Scuola Normale Superiore di Pisa

CLASSE DI SCIENZE

Corso di Perfezionamento in Fisica

Tesi di Dottorato

Search for WZ/ZZ Production in the Lepton(s) + MET + Jets Channel with the CDF Experiment at the Tevatron Collider

Candidato:
Marco Trovato

Relatore:
Prof. Giorgio Bellettini, Prof. Gigi Rolandi

Correlatore:
Dr. Vadim Rusu

Anno Accademico 2009-2013

Dedicata alla mia mamma

Contents

Acknowledgements **vii**

Abstract **ix**

1 Theoretical Overview and Diboson Landscape **1**

1.1 The Standard Model 1

1.1.1 Fundamental Interactions 2

1.1.2 EW Symmetry Breaking: Higgs Model 7

1.1.3 Lagrangian of the Standard Model 9

1.2 W and Z Decay Modes and Branching Ratios 10

1.3 Diboson Physics 10

1.3.1 Cross-Section Measurements: State of the Art 12

2 Accelerator Complex **19**

2.1 Proton production and first step in the acceleration 19

2.2 Anti-proton production and accumulation 20

2.3 Injections and collisions 21

2.4 Performance 23

3 The Collider Detector at the Fermilab Tevatron **26**

3.1 Coordinates 26

3.2 Detector Overview 28

3.3 Tracking System 30

3.3.1 Inner Tracker 31

3.3.2 Central Outer Tracker 35

3.4 Calorimeters 38
3.5 Muon System 40
3.6 Cherenkov Luminosity Counters 43
3.7 Trigger and Data Acquisition 45
3.7.1 Level 1 45
3.7.2 Level 2 47
3.7.3 Level 3 49

4 Object Identification **50**
4.1 Tracks 50
4.1.1 Tracking Algorithms 51
4.2 Electrons 55
4.3 Muons 57
4.4 Outgoing partons: Jets 59
4.4.1 Jet Reconstruction 62
4.5 Neutrinos: missing transverse energy 64

5 Identification of bottom jets **66**
5.1 b-jets versus other jets 66
5.2 An effective b-jet identification tagger: Jet b-ness 67
5.2.1 Track-by-track ANN - track b-ness 67
5.2.2 Jet-by-jet ANN - jet b-ness 70
5.2.3 Calibration of the b-ness response 72

6 Jet Energy Corrections **78**
6.1 Calorimeter Simulation 78
6.2 Jet Energy Corrections 80
6.2.1 η-dependent correction (L1) 81
6.2.2 Multiple Interaction Corrections (L4) 83
6.2.3 Absolute Jet Energy Scale (L5) 86
6.2.4 Quark/Gluon Jet Energy Calibrations 87

7 Signal and Background Simulation **103**

7.1 Overview on Event generators 103
7.1.1 Parton Distribution Function 104
7.1.2 Matrix Element Calculations 104
7.1.3 Modeling of the initial and final state radiation 106
7.1.4 Fragmentation/Hadronization 108
7.1.5 Beam Remnants and Multiple Interactions 109
7.1.6 Detector Simulation 111
7.2 Signal and Background Definition 114

8 Event Selection and Modeling of the Data Sample 119

8.1 Data Sample 119
8.1.1 Central Electron Trigger 120
8.1.2 Central Muon Triggers 122
8.2 Offline Event Selection 123
8.2.1 Selection of the Leptonically Decaying W 123
8.2.2 Di-jet system selection 124
8.2.3 Additional Cuts 124
8.3 Modeling of the Data Sample 125
8.3.1 Top and Diboson Processes 126
8.3.2 V+jets 126
8.3.3 QCD multi-jets 127
8.4 Validation of the Model 138
8.5 Di-jet mass fit 138
8.5.1 Fit procedure 142
8.5.2 Systematic Uncertainties 143
8.6 Fitter Validation 147
8.7 Results 148
8.7.1 Upper Limit on the Narrow Resonance Cross-Section 150
8.8 Conclusions 151

9 WZ/ZZ Search: Optimization and Results 153

9.1 Extending the Online Acceptance of W+2jets Events 153

9.1.1 Forward Electron Triggers 153
9.1.2 Missing Transverse Energy Plus Jets Triggers 155
9.2 Optimization of the Selection Cuts 159
9.3 Splitting of the Pre-tag Sample: Heavy Versus Light Flavor Jets . . 161
9.3.1 Definition of the Tag region 162
9.3.2 The Importance of the NoTag Region 163
9.4 Fitting Procedure 164
9.4.1 Systematic Uncertainties 165
9.5 Results 166
9.5.1 Cross-section Measurement 168
9.6 Statistical Significance of the WZ/ZZ Measurement 168

10 Conclusions **172**

Appendices **1**

A Other b-jet identification taggers **3**
A.0.1 Secondary Vertex Tagger ("SecVtx") 3
A.0.2 Soft Lepton Tagger ("SLT") 5
A.0.3 Jet Probability Tagger ("JetProb") 5
A.0.4 "RomaNN" b-Tagger 6
A.0.5 Higgs Optimized b-Jet Identification Tagger (HOBIT) 7

B More Details on the MC Event Simulation **11**
B.1 Parton Shower Method 11
B.2 String fragmentation 12

C DifferencesLFandHFShapes **14**

D QCD Rate Uncertainty **15**

E Modeling **18**
E.1 Validation of the Model - Additional distributions 18
E.2 Additional Checks on Di-jet Mass Spectrum in the W+2jets Sample 21

E.2.1 Update of the Previous Result 21
E.2.2 Effect of the Addition in the Selection Criteria: $\Delta R(j_1, j_2) > 0.7$ 22
E.2.3 Effect of the Quark/Gluon Jet Energy Calibrations 24
E.2.4 Effects on the Electron Data of the Corrections on the QCD Multi-jets Background Model 24
E.3 Di-jet mass fit templates 24
E.4 Jet Energy Scale Systematic Uncertainties on the Template Shapes . 24
E.4.1 Diboson 24
E.4.2 Other MC processes 29
E.5 Q2 Systematic Uncertainty 29
E.6 Fit Validation 29

F Top-AntiTop Veto **33**

Bibliography **36**

Acknowledgements

I like to think at this thesis as a long journey which contributed to my growth as a person. During this journey I met numerous people. Some of them had a relatively big impact on my life. This section, a collection of flashbacks during my graduate student life, is dedicated to them.

I would like to start from my family. My mother Pia, my father Mario, my brother Fabio and my grandmother Sofia have always cheered me up no matter how crabby I was. Although I spent most of the time away from Gaeta (my home town), I talked to them on a daily basis. The number of times that my mom listened (or pretended to listen) to my complaints are too many to be counted.

I cannot thank enough my high school friends, who were a great company when I was back in Gaeta for Christmas or Easter holidays. A long time passed by since our teenager years, but hanging out with Flavio (who is Flavio? Ask Nino), Erik (a.k.a. Enrico), Parasmo (a.k.a. Erasmo), and Ampiletti ("l'uomo che fa 31 senza il 30") is always the same cheerful fun. Road trips around Lazio with Floriana, Chiara and Serena also deserve to be mentioned.

My Ph.D experience started in Pisa when Prof. Bellettini and I were deciding my thesis project. Prof. Bellettini, who has been my advisor since the master's thesis (6 years ago) was always patient and kind to me. My strongest memory about him relates to when I had a car accident near Fermilab: Prof. Bellettini came twice to the hospital to bring me company and cheer me up. Sitting in the INFN office with Giuseppe Latino brings a lot of memories. While mentoring me in jet spectroscopy, Giuseppe and I underwent extremely long conversations. I owe him all I know about jets. Attending Standard Model classes with Prof. Barbieri was hard, but I probably learned from him in one class more than I have ever learned before on theoretical physics. A big part of my first year of graduate school consisted in preparing the seminar for Prof. Rolandi's class: that experience contributed to strengthen my knowledge on jets and other topics of experimental physics.

The period in Pisa was not always work and study. Clubbing and wondering around at night time with Daniele, Fons (a.k.a. Piero), Marco ("uno dei due secchi...Io sono l'altro"), Luca and Natalino was a routine: an extremely fun routine.

After that I moved to Fermilab. All my professional experience is connected to one outstanding name: Vadim (a.k.a. Dr. Vadim Rusu). Although the beginning was not pleasant [1], it became very challenging and eventually amusing. In the "PHOS" adventure (Vadim knows) I had great conversations with him. My growth in physics is mostly due to him. During my work I also had the pleasure to work with the brilliant Wesley. His skills and his rapidity in performing complicated studies were impressive. Qiuguang was part of the diboson group, appointed to save the Standard Model against the new physics enemies. How can I forget about the Higgs meetings, where Wes, Qiuguang and I were constantly under the "gun bit"! Nonetheless the Higgs group, mainly represented by Eric, Tom and Homer, was a great help for my analysis. All of them were vital for my success. Last but not least, in the Fermilab period I had the pleasure to work and supervise Sabatino, Caterina, Matteo, Marco, and Ottavia. For the first time I experienced how it feels to be "on the other side of the barricade".

I also spent great moments outside of the lab. Il ciociaro (a.k.a. Matteo) and Jesus were my room-mates in the Maplewood house. Matteo and I often spent time together in the freezing winters in Batavia and, when back in Italy, either in Morolo or Gaeta. Matteo is a good friend of mine. Last in time but not in importance I am grateful to the "Baccaglio" group. I like to think about the members of such a group as the roles of the soccer players in the "figurina Panini del A.C Milan" that I had in my wallet when I was a kid. Giggino (a.k.a. Luigi) is the number 10 of the group: "il fantastista". Taurus (a.k.a. Andrea), the goalkeeper, il Cane (a.k.a. Giani), the right back, Renato (a.k.a. Stefano), the midfielder. ViXx (a.k.a. Vincenzo), Damiano, Manfredi and Ludovico are also part of this team. We often went to Chicago at night time clubbing. I am sure that nobody will ever forget the artificial bump in front of BØ when coming back from Chicago at 4 a.m. Right ViXx? I would like to thank also Donato, with whom I am starting a truthful friendship, Alessandra, Gianfranco, Ivano, and Margherita. And I do not want to forget about my sister Kourtney, who I have known for a long time.

Let me conclude with a proverb: "quando si chiude una porta si apre un portone". The "portone" is named Ana.

[1] Marco: "Vadim, this plot does not seem right". Vadim: you "have a bug".

Abstract

In this thesis we present a search for the WZ and ZZ production in a final state ("W+2 jets") with a leptonically-decaying W and two energetic jets. We use the full dataset ($\int \mathcal{L}dt = 8.9\ fb^{-1}$) recorded with the CDF detector at Fermilab. The challenge consists in extracting the small Z-hadronic peak from the large amount of background processes. Those processes also include the WW, whose hadronic peak cannot be distinguished from the Z peak, due to the poor calorimeter resolution. In the past such a signature was used to measure the diboson cross section, which is highly dominated by the WW cross section ([1]).

A theoretical overview of the Standard Model (SM) of particle physics is presented in Chap. 1. In this chapter, we also explain the reasons for measuring the WZ/ZZ production cross section.

In Chap. 2 and 3 we describe the accelerator complex and the detector (CDF). In Chap. 4 we describe the standard CDF algorithms for the identification of the final-state objects: electrons, muons, neutrinos, and jets. Since a particular attention in this analysis is given to the sub-sample with jets carrying b-flavor ("b-jets"), we spend Chap. 5 for describing the b-jet identification algorithm used in this analysis. The efficient identification of b-jets was the key for discovering the top quark ([2]) and for achieving evidence in the search of the SM light Higgs boson-like particle at the Tevatron ([3]). The CDF tremendous efforts devoted to develop the optimal b-tagger for the light Higgs searches is reported in App. A. The result of this effort produced the HOBIT b-tagger (App. A.0.5), of which I have been one of the developer [2].

Because of detectors inefficiencies and physical effects the measured jet energy may deviate from the primary parton energy. For this reason, the jet energy must be corrected. In Chap. 6 we describe the standard CDF jet corrections.

[2]Because of timing issues we were not able to exploit the more perfomant HOBIT b-tagger in this thesis.

When checking the balancing of the corrected jets against the γ and Z bosons in the γ+jet and Z+jet samples, a significant discrepancy between data and simulations are found. Under the assumption that this discrepancy originates from differences in the modeling of the jet response for quarks and gluons, we derive personalized corrections for quark and gluon jets (Sec. 6.2.4). Together with a graduate student of University of Chicago, I derived these corrections: I was in charge of comparing the jet balancing in data and simulation in the full CDF γ+jet dataset.

Before describing the actual measurement, we devote Chap. 7 to describe the basic MonteCarlo (MC) techniques which are exploited in this thesis. Although I am not the author of any of the described MC generators, I think that understanding the details of the MC simulations should be a part of every HEP physicist's education.

As a preliminary step for measuring the WZ/ZZ cross section, it is imperative to model the W+2 jets dataset really well. Two years ago, the CDF collaboration published a discrepancy in the W+2 jets data which was not described by the theoretical predictions within the known statistical and systematic uncertainties ([119]). We investigated two main systematic effects, which in control samples were found to spoil the agreement between data and predictions: modeling of the jet energy scale, and of the QCD multi-jets background. After all those effects were corrected a good agreement between data and predictions was achieved. Details on the procedure to model the W+2 jets dataset are presented in Chap. 8. The effect on each single correction is reported in App. E.

In order to accurately measure the WZ/ZZ cross-section, we optimize the dataset to be investigated: a number of studies to do so are described in Chap 9. We measure the WZ/ZZ cross-section via a simultaneous fit to the di-jet mass in the b-jet and light-quark jets enriched samples. We find a modest excess of events in di-jet mass, corresponding to a cross section $\sigma_{WZ/ZZ} = 4.7 \pm^{3.0}_{2.5}$ pb, which is consistent the SM cross-section (5.1±0.2 pb). Such a result corresponds to a limit of $\sigma_{WZ/ZZ} < 12.2$ pb at 95% C.L. Details on the fitting procedure and results are given in Chap. 9.

A summary of the results and conclusion remarks are reported in Chap. 10.

Chapter 1

Theoretical Overview and Diboson Landscape

The present understanding of fundamental constituents of matter and of fundamental interactions, which mediate the dynamics of the subatomic particles, is provided by a theory called Standard Model. The Standard Model has been developed in the 1960's and 70's and has been experimentally tested in these last 50 years.

In this chapter an overall description of the Standard Model is given in Sec. 1.1. In Sec. 1.2 we focus on the properties of the W and Z bosons. Finally, in Sec. 1.3, we highlights the importance of the diboson measurements and we give the most up-to-date status on such measurements.

1.1 The Standard Model

The Standard Model (SM) incorporates two gauge theories: the theory of Quantum Chromodinamics (QCD, [4]), which describes the strong interactions, and the Glashow-Salam-Weinberg (GSW) theory of electroweak interactions ([5]), which unifies the weak and electromagnetic interactions. The problem to incorporate gravity in the theory is still open.

The SM describes two categories of fundamental particles [1]: fermions and bosons, which have different spins. The former category, which has half-integer spin, contains six quarks and six leptons, arranged in generations of weak isospin doublets. The gauge bosons, which have integer spins, are the interaction carriers. The Higgs boson permeates the physical vacuum breaking the symmetry of the the-

[1] A particle is considered fundamental if no internal structure is revealed.

ory and giving masses the aforementioned elementary particles. Figure 1.1 shows the fundamental particles in SM, along with the masses and the spins. Each of these particles is associated to its antiparticle.

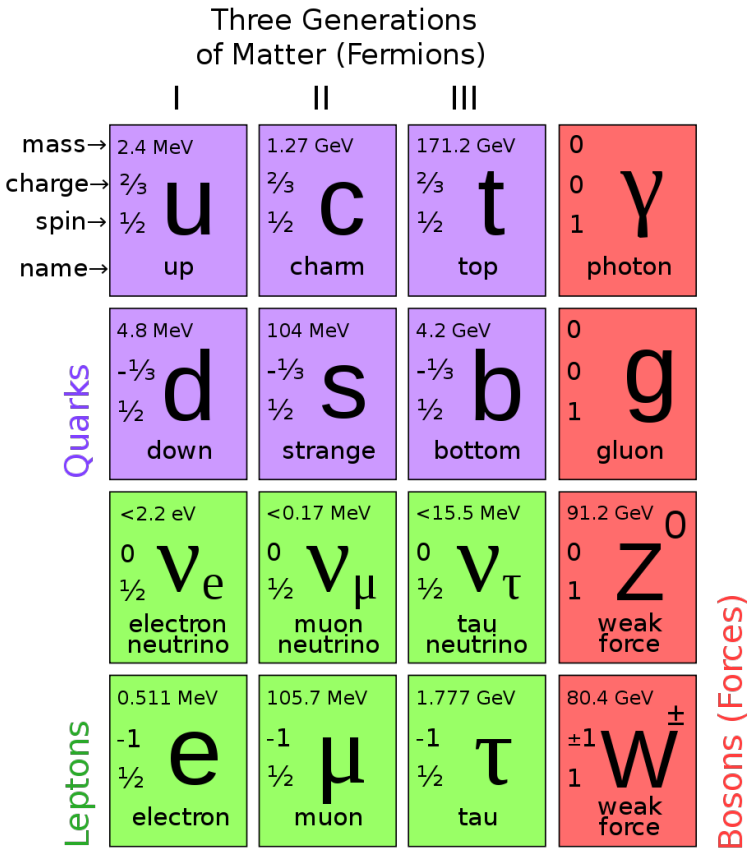


Figure 1.1: *Particles and their properties in the Standard Model. The three generation of quarks and fermions and the four bosons carrying the forces are shown. The recently discovered Higgs boson is not shown.*

1.1.1 Fundamental Interactions

The Standard Model is a local gauge theory. In a gauge theory the fundamental particles are described by quantized fields ψ, characterized by quantum numbers (e.g: spin). The interactions among particles are elegantly described via local *Gauge symmetries*: according to the Noether's theorem ([6]), every differentiable symmetry of the action $S = \int \mathcal{L}(\psi, \partial_\mu \psi) d^4x$ of a physical system [2] has a corresponding conservation law.

[2] $\mathcal{L}$ is the Lagrangian density, which describes the dynamics of the system.

We review below the interactions in the SM.

The Electromagnetic Interaction

The Lagrangian density of a free fermion with mass m is:

$$\mathcal{L}_{Dirac}(x) = \overline{\psi(x)}(i\partial\!\!\!/ - m)\psi(x) \tag{1.1}$$

where ψ is the Dirac spinor, $\partial\!\!\!/ \equiv \gamma^\mu \partial_\mu$, γ^μ being the Dirac matrices.

The Lagrangian density yielding the Maxwell equations for the electromagnetic scalar and vectorial potential A^μ, associated to the photon field, can be expressed as:

$$\mathcal{L}_{Maxwell}(x) = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{1.2}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field tensor.

Eq. 1.1 is invariant under global $U(1)$ rotations ($\psi \to e^{iQ\alpha}\psi$), α being the parameter of the rotation and Q an arbitrary number. If the system is further required to be invariant under a local rotation (i.e.: $U(1)$ with a space-time dependent generator $\alpha(x)$), Eq. 1.1 has to be re-written as follow:

$$\mathcal{L}_{Dirac}(x) = \overline{\psi}(x)(iD\!\!\!/ - m)\psi(x) \tag{1.3}$$

where $D\!\!\!/ \equiv \gamma^\mu D_\mu$, $D_\mu \equiv \partial_\mu + iQA_\mu$ (D_μ is named "covariant derivative"), and $A_\mu \to A_\mu - \frac{1}{Q}\partial_\mu\alpha(x)$ under the local aforementioned rotation [3]. Assembling altogether, the QED Lagrangian density is derived:

$$\mathcal{L}_{QED}(x) = \overline{\psi}(x)(i\partial\!\!\!/_\mu - m)\psi(x) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - QeA^\mu\overline{\psi}\gamma_\mu\psi \tag{1.4}$$

where the new term $QeA^\mu\overline{\psi}\gamma_\mu\psi$ describes the interaction between a fermion of charge Q in units of the electron charge [4] and the photon field. The corresponding current, which is conserved ($\partial_\mu J_\mu^{QED} = 0$, see Noether's theorem), is defined as:

$$J_\mu^{QED} \equiv Q\overline{\psi}\gamma_\mu\psi \tag{1.5}$$

In a classic analogy the conservation of J_0^{QED}, integrated over the space, represents the time-conservation of the electric charge, on which the electromagnetic force depends.

Starting from Eq. 1.5, $\mathcal{L}_{QED}$ in Eq. 1.4 can be re-written as:

$$\mathcal{L}_{QED}(x) = \overline{\psi}(x)(i\partial\!\!\!/_\mu - m)\psi(x) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - eA^\mu J_\mu \tag{1.6}$$

[3] It can be shown that equations 1.2 and 1.3 are respectively invariant under $A_\mu \to A_\mu - \frac{1}{Q}\partial_\mu\alpha(x)$ and $\psi \to e^{iQ\alpha}\psi$. The latter invariance happens since $D_\mu\psi(x) \to e^{iQ\alpha(x)}D_\mu\psi(x)$.

[4] e is the charge of the electron. $e = 1.60217646 \times 10^{-19}$ Coulomb.

The Electroweak Interactions

Weak interaction phenomena show common properties between doublets of fermions ($\binom{\nu_e}{e}$,..., $\binom{u}{d}$,..., Fig. 1.1). Because of this reason the weak isospin is defined and the weak interaction Lagrangian density is built in order to be invariant under a local $SU(2)_L$ rotation in the weak isospin space. Such a rotation is given by $\psi(x) \rightarrow e^{i\boldsymbol{\alpha}(\boldsymbol{x})\boldsymbol{\tau}/2}$, $\boldsymbol{\alpha}$ being the parameters of the rotations, and $\boldsymbol{\tau} = \boldsymbol{\sigma}/2$ the generators of the rotations, with $\boldsymbol{\sigma} = (\sigma^1, \sigma^3, \sigma^3)/2$ being the Pauli matrices.

According to the proposal presented by Glashow in 1961 [5], these two theories, QED and weak, have been unified in the *electroweak (EW) theory*, represented by the group $SU(2)_L \times U(1)_Y$, Y being the hypercharge, as defined by the Gell-Mann-Nishijima formula below:

$$Y = Q - T^3 \tag{1.7}$$

$T^3 \equiv \tau^3$ is the third component of the weak isospin. The invariance under such a transformation implies that two currents are conserved. These currents, $\mathbf{J}_\mu^{weak}$ and J_μ^Y, are defined in analogy with Eq. 1.5:

$$\mathbf{J}_\mu^{weak} \equiv \overline{\chi}\gamma_\mu \frac{\boldsymbol{\sigma}}{2}\chi \tag{1.8}$$

$$J_\mu^Y = \overline{\psi}\gamma_\mu Y \psi = J_\mu^{QED} - J_\mu^3 \tag{1.9}$$

Once the conserved currents are defined, it is straightforward to write the $SU(2)_L \times U(1)_Y$-invariant Lagrangian density:

$$\mathcal{L}_{EW} = \overline{\psi}(i\partial\!\!\!/ - m)\psi - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}\mathbf{W}^{\mu\nu}\mathbf{W}_{\mu\nu} + g\mathbf{J}_\mu^{\mathbf{weak}}\mathbf{W}^\mu + \frac{g^{'}}{2}J_\mu^Y B^\mu \tag{1.10}$$

where:

- $\mathbf{W}^{\mu\nu} = \partial_\mu \mathbf{W}_\nu - \partial_\nu \mathbf{W}_\mu + g\mathbf{W}_\mu \times \mathbf{W}_\nu$, $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ describe the propagation of the $\mathbf{W}_\mu = (W_\mu^1, W_\mu^2, W_\mu^3)$ and B_μ electroweak fields. They are defined in analogy with the QED case. The definition of $\mathbf{W}^{\mu\nu}$ contains an extra term with respect to $F_{\mu\nu}$: it accounts for the non-abelianity of the $SU(2)$ group.

- g, $g^{'}$ are two different coupling constants arising from the the fact that $SU(2)$ and $U(1)$ commute. The coupling constants can be expressed as:

$$e = g\sin(\vartheta_W) = g^{'}\cos(\vartheta_W) \tag{1.11}$$

  ϑ_W being the Weinberg angle [5].

[5] $\sin^2\vartheta_W = 0.23116 \pm 0.00012$ ([10]).

Eq. 1.10 can be re-written as:

$$\mathcal{L}_{EW} = \overline{\psi}(i\not{D} - m)\psi - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}\mathbf{W}^{\mu\nu}\mathbf{W}_{\mu\nu} \tag{1.12}$$

where:

$$D_\mu = \partial_\mu - ig\mathbf{W}_\mu\frac{\boldsymbol{\sigma}}{2} - ig^{'}YB_\mu \tag{1.13}$$

The EW Lagrangian density (Eq. 1.10) uniquely determines the coupling of the boson fields to the fermions, once the quantum number $(T^+, T^-, T^3) \equiv \frac{1}{2}(\sigma^1 + i\sigma^2, \sigma^1 - i\sigma^2, \sigma^3)$ of the fermion fields are specified. To determine these quantum numbers, one must take into account that the W bosons couple only to left-handed chirality states of quarks and leptons [6]. For simplicity fermions are therefore grouped into doublets and singlets as follows:

$$\chi_L = \begin{pmatrix}\psi_u \\ \psi_d\end{pmatrix}_L \tag{1.14}$$

$$\chi_R = \psi_R \tag{1.15}$$

where "L" and "R" mean respectively left-handed and right-handed chirality states. "u" and "d" are the labels up and down fermions in the weak isospin doublet. Such a grouping uniquely defines the quantum numbers of the fermions.

The electroweak theory must also take into account the observed *flavor changing charged currents* in the quark sector [7]. $\psi_d^{quark} \equiv (d, s, b) \rightarrow \psi_d^{'quark} \equiv (d^{'}, s^{'}, b^{'})$ where $(d^{'}, s^{'}, b^{'})$ is expressed according to the following:

$$\begin{pmatrix} d^{'} \\ s^{'} \\ b^{'} \end{pmatrix} = V_{CKM}\begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}\begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.16}$$

where V_{CKM} is the Cabibbo-Kobayashi-Maskawa quark mixing matrix.

Quantum-Chromo-Dynamics

In the SM the strong interaction is described by the Quantum-Chromo-Dynamics (QCD). This force is responsible for quarks "sticking" together to form composite particles (hadrons). QCD is described by a non-abelian local $SU(3)_C$ group.

[6] No right-handed neutrinos have been observed yet. This statement is considered valid for quarks as a natural extension in the electroweak unification.

[7] The first evidence came from strangeness changing charged current. An example of this phenomenon is $\Lambda \rightarrow pe\overline{\nu_e}$.

The relative Lagrangian density, invariant under a $SU(3)$ local transformation, is reported below:

$$\mathcal{L}_{QCD} = \overline{\psi}(i\partial\!\!\!/_\mu - m)\psi - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a - g_S\overline{\psi}^a\gamma_\mu T_S^c\psi_b A_c^\mu \tag{1.17}$$

where:

- a,b,c=1,...,8
- g_S is the strong coupling constant
- T_S^a are the generators of the $SU(3)$ group;
- $G_a^{\mu\nu} = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g_S f_{abc} A_\mu^b A_\nu^c$ describes the propagation of the strong A_μ^a field. It is defined in analogy with the aforementioned $\mathbf{W}^{\mu\nu}$. f_{abc} are the structure constants of the $SU(3)$ group: $[T_S^a, T_S^b] = i2f_{abc}T_S^c$

As for the QED and weak interactions, gauge invariance of the QCD gauge symmetry group implies a new degree of freedom to be conserved: the color. However, unlike QED, the gauge symmetry group is non-abelian, causing gluons (carriers of strong interaction) to possess color charge and interact with each other.

A consequence of the interactions among gluons is that the coupling "constant" $\alpha_s \equiv g_S^2/4\pi$ depends on Q^2 (the interaction momentum transfer scale). To a first approximation in Q^2/λ^2 one has ([7]):

$$\alpha_s(Q^2) = \frac{12\pi}{(33-2f)ln(Q^2/\lambda^2)} \tag{1.18}$$

where f is the number of different flavored quarks with mass less than Q^2. Fits to experimental data set λ around 200 MeV ([6]).

QCD features two very important properties:

- **Asymptotic freedom**. $\alpha_s(Q^2)$ becomes small at large Q^2 (Eq. 1.18). This means that quarks and gluons interact weakly if they are within a short range. This property allows perturbation theory to be used in theoretical calculations to produce experimentally verifiable predictions for hard scattering processes.
- **Confinement**. Colored particles are confined into colorless singlets (hadrons) by an increasing QCD potential with increasing relative distance. If quarks are forced to large relative distances the energy density in the binding color string increases and energy is materialized into colored quark pairs. Thus a hard scattered parton evolves into a shower of partons and finally into hadrons

(hadronization). Even if analytically unproven, confinement is widely believed to be true because it explains the consistent failure of free quark searches.

In the $SU(3)$ representation of QCD, the gluons have an effective color charge that is larger than of the quarks. The effective gluon and quark color charges are proportional to the square roots of C_A and C_F respectively, where $C_A = 3$ and $C_F = 4/3$ [8]. The larger effective gluon color charge with respect to the quark one results in a number of properties (e.g: broadness, particle multiplicity with the jet) distinguishing gluon jets from quark jets ([8]).

1.1.2 EW Symmetry Breaking: Higgs Model

The $SU(2)$ mentioned earlier would be exact if fermions, W and Z^0 bosons were massless. In order for the standard model to be compatible with the large observed masses of W and Z^0 bosons [9], symmetry breaking must occur.

Spontaneous electroweak symmetry breaking (EWSB) can be accomplished by the introduction of a scalar field ϕ which interacts with both fermions and gauge bosons. Since the EW interactions are local and the masses of the particles are different from zero also when non interacting (during free propagation), the scalar field must be always locally present. The scalar field is then usually said to have a non vanishing vacuum expectation value. This mechanism is called Higgs mechanism, after the name of the scientist who proposed an interaction term in the Lagrangian known as Higgs Field ([9]). The existence of a massive boson, the Higgs boson, would be associated with the Higgs field [10]. The Lagrangian density for such a scalar field is:

$$\mathcal{L}_H = |D_\mu \phi|^2 - V(\phi) \tag{1.19}$$

where:

- the covariant derivative D_μ (Eq. 1.13) takes care of the interaction between the Higgs field and the electroweak bosons.
- the potential $V(\phi)$ [11], which includes the self interaction of the Higgs field, is

[8] It can be shown that $C_A = N$ and $C_F = (N^2 - 1)/2N$ for a $SU(N)$ invariant theory, N being a given integer ([7]).

[9] The quantum fields of the $W^\pm$ and Z^0 mass eigenstates are defined as $W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu)$, $Z^0_\mu = \frac{1}{\sqrt{g^2+g'^2}}(gW^3_\mu - g' B_\mu)$, with $\mathbf{W}_\mu$, B_μ, g, g$'$ defined in the last section. $M_{W^\pm} = 80.385 \pm 0.015$ GeV/c^2 and $M_{Z^0} = 91.1876 \pm 0.0021$ GeV/c^2 ([10])

[10] A SM Higgs-like boson was recently discovered at CERN [11].

[11] In order to preserve the renormalizability of the theory terms up to dimension 4 are kept.

given by:

$$V(\phi)_H = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{1.20}$$

with $\mu^2 < 0$ and $\lambda > 0$

The Higgs mechanism predicts two complex scalar fields, such that:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.21}$$

where the field ϕ^+/ϕ^0 are respectively the charged/neutral component of the doublet. The symmetry is broken since the minimum of $V(\phi)$ occurs at $\sqrt{<\phi^\dagger\phi>} = \sqrt{\frac{-\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}} \neq 0$. The gauge freedom allows to conveniently choose the scalar field in its ground state, as follows:

$$<\phi>= \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.22}$$

The gauge boson mass terms come directly from substituting Eq. 1.22 into Eq. 1.19:

$$M_{W^\pm} = g\frac{v}{2} \tag{1.23}$$

$$M_{Z^0} = \frac{v}{2}\sqrt{g^2 + g'^2} \tag{1.24}$$

while the photon remains massless (Eq. 1.22 does not break the $U(1)_{EM}$ symmetry [12]).

From Eq. 1.11 one can see that the masses of the weak bosons are related by the weak missing angle, as follows:

$$sin^2\vartheta_W = 1 - \frac{M_W^\pm}{M_{Z^0}} \tag{1.25}$$

The ϕ doublet in its ground state can be parameterized also in terms of the Higgs boson H:

$$<\phi(x)>= \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{1.26}$$

From Eq. 1.20 and 1.26 it is possible to derive the Higgs mass term:

$$M_H = 2v\sqrt{\lambda} = \mu\sqrt{2} \tag{1.27}$$

The fermion masses can also be generated if a Yukawa couplings for the upper (U) and lower (D) components of fermion isospin doublets (Eq. 1.15) is considered:

$$\mathcal{L}_{\text{Yukawa},D} = -g_D[(\overline{\psi}_L^U, \overline{\psi}_L^D)\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}\psi_R^D + \overline{\psi}_R^D(\phi^{+\dagger}, \phi^{0\dagger})\begin{pmatrix} \psi_L^U \\ \psi_L^D \end{pmatrix}] \tag{1.28}$$

[12] $SU(2)_L \times U(1)_Y$ is spontaneously broken down to the electromagnetic $U(1)_{EM}$ symmetry

and a similar expression for $\mathcal{L}_{\text{Yukawa},U}$. $g_{U/D}$ are the coupling of the fermions to the Higgs boson. Those Lagrangians lead to:

$$M_U = g_U v/\sqrt{2} \tag{1.29}$$

$$M_D = g_D v/\sqrt{2} \tag{1.30}$$

where the constants $g_{U/D}$ can be extracted by the measurements of the fermion masses.

1.1.3 Lagrangian of the Standard Model

From the previous sections we learned that the SM describes all interactions in nature, but gravity, by exploiting a local gauge $SU(3)_C \times SU(2)_L \times U(1)_Y$ structure. When putting all pieces together, the SM Lagrangian density is defined as follows ([12]):

$$\begin{aligned}
\mathcal{L}_{SM} = \; & i\overline{\psi}_L \partial\!\!\!/ \psi_L + i\overline{\psi}_R \partial\!\!\!/ \psi_R + \\
& -\frac{1}{2}W^{+\mu\nu}W^-_{\mu\nu} - \frac{1}{4}Z^{\mu\nu}Z_{\mu\nu} - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \\
& -igcos(\vartheta_W)[(W^-_\mu W^+_\nu - W^-_\nu W^+_\mu)\partial^\mu Z^\nu + W^+_{\mu\nu}W^{-\mu}Z^\nu - W^-_{\mu\nu}W^{+\mu}Z^\nu] \\
& +ie[(W^-_\mu W^+_\nu - W^-_\nu W^+_\mu)\partial^\mu A^\nu + W^+_{\mu\nu}W^{-\mu}A^\nu - W^-_{\mu\nu}W^{+\mu}A^\nu] + \\
& +g^2cos^2(\vartheta_W)(W^+_\mu W^-_\nu Z^\mu Z^\nu - W^-_\mu W^{+\mu}Z_\nu Z^\nu) + \\
& +g^2(W^+_\mu W^-_\nu A^\mu A^\nu - W^-_\mu W^{+\mu}A_\nu A^\nu) + \\
& +\text{eg}cos(\vartheta_W)[(W^+_\mu W^-_\nu (Z^\mu A^\nu + Z^\nu A^\mu) - 2W^+_\mu W^{-\mu}Z_\nu A^\nu] + \\
& +\frac{\text{g}^2}{2}[(W^+_\mu W^-_\nu (W^{+\mu}W^{-\nu} - W^{+\nu}W^{-\mu})] + \\
& +\frac{\text{g}}{\sqrt{2}}[W^+_\mu(\overline{\nu}_L\gamma^\mu l_L + V_{CKM}\overline{u}_L\gamma^\mu d_L) + \text{h.c.}] + \\
& +\frac{\text{g}}{cos\vartheta_W}Z_\mu(T^3 - Qsin^2\vartheta_W)\overline{\psi}_{L,R}\gamma^\mu\psi_{L,R} - eQA_\mu\overline{\psi}\gamma^\mu\psi + \\
& -\sqrt{2}(M\overline{\psi}_L\psi_R + \text{h.c.}) + M^2_W W^{+\mu}W^-_\mu + \frac{M^2_{Z^0}}{2}Z^\mu Z_\mu + \\
& +\frac{1}{2}\partial_\mu H\partial^\mu H + \frac{\text{g}^2}{4}(2v+H)H(W^{+\mu}W^-_\mu + \frac{M^2_{Z^0}}{2}Z^\mu Z_\mu) - \sqrt{2}\frac{M}{v}H\overline{\psi}\psi + \\
& +\lambda(\frac{3}{4}v^4 + 2v^3H + 2v^2H^2 + vH^3 + \frac{1}{4}H^4) + \\
& -\frac{1}{4}G^{\mu\nu}_a G^a_{\mu\nu} - g_S\overline{\psi}^a\gamma_\mu T^c_S\psi_b A^\mu_c
\end{aligned} \tag{1.31}$$

where $l = e, \mu, \tau$, $\nu = \nu_e, \nu_\mu, \nu_\tau$, u/d each up/down quark in Fig. 1.1, M is either M_U or M_D.

1.2 W and Z Decay Modes and Branching Ratios

As we mentioned earlier, the W boson couples to left handed quarks and leptons in the same way [13] and does not couple to right handed fermions. However, because of the color factor and because the W cannot decay to top (W mass smaller than top mass) the overall branching ratio of the W boson into hadrons is two times higher than into leptons (Fig. 1.2).

The couplings of the Z boson to fermions are slightly more complicated, since it couples differently to left and right fermions. The difference depends on $T_3 - Q\sin^2\vartheta_W$, as written in Eq. 1.31. In Table 1.1 we summarize some of the fermion quantum numbers. Given those numbers, and $\sin^2\vartheta_W \sim 0.23$, the different Z branching ratios can be derived (Fig. 1.2).

Fermion	Q	T^3	Y	EM	weak	strong
leptons						
ν_l	0	+1/2	-1/2	✗	✓	✗
l^-	-1	-1/2	-1/2	✓	✓	✗
quarks						
u	+2/3	+1/2	+1/6	✓	✓	✓
d	-1/3	-1/2	+1/6	✓	✓	✓

Table 1.1: Some of the lepton and quark main quantum numbers, along with the force experienced by the fermions. Gravity, which is experienced by all of them, is not shown.

1.3 Diboson Physics

There are numerous reasons to study diboson physics. One of them is to test the electroweak sector of the SM or, equivalently, search for physics beyond the SM. From the third and fourth line of Eq. 1.31, one can see that the SM allows for WWZ and $WW\gamma$ triple gauge couplings (TGC). These couplings can be measured by studying the diboson (WW, WZ, and $W\gamma$) production properties. Anomalies in the coupling strength may alter rates and kinematic distributions (e.g: W or Z P_T), thus hinting to new physics. The virtual W exchange in WZ production (Fig. 1.3)

[13] In this picture we are not considering the mixing of the quark generations parameterized by V_{CKM} matrix, since the couplings are averaged over the different quark generations.

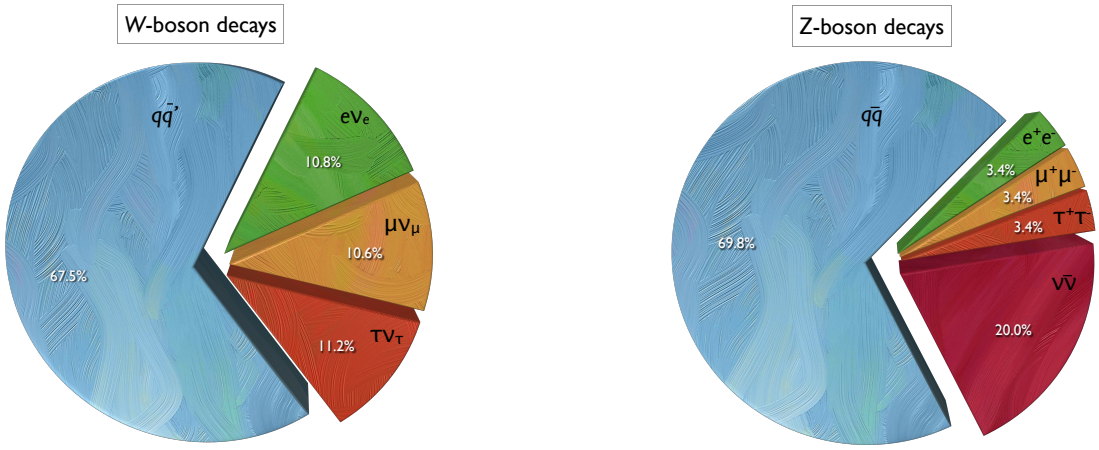


Figure 1.2: *Percentage of the main W (left) and Z (right) decay modes, as predicted by the Standard Model ([10]). q indicates any quark (and anti-quark) in Fig 1.1, When not shown, the sum over charge is considered.*

is very sensitive to the WWZ coupling. An effective theory accounting for possible non-SM physics was proposed in [13]. Such a theory parameterizes WWZ TGC in terms of 7 Anomalous Triple Gauge Couplings (ATGC). In the case of vector bosons coupled to fermions, whose masses are much smaller than the boson ones:

$$\begin{aligned}\mathcal{L}_{WWZ} &= -g\cos\vartheta_W\{ig_1^Z(W^+_{\mu\nu}W^{-\mu}Z^\nu - W^+_\mu Z_\nu W^{-\mu\nu}) + i\kappa_Z W^+_\mu W^-_\mu Z^{\mu\nu} + \\ &+i\frac{\lambda_Z}{M_W^2}W^+_{\lambda\mu}W^{-\mu}_\nu Z^{\nu\lambda} - g_4^Z W^+_\mu W^-_{\nu+}(\partial^\mu Z^\nu + \partial^\nu Z^\mu) + \\ &+g_5^Z\varepsilon^{\mu\nu\rho\sigma}[W^+_\mu(\partial_\rho W^-_\nu) - (\partial_\rho W^+_\mu)W^-_\nu]Z_\sigma + \\ &+i\tilde{\kappa}_Z W^+_\mu W^-_\nu\frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}Z_{\rho\sigma} + \frac{i\tilde{\lambda}_Z}{M_W^2}W^+_{\lambda\mu}W^{-\mu}\frac{1}{2}\epsilon^{\nu\lambda\rho\sigma}Z_{\rho\sigma}\}\end{aligned} \tag{1.32}$$

$\mathcal{L}_{WWZ}$ assumes the SM appearance at tree level if $g_1^Z = \kappa_Z = 1$ and $\lambda_Z = g_4 = g_5 = \tilde{\kappa}_Z = \tilde{\lambda}_Z = 0$. The same theory can be extended to $WW\gamma$ couplings: in that case the above Lagrangian density has the same form. One can show that that $\tilde{\lambda}_{Z,\gamma} = g_4^{Z,\gamma} = g_5^{Z,\gamma} = 0$ by assuming CP and gauge invariances. By using the experimental constraint on the electric dipole moment of the neutron one can impose $\tilde{\kappa}_\gamma = 0$ and expects that $\tilde{\kappa}_Z = 0$. If the electromagnetic gauge invariance is further assumed $g_1^\gamma = 0$ [14]. Therefore, after the aforementioned assumptions, the independent couplings to be measured are five: g_1^Z, $\kappa_{Z,\gamma}$, and $\lambda_{Z,\gamma}$. Those couplings,

[14] For a detailed description on the reduction of the number of parameters in $\mathcal{L}_{WWZ}$ (Eq. 1.32) and $\mathcal{L}_{WW\gamma}$ see [13], [15].

as a function of the partonic center-of-mass collision energy $\sqrt{\hat{s}}$, are defined below:

$$\lambda_{Z,\gamma}(\hat{s}) = \frac{\lambda_{Z,\gamma}}{(1+\hat{s}^2/\Lambda^2)^2} \tag{1.33}$$

$$\Delta\kappa_{Z,\gamma}(\hat{s}) \equiv \kappa_{Z,\gamma}(\hat{s}) - 1 = \frac{\kappa_{Z,\gamma} - 1}{(1+\hat{s}^2/\Lambda^2)^2} \tag{1.34}$$

$$\Delta g_1^Z(\hat{s}) \equiv g_1^Z(\hat{s}) - 1 = \frac{g_1^Z - 1}{(1+\hat{s}^2/\Lambda^2)^2} \tag{1.35}$$

where a form factor Λ has been introduced to preserve unitarity. The most up-to-date limits on ATGC are reported in Fig. 1.4. All measured values show consistency with the SM. From Fig. 1.4 one can see that the major part of the limits are extracted by looking at fully leptonic decay modes (e.g: $WW \to l\nu l\nu$, $WZ \to l\nu ll$), which provide very clean signatures. However, despite of the large background rates, signatures with jets in the final state (WV, $V = W, Z$) are also used for studying ATGC, since the majority of the diboson yield is expected there (Fig. 1.2). Moreover, it is possible that new physics may be hiding in decays with hadronic final states ([14]).

Another reason for studying diboson production in a final state with leptons, large missing transverse energy and jets is to pave the ground for low mass SM Higgs searches. For a standard model Higgs with a mass of 125 GeV/c^2 one of the most fruitful search channels at the Tevatron is the associated WH production in the above signature. As it can be seen from Fig. 1.5, a Higgs boson of this mass is most likely to decay in $b\bar{b}$. However, due to large non-resonant $b\bar{b}$ background, the inclusive $H \to b\bar{b}$ signal is very difficult to isolate: that is why the Higgs signal at the Tevatron is preferred to be searched for in the associated production (e.g: $WH \to l\nu b\bar{b}$) mode. Searches for diboson production fit well in this context. Being able to extract the diboson signal, which was already observed in leptonic signatures at Tevatron and/or other experiments (e.g: LEP), provide a test bench of our understanding of the background processes. Moreover the diboson signal, which is characterized by larger cross section times branching ratios than WH, can be used to optimize the search tools on the data, despite of the small differences (e.g: mass, spin) which characterize the two processes. Last but not least, diboson processes are one of the major backgrounds in the light Higgs boson searches.

1.3.1 Cross-Section Measurements: State of the Art

In this section we give a brief history of the diboson measurements in the various final states, and then we present the most up-to-date diboson cross-section measurements

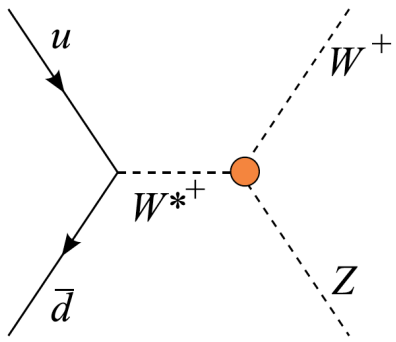


Figure 1.3: *Leading order Feynman diagram for the WZ production in the s-channel*

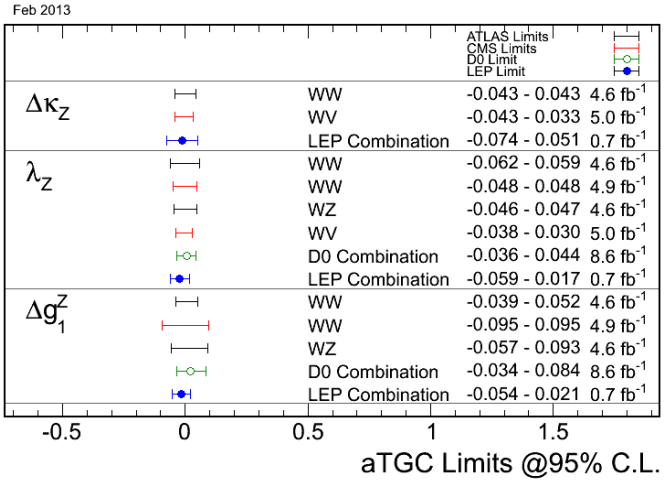
at the Tevatron and LHC. We focus on the production of the massive boson pairs (WW, WZ, ZZ). More details on diboson physics can be found in the recent review [17].

Production cross sections for heavy boson pairs, as computed at next to leading order, are reported in Table 1.2. Since WW and WZ can also be produced via the s-channel (Fig. 1.8, 1.6, Eq. 1.31) WW and WZ boson pairs are produced more often than ZZ. Moreover, because of the additional $WW\gamma$ coupling (Fig. 1.8, Eq. 1.31) and the difference between the W and Z masses, the WW production has the most prominent cross section.

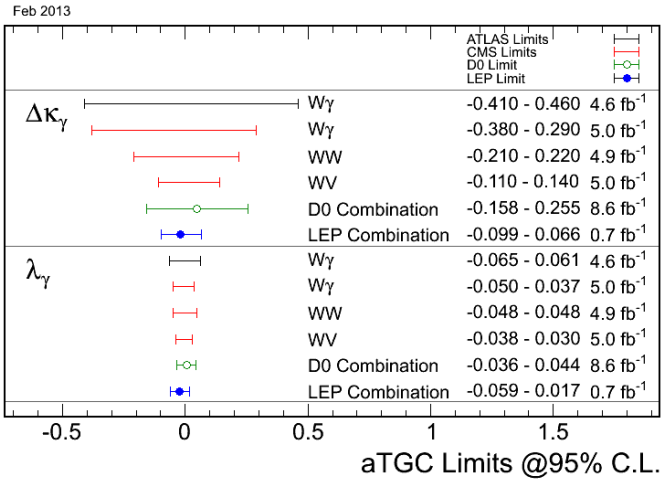
Process	Cross-section (pb) (Tevatron at $\sqrt{s} = 1.96$ TeV)	Cross-section (pb) (LHC at $\sqrt{s} = 7$ TeV)
W^+W^-	11.3 ± 0.7	$47.0 \pm^{2.0}_{1.5}$
W^+Z	3.7 ± 0.2	$11.9\pm^{0.7}_{0.5}$
W^-Z		$6.7\pm^{0.4}_{0.3}$
ZZ^*	1.4 ± 0.1	$6.5\pm^{0.3}_{0.2}$

Table 1.2: Diboson cross sections computed at next to leading order for Tevatron and LHC at of center of mass energies of 1.96 and 7 TeV respectively. The W^+Z and W^-Z cross sections are combined in the case of the Tevatron since, because of the CP-symmetric initial $p\bar{p}$ state, they are equal.

The first evidence of WW, WZ, and ZZ production at the Tevatron came from the fully leptonic signatures $l\nu l\nu$, $l\nu ll$, $llll/ll\nu\nu$, l being an electron or muon and ν a neutrino. Such signatures are very clean. The observation of WW was achieved by CDF and DØ with $\int \mathcal{L}dt$ = 224-252 pb^{-1} ([20]) and $\int \mathcal{L}dt$ = 184 pb^{-1} ([21])



(a)



(b)

Figure 1.4: *A summary of anomalous charged trilinear gauge coupling (ATGC) limits at 95% C.L. Limits are measured for different diboson processes and decay channels at LEP, DØ , CMS, ATLAS experiments ([16]). A number of assumptions are made before extracting the limits. These assumptions as well as the definition of the symbols are described in the text.*

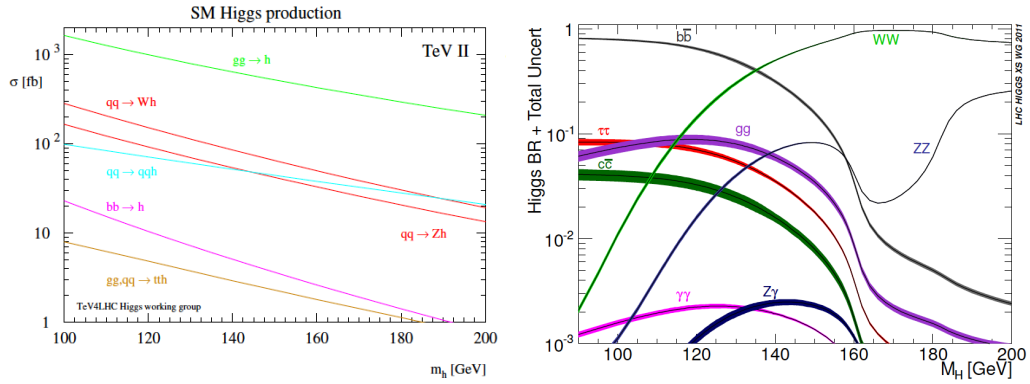


Figure 1.5: *Higgs boson production cross sections in different modes at the Tevatron $p\bar{p}$ collider at a center of mass energy of 1.96 TeV (left, [18]). Higgs boson decay branching ratios (right, [19].*

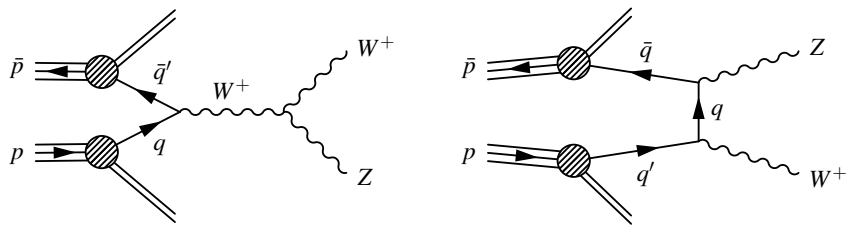


Figure 1.6: *Example of leading order Feynman diagram for the WZ production*

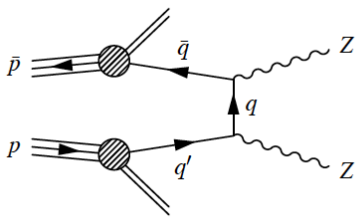


Figure 1.7: *Example of leading order Feynman diagram for the ZZ production.*

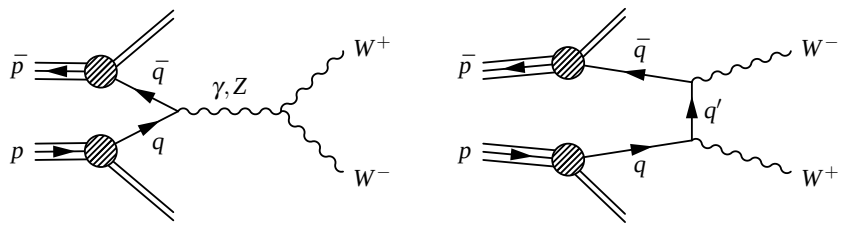


Figure 1.8: *Example of leading order Feynman diagram for the WW production*

respectively. WZ was observed with $\int \mathcal{L}dt = 1.1$ fb^{-1} at CDF ([22]) and $\int \mathcal{L}dt = 1.0\ fb^{-1}$ at DØ ([23]). Evidence of the ZZ production was achieved by CDF by using $\int \mathcal{L}dt = 1.9$ fb^{-1} ([24]), while DØ used $\int \mathcal{L}dt = 1.7$ fb^{-1} to achieve the observation of ZZ production ([25]). WW and ZZ were previously observed at LEP ([26]). The WZ charged final state was inaccessible from the neutral e^+e^- initial state at LEP.

As it can be seen from Fig. 1.2, the largest amount of diboson pairs decays in a final state with jets. However, at hadron colliders those signatures are contaminated by severe W/Z+jets and QCD multi-jets background. Below we will discuss the observations of the diboson processes in signature where one weak boson decays into leptons and the other bosons decays hadronically. Diboson observation in these signatures, which are referred as "semi-leptonic", were already performed at LEP ([26]). Diboson production in the fully hadronic final states are yet to be observed.

Both CDF and DØ measured the diboson cross section in the following two semileptonic final states:

- large momentum imbalance plus jets: "$METjj$"
- high-energy lepton and large momentum imbalance plus jets: "$l\nu jj$"

the former has a large acceptance to diboson decays, as it allows for contributions from both $W \to l\nu$ with an undetected lepton, and $Z \to \nu\nu$. Both searches were performed by exploiting advanced analysis techniques, as the large background contaminations in the samples do not allow for a simple event counting above the background, as done for the fully leptonic signatures. Those techniques heavily rely on the di-jet mass distribution in order to discriminate the signal from the background. However, because of the poor di-jet mass resolution provided by the CDF and DØ calorimeters, these searches are in practice combined measurements of the $WW + WZ + ZZ$ in the METjj signature and $WW + WZ$ in the $l\nu jj$ one.

The most significant background to diboson signal in the METjj sample are the $W(\to l\nu)$ + jets, $Z(\to \nu\nu)$+jets, and QCD multi-jets, where fake large momentum imbalance arises from calorimeter mis-measurements. By exploiting a dataset with $\int \mathcal{L}dt = 3.5$ fb^{-1} CDF achieved a signal significance of 5.3 σ [27].

By exploiting a dataset with $\int \mathcal{L}dt = 1.1$ fb^{-1} DØ was able to observe the $WW+WZ$ signal (4.4σ signal significance). A multivariate discriminant, composed of 12 sensitive kinematic variables besides the di-jet mass, was employed to extract the signal from the large background ([28]). On the other side, CDF employed

two methods to extract the $WW + WZ$ signal. In the first method, a fit to the di-jet invariant mass was performed on a dataset with $\int \mathcal{L}dt = 3.9$ fb^{-1} : the signal significance was 4.6 σ. In the second method CDF took advantage of more kinematic information in the event, by building a discriminant based on the calculations of the differential cross sections of the signal and background processes. Such a method, named matrix-element method, achieved a higher significance of 5.4 σ, despite the smaller dataset ($\int \mathcal{L}dt = 2.7$ fb^{-1}). In addition to the aforementioned dedicated diboson measurements, both CDF and DØ have produced evidence of $WZ + ZZ$ production by exploiting the same analysis technique developed for the low mass Higgs searches. The only difference is that the final discriminant was retrained to discriminate the diboson signal against the background. By exploiting different signatures ($METjj$, $l\nu jj$, $lljj$), and after requiring the jets to originate from b-quarks CDF was able to achieve a $WZ + ZZ$ signal significance of 3.2 σ. The dataset has a $\int \mathcal{L}dt = 9.5$ fb^{-1} ([29]). DØ , by exploiting the same signatures in a dataset with $\int \mathcal{L}dt = 9.5$-9.7 fb^{-1}, achieved a 2.5 σ significance ([30]).

Some of the above Tevatron results were recently (Summer 2012) updated, thanks to the improved techniques and larger available integrated luminosities. The most up-to-date cross section results at the Tevatron are shown in Table 1.3 together with those at the LHC. From that table one can see that the WZ signal is yet to be extracted in the $l\nu jj$ signature alone: this was the goal of this thesis.

Cross section (pb)					
Process	Signature	Theory	Measured	Experiment	$\int \mathcal{L}dt$
WW	$l\nu l\nu$	11.3 ± 0.7 ([31])	$12.1 \pm^{1.8}_{1.7}$	CDF ([32])	3.6
			11.5 ± 2.2	DØ ([33])	1.0
		47.0 ± 2.0 ([31])	52.4 ± 5.1	CMS ([34])	4.9
			54.4 ± 5.9	ATLAS ([35])	1.0
WZ	$lll\nu$	3.7 ± 0.3 ([31])	3.9 ± 0.8	CDF ([36])	7.1
			$4.5 \pm^{0.6}_{0.7}$	DØ ([37])	8.6
		$17.3 \pm^{1.3}_{0.8}$ ([31])	17.0 ± 2.8	CMS ([38])	1.1
			$20.5 \pm^{3.5}_{3.2}$	ATLAS ([39])	1.0
ZZ	$llll$, $ll\nu\nu$	1.4 ± 0.1 ([31])	1.6 ± 0.4	CDF ([40])	6.1
			1.4 ± 0.3	DØ ([37])	8.6
	$llll$	$6.5 \pm^{0.3}_{0.2}$ ([31])	$3.8 \pm^{1.5}_{1.2}$	CMS ([38])	1.1
			$8.5 \pm^{2.7}_{2.3}$	ATLAS ([41])	1.0
WV	$l\nu jj$	16.1 ± 0.9 ([31])	18.1 ± 4.1	CDF ([1])	4.3
			$16.5 \pm^{3.3}_{3.0}$	CDF ([42])	4.6
			$19.6 \pm^{3.2}_{3.0}$	DØ ([43])	4.6
VV	$\nu\nu jj$	16.8 ± 0.5 ([31])	18.0 ± 3.8	CDF ([27])	3.5
VZ	$ll/l\nu/\nu\nu$ + HF	4.4 ± 0.3 ([31])	$4.1 \pm^{1.8}_{1.3}$	CDF ([44])	9.5
			5.0 ± 1.6	DØ ([45])	8.4

Table 1.3: A summary of the most up-to-date cross-section measurements of the massive boson pair production at the Tevatron and LHC. References to the individual measurements are reported. The errors are the sum in quadrature of the statistical and systematic errors. All the measured cross sections are in agreement with the theoretical expectations, calculated with NLO precision. The integrated luminosity and the final state signatures are also shown. We refer to V as W or Z, l as an electron or muon, ν as a neutrino

Chapter 2

Accelerator Complex

The Tevatron in Batavia, Illinois, US was the first large-scale superconducting synchrotron in the world. Originally named the Energy Doubler since, as a proton synchrotron, it was reaching twice the energy of the original Fermilab facility (the "Main Ring"), it began operation in 1983 in fixed target mode and in 1985 as a proton-antiproton collider. From 1985 to 2011 periods of colliding protons and antiproton alternated with periods of inactivity (shut downs) for upgrading the machine. From 2001 to end of the operations on September 30, 2011, the $p\overline{p}$ energy in center of mass system was 1.96 TeV.

In this chapter we will briefly recall the proton (Sec. 2.1) and anti-proton (Sec. 2.2) production, the accelerator complex (Sec. 2.3) eventually bringing protons and anti-protons to collide at the above center of mass energy. Finally (Sec. 2.4) we will highlight the accelerator-complex performances.

2.1 Proton production and first step in the acceleration

Hydrogen gas is introduced in a container, where strong ionization induced by pulsed electric field in presence of a containing magnetic field produces occasionally a number of negative ions. These are used for accelerating beam up to GeV energy with negligible electron-capture losses. A pulsed electrostatic extractor then accelerates the negative ions out of the source at a repetition rate up to 15 Hz and at a energy of 15-22 KeV . After that, ions are further accelerated by a *Cockroft-Walton electrostatic accelerator* to an energy of 750 KeV. Ions, segmented into bunches, are fed into the *Linear Accelerator (Linac, [46])*. The Linac (Fig. 2.1) is approximately 150 m long and comprises two sections. In the first one, five accelerating cavities

with a drift tube inner core, fed by a single RF generator resonating at 201.25 MHz, accelerate ions to approximately 116 MeV. The second one, comprising 7 RF cavities at 805 MHz, fed by a set of Klystron amplifiers, ramps ions to 401.5 MeV. At the Linac exit the negative ion beam strikes a thin carbon target and turns into a proton beam by electron stripping.

Stripped protons enter the *Booster* (Fig. 2.1), a 8 GeV synchrotron whose diameter is about 150 m. To maintain a constant circular orbit the dipole magnetic field in the Booster increases from 0.74 Tesla to 7 Tesla during acceleration. To reach 8 GeV the injected protons circulate in the booster for about 33 milliseconds.

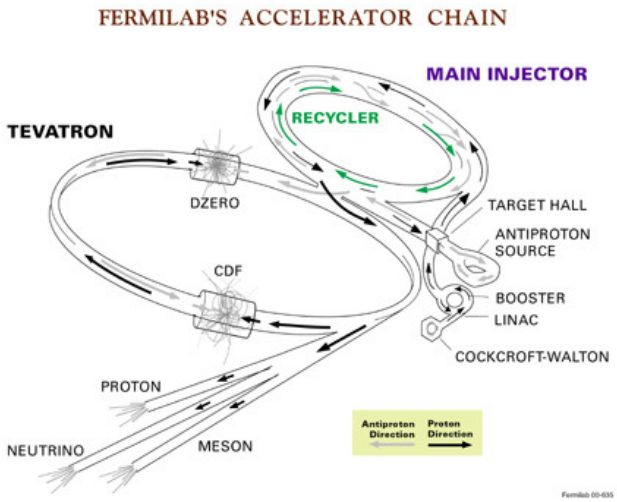


Figure 2.1: *The accelerator complex of the Fermi National Accelerator Laboratory.*

2.2 Anti-proton production and accumulation

The *Main Injector* ([47]) is the next link in the accelerator chain. In accumulation mode, a pulse of 8×10^{12} protons is extracted from the Booster and injected every 2.2 seconds. The main injector is a 53.1 MHz circular synchrotron of a 528.5 m radius, with 18 accelerating cavities and conventional magnets. The protons are accelerated to 120 GeV and then directed to the *anti-proton station*, which is a rotating 7 cm-thick target made of nickel alloys containing chromium, iron and other metals. The resulting particles spray contains some anti-protons with a broad momentum and wide-spread spatial distribution. A cylindrical lithium lens (760 T/m) focuses the particles produced around the forward direction. Negative particles in a 35 mrad cone about the forward direction are selected by a 1.5 T pulsed dipole magnet, focused by strong magnetic lenses and injected in the Debuncher Storage Ring.

Typically, 21 anti-protons per 106 protons on target are collected. Anti-protons are at 8 GeV energy. The stacking rate is approximately $10 - 20$ mA/hour.

In the *Debuncher ring*, a rounded triangular-shaped synchrotron with mean radius of 90 meters stochastic cooling and bunch phase rotation are used to reduce momentum spread while increasing time spread. After each beam pulse the Debuncher is emptied. The anti-proton bunches (8 ± 0.018 GeV) are transferred with a 60%-70% efficiency to the *Anti-proton Accumulator*, a 75 m mean radius storage ring of larger acceptance housed in the same tunnel as the Debuncher (see a sketch in Fig. 2.2). In the Accumulator multiple beam pulses are stacked and $\overline{p}$ are further cooled to increase the anti-proton phase space density.

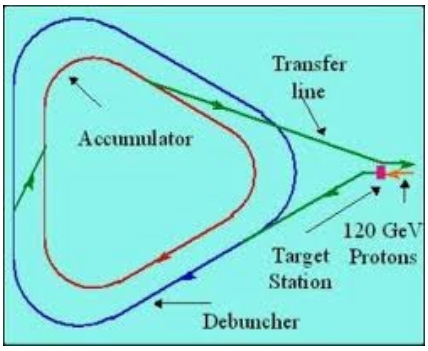


Figure 2.2: *Fermilab Debuncher, Accumulator: a zoom of Figure 2.1.*

Besides the small anti-protons production cross section, problems in anti-proton collection, cooling and stacking are among the main causes limiting the final Tevatron luminosity. A further improvement of the anti-proton source is the *Recycler* (see Fig. 2.1), a post-accumulator storage ring of constant 8 GeV energy, located in the Main-Injector enclosure and composed of permanent magnets. Because of the larger acceptance of the Recycler (it can store an anti-proton current up to over 2.5 Amps, much larger than the Accumulator), its role is to store anti-protons, which are periodically transferred from the Accumulator (95% transfer efficiency). Anti-protons are also further cooled to increase the storing capacity of the recycler and the beam transfer efficiency to the Tevatron.

2.3 Injections and collisions

In normal conditions every 10-20 hours the recycler was fully loaded and anti-proton accumulation was stopped. Protons from the Booster were injected into the Main Injector, accelerated to 150 GeV, coalesced into single bunches of about 300

$\times$ 10^9 protons, and then injected into the *Tevatron*, a large synchrotron of 1 Km radius. The entire process is repeated until 36 bunches are transferred with a timing separation of 396 ns from each other. Typically, the transferring efficiency from the Main Injector to the Tevatron is about 65%.

After the protons are loaded, 7-11 anti-proton bunches are extracted from the either Accumulator or Recycler to the Main Injector, accelerated to 150 GeV, coalesced into four 30×10^9 $\overline{p}$ separated by 396 ns, and then injected into the Tevatron. The anti-proton injection process is repeated until 36 anti-proton bunches circulate in the Tevatron.

Protons and anti-protons circulate in the same vacuum pipe. Electrostatic separators reduce to a negligible amount the unwanted interactions, by keeping the beams away from each other at all points in the orbit helix [1], except at the collision points. Protons and anti-protons are accelerated to 980 GeV . A tour of the Tevatron takes about 21 μs. About one minute is needed to reach the final beam energy [2].

High-gradient focusing quadrupole magnets ("low-β squeezers") reduce the transverse spatial spread to minimize the beam section at the interaction regions and therefore maximize the collision rate. Interactions regions are located where the DØ and the CDF II detector are placed. The resulting transverse beam distributions are approximated by 2D Gaussian functions, with $\sigma = 30$ μm. The typical longitudinal dimension of a bunch is 60-70 cm. The event source is roughly distributed longitudinally as a Gaussian with $\sigma_z = 28$ cm [3].

The 36 bunches of protons or anti-protons are collections of buckets adding up to 1113 in three equispaced trains (see Fig. 2.3). Within a train the inter-bunch time is 396 ns (21 buckets) while inter-train time is 2.6 μs (139 buckets). The intra-train empty sectors allow for anti-proton injection without perturbing the orbiting protons and allow enough time for fast kicker magnets to abort the beam into a dump before the arrival of the next train in case of emergency. As a consequence of this configuration, the average bunch crossing rate is 1.7 MHz.

[1] Intra-beam distance is typically 5 times the sum of the beam widths (in a Gaussian approximation)

[2] The Tevatron comprises about 1000 superconducting magnets including 772 dipoles. Each dipole is approximately 6 m in length and 4 tons in weight. The superconducting coils are made up of niobium-titanium wires embedded in copper. A 4400 A current in the dipoles provides a 4.2 T magnetic field. All superconducting magnets are kept at 4 K temperature

[3] The interaction region is approximately Gaussian in z with $\sigma \sim 28$ cm, as determined by the overlap of the two approximately longitudinally Gaussian bunches.

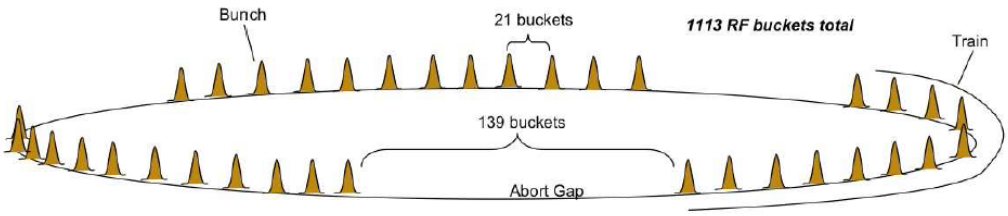


Figure 2.3: *Bunch structure of the Tevatron beams in Run II*

The transverse beam profile is shaped by a number of collimators to avoid detector damages from the tails of the proton or anti-protons interacting with the beam pipe (e.g: beam halo). When the beam profile is narrow and the condition are stable, the detectors are powered and data taking can start.

2.4 Performance

Performance of the Tevatron is qualified primarily by two parameters:

- center of mass energy ($\sqrt{s}$);
- integrated luminosity ($\int \mathcal{L}$dt) in a given time period.

The center of mass energy determines the reach in search for new beyond-the-standard-model phenomena. The latter is directly proportional to the number of events (N) of a given process with a cross section σ

$$N[events] = \int \mathcal{L}dt[cm^{-2}] \times \sigma[cm^2] \tag{2.1}$$

In the absence of a crossing angle or position offset, the luminosity in the Tevatron is given by the expression below ([48]):

$$\mathcal{L} = \frac{fBN_pN_{\overline{p}}}{2\pi(\sigma_p^2 + \sigma_{\overline{p}}^2)} F(\sigma_l/\beta^*) \tag{2.2}$$

where f is the revolution frequency, B is the number of bunches in each beam, N_p ($N_{\overline{p}}$) is the number of protons (anti-protons) in a bunch, σ_p ($\sigma_{\overline{p}}$) is the rms proton (anti-proton) beam size at the interaction point, and F is a form factor that depends on the ratio of the bunch length, σ_l, to the beta function at the interaction point, β^*. The values of the above parameters are reported in [48]. It can be shown that fundamental limitations on the Tevatron luminosity are due to N_p/ϵ_{N_p} and $BN_{\overline{p}}$,

ϵ_{N_p} being the normalized transverse emittance containing 95% of the proton/anti-proton beam ([48]). Higher luminosities were achieved over time thanks to the anti-proton stack rate.

Fig. 2.4, 2.5 show respectively the instantaneous luminosity at the beginning of each data taking period [4], and the integrated luminosity ($\mathcal{L}dt$), as a function of time: the constant progress in the performances of the machine leads to increased instantaneous and integrated luminosities. Blank parts in the figures correspond to periods of time when the Tevatron was not running. In total Tevatron delivered $\sim$ 12 fb^{-1} to the CDF and DØ experiments. CDF acquired about 85% of it ($\sim$ 10 fb^{-1}), because of inefficiencies in the detector, and dead time due to a number of reasons related to operations (e.g: detector calibrations, etc.).

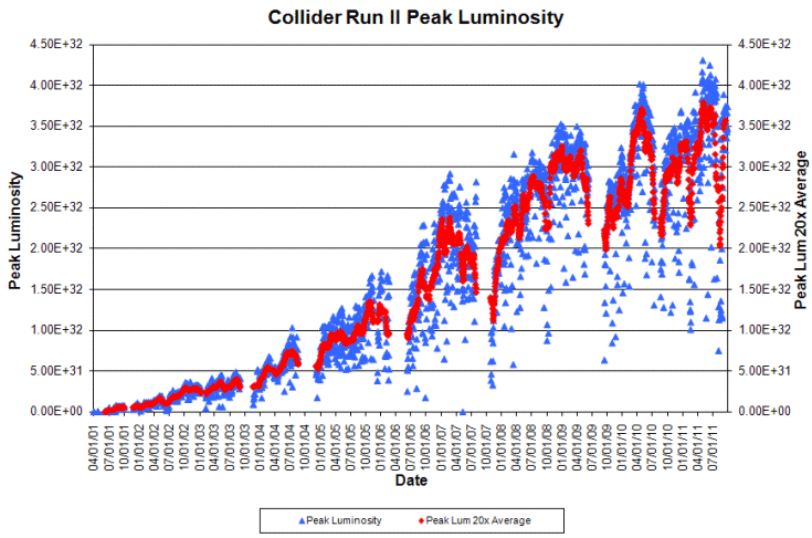


Figure 2.4: *The peak instantaneous luminosity during each store (blue triangles) and as average of 20 subsequent stores (red diamonds) over time in the Run II of the Tevatron.*

[4] A continuous period of collider operation using the same collection of protons and anti-protons is named store

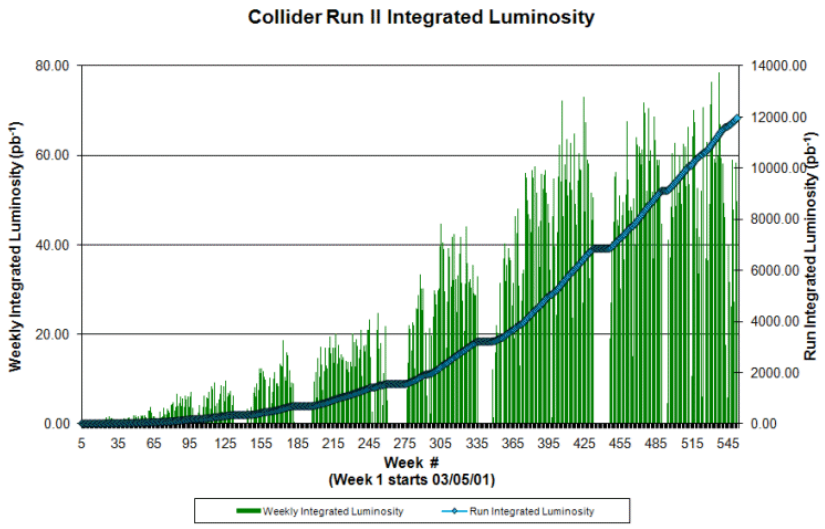


Figure 2.5: *The luminosity integrated during each week (green bars) and in total (cyan diamonds) during the Run II of the Tevatron.*

Chapter 3

The Collider Detector at the Fermilab Tevatron

The Collider Detector at Fermilab (CDF) was designed to study $p\overline{p}$ collisions at the Tevatron. Commissioned in 1985 it was upgraded in 1989 and again in 2001 in order to operate at the expected increased instantaneous luminosity of the accelerator. Each upgrade is considered a "Run". Run 0 was the run before any upgrades, Run I was after the first upgrade and Run II was after the second upgrade. We describe in this section the upgraded detector, CDF II. Further details are available at [49].

3.1 Coordinates

A right-handed Cartesian coordinate system with origin in the B0 nominal interaction point is used for the CDF II detector. The positive z-axis is parallel to the nominal beam line and points toward the proton direction (east). The y-axis points vertically upward, while the x-axis lies in the same plane as the Tevatron and points radially outward with respect to the center of beam (see Fig. 3.2). Due to the geometry of the infrastructure (CDF II detector plus the beams from the Tevatron) there is an obvious cylindrical symmetry around the beam line axis. We use a cylindrical coordinate system ($r \equiv \sqrt{x^2 + y^2}$, $\phi \equiv tan^{-1}x/y$, z) to locate a point on the CDF detector. In this coordinate system the z direction and a direction in the (r, ϕ) plane we will labeled as "longitudinal" and "transverse" respectively. Sometimes when describing the particle trajectory, rather than using z, it is convenient to use the polar angle ϑ. The polar angle is defined relative to the z-axis, the azimuthal angle to the x-axis.

Since hadrons are composite particles, hard interactions at the Tevatron happen with an unknown center of mass energy. Therefore, the overall longitudinal momentum in the initial state is unknown on a event-by-event basis. In the transverse plane the interacting partons are almost at rest since the beams are collimated along the z direction. For this reason it is convenient to use variables which are invariant under boosts along the longitudinal direction. Therefore, rather than using the polar angle $\vartheta \equiv tan^{-1}r/z$, we introduce the *pseudo-rapidity*

$$\eta = -ln(tan(\vartheta/2)) \tag{3.1}$$

which is the approximate expression of the rapidity in the ultra-relativistic limit. The rapidity is defined as follows:

$$y = \frac{1}{2}ln(\frac{E + pcos(\vartheta)}{E - pcos(\vartheta)}) \tag{3.2}$$

$$\tag{3.3}$$

where E and p are respectively energy and momentum of the considered particle. It can be shown that under a boost to an inertial frame with velocity β_z along the z direction $y \rightarrow y^{'} = y + tanh^{-1}(\beta_z)$, therefore $dy = dy^{'}$, meaning that the rapidity (and thus the pseudo-rapidity) difference between two physical systems is invariant under a boost along the longitudinal direction.

Since the longitudinal position of event vertex is distributed around the CDF geometrical center with a $\sim$ 30 cm r.m.s width, it is useful to distinguish the detector pseudo-rapidity, η_{det}, measured with respect to the geometrical center, from the actual pseudo-rapidity, η, measured with respect to the real interaction point.

Other common variables which are invariant under a boost along the longitudinal direction are the following:

$$\begin{aligned} P_T &= Psin\vartheta \\ E_T &= Esin\vartheta \end{aligned} \tag{3.4}$$

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.5}$$

P_T and E_T are respectively the transverse momentum and transverse energy. ΔR can be thought as the separation between two particles in the (η, ϕ) plane.

3.2 Detector Overview

The Run II Detector (see Fig. 3.1, fig. 3.2) is composed of several components, each optimized for a specific task.

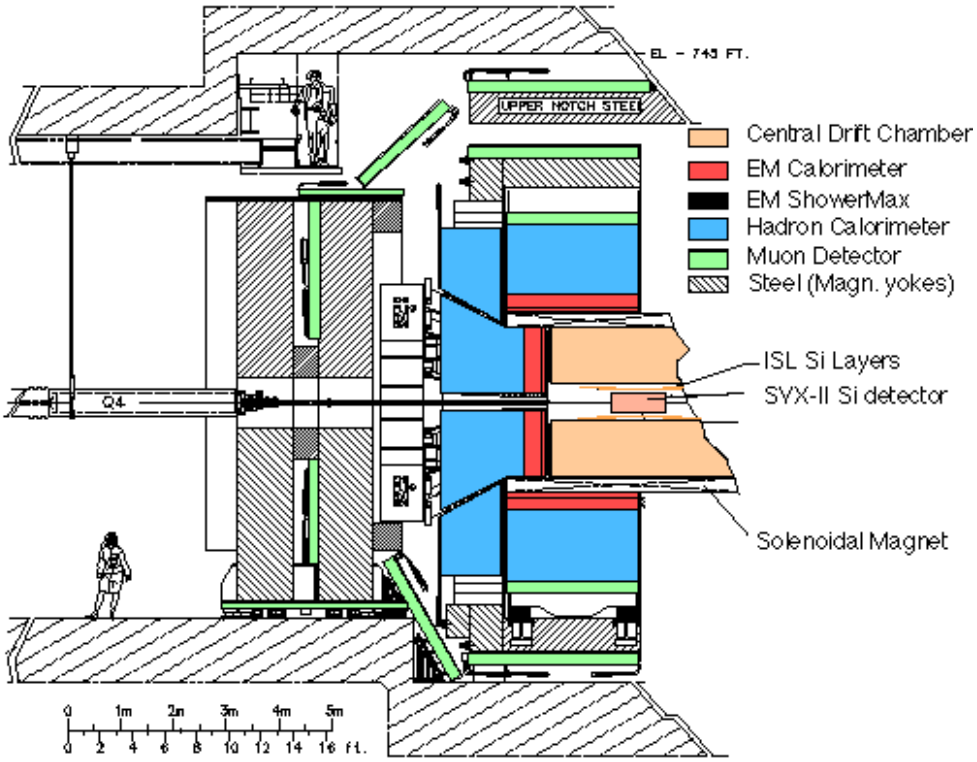


Figure 3.1: *Elevation view of the CDF Run II detector.*

Starting from the interaction point and following the path of an outgoing particle within acceptance there are:

- a tracking system enclosed by a superconducting solenoid (1.5 m in radius and 4.8 m in length), which generates 1.4 T magnetic field parallel to the beam axis. The magnetic field is nearly uniform within the tracking volume.
- Electromagnetic and hadronic plastic scintillator calorimeters split into projective towers.
- planar drift chambers backed by scintillation counters (muon detectors).

In the next sections a number of CDF II sub-systems will be discussed. Some of the components (the time-of-flight detector, etc.) of the detector have been

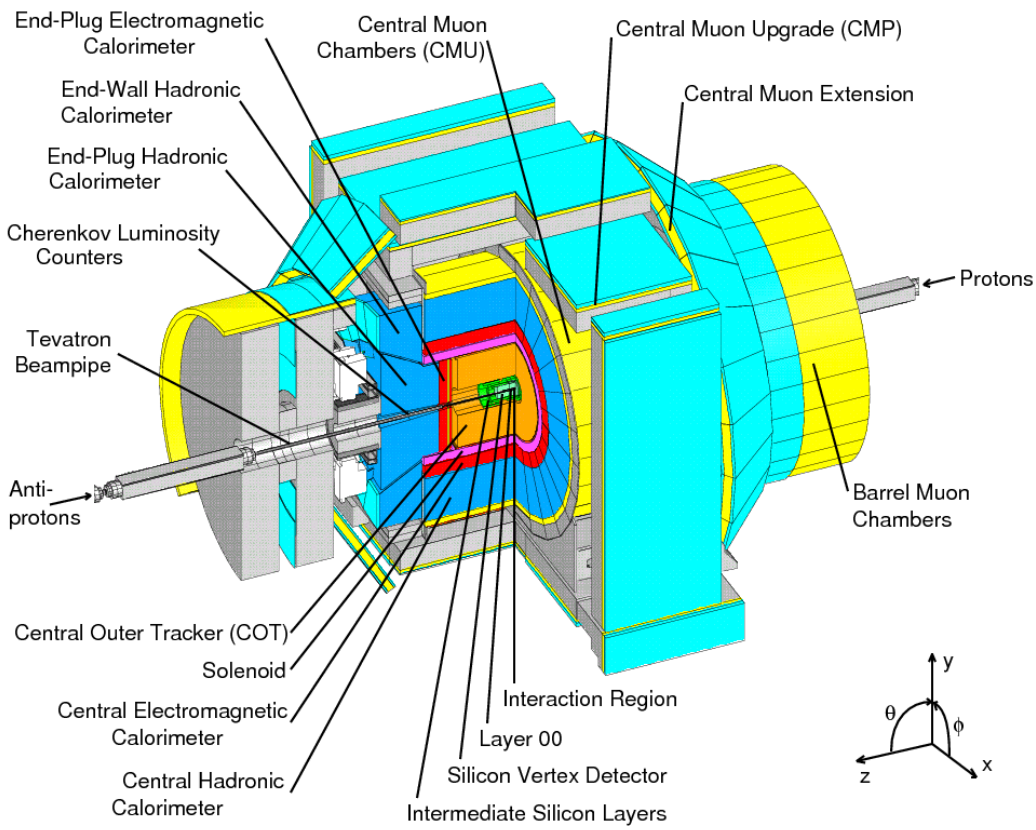


Figure 3.2: *Artistic view of the CDF Run II detector.*

neglected since they are not directly related with the topic of this thesis. A detailed description of the upgraded detector can be found in [49].

3.3 Tracking System

Direction and momentum of charged particles can be measured through a tracking system consisting of three silicon sub-detectors and a large outer drift-chamber (Fig. 3.3). These sub-detectors are contained in a superconducting solenoid, which creates a magnetic field of about 1.4 T over a 1.4 m radial distance. The coil is 0.85 electron radiation lengths thick [1].

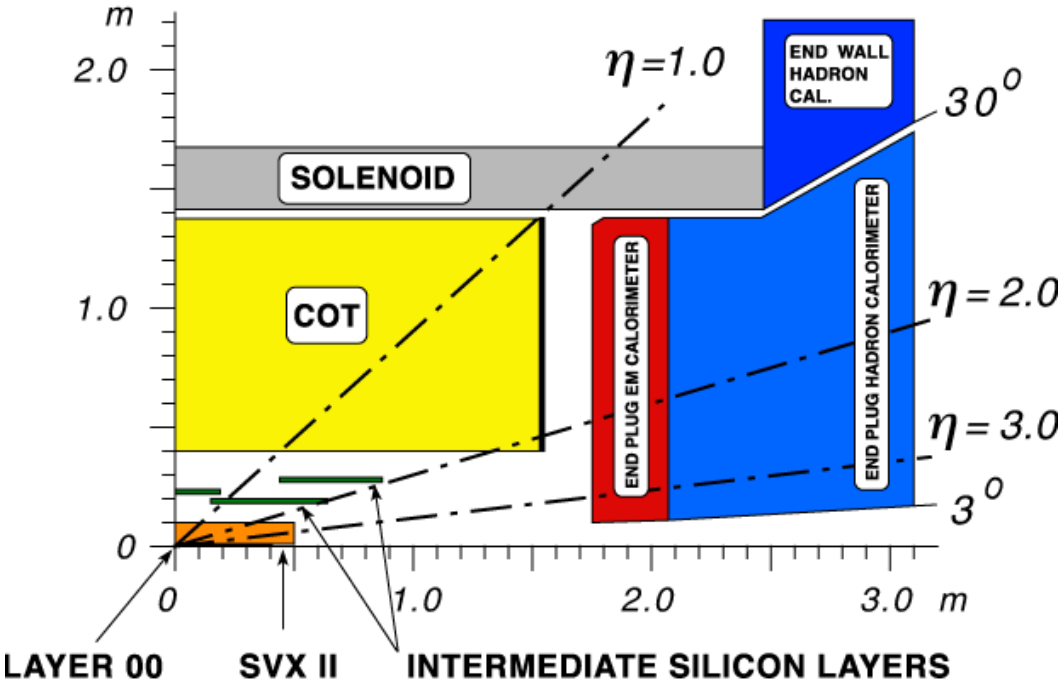


Figure 3.3: *Cut-away view along the beam of one quadrant of the Tracking System of the CDF II detector.*

The tracker is a two component system, comprising a silicon detector at small radii, and a large open-cell drift chamber at larger radii. Although this system is similar to the one used in Run I, it was upgraded in order to overcome some previous limitations (e.g: length not enough to cover the luminous region, non-optimal num-

[1] High-energy electrons predominantly lose energy in matter by bremsstrahlung. The mean distance per unity of density material over which an electron loose all its energy but $1/e$ is called radiation length X_0. X_0 and can be approximated by $\frac{716.4 \cdot A}{Z(Z+1)ln(287/\sqrt{Z})}[g \cdot cm^{-2}]$, where A, Z are the mass and atomic numbers of the nuclei composing the material traversed by the electron.

ber of layers, no stereo or weak stereo layers). It was to some extent over-designed in order to face the expected deterioration with the accumulated radiation dose at higher luminosities. The inner detector provides excellent impact parameter, and z resolutions, while the outer detector provides excellent resolution on the curvature, and pseudo-rapidity. Together they provide an accurate measurement of the azimuthal angle. The two components of the tracking system are described below.

3.3.1 Inner Tracker

With an inner tracker made of silicon, CDF has established the viability and excellent performance of silicon tracking in hadron colliders. The inner-most tracker is also called "vertex detector" in CDF jargon, since its primary role is to provide precise tracking information near the interaction in order to identify displaced tracks, and reconstruct displaced decay vertices produced by long life-time particles [2].

The Inner Tracker is composed of eight layers (seven at $\vartheta = 90°$) of silicon sensors arranged in approximately cylindrical sub-systems coaxial with the beam-pipe: Layer 00 (L00), the Silicon Vertex Detector (SVXII), and the Intermediate Silicon Layers (ISL). Fig. 3.4 zooms on the Inner Tracker from two different point of views. All silicon microstrip sensors have a space resolution of 12 μm in the

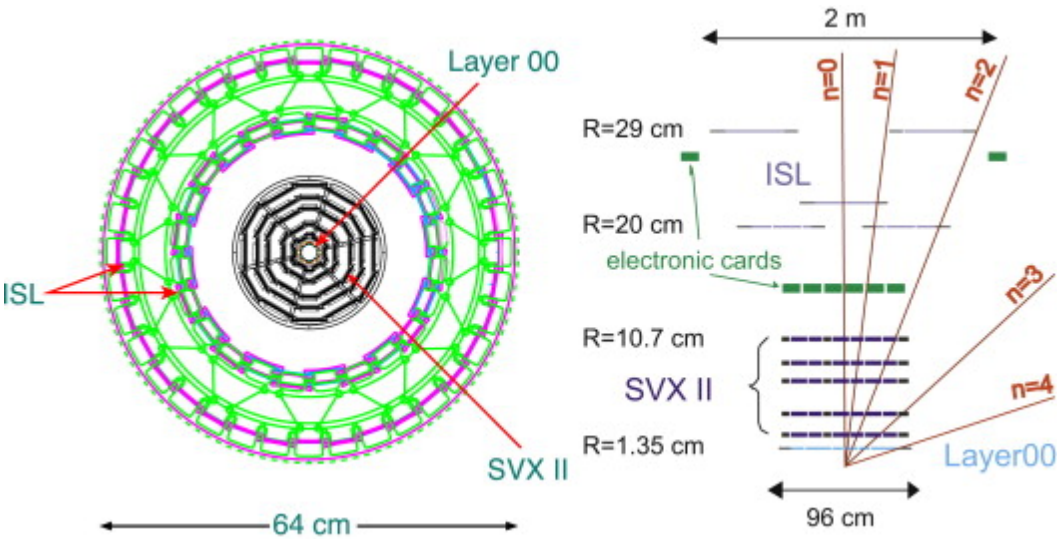


Figure 3.4: *Left: cutaway transverse to the beam of the three inner tracker subsystems. Right: sketch of the silicon detector in a x/y projection.*

[2]The displaced track identification played a major role in the discovery of the top quark by CDF ([50]).

direction transverse to the beam. SVXII and ISL also provide z-measurements with reduced accuracy.

- **Layer 00** ([51])

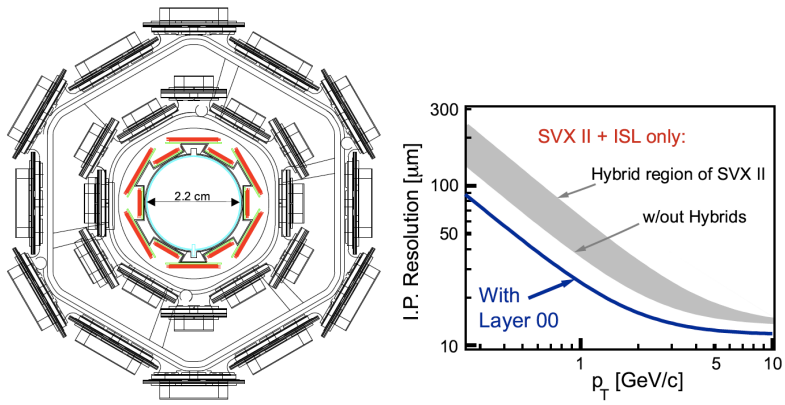
  Layer 00 (L00) is an innovative detector made of a single-sided castellated layer built directly onto the beam pipe. The innermost (128 strip) and outermost (256 strip) sensor layers are located at radii of 1.35 cm and 1.62 cm (Fig. 3.5a). The strips are parallel to the beam axis providing samples of the tracks in the (r, ϕ) plane. The implant (readout) pitch is 25 (50) μm. There are 12 sensors along the beam line for a total length of 94 cm. L00 provides full azimuthal coverage and $|z| \lesssim 47$ cm longitudinal coverage. The front-end electronics is located outside the tracking volume to minimize the multiple scattering from inactive material.

  L00 was added in 2001 to the inner tracking system for two reasons.

  - Extend the lifetime of the silicon system: the inner layer of SVXII were expected to have a limited lifetime because of radiation damage. L00 uses radiation hard silicon, which by standing a significantly higher bias voltage can operate at a higher absorbed radiation dose, thus compensating for the SVXII damaged layers.
  - Further improve the impact resolution of the tracking system (Fig. 3.5b): SVXII readout electronics is located inside the tracking volume, thus degrading the impact resolution because of multiple scattering. This effect is more enhanced for low-momentum particles. Having a minimal silicon material at smaller radii provides a precision measurement on the impact parameter which helps recovering the lost resolution.

  During data taking significant noise was observed in L00. Such a noise was characterized by non-uniform pedestals across the strip sensors. The noise was varying event-by-event. It was decided to employ an offline event-by-event pedestal subtraction. Such a procedure was tested in Monte Carlo $p\overline{p}$ and it was found a 95% efficiency with a 95% purity. However, because of the needed pedestal subtraction and of the slow readout (∼30 KHz), L00 could not be used in the online triggers.

- **Silicon Vertex Detector** [49], [52]



(a) (b)

Figure 3.5: *End-view of the layer 00 of the CDF inner detector (left). This layer is mounted directly on the beam pipe (cyan). The expected impact parameter resolution of the reconstructed track with and without the layer 00 is also shown as a function of the track P_T (right).*

The silicon vertex detector (SVXII) is built in three cylindrical barrels (Fig. 3.6) with a total length of 96 cm, providing coverage over $|\eta_{det}| \leq 2.0$ from the detector center. Each barrel supports five layers of double-sided microstrip detectors at radii between 2.4 cm to 10.7 cm. SVXII has a cylindrical geometry coaxial with the beam: twelve 30° azimuthal sectors ("wedges") provide full coverage over ϕ. A small overlap between adjacent wedges is present (Fig. 3.6).

All five layers provide $r - \phi$ measurement on one side. Three out of the five layers combine such a measurement with a 90° stereo measurement on the reverse side. The remaining two layers provide a small angle stereo at 1.2° on the reverse side. All layers are arranged into twelve concentric independent readout units ("ladders"). A total of 405,504 channels are used for SVXII. Water, cooling the channels, is carried between the ladders.

This sub-detector has a 12 μm resolution on the single hit in the direction transverse to the beam, and provides also some dE/dx information.

With respect to the progenitor SVX in Run I [53], SVXII features an increased length along the z direction. The increase length allows covering 2.5 σ of the luminous region, thus increasing the geometrical and angular acceptances for single tracks. This improvement was designed to enhance the efficiency in the reconstruction of displaced tracks from heavy meson decays.

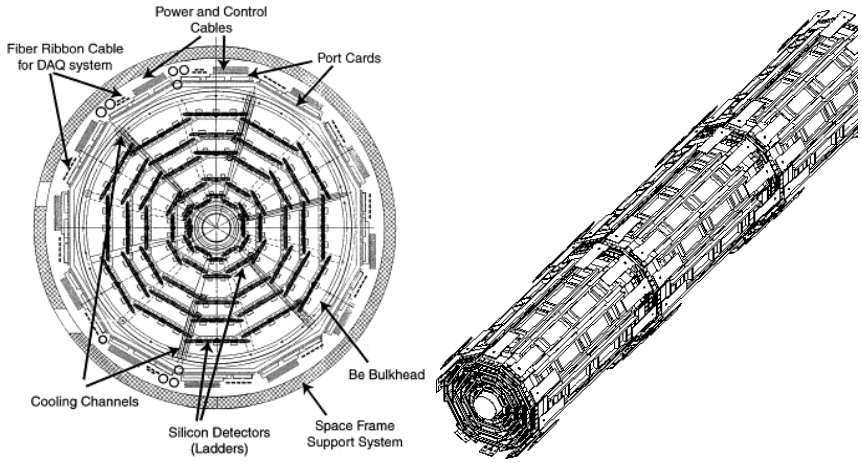


Figure 3.6: *End-view (left) and isometric (right) of the CDF Silicon Vertex Detector.*

- **ISL** [54]

  The main ISL purpose is to compensate for incomplete coverage of other sub-detectors in the region $|\eta_{det}| > 1$ by providing precision tracking at $1 < |\eta_{det}| < 2$. Together with SVXII ISL provides 3D tracking information. ISL is placed between SVXII and the drift chamber (Fig. 3.4) at radii 20-28 cm. ISL is composed of five barrels in total, one central and of two inner and outer barrels in the backward/forward regions. Each barrel has twelve 30° azimuthal sectors to match the SVXII segmentation.

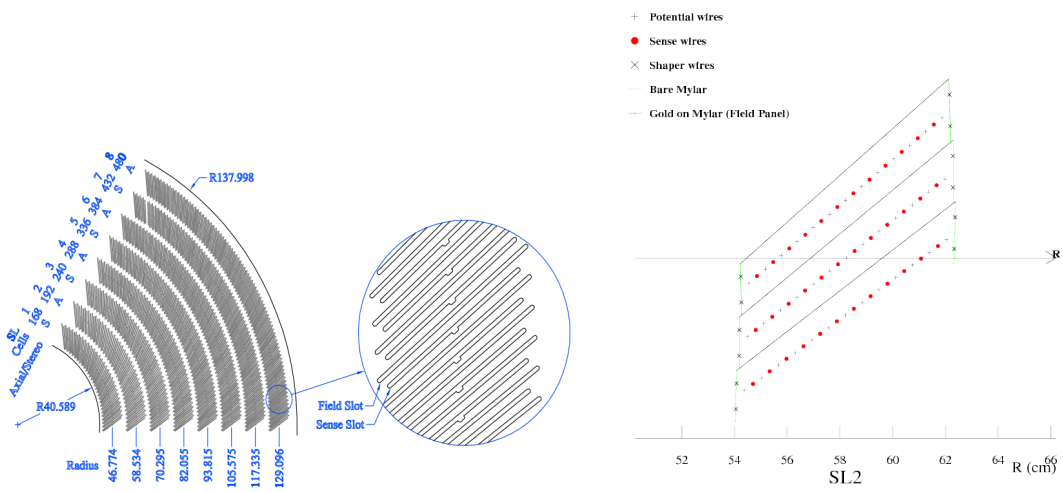
  The basic ISL readout is a "half ladder" module. Such a module is made of three silicon sensors ganged together forming a single electrical unit. Sensors are, as in the case of SVX II, double-sided AC coupled microstrip detectors. Each sensor has 512 readout strips along the beamline of 112 μm pitch on one side, and 1.2°-tilted 146 μm pitch on the other side. The total number of channels is 268,800.

The total amount of material in the silicon system (averaged over azimuthal angle and z) is roughly $10\%/sin\vartheta$ times the electron radiation length. Therefore in the forward/backward region the average material traversed by particles increases by roughly twice with respect the central region.

The combined resolution of the CDF inner trackers for high momentum tracks is 40 μm in impact parameter and 70 μm along the z direction [55].

3.3.2 Central Outer Tracker

The main tracker at CDF II is the Central Outer Tracker (COT). The COT has a cylindrical shape and is radially right outside the ISL. Its active volume spans from radii ∼44 to ∼132 *cm* and $z \lesssim 155$ cm. COT provides full tracking in the central region ($|\eta_{det}| < 1$). With reduced acceptance, its coverage extends to $|\eta_{det}| < 2$ (see Fig. 3.3). Radially, the COT is arranged into 4 axial and 4 stereo superlayers, containing 96 planes of wires (see Fig. 3.7a). Each superlayer is composed of azimuthal cells. Each cell has alternated sense and field shaping wires (see Fig. 3.7b). The latter control the gain on the sense wires optimizing the electric field intensity. Axial superlayers employ sense-wires parallel to the beam axis, while stereo superlayers have the wires alternatively tilted at $\pm 2°$ with respect to the beam-line. Axial superlayers provide measurements of the hit coordinate in the (r, ϕ) plane, while the stereo superlayer also measure the hit coordinate along the z axis.



(a) (b)

Figure 3.7: *Slots housing the wire-holding fixtures of a 1/6 section of the COT endplate (left). For each super-layer the total number of cells, the wire orientation (stereo r axial), and the average radius [cm] is given. Drift and field wires in three cells (right). The horizontal arrow shows the radial direction.*

Within a cell the ionization released by charged tracks is sampled 12 times (every 0.583 cm) by sense wires. Inside the solenoid magnetic field, the drifting electrons experience a Lorentz force which rotates their path. The cells are tilted by 35°(Fig.

3.7b) with respect to the radial direction in order to make the electrons drifting perpendicularly to the radius for optimal momentum resolution. The single hit position has been measured with an uncertainty of 180 μm. This translates into an overall resolution $\sigma(P_T)/P_T = 0.17\%\ P_T\ [GeV/c]$ [56], P_T being the transverse momentum of the tracked particle.

The main parameters of the tracking system are summarized in table 3.1. Detailed studies on the inner detector performance and aging are described in [58].

COT	
Radial Coverage	44 to 132 cm
Number of superlayers (SLs)	8
Readout coordinates of SLs	+3° 0 -3° 0 +3° 0 -3 0°
Maximum drift distance	0.88 cm
Resolution per measurement	140 μm
Rapidity coverage	$\lvert\eta_{det} \leq 1.0\rvert$
Number of channels	30,240
Material thickness at 90°	1.3% X_0
SVX II	
Radial Coverage	2.4 to 10.7 cm, staggered quadrants
Number of layers	5
Readout coordinates	$r - \phi$ on one side of all layers
Stereo side	r-z, r-z, r-uv, r-z, r-uv (uv $\equiv$ 1.2° stereo)
Readout pitch	60-65 μm r-ϕ; 60-150 μm stereo
Resolution per measurement	12 μm (axial)
Total length	96.0 cm
Rapidity coverage	$\lvert\eta_{det}\rvert \leq 2.0$
Number of channels	405,504
Material thickness at 90°	3.5% X_0
ISL	
Radial Coverage	20 to 28 cm
Number of layers	one for $\lvert\eta_{det}\rvert$ <1; two for 1< $\lvert\eta_{det}\rvert$ <2
Readout coordinates	r-ϕ and r-uv (1.2° stereo) (all layers)
Stereo side	r-z, r-z, r-uv, r-z, r-uv (uv $\equiv$ 1.2° stereo)
Readout pitch	110 μm (axial); 146 μm (stereo)
Resolution per measurement	16 μm (axial)
Total length	174 cm
Rapidity coverage	$\lvert\eta_{det}\rvert \leq 1.9$
Number of channels	268,800
Material thickness at 90°	2% X_0

Table 3.1: Design parameters of the baseline tracking system (no L00) [49].

3.4 Calorimeters

The CDF calorimeter measures the particle energy by absorbing their total energy and providing a signal proportional to it. Calorimeter information is also used to estimate the transverse energy of weakly interacting particles, such as neutrinos, by computing the imbalance in the total transverse momentum. CDF uses scintillator sampling calorimeters divided into a front electromagnetic and a rear hadronic compartment. Both calorimeters are segmented into projective towers. Each tower consists of alternating layers of passive absorber material (lead in the front and iron in the rear compartment) and plastic scintillator for shower sampling. The light from the scintillator plates is read out through wavelength-shifting bars or plates and light guides by photo-multiplier tubes (see Fig. 3.8).

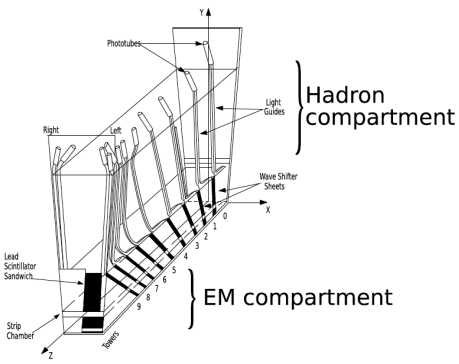


Figure 3.8: *Light-shifter plates connected to light guides and to photomultipliers of the front electromagnetic compartment of a central calorimeter wedge.*

High energy electrons and photons generate an electromagnetic shower, which is mostly absorbed in the front calorimeter compartment.

Hadrons interact with the detector matter mostly through inelastic collisions with nuclei of the absorbing medium. Particles produced in the nuclear interactions can loose their energy by ionization and secondary nuclear interactions. Mixed electromagnetic and hadron showers that originate in this process are largely absorbed in the entire (front + rear compartments) calorimeter. The energy-dependent rear leakage is of the order of a few %.

The coverage of the CDF calorimeter extends up to $|\eta_{det}| = 3.6$ and is complete in azimuth (see Fig.3.2). They include the Central Electromagnetic Calorimeter (CEM) and the Hadronic Calorimeter (CHA) in the $|\eta_{det}| < 0.9$ region, the Endwall Hadronic Calorimeter (WHA) at $0.9 < |\eta_{det}| < 1.3$ and the electromagnetic and hadronic plug calorimeters (PEM, PHA) at $1.1 < |\eta_{det}| < 3.6$ (see Fig. 3.1, 3.3).

Central Calorimeters

The central calorimeters, CEM, CHA and WHA are composed of two parts joining on the median plane of the detector at $z = 0$ [3]. Central calorimeters are azimuthally divided into 24 wedges, covering 15° in ϕ each. Each wedge is divided into projective towers of size $\delta\eta_{det} = 0.11$.

The **CEM** calorimeter ([59]) is made of 31 alternate layers of 0.5 cm thick plastic scintillator plates and 0.32 cm thick lead absorbers: the total amount of material is $18 \cdot X_0$. The CEM energy resolution is [4]:

$$\sigma_{E_T}/E_T = \frac{13.5\%}{\sqrt{E_T[GeV]}} \oplus 1.5\% \tag{3.6}$$

the stochastic factor 13.5 is determined primarily by the sampling structure of the calorimeter.

CEM also includes two additional specialized detector: the Central Electron Strip Chambers (CES) and the Central Preshower (CPR, [60]). CES is a combined strip/wire gas proportional chamber embedded in CEM at about $6 \cdot X_0$ [5]. The CES purpose is to measure the position and the shape of electro-magnetic showers in the transverse plane. CES resolution is about 1 cm in z and 1 mm in $r - \phi$. CPR is a set of scintillator tiles located in front of the calorimeter wedges which help distinguishing electrons from charged hadrons because of their probability of showering in the detector material prior to entering the calorimeter.

The **CHA** calorimeter ([61]), surrounding the CEM, is composed of 32 alternate layers of 1 cm thick plastic scintillator and 2.5 cm thick steel. The **WHA** calorimeter employs the same technology as CHA, except for the smaller number of layers (15) and the larger thickness of the radiator plates (5 cm). The total calorimeter thickness at normal incidence is 4.7 λ_0 (λ_0 is the charged pion absorption length) for both CHA and WHA.

Resolutions of CHA and WHA for perpendicular particle entrance are approximately:

[3] In this zone, $\eta_{det} = 0$, there is an un-instrumented area about 20 cm thick in the z-direction ("crack")

[4] The energy resolution is expressed as a function of E_T rather than E to account for the larger amount of absorber material encountered at higher ϑ.

[5] The maximum energy density in the longitudinal development of the electromagnetic shower is expected at about $6 \cdot X_0$

$$CHA : \sigma_{E_T}/E_T = \frac{50.0\%}{\sqrt{E_T[GeV]}} \oplus 3.0\% \tag{3.7}$$

$$WHA : \sigma_{E_T}/E_T = \frac{75.0\%}{\sqrt{E_T[GeV]}} \oplus 4.0\% \tag{3.8}$$

Plug Calorimeter

The **PEM** calorimeters (see Fig. 3.9) have the same tower segmentation as CEM in η_{det}, and ϕ for $2.11 < |\eta_{det}| < 3.6$ (15° wide ϕ bins). The segmentation in ϕ for $|\eta_{det}| < 2.11$ is finer than CEM (7.5° wide ϕ bins). PEM is composed of 22 layers of 4.5 mm thick lead alternate with 22 layers of 4 mm thick scintillator ([62]). The total thickness is about 21 X_0. The PEM energy resolution is:

$$\sigma_{E_T}/E_T = \frac{16.0\%}{\sqrt{E_T[GeV]}} \oplus 1.0\% \tag{3.9}$$

As for CEM, PEM is equipped with a shower maximum detector (PES). PES, as well as CES, measures the shower profile to distinguish electrons from charged hadrons. PES consists of two layers of 200 scintillating bars each. Bars are oriented at crossed relative angles of 45°. The position of the shower is measured with an accuracy of about 1 mm.. Further details are described in [63].

PHA, surrounding PEM, has its same tower segmentation. The technology is the same as for CHA, with 23 layers of 2 cm thick steel absorber alternating with 6 mm thick scintillator. The total amount of material corresponds to 4.7 λ_0. PHA resolution is:

$$\sigma_{E_T}/E_T = \frac{80.0\%}{\sqrt{E_T[GeV]}} \oplus 5.0\% \tag{3.10}$$

3.5 Muon System

Although muons interact mostly electromagnetically as the electrons, because of their much larger mass they can cross a much larger amounts of material before losing a significant fraction of their energy [6]. For this reason systems dedicated to detect muons are located in the outermost shell of the detector. Muon momenta are measured in the tracker.

[6] At Tevatron energies muons interact in calorimeters as minimum ionizing particles (MIP).

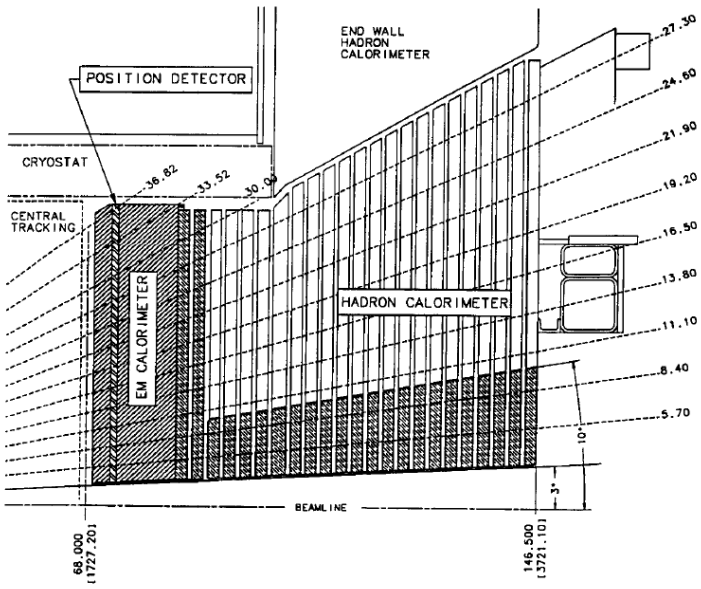


Figure 3.9: *Longitudinal view of Plug Calorimeters.*

Four independent systems are used to detect muons in the $|\eta_{det}| < 1.5$ region: the Central Muon Detectors (CMU), the Central Muon Upgrade Detectors (CMP), the Central Muon Extension (CMX), the Intermediate Muon Detectors (IMU). The (η_{det}, ϕ) coverage of the largest Run II muon detectors is shown in figure 3.10. Muon detectors share common features ([64]). They consist of stacks of rectangular drift chamber modules [7], composed of single-wire cells. Stacks are four layers deep with laterally displaced cells from layer to layer to compensate for cell edge inefficiencies. The difference in drift-electrons arrival-times between neighbor cells provides a typical resolution of 250 μm for the hit position in the transverse plane. Charge division at the wire ends measures the coordinate along the wire with a 1.2 mm resolution. Chambers are coupled with scintillator counters in order to suppress backgrounds due to secondary interactions in the beam pipe material and to cosmic rays. A muon candidate is reconstructed when a short track segment (stub) in the muon chambers corresponds to the extrapolation of a COT track.

The **CMU** detector is behind CHA at a radius of 347 cm from the beam axis and covers the $|\eta_{det}| < 0.6$ region. CMU consists of 144 modules with 16 cells each. The CMU box is arranged in 12.6° wedges. Cells composing CMU are 266 cm long, 2.68 cm thick, and 6.35 wide with a single 50 μm steel wire at their center (Fig. 3.11) parallel to the beam direction. The system is filled with Argone-Ethane gas

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

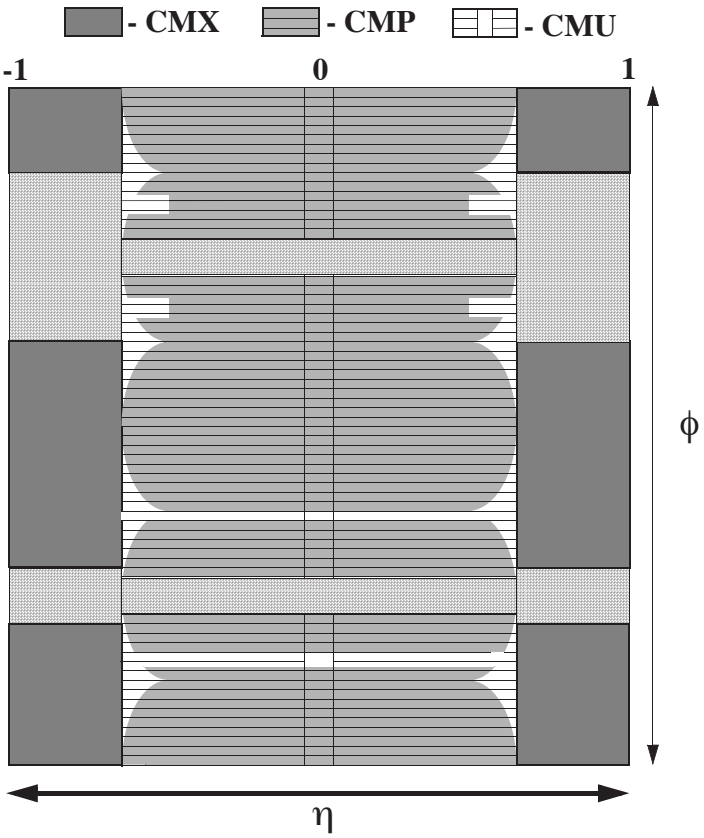


Figure 3.10: *Coverage of muon detectors in the* (η, ϕ) *space.* η *is computed with respect to the CDF center.* ϕ *ranges from -*π *to* π *radians.*

mixture and alcohol like the COT.

The **CMP** detector is arranged to enclose the $|\eta_{det}| < 0.6$ region in an approximately central box (see Fig. 3.10). Scintillator layers (CSP) on the outermost side of the CMP chambers allow identifying bunch crossing. The CMU/CMP system is called CMUP. It detects muons with a minimum energy of about 1.4 GeV.

The **CMX** detector extends the muon identification in the $0.6 < |\eta_{det}| < 1$ region. As for CMP cells are sandwiched to scintillators (CSX).

The forward region of muon system is the **IMU** detector ($1.0 < |\eta_{det}| < 1.5$). The associated scintillator counters sub-system is named BSU.

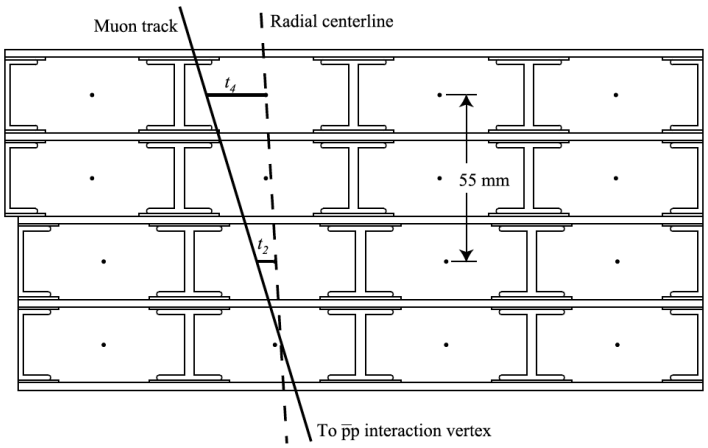


Figure 3.11: *Cross section of CMU single wire cells.*

3.6 Cherenkov Luminosity Counters

The Tevatron collider luminosity is estimated by measuring the known interaction rate of inelastic $p\bar{p}$ events. A detector consisting of two CLC ("Cherenkov luminosity counters") is used to do so ([65]). The two modules, placed inside the CDF end-plug calorimeters in the forward/backward region at $3.7 < |\eta_{det}| < 4.7$, consist of 48 thin, long, gas-filled Cherenkov counters. These counters are grouped in three concentric conical layers (16 counters each) around the beam pipe. Counters point to the center of the interaction region (See fig. 3.12). Counters in the outer two layers are 180 cm long, while the counters in the inner layer are 110 cm long. Counter transverse dimensions range between 2 and 6 cm. This geometry ensures that only charged particles originating from around the nominal interaction point give a full-height Cherenkov signal, and allow distinguishing them from prongs of beam-gas events

and from stray machine background.

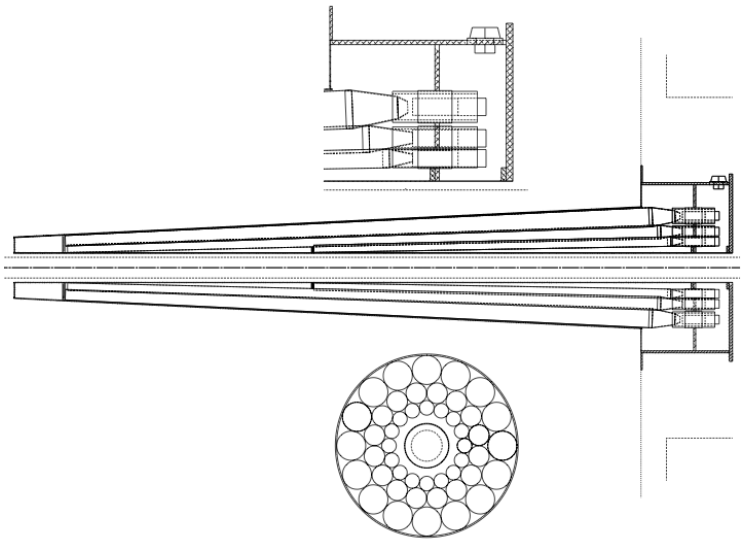


Figure 3.12: *The CDF Cherenkov Luminosity Counters Design.*

The Cherenkov light is collected by conical mirrors at the rear end of the cones and detected by high gain (2×10^6) photomultipliers. The counters are mounted on long aluminum tubes surrounding the beam pipe. Everything is plunged in Isobuthane and kept at the atmospheric pressure. In this condition the Isobuthane refraction index is relatively high ($n = 1.0043$) and the gas is transparent to UV radiation. Around the tube a magnetic field shield absorber is placed in order to reduce the fringe CDF solenoid field. This is done to allow photomultipliers to operate at constant gain.

Prompt particles from the $p\overline{p}$ primary interaction will traverse the full counter length and will generate a large amount of Cherenkov photons. Secondary background particles are mainly electrons or positrons from electromagnetic showers initiated by $\pi^0 \to \gamma\gamma$ decays in the beam pipe or detector material [8]. They may contribute to low-amplitude signals (when they traverse the counters at large angles and with shorter path lengths) or anomalous high-amplitude signals (several particles hit the same counter). Therefore the signal from the secondary particles can be discriminated by setting pulse-height cuts on-line and off-line. The counter time resolution is excellent (< 100 ps) and helps checking the multiple interaction rate as a function of instantaneous luminosity, thus reducing the systematic uncertainties.

[8] Beam halo particles often fall under the Cherenkov threshold, thus producing no light.

In order to measure the luminosity we use the formula below:

$$\mathcal{L} = \frac{R_{p\overline{p}}}{\epsilon_{CLC}\sigma_{in}} \tag{3.11}$$

where $R_{p\overline{p}}$ is the rate of the inelastic $p\overline{p}$ events measured with CLC, ϵ_{CLC} is the the CLC acceptance, and σ_{in} is the inelastic cross $p\overline{p}$ section [9].

The CLC acceptance is measured from simulation: $\epsilon_{CLC} = 60.2 \pm 2.6\%$. The inelastic cross section is obtained by extrapolating the CDF, E811 combined measurements at $\sqrt{s} = 1.8$ TeV ($\sigma_{in} = 59.3 \pm 2.3$) [66]. The cross-section extrapolated at 1.96 TeV (assuming a $ln^2 s$ dependence) is 60.7 ± 2.4 mb.

The uncertainty on the measured luminosity is ∼6% and is largely dominated by systematic uncertainties. Main contributors are uncertainties on the detector stability and calibration (≤ 2.5 %), ϵ_{CLC} (4.4%), σ_{in} (4.0%). Further details about the luminosity calculations are described at [56].

3.7 Trigger and Data Acquisition

At Tevatron Run II the interaction rate is typically 2.5 MHz (the bunch crossing frequency), which is much higher than any possible event recording rate (order of 100 Hz). However, $p\overline{p}$ interactions are mostly inelastic, elastic and diffractive with no significant momentum transfer. More interesting events have cross-sections from 10^3 to 10^{12} times smaller than inclusive $p\overline{p}$ cross section (for example $\sigma(p\overline{p} \to WZ) \sim 4$ pb). The identification of the interesting events is accomplished by dedicated fast online electronics, called the Trigger System, which evaluates the information from the detector and makes an accept/reject decision in real time. The trigger system (see Fig. 3.13) is a three-tier system, where each higher level of electronics performs a slower but more accurate event reconstruction and applies a tighter filter with increasing trigger level according to a set of pre-defined conditions.

3.7.1 Level 1

The level 1 (L1) is a synchronous pipeline system, where up to 42 subsequent events can be stored for 5.5 μs while custom hardware is making a decision. If no acceptance decision is made up to that time the event is rejected. L1 decisions are made in about 4 μs average time: no dead time is expected from this level. L1 roughly accepts one out of 3×10^4 inelastic events and the typical output rate is 20 KHz.

[9] At large angles the rate of elastic events is negligible with respect to the one of inelastic events.

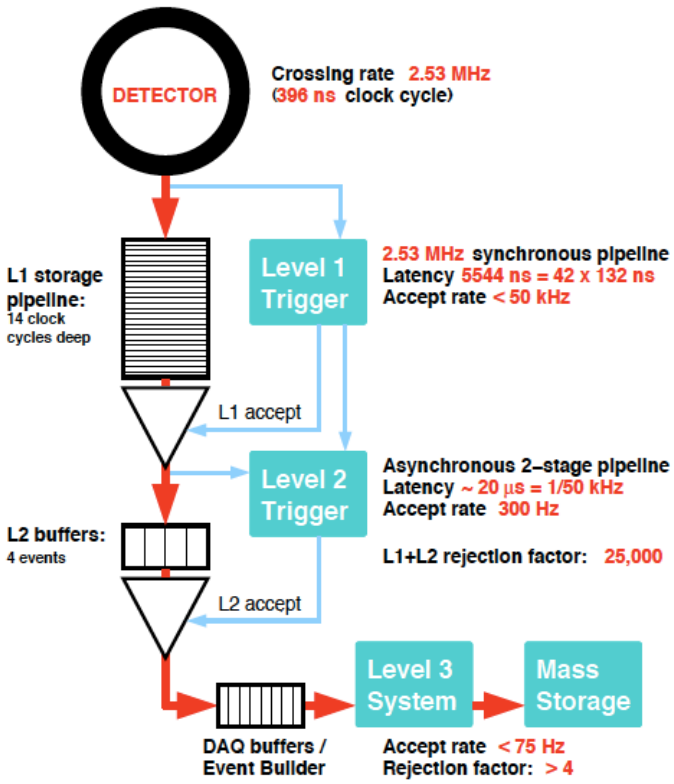


Figure 3.13: *CDF trigger block diagram.*

The L1 decision is generated based on the information below.

- Reconstruction of calorimeter objects (electrons, photons, jets) or calculation of global-event calorimetric observables (e.g: transfer momentum imbalance, total transverse scalar energy, etc.). At this level electron, photon and jet candidates are defined as single-tower ("seed") energy deposit. The nominal CDF geometrical center is used as coordinate origin.

- Track reconstruction [10]: the eXtremely Fast Tracker (XFT, [67]) hardware examines hits on the COT axial superlayers and combine them into track segments. The segments are then linked to form a track candidate with $P_T >$ 1.5 GeV/c.

- Muon identification: the observation of hits in the muon detector wire chamber and scintillator systems hits matched to a XFT track, extrapolated to the muon chamber, defines a muon. A loose P_T threshold is applied based on differences in signal arrival times on pairs of wires in the CMU and CMX chambers.

The final L1 decision is made based upon the number of reconstructed objects or upon the calculated observables mentioned above. Events accepted at L1 are stored in one of the four buffers in the front-end readout hardware.

3.7.2 Level 2

The level 2 (L2) trigger system makes use of dedicated hardware to select events. The two main pieces of dedicated hardware are the following:

- cluster finder: add the energy deposited in the towers neighboring the L1 seeds to form a cluster. The cluster energy is an approximate measure of an electron or jet energy

- silicon vertex tracking (SVT, [68]): use the information from the SVX II detector to improve the parameter (e.g: ϕ, P_T, d_0) resolutions of the XFT tracks. Hereby SVT allows to trigger on secondary vertexes from decay of long-lived beauty hadrons.

These two systems work in an asynchronous way since the processing time is highly correlated to the amount of data to be processed.

[10] This additional information in the trigger was introduced during the Run II upgrade.

The outputs of these systems is transferred to the global L2 processor, along with L1 data, and with additional information from the CES detector to help in low E_T electron selection. The processor decides whether the event is passed to the next step by exploiting simple selection algorithms. The processor board has been designed to read-in one event while processing another one in order to reduce the deadtime. The maximum L2 acceptance rate is ∼300 Hz (rejection factor ∼ 150). The block diagram of L1 and L2 with the involved sub-detectors is schematized in figure 3.14.

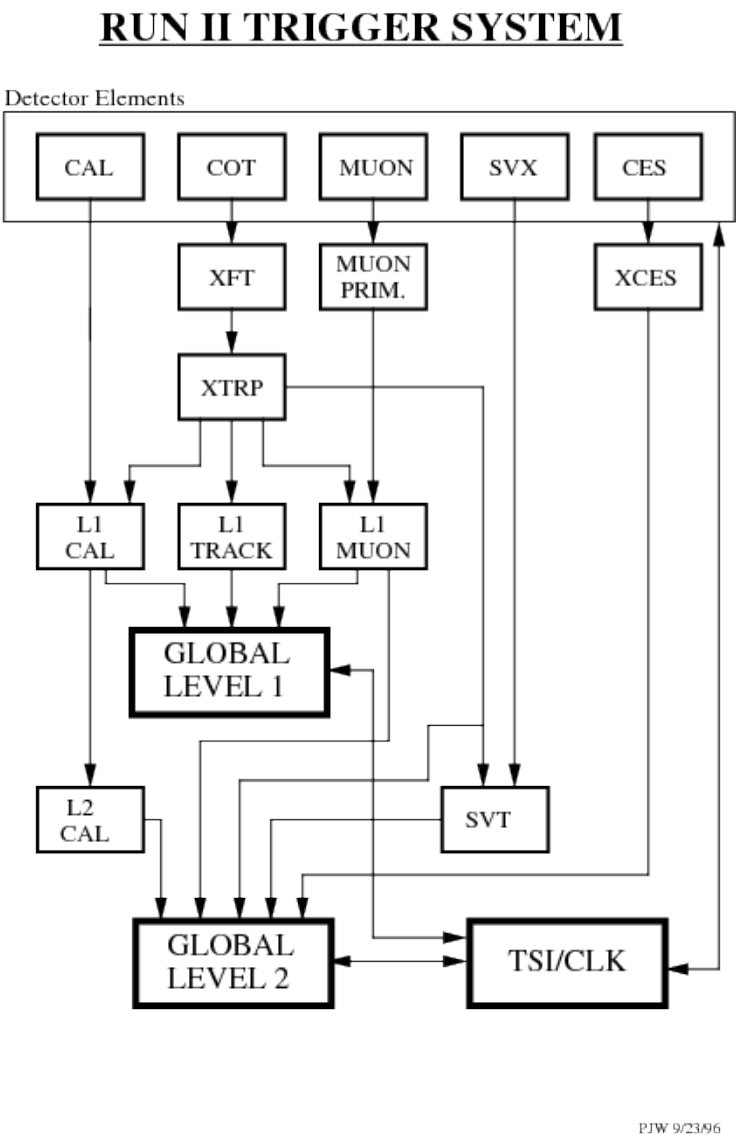


Figure 3.14: *Block diagram of level 1 and level 2 triggers. The involved sub-detectors are indicated.*

As the Tevatron luminosity increased during the Run II, the performance of the level 2 trigger started to degrade since:

- the SVT processing time (and therefore the deadtime) increased because of the higher event complexity
- the background rate accepted by the calorimetric trigger increases

In order to cope with the former issue, a faster track fitter and other changes described in [69] were implemented. In order to deal with the latter issues the L2 calorimetric trigger was upgraded such that the full trigger tower information (same accuracy as the offline one) would be available [70].

3.7.3 Level 3

Level 3 (L3) is a software trigger. The L3 farm is made of about 300 commercial dual processor computers. L3 addresses event components delivered by L2 to the Event Builder (EVB), which reconstructs the entire event with the same accuracy as in the offline analysis, which was unavailable to the lower trigger levels. The L3 algorithms makes a full three-dimensional track reconstruction. The final decision to accept an event is made on the basis of a list of required observables indicating candidate events of physical interest (top production events, W/Z events, Drell-Yan events, etc.). Accepted events exit L3 at a rate of up to 100 Hz and are permanently stored on tapes. Store events will be analyzed offline.

A set of requirements to be fulfilled by an event at L1, L2, and L3 is called trigger path. While at CDF II about 150 trigger paths exist, we will be using the trigger paths described in Chap. 8, 9.

Chapter 4

Object Identification

Information of the CDF sub-detectors are combined in order to reconstruct the "physical objects" (tracks, electrons, muons, jets, missing transverse energy, etc.) of interest for the analysis. The measurement presented in this thesis strongly relies on the identification of those objects. We describe them in the next sections.

4.1 Tracks

The trajectory of a charged particle in a homogeneous magnetic field is described by a helix depending on 5 parameters $\overline{\alpha} = (\lambda, C, z_0, d_0, \phi_0)$, where:

- λ: the helix pitch, i.e. $\cot(\vartheta)$, where ϑ is the polar direction of the track at its closest approach to the z-axis
- C: signed (depending on the charge sign) helix half-curvature, whose absolute value is defined as $C \equiv 1/2r$, where r is the radius of the helix
- z_0: the z coordinate of the point of closest approach to the z-axis
- d_0: signed (depending on the azimuthal hemisphere) impact parameter, i.e. the distance in the transverse plane of closest approach to the z-axis
- ϕ_0: the azimuthal angle at the closest approach.

The trajectory of a charged particle satisfies the following equations [71]:

$$x(\phi) = rsin(\phi) - (r + d_0)sin(\phi_0) \tag{4.1}$$

$$y(\phi) = -rcos(\phi) + (r + d_0)cos(\phi_0) \tag{4.2}$$

$$z(\phi) = z_0 + s\lambda \tag{4.3}$$

where s is the projected length along the track in the transverse plane ($\phi = 2Cs + \phi_0$). The track reconstruction consists in determining the above helix parameters from a set of spatial measurements ("hits") reconstructed in the tracking detectors by the clustering algorithm. The fit to the hits with an helix functional form takes into account field non-uniformities and scattering in the detector material. In the next sub-section we describe all the different tracking algorithms used in this analysis.

4.1.1 Tracking Algorithms

COT Tracking

The COT drift chamber efficiently reconstructs tracks and measure their momenta up to $|\eta_{det}| = 1$. For each sense wire signal above threshold, the COT front-end provides the information on the integrated charge and the arrival time of the avalanche with a few nanoseconds resolution. The hit time is synchronized with the time of the event occurrence with respect to Tevatron clock. The hit space and time information is used by the tracking pattern recognition as described below.

The tracking pattern recognition works in the following steps. Further details are given in [72].

- **Segment finding**: in each of the eight superlayers, triplets of hits consistent with a straight line are used as seeds. Possible other hits within a fixed 1 mm window are added to the straight line fit in an iterative procedure. Segment-seeds are allowed to cross cell boundaries.

- **Axial tracking reconstruction:** at the beginning tracks are reconstructed in the transverse plane (2D) by using axial superlayers only.

- **Stereo tracking reconstruction:** once an axial track is found, this algorithm tries to add hits in the stereo angle superlayers to the axial track.

- **Re-fit**: at the end, tracks are re-fit by taking into account the detailed magnetic field map and the detailed electron drift path in the cells.

- **Duplicates removal**: the set of reconstructed COT tracks include numerous duplicates. Two tracks are called duplicates based upon the fraction of shared hits. If this fraction is more than 15% of the total, the track is considered a duplicate. Among the duplicates, the track with at least 20 axial and 20

stereo hits is kept. If no such a track is found, the duplicate with the most COT hits is kept.

If all channels of the COT are properly working, the tracking efficiency is typically close to 100% for high P_T tracks ($P_T > 1.5$ GeV/c), while drops to 95% for tracks with $P_T < 500$ MeV, since they may miss the outer layers. The nominal single hit resolution is 180 μm. The typical resolutions on track parameters are the following ([52]): $\sigma_{P_T}/P_T^2 \sim 0.0015$ (GeV/c)$^{-1}$, $\sigma_{\phi_0} \sim 0.035°$, $\sigma_{d0} \sim 250$ μm, $\sigma_\vartheta \sim 0.17°$, and $\sigma_{z_0} \sim 0.3$ cm for tracks fit with no silicon information or beam constraint.

Silicon Detector Tracking

A charged particle going through a silicon sensor often deposits charge in several contiguous strips. Only strips above a certain threshold and their immediate neighbors are read out. The purpose of the hit clustering algorithm is to assign sets of strips that belong to the same track, gather them into a strip cluster ("hit"), and estimate the position of the track impact. Compared to COT information, adding the silicon information improves the impact parameter resolution of tracks, which may improve to $\sigma_{d0} \sim 20$ μm if enough silicon hits are attached to the tracks. The silicon tracker improves also the stereo resolution ($\sigma_\vartheta \sim 0.06°$, $\sigma_{z_0} \sim 70$ μm), while keeping the transverse momentum and azimuthal resolutions approximately the same as for COT-only tracks.

Tracking Algorithms

Tracking algorithms exploit the information from the COT and silicon sub-detectors. The following tracking algorithms are used at CDF [1]:

- Stand-alone (SA) [73]. The standalone tracking algorithm reconstructs the tracks by making use of silicon information only. Its main purpose is to capture tracks with P_T as low as 0.1 GeV/c and to improve the track reconstruction efficiency for forward tracks where the COT has a limited reach. The algorithm starts with a $R\phi$ fit. Since we have to extract 3 track parameters (d_0, R, ϕ_0) track candidates with less than 4 $R\phi$ hits are not considered. Once the $R\phi$ fit is performed, the Rz hits to be attached to the track candidate

[1] In the following we describe carefully the different CDF tracking algorithms which are used especially for isolating the jets carrying b-flavor (Chap. 5). Improving the efficiency in recognizing those jets was one of my main task at CDF.

are searched for. A new fit is performed, thus obtaining the two remaining longitudinal parameters (λ and z_0). The transverse parameter are also recalculated. Run dependent simulations show that the overall tracking efficiency is ∼85% in the region $|\eta| < 0.75$ and $|z_0| < 35$ cm, which is fully covered by SVXII.

- Inside-Out (IO) [74] : silicon response is used to find seed tracks and COT hits are linked. This algorithm allows tracking also beyond the $|\eta| = 1$ region where the COT-only tracking efficiency degrades.

- Outside-In (OI) [75]: COT tracks are used as input and the available SVX hits are linked. This algorithm allows best tracking in the $|\eta| \leq 1$ region. The OI tracking algorithm takes full advantage of the excellent COT track recognition and of the high spatial resolution provided by the silicon detectors. Both COT and silicon detectors allow for 3D reconstruction. However, the longitudinal resolution in the COT is one order of magnitude worse than the axial one, therefore making the reconstruction of 3D tracks more challenging. The outside-in tracking consist of two strategies. The first, called "OIT", search for $R\phi$ hits in the silicon layers to be associated to the COT track. The second strategy, called "OIZ", takes as input the OIT tracks and tries to add silicon z hits where the $R\phi$ hits were already found. Algorithm-wise, OIZ is very similar to OIT. The only difference is that OIZ consider the z hits, which are not accounted for in the OIT algorithm. We are going to describe only the OIT strategy in this thesis. Details about the OIZ algorithm are described elsewhere ([75]).
  OIT extrapolates COT tracks in the silicon detectors [2]. Axial hits are attached to the track from the outermost layer moving inward in an iterative procedure. Silicon hits are searched for in a 4σ wide road, where σ is the error matrix on the track parameters. At each iteration, i.e at each layer, the track is re-fit in order to improve the pointing resolution. The error matrix is adjusted to account for the increased amount of material encountered.
  A new track candidate is generated for each hit in the road (Fig. 4.1). The candidates are ranked based on the number of hits found so far. Only the top two are saved for the following iteration.
  At the end of this process the candidate that has the largest number of hits is chosen. If two or more candidates have the largest number of hits, the one

[2] An excellent alignment between COT and silicon is required.

with the smallest reduced χ^2 from the fit is chosen.

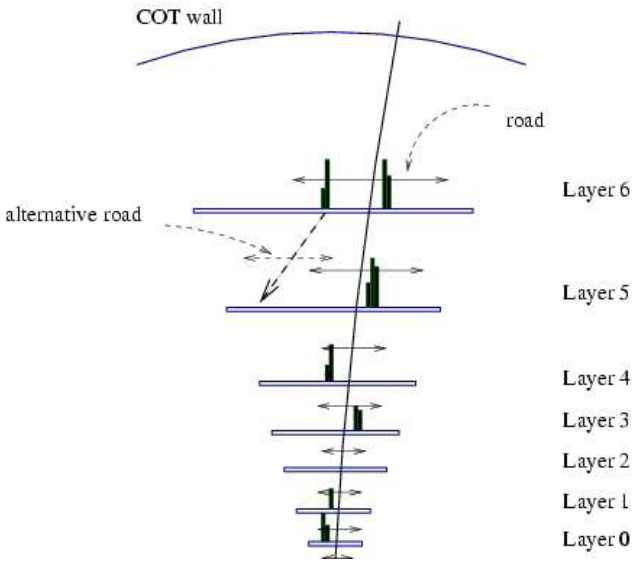


Figure 4.1: *Given a set of silicon hits and a COT-only reconstructed track, the ambiguities in choosing the silicon hits to be associated to the tracks are illustrated.*

The OI algorithm is the most commonly used tracking algorithm at CDF.

- Phoenix forward tracking algorithm [76]: the Plug region ($1 < |\eta_{det}| < 3.6$) is not covered by the COT: a calorimetry-seeded outside-in tracking algorithm, named "Phoenix algorithm", is needed to identify tracks. Phoenix algorithm reconstructs the track of an electron (Phoenix electron or PHX) by using:

  1. the reconstructed 3D point where the $p\overline{p}$ interaction took place (primary vertex);
  2. the 3D position of the EM shower in the PES;
  3. the transverse energy deposited in the calorimeter.

  The primary vertex and the shower position are used as seed for the track helix and the helix curvature is varied to determine a momentum, which matches the transverse energy measured by the calorimeter. If such a match is found the available SVX hits are added to allow for a better fit.

- Event vertex finding algorithm [77]: the position of the interaction point (x_V, y_V, z_V) is reconstructed by the following iterative algorithm. Starting from an initial guess (x_V^I, y_V^I, z_V^I) [3] all OI tracks with z_0 within 1 cm from z_V^I, $P_T \geq$

[3] Typically from the point of nearest approach to the beam line of the highest-P_T track.

500 MeV, $d_0 \leq 1$ cm (with respect to the beamline), $d_0/\sigma_{d0} \leq 3$, and at least 6 hits in the COT (at least 2 axial and stereo segments), are fitted to a common vertex. The tracks with $\chi^2 > 10$ are pruned from the fit and the procedure is iterated until all considered tracks have $\chi^2 \leq 10$. The longitudinal position of each vertex is the mean z_0 of its associated tracks, weighted on the respective σ_{z_0}. Typical resolution is $\sim$100 μm. The vertex associated with the highest sum P_T of the tracks is defined as the primary vertex of the event, where the hard interaction occurred. The other reconstructed vertices are typically associated to minimum bias events.

4.2 Electrons

Electron "objects" are formed from energy clusters in neighboring towers of the calorimeter. A central electron cluster ($|\eta_{det}| < 1.1$) is composed of an electromagnetic seed tower ($E_T > 2$ GeV) and at most one additional tower ($E_T > 0.1$ GeV), which is adjacent to the seed tower in η_{det} and within the same wedge. It is also required that a COT track extrapolates, assuming helicoidal trajectory, from the beam line to the seed tower. The highest-P_T matching track is used to derive the direction of the electron. The energy E is obtained from the calorimeter deposition. Forward electron candidate clusters in the plug calorimeter are made of 3×3 $\eta_{det} \times \phi$ neighboring towers. The plug electron is required to be in the region $1.1 < |\eta_{det}| < 2.8$, and a track reconstructed by the Phoenix forward tracking algorithm is required to extrapolate from the beam line to the position of the cluster.

Electron candidates for these measurements must lie within well-instrumented regions of the calorimeter. Selection requirements are listed in Table 4.1. Note that central electrons (TCE) and forward electrons (PHX) have different identification cuts. The meaning of the identification variables in the table is explained below.

- E_{had}/E_{em} is the ratio between the energy deposited in the hadronic and in the electromagnetic calorimeter. The linear term accounts for the increased shower leakage into the hadronic calorimeter.
- Isolation is the ratio between the additional transverse energy in a cone of radius $R = 0.4$ around the cluster and the transverse energy of the cluster.
- Track z_0 is the position along the beamline of the track at the point of minimal approach to the beamline.

- Track P_T is the transverse momentum of the track.
- COT Axial and Stereo Segments are, respectively, the number of axial and stereo superlayers in the COT which have at least 5 hits attached to the track.
- L_{shr} ([78]) measures how well the lateral shower profile of the cluster compares with that of the test beam electrons. The electron towers adjacent to the cluster seed are exploited to build such a variable.
- E/p is the ratio between the cluster energy and the track momentum. In principle this ratio should be 1, but bremsstrahlung photons may be radiated by the electron in the material of the inner tracking volume, thus lowering the momentum measured by the COT with respect to the corresponding clustered energy. For high energy electrons the photons and the electrons often end up in the same calorimeter tower.
- QΔX is the measured charge of the particle times the signed difference in x between the track and the cluster associated to the electron when the track is extrapolated to the position of the shower max. The cut accounts for asymmetric tails in the distribution due to the bremsstrahlung radiation.
- $|\Delta z|$ is the absolute value of the difference in z position between the cluster and the extrapolated track.
- PEM 3x3 χ^2: the distribution of tower energies in a 3×3 array around the seed tower is compared to distributions from electron test-beam data. An appropriately defined χ^2 measures how well the former compares with the latter.
- PES 5x9 U/V: showering information from the PES detectors is also considered by looking at how well the shower profile matches the energy distribution in the PEM towers. This can be checked in both sets of PES strips (labeled "U" and "V").
- ΔR between PES and PEM centroids: is the distance in the η, ϕ plane between the reconstructed cluster in the PEM calorimeter and the shower maximum detector.
- Silicon hits is the number of hits in the silicon detector associated to the track which matches the calorimeter cluster.

Photons may convert to electron-positron pairs as they traverse the material. A $\gamma \to e^+e^-$ veto ([79]) is implemented to reduce the $\gamma \to e^+e^-$ misidentified as electrons.

Further details are describe in [8], [56], [80].

Identification Cut	TCE	PHX
E_{had}/E_{em}	$< 0.055 + 0.00045 \cdot E$ (GeV)	0.05
Isolation	≤ 0.1	
Track $\lvert z_0 \rvert$ (cm)	≤ 60	
Track P_T	≥ 10 GeV/c	-
COT Axial Segments	≥ 3	-
COT Stereo Segments	≥ 2	-
L_{shr}	≤ 0.2	-
E/p	$< 2.5 + 0.015 \cdot E_T$ (GeV)	-
$Q \cdot \Delta X$	≤ 1.5 ≥ -3.0	-
$\lvert \Delta z \rvert$ (cm)	≤ 3.0	
PEM 3x3 χ^2	-	< 25
PES 5x9 U	-	> 0.65
PES 5x9 V	-	> 0.65
ΔR between PES and PEM centroids	-	< 3.0 cm
Silicon Hits	-	≥ 3

Table 4.1: Central (TCE) and forward (PHX) electron identification cuts.

4.3 Muons

Muons are unlikely to shower in the calorimeter because of their large radiation length ($\sim 4 \times 10^4$ times larger than for electrons). Therefore, high-P_T muons deposit in average only a small fraction of their energy in the calorimeter: they behave as minimum ionizing particles (MIPs).

Real muons, representing a background for this analysis, are cosmic rays and secondary muons originated from decays in flight of kaons or of charged pions or from heavy-flavor semi-leptonic decays within hadronic showers. A veto, based on timing information provided by the muon chambers and the COT, is used to reject cosmic ray muons ([81]). Secondary muons are removed by the requirements listed in Table 4.2 (e.g: d_0).

The types of muons used in this analysis are summarized below. Isolation is computed at calorimeter level as for the electrons (Sec. 4.2), and calorimeter clusters, when available, have to be consistent with the MIP hypothesis.

- **CMUP**: central muons with $|\eta| < 0.6$, isolated COT tracks matched to calorimeter clusters, and with stubs in both CMU and CMP muon detectors.

- **CMX**: central muons with $0.65 < |\eta| < 1.0$, isolated COT tracks matched to calorimeter clusters, and with stubs in the CMX detector.

- **CMU**, **CMP**: central muons with isolated COT tracks matched to calorimeter clusters, hits in the CMU, CMP chambers respectively, but not in both of them.

- **CMIOCES**, **CMIOPES**: respectively central, forward muons with isolated COT tracks matched to calorimeter clusters with $E_T > 0.1$ GeV. No requirements on the presence of stubs in the muon chambers is made. This categories extends the acceptance beyond the coverage of the muon detectors.

- **CRKTRK**: well-reconstructed isolated COT tracks. No requirements on the presence of stubs in the muon chambers and on the minimal energy deposit in the calorimeter are made. This category is used to recover muons (or electrons [4]) pointing to cracks in the calorimeter (e.g: $\eta_{det} \sim 0$).

The aforementioned muon categories may overlap. In order to avoid double-counting we remove from the list the muon candidates that have been already identified. The muon identification follows the above descending priority order.

Selection criteria for identifying the CMUP and CMX muons are summarized in table 4.2. The same identification cuts are applied for the CMU, CMP, CMIOCES, CMIOPES, CRKTRK muon categories, except for the already described differences. The meaning of the identification variables in the table is explained below.

- Track d_0: the distance of closest approach of the track to the beamline (impact parameter). This cut is applied to remove muons from K and π decays or cosmic rays.

- ρ_{COT}: the radial distance at which the track appears to leave the COT.

[4] Simulations show that about 70% are generated muons while the rest are electrons

- χ^2/NDOF: the χ^2 divided by the degrees of freedom of the fit to the track. This cut aims to reject muons which do not originate at the primary interactions points.

- $\Delta X_{CMU,\ CMP,\ CMX}$: stubs in the CMU, CMP, CMX chambers are matched to tracks in the $r-\phi$ plane, and ΔX is defined as the average distance of stub hits from the track.

Further details are describe in [8], [56], [82].

Identification Cut	CMUP	CMX
E_{em}	$\leq 2.0 + max(0, 0.0115(p-100))$	
E_{had}	$\leq 6.0 + max(0, 0.0280(p-100))$	
Isolation	≤ 0.1	
Track $\lvert z_0 \rvert$	≤ 60	
Track d_0	≤ 0.2 (0.02 w/ silicon hits)	
COT Axial Segments	≥ 3	
COT Stereo Segments	≥ 2	
χ^2/NDOF	≤ 3 (4 for earlier data periods)	
ρ_{COT}	> 140	-
ΔX_{CMU}	< 7	-
ΔX_{CMP}	$<$ max(6, $150/P_T$)	-
ΔX_{CMX}	-	$<$ max(6, $125/P_T$)

Table 4.2: Some of the central muon identification cuts. Energies are listed in GeV, momenta in GeV/c, lenghts in cm. More identification cuts are specified in [83].

4.4 Outgoing partons: Jets

Quarks and gluons involved in hard scatterings at hadron colliders undergo showering processes which create cascades of partons. In the world of MC generators outgoing partons keep branching until a cut-off momentum scale of order of 1 GeV is reached. After that we enter in the confinement regime and the hadronization/fragmentation process takes place: the colored partons merge into colorless hadrons. Fragmentation hadrons or their decay products reach the detector: this process is sketched in Fig. 4.2. The resulting cascade of particles is clustered in confined areas of the calorimeter: the **jet**.

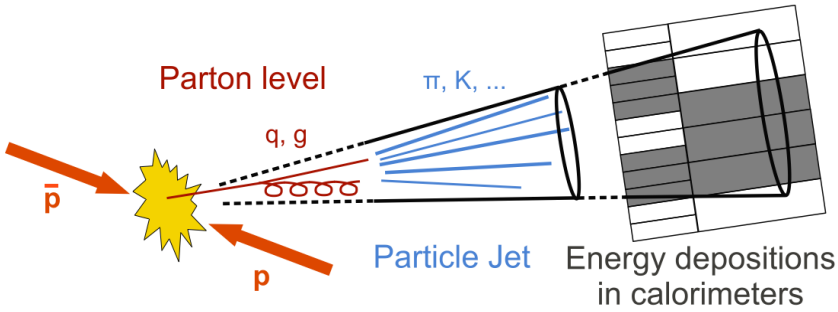


Figure 4.2: *From initial partons to the final jets.*

One the most intriguing challenges at hadronic colliders is to be able to extract from the measured calorimetric energy the four-momentum of the hadronized outgoing partons. At CDF this is done as follows.

The energy scale of CEM/PEM the calorimeter towers [84] is calibrated such that the measured invariant mass of electron pairs from Z's is consistent with the mass measured at LEP [85]. Photon radiation is taken into account and the resolution of the electron-pair invariant mass is improved by making use of the track measured momenta when possible.

The energy scale of the CHA/WHA/PHA calorimeter towers [84] is determined from their response to a test beam of 50 GeV/c charged pions, which did not interact in the front electromagnetic compartments ([86]).

Since the energy scale of the calorimeters decreases with time due to the aging of the scintillators and PMTs, the stability of the calorimeter response is monitored online using a number of methods [87]. The stability of the electromagnetic scales is then restored using the time dependence of the reconstructed Z-boson mass in $Z \to e^+e^-$ events as shown in Fig. 4.3. A systematic uncertainty on the electromagnetic scale of 0.3% is assessed to cover the residual scale fluctuations. The stability of the WHA and CHA scales is verified with the mean energy deposited by muons with $|\eta| < 1$ in $W \to \mu\nu_\mu$ events (Fig. 4.3). A systematic uncertainty on the hadronic scale of 1.5% is assessed. For PHA, samples of muons and generic jets are both used to ensure the stability of the scale. By considering the above uncertainties and the typical jet the average deposit in the CEM calorimeter, which is about 70%, the systematic uncertainty on the energy scale stability is about 0.5%.

The energy of a jet is computed from the energy of its towers. The way to associate a collection of towers to a jet is not unique since there is no control on how the QCD radiation and hadronization take place. However, the jet definition

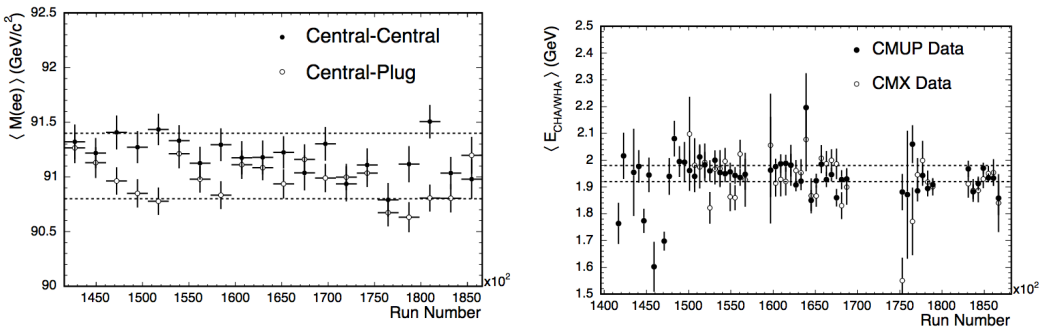


Figure 4.3: *Left plot: mean invariant mass of $Z \to e^+e^-$ candidates versus run number for events with invariant mass between 86 GeV/c^2 and 98 GeV/c^2. Events with two electrons in the central calorimeter or one electron in the central and one in the plug calorimeters are shown with full or empty circles respectively. Right plot: mean energy observed in the CHA/WHA for muons with $P_T > 20$ GeV/c from $W \to \mu\nu_\mu$ candidate event versus run run number. In both plots the dashed line indicate the estimated uncertainties on the stability of the energy scale (0.3%/1.5% in the left/right plots)*

should satisfy the following requirements:

1. the simulated parton-to-jet process should be infrared and collinear safe: a measured jet cross section should not change if the original parton radiated a soft parton or splitted into two collinear partons;

2. An optimal algorithm must have small uncertainties associated with the different theoretical parameters (e.g: renormalization scale. See Chapter 7), which are used as inputs for the simulations. Such a requirement gives enough confidence that missing higher-order QCD contributions do not change significantly the fixed order predictions;

3. close correspondence to the original parton direction: the algorithm should give consistent results if applied at parton, particle or calorimeter level (see Fig. 4.2);

4. should not be sensitive to longitudinal Lorentz boosts.

5. easy to use in experimental analyses and theoretical calculations: the same jet algorithm must be reproducible in the theoretical calculations;

6. small experimental uncertainties;

7. application flexibility to experimental conditions:

   - performances stable against multiple interactions in the same bunch crossing, and independent of the kinematics of the event;
   - the jet reconstruction code should allow for an easy calibration of the jet four momenta.

4.4.1 Jet Reconstruction

A number of jet reconstruction algorithms have been adopted in HEP experiments. Jet clustering algorithms belong to two main classes ([88]) described below.

- Sequential recombination: closest towers are pair-wise recombined until there are available adjacent towers. The resulting clusters are called jets.
- Cone: find coarse regions of large energy flow and call them jets.

The jet algorithm used for this analysis belongs to the cone algorithm class and is called JETCLU [89]. The cone radius is defined as follows:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{4.4}$$

where $\Delta\eta$ and $\Delta\phi$ are the distances in pseudorapidity and azimuthal angle between a tower center and a cluster axis. JETCLU proceeds through the following steps:

1. searching for towers with $E_T \equiv E^{em} sin\vartheta^{em} + E^{had} sin\vartheta^{had} > 1GeV$, where ϑ^{em} (ϑ^{had}) is the polar angle of the vector pointing from the primary vertex [5] to the geometrical center of the electromagnetic (hadron) tower.

2. The towers above 1 GeV are marked as pre-cluster seeds and ordered with decreasing E_T. All seeds in a 49-towers square centered on the highest E_T seed define a pre-cluster. Seeds cannot belong to more than one cluster. The centroid of the pre-cluster is calculated as the E_T-weighted center of the seeds.

3. A $R = 0.4$ cone is drawn around the centroid and all $E_T > 0.1\ GeV$ towers in the cone are added. The centroid is then recalculated as follows:

$$(\eta^{bPC}, \phi^{bPC}) = (\frac{\sum_j \eta_j E_{T_j}}{\sum_j E_{T_j}}, \frac{\sum_j \phi_j E_{T_j}}{\sum_j E_{T_j}}) \tag{4.5}$$

[5] reconstructed position of the $p\overline{p}$ collision

4. The step above is iterated until $(\eta^{bPC}, \phi^{bPC})_i \simeq (\eta^{bPC}, \phi^{bPC})_{i+1}$, i being the ith iteration. Jet candidates are now defined.

5. Since jet candidates may share some amount of transverse energy (see figure fig. 4.4)

   - jet candidates are merged if the shared E_T is more than 75% of the less energetic E_T jet. Jet centroid is then recalculated.
   - if not, shared E_T is assigned to the closest jet candidate.

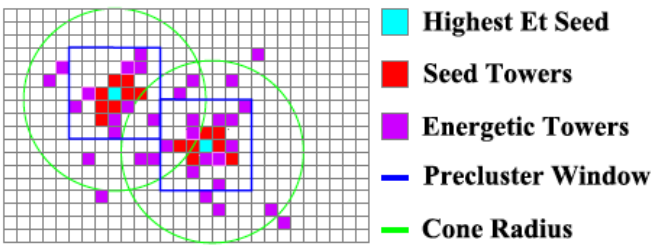


Figure 4.4: *Jet reconstruction by the JETCLU algorithm.*

A jet is uniquely identified by this procedure. The four-momentum of the jet is computed from its towers by using the so called "E-scheme" [89], which is manifestly Lorentz invariant:

$$E = \sum_{i}^{N_{towers}} (E_i^{em} + E^{had})_i \tag{4.6}$$

$$P_x = \sum_{i}^{N_{towers}} (E_i^{em} sin\vartheta_i^{em} + E_i^{had} sin\vartheta_i^{had}) cos\phi_i^{em} \tag{4.7}$$

$$P_y = \sum_{i}^{N_{towers}} (E_i^{em} sin\vartheta_i^{em} + E_i^{had} sin\vartheta_i^{had}) sin\phi_i^{em} \tag{4.8}$$

$$P_z = \sum_{i}^{N_{towers}} (E_i^{em} cos\vartheta_i^{em} + E_i^{had} cos\vartheta_i^{had}) \tag{4.9}$$

where x, y are the Cartesian coordinates in the transverse plane. In this scheme the tower signals are handled as massless particles. Therefore:

$$P = \sqrt{P_x^2 + P_y^2 + P_z^2} \tag{4.10}$$

$$P_T = \sqrt{P_x^2 + P_y^2} \tag{4.11}$$

$$E_T = E \times \frac{P_T}{P} \tag{4.12}$$

$$y = \frac{1}{2} ln \frac{E + p_z}{E - p_z} \tag{4.13}$$

$$\phi = arctan(\frac{P_y}{P_x}) \tag{4.14}$$

4.5 Neutrinos: missing transverse energy

Neutrinos interact with the matter at very low rate. Because of their extremely small interaction cross section, neutrinos escape the CDF detector, carrying away part of the energy associated with the hard interaction, and thus creating energy imbalance in the event. In hadronic colliders the longitudinal momentum of the interacting partons is unknown, while their transverse momentum can be neglected with respect to the hard interaction energy scale. By requesting that the total observed transverse momentum in an event be zero, one can define the transverse missing momentum and, under the hypotheses of momentum conservation and that no other particles escape the detector, associate this quantity to the neutrino. The transverse missing momentum, or better known as transverse missing energy is defined as follows:

$$\vec{\not{E}}_T^{\,raw} = -(\sum_{towers} E_i sin\vartheta_i)[cos\phi_i, sin\phi_i] \tag{4.15}$$

where E_i is the energy released in the ith massless tower, ϑ_i the polar angle defined from the interaction vertex (Sec. 4.1.1) to the tower position, ϕ_i the tower azimuthal angle. All towers with energy deposit above the pedestals and with $|\eta| < 3.6$ are used.

The missing transverse energy defined in 4.15, which we call "raw $\not{E}_T$", is the one used by a number of triggers at CDF [6]. Offline, we correct it because of the two reasons described below. The first $\not{E}_T$ correction is due to known jet mismeasurements. Since jet energies are corrected to reproduce the energies of the

[6] Online, ϑ_i is the polar angle defined from the CDF geometrical center.

hadronizing partons (Chap. 6), $\not\!\!E_T$ has to be updated accordingly. We do that as follows:

$$\vec{\not\!\!E}_T^{\,JEScorr} = \vec{\not\!\!E}_T^{\,raw} - \sum_{jets} (\Delta J\vec{E}S)_i \tag{4.16}$$

where $\Delta J\vec{E}S_i$ is the difference between the corrected and uncorrected transverse energy of the ith jet. Moreover we correct the $\not\!\!E_T$ when muons are identified. At the investigated energies muons pass through the calorimeter as minimum ionizing particles, hence causing an apparent missing transverse energy in the calorimeter response. However, muon momentum is accurately measured by the tracking system. We can use the muon momentum (P^μ) rather the small calorimeter deposit ($E_{CAL} \sim 2$ GeV) to recalculate $\not\!\!E_T$. In formula

$$\vec{\not\!\!E}_T \rightarrow \vec{\not\!\!E}_T = \vec{\not\!\!E}_T^{\,JEScorr} - \vec{P}_T^{\mu}(1 - \frac{E^\mu_{CAL}}{P^\mu}) \tag{4.17}$$

In summary, in the data analysis the missing transverse energy is defined as follows:

$$\vec{\not\!\!E}_T = \vec{\not\!\!E}_T^{\,raw} - \vec{P}_T^{\mu}(1 - \frac{E^\mu_{CAL}}{P^\mu}) - \sum_{jets} (\Delta J\vec{E}S)_i \tag{4.18}$$

In electro-weak processes, where heavy bosons are created, $\not\!\!E_T$ is likely to be associated to neutrino's transverse energy. On the other side, when dealing with multi-jet events large $\not\!\!E_T$ arises from calorimeter mis-measurements. This is because the jet energy corrections are estimated in average and cannot account for event-by-event mis-measurements. Therefore a "fake" momentum imbalance can mimic the neutrino signature.

The resolution of the $\not\!\!E_T$ depends on the fluctuations in the calorimeter response. It was estimated by using minimum bias events with no requirements on the amount of energy deposited in the calorimeter. In minimum bias events, which are dominated by inelastic $p\bar{p}$ collisions, no real $\not\!\!E_T$ is expected. $\not\!\!E_T$ resolution is parametrized in terms of total scalar E_T, $\sum E_T$ (Eq. 4.19). Coefficients of such a parametrization are fitted to data [90].

$$\sigma(\not\!\!E_T) = 0.582 + 0.7418\sqrt{\sum E_T(GeV)} \tag{4.19}$$

Chapter 5

Identification of bottom jets

The identification of b-hadrons in jets was fundamental for the discovery of the top quark in 1995 ([2]) and is one of the crucial features for the evidence of the light Higgs boson-like particle at the Tevatron collider ([3]). While in the top or light Higgs decay chains jets carrying b-flavor ("b-jets") are produced, a large portion of the major background processes (e.g. W+jets) is composed of light-quark jets.

The discrimination of b-jets is performed by exploiting their unique features (Sec. 5.1). Several methods to identify b-jets exist at CDF. We will describe the one ("jet b-ness") used in the analysis reported in this thesis in Sec. 5.2, while we will also summarize the main features of other b-taggers developed at CDF in App. A.

5.1 b-jets versus other jets

Jets carrying b-flavors feature a number of differences from light-quark jets. The proper lifetime of b-flavored hadrons (B^0, $B^{\pm}$, Λ_b), products of the b-quark hadronization, is $\sim$ 1.4-1.6 ps. If, for example, we have a B^0 hadron with an energy of 40 GeV, it will travel on average about 3.5 mm before decaying. Such a distance can easily be resolved by the CDF tracking system (see Sec. 3.3). Most of the light quark-jets are composed of hadrons whose lifetime is so large that they are identified before decaying (e.g: $\pi^{\pm}$), or by hadrons (e.g: π^0) decaying extremely fast in long-lived or stable particles. Furthermore, due to the large mass of the b-flavored hadrons, their decay products tend to have a large momentum transverse to the b-quark direction. This causes b-jets to be wider and to have larger invariant masses than the jets not containing b-quarks. Also, since the B-hadrons are normally the leading secondaries within the jet, they have a larger forward boost and the average

momentum of their decay particles is larger than the average track momentum in light-quark jets. Finally, particle multiplicity is larger for b-jet. This is because B-hadrons decay chains can easily produce a large number of particles.Among the decay products, electrons or muons plus neutrinos can be found in about 22% of the times. Leptonic decays are yet another feature which can help distinguishing b-jets from light quark-jets.

5.2 An effective b-jet identification tagger: Jet b-ness

In the analysis described in this document we employ the jet "b-ness" b-tagger ([91]). Jet b-ness is a multivariate, feed-forward artificial neural-network (ANN) tagger whose output provides a figure of merit indicating how b-like a jet appears to be. What distinguishes this tagger from other multivariate taggers used at CDF is its emphasis on studying individual tracks for characteristics indicating that they may have come from a b-hadron decay. By exploiting a track-by-track ANN, tracks from B-hadron decays are discriminated from other tracks. For each track within the jet a "track b-ness" value is assigned based on the track-by-track ANN output. Such a value is fed into the final ANN ("jet b-ness"), which, together with other jet variables, aims at distinguishing b-jets. The architecture of the track-by-track ANN has two hidden layers of 15 and 14 nodes. The final ANN is composed of two hidden layers of 15 and 16 nodes. Both of them are trained by using the MLP algorithm from the TMVA package [92]. A flow chart, describing the jet b-ness tagger, is shown in Fig. 5.1.

5.2.1 Track-by-track ANN - track b-ness

The track b-ness is evaluated for each track within a cone of $\Delta R < 0.4$ about the jet axis. Tracks with hits from the COT only are rejected, as the COT alone does not have sufficient resolution to signal a possible displacement between a B-hadron decay point and the primary vertices. Only tracks with $p_T > 0.4\ GeV/c$ are considered. Finally, track pairs are removed if they are oppositely charged, form an invariant mass in a 10 MeV window around the K_S (0.497 GeV/c^2) or Λ (1.115 GeV/c^2) mass value, and can be fit into a two-track vertex. It is important to remove the long-lived particles contained in light quark-jets since they can mimic the b-jet signature.

The track b-ness ANN is trained with two categories of input variables. One

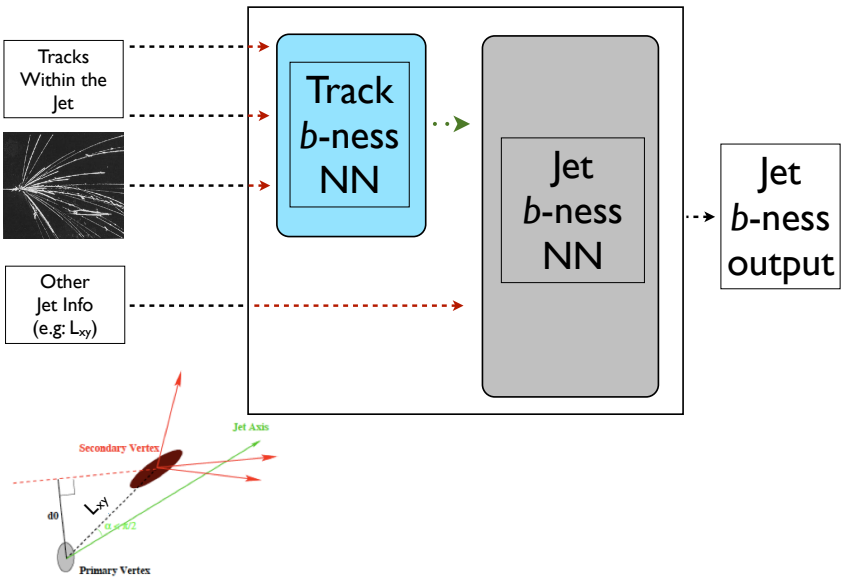


Figure 5.1: *A flow chart of the jet b-ness tagger.*

category relies on the displacement with respect to the primary vertices that B-hadron decay prongs may have. Observables in this category include the track's displacement along the beam direction z_0 and the track's signed impact parameter d_0 (Sec. 4.1). The significance on z_0 and d_0 are also used: they are defined as z_0/σ_{z_0} and d_0/σ_{z_0}, the σ's being the uncertainties on the considered parameter. The second set of observables takes advantage of the different kinematics expected for B-hadron tracks: we use track's P_T, pseudo-rapidity (η_{axis}) and perpendicular momentum (P_{perp}). η_{axis} and P_{perp} are defined relative to the jet axis. Input variables for the track-by-track ANN are summarized in Table 5.1. In Fig. 5.2 those input variables are displayed by separating tracks from B-hadron versus the rest. Input variables are built from PYTHIA MC ([93]) $ZZ \rightarrow 4\ jets$. The training of the track-by-track ANN is performed in the same sample. B-hadron tracks peak at high ANN output.

It should be noticed that, when the training was performed, a weighting was applied such that the parent jet E_T distributions were the same for both tracks from B-hadrons and the rest: this was to ensure that the application of the tagger would not introduce a kinematic bias in the analysis. The parent jet E_T is further used as input variable in order to exploit the residual discrimination power which comes from the correlation between track observables and jet E_T.

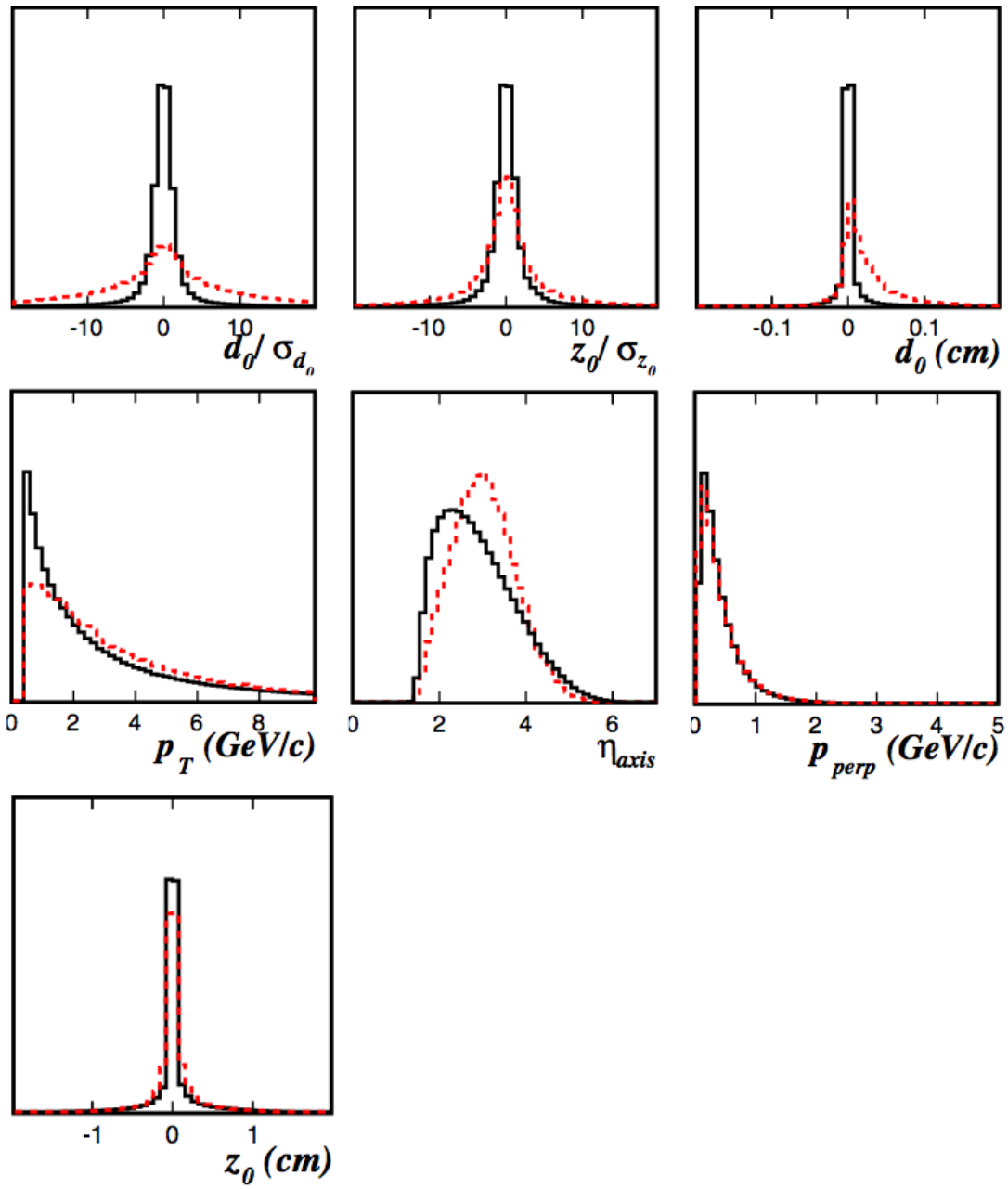


Figure 5.2: *Input variables used to train the track-by-track ANN. Jet tracks matched to B-hadron prongs (red dashed) are compared to the remaining tracks (black). Matching is performed in $ZZ \rightarrow 4$ jets PYTHIA MC [93] by requiring that the track and the generated particle to have $\Delta R < 0.14$.*

Input variable
d_0
z_0
d_0/σ_{d_0}
z_0/σ_{z_0}
P_T
η_{axis}
P_{perp}
parent jet E_T

Table 5.1: Input variables used to train the track-by-track ANN.

5.2.2 Jet-by-jet ANN - jet b-ness

A final ANN is used to determine whether a jet contains B-hadrons or not. Inputs of the final ANN are: the five highest track-by-track ANN outputs, as well as the number of tracks (N_{trk}) with ANN output larger than 0 (ANN output ranges from -1 to 1). As additional input, the significance of the displacement ($L_{xy}/\sigma_{L_{xy}}$, see Fig. A.1) of the reconstructed secondary vertices from the primary one is used. Tracks within the jet with a track-by-track ANN output larger than -0.5 are used to reconstruct secondary vertices. When fitting for the vertices, the contribution of each track to the fit χ^2 is checked. If this contribution exceeds a value of 50, the track is removed and the fit is performed again. This process is iterated either until there are less than two tracks left, or the contribution of each track to the fit is below 50. In the former case, no inputs from displaced vertices can be used. In the latter case, a secondary vertex is found and, besides L_{xy}/σ_{xy}, the invariant mass ("M_{vtx}") of the tracks employed in reconstructing the secondary vertex is used as input of the final ANN.

Since b-jets are expected to contain more K_S particles than non-b-jets, the number of K_S candidates within the jet cone is used as input variable, as well as the number of muon candidates. Muons are identified by using the soft muon tagger (App. A). As for the track-by-track ANN, b-jets and non-b-jets are weighted to have the same the jet E_T spectrum. Jet E_T is then used as input variables. A summary of the input variables used for the training of the jet b-ness ANN is given in Table 5.2. Input variables, as derived from ZZ PYTHIA MC, are shown in Fig. 5.3: b-jets are defined by matching the generated b-quark with the jet. Matching is

obtained if $\Delta R(b, jet) < 0.4$.

Input variable
five highest b-ness tracks
N_{trk}
L_{xy}/σ_{xy}
M_{vtx}
N_{K_S}
N_{μ}
parent jet E_T

Table 5.2: Input variables used to train the jet b-ness ANN.

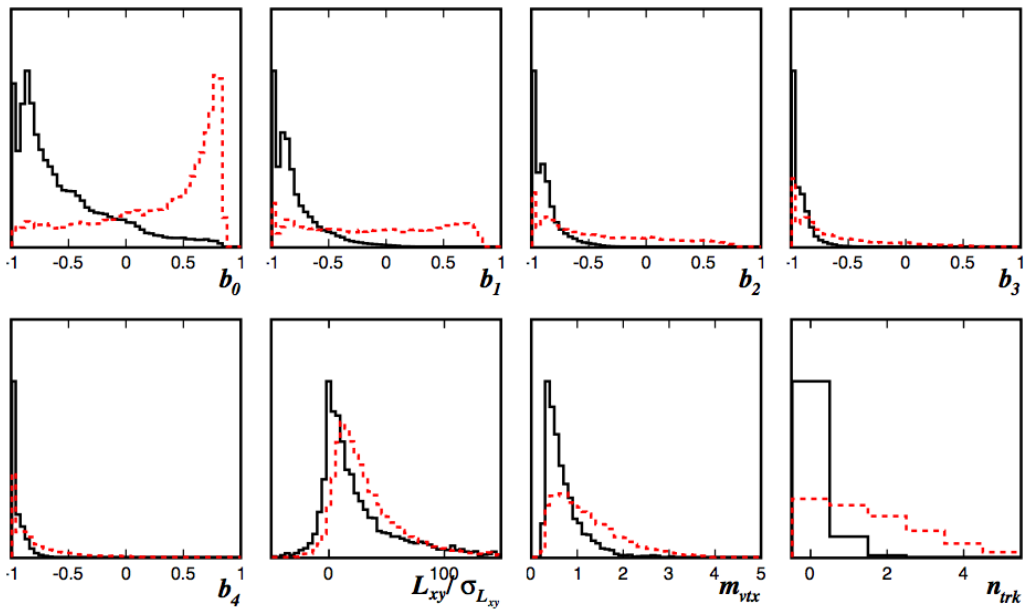


Figure 5.3: *Input variables used to train the jet b-ness ANN. Jets matched to b-quarks (red dashed) are compared to non-b jets (black) in ZZ PYTHIA MC [93]. Matching is performing by requiring that the b-quark is within ΔR of 0.4 about the jet axis. $b_0 \ldots b_4$ refer to the five highest b-ness tracks. The other symbols are defined in the text.*

.

Finally, in Fig. 5.4, we compare the jet b-ness output for jets matched to b quarks to non-b jets. Overall, a good discrimination power is observed. The only region where we do not see a good discrimination between b-jets and non-b-jets is at about

-0.8: that region is dominated by jets with no tracks with positive track-by-track b-ness (short-range decays of B-hadrons or jets produced at large $|\eta|$), no identified K_S, and no secondary vertices. All sharp peaks in the jet b-ness distributions are the result of the discrete inputs to the ANN.

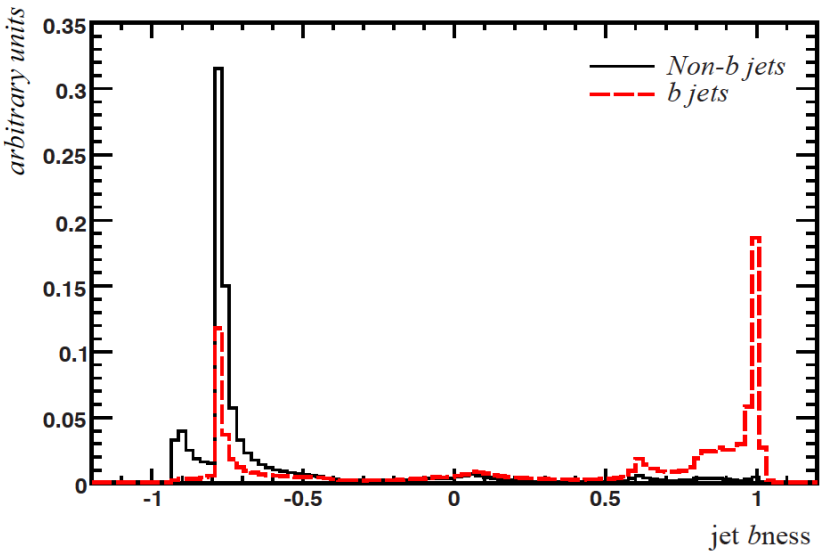


Figure 5.4: *Jet b-ness distribution for jets matched to b quarks (red) versus jets not matched to b quarks (black). These plots are made on ZZ simulated events.*

5.2.3 Calibration of the b-ness response

In order to be used in physics analyses the jet b-ness response must be calibrated. Historically, it is known ([77]) that MC does not predict well the probability of identifying b-jets (b-tag efficiency, named "e") and the rate of mis-identified non-b-jets (false tag or "mistag", named "m"). We estimate efficiency and mistag rate as a function of a minimal jet b-ness requirement "b": $e(b)$, $m(b)$. The mistag rate and the b-tag efficiency are compared in data of known flavor composition (see below) and in MC by looking at the scale factors $s_e \equiv e^{Data}(b)/e^{MC}(b)$, $s_m \equiv m^{Data}(b)/m^{MC}(b)$, and the simulations are corrected accordingly.

Mistag rate and b-tag efficiency are computed as shown below:

$$m(b) = \frac{N(b) - N_B(b)}{N - N_B} = \frac{m_{raw}(b) - s_e(b)e_{MC}(b)f_B}{1 - f_B} \tag{5.1}$$

$$e(b) = \frac{N(b) - N_L(b)}{N - N_L} = \frac{e_{raw}(b) - s_m(b)m_{MC}(b)f_L}{1 - f_L} \tag{5.2}$$

where N is the total number of jets, $N(b)$ the total number of jets above b-ness threshold b. $N_{B/L} \equiv f_{B/L} \cdot N$, where $f_{B/L}$ are the fractions of b and non-b-jets estimated by matching the jets to the generated partons in MC. $m_{raw}(b)$ and $e_{raw}(b)$ are the mistag rate and b-tag efficiency when neglecting respectively the contamination of b-jets ($N_B = 0$ in Eq. 5.1) and non-b-jets ($N_L = 0$ in Eq. 5.2). Although both of them are equal to $N(b)/N$, we use different symbols for them because they will be estimated in different samples. $m_{MC} \equiv N_L(b)/N$ and $e_{MC} \equiv N_B(b)/N$ are the MC mistag rate and b-tag efficiencies for the given jet b-ness cut. It should be noticed that $m(b)$ (Eq. 5.1) depends on $e(b)$ (via $s_e(b)$) and viceversa (Eq. 5.2). We solve for them iteratively. [1]

In order to solve equations 5.1 and 5.2 for the two unknowns e and m we need to measure the parameters of the equations in two independent regions. To reduce uncertainties on the two unknowns and correlations between the two sets of uncertainties, we decide to investigate one region dominated by light-quark jets (Z + 1 jet), and one region dominated by events with b-jets ($t\bar{t} \to l\nu$ + $\geq$4 jets). The two regions are selected as summarized in Table 5.3. In order to discriminate events with neutrinos in the final state from events with fake $\not{E}_T$ due to instrumental mis-measurements $\not{E}_T$ significance is used. Details about this tool are available in [94].

Data events collected by the high-P_T electron and muon triggers (Chap. 8) are investigated: the data sample corresponds to an integrated luminosity of 4.8 fb^{-1}. To model W/Z + jets we use ALPGEN [95], interfaced with PYTHIA [93] for the parton showering and hadronization. PYTHIA is used to model $t\bar{t}$ and the other contributions. Number of selected data events along with the predicted rates are given in Table 5.4. The b-jets purity (N_{b-jets}/N_{jets}), as predicted by the MC, for the Z + jet and $t\bar{t}$ samples are shown in Fig. 5.5a and 5.5b respectively. The aforementioned purity is plotted as a function of minimum jet b-ness in the former case, and as a function of the minimum jet b-ness of the highest b-ness jet in the latter case. Although the b-jet purity is overall low in the Z + 1 jet, it can reach high values ($\sim$60%) at high jet b-ness thresholds. This effect translates in larger uncertainties on the mistag rates when the b-ness threshold increases because of both the small number of available non-b-jets and the high b-jet contamination, which are affected by large uncertainties. On the other hand the b-jet purity on the $t\bar{t}$ sample is relatively large at any jet b-ness cut.

[1] We start from $s_e(b) = 1$, we solve for $m(b)$, then we calculate $s_m(b)$ and we solve for $e(b)$. We iterate the process few times until we reach stable values. The process converges very quickly.

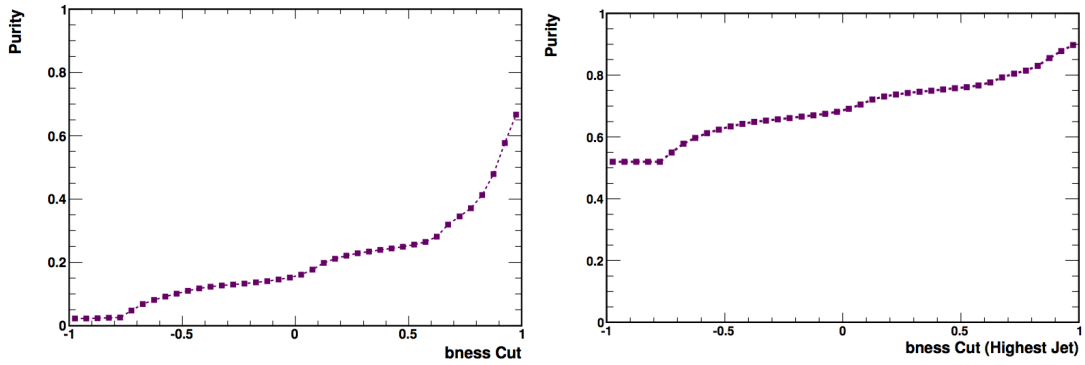
Z + 1 jet selection	$t\bar{t} \to l\nu$ + ≥4 jets selection
$N_l = 2$ (l being either electron or muon)	$N_l = 1$
Leptons with opposite charge	$P_T^l > 20$ GeV/c
Δz_0 between lepton < 5 cm	$\not{E}_T > 20$ GeV
$P_T^l > 20$ GeV/c	$\not{E}_T$ significance > 1 (3) for $l = \mu$ (e)
$75 < M_{ll}/\mathrm{GeV/c^2} < 105$	$M_T^W > 28$ GeV/c^2
$\not{E}_T < 25$ GeV	$E_T > 20$ GeV of the two highest b-ness jets
$P_T^{ll} > 10$ GeV/c	N_{jets} ($E_T > 15$ GeV) ≥ 4
N_{jets} ($E_T > 10$ GeV) $= 1$	Total sum $E_T > 300$ GeV
Jet $E_T > 20$ GeV, $\lvert\eta\rvert < 2.0$	

Table 5.3: Criteria to select the Z+1 jet and $t\bar{t} \to l\nu$ + ≥4 regions. The invariant mass of the $\not{E}_T$ and lepton system in the transverse plane is named M_T^W. See text for significance of symbols.

	Electrons	Muons
Z + 1 jet		
Data Events	9512	5575
MC Events	9640 ± 880	5540 ± 490
$t\bar{t} \to l\nu$ + ≥ 4 jets		
Data Events	507	835
MC Events	542 ± 56	862 ± 85

Table 5.4: Number of events in data and MC for the Z + 1 jet and $t\bar{t} \to l\nu$ + ≥ 4 jets regions. Numbers are listed according to the used trigger. Predictions are corrected for trigger efficiency and lepton identification scale factors. Listed uncertainties on MC are both statistical and systematic. Overall the agreement in number of events is good. Further details at [91].

With the inputs from the Z + 1 jet and $t\bar{t}$ regions we are able to solve Eq. 5.1 and 5.2 in the data and MC. In Fig. 5.6, 5.7 we show the results along with the uncertainties in the data. The uncertainty on the mistag rate is computed as follows:



(a) (b)

Figure 5.5: *Fractional number of b-jets in the Z+jet (left) and $t\bar{t} \to l\nu + \geq 4$ jets (right) samples as a function of the jet b-ness cut. Plots are made in ALPGEN+PYTHIA and PYTHIA MC respectively. A b-jet is identified when a generated b-quark is found in the jet cone.*

$$\sigma_m^2(b) = \frac{1}{(1-f_B)^2}\Big[\frac{m_{raw}(b)(1-m_{raw})}{N} + (\sigma_e(b)f_B)^2 + (\sigma_{f_B}[s_e(b)e(b) - m(b)])^2\Big] \tag{5.3}$$

where the first term is a binomial uncertainty on the raw mistag rate of the sample and depends on the statistical power of the chosen sample (N); the second term comes from the uncertainty on the measured b-tag efficiency (Eq. 5.5) and depends on the contamination of b-jets ($\propto f_B$) of the chosen sample; the third term is due to the uncertainty on f_B ($\sigma_{f_B} = 20\%$).

Similarly, we can write the expression for the uncertainty on the b-tag efficiency as follows [2]:

$$\sigma_e^2(b) = \frac{1}{(1-f_L)^2}\Big(\frac{e_{raw}(1-e_{raw})}{N} + (\sigma_m f_L)^2\Big) + + \sum_X \frac{\sigma_X^2}{[N(1-f_L)]^2} \times [(e+s_m)(f_L - f_L^X) + f_B^X(e_{MC} - e_X)]^2 \tag{5.4}$$

where

$$f_L = \sum_X \frac{f_L^X \cdot N_X}{N} \tag{5.5}$$

where X indicates the most important processes contributing in the $t\bar{t}$ samples:

[2] σ_e^2 is calculated from the uncertainty on $e - e_{MC}$. e_{MC} is affected by negligible uncertainties, since it is derived from the MC.

W+non-b-jets ($\sim 38\%$), W+b-jets ($\sim 4\%$), $t\bar{t}$ ($\sim 58\%$). Each term carries the same meaning as in Eq. 5.4, although the last term is slightly more complicated since it reflects the relative uncertainty on the fraction of each contributing process. Those uncertainties are assigned as follows: $\sigma_{W+b-jets} = 20\%$, $\sigma_{W+\text{non } b-jets} = 8.72\%$, $\sigma_{t\bar{t}} = 6.78\%$, based on a fit to the distribution of the sum of the two highest b-ness jets in the selected $t\bar{t}$ sample.

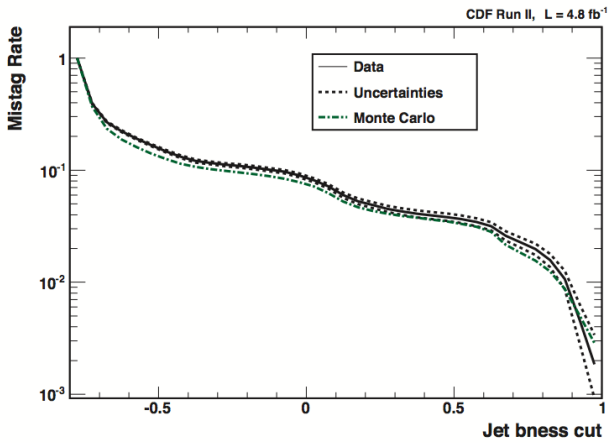


Figure 5.6: *The mistag rate in data as a function of the jet b-ness (solid black line). The dashed lines define the uncertainty band) and in simulation (solid green line) as a function of the jet b-ness. Since the simulation slightly under-predicts the mistag rate measured in data, a correction will be applied to the MC predictions.*

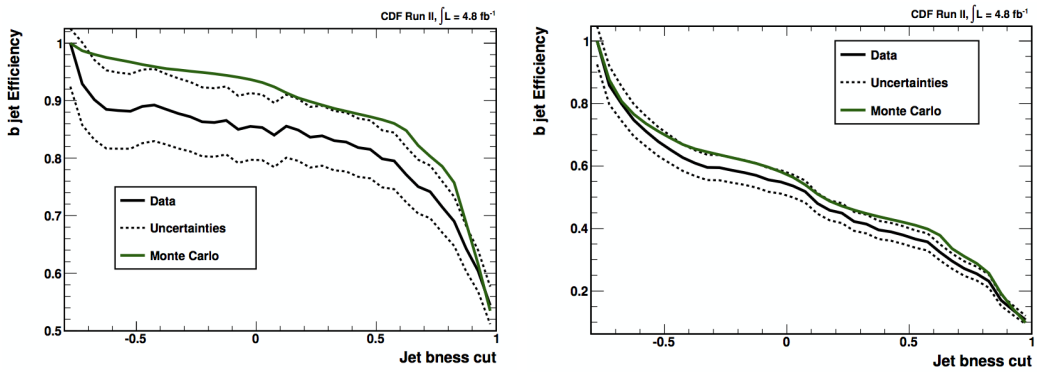


Figure 5.7: *The efficiency of the jet b-ness cut in data (solid black line, dashed lines represent the uncertainty bands) and Monte Carlo (solid green line) as a function of the cut on jet b-ness for the highest (left) and 2nd highest (right) b-ness jets in an event. Our simulation over-predicts the efficiency measured in data, and thus needs to be corrected for.*

From Fig. 5.6 (5.7) we observe that the MC simulation under (over)-predict the mistag rate (b-tag efficiency) measured in data, and thus must be corrected by applying appropriate scale factors. In Fig 5.8 we plot $s_m(b) - 1$ as a function of the jet b-ness. In Fig 5.9 we plot $s_e(b) - 1$ as a function of the b-ness of the highest and 2nd-highest jets. The plotted quantities represent the relative correction factors which will be applied in non-b-jets and b-jets MC respectively. In both cases the matching in direction ($\Delta R < 0.4$) between jets and generated partons will be used to decide the jet flavor.

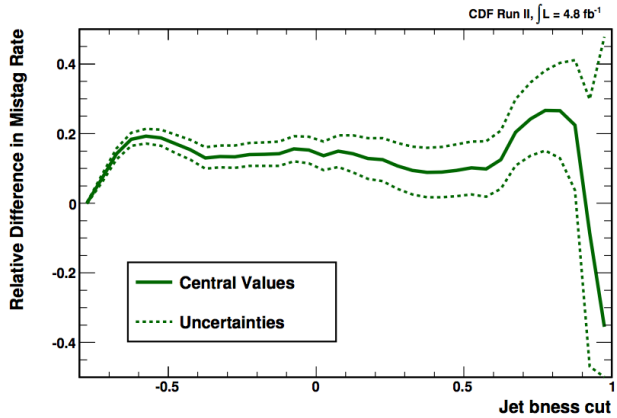


Figure 5.8: *The calculated MC scale factor on the mistag rate (solid line) and its uncertainty (dashed lines) relative to the mistag rate in the Monte Carlo as a function of the cut on jet b-ness.*

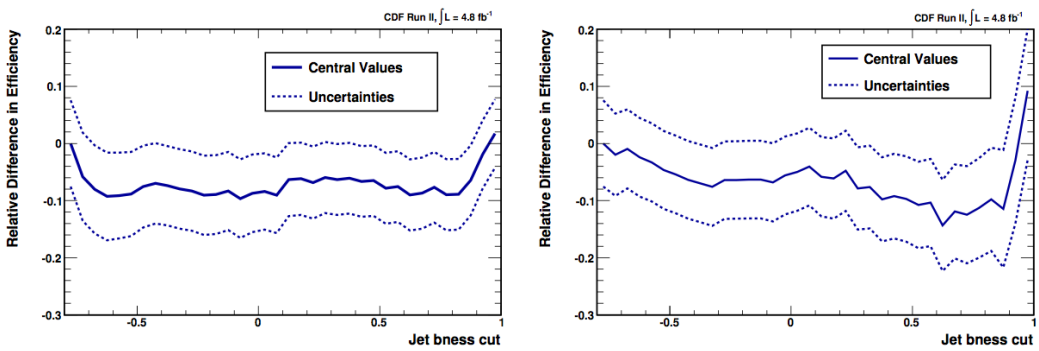


Figure 5.9: *The relative difference in efficiency between data and Monte Carlo (center solid line) and its uncertainty (dashed lines) relative to the efficiency in the Monte Carlo as a function of the cut on jet b-ness for the highest (left) and 2nd highest (right) bness jets in an event.*

Chapter 6

Jet Energy Corrections

Jets of particles resulting from the showering and fragmentation of partons are observed as clusters of calorimetric towers (Sec. 4.4). Because of detectors inefficiencies and physical effects the measured jet energy may deviate from the primary parton energy. In order to account for these effects a transfer function, parameterized as a function of jet transverse energy and pseudo-rapidity, is built and applied to both data and MonteCarlo. This is a standard CDF procedure. When checking the balancing of the corrected jets against the γ and Z bosons in the γ+jet and Z+jet samples, a significant discrepancy between data and simulations is found. Under the assumption that this discrepancy originates from differences in the modeling of the jet response for quarks and gluons, we derive specific corrections for quark and gluon jets. These corrections, which I helped deriving during my Ph.D period, are applied only to MonteCarlo.

In Sec. 6.1 we describe the CDF calorimeter simulation. All CDF jet corrections are described in Sec. 6.2.1-6.2.3. The personalized corrections for quark and gluon jets are described in Sec. 6.2.4.

6.1 Calorimeter Simulation

In this section the simulation of the response of the CDF calorimeter is described. The matrix element of SM processes at leading order, the underlying event, parton showering and fragmentation are computed and simulated with PYTHIA. The modeling of the fragmentation and parton showering in PYTHIA is tuned primarily to e^+e^- data. For the underlying event we use a tuning optimized to describe CDF Run I data (Chap. 7).

The hadronic shower development is simulated using the measured calorimeter response to isolated charged particles, to particles of minimum bias events, and to charged pions in a test beam. In these samples tracks are extrapolated to the position of the CES or PES detectors taking into account their bending in the magnetic field. Figure 6.1 shows the comparison in the central calorimeter between the mean value of the E/p distributions, $< E/p >$, between data and MC as a function of incident particle momenta.

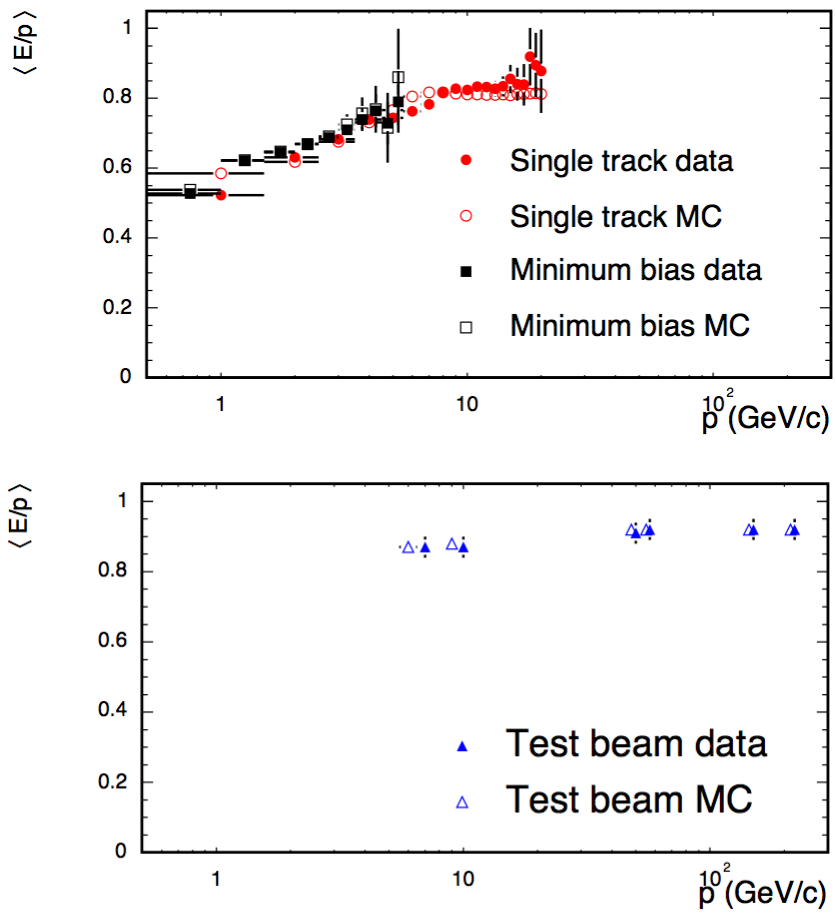


Figure 6.1: *$< E/p >$ in the central calorimeter versus incident particle momentum. In the top plot the minimum bias (circle) and single track data (squared) are compared to the simulation (open symbols). In the bottom plot the test beam data (closed triangles) are compared to the simulation (open triangles).*

Similar studies were performed for the plug calorimeters: they are documented in [87].

The same techniques were used to understand the calorimeter response to elec-

tromagnetic particles. Comparisons between data and simulation in $W \to e^{\pm}\nu_e$ and $J/\psi \to e^+e^-$ datasets are shown in Fig. 6.2. The simulation reproduces the data to better than 1% accuracy.

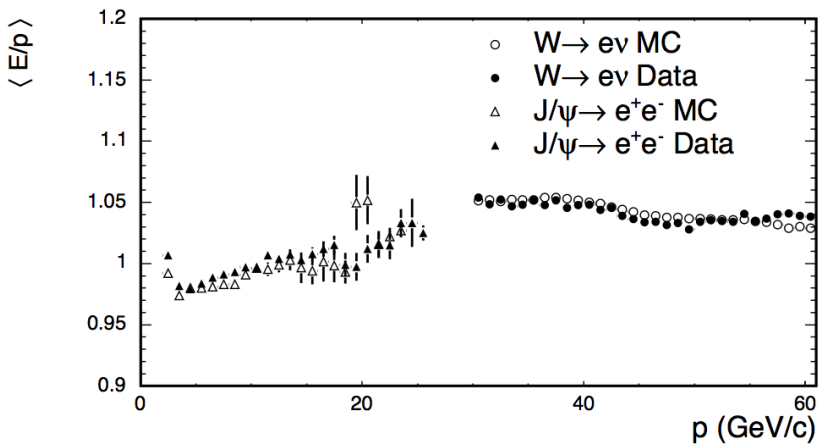


Figure 6.2: $< E/p >$ *versus incident particles momentum for electrons and positrons from* $W \to e^{\pm}\nu_e$ *and* $J/\psi \to e^+e^-$ *data (closed triangles and circles) and MC (open triangles and circles).*

Differences in the calorimeter responses to charged hadrons or electrons and positrons between data and MC are used to estimate the systematic uncertainty on the calorimeter energy scale. A summary of the relative uncertainties is reported in Table 6.1.

p (GeV/c)	0-12	12-20	>20
Hadrons (%)	2.5	3.0	4.0
electromagnetic particles (%)	1.7	1.7	1.7

Table 6.1: Overall relative uncertainties on the modeling of the calorimeter response to charged hadron and electromagnetic showers. Uncertainties on the charged hadron response for $p > 20$ GeV/c and $p < 20$ GeV/c were obtained from studies of isolated tracks in high-P_T events and of minimum bias samples, respectively. The breakdown of the individual uncertainties is described in [87].

6.2 Jet Energy Corrections

Jet energy corrections have been developed in order to correct separately for each source of energy calibration systematic error. These correction factors are applied

to the raw transverse energy E_T^{raw} as follows:

$$E_T(E_T^{raw}, \eta) = (E_T^{raw} \cdot L1(E_T^{raw}, \eta) - L4) \cdot L5(E_T^{raw}) \tag{6.1}$$

where $L1$, $L4$, $L5$ are the "correction level" factors. We briefly summarize below the bias which these factors correct for, while we provide more details about the corrections in the next sections.

- $L1$: non-uniformities in calorimeter response along η.
- $L4$: extra activity in the calorimeter due to additional interactions ("pile-up") which may contribute additional energy within the jet cone.
- $L5$: non-linearity response of the calorimeter. After this correction the jet energy becomes the energy of the physical particles belonging to the jet (the calorimeter jet is "corrected to the particle level").

Two more levels of correction were developed at CDF: $L6$ and $L7$. The goals of those corrections are:

- remove the additional energy from particles created by the soft interaction of spectator partons [1] ("Underlying event");
- add the jet energy leakage outside the jet cone ("Out-of-cone").

Since no appreciable improvement in the di-jet mass resolution is observed when applying the L6 and L7 corrections, they are not used in the analysis described in this thesis.

6.2.1 η-dependent correction (L1)

The energy calibrations described in Sec. 4.4 shows that the response of the calorimeter is not uniform in pseudo-rapidity. The η-dependent correction are meant to flatten the η-dependence of the calorimeter response. This correction is obtained by studying the P_T-balancing in di-jet events. Di-jet events are selected requesting one jet ("trigger jet") in the $0.2 < \eta < 0.6$ region [2], and another jet ("probe jet") free to span over the $\eta < 3$ region (Fig. 6.3). Some additional cuts are also applied:

[1] Spectator partons are the partons within $p/\overline{p}$ which were not involved in the hard interaction

[2] This region is away from detector non-instrumented regions and has a response that can be accurately calibrated.

- jets are required to be back to back in the azimuthal plane: $\Delta\phi(jet_{probe}, jet_{trigger}) > 2.7$;
- the P_T of any additional jet must not exceed a given threshold (E_T^{max,j_3}). Depending on the used trigger path, E_T^{max,j_3} is varied between 7 and 10 GeV/c. This cut depletes the QCD radiation in the event, therefore ensuring a better balancing of the two jets [3];
- other minor cuts, as listed in [97].

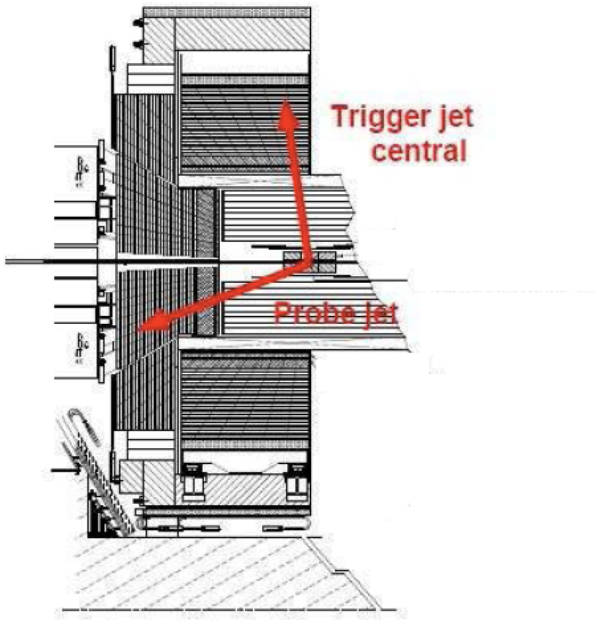


Figure 6.3: *Schematic drawing of the di-jet balancing.*

The transverse momentum (P_T) balancing fraction β_{di-jet} is defined as:

$$\beta_{di-jet} \equiv \frac{P_T^{probe}}{P_T^{trigger}} \tag{6.2}$$

In Fig. 6.4 β_{di-jet} distributions are shown as a function of η in a number of transverse momentum bins. From these plots it can be seen that $\beta_{di-jet} \sim 1$ for $0.2 < |\eta| < 0.6$ and $\beta_{di-jet} \sim 1.05 - 1.1$ for $1.2 < |\eta| < 2.4$. The lower calorimeter

[3] Different trigger paths correspond to different minimum E_T thresholds (E_T^{min,j_1}) on the most energetic jet in the event. E_T^{max,j_3} is increased as E_T^{min,j_1} is increased to allow for more events in the sample, without spoiling the balancing of the two jets

responses where the two halves of the central calorimeter join ($\eta \sim 0$), and where the plug and the central calorimeter are connected ($\eta \sim 1$) are also evident.

PYTHIA reproduces fairly well the data for $|\eta| < 1.4$. For $1.4 < |\eta| < 2.4$ the difference between the two is equal or less than 4%.

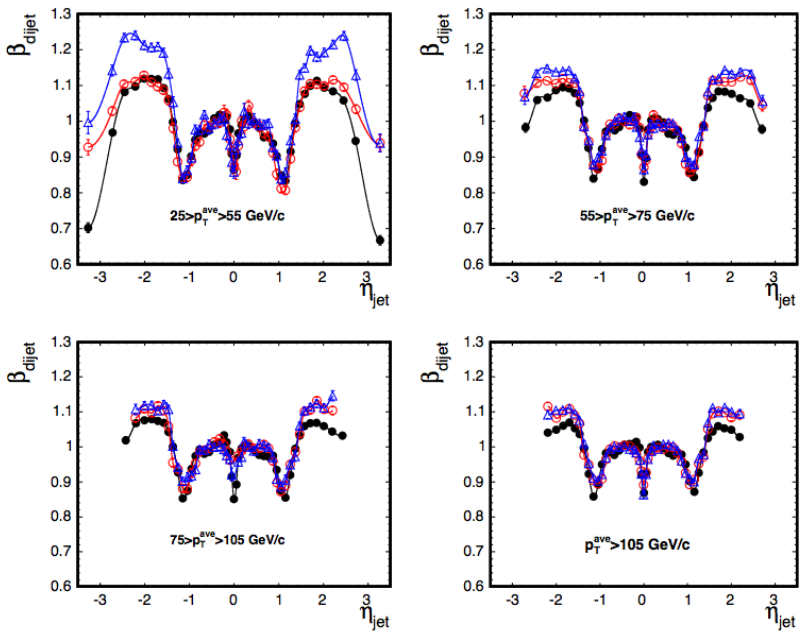


Figure 6.4: *Di-jet balance,* $\beta_{di-jet} \equiv \frac{P_T^{probe}}{P_T^{trigger}}$, *as a function of* η *in data (black circle), PYTHIA MC [93] (red circle), and HERWIG MC [98] (blue triangle). We show the corrections for* $25 < P_T^{ave} < 55$ *GeV/c,* $55 < P_T^{ave} < 75$ *GeV/c,* $75 < P_T^{ave} < 105$ *GeV/c,* $P_T^{ave} > 105$ *GeV/c,* $P_T^{ave} \equiv (P_T^{probe} + P_T^{trigger})/2$. *The lines show the interpolation between the individual measurements and are used to apply* $\eta-$*dependent continuous corrections.*

L1 corrections are defined as $1/\beta_{di-jet}$. They depend on η and P_T.

The uncertainties on this correction come from the limitations of the η and P_T parameterization, and from the observed β_{di-jet} differences when varying the jet sample selection cuts. Uncertainties are shown in Fig. 6.5.

6.2.2 Multiple Interaction Corrections (L4)

At high instantaneous luminosity more than one $p\overline{p}$ collision may occur in the same bunch crossing at the Tevatron. The average number of $p\overline{p}$ collision depends linearly on the instantaneous luminosity, as shown in Fig. 6.6.

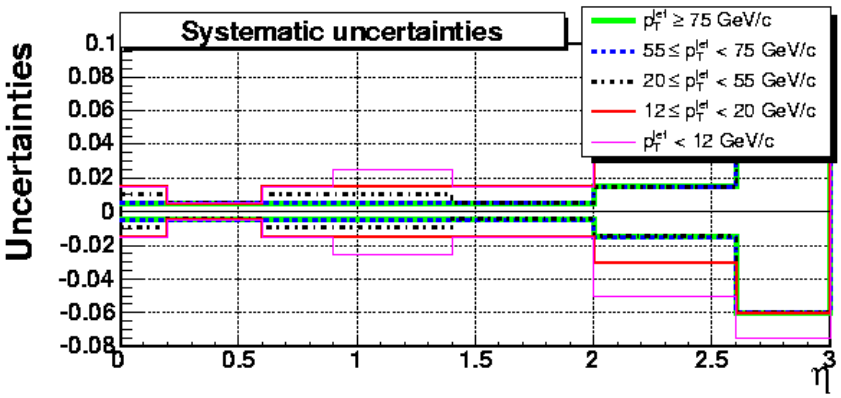


Figure 6.5: *Fractional systematic uncertainties on the L1 corrections for different P_T bins.*

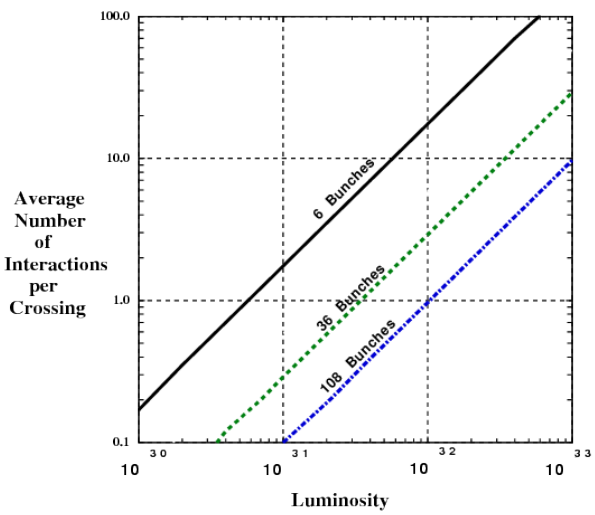


Figure 6.6: *Average number of interactions versus the instantaneous luminosity. As described in Chap. 2, 36 bunches are circulating in the Tevatron.*

Energy from additional interactions may fall into the jet cone and increase the apparent jet energy. This extra energy must be subtracted. In order to estimate the energy from additional interactions a minimum bias data sample is studied. The average transverse energy ($< E_T >$) in a R=0.4 cone is measured. The cone is defined using a seed tower randomly selected in the $0.2 < |\eta| < 0.6$ region. Then, such a transverse energy is parameterized as a linear function of the reconstructed vertices (Fig. 6.7): $< E_T >$ (GeV) $= 0.006 + 0.356 \times N_{vtx}$. The correction value for $N_{vtx} = 0$ is significantly larger than 0 because of inefficiencies in the vertex finding algorithm.

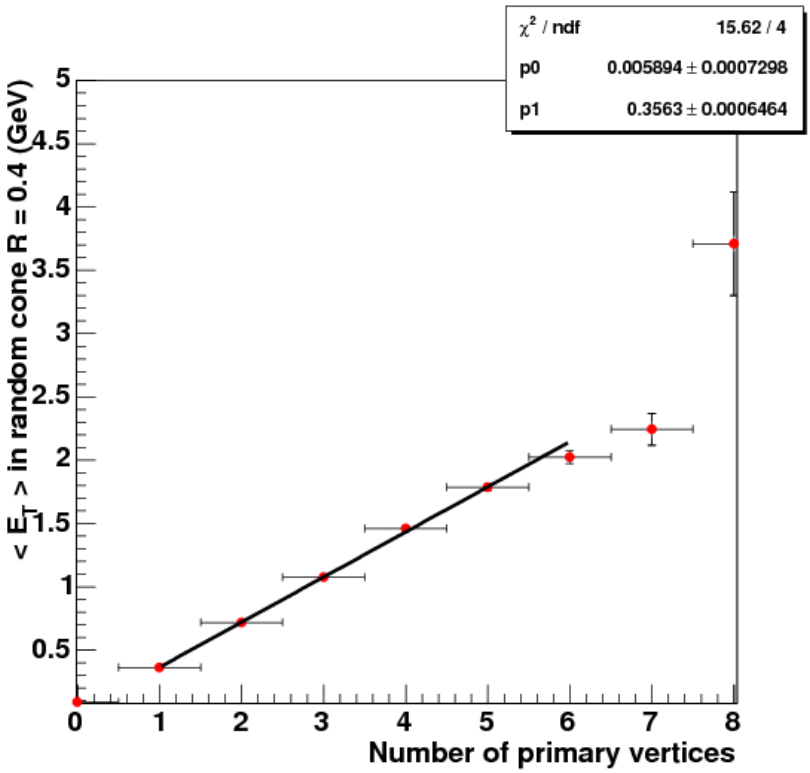


Figure 6.7: *Average E_T as a function of the number of vertices for R=0.4 cone jets in minimum bias events. Such a dependence has been parameterized with a linear function.*

From Fig. 6.7 one can deduce that the energy from additional interactions is about 400 MeV energy per vertex. That energy is subtracted from the jet energy.

This method heavily relies on the vertex finding efficiency. This efficiency, as well as the probability of finding a vertex where no interaction took place, depends on the topology of the event (e.g: number of tracks). The impact of the vertex finding

efficiency is studied by cross-checking the energy from additional interactions in different samples: the systematic uncertainty is assessed to be about 50 MeV (15% relative uncertainty) per additional interaction.

6.2.3 Absolute Jet Energy Scale (L5)

The absolute corrections aim to transform the jet energy from calorimeter level to particle level. After these corrections are applied, the jet energy is independent of the CDF detector.

The L5 corrections are derived on di-jet PYTHIA MC, where the calorimeter simulation has been optimized to reproduce the single particle response (Sec. 6.1). The procedure to estimate the L5 corrections is as follows.

- Jets are reconstructed at both calorimeter and particle levels by using JET-CLU with R=0.4 cone. Only stable particles [4] are used to reconstruct particle jets.
- Each calorimeter jet is associated to a particle-level jet if the ΔR between the two is less than 0.1.
- The probability to measure a calorimeter jet with P_T^{jet}, given a particle jet with $P_T^{particle}$ is parameterized by using a two Gaussian functions of $P_T^{jet} - P_T^{particle}$. Those functions describe the resolutions and tails. The assumed linear dependence on $P_T^{particle}$ of the parameters of the Gaussians are extracted by maximizing the logarithm of an un-binned likelihood function.
- The maximum P_T^{jet} value of this probability is used to extract L5 corrections.

Only jets in the $0.2 < |\eta| < 0.6$ are used. L5 corrections are shown in Fig. 6.8. At $P_T^{jet} = 25$ GeV/c corrections are about 1.30 of the uncorrected energy. The corrections decrease at high transverse momentum to an asymptotic value of 1.12.

Since L5 corrections are derived in MC, the extrapolation of these corrections depends upon how well the MC model the jet response in data. Three sources of systematic uncertainties are investigated:

- the difference in calorimeter response to single particles between data and MC (Sec. 6.1);

[4] Particles with a lifetime such that they can be interact in the calorimeter before decaying.

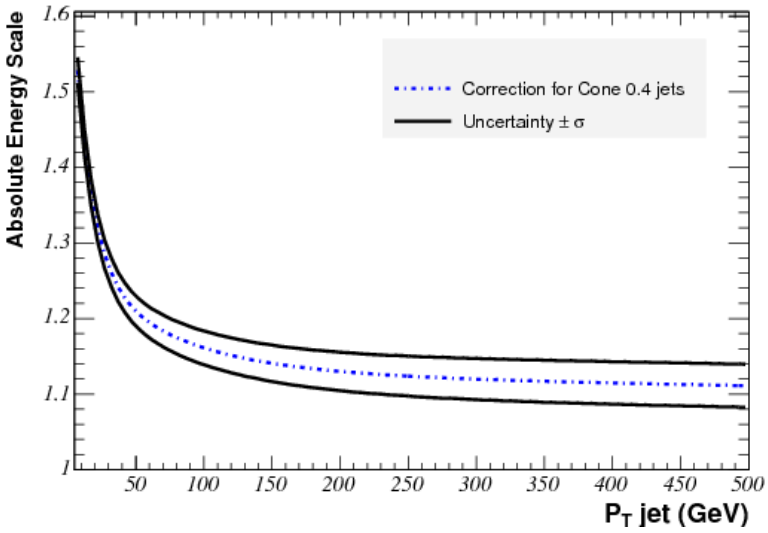


Figure 6.8: *Absolute corrections as a function of the jet transverse momentum. The uncertainties are also shown. More details about the uncertainties are given in Fig. 6.9*

- the differences in the momentum spectrum and particle multiplicity which may be due to inaccurate fragmentation models in our MC;
- the stability of the calorimeter calibrations in data: see Sec. 4.4.

Details on how the above systematic uncertainties are estimated are given elsewhere ([87]). A summary of the impact of the uncertainties on the overall L5 uncertainty is shown in Fig. 6.9.

6.2.4 Quark/Gluon Jet Energy Calibrations

The jet corrections described in the previous sections have been validated by looking at γ/Z+jet samples with a 200 pb^{-1} dataset ([87]). In those samples the balancing between the jet and the well measured boson is found to be well modeled by the MC. Another sample which is used for validating the jet energy corrections is the $t\bar{t}$ sample in the final state with a lepton, neutrino, two b-jets and two light-quark jets. When reconstructing the hadronically-decaying W, a good agreement in the jet energy scale between data and MC is observed by imposing the known W-mass value ([99]).

Because of the larger effective color charge carried by the gluons (Chap. 1), a gluon jet is expected to be broader and to have larger particle multiplicity than a

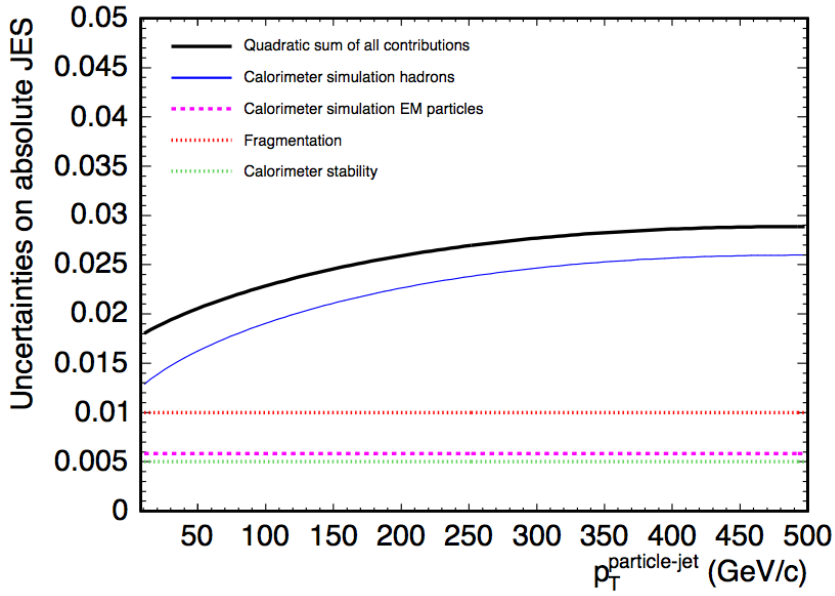


Figure 6.9: *Systematic uncertainty (solid line) on the absolute jet energy scale due to the calorimeter calibration and simulation. Colored lines are the individual contributions to the the overall uncertainty. See text for the significance of the symbols*

quark jet. These differences may not be visible in the samples used for validation either because of the limited statistics (Z+jet samples) or because of the large quark-jets component (γ+jet and $t\bar{t}$ samples). Therefore, additional investigations would be required to understand whether the above differences in the fragmentation of quark and gluon-jets may result in different calorimeter responses, which are currently not accounted by the existing systematic uncertainties on the jet energy scale.

Procedure to estimate the quark/gluon jet energy corrections

The goal is to compare the quark and gluon jet energy responses in data and MC. Given the balancing information ($K_\gamma = (jet\ E_T/\gamma\ E_T)-1$, $K_Z = (jet\ E_T/Z\ E_T)-1$) from two orthogonal samples (γ+jet and Z+jet) with different quark fractions ($F_Q^{\gamma/Z}$), it is possible to extract the quark and gluon jet response in data and MC (K_Q, K_G). K_Q and K_G can be found by solving the following system of equations:

$$\begin{aligned} K_Z &= F_Z^Q K_Q + [1 - F_Z^Q] K_G \\ K_\gamma &= F_\gamma^Q K_Q + [1 - F_\gamma^Q] K_G \end{aligned} \tag{6.3}$$

We solve the above system of equations by dividing the γ/Z+jet samples in a

number of jet E_T bins. By doing so, K_Q and K_G will be function of jet E_T.

Balancing in the Z+jet sample

The event selection for the Z+jet dataset is summarized in Table 6.2. In words, we require two leptons with $P_T > 20$ GeV/c from the Z decay. The lepton pair system must have a $P_T > 10$ Gev/c. No large missing transverse energy is allowed in the sample. The amount of extra activity is greatly reduced by requiring exclusively one jet with $E_T > 3$ GeV and that the reconstructed Z and the jet are back-to-back in the transverse plane.

Z+jet selection	
# vertices	> 0
# leptons	=2
lepton P_T	> 20 GeV/c
di-lepton Mass	[76,106] GeV/c^2
di-lepton P_T	> 10 GeV/c
$\not{E}_T$	< 20 GeV
# jets $E_T > 3$ GeV	=1
$\Delta\phi$ (di-lepton, jet)	>2.8 rad

Table 6.2: Event selection cuts for the Z+jet sample

In Figs. 6.10 to 6.13 the K_Z distributions are shown in a number of jet E_T bins. These distributions are fitted with Gaussian distributions in appropriate K_Z windows. The means and standard deviations of the Gaussian distributions will be used as average K_Z's and their uncertainties. The ALPGEN MC, showered and hadronized by PYTHIA is used to model the sample [5]. Significant shifts between the peaks in the simulations and data are observed for all jet E_T bins [6].

Balancing in the γ+jet sample

We use a dataset collected with the high-E_T photon trigger, which requires an isolated photon with $E_T > 25$ GeV. We apply the following offline selection:

[5] The contaminations of $t\bar{t}$, diboson, and fake leptons from W+jets and QCD multi-jets background are negligible.

[6] It can be seen that both data and simulations do not peak at zero. This is because a significant fraction of the jet energy is lost outside the cone. When using the L7 corrections (Sec. 6.2), jets appear to be better balanced to the Z's ([8]).

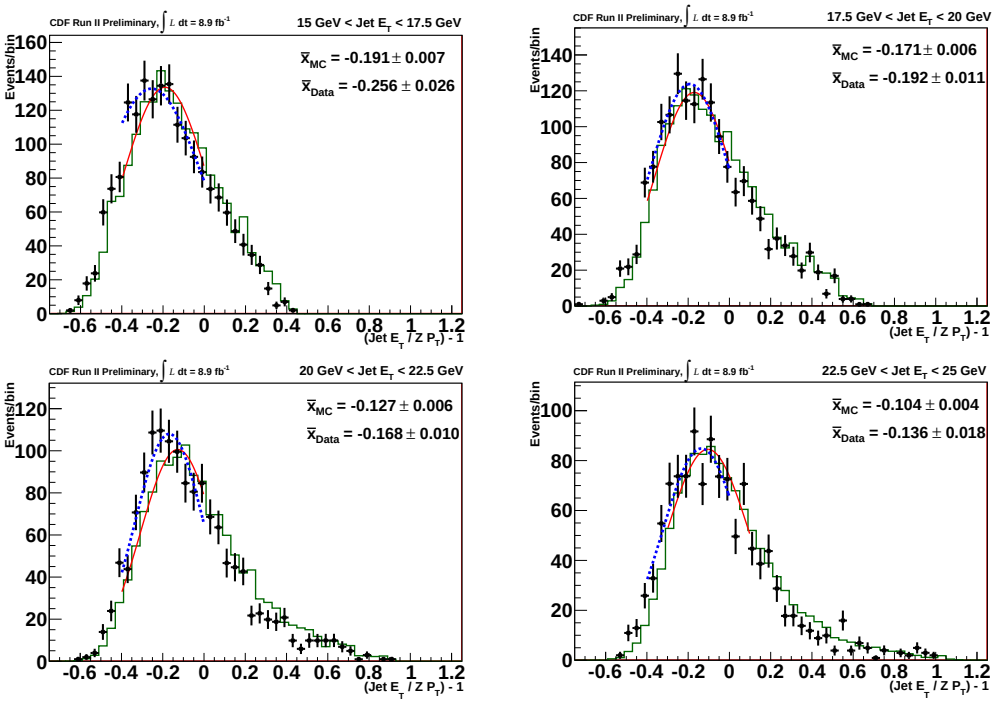


Figure 6.10: $K_Z \equiv (jet\ E_T/Z\ E_T) - 1$ *distributions in data (black) and MC (green) in the jet transverse energy range (GeV) 15-17.5 (upper left), 17.5-20.0 (upper right), 20.0-22.5 (bottom left), 22.5-25.0 (bottom right). Gaussian fitted distributions to data (blue) and MC (red) are shown.*

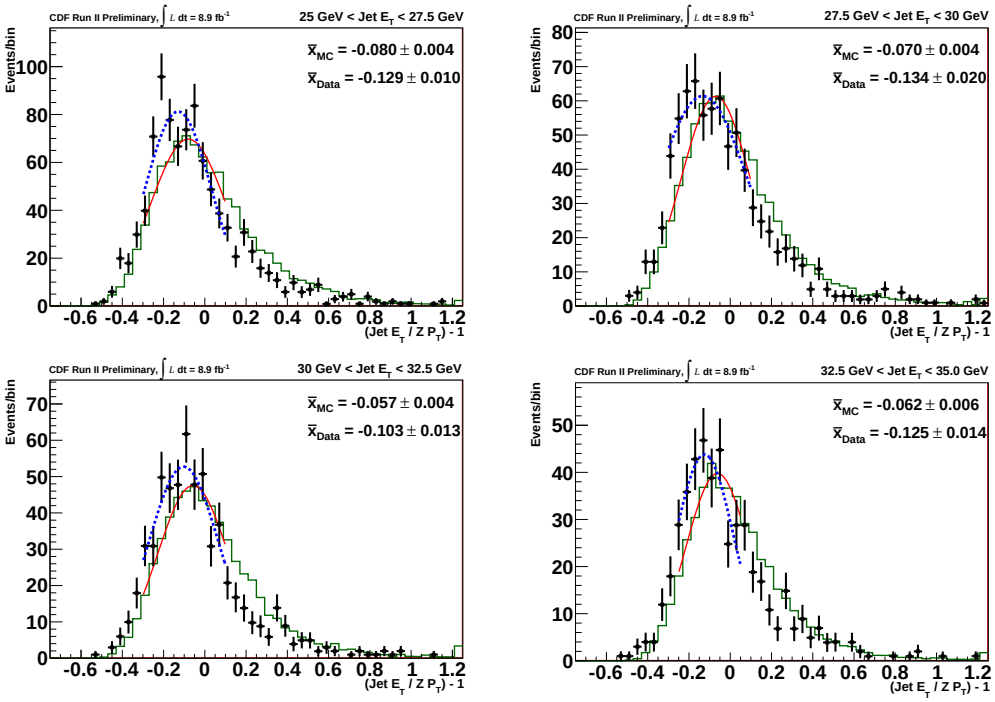


Figure 6.11: $K_Z \equiv (jet\ E_T/Z\ E_T) - 1$ *distributions in data (black) and MC (green) in the jet transverse energy range (GeV) 25-27.5 (upper left), 27.5-30.0 (upper right), 30.0-32.5 (bottom left), 32.5-35.0 (bottom right). Gaussian fitted distributions to data (blue) and MC (red) are shown.*

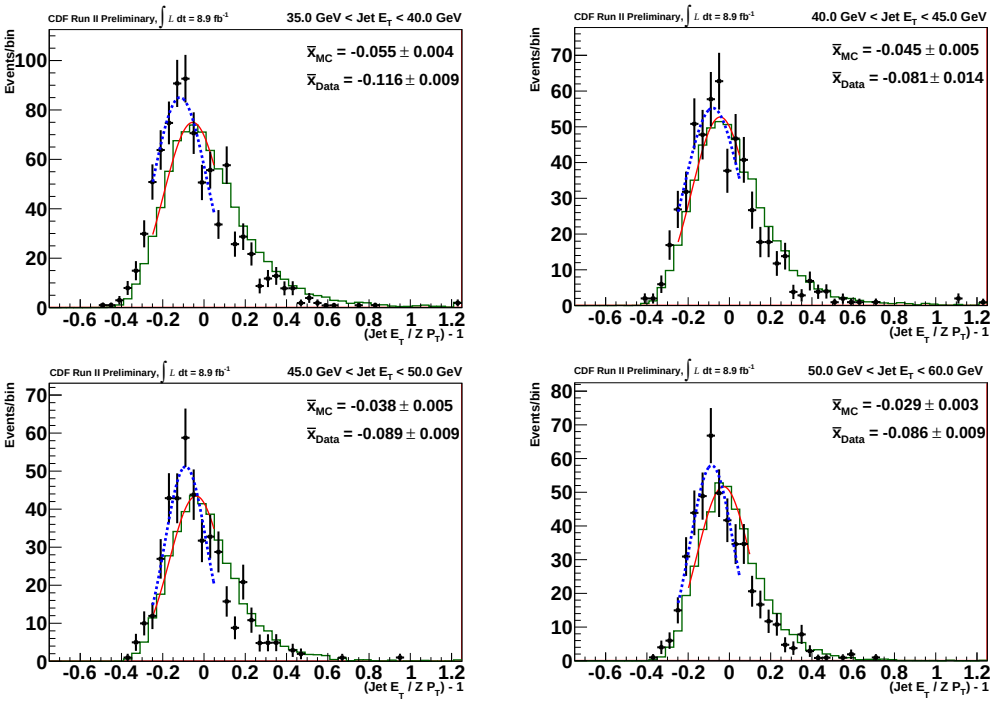


Figure 6.12: $K_Z \equiv (jet\ E_T/Z\ E_T) - 1$ *distributions in data (black) and MC (green) in the jet transverse energy range (GeV) 35.0-40.0 (upper left), 40.0-45.0 (upper right), 45.0-50.0 (bottom left), 50.0-60.0 (bottom right). Gaussian fitted distributions to data (blue) and MC (red) are shown.*

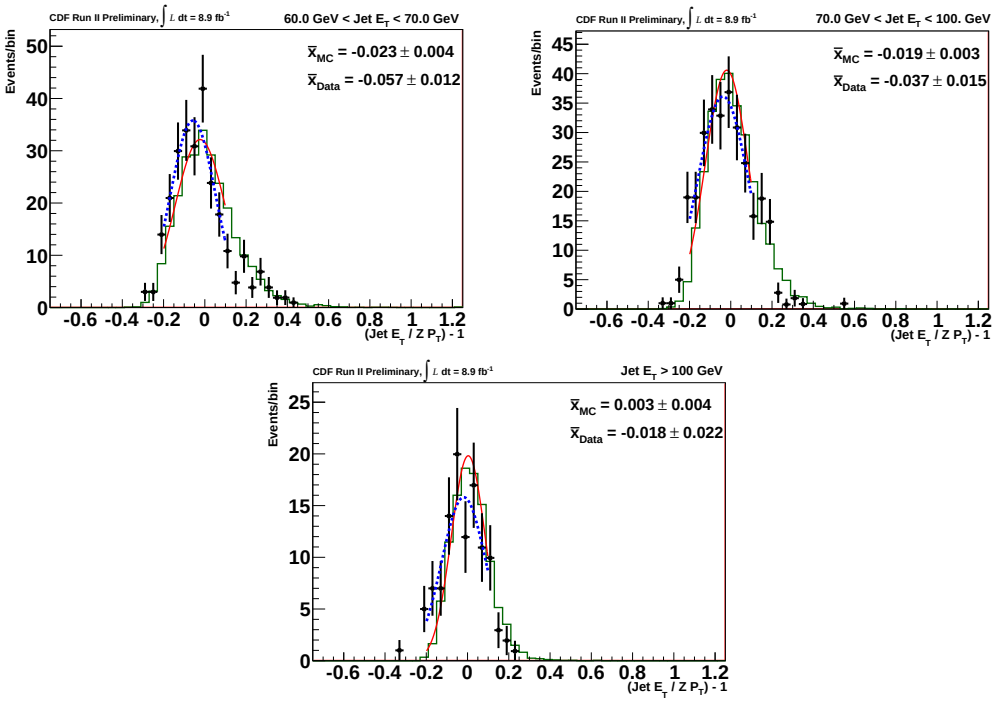


Figure 6.13: $K_Z \equiv (jet\ E_T/Z\ E_T) - 1$ *distributions in data (black) and MC (green) in the jet transverse energy range (GeV) 60.0-70.0 (upper left), 70.0-100.0 (upper right), >100.0 (bottom). Gaussian fitted distributions to data (blue) and MC (red) are shown.*

- one well identified photon candidate ("γ") [7] with γ $E_T > 27$ GeV and $0.2 < |\eta^\gamma| < 0.6$.
  - γ $E_T > 27$ GeV to avoid trigger biases;
  - $0.2 < |\eta^\gamma| < 0.6$ is required to make sure that the photon energy is well measured (the photon is far away from the un-instrumented regions of the CDF calorimeter);
  - "isolation transverse energy" is the amount of additional activity in a R=0.4 cone around the photon direction. This extra activity is either measured by the calorimeter or by the tracker. In order for the photon to be isolated the extra activity have be less than 1 GeV in the calorimeter and less than 2 GeV/c in the tracker.
- Exclusively one jet with $E_T > 3$ GeV and $|\eta| < 2.4$.
- $\delta\phi(\gamma, jet) > 3.0$.
- $\not{E}_T/E_T^\gamma < 0.8$.
- Exclusively one reconstructed primary vertex
  - the above four cuts are meant to reduce the amount of extra activity in the event.

All the selection cuts are summarized in Table 6.3.

γ+jet selection	
# vertices	$= 1$
E_T^γ	> 27 GeV/c
$\lvert\eta\rvert^\gamma$	[0.2,0.6]
$\not{E}_T/E_T^\gamma$	< 0.8
# jets $E_T > 3$ GeV	=1
$\Delta\phi(\gamma, jet)$	>3.0 rad

Table 6.3: Event selection cuts for the γ+jet sample

In Figs. 6.14 to 6.16 the K_γ distributions are shown in a number of jet E_T bins [8]. These distributions are fitted with Gaussian distributions as described earlier

[7] Photon identification cuts are described in [100].

[8] It is evident that for low jet energies the $E_T^\gamma > 27$ GeV/c cut is sculpting the distributions at high K_γ values.

for K_Z. The PYTHIA MC is used to model the γ+jet sample. The peaks in the simulations and data agree within the statistical uncertainties throughout [9].

The above studies were performed by assuming that the only contribution to the selected dataset comes from prompt photons. However, it is possible that photons are produced during the parton hadronization (e.g: photons from π^0 or η decays) in di-jet events. The background events will contaminate the selected sample and populate the region at high K_γ. Therefore, we do not expect variations in the fitted Gaussian means once the background model is included. Nonetheless, we estimated from simulations that the effect of this background is negligible and does not bias our results. These studies are described in [101].

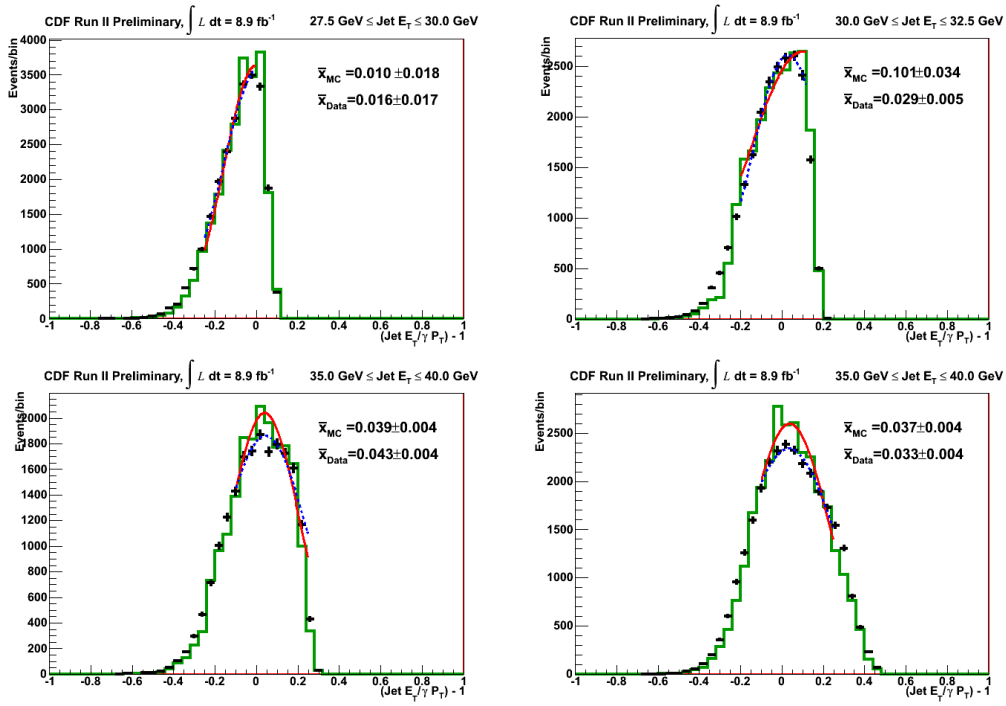


Figure 6.14: *$K_\gamma \equiv$ (jet E_T/γ E_T)-1 distributions in data (black) and MC (green) in the jet transverse energy range (GeV) 27.5-30.0 (upper left), 30.0-32.5 (upper right), 32.5-35.0 (bottom left), 35.0-40.0 (bottom right). Gaussian fitted distributions to data (blue) and MC (red) are shown.*

[9] As opposed to the Z+jet, the fitted Gaussian means in the γ+jet are consistent with zero. This is interpreted as being due to the fact that γ+jet is largely dominated by quark-jets (Fig. 6.18), which are expected to be collimated (Chap. 1): the energy lost outside the cone is negligible.

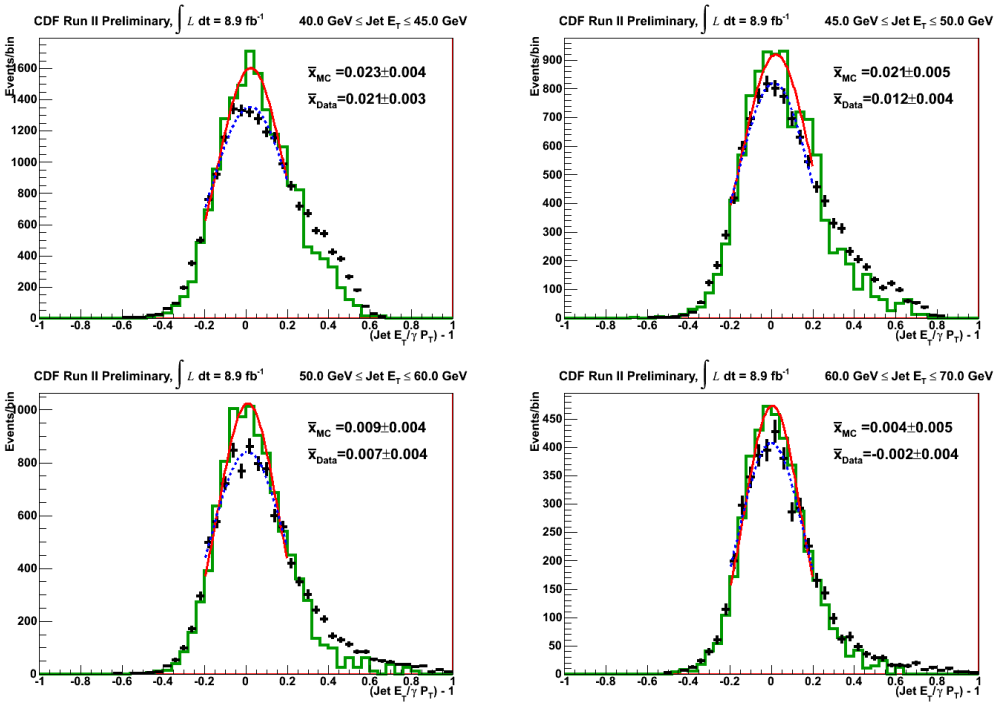


Figure 6.15: $K_\gamma \equiv$ (*jet* E_T/γ E_T)*-1 distributions in data (black) and MC (green) in the jet transverse energy range (GeV) 40.0-45.0 (upper left), 45.0-50.0 (upper right), 50.0-60.0 (bottom left), 60.0-70.0 (bottom right). Gaussian fitted distributions to data (blue) and MC (red) are shown.*

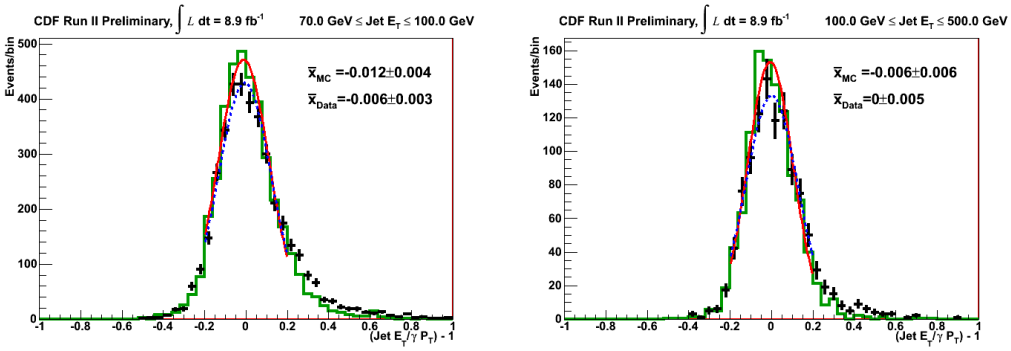


Figure 6.16: $K_\gamma \equiv$ (*jet* E_T/γ E_T)*-1 distributions in data (black) and MC (green) in the jet transverse energy range (GeV) 70.0-100.0 (left), and for jet* $E_T > 100.0$ *GeV. Gaussian fitted distributions to data (blue) and MC (red) are shown.*

Extraction of the quark/gluon jet energy corrections

In Fig. 6.17 we combine the results obtained in the previous two sections. Since we want to investigate the compatibility between data and simulations, we plot their difference in the balancing in the γ+jet and Z+jet samples. A striking difference between the modeling in the γ+jet and Z+jet samples is visible: the MC models well the calorimeter response in the γ+jet sample, while the MC does not model well the data in the Z+jet sample. Such a disagreement in the Z+jet sample is not covered by the current systematic uncertainties (2.0%-2.5%, see Fig. 6.9) [10].

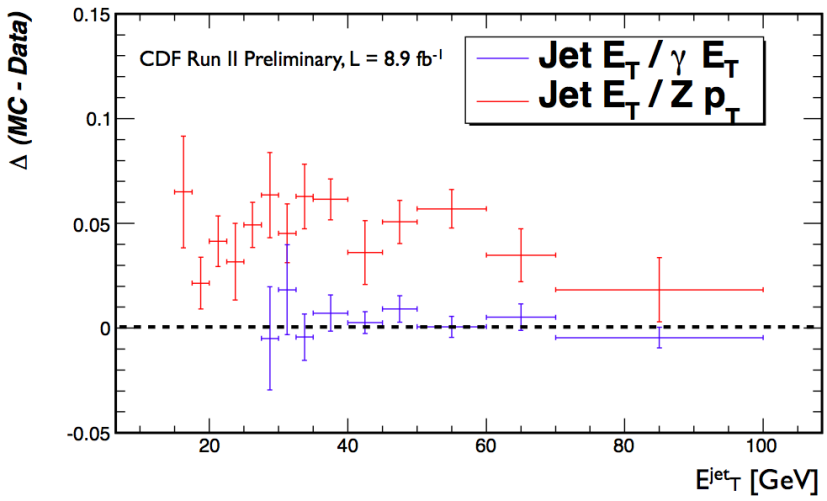


Figure 6.17: *Difference in the balancing between MC and data for the Z+jet (red) and γ+jet (blue) samples as a function of the jet E_T. The dashed black line means agreement between data and MC. It's clear that the simulations disagree with the Z+jet data while they agree with γ+jet data.*

One of the biggest differences in the samples used for the balancing is the nature of the outgoing partons which originate the jets. As shown in Fig. 6.18 and 6.19 MC predicts larger/smaller number of quark jets in the γ+jet/Z+jet. The different quark fractions in the two samples and the differences featured by jets originated by quarks and gluons (e.g: track multiplicity, different spacial distribution of the energy deposition [8]) could explain the observed differences in the modeling in the γ+jet and Z+jet samples. We will make the assumption that the quark fractions predicted

[10] Even if the current systematic uncertainties were larger and we could cover the observed discrepancy in the Z+jet by varying the jet energy response in MC, the same variation would spoil the modeling in γ+jet sample (and also in the $t\bar{t}$ sample ([99])). Therefore the modeling problem seen in the Z+jet sample cannot be resolved by simply saying that the jet energy scale uncertainties are underestimated.

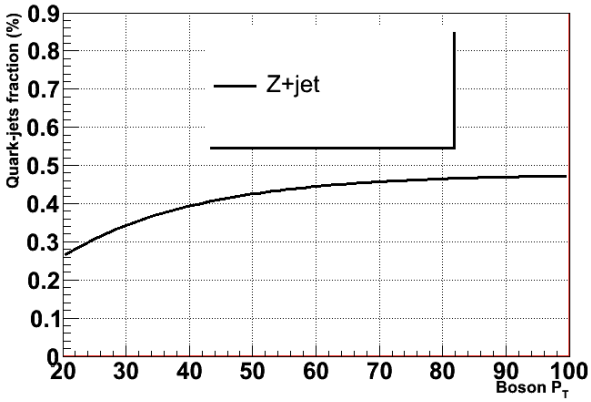


Figure 6.18: *Quark fraction in the Z+1 jet sample as a function of the Z P_T. Such a fraction is obtained from the MC, by looking for the highest-P_T generated parton within the jet cone. The relative number of jets with a quark as the highest-P_T parton is taken as to the quark fraction of the sample.*

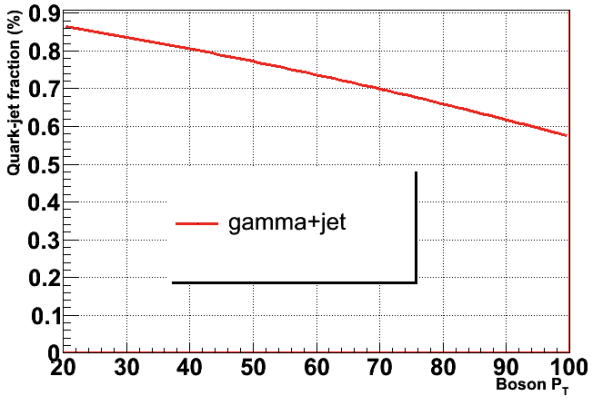


Figure 6.19: *Quark fraction in the γ+1 jet sample as a function of the γ E_T. Such a fraction is obtained from the MC, by looking for the highest-P_T generated parton within the jet cone. The relative number of jets with a quark as the highest-P_T parton is taken as to the quark fraction of the sample.*

by the MC are correct and the observed discrepancies between predictions and data originate from differences in the modeling of the jet response for quarks and gluons and we will derive specific corrections for quark and gluon jets. The Monte Carlo predicted quark and gluon fractions will be later certified by a comparison with data. The remaining uncertainties on the quark/gluon fractions will conservatively accounted for in the error analysis.

With $F^Q_{Z/\gamma}$ (Fig. 6.18, 6.19), $K_{Z/\gamma}$ (Fig. 6.10-6.13, 6.14-6.16) we can solve Eq. 6.3 for K_Q and K_G:

$$\begin{aligned} K_Q &= \frac{1}{F^Q_\gamma - F^Q_Z}[(1-F^Q_Z)K_\gamma - (1-F^Q_\gamma)K_Z] \\ K_G &= \frac{1}{F^Q_\gamma - F^Q_Z}[F^Q_\gamma K_Z - F^Q_Z K_\gamma] \end{aligned} \tag{6.4}$$

It should be noticed that, due to the photon trigger requirements, no accurate K_γ information are available for $E_T^{jet} < 27.5$ GeV. We decided to extrapolate down to $E_T^{jet} < 27.5$ GeV the extracted K_Q value obtained for $E_T^{jet} > 27.5$ GeV. By doing so we can solve Eq. 6.4 for all jet energies of interest (down to $E_T^{jet} > 15$ GeV). In Fig. 6.20 we show the corrections that need to be applied to quark and gluon jets in MC in order to match the data. Both corrections for quark and gluon jets in MC appear to be flat over the investigated jet energy range ($E_T^{jet} > 15$ GeV): the quark jet energy in MC has to be increased by $\sim (1.4 \pm 2.7)\%$, while the gluon jet energy has to be decreased by $\sim (7.9 \pm 4.4)\%$. The former value is in agreement with the nominal corrections (Sec. 6.2.3) and with the ones obtained in the $t\bar{t}$ sample ([99]). The small quark jets corrections are consequences of the previously shown agreement in the γ+jet sample, which is heavily quark dominated, while the disagreement shown in the Z+jet sample directly translates in large gluon-jet corrections.

Uncertainties on the quark/gluon jet energy corrections

The impact of each source of systematic uncertainty is evaluated by varying the relative central value within the given a-priori uncertainty and re-deriving the corrections, as described in the previous section. The other parameters in the equation 6.4 are kept equal to the nominal values. The difference between the corrections obtained with the varied parameter and the nominal one is the systematic uncertainty on the jet energy corrections associated to that particular source. This procedure

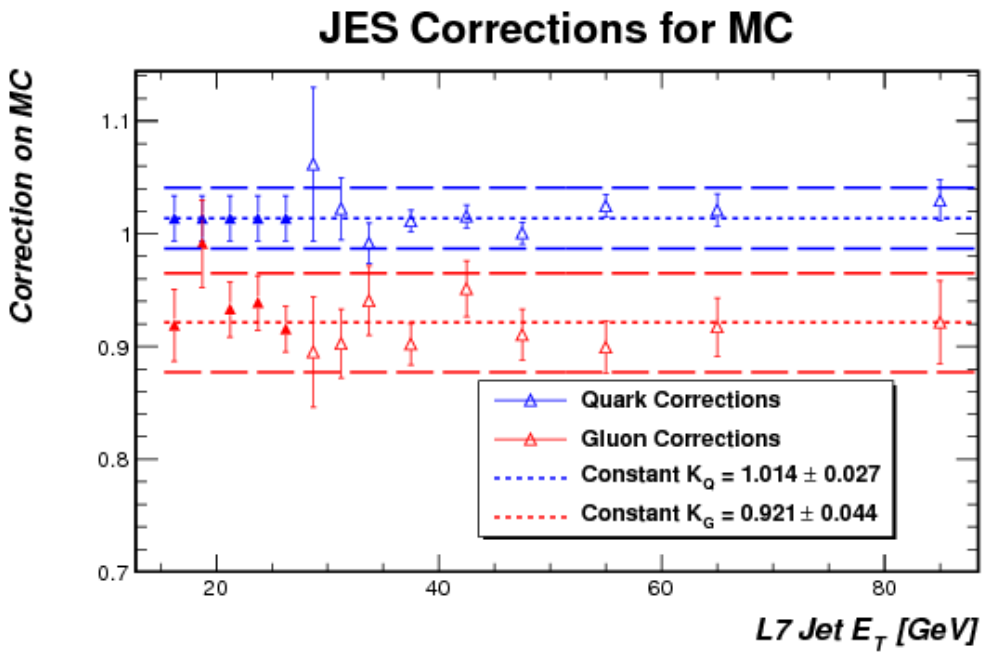


Figure 6.20: The derived corrections for simulated quark jets (*blue*) and gluon jets (*red*) as a function of jet E_T. The open triangles represent corrections derived by using the γ-jet and Z-jet balancing samples, while the filled triangles represent the assumed flat correction for quarks and the corresponding correction for gluons calculated from the Z-jet balancing sample alone. The error bars account only for the statistical uncertainty (Table 6.4). The short dashed lines are the fits of the correction to a constant across jet E_T, and the long dashed lines represent the total (statistical+systematic) uncertainty bands on that constant correction, further described in text.

is repeated for each source of systematic uncertainty. We list below the source of the known systematic uncertainties on the aforementioned MC jet energy scale corrections.

- *Fit/Statistical Uncertainty*: we use the standard deviations from the Gaussian fits which were used to extract the most probable K_γ, and K_Z values. These uncertainties contributes to $\pm 2\%$ for the quark-jet corrections and $\mp 2.5\%$ for gluon-jet corrections.

- F_Q^γ, F_Q^Z: we vary the MC-predicted quark fraction by an absolute value of 10% in an uncorrelated fashion in the γ+jet and Z+jet samples. The 10% variation is decided by fitting the quark fraction to the data. To do this fit, a multi-variate discriminant ([8]), which efficiently separates quark from gluon jets by looking at the jet shapes, is used. The quark fraction variation in the γ+jet sample translates to an uncertainty of $\pm 1.8\%$ on the quark jet corrections, and $\mp 2.7\%$ on the gluon jet corrections. When varying the quark fraction in the Z+jet sample, the uncertainties are respectively $\pm 0.6\%$ and $\mp 2.1\%$.

- *Extrapolation at low energy*: we vary the extrapolated K_Q value by $\pm 2\%$ in the $E_T^{jet} < 27.5$ GeV region and we observe that the the effect on gluon jet energy corrections is $\mp 0.4\%$.

- *Dependence on the number of interaction vertices*: the selection criteria of the γ+jet and the Z+jet samples differ on the requirement of the number of vertices (Tables 6.3, 6.2). We investigate possible biases due to such a difference, by requiring the Z+jet sample to have exactly one vertex as the γ+jet sample. The absolute shifts in quark-jet and gluon-jet corrections from the nominal values are 0.2% and 1.2% respectively: we conservatively use as systematic uncertainty $\pm 0.2\%$ for the quark-jet corrections and $\mp 1.2\%$ for the gluon-jet corrections.

The uncertainties described above together with their quadrature sum are summarized in Table 6.4. The estimated uncertainties on quark and gluon jet energy corrections are anti-correlated, as the corrected jet energy are required to match the data. This anti-correlation property was considered in this thesis for the first time at CDF.

Uncertainty	Quark-jet corrections	Gluon-jet corrections
Fit (Statistical)	0.020	0.025
F_Q^γ	0.018	0.027
F_Q^Z	0.006	0.021
Extrapolation at low energy	-	0.004
# vertex dependence	0.002	0.012
Total	0.027	0.044

Table 6.4: Summary of the statistical (first row) and systematic (other rows) uncertainties on the quark and gluon jet energy corrections applied in MC. The total uncertainties are the sum in quadrature of each contribution. The estimated uncertainties on quark and gluon jet energy corrections are anti-correlated.

Conclusions

We derived specific MC corrections for quark and gluon jets in order to match the boson versus jet balancing distributions in the γ+jet and Z+jet data samples. After the absolute correction (Sec. 6.2.3) MC quark-jet energies are increased by (1.4±2.7)%, while gluon-jet energies are decreased by (7.9±4.4)%. Quark and gluon jets are identified in MC by looking at the type of the highest-E_T parton within the jet cone. The MonteCarlo predicted fractions of quark and gluon jets in the two processes were certified by comparing with data samples, and the uncertainties determined accordingly.

Uncertainties on quark and gluon jet energies are of the same order of magnitude of L5 corrections (Sec. 6.2.3). We decided to use only the above uncertainties and to neglect the previously assessed systematic uncertainties (Sec. 6.2.1, 6.2.2, 6.2.3) for the following reasons:

- the new specific jet energy corrections account for possible discrepancy between data and MC. The goal of the previous systematic uncertainties was to cover those discrepancies;
- the cross section measurement is sensitive to relative differences in the jet energy scale between data and MC, rather than to the absolute jet energy scale value [11].

[11] We threw pseudo-experiments with jet energy scales varied within L5 uncertainties. The average extracted cross section was found to have a negligible dependence on the chosen jet energy scale value.

Chapter 7

Signal and Background Simulation

In high energy physics the large particle multiplicity makes events too complex to be modeled analytically. Moreover, the physics behind those events is not understood well enough to provide an exact description. This is why event generators were developed. By using standard MC techniques event generators provide an output which has the same form as the real data, including the same quantum mechanics fluctuations. Once available, generated events can be used for a number of purposes, such as for having a feeling on what to expect, for designing a new detector, and so on. In this chapter we give an overview of the event generation process (Sec. 7.1), and we introduce the signal and background processes which are expected to contribute in the dataset under investigation (Sec. 7.2).

7.1 Overview on Event generators

The event simulation is made practically feasible at hadron colliders by assuming "factorization". This means that the calculation of the cross section for a particular process can be decoupled into subsequent steps, which are easier to describe numerically. The generation is subdivided into the following three steps:

1. the initial state is described by parton distribution functions of general validity, giving the flux on the incoming partons;

2. the parton-parton interaction is described by the SM matrix-element of the hard elementary process,

3. including QED and QCD radiations. The ansatz is made that the physical process can be represented ("factorized") by 1. times 2. times 3.;

4. the simulation is completed by applying semi-phenomenological hadronization routines describing parton hadronization.

The generated physical events can be fed into the simulation of a detector to obtain a detailed description of the events in a particular experiment.

As opposed to lepton colliders, in hadron colliders the nature and energy of the interacting partons is unknown. However, the structure functions are considered process-independent. Therefore, the cross sections can be easily (in principle) computed.

7.1.1 Parton Distribution Function

In the factorization approximation, the cross section for a process $ij \to k$, i, j, k being generic particles is given by

$$\sigma_{ij\to k} = \int_0^1 dx_1 \int_0^1 dx_2 \; PDF_i^1(x_1, Q^2) \; PDF_j^2(x_2, Q^2) \; \hat{\sigma}_{ij\to k} \tag{7.1}$$

where $\hat{\sigma}$ is the cross section at parton level. $PDF(x, Q^2)$ is the parton distribution function, which describes the probability to find a parton carrying a x fraction of the beam energy when the beam hadron is probed by a hard scattering at virtuality scale Q^2. At Tevatron PDF's for quarks and gluons are being used. Based on the correlated information by a large set of experiments, PDF's are available for each quark/anti-quark flavor. Although some progress are being made in lattice QCD studies, PDF's are yet to be understood from first principles. Therefore, PDF's come from parameterizations obtained by fitting a number of datasets, used together with the expected evolution equations for the Q^2 dependence.

The PDF used in this analysis is the CTEQ5L [102], a leading order parameterization, although next-to-leading order PDF's exist in the market. The PDF's are evolved, which means that the inclusive effects of the initial state radiation have been already accounted for.

7.1.2 Matrix Element Calculations

The calculation of a hard scattering at hadron colliders is based on tree-level Feynman diagrams. Calculations at higher orders, which involve a combination of loop

graphs and the soft parts from the QED and QCD radiation, are very difficult and are slowly becoming available.

Which generator, or combination of generators, should better be used in the analysis of the CDF data has been a reason of concern and eventually a difficult choice. Depending on the specific analysis different generators have been used. Because of this, we include below some information and a number of comments on the generators used in the analysis described in this thesis.

PYTHIA

PYTHIA [93] contains a rich selection (~ 300) of different hard processes. They are often classified based on the number of final state objects. PYTHIA has been optimized for $2 \to 1$ and $2 \to 2$ processes. There is no generic treatment of processes with three or more particle in the final state. PYTHIA also provides treatment of processes at softer scale: QCD radiation via the parton showering (See sec. 7.1.3) and the fragmentation/hadronization (see Sec. 7.1.4).

The advantage of PYTHIA is its speed, the capability to generate a desired number of jets via the parton shower, and the versatility. The downside of this generator is the lack of color flow information and to some extent spin correlations (included for example by other generators, such as HERWIG [98] and MADGRAPH [103]) at the level of matrix element.

ALPGEN

The amount of radiation in events with jets produced in association with massive bosons is large. Calculations involving large number of partons in the final state are hard to compute, since the number of diagrams grows as the factorial of the number of objects in the final state. ALPGEN [95] is able to compute the complete tree-level matrix element by using the ALPHA algorithm [96], which only grows with the power of the number of particles in the final state. Moreover, ALPGEN includes the color flow information between jets by modeling the soft gluon radiation, via the prescription of angular ordering borrowed by the parton showers algorithms [95]. ALPGEN, designed specifically for processes whose final state contains a W or Z boson, can generate events with a fixed number of partons. As opposed to PYTHIA, it includes the b, c, and t masses, as well as helicity correlations in top quark and gauge boson decays. ALPGEN relies on other generators (PYTHIA is chosen at CDF) for the QED and QCD radiation and hadronization.

MADEVENT

MADEVENT [104] is a versatile tree-level generator which is powered by the matrix element generator MADGRAPH [103]. MADGRAPH generates the sub-processes amplitude and produces the mappings for the integration over the phase space. It can calculate diagrams with color and spin polarization information fully included. Since MADEVENT generates event only at parton level, in connection with this generator we use in CDF PYTHIA (via the interface PYXTRA) to simulate the QED and QCD radiation as well as the hadronization.

7.1.3 Modeling of the initial and final state radiation

In every process containing colored and/or charged objects in the initial or final state, the topology of the event may be modified from gluon (QCD) and/or photon (QED) radiation. These effects can still be treated perturbatively as much as they belong to a relatively hard phase of the process. Two approaches for modeling these corrections exist. The first one is the matrix element method. Since this method calculates the Feynman diagrams at all orders, it becomes formally very difficult because of the too many loop diagrams involved. In addition, from the point of view of computing time this method is in general prohibitive. The second method, which we will be using, is the so called parton shower method. Both initial and final state radiation can be viewed as parton showers with an arbitrary number of involved partons. The essential piece in the shower formalism is the parton branching $a \to bc$, specifically $q \to qg$, $g \to gg$ and $g \to q\bar{q}$ in case of QCD radiation and $e \to e\gamma$, $q \to q\gamma$ in case of QED radiation [1]. Each of the above branching scenario are parameterized by a splitting kernel $P_{a\to bc}(z)$, z being E_b/E_a where E is the energy ($1 - z = E_c/E_a$). $P_{a\to bc}(z)$ is also known as evolution or DGLAP equation [105]. From a chosen evolution ordering variable t (e.g: $t \propto m_a^2 = E_a^2 - \mathbf{p}_a^2$, $\mathbf{p}$ being the vector momentum) and from the DGLAP equation one can define the Sudakov form factor $\Delta(t)$, as the probability for a particle of evolving from t_0 to t without branching ([7]):

$$\Delta(t) \equiv exp[-\int_{t0}^{t} dt'/t' \int_0^1 dz \frac{\alpha_s}{2\pi} P(z)] \tag{7.2}$$

More technical details on the parton shower method are described in App. B.1.

[1] Further branchings, such as $\gamma \to e^+e^-$ and $\gamma \to q\bar{q}$ are not added because have not yet been of interest.

Choice of evolution ordering variable

The version of PYTHIA that we use for the W/Z+jets background simulation supports both ordering variables $t \propto m^2$ and $t \propto P_T^2$, m and P_T being the mass and transverse momentum of the branching parton. For the simulation of W/Z+jets major background we are using the transverse momentum-ordered showers. One of the main advantages of using P_T^2-ordered showers is that the coherent radiation in between two jets in automatically treated in a correct way, while the m^2-ordered showers can only model such a coherence in approximate way. The other known advantages are listed in [106].

Matching to the hard interaction: avoiding double counting extra-jets

A given $N+1$-jet event can be obtained either from a $N+1$-parton final state with collinear/soft radiation evolution, or by a N-parton final state with an extra jet from the shower. In this case a factorization prescription (also called "matching scheme") is required to understand which path to follow for avoiding a possible double counting [2]. As described in Sec. 7.1.3, the factorization scheme puts a boundary on where the matrix element calculations take over the parton showering. However, physics is independent of the separation between the two phases only if a control over the perturbative expansion is achieved. Therefore, exploring different matching schemes is a good sanity check on how we are modeling the data.

In a simplistic picture of the matching scheme the only problem is to constrain the parton shower in order not to produce extra jets. Two major matching schemes have been used for this analysis: CKKW [107] and MLM [95]. The former proceeds as follows:

- the phase space separation for multi-parton event is achieved by requiring that the k_T^{ij} (Eq. 7.3) of the two partons i and j is larger than a chosen critical value $k_{T,0}$;

$$k_T^{ij} = 2\ \min\{P_T^i, P_T^j\}^2\ [cosh(\eta^i - \eta^j) - cos(\phi^i - \phi^j)] \tag{7.3}$$

  In addition the transverse momentum of each jet has to be larger than $k_{T,0}$;

- a "parton shower history" is assigned to the multi-parton event: the computed matrix element is reweighted by the appropriate Sudakov factor (Sec. 7.1.3).

[2] Requirements for an optimal matching scheme are also to avoid "dead regions" (final state configurations with no initial path), and to guarantee the best possible approximation of the event kinematics.

- events are showered. Showers are not allowed at scales $k_T < k_{T,0}$.

The MLM approach ([95]) is completely different. Rather than smoothly merging the matrix element calculation with the parton shower by assigning a parton shower history, MLM acts after the parton shower has been generated. The MLM method works as follows:

- at matrix element level partons are generated with a minimum transverse energy threshold E_T^{min} and a minimum separation ΔR_{min};
- events are showered as described in Sec. 7.1.3;
- starting from the hardest parton (matrix element) a jet-to-parton matching is performed. A jet and a parton are considered matched if $\Delta R(jet, parton) < \Delta R_{min}$. The jet and the parton are removed from the list and the matching procedure continues;
- events with one or more not matched partons are vetoed (e.g: soft or collinear partons). Events with one or more not matched jet are also vetoed. On comparing with the CKKW approach, the latter veto is somehow equivalent to the Sudakov reweighting.

The distributions obtained from the two schemes should not depend on the choice of E_T^{min}, dR_{min} and $k_{T,0}$ respectively.

7.1.4 Fragmentation/Hadronization

After the parton shower has ended, a number of partons with virtualities (either squared masses or P_T^2) of order of the chosen t_0 [3] is created. After this point we enter in the confinement regime. In such a regime, colored partons are transformed in colorless hadrons. Such a process is called either fragmentation or hadronization. The fragmentation has yet to be understood from first principles. Therefore, only approximated phenomenological models exist. Three main types of model exist: string fragmentation (SF, [108]), independent fragmentation (IF, [109]), and cluster fragmentation (CF, [110]). Being phenomenological approximations, one cannot say which one is better a-priori. However, one can judge based on the representation of

[3] Physics should not depend on the t_0. In practice, when combining the parton shower and the hadronization a residual dependence on t_0 remains. For this reason t_0 is tuned to the data (the favored one is a few times Λ_{QCD}) rather than being chosen arbitrarily.

the existing data. We will briefly describe in App. B.2 the SF model, also known as the "Lund model", which is used throughout this analysis.

Decays

Hadrons produced during the fragmentation are often unstable and decay promptly into stable [4] particles that are detected in the experiment. Therefore, in order to model correctly this process it is vital to include the particles with their proper mass distribution and decay properties. Because of the limited knowledge on the composition of the final state the above task is everything but trivial. One assumes for simplicity that no dynamic is involved for modeling these decays [5].

7.1.5 Beam Remnants and Multiple Interactions

In the present measurement, simulating the hard scattering is not enough to understand the whole picture. The so called beam remnants and multiple interactions play a not negligible role. In order to understand the former, let us consider a proton beam, where a u-quark, responsible of the hard interaction, would leave behind a ud di-quark beam remnant. Such a di-quark is color connected to the hard interaction and is therefore considered part of the same fragmenting system. It is customary to assign an initial transverse momentum ("$k_\perp$") to the shower initiator (the u), whose recoil is assumed to be taken up by the beam remnants. This transverse momentum is chosen to be compatible with the expected quark motion inside the original hadron (i.e: from the uncertainty principle by the proton size). Often the remnant scheme is more complex ([93]).

Besides the interacting hadrons, other hadrons in the beam may interact. The most probable scenario for extra interaction to arise is because of separate $2 \to 2$ interactions. This mechanism, whose rate is expected to diverge at low exchanged momenta, is modeled by PYTHIA down to a chosen exchanged $p_{T,min}$. Below $p_{T,min}$ the created activity is not expected to be resolved by the available detectors. Therefore, the rate of these events is either zeroed or strongly damped. Phenomenologically, $p_{T,min}$ is $1.5 - 2.5$ GeV.

It should be noticed that the understanding of the multiple interaction is still

[4] Long-lived enough to be detected.

[5] This assumption is obviously inaccurate when, for example, a hadron is produced polarized, therefore leading to a non-isotropic distribution of its decay products. Many of these effects are not included in the program: they have to be reconstructed explicitly when needed.

very primitive. However, their rate has been measured in dedicated searches, and it is low enough for being disregarded as a background in the present measurement.

We refer to [93] for further details about the beam remnants and multiple interactions.

Data validation

Both the modeling of the beam remnant and the multiple interactions do not come from first principles. Therefore, it is mandatory to check if the chosen models can mirror the data.

Two datasets are used to validate the models. Datasets are selected as follows:

- at least one jet with $E_T > 15$ GeV and $|\eta| < 2$
- at least two jets with $E_T > 15$ GeV and $|\eta| < 2$; jets have to be back nearly back-to-back ($\Delta\phi > 150°$)

We refer to the former/latter sample as "leading jet/back-to-back" sample. The topological structure of these collisions is exploited to validate the modeling of the beam remnants and multiple interactions. We define the "TransMax" and "TransMin" as the regions orthogonal to the highest-E_T jet direction in the azimuthal plane ($60° < \Delta\phi < 120°$ and $60° < -\Delta\phi < 120°$ respectively, $\Delta\phi$ being the angle relative to highest-E_T jet direction) and spanning a $\Delta\eta = 2$ (see Fig. 7.1): they contain the largest and smallest numbers of charged particles. One expects that the TransMax region is invested by the hardest initial and final state, while both the TransMax and TransMin are sensitive to the beam remnant activity. In Fig. 7.2, 7.3 the density of the charged particles and of the charged P_T scalar sum are compared in the two samples between data and particular configurations of the PYTHIA and HERWIG [98] MC. As expected, the TransMin and TransMax regions behave quite differently. In the leading jet sample the rise as a function of jet P_T of both distributions in the TransMax region can be attributed to the initial and final state radiation, which is suppressed in the back-to-back sample. The TransMin shows a decrease with jet P_T in the two distributions. This interesting and unexpected effect may be due to a saturation of the multiple interactions at small impact parameters. The shown PYTHIA configuration ("Tune A") with slightly modifications is used in the analysis described in this thesis. PYTHIA describes fairly well the data in both the back-to-back and leading jet samples. We conclude that the model of the beam remnants, multiple interactions (and initial/final state radiation) is under control.

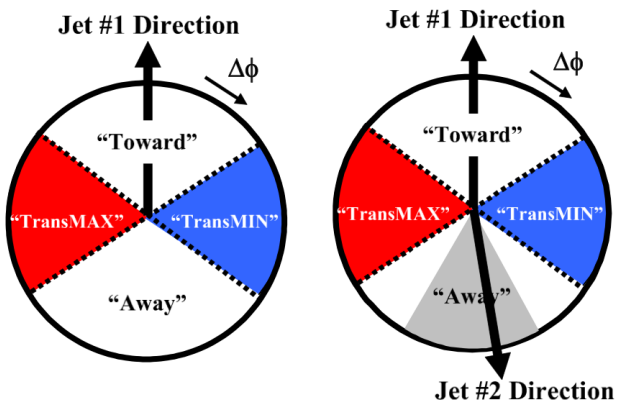


Figure 7.1: *Illustration of correlations in azimuthal angle relative to the direction of the highest-E_T (jet#1) in the event. The angle $\Delta\phi = \phi - \phi_{jet\#1}$ is the relative azimuthal angle between charged particles and the direction. "TransMax" and "TransMin" regions are defined by requiring $60° < \Delta\phi < 120°$ and $60° < -\Delta\phi < 120°$. TransMax and TransMin are defined to contain the largest and the smallest number of charged particles respectively. Both regions have an area $\Delta\eta\Delta\phi = 4\pi/6$.*

7.1.6 Detector Simulation

Once all the stable hadrons are generated, their propagation through the detector material is simulated using the GEANT3 package [111]. GEANT is a platform for handling:

- Geometry: physical layout of the experiment, including detectors, absorber and how the layout would affect the path of the particles.
- Tracking: mathematical models are used to simulate the passage of a particle through matter. Those models include particles interactions, which cause showering to secondary and tertiary particles, and decays.
- Detector response: simulate how the detector would respond as the particle passes through its volume.
- Detector conditions: the conditions of the detector for each time period can be inputted in order to mimic as close as possible the real data. The status of the different sub-detectors is taken into account.

The modeling of the particle interaction at the passage through matter is computationally intensive. Therefore the simulation of the calorimeter response is performed using parametrized response function tuned to the data. This is done by employing

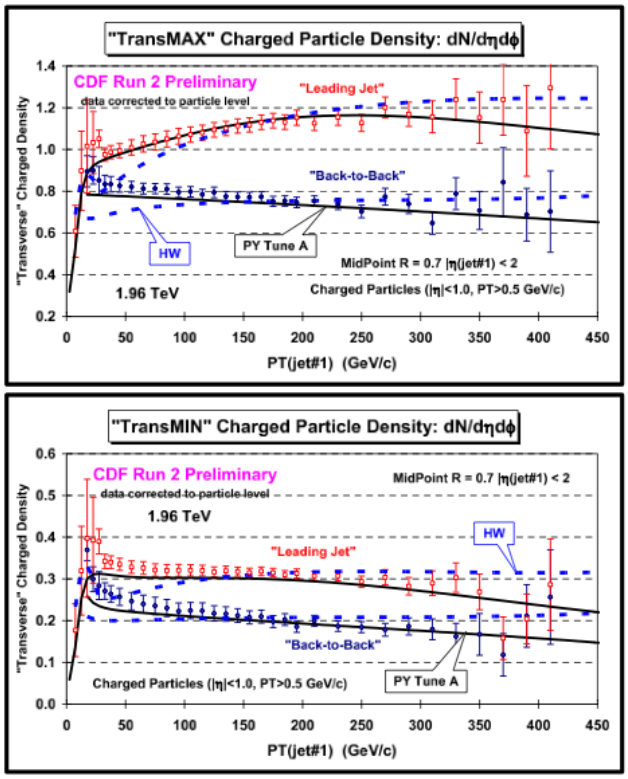


Figure 7.2: *Density of the charged particles* $dN_{chg}/d\phi d\eta$ *with* $p_T > 0.5$ *GeV/c and* $|\eta| < 1$ *in the "TransMax" and "TransMin" regions (see Fig. 7.1) for "leading jet" and "back-to-back" events (see text for the significance of symbols) as a function of the leading jet* P_T *compared with PYTHIA Tune A (black line) and HERWIG (dashed blue)*

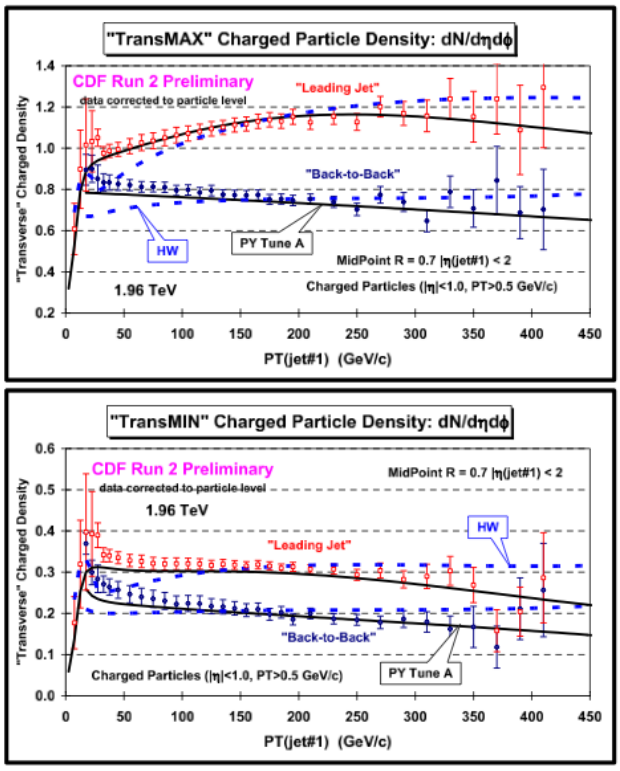


Figure 7.3: *Scalar total P_T density of charged particles, $dP_{T_{sum}}/d\phi d\eta$ with $p_T > 0.5$ GeV/c and $|\eta| < 1$ in the "TransMax" and "TransMin" regions (see Fig. 7.1) for "leading jet" and "back-to-back" events (see text for the significance of symbols) as a function of the leading jet P_T compared with PYTHIA Tune A (black line) and HERWIG (dashed blue)*

a program called GFLASH [112]. The result is a rapid and accurate response of the individual calorimeter towers to the energy deposited by the outgoing particles.

7.2 Signal and Background Definition

As mentioned earlier in this document our goal is to isolate signal events (WZ/ZZ) from the large background when investigating a final state signature of at least one high-P_T isolated electron or muon, large missing transverse energy, and two energetic jets. Leading order Feynman diagram for WZ and ZZ production are shown in Fig. 1.6, 1.7.

A number of other processes can mimic the above signature. We will refer to them as background for our search. There are mainly two categories of background processes: "physics background" processes, where all identified objects are true but originate from other processes and "instrumental" background processes, where some object is fake due to mis-measurements in the detector. The following backgrounds are expected to mostly contribute in the investigated dataset:

- WW: leading order Feynman diagram for WW production are similar to the signal ones (Fig. 1.8). Because of the poor calorimeter resolution the W and Z hadronic peaks cannot be distinguished from each other. In the W, Z hadronic modes, another difference between the WZ/ZZ signal and the WW process is in the decay modes: while the Z-boson can decay into quark of any flavors allowed by the phase-space, the W is not allowed to decay into b-quarks. Other than that, WW is an irreducible background for this analysis [6]. The WW cross section, as estimated from NLO calculations, is 11.3 ± 0.7 pb ([31]).

- W+jets: a leptonically decaying W boson is produced in association with jets by the annihilation of a quark from the proton and an anti-quark from the anti-proton. The W is produced along with a number of partons from the QCD radiation of the interacting partons (Fig. 7.4). If we exclude the absence of a hadronic peak, the W+jets is an irreducible background of this analysis. The inclusive cross section times the branching ratio of $p\overline{p} \rightarrow W + jets \rightarrow l\nu + jets$ is about 2900 pb. The given inclusive W+jets cross section is derived from leading order calculations and scaled up by a k-factor of 1.4 to account for the

[6] No other known method is available to isolate the signal from such a background.

next to leading order effects [7]. Moreover, the aforementioned cross section takes into account a 1.4 ± 0.4 correction factor in the $W + b\bar{b}$ and $W + c\bar{c}$ normalizations with respect to the NLO calculations. Such an increase was derived by fitting the heavy flavor composition of the W+1 jet data ([113]). Because of the large uncertainty carried by the theoretical NLO calculations ($\sim$ 40%) and by the 1.4 correction factor ($\sim$ 30%), both the W+jets and W+b/c-jets cross sections will be estimated from data.

- Z+jets: a Z boson is produced in association with jets in the same way as W+jets (see Fig. 7.4). Z+jets events contaminating our sample are those where a Z decays in charged lepton pairs and the large missing transverse energy arises by one lepton escaping the detection. The inclusive cross section times the branching ratio of $p\bar{p} \rightarrow Z \ + \ jets \rightarrow l^+l^- \ + \ jets$ from NLO calculations is about 260 pb, after accounting for corrections amounting to a 1.4 increase in the total Z+jets rate, and a 2.0 increase in the $Z + b\bar{b}/Z$+jets ([114]). Because of the large uncertainty carried by the theoretical calculations and by the 2.0 correction factor ($\sim$ 40%), both the Z+jets and Z+$b-$jets cross sections will be estimated from data.

- $t\bar{t}$: according to the leading order calculations, top pairs are expected to be produced at the Tevatron via $q\bar{q}$ annihilation approximately in the 85% of the cases and via gluon fusion in the remainder ([115]). The main Feynman diagram for the top pair production are shown in Fig. 7.5. Depending on the decays of the two $W's$, top pair production generates the following three final states:

  1. all-hadronic (BR = 44.4%): both $W's$ decay into jets, resulting in a final state with six jets;
  2. lepton+jets (BR = 44.4%): one hadronically and one leptonically decaying W, resulting in a final state of one charged lepton, large $\not\!\!E_T$, and 4 jets;
  3. di-lepton (BR = 11.1%): both leptonically decaying Ws, resulting in a final state with two leptons, large $\not\!\!E_T$, and 2 jets.

  The contribution of the first decay mode is negligible in our sample, since no charged leptons and no significant $\not\!\!E_T$ are expected. The contribution of the lepton+jets and di-lepton mode are almost equivalent when requiring two

[7] It was observed that such a correction was needed to model the data ([116]).

jets in the final state. This is due to the lower acceptance of the former with respect of the latter, despite of the higher branching ratio. Top pair cross section, as estimated from NNLO calculations for a top mass of 172.5 GeV/c^2, is 7.04 $\pm$ 0.49 pb ([115]).

- Single-top: the final state of single top production events have a leptonically decaying W boson and two quarks. Single-top can be produced either in the s-channel or t-channel (Fig. 7.6). Depending on the channel, one or two b-quarks will be produced in the final state. The s-channel and t-channel cross sections for the production of single-top quarks are respectively 1.02 $\pm$ 0.06 pb and 2.04 $\pm$ 0.18 pb from approximate NNLO+NNLL calculations ([117]) and approximate NNNLO+NLL calculations ([118]). The total single-top cross section for a top mass of 172.5 GeV/c^2 is 3.06 $\pm$ 0.19 pb.

- QCD multi-jets: multi-parton production (Fig. 7.7) can mimic our final signature if a jet is mis-identified as a lepton, and calorimeter mis-measurements fake a large missing transverse energy. A fake lepton may appear because one hadron is confused as a lepton. Real background leptons can also be produced during the hadronization (e.g: semi-leptonically decaying B-hadrons). QCD events contaminating our sample are mostly three jet events. Despite of the tiny lepton faking probability ($\sim 10^{-4}$) the contamination of such a background is relatively large because of the huge production cross sections (0.1-10 μb for exchanged P_T of 60 GeV/c down to 18 GeV/c). We do not rely on the MC to model the QCD multi-jets background. The main reason is that, because of the huge cross sections, it would be necessary to generate an extremely large number of events in order to have a statistically significant sample [8]. The other reason is that the lepton faking and the $\not\!\!E_T$ mechanisms are not understood very well and are therefore not expected to be modelled by our current detector simulations. We therefore derive the modeling of such a process from the data, as discussed in the next chapter.

The cross-sections for the MC-based processes described above are listed in Table 7.1. W/Z+jets cross section will be estimated from data (Sec. 8.3).

[8] if we define $\mathcal{L}^{eff} = N/\sigma$, N being the number of generated event, $\sigma \sim 10\mu b$ the production cross section, we would need to generate $N \gg 10^9$ in order to have $\mathcal{L}^{eff} \gg \mathcal{L}^{data} \sim 8.9\ fb^{-1}$. With the available computing resources this task is prohibitive.

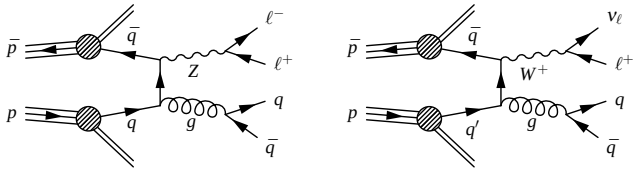


Figure 7.4: *Example of leading order Feynman diagram for the Z+jets (left) and W+jets (right) productions.*

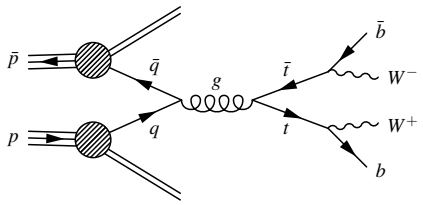


Figure 7.5: *Example of leading order Feynman diagram for the $t\bar{t}$ production*

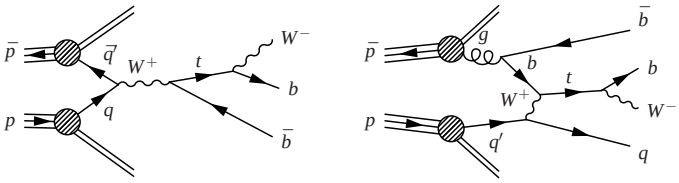


Figure 7.6: *Example of leading order Feynman diagram for the single-top production. Left: t-channel, Right: s-channel*

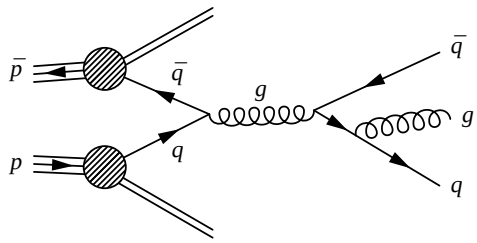


Figure 7.7: *Example of leading order Feynman diagram for the QCD multi-jets production.*

Process	Cross-section (pb)	Source
WW	11.3 ± 0.7	NLO ([31])
WZ	3.7 ± 0.2	NLO ([31])
ZZ	1.4 ± 0.1	NLO ([31])
$t\bar{t}$	7.7 ± 0.5	Data ([133])
Single-top, s-channel	1.0 ± 0.1	NNLO+NNLL ([117])
Single-top, t-channel	2.0 ± 0.2	NNNLO+NLL ([118])

Table 7.1: Cross-sections of diboson and top processes.

Chapter 8

Event Selection and Modeling of the Data Sample

In searches for small signals over large backgrounds it is imperative to understand the background shapes really well. Two years ago, the CDF collaboration published an excess of events in the di-jet mass spectrum, which was not described by the theoretical predictions within the known statistical and systematic uncertainties ([119]). Such a discrepancy was found in the data sample selected by requiring a high-P_T isolated electron or muon, large $\not\!\!E_T$ and two energetic jets. Since such a sample is very similar to the one that we were analyzing for measuring the WZ/ZZ cross section, it was mandatory to understand the reason of the aforementioned discrepancy.

In Sec. 8.1, 8.2 we describe the online and offline selections of the W+2jets sample. The modeling of such a sample is described in Sec. 8.3. In Sec. 8.4 we describe the accurate and careful studies that were performed to improve the agreement between our models and the data. In Sec. 8.5, 8.6 we introduce the fitting procedure and its validation. Finally, we present the final results in Sec. 8.7, with a new fit to the W+2jets sample (Sec. 8.8).

8.1 Data Sample

Candidate events for this analysis are required to have an event topology consistent with a leptonically decaying W, produced in association with two jets (Fig. 8.1). Online, these events are isolated with high efficiency, while keeping the QCD multi-jets background low, by triggering on the presence of a high-P_T electron or muon

in the final state [1]. Those triggers will be described in the next sections.

The whole CDF dataset is used. We require that all the sub-detectors of interest (e.g: silicon innermost and outer tracker, calorimeters, and muon chambers) are properly working. After these requirements the integrated luminosity is about 8.9 fb^{-1}.

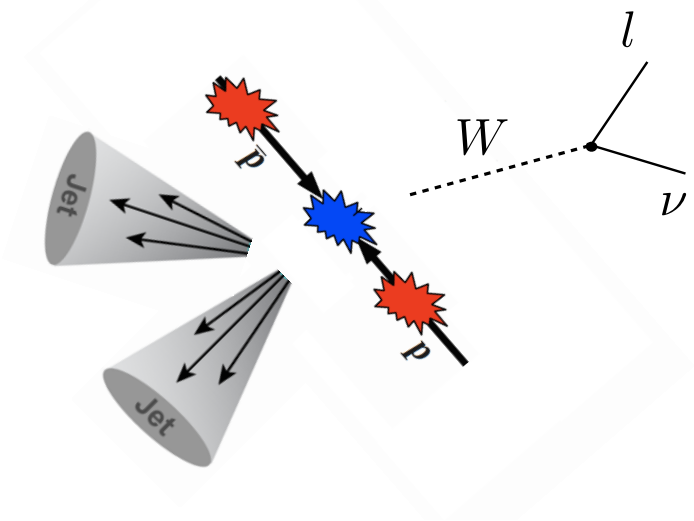


Figure 8.1: *Sketch of W +2jets event at the Tevatron.*

8.1.1 Central Electron Trigger

The trigger path ELECTRON_CENTRAL_18 is used to select central ($|\eta| < 1.1$) electrons. The selection is performed through three subsequent trigger levels. The requirements of each trigger level are summarized below.

- L1:
  - one "trigger tower" (corresponding to two calorimeter towers adjacent in η_{det}) with $E_T > 8$ GeV in the central calorimeters;
  - $P_T^{XFT} > 8.34$ GeV/c, where P_T^{XFT} is the transverse momentum of the COT track found by the XFT hardware (Sec. 3.7) and pointing to the tower;

[1] In the next chapter we will describe two more trigger paths, which rely on features of the W+2 jets which are approximately orthogonal to those presented in this chapter.

  - E_{had}/E_{em} must not exceed 0.125 (for $E_T < 14$ GeV).

- L2:
  - one electromagnetic cluster with $E_T > 16$ GeV [2] and $E_{had}/E_{em} < 0.125$. The cluster is formed by adding adjacent towers with $E_T > 7.5$ GeV to the tower above threshold at L1; only electromagnetic towers are used for the clustering purpose;
  - $|\eta^{XFT}| < 1.317$ and confirm of $P_T^{XFT} > 8.34$ GeV/c by using higher-accuracy algorithms [3].

- L3:
  - confirm of one electromagnetic calorimeter cluster of $E_T > 18$ GeV and $E_{had}/E_{em} < 0.125$,
  - a fully reconstructed three-dimensional track with $P_T > 9$ GeV/c, pointing to the cluster;
  - other requirements as those to be applied offline (Sec. 4.2).

The central electron trigger efficiency is measured in two steps: tracker and calorimeter efficiencies. The former is measured for each level by using a trigger with the same calorimeter requirements, but no tracking requirements. The tracking efficiency $\epsilon^{tracking} = 0.957 \pm 0.003$, is dominated by the L1 efficiency. The calorimeter efficiency is measured by exploiting the high-P_T muon trigger (see Sec. 8.1.2) for the L1 and calorimeter triggers with electron E_T thresholds lower than ELECTRON_CENTRAL_18 for L2, L3. The calorimeter efficiency is close to 100%. Tailored studies have shown that the trigger efficiencies do not depend on the most part of investigated observables, except for the electron η for the L1 tracking efficiency and the electron E_T for the calorimeter trigger efficiency [4]. These dependencies are not included in this analysis as they will be covered by the systematic uncertainty on the trigger efficiencies that we shall place in the final fit (Sec. 8.5).

Further details about the central electron trigger are available at [56], [120], [121].

[2] Thanks to the higher accuracy achieved with trigger upgrade (Sec. 3.7) such a cut was lowered from 18 GeV in the latest data taking periods.

[3] The high accuracies on some reported numbers are due to the CDF convention: these accuracies do not carry particular physics meanings.

[4] Lower tracking and calorimeter efficiencies are observed for $\eta_{det} \sim 0$ and $E_T < 30$ GeV respectively.

8.1.2 Central Muon Triggers

The trigger paths MUON_CMUP_18 and MUON_CMX_18 are used to select respectively CMUP and CMX muons (Sec. 4.3). The η_{det}, ϕ coverage of the CMUP and CMX triggered muons is shown in Fig. 8.2. The requirements of the three-level triggers are summarized below.

- L1:
  - Hits in different layers of the CMU (MUON_CMUP_18) or CMX (MUON_CMX_18) wire chambers. The arrival times of hits in different layers must be within 124 ns of each other;
  - a pattern (stub) of CMP hits on three out of four layers, consistent in ϕ with the observed CMU hits (MUON_CMUP_18);
  - a COT track found by the XFT hardware is required to match in the transverse plane the reconstructed stubs. The transverse momentum threshold is 4.09 GeV/c for MUON_CMUP_18 and 8.34 for MUON_CMX_18.
- L2:
  - a COT reconstructed track $P_T^{XFT} > 14.77$ GeV/c.
- L3:
  - a fully reconstructed three-dimensional COT track with $P_T > 18$ GeV/c matched to reconstructed stubs in the corresponding muon chambers is required. Other requirements similar to the offline ones (Sec. 4.3) are applied.

The trigger efficiency is measured in the $Z \to \mu\mu$ sample, by requiring one muon to pass the aforementioned MUON_CMUP_18 or MUON_CMX_18 requirements. The other track is checked whether it passes the MUON_CMUP_18/MUON_CMX_18 trigger requirements. These efficiencies are computed for the L1 and L2+L3. The overall measured efficiency for MUON_CMUP_18 and MUON_CMX_18 are respectively $\epsilon^{CMUP} = 0.864 \pm 0.004$ and $\epsilon^{CMX} = 0.897 \pm 0.003$.

Further details about the central muon triggers are available in [56], [121], [122].

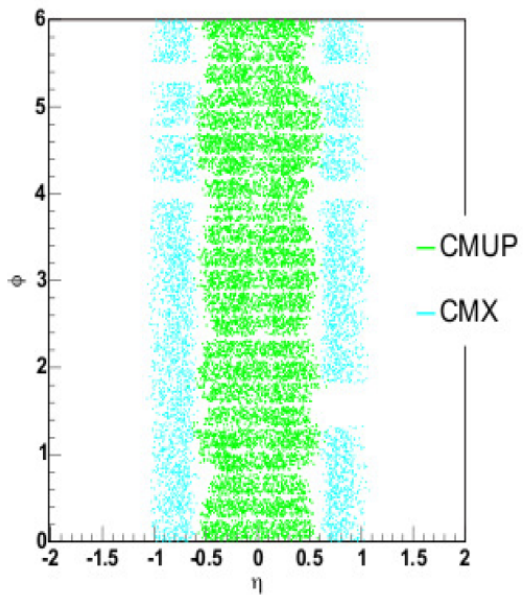


Figure 8.2: η_{det}, ϕ *coverage of the CMUP and CMX triggered muons.*

8.2 Offline Event Selection

We shall describe first the cuts to isolate the leptonically decaying W, and next the cuts meant to select the di-jet system recoiling against the W. Finally, we shall introduce some additional cuts to reproduce what was done in the previous analysis, where an anomalous excess of events in di-jet mass was observed ([119]). The jet energy is corrected for detectors inefficiencies, for physical effects in data and MC, and in the MC for improving the modeling in the $Z/\gamma+$ jet samples (Chap. 6). We apply those corrections only to jets with clustered energy more than 8 GeV.

8.2.1 Selection of the Leptonically Decaying W

The triggered electrons or muons are required to pass respectively the TCE (Sec. 4.2) or CMUP/CMX (Sec. 4.3) identification cuts. Those cuts are in general tighter than the online cuts (see Sec. 8.1), since we want to further remove jets misidentified as leptons at trigger level. The lepton E_T (P_T) is required to be larger than 20 GeV (GeV/c). Finally, we require $\not\!\!E_T > 25$ GeV. The missing transverse energy is computed as in Eq. 4.18. No quark/gluon-jet energy corrections (Sec. 6.2.4) are used for correcting $\not\!\!E_T$[5].

[5]The agreement between data and predictions in a number of observables is better when no quark/gluon-jet energy corrections are used for correcting $\not\!\!E_T$. This is a hint that the observed

8.2.2 Di-jet system selection

We require two jets with $E_T > 30$ GeV and $|\eta| < 2.4$. We do not consider jets with:

- an identified electron or muon in a 0.4 cone about the jet axis. For this purpose, all lepton categories listed in Sec. 4.2 and 4.3 are considered;
- electromagnetic fraction more than 0.9, since we do not want to confuse electrons with jets.

8.2.3 Additional Cuts

Although in this chapter we are studying the modeling of the W+2jets sample, our ultimate goal is to isolate the WZ/ZZ from the rest of the processes. To enhance our chances, a number of additional cuts are implemented to reject particular background processes, which mimic the diboson signature in the selected sample, or reject a not modeled portion of the phase-space. We list below the cuts against each particular background:

- Z+jets/$t\bar{t}$ veto: we reject events with more than one identified lepton. All lepton categories listed in Sec. 4.2 and 4.3 are considered.
- QCD multi-jets veto:
  - We request $M_T^W > 30$ GeV. M_T^W, the W transverse mass, is defined as follows:
  $$M_T^W \equiv \sqrt{2 \cdot E_T^l \not{E}_T \cdot (1 - cos\Delta\phi)} \tag{8.1}$$
  where E_T^l is the electron transverse energy or muon transverse momentum, $\Delta\phi$ is the difference in the azimuthal angles between the lepton and the missing transverse energy. Since no real W is produced in QCD multi-jets events, M_T^W is expected to peak at low values.
  - $\Delta\phi(j_1, \not{E}_T) > 0.4$, where j_1 is the leading E_T jet in the event. Since no neutrino is produced, fake $\not{E}_T$ is expected to be aligned in the transverse plane along a jet [6].

disagreement between data and simulations in the jet response may due to the the parton energy leaked outside the jet cone, which is not properly modeled. This energy was already considered when defining the $\not{E}_T$ (Eq. 4.15).

[6] Although the missing transverse energy is often aligned to the sub-leading jet, $\Delta\phi(j_1, \not{E}_T)$ was chosen for its greater discrimination power.

Three more cuts are applied: di-jet $P_T > 40$ GeV, $\Delta\eta(j_1, j_2) < 2.5$, and $\Delta R(j_1, j_2) > 0.7$. The latter cut was implemented because of the some observed discrepancies between predictions and data at low ΔR (see App. E.2 for further details). Such a cut was not implemented in [119]. We summarize all the applied cuts in Table 8.1.

Variable	Requirement
# vertices	> 0
Lepton type	TCE or CMUP/CMX
Lepton $E_T > (P_T)$	> 20 GeV (GeV/c)
# leptons	$= 1$
$\not\!\!E_T$	> 25 GeV
M_T^W	> 30 GeV
# jets	$= 2$
jet E_T	> 30 GeV
jet $\lvert\eta\rvert$	< 2.4
$\Delta\phi(j_1, \not\!\!E_T)$	> 0.4
$P_T(j_1, j_2)$	> 40 GeV
$\Delta\eta(j_1, j_2)$	< 2.5
$\Delta R(j_1, j_2)$*	> 0.7

Table 8.1: Summary of the event selection requirements in this analysis. All requirements but the ones marked with a * were applied in [119].

8.3 Modeling of the Data Sample

The expected sample composition is estimated in a sequence of steps by using a method similar to "Method2", which was used in several analyses at CDF investigating the W+jets signature in the heavy flavor enriched sample ([77]). The employed method assumes that the processes described in Sec. 7.2 are the only ones contributing to the selected sample. The steps implemented by the method are the following:

1. the contribution of well known processes (top, diboson) is estimated by using either theoretical or measured cross sections;

2. the contribution of QCD fakes and of V+jets is estimated with a data-based method, V being either W or Z.

8.3.1 Top and Diboson Processes

Theoretical calculations accurately predict the cross sections for top and diboson processes. We determine the yields of those processes by generating N_{gen} events, and performing a weighted sum of all selected events. In formula:

$$N_X = \sum_{i=1}^{N_{sel}} w_i \tag{8.2}$$

where $N_{sel} < N_{gen}$ is the number of events selected with the same selection cuts described in Sec. 8.2, and w_i is the weight of the ith events. For a given process, such a weight is computed as in Eq 8.3.

$$w_i = \sigma_{p\bar{p}\to X} \cdot \epsilon_i^{trigger} \cdot SF_i^{LEP-ID} \cdot \frac{1}{N_{gen}} \cdot \int \mathcal{L}dt \tag{8.3}$$

where

- $X = WW, WZ, ZZ, t\bar{t}$, single-top;
- $\sigma_{p\bar{p}\to X}$ refers to the cross section of the process (see Table 7.1);
- $\epsilon_i^{trigger}$ is the trigger efficiency, which is not included in the event simulation. The estimation of the trigger efficiencies is described in Sec. 8.1.1, 8.1.2;
- SF_i^{LEP-ID} is the scale factor, accounting for possible discrepancies in lepton identification efficiencies between data and MC. The estimation of the scale factors is described in [56].

By using Eq. 8.3, we can rewrite Eq. 8.2 as:

$$N_X = \sigma_{p\bar{p}\to X} \cdot A \int \mathcal{L}dt \tag{8.4}$$

where $A \equiv \epsilon_{trigger} \cdot SF_{LEP-ID} \cdot \frac{N_{sel}}{N_{gen}}$ is the MC-derived acceptance for the process X. Top and diboson yields in the W+2jets sample are reported in Table 8.3.

Diboson and $t\bar{t}$ samples are generated with PYTHIA; single-top samples are generated by using MADEVENT interfaced with PYTHIA (Chap. 7). A top mass of 172.5 GeV/c^2 is used for generating top samples.

8.3.2 V+jets

V+jets processes are harder to simulate because of the large number of contributing Feynman diagrams. At CDF, ALPGEN is used to simulate the associated

production of V bosons along with a fixed number of partons. The showering and hadronization are performed by PYTHIA (Chap. 7). We generate two kinds of processes:

- W/Z+n partons (n=0,1,2,3,$\geq$4);
- W/Z+$b\bar{b}/c\bar{c}$+m partons (m=0,1,$\geq$2)

where partons indicate the light flavor u,d,s,g partons ("LF"), while b, c are named heavy-flavors ("HF"). A set of kinematic cuts is applied at the generation level (e.g: $P_T^{LF-parton} > 15$ GeV, $P_T^{HF-parton} > 8$ GeV) to avoid infrared divergences in the calculation ([123]).

One known problem is the non-negligible overlap between the LF and the HF samples, since charm or bottom quarks can also be produced in the LF sample at the level of parton showering ([124]). This may lead to an overestimate of the HF flavor contribution if the overlap is not removed properly. However, since the V+jets rate is fitted to the data (see next), and there is no observed difference in shapes between the HF and LF samples (App. C), the aforementioned overlap is not expected to produce any bias in our results.

8.3.3 QCD multi-jets

As mentioned in Sec. 7.2, we model the QCD multi-jets background from the data. The mechanism for the QCD multi-jets to fake the W+2jets signature is different depending on the lepton type. It is customary at CDF to use the two datasets described below as QCD models for the background muon and electron contributions to the W+2jets samples.

To model CMUP or CMX muons from QCD, we use the non-isolated muon sample. Such a sample is selected in the same way as described in Sec. 8.2, except that the requirement on the isolation energy is inverted (Isolation > 0.1 rather than ≤ 0.1, Sec. 4.3). The kinematic of the event is not expected to change when flipping the isolation cut. Moreover, in order to closely replicate what is done in W+2jets sample, the jet containing the non-isolated muon is excluded from the list of jets accounted for correcting the $\not\!\!E_T$ (see Sec. 4.5).

To model identified electrons from QCD, we use the so called "non-electron (NE)" sample ([125]). Such a sample is selected with the same selection criteria described in Sec. 8.2, except that exclusively two of the following TCE-electron identification cuts (Sec. 4.2) are required to fail: E_{had}/E_{em}, L_{shr}, $Q \cdot \Delta X$, Δz.

These cuts are not expected to change the kinematics of the event in any way: the non-electron sample is thus expected to have similar kinematic features as the TCE sample [7].

The aforementioned QCD models must be tested. We report below the studies that were performed aiming to either validate or improve the models. When studying the QCD modeling in the electron sample we will often refer to a particular control region, selected in same way as the signal region (Sec. 8.2), except that one out of the $\not\!\!E_T$ or M_T^W cuts has been inverted. Because of the large QCD contamination ($\sim 84\%$), and the orthogonality with the signal region, the control regions can be used either to validate the non-electron model or to derive the corrections that such a model may need.

In the last subsection of this chapter we shall describe the data-driven method to derive the QCD multi-jets expected rate in the W+2jets selected sample.

Removal of real leptons from the QCD samples

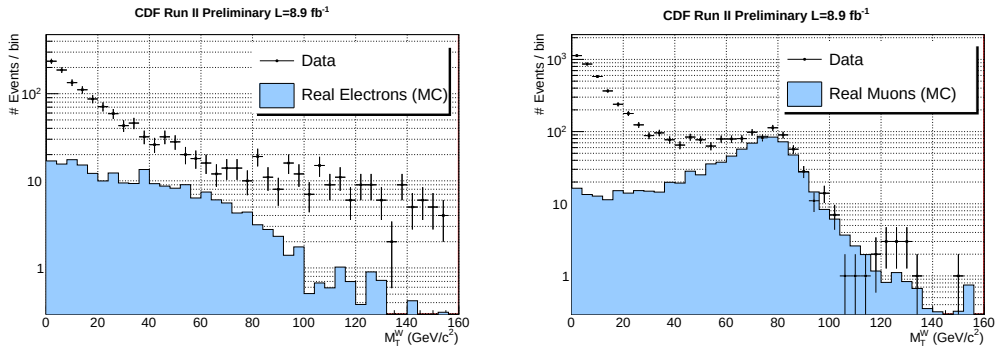
Since the lepton identification cuts mentioned above are not fully efficient, real leptons contaminate the QCD samples (Fig. 8.3). Such contamination is quantified in Table 8.3: the total fraction of real lepton is about 20% per sample. In figure 8.3b one can see that the Jacobian peak around 80 GeV/c^2 is due to real W's in the non-isolated muon sample. In order to avoid double counting between the models of the QCD multi-jets and the other background processes, we subtract the contribution of the MC-simulated processes from the QCD multi-jet models. This is done bin-by-bin for any variable of interest by taking the data content of a bin and subtracting the Monte Carlo prediction in that bin.

Corrections to the non-electron energy scale

The non-electrons may be different than TCE electrons (Sec. 4.2) for a number of reasons. For example, because L_{shr} cut is inverted, a larger amount of energy around the non-electron calorimeter cluster than the TCE one is expected (Fig. 8.4). Therefore, one cannot expect that the non-electrons model appropriately reproduces the electron energy spectrum.

We investigate QCD multi-jets simulated events and compare the measured

[7] One limitation of the non-electron sample is the poor statistics. However, the non-isolated electron sample, which is not statistic limited, cannot be used easily because the isolation is heavily correlated to the kinematics of the event.



(a) (b)

Figure 8.3: M_T^W *distribution in a log scale for the non-electron (a) and the non-isolated muon (b) samples. The contamination of real electrons and muons respectively (solid blue) is overlaid. Such a contamination is derived from simulating the MC processes described in the text. MC processes are normalized to the theoretical cross sections (Sec. 7.2).*

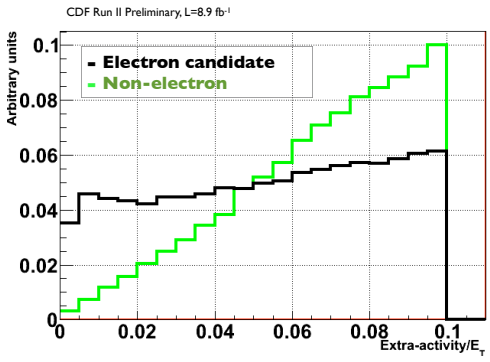


Figure 8.4: *Ratio between the extra energy around the calorimeter cluster (in a cone of radius $R = 0.4$) and the transverse energy of the cluster. TCE identified electrons (black) are compared to non-electrons (green) in the control region defined in the text ($\not\!\!E_T$<25 GeV or M_T^W <30 GeV/c^2). A < 0.1 cut on the ratio is applied.*

Process	Rate (Non-electron sample)	Rate (Non-isolated muon sample)
Diboson	12.3 ± 0.6	36.4 ± 1.0
Top	16.4 ± 0.3	89.3 ± 0.8
V+jets	191.1 ± 6.9	701.0 ± 9.2
Total	219.8 ± 8.0	826.7 ± 11.4
Data	1365	4784

Table 8.2: Expected rates from processes containing real leptons in the non-electron and non-isolated muon samples. Theoretical cross sections from Sec. 7.2 have been used. Errors are statistical only. The last line shows the observed total background rates.

energy of the jets matched to the TCE/non-electron to the energy of the highest-E_T generated parton within the jet cone [8]. Matching is achieved if the ΔR between the two objects is less than 0.4. The ratio between the parton E_T and the uncorrected E_T of the matched jet to TCE/non-electron is compared in six bins of jet uncorrected E_T. An example is shown in Fig. 8.5. One can see that in both cases (TCE and non-electrons) the jet underestimates in average the parton energy. Two different transfer functions, "TF^{TCE}" and "$TF^{non-electron}$", are needed to correct the jet energy and reproduce the primary parton energy in the TCE and non-electrons samples.

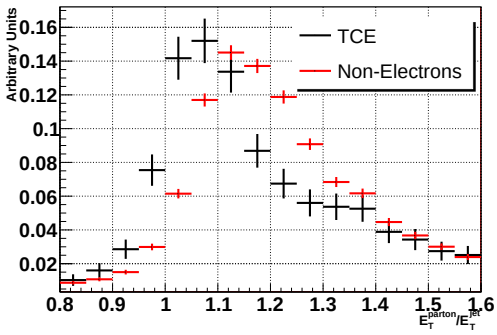


Figure 8.5: *Ratios between the parton E_T and the uncorrected E_T of the jets matched to the TCE (black) and non-electrons (red). Those jets are required to have transverse energy between 100 GeV and 150 GeV.*

[8] In such a comparison the jet rather than the TCE/non-electron cluster is used in order to recover the energy lost outside the cluster. We assume that the matched parton is a good representation of the true jet energy scale.

However, since our goal is to improve the non-electrons model, we correct the energy of the jets matched to non-electrons such that the jet response in the non-electrons and the TCE samples agree in average. The applied correction is $K \equiv \frac{TF^{TCE}}{TF^{non-electron}}$. The dependence of K upon the jet uncorrected E_T is shown in Fig. 8.6. The $\not{E}_T$ in the non-electron sample will be corrected accordingly:

$$\vec{\not{E}}_T \rightarrow \vec{\not{E}}_T - K \cdot \hat{P_T}^{NE} \tag{8.5}$$

where $\vec{\not{E}}_T$ is defined in Eq. 4.18 and $\hat{P_T}^{NE}$ is the direction of the non-electron in the transverse plane.

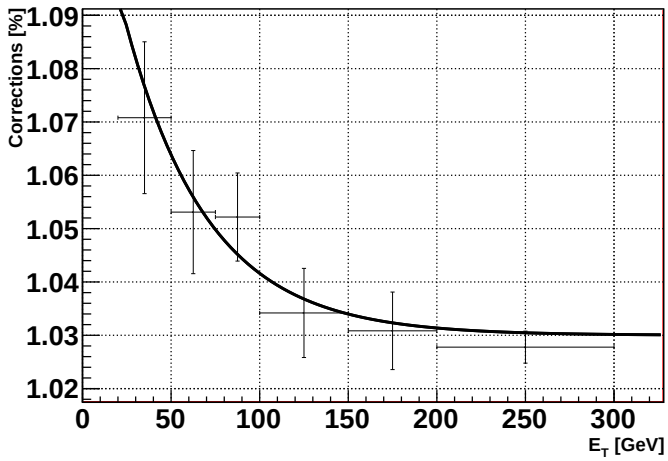


Figure 8.6: *Corrections to the uncorrected energy of the jets matched to non-electrons. The dependence of the corrections upon the jet energy is shown. Such a dependence is parameterized as* $A \cdot e^{B \cdot (x-20)} + C$, *where* $A = 0.064 \pm 0.034$, $B = -0.021 \pm 0.000$, $C = 1.030 \pm 0.001$.

Trigger bias removal in the non-electron sample

When comparing in the control region the energy distribution of the jets matched to TCE and non-electrons we find large discrepancies (Fig. 8.7a): while the TCE transverse energy spectrum exponentially decreases, as expected because of the non-resonant nature of the QCD multi-jets process, the non-electron spectrum exhibits an artificial rise for $E_T < 35$ GeV. The latter behavior is due to the trigger efficiency being energy dependent rather than constant, as for the TCE (Sec. 8.1.1). This happens because of:

- the trigger cuts on the cluster electromagnetic energy only;

- a sizable percentage of the non-electron energy is hadronic ($\lesssim 12.5\%$).

Obviously, such an effect is not present in the TCE sample, since the TCE energy is almost purely electromagnetic.

In order to remove the trigger bias, we reweight non-electron events bin-by-bin of matched jet E_T such that the predicted spectrum agrees with the data (8.7b). From now on such a set of weights will be used for non-electrons in the control region as well as in the signal region.

In order to test the effect of these corrections we check the modeling of crucial kinematic distributions for this analysis: the P_T and ΔR of the di-jet system. Fig. 8.8, 8.9 show these quantities in the control region, before and after applying the non-electron energy corrections and removing the trigger bias. The improvement in the modeling is striking. The same level of improvement can be seen in other kinematical variables as well.

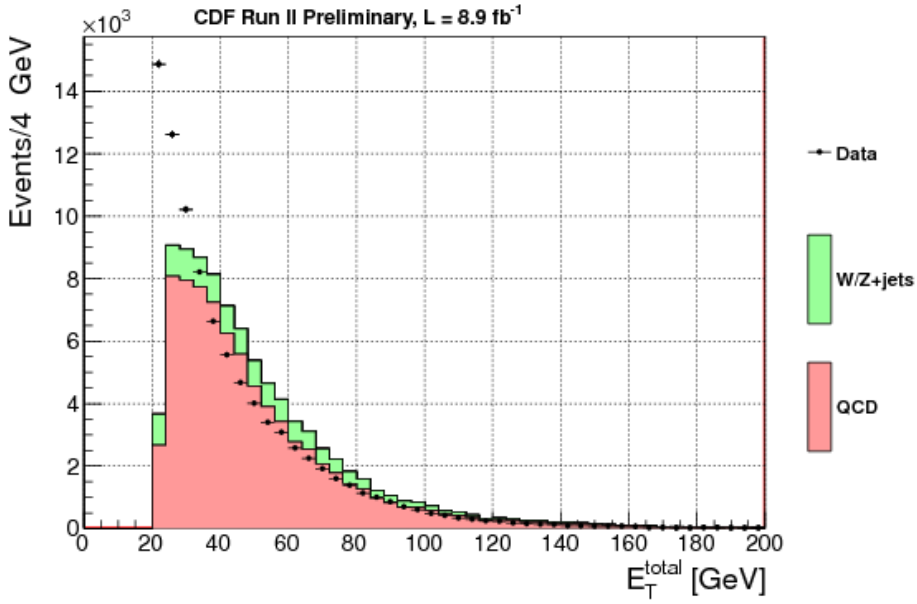
Studies to investigate possible problems in the non-isolated muon model have been considered. However, since the QCD multi-jets background is tiny in the muon sample (Table. 8.3), any imperfection in the modeling of the QCD multi-jets background in the muon sample cannot possibly bias the final result. Therefore, we decided not to finalize those studies.

Estimation of the QCD multi-jets rate

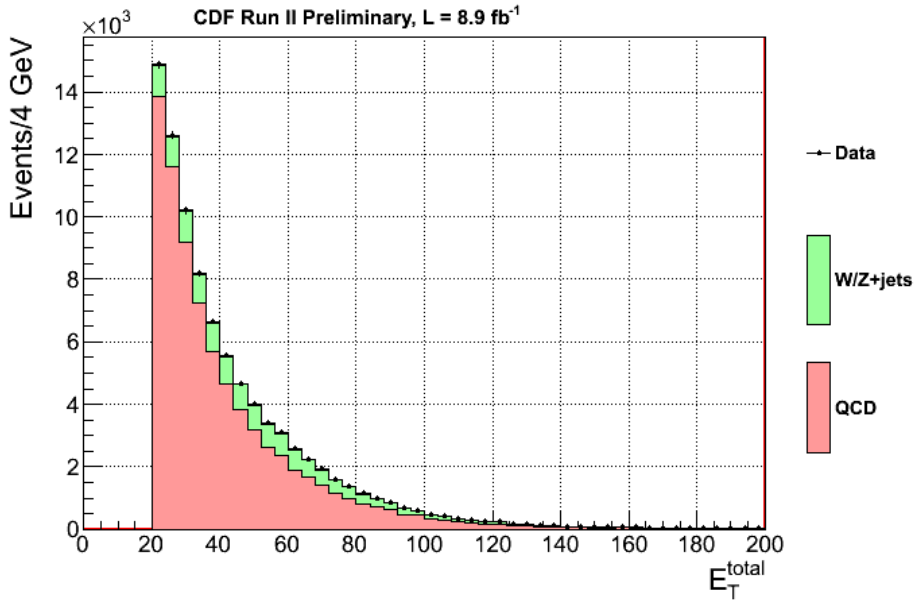
The QCD multi-jets rate in the W+2jets sample is derived by fitting the data. The fit is performed in missing transverse energy because of the large difference in this sample between the shape of the QCD multi-jets background, where the imbalance is mostly caused by instrumental mis-measurements, and the simulated MC processes, where energetic neutrinos escape the detector. The data sample used for the fit is selected as in Sec. 8.2 except for the $\not\!\!E_T$ cut, which is removed. This is done to increase the statistical power of the sample, since most of the QCD is characterized by low $\not\!\!E_T$. Three templates [9] are used for the fit:

1. EW: top and diboson templates are combined. Their rates are estimated from the theoretical cross-sections, as described earlier. The normalization of the EW template is constrained to the expectation within a rate uncertainty ($\sim$ 6%) from the cross sections in Table 7.1.

[9] A template is a distribution of events in a given variable. If the template is built by simulations, the events are weighted as defined in Sec. 8.3.1. Unless differently specified, the data template is composed of events with unitary weight.

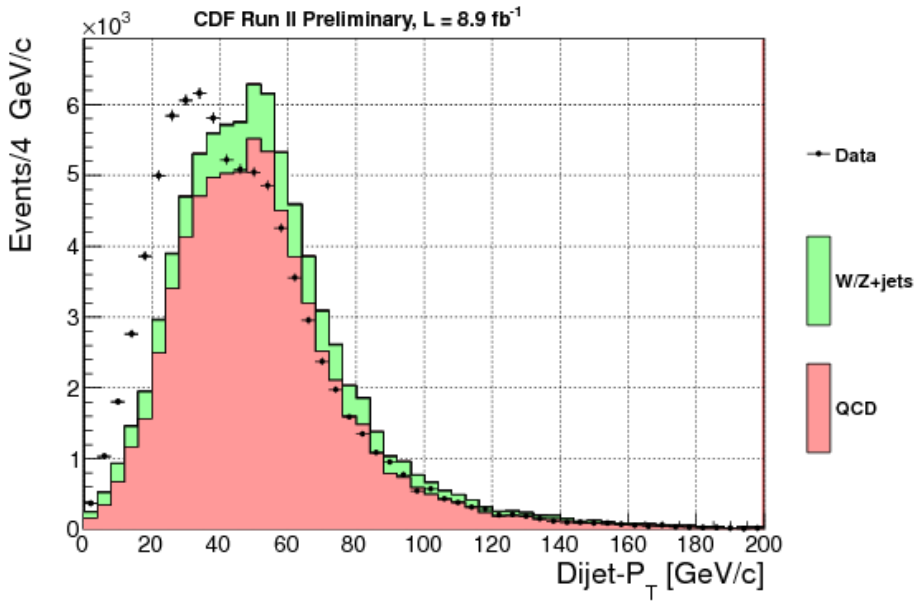


(a)

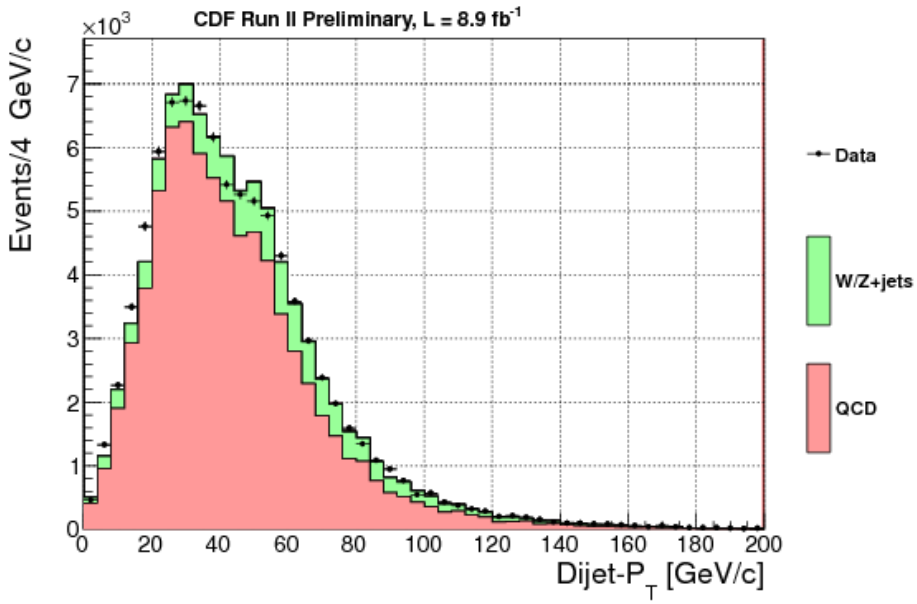


(b)

Figure 8.7: *E_T of non-electron jets from the dominating QCD multi-jets (pink) and from W/Z+jets (green, stacked) in the control region. The TCE exponentially decreasing spectrum (data points) is shown for comparison. (a) At low momenta there is a strong inefficiency in non-electron relative to TCE, which should be attributed to a trigger bias (see text). (b) The same distributions are shown after applying the correction to the QCD contribution (see text) in order to recover the inefficiencies. In both plots the W/Z+jets processes are normalized to the theoretical cross sections (Sec. 7.2). The QCD multi-jets is normalized to the observed rate after removing the expected W/Z+jets rate.*

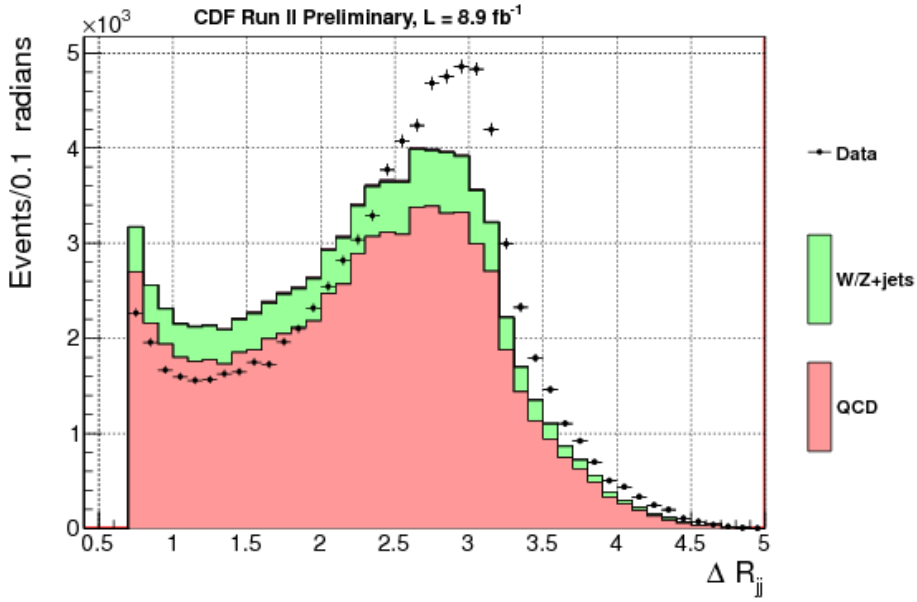


(a)

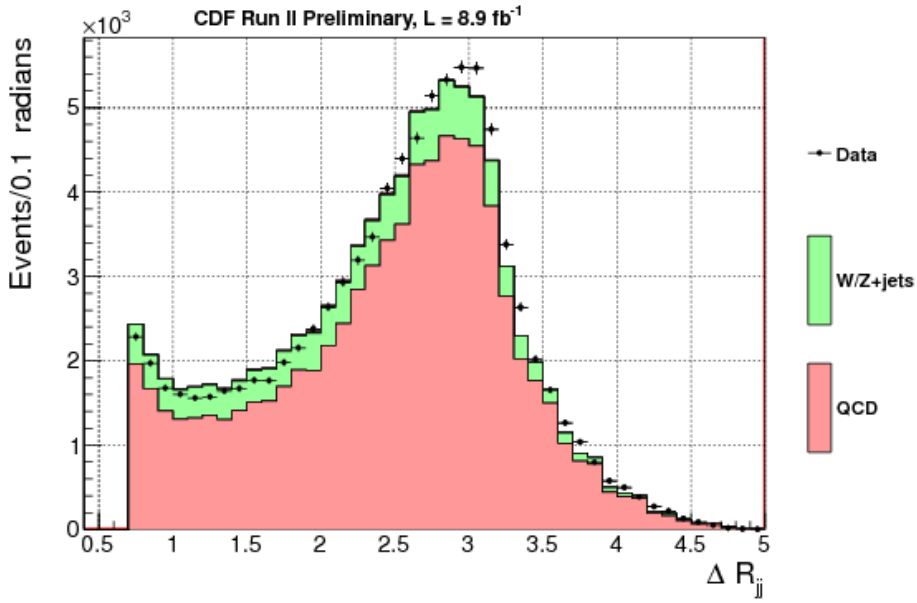


(b)

Figure 8.8: P_T of the di-jet system in the control region before (a) and after (b) applying the corrections to the QCD model (see text). The W/Z+jets processes are normalized to the theoretical cross sections (Sec. 7.2). QCD multi-jets is normalized to the observed rate after removing the expected W/Z+jets rate.



(a)



(b)

Figure 8.9: $\Delta R(jet_1, jet_2)$ of the di-jet system in the control region before (a) and after (b) applying the corrections to the QCD model (see text). The W/Z+jets processes are normalized to the theoretical cross sections (Sec. 7.2). The QCD multi-jets is normalized to the observed rate after removing the expected W/Z+jets rate.

2. V+jets: the ratio between the W+jets and Z+jets normalizations is assumed from the theoretical cross sections (Sec. 7.2). The W+jets and Z+jets templates are combined into the V+jets template, whose normalization is a free parameter of the fit.

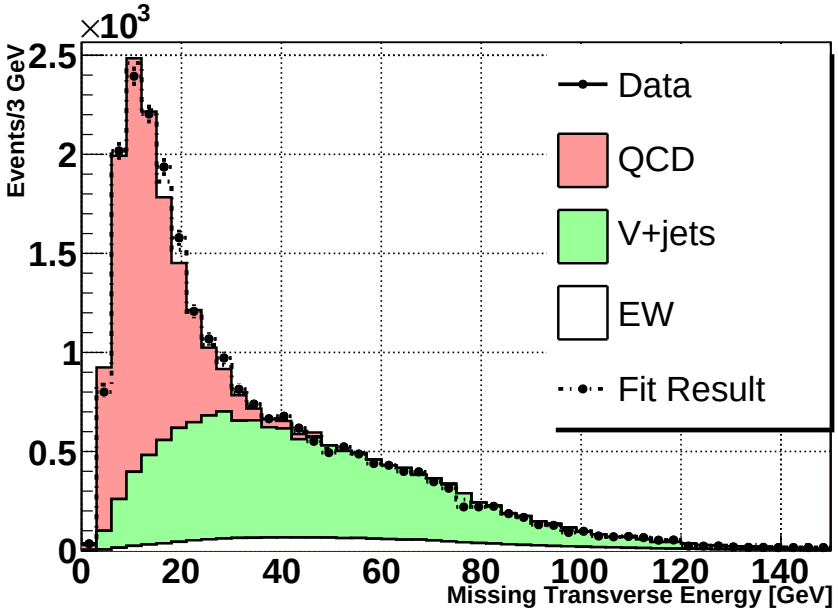
3. QCD multi-jets: non-electron or non-isolated muon data events are used to build the QCD multi-jets templates for the electron and muon sample (Sec. 8.3.3). The non-electron and non-isolated muon models are corrected as described earlier in this section.

The fit is performed by maximizing a binned likelihood function. When performing the maximization, each bin of the templates is allowed to vary within its error. This is done to account for the limited statistics of the data and of the templates. Such a fit is performed within the *TFractionFitter* package ([126]).

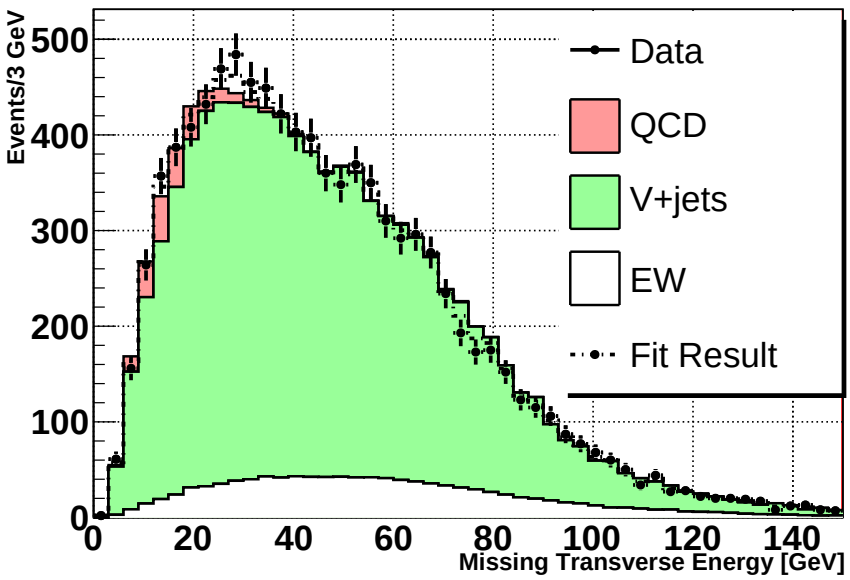
The fit results in the electron and muon samples are shown in Fig. 8.10. The fitted QCD fractions are 0.3% and 7.8% for the muon and electron samples respectively. Those fractions are affected by a relative uncertainty of 14.1%. This uncertainty was derived with a number of studies documented in App. D. The resulting uncertainty improves significantly relative to the findings of the CDF Higgs group ([127]).

From the $\not\!\!E_T$ fit we also obtain an initial estimate of the V+jets rates (Table 8.3): they correspond to about 83% (electron sample) and 84% (muon sample) of the rates predicted when using the NLO cross sections [10]. The above V+jets fractions will be used as starting values of the final di-jet invariant fit, which will be described in Sec. 8.5.

[10] As stated in Sec. 7.2, the theoretical NLO calculations carry large uncertainties ($\sim$ 40%). Therefore, the fitted V+jets rates are compatible with the theory.



(a)



(b)

Figure 8.10: $\not\!\!E_T$ distribution in the electron (a) and muon (b) samples. EW (white), V+jets (green), and QCD multi-jets (pink) templates are normalized according to the fit results, stacked together and superimposed to the data (dots). While fitting, each bin of the stacked histogram is varied within its error. This is shown in the distribution named "Fit Result".

Process	Rate (electrons)	Rate (muons)
Diboson	589.3 ± 35.9	392.0 ± 23.9
$t\bar{t}$	670.0 ± 43.6	430.5 ± 28.0
single-top	160.8 ± 10.1	106.3 ± 6.7
V+jets	9147.8 ± 121.9	6430.9 ± 101.8
QCD multi-jets	898.1 ± 126.6	20.2 ± 2.8
Total expected	11466.0 ± 184.9	7380.0 ± 108.5

Table 8.3: Breakdown of the expected rates in the electron and muon samples. The total expected rates are constrained to be equal to the number of events in the data. The reported uncertainties are the uncertainties of the theory for diboson and single-top processes, and of the measured CDF cross-section for $t\bar{t}$ process. The QCD multi-jets and V+jets rate uncertainties are derived from the $\not\!\!E_T$ fit.

8.4 Validation of the Model

After correcting the MC to match the instantaneous luminosity in the data [11], we proceed to investigate the agreement between data and the predictions in the W+2jets sample. Since the expected rates are normalized to the total number of data events [12], we compare the agreement between the predicted and observed shapes of a number distributions (Fig. 8.11-8.13), which are important for this analysis [13]. The electron sample is kept separated from the muon sample, since the contributions of different processes (e.g: QCD, V+jets) are different in the two samples. Overall, the agreement is good. We conclude that the shapes and the relative rates of each process are under control.

8.5 Di-jet mass fit

We test how well the SM predictions describe the W+2jets sample by performing a fit in the invariant mass of the two jets. Such a variable is the most sensitive variable for extracting the cross-section of hadronically-decaying resonances.

[11] MC samples are weighted to match the distribution of the number of vertices in the data. This is a standard CDF procedure ([128]).

[12] $\not\!\!E_T$ fits constrain the predicted rates to be equal to the total number of data events (Sec. 8.3.3). After requiring $\not\!\!E_T >$ 25 GeV, the V+jets is further adjusted to cover for the minor rate discrepancies between predictions and observations.

[13] Other distributions are shown in App. E.1.

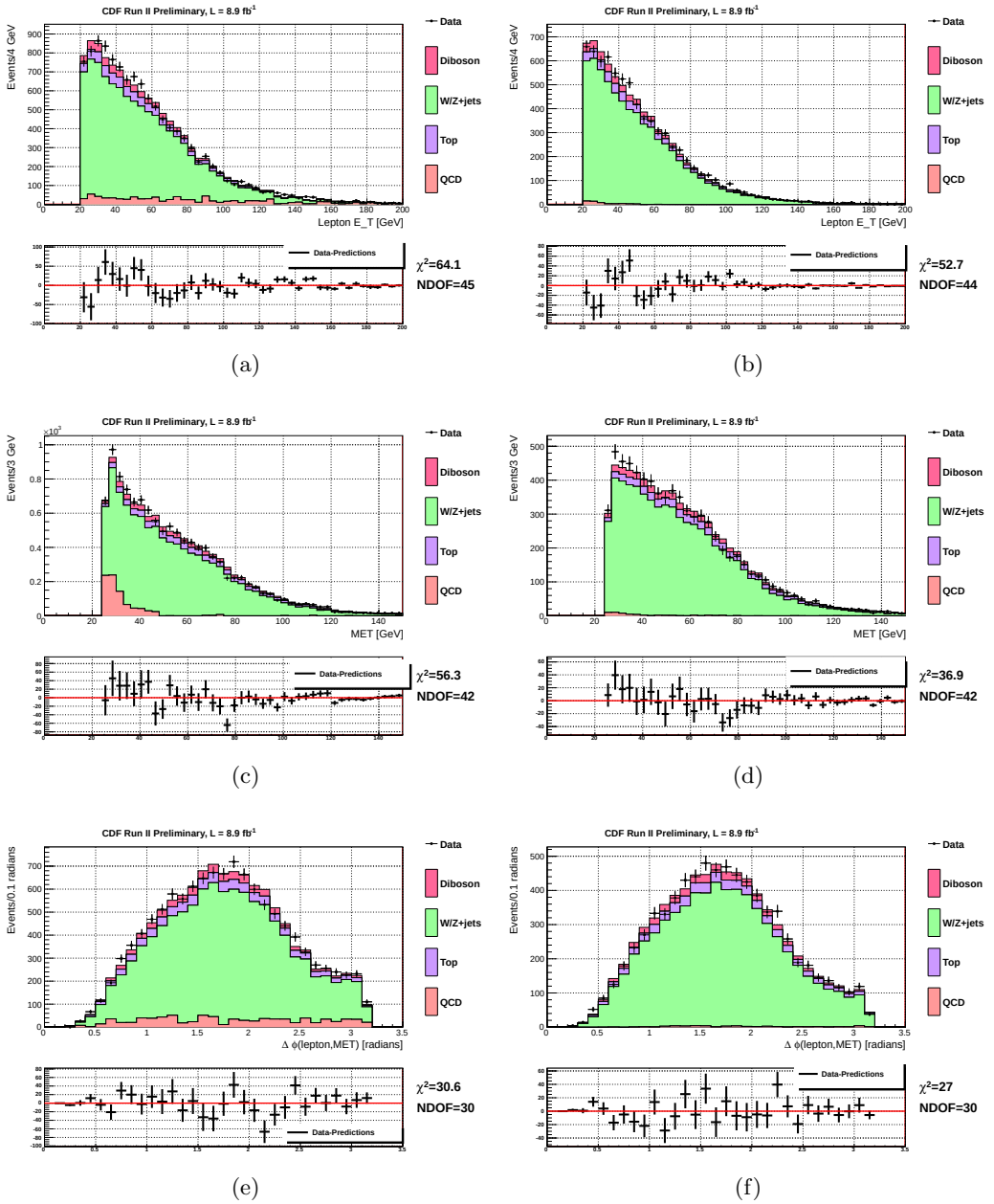


Figure 8.11: Lepton E_T (a-b), and $\not\!\!E_T$ (c-d), $\Delta\phi$(lepton,$\not\!\!E_T$) (e-f) distributions in the electron (left) and muon (right) samples, lepton being the triggered lepton. In the upper pads data (dots) is superimposed to the predictions, which have been properly normalized (see text). In the lower pads the data minus the predictions is plotted: the red horizontal line at 0 represents the ideal agreement between the two. The χ^2 and the number of non-zero bins (NDOF) are also reported to quantify the agreement. The jet matched in direction to the electron is used when plotting the above variables in the electron sample.

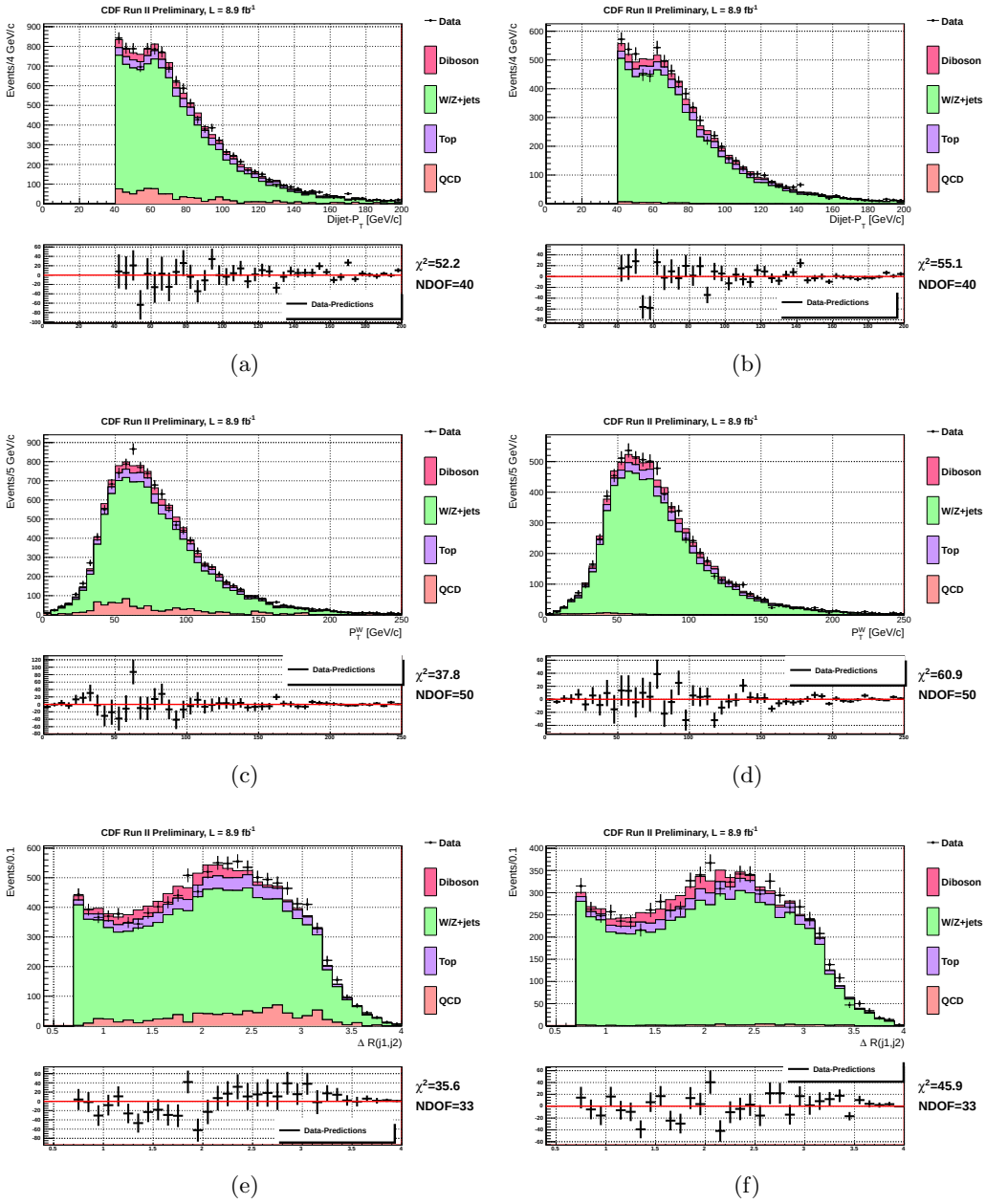
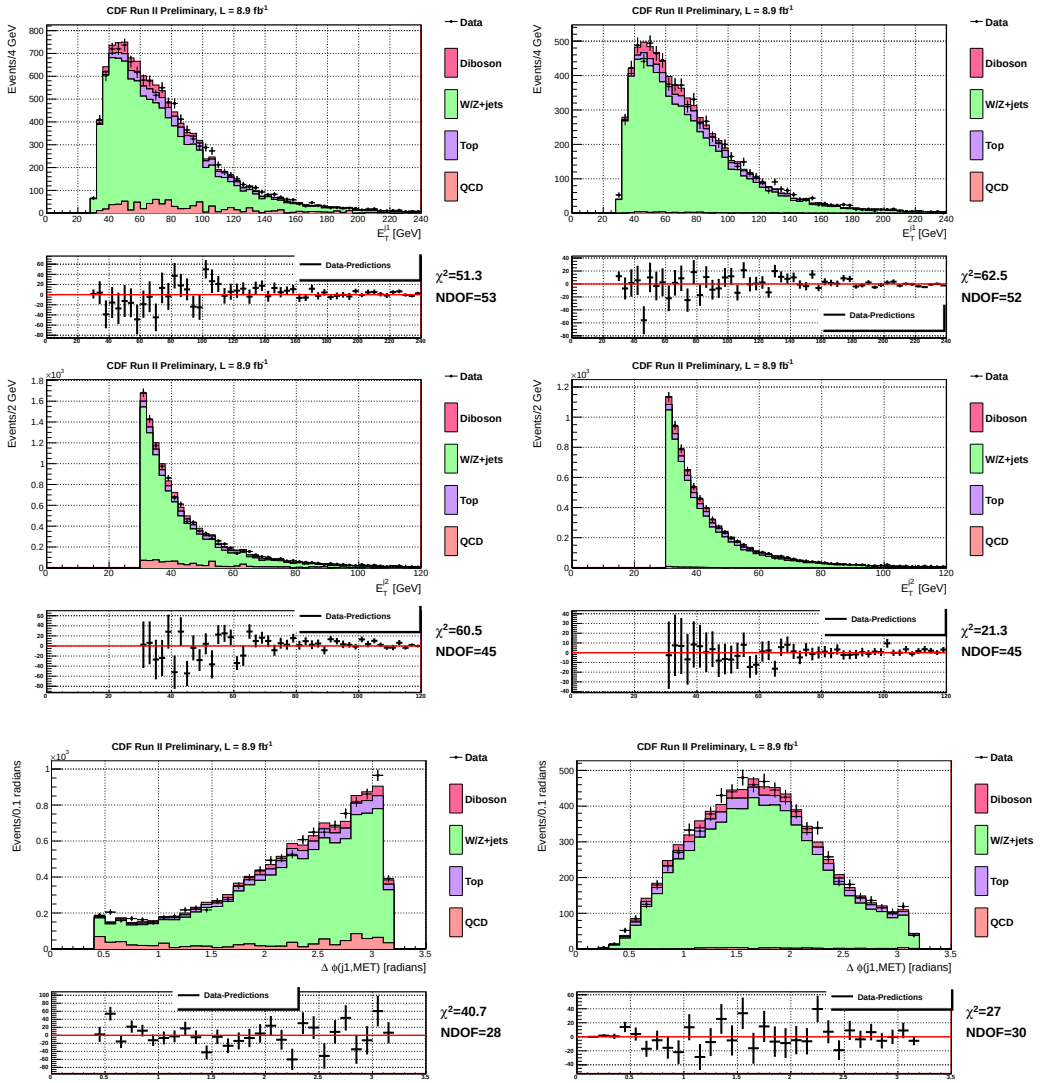


Figure 8.12: Di-jet P_T (a-b), P_T(lepton,$\not\!\!E_T$) (c-d), $\Delta R(j_1, j_2)$ (e-f) distributions in the electron (left) and muon (right) samples. $j_{1,2}$ are the highest and second highest-E_T jets. In the upper pads data (dots) is superimposed to the predictions, which have been properly normalized (see text). In the lower pads the data minus the predictions are plotted: the red horizontal line at 0 represents the ideal agreement between the two. The χ^2 and the number of non-zero bins (NDOF) are also reported to quantify the agreement.



(a) (b)

Figure 8.13: $E_T^{j_1}$ (a-b), $E_T^{j_2}$ (c-d), $\Delta\phi(j_1,\not{E}_T)$ (e-f) distributions in the electron (left) and muon (right) samples. In the upper pads data (dots) is superimposed to the predictions, which have been properly normalized (see text). In the lower pads the data minus the predictions are plotted: the red horizontal line at 0 represents the ideal agreement between the two. The χ^2 and the number of non-zero bins (NDOF) are also reported to quantify the agreement.

8.5.1 Fit procedure

The fit to the data is performed by maximizing the binned likelihood function $L(data|\vec{\vartheta},\vec{\nu})\pi(\vec{\nu})$, which express the probability of observing the data, given the model parameter $\vec{\vartheta}$ and the nuisance parameters $\vec{\nu}$. The function

$$L = \prod_{i=1}^{n_{bins}} \mu_i^{d_i} \frac{e^{-\mu_i}}{d_i!}$$

is the product of Poisson probabilities over n_{bins} bins to observe d_i events in the ith bin, where the predicted rate is μ_i. $\mu_i = \sum_{k=1}^{N_{process}} c_{ik}$, c_{ik} being the rate prediction for the kth process in the ith bin (i.e.: the kth-template bin content), and $N_{process}$ is the number of processes in the model. The function π is a product of Gaussian constraints (one per each nuisance parameter), which incorporates external information as measured in control samples, other measurements or theoretical predictions. Restrictions are placed on the allowed ranges of the fitted nuisance parameters to ensure that all predictions are non-negative. The likelihood function maximization, or χ^2 minimization [14], is performed within the MCLIMIT package [129], which exploits the program MINUIT [130]. The associated uncertainties are computing using MINOS, part of MINUIT. The minimized χ^2 obeys to a χ^2 distribution with an effective number of degrees of freedom, which is n_{bins} minus the number of nuisance parameter which we want to extract (i.e: unconstrained nuisance parameters). A more detailed description of the fit is given in [113].

In our analysis we build the likelihood with a model composed of the following five templates:

1. diboson
2. $t\bar{t}$
3. single-top
4. V+jets
5. QCD multi-jets

When the template is the sum of two or more processes (eg: diboson, V+jets), the fraction of each process with respect to the total is fixed to the theoretical predictions (Eq. 8.4). The fit is performed simultaneously in the electron and

[14] $\chi^2 \equiv -2lnL$

muon samples, because shapes and process contaminations are different in the two samples. Templates (Fig. E.8, E.9) are initially normalized to the rates provided in Table 8.3. The treatment of the nuisance parameters, describing the systematic uncertainties on the parameters of our models, is discussed in the next section.

8.5.2 Systematic Uncertainties

When simulating events a number of assumptions were made. Since this analysis relies heavily on MC simulations, a number of systematic uncertainties are considered in order to make sure that those assumptions do not bias the final results. These systematic uncertainties are described by nuisance parameters, which are of three forms:

- bin-by-bin uncertainties: they are considered uncorrelated for each bin and each process;
- shape uncertainties: coherent distortions of the contents of each bin of a template, parameterized by a single nuisance parameter;
- rate uncertainties: they affect the overall normalization of one or more templates.

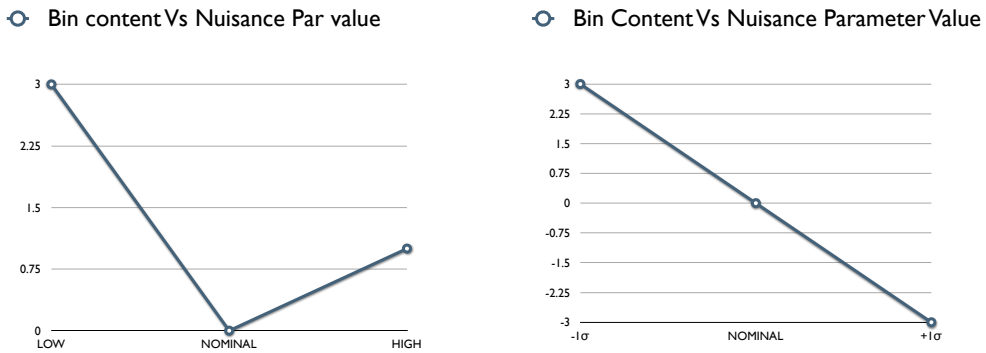
The first category of nuisance parameters is similar to the one considered when performing the $\not{E}_T$ fit for the QCD multi-jets rate (Sec. 8.3.3). It is introduced to account for the limited statistics of the non-electron template (Fig. E.8). In the case of the shape and rate nuisance parameters, the variation of each bin for a given template is totally correlated.

In principle the rate or shape systematic uncertainties are handled in the same way ("morphing"):

1. the impact on the rate of the bin content (shape systematic) due to $\pm 1\sigma$ variation of the nuisance parameter is estimated;
2. linear interpolations from the nominal rate/bin contents to the $\pm 1\sigma$ varied ones are performed: this interpolation allows for an analytic description of the dependence of the variation on the nuisance parameter value;
3. the linear interpolation is extrapolated beyond $\pm 1\sigma$.

The analytic description is needed to fit the nuisance parameters to the data. However, the morphing procedure may produce a discontinuity in the rate change (Fig.

8.14a) with the subsequent risk of producing a false maximum in the fitter. To avoid such a technical problem we symmetrize the $\pm 1\sigma$ rate and shape uncertainties around the nominal values before the morphing takes place. An example of such a symmetrization is shown in Fig. 8.14b [15].



(a) (b)

Figure 8.14: Impact of the nuisance parameter variation (in unit of given uncertainty σ) from the nominal value on either the rate or the bin content of a template. The scenarios before (left) and after (right) the symmetrization procedure are shown. Arbitrarily units are used on the Y-axis.

The handling of the shape systematic uncertainties on the diboson templates is very different:

- we parameterize the di-jet mass peak with a Gaussian function on top of a 4th-degree polynomial [16],
- we fit the $\pm 1\sigma$ varied shapes with the same function, only allowing the Gaussian parameters to change.
- we interpolate the resulting Gaussian parameters with straight lines, by forcing them to pass through the nominal values.

This procedure allows for an analytic parameterization of the diboson shape dependence on the nuisance parameter. Such a parameterization, rather than the

[15] If the nominal value is already contained within the $\pm 1\sigma$ uncertainties, we use the average of the two to redefine the new uncertainties.

[16] The polynomial accounts for invariant mass tails due to either detector or physics effects (eg: initial and final state radiations)

aforementioned morphing is used in the binned likelihood fit. All fits used to derive the above parameterization are shown in App. E.4.1.

A detailed description of the considered systematic uncertainty is given below, while their values are summarized in Table 8.4. If not stated otherwise, the nuisance parameters described below are treated as fully correlated between the electron and the muon samples.

Cross-section/Normalization Uncertainties

The event rate for the top and diboson processes is derived from Eq. 8.4: uncertainties based on the production cross-section directly translates into rate uncertainties. Therefore we consider the following systematic uncertainties:

- diboson, Single-top: respectively 6.1%, 6.3% rate uncertainties from the theoretical cross-sections (Table 7.1);
- $t\bar{t}$: 6.5% rate uncertainty from the CDF measured cross-section (Table 7.1. Such a cross-section is slightly more accurate than the one derived from NNLO calculations ([115])

A 14.1% rate uncertainty is used for the QCD multi-jets rate in both the electron and muon samples [17]. When fitting the QCD rate, the nuisance parameters in the two samples are treated as uncorrelated to be as conservative as possible.

Because of the large theoretical uncertainties on the V+jets cross-section, the V+jets rate will be treated as an unconstrained nuisance parameter in the fit.

Jet Energy Uncertainty

The modeling of the jet energy scale (JES) by our MC is affected by uncertainties. As we discussed in Sec. 6.2.4, we will use a $\sigma_{E_T}^{quark}$ ≡2.7% and $\sigma_{E_T}^{gluon}$ ≡ 4.4% uncertainties respectively for quark and gluon jets. These uncertainties are 100% anti-correlated. Therefore, the $\pm 1\sigma$ JES uncertainties are obtained by shifting by $\pm 1\sigma_{E_T}^{quark}$ the quark jet energy scale, and by $\mp 1\sigma_{E_T}^{gluon}$ the gluon jet energy scale. These uncertainties will be applied to all MC based processes ($V + jets$, top, and diboson). Because of the V+jets rate is pinned to the data ($\not\!\!E_T$ fit - Sec. 8.3.3), the V+jets JES uncertainty is shape only. The JES rate systematic uncertainties

[17] We assume that the same QCD rate uncertainty for the electron sample (App. D) applies to the muon sample

are listed in Table 8.4. The shape distortions for all considered processes due of the JES variations are displayed in App. E.4.2.

A wrong modeling of the jet energy resolution (JER) may affect the predicted rates and shapes. We evaluate such a systematic uncertainty by smearing the MC jet energies with a Gaussian distribution with mean of 1.0 and width of $0.03 \cdot E_T + 1.7$ (GeV), E_T being the jet transverse energy in GeV ([131]). No significant effect is observed, therefore no JER systematic uncertainty is considered in this measurement.

$\mathbf{Q^2}$

As mentioned in Chap. 7, we simulate events by factorizing the PDF and the finite-order matrix element calculations. Such a method involves the choice of a "boundary" Q^2 scale, which separates the perturbative (matrix element) from non-perturbative regions, such that collinear and infrared divergences neglected in the matrix element can be properly accounted for by the Q^2 running of the PDF. The scale, named factorization scale (Q^2), naively can be thought as the scale at which the proton or anti-proton is being probed. Because the hadronic cross-sections are computed at some finite order, there is an explicit dependence on Q^2 of the kinematic of the generated process. Such a dependence is embedded in the formulation of the strong coupling constant α_s, as reported in Eq. 1.18. With respect to that equation, ALPGEN ([95]) uses a definition of α_s which includes terms of higher orders in $ln(Q^2/\lambda^2)$. We report it below:

$$\alpha_s(Q) = \frac{1}{b_5 log(Q^2/\Lambda_5^2)} - \frac{b_5^{'}}{b_5 log^2(Q^2/\Lambda^2) loglog(Q^2/\Lambda^2)} \tag{8.6}$$

where b_5 and $b_5^{'}$ are respectively the 1- and 2-loop coefficients of the QCD β function for 5 flavors. The chosen factorization scale Q^2 controls the strong coupling. Therefore, different choices of the factorization scale may change the amount and the spectrum of the initial state radiation. In V+jets events a natural choice for the factorization scale is:

$$Q^2 = (M^V)^2 + P_T^2 \tag{8.7}$$

where P_T^2 is the squared sum of transverse energies of all final state partons.

The $\pm\ \sigma$ systematic uncertainties on the factorization scale are defined by respectively doubling and halving the value in Eq. 8.7.

Acceptance Uncertainties

The event rate for MC-based processes depends on the acceptance estimation (Eq. 8.4): a wrong acceptance will bias the rate predictions of the processes in the sample. We consider the following acceptance uncertainties:

- *Luminosity:* A 6% systematics is considered (Sec. 3.6).
- *ISRFSR:* Following [132], we assign a 2.5% systematic uncertainty due to more/less ISR and FSR.
- *PDF:* We quote the same systematics determined in the search for the process $WH \to l\nu b\bar{b}$ that was found to be 2.0% ([127]).
- *Lepton acceptance*: We quote 2.2% as systematics due to uncertainties in the trigger acceptance and in the scale factors applied on the lepton reconstruction efficiencies (Sec. 8.3.1).

The above uncertainties are added in quadrature: an overall 7.1% is used for single-top and diboson processes. They are not applied to the V+jets rate since such a rate is derived from fitting the data ($\not\!\!E_T$ fit - Sec. 8.3.3), and they are not applied to the $t\bar{t}$ to avoid double-counting (the $t\bar{t}$ rate is normalized by using the CDF measured cross-section, which already accounts for such the acceptance uncertainty [133]).

8.6 Fitter Validation

Before looking at the data, the fitting procedure must be validated. Validation is performed over a large number of pseudo-data samples (also called pseudo-experiments). Each pseudo-data sample is created as follows.

1. Generate the nuisance parameters in the fit (Tab. 8.4) by using Gaussian distributions with nominal values and uncertainties equal to the means and standard deviations respectively. Some generated nuisance parameters include the normalization N^i of the template $i = 1, \ldots, 5$ used in the fit (Sec. 8.5.1).
2. A new distributions is generated by drawing N^i di-jet masses from each ith template. The drawing is performed bin-by-bin, by throwing random numbers according to a Poisson distribution, whose mean is the bin content of the ith template.

Systematic	Diboson	$t\bar{t}$	single-top	V+jets	QCD
Cross-section Normalization	± 6.1%	± 6.5%	6.3%	uncostrained	14.1%(*)
JES shape	✓	✓	✓	✓	-
JES rate	± 4.3%	± .5%	±0.5%	-	-
Q^2 (shape)	-	-	-	✓	-
Luminosity	$\pm^{6.0\%}_{6.0\%}$				-
ISR/FSR	$\pm^{2.5\%}_{2.5\%}$				-
PDF	$\pm^{2.0\%}_{2.0\%}$				-
Lepton Acceptance	$\pm^{2.2\%}_{2.2\%}$				-
Overall Acceptance	$\pm^{7.1\%}_{7.1\%}$				-

Table 8.4: Summary of the systematic uncertainties considered for the di-jet mass simultaneous fit in the electron and muon samples. Uncertainties marked with * are treated as uncorrelated in the the two samples. The other uncertainties are treated as 100% correlated. Both shape and rate uncertainties with the same label (e.g: JES) are also considered 100% correlated. Rate uncertainties have been symmetrized.

We chose to generate 20,000 pseudo-data samples. We check the differences between the fitted and the generated values in unities of fitted uncertainties (pulls). Pull distributions, shown in App. E.6 are compatible with normal distributions, as expected for unbiased fits.

8.7 Results

The fit to the data is performed in the electron and muon channels simultaneously, as described above. The fit results are shown in Fig. 8.15: the di-jet mass distributions in the electron and muon samples are stacked together and superimposed to the model. Each template in the model is adjusted to account for the variations in the post-fit nuisance parameters with respect to the nominal values (Fig. 8.16). We observe a good agreement between the predictions and the data. The $\chi^2/NDOF$ [18] $= 67.21/55 = 1.20$, corresponding to a goodness of fit probability of about 12.5%. The post-fit nuisance parameters are in good agreement with the pre-fit

[18] NDOF is the number of degrees of freedom, computed as the number of bins in the data histogram (56), minus the number of unconstrained nuisance parameters in the fit (1).

expectations. The JES nuisance parameter shows the largest deviation from the expectations: the fit prefers to pull the JES to negative values, thus

- decreasing/increasing the quark/gluon JES,
- accepting less/quark-jets in the sample.

Both effects contribute to broaden the di-jet mass shape of the model, as it can be seen from Fig E.12 .

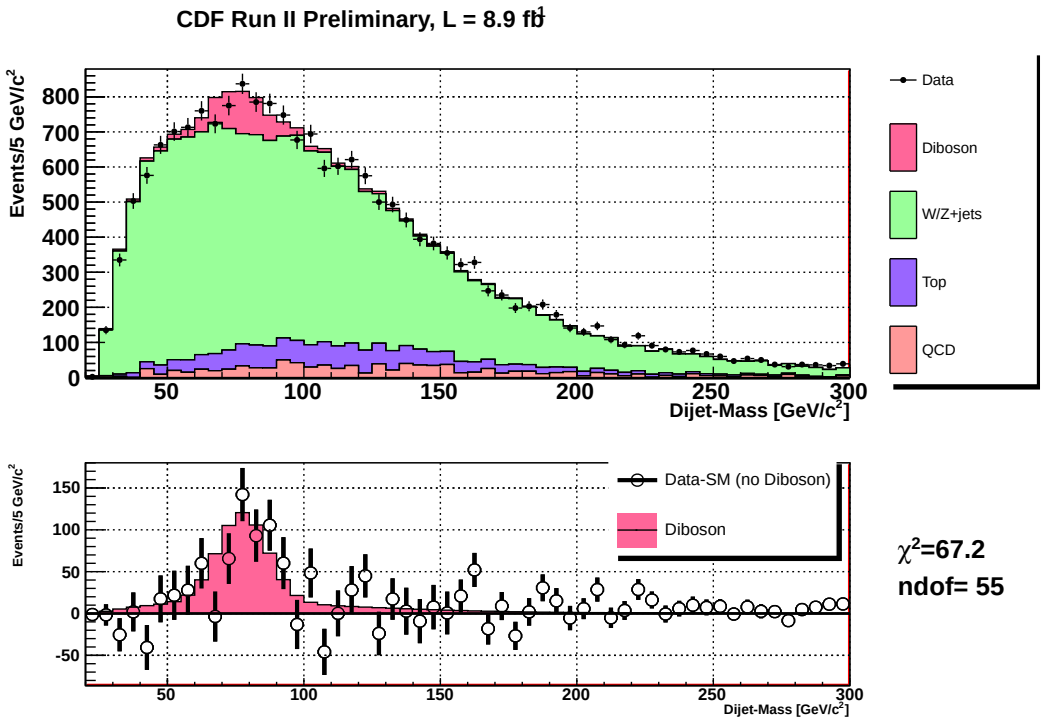


Figure 8.15: *Di-jet mass distributions in the electron and muon data samples. The fitted test hypothesis is superimposed. Each template in the model is adjusted to account for the variation in the nuisance parameters with respect to the nominal values. The difference between data and the model is shown along with the fitted diboson distribution (lower pad).*

Since our goal is to compare with the previous CDF result ([119]), we repeat the same fitting procedure with the inclusion of an additional template, modeled with a Gaussian distribution centered at 145 GeV/c^2 and with a width 14.3 GeV/c^2 [19].

[19]The Gaussian is chosen as the simplest hypothesis compatible with a two-jet decay of a narrow resonance "X" with definite mass produced in association with a leptonically-decaying W. $\sigma = \sigma_W\sqrt{145/M_W^{reco}}$ =14.3 GeV/c^2, where σ_W = 10.7 GeV/c^2 and M_W^{reco} = 77.7 GeV/c^2 are the W mass and width from a Gaussian fit to the reconstructed di-jet mass in simulated WW events.

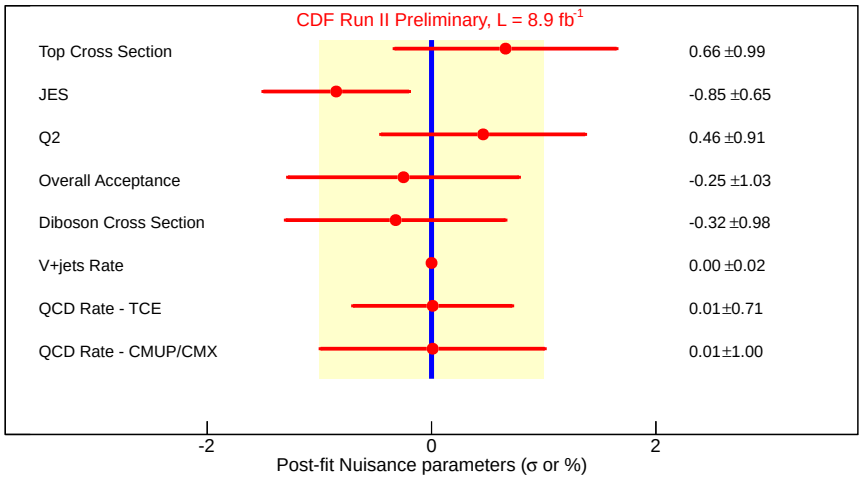


Figure 8.16: *Variations of the fitted nuisance parameters with respect to the nominal values. The variations are shown in percentage of the pre-fit values for unconstrained parameters, or in percentage of the pre-fit uncertainty (σ) for constrained parameters. The $\pm 1\sigma$ band is also shown in light yellow.*

The normalization of the Gaussian template (i.e.: WX rate) is left unconstrained in the fit. When fitting with such a model, the fitted values are unchanged with respect to those shown in Fig. 8.16, and the fitted WX rate is exactly 0.

8.7.1 Upper Limit on the Narrow Resonance Cross-Section

We found no excess at 145 GeV/c^2 in the di-jet mass distribution of the W+2jets sample. We proceed to estimate the upper limit on the the cross section (σ_{WX}) of a resonance of that mass produced in association with a W and decaying in two jets. In order to do that we assume that $A_{WX} = A_{WH}$, A being the MC-derived acceptance, and H the SM Higgs with 150 GeV/c^2 mass. Such an acceptance is estimated from a PYTHIA generated sample.

We estimate the limit by constructing the Feldman-Cousin (FC) bands ([134]), as shown in Fig. 8.17. Each band is drawn as follows:

- $10,000$ pseudo-experiments are generated with a given scale-factor $k \equiv \sigma_{generated}[pb]/3.1$ for σ_{WX}. A scale factor of 1 means a $\sigma_{WX} = 3.1$ pb, which is our reference value [20].

[20] In an updated version of the analysis presented in [119], a 3.1 pb cross-section was estimated for the narrow resonance. A WH ($M_H = 150$ GeV/c^2) simulated sample was used to estimate A_{WH} and derive the above cross-section.

- The distribution of the measured cross-sections is used to find the range that meets the desired coverage threshold (68% or 95% probability).

We build 12 bands with k ranging from 0.1 to 1.2 in steps of 0.1. We estimate an upper limit $\sigma_{WX} < 0.9$ pb at the 95% C.L (Fig. 8.17).

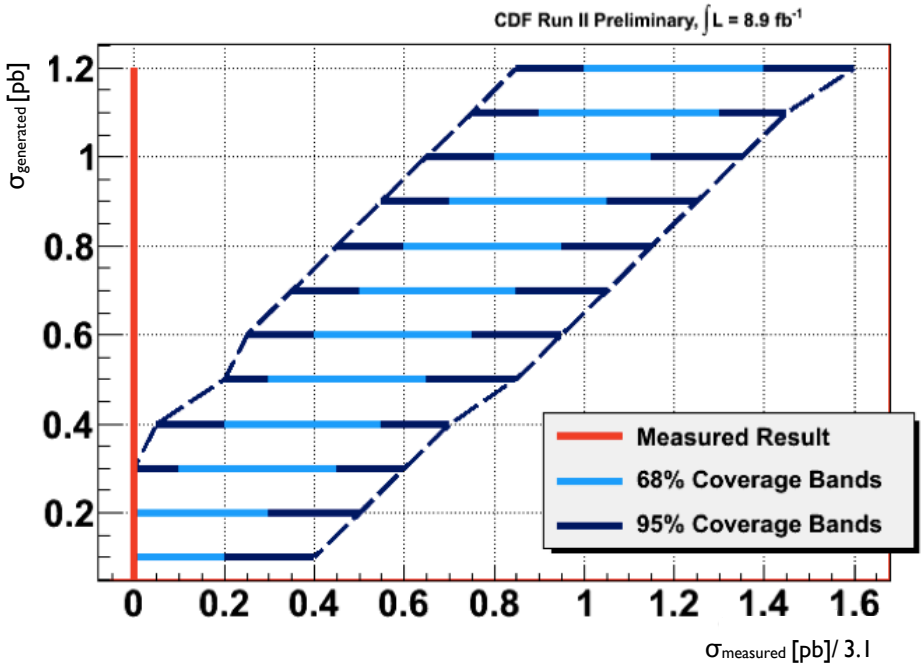


Figure 8.17: Feldman-Cousins bands showing the expected range of measured cross sections as a function of the true cross section, with 68% CL (light blue region) and 95% CL (dark blue region). The measured cross-section is (0.0 ± 0.7) pb. The intersection of the measured cross-section and the 95% CL bands corresponds to the upper limit $\sigma_{WX} < 0.9$ pb at the 95% C.L.

As uncertainty of the WX process we use only the acceptance systematic uncertainties on the rate (7.1%), which was mentioned earlier.

8.8 Conclusions

In this chapter we studied the invariant mass spectrum of two exclusive jets produced in association with a leptonically-decaying W boson. Since the last CDF publication on this topic ([119]), a number of systematic effects were investigated.

The most important ones, affecting the jet energy scale and the QCD-multi jets background, resulted in significant corrections to be applied to our background models. Using all data collected by CDF, and after the above effects have been corrected, we observe a good agreement between data and the Standard Model predictions.

Additional studies on the W+2jets sample and further details about the impact on the di-jet mass spectrum of each of the above systematic effect are described in App. E.2.

Chapter 9

WZ/ZZ Search: Optimization and Results

In the previous chapter we were able to achieve a good agreement between the SM predictions and the data in the W+2jets sample. Therefore, we can focus in measuring the WZ/ZZ cross section in the sample. In this chapter we first present a number of studies aiming to optimize this search (Sec. 9.1-9.3), and then we describe the fitting procedure (Sec. 9.4), which was slightly changed with respect to the one described in the previous chapter. In Sec. 9.5 we extract the WZ/ZZ cross section. Finally, in Sec. 9.6, we estimate the significance of our measurement.

9.1 Extending the Online Acceptance of W+2jets Events

So far, we analyzed events where a central electron or muon fires the trigger. Other trigger paths at CDF are thought to isolate W+2jets events with fairly low background rates. These triggers exploit features of the W+2jets events somehow orthogonal to the central lepton triggers. In the following sections we describe the two additional triggers that were investigated for this analysis. When listing the trigger requirements, we will refer to the raw missing transverse energy (Eq. 4.15), rather than the fully corrected one.

9.1.1 Forward Electron Triggers

From simulations we expect a large number of leptonically-decaying Ws to direct the lepton in the forward region (see Fig. 9.1). Therefore, we decide to investigate

the standard CDF forward electron trigger, named MET_PEM, whose requirements are described below.

- L1:
  - an electromagnetic tower with $E_T > 8$ GeV in the forward region (1.1 $< |\eta_{det}| <$3.6)
  - $\not\!\!E_T >$ 15 GeV.
- L2:
  - a reconstructed electromagnetic cluster of $E_T > 20$ GeV
  - confirm of $\not\!\!E_T >$ 15 GeV.
- L3:
  - confirmation of L2 requirements

Triggering on events in the forward region is particularly challenging because of the following two reasons: no tracking information can be used efficiently because of the limited COT coverage (see Sec. 3.3), and a larger rate of QCD multi-jets background is expected [1]. Such a large QCD rate dictates tighter online cuts in order to keep the trigger rate low.

The MET_PEM trigger efficiency is measured in two steps. The first step measures the trigger efficiency due to the $\not\!\!E_T$ requirement. This is done in datasets triggered with looser $\not\!\!E_T$ cuts, but similar cuts on the electro-magnetic cluster. The second step is a measure of the efficiency due to the electromagnetic cluster. It is measured in the same way as for central muon triggers (Sec. 8.1.2). The exploited dataset is enhanced of $Z \to e^+e^-$ events, where one electron is required to pass the central electron trigger requirements (Sec. 8.1.1). The MET_PEM trigger efficiencies $\epsilon(x)$ are parameterized as follows ([135]):

$$\epsilon(x) = \frac{1}{1 + e^{-A(x-B)}} \tag{9.1}$$

where $x \equiv \not\!\!E_T$ or cluster E_T. A, B are reported in Table 9.1

Further details about the forward electron trigger are described in [135], [136].

[1] Because of the possible large longitudinal boost in the initial state and the energy-momentum conservation, jets have higher chance to be directed more forward when no heavy resonance are produced in the final state. Moreover, jets from the initial state radiation are more likely to be produced along the beam line because of the collinear nature of QCD gluon radiation.

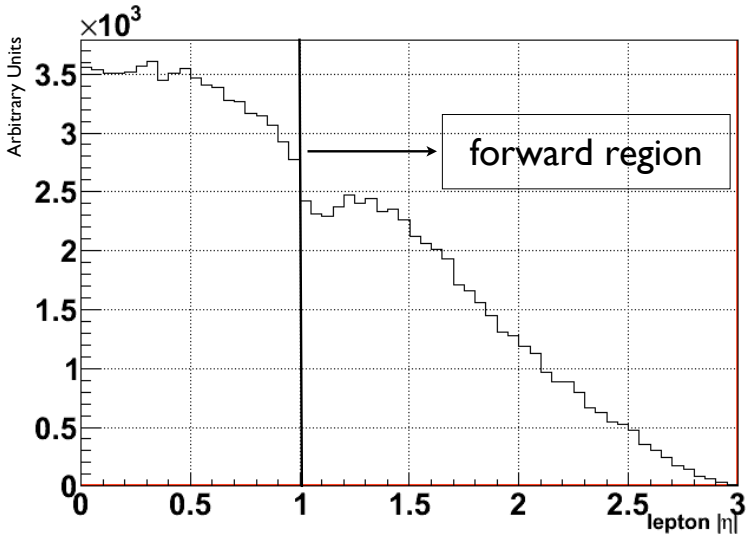


Figure 9.1: $|\eta|$ *distributions of decay electrons within the CDF detector acceptance in* $W(\rightarrow e\nu)$*+2jets ALPGEN+PYTHIA events. The region covered by the forward calorimeters is indicated.*

Observable	A	B
$\not\!\!E_T$	15.3 ± 0.2	0.336 ± 0.013
Cluster E_T	22.6 ± 0.1	0.802 ± 0.044

Table 9.1: Coefficients of the MET_PEM parameterization defined in Eq. 9.1.

9.1.2 Missing Transverse Energy Plus Jets Triggers

By looking at Fig. 8.2, one can notice that some of the $W \rightarrow \mu\nu$ acceptance is lost. This is due to non-instrumented regions, limiting the muon identification efficiency. A way to recover those events is to trigger on information not requiring the presence of an energetic charged lepton: $\not\!\!E_T$ or high-energy jets. The three available trigger paths exploiting that information are named MET45 and MET + 2jets (the latter contains two trigger paths). At least two out of the three trigger paths were active at the same time: this makes the investigation of all of them a slightly involved procedure. An optimal way to combine these three triggers was proposed within the Higgs group at CDF ([137]). However, we decided to simplify the problem and to investigate the only active $\not\!\!E_T$ + jets trigger path (MET2J) for the first set of data ($\mathcal{L} \sim 2.3$ fb^{-1}), and the other one, named METDI, for the last part ($\mathcal{L} \sim$

6.6 fb^{-1}). Since METDI trigger was tailored for Higgs boson searches, it provides higher sensitivity on Higgs boson searches than MET2J ([138]) [2]. We did not use MET45 because of the relatively limited WZ/ZZ acceptance.

The MET2J trigger has been active for the whole CDF Run II. MET2J trigger was dynamically prescaled at L2 with a factor between 1 and 40 [3]. Trigger requirements are listed below [4].

- L1:
  - $\not\!\!E_T >$ 28 GeV.
- L2:
  - $\not\!\!E_T >$ 30 GeV;
  - $\geq$1 jet with $E_T > 20$ (25) GeV, $|\eta_{det}| < 1.1$, and the other jet with $E_T >$ 15 (10) GeV, $|\eta_{det}| < 3.6$.
- L3:
  - $\not\!\!E_T >$ 35 GeV.

We list the METDI trigger requirements below.

- L1:
  - Hadronic tower with $E_T > 10$ GeV;
  - $\not\!\!E_T > 28$ GeV.
- L2:
  - confirm of $\not\!\!E_T >$ 28 GeV;
  - at least one hadronic cluster with $E_T > 3$ GeV and $|\eta_{det}| < 3.6$, in addition to the L1 tower.
- L3:

[2] METDI takes advantage of the L2 calorimeter upgrade providing better energy resolution (Sec. 3.7) .

[3] The prescale is needed to maintain the trigger rate below some maximum frequency. The prescale factor is defined as the reduction factor of the trigger rate. A dynamically prescaled trigger is realized with a prescale factor which depends on the instantaneous luminosity.

[4] A number of trigger paths contribute to MET2J. Different trigger paths may have slightly different jet E_T thresholds. Additional thresholds are shown in parentheses.

– $\not\!\!E_T >$ 30 GeV.

With respect to MET2J trigger requirements, the $\not\!\!E_T$ and jet E_T thresholds of the METDI were varied. Moreover, a jet E_T requirement was moved at L1, and no centrality on the jet was required any longer. It was possible to vary these requirements mainly because of the improved energy resolution provided by the upgraded L2 calorimetric trigger (Sec. 3.7). Because of these changes the METDI was never prescaled.

Trigger parameterization

The MET2J and METDI trigger efficiencies are measured in the central muon triggered sample (Sec. 8.1.2). Such a sample is well understood, thus providing a good handle for the $\not\!\!E_T$+jets trigger efficiency studies. With respect to Sec. 8.2, we added the selection cuts reported in Table 9.2 in order to reduce possible effects due to trigger biases [5]. These cuts will be used also when extracting the WZ/ZZ signal.

Trigger	Additional Requirement
MET2J	$\Delta R(j_1, j_2) > 0.9$
	at least one jet with $\eta_{det} < 0.9$
METDI	at least one jet with $E_T > 40$ GeV

Table 9.2: *Additional requirements to the ones listed in Table 8.1 when analyzing the $\not\!\!E_T$ + jets triggered data. Different cuts are given for the different used trigger paths.*

Offline, we parameterize the trigger efficiency curves as a function of the missing transverse energy and of the transverse energy of the leading jet ($E_T^{jet_1}$). In order to do that, we need offline variables as close as possible to those exploited by the trigger. The natural choice for missing transverse energy would be the uncorrected one ($\not\!\!E_T{}^{raw}$, Eq. 4.15), but such a variable is not well modeled by our MC. A better

[5] As it can be seen above the MET2J triggers requires at least one jet with $|\eta| < 1.1$, while the METDI requires one hadronic tower with $E_T > 10$ GeV. Offline, we require the second and third cut reported in Table 9.2 to avoid biassed distributions respectively in the MET2J and METDI triggered data. $\Delta R(j_1, j_2) > 0.9$ is applied since, as the Tevatron luminosity increased, energetic towers from multiple interactions caused two clusters to merge into a single jet, thus resulting in a lower online efficiency for two-jets events ([139]). Thanks to the L2 calorimeter trigger upgrade (Sec. 3.7), the efficiency loss was resolved.

modeling is observed when the $\not\!\!E_T{}^{raw}$ is recalculated by using the z position of the primary interaction vertex and when jet corrections are applied ($\not\!\!E_T{}^{JEScorr}$, Eq. 4.16). If we neglect the small muon calorimeter deposit, this variable can be thought as the transverse energy of the W-boson candidate. The trigger efficiencies are parameterized as:

$$\epsilon(x) = \frac{c}{1 + e^{\frac{(a-x)}{b}}} \tag{9.2}$$

where $x \equiv \not\!\!E_T^{JEScorr}$ or $E_T^{jet_1}$. Efficiency curves along with the parameterizations are shown in Fig. 9.2. The parameters extracted from the fit are summarized in Table 9.3. The slight inefficiency at the plateau is due to the saturated towers at L1, which is a known feature of the calorimeter trigger before the upgrade (Sec. 3.7).

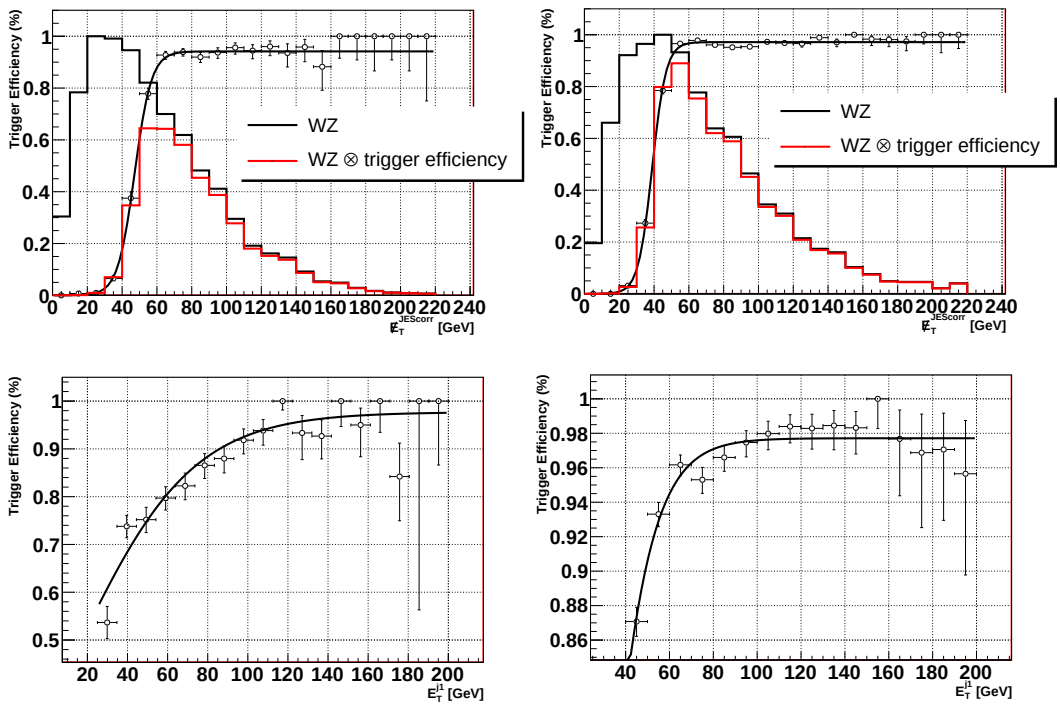


Figure 9.2: *Turnon efficiency curve parameterizations for MET2J (left) and METDI (right) triggers. Data collected until period 15 (left) and in later periods (right) were parameterized independently as* $\epsilon(x) = \frac{c}{1+e^{\frac{(a-x)}{b}}}$, *where a, b, c are given in Table 9.3, and x is* $\not\!\!E_T^{JEScorr}$ *(top) or* $E_T^{j_1}$ *(bottom). In the upper plots WZ MC distributions are shown before (black) and after the convolution with the trigger efficiency parameterization (red).*

The trigger efficiency parameterizations in Eq. 9.1 will be used for modeling the

Trigger path	Variable	a	b	c
MET2J	$\not{E}_T{}^{JEScorr}$	47.17 ± 0.42	4.85 ± 0.27	0.94 ± 0.01
	$E_T^{jet_1}$	15.63 ± 4.08	28.59 ± 4.75	0.98 ± 0.02
METDI	$\not{E}_T{}^{JEScorr}$	38.90 ± 0.22	3.99 ± 0.12	0.97 ± 0.00
	$E_T^{jet_1}$	19.32 ± 5.61	12.06 ± 2.47	0.98 ± 0.00

Table 9.3: Coefficient of the MET2J and METDI trigger parameterizations $\epsilon(x) = \frac{c}{1+e^{\frac{(a-x)}{b}}}$.

W+2jets sample. In order not to correct the MC twice, the $E_T^{jet_1}$ parameterization of the trigger will only be used to modify the shapes. The rate does not change after applying such a parameterization. The trigger efficiency for our WZ/ZZ signal, estimated by convoluting the $\not{E}_T$-efficiency curves with the signal $\not{E}_T$spectrum from MC (Fig. 9.2), is about 51% for MET2J and 68% for METDI. The uncertainties of these values will be accounted for in the final fit.

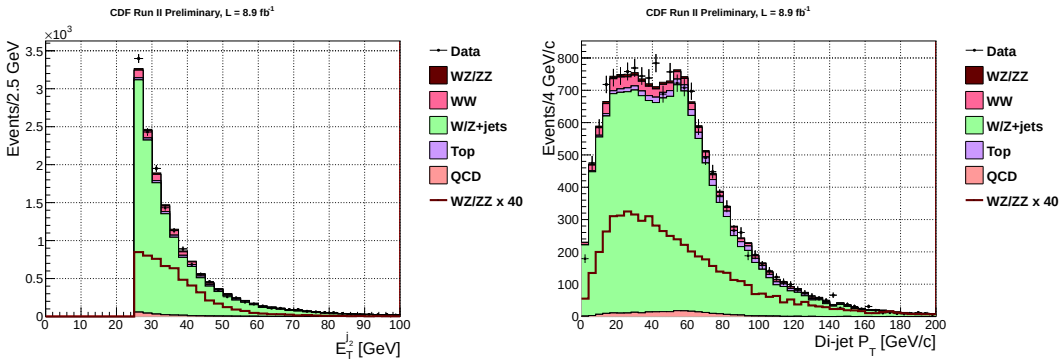
9.2 Optimization of the Selection Cuts

The selection cuts described in Sec. 8.2 are not optimal for this search. In Fig. 9.3 we show the $E_T^{jet_2}$ and di-jet P_T distributions in data and simulations: it's clear that cuts at $E_T^{jet_2} > 30$ GeV and at di-jet $P_T > 40$ GeV would reject too many signal events [6].

For this reason we lower the former cut and we remove the latter. A number of other studies are performed in order to accept as much signal as possible, while keeping the background low. The final selection requirements are summarized in Table 9.4. The cut on the number of leptons is removed to increase the ZZ signal acceptance. The cut on the number of jets is tightened compared to what done previously: extra jets with $E_T > 13$ GeV and $|\eta| < 2$ are not allowed. This is done in order to keep the 2-jet and 3-jet regions separated from each other, since they are characterized by a different signal over background ratio and by different background contaminations [7]. With respect to Table 8.1, we remove $\Delta\phi(j_1, \not{E}_T)$ >0.4 cut to further increase the signal rate. We are not worried about allowing for more QCD

[6] The $S/\sqrt{B}$, S and B being respectively the signal (WZ/ZZ) and the background (data minus signal) rate in the sample, decreases by 10% (25%) when applying the $E_T^{jet_2} > 30$ GeV (dijet $P_T > 40$ GeV) cut.

[7] The investigation of the 3-jet region is in progress.



(a) (b)

Figure 9.3: $E_T^{jet_2}$ *(a) and di-jet* P_T *(b) predicted versus observed distributions in the CMUP/CMX muon sample. The normalization of each process is described in Chap. 8. Compared to the selection cuts described in Sec. 8.2, the* $E_T^{jet_2}$ *cut is lowered and the* $P_T(j_1, j_2)$ *cut is removed. The* WZ/ZZ *signal (dark red) is magnified by 40.*

multi-jets events in the sample, since this background can be accurately subtracted (the relative systematic uncertainty is 14.1%) and is well modeled (Sec. 8.3.3).

Our QCD multi-jets veto, depending upon the lepton category, is described below.

- CMUP,CMX: $M_T^W > 10$ GeV;
- TCE: $M_T^W > 30$ GeV;
- PHX: $M_T^W > 30$ GeV, $\not\!\!E_T$>25 GeV, $\Delta\phi^{min}(jet,\not\!\!E_T) > 0.3$, where all jets with $E_T > 5\ GeV$ are considered;
- CMU,CMP,CMIOPES: $M_T^W > 10$ GeV;
- CMIOCES: $M_T^W > 10$ GeV, $\delta\phi(jet_2,\not\!\!E_T) > 0.3$;
- CRKTRK: $M_T^W > 20$ GeV, $\delta\phi(jet_2,\not\!\!E_T) > 0.3$;

As shown above, the QCD veto is looser for CMU, CMP, CMUP, CMX and tighter for PHX, CRKTRK: different lepton identification cuts (Chap. 4) allow for different QCD background contamination in the sample. Moreover, for the PHX leptons, the cut is even tighter because a larger QCD background is expected in the forward region.

Variable	Requirement
# vertices	> 0
Lepton type	TCE, CMUP/CMX, EMC, PHX
Lepton $E_T > (P_T)$	> 20 GeV (GeV/c)
$\not\!\!E_T$	$>$20 GeV
# jets	$= 2$
jet E_T	> 20 GeV
jet η	< 2
QCD veto	
$t\bar{t}$ veto (Tag region only)	
$\not\!\!E_T$ + jets trigger cuts (EMC only)	

Table 9.4: Summary of the event selection. EMC is one of the following muon categories: CMU, CMP, CMIOCES, CMIOPES, CRKTRK (see Sec. 4.3). We identify them in the $\not\!\!E_T$ +jets triggered data (Sec. 9.1.2), while TCE, CMUP/CMX come from the central high-P_T lepton triggers (Sec. 8.1), and PHX from the forward electron trigger (Sec. 9.1.1). The QCD veto is described in the text. The $t\bar{t}$ veto, which is applied only in the Tag region, is described in App. F. The $\not\!\!E_T$ + jets trigger cuts (Table 9.2) are applied only to EMC leptons.

9.3 Splitting of the Pre-tag Sample: Heavy Versus Light Flavor Jets

The selection described in Table 9.4 defines the inclusive "pre-tag" sample. However, the signal over background ratio in a b-jets enriched sample is expected to be larger than in the inclusive one [8]. When searching for the WZ/ZZ signal, a common attitude at the Tevatron (and LHC) has been to isolate the b-jets sample and not to consider the rest of the pre-tag sample any more [9].

We report below some studies proving that the light-jets sample can be efficiently used for the WZ/ZZ search and should not be neglected. We proceed in two steps: we start from defining the b-jets enriched region ("Tag region"), and then we prove that more accurate results can be achieved by investigating also the light-jets enriched region ("NoTag region"). Finally, we demonstrate that higher

[8] The BR($Z \to b\bar{b}$)/BR($Z \to$hadrons)~ 0.2 ([10]), while $\sigma_{W+b-jets}/\sigma_{W+jets} \sim 3 \times 10^{-3}$.

[9] This was done to pave the way to searches of the SM light Higgs boson, which features the exclusive $b\bar{b}$ decay signature.

performances are obtained by analyzing separately the two regions.

9.3.1 Definition of the Tag region

We identify b-jets by using the jet b-ness algorithm (Sec. 5.2). Since the jet b-ness provides a continuous output, the b-tagging requirements on the jets can be optimized against the significance of the measurement. This is done by considering the WZ with the $Z \to b\bar{b}$ as signal [10]. Jets are ordered in decreasing jet b-ness rather than in decreasing E_T. The used figure of merit for the optimization is the probability of a 2σ signal evidence ("P2σ"). P2σ relies on the definition of the test (H_1) / null (H_0) hypotheses, which are models built with the SM processes listed in Sec. 7.2, when including / excluding the WZ/ZZ signal. The computation of P2σ is as follows.

1. $50,000$ pseudo-experiments (PE's) are generated under the null hypothesis and the value of $\Delta\chi^2 \equiv \chi^2(H_1) - \chi^2(H_0)$ is computed, the χ^2 being the result from the fit described in Sec 8.5. The only difference is that the WZ/ZZ is now separated from the WW and is allowed to vary unconstrained. No systematic uncertainties other than the cross-section/normalization ones are used at this point.

2. The $\Delta\chi^2_{2\sigma}$ value is located from the distribution of the $50,000$ $\Delta\chi^2$ such that the probability to have $\Delta\chi^2 < \Delta\chi^2_{2\sigma}$ is 2.3% [11].

3. $50,000$ PEs are generated under the test hypothesis and P2σ is computed as the integral from $-\infty$ to $\Delta\chi^2_{2\sigma}$.

The result of the optimization procedure for this WZ/ZZ search is plotted in Fig. 9.4: the most optimal b-ness requirements for the first and second jets are respectively $b\text{-ness}^{jet_1} > 0.75$ and $b\text{-ness}^{jet_2} > -0.2$. By applying these requirements we find P2$\sigma \sim 7\%$. The fraction of events with at least one b-jet is expected to be about 60% for this operation point.

[10] If we consider the inclusive WZ as signal, our procedure would choose the pre-tag region to be the most optimal "tag" region, in order not to lose the large amount of Z bosons decaying in light-quarks. This suggests that the best strategy would be to perform the analysis in the single full data sample. However, the most correct optimization procedure, which is not feasible because of the large required computing time, should define the two regions (Tag and NoTag) at the same time. At the end of the section it would be clear that the optimal scenario is separating the inclusive data sample in the NoTag and Tag regions.

[11] 2.3% is the integral from 0 to 2 and of a normal distribution

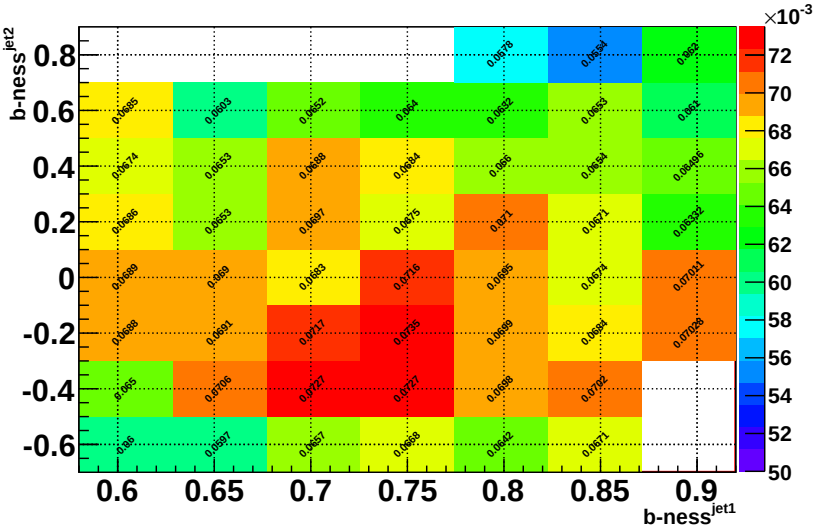


Figure 9.4: *P2σ as a function of the minimum requirements on the jet b-ness of the first and second jets. P2σ is estimated from 50,000 PEs, and by assuming that WZ with the $Z \to b\bar{b}$ is our signal (see text for further details). The b-ness$^{jet_1} > 0.75\times$ b-ness$^{jet_2} > -0.2$ point scores slightly higher than the neighbor points.*

9.3.2 The Importance of the NoTag Region

The NoTag region is defined from the pre-tag region after the subtraction of the Tag region. In order to show that the inclusion of the NoTag region improves the significance of the measurement, we compare the P2σ value obtained when fitting the tag region only, versus fitting the Tag and NoTag region simultaneously. The fits are performed as described in the previous section. When fitting the two regions simultaneously, the nuisance parameters are considered fully correlated across the regions. We find that by including the no-tag region P2$\sigma \sim 19\%$, which is almost a factor of 3 higher than the one obtained when fitting the tag-region only.

Moreover, we estimate that P2$\sigma \sim 14\%$ for the inclusive pre-tag region. By comparing such a result with the previous one, we conclude that the splitting of the pre-tag region in the NoTag and Tag regions improves the significance of the measurement. This result was expected, given that the signal over background ratios of the two regions are very different.

9.4 Fitting Procedure

The fitting procedure to extract the WZ/ZZ cross-section is very similar to the one described in Sec. 8.5.1. We describe the differences below.

The model that we will be using as test hypothesis contains the following templates:

1. $WZ + ZZ$ (signal);
2. WW;
3. Single-top;
4. $t\bar{t}$;
5. V+jets;
6. QCD multi-jets;

The WW is now treated as additional template, while such a template was previously added to the WZ and ZZ into the diboson template.

The null hypothesis is performed as the test hypothesis, after excluding the signal template.

Under each hypothesis, the fit is performed simultaneously in eight regions, separated according to the triggered lepton type (TCE, CMUP/CMX, PHX, EMC) and jet flavor (Tag, NoTag regions). When defining the tag regions i) the $t\bar{t}$ veto (App. F) and ii) the optimized b-ness$^{jet_1} > 0.75$ and b-ness$^{jet_2} > -0.20$ cuts are applied to pre-tag data. In simulated events, rather than applying a renormalization of b and non-b jets based on the calculated scale factors for the efficiency and mistag rate (Sec. 5.2.3), we locate the equivalent cuts in the Monte Carlo that matches respectively the measured b-tag efficiency and mistag rates in the data: respectively b-ness$^{jet_1} > 0.83$, b-ness$^{jet_2} > 0.07$ and b-ness$^{jet_1} > 0.68$, b-ness$^{jet_2} > -0.33$. The former are applied to MC samples containing generated b-quarks, the latter to the remaining samples. The criteria to isolate the Tag region from the pre-tag region in data and MC are summarized in Table 9.6. As already mentioned, the NoTag region is defined as the pre-tag region minus the Tag region. The predicted rates for each region are listed in Table 9.5.

Process	Region	Rate			
		TCE	CMUP,CMX	PHX	EMC
Signal	NoTag	165.1 ± 7.6	138.9 ± 6.4	60.6 ± 2.8	63.8 ± 2.9
	Tag	9.0 ± 0.4	7.6 ± 0.3	3.9 ± 0.2	3.1 ± 0.1
WW	NoTag	872.2 ± 50.6	589.2 ± 34.2	277.6 ± 16.1	341.2 ± 19.8
	Tag	5.7 ± 0.3	3.5 ± 0.2	2.1 ± 0.1	1.3 ± 0.1
tt	NoTag	302.7 ± 19.7	214.2 ± 13.9	52.3 ± 3.4	144.8 ± 9.4
	Tag	35.5 ± 2.3	27.9 ± 1.8	7.9 ± 0.5	15.4 ± 1.0
Single-top	NoTag	137.7 ± 8.7	98.3 ± 6.2	28.5 ± 1.8	71.5 ± 4.5
	Tag	22.6 ± 1.4	16.7 ± 1.1	4.5 ± 0.3	8.2 ± 0.5
Vjets	NoTag	18539.7 ± 185.4	13683.4 ± 136.8	7008.1 ± 140.2	7125.2 ± 142.5
	Tag	169.9 ± 1.7	137.6 ± 1.4	38.4 ± 0.8	63.4 ± 1.3
QCD	NoTag	3451.6 ± 486.7	126 ± 17.8	2972.8 ± 535.1	130.7 ± 18.4
	Tag	58.3 ± 8.2	5.7 ± 0.8	22.1 ± 4.0	4.6 ± 0.6
Data	NoTag	23469	14850	10400	7825
	Tag	301	199	79	96

Table 9.5: Predicted and observed number of events in the NoTag and Tag samples selected according to the requirements described in Sec. 9.2. W+jets and QCD rates are estimated from the $\not\!\!E_T$ fit in data (Sec. 8.3.3). The expected rates are separated according to the triggered lepton type. We also require the di-jet mass to be within [20,300] GeV/c^2. By construction the overall expected rates are the same as the observed ones.

9.4.1 Systematic Uncertainties

All the systematic uncertainties described in Sec. 8.5.2 are implemented in the fit. Differences and additions are described below.

Cross-Section/Normalization Uncertainties

The WW and $WZ + ZZ$ rates are now uncorrelated and float separately in the fit. The uncertainty of the former comes from the theoretical cross-section: it is 6.2% (Table 7.1). The rate of the latter is left fully unconstrained in the fit.

The other difference concerns the V+jets rate. At CDF, other analyses investigating samples highly contaminated by W/Z+b-jets take an additional 30-40% rate uncertainty on the cross-section ratio (W/Z+b-jets)/(W/Z+jets) (Sec. 7.2). We decided to be more conservative and fully decorrelate the V+jets rates in the Tag and NoTag regions (Table 9.7).

b-tag Efficiency and Mistag Rate Uncertainties

In Sec. 5.2.3 we described the calculation of the b-tag efficiency and mistag rate of the jet b-ness tagger, along with their uncertainties. In the same manner as for the nominal values, we locate the equivalent cuts in the Monte Carlo that matches the $\pm 1\sigma$ uncertainties value on the measured b-tag efficiency and mistag rates in the data. These values are summarized in Table 9.6. The effect of changing the requirement on the b-ness of the two jets is to migrate MC events from the NoTag to the tag regions and viceversa. The differences in rates due to the migration are reported in 9.7. No difference in the template shapes is observed when varying the b-ness cuts. Therefore we consider only the systematic uncertainty associated to the jet b-ness tagger rate.

Process	Shift	$bness^{jet_1}$	$bness^{jet_2}$
b-Jets	Nominal	> 0.83	> 0.07
Non b-Jets		> 0.68	> -0.33
b-Jets	-1σ	> 0.79	> -0.06
Non b-Jets		> 0.65	> -0.43
b-Jets	$+1\sigma$	> 0.86	> 0.18
Non b-Jets		> 0.72	> -0.35
Data		> 0.75	> -0.20

Table 9.6: Jet b-ness requirements applied to the pre-tag region to select the tag region. Different cuts are applied to jets originated from b-quarks (top, V+b-jets, WZ, ZZ) and not b-quarks (WW, V+jets) in data and MC.

9.5 Results

We fit the test hypothesis model to the data. The fit is performed simultaneously in eight channels, as described in the previous section. The post-fit nuisance parameter values are shown in Fig. 9.5. The nuisance parameters are in good agreement with the expectations, even though few of them (JES, Q^2, WW Cross Section, Overall Acceptance) differ from the expectations of $1\sigma - 1.5\sigma$, σ being the expected uncertainty. [12]. The effect on di-jet mass spectrum of the fitted jet energy scale,

[12]The observed deviations in the nuisance parameters are not expected to affect much the measured WZ/ZZ cross-section, since the Gaussian constraints on the nuisance parameters (Sec. 8.5.1) reduce their impact for deviations less than 3σ.

Systematic	Region	WZ/ZZ	WW	$t\bar{t}$	single-top	V+jets	QCD
Cross-section/ Normalization	All	UC	± 5.8%	± 6.5%	6.3%	UC	14.1-18%
JES shape	All	✓	✓	✓	✓	✓	-
JES rate	No-tag	± 4.2%	± 4.3%	± 0.5%	± 0.5%	-	-
	Tag	± 1.8%	± 1.7%	± 0.4%	± 0.4%		
b-tag	No-tag	± 0.5%	± 0.1%	± 4.3%	± 4.3%	-	-
	Tag	± 9.0%	± 3.2%	± 12.2%	± 12.2%		
Q^2 (shape)	All	-	-	-	-	✓	-
Luminosity	All			$\pm^{6.0\%}_{6.0\%}$			-
ISR/FSR				$\pm^{2.5\%}_{2.5\%}$			-
PDF				$\pm^{2.0\%}_{2.0\%}$			-
Lepton Acceptance				$\pm^{2.2\%}_{2.2\%}$			-
Overall Acceptance				$\pm^{7.1\%}_{7.1\%}$			-

Table 9.7: Summary of the systematic uncertainties considered for the di-jet mass simultaneous fit in the no-tag and tag regions. These regions are further separated according to the lepton type, as described in the text. The QCD uncertainties (14.1% and 18% respectively for the TCE, CMUP/CMX, EMC and for the PHX samples) are treated as uncorrelated in all samples. WZ/ZZ and V+jets rates are treated as unconstrained (UC) nuisace parameters. The other uncertainties are treated as fully correlated. Both shape and rate uncertainties with the same label (e.g: JES) are also considered correlated. Rate uncertainties have been symmetrized.

which is lower than the nominal value, was already discussed in Sec. 8.7. The fitted low Q^2 value does not cause of major concern since it does not have a large effect on the di-jet mass spectrum (Fig. E.13).

The most important parameter of interest is the WZ/ZZ rate: we fit (0.93±0.67)% of the expected rate, in agreement with the expectations.

The di-jet mass distributions in the NoTag and Tag regions are superimposed to the predictions, after adjusting the templates and the rates for the observed nuisance parameters (Fig. 9.6). We observe a good agreement between data and predictions in the tag region and a fair agreement in the NoTag region. The overall $\chi^2/NDOF = 334.6/294 = 1.1$ [13], corresponding to a goodness of fit probability of

[13]The minimization procedure described in Sec. 8.5.1 returns such a χ^2.

about 5.1%.

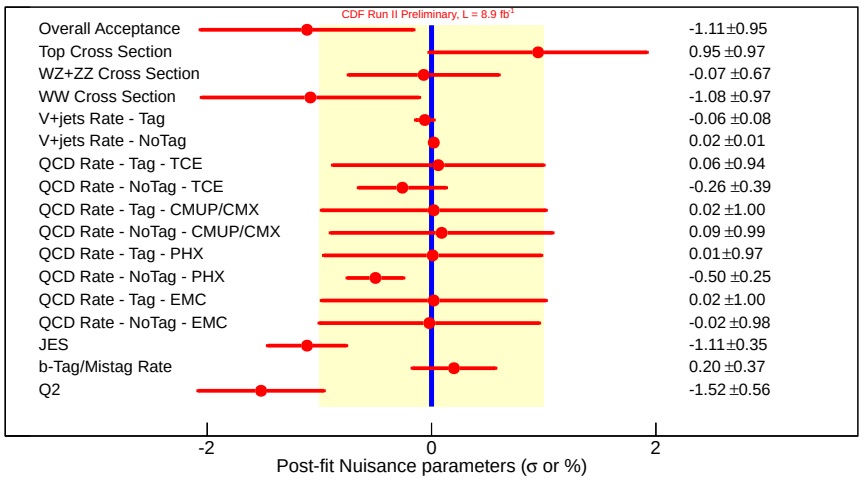


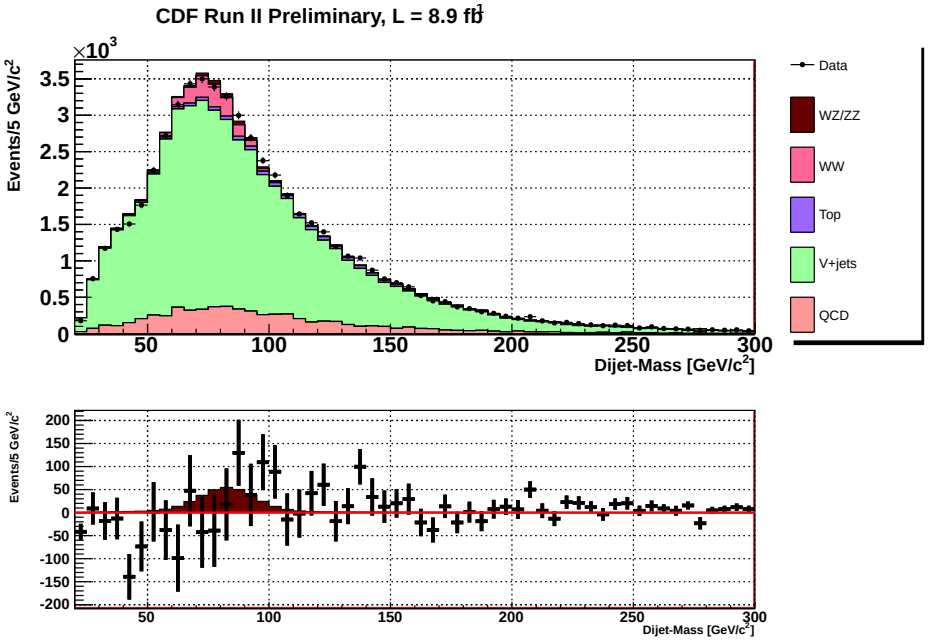
Figure 9.5: *Variations of the fitted nuisance parameters with respect to the nominal values. The variations are shown in percentage of the pre-fit values for unconstrained parameters, or in percentage of the pre-fit uncertainty (σ) for constrained parameters. The $\pm 1\sigma$ band is also shown in light yellow.*

9.5.1 Cross-section Measurement

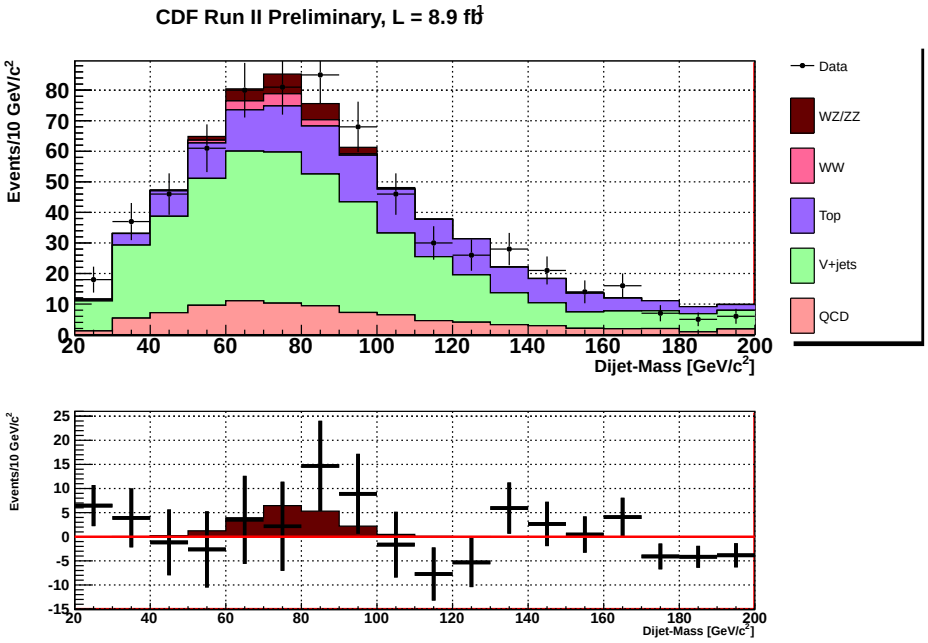
From the fits to the data shown in the previous section, we can extract the cross-section of WZ/ZZ production. We do that by generating the FC bands (Sec. 8.7.1). Each band is generated for a given scale-factor $k \equiv \sigma_{generated}/\sigma^{SM}_{WZ/ZZ}$, $\sigma^{SM}_{WZ/ZZ} = 5.1$ pb (Tab. 7.1). We generate 40 bands with k ranging from 0.1 to 4 at a step of 0.1. The number of pseudo-experiments per band is 300,000. The results are shown in Fig. 9.7. The vertical line indicates the favorite cross section value. We can use the 1σ bands to quote the preferred cross section value as the result of the measurement. We measure a cross-section of $\sigma_{WZ/ZZ} = 4.7 \pm^{3.1}_{2.5}$ pb, 93% of the SM value. This result corresponds to a limit of $\sigma_{WZ/ZZ} < 12.2$ pb ($2.4 \times \sigma^{SM}_{WZ/ZZ}$) at 95% C.L. The expected result, based on drawing a vertical line at 1 on the FC bands, is $\sigma^{exp}_{WZ/ZZ} = 5.1 \pm^{3.6}_{2.5}$ pb.

9.6 Statistical Significance of the WZ/ZZ Measurement

The fitting procedure presented in the previous section is also used to compute the p-value of the measurement, which quantifies our ability to isolate the WZ/ZZ



(a)



(b)

Figure 9.6: *Di-jet mass distributions in all considered lepton categories in the No-Tag (a) and Tag (b) data samples. The fitted test hypothesis is superimposed. Each template in the model is adjusted to account for the variation in the nuisance parameters with respect to the nominal values. In each plot the difference between data and the model is shown along with the fitted WZ/ZZ distribution (lower pad).*

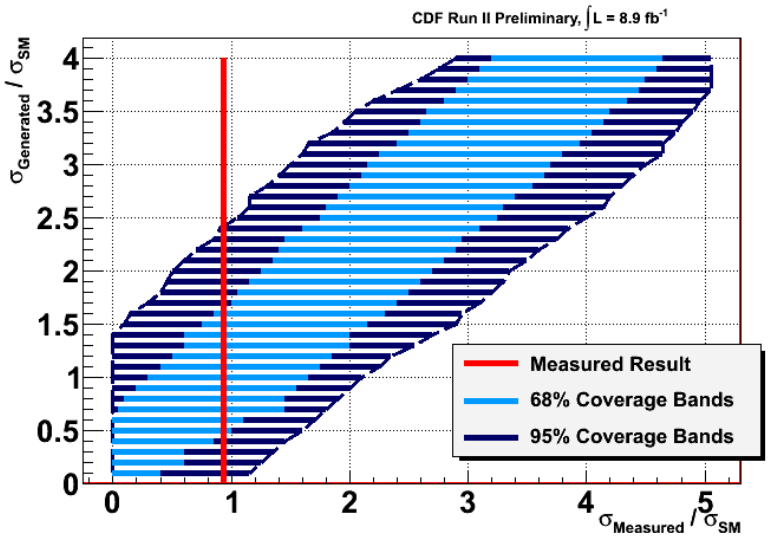


Figure 9.7: *Feldman-Cousins bands showing the expected range of measured cross sections as a function of the SM WZ/ZZ cross-section, with 68% CL (light blue region) and 95% CL (dark blue region). The measured cross section is* $\sigma_{WZ/ZZ} = 4.7 \pm^{3.1}_{2.5}$ *pb. This result corresponds to a limit of* $\sigma_{WZ/ZZ} < 12.2$ *pb at 95% C.L.*

signal from the large background in the W+2jets data. The p-value is computed as follows.

- A number N of pseudo-data samples are generated under the null hypothesis.
- The null and test hypothesis models are fitted to each pseudo-data sample, thus obtaining χ^2_{null} and χ^2_{test} respectively.
- A $\Delta\chi^2 \equiv \chi^2_{test} - \chi^2_{null}$ distribution is built over the N pseudo-data samples.
- $\Delta\chi^2_{data}$ is obtained by fitting the data sample in the same way as the pseudo-data sample.
- The observed (expected) p-values is defined as the integral of the pseudo-data $\Delta\chi^2$ distribution from $-\infty$ to $\Delta\chi^2_{data}$ ($\Delta\chi^2_{med-test}$). $\Delta\chi^2_{med-test}$ is the median of $\Delta\chi^2$ distribution obtained from fitting pseudo-data samples, which are generated under the test hypothesis.

The result of the aforementioned procedure for $N = 300,000$ is shown in Fig. 9.8: the observed (expected) p-value is 0.0807 (0.0665), corresponding to a signal significance of ~ 1.4 (1.5) σ.

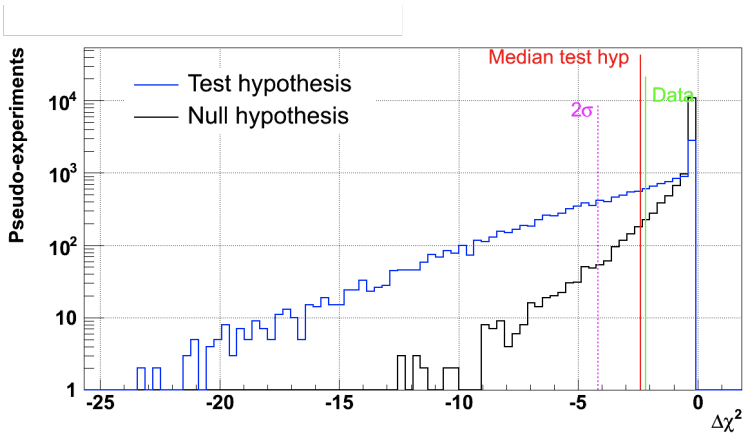


Figure 9.8: $\Delta\chi^2$ *distribution in 300,000 pseudo-data samples generated under the null (black) and test (blue) hypotheses. The median of test hypotheses,* $\Delta\chi^2_{data}$*, and* $P2\sigma$ *are also shown. See text for the significance of the symbols.*

Chapter 10

Conclusions

In this thesis we described a search for WZ and ZZ production in a final state with a high-P_T lepton, large missing transverse energy, and two jets with the full CDF dataset ($\int \mathcal{L} = 8.9$ fb^{-1}). As a consequence of the small signal and large background, a very considerable effort was necessary for getting just a hint of a WZ/ZZ component in the W+2 jets data. The effort included implementing a number of novel techniques for improving the modeling of the dataset and increasing the acceptance to our diboson signal.

In order to improve the modeling of the jet energy scale, we derived specific quark and gluon jet energy corrections to be applied to our simulations additionally to the standard CDF jet energy corrections. In the course of this process we found and remedied a number of flaws in the standard CDF technique to model the QCD multi-jets background in the high-P_T electron sample. The achieved improvements were able to clarify the anomaly in the di-jet mass spectrum previously observed by CDF around 145 GeV/c^2: agreement between data and standard model prediction was achieved.

In order to accept the largest amount of WZ/ZZ signal events, we exploited four different trigger paths (high-P_T electron and muon, forward electron, and missing transverse energy plus jets trigger paths). We also made use of neural network discriminants to tag jets coming from heavy quarks. By doing so, we were able to divide the W+2jets sample in a b-jets enriched sub-sample and the remaining sub-sample. Because of the different signal over background ratios and background compositions, higher performances are obtained by analyzing separately the two samples.

We measure the WZ/ZZ cross-section via a fit to the di-jet mass:

$$\sigma_{WZ/ZZ} = 4.7 \pm^{3.0}_{2.5} \text{ pb}$$

which is consistent the SM cross-section ($\sigma^{SM}_{WZ/ZZ} = 5.1 \pm 0.2$ pb), corresponding to a significance of about 1.5 σ. Such a result corresponds to a limit of $\sigma_{WZ/ZZ} < 12.2$ pb at 95% C.L. The significance of the result is about 1.4 σ.

The combination of this result with similar searches exploiting the orthogonal signatures of two charged leptons plus two jets ("$lljj$"), large missing transverse energy plus two jets ("$METjj$") will lead to a more significant measurement of the WZ/ZZ diboson cross section. If we assume combining the information by uncorrelated treatment of the systematic uncertainties and we consider the expected significances for SM diboson production of about 2.6 σ in the the $lljj$ sample ($\mathcal{L}dt = 8.9\ fb^{-1}$ [8]) and 1.4 σ in the $METjj$ sample ($\mathcal{L}dt = 5.2\ fb^{-1}$ [132]), CDF would be able to achieve a 3σ evidence for the WZ/ZZ signal. It is fair to expect that future work based on the whole Tevatron and LHC data will lead to the observation of this diboson process. At the same time, a much more detailed study of associated production of dibosons and jets will be performed, and our knowledge of an area of physics where discoveries might still be hiding will be greatly extended.

Appendices

Appendix A

Other b-jet identification taggers

Although the main focus in this chapter was set on the jet b-ness b-tagger, we describe briefly also some other b-taggers to illustrate the progress obtained by CDF in the tremendous efforts devoted to developing an optimal b-tagger. In this section we describe the four most used b-tagger algorithms in CDF and highlight their peculiarities.

A.0.1 Secondary Vertex Tagger ("SecVtx")

For historical reasons (top quark discovery [2]) and for the subsequent large use, SecVtx ([140]) is the most famous CDF b-tagger. It was an important component of all subsequent more efficient b-taggers, including b-ness. SecVtx aims to identify and locate secondary displaced vertices. In order to do that, displaced tracks are reconstructed with an accuracy such that they can be distinguished from tracks originating from the primary interaction. Tracks within a jet are selected as follows [1]:

1. $P_T > (P_T)_{min}$
2. $d_0 < .3\ cm$
3. $|z - z_{PV}| < 5\ cm$, PV being the primary vertex
4. $S_{d_0} \equiv d_0/\sigma_{d_0} > (S_{d_0})_{min}$, σ_{d_0} being the uncertainty on the d_0
5. Other cuts based on the track quality ([141]).

[1] Selection criteria has been optimized to enhance the purity of B-hadron tracks

Moreover track pairs whose invariant massed is within appropriate K_s or Λ-mass windows are removed.

Given the list of selected tracks within a jet, SecVtx uses a two pass approach to find vertices:

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

2. Pass 2: if a secondary vertex is not found at the end of the first pass, only two displaced tracks are required to reconstruct the vertex. Compared to Pass 1, these tracks have to pass tighter requirements: the first (second) track has to have $(P_T)_{min} = 1\ GeV/c$ $(= 1.5\ GeV/c)$. Both tracks are required to have $(S_{d_0})_{min} = 3$.

If the vertex is found in a jet, the two dimensional decay length L_{xy} is calculated as the projection in the transverse plane of the vector pointing from the primary to the secondary vertex (see Fig. A.1). Secondary vertices corresponding to the decay of heavy-flavor hadrons are expected to have large (positive) L_{xy}, while secondary vertices from mis-measured tracks are expected to be less displaced and to feature with the same probability positive and negative L_{xy}. Therefore SecVtx requires $L_{xy}/\sigma_{L_{xy}} > 3$, where $\sigma_{L_{xy}}$ is the estimated uncertainty on L_{xy}. Other cuts to reject secondary vertices due to interactions in the detector material are described in [141].

One of the downsides of this procedure is that C hadrons cannot be distinguished from B hadrons. Although C hadrons decay with a lower track multiplicity and in a shorter average time than b-hadrons [2], they are tagged with a relativity high efficiency as well. Therefore SecVtx is actually a "heavy-flavor tagger".

SecVtx allows for two main operation points: "tight", and "loose". These operation points are defined by changing the requirements on the displacement of the tracks and the secondary vertices, albeit keeping the same two-pass approach. The tight and loose operating points feature different b-tag efficiencies and mistag rates, as shown in Fig. A.5.

SecVtx b-tag efficiency was calibrated in a low-P_T inclusive electron data sample, which is enriched in semi-leptonic decays of bottom and charm hadrons. The mistag rate was calibrated by counting the number of negative tags (Fig. A.1) in inclusive jet trigger samples (jet E_T >20 GeV). Such a mistag rate is parametrized as a

[2] The D^0 proper lifetime is about 0.4 ps

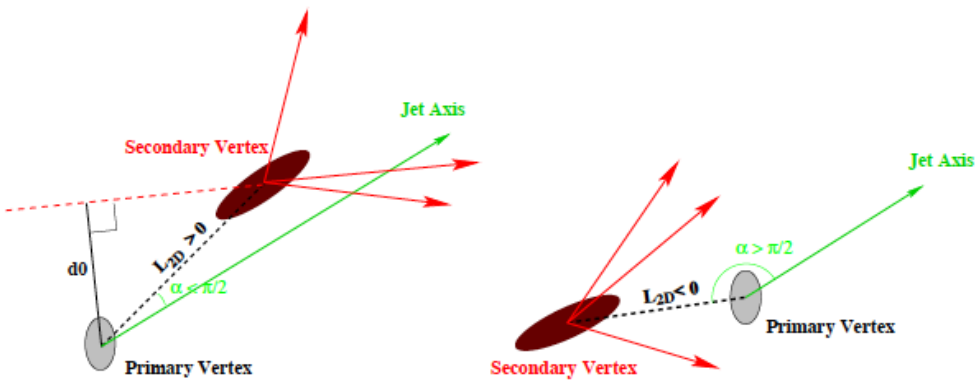


Figure A.1: *Schematic view of a positive (left) and a negative (right) SecVtx tag.*

function of jet E_T, ϕ, η, track multiplicity, and the scalar summed E_T of all jets in the event with $E_T > 10$ GeV and $|\eta| < 2.4$.

A.0.2 Soft Lepton Tagger ("SLT")

The soft lepton tagger ([142]) takes a different approach with respect to the other CDF b-taggers: SLT focuses on relatively low-P_T charged leptons in the jet cone. Those leptons originate from the semileptonic decay of the b-hadrons. Since the fraction of those decays is about 11% per lepton flavor, SLT cannot compete in performances with the other taggers. However, since SLT inputs are somehow independent of the other taggers (no displaced track info is used), SLT can still tag a jet when the other methods have failed. In practice, CDF uses only the soft muon tagger ("SLT$_\mu$"), as the purities of the electron and τ samples are very low.

SLT$_\mu$ is calibrated in Z^0 and J/ψ samples.

A.0.3 Jet Probability Tagger ("JetProb")

The JetProb tagger ([143]) does not look for secondary vertices, but it assigns to each track a probability of coming from the primary vertex based on its signed impact parameter significance. Combining the probabilities for the well-identified

tracks within a jet, it is possible to evaluate the probability of the jet to be fully composed of particles originated from the primary vertex. The distribution of this probability is flat for jets originated by light quarks, and is peaked at small values for b and c jets (see Fig. A.2). Although this algorithm produces a continuous output, in practice three operation point are often chosen (JetProb <0.5%, 1%, 5%) to b-tag a jet.

Calibrations of JetProb response were performed with the same procedure as SecVtx (Sec. A.0.1) for both b-tag efficiency and mistag rate.

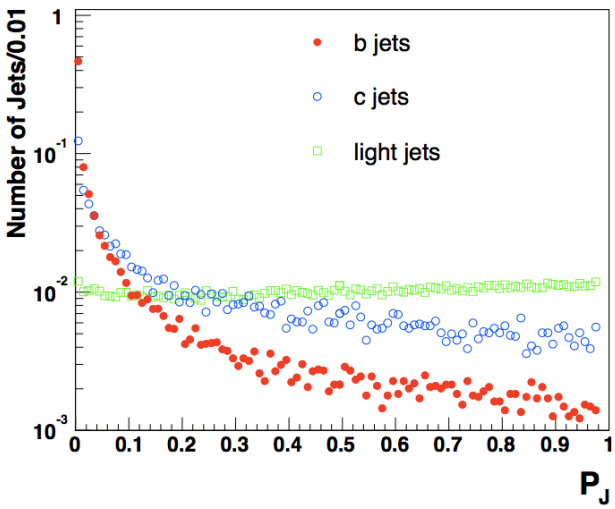


Figure A.2: *Distributions of the probability (JetProb) of jets to be composed of particles originated from the primary vertex. Distributions for jets matched to b (full circles), c (empty circles), and light (empty squares) quarks in MonteCarlo events are displayed.*

A.0.4 "RomaNN" b-Tagger

The RomaNN tagger [3] is a multi-variate tagger which employs neural network architectures. The RomaNN tagger is unique in its emphasis on finding multiple displaced vertices, as it may happen in the case of multiple decays, such as $B \to D \to K$. This feature distinguishes the RomaNN tagger from jet b-ness and from SecVtx, which search for no more than one displaced vertex per jet. RomaNN tagger uses the four subsequent neural networks (NNs) described below (Fig. A.3):

[3]The RomaNN tagger takes the name after "Universitá La Sapienza di Roma", where such a tagger was developed ([144]).

- the first NN is used to distinguish vertices due to heavy flavored (b or c) hadron decays from false vertices, or vertices from other hadron decays;
- the second NN aims to identify (and then recover) unvertexed tracks which come from a heavy flavor hadron decay;
- the third NN which takes as inputs the first two NNs along with other inputs, such as the output from other taggers (SecVtx, JetProb, SLT_μ), aims at distinguishing b/c/light-flavor jets from each another (Fig. A.3)[4];
- the output of these three flavor-separating NNs are then used to train a final NN which provides a three flavor (b/c/light flavor) discriminant output.

The RomaNN tagger has been used at CDF in light Higgs boson searches [127] not only for the superior performances with respect to SecVtx (Fig. A.5) but also because it allows for operation points ("ultra-loose") yielding greater efficiency, which is vital in the context of rare signals.

However, the downside of the RomaNN tagger is that it does not guarantee to fit for a displaced vertex or to always have sufficient information to reliably tag a jet ($\sim$ 21% of the MC b-jets).

The calibration of the RomaNN tagger b-tag efficiency was performed in the same way as SecVtx (Sec. A.0.1). The calibration of the mistag rate was performed with a slightly more complicated procedure, since RomaNN cannot rely on the concept of negative tags. See [145] for further details.

A.0.5 Higgs Optimized b-Jet Identification Tagger (HOBIT)

HOBIT is a multivariate b-jet identification algorithm, which employs feed-forward neural networks (multilayer perceptron) [146]. The strategy used in developing HOBIT, chronologically the last CDF b-tagger, is to build upon the strength of the previous CDF b-taggers and construct a new tagger that is optimized specifically for finding light Higgs boson decays ($H \to b\bar{b}$) [5]. The training was performed by exploiting the samples used by light Higgs analyses at CDF: b jets in Pythia ([93]) 120 GeV/c^2 Higgs boson Monte Carlo (MC) and light jets from W+jets MC generated by Alpgen ([95]) and showered by Pythia. Jets were required to have an $E_T > 15$ GeV and $|\eta| < 2$.

[4] by using a 3 flavor NN jets containing two heavy quarks ($b\bar{b}$, $c\bar{c}$) are not considered.

[5] Also thanks to HOBIT CDF in combination with DØ achieved a sensitivity at the 3σ-level ([3]).

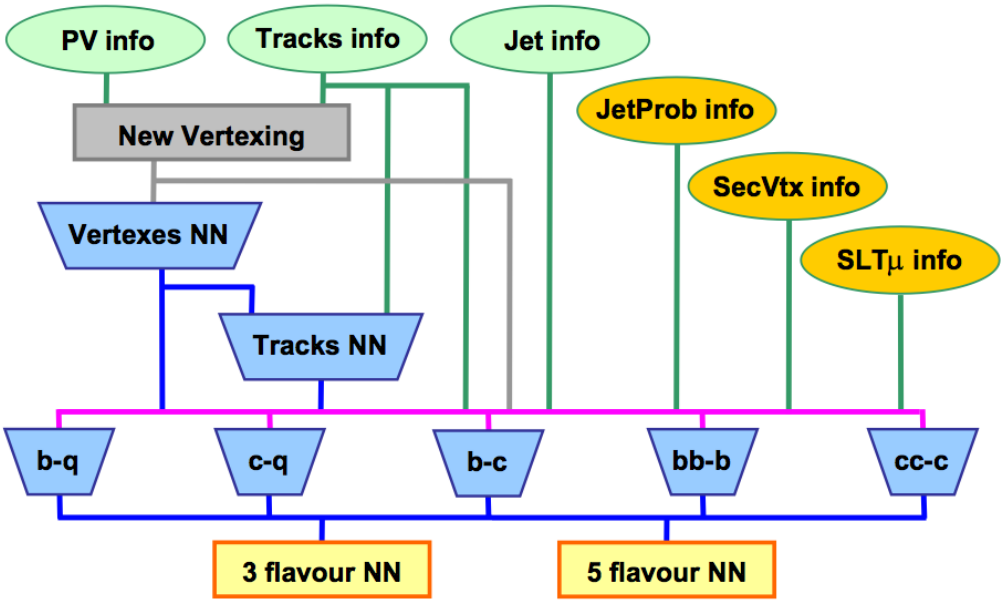


Figure A.3: *Flow chart of the RomaNN b-tagger. See text for further explanations.*

HOBIT employs twenty-five inputs: fourteen of these inputs are also inputs to the RomaNN (Sec. A.0.4) tagger, further ten inputs to HOBIT are the ten highest track-by-track b-ness (Sec. 5.2) in the jet cone. When less than ten tracks are found within a jet, the value of the remaining track-by-track NN inputs are set to 1 as this is the light-jet-like value of the NN output. The number of tracks which pass the track-by-track NN selection criteria is found to have additional discriminating power and is also used as an input to HOBIT. As for the jet b-ness tagger kinematic biasing of HOBIT is prevented by weighting the light jet training sample to have the same E_T distribution as the b-jet training sample. In Sec. A.0.4 we mentioned one potential weakness of the RomaNN tagger, its inability to produce a usable output when there is insufficient input information for reconstructing a secondary vertex ($\sim$21% of the total b-jets). In HOBIT, thanks to the jet b-ness inputs, we were able to reduce those cases to 3%.

HOBIT output is shown in Fig. A.4. In Fig. A.5 HOBIT performances are compared to the other taggers. For a given purity level the improvement in the absolute efficiency is about 10% with respect to the jet b-ness and RomaNN taggers, and about 15% with respect to SecVtx.

HOBIT response for b-tag efficiency and mistag rate was calibrated by exploiting the low-P_T inclusive electron data sample (Sec. A.0.1) and the $t\bar{t}$ sample. In the latter sample, a simultaneous measurement of the b-tag efficiency and $t\bar{t}$ cross-

section is performed. When computing the mistag rate, the SecVtx procedure was slightly changed, since HOBIT cannot rely on the concept of negative tags.

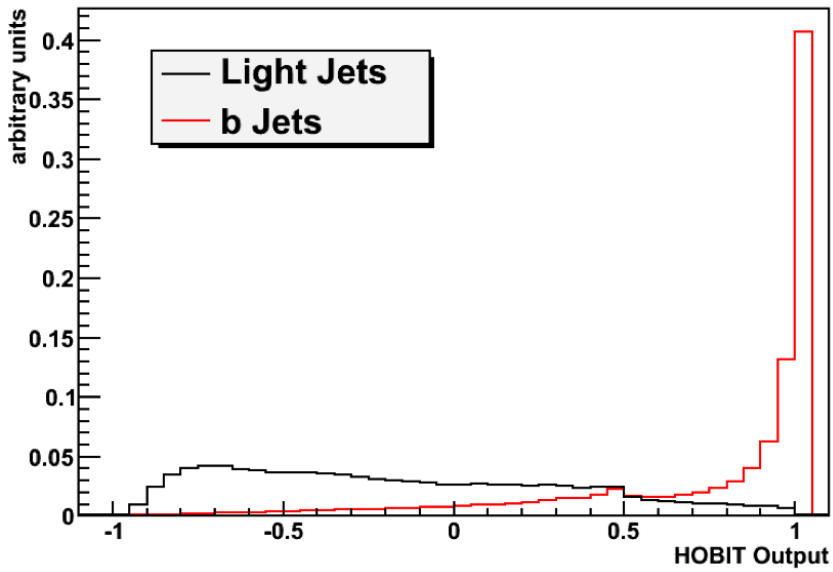


Figure A.4: *HOBIT outputs. The output is trained so that 1 is b jet-like and -1 is targeted to be light jet-like. The black histogram is for light quark jets and the colored histogram is for b jets. Taken from MC, the distributions are normalized to one another.*

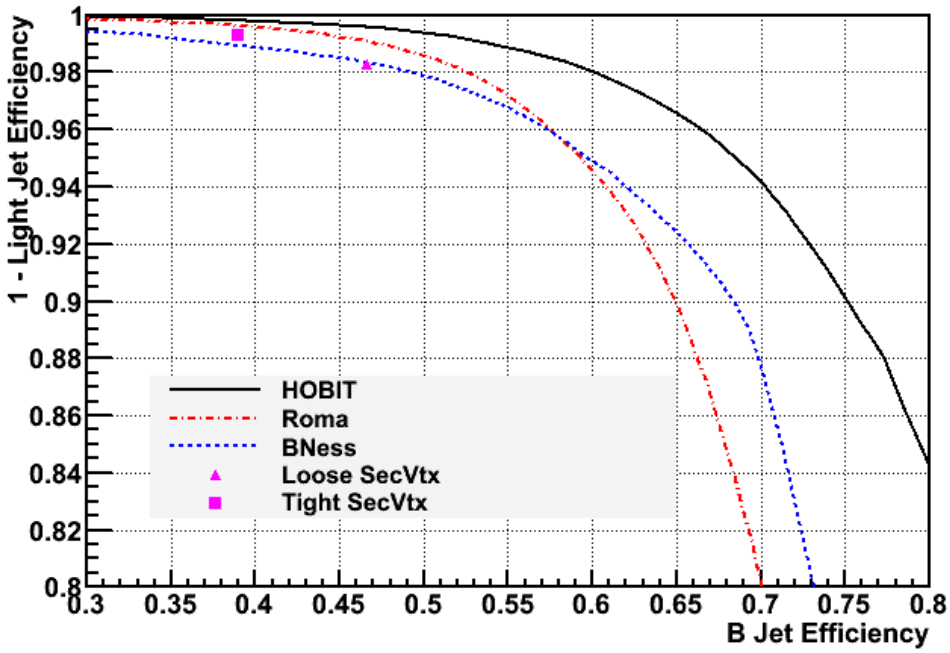


Figure A.5: *A comparison of the purity-efficiency tradeoffs for HOBIT versus RomaNN (Sec. A.0.4), Bness (Sec. 5.2), and SecVtx (Sec. A.0.1) loose and tight operation points. A significant improvement over prior multivariate taggers is seen.*

Appendix B

More Details on the MC Event Simulation

B.1 Parton Shower Method

In Chap. 7 we introduced the parton shower method to model the initial and final state radiations in HEP events. In this section we give more technical details.

The formulation of parton branching in term of Sudakov form factor ([7]) is well suited to an implementation in MC programs. Briefly the MC generation of the initial state radiation shower works in the following way (see Fig. B.1):

- The probability of evolving from t_1 to t_2 without resolvable branching is $\Delta(t_1)/\Delta(t_2)$. Given t_1, t_2 can be generated with the correct probability distribution by solving the equation

$$\Delta(t_2)/\Delta(t_1) = R, \tag{B.1}$$

  where R is a uniform number generated between $[0, 1]$

- if $t2 \geq Q$ no further branching occurs. Q defines the hard interaction scale [1],

- otherwise, the momentum fraction $z = x_2/x_1$ released by radiation is found

[1] When calculated to all orders in perturbative QCD, the hadronic cross sections are independent of the Q scale. However, at finite order, the hadronic cross sections depend on the Q scale. Therefore, the choice of the scale may strongly affect the kinematics itself. As it will be described in the Chap. 8, we will assess an appropriate systematic uncertainty to cover for all kinematic differences due to a different choice of the Q scale.

by solving:

$$\int_{\epsilon}^{x_2/x_1} dz \frac{\alpha_s}{2\pi} P(z) = R' \int_{\epsilon}^{1-\epsilon} dz \frac{\alpha_s}{2\pi} P(z) \tag{B.2}$$

where R' is a uniform number generated between $[0, 1]$, and ϵ is the infra-red cut-off for resolvable branching.

For the final state radiation the idea is the same, except that the evolution is "downwards" (decreasing t) rather than "upward" and stops when the cut-off scale t_0 (typically ~ 1 GeV) is reached. All this mechanism creates a parton cascade.

It should be noticed that, although the initial state radiation evolves upward, PYTHIA handles it within the backward evolution scheme. The hard scattering makes use of PDF's with initial state radiation already included (Sec. 7.1.1). Starting from the partons in the PDF's, one has to find the initial partons before any branching occurred, thus reconstructing the shower backwards.

In the above description we did not distinguish between gluon and photon emissions. Some obvious differences do exist when simulating the two processes. For this and other details we refer to [93].

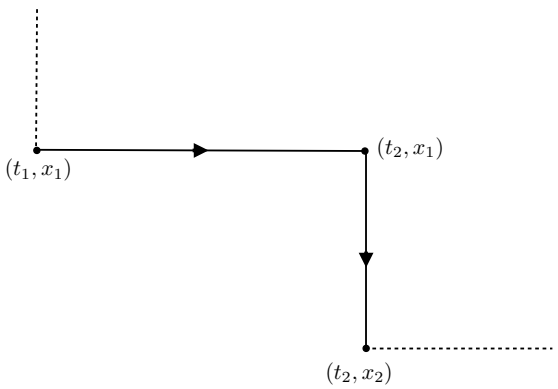


Figure B.1: *Fundamental step in (t, x)-space in DGLAP evolution.*

B.2 String fragmentation

The string fragmentation model (SF) can be better understood with an example: $e^+e^- \to q\bar{q}$. Here, lattice QCD studies support a linear confinement picture, i.e. the energy in the color dipole field between charge and anticharge increases linearly with the distances of the charges. The Coulomb term is neglected as well as the

possibility of gluon soft radiation from one of the quarks. The SF model is developed under the linear confinement assumption. As the q and $\overline{q}$ move apart from their common origin, a color flux tube or string (See. Fig. B.2) is stretched between them. The typical transverse dimension of the tube is of the order of the hadron size (~ 1 fm). The string has 1 GeV/fm energy density. Such a value was extrapolated from quarkonium spectroscopy. As the two charges keep moving apart, the potential stored in the string increases, and the string may break by producing a new $q'\overline{q}'$ pair. The system is thus split into two color singlets: $q\overline{q}'$ and $q'\overline{q}$. If the invariant mass of either of these strings is large enough further break-ups may occur. The process continues iteratively until only on-mass-shell hadrons remain.

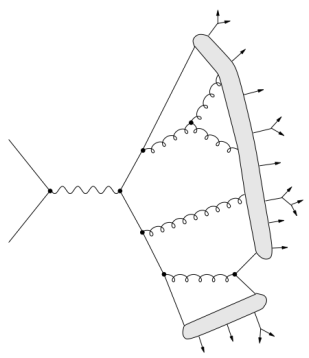


Figure B.2: *Parton shower with string hadronization model for $e^+e^- \to q\overline{q}$. Grey regions connecting different partons are a pictorial representation of the strings.*

The generation of $q\overline{q}$ pairs in the color field of the string happens via the quantum mechanical tunneling. In such a picture the production of heavy-flavor quarks is suppressed ($u : d : s : c \sim 1 : 1 : 0.3 : 10^{-11}$). Therefore, very little charm and even less heavier quarks are expected to be produced at the level of the hard interaction or parton shower.

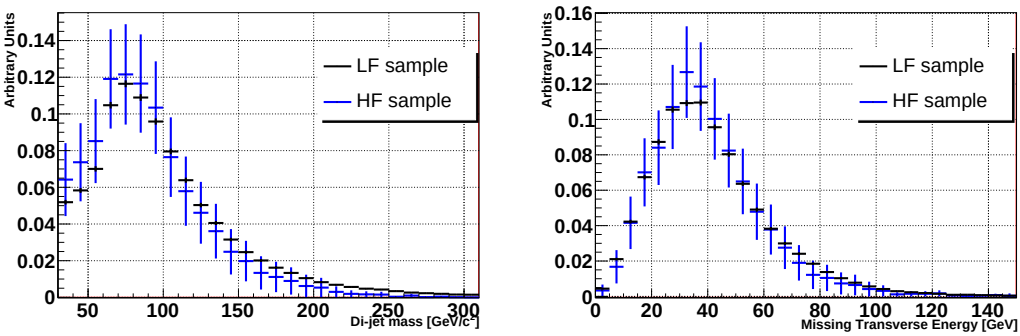
When a gluon is radiated off of a quark the picture becomes more complicated. The gluon is a kink on the string which connects the q and the $\overline{q}$. The $q\overline{q}g$ will fragment along its length. The modeling of such a fragmentation is crucial to predict for example the angular distribution of the jets in $e^+e^- \to 3jets$. The Lund model fits the data better than other fragmentation models in this final state [147].

Further details about fragmentation are available in [7], [93].

Appendix C

DifferencesLFandHFShapes

Shape differences between the di-jet mass and the missing transverse energy distributions in W+light-flavor jets (LF) and W+heavy-flavor jets (HF) are shown in Fig C.1. LF and HF samples were defined in Sec. 8.3.2. We see no difference in the LF and HF shapes, validating our assumption that the overlap of b and c quarks in LF and HF samples does not bias our results



(a) (b)

Figure C.1: *Di-jet mass (a) and missing transverse energy (b) distributions in simulated W+light-flavor jets and W+heavy-flavor jets simulated samples.*

Appendix D

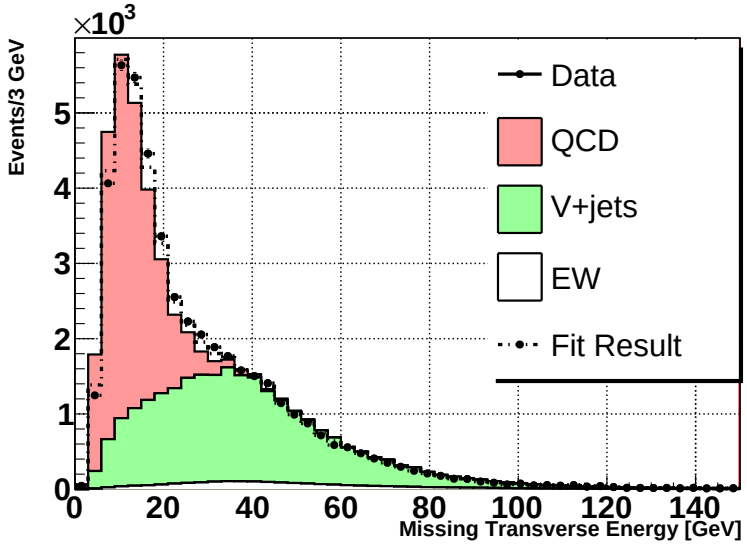
QCD Rate Uncertainty

The QCD rate systematic uncertainty for the TCE electron sample is derived in the sample selected as described in Table 9.4. The uncertainty is the sum in quadrature of different sources, which are summarized below:

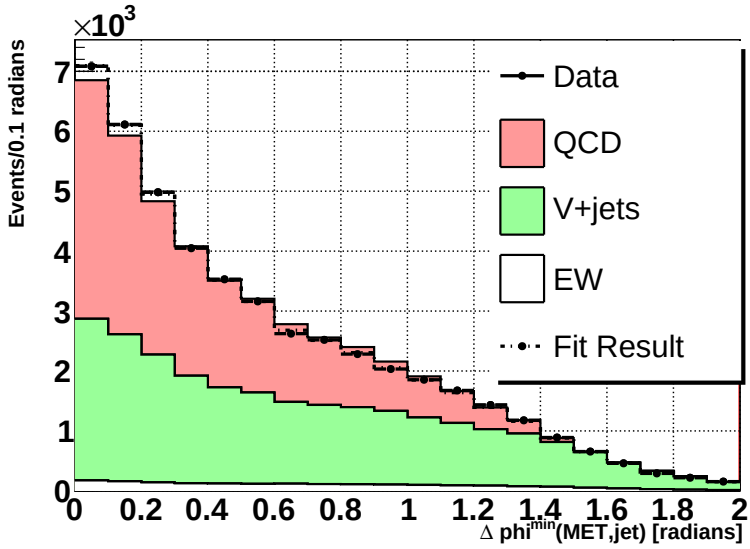
- Statistical: 1.4%. These uncertainties are the ones returned by the fit.
- Systematics:
  1. MC templates: .9%. We vary the jet energy scale by $\pm 1\sigma$ (Sec. 6.2.4) when building the MC-based templates (V+jets, diboson, top) and repeat the $\not\!\!E_T$ fit. The newly obtained rate compared to the nominal rates.
  2. QCD templates: 4.4%. We compare the nominal rates with the number of data events after subtracting the MC-based predictions. These predictions are normalized to the fit outcomes.
  3. Fit variable choice: 13.1%. A number of different variables rather than $\not\!\!E_T$ are used for the fit. Compared to the nominal value, the largest discrepancy in QCD fraction is obtained when fitting in $d\phi_{min}(\not\!\!E_T\ , jet)$, which is the difference in azimuthal angle between the missing transverse energy and the closest jet with $E_T > 5\ GeV$. The result of such a fit is shown in Fig. D.1.

The overall systematic uncertainty for the QCD rate in the TCE electron sample is 14.1% [1].

[1] We assume that the systematic uncertainty on the QCD rate does not depend on the selection criteria. Therefore we have the same uncertainty on the sample selected as described in Table 8.1. We further assume that the systematic uncertainties on the QCD fractions in the CMUP, CMX,



(a)



(b)

Figure D.1: $\not\!\!E_T$ (a) and $d\phi_{min}(\not\!\!E_T\ , jet)$ (b) distributions in the electron (left) and muon (right) samples. EW (white), V+jets (green), and QCD multi-jets (pink) templates are normalized according to the fit results, stacked together and superimposed to the data (dots). While fitting, each bin of the stacked histogram is varied within its error. This is shown in the distribution named "Fit Result". The details of the fit are described in Sec 8.3.3.

EMC samples are the same as the one in the TCE sample. The uncertainty in the PHX sample is recalculated by using the same method shown in this Section: the QCD rate uncertainty is 18.0%.

Appendix E

Modeling

E.1 Validation of the Model - Additional distributions

Additional distributions in the electron and muon samples described in Chap. 8 are shown in Fig. E.1-E.3.

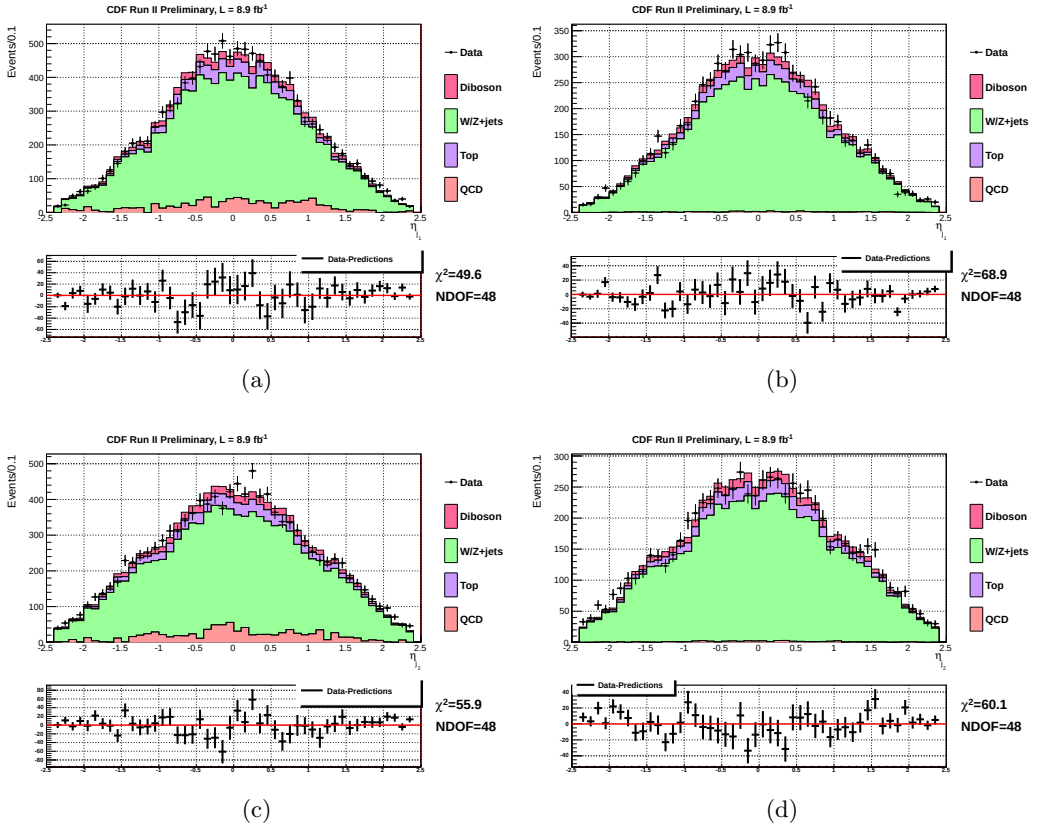


Figure E.1: *η distributions of the first leading (a-b) and second leading (c-d) jets in the electron (a-c) and muon (b-d) samples. In the upper pads data (dots) is superimposed to the predictions, which have been properly normalized. In the lower pads the data minus the predictions are plotted: the red horizontal line at 0 represents the ideal agreement between the two. The χ^2 and the number of non-zero bins (NDOF) are also reported to quantify the agreement. See Sec 8.4 for further details.*

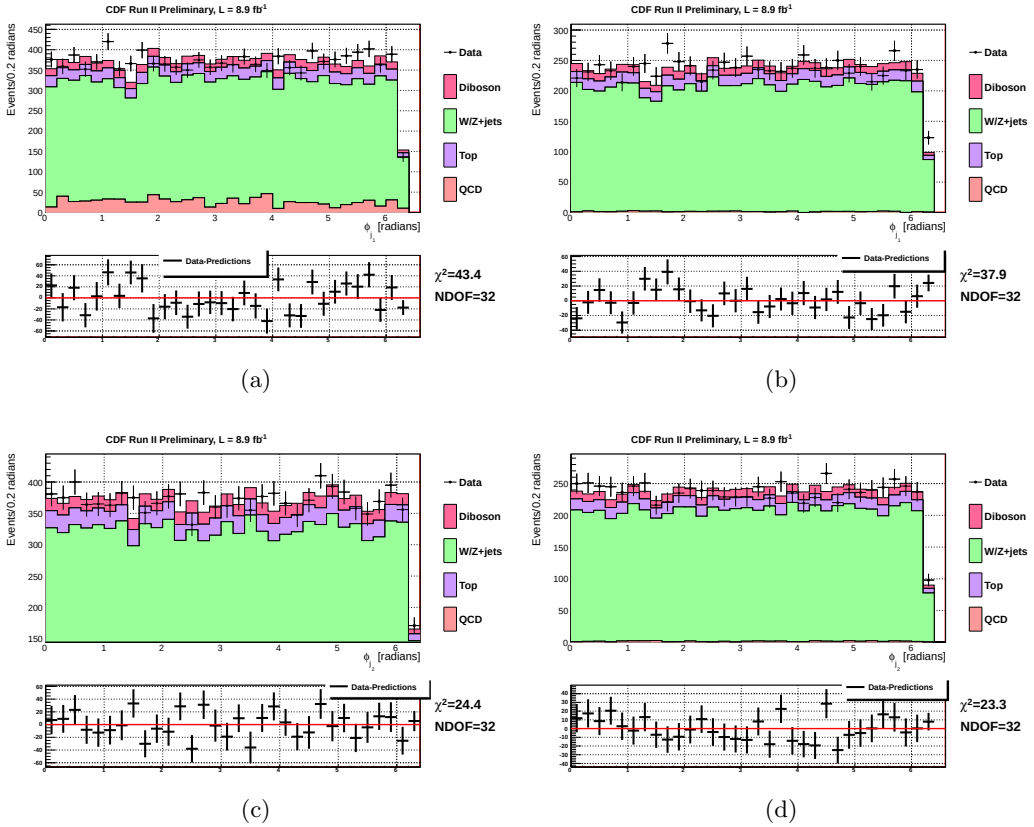


Figure E.2: *ϕ distributions of the first leading (a-b) and second leading (c-d) jets in the electron (a-c) and muon (b-d) samples. In the upper pads data (dots) is superimposed to the predictions, which have been properly normalized. In the lower pads the data minus the predictions are plotted: the red horizontal line at 0 represents the ideal agreement between the two. The χ^2 and the number of non-zero bins (NDOF) are also reported to quantify the agreement. See Sec 8.4 for further details.*

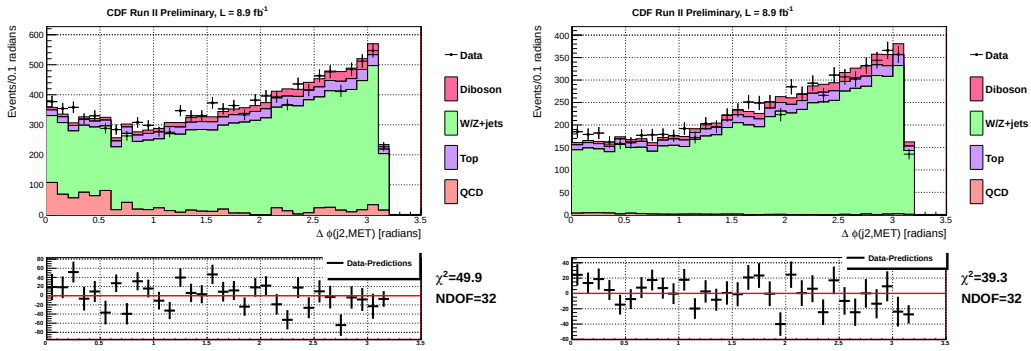


Figure E.3: $\Delta\phi(j_2,\not\!\!E_T)$ *distributions in the electron (a) and muon (b) samples. In the upper pads data (dots) is superimposed to the predictions, which have been properly normalized (see text). In the lower pads the data minus the predictions are plotted: the red horizontal line at 0 represents the ideal agreement between the two. The* χ^2 *and the number of non-zero bins (NDOF) are also reported to quantify the agreement. See Sec 8.4 for further details.*

E.2 Additional Checks on Di-jet Mass Spectrum in the W+2jets Sample

In this section we present studies on the di-jet mass spectrum in the W+2jets sample in addition to what described in Chap. 8.

E.2.1 Update of the Previous Result

The analysis reporting an excess of events in the di-jet mass spectrum in the W+2jets sample was published with about half of the available CDF dataset ([119], $\int \mathcal{L}dt = 4.3$ fb^{-1}). In Fig. E.4 we show the di-jet mass spectrum with the whole CDF dataset ($\int \mathcal{L}dt = 8.9$ fb^{-1}): large discrepancies between data and the SM predictions throughout the mass spectrum can be seen.

Although the excess in the data above the diboson peak does not seem consistent with a Gaussian distribution, we repeat the fit by including the template for the narrow resonance X at 145 GeV/c^2. Results are shown in Fig. E.6a. The cross-section for the resonance X that we measure is $\sigma_{WX} = (2.4 \pm 0.6)$ pb. Such result is lower than the previous result, but still compatible with it ($\sigma_{WX} = (3.1 \pm 0.8)$ pb).

As additional check, we fit separately in the electron (Fig. E.7a) and muon (E.7b) channels: the same sort of discrepancies are seen.

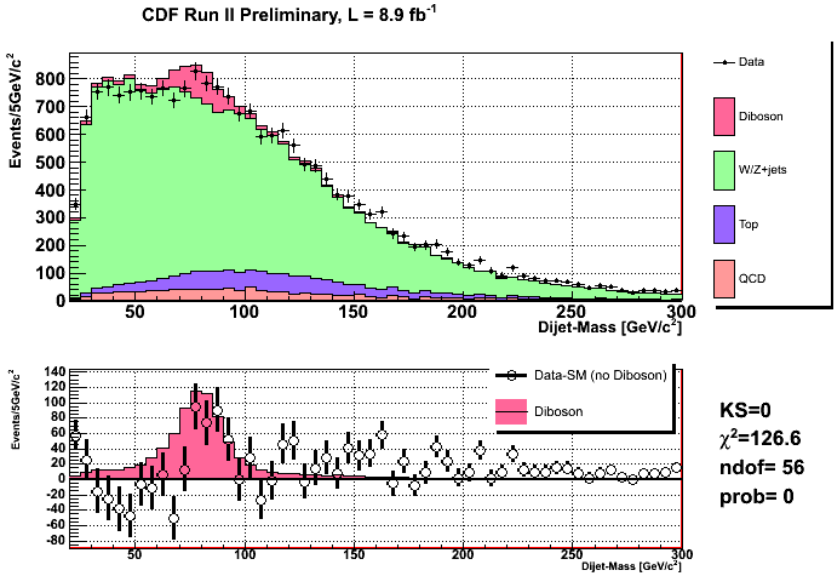


Figure E.4: Using the full CDF data set, the di-jet mass distributions in the electron and muon samples are stacked together and superimposed to the model. Compared to [119] no changes are applied. Details of the fit are explained in Sec. 8.5. In the lower pads of each plot the differences between data and the models (no diboson) are shown.

E.2.2 Effect of the Addition in the Selection Criteria: $\Delta R(j_1, j_2) > 0.7$

Fig. E.5 shows that the SM predictions, which were described in Chap. 8, do not model properly the low $\Delta R(j_1, j_2)$ region. Although no quantitative studies were performed on this matter, a cause of such a discrepancy could be the clustering efficiency, whose dependency on the relative jet angle is not well modeled by the simulations (e.g: wider jets in MC than data).

However, a jet pair from heavy particles is expected to be produced at large angle. Therefore, implementing a $\Delta R > 0.7$ is not expected to bias a search for heavy particles. Nonetheless, we investigate the effect of this cut on the final result. Fig E.6b shows that, although the region at low masses is now better modeled, the same discrepancies as in Fig. E.6a around and above the diboson peak are present. We extract a $\sigma_{WX} = (2.3 \pm 0.5)$ pb, which is compatible with the one extracted with no $\Delta R(j_1, j_2)$. Therefore, we confirm our initial guess that such a cut would not influence the result.

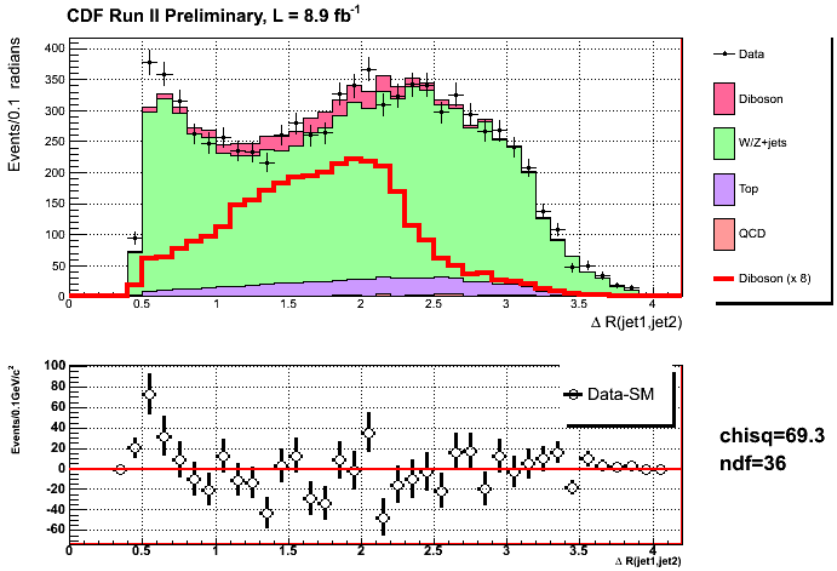
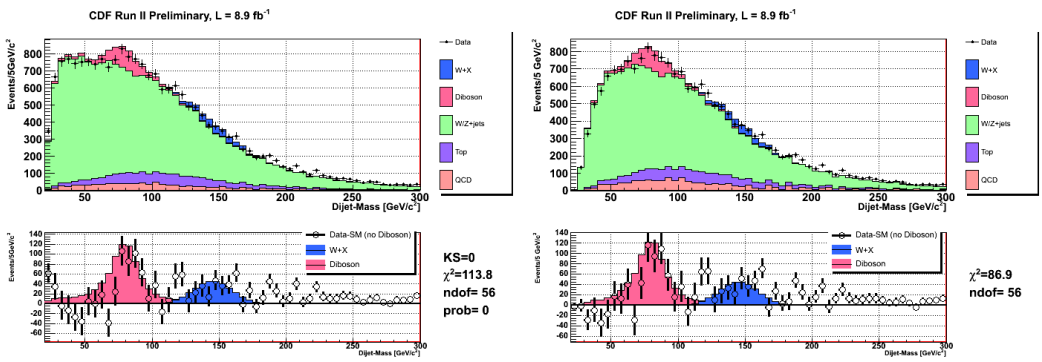


Figure E.5: $\Delta R(j_1, j_2)$ distributions in the muon samples, selected by using all the cuts described in Sec. 8.2, except for the $\Delta R(j_1, j_2)$ cut. In the upper pads data (dots) is superimposed to the predictions, which have been properly normalized. The diboson (red line) is magnified by a factor of 8. In the lower pads the data minus the predictions are plotted: the red horizontal line at 0 represents the ideal agreement between the two.



(a) (b)

Figure E.6: Using the full CDF data set, the di-jet mass distributions in the electron and muon samples are stacked together and superimposed to the model. Compared to [119], no changes (a), $\Delta R(j_1, j_2) > 0.7$ cut (b) are applied. Details of the fit are explained in Sec. 8.5. In the lower pads of each plot the differences between data and the models (no diboson) are shown.

E.2.3 Effect of the Quark/Gluon Jet Energy Calibrations

In the 4.3 fb^{-1} result the quark and gluon jets energy scales are corrected in the same way in the data as well as in the MonteCarlo simulations. In this thesis we introduced separate correction for MC gluon and quark jets (Sec. 6.2.4). The effect on the di-jet mass spectrum after applying those MC corrections in the electron and muon samples is shown respectively in Fig. E.7e and E.7f. With respect to the case where no MC corrections were applied (Fig. E.7c and E.7c) the predictions model the data better. However, while the SM predictions properly describe the muon data, the electron sample still show some discrepancies.

E.2.4 Effects on the Electron Data of the Corrections on the QCD Multi-jets Background Model

The di-jet mass spectrum in the electron sample, after the QCD multi-jet background model has been corrected (Sec. 8.3.3), is shown in Fig. E.7g: the predictions now agree with the data.

E.3 Di-jet mass fit templates

The templates used in the di-jet mass fit, which were described in Sec. 8.5, are shown in Fig. E.8, E.9.

E.4 Jet Energy Scale Systematic Uncertainties on the Template Shapes

In this section we discuss the dependence of the template shapes on the JES value. The JES and its uncertainty was discussed in Chap. 6. As mentioned in Sec. 8.5.2, the effect of the JES variation on the diboson templates and the other MC templates is handled differently in the fit.

E.4.1 Diboson

Diboson templates are parameterized with a Gaussian function on top of a 4th-degree polynomial (Fig E.11). The obtained parameterization as well as the dependence of the Gaussian parameters on the JES are used in the final fit. Further details were given in Sec. 8.5.2.

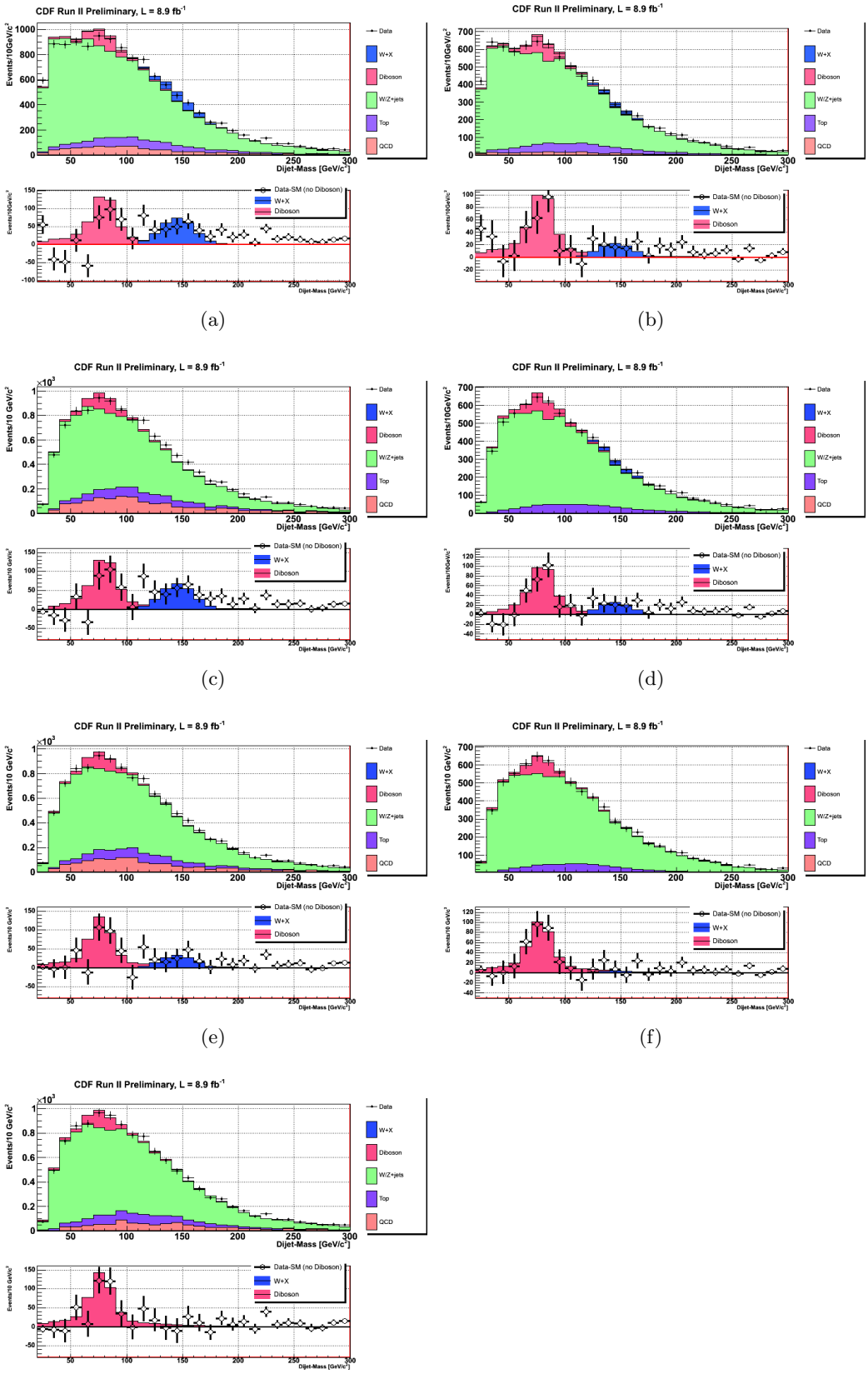
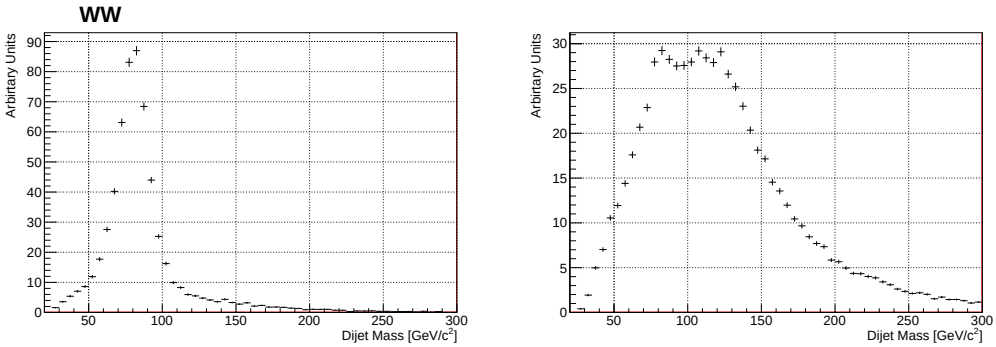
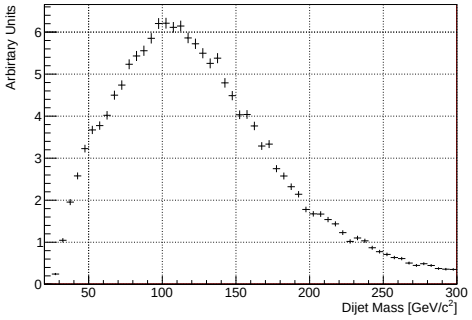


Figure E.7: Using the full CDF data set, di-jet mass distributions in the electron (left), and muon (right) samples are superimposed to the model. Compared to [119], no changes (a-b), $\Delta R(j_1, j_2) > 0.7$ cut (c-d), the energy corrections to MC gluon and quark jets (e-f, Sec. 6.2.4), and the corrections to the QCD multi-jet background model (g, Sec. 8.3.3) are subsequentially applied. In the lower pads of each plot the differences between data and the models (no diboson) are shown.

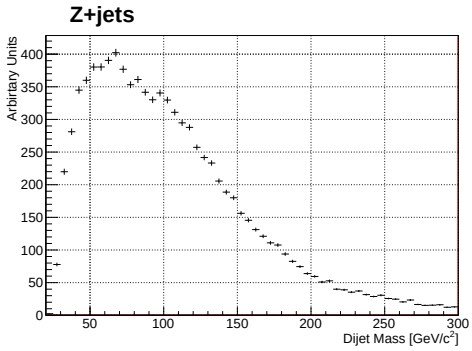


(a)

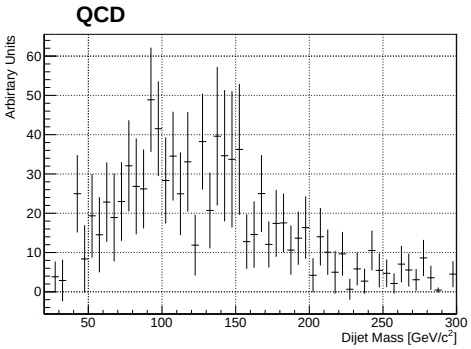
(b)



(c)

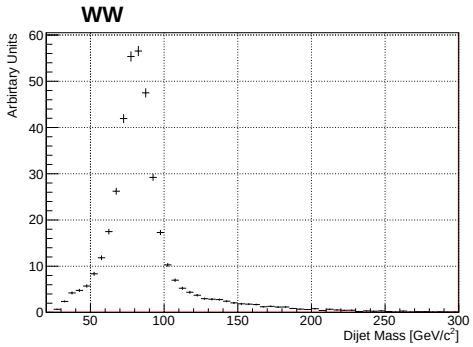


(d)

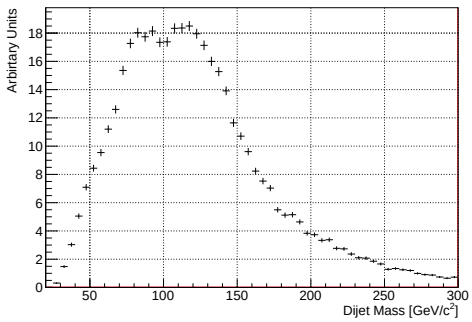


(e)

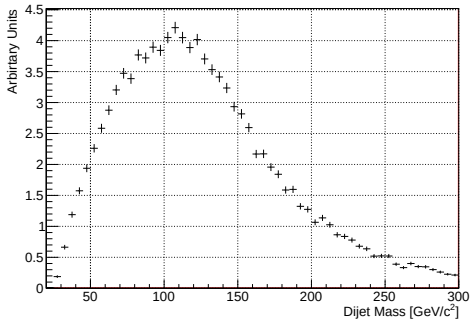
Figure E.8: Diboson (a), $t\bar{t}$ (b), single-top (c), V+jets (d), and QCD (e) templates used in the di-jet mass fit in the electron sample (Sec. 8.5). Templates are initially normalized to the rates provided in Table 8.3. Statistical errors are also shown.



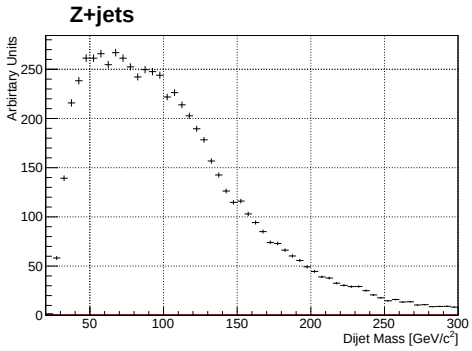
(a)



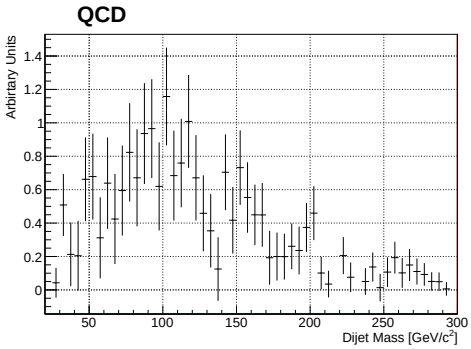
(b)



(c)



(d)



(e)

Figure E.9: Diboson (a), $t\bar{t}$ (b), single-top (c), V+jets (d), and QCD (e) templates used in the di-jet mass fit in the muon sample (Sec. 8.5). Templates are initially normalized to the rates provided in Table 8.3. Statistical errors are also shown.

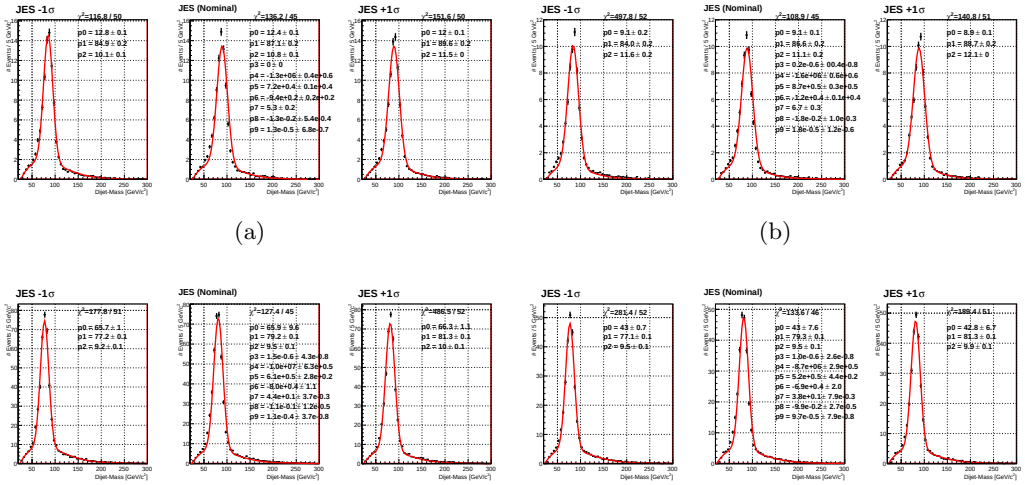


Figure E.10: WZ/ZZ (a-b) and WW (c-d) templates, obtained with the nominal and the $\pm 1\sigma$ varied JES, in the electron (a-c) and muon (b-d) samples. The parameterizations for the templates are also shown. See text for further details.

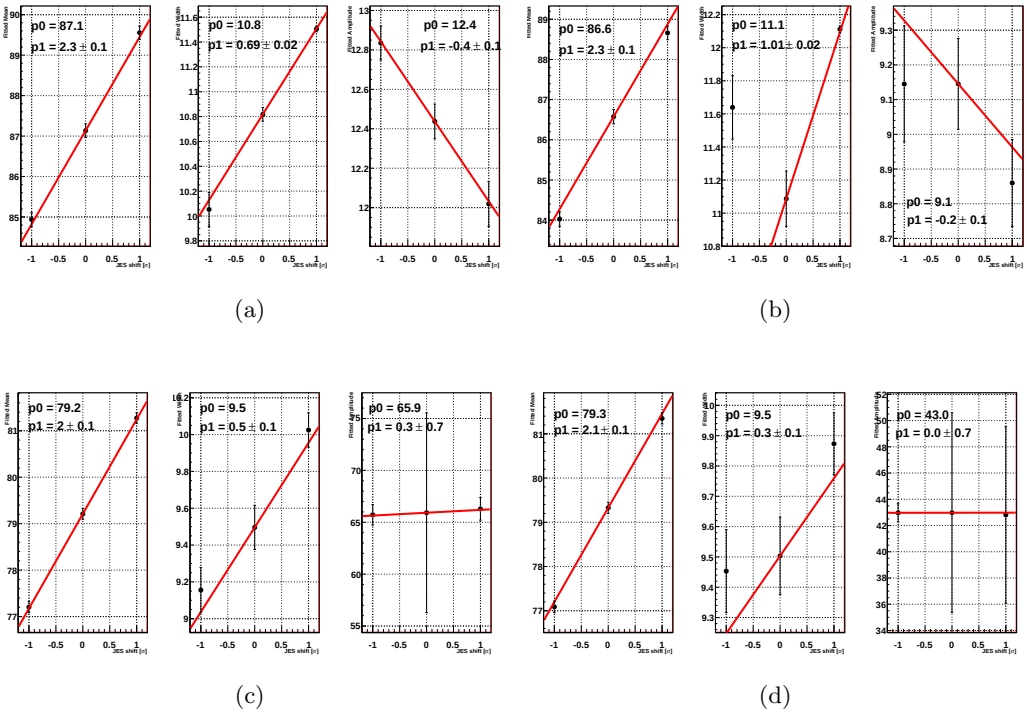


Figure E.11: Linear parameterizations of the dependence of the mean, width and amplitude of the Gaussian fitted function to the WZ/ZZ (a-b) and WW (c-d) templates in the electron (a-c) and muon (b-d) samples. See text for further details.

E.4.2 Other MC processes

In Fig. E.12 we show the effect of the JES variation on the W/Z+jets, the $t\bar{t}$, and single-top template shapes. All the shape variations have been derived in a MC samples with slightly looser selection cuts [1]. The W+jets shape variation is parameterized with a 5th-degree polynomial. The chosen functional form was tuned to the di-parton distribution in W+jets ALPGEN+PYTHIA events.

E.5 Q2 Systematic Uncertainty

In Fig. E.13 we show the effect of Q^2 scale systematic (Sec. 8.5.2) on the W+jets shape. Such a variation was parameterized with a 5th-degree polynomial/exponential function for di-jet mass less/more than 150 GeV/c^2. The chosen functional form was tuned to the di-parton distribution in W+jets ALPGEN+PYTHIA events with no detector simulation.

E.6 Fit Validation

In this section we validate the fit described in Sec. 8.5. Pull distributions (Fig. E.13) for all the nuisance parameters in the fit are built with 20,000 pseudo-experiments. Pulls are defined as the differences between the fitted and the generated values in unities of fitted uncertainties.

All pulls are compatible with Gaussian distributions centered at 0, thus proving that the fitted values in the data (Fig. 8.16) are unbiassed. The standard deviations of the JES, Q^2, QCD Rate - TCE, Diboson Cross Section, V+jets Rate, and Overall Acceptance parameters are lower than 1, indicating that the fit has greater constraining power than the input systematic uncertainties.

The signal sensitivity (Sec. 9.6) is calculated with the input systematic uncertainties for all parameters other than the V+jets Rate parameter to be conservative. For the V+jets Rate parameter we use the fitted uncertainty (Fig. 8.16).

[1] No requirements on the leptonic W have been made to increase the statistical power of the sample. We do not expect that the shape variations in electro-weak MC samples due to the JES are sensitive to the selection requirements on the leptonic leg of the event.

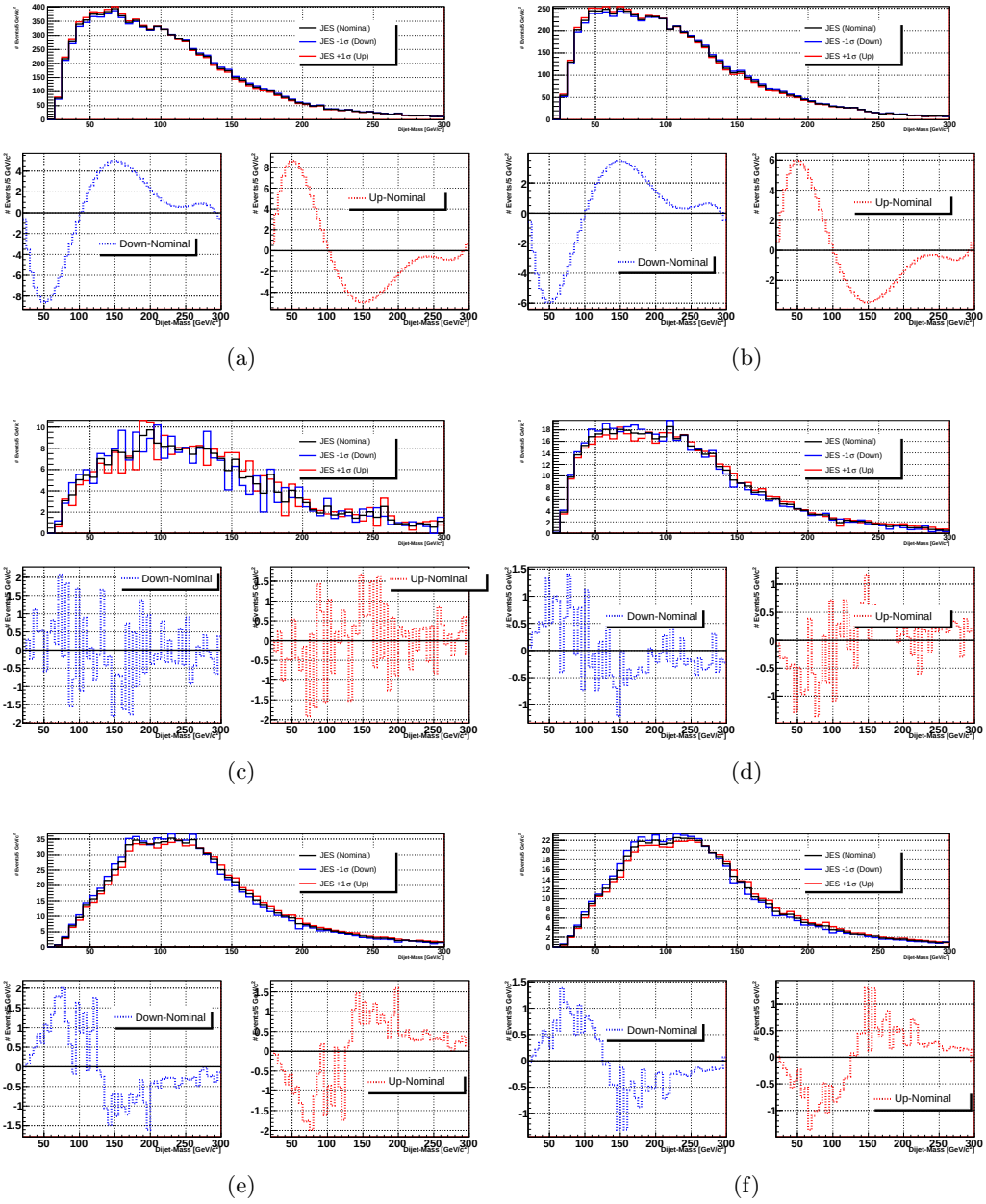
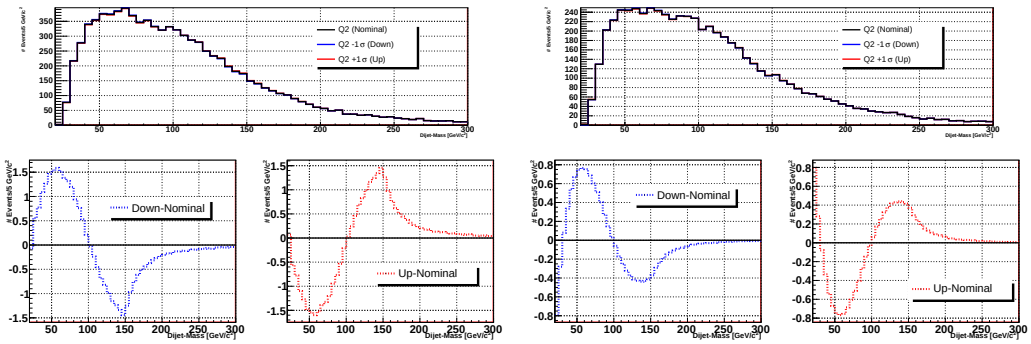
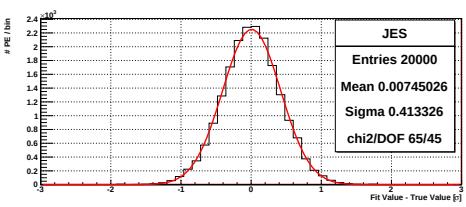


Figure E.12: Di-jet mass distributions in the W+jets (a-b), Z+jets (c-d), and top events (e-f) in the electron (left) and muon (right) samples. Those distributions are obtained with the nominal JES (black), $+1\sigma$ (red), -1σ (blue). In the lower pads we show the differences between the varied and the nominal shapes. The two shape variations have been symmetrized. Relative to the nominal values the shape variations are small.

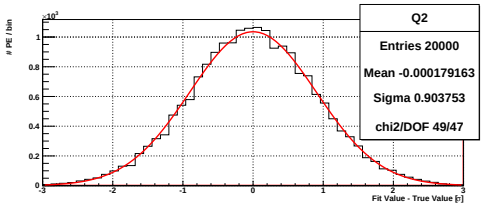


(a) (b)

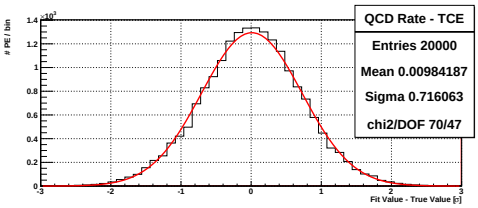
Figure E.13: *Di-jet mass distributions in the W+jets in the electron (left) and muon (right) samples. Those distributions are obtained with the nominal Q2 (black), $+1\sigma$ (red), -1σ (blue). In the lower pads we show the differences between the varied and the nominal shapes. The two shape variations have been symmetrized.*



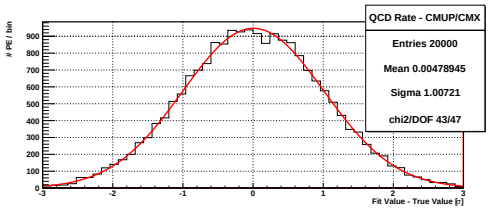
(a)



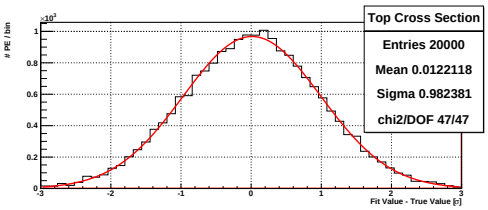
(b)



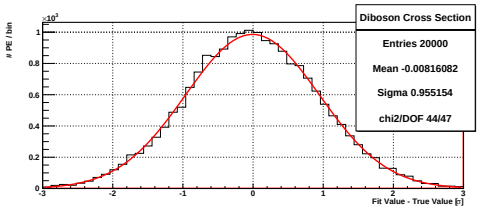
(c)



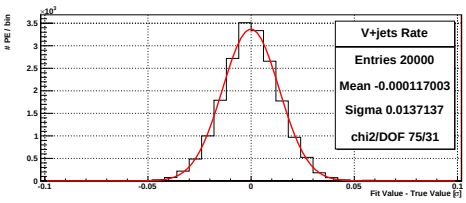
(d)



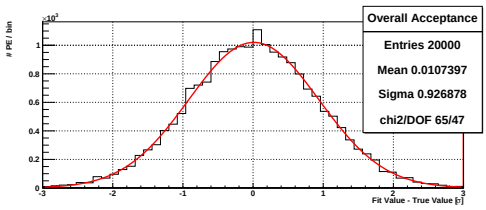
(e)



(f)



(g)



(h)

Figure E.14: *Pull distributions for the JES (a), Q^2 (b), QCD Rates - TCE and CMUP/CMX (c-d), Top Cross and Diboson Cross Sections (e-g), V+jets Rate (h), Overall Acceptance nuisance parameters. These nuisance parameters are part of the fit described in Sec. 8.5.*

Appendix F

Top-AntiTop Veto

In Sec. 9.2 we described the optimization of the selection cuts of the sample used to extract the WZ/ZZ signal. In order to reject one of the largest background ($t\bar{t}$) in the Tag region we develop a veto, which consists of two steps:

- a series of preselection cuts, applied when the number of lepton is two [1]: the goal is also to reject as little ZZ signal as possible;
- the training of Neural Network (NN) on the events that pass these cuts.

The first step rejects events if the two leptons do not have the same flavor (one electron and one muon), or if the invariant mass of the two leptons is not within a $[60, 120]$ Gev/c^2 window. This mass window was optimized against $S/\sqrt{B}$, S being the number of WZ/ZZ signal events and B number of $t\bar{t}$ background events. After the first step we keep about 97% of the signal events while rejecting about 36% of $t\bar{t}$ background events. These numbers are estimated from simulations.

The second step consists in training a NN on the events passing the preselection cuts. The NN has been trained by using the MLP method ([92]) with 9 variables that, apart from di-jet mass, give the best separation between signal and $t\bar{t}$ background [2]. The NN combines kinematical information about the processes: we consider variables related to the angles between the reconstructed objects, to the transverse energy or to the transverse momentum. The list of the exploited vari-

[1] The considered leptons are listed in Sec. 4.2 and 4.3.

[2] When training was performed a weighting was applied in $t\bar{t}$ simulated event to have the same di-jet mass spectrum as signal events. By doing so we decorrelate the NN output from the numerical value of di-jet mass, which is used as final discriminant to extract the signal. Therefore, cutting on the NN output is not expected to modify the background di-jet mass shape.

ables is reported below. This list is ordered in decreasing "importance", as ranked by the TMVA package ([92]).

- H_T: the sum of the transverse energy of all reconstructed objects (leptons, $\not\!\!E_T$ jets);
- $M_T(l_1,\not\!\!E_T)$: the transverse mass of the system composed of the highest-P_T lepton (l_1) and missing transverse energy ($\not\!\!E_T$);
- $P_T(l_1, j_{12})$: the transverse momentum of the system composed of l_1 and the two jets (j_{12});
- P_T^{all}: the transverse momentum of the system of all reconstructed objects;
- $\not\!\!E_T$;
- $P_T^{imbalance}$: module of component-by-component difference between the transverse momenta of j_{12} and of the (l_1, $\not\!\!E_T$) system;
- $\Delta\eta(l_1, j_{12})$: η difference between l_1 and j_{12};
- $P_T(l_1,\not\!\!E_T)$: the transverse momentum of the (l_1, $\not\!\!E_T$) system;
- dijet P_T: the transverse momentum of j_{12};
- di-jet mass: invariant mass of j_{12} [3]

The NN output ("NN^{ttbar}") normalized distributions in signal and background events are shown in F.1: a large separation between the two processes can be seen. By applying an optimal $NN^{ttbar} > 0.15$ cut we keep about 97% of the WZ/ZZ signal events, while rejecting about 47% of the $t\bar{t}$ background events.

In summary, the whole two-step $t\bar{t}$ veto allows us to keep about 94% of the WZ/ZZ signal events while rejecting about 66% of $t\bar{t}$ background events: the signal significance in the Tag region, quantified with the figure of merit $S/\sqrt{B}$, improves by about 10%. These numbers are estimated from simulations.

[3] Although this variable is the same for signal and background because of the aforementioned weighting, we want to exploit the residual discrimination power which comes from the correlation between di-jet mass and the other variables.

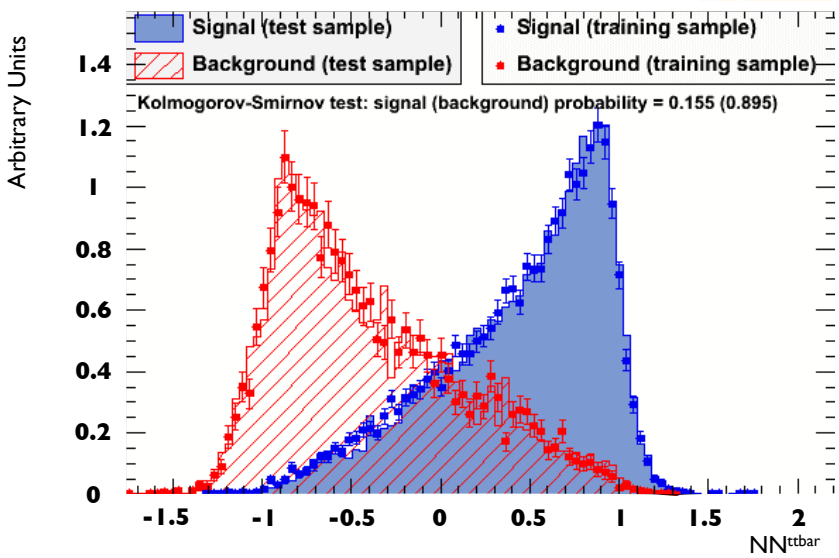


Figure F.1: NN^{ttbar} *distributions in simulated* WZ/ZZ *(blue) and* $t\bar{t}$ *events. These distributions have been normalized to the same area. The distribution are plotted within the TMVA package. See [92] for futher details.*

Bibliography

Abstract

[1] T. Aaltonen *et al.* (CDF Collaboration), "Measurement of the $WW + WZ$ Production Cross Section Using the lepton + jets Final State at CDF II", Phys. Rev. Lett. **104**, 101801 (2009).

[2] F. Abe *et al.*, "Observation of Top Quark Production in ppbar Collisions with the Collider Detector at Fermilab", Phys. Rev. Lett. **74**, 2626-2631 (1995).

[3] T. Aaltonen *et al.* (CDF Collaboration), "Combination of searches for the Higgs boson using the full CDF data set", [arXiv:1301.6668 (2013)].

Chapter 1

[4] D. J. Gross and F. Wilczek, "Asymptotically free gauge theories" Phys. Rev. D **8**, 3633 (1973);
H. D. Politzer, "Asymptotic freedom: An approach to strong interactions" Phys. Rev. Lett. **14**, 129 (1974).

[5] S. L. Glashow, "Partial symmetries of weak interactions" Nucl. Phys. **22**, 579-588 (1961);
A. Salam and J. C. Ward, "Electromagnetic and weak interactions" Phys. Lett. **13**, 168-171 (1964);
S. Weinberg, "A model of leptons" Phys. Rev. Lett. **19**, 1264 (1967).

[6] I. J. R. Aitchison & A.J.G. Hey, "Gauge Theories in Particle Physics (second edition)", Douglas F. Brewer, M.A., D.Phil. (1989).

[7] R.K. Ellis, W.J. Stirling and B.R. Webber, "QCD and Collider Physics", Cambridge University Press (1996).

[8] Wesley Robert Ketchum, "Search for $ZW/ZZ \to l^+l^- +$ Jets Production in $p\bar{p}$ Collisions at CDF", University of Chicago, Chicago U., FERMILAB-THESIS-2012-36 (2012).

[9] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", Phys. Rev. Lett. **13**, 508 (1964).

[10] Beringer, J. and others "Review of Particle Physics (RPP)", Phys. Rev. D **86**, 010001 (2012).

[11] Aad G. *et al.* "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", arXiv:1207.7214. Chatrchyan S. *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", arXiv:1207.7235.

[12] Michael E. Peskin, Daniel V. Schroeder, "An Introduction to Quantum Field Theory", Westview Press (1995).

[13] K. Hagiwara *et al.* "Measuring The WWZ Coupling At The Tevatron" Phys. Rev. D **41**, 2113-2119, (1990).

[14] Matthew R. Buckley, "What the Tevatron Found?", FERMILAB-PUB-11-359-A-T (2011).

[15] T. Corbett *et al.* "Determining Triple Gauge Boson Couplings from Higgs Data" Phys.Rev.Lett. **111**, 011801, (2013).

[16] CMS Public (2013).

[17] Neubauer, M. S. "Diboson production at colliders", Ann.Rev.Nucl.Part.Sci., **61** 223–250, (2011).

[18] Tev4LHC Higgs Working Group. Standard Model Higgs cross sections at hadron colliders. url.

[19] LHC Higgs Cross Section Working Group. "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", CERN-2011-002 [arXiv:1101.0593] (2011).

[20] Acosta D. *et al.*, "Measurement of the W^+W^- production cross section in $p\bar{p}$ collisions at $\sqrt{s}$ =1.96 TeV using dilepton events", Phys. Rev. Lett. **94**, 211801 (2005).

[21] Abazov VM, *et al.*, "Measurement of the WW production cross section in $p\overline{p}$ collisions at $\sqrt{(s)}$ =1.96 TeV", Phys. Rev. Lett. **94**, 151801 (2005).

[22] Abulencia A, *et al.*, "Observation of WZ Production", Phys. Rev. Lett. **98**, 161801 (2007).

[23] Abazov VM, *et al.*, "Measurement of the $p\overline{p} \to WZ + X$ cross-section at $\sqrt{s} =$ 1.96 TeV and limits on WWZ trilinear gauge couplings", Phys. Rev. D, **76** 111104 (2007).

[24] Aaltonen T, *et al.*, "First Measurement of ZZ Production in panti-p Collisions at $\sqrt{s} =$ 1.96-TeV ", Phys. Rev. Lett. **100**, 201801 (2008).

[25] Abazov VM, *et al.*, "Observation of ZZ production in $p\overline{p}$ collisions at $\sqrt{s} =$ 1.96-TeV ", Phys. Rev. Lett. **101**, 171803 (2008).

[26] The LEP collaborations and LEP WW working group, LEP W-pair, Z-pair and Single W Cross Section Results for Summer 2000 Conferences, LEPEWWG/XSEC/2000-01.

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

[28] Abazov VM, *et al.* Phys. Rev. Lett. **102**, 161801 (2009).

[29] Aaltonen, T. *et al.* "Combined search for the standard model Higgs boson decaying to a bb pair using the full CDF data set.", Phys. Rev. Lett. **109**, 111802 (2012).

[30] Abazov, V. M. *et al.* "Combined search for the standard model Higgs boson decaying to $b\overline{b}$ using the D0 Run II data set.", Phys. Rev. Lett. **109**, 121802 (2012) [arXiv:1207.6631].

[31] J. M. Campbell and R. K. Ellis, Phys. Rev. D **60**, 113006 (1999).

[32] T. Aaltonen *et al.* (CDF Collaboration), "Measurement of the W^+W^- Production Cross Section and Search for Anomalous $WW\gamma$ and WWZ Couplings in $p\overline{p}$ Collisions at $\sqrt{s} =$ 1:96 TeV", Phys. Rev. Lett. **104**, 201801 (2010).

[33] V. M. Abazov *et al.* (D0 Collaboration). "Measurement of the WW Production Cross Section with Dilepton Final States in $p\overline{p}$ Collisions at $\sqrt{s} =$ 1.96 TeV and Limits on Anomalous Trilinear Gauge Couplings", Phys. Rev. Lett. **103**, 191801 (2009).

[34] A. Adam *et al.* (CMS Collaboration). "Measurement of WW production rate in pp collision at $\sqrt{s} = 7$ TeV", CMS PAS SMP 12-005 (2012).

[35] G. Aad *et al.* (ATLAS Collaboration). "Measurement of the WW cross section in $\sqrt{s}$ =7 TeV pp collisions with the ATLAS detector and limits on anomalous gauge couplings", Phys. Lett. B **712**, 289-308 (2012).

[36] T. Aaltonen *et al.* (CDF Collaboration). "Measurement of the WZ Cross Section and Triple Gauge Couplings in $p\overline{p}$ Collisions at $\sqrt{s} = 1.96$ TeV", arXiv:1202.6629 (2012).

[37] V. M. Abazov *et al.* (D0 Collaboration). "A measurement of the WZ and ZZ production cross sections using leptonic final states in 8.6 fb^{-1} of $p\overline{p}$ collisions", arXiv:1201.5652v1 (2012).

[38] A. Adam *et al.* (CMS Collaboration), "Measurement of the WW, WZ and ZZ cross sections at CMS", CMS PAS EWK-11-010 (2011).

[39] G. Aad *et al.* (ATLAS Collaboration). "Measurement of the WZ production cross section and limits on anomalous triple gauge couplings in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector." Phys. Lett. B, **709** 341-357 (2012).

[40] T. Aaltonen *et al.* (CDF Collaboration). "Measurement of ZZ Production in Leptonic Final States at $\sqrt{s}$ of 1.96 TeV at CDF", Phys. Rev. Lett. **108**, 101801 (2012). arXiv:1112.2978.

[41] G. Aad *et al.* (ATLAS Collaboration), "Measurement of the ZZ Production Cross Section and Limits on Anomalous Neutral Triple Gauge Couplings in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector", Phys. Rev. Lett. **108**, 041804 (2012).

[42] T. Aaltonen *et al.* (CDF Collaboration). "Measurement of the $WW + WZ$ Production Cross Section Using a Matrix Element Technique in Lepton + Jets Events ", Phys. Rev. D **82**, 112001 (2010).

[43] V. M. Abazov *et al.* (D0 Collaboration). "Measurements of WW and WZ production in $W+$ jets final states in ppbar collisions", Phys. Rev. Lett **108**, 181803 [arXiv:1112.0536v1] (2011).

[44] T. Aaltonen *et al.* (CDF Collaboration). "Evidence for WZ and ZZ production in final states with b-tagged jets at CDF", CDF Public Note 10805 (2012).

[45] V. M. Abazov *et al.* (D0 Collaboration). "Evidence for WZ and ZZ production in final states with b-tagged jets", DØ Conf. Note 6260 [arXiv:1204.4496v1] (2011).

Chapter 2

[46] Link to Fermilab Beam Division (Linac)

[47] Link to Fermilab Beam Division (Main Injector)

[48] Link to Fermilab Beam Division

Chapter 3

[49] D. Acosta *et al.*. (CDF Collaboration), "The CDFII Detector Technical Design Report", Fermilab Report No. Fermilab-Pub-96/390-E
Phys. Rev. D **71**, 032001 (2005).

[50] F. Abe *et al.* (CDF Collaboration), "Evidence for top quark production in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV", Fermilab-Pub-94/097-E.

[51] C.S. Hill (CDF Collaboration), "Initial Experience with the CDF Layer 00 Silicon Detector" Nucl. Instrum. Methods A **511**, 118 (2004).

[52] Diego Tonelli, "First observation of the $B_s \to K^+K^-$ decay mode, and measurement of the B^0 and B_s mesons decay-rates into two-body charmless final states at CDF.", Scuola Normale Superiore, Pisa, FERMILAB-THESIS-2006-23 (2006).

[53] D. Amidei *et al.* (CDF Collaboration), " The Silicon Vertex Detector of the Collider Detector at Fermilab", Nucl. Instr. and Meth. A **350** 73-130 (1994);
S. Cihangir *et al.*, "The SVX-II silicon vertex detector upgrade at CDF", Nucl. Instr. and Meth. A **360**, 118-124 (1995).

[54] A. Affolder *et al.*, "Intermediate silicon layers detector for the CDF experiment", Nucl. Instrum. Methods, A **453**, 84-88 (2000).

[55] C. S. Hill (CDF Collaboration), "Operational experience and performance of the CDFII silicon detector", Nucl. Instrum A **530**, 1-6 (2004).

[56] A. Abulencia *et al.* (CDF Collaboration), "Measurements of Inclusive W and Z Cross Sections in $p\overline{p}$ Collisions at ps = 1.96 TeV", J. Phys. G: Nucl. Part. Phys. **34**, 2457-544 (2007).

[57] T. Affolder *et al.* (The CDF Collaboration), "CDF central outer tracker", Nucl. Instrum. Meth. A **526**, 249-299 (2004).

[58] T. Aaltonen *et. al* (CDF Collaboration), "Operational Experience, Improvements, and Performance of the CDF Run II Silicon Vertex Detector", FERMILAB-PUB-13-015-E.

[59] L. Balka *et al.*, "The CDF Central Electromagnetic Calorimeter", Nucl. Instrum. Methods A **267**, 272 (1988).

[60] M. Gallinaro *et al.* (CDF Collaboration), "New Scintillator TileFiber Preshower Detector for the CDF Central Calorimeter", IEEE Trans. Nucl. Sci. **52**, 879 (2005) [arXiv:physics/0411056];
M. Albrow *et al.* (CDF Collaboration), "A preshower detector for the CDF plug upgrade: Test beam results", Nucl. Instrum. Meth. A **431**, 104-111 (1999);
A. Artikov *et al.*, "CDF Central Preshower and Crack Detector Upgrade.", FNAL-PUB-07-023-E (2007).

[61] S. Bertolucci *et al.*, "The CDF Central and Endwall Hadron Calorimeter", Nucl. Instrum. Methods A **267**, 301-314 (1988).

[62] M. Albrow *et al.*, "The CDF plug upgrade electromagnetic calorimeter: test beam results.", Nucl. Instr. Meth. A 48Nucl. Instrum. Methods A **480**, 524-546 (2002).

[63] G. Apollinari *et al.*, "Shower maximum detector for the CDF plug upgrade calorimeter" Nucl. Instrum. Methods A **412**, 515-526 (1998).

[64] C. M. Ginsburg for the CDF COllaboration, "CDF Run 2 muon system", Eur. Phys. J. C **33**, S1002-1004 (2004).
T. Dorigo (CDF Collaboration), "The muon system upgrade for the CDF II experiment", Nucl. Instrum. Meth. A **461**, 560-2 (2001).

[65] D. Acosta *et al.* (CDF Collaboration), "The CDF Luminosity Monitor", Nucl. Instrum. Meth. A **461**, 540-44 (2001);

D. Acosta *et al.* (CDF Collaboration), "The performance of the CDF luminosity monitor", Nucl.Instrum. Meth. **A**, 57-62 (2002).

[66] Klimenko S., Konigsberg J. and Liss T. M. "Averaging of the inelastic cross sections measured by the CDF and the E811 experiments", FERMILAB-FN-0741.

[67] E.J. Thomson *et al.*, IEEE Trans. on Nucl. Science. **49**, 1063-1070 (2002).

[68] B. Ashmanskas *et al.* (CDF Collaboration), "The CDF Silicon Vertex Trigger", Nucl. Instrum. Meth. A **518**, 532-36 (2004);
I. Vila for the CDF Collaboration, "Performance and First Physics Results of the SVT Trigger at CDF II", ECONF C0304052:FO008 (2003).

[69] A. Adelman, *et al.*, "The Silicon Vertex Trigger upgrade at CDF", Nucl. Instrum. Methods A **572**, 361-64 (2007).

[70] A. Canepa *et al.*, "Level-3 Calorimetric Resolution available for the Level-1 and Level-2 CDF Triggers", [arXiv:0810.3738] (2008).

Chapter 4

[71] J. Mariner, "Secondary Vertex Fit With Mass and Pointing Constraints (CTVMFT)", CDF Internal Note 1996 (1993).

[72] C. Hays *et al.*, "The COT Pattern Recognition Algorithm and Offline Code" CDF Internal Note 6992.

[73] B.R. Ko *et al.* "SVXII Stand-alone Tracking" CDF internal note 6440.

[74] Y. Huang *et al.* "Inside-Out Tracking" CDF internal note 6707.

[75] K. Bloom *et al.* ""Outside-In" Silicon Tracking at CDF" CDF internal note 5991.

[76] J. Goldstein *et al.* "Silicon Tracking for Plug Electrons" CDF internal note 5970.

[77] D. Acosta *et al.* (CDF Collaboration), "Measurement of the $t\bar{t}$ production cross section using lepton+jets events with secondary vertex b-tagging", Phys. Rev. D **71**, 052003 (2005).

[78] Abe F *et al*, "Evidence for top quark production in $p\overline{p}$ collisions at $\sqrt{s}$ =1.8 TeV", Phys. Rev. D **50**, 2966-3026 (1994).

[79] D. Acosta *et al.* (CDF Collaboration), "Direct Photon Cross Section with Conversions at CDF", Phys. Rev. D **70**, 074008 (2004);
Alon Attal, Anadi Canepa, "Photon Conversion Removal Efficiency", CDF Internal Note 8073.

[80] Spreitzer *et al*, "Electron ID Efficiency and Scale Factor with TopNtuple 6.1.4 for Summer 2006" CDF Internal Note 8274.

[81] A. Taffard, "Run II Cosmic Ray Tagger Performances", CDF Internal Note 6255.

[82] Grundler *et al*, "High-P_T muons recommended cuts and efficiencies for Summer 2006", CDF Internal Note 8262.

[83] Ulysses, Grundler, Lubomir, Lovas, Anyes, Taffard, "High-Pt muons recommended cuts and efficiencies for Winter 2007" CDF Internal Note 8618.

[84] F. Abe *et al.* (The CDF Collaboration), "Measurement of the W Boson Mass" Phys. Rev. D **52**, 4784-4827 (1995).

[85] S. Eidelman *et al.*, "Review of particle physics. Particle Data Group ", Phys. Lett. B **592**, 1 (2004).

[86] S.R. Hahn *et al.*, Nucl. Instrum. Meth. A **267**, 351-366 (1988);
S. Bertolucci *et al.*, "The CDF Central and Endwall Hadron Calorimeter", Nucl. Instrum. Meth. A **267**, 301 (1988).

[87] A. Bhatti *et al.*, "Determination of the jet energy scale at the collider detector at Fermilab", Nucl. Instrum. Meth. A **566**, 375-412 (2006).

[88] S.V. Chekanov, "Jet algorithms: a minireview", [arxiv:0211298v1] (2002).

[89] G. C. Blazey *et al.*, "Run II jet physics" [arXiv:0005012] (2000).

[90] Maria Spiropopulu, "A Blind search for supersymmetry in $p\overline{p}$ collisions at $\sqrt{\mathrm{s}}$ = 1.8TeV using the missing energy plus multijet channel", Harvard U. FERMILAB-THESIS-2000-16 (2000).

Chapter 5

[91] J. Freeman *et. al*, "An Artificial neural network based b jet identification algorithm at the CDF Experiment", Nucl. Instrum. Meth. A **663**, 37-47 (2012).

[92] A. Hoecker *et al.* PoS ACAT **040** (2007).

[93] T. Sjöstrand *et al.*, J. High Energy Phys. **05**, 26 [arXiv:hep-ph/0603175] (2006).

[94] M. Goncharov *et al.*, Nucl. Instrum. Methods A **565**, 5643 (2006). SBAGLIATA

[95] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, "a generator for hard multiparton processes in hadronic collisions." JHEP **0307**, 001 (2003);
Michelangelo L. Mangano, Mauro Moretti, and Roberto Pittau. "Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b} + n$ jets as a case study", Nucl.Phys. B **632**, 343–362 (2002).

[96] F. Caravaglios, Michelangelo L. Mangano, M. Moretti, R. Pittau, "A new approach to multi-jet calculations in hadron collisions", Nucl.Phys. B **539**, 215-232 (1999).

Chapter 6

[97] A.A. Bhatti, K. Hatakeyama, "Relative jet energy corrections using missing ET projection fraction and djet balancing", CDF Internal Note 6854 (2004);
A.A. Bhatti, K. Hatakeyama, "Update on the relative jet energy corrections", CDF Internal Note 7354 (2005).

[98] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, *et al.*, "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)", JHEP **0101** 010 (2001).

[99] T. Aaltonen *et al*, "Precision Top-Quark Mass Measurements at CDF ", Phys. Rev. Lett. **109**, 152003 (2012).

[100] C. DeLuca Silberberg, "Measurement of the inclusive isolated prompt photon production cross section at the Tevatron using the CDF detector", Autonoma U. (Barcelona), FERMILAB-THESIS-2009-18.

[101] W. Ketchum, V. Rusu, M. Trovato, "New Jet Energy Scale Corrections for Quarks and Gluons", CDF Internal Note 10829.

Chapter 7

[102] H.L. Lai *et al.* (CTEQ Collaboration), "Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions", Eur. Phys. J. C **12**, 375-392 (2000).

[103] T. Stelzer and W.F. Long, "Automatic generation of tree level helicity amplitudes", Comput. Phys. Commun. **81**, 357–371 (1994).

[104] Fabio Maltoni and Tim Stelzer, "MadEvent: Automatic event generation with MadGraph", JHEP **0302**, 027 (2003);
Johan Alwall *et al.*, "MadGraph/MadEvent v4: The New Web Generation", JHEP **09** 028 (2007).

[105] V.N. Gribov and L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 438-450 (1972);
Yu. L. Dokshitzer, Sov. J. Phys. JETP **46**, 641 (1977);
G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", Nucl. Phys. B **126**, 298 (1977).

[106] Link to P. Skands' talk (1/2);
Link to P. Skands' talk (2/2)

[107] S. Catani, F. Krauss, R. Kuhn, B.R. Webber, "QCD Matrix Elements + Parton Showers", CERN-TH/2000-367 [hep-ph/0109231] (2001).

[108] X. Artru and G. Mennessier, "String model and multiproduction", Nucl. Phys. B **70**, 93-115 (1974);
M.G. Bowler, "e^+e^- Production of Heavy Quarks in the String Model", Zeit. Phys. C **11**, 169 (1981);
B. Andersson, G. Gustafson and B Söderberg, "A General Model for Jet Fragmentation", Zeit. Phys. C **20**, 317 (1983); Nucl. Phys. B **264**, 29 (1986).

[109] R.D. Field and R.P. Feynman, "Quark Elastic Scattering as a Source of High Transverse Momentum Mesons ", Phys Rev. D **15**, 2590-2616 (1977);

[110] R.D. Field and S. Wolfram, "A QCD Model for e^+e^- Annihilation", Nucl. Phys B **213**, 65 (1983);
B.R. Webber, "A QCD Model for Jet Fragmentation Including Soft Gluon Interference", Nucl. Phys. B **238**, 492 (1984).

[111] Rene Brun, Federico Carminati, and Simone Giani, "GEANT Detector Description and Simulation Tool. 1994. CERN Program Library Long Writeup".

[112] Guenter Grindhammer, M. Rudowicz, and S. Peters, "The Fast Simulation Of Electromagnetic And Hadronic Showers", Nucl. Instrum. Meth. A **290**, 469 (1990).

[113] Aaltonen T. *et al.* (CDF Collaboration), "Observation of Single Top Quark Production and Measurement of $|V_{tb}|$ with CDF", Phys. Rev. D **82**, 112005 (2010).

[114] Aaltonen T. *et al.*, "Measurement of cross sections for b jet production in events with a Z boson in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV", Phys. Rev. D **79**, 052008 (2009).

[115] U. Langenfeld, S. Moch, and P. Uwer. "Measuring the running top-quark mass.", Phys. Rev. D **80**, 054009 (2009);
A. Abulencia *et al.* (CDF Collaboration), "Measurement of the top quark mass using template methods on dilepton events in proton antiproton collisions at $\sqrt{s} = 1.96$ TeV" Phys. Rev. D **73**, 112006 (2006).

[116] Aaltonen T. *et al.* (CDF Collaboration), "Measurement of the cross section for W^- boson production in association with jets in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV", Phys. Rev. D **77**, 011108 (2008).

[117] N. Kidonakis, "Single top quark production cross section at hadron colliders", arXiv:1005.3330 (2010);
N. Kidonakis, "NNLL resummation for s-channel single top quark production", Phys. Rev. D **81**, 054028 (2010).

[118] N. Kidonakis, "Single top production at the Tevatron: Threshold resummation and finite-order soft gluon corrections", Phys. Rev. D **74**, 114012 (2006).

Chapter 8

[119] T. Aaltonen *et. al* (CDF Collaboration), "Invariant Mass Distribution of Jet Pairs Produced in Association with a W Boson in $p\overline{p}$ Collisions at $\sqrt{s}$ 1.96 TeV", Phis. Rev. Lett. **106**, 171801 (2011);
A. Annovi, P. Catastini, V. Cavaliere, R. Ristori, "Invariant Mass distribution of Jet Pairs Produced in Association with a W in 7.3 fb^{-1}", CDF Public Note 10518.

[120] V. Boisvert, "Trigger Efficiencies for the High E_T Central Electrons in the Gen6 data", CDF Internal Note 7939 (Perfidia).

[121] S. Rolli and D. Glenzinski (CDF Collaboration). PerfIDia: an automatic tool for trigger efficiency and SF evaluation.

[122] U. Grundler, A. Taffard, X. Zhang, "High-P_T muons recommended cuts and efficiencies for Summer 2006" CDF Internal Note 8262.

[123] Link To CDF MC Page.

[124] M. Franklin *et al.* "Heavy-Flavor Content of the W+Jets Sample" CDF Note 8765 (2007).

[125] B. Cooper and A. Messina, "Estimation of the Background to $W \to e\nu$ jets Events", CDF Internal Note 7760.

[126] Link To FractionFitter WebPage;
R. Barlow and C. Beeston, Comp. Phys. Comm. **77**, 219-228 (1993).

[127] T. Aaltonen *et al.* (CDF Collaboration), "Search for the standard model Higgs boson produced in association with a $W^{\pm}$ boson with 7.5 fb^{-1} integrated luminosity at CDF", Phys. Rev. D **86** 032011 (2012).

[128] Marco Trovato, "A New Top Mass Measurement in The Dilepton Channel" Master's Thesis, University of Pisa (2008) FERMILAB-MASTERS-2008-10.

[129] T. Junk, "Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties." CDF Public Note 8128 (2007).

[130] F. James and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations", Comput. Phys. Commun. **10**, 343 (1975).

[131] F. Canelli *et al.*, "Jet Energy Resolution", CDF Internal Note 7856.

[132] T. Aaltonen *et al.* (The CDF Collaboration), "Search for WZ+ZZ production with missing transverse energy+jets with b enhancement at $\sqrt{s} = 1.96$ TeV", Phys. Rev. D **85**, 012002 (2012).

[133] Evelyn Thomson *et. al*, "Combination of CDF and D0 ttbar cross-sections", CDF Internal Note 10916.

[134] Feldman, G. J. and R. D. Cousins. "Unified approach to the classical statistical analysis of small signals.", Phys. Rev. D**57**, 3873-3889 (1998).

Chapter 9

[135] G. Chiarelli, F. Sforza "MET_PEM Trigger Efficiency up to Summer 2008", CDF Internal Note 9493.

[136] Federico Sforza, "Evidence for Diboson Production in the Lepton plus Heavy Flavor Jets Final State at CDF", Pisa University, FERMILAB-THESIS-2012-41 (2012).

[137] A. Buzatu. "Search for the Standard Model Higgs boson produced in association with a W Boson in the isolated-track charged-lepton channel using the Collider Detector at Fermilab", McGill U.,FERMILAB-THESIS-2011-24 (2011).

[138] B. Heinemann and K. Pitt (for the Higgs Trigger Task Force). "Report of the Higgs Trigger Task Force". CDF Internal Note 8875 (2007).

[139] Karolos J. Potamionos, "Search for the Higgs Boson and Rare Standard Model Processes in the MET+b-jets Signature at the Collider Detector at Fermilab", Purdue U., FERMILAB-THESIS-2011-34 (2011).

Appendix A

[140] D. Acosta *et al.* (CDF Collaboration), "Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s}$ =1.96 TeV using Lepton + Jets Events with Secondary Vertex b-tagging SecVtx Tagger", Phys. Rev. D **71**, 052003 (2005).

[141] Henry Bachacou, "A Measurement of the Production Cross Section of tt-bar Pairs Using Secondary Vertex b-tagging", FERMILAB-THESIS-2004-48, UC. Berkeley, (2004).

[142] D. Acosta *et al.* (CDF Collaboration), "Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96 TeV Using Lepton+Jets Events with Semileptonic B Decays to Muons", Phys. Rev. D **72**, 032002 (2005).

[143] A. Abulencia *et al.*, "Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using Lepton + Jets Events with Jet Probability b-tagging", Phys. Rev. D **74**, 072006 (2006).

[144] C. Ferrazza, "Identificazione di quark pesanti in getti adronici in interazioni $p\bar{p}$ con il rivelatore CDF al Tevatron, Master's thesis", Universita "La Sapienza" Roma (2006).
P. Mastrandrea, "Study of the heavy flavour fractions in Z+jets events from $p\bar{p}$ collisions at energy = 1.96 TeV with the CDF II detector at the Tevatron collider", Universita "La Sapienza" Roma, FERMILAB-THESIS-2008-63 (2008).

[145] Justing Kien Keung, "Search for the production of the standard model Z boson in association with W boson in p - anti-p collisions at 1.96 TeV center of mass energy", Pennsylvania U., FERMILAB-THESIS-2010-32 (2010).

[146] J. Freeman *et al.*, "Introduction to HOBIT, a b-Jet Identification Tagger at the CDF Experiment Optimized for Light Higgs Boson Searches", Nucl. Instr. Meth. A **697** 64-76 (2013).

Appendix B

[147] JADE collaboration, W. Bartel *et al.*, "Comparison Of Three Jet Events With Qcd Shower Models", Phys. Lett. B **157**, 340 (1985);
TPC collaboration, H. Aihara *et. al.*, "Tests of models for quark and gluon fragmentation in e^+e^- annihilation $\sqrt{s}$ =29 GeV", Zeit. Phys. C **28**, 31 (1985);
TASSO collaboration, M. Althoff *et al.*, "A Study of Three Jet Events in e^+e^- Annihilation Into Hadrons at 34.6-GeV Center-Of-Mass Energy" , Zeit. Phys. C **29**, 29 (1985).