to derive the five parameters from the signals (“hits”) detected by the tracking subsystems (COT and silicon detectors, see section 2.2.2). The experiment exploits several tracking algorithms[43], each optimized for the information available in different detector region. In the following paragraphs we describe the main features of the four tracking algorithms relevant to this work: the Outside-In algorithm (OI), the Silicon-Stand-Alone (SiSA) algorithm, the Inside-Out (IO) algorithm and the new BackWard tracking (BW) algorithm.

#### Outside-In Algorithm

The Outside-In is the most reliable of CDF tracking algorithms as it is based on COT coverage and its 96 sampling planes of wires arranged in 8 superlayers, the short come is the (limited) region covered, that extends up to  $\eta = 1$ , the central region fully covered by all 8 COT superlayers.

Track pattern recognition starts in the COT outer layers (lower hit density) and proceeds through four steps: first each superlayer is searched for groups of three aligned hits that are fit to a straight line with the least squares method. Then the tracks are reconstructed from the information of the “axial” superlayers that are linked by two different algorithms (“segment linking” and “histogram linking” algorithms[44]). During the third step, the information of the stereo layers is added and the algorithm searches for the vertex of the track. As final step a global refit of the track is performed taking into account corrections for the non-uniformity of the magnetic field and for the modeling of the electrons drift.

At second stage of reconstruction, the track found in the COT is propagated into the silicon system. A road around a track is defined using the uncertainties on the COT track parameters and silicon hits are added if they lie inside this predefined road. When a hit is added, the track parameters are recalculated and the search is performed again. The impact parameter resolution of COT + SVX tracks is found to be  $\sigma_{d_0} \simeq 20 \mu\text{m}$ .

### Silicon-Stand-Alone Algorithm

The hits in silicon subdetectors not used by OI tracking are available to the Silicon-Stand-Alone algorithm[45], it covers the region  $|\eta| < 2$  with few residual capability up to  $|\eta| \simeq 2.8$ .

The SiSA algorithm starts from a collection of at least four hits in the SVXII detector in the  $r - \phi$  plane (SVX has five axial layers, three  $90^\circ$  layers and two small angle layers) and fits the  $C$ ,  $D_0$  and  $\lambda$  parameters to obtain a projection of the helix on the transverse plane. Then the algorithm creates a 3-D seed track adding small angle hits and the primary vertex information. At this point the  $90^\circ$  stereo hits are added and a global refit is performed.

SiSA tracks reconstructed only with SVXII have a poor resolution for high  $p_T$  tracks so hits are searched in LØØ and ISL with the SVXII track as seed. The track is refit if other layers can be added. However, the performances on momentum and impact parameter resolution are limited and indeed SiSA tracks are not used for secondary vertexing.

### Inside-Out Algorithm

The third tracking algorithm, the Inside-Out[46], tries to recover efficiency and  $p_T$  resolution in the region  $1.2 < |\eta| < 1.8$  where the COT coverage is limited. Practically SiSA tracks are used as seeds which are extrapolated to the COT inner cylinder. Matching hits in the COT are added, track is refitted and all duplicates are removed.

### New Backward tracking Algorithm

The last algorithm (BackWard tracking or BW) is a new feature of the Gen7 release of the CDF software production package and it is still in its testing phase. This work is part of the evaluation by CDF and it is the first physics analysis using it.

While the IO and SiSA algorithms start track reconstruction from the silicon system, Backward tracking uses the COT segments as seeds (see Fig. 3.2). The lower track density in the region of the seed track and the lower hits occupancy in the COT make BW tracking appealing to reconstruct tracks in the forward region<sup>2</sup>. The main problem for forward and partial COT seeded tracks is a careful calibration of electron drift times. After a seed is found, the algorithm attempts to add silicon hits using the same pattern recognition code of OI. The resulting track is then refit using the IO code. In this way the region up to  $\eta \sim 1.8$  can be further exploited.

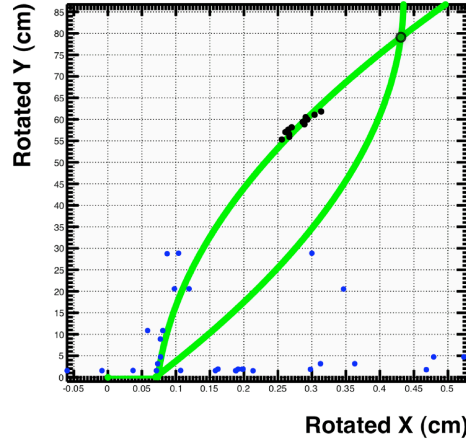


Figure 3.2: Concept of the BackWard tracking algorithm. COT seeds are used to extrapolate track to the silicon sensors.

## 3.2 Calorimeter Electron Identification

The calorimeter system plays a fundamental role in particle detection. The first hint of an outgoing electron is a energy deposit (“cluster”) in the EM section of the calorimeters.

---

<sup>2</sup>As it was said, the SiSA tracking has poor resolution for high momentum tracks.

The CDF EM clustering algorithm[40] works in a simple but efficient way. The physical space corresponding to the calorimeter towers is mapped in an  $\eta - \phi$  plane, the algorithm creates two lists of the calorimeter towers ordered by decreasing energy revealed on them: the “usable list” (working towers with energy  $> 100$  MeV) and the “seed list” (towers with energy  $> 2$  GeV). It then takes the first seed tower and creates an  $\eta - \phi$  cluster by adding the neighboring towers to form a 2x2 or 3x3  $\eta - \phi$  area.

An EM cluster is found if  $E^{Had}/E^{EM} < 0.125$ , where  $E^{Had}$  is the energy deposited in the hadronic calorimeter and  $E^{EM}$  is the corresponding quantity for the EM section. As final step the  $\eta - \phi$  centroid of the cluster is calculated and the used towers are removed from the lists. The algorithm selects the next seed tower and iterates the process until all the seed towers have been used.

Usually the 3x3 clustering is used in the CEM region while 2x2 clusters are used in the PEM region, this reduces the probability to overlap the clusters of two different electrons. A cluster is not allowed to cross the boundary between different subdetectors. Several corrections are applied to reconstruct the initial energy of the EM object. Clusters are corrected for lateral leakage, location inside the physical tower, on-line calibration and response curve drawn by the test beam data. Also the energy measured in the shower max (PES) and pre-shower (PPR) detectors is added to the final reconstructed energy. PES is also used to compare the shower profile of electrons or photons and, last but not least, it is used to measure the spatial position of the EM shower centroid.

Beyond the raw EM energy measurement, the calorimeter information can be further exploited for a better particle identification. The  $E^{Had}/E^{EM}$  ratio is used to identify electrons, indeed studies performed with candidate  $Z^0 \rightarrow e^+e^-$  events[41] show that electrons detected in the central or in the plug region have a little deposit in the hadronic part of the calorimeter (Fig. 3.3).

The *IsoRel* (or isolation) is another quantity derived from calorimeters. It is defined as:

$$- IsoRel \equiv E_T^{iso}/E_T^{cluster} < 0.1,$$

where  $E_T^{iso} = E_T^{0.4} - E_T^{cluster}$  and  $E_T^{0.4}$  is the energy collected by the calorimeters within a radius  $\Delta R = 0.4$  from the centroid of the EM cluster. Isolation is used in analysis involving a  $W^\pm$  or  $Z^0$  boson, indeed the kinematic region allowed to leptons coming from the bosons decay is usually far from jets or other particles (see Fig. 3.3).

If a track is matched to the EM cluster, also the  $E^{cluster}/p^{trk}$  ratio can be used for a better electron identification. The  $E/p$  distribution is peaked at 1 but it has large tails because the electrons can radiate bremsstrahlung



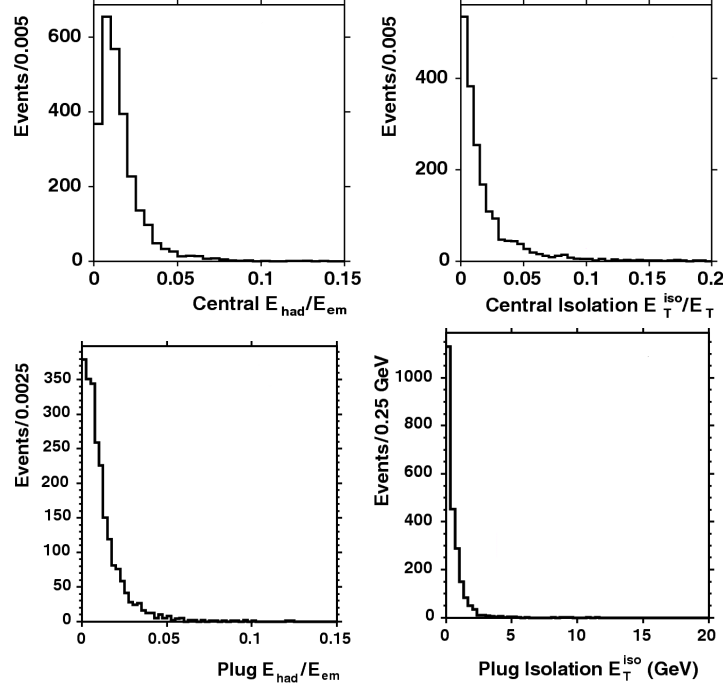


Figure 3.3:  $E^{Had}/E^{EM}$  (left) and isolation (right) distribution of central (top) and plug (bottom) calorimeter electron selection from unbiased, second legs of  $Z^0 \rightarrow e^+e^-$  candidate events in data[41].

collinear photons in the passage through matter in the tracking volume. The EM energy measurement is weakly influenced by that (the photon usually deposits energy in the same EM cluster), but the momentum measurement decreases.

### 3.3 Muon Identification

Muons play a secondary role in this analysis because the selected events require an electron, anyway their identification is necessary to avoid a background source.

The base concept for muon identification[47] is that high energy muons pass through the calorimetric system as Minimum Ionizing Particles (MIP) so they are identified by the presence of a track inside COT, a deposit of a MIP inside EM and HAD calorimeters and a track segment (“stub”) in the outer muon chambers. Muons can be faked by cosmic rays or hadrons showering deep inside calorimeters or not showering at all.

### 3.4 Primary Vertex Identification

The primary vertex is the position of the event interaction point. Precise reconstruction of the primary vertex makes possible to correct  $E_T$  and  $\cancel{E}_T$  and, above all, permits the individuation of displaced secondary vertices in an event, the signature of long living heavy hadrons.

The algorithm used here to reconstruct primary vertices is **PrimVtx**: a seed vertex is calculated as the average  $z$  position measured during collisions and is provided as input, then all tracks with  $|z_0 - z_{vtx}| < 1$  cm,  $|d_0| < 1$  cm and  $|d_0/\sigma_{d_0}| < 3$  are collected and ordered in decreasing  $p_T$ . They are fitted to a new vertex and the tracks with  $\chi^2 > 10$  are removed. The procedure is iterated until all accepted tracks have  $\chi^2 < 10$ . A quality index is assigned to the primary vertex depending on parameters like the number of final tracks and their provenience.

### 3.5 Neutrino Identification

Momentum conservation is the only way to reveal the presence of neutrinos since they do not interact inside the detector components. Although it is impossible to know the exact momentum of the colliding partons, the transverse component,  $p_T$ , should be approximately zero in the detector frame. All the detected (transverse) energy is vectorially summed, and if greater than zero, we assume a candidate neutrino is revealed. The missing transverse energy  $\cancel{E}_T$  gives a measurement of the neutrino transverse momentum<sup>3</sup>. It is defined as:

$$\vec{\cancel{E}}_T \equiv - \sum_i \vec{E}_T^i \quad (3.7)$$

where  $\vec{E}_T^i$  is a vector with magnitude equal to the transverse energy collected by the  $i$ -th calorimeter tower and pointing from the interaction vertex to the center of the tower. The sums involve all the towers with total energy above 0.1 GeV in the region  $|\eta| < 3.6$ . At offline level, the algorithm corrects for the position of the reconstructed event vertex and for any reconstructed muon (their energy is calculated using track information).

---

<sup>3</sup>For a massless neutrino  $p_T = E_T$ .

### 3.6 Jet Identification

QCD tells us that the partons composing the (anti)proton can be treated perturbatively as free particles if they are struck by an external probe<sup>4</sup> with sufficient high energy (“hard scattering”). However partons resulting from the interaction can not exist as free particles because at longer distances (i.e. lower energies) the *strong potential* can not be treated perturbatively and partons must form colorless hadrons. This process is called “hadronization” or “showering” and produces a collimated cluster of stable particles named “jet”. A jet approximately retains the total momentum and direction of the initial parton (for a pictorial representation see Fig. 3.4). From an experimenter’s

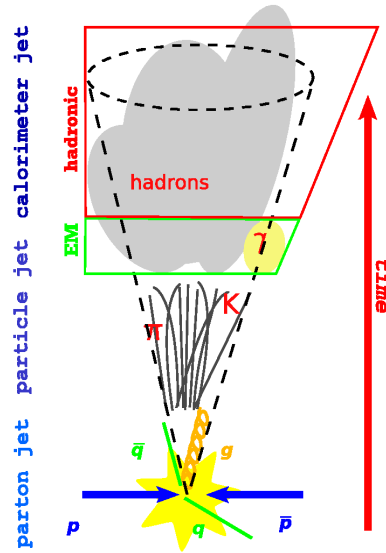


Figure 3.4: A parton originating from a hard scattering hadronizes and generates a collimated spray of particles, a jet.

point of view a jet is defined as a large energy deposit in a localized area of the detector (see Fig. 3.5). The challenge of a physics analysis is to recover from detector information the initial energy, momentum and, possibly, the kind of the parton produced in the original interaction. A “jet algorithm” is a tool to reconstruct such information and it must satisfy at best the following requirements[48]:

- *Infrared safety*: the presence of soft radiation between two jets may cause a merging of the two jets. This should not occur to avoid an

<sup>4</sup>I.e. a lepton or a parton from an other hadron.

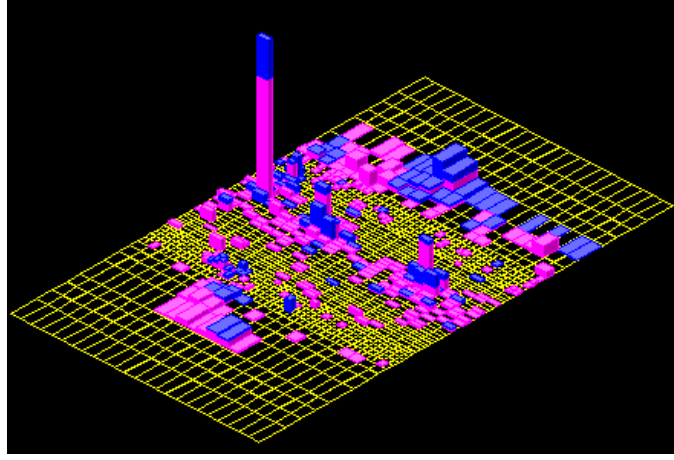


Figure 3.5: Calorimetric deposit in the  $\eta-\phi$  plane as represented in the CDF event display. EM deposits are red while HAD deposits are blue.

incorrect parton attribution.

- *Collinear safety*: the jet reconstruction should be independent from any collinear radiation in the event, i.e. different energy distribution of particles inside calorimetric towers.
- *Invariance under boost*: the same jets should be found independently from boosts in longitudinal direction.
- *Boundary stability*: kinematic variables should be independent from the details of the final state.
- *Order independence*: the same reconstructed quantities should appear looking at parton, particle and detector levels.
- *Straightforward implementation*: algorithm should be easy to implement in perturbative calculations.

Beyond this theoretical aspects a jet algorithm should be experimentally efficient with a high reconstruction efficiency, good resolution and stable at different luminosity.

### 3.6.1 CDF Cone Algorithm

CDF uses several algorithms, none of them really satisfying all the above requirements. The most common one, that is also the one used in this ana-

lysis, is JETCLU[49], an iterative fixed cone jet reconstruction algorithm based only on calorimetric information.

The algorithm begins by creating a list of the seed towers from all the calorimeter towers with transverse energy above the threshold of 1 GeV. Starting with the highest- $E_T$  seed tower, a precluster is formed by combining together all adjacent seed towers within a cone of given radius  $R^5$ . This procedure is repeated, starting with the next unused seed tower, until the list is exhausted. The  $E_T$ -weighted centroid is then formed from the towers in the precluster and a new cone of radius  $R$  is formed around this centroid. All towers with energy above the lower threshold of 100 MeV within this new cone are added to the cluster. Then, a new centroid is calculated from the set of towers within the cluster and a new cone drawn. This process is iterated until the centroid of the energy deposition within the cone is aligned with the geometric axis of the cone (stable solution).

The initial clusters found can overlap so the next step is to merge or separate overlapping clusters, since each tower may belong only to one jet, each particle should not be assigned to more than one jet. Two clusters are merged if the total energy of the overlapping towers is greater than 75% of the energy of the smallest cluster. If the shared energy is below this cut, the shared towers are assigned to the cluster that is closer in  $\eta - \phi$  space. This process is iterated again until the list of clusters remains fixed.

Massless four-vector momenta are assigned to the towers in the clusters for EM and HAD components with a magnitude equal to the energy deposited in the tower and the direction defined by a unit vector pointing from the event vertex to the center of the calorimeter tower at the depth that corresponds to the shower maximum. A cluster four-vector is then defined summing over the towers in the cluster:

$$E = \sum_{i=1}^N (E_i^{EM} + E_i^{HAD}), \quad (3.8)$$

$$p_x = \sum_{i=1}^N (E_i^{EM} \sin \theta_i^{EM} \cos \phi_i^{EM} + E_i^{HAD} \sin \theta_i^{HAD} \cos \phi_i^{HAD}), \quad (3.9)$$

$$p_y = \sum_{i=1}^N (E_i^{EM} \sin \theta_i^{EM} \sin \phi_i^{EM} + E_i^{HAD} \sin \theta_i^{HAD} \sin \phi_i^{HAD}), \quad (3.10)$$

$$p_z = \sum_{i=1}^N (E_i^{EM} \cos \theta_i^{EM} + E_i^{HAD} \cos \theta_i^{HAD}). \quad (3.11)$$

where the index  $i$  runs over the towers in the cluster. Other variables are

---

<sup>5</sup>CDF reconstructs jets using radii 0.4, 0.7 and 1.0.

added to the final jet-object used in the analysis: kinematic one, like  $E_T$ ,  $\eta$ , and  $\phi$ , or other useful information like the number of tracks reconstructed inside the jet cone, the vertex quality or the energy deposited in the HAD and EM calorimeter.

### 3.6.2 Jet Corrections

The ultimate goal of the jet reconstruction algorithm is the determination of the exact energy of the outgoing partons coming from the hard interaction. Clearly many factors produce a mismatch between the raw energy measured by the algorithm and the one of the parton before the hadronization.

CDF developed a set of jet energy corrections depending of  $\eta$ ,  $E_T^{raw}$  and  $R$  of the jet reconstructed by JETCLU algorithm. The corrections are divided into five levels<sup>6</sup> (“ $L$ -levels”) so that can be applied in a standard way to different analyses[50, 51]:  $\eta$ -dependent response( $L1$ ), effect of multiple interactions ( $L4$ ), absolute energy scale ( $L5$ ), underlying event ( $L6$ ) and out-of-cone ( $L7$ ) corrections. In this analysis jets are corrected up to Level-5. The correction  $L1$  and  $L5$  are multiplicative factors ( $f_{L1}$  and  $f_{L5}$ ) on the raw  $E_T$  of the jet, the others are additive constants ( $A_{L4}$ ,  $A_{L6}$  and  $A_{L7}$ ). The general equation to apply all corrections is:

$$E_T^{corr}(\eta, E_T^{raw}, R) = (E_T^{raw} f_{L1} - A_{L4}) f_{L5} - A_{L6} + A_{L7}. \quad (3.12)$$

#### Level-1: $\eta$ dependent corrections

$L1$  correction is applied to raw jet energy measured in the calorimeters to make the detector response uniform in  $\eta$ , it takes into account aging of the subdetectors<sup>7</sup> and other “hardware” non-uniformities (for example the presence of cracks). This correction is obtained using a large di-jet sample: events with one jet (*trigger jet*) in the central region of the calorimeter ( $0.2 < |\eta| < 0.6$ ), where the detector response is well known and flat in  $\eta$ , and a second jet (*probe jet*), allowed to range anywhere in the calorimeter ( $|\eta| < 3.6$ ). In a perfect detector the jets should be balanced in  $p_T$ , a balancing fraction is formed:

$$f_b \equiv \frac{\Delta p_T}{p_T^{ave}} = \frac{p_T^{probe} - p_T^{trigger}}{(p_T^{probe} + p_T^{trigger})/2}, \quad (3.13)$$

---

<sup>6</sup>The actual naming skips  $L2$ , because it is absorbed in  $L1$ , and  $L3$ , as it was introduced as a temporary MC calibration in Run II.

<sup>7</sup>This was the  $L2$  correction during Run I

the average of  $f_b$  in the analyzed  $\eta$  bin is used to define the  $\beta$  factor<sup>8</sup> (Fig. 3.6 shows  $\beta$  distribution for different cone radii):

$$\beta \equiv \frac{2 + \langle f_b \rangle}{2 - \langle f_b \rangle}. \quad (3.14)$$

The final  $L1$  correction is defined as  $f_{L1}(\eta, E_T^{raw}, R) = 1/\beta$  and reproduces an approximately flat response in  $\eta$  with an error varying from 0.5% to 7.5%.

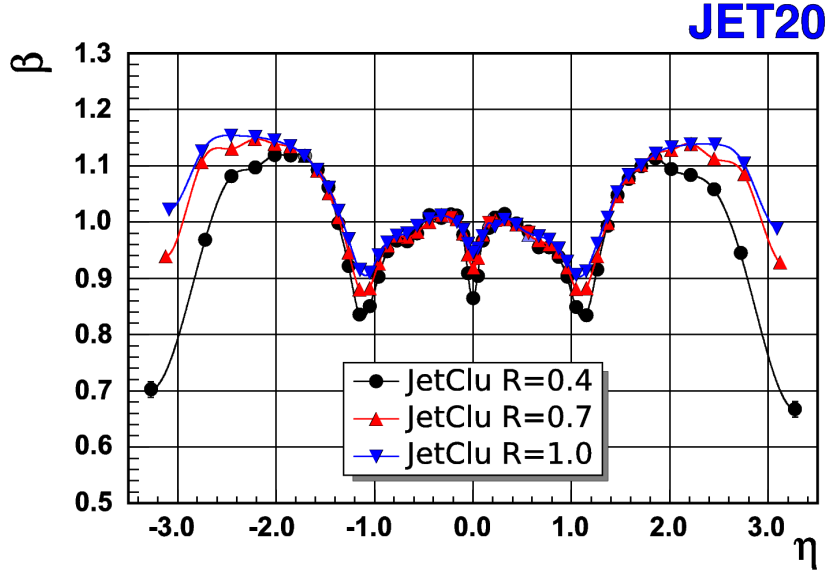


Figure 3.6:  $\eta$ -dependence of  $\beta$  factor for cone radii  $R = 0.4, 0.7$  and  $1.0$ , measured in the di-jet component of jet20 sample.

#### Level-4: multiple interactions corrections

Jet energy measurement is also degraded by the presence of minimum-bias events that come from multiple  $p\bar{p}$  interactions. This correction becomes more relevant at high luminosity, indeed the number of  $p\bar{p}$  interactions is Poisson distributed with mean value almost linear with instantaneous luminosity:

$$\langle N(\mathcal{L} \simeq 10^{32} \text{ cm}^{-2}\text{s}^{-1}) \rangle \simeq 3, \quad \langle N(\mathcal{L} \simeq 3 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}) \rangle \simeq 8. \quad (3.15)$$

<sup>8</sup>The definition of Eq. 3.14 has a average value equal to  $p_T^{probe}/p_T^{trig}$  but is less sensitive to presence of non-Gaussian tails in the usual  $p_T^{probe}/p_T^{trig}$  ratio.

The energy of particles coming from those processes is estimated from minimum-bias events drawing a cone in a random position in the region  $0.1 < \eta < 0.7$ . Figure 3.7 shows that the measured minimum-bias  $E_T$  grows linearly with the number of primary vertices<sup>9</sup>, such quantity,  $A_{L4}$ , must be subtracted by jet raw energy. The total uncertainty is about 15%, it mostly depends on luminosity and event topology.

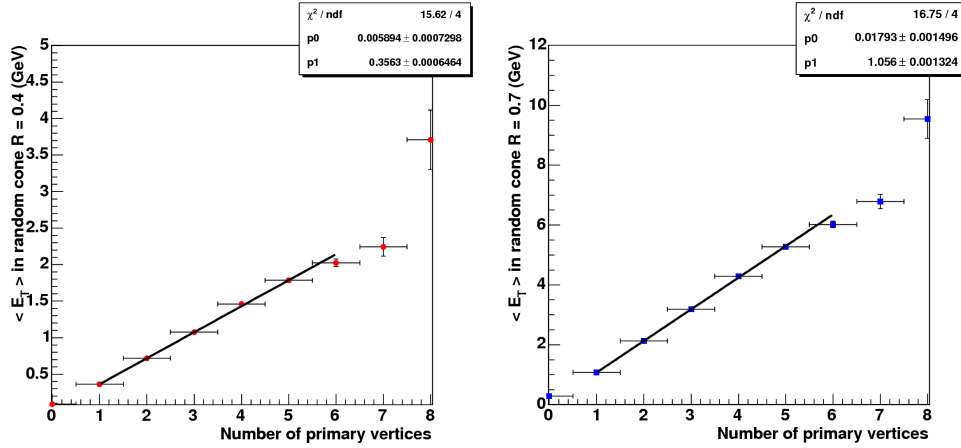


Figure 3.7: Multiple interaction  $E_T$  correction as a function of primary vertex number for cones with  $R = 0.4$  (right) and  $R = 0.7$  (left).

### Level-5: absolute energy scale corrections

While  $L1$  and  $L4$  make jet reconstruction uniform over the whole detector and over the global behavior of  $p\bar{p}$  beam interaction,  $L5$  correction ( $f_{L5}$ ) cures the difference between the final jet energy measurement of the event with respect to the parton level physics.

The study is MC driven: first jet events are generated with full CDF detector simulation, then jets are reconstructed both at calorimeter and hadron generation levels (HEPG) with the use of same clustering algorithm. A calorimeter jet (C) is associated to the corresponding hadron jet (H) if  $\Delta R < 0.1$ . For both HEPG and detector jets the transverse momentum,  $p_T^C$  and  $p_T^H$ , is calculated. The absolute jet energy is defined as  $\mathcal{P}(p_T^C | p_T^H)$ , the probability to measure  $p_T^C$  with a given  $p_T^H$ <sup>10</sup>.

Figure 3.8 shows the correction factor  $f_{L5}$  for different cone sizes as function of the different jet transverse energies. The total uncertainty is about

<sup>9</sup>Good quality primary vertices are reconstructed through at least 2 COT tracks.

<sup>10</sup>Different  $p_T^H$  can give the same  $p_T^C$ , in this case the maximum is taken.



3% and it mainly arises from the determination of calorimetric response to single particles and MC fragmentation modeling.

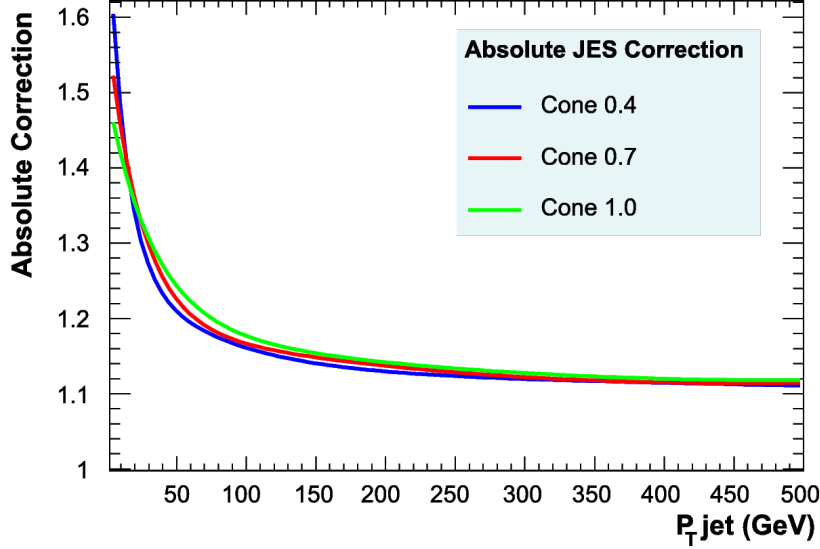


Figure 3.8: Absolute jet energy scale correction ( $f_{L5}$ ) for different cone sizes.

### Level-5 & Level-6: underlying event and out-of-cone corrections

The underlying event correction ( $L6$ ) takes into account the interaction processes which can occur between spectator partons or that originates from initial state radiation (usually soft gluon radiation) while the out-of-cone correction ( $L7$ ) considers the fraction of particles coming from the original parton that fall outside the jet cone.

The underlying event energy ( $A_{L6}$ ) must be subtracted to the total jet energy. It was measured studying minimum-bias events during Run I and is parametrized with a constant value that scale with the cone radius. Out of cone energy ( $A_{L7}$ ) must be added to the total jet energy, studies are carried out with the same jet-parton matching method of  $L5$ .

## 3.7 Secondary Vertex Tagging

The identification of heavy flavor jets (i.e jets containing a bottom or charm hadron) is of fundamental importance in this and in many other analyses.

Both quark top and the SM Higgs boson have large branching fraction in  $b$ -quark, furthermore the exclusive identification of  $b$ -hadrons is a methodology to reduce background because many uninteresting physical processes contain only light flavor hadrons in their final state. An algorithm able to select a jet coming from a  $b$ -hadron is called “ $b$ -tagger” or “heavy flavor tagger”.

The Secondary Vertex Tagger algorithm (**SecVtx**) is one of the main  $b$ -taggers used at CDF<sup>11</sup>. It takes advantage of the long life time of  $b$ -hadrons: a  $c\tau$  value of about  $450\ \mu\text{m}$  together with a relativistic boost due to a momentum of several GeV permit to a  $b$ -hadron to fly several millimeters<sup>12</sup> away from the primary interaction vertex. The relevant quantity is the  $c\tau$  which is approximately the average impact parameter of the outgoing debris of  $b$ -hadron decays. The decay produces small sub-jets composed by tracks with large impact parameter ( $d_0$ ). The silicon detectors (see section 2.2.2) are able to reconstruct  $d_0$  with adequate precision to separate displaced tracks from the prompt tracks coming from the primary interaction. Figure 3.9 shows as a  $W$ +jets candidate event with two displaced secondary vertices is identified by **SecVtx** and reconstructed by the CDF event display.

The richness of  $b$ -hadrons decay channels makes very difficult for a single algorithm to achieve optimal efficiency. For example **SecVtx** has low efficiency in detecting semileptonic decays, or in decays with low charged tracks multiplicity. Moreover the presence of  $D$  (charm) hadrons that may produce ternary vertices makes the kinematic even more complex.

To improve its efficiency, and to add tools to its searches, CDF developed other tagger algorithms:

- the “soft-lepton-tagger” algorithm which looks for semileptonic heavy flavor decays (its main difficulty is the identification of low energy leptons inside high occupancy jets);
- the “jet-probability” algorithm assigns to each track the probability to come from the primary interaction vertex, so, setting different thresholds heavy flavor tagging is possible.

As we write a strong effort is being produced to create a new, multivariate neural network  $b$ -tagger (“Roma- $b$ -tagger”) capable of exploiting all information, i.e. soft leptons, ternary vertices,  $c/b$  hadron differences. So far this algorithm has been tested on a single physics channel[52].

<sup>11</sup>Historically it was the most important component in top discovery in 1995.

<sup>12</sup>The average transverse momentum of a  $b$ -hadron coming from a  $WH$  events is about  $40\ \text{GeV}/c$  for a Higgs boson mass of  $115\ \text{GeV}/c^2$ ; in that condition a neutral  $B^0$  meson of mass  $5.28\ \text{GeV}/c^2$  undergoes a boost  $\beta\gamma = 7.6$  and the average decay length is  $3.5\ \text{mm}$ .

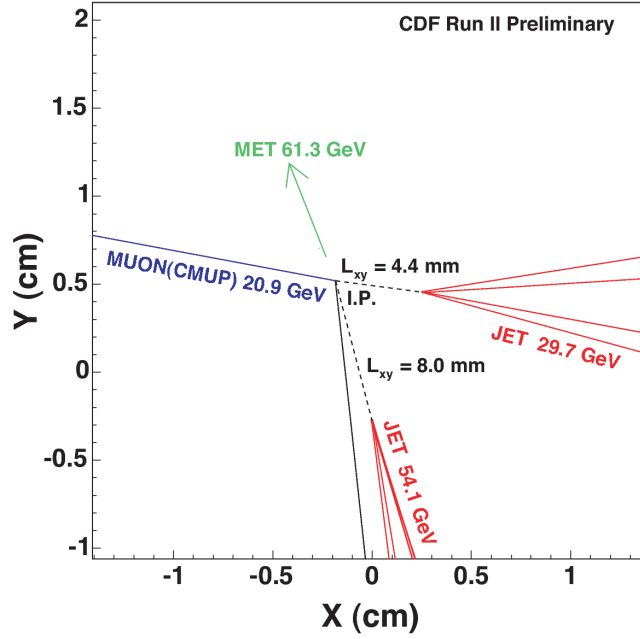


Figure 3.9:  $W$ +jets candidate event with two secondary vertices tagged by **SecVtx** (run 166063, event 279746). The  $\cancel{E}_T$  direction, a muon track, a prompt track and tracks from the secondary vertices are shown.

### 3.7.1 The SecVtx Algorithm

**SecVtx**  $b$ -tagging is performed for all the jets with  $|\eta| < 2.4$  in a event. The algorithm searches for secondary vertices using the tracks within the jet cone of radius  $\Delta R = 0.4$ . The usable tracks must satisfy the following requirements:

- $p_t > 0.5 \text{ GeV}/c$ ,
- $|d_0| < 0.15 \text{ cm}$  and  $|d_0/\sigma_0| > 2.0$ ,
- $|z_0 - z_{prmVtx}| < 2.0 \text{ cm}$ ,
- have a minimum number of hits in the silicon detector, the number depends on track reconstruction quality and position,
- be seeded or confirmed in the COT.

A “taggable” jet is defined as a jet containing at least two usable tracks.

A special note must be posed on the last request, in fact it excludes the possibility to use SiSA tracking but allows the use of the new “Backward tracking” present in the Gen7 production. This surely brings an increase in **SecVtx** efficiency. However, the officially quoted values for efficiency, purity and scale factors have been evaluated only for Gen6 reconstruction version.

The algorithm works on a two step basis and has three operating modes, “tight” (the standard one), “loose” and “ultra-tight”, that depend of the applied thresholds.

In the *Pass 1* at least three tracks are required to pass loose selection criteria, i.e the general one for **SecVtx** tracks, but at least one of the tracks used is required to have  $p_T > 1.0$  GeV/c. The selected tracks are combined two by two until a seed secondary vertex is reconstructed, then all the others are added one by one and a quality  $\chi^2$  is computed, tracks are added or removed depending of their contribute to the  $\chi^2$ .

The *Pass 2* begins if *Pass 1* gives a negative result. Now only two tracks are required to form a secondary vertex but they must pass tighter requirements:  $p_t > 1.0$  GeV/c,  $|d_0/\sigma_0| > 3.5$  and one of the tracks must have  $p_T > 1.5$  GeV/c.

If a secondary vertex is identified in a jet, the jet is “tagged”. The two dimensional decay length  $L_{xy}$  is calculated as the projection into the jet axis, in the  $r - \phi$  plane, of the **SecVtx** vector, the one pointing from the primary vertex to the secondary. The sign of  $L_{xy}$  is defined by the angle  $\alpha$  between the jet axis and the **SecVtx** vector. Figure 3.10 explains the geometry.

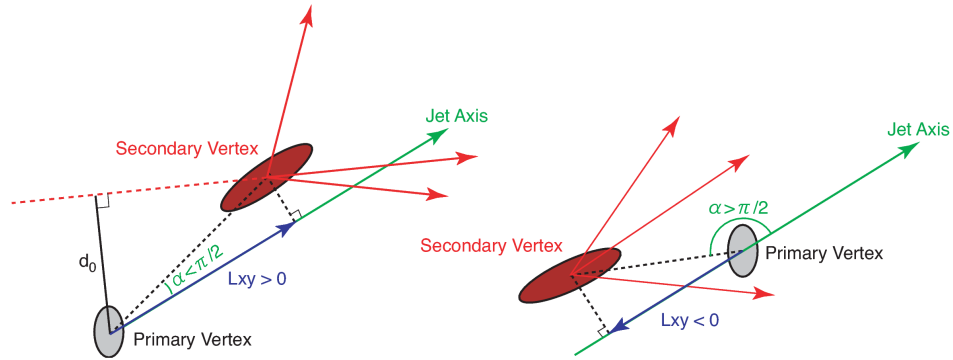


Figure 3.10: **SecVtx** variables. Left: true reconstructed secondary vertex. Right: negative **SecVtx** tag, falsely reconstructed secondary vertex.

A secondary vertex coming from a heavy flavor hadron is expected to have large  $L_{xy}$ . To reduce background due to mismeasured tracks  $|L_{xy}/\sigma_{L_{xy}}| > 7.5$

is required<sup>13</sup>. Other cuts are made on the invariant mass of the pair of tracks, to avoid  $K$  and  $\lambda$  decays, and on vertex multiplicity and impact parameter to reject secondary vertices due to interaction with material inside the tracking volume.

### 3.7.2 Tagging Performances and Scale Factors

The performances of a  $b$ -tagger are evaluated on its efficiency, i.e the rate of correctly identified  $b$ -hadrons over all the produced  $b$ -hadrons, and on its purity, i.e the rate of falsely identified  $b$ -hadrons in a sample with no true  $b$ -hadrons. CDF uses  $t\bar{t}$  MC to evaluate **SecVtx** efficiency relying on detector and physical processes simulation. Figure 3.11 shows the  $b$ -tagging efficiency as a function of jet  $\eta$  and  $E_T$  for the three **SecVtx** modes, tagging efficiency drops essentially because of lower track resolution.

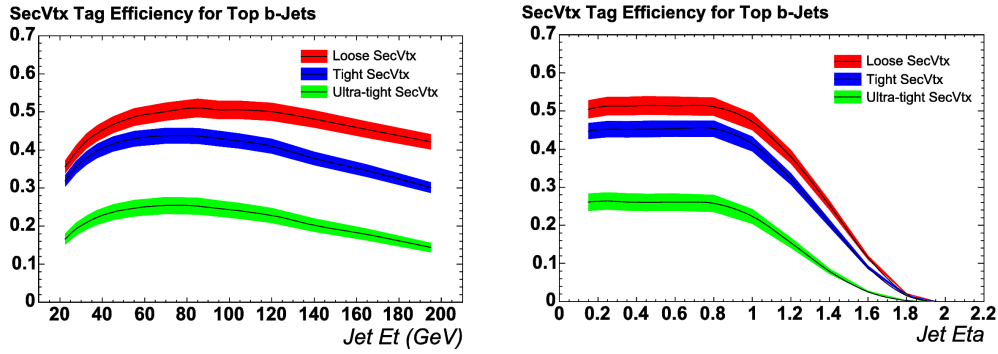


Figure 3.11:  $b$ -tagging efficiency for the three operation modes of the **SecVtx** algorithm.

As MC does not reproduce the exact  $b$ -tagging efficiency of **SecVtx** a “Scale Factor” (SF or  $\Phi$ ) is introduced to account for data/MC difference in the form:

$$\Phi \equiv \frac{\epsilon_{data}}{\epsilon_{MC}}. \quad (3.16)$$

CDF uses two methods to calculate SF: the first looks in the inclusive jet sample for jet pairs consistent with heavy flavor pair production: for example an event can contain two back-to-back jets one matched to a high  $p_T$  muon and the other one  $b$ -tagged by **SecVtx**, we can infer that the jet matched to

<sup>13</sup>Negative  $L_{xy}$  has no physical meaning but it is important to estimate the mistag probability due to resolution effects.

mode	$\Phi$	$\sigma_\Phi$ (stat)	$\sigma_\Phi$ (syst)
Loose	0.95	0.01	0.05
Tight	0.95	0.01	0.04
Ultra-tight	0.88	0.01	0.05

Table 3.1: **SecVtx** Scale Factors ( $\Phi$ ) for the three different **SecVtx** operating modes.

the muon originate from a  $b$ -hadron semileptonic decay but was not tagged. So it is possible to compare  $b$ -tagging efficiency on real data and derive the SF. The second method exploits electrons instead of muons. Figure 3.12 shows the SF determination with the two methods, Table 3.1 reports the SF for the three **SecVtx** operation modes.

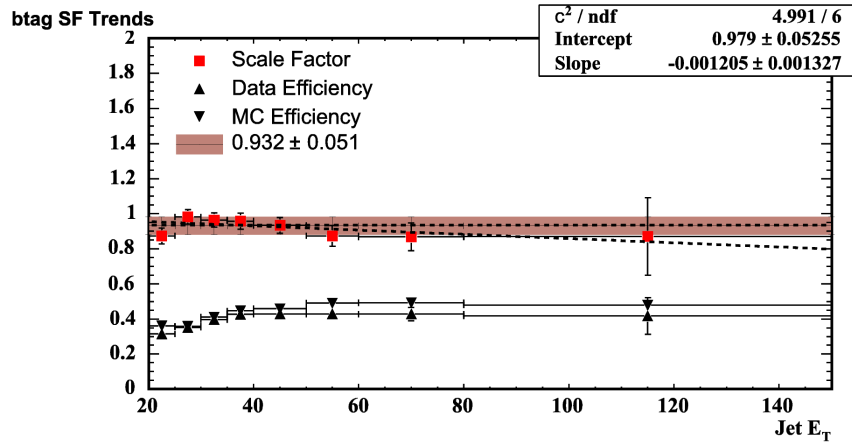


Figure 3.12: Scale Factor calculation with muons and electrons methods and its dependence over the  $E_T$  of the jet.

The number of falsely **SecVtx** tagged jets is dubbed *mistags*. Mistags can be due to track resolution, long living light-flavor hadrons or secondary interactions with detector material. They are measured in an inclusive jet sample (with low or none heavy flavor components) and the reconstruction parametrizes mistags assigning a *mistag probability* to each event depending on  $E_T$ ,  $\eta$ ,  $\phi$  and track multiplicity of the considered jet (more details can be found in section 5.4). Figure 3.13 shows the mistag rate as function of  $E_T$  and  $\eta$  for a jet sample and two different **SecVtx** operating modes.

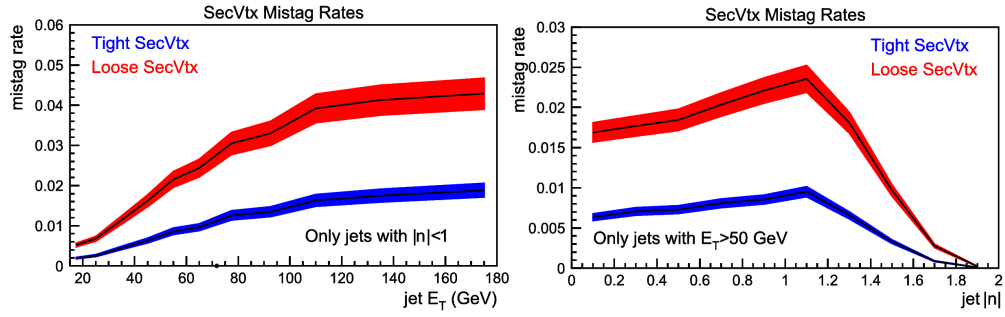


Figure 3.13: Rate of wrongly SecVtx tagged jets (*mistags*) as a function of  $E_T$  and  $\eta$  for tight and loose SecVtx operation mode. The rate is derived from an inclusive jet sample.





# Chapter 4

## Event Selection and Signal Acceptance

We search for events compatible with  $WH$  associated production and decay in the form:

$$p\bar{p} \rightarrow W^\pm + H \rightarrow e^\pm \nu_e + b\bar{b}, \quad (4.1)$$

with the electron going into the forward region of CDF detector. This section describes the event selection criteria for  $W$  reconstruction, jet identification and  $b$ -tagging.

The analysis is performed on physics quality data corresponding to an integrated luminosity:

$$\int \mathcal{L} dt = 1360 \pm 80 \text{ pb}^{-1}. \quad (4.2)$$

Data are collected by the **MET\_PEM** trigger path: it requires the presence of  $\cancel{E}_T$  and of an EM deposit in the forward calorimeters ( $1.2 < |\eta| < 2.8$ ).

Electron selection criteria involve calorimetric cuts and track selection. The tracking efficiency in the forward region has been studied in Gen6 and Gen7 framework with a  $Z^0 \rightarrow e^+e^-$  sample.

The presence of missing transverse energy is the signature of a neutrino in the event, two different thresholds ( $\cancel{E}_T > 20 \text{ GeV}$  and  $\cancel{E}_T > 25 \text{ GeV}$ ) were applied to study background composition (see chapter 5).

Jets reconstructed in the region  $|\eta| < 2.0$  are selected and **SecVtx**  $b$ -tagging algorithm is then applied.

### 4.1 Data Sample and Run Interval

The data sample used corresponds to the three data taking periods:

- 0d: from March 2002 to August 2004, it collects runs from 138425 to 186598,  $\int \mathcal{L} dt = 336 \pm 20 \text{ pb}^{-1}$ ;
- 0h: from December 2004 to September 2005 it collects runs from 190697 to 203799,  $\int \mathcal{L} dt = 415 \pm 25 \text{ pb}^{-1}$ ;
- 0i: from September 2005 to November 2006 it collects runs from 203819 to 228596,  $\int \mathcal{L} dt = 612 \pm 37 \text{ pb}^{-1}$ .

the dataset `bpe1` is used requiring only “Good Run” data, i.e. runs with at least  $10 \text{ nb}^{-1}$  of continuous data taking, and marked as good for EM calorimeters and silicon subdetector use.

The considered data period contains nearly  $60 \cdot 10^6$  recorded events, and the total corresponding luminosity is:

$$\int \mathcal{L} dt = 1360 \pm 80 \text{ pb}^{-1}, \quad (4.3)$$

the quoted 6% error is dominated by the uncertainty on CLC acceptance (see section 2.2.5).

All the `bpe1` datasets were entirely reprocessed with the Gen7 development version (6.1.4int6). The analysis has been performed using both the standard CDF Gen6 software release and the Gen7 release which contains the new Backward Tracking (see section 3.1.1).

## 4.2 Event Selection Requirements

The events used in this analysis are collected by the three-levels trigger path (see appendix A) called `MET_PEM` trigger, developed for the online selection of physics processes involving  $W^\pm \rightarrow e^\pm + \nu$  in high pseudo-rapidity regions. The three-levels requirements are:

**Level-1** at least an EM deposit with  $E_T > 8 \text{ GeV}$  and  $Had/EM < 0.125$  and a L1  $\cancel{E}_T > 15 \text{ GeV}$ ;

**Level-2** a Level-2 EM cluster with  $E_T > 20 \text{ GeV}$  and  $Had/EM < 0.125$  in the plug region;

**Level-3:** a Level-3 plug EM object with  $E_T > 20 \text{ GeV}$  and a L3  $\cancel{E}_T > 15 \text{ GeV}$ .

In the following we list the criteria applied in cascade for offline identification of the physics objects of interest. They can be classified in four main steps: calorimetric cuts to identify forward EM cluster and neutrino, track matching to confirm the presence of a *plug* electron candidate, other lepton vetoes and jet selection with the secondary vertices identification.

### Calorimeter Selection

Calorimetric cuts tighten trigger requests on the online-selected high energy EM cluster and introduce isolation to clean the  $W$  sample (variables definition can be found in section 3.2). The requests are:

- at least one EM cluster in fiducial region;
- EM cluster in the plug region:  $1.2 < |\eta^{cluster}| < 2.8$ ;
- offline corrected  $E_T^{cluster} > 20$  GeV;
- $Had/EM < 0.05$ ;
- $IsoRel < 0.1$ ;
- offline corrected  $\cancel{E}_T > 20$  GeV or  $\cancel{E}_T > 25$  GeV.

This work analyzes two  $\cancel{E}_T$  scenarios, indeed a better understanding of the selected sample is obtained from studying the different background composition produced (see chapter 5).

### Track Matching

A track matched to the EM cluster is a commonly used strategy to reject events with photons (or more generically jets) faking electrons. Cuts optimization and variables description are reported in appendix B and section 3.1. The requirements are:

- $p_T^{trk} > 1.0$  GeV/ $c$ ;
- $\Delta_{x,y} < 3$  cm between the PES cluster centroid and the extrapolated track;
- track vertex  $|z_0^{trk}| < 60$  cm;
- $E^{cluster}/p^{trk} < 2$ .

### Other Lepton Vetoes

All the previous cuts select a  $W^\pm \rightarrow e^\pm + \nu$  candidate containing a *plug* electron. Since of  $WH \rightarrow l\nu + b\bar{b}$  decay contains only one lepton, the following vetoes are applied to reject events with more than one lepton:

- second *plug* electron veto (mostly Z plug-plug removal);

- central *tight* electron veto (mostly Z central-plug removal);
- events tagged as cosmic rays are vetoed<sup>1</sup>;
- *tight* muon veto.

Details on *tight* central electrons and *tight* muons are in appendix C.

### ***b*-jets Candidate Selection**

Jet identification is the last step in the  $W$ +jets sample construction. Jets are identified by the following requests:

- reconstruction cone radius  $R = 0.4$ ;
- identification in the calorimeter region  $|\eta_{jet}| < 2$ ;
- $E_T^{corr} > 20$  GeV, with raw  $E_T$  corrected to Level-5 generic jet-correction;
- Two or more non SiSA tracks with  $\Delta R_{jet-trk} < 0.4$ ;
- One or more jet with **SecVtx** tagging.

The sample is divided in four jet-bins:  $W + 1jet$ ,  $W + 2jets$ ,  $W + 3jets$  and  $W + Njets$  with  $N \geq 4$  (sometimes the last bin is simply named  $W + 4jets$ ). Requests up to  $E_T^{corr} > 20$  GeV define the “pretag” jet sample; the request of two or more non SiSA tracks<sup>2</sup> define the “taggable” jet sample and, finally, the **SecVtx** tag defines the “tagged”  $W$ +jets sample. The final  $W$ +jets selection includes four “Scenarios”:

**Scenario I:** Generation 6 of CDF software release with  $\cancel{E}_T > 20$  GeV,

**Scenario II:** Generation 6 of CDF software release with  $\cancel{E}_T > 25$  GeV,

**Scenario III:** Generation 7 of CDF software release with  $\cancel{E}_T > 20$  GeV,

**Scenario IV:** Generation 7 of CDF software release with  $\cancel{E}_T > 25$  GeV.

The pretag and tagged samples in the different scenarios are summarized in Table 4.1.

---

<sup>1</sup>Track distance from primary vertex, timing and calorimeters information is used to identify cosmic rays[54].

<sup>2</sup>It is required also good quality reconstruction of the primary vertex.

Selected Events (Scenario I) $\cancel{E}_T > 20$ GeV, Generation 6				
	1 jet	2 jet	3 jet	4 jet
pretag events	31890	5815	1047	207
tagged events ( $\geq 1tag$ )	415	231	66	35
Selected Events (Scenario II) $\cancel{E}_T > 25$ GeV, Generation 6				
	1 jet	2 jet	3 jet	4 jet
pretag events	23879	4009	681	148
tagged events ( $\geq 1tag$ )	272	161	45	29
Selected Events (Scenario III) $\cancel{E}_T > 20$ GeV, Generation 7				
	1 jet	2 jet	3 jet	4 jet
pretag events	38753	7156	1303	259
tagged events ( $\geq 1tag$ )	876	418	137	57
Selected Events (Scenario IV) $\cancel{E}_T > 25$ GeV, Generation 7				
	1 jet	2 jet	3 jet	4 jet
pretag events	27867	4735	835	172
tagged events ( $\geq 1tag$ )	576	269	89	45

Table 4.1: Events passing pretag and tag selection divided into jet-bins in the four scenarios.

### 4.3 Forward Tracking Efficiency Study

Before calculating  $WH$  signal acceptance it is important to study the efficiency for track matching and how well it is described by MC. Selection of the  $W$  candidates relies on a reconstructed track matched to the plug EM cluster. The new aspect of this analysis is the use of 3-D tracks reconstructed in forward region of the detector with the improved algorithms available in Generation 7 of the CDF software release. A similar work was done on data produced using Gen5[55] but this is the first time it is reproduced using Gen7-produced data.

The efficiency is defined as the number of signal events after the cut divided by the number of signal events before the cut. Track matching efficiency and the  $E/p < 2$  cut efficiency were studied both on a  $Z^0 \rightarrow e^+ + e^-$  data sample and MC sample. Agreement between data and MC was tested.

The  $Z^0 \rightarrow e^+ + e^-$  candidates are collected by the central high energy electron trigger `ELECTRON_CENTRAL_18`<sup>3</sup> and are stored in `zewk` dataset. The offline requests are a *tight central* electron (see appendix C) and an inclusive selection on plug candidates:

- EM cluster reconstructed in the Plug region ( $1.2 < |\eta| < 2.8$ );
- cluster  $E_T > 20$  GeV;
- $Had/EM < 0.125$ .

The cuts on the second leg reproduce the `MET_PEM` trigger requirements for the electron, in this way an independent control sample is created as the plug leg is not involved in the trigger selection. The purity of the sample is increased by requiring that the invariant mass of the two electrons is within the  $Z^0$  mass window:

$$80 < M_{ee} < 100 \text{ GeV}/c^2. \quad (4.4)$$

---

<sup>3</sup>Trigger path requirements are:

- L1: one CEM tower with  $E_T > 8$  GeV and one associated XFT track with  $p_T^{XFT} > 8.34$  GeV;
- L2: L1 confirm, a CEM cluster with  $E_T > 16$  GeV, track  $|\eta^{XFT}| < 1.317$  GeV and  $p_T^{XFT} > 8.34$  GeV;
- L3: one CEM electron object with  $E_T > 18$  GeV and a COT reconstructed track with  $p_T^{COT} > 9$  GeV.

Also the cuts  $Had/EM < 0.05$  and  $IsoRel < 0.1$  are applied as we are interested only in tracking and  $E/p$  efficiency. The same cuts are also applied to the  $Z^0 \rightarrow e^+e^-$  MC<sup>4</sup>.

The number of selected  $Z^0$  on data is estimated for each cut by fitting the  $M_{ee}$  peak and extrapolating the background ( $\lesssim 1\%$ ) from the sidebands regions (see Figure 4.1). The calculation is straightforward in the MC sample, where, of course no background is present. Table 4.2 summarizes the calculated efficiencies. The low efficiency of the  $E/p$  for Gen7 is due to a large number of tracks reconstructed as silicon standalone (SiSA) (see appendix B). All MC results well reproduce data. It is important to notice that here, the tracking efficiency itself is convoluted with the spectrum of electrons from  $Z$  decays (which is known to be similar to the  $W$  one) and the detector performances (electronics dead channels, readout problems etc).

	Gen6 (data)	Gen 6 (MC)	Gen7 (data)	Gen 7(MC)
$\epsilon_{trk}$	$0.480 \pm 0.008$	$0.498 \pm 0.001$	$0.613 \pm 0.007$	$0.612 \pm 0.001$
$\epsilon_{E/p}$	$0.72 \pm 0.01$	$0.710 \pm 0.002$	$0.680 \pm 0.008$	$0.666 \pm 0.001$

Table 4.2: Track matching and  $E/p$  efficiency in Gen6 and Gen7 for  $Z \rightarrow ee$  events. Data and MC.

## 4.4 WH Signal Acceptance

The aim of this work is to increase the acceptance for  $WH$  events at CDF. The number of expected  $WH$  signal events and the acceptance are evaluated using a PYTHIA MC sample of  $WH \rightarrow l\nu b\bar{b}$  with  $m_H = 120 \text{ GeV}/c^2$ .  $WH$  signal events are calculated with the following equation:

$$N_{WH \rightarrow l\nu b\bar{b}} = \epsilon_{WH \rightarrow l\nu b\bar{b}} \cdot \int \mathcal{L} dt \cdot \sigma(p\bar{p} \rightarrow WH) \cdot BR(H \rightarrow b\bar{b}), \quad (4.5)$$

where  $\epsilon_{WH \rightarrow l\nu b\bar{b}}$  is the detection efficiency for signal,  $\int \mathcal{L} dt = 1360 \pm 80 \text{ pb}^{-1}$  (Eq. 4.2) is the integrated luminosity,  $\sigma(p\bar{p} \rightarrow WH) = 0.16 \text{ pb}$  is the next-to-leading order  $WH$  production cross section for a  $m_H = 120 \text{ GeV}/c^2$  (Fig. 1.8), and  $BR(H \rightarrow b\bar{b}) = 0.68$  is the next-to-leading order branching fraction to  $b$  quarks for  $m_H = 120 \text{ GeV}/c^2$  (Fig. 1.9). Factor  $\epsilon_{WH \rightarrow l\nu b\bar{b}}$  is calculated as:

$$\epsilon_{WH \rightarrow l\nu b\bar{b}} = \epsilon_{trig} \cdot \sum_{l=e,\mu,\tau} BR(W \rightarrow l\nu) \cdot \epsilon_{WH \rightarrow l\nu b\bar{b}}^{MC}, \quad (4.6)$$

<sup>4</sup>MC are generated with Pythia, further details on MC can be found in section 2.4.

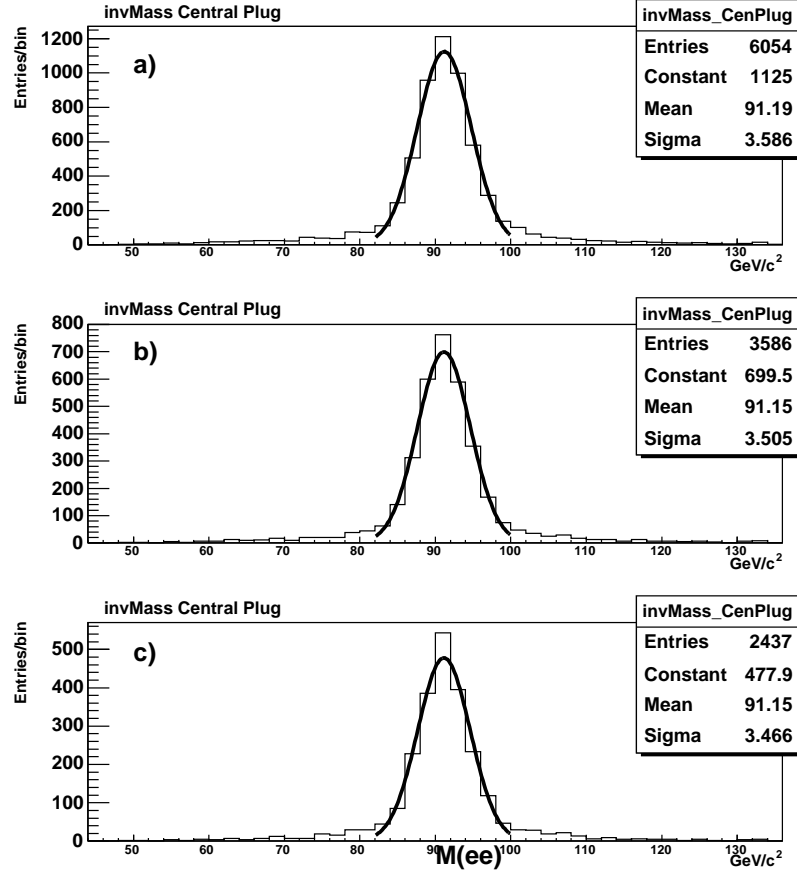


Figure 4.1: Invariant mass of  $Z_{CP}$  candidates: a) before tracking requirement; b) after track matching requirement; c) after  $E/p$  cut.

where  $\epsilon_{trig} = 0.961 \pm 0.004$  is the trigger efficiency (appendix A),  $\sum_{l=e,\mu,\tau} BR(W \rightarrow l\nu) = 0.32$  is the  $W$  to leptons branching ratio and finally  $\epsilon_{WH \rightarrow l\nu b\bar{b}}^{MC}$  is the MC calculated acceptance, i.e. the fraction of signal events which pass selection requirement at pretag and  $b$ -tagging levels, values are corrected for *tight* SecVtx data/MC scale factor (Table 3.1).

The  $\epsilon_{WH \rightarrow l\nu b\bar{b}}^{MC}$  values in the four different scenarios for each jet-bin are reported in Table 4.3. The number of expected signal events,  $N_{WH \rightarrow l\nu b\bar{b}}$ , is reported in Table 4.4. The quoted uncertainty includes several contributions: statistical, trigger efficiency, the SecVtx scale factor and (on  $N_{WH \rightarrow l\nu b\bar{b}}$ ) the luminosity measurement. A full calculation of systematics would also include a data driven measurement on lepton ID, and other data samples for the measurement of initial state radiation, final state radiation, parton distribution



Pretag Signal Acceptances (%)				
Scenario	1 jet	2 jet	3 jet	4 jet
I	$.132 \pm .002$	$.320 \pm .003$	$.083 \pm .001$	$.0182 \pm .0007$
II	$.116 \pm .002$	$.286 \pm .003$	$.075 \pm .001$	$.0169 \pm .0006$
III	$.178 \pm .003$	$.422 \pm .004$	$.105 \pm .002$	$.0217 \pm .0009$
IV	$.157 \pm .002$	$.375 \pm .004$	$.096 \pm .003$	$.0199 \pm .0009$
Tagged Signal Acceptances (%)				
Scenario	1 jet	2 jet	3 jet	4 jet
I	$.0427 \pm .002$	$.1732 \pm .00007$	$.0442 \pm .002$	$.0095 \pm .0006$
II	$.0380 \pm .002$	$.1555 \pm .00006$	$.0395 \pm .002$	$.0088 \pm .0004$
III	$.0687 \pm .003$	$.2584 \pm .00010$	$.0636 \pm .003$	$.0133 \pm .0009$
IV	$.0617 \pm .003$	$.2288 \pm .00009$	$.0580 \pm .003$	$.0121 \pm .0008$

Table 4.3:  $WH$  signal acceptances ( $\epsilon_{WH \rightarrow l\nu b\bar{b}}^{MC}$ ) for pretag and tag selection divided into jet-bins in the four scenarios. Statistical uncertainty only.

functions and jet energy scale. Such study was not performed because many samples are not yet available for Gen7 release, furthermore -at this stage- we are not aiming to a cross section limit measurement. Anyway, for reference, previous analyses[56] find a total systematic uncertainty between 6% and 10% depending of the used tagging category. The luminosity uncertainty of 6% is the dominant contribution.

At the moment the  $WH$  search at CDF is performed exclusively in the jet-bin 2 with jet-bin 1, 3 and 4 being considered control samples.

Pretag Expected $WH$ Signal Events				
Scenario	1 jet	2 jet	3 jet	4 jet
I	$0.196 \pm 0.003$	$0.474 \pm 0.004$	$0.123 \pm 0.002$	$0.027 \pm 0.001$
II	$0.172 \pm 0.002$	$0.423 \pm 0.004$	$0.111 \pm 0.002$	$0.025 \pm 0.001$
III	$0.263 \pm 0.004$	$0.625 \pm 0.006$	$0.155 \pm 0.003$	$0.032 \pm 0.001$
IV	$0.232 \pm 0.004$	$0.555 \pm 0.005$	$0.142 \pm 0.005$	$0.029 \pm 0.001$
Tagged Expected $WH$ Signal Events				
Scenario	1 jet	2 jet	3 jet	4 jet
I	$0.063 \pm 0.003$	$0.26 \pm 0.01$	$0.065 \pm 0.003$	$0.014 \pm 0.001$
II	$0.056 \pm 0.003$	$0.23 \pm 0.01$	$0.058 \pm 0.003$	$0.013 \pm 0.001$
III	$0.102 \pm 0.005$	$0.38 \pm 0.02$	$0.094 \pm 0.004$	$0.020 \pm 0.001$
IV	$0.091 \pm 0.004$	$0.34 \pm 0.01$	$0.086 \pm 0.004$	$0.018 \pm 0.001$

Table 4.4:  $WH$  MC events passing pretag and tag selection divided into jet-bins in the four scenarios. Statistical and luminosity uncertainty included.

# Chapter 5

## Background Analysis

The increased  $WH$  signal acceptance is useful only if the background is understood.

The  $W$ +jets sample composition with the  $W$  lepton identified in the central region of the CDF detector has been extensively studied in other analyses[57, 58, 56]. The background composition is obtained using a combination of data driven estimates and MC samples<sup>1</sup>. This analysis is a benchmark to extend the same technique to the forward region of CDF detector, indeed the plug selection we use relies on criteria very close to what used for central electron identification.

The background composition is estimated and compared to data in the four scenarios:

**Scenario I:** Generation 6 of CDF software release with  $\cancel{E}_T > 20$  GeV,

**Scenario II:** Generation 6 of CDF software release with  $\cancel{E}_T > 25$  GeV,

**Scenario III:** Generation 7 of CDF software release with  $\cancel{E}_T > 20$  GeV,

**Scenario IV:** Generation 7 of CDF software release with  $\cancel{E}_T > 25$  GeV.

It is important to realize that the new Backward tracking algorithm (see section 3.1.1) is included in Gen7 both at selection level and at  $b$ -tagging level.

### 5.1 Overview of the Background Composition

The components of the background to  $WH$  in the  $W$ +jets sample can be classified in four categories:

---

<sup>1</sup>The CDF naming convention for the overall technique described in this chapter is “Method 2” background determination.

**Non- $W$  QCD:** events without a real  $W$  but passing selection. They are due to QCD production of jets faking the charged lepton and the  $\cancel{E}_T$  of the neutrino.

**$W$ +Heavy Flavors:** events with a  $W$  produced in association with one or more Heavy Flavor (HF) jets, those containing real bottom and charm hadrons. In the following we will call it simply HF.

**$W$ +Light Flavors/Mistags:** events where the  $W$  is produced in association with one or more Light Flavor (LF) jets. This component is drastically reduced by the request of one or more **SecVtx**  $b$ -tagged jets, the remaining events are due to mismeasured jet tracks and long living light flavors hadrons. In the following we will name this as LF when talking of the pretag sample or “mistags” when talking of the tagged sample.

**Electroweak and Top:** events due to other physical processes known contribution:  $t\bar{t}$  production, single-top, diboson production and  $Z$ +jets component.

The four components are combined and compared to data passing selection both at “pretag” and “tagged” level in each jet-bin (1, 2, 3 and  $\geq 4$  jets).

### 5.1.1 Background Estimate

The background estimate relies on data to normalize or constrain the sample composition as much as possible. This is necessary because the LF and HF contribution, which dominate the pretag and the tagged sample, are not known a priori, due to large corrections to the calculable cross section at high order.

The normalization is done independently in each jet-bin ( $N_j^W$  with  $j = 1, 2, 3, 4$ ) and it is obtained from the pretag data sample ( $N_j$ ) after the subtraction of fake  $W$  events (the non- $W$  fraction  $F_j^{nonW}$ ), electroweak, *single-top* and  $t\bar{t}$  production ( $N_j^{EWK}$ ,  $N_j^{s-top}$  and  $N_j^{t\bar{t}}$ ):

$$N_j^W = N_j(1 - F_j^{nonW}) - N_j^{t\bar{t}} - N_j^{EWK} - N_j^{s-top}. \quad (5.1)$$

While the LF contribution to the tagged sample is parametrized together with the mistags, the HF fractions and the tagging efficiencies in each jet-bin (respectively  $f_j^{HF}$  and  $\epsilon_j^{HF}$ ) are evaluated using leading-order MC samples which take into account  $W + b\bar{b}$ ,  $W + c\bar{c}$  and  $W + c$  processes. Indeed we

assume that the fractions are more stable at higher order than the absolute cross sections. The final HF tagged contribution can be written as:

$$\tau_j^{HF} = N_j^W f_j^{HF} K \epsilon_j^{HF} \Phi, \quad (5.2)$$

where  $\Phi$  is the **SecVtx** scale factor (see Table 3.1) and “ $K$ ” is an overall HF scale calibrated in generic jets, that is meant to cover a misestimate of  $g \rightarrow b\bar{b}$  and  $g \rightarrow c\bar{c}$ , processes which can contribute to the heavy flavor content.

## 5.2 Non- $W$ QCD Background

The pretag event selection (see section 4.2) is optimized for  $W^\pm \rightarrow e^\pm \nu$  identification but QCD jet production can fake the signature of a true  $W$  event. The high energy isolated electron can be faked by conversions or by jets containing semileptonic decays of  $b$  or  $c$  hadrons which also fluctuate in low isolation region, while the  $\cancel{E}_T$  of the neutrino can be faked by jet energy mismeasurements, detector effects or wrong event vertex reconstruction.

The QCD contribution to the pretag sample is estimated with the “Isolation versus  $\cancel{E}_T$ ” method: we select an inclusive sample removing the isolation and missing energy requirements and all collected events are disposed in a “ $IsoRel$  vs  $\cancel{E}_T$ ” plot divided in the following regions (see Fig. 5.1):

**region A:**  $IsoRel > 0.2$  and  $\cancel{E}_T < 10$  GeV,

**region B:**  $IsoRel \leq 0.1$  and  $\cancel{E}_T < 10$  GeV,

**region C:**  $IsoRel > 0.2$  and  $\cancel{E}_T \geq 20$  GeV (or  $\cancel{E}_T \geq 25$  GeV in scenario II and IV),

**region D:**  $IsoRel \leq 0.1$  and  $\cancel{E}_T \geq 20$  GeV (or  $\cancel{E}_T \geq 25$  GeV in scenario II and IV),

the region D is the signal region, the regions A, B and C are sidebands used to calculate background contribution. Figure 5.2 shows the  $IsoRel$  vs  $\cancel{E}_T$  plane for events passing selection and with at least one jet, for the different scenarios.

If we assume that the lepton isolation and the  $\cancel{E}_T$  are uncorrelated in QCD processes, the fraction of non- $W$  events in the region D in the different jet-bins can be calculated using:

$$f_j^{nonW} = \frac{N_j^B \times N_j^C}{N_j^A \times N_j^D}, \quad (5.3)$$

$$N_j^{nonW} = f_j^{nonW} N_j^D. \quad (5.4)$$

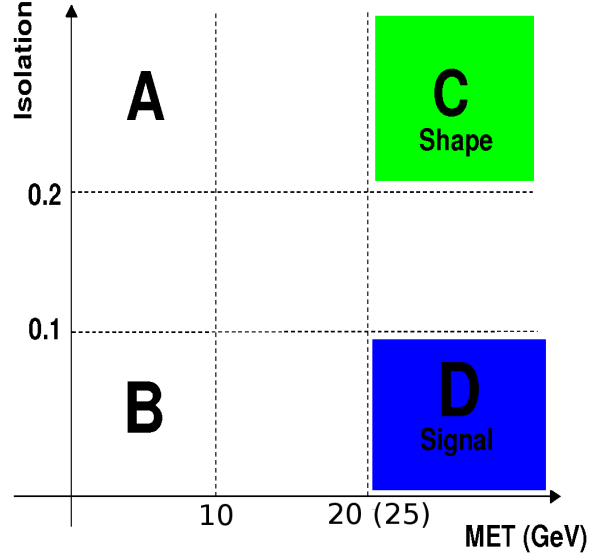


Figure 5.1: Location of the four regions in the  $IsoRel$  vs  $E_T$  plane.

Scenario	Non $W$ Fractions			
	1 jet	2 jet	3 jet	4 jet
I	$0.30 \pm 0.09$	$0.32 \pm 0.09$	$0.48 \pm 0.15$	$0.44 \pm 0.14$
II	$0.10 \pm 0.03$	$0.17 \pm 0.05$	$0.3 \pm 0.1$	$0.36 \pm 0.12$
III	$0.40 \pm 0.12$	$0.38 \pm 0.11$	$0.52 \pm 0.16$	$0.54 \pm 0.17$
IV	$0.15 \pm 0.05$	$0.21 \pm 0.06$	$0.35 \pm 0.11$	$0.41 \pm 0.13$

Table 5.1: Summary of non- $W$  fractions ( $f_j^{nonW}$ ) in the pretag sample in the four different scenarios.

This assumption has been extensively tested at CDF and is used in electroweak and top studies and in  $WH$  searches[59]. It was tested and used also in the forward region[60] measuring the inclusive  $W$  production. The systematic error on the non- $W$  fraction is derived by studies performed on jet-enriched samples and it is conservatively quoted as 30%.

Before computing the final non- $W$  fraction, events in each side band region are corrected by subtracting the amount of electroweak physics processes (mostly diboson and  $t\bar{t}$  production, see section 5.5). The final non- $W$  fractions in the different categories are summarized in Table 5.1. Of course the fractions are the key element to evaluate the  $W$ +jets contribution.

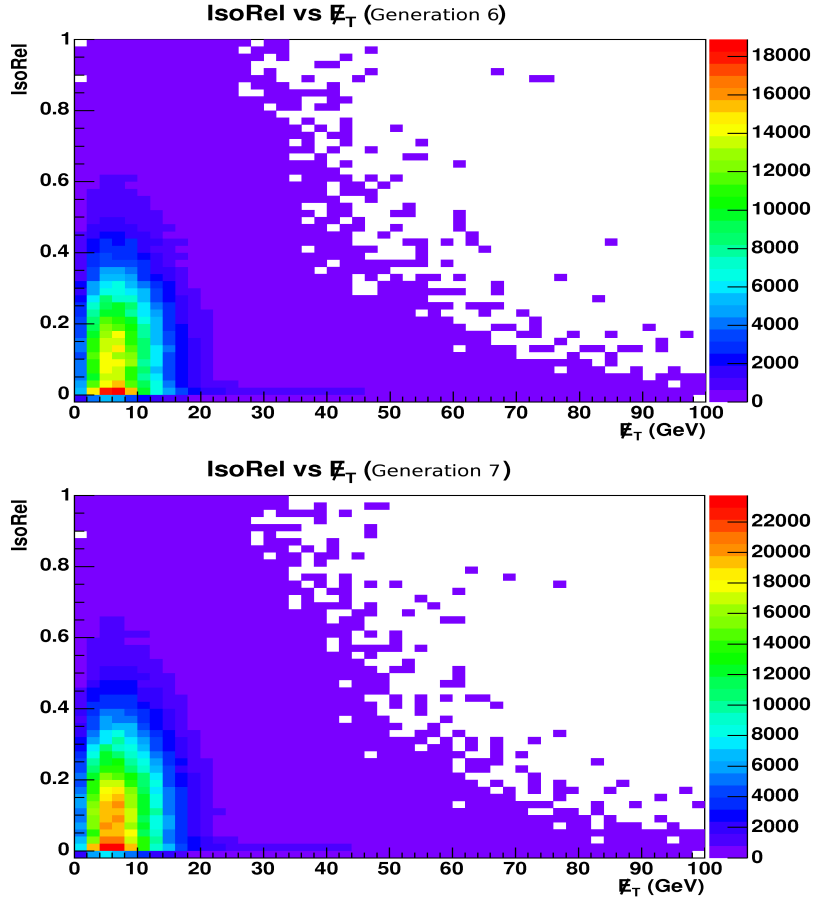


Figure 5.2: *IsoRel* vs  $\cancel{E}_T$  plots for events with at least one jet passing pretag selection. Top: Gen6 reconstruction. Bottom: Gen7 reconstruction.

### 5.2.1 Tagged non- $W$ Background

QCD events also contribute to the composition of the final tagged sample. It is possible to use the sideband method to estimate this contribution in two different ways: the “Tagged Method” and the “Tag Rate Method”.

The “Tagged Method” applies exactly the same criteria of the pretagged sample to infer the number of tagged QCD events in the signal region but the ratio of Eq. 5.3 is computed using the tagged events of the sidebands: ( $\tau_j$ ):

$$\tau_j^{nonW(I)} = \frac{\tau_j^B \times \tau_j^C}{\tau_j^A}. \quad (5.5)$$

This method is affected by the larger contents of true  $b$ -hadrons inside the

Expected Tags (Tagged Method)				
Scenario	1 jet	2 jet	3 jet	4 jet
I	$104 \pm 31$	$45 \pm 14$	$16 \pm 5$	$5 \pm 2$
II	$25 \pm 8$	$16 \pm 5$	$6 \pm 2$	$2.3 \pm 0.9$
III	$292 \pm 88$	$94 \pm 28$	$39 \pm 12$	$28 \pm 9$
IV	$79 \pm 24$	$33 \pm 10$	$15 \pm 5$	$13 \pm 4.6$
Expected Tags (Tag Rate Method)				
Scenario	1 jet	2 jet	3 jet	4 jet
I	$234 \pm 71$	$69 \pm 22$	$15 \pm 5$	$n/a$
II	$61 \pm 19$	$27 \pm 9$	$6.4 \pm 2.5$	$n/a$
III	$523 \pm 158$	$174 \pm 53$	$64 \pm 21$	$23 \pm 9$
IV	$138 \pm 43$	$69 \pm 22$	$34 \pm 11$	$13 \pm 5$

Table 5.2: Expected tagged non- $W$  events evaluated with the “Tagged Method” (upper part), and with the “Tag Rate Method”(lower part).

QCD sample, which might bias upward the background estimate. It also has a large statistical uncertainty.

The “Tag Rate Method” use the pretag non- $W$  fraction ( $f_j^{nonW}$ ) and the expected tag rate of QCD events in the signal region, where the tag rate of region D is inferred from control region B assuming that the tagging efficiency on QCD events do not depends by  $\cancel{E}_T$ .

The tag rate is defined as the number of tagged jets divided by the number of “taggable” jets<sup>2</sup>. So the number of tagged QCD events in region D is:

$$\tau_j^{nonW(II)} = f_j^{nonW} \times r_j^B \times N_j^D(taggable), \quad (5.6)$$

with  $r_j^B = \tau_j^B / N_j^B(taggable)$ . Because of the presence of semileptonic decays in true HF QCD events, the tag rate is not completely independent from  $\cancel{E}_T$  and the “Tag Rate Method” seem to give lower than expected values.

Tables 5.2 show the number of tags obtained from the two methods. Following the standard CDF approach, the final estimate used is the weighed average of the two (see the summary Tables 5.7, 5.8, 5.9 and 5.10).

### 5.3 Heavy Flavor Background

The associated  $W$ +HF production contributes significantly to the total composition of the  $W$ +jets tagged sample[63]. It would be possible to obtain a

<sup>2</sup>A jet is “taggable” if `SecVtx` algorithm can be applied to it, see section 3.7.1.



good representation of  $W$ +HF sample combining few MC generated processes, like  $W + c\bar{c}$ ,  $W + b\bar{b}$ ,  $W + c$ , but the available simulations can produce only next-to-leading-order calculation of such matrix elements. As a result, the overall normalization of these calculations has large theoretical uncertainty.

The handle to parametrize  $W$ +HF composition is to measure the relative fractions,  $f_j^{HF}$ , of each component using MC and to derive the overall normalization,  $N_j^W$ , from an independent estimate (see Eq. 5.1). Some care must be used in the recombination of the sample, in order to avoid double-counting (see below). However much of the systematic uncertainties cancel in the ratio and the remaining contribution (“K” correction) can be estimated from the generic jet samples. After the calculation of the HF tagging efficiency,  $\epsilon_j^{HF}$ , the final HF contribution to the tagged sample is given by Eq. 5.2. In the following we will explain how to calculate each component of Eq. 5.2.

A reasonable representation of  $W$ +jet sample is obtained combining together four different processes generated with a given multiplicity of light flavor partons:

- $W$  plus 0, 1, 2, 3, 4 partons (to describe the generic  $W + \geq 0$  jets),
- $W + b\bar{b}$  plus 0, 1, 2 partons,
- $W + c\bar{c}$  plus 0, 1, 2 partons,
- $W + c$  (or  $W + \bar{c}$ ) plus 0, 1, 2, 3 partons,

A total of 15 MC sample is used. The  $W$  is forced to decay into electrons as acceptance for  $W^\pm \rightarrow \tau^\pm \nu$  has been estimated from MC to be less than 1%. All samples are produced at leading-order with ALPGEN[36] as the Matrix Element (ME) generator, interfaced with PYTHIA[37] to produce the Parton Shower (PS). Table 5.3 summarizes the samples used along with the respective leading-order cross sections and the approximate number of events initially generated.

Before the determination of HF fraction, some care must be used to recombine the 15 samples together to avoid double counting. In fact in the ALPGEN-PYTHIA interface there is no attempt to match explicitly the final state heavy flavors hadrons to the initial ME  $b$  and  $c$  quarks<sup>3</sup>. This implies that in some cases the same final state arises in two different samples<sup>4</sup> producing a double counting that degrades the relative weight of the samples.

<sup>3</sup>The same consideration is valid also for LF but we are not interested in their distinction.

<sup>4</sup>Specifically the radiation from a parton shower in a  $W + n$  partons MC can produce jets which cover also part of the phase space described by the  $W + (n + 1)$  partons MC sample.

Sample	Cross Section (pb)	Generated Events
$W + 0p$	1810	1 M
$W + 1p$	225	1 M
$W + 2p$	35.3	1 M
$W + 3p$	5.59	1 M
$W + 4p$	1.03	0.5 M
$Wb\bar{b} + 0p$	2.98	1.5 M
$Wb\bar{b} + 1p$	0.888	1.5 M
$Wb\bar{b} + 2p$	0.287	1.5 M
$Wc\bar{c} + 0p$	5	2 M
$Wc\bar{c} + 1p$	1.79	2 M
$Wc\bar{c} + 2p$	0.628	2 M
$Wc + 0p$	17.1	2 M
$Wc + 1p$	3.39	2 M
$Wc + 2p$	0.507	2 M
$Wc + 3p$	0.083	2 M

Table 5.3: MC samples used to estimate the various contributions to the  $W$ +HF background. The second column lists the leading order cross section of the samples, the third column shows the approximate number of generated events.

The “overlap removal” technique used in this work consists in the selection of HF production from the shower or matrix element on the base of quantities reconstructed at detector level, i.e. jets. In fact the shower model is tuned on the more collinear gluon splitting pairs but fails in the limit of large opening angles. So ME and shower model predict very different  $\Delta R$  (separation) between heavy quarks. We explicitly veto events where the ME heavy flavor quarks wind up in the same jet, and we also remove events where the HF quark pair from the shower is divided in two jets. In both cases, the distinction is made with simple geometric cuts, defining “in a jet” to be within  $|\Delta R| < 0.4$  of a jet with a Level-5 corrected  $E_T$  above 20 GeV. The final procedure for the determination of the  $f_j^{HF}$ ’s is the following:

- each MC sample is assigned a weight:

$$w^m = \frac{\sigma^m}{N_{z_0}^m}, \quad (5.7)$$

where  $m$  varies from 1 to 15 (the different samples),  $\sigma^m$  is the cross section of a given sample (see Table 5.3) and  $N_{z_0}^m$  is the total number of good run events with  $|z_0^{vtx}| < 60$  cm.

- After pretag selection, the denominators of the HF fractions are calculated in each jet-bin:

$$Den_j = \sum_{m=1}^{15} w^m N_j^m, \quad (5.8)$$

where  $N_j^m$  are the “unique” events (i.e. not vetoed by the “overlap removal”) falling in each jet-bin for sample  $m$ .

- Each selected jet is defined as HF jet or not: HF jets have a  $b$  or  $c$  parton (at generation level) laying inside the jet cone ( $|\Delta R(jet, parton)| < 0.4$ ). We distinguish four HF categories depending on the number of jets containing HF partons (1 or  $\geq 2$  HF) and on the kind of HF parton ( $b$  or  $c$ ). The categories are:  $1B$ ,  $2B$ ,  $1C$  and  $2C$ .
- The numerators are defined by the sum of the events in each HF category (with weights given by Eq. 5.7) over all the samples in each jet-bin:

$$Num_j^{HF} = \sum_{m=1}^{m=15} w_j^m N_j^{m, HF}. \quad (5.9)$$

The computation of the  $f_j^{HF}$ 's is straightforward but it needs the  $K$  factor correction:

$$K^{HF} = \frac{f_{data}^{HF}}{f_{MC}^{HF}}, \quad (5.10)$$

that takes into account the possible dependence of HF fractions from NLO (and beyond) effects.

The calibration[64] is carried out in an independent sample of generic jets and on ALPGEN QCD samples with LF and HF final states. Thanks to the large statistics of QCD data samples it is possible to build up “tag-mass” template fits of the heavy flavor content of the jets. The discrepancy between the HF composition inferred in data and the values resulting by MC samples is the  $K$  correction (see Eq. 5.10). For all the HF categories the  $K$  correction is consistent with unity but a systematic uncertainty of 0.3 is conservatively quoted to take into account discrepancies in the lower jet-bins. Table 5.4 summarizes the final HF fraction:

$$f_j^{HF} = K \frac{Num_j^{HF}}{Den_j}. \quad (5.11)$$

The HF fractions can be used in the construction of the pretag sample shapes but are of marginal importance because of the overwhelming LF contribution. On the other hand HF hadrons have a fundamental role in the tagged sample composition.

The last ingredient of Eq. 5.2 is the tagging efficiency of each HF category ( $\epsilon_j^{HF}$ ). It is derived from the 15 MC samples, using the same procedure of weighed sums but with the identification of **SecVtx** tagged jets. The final computation of the tagging efficiency is corrected for the **SecVtx** data/MC scale factor (see Table 3.1) where we double the systematic uncertainty when tagged  $c$ -jets are considered. Tagging efficiencies for each HF category are summarized in Table 5.5.

## 5.4 Light Flavors/Mistags

The  $W$ +jet events coming from a light flavor quark are the dominant contribution to the pretag sample. The LF pretag fraction ( $f_j^{LF}$ ) can be derived in the same way as the HF fraction. Anyway after the  $b$ -tagging request only a very small fraction of LF jets is  $b$ -tagged so the parametrization of this background is included in the study of the “mistag” component. The latter is generated from wrongly tagged jets due to track misidentification or interaction with the detector material and the beam pipe.

Heavy Flavor Fractions				
Scenario I: $\cancel{E}_T > 20$ GeV, Generation 6				
	1 jet	2 jet	3 jet	4 jet
$f_j^{1B}$	$0.0054 \pm 0.0016$	$0.014 \pm 0.004$	$0.023 \pm 0.007$	$0.029 \pm 0.009$
$f_j^{2B}$	$n/a$	$0.0065 \pm 0.0020$	$0.015 \pm 0.005$	$0.028 \pm 0.008$
$f_j^{1C}$	$0.050 \pm 0.015$	$0.081 \pm 0.024$	$0.114 \pm 0.034$	$0.122 \pm 0.037$
$f_j^{2C}$	$n/a$	$0.011 \pm 0.003$	$0.031 \pm 0.009$	$0.055 \pm 0.016$
Scenario II: $\cancel{E}_T > 25$ GeV, Generation 6				
	1 jet	2 jet	3 jet	4 jet
$f_j^{1B}$	$0.0053 \pm 0.0016$	$0.015 \pm 0.005$	$0.023 \pm 0.007$	$0.029 \pm 0.009$
$f_j^{2B}$	$n/a$	$0.0066 \pm 0.0020$	$0.015 \pm 0.005$	$0.028 \pm 0.008$
$f_j^{1C}$	$0.048 \pm 0.015$	$0.079 \pm 0.024$	$0.111 \pm 0.034$	$0.120 \pm 0.036$
$f_j^{2C}$	$n/a$	$0.0112 \pm 0.0034$	$0.031 \pm 0.009$	$0.054 \pm 0.016$
Scenario III: $\cancel{E}_T > 20$ GeV, Generation 7				
	1 jet	2 jet	3 jet	4 jet
$f_j^{1B}$	$0.0055 \pm 0.0016$	$0.014 \pm 0.004$	$0.024 \pm 0.007$	$0.031 \pm 0.009$
$f_j^{2B}$	$n/a$	$0.006 \pm 0.002$	$0.015 \pm 0.004$	$0.028 \pm 0.009$
$f_j^{1C}$	$0.050 \pm 0.015$	$0.080 \pm 0.024$	$0.112 \pm 0.034$	$0.123 \pm 0.037$
$f_j^{2C}$	$n/a$	$0.011 \pm 0.003$	$0.031 \pm 0.009$	$0.055 \pm 0.017$
Scenario IV: $\cancel{E}_T > 25$ GeV, Generation 7				
	1 jet	2 jet	3 jet	4 jet
$f_j^{1B}$	$0.005 \pm 0.002$	$0.014 \pm 0.004$	$0.023 \pm 0.007$	$0.030 \pm 0.009$
$f_j^{2B}$	$n/a$	$0.006 \pm 0.002$	$0.015 \pm 0.005$	$0.028 \pm 0.009$
$f_j^{1C}$	$0.048 \pm 0.015$	$0.078 \pm 0.024$	$0.109 \pm 0.033$	$0.121 \pm 0.036$
$f_j^{2C}$	$n/a$	$0.011 \pm 0.003$	$0.030 \pm 0.009$	$0.054 \pm 0.016$

Table 5.4: Heavy Flavor fractions classified in the different categories.  $1B$ : events with 1  $b$ -jet.  $2B$ : events with  $\geq 2$   $b$ -jet.  $1C$ : events with 1  $c$ -jet.  $2C$ : events with  $\geq 2$   $c$ -jet. The correction  $K = 1 \pm 0.3$  is included.

Heavy Flavor Tagging Efficiency				
Scenario I: $\cancel{E}_T > 20$ GeV, Generation 6				
	1 jet	2 jet	3 jet	4 jet
$\epsilon_j^{1B}$	$0.28 \pm 0.02$	$0.28 \pm 0.03$	$0.306 \pm 0.017$	$0.330 \pm 0.026$
$\epsilon_j^{2B}$	$n/a$	$0.504 \pm 0.024$	$0.506 \pm 0.026$	$0.523 \pm 0.028$
$\epsilon_j^{1C}$	$0.053 \pm 0.005$	$0.056 \pm 0.005$	$0.063 \pm 0.006$	$0.067 \pm 0.008$
$\epsilon_j^{2C}$	$n/a$	$0.117 \pm 0.012$	$0.119 \pm 0.013$	$0.117 \pm 0.012$
Scenario II: $\cancel{E}_T > 25$ GeV, Generation 6				
	1 jet	2 jet	3 jet	4 jet
$\epsilon_j^{1B}$	$0.284 \pm 0.020$	$0.282 \pm 0.028$	$0.306 \pm 0.017$	$0.334 \pm 0.027$
$\epsilon_j^{2B}$	$n/a$	$0.504 \pm 0.025$	$0.511 \pm 0.026$	$0.523 \pm 0.029$
$\epsilon_j^{1C}$	$0.054 \pm 0.006$	$0.056 \pm 0.005$	$0.064 \pm 0.006$	$0.068 \pm 0.008$
$\epsilon_j^{2C}$	$n/a$	$0.114 \pm 0.011$	$0.12 \pm 0.013$	$0.118 \pm 0.012$
Scenario III: $\cancel{E}_T > 20$ GeV, Generation 7				
	1 jet	2 jet	3 jet	4 jet
$\epsilon_j^{1B}$	$0.334 \pm 0.021$	$0.340 \pm 0.027$	$0.376 \pm 0.020$	$0.390 \pm 0.026$
$\epsilon_j^{2B}$	$n/a$	$0.597 \pm 0.028$	$0.594 \pm 0.029$	$0.602 \pm 0.031$
$\epsilon_j^{1C}$	$0.078 \pm 0.007$	$0.085 \pm 0.007$	$0.092 \pm 0.008$	$0.107 \pm 0.011$
$\epsilon_j^{2C}$	$n/a$	$0.153 \pm 0.014$	$0.169 \pm 0.017$	$0.170 \pm 0.016$
Scenario IV: $\cancel{E}_T > 25$ GeV, Generation 7				
	1 jet	2 jet	3 jet	4 jet
$\epsilon_j^{1B}$	$0.331 \pm 0.021$	$0.347 \pm 0.030$	$0.375 \pm 0.020$	$0.396 \pm 0.027$
$\epsilon_j^{2B}$	$n/a$	$0.598 \pm 0.028$	$0.597 \pm 0.029$	$0.600 \pm 0.031$
$\epsilon_j^{1C}$	$0.079 \pm 0.007$	$0.085 \pm 0.007$	$0.093 \pm 0.008$	$0.105 \pm 0.011$
$\epsilon_j^{2C}$	$n/a$	$0.157 \pm 0.014$	$0.167 \pm 0.017$	$0.171 \pm 0.016$

Table 5.5: Heavy Flavor tagging efficiency  $\epsilon_j^{HF}$  in each category. 1B: events with 1  $b$ -jet. 2B: events with  $\geq 2$   $b$ -jet. 1C: events with 1  $c$ -jet. 2C: events with  $\geq 2$   $c$ -jet. The SecVtx tagging scale factor  $\Phi = 0.95 \pm 0.04$  is included.

The rate of  $W$ +mistag jets is derived from a sample of events collected with an inclusive jet-based trigger with no heavy flavor requirement<sup>5</sup>. The mistag parametrization[61] is obtained using “negative tags” (see Fig. 3.10), i.e.  $b$ -jets that appear travelling back toward the primary vertex. Resolution and material effects are expected to produce false tags in a symmetric pattern around the primary interaction vertex. The mistag rate is corrected for the effects of long-lived light flavor hadrons to take into account the LF contamination. The complete parametrization is available only for Gen6 release. In the estimate for Scenario III and IV (Gen7) this could produce an underestimate of the mistag background due to new tracks reconstructed with Backward Tracking algorithm.

The mistag rate obtained from negative tags is parametrized<sup>6</sup> in bins of:

- $N$ -jets,
- jet  $E_T$ ,
- jet  $\eta$ ,
- track multiplicity within a jet,
- total  $E_T$  of the event,
- number of interaction vertices,
- the  $z$  vertex position.

The total mistag probability of an event is obtained adding the mistag probability computed for each taggable jet.

The uncertainty on the mistag estimate includes the statistical uncertainties from the parametrization, accounting for correlations between jets which fall in the same jet-bin of the mistag matrix. For Gen6 data an additional 5.9% contribution from all systematic uncertainties was estimate. Due to the lack of information for Gen7 we decided not to use this additional contribution in order to avoid the risk of covering up other effects. Therefore -for the sake of comparison- the systematics was not included in the final estimate for all Scenarios.

---

<sup>5</sup>The presence of a small HF contamination in the sample is a source of systematic uncertainty.

<sup>6</sup>The naming convention for this parametrization is “Mistag Matrix”.

Sample	Cross Section (pb)	Generated Events
$WW$	$12.4 \pm 0.25$	2.2 M
$WZ$	$3.96 \pm 0.06$	2.2 M
$ZZ$	$1.58 \pm 0.05$	2.2 M
$s\text{-top}(t\text{-ch})$	$0.88 \pm 0.05$	4.5 M
$s\text{-top}(s\text{-ch})$	$1.98 \pm 0.08$	1 M
$t\bar{t}$	$6.7 \pm 0.8$	1 M
$Z^0 \rightarrow \tau^+\tau^-$	$265 \pm 30$	10 M

Table 5.6: MC samples used to derive electroweak and top background. The second column lists the next-to-leading order cross section of the samples, the third column the approximate number of generated events. *Single*-top production is evaluated in the  $t$  and  $s$  channels. The  $Z^0 \rightarrow \tau^+\tau^-$  sample is composed by  $Z^0 \rightarrow \tau^+\tau^- + [0 - 2]\text{jets}$  samples.

## 5.5 Electroweak and Top Background

Besides the contribution of QCD,  $W$ +LF and  $W$ +HF events, several other physics processes involving a real  $W$  enter into the final tagged sample. The main processes are diboson production (i.e.  $WW$ ,  $WZ$  and  $ZZ$  with a misidentified lepton) with a  $b\bar{b}$  or  $c\bar{c}$  in the final state,  $t\bar{t}$  production with dilepton decay with a missed lepton or single lepton decay with a merged jet, single top production and  $Z^0 \rightarrow \tau^+\tau^-$ , with a  $\tau \rightarrow l + \nu$  and a **SecVtx** tag on the opposite-side three-prong hadronic decay.

Both diboson and top production processes have well defined theoretical cross sections and the MC samples used in their simulation are calculated to next-to-leading order. The  $Z^0 \rightarrow \tau^+\tau^-$  processes should be evaluated with the same procedure of  $W + Q\bar{Q}$  decays but the small acceptance and cross section measure performed at CDF[62] make reliable to derive this contribute from MC. Table 5.6 display the cross sections used.

The contribution of each physics process is derived in each jet bin the same way as  $WH$  signal (see section 4.4):

$$N_j^{phy} = \mathcal{A}_j^{phy} \sigma^{phy} \int \mathcal{L} dt. \quad (5.12)$$

where  $N_j^{phy}$  indicates the pretag or the tagged values depending of the situations and  $\mathcal{A}$  is the total detection efficiency: inclusive of MC derived acceptance, trigger efficiency and scale factors.



Background Summary (Scenario I) $\cancel{E}_T > 20$ GeV, Generation 6				
	1 jet	2 jet	3 jet	4 jet
mistag	$175 \pm 2$	$66.7 \pm 1.4$	$17.0 \pm 0.8$	$4.8 \pm 0.4$
non- $W$ (Ave)	$124 \pm 29$	$52 \pm 12$	$15 \pm 4$	$4.96 \pm 1.91$
$t\bar{t}$	$1.01 \pm 0.15$	$7.7 \pm 1.1$	$16.3 \pm 2.3$	$22.0 \pm 3.1$
<i>single</i> -top	$8.3 \pm 0.6$	$15.2 \pm 1.0$	$5.1 \pm 0.3$	$1.41 \pm 0.11$
diboson	$2.1 \pm 0.2$	$4.7 \pm 0.3$	$1.29 \pm 0.11$	$0.39 \pm 0.05$
$Z^0 \rightarrow \tau^+ \tau^-$	$1.9 \pm 0.3$	$1.0 \pm 0.1$	$0.36 \pm 0.06$	$0.09 \pm 0.02$
$W + b\bar{b}$	$33 \pm 11$	$27 \pm 7$	$6 \pm 2$	$1.0 \pm 0.5$
$W + c, W + c\bar{c}$	$57 \pm 20$	$21 \pm 6$	$4.7 \pm 1.6$	$0.6 \pm 0.3$
Total expected	$403 \pm 36$	$195 \pm 15$	$67 \pm 5$	$35 \pm 4$
$WH$ expected	$0.063 \pm 0.003$	$0.26 \pm 0.01$	$0.065 \pm 0.003$	$0.014 \pm 0.001$
Data events	415	231	66	35

Table 5.7: Summary of expected tagged background in Scenario I. The non- $W$  component is the weighed average of values in Tables 5.2. The 1 $B$  and 2 $B$  HF categories are summed in the total  $W + b\bar{b}$  contribution, the same for 1 $C$ , 2 $C$  and  $W + c, W + c\bar{c}$ .

## 5.6 Background Summary

All the analyzed contributions to the  $W$ +jets sample, with the  $W^\pm \rightarrow e^\pm \nu$  in the forward region of the CDF detector, are listed in Tables 5.7, 5.8, 5.9 and 5.10. They must be compared to the events selected on data summarized in Table 4.1 and reported in the last line of the tables. Figures 5.3, 5.4, 5.5 and 5.6 show background overlaid to data on a log-scale. The agreement is good in all bins for all Gen6 scenarios. For Gen7 scenarios the situation is somewhat different. As bin 1 and 2 are the ones mostly affected by mistag contribution, this is reflected in some discrepancy in those bins. In high multiplicity bins, where top contribution becomes dominant, the agreement improves.

Background Summary (Scenario II)				
$\cancel{E}_T > 25$ GeV, Generation 6				
	1 jet	2 jet	3 jet	4 jet
mistag	$118 \pm 2$	$44.7 \pm 1.2$	$11.5 \pm 0.7$	$3.8 \pm 0.4$
non- $W$ (Ave)	$30 \pm 7$	$18 \pm 4$	$6 \pm 1.5$	$2.3 \pm 0.9$
$t\bar{t}$	$0.94 \pm 0.14$	$7.2 \pm 1.0$	$15.1 \pm 2.1$	$20.4 \pm 2.9$
<i>single</i> -top	$7.6 \pm 0.5$	$13.9 \pm 0.9$	$4.67 \pm 0.32$	$1.27 \pm 0.10$
diboson	$1.89 \pm 0.15$	$4.3 \pm 0.3$	$1.09 \pm 0.10$	$0.35 \pm 0.05$
$Z^0 \rightarrow \tau^+ \tau^-$	$0.93 \pm 0.18$	$0.69 \pm 0.10$	$0.29 \pm 0.05$	$0.07 \pm 0.02$
$W + b\bar{b}$	$32 \pm 10$	$23 \pm 5$	$5.4 \pm 1.3$	$0.6 \pm 0.3$
$W + c, W + c\bar{c}$	$55 \pm 17$	$17 \pm 4$	$4 \pm 1.$	$0.2 \pm 0.2$
Total expected	$246 \pm 21$	$129 \pm 8$	$48 \pm 3$	$29 \pm 3$
$WH$ expected	$0.056 \pm 0.003$	$0.23 \pm 0.01$	$0.058 \pm 0.003$	$0.013 \pm 0.001$
Data events	272	161	45	29

Table 5.8: Summary of expected tagged background in Scenario II. The non- $W$  component is the weighed average of values in Tables 5.2. The  $1B$  and  $2B$  HF categories are summed in the total  $W + b\bar{b}$  contribution, the same for  $1C$ ,  $2C$  and  $W + c, W + c\bar{c}$ .

Background Summary (Scenario III)				
$\cancel{E}_T > 20$ GeV, Generation 7				
	1 jet	2 jet	3 jet	4 jet
mistag	$272 \pm 3$	$107 \pm 2$	$32 \pm 1$	$8.8 \pm 0.6$
non- $W$ (Ave)	$346 \pm 77$	$111 \pm 25$	$45 \pm 10$	$25 \pm 6$
$t\bar{t}$	$1.45 \pm 0.21$	$11.1 \pm 1.6$	$23.9 \pm 3.4$	$32.1 \pm 4.6$
<i>single</i> -top	$13.1 \pm 0.9$	$22.8 \pm 1.5$	$7.5 \pm 0.5$	$1.90 \pm 0.14$
diboson	$3.79 \pm 0.26$	$8.63 \pm 0.53$	$2.46 \pm 0.18$	$0.79 \pm 0.08$
$Z^0 \rightarrow \tau^+ \tau^-$	$3.96 \pm 0.55$	$2.21 \pm 0.25$	$0.53 \pm 0.08$	$0.15 \pm 0.03$
$W + b\bar{b}$	$41 \pm 15$	$35 \pm 9$	$9 \pm 3$	$0.9 \pm 0.8$
$W + c, W + c\bar{c}$	$88 \pm 33$	$34.23 \pm 10$	$7 \pm 2$	$0.7 \pm 0.6$
Total expected	$770 \pm 85$	$333 \pm 29$	$128 \pm 12$	$71 \pm 8$
$WH$ expected	$0.102 \pm 0.005$	$0.38 \pm 0.02$	$0.094 \pm 0.004$	$0.020 \pm 0.001$
Data events	876	418	137	57

Table 5.9: Summary of expected tagged background in Scenario III. The non- $W$  component is the weighed average of values in Tables 5.2. The  $1B$  and  $2B$  HF categories are summed in the total  $W + b\bar{b}$  contribution, the same for  $1C$ ,  $2C$  and  $W + c, W + c\bar{c}$ .

Background Summary (Scenario IV)				
$\cancel{E}_T > 25$ GeV, Generation 7				
	1 jet	2 jet	3 jet	4 jet
mistag	$178 \pm 2$	$68 \pm 2$	$21.6 \pm 0.9$	$6.3 \pm 0.6$
non- $W$ (Ave)	$93 \pm 21$	$39 \pm 9$	$18 \pm 4$	$13 \pm 3$
$t\bar{t}$	$1.36 \pm 0.20$	$10.4 \pm 1.5$	$22.1 \pm 3.1$	$29.6 \pm 4.2$
single-top	$12.0 \pm 0.9$	$20.9 \pm 1.4$	$6.89 \pm 0.47$	$1.73 \pm 0.13$
diboson	$3.50 \pm 0.24$	$7.89 \pm 0.50$	$2.16 \pm 0.16$	$0.71 \pm 0.08$
$Z^0 \rightarrow \tau^+ \tau^-$	$1.9 \pm 0.31$	$1.39 \pm 0.16$	$0.41 \pm 0.07$	$0.12 \pm 0.03$
$W + b\bar{b}$	$41 \pm 13$	$30 \pm 7$	$7 \pm 2$	$0.4 \pm 0.4$
$W + c, W + c\bar{c}$	$88 \pm 28$	$28 \pm 7$	$6 \pm 2$	$0.3 \pm 0.3$
Total expected	$419 \pm 40$	$207 \pm 14$	$84 \pm 6$	$51.8 \pm 5.5$
$WH$ expected	$0.091 \pm 0.004$	$0.34 \pm 0.01$	$0.086 \pm 0.004$	$0.018 \pm 0.001$
Data events	576	269	89	45

Table 5.10: Summary of expected tagged background in Scenario IV. The non- $W$  component is the weighed average of values in Tables 5.2. The  $1B$  and  $2B$  HF categories are summed in the total  $W + b\bar{b}$  contribution, the same for  $1C$ ,  $2C$  and  $W + c$ ,  $W + c\bar{c}$ .

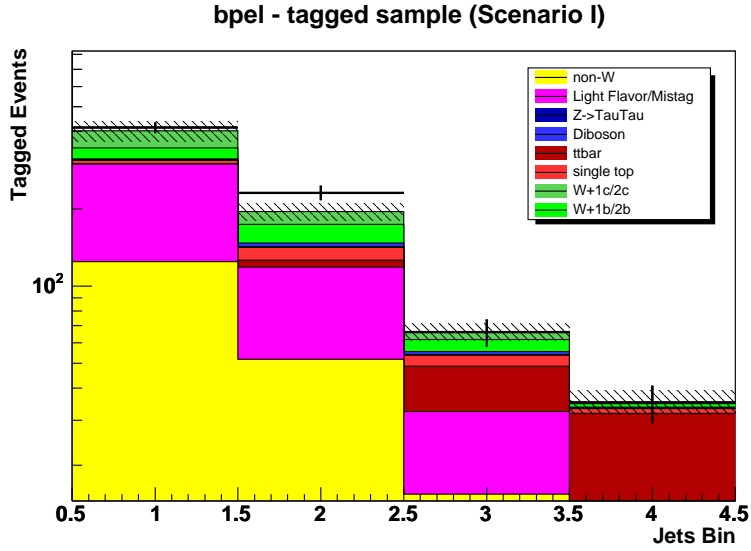


Figure 5.3: Expected background events in tagged sample compared to data (crosses) in the different jet-bins for Scenario I:  $\cancel{E}_T > 20$  GeV, Generation 6.

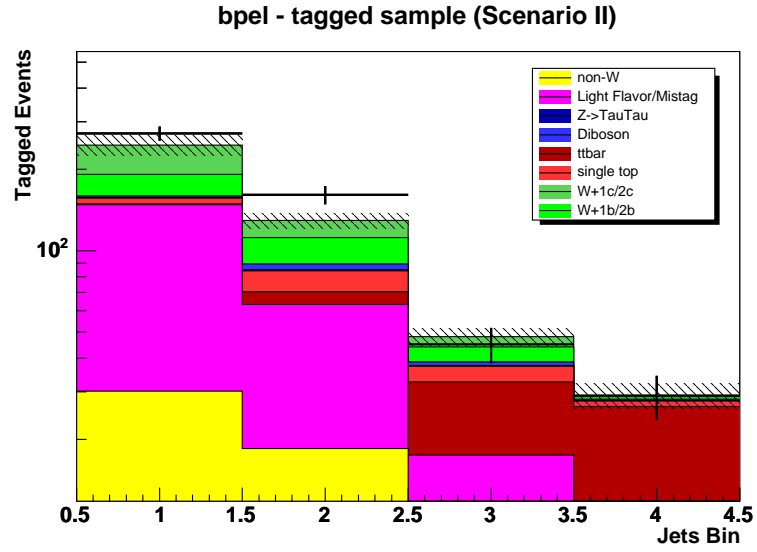


Figure 5.4: Expected background events in tagged sample compared to data (crosses) in the different jet-bins for Scenario II:  $\cancel{E}_T > 25$  GeV, Generation 6.

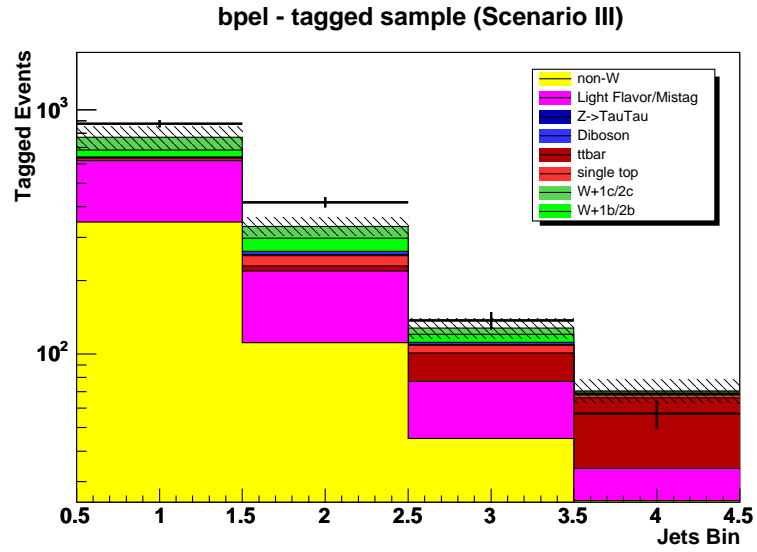


Figure 5.5: Expected background events in tagged sample compared to data (crosses) in the different jet-bins for Scenario III:  $\cancel{E}_T > 20$  GeV, Generation 7.

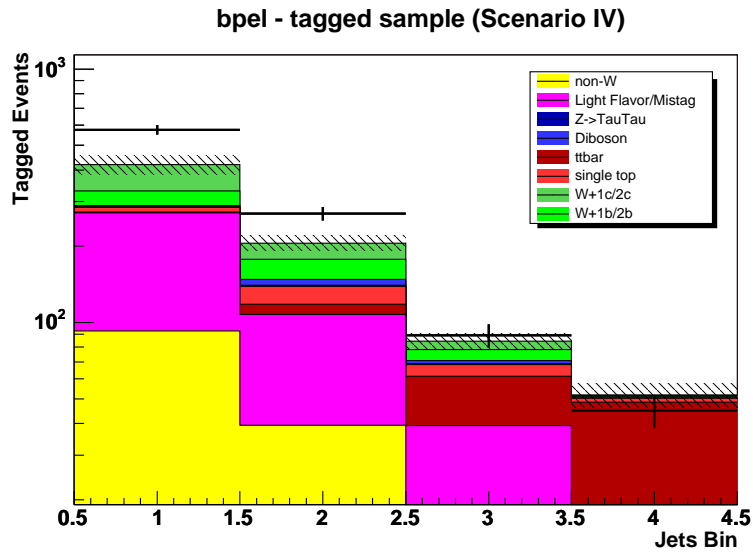


Figure 5.6: Expected background events in tagged sample compared to data (crosses) in the different jet-bins for Scenario IV:  $\cancel{E}_T > 25$  GeV, Generation 7.



## Chapter 6

# Comparison of Kinematic Quantities

It is obvious, from previous chapter, that there is the need to improve  $S/\sqrt{\text{Bkg}}$  ratio ( $S$  is the signal expectation and Bkg is the background expectation). CDF explored several techniques, essentially all based on the use of kinematic variables. Therefore it is fundamental that the background be well reproduced. As in our data set the signal is negligible, the comparison of kinematic variables in data with expected background should provide us with some feedback.

For the purpose of this work we will be comparing results for Scenario II ( $E_T > 25$  GeV, Gen6) and Scenario IV ( $E_T > 25$  GeV, Gen7). The Kolmogorov-Smirnov test (see appendix D) is used to check distributions of variables commonly used by CDF.

### 6.1 Kinematic Comparison

Distributions of electron  $E_T$ , electron  $\eta$ ,  $\cancel{E}_T$ ,  $W$  transverse mass, leading jet  $E_T$  and leading jet  $\eta$  are compared in Scenario II and Scenario IV in the four jet-bins (Figure from 6.1 to 6.12). The shape comparison probability calculated by the Kolmogorov-Smirnov test is reported on top of each histogram.

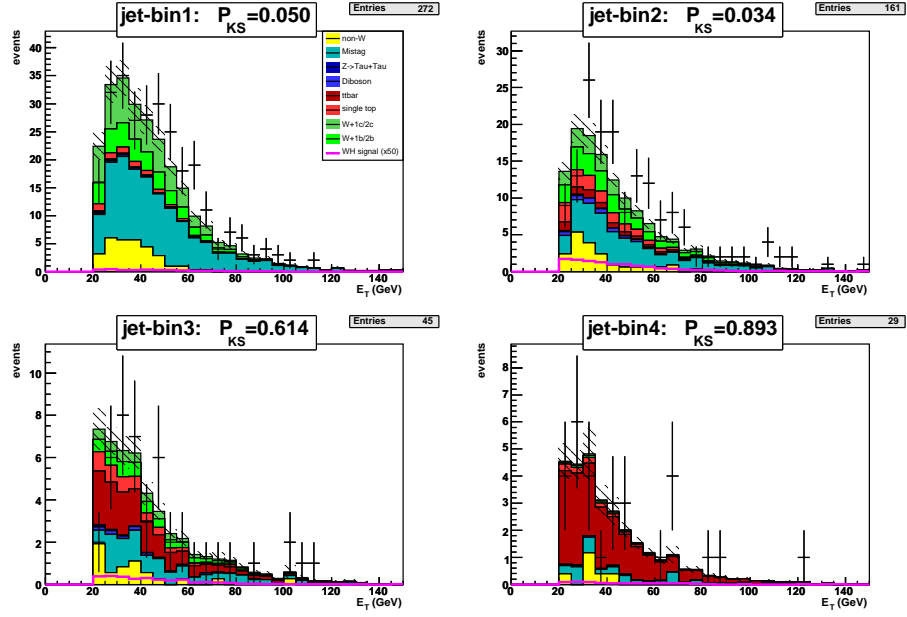


Figure 6.1:  $E_T$  distribution of the  $W$ -electron in the four jet-bins; Scenario II ( $E_T > 25$  GeV, Generation 6 reconstruction).

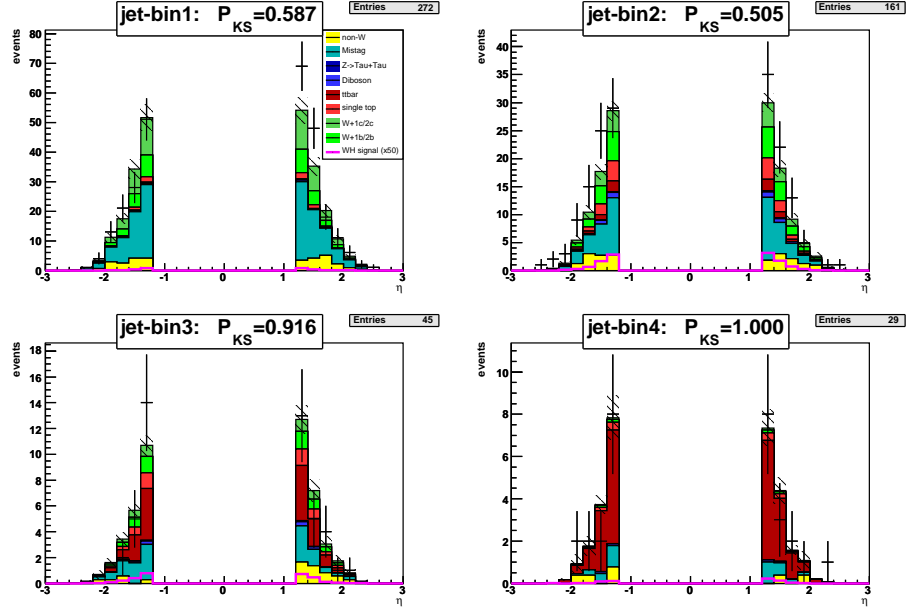


Figure 6.2:  $\eta$  distribution of the  $W$ -electron in the four jet-bins; Scenario II ( $E_T > 25$  GeV, Generation 6 reconstruction).



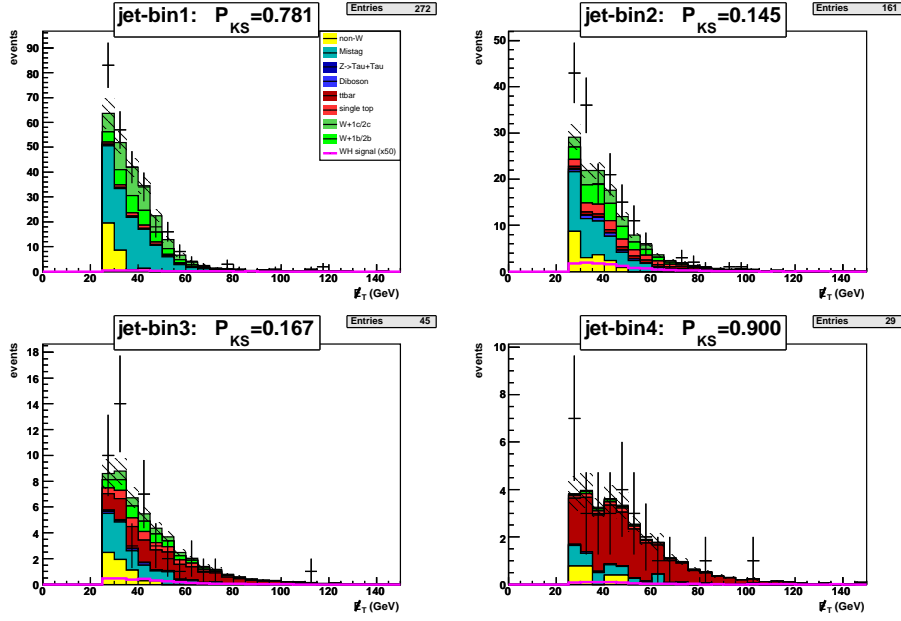


Figure 6.3:  $\cancel{E}_T$  distribution in the selected events in the four jet-bins; Scenario II ( $\cancel{E}_T > 25$  GeV, Generation 6 reconstruction).

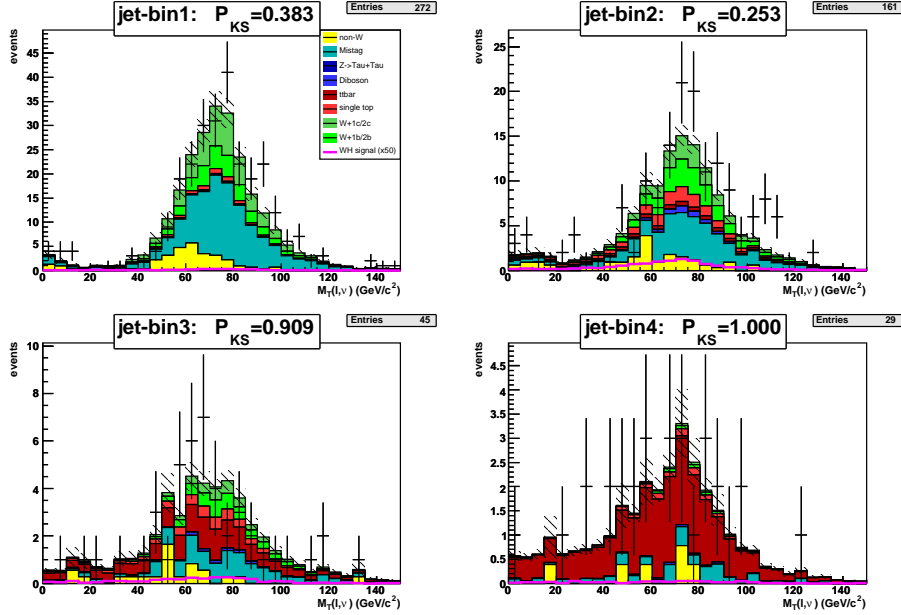


Figure 6.4:  $M_T(l, \nu)$  distribution in the selected events in the four jet-bins; Scenario II ( $\cancel{E}_T > 25$  GeV, Generation 6 reconstruction).

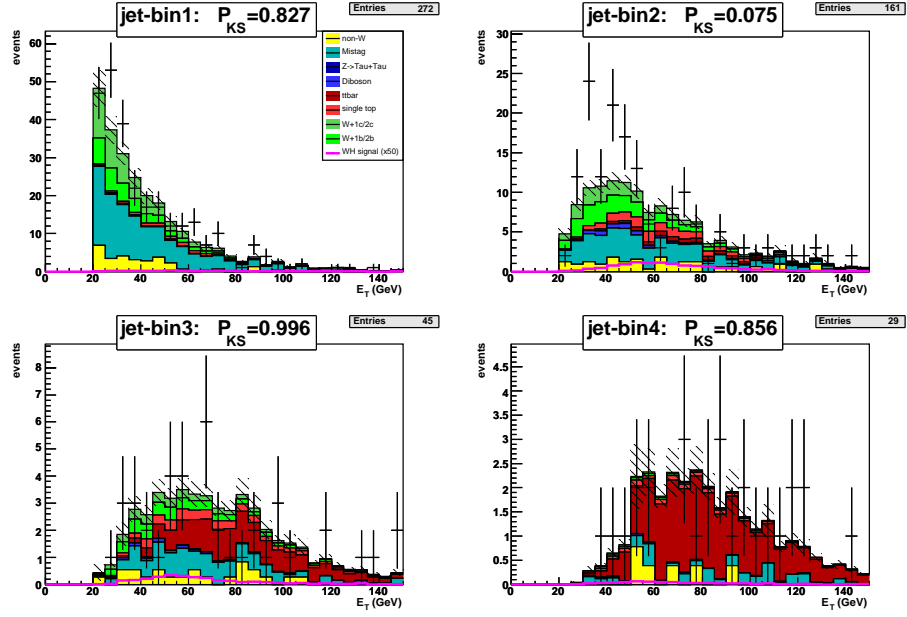


Figure 6.5:  $E_T$  distribution of the leading jet in the four jet-bins; Scenario II ( $E_T > 25$  GeV, Generation 6 reconstruction).

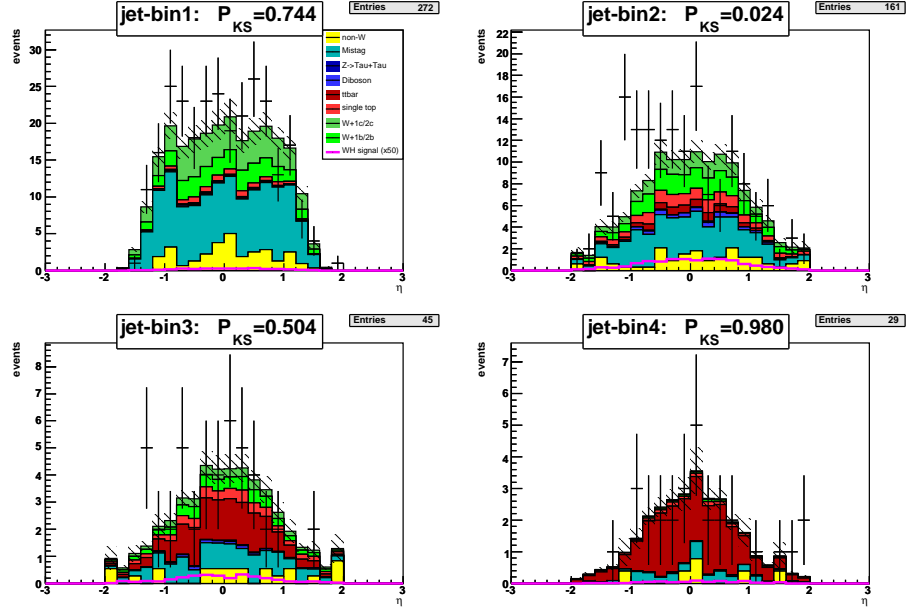


Figure 6.6:  $\eta$  distribution of the leading jet in the four jet-bins; Scenario II ( $E_T > 25$  GeV, Generation 6 reconstruction).

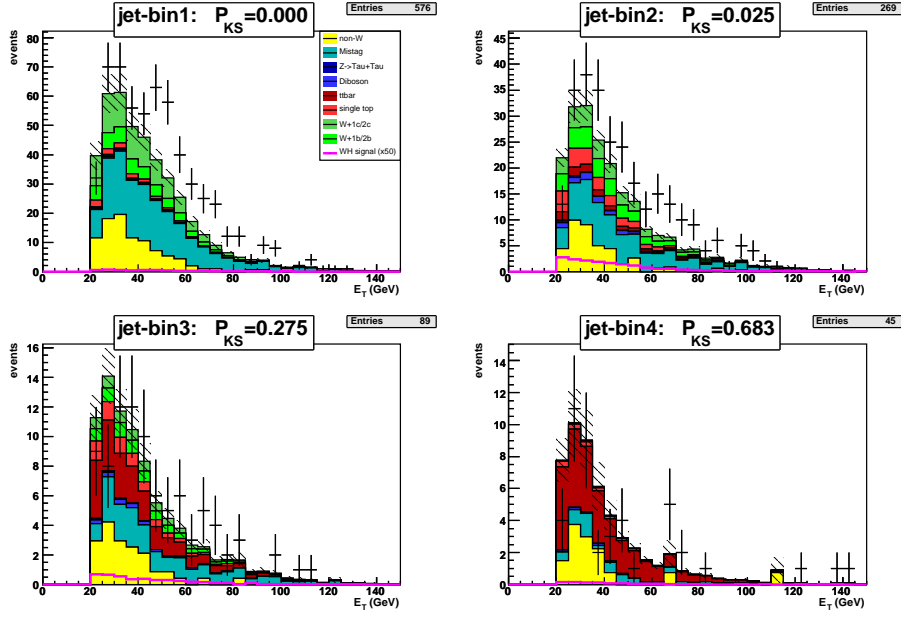


Figure 6.7:  $E_T$  distribution of the  $W$ -electron in the four jet-bins; Scenario IV ( $E_T > 25$  GeV, Generation 7 reconstruction).

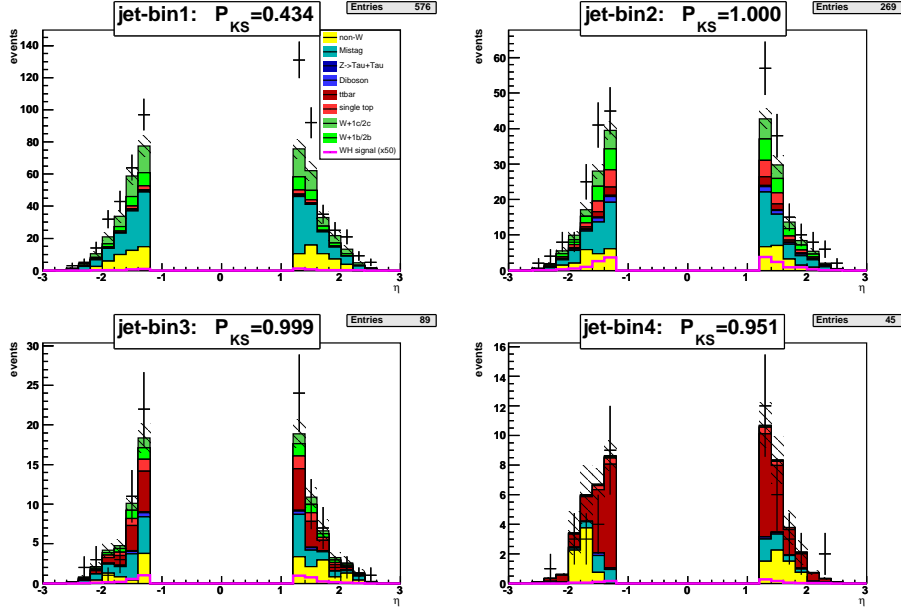


Figure 6.8:  $\eta$  distribution of the  $W$ -electron in the four jet-bins; Scenario IV ( $E_T > 25$  GeV, Generation 7 reconstruction).

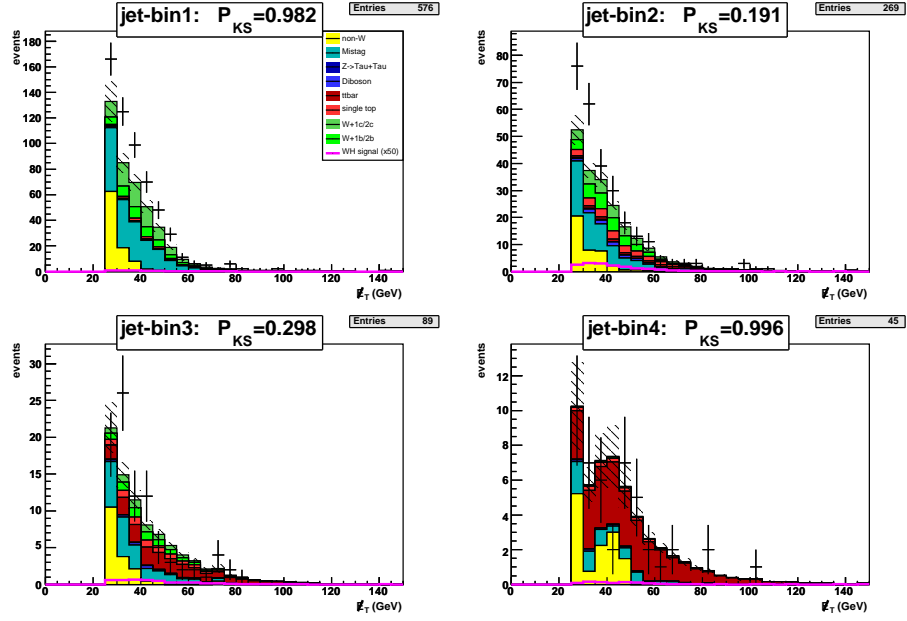


Figure 6.9:  $\cancel{E}_T$  distribution in the selected events in the four jet-bins; Scenario IV ( $\cancel{E}_T > 25$  GeV, Generation 7 reconstruction).

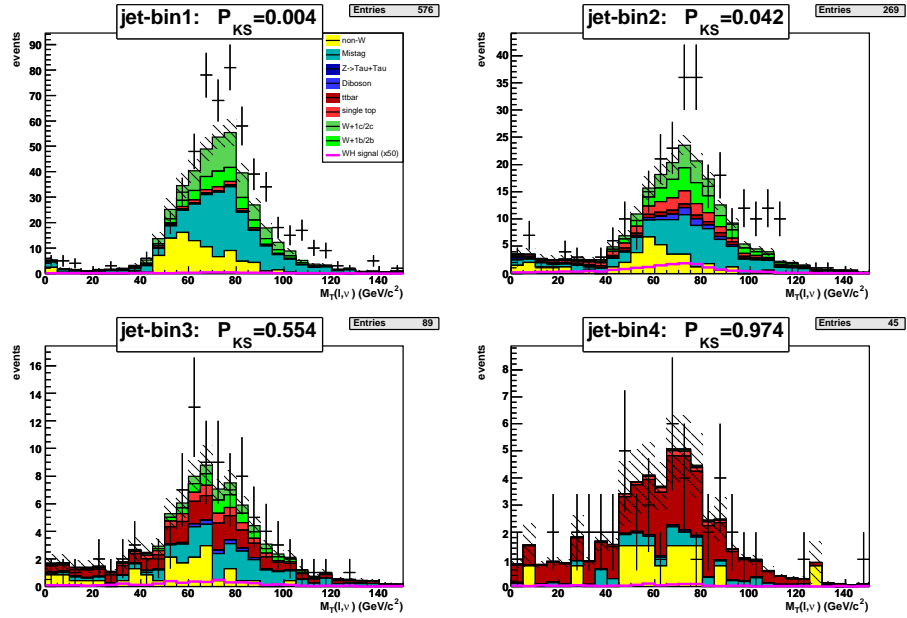


Figure 6.10:  $M_T(l, \nu)$  distribution in the selected events in the four jet-bins; Scenario IV ( $\cancel{E}_T > 25$  GeV, Generation 7 reconstruction).

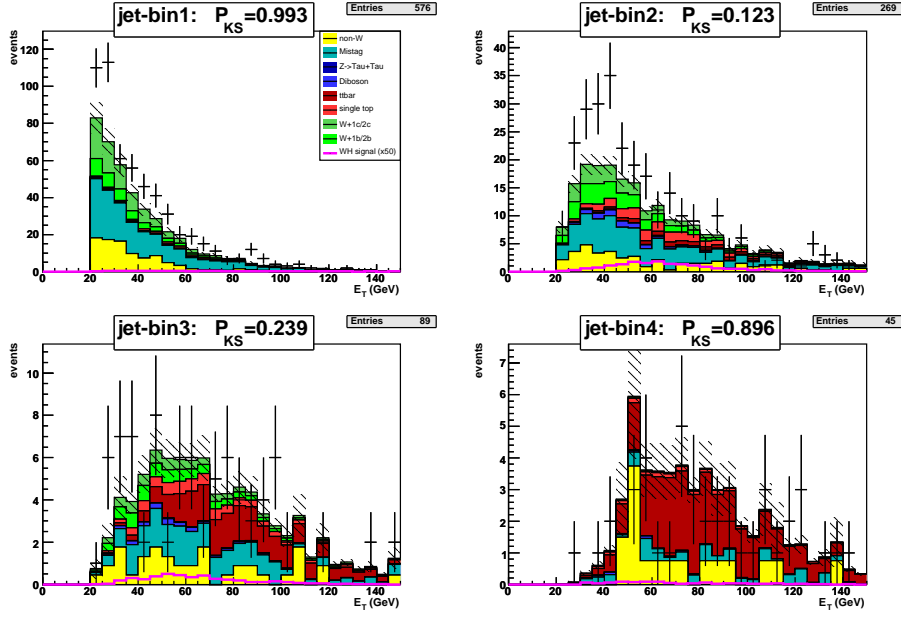


Figure 6.11:  $E_T$  distribution of the leading jet in the four jet-bins; Scenario IV ( $E_T > 25$  GeV, Generation 7 reconstruction).

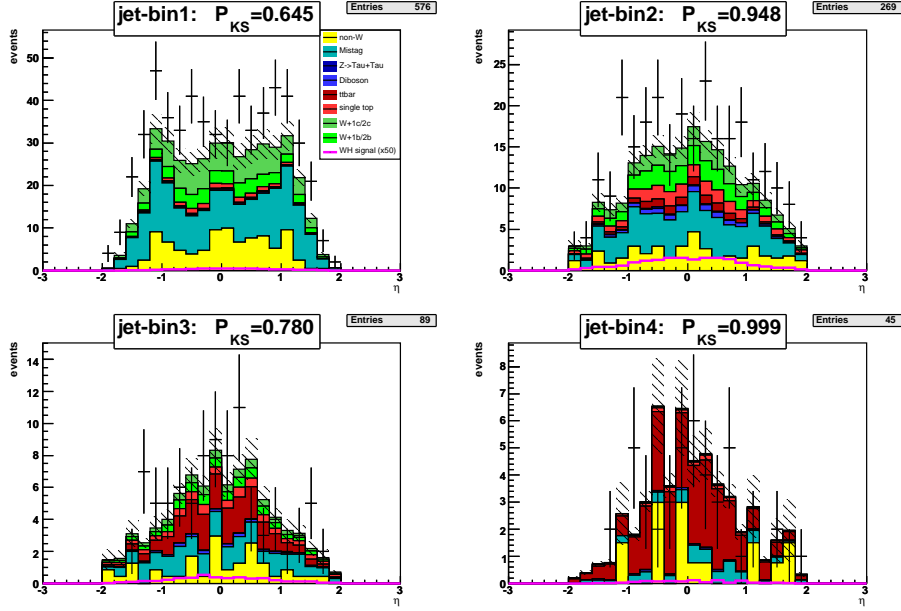


Figure 6.12:  $\eta$  distribution of the leading jet in the four jet-bins; Scenario IV ( $E_T > 25$  GeV, Generation 7 reconstruction).

For the two Scenarios the situation is quite different. In Scenario II (Gen6 reconstruction) the KS test shows a good agreement in most cases. The exception is the  $E_T$  distribution of electrons<sup>1</sup> and of the leading jet for the low (1, 2) jet-bins.

The discrepancy in the  $E_T$  distribution of electrons is probably due to the incorrect assumption that the trigger turn-on curve could be extrapolated from period 0d to the entire sample. Although the overall normalization of the background is in good agreement (see Table 5.8) an incorrect weighing of the events in the low  $E_T$  region produces a distortion in the shape. About the discrepancy in the leading jet  $E_T$ , we believe that this is due to the underestimate of uncertainties for the light flavor component (on purpose we zeroed the systematica contribution). The bin with  $N_{jet} > 2$  have a lower contribution from mistags and are not affected by this problem.

In Scenario IV (Gen7 reconstruction) both the effects present in Gen6 reconstruction can be seen again. KS tests show a poor agreement in the low (1, 2) jet-bins for the leading jet and the electron  $E_T$ . The use of the mistag parametrization tuned without the Backward tracking, makes the problem even more clear.

We also show the distributions of  $\Delta R$  between the jets (Fig. 6.13) and the di-jet invariant mass ( $M_{jj}$ , Fig. 6.14) in the 2jet-bin sample for both Scenario II and Scenario IV. They have a particular importance in background rejection and in the estimate of the  $WH$  cross section limit as in previous analysis[56] the 2-jets bin is used as signal region. Those variables appear to be less sensitive to the difference data-expected background than the ones previously discussed.

---

<sup>1</sup>The electron  $E_T$  distortion influences also the  $M_T(l, \nu)$  distribution.

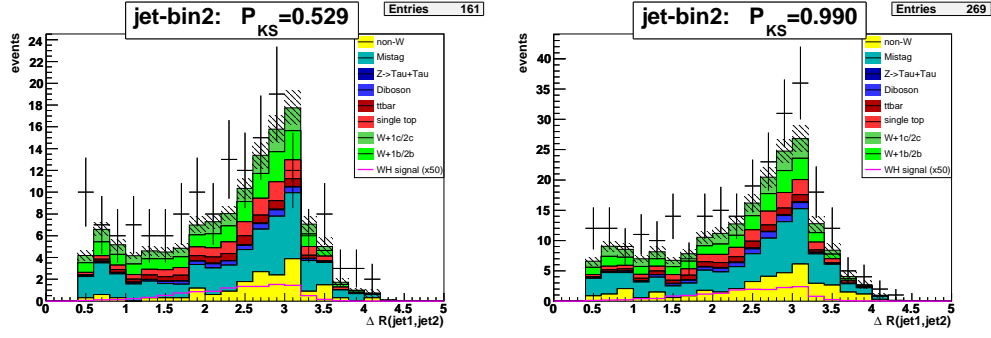


Figure 6.13:  $\Delta R$  distribution of the jets in the jet-bin2 for Scenario II (left) and Scenario IV(right).

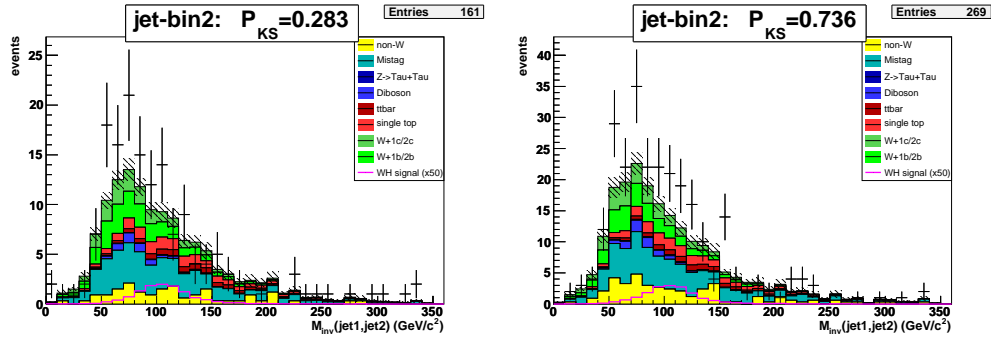


Figure 6.14: Di-jet invariant mass distribution of the jets in the jet-bin2 for Scenario II (left) and Scenario IV(right).





# Chapter 7

## Conclusions

Our findings bring us to two conclusions which are relevant to the CDF capability to search for the Higgs boson in WH events (and possibly even in other channels).

First of all it is possible to use the forward region of the CDF detector using electrons identified with a EM cluster matched to a reconstructed track. The background seems to be under control. Although a very simple requirement of  $\cancel{E}_T > 25$  GeV already strongly reduces the background, we studied different options that can still be optimized to reduce the background without depleting the signal.

As a second instance we demonstrate that the use of the Generation 7 release of CDF software produces a twofold effect: it increases the acceptance at the pretag level, thanks to the larger number of good primary leptons identified, and at the tagged level as the new tracking algorithm provides good quality tracks that can be used by the standard secondary vertex algorithms.

On the first aspect, it is worth noticing that a new version of the tracking is now available, within the framework of Generation 7. Many tracks that were reconstructed as SiSA (silicon stand-alone) are now reconstructed as Inside Out. The linking to COT improves the track quality and adds further improvement in the acceptance of the primary lepton as discussed in section 3.1.

As for the use of BackWard tracks within **SecVtx** we believe that, once a new parametrization of the mistag background will be available, they could be safely used in any analysis.

## 7.1 Shapes Comparison

Almost all the histograms of the preceding section display a good agreement between data and expected background. The main discrepancy arises from the overall background expectation in the Scenario IV (see Table 5.10, particularly in 1-jet and in 2-jets bin), this is due to the lacking information on the “mistag” background. Indeed the jet study that parametrizes the mistag rate (section 5.4) was carried on only in Generation 6 release without the inclusion of the Backward tracking. Furthermore a better study of QCD background would reduce the uncertainty on the overall normalization and on the final shape reconstruction.

## 7.2 Signal Improvements

As shown in section 4.4, going from Scenario II to Scenario IV, the signal acceptance increases by about 30% in the pretag sample:

$$\epsilon_{WH \rightarrow l\nu b\bar{b}}^{MC, Scenario II} = 0.286 \pm 0.003\%, \quad (7.1)$$

$$\epsilon_{WH \rightarrow l\nu b\bar{b}}^{MC, Scenario IV} = 0.375 \pm 0.004\%, \quad (7.2)$$

and by about 45% in the tagged sample:

$$\epsilon_{WH \rightarrow l\nu b\bar{b}}^{MC, Scenario II} = 0.1555 \pm 0.0006\%, \quad (7.3)$$

$$\epsilon_{WH \rightarrow l\nu b\bar{b}}^{MC, Scenario IV} = 0.2288 \pm 0.0009\%. \quad (7.4)$$

An idea of the improvement on the global acceptance can be obtained by looking at Figure 1.10. We gain an increase of more than 10% for  $WH$  selection, with  $W$  decaying to electron or muon and with at least one  $b$ -tag in the 2jet-bin (for a Higgs mass of  $M_H = 120 \text{ GeV}/c^2$ ).

If we look only to the plug electron selection, the simple passage from Gen6 to Gen7, with the preceding selection requirements, produces an increase in the significance ( $S/\sqrt{Bkg}$ ) of about 20%:

$$\left(\frac{S}{\sqrt{Bkg}}\right)^{Scenario II} = \frac{0.23 \pm 0.001}{\sqrt{129 \pm 8}} = 0.020 \pm 0.001, \quad (7.5)$$

$$\left(\frac{S}{\sqrt{Bkg}}\right)^{Scenario IV} = \frac{0.34 \pm 0.001}{\sqrt{207 \pm 14}} = 0.024 \pm 0.001. \quad (7.6)$$

A further reduction of the background can be obtained in several ways. One technique is to use exactly two **SecVtx** tagged events, a requirement that drastically reduces the background. Another possibility is to study

kinematic quantities trying to reduce the most serious backgrounds. The QCD background can be reduced by a requirement on a minimal  $M_T(l, \nu)$  (as can be seen in Fig. 6.4 and Fig. 6.10) and by cutting on the angle between the  $\cancel{E}_T$  and the jets, in fact mismeasured jets are one of the main sources of uncorrect missing transverse energy. We studied the latter option and it looks promising but we decided to postpone its optimization at a later stage.

### 7.3 Future Prospect

This is the first study performed using the complete set of tracking algorithms available to CDF in the forward region. It is quoted in the past that in this region there is the possibility to gain a 30% in acceptance with respect to the central region, and we believe that with Generation 7 this can be done. This analysis will be included in the  $WH$  limit calculation as soon as Gen7 will be fully tested and completely adopted by the CDF collaboration.

In any case the level of  $S/\sqrt{Bkg}$  makes very difficult to pose a limit on  $WH$  production with a simple “counting experiment” (i.e. to simply count an excess over background). CDF is now currently developing tools to kinematically select signal-enriched regions (such as neural-network and multi-variate analysis) and, all that, together with the growing integrated luminosity, makes plausible that CDF will explore the low-mass region of the Higgs with good chances to set a strong limit or even an evidence for the production of this fundamental particle.



# Appendix A

## Trigger Efficiency Studies

### A.1 MET\_PEM Trigger Path

The events used in this analysis are collected by the three-level trigger path (see section 2.3) called **MET\_PEM** trigger<sup>1</sup>. It is developed for the online selection of physics processes involving  $W^\pm \rightarrow e^\pm + \nu$  in high pseudo-rapidity regions. Its requests are:

- **Level-1:** at least a single EM trigger tower must be fired with  $E_T > 8$  GeV, the ratio of hadronic (*Had*) deposit with respect to electromagnetic one (*EM*) must be less than 0.125, i.e.  $Had/EM < 0.125$ , global imbalance in transverse momentum must be greater than 15 GeV, i.e.  $\cancel{E}_T > 15$  GeV. This L1 trigger is named “L1\_EM8\_MET15”.
- **Level-2:** L1\_EM8\_MET15 confirmation and a Level-2 EM cluster with  $E_T > 20$  GeV and  $Had/EM < 0.125$  in the plug region. This L2 trigger is named L2\_PEM20\_L1\_EM8\_MET15.
- **Level-3:** L1 and L2 trigger confirmation, an EM object with  $E_T > 20$  GeV and  $\cancel{E}_T > 15$  GeV (as L1). The EM object is build using PES information and a  $\chi^2$  comparison with test beam data. This L3 trigger is named L3\_PEM\_20\_MET\_15.

All events recorded by this trigger are stored in the dataset **bpe1**, it collects events with high energy electromagnetic objects going into the plug region.

The efficiency of **MET\_PEM** trigger path was evaluated in[55] for the run period 0d and we extrapolate the results also in periods 0h and 0i.

The efficiency measure was performed with the help of the back up trigger **PLUG\_ELECTRON20**. Its requests are very similar to **MET\_PEM** trigger path:

---

<sup>1</sup>Other backup triggers are used to calculate efficiency and for background studies

- **Level-1:** L1\_EM8 (at least an EM deposit with  $E_T > 8$  GeV and  $Had/EM < 0.125$ ).
- **Level-2:** L2\_PLUG20 (Level-2 EM cluster with  $E_T > 20$  GeV and  $Had/EM < 0.125$  in the plug region).
- **Level-3:** L3\_PLUG20 (Level-3 plug EM object).

The only differences with MET\_PLUG are the missing energy request and a prescale factor equal to  $\Gamma$  in the Level-2 accept, i.e. only one event every  $\Gamma$  events that pass L2 criteria is accepted (otherwise the rate of this trigger would be excessive).<sup>2</sup>

It is possible to calculate efficiency,  $\epsilon_{\text{MET}}$ , of the combined requests L1\_MET15 & L3\_MET15 trough:

$$\epsilon_{\text{MET}} = \frac{\text{PLUG\_ELECTRON20 \& MET\_PEM}}{\text{PLUG\_ELECTRON20}}, \quad (\text{A.1})$$

where the name of the trigger path indicates the number of events passing such trigger.

The efficiency of PLUG\_ELECTRON20 trigger path and in particular of the L2\_PLUG20 & L3\_PLUG20 requests is measured using a  $Z^0 \rightarrow e^+e^-$  Central Plug (CP) sample, with one of the two electrons (“legs”) going into the central region and collected with an independent central trigger (see section 4.3 for further details on the  $Z^0 \rightarrow e^+e^-$  CP sample). In turn the central trigger and the L1\_EM8 request were studied with test beam data. The final MET\_PEM efficiency is:

$$\epsilon_{\text{MET\_PEM}} = \epsilon_{\text{L2\_PLUG20}} \times \epsilon_{\text{MET}} \times \epsilon_{\text{L3\_PLUG20}}, \quad (\text{A.2})$$

The  $\eta$ ,  $E_T$  and run dependency has been measured. Figure A.1 display the turn-on curve for the L1\_L3\_MET15, L2\_PLUG20 and L3\_PLUG20 triggers. We observe that the slope of the L2\_PLUG20 trigger is not so steep and the plateau appears after 25 GeV.

The value averaged in all variables for  $W \rightarrow e + \nu$  events is:

$$\epsilon_{\text{MET\_PEM}} = 0.9610 \pm 0.0006(\text{stat})_{-0.0041}^{+0.0034}(\text{syst}). \quad (\text{A.3})$$

This analysis use the average efficiency of Eq A.3 to scale the expectation value of MC samples, while turn-on curves are used to weigh the MC contribution inside the shape composition of the kinematic variables.

---

<sup>2</sup> $\Gamma$  is equal to 10 before run number 183877 and equal to 25 successively.

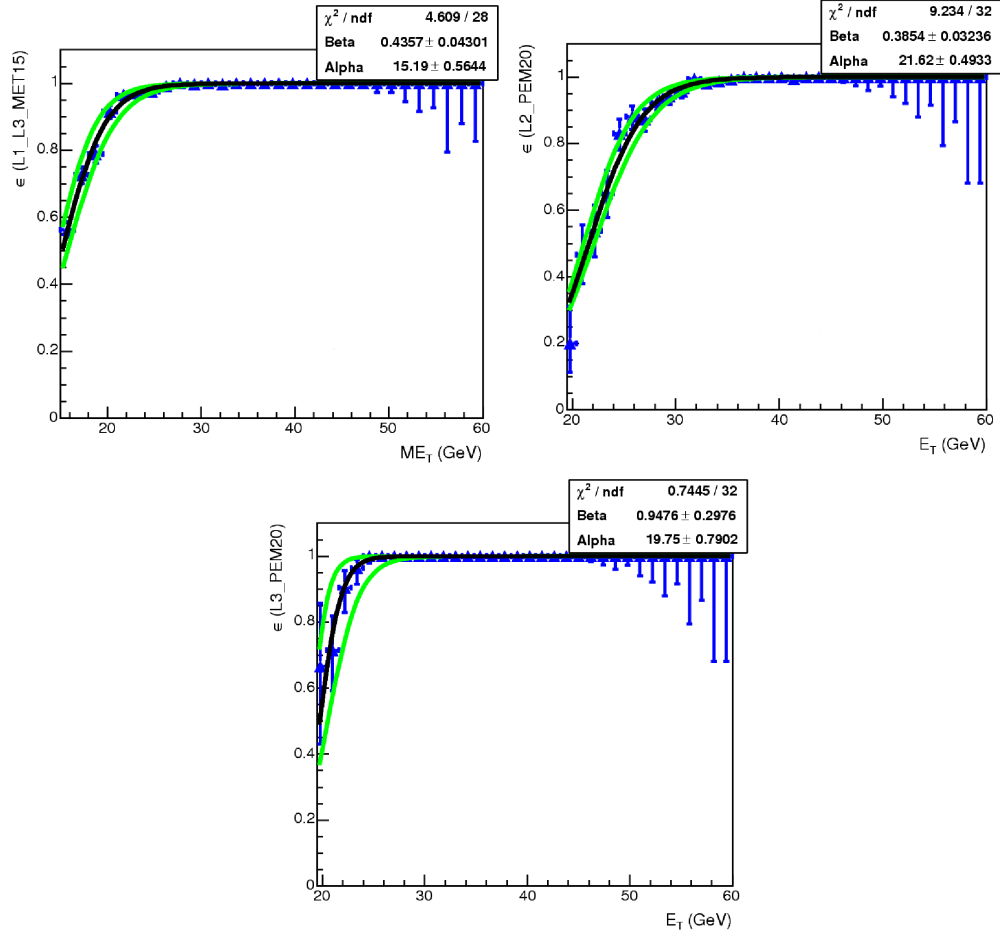


Figure A.1:  $E_T$  dependency of efficiency turn-on curves for L1\_L3\_MET15 (top), L2\_PLUG20 and L3\_PLUG20 triggers. The green lines show the effect of varying by  $\pm 1\sigma$  the fitting parameters.





# Appendix B

## Track Matching Optimization

A track matched to a given EM cluster is a clear signature of an electron candidate. The tracks used in this work to identify plug electrons can be reconstructed by any of the CDF tracking algorithms (see section 3.1.1) covering the forward region, this implies an heavy use of SiSA, Inside-Out and, in the Gen7 datasets, of the new Backward tracking algorithms. The optimization of track selection cuts has been done with a  $Z^0 \rightarrow e^+e^-$  Central-Plug sample (see section 4.3).

In practice each reconstructed 3D track is considered “usable” if the transverse momentum satisfy  $p_T > 1.0$  GeV/ $c$ . Successively all the tracks are ordered in decreasing  $p_T$  and extrapolated up to the PES plane. Each track intersection point  $(x^{trk}, y^{trk})$  is checked against the PES reconstructed cluster coordinate  $(x^{PES}, y^{PES})$ , both  $x$  and  $y$  coordinates must be within a maximum distance of 3 cm:

- $\Delta x = |x^{PES} - x^{trk}| < 3$  cm,
- $\Delta y = |y^{PES} - y^{trk}| < 3$  cm.

The range of the matching window is estimated from  $\Delta x$  and  $\Delta y$  distribution of tracks selected in the  $Z^0 \rightarrow e^+e^-$  CP sample, see Figure B.1[55]. A matched track is rejected if its vertex falls outside the beam fiduciality region ( $|z_0^{trk}| < 60$  cm). As last cleaning cut a rejection threshold is applied on the ratio between the total energy of the EM cluster and the momentum of the matched track ( $E^{cluster}/p^{trk}$ ). This cut reduce background due to unidentified charged particles collinear to photons, or tracks with wrong reconstruction of the momentum. Figure B.2 displays the  $E/p$  distribution of reconstructed tracks for different algorithms. A threshold of  $E^{cluster}/p^{trk} < 2$  is applied.

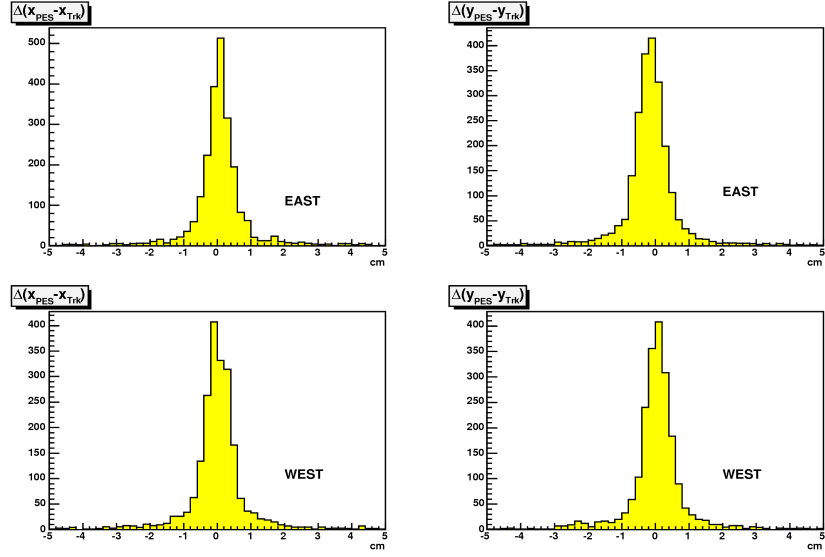


Figure B.1: Residuals along the  $x$  and  $y$  directions (plane orthogonal to the beam line) for tracks extrapolated to the PES plane in the  $Z^0 \rightarrow e^+e^-$  sample. Top plots show East side plug, bottom plots show West side plug.

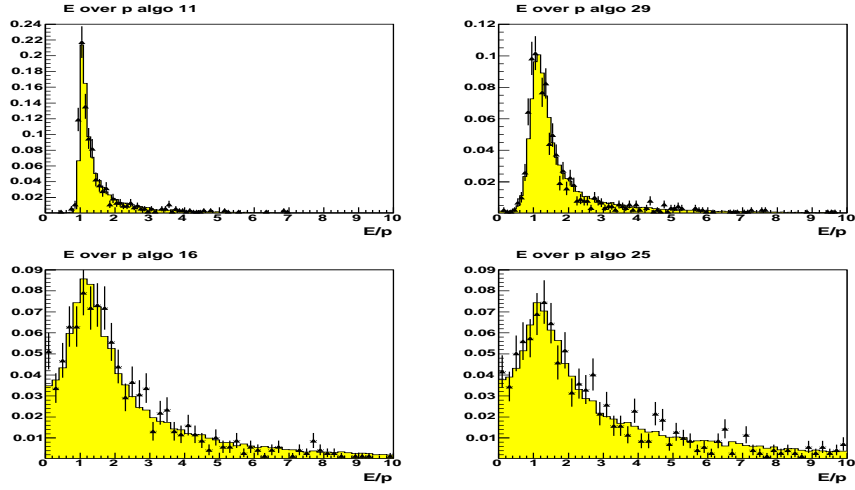


Figure B.2:  $E/p$  distribution of tracks reconstructed in the forward region with different tracking algorithms. Inside-Out (top left), Backward tracking (top right) and two kinds of Silicon-Stand Alone tracking, central SiSA (bottom left), forward SiSA (bottom right).

# Appendix C

## Tight Electrons and Muons

CDF collaboration often uses standard selection cuts optimized for high  $p_T$  physics analysis. Definition of *tight central electron* and of *tight muon* are used in this work.

*Tight electron* requirements[53] are based on the CEM and COT coverage. Variables used for central electron analysis are mostly the same to those used in the plug region but with different name (as example PEM and PES are replaced by CEM and CES), only two new variables need explanation:

- the  $\chi^2_{strip}$  is the  $\chi^2$  of the fit between the energy deposited on the  $z$  strips of the CES detector and the shape obtained with test beam data;
- the Lateral Shower Sharing variable ( $L_{shr}$ ) compares the *revealed* sharing of energy deposition between the towers in the CEM to that *expected* in true electromagnetic showers taken with test beam data.

$L_{shr}$  is defined as:

$$L_{shr} = 0.14 \sum_i \frac{E_i^{adj} - E_i^{expected}}{\sqrt{(0.14\sqrt{E_i^{adj}})^2 + (\Delta E_i^{expected})^2}}, \quad (C.1)$$

where the sum is over the towers adjacent (*adj*) to the seed tower of the cluster,  $0.14\sqrt{E_i^{adj}}$  is the error on the energy measure and  $\Delta E_i^{expected}$  is the error on the energy estimate. The cuts for *tight central electron* are summarized in Table C.1.

Also the *tight muon* cuts are optimized and standardized at CDF[47]. All kinds of *tight muons* need a reconstructed track and a stub in the muon chambers, requests vary slightly depending on the considered subdetector. *Tight muon* cuts are summarized in Table C.2

EM cluster in CEM fiducial region.
$E_T^{cluster} > 20 \text{ GeV}$
$Had/EM < 0.055 + 0.0045 E^{cluster}$
$IsoRel < 0.1$
$L_{shr} < 0.2$
$\chi_{strip}^2 < 10$
COT track with >3 axial >3 stereo Super Layers segments
Track matching with track coordinates extrapolate to CES plane: $\Delta z =  z^{CES} - z^{trk}  < 3 \text{ cm},$ $-3.0 < q\Delta x < 1.5 \text{ cm},$ with $\Delta x =  x^{CES} - x^{trk} $ and charge $q$
Track $z_0 < 60 \text{ cm}$
Track $p_T^{trk} > 10 \text{ GeV}/c$ , accept if $E_T^{cluster} > 100 \text{ GeV}$
$E/p^{trk} < 2.0$ , accept if $E_T^{cluster} > 100 \text{ GeV}$

Table C.1: Summary of *tight central electron* requirements.

COT track with >3 axial and >2 stereo Super Layers segments
Track $p_T > 20 \text{ GeV}/c$
$E_T^{EM} < 2 + \max(0, ((p^{trk} - 100)0.0115)) \text{ GeV}$
$E_T^{Had} < 6 + \max(0, ((p^{trk} - 100)0.028)) \text{ GeV}$
$IsoRel < 0.1$ , with $IsoRel \equiv E_T^{isol}/p_T$
Track $z_0 < 60 \text{ cm}$
Track $ d_0  < 0.2 \text{ cm}$ (track w/o silicon)
Track $ d_0  < 0.02 \text{ cm}$ (track with silicon)
Track extrapolated and matched to a muon stub: stub in fiduciality region CMUP: track matched to both CMU and CMP CMX: track $r^{trk} > 140 \text{ cm}$

Table C.2: Summary of *tight muon* requirements.

# Appendix D

## Kolmogorov Smirnov Test

The Kolmogorov-Smirnov test (KS test) is a goodness of fit test used to decide if a sample of  $N$  measurements  $x_1, x_2, \dots, x_N$ , of a variable  $X$  comes from a population with a specific known distribution  $F(x) = \int_{-\infty}^x f(y)dy$ , named *null hypothesis*. The KS test is applied to the unbinned distribution of the  $x_i$  measurements and can be used to compare the shape of data against the continuous distribution of the MC sample<sup>1</sup>. KS test is based on the *empirical distribution function*:

$$F_N(x) = \frac{1}{N} \sum_{i=1}^N I_{x_i < x}, \quad (\text{D.1})$$

where  $I = 1$  for each  $x_i < x$  and  $I = 0$  otherwise. The KS distance  $D$  for the given function  $F(x)$  (see Fig. D.1) is defined as:

$$D = \text{Max}(F_N(x_i) - F(x_i)) \quad i = 1, 2, \dots, N. \quad (\text{D.2})$$

The value of  $D$  is a random variable but for large  $N$  ( $N = 20 \sim 30$  is usually enough) and under the *null hypothesis* the limit distribution is:

$$K(\lambda) = 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 \lambda^2}, \quad (\text{D.3})$$

where  $\lambda = \sqrt{N}D$ . Given  $\lambda$  the  $K$  function returns the probability that two distributions are compatible.

---

<sup>1</sup>Usually MC is created with a very large statistics and has a continuous distribution.

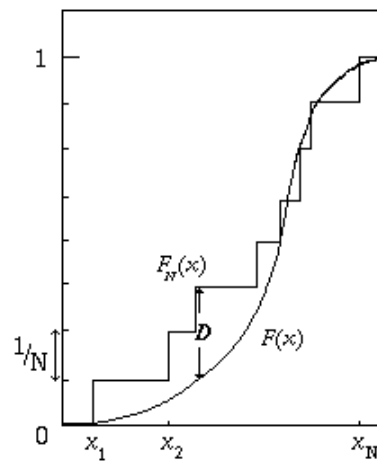


Figure D.1: The *empirical distribution function* compared with the known distribution  $F(x)$  (the *null hypothesis*).

# Bibliography

- [1] Particle Data Group, *Review of Particle Physics*, Jour. Phys. G **33**, 1 (2006)
- [2] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22**, 579 (1961)
- [3] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19**, 1264 (1967)
- [4] S. Weinberg, *The Quantum Theory of Fields, Vol I Foundation*. Cambridge Edition 2005
- [5] P. W. Higgs, *Broken Symmetries, Massless Particles and Gauge Fields*, Phys. Lett. **12**, 132 (1964)
- [6] J. Conway web page:  
<http://www.physics.ucdavis.edu/~conway/research/higgs/smhiggs-tev.html>.
- [7] ALEPH and DELPHI and L3 and OPAL Collaboration and the LEP Working Group for Higgs Boson Searches, Phys. Rev. Lett. **B 565**, 61 (2003).
- [8] J. Erler, *Global Fits to Electroweak Data Using GAPP*, arXiv:hep-ph/0005084, May (2000).
- [9] ALEPH Collaboration, J. Alcaraz et al., *A combination of preliminary electroweak measurements and constraints on the Standard Model*, arXiv:hep-ex/0612034, December (2006).
- [10] CDF II Collaboration, Top Quark Group Physics Results:  
<http://www-cdf.fnal.gov/physics/new/top/top.html>.

- [11] CDF and D0 Working Group Members, Collaboration, R Barate et. al., *Results of Tevatron Higgs sensitivity studies*. FERMILAB-PUB-03-320-E.
- [12] The DØ Collaboration, Higgs Physics Results:  
<http://www-d0.fnal.gov/Run2Physics/WWW/results/higgs.htm>.
- [13] J. Dittmann et al., *Search for Standard Model Higgs Boson Production in Association with a  $W^\pm$  Boson using  $\int \mathcal{L} dt = 1.9 \text{ fb}^{-1}$* , CDF note 9136, January 24, 2008.
- [14] CDF II Collaboration, Higgs Discovery Group:  
<http://www-cdf.fnal.gov/physics/new/hdg/hdg.html>.
- [15] Fermilab Accelerator Division: *Documentation Web Page*:  
<http://beamdocs.fnal.gov/AD-public/DocDB/ListTopics>.
- [16] Fermilab Accelerator Division: *Details of Acceleration Chain*:  
<http://beamdocs.fnal.gov/public/Chain.html>.
- [17] S. van der Meer, *Stochastic cooling and the accumulation of antiprotons*, Rev. Mod. Phys. **57**, 699 (1985).
- [18] CDF Summary of Stores:  
<http://www-cdfonline.fnal.gov/ops/opshelp/stores/>.
- [19] R. Moore, *Performance and Future of the Tevatron*, Presentation given at Hadron Collider Physics (HCP 2007), La Biodola, Elba, Italy, May 21-25, 2007.
- [20] The CDF Collaboration, *The CDF II Detector Technical Design Report*, Fermilab Publication FERMILAB-PUB-96/390-E, October, 1996.
- [21] The CDF Collaboration, *Summary of the CDF RunII Detector Parameters*  
<http://www-cdf.fnal.gov/internal/detectors/parameters.html>.
- [22] C. S. Hill, *Initial experience with the CDF layer 00 silicon detector*, Nucl. Instrum. Methods **A511**, 118 (2003).
- [23] A. Sill et al., *CDF Run II silicon tracking projects*, Nucl. Instrum. Methods **A446**, 1 (2000).
- [24] A. Affolder et al., *Status report of the intermediate silicon layers detector at CDF II*, Nucl. Instrum. Methods **A485**, 6 (2002).



- 
- [25] C. S. Hill et al., *Operational experience and performance of the CDFII silicon detector*, Nucl. Instrum. Method A **530**, 1 (2004).
- [26] A. Affolder et al., *CDF Central Outer Tracker*, Nucl. Instrum. Methods **A447**, 1 (2000).
- [27] D. Acosta et al., *The CDF central electromagnetic calorimeter*, Nucl. Instrum. Methods **A267**, 272 (1988).
- [28] S. Bertolucci et al., *The CDF central and endwall hadron calorimeter*, Nucl. Instrum. Methods **A267**, 301 (1988).
- [29] M. Albrow et al., *The CDF plug upgrade calorimeter: test beam result*, Nucl. Instrum. Methods **A480**, 524 (2002).
- [30] C. M. Ginsburg, *CDF Run 2 Muon System*, Eur. Phys. J., **33**, S100233 (2004).
- [31] D. Acosta et al., *The CDF Cherenkov luminosity monitor*, Nucl. Instr. Methods **A 494**, (2001).
- [32] D. Acosta et al., *The performance of the CDF luminosity monitor*, Nucl. Instr. Methods **494**, 57 (2002).
- [33] C. Avilla et al., (E811 Collaboration), *A measurement of the proton-antiproton total cross-section at  $\sqrt{s} = 1.8$  TeV*, Phys. Lett.**B445**, 419 (1999).  
F. Abe et al. (CDF Collaboration), *Measurement of the anti-proton proton total cross-section at  $\sqrt{s} = 546$  GeV and 1800 GeV*, Phys. Rev. D **50**, 5550 (1994).
- [34] G. Gomez-Ceballos et al., *Event Builder and Level 3 at the CDF experiment*, Nucl. Instr. Methods **A518**, 522 (2004).
- [35] ROOT: *An Object Oriented Analysis Framework*:  
<http://root.cern.ch>.
- [36] M.L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP 0307:001, (2003).
- [37] T. Sjostrand et al. *High-Energy-Physics Event Generation with PYTHIA 6.1*, Comp. Phys. Commun. **135**, 238 (2001).  
Pythia home page: <http://www.thep.lu.se/~torbjorn/Pythia.html>.

- [38] G. Corcella et al., *HERWIG 6.5 release note*, arXiv:hep-ph/0210213, Oct 15, 2002.
- [39] R. Brun and F. Carminati, *CERN Program :Library Long Writeup*, W5013, 1993 (unpublished).
- [40] S. Harper et al., *Reconstructing the Plug Electron Energy in 5.3.3*, CDF note 7687, July 17, 2005.
- [41] CDF Collaboration, *Measurement of the inclusive W and Z cross sections in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV*, Jour. of Phys. G **34**, 2457 (2007)
- [42] J. Marriner, *Secondary vertex fit with mass and pointing constraints*, CDF Internal Note 1996.
- [43] F. D. Snider et al., *Tracking at CDF: Algorithms and Experience from Run I and Run II*, Nucl. Instr. Meth. A **566**, 133 (2006).
- [44] P. Azzi et al., *Histogram tracking in the COT*, CDF note 5562, February 21 (2001).
- [45] C. Hays et al., *Inside-Out tracking at CDF*, Nucl. Instr. Meth. A **538**, 249 (2005).
- [46] Byeong Rok Ko et al., *SVXII Stand-alone Tracking*, CDF note 6440, May 15 (2003).
- [47] U. Grunder et al., *High- $P_T$  muons recommended cuts and efficiency for Summer 2006*, May 31, 2006.
- [48] C. Gerald et al., *Run II Jet Physics*, CDF note 5293, April 18, 2000.
- [49] F. Abe et al., *Topology of three-jet events in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV*, Phys. Rev. **D45**, (1992).
- [50] A. Bhatti et al., *Determination of the jet energy scale at the Collider Detector at Fermilab*, Nucl. Instr. Meth. A **566**, 375 (2006).
- [51] J. Adelman et al., *Generic Jet Scale Correction for Run II*, CDF note 7358, December 12, 2005.
- [52] V. Cavaliere et al., *Measurement of flavor fractions of jets produced in association with a Z boson with Roma NN tagger*, CDF note 8821, May 21, 2007. Also University of Rome Thesis, unpublished.

- [53] D. Hare et al., *Electron ID Efficiency and Scale Factors for Winter 2007 Analysis*, CDF note 8641, December 7, 2006.
- [54] A. Taffard, *Run II Cosmic Ray tagger*, CDF note 6100, February 11, 2003.
- [55] G. Chiarelli et al., *Selection criteria and efficiency studies for the Measurement of the  $\sigma(p\bar{p} \rightarrow W) \times BF(W \rightarrow e + \nu)$  in the Plug Region*, CDF note 7594, June 14, 2006.
- [56] Y. Kusakabe et al., *Search for Standard Model Higgs Boson Production in Association with a W Boson at CDF with  $\int \mathcal{L} dt = 1.9 \text{ pb}^{-1}$* , CDF note 8355, July 3, 2006.
- [57] The CDF Collaboration, *Measurement of the  $t\bar{t}$  production cross section in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96 \text{ TeV}$  using lepton + jets events with secondary vertex b-tagging*, Phys. Rev. D **71**, 052003 (2005).
- [58] Franklin et al., *Method 2 Background for 1.12/fb Lepton+Jets Analysis*, CDF note 8766, April 25, 2007.
- [59] H. Bachacou et al., *Non-W background in leptons+jet SECVTX tagged sample*, CDF note 6569, August 2, 2004.
- [60] The CDF Collaboration, *First Measurement of the Ratio of Central-Electron to Forward-Electron W Partial Cross Section in  $p\bar{p}$  Collision at  $\sqrt{s} = 1.96 \text{ TeV}$* , Phys. Rev. Lett. D **98**, 251801 (2007).
- [61] S. Grinstein et al., *SecVtx Scale Factors and Mistag Matrices for Winter 2007*, CDF note 8910, July, 2007.
- [62] CDF Collaboration, *Physics with Taus at CDF*, Nucl. Phys. Proc. Suppl. **144**, 2005.
- [63] Franklin et al., *Heavy-Flavor Content of the W+Jets Sample*, CDF note 8765, April 25, 2007.
- [64] Franklin et al., *Calibration of Heavy-Flavor Production in QCD Data*, CDF note 8768, April, 2007.