UNIVERSITÀ DEGLI STUDI DI UDINE

DOTTORATO DI RICERCA IN MATEMATICA E FISICA

XXVII CICLO

Tesi di Dottorato

Measurement of the $W+$ jets differential cross-sections in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using the CDF II Detector

Supervisore:
Prof. Giovanni Pauletta

Dottoranda:
Anna Driutti

Co-supervisori:
Dr. Vadim Rusu
Dr. Marco Trovato

Abstract

In this thesis the measurements of differential cross sections for the production of the W-boson in association with jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,\text{TeV}$ are presented. The measurements are based on $9.0\,\text{fb}^{-1}$ of CDF Run II data (*i.e.*, the full dataset). Only events in which the W-boson decays leptonically (*i.e.*, $W \to e\nu$ and $W \to \mu\nu$) and at least one jet is present are considered. The lepton candidates are required to have a transverse energy $E_T^\ell > 25\,\text{GeV}$ and pseudorapidity in the range $|\eta^\ell| < 1$ whereas, the jets are reconstructed using the JETCLU algorithm with a radius of 0.4 requiring transverse energy $E_T^{jet} > 25\,\text{GeV}$ and pseudorapidity in the range $|\eta^{jet}| < 2$. The reconstructed W-boson transverse mass should be greater than $40\,\text{GeV/c}^2$.

The differential cross sections as a function of the jet multiplicity ($N \geq 1,\ 2,\ 3,\ 4$) and the leading jet transverse energy, are measured separately for each decay channel and then combined.

For a meaningful comparison with theory the measured cross-sections are unfolded to remove detector effects. The resulting particle-level cross-sections are compared to theoretical predictions.

Contents

List of Figures

Introduction

Collider experiments have been one of the most efficient way of testing the predictions of the Standard Model (SM) of particle physics. The Tevatron, located near Chicago (USA), was the highest energy proton-antiproton collider until the recent startup of the LHC. The center-of-mass energy of the collisions was 1.96 Tev and during the RunII (February 2002 - September 2011) about $12\,\mathrm{fb}^{-1}$ of integrated luminosity were delivered. The CDF (Collider Detector at Fermilab) experiment recorded about $10\,\mathrm{fb}^{-1}$ of this integrated luminosity and the analysis presented in this thesis is based on this dataset.

We search for events in which a W boson is produced in association with jets and decays leptonically. In particular, we look for a final state with a lepton (e or μ), a neutrino (missing transverse energy, $\not\!\!E_T$) and one (or more) jets. A precise understanding and modeling of this process is essential for precision measurements of Higgs, Top-quark and beyond SM physics since the W+jets is one of the main backgrounds for these physics processes. In addition, the $W+$ jets process is a testing ground for calculations based on the perturbative Quantum ChromoDynamics (pQCD), the model of the strong interactions in the high-energy regime.

The main challenges of the analysis are the accurate modeling of the backgrounds that need to be subtracted from the data in order to measure the W+jets cross-section, the precise reconstruction and measurement of the jets and the careful estimate of the systematic uncertainties. The analysis presented here is unique with regard to the method used the model the QCD multi-jet background and the use of a jet energy scale (JES) that accounts for the difference of the quark-jets and the gluon-jets. The corrections to the multijet model and the quark/gluon JES were previously derived by [84] for the analysis of W+2 jets production and are here, for the first time, tested and used in the more generalized $W+$ jets process. Moreover, it is the first W+jets analysis that uses the complete dataset delivered in Run II by the Tevatron.

In Chapter 1 after an overview of the Standard Model (SM) of particle physics, the fundamentals of the Quantum Chromodynamics (QCD) are summarized. Chapter 2 is dedicated to the production of a W boson in association with jets at hadron colliders. Motivations for the study of the W+jets process are expanded on and the details of the theoretical predictions and previous results are given. Chapter 3 contains the description of the Tevatron accelerator complex, the CDF detector and the general procedures for reconstructing objects of interest to the analysis: leptons, neutrinos, and jets. The analysis technique and the experimental measurement are described in Chapter 4, where the datasets used to perform the measurement, the

triggers, and the event selection requirements are discussed. Most of this chapter is devoted to explaining the method used to estimate and subtract the background, with particular attention to the modeling of the QCD multi-jet background. In addition, detector-level cross sections obtained are unfolded to particle-level, the systematic uncertainties of the measurement are evaluated and the measurements in the electron and muon channels are combined. The ensuing results are discussed in Chapter 5 and compared with theoretical predictions.

Chapter 1

The Standard Model and The QCD Theory

The Standard Model (SM) is the theory that describes the properties of the elementary particles and the fundamental interactions between them. Through this theory the electromagnetic, weak and strong interactions have been unified. Since its formulation (mid-1970s), many of the Standard Model predictions have been successfully verified. However although it is presently the best model of the phenomenology of particle physics, the theory has some limits and leaves some phenomena unexplained.

A brief review of the particles and the interactions included in the Standard Model is presented in section 1.1 of this chapter. The theory of the strong interaction, Quantum ChromoDynamics (QCD), which is the theoretical framework for the interpretation of the measurement presented in this thesis, is introduced in section 1.2 while its application to hadronic collisions is described in section 1.3. Section 1.4.1 introduces the concept of jets which is fundamental to this thesis.

1.1 The Standard Model of Particle Physics

The Standard Model (SM) is a quantum field theory (QFT) that describes the fundamental constituents (*i.e* particles that do not have an internal structure) of matter and the interactions between them. The SM is based on the gauge group of symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$. The gauge group $SU(3)_C$ defines the theory of the strong interactions known as Quantum ChromoDynamics, QCD whereas the symmetry $SU(2)_L \times U(1)_Y$ provides, a theoretical description of the electroweak (electromagnetic and weak) interactions [2] [3] [4] [5].

The Standard Model theory includes twelve elementary fermions, with half-integer spin and their twelve antiparticle counterparts. Particle and antiparticles have the same quantum numbers but opposite electric charge. Neutral particles are their own antiparticles. Fermions obey the Fermi Dirac statistics and are constrained by the Pauli exclusion principle. They are further grouped into three lepton and three quark families. Leptons consist of electrons (e), muons (μ), taus (τ) and their associated neutrinos (ν_e, ν_μ and ν_τ). The six types (conventionally referred to as flavors) of quarks are: up (u), down (d), charm (c), strange (s), top (t) and bottom (b).

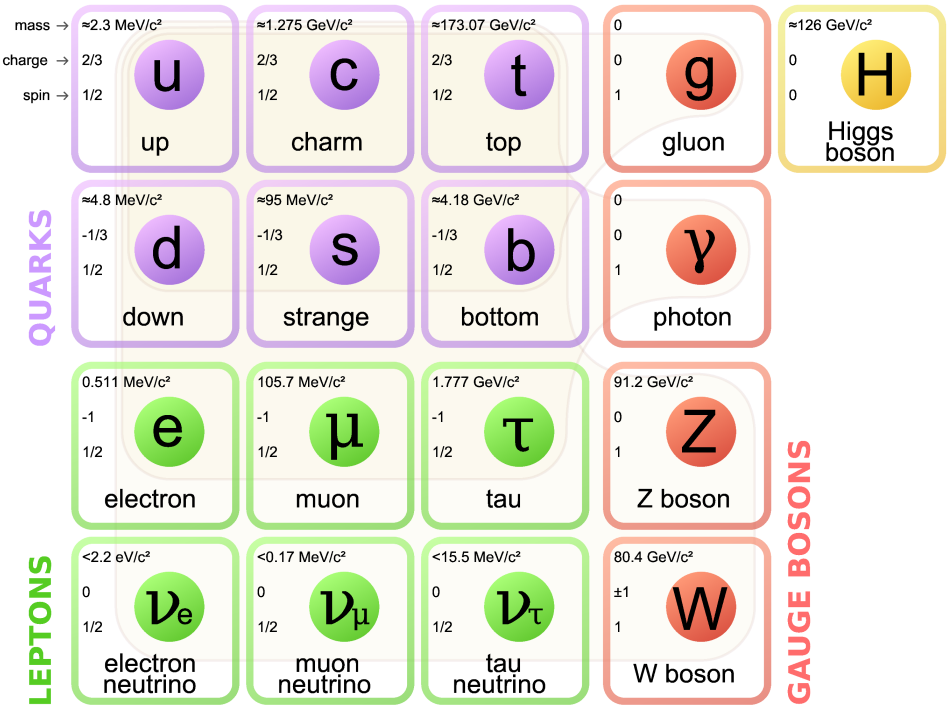


Figure 1.1: Schematic depiction of the Standard Model particle content. The three generations of leptons and quarks are in green and purple respectively, the gauge bosons are in the fourth column, and the Higgs boson in the fifth. Figure from [6]

Interactions between particles occur through the exchange of vector bosons. The force carrier particles (gauge bosons) are integer spin particles that obey the Bose-Einstein statistics. These force mediators are the photon (γ) for the electromagnetic interaction, the gluon (g) for the strong force, the $W^{\pm}$ and the Z^0 for the weak force. The photon and the gluon are massless whereas the W-boson and the Z-boson are massive. The mass of the Standard Model particles is introduced by the spontaneous symmetry breaking triggered by the Higgs mechanism [7][8] whereby, when the particles interact with the Higgs field, they acquire a mass proportional to their coupling with the field. The Higgs mechanism requires the existence of another boson as a carrier particle of the scalar Higgs field: the Higgs boson. The fundamental particles of the Standard Model and some of their properties are summarized in table 1.1. Each of these particles have been experimentally detected and studied.

The Standard Model is an incomplete theory in the sense that it does not explain some phenomena (*e.g.*, neutrino's masses) and it does not include gravitation [9].

1.2 Quantum Chromodynamics (QCD)

Quantum Chromodynamics (QCD) is a non-Abelian gauge theory based on the gauge group $SU(3)$ and it is the part of the Standard Model that describes the strong interactions binding quarks in hadrons. Quarks are the elementary constituents of the hadronic matter, and they are the only SM elementary particles which experience the strong force. All observed hadrons are either mesons (quark-anti-quark bound states) or baryons (bound states of three quarks). To form bound states of three quarks and, at the same time satisfy the requirement of a total antisymmetric wave function imposed by the Fermi-Dirac statistics, quarks carry an additional quantum number called color. It is been experimentally confirmed that there are three colors: red (R), green (G) and blue (B) and three anti-colors: anti-red ($\bar{R}$), anti-green ($\bar{G}$) and blue ($\bar{B}$). The carrier of the strong force, the gluon, carries color and anti-color charge itself because the non-abelian nature of the theory. Moreover, since $SU(3)$ has eight generators, there are eight color-variated gluons. As a consequence, a propriety of the gluons is that they can self-interact. This constitutes a substantial difference between the strong (QCD) and electromagnetic (Quantum ElectroDynamics, QED) interactions. The force carrier of the latter, the photon, is electrically neutral thereby photon self-interactions are forbidden.

The QCD Lagrangian can be written as:

$$\mathcal{L} = \sum_q \bar{\psi}_i^q (i\gamma_\mu D_{ij}^\mu - m_q \delta_{ij}) \psi_j^q - \frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu} \tag{1.2.1}$$

where the quark (anti-quark) fields are denoted by ψ_i^q ($\bar{\psi}_i^q$) indexed according to the color with i or j and flavor q, m_q is the mass of the quark which is a free parameter of the theory, γ^μ is the Dirac matrix which expresses the vector nature of the strong interaction, $F_{\mu\nu}^a$ is the gluon field strength tensor for a gluon with color index a, and D^μ is the so-called covariant derivative,

$$D_{ij}^\mu = \partial_\mu \delta_{ij} + i g_s t_{ij}^a A_a^\mu \tag{1.2.2}$$

where g_s is the strong coupling constant, A_a^μ the gluon field with color index a and t_{ij}^a the generators of $SU(3)$. The QCD analog of the fine structure constant, α_s, can be defined from the strong coupling constant g_s by $g_s^2 = 4\pi\alpha_s$. Since gluons carry color charge, besides fermionic (quark) loops, gluonic loops are also present. Calculations beyond the tree level, involving gluon and quark internal loops diverge logarithmically with coefficients of opposite sign. To handle such divergences a renormalization procedure is necessary and a renormalization scale μ, related to the momentum transfer Q in a given process, is introduced. As a consequence the strong coupling constant α_s becomes a function of the scale of the process *i.e.*, a running constant and must be specified at a given reference scale, $\alpha_s(Q^2)$ in order to have an indication of the effective strength of the strong interaction in that process [10]. Figure 1.2 shows the theoretical and experimental results (as of 2013) for the running coupling constant as a function of Q in which the reference scale is chosen to be $Q^2 = M_Z^2$.

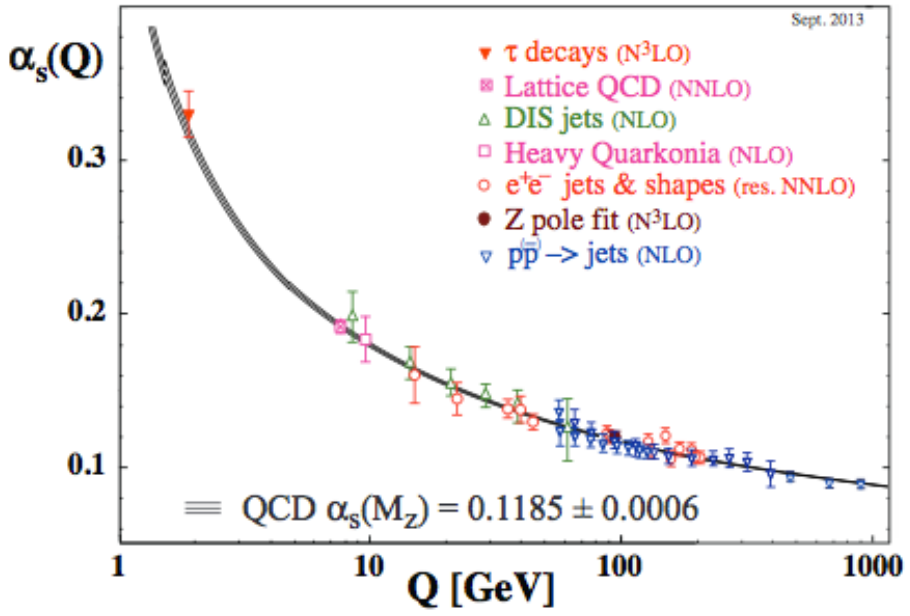


Figure 1.2: Summary of the QCD running coupling constant α_s measurements as a function of the energy scale Q. Figure from [10]

1.2.1 Asymptotic Freedom and Confinement

The running constant for the strong interactions, as shown in figure 1.2, is significant at low Q^2 and decreases as Q^2 increases. This property, whereby the strong coupling becomes small for "hard" (*i.e.*, high-momentum transferred processes) processes, is known as asymptotic freedom whereas the increase of the strong coupling constant with decreasing momentum transfer (*i.e.*, increasing distance), is referred to as "confinement". As a result this confining propriety, with the exception of

the top quark which decays before it has time to hadronize, free quarks are not observed since they hadronized on a time scale $\sim 1/Q^2$. This implies that only colorless (color-singlet) hadronic states can be observed in nature. The low-energy regime is described by lattice gauge theories which provides a way to compute the hadron mass spectrum directly from the QCD Lagrangian [11]. On the other hand, as consequence of the asymptotic freedom, the perturbation approach becomes applicable for high momentum transfers, where the coupling is weak.

1.3 QCD in hadronic collisions

As momentum transfers increase, interactions between the individual constituents (quarks and gluons) become predominant in hadronic collisions. Understanding the internal structure of the hadrons is therefore important for calculation of their cross-sections at hadron-hadron colliders. Deep inelastic lepton-hadron scattering (DIS) experiments have been the basic tools used to investigate the structure of hadrons in detail. At very high momentum transfers the lepton interacts with the elementary constituents of the hadron so that these measurements contain the information about the hadron's structure. These studies leads to the so called parton model (Bjorken and Feynman 1969) [12][13]. The model describes the fast moving hadrons as a beam of almost free collinear point-like particles named partons moving in the same direction of the hadron and sharing its momentum so that the fraction of the hadron momentum (p) carried by the parton i is:

$$p_i = x_i p, \tag{1.3.1}$$

where the dimensionless variable x_i can assume a value between 0 and 1 and represents the fraction of the hadron momentum carried by the parton, so that $\sum_i x_i = 1$.

In the framework of the parton model, the cross-sections of processes where a proton collides with an antiproton, $\sigma_{p\bar{p}\to X}$, at high centre-of-mass energy, s, can be expressed as:

$$\sigma_{p\bar{p}\to X}(s) = \sum_{ij} \int_0^1 dx_1 dx_2 \; f_i^p(x_1, \mu_F) f_j^{\bar{p}}(x_2, \mu_F) \; \sigma_{ij\to X}(x_1, x_2, s, \alpha_s(\mu_R)). \tag{1.3.2}$$

which is factorized into a partonic hard scattering cross-sections ($\sigma_{ij\to X}$) that can be described with a perturbative approach, and probability densities to find a parton i with fractional momentum x_i in the proton (or antiproton), described by parton distributions functions (PDFs) $f_i(x_i, \mu_F)$. The factorization scale μ_F is introduced to absorb collinear divergences, and usually it is set at the renormalization scale $\mu_F^2 = Q^2$ (improved parton model).

This aspect of the theory is known as the QCD factorization theorem, which states that any QCD process can be written as the convolution of its high-energy (perturbative QCD) and low-energy (non-perturbative QCD) components [15]. The factorization scale μ_F is the value that distinguishes these two phases of the process, *i.e.*, the complex structure of the hadrons in which quarks and gluon are confined

and the hard scattering. Pictorially the main phases of the hadron-hadron collision can be described as in figure 1.3.

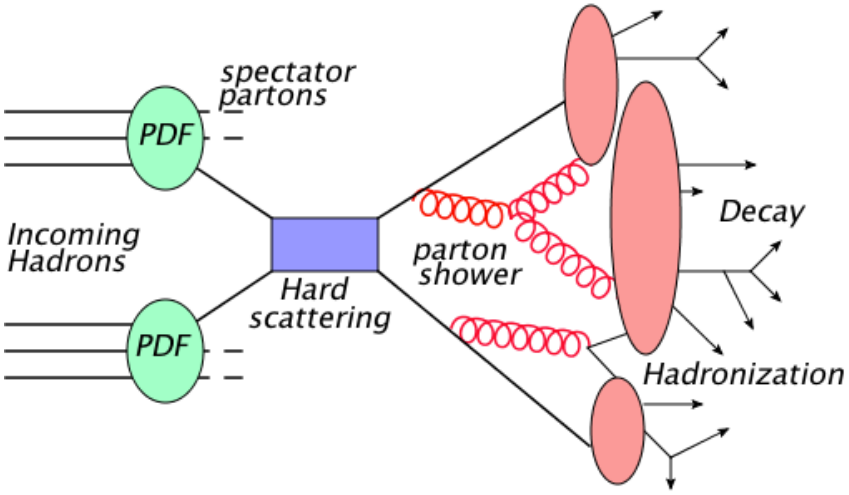


Figure 1.3: Pictorial representation of an hadron-hadron collision.

1.3.1 Description of the Incoming Hadrons with PDFs

The proton and anti-proton structures before the interaction are described by the PDFs. Perturbative QCD can be used to describe their evolution with the variation of the factorization scale. The quark PDF $q(x, \mu_F^2)$ and gluon PDF $g(x, \mu_F^2)$ evolve following the DGLAP equations (called after Dokshitzer, Gribov, Lipatov, Altarelli and Parisi [14]):

$$\frac{dq(x,\mu_F^2)}{d\ln\mu_F^2} = \frac{\alpha_s(\mu_R^2)}{2\pi}\int_x^1 \frac{dz}{z}\Big[P_{qq}(z)q(x/z,\mu_F^2) + P_{qg}(z)g(x/z,\mu_F^2)\Big] \tag{1.3.3}$$

$$\frac{dg(x,\mu_F^2)}{d\ln\mu_F^2} = \frac{\alpha_s(\mu_R^2)}{2\pi}\int_x^1 \frac{dz}{z}\Big[P_{gg}(z)q(x/z,\mu_F^2) + P_{qg}(z)g(x/z,\mu_F^2)\Big] \tag{1.3.4}$$

where $P_{qq}(z)$, $P_{gq}(z)$, $P_{gg}(z)$ and $P_{qg}(z)$ are the probabilities that a quark splits into a quark via gluon radiation ($q \to qg$), that a gluon splits to a quark anti-quark pair ($g \to q\bar{q}$), that a gluon splits into a gluon pair ($g \to gg$) and that a quark splits into a gluon ($q \to gq$). After the splitting, each parton carry a fraction z of momentum. At leading order the splitting functions can be expressed as:

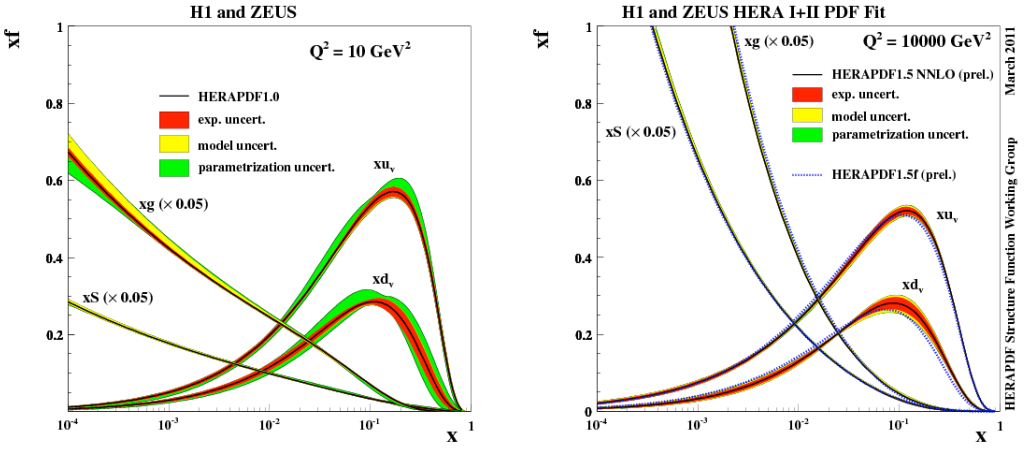
$$P_{qq}(z) = \frac{4}{3}\left[\frac{1+z^2}{1-z}\right] \tag{1.3.5}$$

$$P_{gq}(z) = \frac{4}{3}\left[\frac{1+(1-z)^2}{z}\right] \tag{1.3.6}$$

$$P_{gg}(z) = 6\left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z)\right] \tag{1.3.7}$$

$$P_{qg}(z) = \frac{1}{2}\left[z^2(1-z)^2\right] \tag{1.3.8}$$

PDFs are process-independent functions which are experimentally determined, and then used in all the calculations. Quark and anti-quark PDFs can be directly constrained by fitting the experimental data at different scales, for example in electron-proton DIS experiments as shown in figure 1.4, while the gluon PDFs can be indirectly constrained by assuming that the DGLAP evolution equation is valid (also in figure 1.4).



(a) (b)

Figure 1.4: Examples of parton distribution functions of the proton at $Q^2 = 10$ GeV^2 (a) and $Q^2 = 10^4$ GeV^2 from DIS experiments. Figure from [16]

The evolution equations can be also tested with nucleon DIS experiments by probing the structure function $F_2(x, Q^2)$ as a function of Q^2. At leading-order $F_2(x, Q^2)$ can be expressed as the sum over all the quarks in the hadrons:

$$F_2(x, Q^2) = \sum_{i=q,\bar{q}} e_i^2 x f_i(x, Q^2). \tag{1.3.9}$$

where e_i is the quark's charge. Experiment has confirmed the QCD predictions. As an example figure 1.5 shows the comparison of the measured structure function of the proton as a function of Q^2 for different values of x with the global fit (DGLAP predictions) by the ZEUS Collaboration.

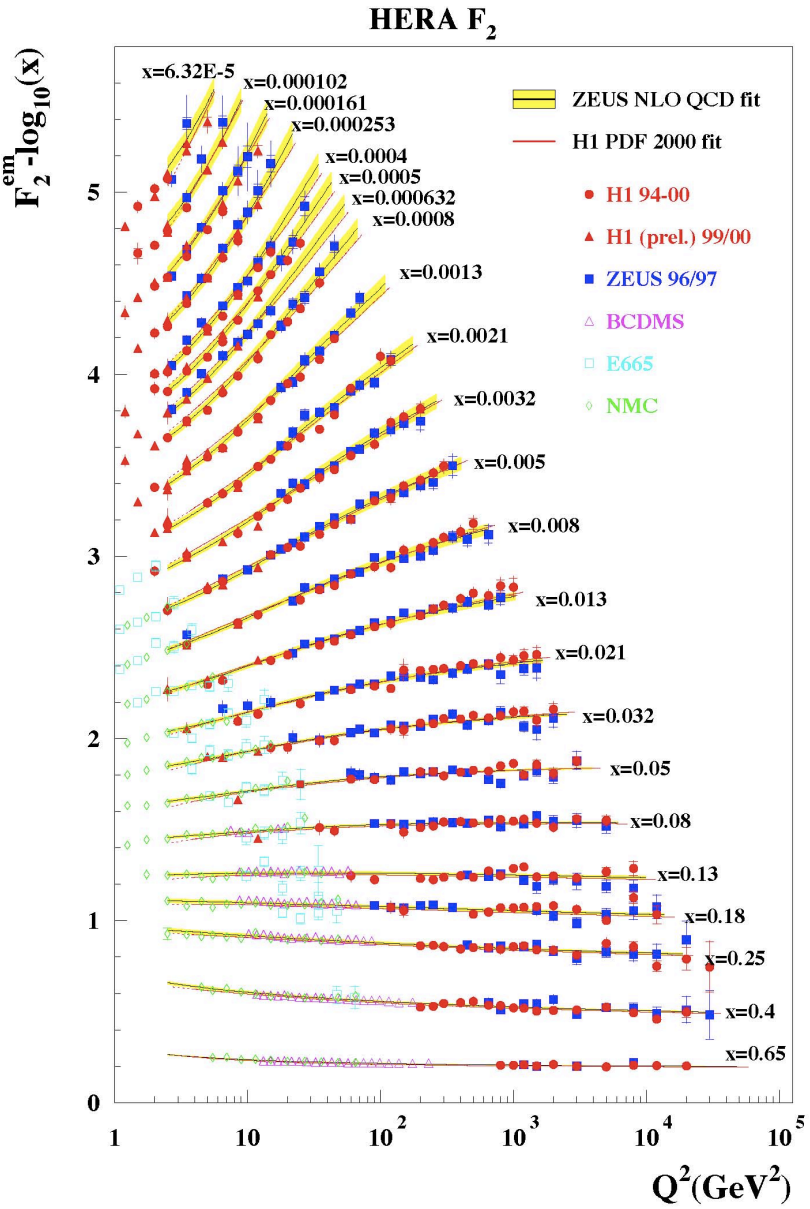


Figure 1.5: Experimental results for the proton structure function F_2 as a function of Q2 for many different x values, compared with a global fit by the ZEUS collaboration. Figure from [17]

The incoming hadrons are described by PDFs obtained by performing global fits to experimental data. There are several aspects that can lead to different results such as the choice of the experimental data set used as input (DIS, Drell-Yan, TEVATRON jets), the value of the strong coupling α_s, the treatment of the heavy quarks and their masses, fitting procedure and uncertainties treatment of the PDFs. For this reason there are various PDF sets available. At present, the standard library of PDF distributions is the LHAPDF (Les Houches Accord Parton Density Func-

tion) package [18]. The library contains the PDFs as determined by several groups. The most common used sets at the Tevatron and LHC experiments are MSTW (Martin-Stirling-Thorne-Watt) [19], CTEQ (Coordinate Theoretical-Experimental project on QCD) [20] and also the recent NNPDF (Neural Net PDF) [21].

1.3.2 Hard Scattering and Parton Showers

As previously discussed, the scattering of two hadrons can be described as the scattering between beams of collinear partons. Because of asymptotic freedom, the evaluation of the short-distance cross-section for the hard scattering of partons ($\sigma_{ij\to X}$ of equation 1.3.2) can be performed with perturbative QCD calculations using Feynman diagrams.

The perturbative approach is feasible only within a few orders in α_s and for processes with a limited number of particles in the final states, because the complexity and the number of the diagrams increases rapidly with the number of particles involved. As the transverse momenta of the shower become smaller and the running coupling constant becomes stronger a progressively larger number of diagrams are involved and the perturbative calculations become unreliable. Moreover, fixed order predictions diverge for soft gluon emission and collinear splitting. Soft gluon emissions refers to the presence of a gluon emitted with a very low energy, whereas collinear splitting refers to the cases in which a quark or a gluon splits into two partons that are almost collinear.

To overcome these problems an approximate description of the QCD dynamics, the parton shower (PS) was developed. The simulation chain begins with a simple $2 \to n$ process, where the particle multiplicity in the final state, n, is low. In the course of an iterative process initial and final states partons are permitted to radiate (or "branch") by $1 \to 2$ QCD splitting, *i.e.*, $q \to qg$, $q \to gq$, $g \to gg$ and $g \to q\bar{q}$, thereby producing the final cascade with large particle multiplicity. The probability that a branching occurs is evaluated with equations 1.3.5, 1.3.6. 1.3.7 and 1.3.8, whereas the probability that no branch occurs for a parton of flavor i is given by the Sudakov form factor:

$$\Delta_i(T, t_0) = \exp\left[-\sum_j \int_{t_0}^{T} \frac{dt}{t} \int_0^1 dz \frac{\alpha_s}{2\pi} P_{ji}(z)\right] \tag{1.3.10}$$

where t_0 is an infrared cut-off for an evolution variable t which starts at some high value T [18]. The introduction of the Sudakov form factor ensures that the total parton branching probability does not exceed one and handles the cancellation between real and virtual divergencies [23]. The parton shower describes the evolution of the final state partons until the square of the momentum transfer is higher then $Q_0^2 \sim 1\,\mathrm{GeV}^2$ (infrared cut-off). Parton showers are simulated by Monte Carlo for both final and initial-state radiation. Final-state radiation (FSR) refers to the showers produced after the hard scattering while initial-state radiation (ISR) refers to partons radiated before hard scattering.

1.3.3 Final States of Hadronic Collisions

The physics that governs when the particles in the showers have a momentum transfer-square lower then 1 GeV^2 is non perturbative, and pertains to hadronization (also called fragmentation). At this stage the partons group into high-mass color-neutral states which subsequently decay into the final-state particles. The parton to hadron transitions are modeled under the assumption of the Local Parton Hadron Duality [24], which states that the hadronization is a local process. It follows that hadronization involves only nearby partons and that the observable hadrons carry the kinematic and flavor informations of the original partons. At present, the mechanism of hadronization is described using phenological models. The main models implemented in the MC generators which used to predicted the final states hadronic collisions are the Independent Fragmentation, the String Model and the Cluster Model [23].

1.3.4 Spectator Partons

Additional interactions between the partons not involved in the primary interaction are not negligible and their contribution to the final state of the hard scattering process is known as the "underlying event" (UE). In order to incorporate their contribution, they are modeled as multiple parton-parton $2 \to 2$ QCD interactions with $\hat{p}_T \geq \hat{p}_T^{min}$ and uniformly distributed in rapidity. Below $\hat{p}_T^{min}$ the particles from the underlying events are not expected to be resolved by the detector. Phenomenologically, $\hat{p}_T^{min}$ has a typical value of 1.5-2.5 GeV/c.

1.4 Jets

The final state of the hadronic collisions are streams of hadrons collimated into a small angular region with the same direction as that of the primary parton. They are known as "jets". Jets can be reconstructed using a large variety of algorithms that combine the four-momenta of the final state particles with a set of rules in order to satisfy specific attributes [25].

From a theoretical point of view jets should be invariant under boost transformations as well as infrared-safe and collinear-safe. "Infrared-safe" means that the number of jets reconstructed in the event does not change if a soft gluon is emitted, whereas "collinear-safe" implies that jets remain unchanged if a hard particle splits into two collinear ones. From an experimental point of view jet algorithms should be detector independent, and computationally feasible in a minimum of computer time.

The algorithms can be either applied to the partons in a fixed order prediction (parton level jets) or to the stable particles, *i.e.*, particle with a lifetime of at least 10 ps in events simulated with MC generators (particle level jets), or to the calorimeter towers of the detector (detector level jets). The different type of jets are reported in figure 1.6.

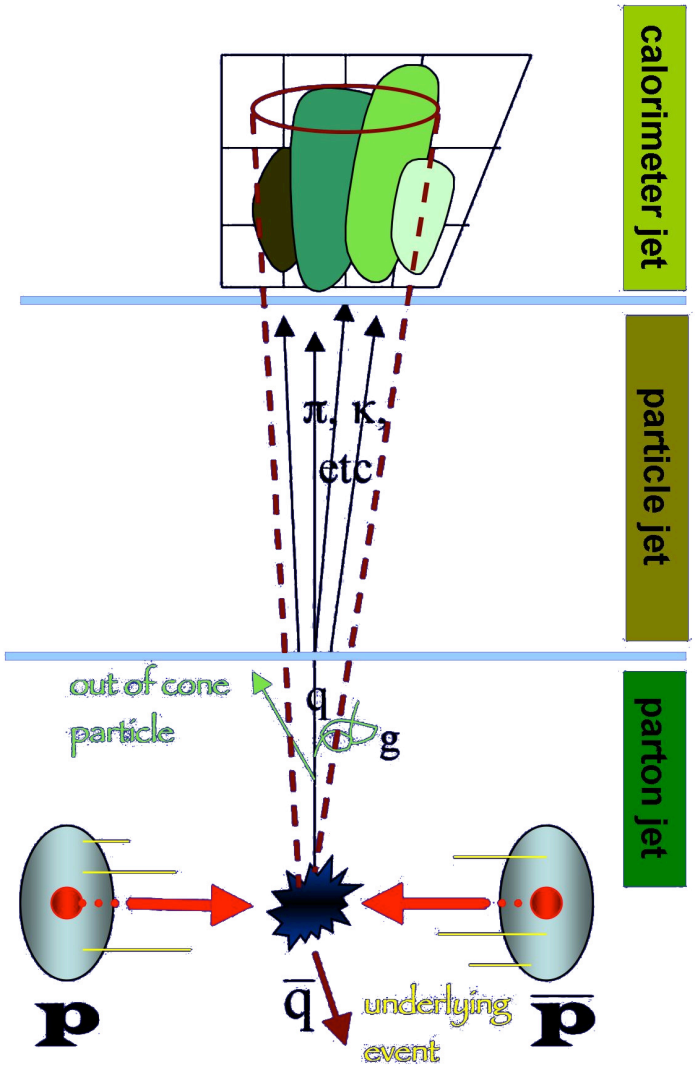


Figure 1.6: Different level jets. Figure from [50]

There are several jet algorithms in use [27], which are classified into two major categories: cone algorithms and sequential recombination algorithms.

1.4.1 Cone Algorithms

Cone algorithms are the traditional way to form the jets by association of particles with trajectories inside a cone of radius R in (η, ϕ)-space around trajectory (seed particles) which exceeds a given threshold. Most of the cone algorithms have iterative features. In these algorithms cone jets are built around an initial set of seed particles. For each of these jets, the four-momentum of the particles contained are then combined to form a new seed point. Jets are then reconstructed using these new seeds and the procedure is iterated until a stable configuration is reached.

To avoid the presence of particles that belong to two or more jets, a split-merge procedure is run after the cone clustering. Cones that share more than a fraction f of their energy are merged, otherwise they are split by assigning the particles to the nearest cone.

The use of an initial set of seeds defined with an energy threshold makes the cone algorithms computationally efficient but not collinear and infrared safe because the splitting and the soft emission can change the number of initial seeds and therefore the number of final jets. A partial solution is to introduce seeds in the "midpoints" of the stable cones and iterate the algorithm. This procedure is called the midpoint algorithm [28]. With these modification infrared safety is established for configurations of 2 nearby stable cones with soft emission in between but not for configuration with 3 nearby hard particles and a soft one. A definitive solution that assures infrared and collinear safety is represented by the Seedless-Cone algorithms which avoid the use of seeds and iterations. However these algorithms are computationally time-consuming.

1.4.2 Sequential Recombination Algorithms

Another category of jet algorithm is based on a sequential recombination (SR) of particles approach where jets are built by clustering nearby objects. The definition of the distance d between two objects i and j distinguishes different algorithms:

- **$\mathbf{k_T}$ algorithm:**

$$d_{ij} = \min(k_{T,i}^2, k_{T,j}^2)\frac{\Delta R_{ij}^2}{R^2} \tag{1.4.1}$$

- **anti-$\mathbf{k_T}$ algorithm:**

$$d_{ij} = \min(1/k_{T,i}^2, 1/k_{T,j}^2)\frac{\Delta R_{ij}^2}{R^2} \tag{1.4.2}$$

- **Cambridge-Aachen algorithm:**

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \tag{1.4.3}$$

where $\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$, and ϕ_i, y_i and $k_{T,i}$ are the azimuthal angle the rapidity and the transverse momentum of the particle i. The parameter R is an adjustable distance parameter called jet-radius. A beam distance, $d_{iB} = k_{T,i}^{2p}$ with $p = 1, -1$ and 0 for the k_T, anti-k_T and Cambridge-Aachen algorithms, respectively, is also defined. Particles i and j are recombined if $\min\{d_{ij}, d_{iB}\} = d_{ij}$ otherwise the particle itself is considered a jet and removed from the list of particles. The process is iterated as long as there are particles in the list. All of these algorithm are infrared and collinear safe.

Chapter 2

Production of $W(\to \ell\nu)+\geqslant N$ jets at Hadron Colliders

The physics program at hadron colliders has for some time focussed on the search and study of fundamental particles predicted by the Standard Model and by theories which go beyond. These particles often decay to electroweak gauge bosons, Z and W, and jets. It follows that an understanding of the production of jets in association with the W-boson is a priority. The analysis of data related to this process allows for testing of the perturbative QCD predictions.

In this chapter, after expanding on the motivations for this work in section 2.1, the theoretical understanding of the W+jets cross section is summarized in section 2.2 and other recent measurement performed at the Tevatron and LHC are reviewed in section 2.3.

2.1 Motivation for the study of the W+jets differential Cross Section

The production of jets in association with a W-boson is one of the most important Standard Model processes in high-energy $p\bar{p}$ collisions because it is expected to mimic a variety of physics processes beyond the SM. Moreover, the W+jets final state is one of the main backgrounds for important SM processes like, for example, $t\bar{t}$, single-top and Higgs boson production. Therefore, it is crucial to accurately measure this process, in order to understand its background contribution to these signals. Furthermore it makes precision tests of perturbative Quantum ChromoDynamics (QCD) possible because the production of the W-boson during the hard-scattering assures an interaction with high momentum transfer Q^2, so that the perturbative approach is justified. Perturbative predictions can be investigated in detail by comparing them with inclusive observables, such as the number of jets produced above an energy threshold as well as by studying the jet kinematics in the event.

2.2 W+ jets Theoretical Predictions

The lowest-order, also called Born-level, diagram that describes the cross-section for inclusive W production at a $p\bar{p}$ collider is illustrated in figure 2.1: the interaction of a quark from the proton with the antiquark from the antiproton with sufficient energy to produce a W-boson.

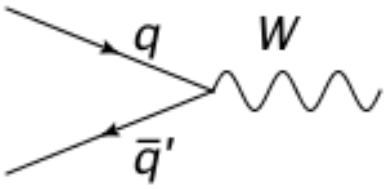


Figure 2.1: Born-Level diagram for inclusive W production at hadron-hadron colliders.

The vertex present in the diagram is described by the EWK theory. No QCD processes are involved so that the leading order cross-section for the $q\bar{q}' \to W$ process with a center-of-mass energy of the $q\bar{q}'$ interaction $\sqrt{\hat{s}}$, is:

$$\hat{\sigma}_{q\bar{q}'\to W} = \frac{\pi}{3}\sqrt{2}G_F M_W^2 |V_{qq'}|^2 \delta(\hat{s} - M_W^2) \tag{2.2.1}$$

where G_F is the the Fermi coupling constant, M_W is the mass of the W-boson and $V_{qq'}$ is the CKM matrix element for the quarks involved [15]. The formula is valid in the limit of narrow width *i.e.*, $\Gamma_W \ll M_W$, so the W-boson can be treated as a stable particle and the δ function ensures the kinematical restriction $\hat{s} = M_W^2$.

In the case of a $p\bar{p}$ collision at large center-of-mass $\sqrt{s}$, recalling equation 1.3.2,

the leading order cross-section for the $p\bar{p} \to W$ process can be calculated as:

$$\sigma_{p\bar{p}\to W} = \sum_{q,q'} \int_0^1 dx_p dx_{\bar{p}} \left[f_q(x_p) f_{\bar{q}'}(x_{\bar{p}}) + f_{\bar{q}'}(x_p) f_q(x_{\bar{p}}) \right] \hat{\sigma}_{q\bar{q}'\to W} \tag{2.2.2}$$

where the proton and anti-protons are described as parton distributions using the PDF functions f [29]. The equation accounts for both the probabilities that the quark and anti-quark involved are from the proton and the antiproton, respectively (first term inside the square brackets) or from the antiproton and the proton, respectively (second term inside the square brackets).

After production, the W-boson decays into a quark-antiquark pair or into a lepton-antineutrino (antilepton-neutrino) pair. The final cross-section can be calculated by multiplying the cross-section of formula 2.2.4 by the appropriate final state branching ratios (BR), *e.g.*:

$$\sigma_{p\bar{p}\to W(\to \ell\nu)} = \sigma_{p\bar{p}\to W} \cdot BR(W \to \ell\nu) \tag{2.2.3}$$

In this analysis only the leptonic decay of the W-boson, *i.e.* $W \to \ell\nu$ where the lepton can be either an electron or a muon are considered since their experimental signature are distinctive. The branching ratios for the $W \to e\nu$ and $W \to \mu\nu$ are both $\sim 10.7\%$.

The total W production cross section can be decomposed, using infrared-safe techniques, into its multijet components [15]:

$$\sigma_W = \sigma_{W+0\,jets} + \sigma_{W+1\,jet} + \sigma_{W+2\,jets} + \sigma_{W+3\,jets} + \cdots \tag{2.2.4}$$

The probability of producing a parton in addition to the W-boson is in the first order approximation, of the order of α_s.

2.2.1 Leading Order (LO) Predictions

The diagrams that contribute to the $W + 1$ parton cross section calculation are reported in figure 2.2. These are the leading order diagrams of the $W+$ jets cross-section since the requirement of the presence of at least one parton in the final state excludes the diagram of figure 2.1.

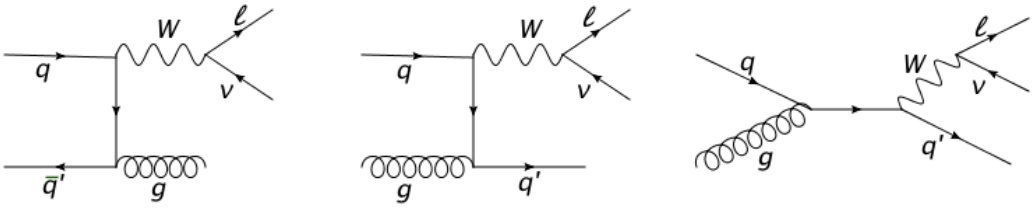


Figure 2.2: Leading order diagrams for the production of $W+$jets at hadron colliders.

The Monte Carlo generator most frequently used to calculate the LO predictions is ALPGEN [30].

ALPGEN is designed to calculate tree-level matrix elements for processes with a massive vector boson, *i.e.* W and Z, associated with the production of a fixed number (up to six) of partons. Calculation with several partons in the final states are difficult to compute even at low order since the number of Feynman diagrams involved grows factorially with the amount of particles in the final states. ALPGEN relies on the ALPHA [32] algorithm for these calculations since it is computationally convenient because its complexity is only proportional with the power of the number of final state objects. ALPGEN also includes information regarding the color flow by using a parton shower algorithm technique to model the gluon radiation. Parton shower and hadronization are, instead, not included, so it is common to interface ALPGEN with an external generator. In this analysis ALPGEN is interfaced with PYTHIA [31] for the modeling of parton showering and hadronization.

PYTHIA is one of the most used generators since it contains a great number of different hard processes. PYTHIA performs leading order matrix calculations for $2 \to 1$ and $2 \to 2$ processes. In addition, it uses parton showering and the Lund hadronization model, an application of the string model mentioned in section 1.3.3, to describe the final part of the hadronic collision. However, to some extent, it lacks color-flow and spin-correlation information.

The matching between the matrix elements generated by ALPGEN and the parton shower evolution performed with PYTHIA is done using the the MLM matching scheme [33]. The matching is performed after that the n matrix element level partons have showered. Starting from the most energetic parton, all the matrix element partons generated are matched with the particle level jets. Events with an exact parton-jet correspondence are accepted, whereas events with unmatched jets or partons are rejected. Other matching schemes are possible, for example the CKKW [34] scheme is commonly used as an alternative technique by generators other then ALPGEN.

The LO predictions generally describe the principal kinematical features of the processes and, consequently the shaped of related distributions with a good approximation. The normalization is, instead, usually badly predicted because of the lack of contribution from the higher orders diagrams to the total cross sections. In order to compensate for this problem it is common to introduce K-factors when these prediction are compared to data. The K-factor is determined by calculating the ratio of the NLO to LO cross section and usually depends on the kinematical region for which they are performed.

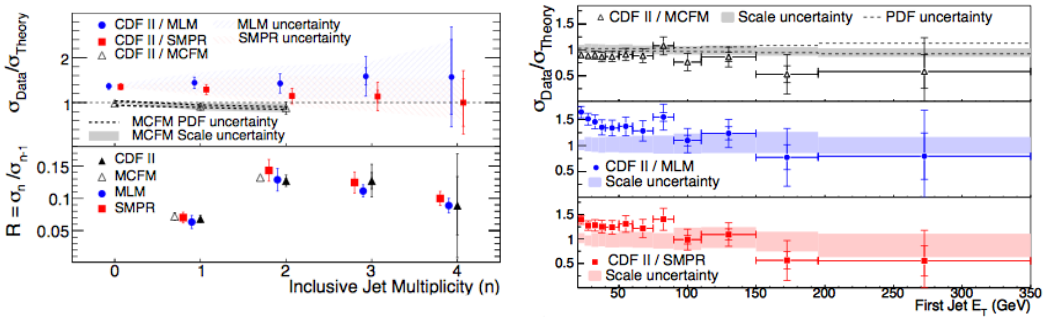
2.2.2 Higher Order Predictions

The next-to-leading order (NLO) calculations in perturbative QCD include diagrams with virtual contributions, such as particle loops and internal gluons and the presence of an additional parton. These calculations are computing time-intensive and are currently are available for processes with a maximum of 3 partons [35]. Such predictions are implemented, *e.g.*, in the MC program MCFM [36]. Although MCFM predictions are very accurate, they performed at parton-level. Hadronization and the contribution from underlying events are not included. NLO predictions which include also the modeling of the parton shower are also performed, as for example with the MC@NLO [37] package.

2.3 Previous Experimental Results

The most recently published measurements of $p\bar{p} \to (W+\text{ jets})$ production at the Tevatron were performed by both the CDF and the DØ collaborations at a center-of-mass of $\sqrt{s} = 1.96$ TeV, so where the dominant production channel for the W was $q\bar{q}' \to W$.

The analysis performed by CDF [38] is based on an integrated luminosity of $320\,\text{pb}^{-1}$ and the W-boson was reconstructed considering only its electron decay channel, *i.e.* $W \to e\nu$, by requiring the presence of an electron with transverse energy $E_T > 20$ GeV and missing transverse energy $\not\!\!E_T > 30$ GeV. Furthermore, the transverse mass of the reconstructed W-boson candidate is required to be greater than 20 GeV/c^2 and the phase space of the W decay is limited to the central part of the CDF detector (pseudorapidity $|\eta| < 1$). Residual background events are estimated using MC simulations and data-driven models and subtracted from the data sample. Events with up to 4 jets produced in association with the W are selected. Jets are defined with the JETCLU [26] cone algorithm of radius 0.4, and are required to have transverse $E_T > 20$ GeV and pseudorapidity range $|\eta| < 2$. The cross sections σ_n of each inclusive n jet multiplicity sample are measured requiring the transverse energy of the n^{th}-jet to be greater than 25 GeV. All the measured cross sections are corrected for detector effects and compared to LO and NLO particle-level theoretical predictions. The ratio between the total cross-sections measured and the theoretical predictions are shown in the top part of figure 2.3a and in the lower plot the σ_{n+1}/σ_n ratios. In addition, the measured to predicted ratio of the differential cross-section for events with $\geqslant 1$ jet as a function of the the leading jet E_T is shown in figure 2.3b. There is a close agreement with the NLO predictions for both the total and differential cross sections.

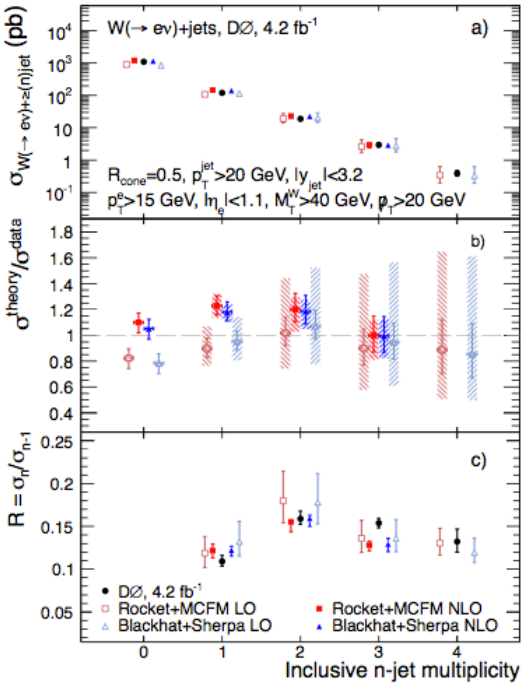


(a) (b)

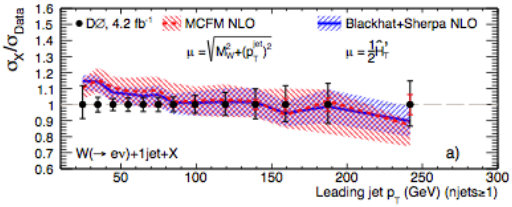
Figure 2.3: CDF results for the $W+ \geqslant n$ jets cross-section: (a) the ratio between the measured and the theoretical cross sections (top) and σ_{n+1}/σ_n ratios (bottom) as a function of jet multiplicity σ_n. (b) the ratio between the measured and the theoretical cross sections as a function of the leading jet E_T. Figure from [38].

The analysis performed by DØ [39] uses an integrated luminosity of $4.2\,\text{fb}^{-1}$. Although, also in this case, only the electron decay channel of the W-boson is considered, the requirements for the reconstruction of the W-boson differ slightly from

those used at CDF: selected events have to contain an electron with transverse momentum $p_T \geqslant 15$ GeV/c, missing transverse momentum $\not{p}_T > 20$ GeV/c and W-boson transverse mass $m_T^W \geqslant 40$ GeV/c^2. The residual background events are removed by subtracting their estimated contributions before performing the cross-section measurement. In addition, to reduce the systematics uncertainties the differential cross-sections are normalized to the inclusive $W+ \geqslant 0$ jets cross-section. Jets are reconstructed with a midpoint cone algorithm (cone radius 0.5) and are require to carry a $p_T \geqslant 20$ GeV/c. Final results refer to cross-section that are independent of the detector and they are compared to the NLO theoretical prediction. The inclusive cross section, the ratio of measured to predicted inclusive cross-section and σ_{n+1}/σ_n are shown in the top and bottom of figure 2.4a as a function of the jet multiplicity, whereas the the measured to predicted ratio of the differential cross-section for events with $\geqslant 1$ jet are shown in figure 2.4b as a function of the the leading jet E_T. In general the measured cross section agree with the NLO calculation.



(a)



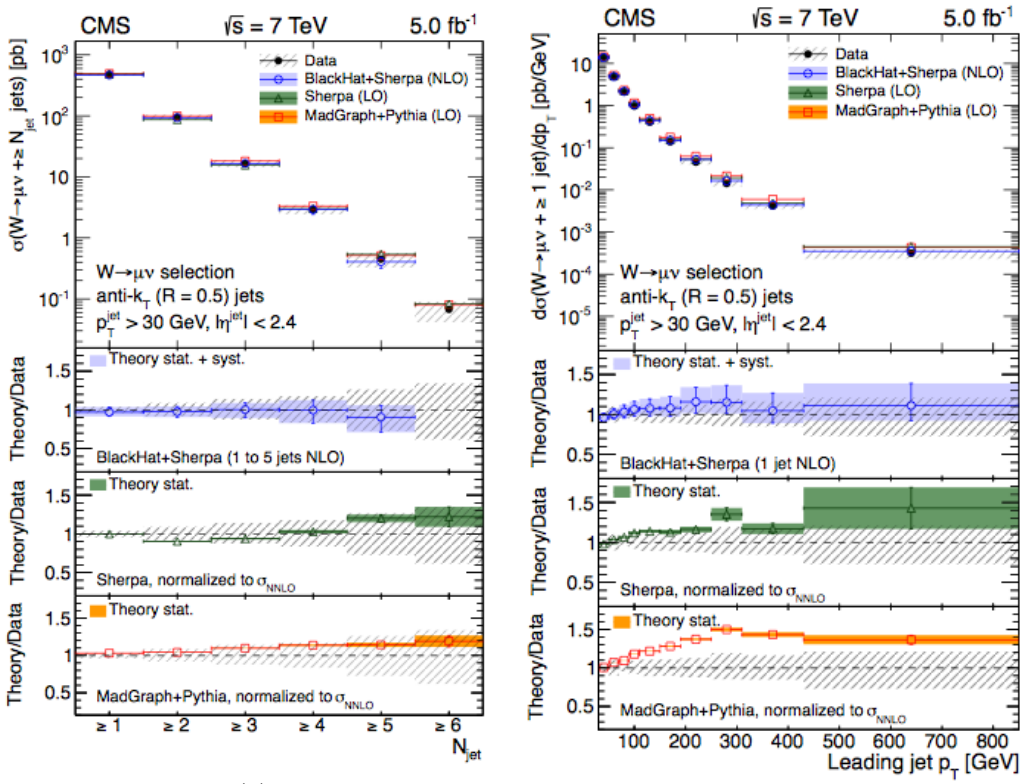
(b)

Figure 2.4: DØ results for the $W+ \geqslant n$ jets cross-section: (a) the total inclusive σ_n (top), the ratio between the measured and the theoretical cross sections (middle) and σ_n(top) and σ_{n+1}/σ_n ratios (bottom) as a function of jet multiplicity. (b) the ratio between the measured and the theoretically predicted cross sections as a function of the leading jet E_T. Figure from [39].

Measurements of W production in association with jets in proton-proton collisions with a center-of-mass energy of $\sqrt{s} = 7$ TeV using an integrated luminosity of 5.0 fb^{-1} and 4.6 fb^{-1} have been reported by the CMS [40] and ATLAS [41] collaborations, respectively. At the LHC the process $W + 1$ jet is mainly produced by

$qg \to Wq$.

The CMS production cross sections for W bosons are determined using the muon decay mode of the W boson. Selected events are required to contain exactly one muon with $p_T > 25$ GeV/c at least one jet with $p_T > 30$ GeV/c clustered using the anti-k_T algorithm with a radius parameter of 0.5 and a W-boson transverse mass above 50 GeV/c^2. The cross sections, corrected for all detector effects, as functions of the jet multiplicity and the transverse momenta of the leading jet are reported in figure 2.5a and 2.5b, respectively. The results have been compared with NLO particle-level simulated predictions from perturbative QCD. All the predictions describe the jet multiplicity within the uncertainties, but some of them overestimate the cross section as a function of the leading jet p_T, especially at high-p_T.

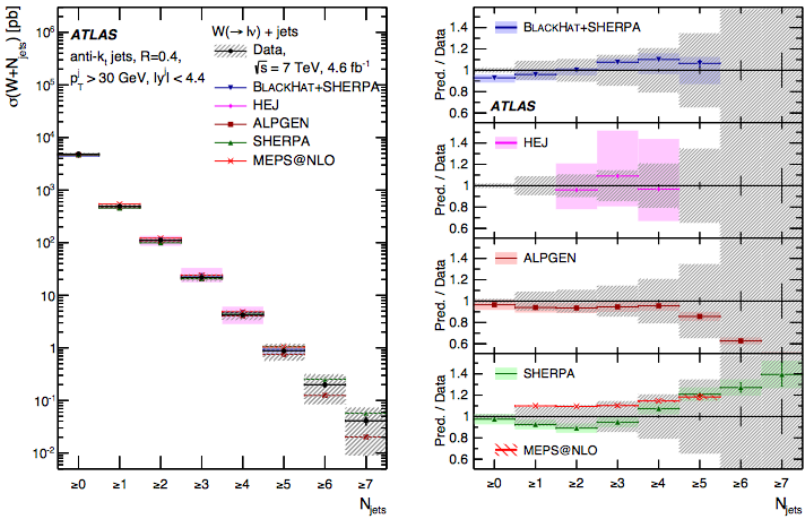


(a) (b)

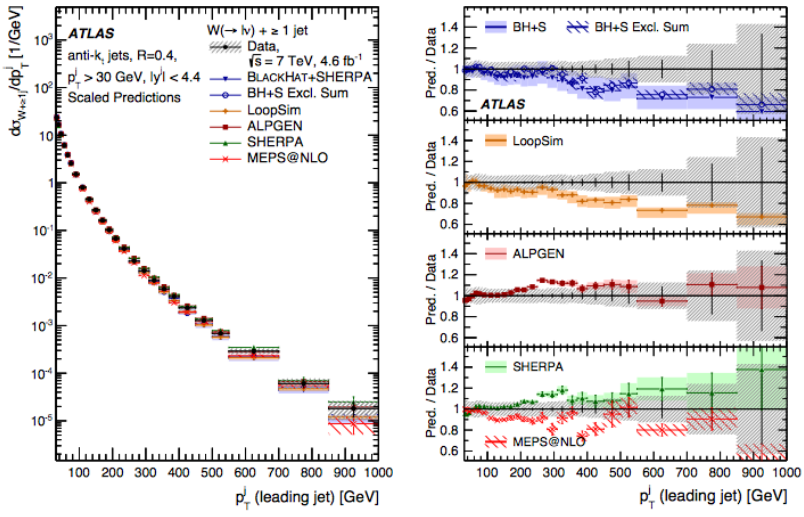
Figure 2.5: CMS results for the cross-section for (a) the production of $W+ \geqslant n$ jets as a function of the inclusive jet multiplicity; (b) the production of $W+ \geqslant 1$ jets as a function of the leading-jet p_T. In each figure, the differential cross sections are shown in the top-plot and the ratios of the predictions are shown in the bottom plot. Figure from [40].

The ATLAS measurements of the production cross sections for W bosons are performed in both the electron and muon decay channels. Then the events are selected by requiring either one electron or one muon with a transverse momentum p_T greater than 25 GeV/c, missing transverse energy $\not{E}_T > 25$ GeV and a W-

boson transverse mass $m_T^W \geqslant 40$ GeV/c^2. Jets are reconstructed using the anti-k_T algorithm with a radius parameter of 0.4 and are required to have a $p_T > 30$ GeV/c and rapidity $|y| < 4.4$. The two decay channels are combined after the correction for the detector effects is performed. Beside the inclusive cross-section for each jet multiplicity considered (up to 7 jets) shown in figure 2.6a, the differential cross sections for many observables are measured including the leading jet p_T shown in figure 2.6b.



(a)



(b)

Figure 2.6: ATLAS results for the cross-section for (a) the production of $W+\geqslant n$ jets as a function of the inclusive jet multiplicity; (b) the production of $W+\geqslant 1$ jets as a function of the leading-jet p_T. In each figure, the differential cross sections are shown in the left-hand plot and the ratios of the predictions to data in the right-hand plot. Figure from [41].

The measurements are compared with numerous QCD predictions at both LO and NLO order and the results are, with a few exceptions, in fair agreement with theory.

Chapter 3

The Tevatron and The Collider Detector at Fermilab (CDF)

The analysis presented in this thesis is based on the entire dataset collected by the Collider Detector at Fermilab (CDF) during Run II of the Tevatron high energy proton-antiproton collider located in Illinois.

This chapter provides a description of the experimental apparatus. The Tevatron accelerator complex and its performances are presented in section 3.1. The CDF detector is described in section 3.2, with particular focus on the relevant sub-detectors important for this analysis. Sections 3.3 and 3.4 contain a description of the online and offline data processing. Finally, in section 3.5 the object reconstruction methods used at CDF are discussed.

3.1 The Tevatron accelerator complex

The Tevatron is a proton-antiproton synchrotron, located at the Fermi National Accelerator Laboratory (FNAL, Fermilab [42]) that completed operations on September 30^{th}, 2011. The proton and antiproton beams, produced and accelerated with a complex chain of components involving different accelerating stages, were collided at center-of-mass energy of $\sqrt{s} = 1.96$ TeV inside the CDF and DØ particle detectors. A schematic view of the Fermilab's accelerator complex is shown in Figure 3.1. The major components of the chain preceding the Tevatron Ring are: the Proton Source complex (pre-Accelerator, LINAC, Booster), the Antiproton Source complex, the Main Injector and the Recycler Ring [43].

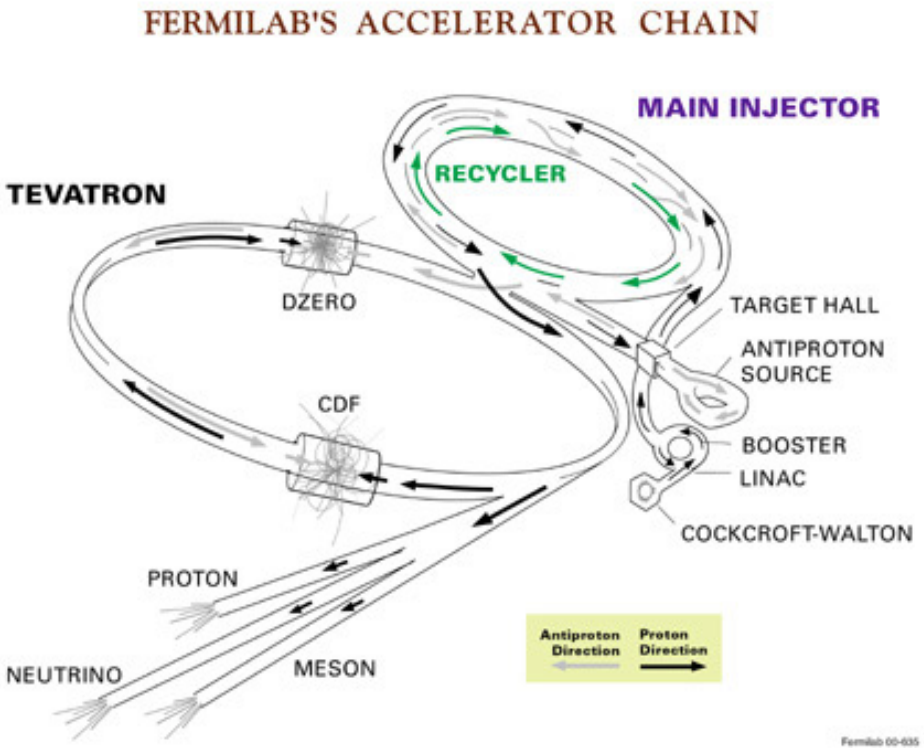


Figure 3.1: The Fermilab accelerator chain. Figure from [42].

3.1.1 The Proton Source Complex and the Main Injector

The first accelerating stage of the proton source complex is the Cockcroft-Walton pre-accelerator which ionizes hydrogen gas and accelerates the negative ions, through positive voltage, to 750 KeV. The negative ions are then fed into the 150 meter long linear accelerator (LINAC) which accelerates them to 400 MeV using oscillating electrical fields [44]. At the end of the LINAC the hydrogen ions are stripped of their electrons by passing through a carbon foil. Acceleration of the resulting protons continues inside the Booster. The Booster is a 75-meter circular synchrotron where the

protons are accelerated up to 8 GeV and compacted into bunches of approximately 9×10^{12} particles [45]. It consists of a lattice of magnets and radio-frequency accelerating cavities. The Booster is the final state of the Proton Source. The bunches of protons are passed into the Main Injector (MI) where they are accelerated to 150 GeV. The Main Injector is composed of a lattice of dipoles, quadrupoles and RF cavities and in addition to accelerating the proton beam for injection into the Tevatron, it furnish a 120 GeV-proton beam for antiproton production [46].

3.1.2 The Antiproton Source Complex and the Recycler Ring

A portion of the proton bunches (a "batch" i.e. 84 bunches) is taken from the Main Injector and used for the antiproton where they are sent to a nickel production target. About 20 antiprotons are produced at the target for every 100 million protons. Immediately downstream of the target, 8-GeV-antiprotons are selected by a pulsed dipole magnet to form a beam with the same bunch structure as that of the proton's beam used to produced it. The antiprotons are then injected into the Debuncher, a storage ring that uses stochastic cooling to reduce the momentum spread of the protons while maintaining their kinetic energy constant at 8 GeV. The antiprotons are then collected in a triangular storage ring, the Accumulator, and when each of 36 bunches has accumulated approximately 3×10^{12} antiprotons they are transferred in the Main Injector or Recycler ring for storage [43].

The Recycler is a 3.3-Km storage ring whose purpose is to build up the luminosity by injecting antiprotons, into each Tevatron store. It is located in the same tunnel of the Main Injector and it was first designed to collect the antiprotons left over in the Tevatron at the end of the collisions for reuse.

Recycling the antiprotons with a kinetic energy of 1 TeV turned out to be too inefficient because the probability to lose them is very high. Its function was therefore restricted to storing and cooling antiprotons from the Antiproton Source. The Recycler is equipped with stochastic and electron cooling whereby the antiproton beam is cooled by injecting a beam of 8 GeV electrons into it. This cooling procedure reduces its spread in momentum and space in order to maximize the probability of collisions. [47].

3.1.3 Tevatron Ring

Finally the 150-GeV proton and the antiproton beams are injected from the Main Injector into the Tevatron which accelerates them up to 980 GeV. The Tevatron is the largest circular accelerator of the Fermilab accelerator chain. Its circumference is approximately 6 kilometers and contains eight accelerating cavities to accelerate the two beams. During acceleration, the proton and antiproton beams travel in opposite direction following two non-intersecting helical orbits. Each beam is composed of 36 bunches of particles, divided into three trains of 12 bunches each, arranged as shown in figure 3.2. The gap that separates the trains is 396 ns long and allows, in case of emergency, to abort the beam into a dump by switching on the fast kicker magnets.

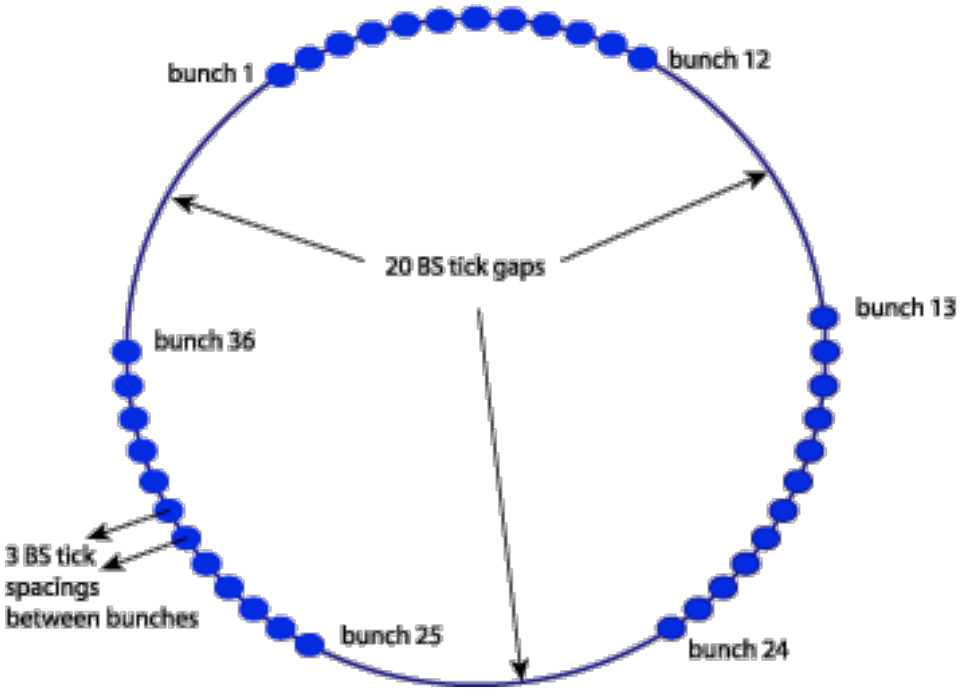


Figure 3.2: Structure of the Tevatron beams in Run II. The 36 bunches of the beams are split into 3 trains separated by 396 ns. In terms of beam sync units (BS) there are 159 BS ticks/turn. Figure from [49].

Moreover the transversal profile of the beam is narrowed to avoid possible interaction between the beams and the beam pipe (*e.g.*, by the beam halo) which can cause detector damage. When the final energy is reached and, the transverse profile of the beam is optimal, the two beams collide in two designated points of the Tevatron, at the center of the CDF and DØ detectors. As a result of the collision a shower of particles are created and then recorded by the detectors. The protons and antiprotons continue circulating and colliding inside the Tevatron until the bunches are too depleted to be useful. The time the protons and antiprotons spend circulating and colliding inside the ring is called a "store" [48].

3.1.4 Tevatron Performance

The performance of a collider can be quantified in terms of: the center-of-mass-energy ($\sqrt{s}$) and the integrated luminosity. The center-of-mass-energy (1.96 TeV for Run II) quantifies the total energy available in the proton-antiproton collision whereas the integrated luminosity correspond to the number of collisions per unit of time which is proportional to the chance of producing new particles. It is obtained by integrating the instantaneous luminosity over the period of time in which the experiment has effectively run:

$$L = \int \mathcal{L} dt \tag{3.1.1}$$

The instantaneous luminosity is a measure of the interactions between the particles inside the bunches and can be expressed as:

$$\mathcal{L} = \frac{f}{2\pi} F N_B \frac{N_p N_{\bar{p}}}{(\sigma_p^2 + \sigma_{\bar{p}}^2)} \tag{3.1.2}$$

where f is the revolution frequency, F is a form factor that depends on the bunch length, N_B is the number of bunches, N_p and $N_{\bar{p}}$ are the number of protons and antiprotons, respectively, in the bunch and σ_p ($\sigma_{\bar{p}}$) the standard deviation of the proton (antiproton) beam size at the interaction point. The instantaneous luminosity achieved by the Tevatron during Run II (June 30, 2001- September 30, 2011) ranged from about 0.1×10^{32} cm^{-2} s^{-1} up to 4×10^{32} cm^{-2} s^{-1}, as shown in Figure 3.3, where the peak luminosity as a function of time is reported.

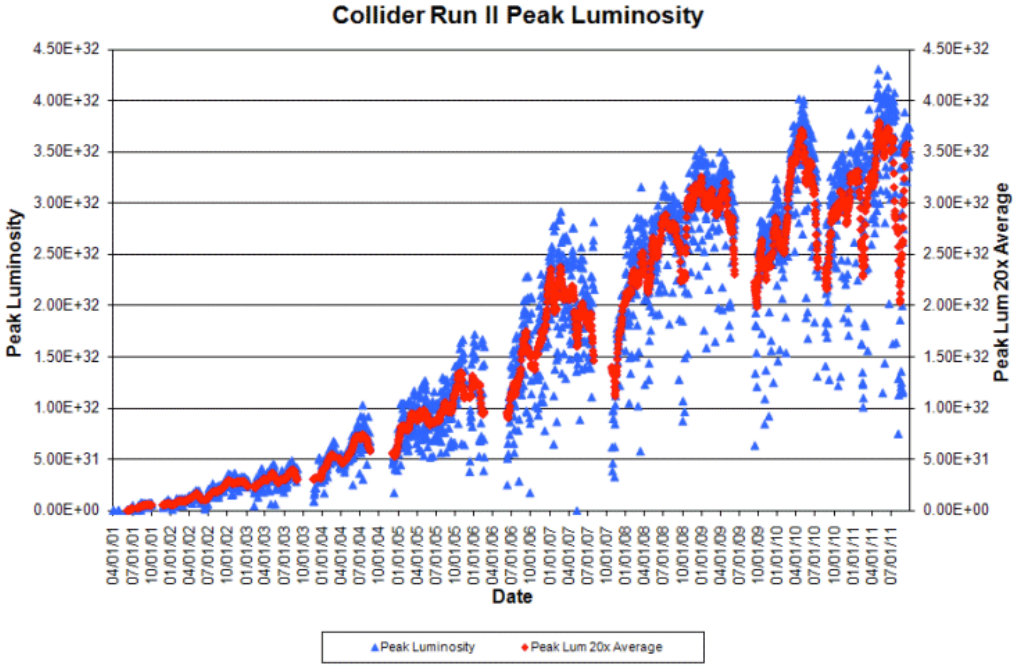


Figure 3.3: Instantaneous luminosity achieved during Run II Tevatron stores. Blue triangles represent the peak of the luminosity at the beginning of each store, whereas the red diamonds correspond to the average over 20 subsequent stores. Figure from [42].

At the beginning of the data acquisition period the instantaneous luminosity (blue triangles of figure 3.3) is maximum and decreases during the store due to particle loss and heating up of the beams. The intensities of protons and antiprotons beams is usually high enough to guarantee good luminosity for 10-15 hours, then it becomes too low and the store is dumped. The red diamonds of figure 3.3 display the average over 20 stores of the peak instantaneous luminosity. The corresponding

weekly and run integrated luminosity during Run II is shown in figure 3.4 The missing periods in figures 3.3 and 3.4 represent the times when the Tevatron was shut down for upgrading. The continuos upgrades over the time in the accelerator complex led to a significant improvement of both the instantaneous and integrated luminosity delivered by the Tevatron to the experiments. In total during Run II the Tevatron delivered approximately 12 fb^{-1} to the CDF and DØ experiments.

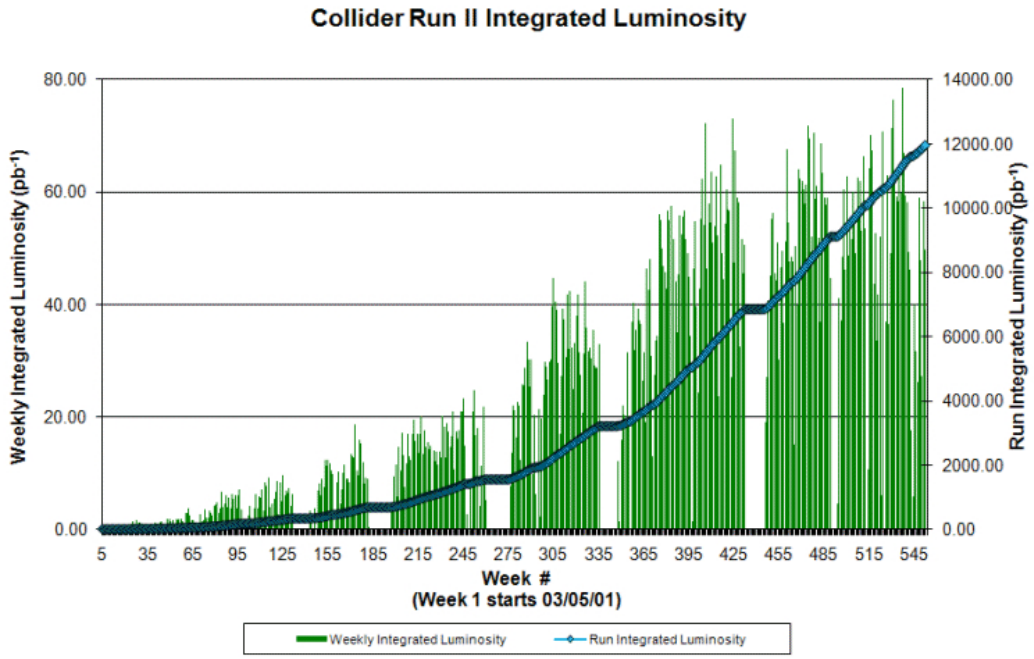


Figure 3.4: The weekly and run integrated luminosity acquired during Tevatron Run II. Figure from [42].

3.2 The CDF Detector

The Collider Detector at Fermilab (CDF) is one of the two general purpose detectors installed to study $p\bar{p}$ collisions at the Tevatron. It was commissioned in 1985 (at that time CDF stood for Collider Detector Facility) and significantly upgraded, first in 1989 and then between 1996 and 2001. The latter upgrade resulted in the detector version referred to as CDF II. This version of the detector has been functional until the end of operations in 2011. During Run II, the CDF Detector recorded about the 85% of the luminosity delivered by the Tevatron, *i.e.*, approximately 10 fb^{-1}. CDFII's physic program has included precise measurement of the top quark and W-boson masses, CP violation and B_s mixing measurements, strong interaction studies, searches for the Higgs boson and physics beyond the Standard Model.

3.2.1 Detector overview

The CDF II Detector has a cylindrical layout around the beamline, with dimensions of approximately 7.5 meter in radius and 15 meter in length. The detector is composed of several sub-detectors, as shown in figure 3.5.

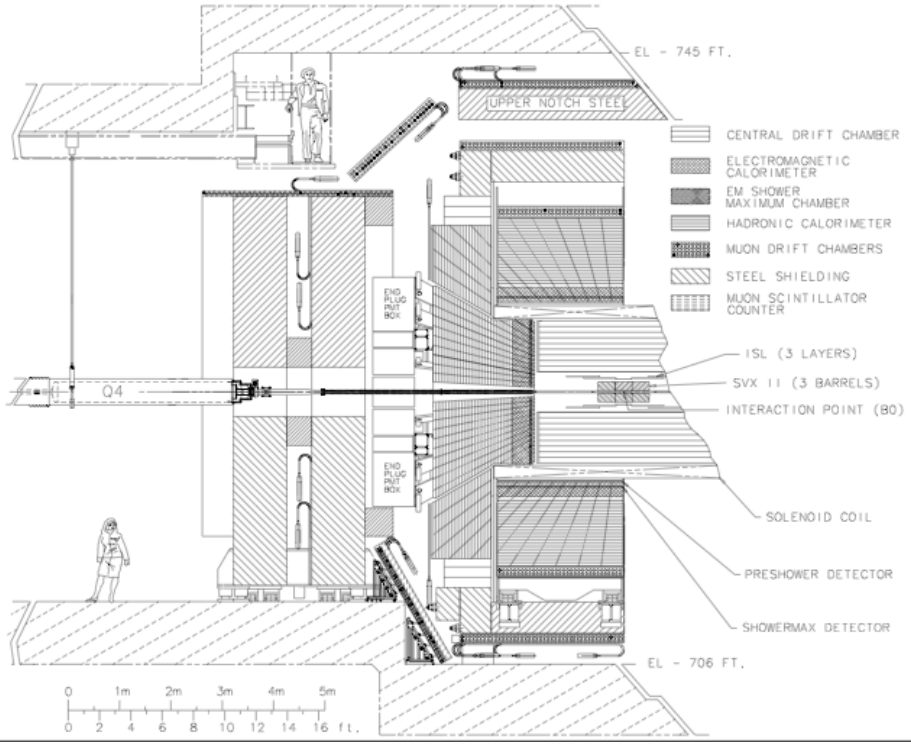


Figure 3.5: Elevation view of the CDF Run II Detector. Figure from [50].

Each of the sub-detectors is designed to perform a specific task. The tracking detectors (silicon and drift chambers) whose purpose is the reconstruction of the charged-particle trajectories and momenta, , are installed closest to the beamline, around the interaction point.

The tracking detectors are encapsulated inside a superconducting solenoid (4.8 m long with a diameter of 1.5 m), which generates a nearly uniform 1.4 T magnetic field parallel to the beam direction. The electromagnetic and hadronic sampling calorimeters, used to measure the energy of both neutral and charged particles, are located outside the solenoid. The outermost detectors are the muon detectors which consist of a series of scintillation counters used to detect the minimum ionizing particle that escape the calorimeter modules.

The sub-detectors directly related with this analysis are described in the following sections, after describing some conventions. A complete and detailed description of the entire Run II detector is given in Ref. [51].

3.2.2 Detector Coordinate System

CDF II used a Cartesian coordinate system with the origin placed on the geometrical center of the detector, assumed to be the interaction point. The z-axis is oriented along the beamline towards the east, parallel to the incoming proton direction. It follows that the (x, y) plane is perpendicular to the beams with the x-axis in the horizontal plane of the Tevatron pointing radially outward with respect to the center of the Tevatron and the y-axis pointing vertically upward. The Cartesian coordinate system is reported in figure 3.6 with a isometric view of the detector.

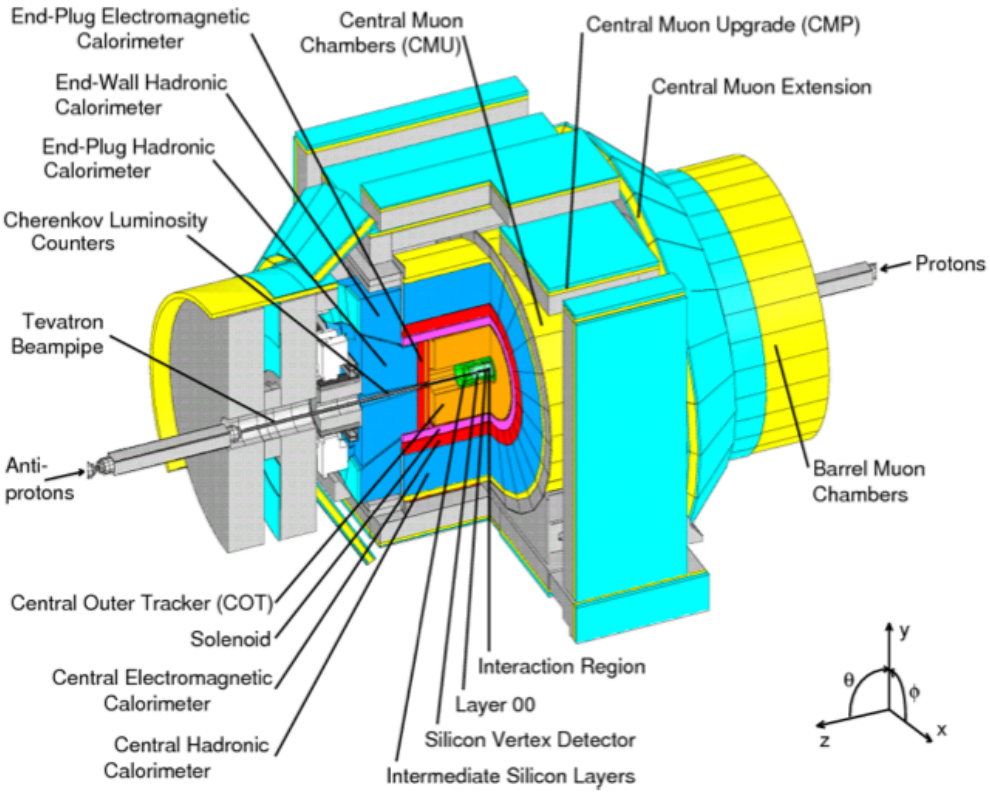


Figure 3.6: Isometric view of the CDF Run II Detector, with a quadrant cut to show the different sub-detectors and the Coordinate system specified. Figure from [52].

In fact, cylindrical coordinate system ($r = \sqrt{x^2 + y^2}$, $\phi = \tan^{-1}(x/y), z$) is more suitable to describe the detector geometry, as suggested by its shape. The z-axis of the cylindrical and cartesian coordinate systems coincide and we refer to it as the "longitudinal" direction (*i.e.*, the proton beam direction), whereas the (r, ϕ)-plane, which corresponds to the (x, y)-plane is referred to as "transverse" direction. Sometimes, it is more convenient describe the coordinates in terms of polar coordinates with azimuthal angle ϕ, measured counterclockwise from the x-axis, and the polar angle θ, measured with respect to the positive z (proton) direction.

As described in chapter 1, the proton and antiproton are composite particles and at high energy the hard interaction occurs between the partons which carry only a fraction of the total momentum. Therefore, the longitudinal component of the overall momentum in the initial state is not known and it is more convenient to describe the collisions in terms of invariant-under-the longitudinal-boost variables. Among these variables we define the transverse energy E_T and the transverse momentum P_T as:

$$E_T = E \sin\theta \tag{3.2.1}$$

$$P_T = P \sin\theta \tag{3.2.2}$$

In addition to the transverse energy and momentum, we also define a quantity called the rapidity, y of a particle with energy E and momentum P as:

$$y = \frac{1}{2} \ln\left(\frac{E + Pcos\theta}{E - Pcos\theta}\right) \tag{3.2.3}$$

Rapidity will be close to 0 when the particle directions tend to be transverse *i.e.*, perpendicular to the beam direction and $y \to \pm\infty$ when the particle is predominantly parallel to the beam axis ($\pm\infty$ if the direction is $\pm z$). For highly energetic particles the rapidity y can be conveniently substitute with the pseudo-rapidity defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{3.2.4}$$

The pseudo-rapidity η can be seen as the ultra-relativistic limit of the rapidity, and at the Tevatron, it is expected that the two variables coincide. Moreover, the interactions can occur within $\sim$ 30 cm from the nominal interaction point. It is useful to distinguish the detector pseudo-rapidity η_{det}, measured with respect the center of the detector rest frame, from the particle pseudo-rapidity, measured with respect to the center of mass frame. The distance between two points within the detector is defined with the approximately Lorentz invariant angular distance ΔR as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.2.5}$$

where $\Delta\eta$ and $\Delta\phi$ are the pseudo-rapidity and azimuthal distances between the two points.

3.2.3 The Tracking System

The tracking system, inside the superconducting solenoid, measures the trajectories and momenta of the charged particles . The uniform magnetic field provided by the solenoid has a magnitude of 1.4 T over a 1.5 meter radial distance and is oriented along the z-axis. The Lorentz force, due to the magnetic field, constrains the particle with electric charge to a helicoidal trajectory. The momentum of the particle can be obtained from the measurement of the trajectory's radius in the (x, y)-plane. The

detector's resolution for the measurement of the transverse particle momentum p_T is given by $\sigma(p_T)/p_T = 0.15\% \cdot p_T/(\text{GeV/c})$ [52].

The tracking system consisted of a multi-wire drift chamber, known as the Central Outer Tracker (COT), and of an Inner Silicon tracking System. A schematic layout of the entire tracking system is shown in figure 3.7.

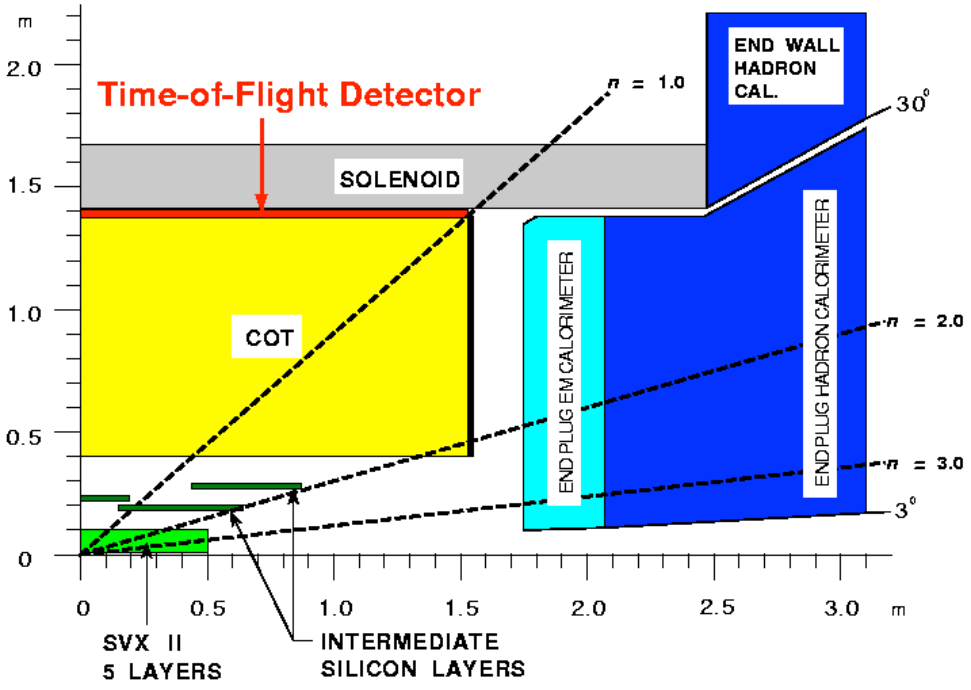


Figure 3.7: Longitudinal view of the Tracking System of the CDF II Detector. Figure from [53].

The Inner Silicon System

The inner silicon detector system is composed of three sub-detectors: the Layer 00 (L00), the Silicon Vertex Detector (SVX), and the Intermediate Silicon Layers (ISL). The principal consideration in its design was the reconstruction of the displaced secondary vertices which is crucial for the identification of events with bottom-quarks, by providing excellent spatial resolution needed to identify secondary vertices. Moreover, the inner silicon system increases the acceptance of the tracking system in the forward regions. All three detectors have a barrel geometry, for a total of eight layers of silicon sensors ($p-n$ junctions) to guarantee a good track reconstruction even in the event of component failure or degradation. Figure 3.8 show a schematic layout of the whole inner silicon system in two different projections [52].

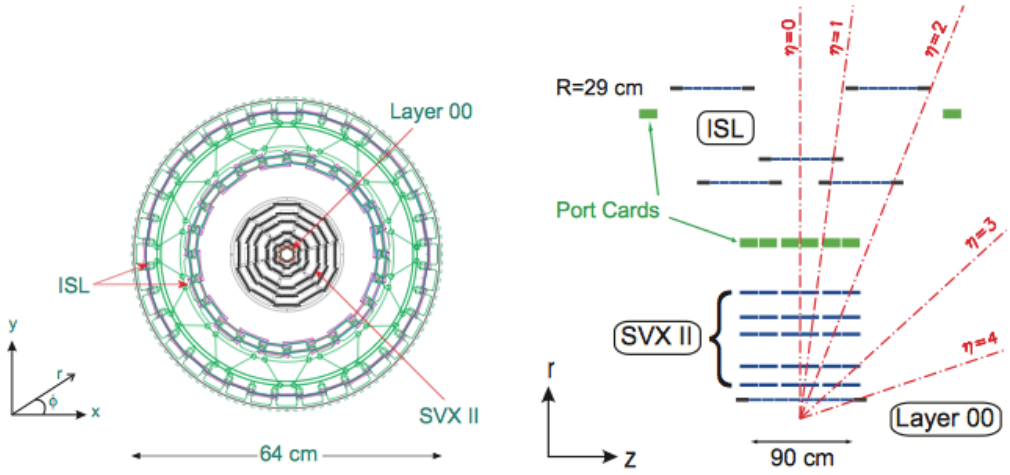


Figure 3.8: CDF II silicon system schematic in the $(x-y)$-projection (left) and in the $(r-z)$-projection (right) views. Figure from [52].

The innermost detector, placed directly on the beam pipe, is the Layer 00. It consisted of a series of silicon microstrip sensors, single-sided for improved radiation resistance with the disadvantage of providing position information in the $(r-\phi)$-plane only. The L00 detector covered a region up to $|\eta_{det}| < 4$. The Silicon Vertex Detector which has a coverage up to $|\eta_{det}| < 2.0$ with radii from 2.1 cm to 17.3 cm, was located immediately outside Layer 00. The SVX system was composed of 5 layers of double sided radiation-hardened silicon wafers able to provide information in both the $(r-\phi)$ and $(r-z)$ planes.

All the layers measured the $(r-\phi)$-component on one side, while the other side supplemented this information with a stereo measurement of 90° for layers 0,1 and 3 and of 1.2 ° for layers 2 and 4. The outermost Intermediate Silicon Layers (ISL) were located between the SVX and the COT. The ISL consisted of three layers at radial distance of 20, 22 and 28 cm from the beampipe. The middle layer is positioned in the central region of the detector ($|\eta_{det}| < 1$) while the closest and the furthest away provided precision tracking in the $1 < |\eta_{det}| < 2$ region. The three layers consist of double-sided strips with axial strips on one side and strips with a 1.2 ° stereo angle with respect to the beam axis on the other side. By combining the information from the SVX and the ISL tracks could be reconstructed in three-dimensions.

The Central Outer Tracker (COT)

The Central Outer Tracker (COT) was located around the inner silicon system. It extended from an inner radius of 44 cm to an outer radius of 132 cm and was 155 cm long. The COT is a cylindrical, open-cell, multiwired drift chamber filled with 50/50 mixture of argon-ethane gas. The chamber provided full coverage for $|\eta_{det}| \leqslant 1$ and a region of reduced acceptance coverage for $1 < |\eta_{det}| \leqslant 2$ (see figure 3.7). It was composed of 96 sense wire layers grouped in 4 axial (parallel with the beam-line) and 4 stereo (forming a $\pm 2°$ angle with respect to the beam axis) planes

for a total of 8 super-layers (see Figure 3.9). The axial superlayers measured the $(r-\phi)$-component of the trajectory, while the stereo superlayers the measured the z-component. The superlayers are further divided in supercells corresponding the ϕ-position. Supercells were 35° tilted with respect to the radial direction in order to account for the drifting of the electrons due to the magnetic field [54].

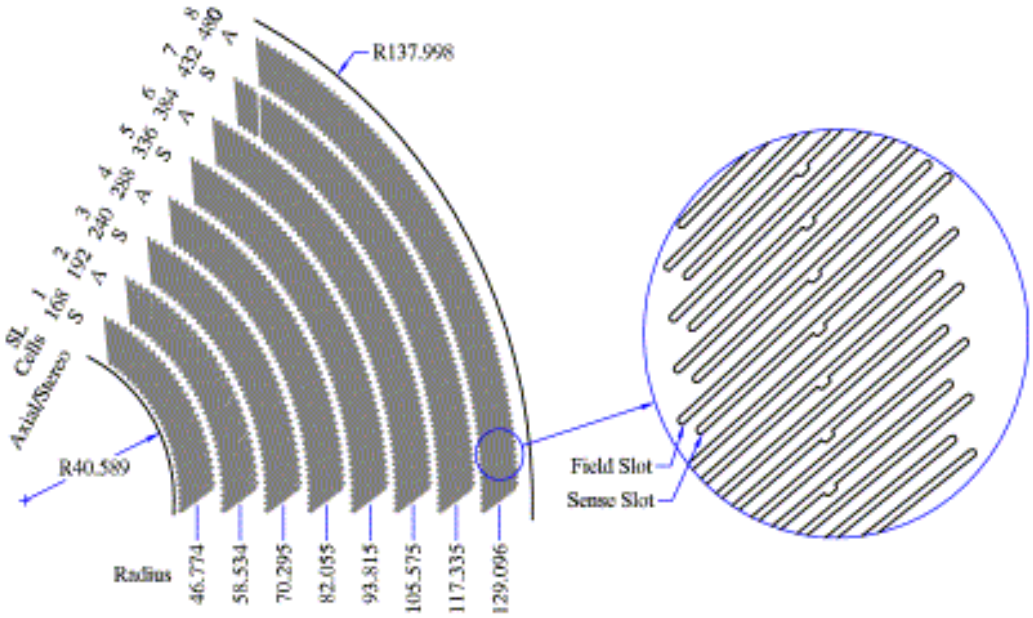


Figure 3.9: Schematic view of the eight COT superlayers (left) and a blow to show the alternation between field and sense wire plane (right). Figure from [54].

3.2.4 Time of Flight Detector

The Time of Flight (ToF) detector is located between the COT and the solenoid (at a radius of 140 cm). The TOF consist of 216 scintillator bars 3 m-long with a photomultiplier tube at both end of each bar. The main purpose of the ToF is used to enrich the information collected by the tracking system by measuring the time that the particle need to travel from the interaction point to the scintillator bars. Since the resolution of the detector is 100 ps, the time-of-flight information can distinguish between low momentum pions, kaons and protons ($K-\pi$ discrimination) [53].

3.2.5 Calorimeter Modules

The CDF calorimeter is designed to measure the energy deposited by the charged and neutral particles. The energy deposition in the calorimeters is a consequence of the electromagnetic, strong and, more rarely, weak forces between particles and the constituents of the calorimeter medium. In the case of high-energy particles, energy deposition most often results in the generation of showers. Showers initialized

by leptons and photons deposit nearly all the energy by ionization and are called electromagnetic showers, while showers initialized by hadrons are called hadronic showers and are much more complicated due to the presence of both an electromagnetic and a non-electromagnetic component. Thanks to the presence of both an inner electromagnetic and an outer hadronic part the CDF calorimeter system is able to detect both particles which interact electromagnetically and those that that interact hadronically. The electromagnetic and hadronic calorimeter parts are arranged in projective towers (also called wedges) that surround the tracking system and the solenoid to guarantee full coverage of the pseudorapidity range $|\eta_{det}$ <3.64.

In the central calorimeter, each wedge is about eight feet long, and the hadronic part is much larger than its electromagnetic counterpart because the depth necessary to fully contain the hadronic showers is much larger. A prospective view of the wedge is presented in figure 3.10.

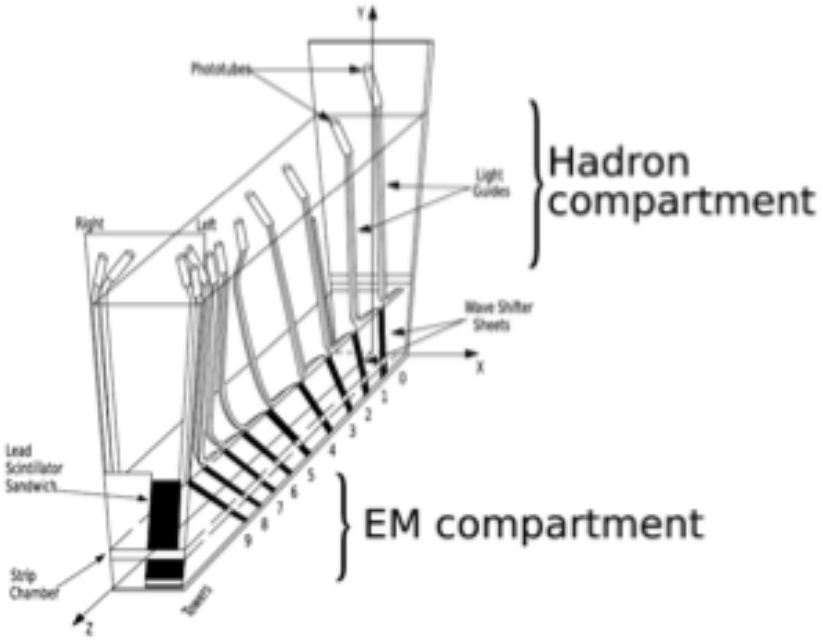


Figure 3.10: Prospective view of a calorimeter wedge. Figure from [50].

The central electromagnetic calorimeter (CEM) covers the pseudorapidity range $|\eta_{det}| < 1.1$ while the central hadronic calorimeter (CHA) covers the pseudorapidity range $|\eta_{det}| < 0.9$ but its coverage is partially supplemented by the end-wall hadronic calorimeter (WHA) in the pseudorapidity range $0.7 < |\eta_{det}| < 1.3$. The plug calorimeter extends the coverage to distances close to the beam line corresponding to a range of $1.3 < |\eta_{det}| < 3.6$ for both the electromagnetic (PEM) and hadronic (PHA) parts. The structure leaves two narrow uncovered regions, "gaps", at $|\eta_{det}| = 1.3$ and $|\eta_{det}| \sim 1.1$. The various calorimeters compartments are displayed in figure 3.11.

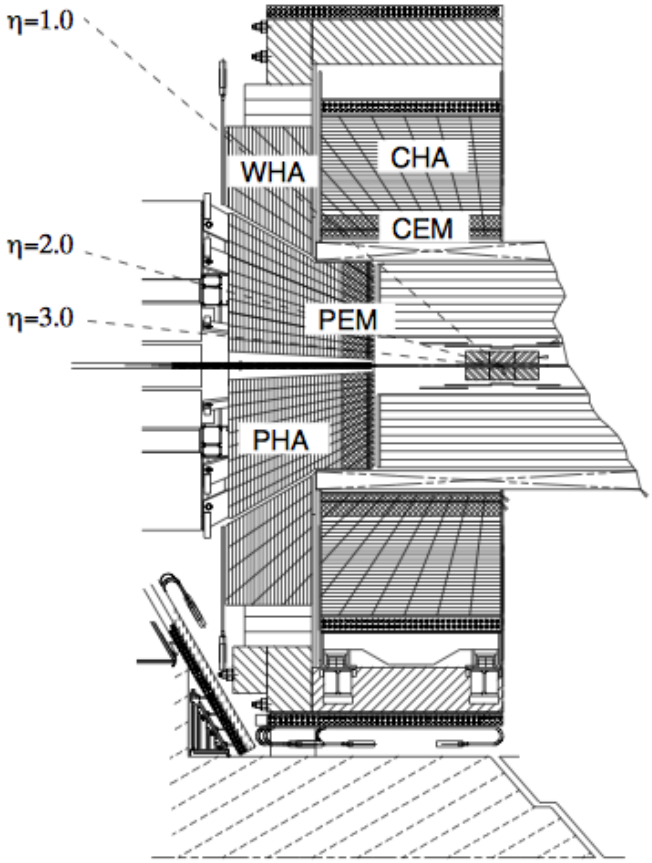


Figure 3.11: Elevation view of the calorimeter components of the CDF detector. Figure from [50].

Central Calorimeter

The central calorimeters, CEM, CHA and WHA, comprise 478 towers, each covering a range of 15° in azimuth and about 0.11 in pseudorapidity.

The CEM [59] is a sampling calorimeter made of 31 layers of 0.5 cm thick polystyrene scintillator tiles (the active media) alternated with absorber layers of 0.32 cm thick lead sheets. The total amount of material corresponds to a depth of about 19 radiation lengths (X_0). A gap (called the "chimney") allows for transfer of liquid helium to the solenoid.

The CHA and WHA are also sampling calorimeters [60]. The CHA surrounds the CEM and consists of 32 layers of 2.5 cm-thick steel alternating with 1 cm of acrylic scintillator. The WHA extends the CHA and employs the same active and passive medium of the CHA but is made of 15 layers with double thickness of lead (5 cm). The total thickness for both the hadronic parts is approximately 4.7 interaction lengths (λ_0).

Plug Calorimeters

The PEM calorimeters are also sampling calorimeters but are built using scintillator-tile technology with wavelength-shifter fiber readout. They are composed by 22 layers that consist of 4.5 mm thick lead and 4 mm thick scintillator. The tower segmentation is $\Delta\eta \times \Delta\phi \sim 0.11 \times 15^\circ$ in the outermost region, $2.11 < |\eta_{det}| < 3.64$ and is finer, $\Delta\eta \times \Delta\phi \sim 0.11 \times 7.5^\circ$ in the $|\eta_{det}| < 2.11$ region. The total thickness is about $21X_0$. The hadronic part, the PHA surrounds the PEM and has the same tower segmentation. It consists of 23 layers of 2 cm thick steel absorber alternating with 6 mm thick scintillator. resulting in a total amount of material that corresponds to $4.7\lambda_0$ [61].

Calorimeter performance

In sampling calorimeters, the high-Z passive media induce the electrons and photons to produce electromagnetic showers. The showers then interact with the scintillator tiles and generate a detectable signal by the production of photon via ionization. Photons are collected by wavelength shifter (WLS) bars or fibers and directed to the photomultiplier tubes (PMTs), that convert them in electrical pulses proportional to the energy deposited in the scintillator. There are two PMTs, located at the low and high-ϕ side of the towers in the central calorimeter and only one on the back plane of each endplug. These pulses are amplified, shaped, digitized, and recorded. In this way, the sample of energy deposited in the active media (*i.e.*, scintillator) is detected. This sample is a measure of of the total electromagnetic energy deposited in the calorimeter. The relation between the measured and total energy is determined by calibration. In the case of hadronic showers the situation is complicated by the fact that a fraction of the non-electromagnetic part is not detectable. The presence of this invisible energy contributes to making the response to the purely hadronic portion of the shower smaller than the response to the electromagnetic. This condition is called non-compensation. CDF central calorimeter is a non-compensated calorimeter.

The energy detected in each tower, E is the sum of the energy deposited in the both the electromagnetic and hadronic parts of the calorimeter: $E = E_{EM} + E_{HAD}$. The transverse energy defined in equation 3.2.1 is the projection of the energy in the (x, y)-plane, calculated by using the polar angle for the given tower. The amount of transverse energy detected by the calorimeter is also used to estimate the transverse energy of the particle that interacted weakly *e.g.*, neutrinos. In fact, the energy assigned to the weakly interacting particles is the difference between the energy at the collision and the energy detected inside the calorimeter with appropriate adjustments (missing transverse energy, $\not\!\!E_T$).

The figure of merit of a calorimeter is the precision with which the energy of the ingoing particle is measured *i.e.*, the energy resolution:

$$\frac{\sigma_{E_T}}{E_T} = \frac{a}{\sqrt{E_T}} \oplus b \tag{3.2.6}$$

where the first term is related to the sampling fluctuations and the photostatistics of PMTs, and the second term to the non-uniform response of the calorimeter. The

transverse energy of the particle E_T is measured in GeV. The energy resolution for the central calorimeters are $14\%/\sqrt{E_T} \oplus 2\%$ and $50\%\sqrt{E_T} \oplus 3\%$ ($75\%\sqrt{E_T} \oplus 4\%$) for the electromagnetic and hadronic (wall) part, respectively, whereas for the plug calorimeter $16\%/\sqrt{E_T} \oplus 1\%$ and $80\%\sqrt{E_T} \oplus 5\%$.

Table 3.1 reports the energy resolution, as well as, the pseudorapidity range of coverage by each of the calorimeter modules.

	CEM	CHA	WHA	PEM	PHA
$\lvert\eta_{det}\rvert$	<1.1	<0.9	$0.7<\lvert\eta\rvert<1.3$	$1.3<\lvert\eta\rvert<3.6$	$1.3<\lvert\eta\rvert<3.6$
$\frac{\sigma_{E_T}}{E_T}$	$\frac{14\%}{\sqrt{E_T}} \oplus 2\%$	$\frac{50\%}{\sqrt{E_T}} \oplus 3\%$	$\frac{75\%}{\sqrt{E_T}} \oplus 4\%$	$\frac{16\%}{\sqrt{E_T}} \oplus 1\%$	$\frac{80\%}{\sqrt{E_T}} \oplus 5\%$

Table 3.1: Summary of the spatial coverage ($|\eta_{det}|$ range) and performance (σ_{E_T}/E_T, energy resolution with E_T measured in GeV) of the CDF Calorimeter modules.

3.2.6 Shower Profile Detectors

The electromagnetic calorimeter includes shower profile detectors in both central and forward region approximately at $6X_0$, where the maximum of the lateral shower profile is expected.

CEM includes the Central Electron Strip Chambers (CES) and the Central Preshower (CPR, [62], [63]). The CES is a proportional chamber with strip and wire readout that measures the position and shape of the electromagnetic showers in the transverse plane with a resolution in (z, r)-plane of about (1 cm,1 mm). The CPR is a set of scintillator tiles located between the CEM and the solenoid to gather information about the early (prior to entry the calorimeter) electromagnetic shower.

As for CEM, PEM is equipped with a shower maximum detector (PES) and a preradiator detector (PPR). PES [64] is composed of two layers of 200 scintillating bars each, oriented at crossed relative angles of 45° and measures the shower position in the transverse plane with a resolution of 1 mm. PPR consists of 20 scintillating tiles 10 mm thick matching the projective tower geometry of the PEM and with a 0.83 mm-diameter-WLS fiber embedded in each of them [65].

CES and PES furnish information regarding transverse shower profile and are used to distinguish between electrons and photons, whereas CPR and PPR measurements are useful in discriminating between electrons and charged hadrons.

3.2.7 Muon Detectors

Muons, like electrons interact with matter only via the electromagnetic force like the electrons, but due to their much larger mass do not shower and they easily penetrate much deeper in material. For this reason the CDF muon detector system is located outside the calorimeter modules where they detect the non-absorbed muons. The momentum of the muons is measured using the tracking system to which the muon detectors contribute. The Muon Detector System [66] [67] consist of four detectors: the Central Muon Detector (CMU), the Central Muon Upgrade Detector

(CMU), the Central Muon Extension Detector (CMX) and the Intermediate Muon Detector Detector (IMU). The detectors are composed of drift chambers and scintillator counters and they cover different region of pseudorapidity. Figure 3.12 shows the spatial coverage of each of the detectors in the (η_{det}, ϕ)-plane.

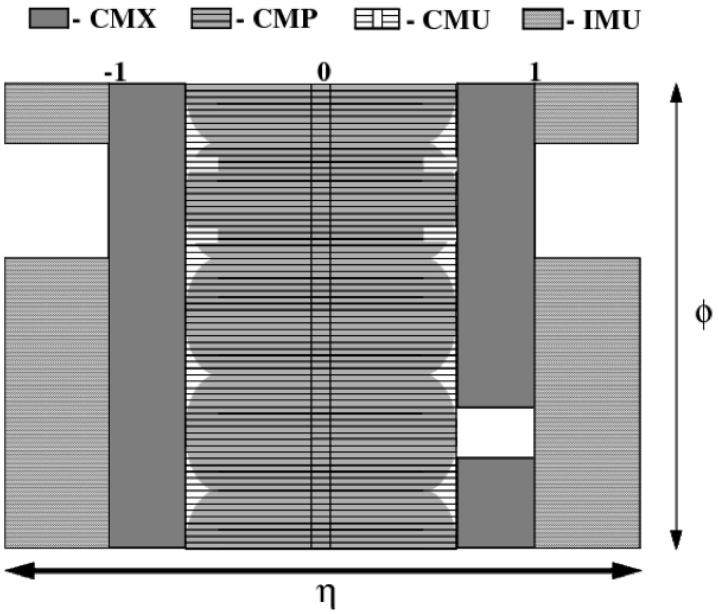


Figure 3.12: Coverage of the muon detector system in the $(\eta_{det} - \phi)$-space. Figure from [50].

The CMU detector is situated right behind the Central Hadron Calorimeter (CHA) and covers the region with pseudorapidity $|\eta_{det}| \leqslant 0.6$. It consist of a layer of drift chambers with four planes of wires and detects muons with a transverse momentum of at least 1.4 GeV/c. Behind the CMU, the magnet return yoke of the solenoid, made of 60 cm of steel, stops particle leaking from the calorimeter before the CMP and the CMX detectors. It is very likely that the only charged particles able to penetrate the shielding are muons. The CMP and CMX detector detect muons with $p_T > 2.2$ GeV/c in the pseudorapidity range $|\eta_{det}| \leqslant 0.6$ and $0.6 < |\eta_{det}| < 1.0$, respectively. The CMP detector comprises drift chambers backed by 2.5-thick scintillator counters (CSP), while the CMX detector consist of large conical sections of drift tubes and scintillation counters (CSX). Finally, the IMU detector consists of two proportional drift chambers barrels (BMU) and backed by layers of scintillation counters (BSU) behind a steel shielding. The scintillator counters yield for timing information.

3.2.8 Cherenkov Luminosity Counters

The luminosity at CDF is measured with the Cherenkov Luminosity Counters (CLC). The CLC consist of two arrays of 48 thin, long, gas-filled Cherenkov counters

placed inside the endplug calorimeters around the beampipe in the pseudorapidity region $3.7 < |\eta_{det}| < 4.7$ and grouped into three concentrical conical layers, as shown in figure 3.13.

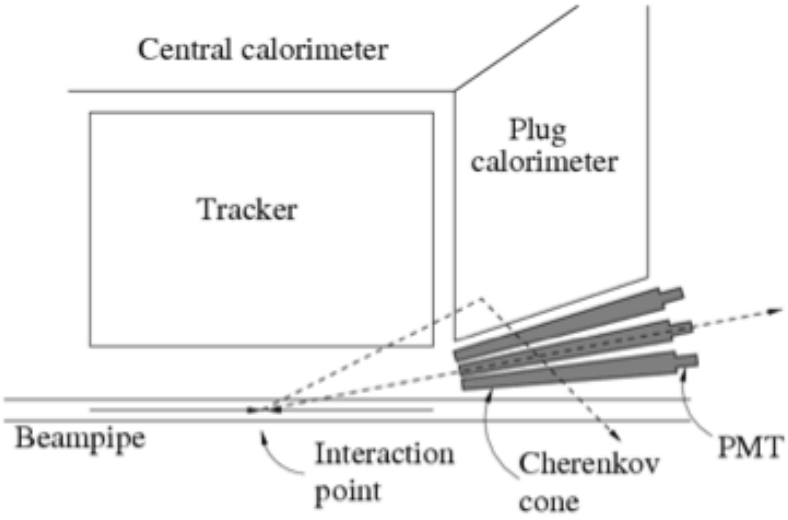


Figure 3.13: Position of Cherenkov Luminosity Counter inside the CDF Detector. Figure from [55].

The CLC measures the average interaction rate of inelastic $p\bar{p}$ collisions, $R_{p\bar{p}}$, which is related to the instantaneous luminosity $\mathcal{L}$ by the following relation:

$$\sigma_{p\bar{p}} \cdot \mathcal{L} = \frac{R_{p\bar{p}}}{\epsilon_{CLC}} \tag{3.2.7}$$

where $\sigma_{p\bar{p}} = 60.7 \pm 2.4$ mb is the inelastic $p\bar{p}$ cross section at $\sqrt{s} = 1.96$ TeV extrapolated from the measurement at $\sqrt{s} = 1.8$ TeV and $\epsilon_{CLC} = 60.2 \pm 2.6\%$ is the CLC acceptance obtained with simulations and cross-check by different method, using both data and simulations. An uncertainty of 6% is assigned to the luminosity measured with this method is assigned. Further details of the luminosity measurement are described in reference [56].

3.3 The CDF II Trigger and Data Acquisition Systems

In order to analyze the data acquired by the CDF detector, all the information from each of the subdetectors needs to be recorded and stored. Collisions at the Tevatron had a rate of approximately 2.5 MHz, which is very large in terms of the bandwidth needed to write the corresponding data in storage media. Moreover, the amount of data produced exceeded the storage capacities. These constrains led to the necessity of selecting the information to be processed, recorded and then stored.

The selection mechanism used to distinguish between interesting and not interesting events is a fast online selection system called trigger system. Only the events accepted by the trigger are saved on a mass storage and subsequently processed and analyzed offline. The CDF trigger is a 3-level system. The first two levels are performed by hardware tools, while the third is software implemented. The CDF trigger system is schematically represented in figure 3.14.

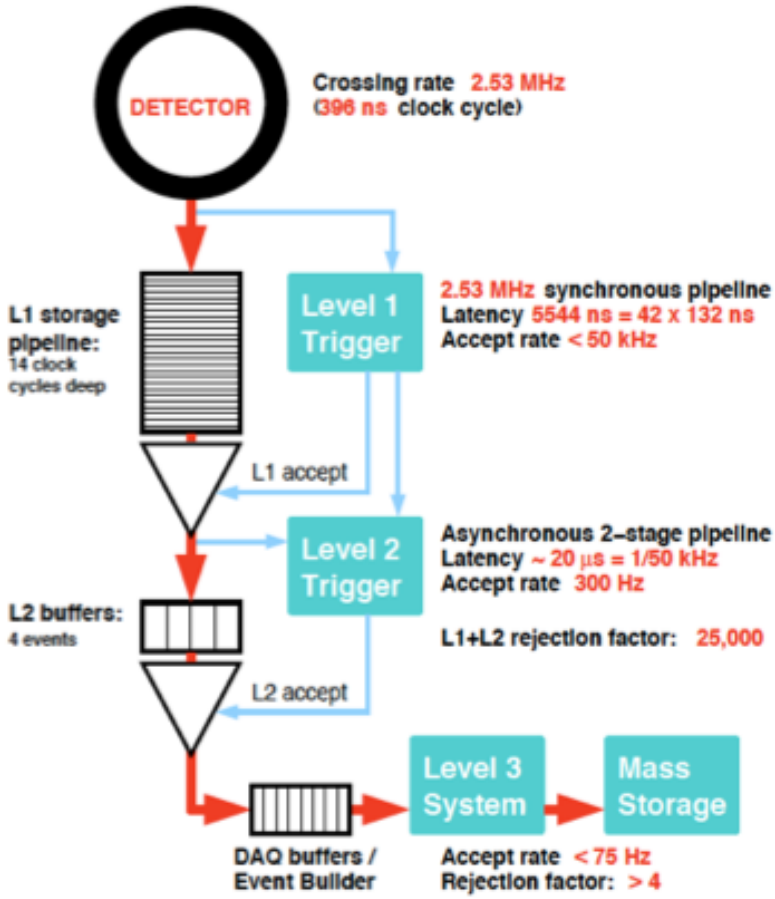


Figure 3.14: Diagram representation of the CDF II trigger system. Figure from [68].

Level 1

Level 1 (L1) consists of a three parallel synchronous systems that are able to store up to 42 subsequent events for 5.5 μs and to take a decision in about 4 μs using the information from the tracking system, the calorimeters and the muon detectors. At this level, the calorimeter objects (electrons, photons and jets) are defined by using the energy deposited inside the single towers. The total transverse energy and the missing transverse energy are obtained by summing over all the single towers and using the detector's center as the collision vertex position. Each object identified

in the L1 calorimeter is matched with a momentum track as shown in figure 3.15. Level 1 tracks are reconstructed with the eXtremely Fast Tracker (XFT), a hardwire algorithm designed to work with the output of the COT detector. Each extrapolated XFT track is then matched by the XTRP to stubs in the muon chambers and calorimeters objects by the XTRP to the L1 Muon and the L1 calorimeter systems, respectively. The Global Level 1 decision is made based on all the final reconstructed objects, the total E_T ($\not\!\!E_T$) of each event. Rejected events are lost while events which passed Level 1 are stored in one of the four buffers in the front-end readout hardware and passed to Level 2. The L1 trigger reduces the event rate to roughly 20 kHz, giving the second level about 50 μs per event to perform an asynchronous event reconstruction before making a decision.

Level 2

At level 2, the calorimeter towers are gathered into clusters by means of a simple clustering algorithm. In the meantime, the L1 XFT track reconstruction is improved by adding the information from the silicon vertex detector, so that it is possible to trigger on displaced tracks from B hadron decays at Level 2. Finally, each of the clusters are correlated with the improved tracks. As a result, approximate information about jets, electrons and photons is available to the Global Level 2 decision hardware thereby allowing for the further reduction of the event rate to about 300 Hz. Events that have passed the Level 1 and Level 2 triggers are stored in a temporary storage area on a hard disk array for Level 3 processing.

DAQ

The Data Acquisition System (DAQ) is responsible for sending the data accepted by the Level 2 trigger to the temporary storage area on the hard disk array. The data from the VME crates are transferred to the Network Switch by "scanner CPU's". The network switch distributes the informations among the processors that form the level 3 trigger.

Level 3

The final trigger decision, the Level 3 trigger, is a software trigger that performs a full event reconstruction on the events that passed the Level 2 trigger using a computer farm. The full event reconstruction uses algorithms similar to the ones employed for the offline analysis. The events delivered by the L2 trigger are addressed to the Event Builder (EVB) which combines the informations acquired by each of the detectors. The output of the EVB contains all the observable required for a Global Level 3 as to whether the event contains a physic process of interest. After filtering, the events are grouped in run numbers which are permanently written to archival mass storage to form the CDF dataset.

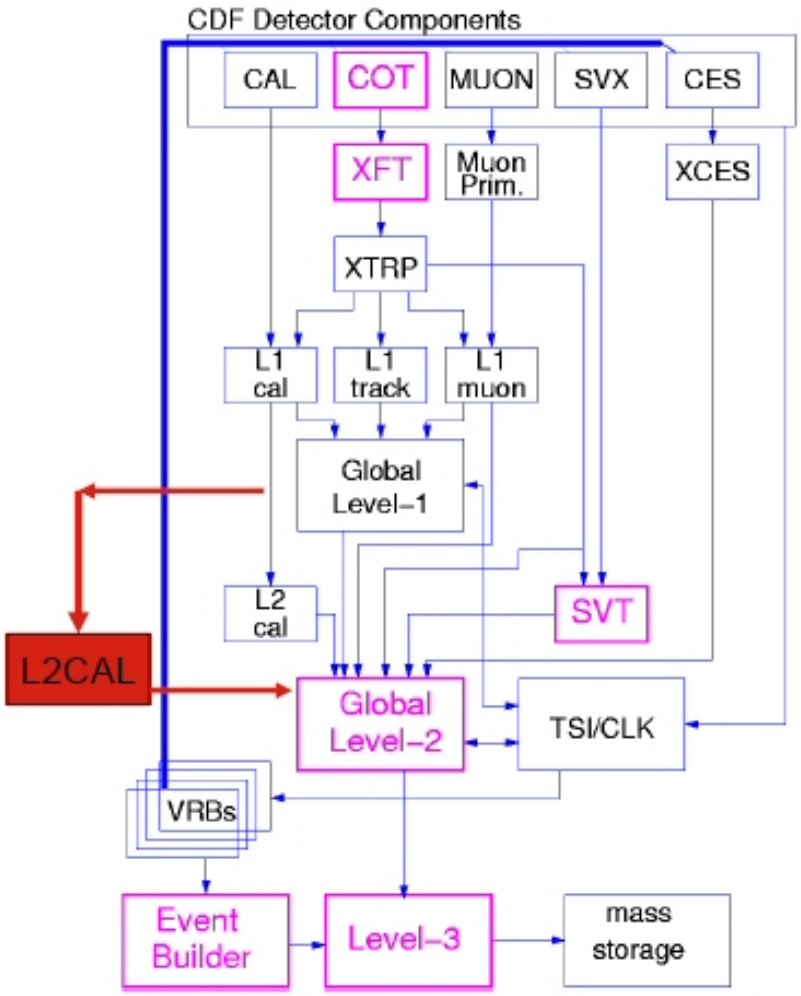


Figure 3.15: Block diagram illustrating the three-level CDF II trigger system. Figure from [57].

3.3.1 Trigger Upgrades

The increasing of the Tevatron instantaneous luminosity during Run II required two major upgrades to improve the performance and the efficiency of the CDF II trigger system. The upgrades regarded the XFT and the L2CAL system [58][69].

The XFT upgrade involved changes in both Level 1 and Level 2 and was necessary to cope with the increase in the number of tracks with the luminosity. At level 1, the upgrade consisted of using information from the COT stereo layers in order to reconstruct the tracks with higher resolution and thereby reject more fake tracks. As a consequence, the trigger rate decreases allowing the level 2 processors to fully reconstruct three-dimensionally the tracks [70].

The L2CAL upgrade system consisted in the use of more sophisticated clustering

algorithms to identify the electrons, photons and jets more accurately. In this way L2 $\not\!\!E_T$, E_T and the position of the jets are nearly equivalent to their offline measurement. The impact of this upgrade is, on the purity and the efficiency of the triggers that are related to the jets and the $\not\!\!E_T$ were, as a consequence, significant.

The block diagram of three level CDF II trigger system with the upgrades is schematize in figure 3.15.

3.3.2 Trigger Paths

Events, in order to be recoded, have to fulfill an ensemble of requirements at each trigger level. This ensemble of requirements is called trigger path. The trigger system of the CDF experiment is composed by 150 trigger paths, as described in Ref. [71]. In each path the measurements of the different detector subsystems are combined to identify and record events characterized by a specific signature. In this way, by using the data collected by different trigger paths CDF was able to study a wide range of physics process involving different objects, such as top quark and weak bosons proprieties, heavy hadrons decays and SM processes in general, or to search physics beyond the SM evidences. The trigger paths used in the analysis described in this thesis require a central high-momentum light lepton, *i.e.*, an electron or a muon, as described later in Sec. 3.5.

3.4 Offline Data Processing

Offline Data Processing refers to all the operations performed on the events selected by the L3 trigger after their real time acquisition. An object reconstruction algorithm, similar to the L3 algorithm but more refined, is used on each event stored on the tapes. The offline reconstruction process identifies physics objects such as tracks, vertices, leptons and jets more precisely than the L3 reconstruction because it employ the most up-to-date detector calibrations. Data processed online are grouped in run numbers. Offline several run numbers are grouped into run periods with integrated luminosities of 100-500 pb^{-1}. This thesis describes the analysis of the entire dataset collected by the CDF high-p_T lepton triggers (trigger paths: ELECTRON_CENTRAL_18, MUON_CMUP_18 and MUON_CMX_18) between February 4^{th} 2002 and September 30^{th} 2011, which corresponds to an integrated luminosity of 10 fb^{-1}.

3.5 Physics Objects Identification

The process of interest in the analysis described in this thesis is the production of a W-boson in association with at least one jet. As described in chapter 2, the objects involved in the final state of hard scattering are leptons, neutrinos and quarks. Photons and gluons can be produced by both the initial and the final state radiations. All these objects are reconstructed from the raw data collected by the triggers by combining the measurements of the different sub-detectors. Since the central region of the CDF II detector is well instrumented and better understood, only central leptons are considered.

In this section a summary of the methods used to reconstruct the objects involved in the process under study, based on their signatures in the CDF detector, is presented.

3.5.1 Vertex Reconstruction

First step of the object reconstruction is to locate the position of the primary interaction vertex. The interaction can occur anywhere the partons of the proton and anti-proton collide, and does not generally coincide with the geometrical center of the detector. Moreover, at high luminosity there are multiple collisions for each bunch crossing (in average 2.3 interactions per bunch crossing when $\mathcal{L} = 10^{32}$ cm^{-2} s^{-1}). The spatial position of the $p\bar{p}$ interaction is called the primary vertex and is fundamental for the correct measurement of the transverse energy, and, as a consequence, the transverse energy of the jets and, the missing transverse energy because the polar angle θ is measured taking as reference the z-axis position of the primary vertex. The position of the vertex is identified with an iterative dedicated vertex reconstruction algorithm which relies on a fit of high quality tracks to a common point of origin. At first, the interaction point coordinates (x_v, y_v, z_v) are guessed to be the average of the possible (x_v^0, y_v^0, z_v^0) of all reconstructed tracks. Then for each the χ^2 is evaluated considering all the tracks within a z window of 1 cm from the guessed position (x_v, y_v, z_v) with $p_T \geq 500$ MeV and an impact parameter significance relative to the beamline, $S_{d_0} = |d_0/\sigma_{d_0}|$[1], lower than 3. A pruning process removes all tracks with $\chi^2 > 10$, and the procedure is iterated until there are no more tracks to remove. Finally, the vertex associated with the highest transverse momentum measured summing over the survived tracks is defined as primary vertex of the event. Otherwise, if no tracks survive the primary interaction vertex is assumed to have the beamline coordinates [72].

3.5.2 Electrons

Electrons candidates are reconstructed by looking for an electromagnetic shower in the EM calorimeter and matching it with very little energy deposited in the hadronic calorimeter matched with a track in the silicon detectors and the COT. Showers in the calorimeter often deposit energy in more than one towers. Ensembles of calorimeter towers are called clusters. A cluster is defined by searching for towers with $E_T > 2$ GeV, called seed towers, and then merging all the nearby towers with at least 0.1 GeV of transverse energy with the seed tower. Since the electromagnetic showers are narrow, they usually deposit energy only in one or two calorimetric towers. So, the clusters that identify an electron object generally contains one or two towers of the electromagnetic calorimeter. The total transverse energy of the electron object is then the transverse energy deposited in the cluster, while its direction is determined by matching track with the highest p_T. The matching is performed assuming that the candidate electron travelled from the beamline to the cluster seed tower with a helicoidal trajectory typical of a charged particle inside a magnetic field.

[1] where d_0 is the impact parameter, define as the shortest distance between the track of the object and the beamline and σ_{d_0} its uncertainty which includes track and beamline position uncertainties.

Electron candidates considered in this analysis, as previously mentioned, are required to be identified in the well-instrumented part of the detector, *i.e.* the central part of the calorimeter ($|\eta_{det}| < 1.1$). Moreover, to identify a high-p_T electron candidate the EM cluster matched with the COT track has to satisfy the following selection requirements:

- $E_{had}/E_{em} < 0.055\ \text{GeV} + 0.00045 \cdot E\ (\text{GeV})$, where E_{had}/E_{em} is the ratio between the energy deposited in the hadronic (E_{had}) and electromagnetic (E_{em}) calorimeters and E is the total energy of the cluster expressed in GeV. Since electrons interact only electromagnetically with the calorimeter it is expected that their energy is almost all ($\sim 95\%$) deposited inside the electromagnetic calorimeter. The linear term accounts for the energy that leaks into the hadronic towers. This requirement distinguishes the electrons candidates from jets, since the latter deposit energy in both the electromagnetic and hadronic calorimeters.

- Isolation $\leqslant 0.1$. This requirement is also used to distinguish between electrons and jets. In fact, Isolation is defined as the fractional amount of energy detected in a cone with radius $\Delta R = 0.4$ around the electromagnetic cluster:

$$\text{Isolation} = \frac{(E_T^{cone} - E_T^{cluster})}{E_T^{cluster}} \tag{3.5.1}$$

  where E_T^{cone} is the transverse energy deposited inside the cone with radius $\Delta R = 0.4$ with the cluster as origin (the cluster energy is included) and $E_T^{cluster}$ is the transverse energy of the cluster. As mentioned above, the electromagnetic showers are narrow, so the electron should deposit its energy in one or two localized towers of the calorimeter.

- Track $|z_0| \leqslant 60$ cm where z_0 is the estimate position of the primary interaction vertex along the beamline. The z_0 coordinate has to be within the fiducial volume of the detector to ensure that the transverse energy of the calorimeter towers is calculated correctly.

- Track $p_T \geqslant 10$ GeV/c, p_T refers to the transverse momentum of the COT track associated to the electromagnetic cluster.

- COT Axial Segments $\geqslant 3$ and COT Stereo Segments $\geqslant 2$. These are track quality requirements which refer to the number of axial and stereo superlayers of the COT track associated to the electromagnetic cluster. Each superlayer should have at least 5 hits.

- $L_{shr} \leqslant 0.2$ where the lateral energy sharing L_{shr} is the difference between the amount of energy E_i^{adj} deposited in the towers adjacent to the cluster and the same quantity measured in the test beam electrons ($E_i^{exp} \pm \Delta E_i^{exp}$):

$$L_{shr} = 0.14 \sum_i \frac{(E_i^{adj} - E_i^{exp})}{\sqrt{(0.14)^2\ E + (\Delta E_i^{exp})^2}} \tag{3.5.2}$$

  where the sum runs over the adjacent towers and $0.14\sqrt{E}$ is the uncertainty on the cluster energy.

- $E/p < 2.5\ \text{GeV} + 0.015 \cdot E_T\,(\text{GeV})$ where E/p is the ratio of the cluster energy measured by the calorimeter and the momentum of the track. The ratio is expected to be one but electrons might radiate photons by bremsstrahlung when passing through the detector material. The emitted photon energy is likely to be accounted for inside the calorimeter cluster but not in the track momentum resulting in an energy-momentum ratio lower than 1.
- $-3.0\ \text{cm} \leqslant Q \cdot \Delta X \geqslant 1.5\ \text{cm}$ and $|\Delta Z| \leqslant 3.0\ \text{cm}$. ΔX and ΔZ are the separation in the $(r-\phi)$-plane between the cluster position extrapolated with the shower max and the track. The separation ΔX is multiplied by the measured charge of the particle to account for the different trajectory of electron and positrons in the magnetic field. The asymmetry in the ΔX requirements. is due to the presence of the magnetic field.
- Strip $\chi^2 < 10$ is the χ^2 resulting from the comparison of the shower maximum profile measured by the CES detector with the shower profile of the test beam electrons.

The selection requirements described above are summarized on table 3.2. We refer to the candidate electrons that met all these requirements as tight central electron (TCE).

Identification cuts	TCE
Region	central ($\lvert\eta_{det}\rvert < 1.1$)
E_{had}/E_{em}	$< 0.055 + 0.00045 \cdot E\,(\text{GeV})$
Isolation	$\leqslant 0.1$
Track $\lvert z_0\rvert$	$\leqslant 60$ cm
Track p_T	$\geqslant 10$ GeV/c
COT Axial Segments	$\geqslant 3$
COT Stereo Segments	$\geqslant 2$
L_{shr}	$\leqslant 0.2$
E/p	$< 2.5 + 0.015 \cdot E_T\,(\text{GeV})$
$Q \cdot \Delta X$	$\geqslant -3.0$ cm and $\leqslant 1.5$ cm
$\lvert\Delta Z\rvert$	$\leqslant 3.0$ cm
Strip χ^2	< 10

Table 3.2: Tight Central electron (TCE) identification cuts.

3.5.3 Muons

The muon candidates are identified by matching tracks reconstructed with COT and silicon subdetectors and a signal in the muon detectors. Muons are unlikely to deposit significant energy in the calorimeter towers due to their large radiation length (they act like minimum-ionizing particles, MIPs), so that the tracks are required to

be associated with only a small energy deposit in the calorimeter. The 4-momentum $(E, \vec{p})$ of the muon object is measured from the p_T and ϕ of the associated track assuming the invariant mass equal to the muon mass (105.7 MeV/c).

The muon candidates in use are from two categories called CMUP and CMX. In the first category muons are central with $|\eta_{det}| < 0.6$ and with stubs in both the CMU and the CMP detectors. In the second category, muons have $0.6 < |\eta_{det}| < 1.0$ and stubs in the CMX detector. The other identification requirements for CMUP and for CMX muons are the same. The energy detected in the electromagnetic (E_{em}) and hadronic (E_{had}) calorimeters must be consistent with that expected from a MIP. The track quality requirements, *i.e.* the minimum number of axial and stereo superlayers and the isolation, as well as the position of the primary interaction vertex along the beamline (z_0), used for the electron candidates are also required to be satisfied by the muon candidates. The impact parameter, d_0, of the track (defined as the the radial distance of closest approach between the track and z-axis) must be less than 0.2 cm, but for tracks with signals in the silicon tracker, this cut is harder due to better track resolution. To reject the muons that do not originate from the primary interaction vertex is required that the track fit's reduced χ^2 does not exceed 3. Finally, the radial distance at which the track leaves the COT (ρ_{COT}) must be at least 140 cm and the separation between the COT track in the $(r-\phi)$-plane and the stubs in the CMU, CMP and CMX detectors (ΔX_{CMU}, ΔX_{CMP}, ΔX_{CMX}) should be minimal. The identification cuts are summarized on table 3.3. These requirements ensure the rejection of the background muons such as cosmic rays and secondary muons originating from other decays, like decays in flight of either kaons or charged pions. In addition, for the cosmic ray rejection, a veto based on the timing information from the muon chambers and the COT is also used.

Identification cuts	CMUP	CMX
$\lvert\eta\rvert$	$\leqslant 0.65$	$\geqslant 0.65$ and $\leqslant 1.0$
E_{em}	$\leqslant 2 + \max(0,\ 0.0115 \cdot (p - 100))$ (GeV)	
E_{had}	$\leqslant 6 + \max(0,\ 0.0280 \cdot (p - 100))$ (GeV)	
Isolation	$\leqslant 0.1$	
Track $\lvert z_0\rvert$	$\leqslant 60$ cm	
Track d_0	$\leqslant 0.2$ cm (0.02 cm with Silicon Tracks)	
COT Axial Seg.	$\geqslant 3$	
COT Stereo Seg.	$\geqslant 2$	
COT χ^2/n.d.f.	$\leqslant 3$ (4 for earlier data periods)	
ρ_{COT}	> 140 cm	
ΔX_{CMU}	< 7 cm	
ΔX_{CMP}	$<\max(6, 150/P_T[\text{GeV/c}])$ cm	
ΔX_{CMX}		$<\max(6, 125/P_T[\text{GeV/c}])$ cm

Table 3.3: Identification cuts for CMUP and CMX muons.

3.5.4 Jets

Quarks and gluons produced in the hard scattering hadron collisions, as already discuss in chapter 1, hadronized due to the color confinement. Jets correspond to the cone of particles resulting from the hadronization process of a single gluon or quark, and they can be reconstructed in several ways (see section 1.4). The jets in this analysis are reconstructed from clusters of calorimetric towers using the JETCLU [26] algorithm with a cone radius in the (η, ϕ)-space of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.4$. At first, all the towers with total E_T^i exceeding 1 GeV, *i.e.*:

$$E_T^i = E_{EM}^i \sin\theta_{EM}^i + E_{HAD}^i \sin\theta_{HAD}^i > 1\ GeV \tag{3.5.3}$$

are considered as seed towers. In equation 3.5.3, E_{EM}^i and E_{HAD}^i are the energies deposited in the electromagnetic and hadronic parts of the calorimeter tower i, which are identified by the coordinates $(\theta_{EM}^i, \phi_{EM}^i)$ and $(\theta_{HAD}^i, \phi_{HAD}^i)$, respectively. Tower coordinates are evaluated considering the primary interaction vertex as origin of the reference frame and using the information from the shower max detectors. Starting with the seed tower with the highest transverse energy, the towers within $\Delta R = 0.4$ from it in the (η, ϕ)-space are gathered together to form a precluster. The center of the precluster is then adjusted to coincide with the position calculated as follows:

$$(\eta^{centroid}, \phi^{centroid}) = \left(\sum_{i=0}^{N_{tow}} \frac{E_T^i \eta^i}{E_T^{centroid}}, \sum_{i=0}^{N_{tow}} \frac{E_T^i \phi^i}{E_T^{centroid}} \right) \tag{3.5.4}$$

where the sums run over all the towers contained in the pre-cluster (N_{tow}) and each tower has coordinates (η^i, ϕ^i). The transverse energy of the centroid is the sum of the energy of the N_{tow} towers:

$$E_T^{centroid} = \sum_{i=0}^{N_{tow}} E_T^i \tag{3.5.5}$$

The procedure is iterated until the position $(\eta^{centroid}, \phi^{centroid})$ does not change. At the end of the iterative process, overlap between clusters is possible (see figure 3.16). Clusters that are fully contained inside another higher-E_T cluster, are not considered as jet candidates, while the clusters that partially overlap are merge together if they share more than 75% otherwise they are split. In the case of splitting the shared towers are assigned to the cluster with nearest centroid.

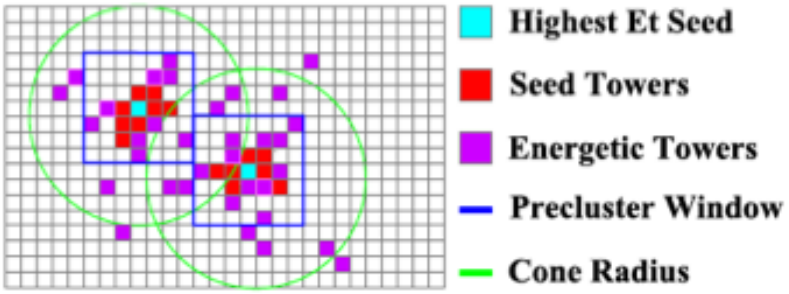


Figure 3.16: Final clusters reconstructed with the JETCLU algorithm. Figure from [73].

Each of the final clusters is considered as a raw jet. The jet Lorentz-invariant four-momentum and rapidity are calculated from the calorimeter towers contained in it according to the so-called "E-scheme" [26]:

$$E^{jet} = \sum_{i=0}^{N_{tow}} (E_i) \tag{3.5.6}$$

$$P_x^{jet} = \sum_{i=0}^{N_{tow}} E_i \sin\theta_i \cos\phi_i \tag{3.5.7}$$

$$P_y^{jet} = \sum_{i=0}^{N_{tow}} E_i \sin\theta_i \sin\phi_i \tag{3.5.8}$$

$$P_z^{jet} = \sum_{i=0}^{N_{tow}} E_i \cos\theta_i \tag{3.5.9}$$

$$y_z^{jet} = \frac{1}{2} \ln\left(\frac{E^{jet} + P_z^{jet}}{E^{jet} - P_z^{jet}}\right) \tag{3.5.10}$$

Since in the E-scheme jets are massive, pseudorapidity η is considered equal to the rapidity y, which is invariant under longitudinal Lorentz boost. From these definitions, it follows that the jet transverse energy $E_{T,raw}^{jet}$ can be expressed as:

$$E_{T,raw}^{jet} = E^{jet} \sin\theta_{jet} = E^{jet} \frac{P_T^{jet}}{\sqrt{(P_x^{jet})^2 + (P_y^{jet})^2 + (P_z^{jet})^2}} \tag{3.5.11}$$

where the transverse momentum of the jet is defined as $P_T^{jet} = \sqrt{(P_x^{jet})^2 + (P_y^{jet})^2}$. The transverse energy is labelled as "raw" transverse energy because it needs corrections for detector and physics effects, such as the presence of inactive material, the non-compensating nature of the calorimeter and multiple $p\bar{p}$ interaction per beam crossing.

Corrections to the Energy of the Jets

Jets reconstructed at the calorimeter level are then corrected to correspond to the parent parton. Corrections used by the CDF collaboration [74], often referred to as jet energy scale (JES) corrections, include all the possible corrections due to both instrumental effects and to radiation and fragmentation effects and can be parametrized as follows:

$$E_T(E^{jet}_{T,raw}, \eta) = (E^{jet}_{T,raw} \cdot L1(E^{jet}_{T,raw}, \eta) - L4) \cdot L5(E^{jet}_{T,raw}) \tag{3.5.12}$$

where $L1$, $L4$ and $L5$ are the "correction level" factors. The corrections carried by each of the correction factors are described in the following.

L1: refers to the correction of the calorimeter's non-uniform response along the η-coordinate. The correction is determined by studying p_T-balancing in dijet events. The dijet events used are required to have one jet, to which we refer as "trigger jet", in the pseudorapidity region $0.2 < |\eta_{det}| < 0.6$ far away from non-instrumented detector parts and the other jet, the "probe jet", free to span over the $|\eta_{det}| < 3$ region. In addition, jets are required to be back-to-back in the (r, ϕ)-plane. *i.e.*, $\Delta\phi(jet^{trigger}, jet^{probe}) > 2.7$ and with an average transverse momentum $P_T^{ave} = (P_T^{jet\,trigger}, P_T^{jet\,probe})/2 > 5$ GeV. A third jet can be present only if its transverse energy does not exceed 7, 8 or 10 GeV depending on the trigger path used. Other minor requirements are listened in Ref. [74]. The correction factor is then the inverse of the p_T balancing fraction β_{dijet} defined as:

$$\beta_{dijet} \equiv \frac{P_T^{jet\,probe}}{P_T^{jet\,trigger}}. \tag{3.5.13}$$

The values of β_{dijet} are evaluated for both data and Monte Carlo simulations (based on HERWING and PYTHIA generators), considering different ranges of P_T^{ave} as shown in figure 3.17. In the well-instrumented region, *i.e.* pseudorapidity range $0.2 < |\eta_{det}| < 0.6$, the correction is ~ 1, because that part of the detector is well understood, while $\beta_{dijet} \sim 1.05 - 1.1$ for $1.2 < |\eta_{det}| < 2.4$. The uncertainty associated with this correction is estimated to be of the order of 1% for central jets and 7.5% for forward jets as shown in figure 3.21. Uncertainties are mainly due to the correlation between the β_{dijet} fraction and the selection requirement applied as well as to the limitations of the parametrization in η and p_T.

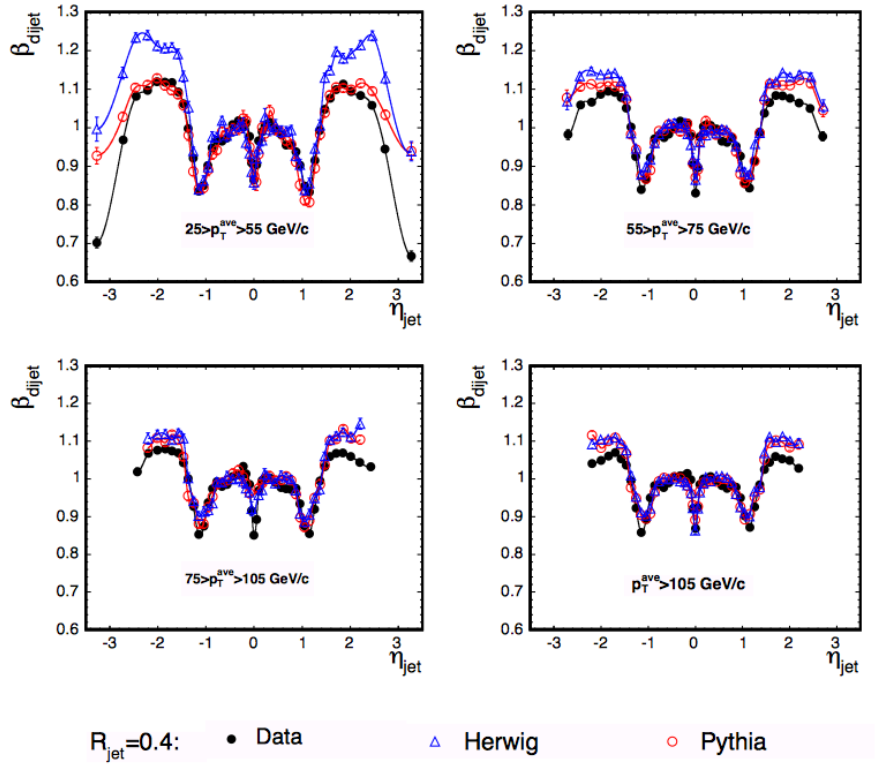


Figure 3.17: Dijet balance, β_{dijet}, as a function of the pseudorapidity $|\eta|$ in data (black circle) and in samples simulated with HERWING (blue triangles) and PYTHIA (red circles) MC generators. Four ranges of P_T^{ave} are shown: $25 < P_T^{ave} < 55$ GeV/c, $55 < P_T^{ave} < 75$ GeV/c, $75 < P_T^{ave} < 105$ GeV/c and $P_T^{ave} < 105$ GeV/c. The lines are the interpolation between the individual measurements used to ensure a continuos corrections along η. Figure from [74].

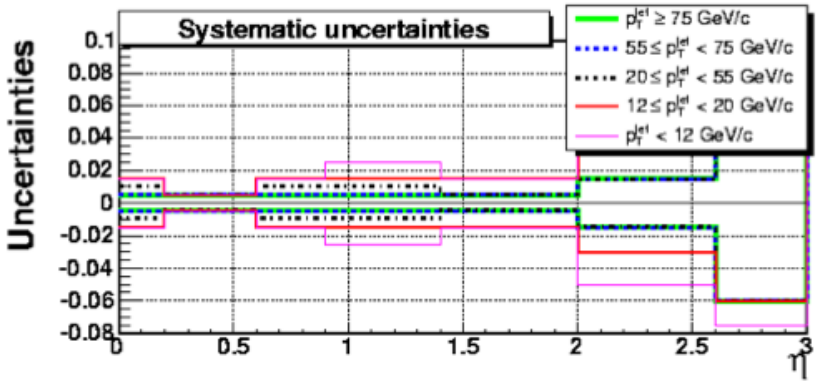


Figure 3.18: Systematic uncertainties of the L1 corrections versus $|\eta|$ and for different P_T^{jet} ranges. Figure from [73].

L4: represents the amount of energy that needs to be subtracted from the jet energy due to additional $p\bar{p}$ interactions in the same bunch crossing ("pile up"). The number of interactions follows a Poisson distribution with average that increases linearly with the instantaneous luminosity. The particles from these interactions may deposit energy in the same calorimetric towers collected inside the jet cone thereby increasing its total energy. The amount of transverse energy that needs to be subtracted is evaluated by using minimum bias events. The correction is obtained by fitting the average amount of transverse energy as a function of the number of reconstructed vertices with a linear function as shown in figure 3.19. As a result of the fit, it is possible to deduce that the amount of transverse energy that needs to be subtracted from the jet is ~ 400 MeV per vertex. The uncertainty on this correction is estimated to be of the order of 15% per each additional interaction and depends on the efficiency on the vertex finding algorithms.

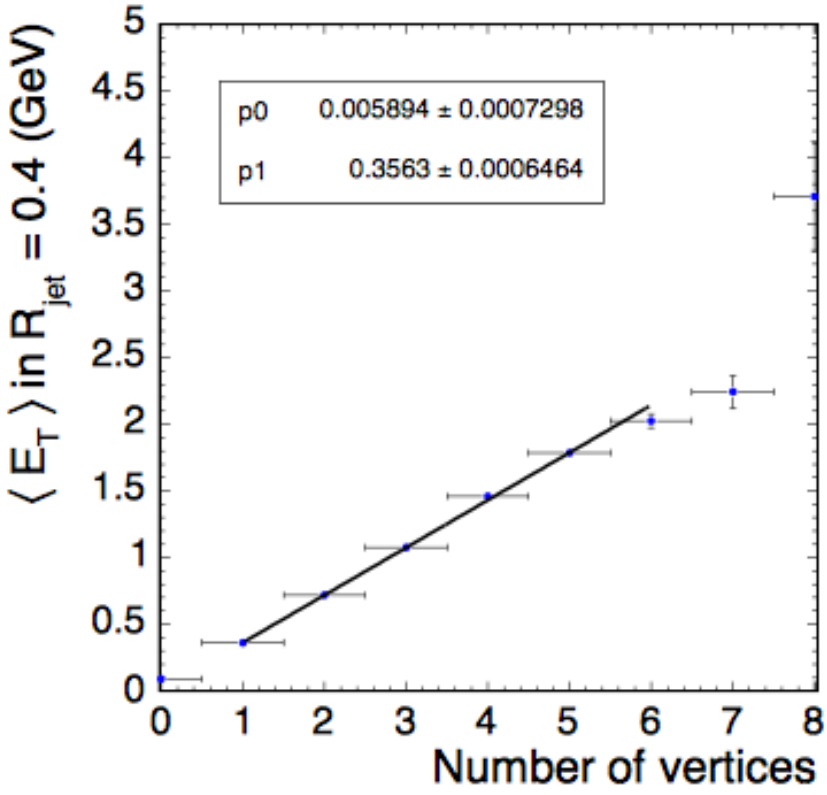


Figure 3.19: Average amount of transverse energy as a function of the number of reconstructed vertices in minimum bias events with linear parametrization superimposed. Figure from [74].

L5: corresponds to the correction for the non-linearity on the calorimeter response. This is also called absolute jet energy scale correction, and is intended to cor-

rect for detector effects. A MC sample of inclusive dijet events, simulated with PYTHIA, is used to estimate these corrections. The corrections are calculating by comparing the energy of the jets at particle level, which are the jets before the detector simulation, with the energy of the jets at calorimeter level, which are reconstructed using the detector simulation. Particle jets and calorimeter jets are matched in (η, ϕ)-space to ensure that the same object is compared. The probability of measuring a calorimeter jet with P_T^{jet}, given a particle jet with momentum $P_T^{particle}$, is parametrized using a distribution function of the difference $P_T^{jet} - P_T^{particle}$. The distribution function is parameterized as a double-Gaussian: one to describe the resolution and the other for the tails [74]. The parameters of the distributions are extracted by maximizing the logarithm of the likelihood. The maximum P_T^{jet} value of the probability function is used to extract the L5 correction. The sources of uncertainties of the correction L5 correction factors are all related to the MC model since corrections are extracted from simulations. The L5 correction factor together with its uncertainties are reported in figure 3.20.

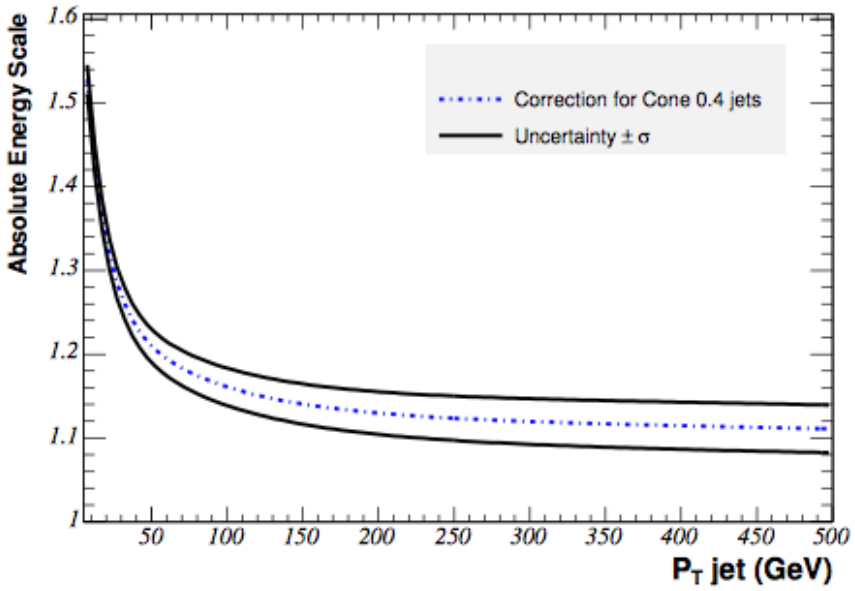


Figure 3.20: L5 correction factors as a function of jet p_T with uncertainty. Figure from [74].

After these correction the calorimeter jet is corrected to particle level, *i.e.* the energy of the jet corresponds to the energy of the physical particles resulting from hadronization process of the parton.

Two more corrections, L6 and L7 are needed to reach the parton level. These two corrections are independent of the CDF detector and correct for the energy from the particles that originate from soft interactions of partons not involved in the hard interaction (called "underlying event") and for the energy that leaks outside the jet

cone.

The total systematic uncertainty on the final correction and the relative fraction of each correction is shown on figure 3.21 as a function of the corrected p_T of the jet.

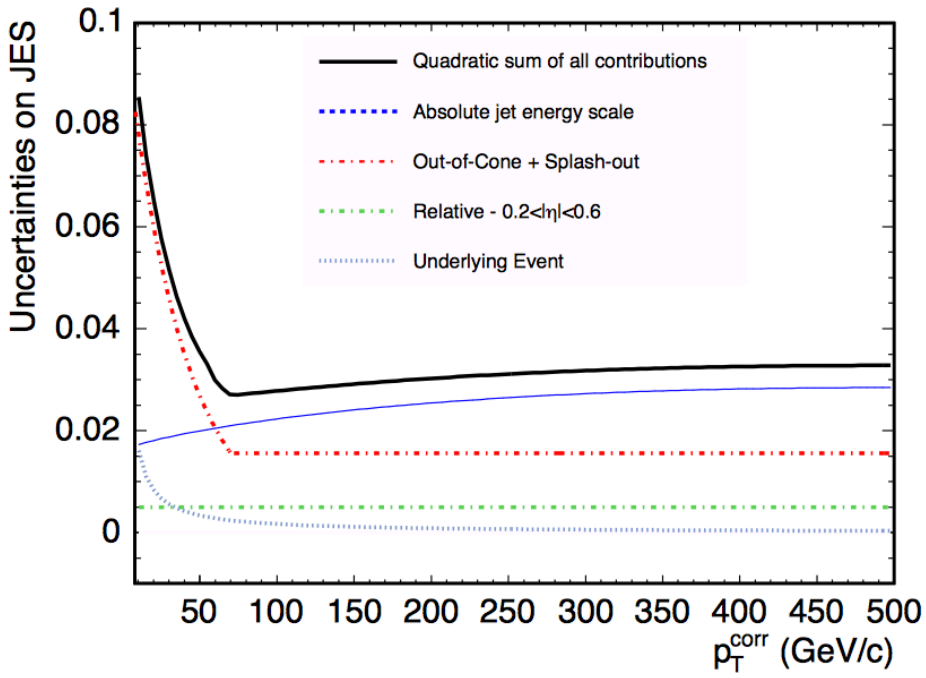


Figure 3.21: Systematic Uncertainties on the JES corrections as a function of the corrected jet p_T in the pseudorapidity range $0.2 < |\eta| < 0.6$. Figure from [74].

Quark/Gluon Jet Energy Corrections

The jet energy scale corrections previous described are the standard CDF corrections and have been validated using $\gamma/Z+$ jets samples with a dataset of 200 pb^{-1} [74] and a $t\bar{t} \to \ell\nu b\bar{b}jj$ sample [75]. However, the corrections do not explicitly account for difference in the calorimeter response to quark and gluon jets. Gluons, because of the larger effective color charge carried by them, evolve to showers with more particles than quarks, so, the gluon jets are expected to be broader than the quarks jets. Since the calorimeter response to particle showers is modeled using MC simulations which rely on different fragmentation and hadronization models, the differences in the response to gluon quark and jets lead to significative discrepancies between MC simulations and Data which are not covered by the systematic uncertainties of the JES corrections, as observed in Ref. [84]. Differences were not visible in the validation samples either because statistical limitations, as in the case of the Z+jets sample, or large quark-jets components, as in the case of γ+jets and $t\bar{t}$ samples. Subsequently, corrections for the response of quark and gluon jets were derived by using larger $Z(\to \ell^+\ell^-)+$ jets (with $\ell = e, \mu$) and $\gamma+$ jets samples. These samples

are used because they have different quark fractions balanced by against objects (the γ and the Z-boson) with energy that can accurately be measured. Sample selections and a detailed description of the procedure can be found in Ref. [73] and Ref. [76]. As a result of these studies, it was seen that the MC simulations model the energy of quark jets well but overestimate the energy of the gluon jets. It was observed that, in order to match the quark and gluon energies in data and MC, the energy of the quark jets should be increased by $\sim (1.4 \pm 2.7)\%$ while the energy of the gluons should be decreased of a $\sim (7.9 \pm 4.4)\%$. The quark jet correction for jet energies lower than 27.5 GeV are extrapolated because data at low energy are not available due to a γ-energy threshold imposed by the photon trigger during online acquisition. Both quark and gluon corrections are not dependent on the energy of the jets as can be seen in figure 3.22.

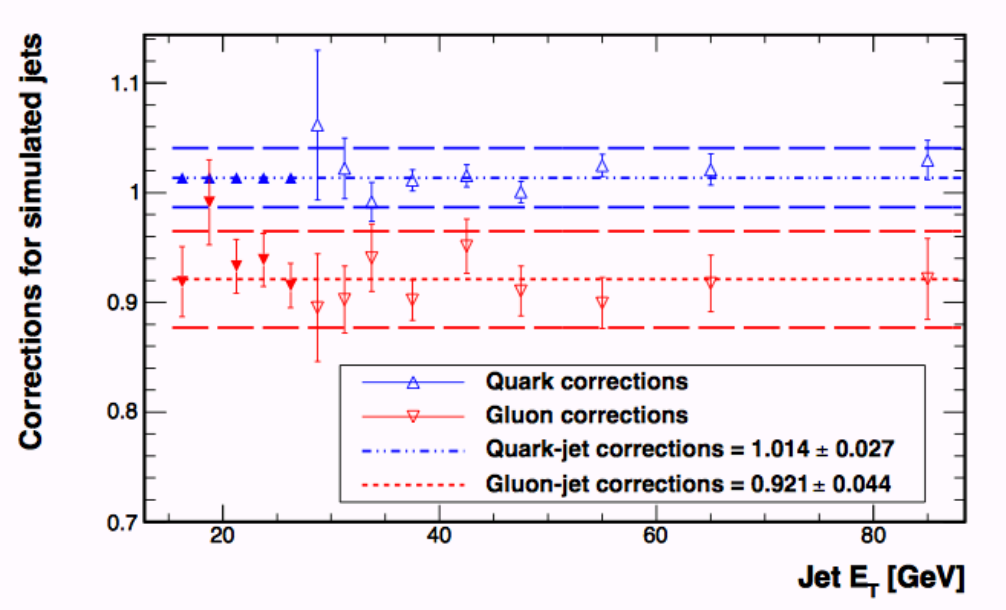


Figure 3.22: Corrections for simulated quark jets (blue) and simulated gluon jets (red) as a function of the transverse energy of the jet. The open triangles represent the corrections derived from the $\gamma/Z+$ jets samples while the filled triangles for the quarks jets (blue) are the extrapolated corrections in the energy region not available due to the photon trigger requirements. For the gluons jets (red) they are derived using only the $Z+$ jets sample. The error bars represent only the statistical uncertainties. The short dashed lines show the resulting fits (with a constant) of the corrections. The long dashed lines the total systematic uncertainty on the corrections from the fits. Figure from [84].

The systematic uncertainties are dominated by the uncertainty on the quark fractions in the $Z+$jet or $\gamma+$jet samples. The quark fractions are determined in the MC simulations by matching the jets at detector level with the partons at generator level ($\Delta R(jet, parton) < 0.4$). Since generator level information is not available

in data, simulated samples are then used to extract the fractions from the data. The uncertainty is estimated by comparing the data distribution of a quark-gluon discriminant parameter [77] with quark and gluon templates from simulation. The systematic uncertainty on the quark fraction, is then the average deviation of the extracted quark fraction from the MC simulation value. The other minor sources of systematic uncertainties include the extrapolation to low energy for the quark correction ($\gamma + photon$ trigger) and the differences in the number of reconstructed interaction vertices required in the γ+jet and $Z+$ jet samples. The total systematic uncertainties are comparable with the statistical ones. Uncertainties in the quark/gluon corrections are necessarily anti-correlated [84] and because the combination of these anti-correlated uncertainties encompasses the uncertainties in the absolute energy-scale corrections, only the former uncertainties are applied to this analysis.

3.5.5 Neutrinos

Neutrinos, as already mention interact only weakly, so they pass through the experimental apparatus undetected. However, their energy can be inferred by using the energy and momenta conservation laws. In hadronic collisions the initial energy of the partons involved in the scattering is unknown but the transverse component of total energy should be negligible respect to the longitudinal component. So, under the assumption that the total momentum is conserved and that all the other particles are detected, the negative of the transverse energy needed for a null total transverse energy is called the missing transverse energy $\not\!\!E_T$ (or MET) and is attributed to the neutrinos. The missing energy can be expressed as:

$$\vec{\not\!\!E}_T^{raw} = -\sum_{i=0}^{N_{tow}} E_{T,i}^{raw}\hat{n}_i \tag{3.5.14}$$

where $E_{T,i}^{raw}$ is the transverse energy of the tower and $\hat{n}_i$ is the unit vector pointing from the interaction vertex to the tower center which projects the onto the planes orthogonal to the beam axis. The sum runs over all the towers (N_{tow}) with $|\eta| < 3.6$ The raw MET has to be corrected for the energy of the minimum ionizing particles (mips), *e.g.* muons. The expected amount of energy left by the mips inside the calorimeter, $\vec{E}_T^{\mu,cal}$, is subtracted and the muon transverse momentum is included ($\vec{p}_T^{\mu}$) Moreover, the final $\not\!\!E_T$ includes also all the energy corrections applied to the jets:

$$\vec{\not\!\!E}_T = \vec{\not\!\!E}_T^{raw} - \vec{E}_T^{\mu,cal} + \vec{p}_T^{\mu} + \Delta\vec{E}_T^{corr,jets}. \tag{3.5.15}$$

where $\Delta\vec{E}_T^{corr,\ jets} = E_T^{jets} - E_{T,raw}^{jets}$ with the exclusions of the additional gluon/quark correction in the MC since the energy added/subtracted to the quark/gluon is already accounted for in the raw $\not\!\!E_T$ [99].

Chapter 4

Measurement of the $W+$ jets Cross-Section

The measurement of the $W+$ jets cross-sections begins with the identification of the W vector bosons. In order to identify the W boson in a multi-jet environment only the $W \to e\nu$ and $W \to \mu\nu$ decay channels are considered in this analysis. Data selection begins with the application of high-p_T electron and muon triggers to build W-enriched data sets. The triggers are described in 4.1. Since the triggers are by themselves not able to select only $W+$ jets events, additional cuts are investigated and applied offline. The physics processes expected to be present in the final sample, besides the signal, and all the selection criteria used to isolate potentially interesting events for the $W+$ jets study, are described and motivated in section 4.2 and 4.3, respectively. The estimate of the irreducible backgrounds that has to be subtracted from the data sample is discussed in detail in sections 4.4 and 4.5.

The cross section measurements are performed in section 4.6. After the subtraction of the background events, the $W(\to e\nu)+$ jets and $W(\to \mu\nu)+$ jets differential cross-sections are unfolded back to particle level in order to have results independent of the detector. This is done using the ALPGEN+PYTHIA Monte Carlo simulation together with the detector simulation. Systematic uncertainties are evaluated in section 4.7 and in section 4.8 the two channels are combined by using an algorithm that accounts for the correlations between uncertainties.

4.1 Data Sample

The whole CDF dataset, *i.e.*, all the data collected between February 2002 and September 2011, is used in the analysis. The corresponding total integrated luminosity is $\sim$10 fb^{-1}. Selection of events with a signature compatible with a W boson that decays into an electron or muon plus a neutrino, produced in association with at least one jet (Fig. 4.1) begins at trigger level where it is required that all the events pass the high-p_T central electron or muon triggers. The high-p_T lepton datasets are commonly used for the vector boson (W or Z) searches because the final state leptons produced in the leptonic decay of vector bosons carry away approximately half of the boson mass. It follows that, the presence of final high-p_T leptons is a characteristic signature of the vector bosons leptonic decay. If an event is triggered by more than one trigger, it is accepted only once, to avoid double-counting.

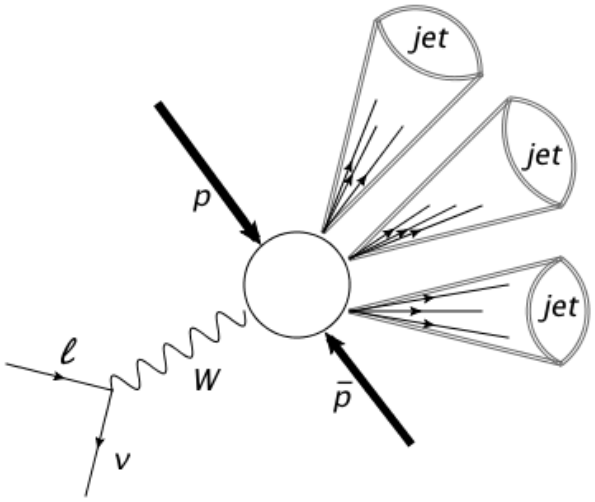


Figure 4.1: Sketch of the signature of the $W+ \geqslant 1$ jet event at the Tevatron.

In addition, it is required that the events considered for analysis be collected in course of runs when all the sub-detectors that play a role in the analysis (*i.e.*, the calorimeter, the muon detectors, and the silicon detectors) were fully operational. After the imposition of these preliminary prerequisites, the residual integrated luminosity of the data sample is $\sim$9.0 fb^{-1} for both the $W \to e\nu$ and $W \to \mu\nu$ decay channels.

4.1.1 The Electron Trigger

The ELECTRON_CENTRAL_18 is the designated trigger path used to select high-p_T electrons. The CDF Stntuple dataset corresponding to this trigger labeled **bhel**. The trigger selects high-p_T central electrons by requiring an energy cluster in the central ($|\eta| < 1.1$) electromagnetic calorimeter with transverse energy, E_T, higher than 18 GeV, matched to a fully reconstructed track. The requirements of

each of the three sequential levels comprising the trigger path are summarized in detail in table 4.1.

High-p_T Central Electron Trigger	
Level 1 (L1)	– $\geqslant 1$ "trigger tower's' (i.e. two nearby calorimeter towers) with $E_T \geqslant 8\,\text{GeV}$ and $\lvert\eta\rvert < 1.1$ – $E_{HAD}/E_{EM} \leqslant 0.125$ (for $E_T < 14\,\text{GeV}$) – a track with $p_T^{trk} \geqslant 8.34\,\text{GeV/c}$ matched to the trigger tower
Level 2 (L2)	– EM cluster with $E_T \geqslant 16\,\text{GeV}$ and $E_{HAD}/E_{EM} \leqslant 0.125$ built by gathering nearby em towers with $E_T^{EM,\,\text{seed}} \geqslant 8\,\text{GeV}$ and $E_T^{EM,\,\text{shoulders}} \geqslant 7.5\,\text{GeV}$ an XFT track with $p_T \geqslant 8\,\text{GeV/c}$ and $\lvert\eta^{trk}\rvert < 1.317$ matched to the EM cluster
Level 3 (L3)	– EM cluster with $E_T > 18\,\text{GeV}$, $E_{HAD}/E_{EM} \leqslant 0.125$ and lateral shower $L_{shr} \leqslant 0.4$ – a fully reconstructed track with $p_T^{trk} \geqslant 9\,\text{GeV/c}$ – $z_0 \leqslant 8\,\text{cm}$ from the primary vertex

Table 4.1: Detailed description of the three-level electron trigger requirements.

4.1.2 The Muon Trigger

Two muon triggers have been used to select events from the central muon sub-detectors CMUP and CMX. The trigger paths are MUON_CMUP_18 and MUON_CMX_18 respectively, and the corresponding StnNtuple dataset is designated **bhmu**. Both the muon triggers require a stub (a track segment in the muon chambers) matched to a COT track with $p_T > 18\,\text{GeV/c}$. Requirements of the three-level triggers are summarized in table 4.2.

	CMUP Trigger	CMX Trigger
Level 1 (L1)	– CMU-stub with $p_T > 6\,\text{GeV/c}$ matched to a CMP-stub – an XFT track with $p_T^{trk} \geqslant 4.1\,\text{GeV/c}$ matched to CMU-stub	– CMX-stub with $p_T > 6\,\text{GeV/c}$ – an XFT track with $p_T^{trk} \geqslant 8.3\,\text{GeV/c}$ matched to CMX-stub
Level 2 (L2)	– an XFT track with $p_T^{trk} \geqslant 14.77\,\text{GeV/c}$ 3D-matched to CMU-stub	– an XFT track with $p_T^{trk} \geqslant 14.77\,\text{GeV/c}$ 3D-matched to CMX-stub
Level 3 (L3)	– a CMUP muon with $p_T \geqslant 18\,\text{GeV/c}$ $\lvert\Delta x_{CMU}\rvert \leqslant 10\,\text{cm}$ $\lvert\Delta x_{CMP}\rvert \leqslant 20\,\text{cm}$	– a CMX muon with $p_T \geqslant 18\,\text{GeV/c}$ $\lvert\Delta x_{CMX}\rvert \leqslant 10\,\text{cm}$

Table 4.2: Detailed description of the two three-level muon trigger requirements.

4.1.3 Efficiency of the Triggers

The trigger efficiency $\epsilon^{trigger}$ is defined as the probability for a reconstructed object to satisfy all the trigger requirements, that, in the context of this analysis, is the ratio of the number of triggered electrons or muons to the number of reconstructed TCE or CMUP/CMX. In this analysis both the electron and the muon trigger efficiencies are evaluated using data-driven estimates.

The central electron trigger efficiency is determined by the tracker and the calorimeter efficiency, evaluated for each trigger level (L1, L2 and L3). The total tracking efficiency, $\epsilon^{tracking} = 0.957 \pm 0.003$, is measured by applying the same calorimeter trigger without the tracking requirements. Studies of the trigger [81] showed that the tracking efficiency is independent from the crucial observables of this analysis, with the exception of a slight dependence (of the L1- tracking efficiency) on the pseudo-rapidity, η, of the electron. This dependence and the detector deterioration due to aging are accounted for in the systematic uncertainties. The calorimeter efficiency depends on the transverse energy of the electron but for electrons with $E_T \geqslant 25$ GeV is 100%. Consequently, since in this analysis only electrons with transverse energy higher than 25 GeV will be considered the overall central electron trigger efficiency coincides with the total tracking efficiency: $\epsilon^{TCE} = 0.957 \pm 0.003$ [84].

The trigger efficiency for the muon sample is measured using the $Z \to \mu\mu$ sample. The number of selected events with the invariant mass of the two identified muons with opposite charge $M(\mu\mu) \in [76, 106]\,\mathrm{GeV/c^2}$ is the denominator of the trigger efficiency. The numerator is the number of the events with one muon tagged as CMUP muon and the other tagged as CMX or conversely. The measured efficiencies are respectively $\epsilon^{CMUP} = 0.864 \pm 0.004$ and $\epsilon^{CMX} = 0.897 \pm 0.003$ [84].

Table 4.3 reports a summary of the total efficiencies for each trigger in used in this analysis. The systematic uncertainty related to the trigger inefficiencies are discussed in section 4.7.

	TCE	CMUP	CMX
efficiency ϵ	0.957 ± 0.003	0.864 ± 0.004	0.897 ± 0.003

Table 4.3: Summary of the trigger efficiency for each trigger in used in this analysis.

Additional informations about the electron and muon triggers used can be found in [81] and [82], respectively.

4.2 Sample Composition

The data sample selected by the the high-p_T lepton triggers does not include only signal. Other processes with the same final objects, *i.e.*, leptons and jets, and processes that, due to experimental inefficiencies have signal-like signatures are included and contribute to the background.

Backgrounds where there is a real high-p_T lepton in the final state are called electroweak-like (EWK-like) backgrounds. These backgrounds have either the same

final state physics objects (*e.g.*, electron or muon plus neutrino, and jets) or, at least, a real high-p_T lepton. The EWK-like backgrounds that we expect to contaminate the data sample of this analysis are:

- **Z($\to \ell\ell$)+ jets production:** is similar to the $W+$jets in that it involves production of a vector boson plus jets. The $Z+$jets events that contribute most to the background are the one's in which the Z decays in two leptons and one of the leptons is not reconstructed, thereby generating a large missing transverse energy ($\not\!\!E_T$). Figure 4.2 shows the leading order Feynman diagrams for $Z+$jet production.

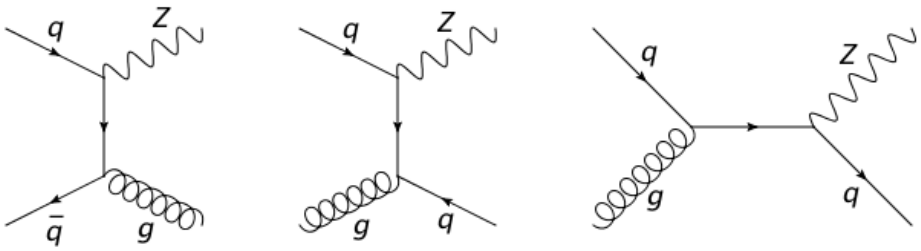


Figure 4.2: Leading order Feynman diagrams for $Z+$jet production.

- **t$\bar{\text{t}}$ production:** the typical decay diagram of this process ($\sim 85\%$ of the cases) is shown in Figure 4.3. Most of the times at least one of the W bosons decays hadronically. If the other W boson decays leptonically ($BR = 44.4\%$), the final state is a lepton-neutrino pair and four jets. This final state is indistinguishable from the signature of the $W+$higher jet multiplicity production. For this reason $t\bar{t}$ is the main source of background at high jet multiplicity. If the other W boson also decays hadronically (all-hadronic channel with $BR = 44.4\%$) the final states results in six jets, and its contribution to the data sample is negligible. When both of the W bosons decay leptonically (di-leptonic mode with $BR = 11.1\%$), it is possible that this background contributes in the $W+$jets sample, especially for lower jet multiplicity. However the contribution is very small, since the expected rate (cross-section) of $W+$jets events is much higher than the $t\bar{t}$ events for lower jet multiplicities.

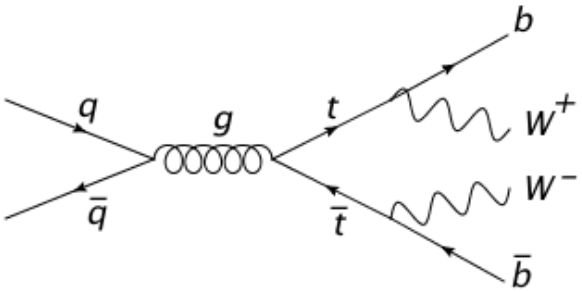


Figure 4.3: Diagram illustrating the main diagram for the $t\bar{t}$ production.

- **single top production:** there are two channels in which the single-top can be produced: the t-channel and the s-channel. The final states result in a lepton-neutrino pair if the W boson produced decays leptonically, and one or two jets depending on the channel. The two channels are illustrated in Figure 4.4.

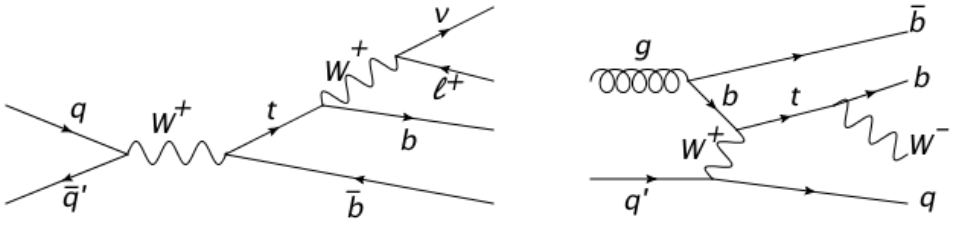


Figure 4.4: Diagram illustrating the single-top t-channel (left) and s-channel (right) production.

- **WW**, **WZ**, **ZZ production:** although the cross section of these processes is 200 times smaller than the W production, these processes can mimic the signature of $W+2$ jets (when a W decays leptonically and the other boson decays hadronically, see Figure 4.6) and make an important background contribution for jet multiplicity equal or higher than 2.

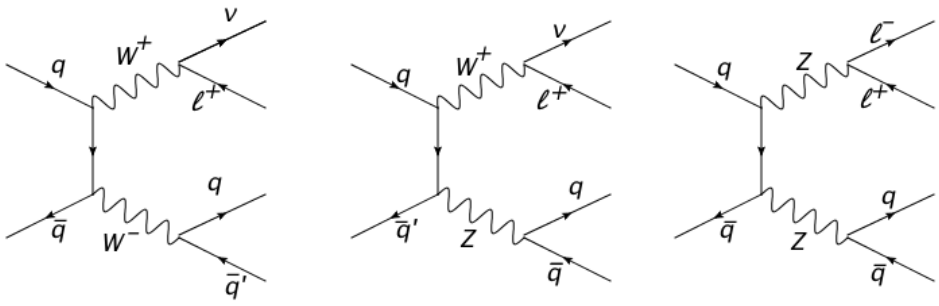


Figure 4.5: Examples of leading order Feynman diagrams for WW, WZ and ZZ processes with signature identical to $W+2$ jets final state.

- **W($\to\tau\nu$)+jets process:** the W boson decays with the same branching ratio in electrons, muons and taus due to the lepton universality. The tau lepton subsequently decays into electron (18% of the times) or muons (18%) plus two neutrinos. Although the processes $W(\to\tau(\to\ell\nu)\nu)+$jets ($\ell = e,\mu$) are not discernible from the other leptonic decays of the W boson, the contribution of this background is lower than the contribution of the signal, since the electrons and muons from the τ's decay often do not meet the high-p_T requirement.

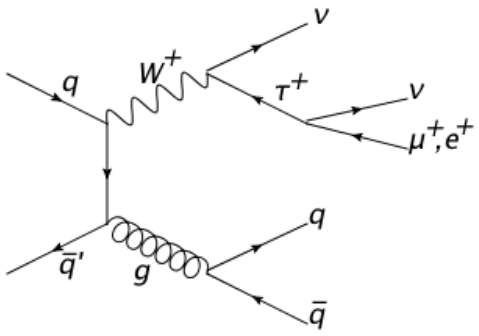


Figure 4.6: Example of Feynman diagram for $W(\rightarrow\tau(\rightarrow e\nu)\nu)+$jets and $W(\rightarrow\tau(\rightarrow \mu\nu)\nu)+$jets plus 2 jets productions.

An important contribution to the data sample comes from the **QCD multi-jet** background. Although the QCD multi-jet process is characterized by the production of hadronic jets, it is possible that a jet be misidentified as an electron when the missing E_T is significant. Figure 4.7 shows Feynman diagrams that describes the multi-jet production.

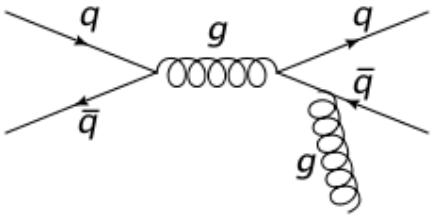


Figure 4.7: Example of QCD multi-jet production.

Although, the probability that an hadronic jet fakes the lepton ID selection is extremely small ($\sim 10^{-4}$) [83], the multi-jet background contamination is significant, because the cross-section for the QCD multi-jet production is very much larger (0.1 and 10 μb for exchange p_T of 60 and 18 GeV/c [73]) than the production cross-section of the inclusive $W+$jets. Contamination from QCD multi-jet events in which real leptons from the decay of heavy flavor quarks are present, is also possible but negligible with respect to the fake leptons rate.

In order to study a pure $W+$jets sample the backgrounds described in this section need to be removed from the data sample. To this end, first a specific event selection has been developed and applied to the data sample, then the residual background events are estimated and subtracted.

4.3 Event Selection

The physical objets, *e.g.* electrons, muons, jets and neutrinos, described in section 3.5 are identified and reconstructed offline from the data events collected with the electron and muon triggers.

To accurately reconstruct the kinematic proprieties of the physical objects, it is necessary to identify the position of the interaction vertex. Whereas the interaction region in the (x, y) plane is defined by the beam width of 30 μm, identification of the z-coordinate of the interaction requires the application of detail vertex reconstruction algorithms (described in section 3.5.1). The z-region for the interaction vertex compatible with the projective geometry of the calorimeter towers is within 60 cm from the center of the detector. Therefore, the first requirement to be fulfilled by all the events, is that at least one reconstructed primary interaction vertex with z position 60 cm from the center of the detector is present.

4.3.1 W boson reconstruction

The W boson is identified by its leptonic (electron or muon) decay. For this reason, events only one electron or only one muon with $E_T > 25$ GeV and $|\eta| < 1.0$ are selected. The requirement of only one lepton in each event guarantees the rejection of most of the $Z+$jets background (Z-Veto) and the others backgrounds involving the production of more than one lepton. The electron and muon candidates are identified in the central part of the detector ($|\eta_{det}| < 1.0$) and they have to satisfy the selection requirements summarized in table 3.2, if electrons, or in table 3.3, if muons, *i.e.*, they have to be tagged as TCE, CMUP or CMX. The minimum energy requirement is higher than the TCE, CMUP or CMX requirement to guarantee that the event selection is not affected by the trigger efficiency (as discussed in section 4.1.3 the calorimeter efficiency for the high-p_T central electron trigger is 100% if the electron-E_T is higher than 25 GeV). The presence of a neutrino is also necessary for the reconstruction of the W-boson and, although they cannot be detected by the CDF apparatus, they can be identified as missing transverse energy $\not\!\!E_T$ as discussed in section 3.5.5. Due to the limitation in the full reconstruction of the z-information of the neutrino-object it is not possible to measure the mass of the W boson. However, the transverse mass of the W boson (M_T^W) can be reconstructed by combining the transverse momentum of the selected lepton ($\vec{p}_T^{\,\ell}$) and the $\vec{\not\!\!E}_T$ as described in equation 4.3.1:

$$M_T^W = \sqrt{2\ p_T^\ell\ \not\!\!E_T\ \{1 - cos[\Delta\phi(\ell, \not\!\!E_T)]\}} \tag{4.3.1}$$

where $\Delta\phi(\ell, \not\!\!E_T)$ is the difference between the azimuthal angles of the lepton (ℓ) and the missing transverse energy. Most of the background events where neutrinos are not expected can be rejected by requesting $M_T^W > 40$ GeV/c^2 (QCD-Veto). This requirement is very efficient in rejecting the multi-jet background. In fact, since in this background no real W is produced, M_T^W is expected to be peak at low values.

4.3.2 Jet reconstruction and Selection

In addition to the W boson the presence of at least 1 jet is required. Jets, as described in section 3.5.4, are defined as calorimetric towers clustered by using the JETCLU algorithm with a cone of radius 0.4 with transverse energy of at least 25 GeV and $|\eta| < 2.0$. To ensure the shower is an hadronic jet, we also require that the fraction of electromagnetic energy ($emfr$) is lower that 0.9. Furthermore, in order to consider only well-separated jets, jets close ($\Delta R < 0.4$) to the lepton from the W-decay are neglected.

4.3.3 Summary of the Selection Criteria

In summary, the selection requirements applied to the data sample in this analysis are:

1. events collected by one of the high-p_T lepton triggers;
2. data events from runs in which the sub-detectors needed were working properly (*i.e.*, runs from the appropriate good run list);
3. cosmic ray veto;
4. at least one vertex with z-position within 60 cm from the center of the detector;
5. only one lepton tagged TCE, CMUP or CMX with $E_T > 25$ GeV and $|\eta| < 1.0$;
6. at least 1 jet (reconstructed with JETCLU algorithm and cone 0.4) with $E_T > 25$ GeV, $|\eta| < 2$, $emfr < 0.9$ and $\Delta R(\ell, jets) > 0.4$.

The number of data that survived these cuts is 477665 for the $W(\rightarrow e\nu)$ channel and 229823 for the $W(\rightarrow \mu\nu)$ channel.

4.4 Signal and Background Modeling

In this analysis backgrounds are modeled by using both simulation and data-driven methods. The EWK-like backgrounds are estimated using Monte Carlo simulations whereas data-driven techniques are used to estimate the QCD multi-jets background contribution. $W+$ jets samples simulated using the MC technique, are used to validate the background models and to perform unfolding.

4.4.1 Monte Carlo Models

The samples used to model the signal ($W(\rightarrow \ell\nu)+$ jets, $\ell = e, \mu$), the electroweak (WW, WZ, ZZ, $Z(\rightarrow \ell\ell)+$ jets, $W(\rightarrow \tau\nu)+$ jets) and top ($t\bar{t}$ and single-top) backgrounds are simulated with a MC-based method that can be summarized in three steps, as follows:

step 1: Simulation of the physics process *i.e.* the proton-antiproton interaction and the particle produced by it. The output of this step is a list of particles, their position's coordinates and their 4-momenta.

step 2: Simulation of the interaction of the particles produced in the previous step with the CDF detector using GEANT3 [85]. This simulation is run-dependent *i.e.* the specific conditions of the detector during data-taking are reproduced run by run.

step 3: Application of the selection criteria reported in section 4.3.3, *i.e.*, the trigger requirements and the analysis selection to the sample generated on step 2.

Step 1: Physics Events Generation

To perform step 1 three different physics process generators are used in this analysis: ALPGEN [30], PYTHIA [31] and MadEvent [86].

ALPGEN version 2.10 interfaced with PYTHIA version 6.325, is used to model $p\bar{p} \rightarrow (Z+\text{jets})$ and $p\bar{p} \rightarrow (W(\rightarrow \ell\nu)+\text{jets})$ processes. ALPGEN which is a leading order (LO) tree-level matrix element calculator, is specifically designed for processes like a vector boson and a fixed number, n, of partons in the final state (*i.e.* $p\bar{p} \rightarrow Z(\rightarrow \ell\ell)+0,1,2,3,\geqslant 4$ partons and $p\bar{p} \rightarrow W(\rightarrow \ell\nu)+0,1,2,3,\geqslant 4$ partons). ALPGEN performs only the generation of the hard process. PYTHIA, instead, performs the parton shower evolution into the final state particles. To mach the partons generated by the initial interaction with those produced in the parton shower and avoiding double-counting MLM matching scheme it is used. Details on ALPGEN and PYTHIA generators and MLM matching scheme are reported in section 2.2.1. The $Z+$ jets sample is built by generating $Z(\rightarrow \ell\ell) + n$ partons samples with three different cuts on the di-lepton invariant mass (Z mass range): $m(\ell\ell) \in [20, 75]$ Gev/c^2, $m(\ell\ell) \in [75, 105]$ Gev/c^2 and $m(\ell\ell) \in [105, 600]$ Gev/c^2. The low and the high mass ranges are needed to include the contribution from "Drell-Yan" production where the mediator is off-shell.

PYTHIA version 6.216, is used to generate the $p\bar{p} \rightarrow WZ, WW, ZZ$ and $p\bar{p} \rightarrow t\bar{t}$ backgrounds. These four samples are inclusive decay samples. The diboson samples include a generated $Z/\gamma*$ mass down to 2 GeV/c^2 while the W boson is required to be on-shell. The $t\bar{t}$ sample is generated assuming a top mass of 172.5 GeV/c^2, and contains all the possible (not only leptonic) decays of the W boson produced.

The $p\bar{p} \rightarrow$ single-top (both s-channel and t-channel) are generated using MadEvent version 6.510 interfaced with PYTHIA. MadEvent is a tree-level generator that uses MadGraph [87] as matrix element generator. MadEvent produces events at the parton level according to Standard Model processes. The shower evolution is performed by PYTHIA.

The Parton Distribution Functions (PDFs), used in the analysis presented here, is the leading order (LO) CTEQ5L set [88].

Step 2: CDF Detector Simulation

In step 2, the propagation of the final particles generated in step 1, through the passive and active materials of the detector and the response of the various sub-detectors is simulated. In order to have a very realistic model, all the experimental and detector variations that have taken place during data taking, *e.g.*, informations

on beam position, detector deterioration and new operating conditions (noise, calibrations, etc.), are implemented run-by-run in the simulation. Also, the simulation accounts for the instantaneous luminosity profile to reproduce multiple interactions in the same $p\bar{p}$ collision. This is a very computer-time consuming step. To speed it up, the particle-shower simulation in the calorimeter is based on the GFLASH [89] parametrization package. The GFLASH program basically uses the parameterizations of longitudinal and lateral shower profiles for both the electromagnetic and hadronic showers. The parameters are tuned to agree with pion testbeam data and $p\bar{p}$ collision data [90].

Step 3: Trigger and Analysis Requirements

The events output of step 2 are treated as real data: the same reconstruction algorithms of section 3.5 and selection criteria of section 4.3 are applied.

To each Monte Carlo event i, that survived the trigger and analysis cuts, a weight (w_i) computed as shown in Eq. 4.4.1 is assigned:

$$w_i = \sigma_{p\bar{p}\to X} \cdot \epsilon_i^{trigger} \cdot w_i^{vertex} \cdot SF_i^{\ell-ID} \cdot \frac{1}{N_{gen}} \cdot \int \mathcal{L}dt \tag{4.4.1}$$

where:

$\sigma_{p\bar{p}\to X}$ is the theoretical cross section for the process simulated *i.e.* $X = WW$, WZ, ZZ, $t\bar{t}$, single top, $Z(\to \ell\ell)+$ jets, $W(\to \tau\nu)+$ jets;

$\epsilon_i^{trigger}$ is the trigger efficiency;

w_i^{vertex} is the weight applied to reproduce the vertex distribution corresponding to the instantaneous luminosity .

$SF_i^{\ell-ID}$ is the scale factor which accounts for possible discrepancies in lepton identification efficiencies between data and MC (the estimation of $SF_i^{\ell-ID}$ is described in [56]);

N_{gen} is the number of events generated for each MC sample;

$\int \mathcal{L}dt$ is the integrated luminosity.

As shown by 4.4.1, each simulated event is scaled to the theoretical cross-section times branching fraction ($\sigma_{p\bar{p}\to X}$) for the background process intended to reproduced being generated. The associated production of W and Z (WW, WZ, ZZ) is scaled to the cross-section calculated to next-to-leading order (NLO) [91], the $t\bar{t}$ sample is normalized using an approximate next-to-next-to-leading order plus next-to-next-to-leading logarithm (NNLO+NNLL) [92] cross-section calculation, the single-top process is normalized to approximate next-to-next-leading order plus next-to-next-to-leading logarithm (NNLO+NNLL) calculations [93] for the s-channel and approximate next-to-next-to-leading order plus next-to-leading logarithm (NNLO+NLL) calculations [94] for the t-channel, and the $Z+$ jets and $W+$ jets are normalized to leading order (LO) calculations and scaled by an additional K-factor of 1.4 to account for next-to-leading order effects [95] and [96].

The event generators, the theoretical cross-sections ($\sigma_{p\bar{p}\to X}$) and their accuracy for each simulated physics process are summarized on tables 4.4, 4.5 and 4.6.

Process (p stands for parton)	σ (pb)	Accuracy	Generator
WW	11.3 ± 0.7	NLO	PYTHIA
WZ	3.7 ± 0.2	NLO	PYTHIA
ZZ	1.4 ± 0.1	NLO	PYTHIA
$t\bar{t}$	7.164 ± 0.16	NNLO	PYTHIA
single-top (s-channel)	1.06 ± 0.06	NNLO+NNLL	MadEvent+PYTHIA
single-top (t-channel)	2.12 ± 0.22	NNLO+NLL	MadEvent+PYTHIA
$W^{\pm}(\to \tau^{\pm}\nu_{\tau}) + 0\,\mathrm{p}$	2520	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to \tau^{\pm}\nu_{\tau}) + 1\,\mathrm{p}$	315	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to \tau^{\pm}\nu_{\tau}) + 2\,\mathrm{p}$	49.42	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to \tau^{\pm}\nu_{\tau}) + 3\,\mathrm{p}$	7.826	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to \tau^{\pm}\nu_{\tau}) + \geq 4\,\mathrm{p}$	1.442	LO×K-factor	ALPGEN+PYTHIA

Table 4.4: Monte Carlo samples used for modeling the $W(\to \tau\nu)+$jets, WW, ZZ, WZ, $t\bar{t}$ and single-top backgrounds. The uncertainty for the $W(\to \tau\nu)+$jets cross-section is 40% to account for the large uncertainties carried by both the theoretical LO calculations and the K-factor

Physic Process (p stands for parton)	σ (pb)	Accuracy	Generator
$W^{\pm}(\to e^{\pm}\nu_{e}) + 0\,\mathrm{p}$	2520	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to e^{\pm}\nu_{e}) + 1\,\mathrm{p}$	315	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to e^{\pm}\nu_{e}) + 2\,\mathrm{p}$	49.42	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to e^{\pm}\nu_{e}) + 3\,\mathrm{p}$	7.826	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to e^{\pm}\nu_{e}) + \geq 4\,\mathrm{p}$	1.442	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to \mu^{\pm}\nu_{\mu}) + 0\,\mathrm{p}$	2520	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to \mu^{\pm}\nu_{\mu}) + 1\,\mathrm{p}$	315	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to \mu^{\pm}\nu_{\mu}) + 2\,\mathrm{p}$	49.42	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to \mu^{\pm}\nu_{\mu}) + 3\,\mathrm{p}$	7.826	LO×K-factor	ALPGEN+PYTHIA
$W^{\pm}(\to \mu^{\pm}\nu_{\mu}) + \geq 4\,\mathrm{p}$	1.442	LO×K-factor	ALPGEN+PYTHIA

Table 4.5: Monte Carlo samples used to perform the unfolding. The uncertainty for the $W(\to \ell\nu)+$jets cross-section is 40% to account for the large uncertainties carried by both the theoretical LO calculations and the K-factor.

Physic Process (p is parton)	Z Mass Range (GeV/c^2)	σ (pb)	Accuracy	Generator
$Z(\to e^+e^-) + 0\,\mathrm{p}$	$[75, 105]$	221.2	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 1\,\mathrm{p}$	$[75, 105]$	30.24	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 2\,\mathrm{p}$	$[75, 105]$	4.844	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 3\,\mathrm{p}$	$[75, 105]$	0.77	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + \geq 4\,\mathrm{p}$	$[75, 105]$	0.139	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 0\,\mathrm{p}$	$[75, 105]$	221.2	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 1\,\mathrm{p}$	$[75, 105]$	30.24	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 2\,\mathrm{p}$	$[75, 105]$	4.844	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 3\,\mathrm{p}$	$[75, 105]$	0.77	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + \geq 4\,\mathrm{p}$	$[75, 105]$	0.139	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + 0\,\mathrm{p}$	$[75, 105]$	221.2	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + 1\,\mathrm{p}$	$[75, 105]$	30.24	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + \geq 2\,\mathrm{p}$	$[75, 105]$	4.844	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 0\,\mathrm{p}$	$[20, 75]$	224	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 1\,\mathrm{p}$	$[20, 75]$	11.75	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 2\,\mathrm{p}$	$[20, 75]$	2.254	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 3\,\mathrm{p}$	$[20, 75]$	0.326	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + \geq 4\,\mathrm{p}$	$[20, 75]$	0.557	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 0\,\mathrm{p}$	$[20, 75]$	224	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 1\,\mathrm{p}$	$[20, 75]$	11.75	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 2\,\mathrm{p},$	$[20, 75]$	2.25	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 3\,\mathrm{p}$	$[20, 75]$	0.326	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + \geq 4\,\mathrm{p}$	$[20, 75]$	0.557	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + 0\,\mathrm{p}$	$[20, 75]$	224	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + 1\,\mathrm{p}$	$[20, 75]$	11.73	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + \geq 2\,\mathrm{p}$	$[20, 75]$	2.55	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 0\,\mathrm{p}$	$[105, 600]$	5.698	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 1\,\mathrm{p}$	$[105, 600]$	0.988	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 2\,\mathrm{p}$	$[105, 600]$	0.164	LO×K-factor	ALPGEN+PYTHIA
$Z(\to e^+e^-) + 3\,\mathrm{p}$	$[105, 600]$	0.026	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 0\,\mathrm{p}$	$[105, 600]$	5.698	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 1\,\mathrm{p}$	$[105, 600]$	0.988	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 2\,\mathrm{p}$	$[105, 600]$	0.164	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \mu^+\mu^-) + 3\,\mathrm{p}$	$[105, 600]$	0.026	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + 0\,\mathrm{p}$	$[105, 600]$	5.698	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + 1\,\mathrm{p}$	$[105, 600]$	0.988	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + 2\,\mathrm{p}$	$[105, 600]$	0.164	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + 3\,\mathrm{p}$	$[105, 600]$	0.026	LO×K-factor	ALPGEN+PYTHIA
$Z(\to \tau^+\tau^-) + \geq 4\,\mathrm{p}$	$[105, 600]$	$4.65{\cdot}10^{-3}$	LO×K-factor	ALPGEN+PYTHIA

Table 4.6: Monte Carlo samples used for modeling the $Z+$jets background. The table reports the cross-sections with the K-factor applied, with an uncertainty of 20% [97].

The number of events, for each background X, that contributes to the data sample is obtained by summing the weights w_i, defined in equation 4.4.1, over the number of the events (N_{sel}) that survived the three selection steps *i.e.*:

$$N_X = \sum_{i=0}^{N_{sel}} w_i \tag{4.4.2}$$

The contribution of each background to the data sample for each inclusive jet multiplicity is reported in table 4.14 and 4.15 of section 4.6.1.

4.4.2 QCD Multi-jet Background Model

The QCD multi-jet background is characterized by hadronic jets that are incorrectly measured and consequently are wrongly identified as leptons. These jets, misidentified as leptons, are often referred to as fake electrons or fake muons. At the same time, the energy of the jet can be underestimated thereby contributing to an overestimate of the missing transverse energy which can be interpreted as a neutrino. The huge QCD multi-jet production cross-section, makes the contribution of such background in the data sample significant, despite the fact that the fake rate, *i.e.*, the probability that a hadron jet fakes the lepton signature, is very small. Another consequence of the high QCD multi-jet production rate is that the Monte Carlo simulation approach is not feasible because it is necessary to simulate an extremely large number of events in order to generate a statistically significant sample. Moreover, the Monte Carlo simulation would not be reliable because the mechanism whereby leptons are faked is not completely understood, and needs a more accurate detector simulation than is available. For these reasons the QCD multi-jet background in this kind of analysis is estimated using a data-driven technique.

Fake lepton samples are defined by choosing events with kinematics which are similar but nevertheless orthogonal to the data. In practice, we change the definition of the TCE, CMUP and CMX leptons slightly by reversing one or more of the selection criteria used for lepton identification (sections 3.5.2 and 3.5.3), while the remaining event selection criteria are still applied.

The fake muon sample, is built using events with *non-isolated muons*, which are objects that pass all the CMUP or CMX selections criteria of table 3.3 with the exception of the isolation requirement.

The fake electron sample is built using a model (called the "antielectron method") developed in previous $W+$jets CDF analyses [78] [98], *i.e.*, by using events in which a number (at least two in the antielectron method) of the so-called identification (ID) cuts, reported in the right half of table 4.7, are reversed. The other criteria that identify a TCE, which are called kinematic cuts and are reported in the left half of table 4.7, still need to be satisfied.

kinematic cuts	Identification (ID) cuts
Fiducial to CEM	$E_{had}/E_{em} < 0.055 + 0.00045 \cdot E$ (GeV)
Track $\lvert z_0 \rvert \leqslant 60$ cm	Strip $\chi^2 < 10$
$E/p < 2.5 + 0.015 \cdot E_T$ (GeV)	$L_{shr} \leqslant 0.2$
Track $p_T \geqslant 10$ GeV/c	-3.0 cm $\leqslant Q \cdot \Delta X \leqslant 1.5$ cm
Isolation $\leqslant 0.1$	$\lvert \Delta Z \rvert \leqslant 3.0$ cm
COT Axial Segments $\geqslant 3$	
COT Stereo Segments $\geqslant 2$	

Table 4.7: Tight Central Electron (TCE) kinematic and identification cuts. Fake electrons are defined reversing at least two of the ID cuts.

The electron requirements that can be reversed are chosen in such a way that the *antielectrons* have kinematic features similar to the electron sample while the statistics are maximized.

Modeling the QCD multi-jet background has always been a challenge, and the *antielectron* and the *non-isolated muon* models are the CDF's customary way if estimating the QCD custom multi-jet background for the analysis with a $W+$ jets-like signature. The reason for using two different models for the description of the QCD-multijet background is because underlays that the "faking mechanism" for the two leptons is different. Besides, the fake electron model requires more accuracy because the QCD multijet production is the dominant background for the $W \to e\nu$ channel, whereas, its contribution is almost negligible for the $W \to \mu\nu$ channel. This because, it is easier for a jet to fake an electron signature since both deposit energy in the electromagnetic calorimeter whereas on the other hand, a jet cannot deposit a significant amount of energy in the muon detectors without depositing considerably more than a MIP signature inside the calorimeter.

The *antielectron* model has been tested in a study, using the full CDF dataset, of the W-boson produced in association of two jets [84]. This study revealed some defects of the model, and led to the development of a new model called *non-electron* model. In the *non-electron* model, the definition of the fake electron sample is similar that of the *antielectron* but only exactly two of the kinematic cuts are reversed. In addition, the model is improved by applying the corrections which can be summarized as follows (see section 4.5 for details):

- the contribution of the EWK-like backgrounds is removed from the fake electron sample. As the background contaminates the data sample due to the selection criteria inefficiencies, we expect that the data-driven QCD multi-jet model has the same problem. The bin by bin subtraction is performed per variable of interest to guarantee that the shape of each distribution is not biased by the other backgrounds.

- in order to correctly model the event kinematic, *non-electrons* are described as jets by using the jet that matches the *non-electron i.e.* the jet with a distance

on the $\eta - \phi$ plane of ΔR(jet,non-ele) < 0.4. In this way, we account for the additional energy in the cone, with radius 0.4, around the *non-electron*. This additional energy is due to *non-electrons* being characterized by a larger amount of energy around the cluster than electrons. The *non-electron* energy is furthermore corrected using a scale correction. This *non-electron* energy scale (neES) correction, consists in multiplying the transverse energy of the jet matched to the *non-electron* by a scale factor in order to make it consistent with the transverse energy of the misidentified electron (i.e. a jet which passed all the TCE requirements of table 3.2). The scale factors are evaluated using QCD di-jet MC events, as described in section 4.5.1.

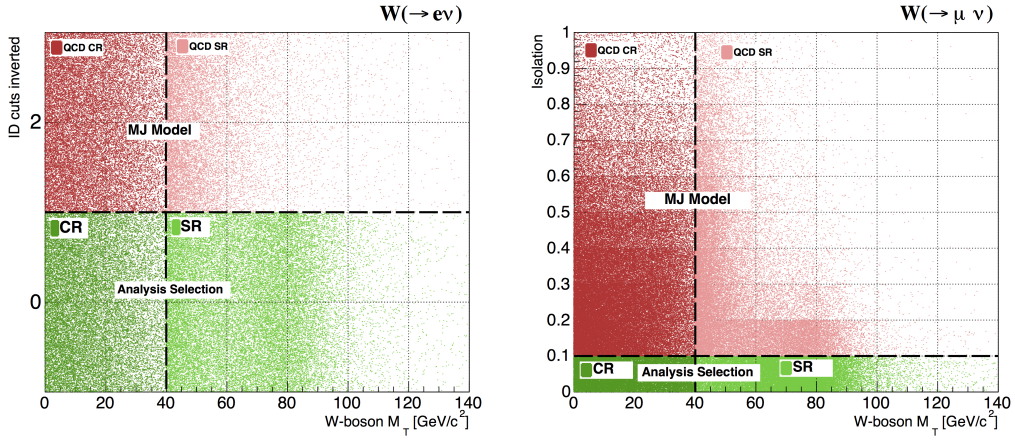
- A weight designed to remove a trigger bias, is assigned to each event of the *non-electron* sample. The weights are calculated by reweighting the *non-electron*-E_T template bin-by-bin so as to reproduce the data transverse energy distributions on the data, in a region where only QCD multi-jet contribution is expected. The need for this correction is due from the fact that the trigger threshold sculpts the E_T spectrum of *non-electrons* being select by cut inversion, their electromagnetic energy fraction is lower that of the misidentified electrons selected in the data sample.

In this analysis, we evaluate the QCD multi-jet background using the *non-electron* model for the $W \to e\nu$ channel and the *non-isolated muon* model for the $W \to \mu\nu$ channel. Since the *non-electron* model was specifically developed for $W(\to e\nu) + 2$ jets channel, we performed studies to either validate or further improve the model for application to the more general $W+$jets. Moreover, we performed studies to investigate possible improvements in the *non-isolated muon* model.

4.5 Validation of the QCD Multi-jet Background

The modeling of the QCD multi-jet background, especially for the $W \to e\nu$ channel is the main challenge of this analysis. Both the *non-electron* and the *non-isolated muon* models are tested for each inclusive jet multiplicity. To this end we define a special region expected to be QCD multi-jet enriched, kinematically close and orthogonal to the region in which we want to perform the measurement (the signal region, SR). This region, to which we refer as Control Region (CR), is defined by reversing the QCD-Veto, *i.e.* by requiring $M_T^W \leqslant 40$ GeV.

Figure 4.8 depicts the four samples into which the events that pass the high-p_T electron and muon triggers are divided. The QCD-Veto (x-axis) defines the control and signal region samples, and the fake lepton conditions (y-axis) distinguish between reconstructed electrons and muons and objects which are used for the QCD multi-jet model (regions labelled QCD).



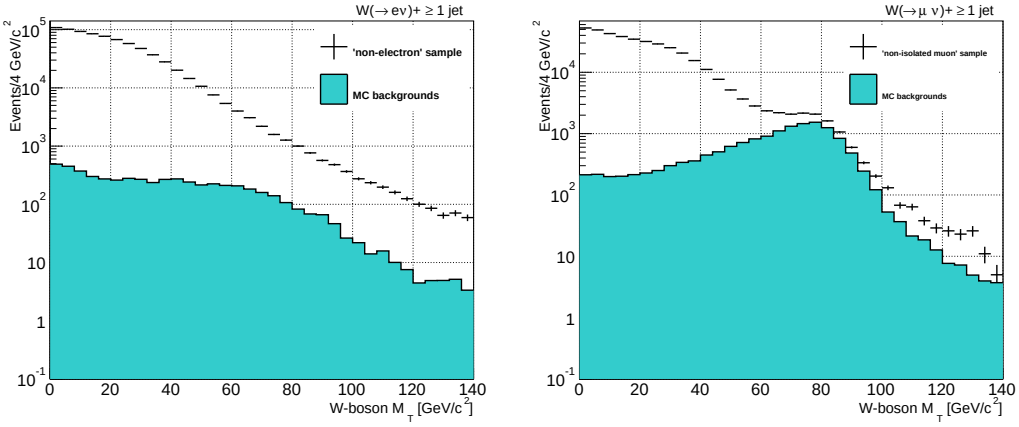
(a) (b)

Figure 4.8: Two-dimensional plot of the events used in this analysis (a) for the $W \to e\nu$ and (b) for the $W \to \mu\nu$ channels, respectively. The y-axis is used to indicate the conditions used to define the *non-electrons*, *i.e.*, two of the ID cuts inverted, and the *non-isolated muons*, *i.e.* isolation $\geqslant 0.1$.

In what follows, signal and control regions will be identified by the subscripts SR and CR for the data selected for the analysis by the subscripts, QCD SR and QCD CR for the data used for QCD multi-jet model (MJ model).

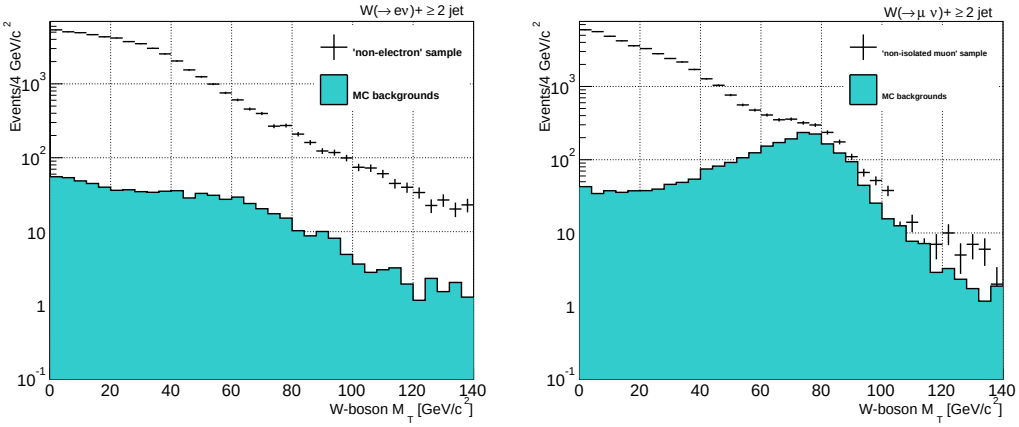
Removal of contaminations from the QCD multi-jet samples

The first step of the procedure performed to obtain reliable QCD multi-jet samples is to subtract the contribution of the EWK-like backgrounds in which a real lepton is present from the data-driven *non-electron* and *non-isolated muon* samples *i.e.*, from the QCD CR and SR of figure 4.8. To simulate the contamination we use the same MC-simulated backgrounds defined in section 4.4.1 and we evaluate the contamination by reversing the same ID criteria as done to define the non lepton samples. The removal procedure, avoids double counting between the MC-simulated backgrounds and the QCD multi-jet model and is performed in both signal and control regions. Figures 4.9, 4.10, 4.11 and 4.12 show the background contribution present in the QCD multi-jet sample for the $W \to e\nu$ channel (left) and $W \to \mu\nu$ channel (right) as a function of the inclusive jet multiplicity. The variable histogrammed is the transverse mass of the W boson, M_T^W, before applying any selection condition to it, *i.e.*, the plots depict signal and control region together. The Jacobian peak around 80 GeV/c^2 confirming the presence of a real vector boson (W or Z, since they are indistinguishable with the CDF calorimeter resolution) is visible in the $W \to \mu\nu$ channel. The peak is instead not well defined in the $W \to e\nu$ channel, probably because the *non-electron* selection is more likely to select backgrounds without ("well reconstructed") real vector bosons since the fake mechanism is different.



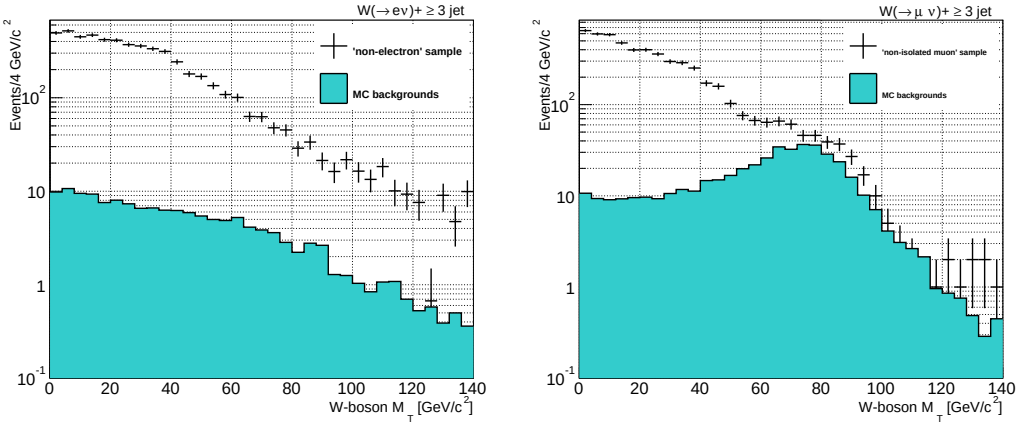
(a) (b)

Figure 4.9: M_T^W distribution (in logarithmic scale) before applying the control/signal region selection for the (a) *non-electron* sample in the $W(\to e\nu)+ \geqslant 1$ jets channel and (b) *non-isolated muon* sample in the $W(\to \mu\nu)+ \geqslant 1$ jets channel with MC background contamination (solid green) overlaid.



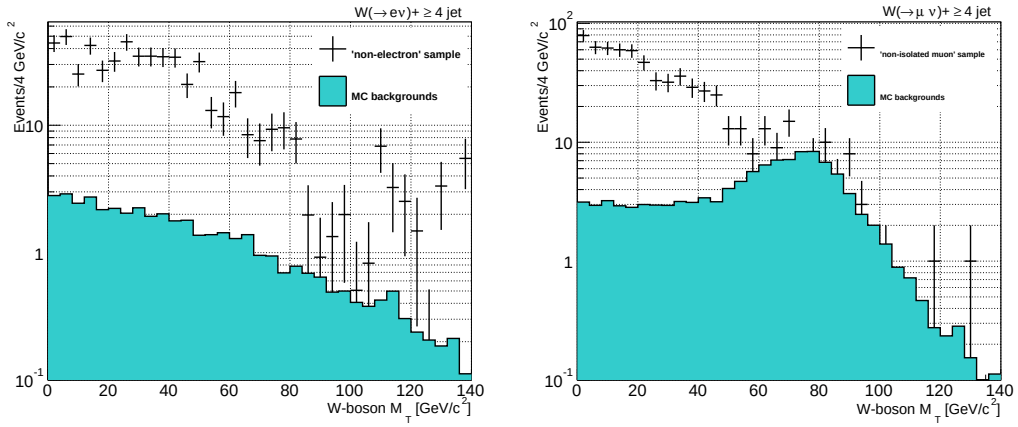
(a) (b)

Figure 4.10: M_T^W distribution (in logarithmic scale) before applying the control/signal region selection for the (a) *non-electron* sample in the $W(\to e\nu)+ \geqslant 2$ jets channel and (b) *non-isolated muon* sample in the $W(\to \mu\nu)+ \geqslant 2$ jets channel with MC background contamination (solid green) overlaid.



(a) (b)

Figure 4.11: M_T^W distribution (in logarithmic scale) before applying the control/signal region selection for the (a) *non-electron* sample in the $W(\rightarrow e\nu)+ \geqslant 3$ jets channel and (b) *non-isolated muon* sample in the $W(\rightarrow e\nu)+ \geqslant 3$ jets channel with MC background contamination (solid green) overlaid.



(a) (b)

Figure 4.12: M_T^W distribution (in logarithmic scale) before applying the control/signal region selection for the (a) *non-electron* sample in the $W(\rightarrow e\nu)+ \geqslant 4$ jets channel and (b) *non-isolated muon* sample in the $W(\rightarrow \mu\nu)+ \geqslant 4$ jets channel with MC background contamination (solid green) overlaid.

The subtraction of the EWK backgrounds predicted by MC is done bin-by-bin for each variable of interest. The number of data-events selected to model the QCD multi-jet backgrounds, $N_{MJ}|_{QCD}$, for each inclusive jet multiplicity and for the two decay mode channels are reported on table 4.8.

	***non-electrons* model**			
jet multiplicity	$\geqslant 1$ jet	$\geqslant 2$ jets	$\geqslant 3$ jets	$\geqslant 4$ jets
events MJ model	777983 ± 882	50965 ± 226	5511 ± 74	573 ± 24
EWK-backgrounds	5566 ± 62	758 ± 14	148 ± 3	43 ± 1
$\mathbf{N_{MJ}\vert_{QCD}}$	772417 ± 884	50207 ± 226	5363 ± 74	530 ± 24

	***non-isolated muons* model**			
jet multiplicity	$\geqslant 1$ jet	$\geqslant 2$ jets	$\geqslant 3$ jets	$\geqslant 4$ jets
events MJ model	385161 ± 620	43151 ± 208	5308 ± 73	665 ± 26
EWK-backgrounds	15152 ± 122	2429 ± 34	467 ± 7	116 ± 1
$\mathbf{N_{MJ}\vert_{QCD}}$	370009 ± 632	40722 ± 211	4841 ± 73	549 ± 26

Table 4.8: Number of events in the *non-electrons* and *non-isolated muons* samples (MJ model) in both control and signal region for each jet multiplicity. In the table are also reported the expected EWK-like background events that contaminate the QCD multi-jet model samples and the number of events after the subtraction of the contamination. All the errors are statistical only.

All the MCs are normalized to their theoretical cross-section. The $W+$jets background is simulated using the MC samples and the normalizations reported in table 4.5. To avoid circularity instead of the theoretical cross-section we should use the $W+$jets cross-section estimated from the signal region. However, since $W+$jets contribution to the subtraction sample is small, we expect that the differences are well within the QCD multi-jet uncertainties.

Estimation of the QCD multi-jet rate

The removal of the real lepton backgrounds described in the previous section is only a shape correction. The QCD multi-jet rate expected in the analysis signal region ($N_{MJ}|_{SR}$) is estimated using equation 4.5.1:

$$N_{MJ}|_{SR} = \frac{N_{MJ}|_{CR}}{N_{MJ}|_{QCD\,CR}} \cdot N_{MJ}|_{QCD\,SR} \tag{4.5.1}$$

where $N_{MJ}|_{QCD\,CR}$ and $N_{MJ}|_{QCD\,SR}$ are the number of events that have passed the *non-electron* or *non-isolated muon* selections after the removal of the contribution from the EWK processes as previously explained and $N_{MJ}|_{CR}$ is the number of QCD multi-jet events expected in the control region and is evaluated by subtracting the Monte Carlo-simulated EWK background ($N_{EWK}|_{CR}$) and signal ($N_{WJETS}|_{CR}$) contribution from the control region data sample ($N_{DATA}|_{CR}$):

$$N_{MJ}|_{CR}=(N_{DATA} - N_{EWK} - N_{WJETS})|_{CR} \tag{4.5.2}$$

To avoid circularity, $N_{WJETS}|_{CR}$ is estimated using the measured cross-section σ^{meas}_{WJETS}, instead of the theoretical calculation. The process is iterative: since, at

first order, the CR is compose entirely of QCD multi-jet events, we start with the approximation:

$$N_{MJ}|_{CR}=N_{DATA}|_{CR} \tag{4.5.3}$$

and we calculate σ^{meas}_{WJETS} with equation 4.5.4:

$$\sigma^{meas}_{WJETS} = \frac{N_{DATA}|_{SR} - N_{EWK}|_{SR} - N_{MJ}|_{SR}}{L\,\mathcal{A}_{SR}\,\epsilon} \tag{4.5.4}$$

where $L\ \mathcal{A}_{SR}\ \epsilon$ is the product of the integrated luminosity, the acceptance in the SR and the trigger efficiency; $N_{EWK}|_{SR}$ is the estimated number of EWK-like background in the signal region and $N_{MJ}|_{SR}$ is evaluated by replacing $N_{MJ}|_{CR}$ of equation 4.5.3 into equation 4.5.1. $N_{WJETS}|_{CR}$ is then calculated by using σ^{meas}_{WJETS}. The next order approximation of $N_{MJ}|_{CR}$ is then calculated with the measured $N_{WJETS}|_{CR}$. The process is iterated until the QCD scale factor f^{ℓ}_{QCD} defined by equation:

$$f^{\ell}_{QCD} = \frac{N_{MJ}|_{CR}}{N_{MJ}|_{QCD\,CR}} \tag{4.5.5}$$

does not change by more than 1%.

Uncertainties in the shape and rate of the EWK-like backgrounds predicted by the Monte Carlo contribute to the total systematic uncertainty of the final QCD multi-jets model prediction. This QCD multi-jet rate uncertainty is evaluated by changing the rate of each MC process by:

- 2.2% to account for the systematic uncertainty in the trigger acceptance [80].
- varying each cross-section within theoretical uncertainties. In the CR, to avoid circularity the measured $W+$ jets cross-section is used as described above (equation 4.5.2). We assigned a conservative 20% uncertainty to this contribution.
- 6% for the standard systematic uncertainty on the luminosity.
- by shifting the JES up and down by 1 sigma. This uncertainty is not only a rate uncertainty but also a shape uncertainty, because it varies from bin-to-bin.

All the systematic uncertainties are summed in quadrature. Table 4.9 summarizes the various contributions, more details about these systematic uncertainties can be found in section 4.7.

Systematic	W+jets	Z+jets	$t\bar{t}$	sing. top	WW,WZ,ZZ
Trigger Acc.		2.2%	2.2%	2.2%	2.2%
Cross-section	20%	20%	3%	11%	6.1%
Luminosity		6%	6%	6%	6%
JES	$\pm1\sigma$	$\pm1\sigma$	$\pm1\sigma$	$\pm1\sigma$	$\pm1\sigma$

Table 4.9: List of the systematic uncertainties considered for the QCD rate estimation.

The QCD scale factor f^{ℓ}_{QCD} is evaluated separately for each exclusive jet multiplicity, since there is no *a priori* reason to assume that the QCD rate does not

vary with jet multiplicity. Only the 4-jet bin is inclusive for statistical reasons. Tables 4.10 and 4.11 report the values of f^{ℓ}_{QCD} together with the expected EWK-like background events for the $W \to e\nu$ and $W \to \mu\nu$ control regions, respectively.

CR W $\to$ eν	+1 jet	+2 jets	+3 jets	+ $\geq$ 4 jets
Data	1352174	94726	9727	1187
Z+jets	56912 ± 12013	4501 ± 949	457 ± 96	44 ± 9
$W \to \ell\nu$+jets	40671 ± 8229	5354 ± 1075	742 ± 149	76 ± 16
$t\bar{t}$	44 ± 3	171 ± 12	241 ± 18	187 ± 16
single top	136 ± 18	169 ± 20	42 ± 21	7 ± 1
WW, WZ, ZZ	364 ± 33	302 ± 31	56 ± 6	9 ± 1
QCD	1254047 ± 14608	84229 ± 1467	8189 ± 204	864 ± 42
$\mathbf{f^{e}_{QCD}}$	$\mathbf{1.90 \pm 0.02}$	$\mathbf{2.29 \pm 0.04}$	$\mathbf{2.21 \pm 0.06}$	$\mathbf{2.50 \pm 0.14}$

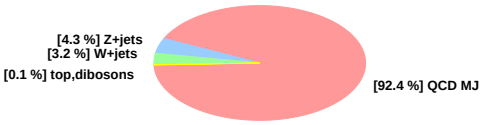
Table 4.10: QCD scale factors f^{μ}_{QCD} used to evaluate the QCD multi-jet contribution to the SR. Data and expected backgrounds events in the CR for the $W(\to e\nu) + n$ jets are also reported. The number of QCD multi-jet events is calculated with equation 4.5.2 and the uncertainties are obtained by considering table 4.9.

CR W $\to \mu\nu$	+1 jet	+2 jets	+3 jets	+ $\geq$ 4 jets
Data	117663	10528	1386	272
Z+jets	23826 ± 5021	2265 ± 477	231 ± 49	22 ± 5
$W \to \ell\nu$+jets	30937 ± 6187	4036 ± 813	472 ± 95	73 ± 16
$t\bar{t}$	33 ± 3	122 ± 9	175 ± 13	133 ± 11
single top	99 ± 12	126 ± 15	29 ± 4	5 ± 1
WW, WZ, ZZ	276 ± 25	218 ± 22	37 ± 4	6 ± 1
QCD	62492 ± 7975	3761 ± 949	442 ± 114	33 ± 26
$\mathbf{f^{\mu}_{QCD}}$	$\mathbf{0.21 \pm 0.03}$	$\mathbf{0.12 \pm 0.03}$	$\mathbf{0.12 \pm 0.03}$	$\mathbf{0.07 \pm 0.05}$

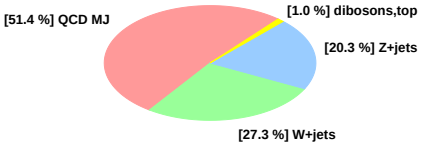
Table 4.11: QCD scale factors f^{μ}_{QCD} used to evaluate the QCD multi-jet contribution to the SR. Data and expected backgrounds events in the CR for the $W(\to \mu\nu) + n$ jets are also reported. The number of QCD multi-jet events is calculated with equation 4.5.2 and the uncertainties are obtained by considering table 4.9.

We observe that f^{e}_{QCD} is a factor 10 higher than f^{μ}_{QCD}.

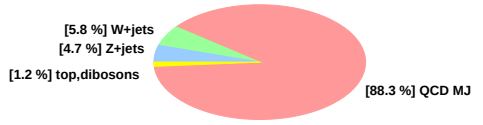
After the estimation of the QCD multi-jet yield in each bin of exclusive jet multiplicity, the inclusive jet multiplicity templates are built. The contributions in the control region of the various processes for each jet multiplicity are reported in figure 4.13.



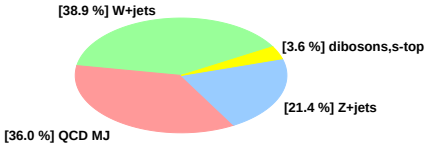
(a) $W(\to e\nu)+ \geqslant 1$ jets control region



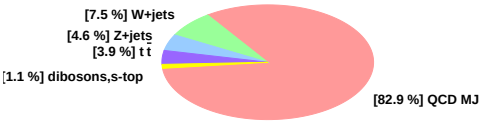
(b) $W(\to \mu\nu)+ \geqslant 1$ jets control region



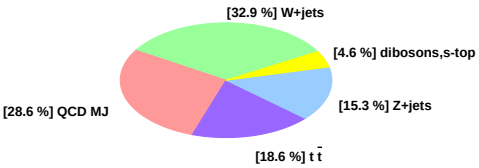
(c) $W(\to e\nu)+ \geqslant 2$ jets control region



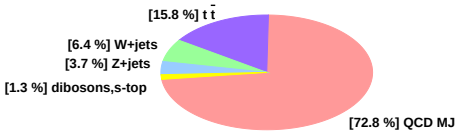
(d) $W(\to \mu\nu)+ \geqslant 2$ jets control region



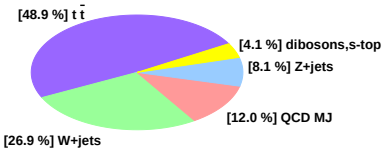
(e) $W(\to e\nu)+ \geqslant 3$ jets control region



(f) $W(\to \mu\nu)+ \geqslant 3$ jets control region



(g) $W(\to e\nu)+ \geqslant 4$ jets control region



(h) $W(\to \mu\nu)+ \geqslant 4$ jets control region

Figure 4.13: Relative composition of the control region for $W \to e\nu$ (left) and $W \to \mu\nu$ (right) as a function of the jet multiplicity. For convenience $t\bar{t}$ and single-top processes are labelled "top" when considered together, and WW, WZ, ZZ processes are labelled as "dibosons".

From figure 4.13 it is evident that the QCD MJ background in the control region define by reversing the QCD-Veto is dominant only for the electron channel. This is a consequence of the, already discussed, diversity between the electron and muon "faking mechanism". The estimated fraction of multi-jet events in the control region

is larger for ≥ 1 jets and decreases with increasing jet multiplicity.

4.5.1 QCD Model Validation for the $W(\to e\nu)+ \geq n$ jets sample

As already anticipated in section 4.4.2, a previous similar analysis [84] concluded that the energy of the *non-electrons* needs to be corrected in order to properly model the QCD multi-jet background. In this section we repeat the tests performed for the $W(\to e\nu) + 2$ jets channel on [84], and see if the same corrections need to be applied. Furthermore, we estimate the systematic uncertainty that each correction carries.

Energy corrections for QCD multi-jet model in the $W(\to e\nu)$ channel

The candidates for the *non-electron* model are defined by inverting the ID requirements reported in table 4.7 and by assuming that the failure of these requirements do not modify the kinematics of the object. However, some properties of the jets that fake the electrons are not reproduced. For example, by inverting the lateral energy sharing, L_{shr}, the energy modeled by the fake electron sample is more diffused in the shoulder towers. Although the isolation is still less than 0.1, as figure 4.14 shows, the energy of the fake electron defined by the *nonelectrons* cuts is almost entirely deposited outside the electromagnetic cluster that defines the electron.

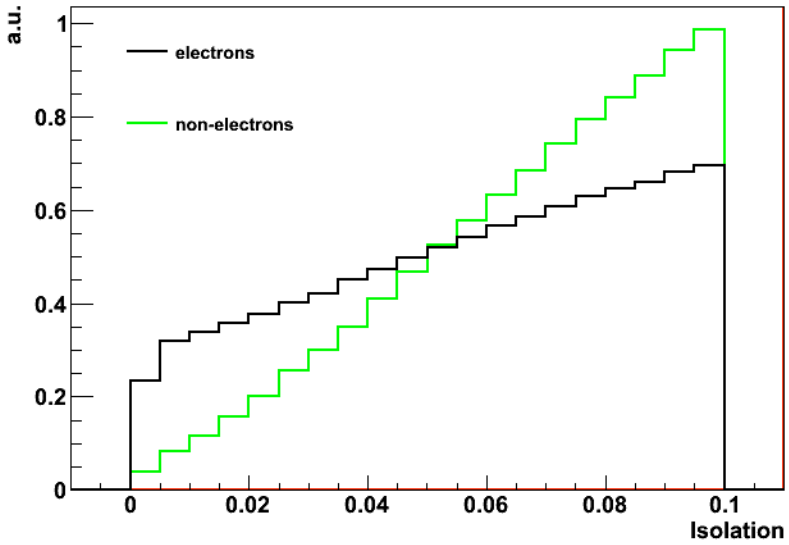


Figure 4.14: Relative isolation for the *non-electrons* and electrons in the control region for $W(\to e\nu)+ \geq 1$ jets sample. Distributions are normalized at the same area.

To account for these spread energy, the *non-electrons* are modeled by using the jet in the $\eta - \phi$ space that contains ($\Delta R \leqslant 0.4$) the *non-electron* candidate.

The "non-electron" energy scale correction for the $W(\to e\nu)$ channel

In the $W(\to e\nu) + 2$ jets channel, a correction to the energy of the *non-electrons* is derived by the comparison to the real mis-identified electrons using QCD dijet Monte Carlo samples generated with PYTHIA 6.203 Tune A [31]. Each sample is

generated with different minimum exchanged parton transverse momentum ($\hat{p}_T$) in the range 18 to 230 GeV/c. The limitations of this method stem from the fact that the final QCD sample energy spectrum is shifted to higher energies than those of ours experimental data sample. For this reason and because the statistics would not be sufficient, these dijet MC samples cannot be used to build the QCD multi-jet background. The statistics are nevertheless sufficient to be useful for the study described. Because the sample contains only jets we can isolate events in which a jet satisfies all the tight central electron requirements and events with jets that satisfies the TCE kinematical requirements but fail two of the ID cuts of table 3.2. We refer to the former with the name *misidentified electrons* (mis-TCE) and the latter as simulated *non-electrons* (non-TCE). The *misidentified electrons* correspond to our simulated QCD multi-jet background so, by comparing the non-TCE to them it is possible to identify defects of the data-driven model. In practice, we build a mis-TCE sample by requiring all the events to pass the analysis selection summarized on section 4.3.3 and a non-TCE sample by requiring the same selection but a *non-electron* instead of an electron. We do not apply the QCD-Veto so as to obtain a sample with significant statistics. Since both the mis-TCE and the non-TCE are generated by Monte Carlo the trigger does not influence the kinematics of the samples.

The main defect of the *non-electron* model that was observed using the MC sample is that the non-TCE transverse energy (E_T) spectrum is on average lower than the mis-TCE E_T spectrum. This discrepancy is observable in figure 4.15, were E_T distributions of the mis-TCE and the non-TCE samples are compared.

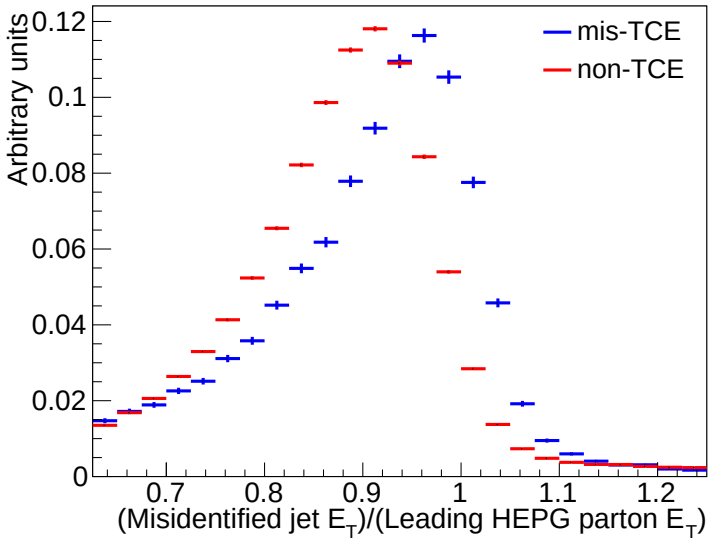


Figure 4.15: Transverse energy of the jet that is misidentified as an electron (blue curve) and as a *non-electron* (red curve) normalized to the energy of the leading HEPG parton that matches them ($\Delta R < 0.4$). Both the distributions are from the di-jet MC sample and are normalized to unitary area.

The E_T of both mis-TCE and non-TCE objects are scaled to the the energy of the leading HEPG parton that matches them in direction (*i.e.* ΔR(parton, mis/non-TCE) $<$ 0.4). In this way we attempt to minimize the effect of differenced in the energy spectra.

To make the *non-electron* energy scale (neES) consistent with the mis-TCE, we evaluate a correction factor as follows:

- we divide the mis-TCE and the non-TCE samples in 6 bins of transverse energy (E_T);
- for each bin we plot the distribution of $E_T^{non-TCE}/E_T^{parton}$ and $E_T^{mis-TCE}/E_T^{parton}$ both normalized to unit area (figure 4.15 is the example of the distribution for the entire sample);
- the correction factor K is the ratio of the mean values of the two distributions in each E_T bin:

$$K(E_T) = \frac{< E_T^{non-TCE}/E_T^{parton} >}{< E_T^{mis-TCE}/E_T^{parton} >} \tag{4.5.6}$$

Figure 4.16 shows the dependence of the correction K on the *non-electron* E_T. In black are reported the results previously obtained for the $W + 2$ jets and the parametrization used for the correction. We will refer to this correction as the nominal non-electron Energy Scale, neES, correction.

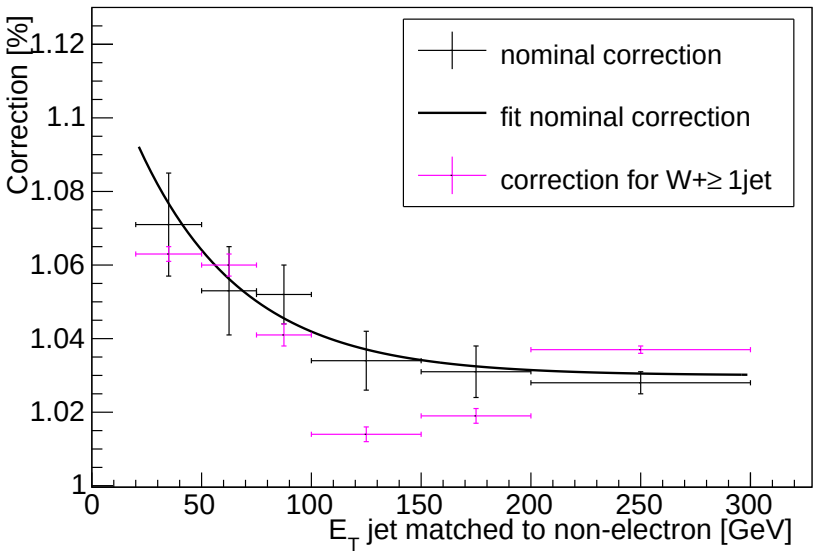


Figure 4.16: Corrections to the *non-electron* E_T as a function of its transverse energy. Black points and fit are from [80]. The fit is performed with an exponential functional form: $p_0 + \exp[p_1 \cdot (E_T - 20) + p_2]$. Pink points show the corrections for $W(\to e\nu)+ \geq$ 1 jets sample. No cut in transverse mass has been made for the estimation of the corrections.

The pink points in figure 4.16, are the correction factors evaluated for the $W(\rightarrow e\nu)+ \geq 1$ jets sample. For *non-electron* energy lower than 100 GeV the correction factor coincides within a 1σ-uncertainty with the nominal neES. Above 100 GeV the difference between the two correction factors is within a 3σ-uncertainty. The reason of the discrepancy can be the slight different definition of the selection phase space between the two corrections. In fact, the event selection applied to the QCD dijet sample in order to obtain the nominal correction includes an additional requirement in addition a minimum distance in the $\eta - \phi$ space of 0.7 between the two jets. Given that there are less than 2% of events with an electron that has $E_T > 100$ GeV in our data-driven QCD background, these discrepancies do not appreciably effect our distributions and we decided to keep using the nominal correction. After the correction the missing transverse energy ($\not\!\!E_T$) is recalculated.

We assign two systematic uncertainties to this correction. The first is obtained by shifting each correction factor up and down by $\pm 1\sigma$ and fitting them again. The two curves are reported in figure 4.17.

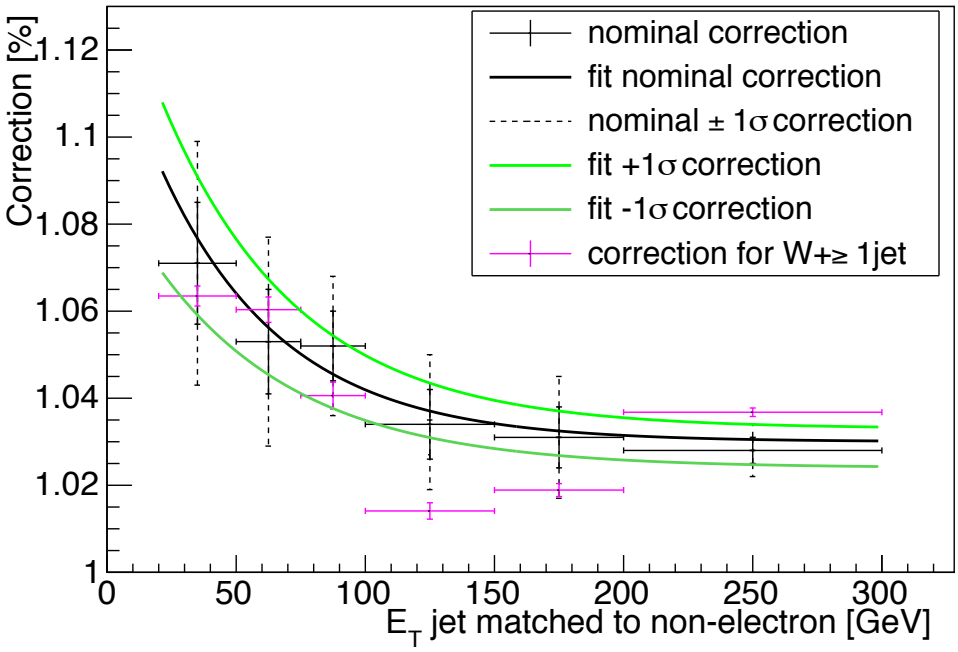
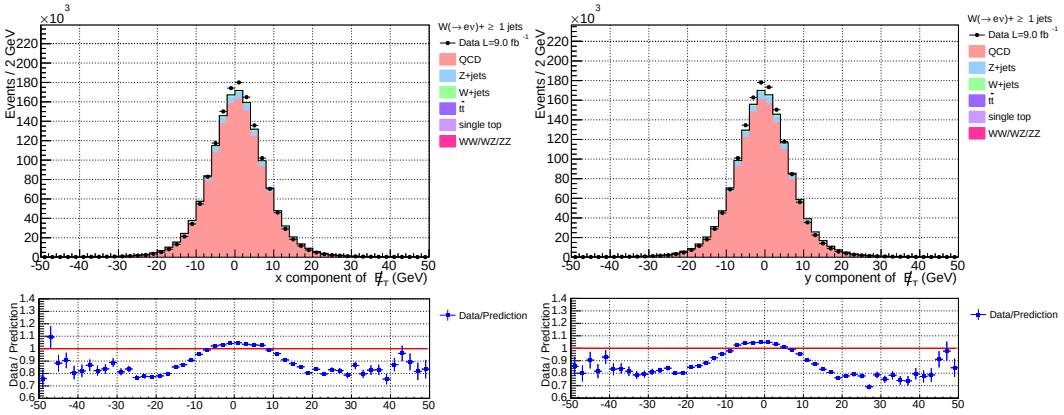


Figure 4.17: Corrections to the *non-electron* E_T as a function of its transverse energy together with the curves used to evaluate the systematic uncertainties. The fit is performed with an exponential functional form: $p_0 + \exp[p_1 \cdot (E_T - 20) + p_2]$ for both nominal and $\pm 1\sigma$ shifted corrections.

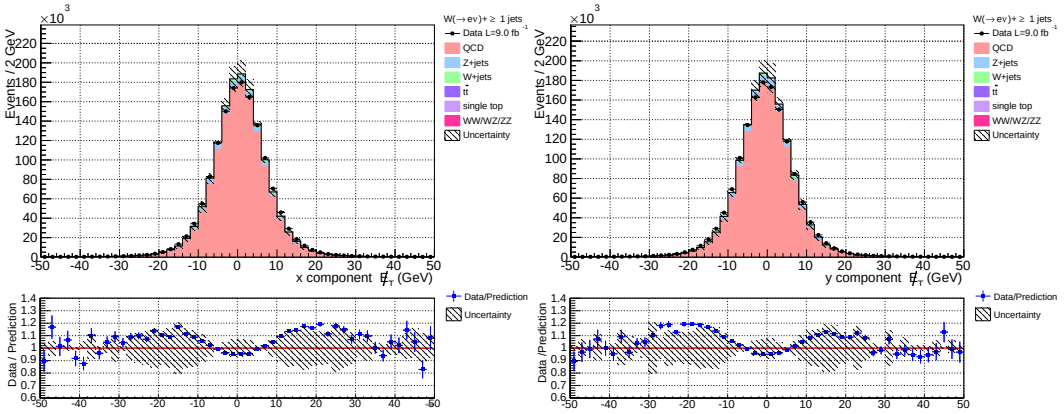
The second systematic uncertainty accounts for the assumption of using the mean values of the $E_T^{non-TCE}/E_T^{parton}$ and $E_T^{mis-TCE}/E_T^{parton}$ distribution. From this assumption it follows that the neES correction is an average correction because it does not consider the energy resolution. This effect is evident in the projections

on the x and y axis of the missing energy distribution as shown in figures 4.18 and 4.19 before and after applying the neES correction, respectively. In the lower pads of figure 4.19 a systematic uncertainty which has been evaluated by smearing each distribution with a gaussian 2.5 GeV wide is superimposed on the ratio between data and prediction.



(a) (b)

Figure 4.18: x (a) and y (b) projections of the missing transverse energy distribution before the neES correction for the $W(\to e\nu)+ \geq 1$ jets sample in the control region. The ratio between data and prediction is plotted in the lower pads.



(a) (b)

Figure 4.19: x (a) and y (b) projections of the missing transverse energy distribution after neES correction for the $W(\to e\nu)+ \geq 1$ jets sample in the CR. The systematic uncertainty obtained by smearing is shown in the lower pads superimposed on the ratio between data and prediction.

The Trigger Bias correction for the $W(\to e\nu)+ \geq n$ jets sample

The distribution of the electron transverse energy in the control region for the QCD multi-jet background after the neES correction is shown in figure 4.20, where

it is summed to other backgrounds, and compared to data. The discrepancies are principally due to the fact that non-TCEs contain a smaller fraction of EM energy than TCEs on account of the inversion of ID cuts [84]. As a result, as is evident in figure 4.20, the trigger (which requires that the electron candidate has at least one seed tower of 8 GeV or a cluster of 7.5 GeV in the electro-magnetic calorimeter) sculpts the distribution and this effect needs to be corrected for.

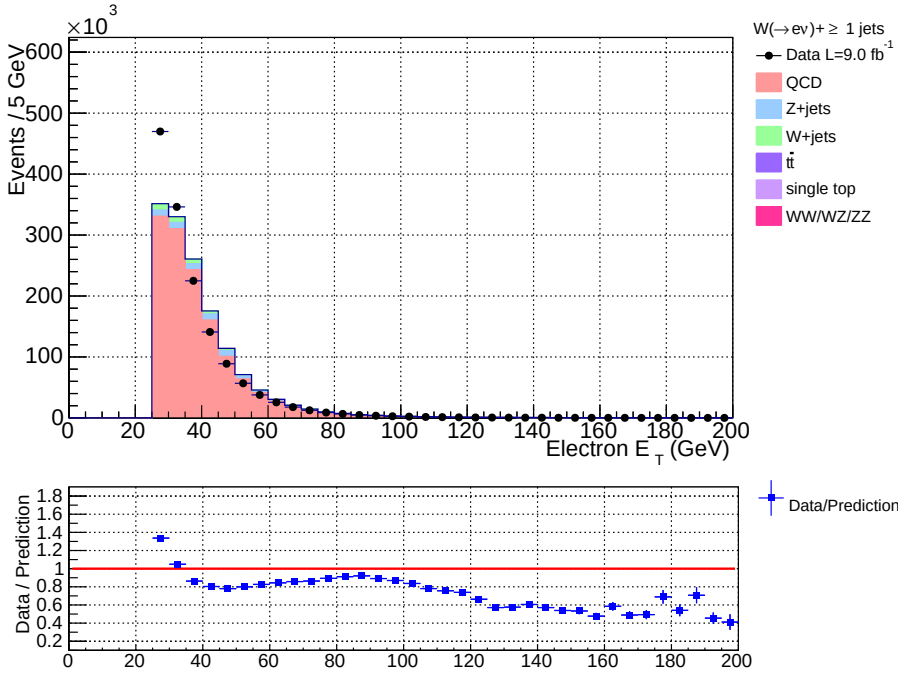


Figure 4.20: The electron transverse energy distribution for the $W(\rightarrow e\nu)+ \geq 1$ jets sample in the control region before the trigger bias correction. It is evident that the trigger is sculpting the *non-electron* distribution.

The correction consists in reweighing the *non-electron* E_T QCD template bin by bin with a weight defined as:

$$w_{TB}(bin\ E_T) = \frac{N_{DATA}(bin\ E_T) - N_{EWK}(bin\ E_T)}{N_{MJ}(bin\ E_T)}\Big|_{CR} \tag{4.5.7}$$

The result is that the electron transverse energy templates for data and background coincide in the control region by construction (see figure 4.21). The same weights are then used in the signal region in the signal region.

The trigger bias weights depend on the choice of the control region. In order to evaluate the dependence of the weights on the choice of region we derived two sets of weights from two orthogonal control sub-regions: $M_T^W < 20$ GeV and 20 GeV $\leqslant M_T^W \leqslant 40$ GeV. The two sets of weights are than applied to the whole CR and the differences with respect to the nominal distribution are assigned as an

uncertainty to each weight. The weights and their uncertainties are evaluated and applied considering each exclusive jet multiplicity individually, as shown in figure 4.21.

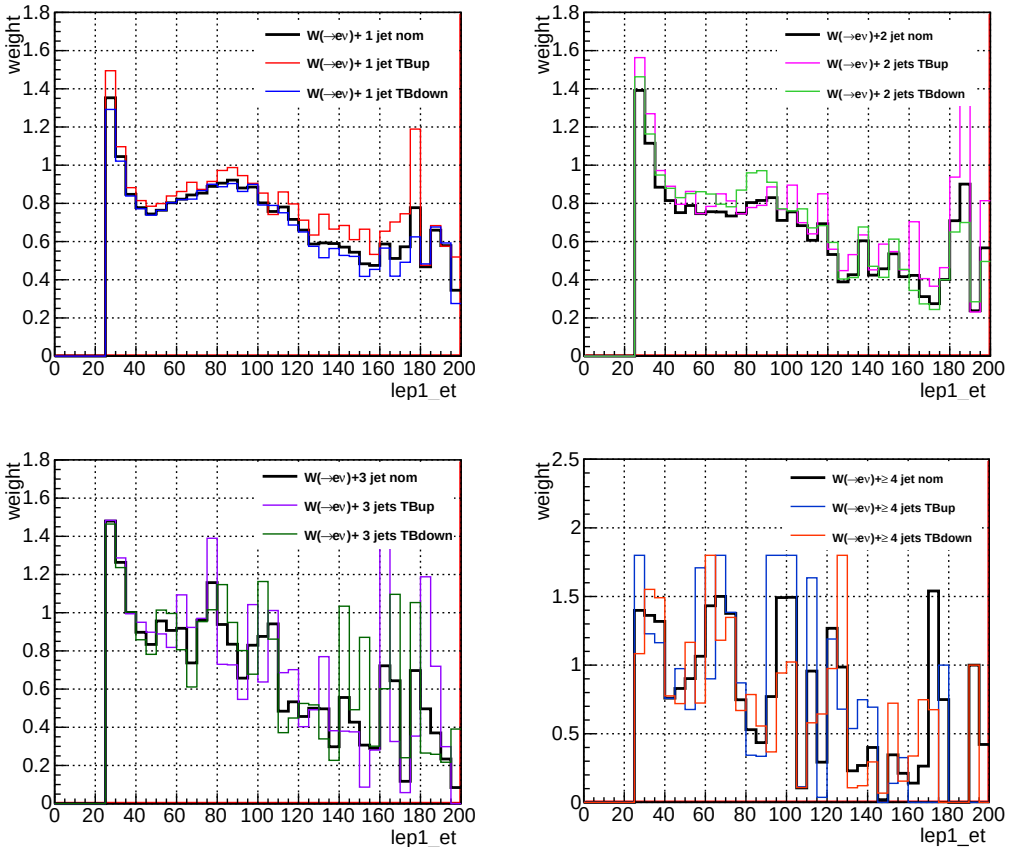


Figure 4.21: Trigger bias weights and their uncertainties for each exclusive jet multiplicity (clockwise from the top left: $W(\to e\nu) + 1\,\text{jets}$, $W(\to e\nu) + 2\,\text{jets}$, $W(\to e\nu) + 3\,\text{jets}$ and $W(\to e\nu) + \geq 4$ jets).

The lepton E_T distributions in the control region after the *non-electron* E_T reweighing to remove the bias caused by the trigger selection in the QCD multi-jet background model, for the inclusive $W(\to e\nu) + \geq 1$ jets samples and the systematic uncertainty resulting from the reweighting procedure are shown in figure 4.22.

Effects of the corrections $W(\to e\nu) + \geq n$ jets

The *non-electron* method can be summarized as follows:

1. build the QCD sample by applying the event selection summarized in section 1.4a but with an electron that satisfies all the kinematic requirements but fails two of the ID conditions (see table 1.4a);

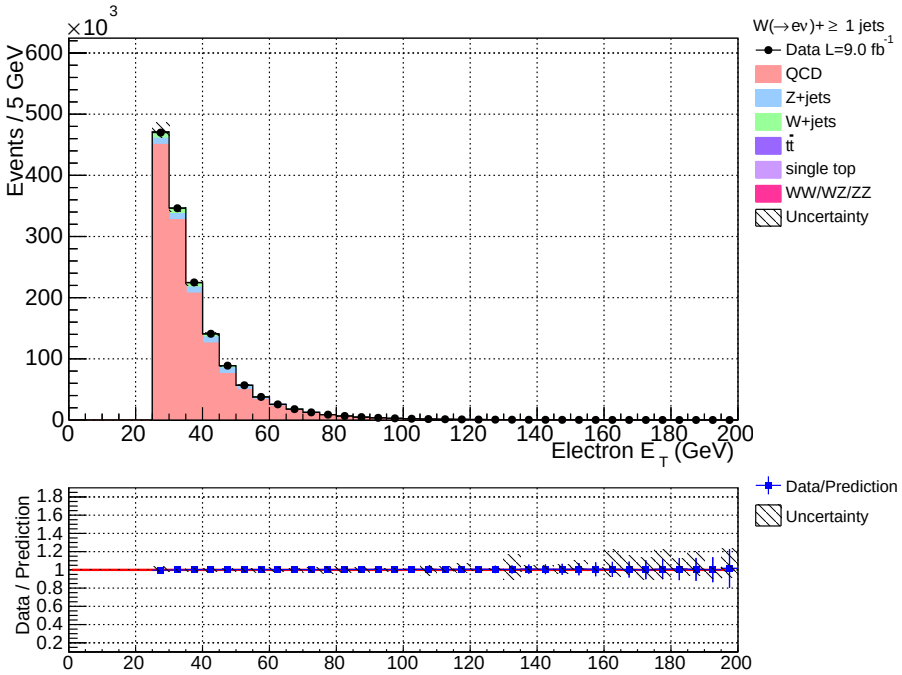
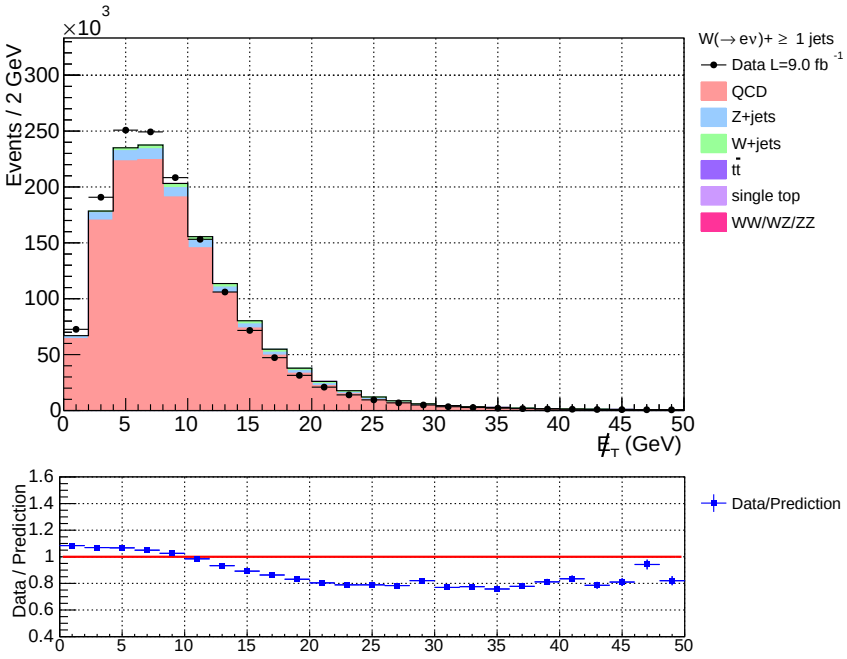


Figure 4.22: The electron energy distribution for the $W(\to e\nu)+ \geq 1$ jets sample in the control region after the trigger bias correction. Data and Prediction coincide by construction, the systematic uncertainty of the weights is superimposed on the ratio of data to prediction in the lower pad.

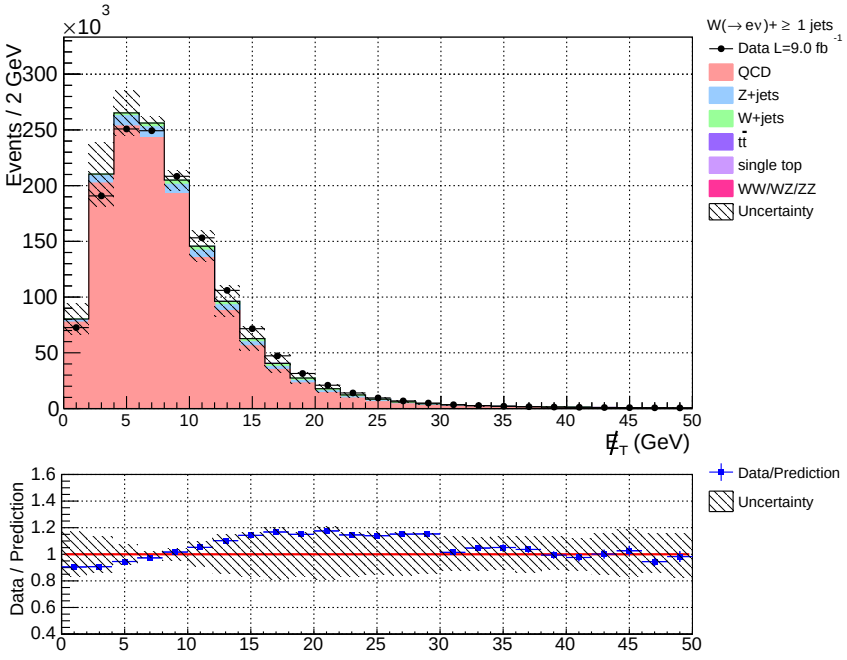
2. remove the EWK contamination from the sample;
3. use the jet that matches the *non-electron* to account for the energy around the electron cluster;
4. correct the *non-electron* E_T with the neES correction, and recalculate the $\not\!\!E_T$ of the event;
5. apply the trigger bias correction.

We investigate the effect of each assumption and correction in order to quote a systematic uncertainty for each of them. All the systematic uncertainties obtained for the QCD model are summed in quadrature. The trigger bias correction, the QCD multi-jet estimate and their systematic uncertainties are evaluated separately for each exclusive jet multiplicity.

Figures 4.23 – 4.31 compare the modeling of several fundamental distributions prior to and after applying the corrections for the $W(\to e\nu)+ \geq 1$ jets sample. In the lower pad of the corrected distributions the systematic uncertainty is superimposed on the ratio between data and predictions.

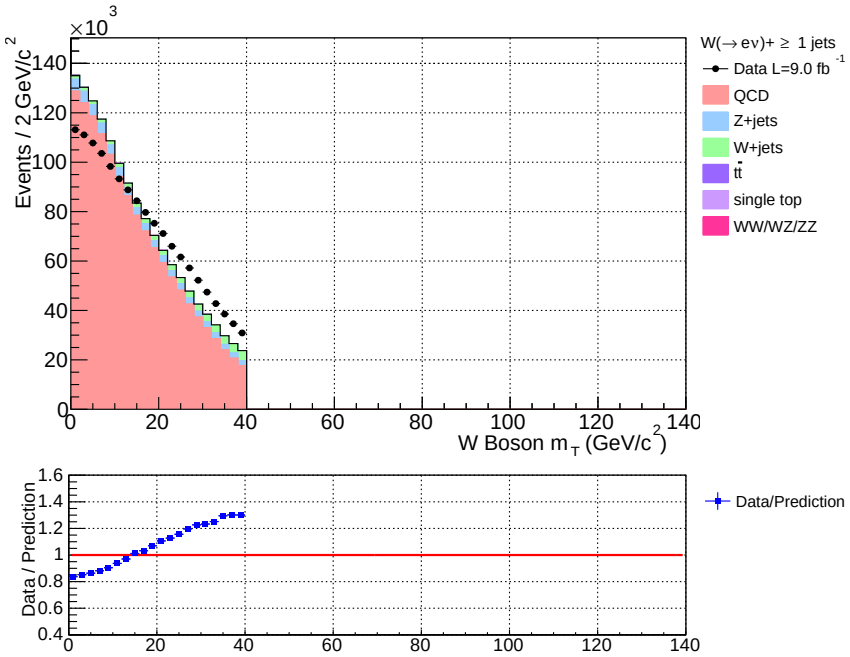


(a)

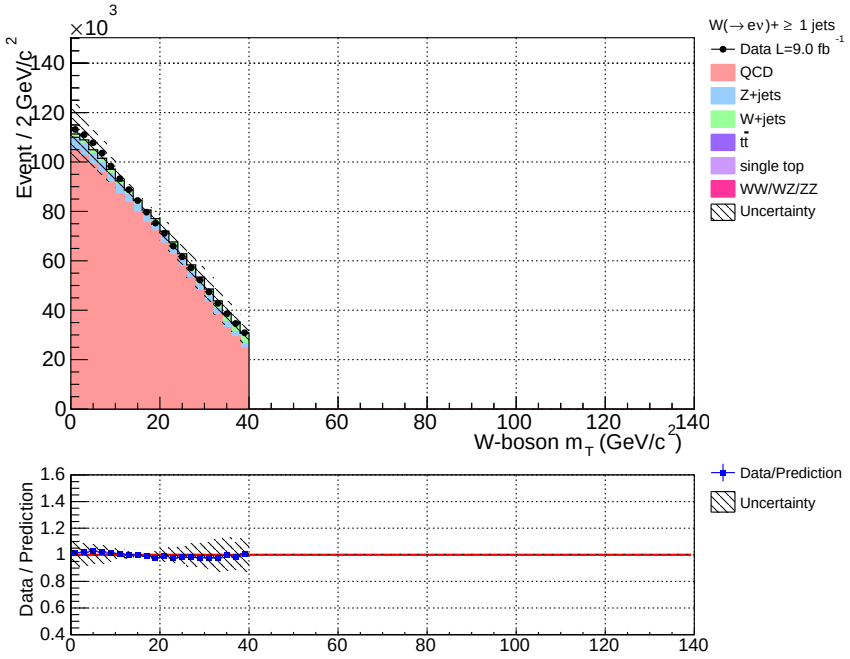


(b)

Figure 4.23: The missing energy distribution for the $W(\rightarrow e\nu)+ \geq 1$ jets sample in the control region as observed (a) before corrections and (b) after corrections (with systematic uncertainties). Lower pads illustrate the ratio between data and prediction.

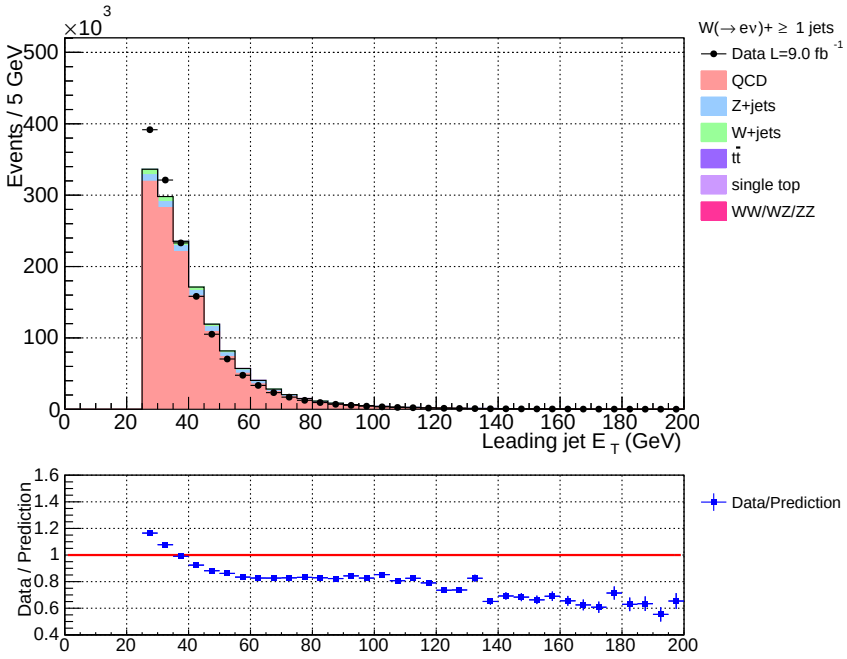


(a)

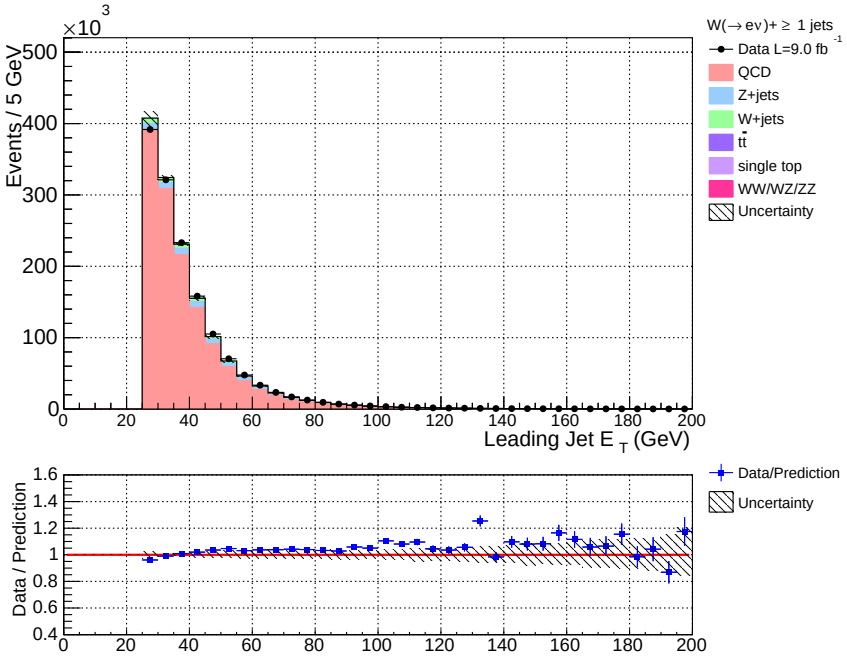


(b)

Figure 4.24: The W-boson transverse mass distribution for the $W(\rightarrow e\nu)+ \geq 1$ jets sample in the control region as observed (a) before corrections and (b) after corrections (with systematic uncertainties). Lower pads illustrate the ratio between data and prediction.

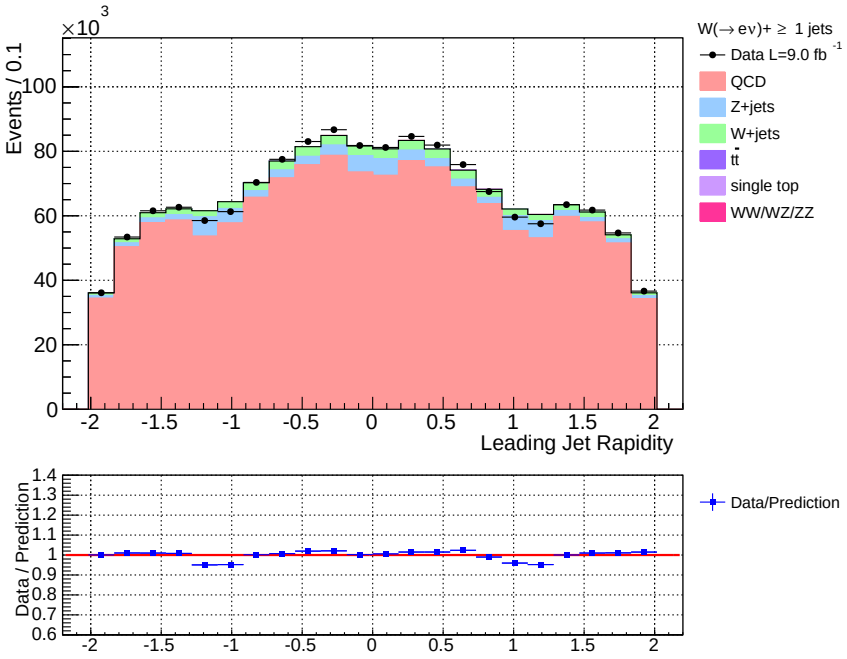


(a)

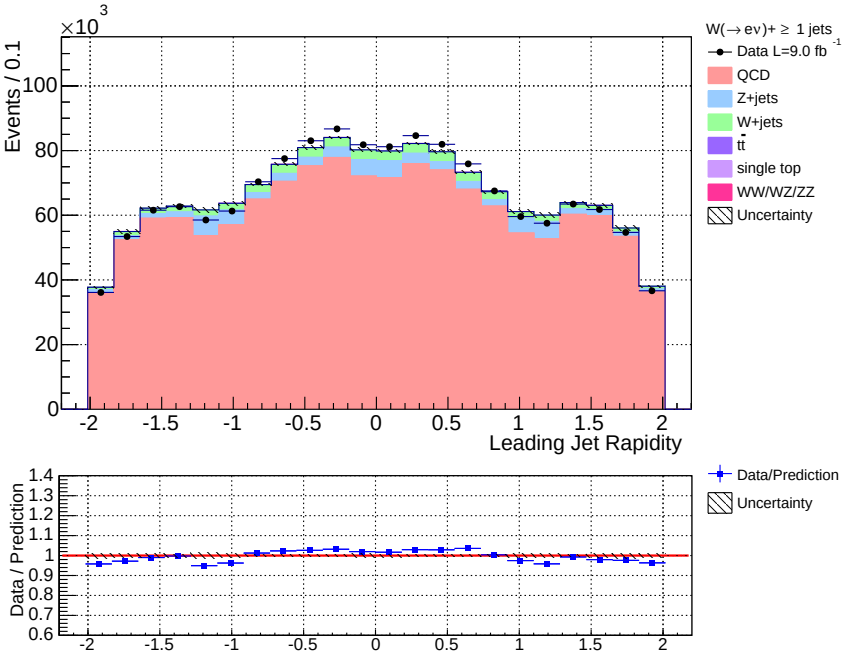


(b)

Figure 4.25: The leading jet transverse energy distribution for the $W(\rightarrow e\nu)+ \geq 1$ jets sample in the control region as observed (a) before corrections and (b) after corrections (with systematic uncertainties). Lower pads illustrate the ratio between data and prediction.

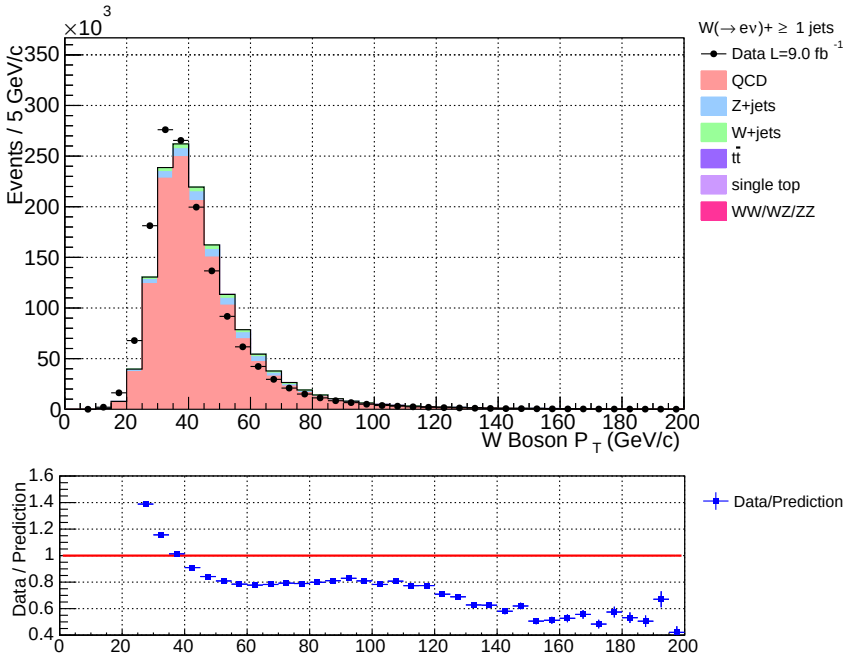


(a)

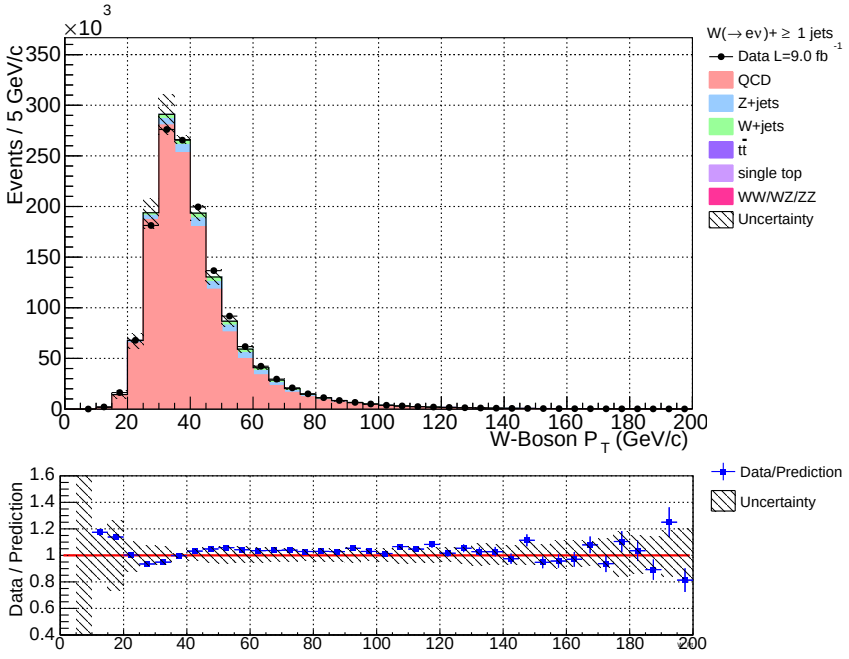


(b)

Figure 4.26: The leading jet rapidity distribution for the $W(\rightarrow e\nu)+ \geq 1$ jets sample in the control region as observed (a) before corrections and (b) after corrections (with systematic uncertainties). Lower pads illustrate the ratio between data and prediction.

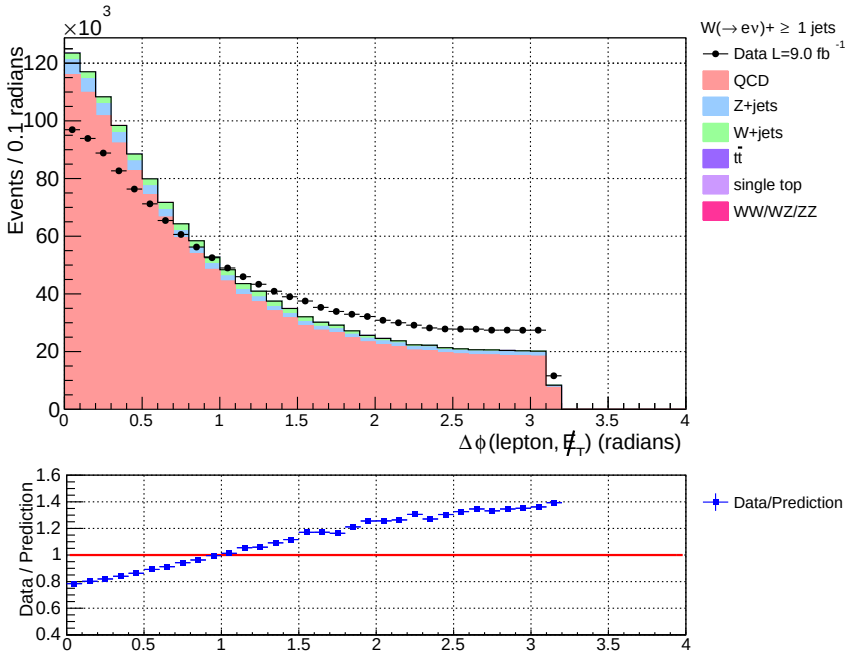


(a)

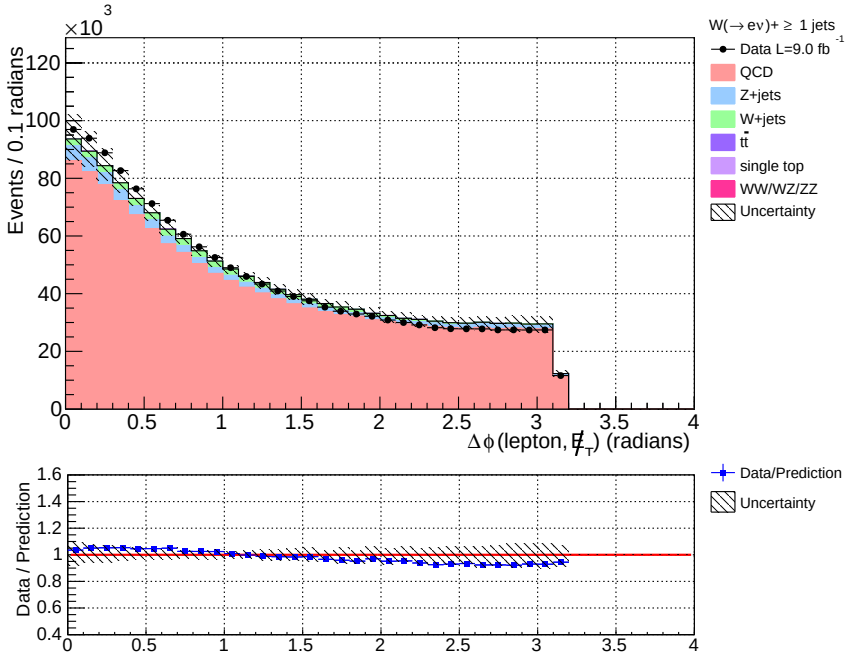


(b)

Figure 4.27: The W-boson transverse momentum distribution for the $W(\rightarrow e\nu)+ \geq 1$ jets sample in the control region as observed (a) before corrections and (b) after corrections (with systematic uncertainties). Lower pads illustrate the ratio between data and prediction.

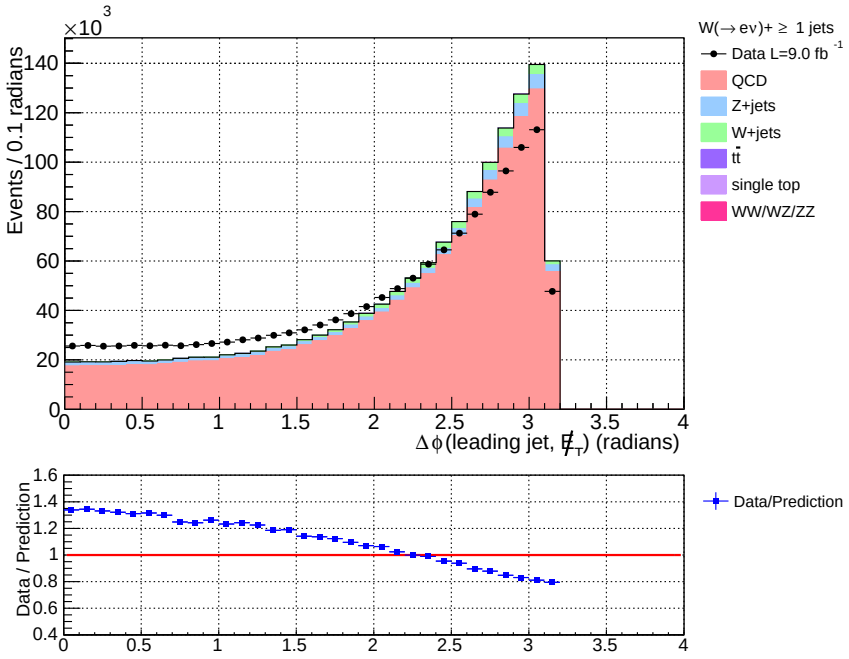


(a)

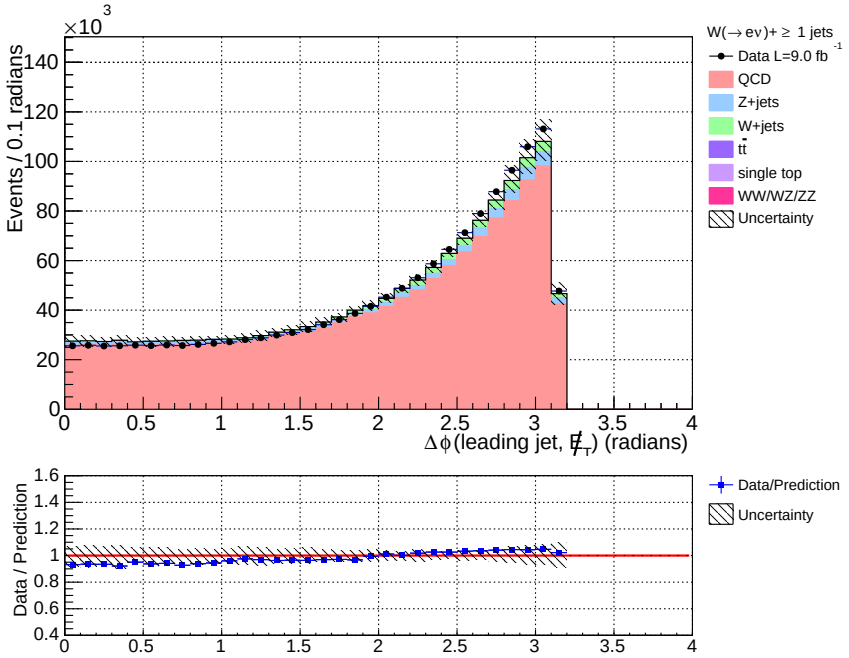


(b)

Figure 4.28: The angle between the electron and the $\not{E}_T$ distribution for the $W(\rightarrow e\nu)+ \geq 1$ jets sample in the control region as observed (a) before corrections and (b) after corrections (with systematic uncertainties). Lower pads illustrate the ratio between data and prediction.

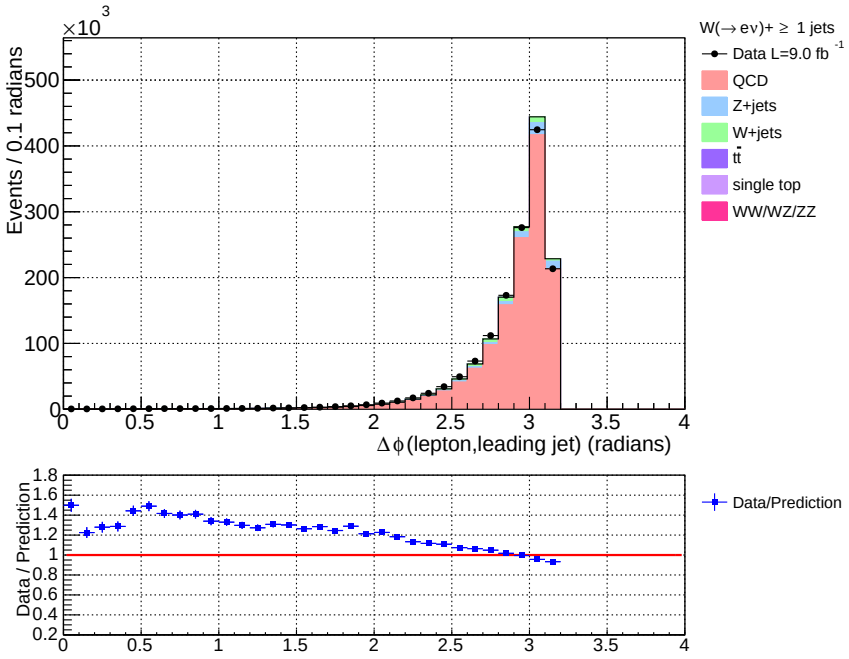


(a)

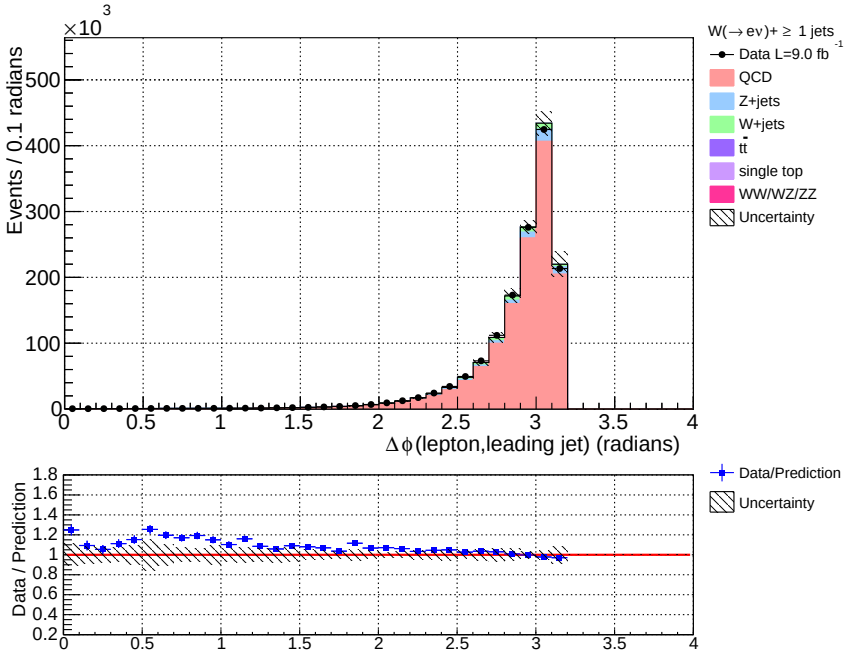


(b)

Figure 4.29: The angle between the leading jet and the $\not{E}_T$ distribution for the $W(\rightarrow e\nu)+ \geq 1$ jets sample in the control region as observed (a) before corrections and (b) after corrections (with systematic uncertainties). Lower pads illustrate the ratio between data and prediction.

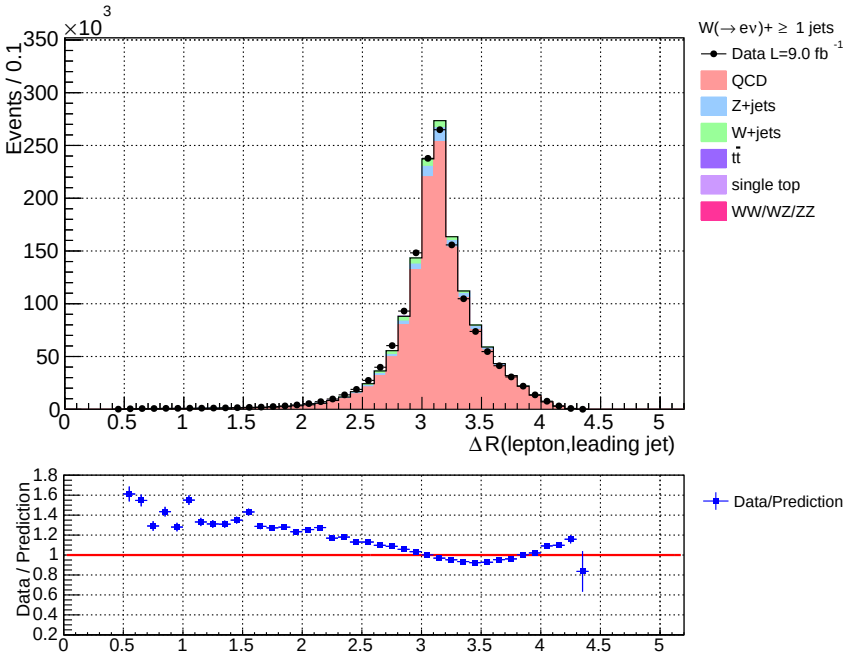


(a)

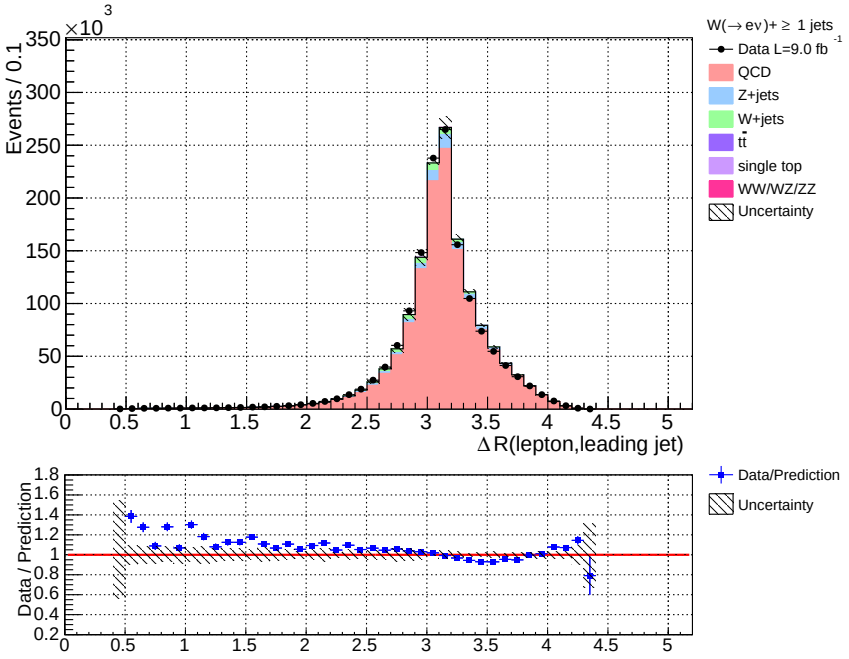


(b)

Figure 4.30: Distribution of the angle between the leading jet and the electron for the $W(\rightarrow e\nu)+ \geq 1$ jets sample in the control region as observed (a) before corrections and (b) after corrections (with systematic uncertainties). Lower pads illustrate the ratio between data and prediction.



(a)



(b)

Figure 4.31: Distribution of the ΔR between the leading jet and the $\not\!\!E_T$ for the $W(\rightarrow e\nu)+ \geq 1$ jets sample in the control region as observed (a) before corrections and (b) after corrections (with systematic uncertainties). Lower pads illustrate the ratio between data and prediction.

Angular Discrepancy between leading jet and *non-electron*

These results show that the *non-electron* model reproduces well the QCD multi-jet background in the control region for all the inclusive $W(\rightarrow e\nu)+ \geq n$ jets channels studied in this analysis. The few remaining discrepancies are absorbed by the systematic uncertainties. In figures 4.30b and 4.31b, a residual discrepancy which is not covered by systematic uncertainty, is visible in the low-statistics tails. From further studies performed using the di-jet MC sample we observed the same discrepancy between the angles of the *misidentified-electron* and *non-electron* with respect to the leading jet. The discrepancy is illustrated in figure 4.32 where the distribution of the ΔR between *non-electron* and the leading jet is compared with that of the ΔR between *misidentified-electron* and the leading jet.

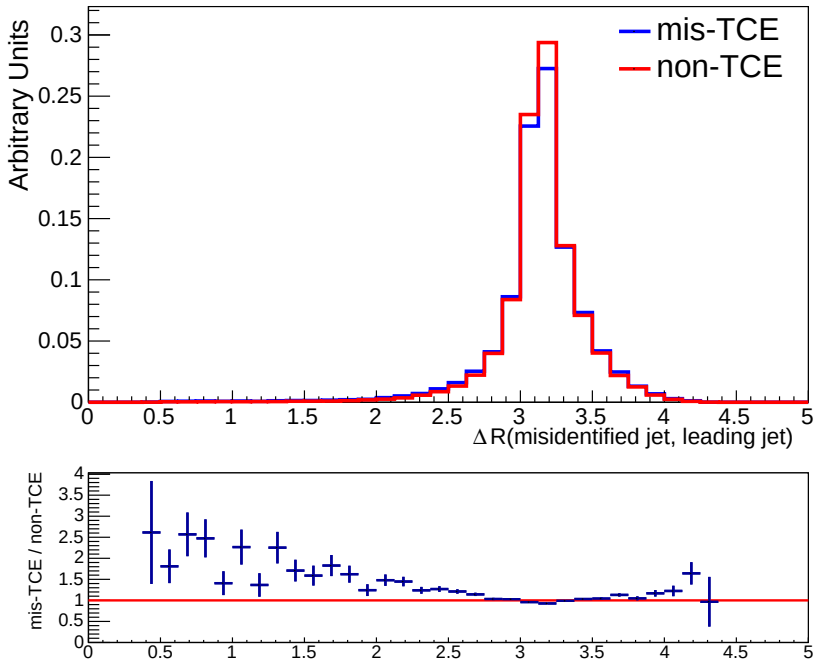
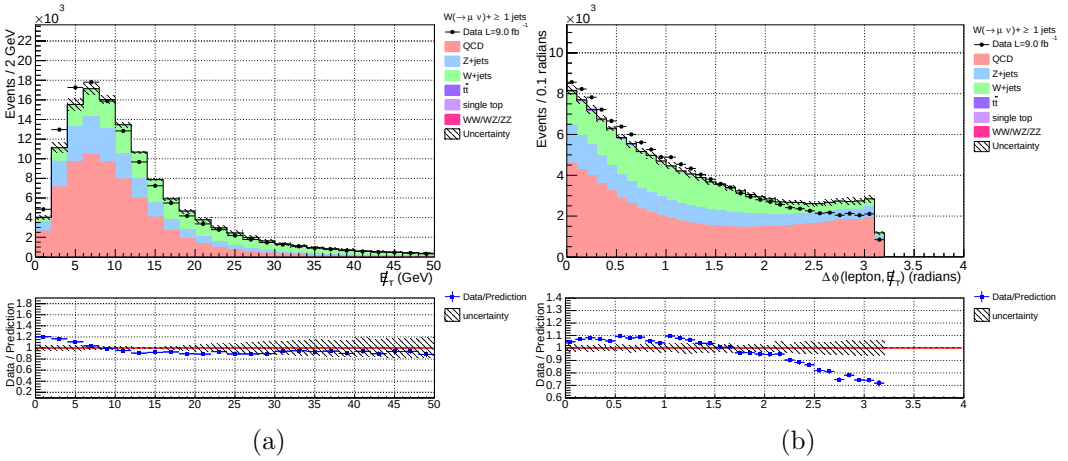


Figure 4.32: ΔR between *misidentified-electron* and the leading jet (blue curve) with ΔR between *non-electron* and the leading jet (red curve) superimposed. Both the distributions are obtained using the MC sample and are normalized to unit area.The lower pad reports the bin-by-bin mis-TCE/non-TCE ratios.

Since the discrepancy affects only the low-statistic tail of these two distributions, it leads to no significant effect on the analysis so we decided to not account for it as correction or as a systematic uncertainty.

4.5.2 QCD Model Validation for the $W(\rightarrow \mu\nu)+ \geq n$ jets sample

In the muon sample the QCD multi-jet background is modeled using *non-isolated muons*, *i.e.*, muons which pass all selections criteria of table 3.3 except the requirement of lepton isolation. The assumption underlying this model is that the jets that are more likely to fake the muons are jets that contain energetic muons and these are typically *non-isolated muons*. As for the $W(\rightarrow e\nu)$ channel, we checked if the *non-isolated muon* model reproduced the kinematics of the QCD multi-jet background by comparing the most important kinematic variable distributions in the control region. Distributions with respect to missing transverse energy and the angle between MET and the muon for $W(\rightarrow \mu\nu)+ \geq 1$ jet in the control region, for *non-isolated muons* are shown in figure 4.33.



(a) (b)

Figure 4.33: (a) Missing transverse energy and (b) the angle between the MET and the muon distributions for $W(\rightarrow \mu\nu)+ \geq 1$ jet in the control region.The multi-jet background is modeled with *non-isolated muons*. The systematic uncertainties assigned to the rate estimates are also shown.

Figure 4.33 clearly shows that the non-isolated muon model does not correctly reproduce the fake muon kinematics and that the systematic uncertainty in the rate does not cover all the discrepancies. A more reliable QCD multi-jet model is obtained by replacing the *non-isolated muons* with *almost-isolated muons* in which the jets that fake the muons are modeled with objects that pass all the muon requirements but with isolation between 0.1 and 0.2. Basically, the *almost-isolated muon* sample is a sub-sample of the *non-isolated muons*, as shown in figure 4.34, where isolation *vs.* transverse mass of the W boson is shown. Events with isolation greater than 0.2 (shaded area in figure 4.34) are used to build a systematic uncertainty.

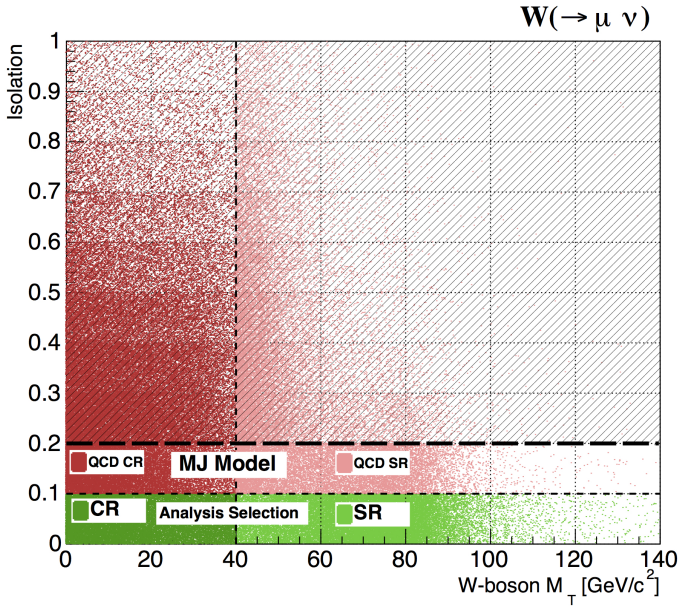


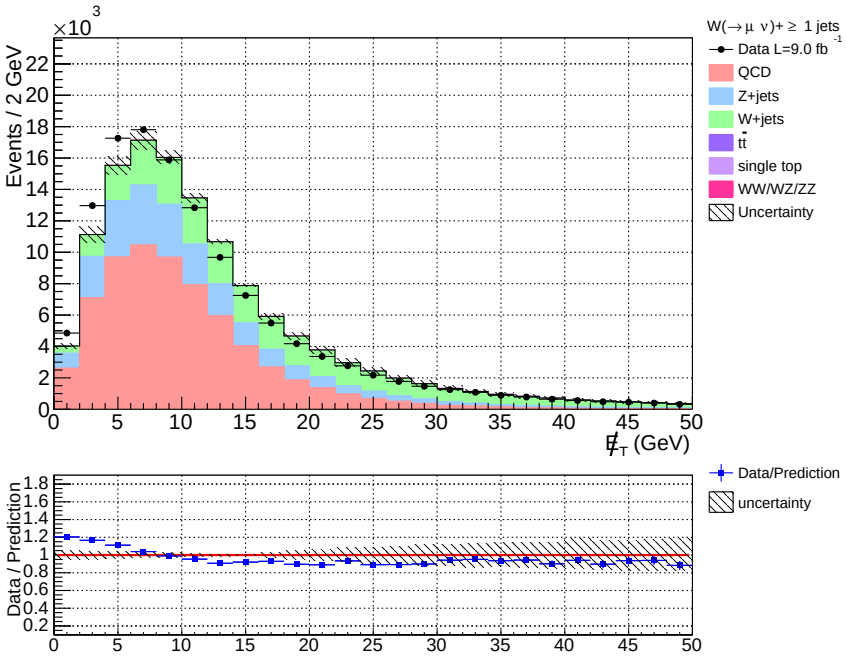
Figure 4.34: Two dimensional plot of the events used in this analysis for the $W \to \mu\nu$ channel. Events with isolation between 0.1 and 0.2 are used for the QCD multi-jet model. Events with isolation higher than 0.2 (shaded region) are used to build a shape systematic uncertainty to account for the residual dependence on the isolation.

The cost of using the *almost-isolated muons* is a deterioration in statistical accuracy. Table 4.12 reports the values of QCD scale factors f^{μ}_{QCD} defined by equation 4.5.5, for each jet multiplicity. Comparing them with the QCD scale factors for the electron sample, f^{e}_{QCD}, of table 4.12 we see that the statistical uncertainty of the contribution used to model the QCD multi-jet background in the $W \to \mu\nu$ channel is still a factor ~ 2 larger than the one used for the $W \to e\nu$ channel.

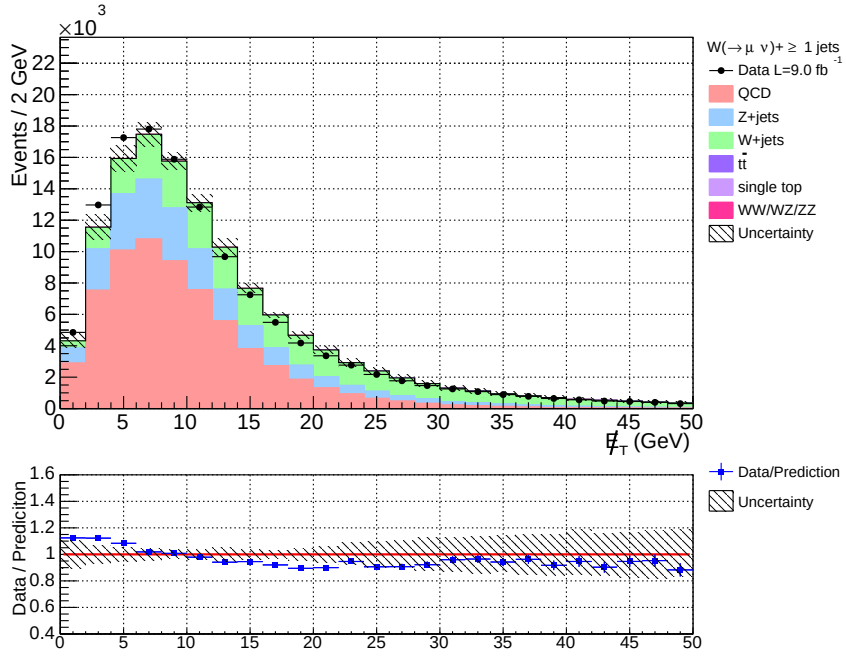
CR $W \to \mu\nu$	$+1$ jet	$+2$ jets	$+3$ jets	$+ \geq 4$ jets
$\mathbf{f^{\mu}_{QCD}}$	$\mathbf{0.80 \pm 0.12}$	$\mathbf{0.67 \pm 0.10}$	$\mathbf{0.79 \pm 0.15}$	$\mathbf{0.58 \pm 0.10}$

Table 4.12: QCD scale factors f^{μ}_{QCD} for the $W(\to \mu\nu) + n$ jets calculated using the *almost-isolated muons* sample with equation 4.5.2. These factors are going to be used to evaluate the QCD multi-jet yields in the SR.

Despite the deterioration in statistics, *almost-isolated muon* model reproduces fake muon kinematics with systematic uncertainties. This can be seen by comparing the distributions for which the multi-jet background is built with *non-isolated muons* (top figure) with the same distributions using *almost-isolated muon* models (bottom figure). The choice of using muons with isolation between 0.1 and 0.2 is the best compromise between statistics and model quality.

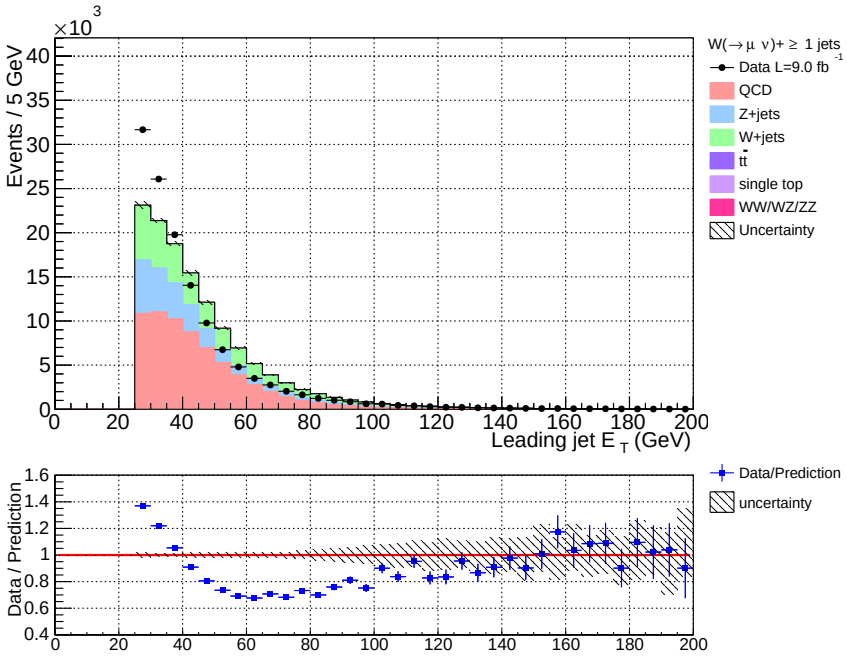


(a)

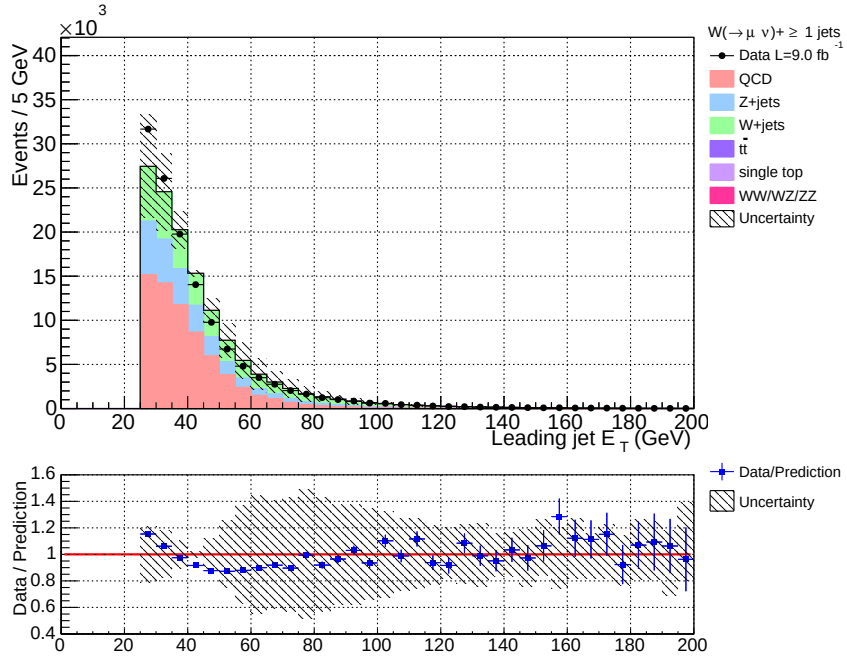


(b)

Figure 4.35: Distributions with respect to (a) missing transverse energy and (b) the angle between the MET and the muon for $W(\rightarrow \mu\nu)+ \geq 1$ jet in the control region. Multi-jet background is modeled with *almost-isolated muons*. Systematic uncertainties include those relative to the choice of isolation cut.

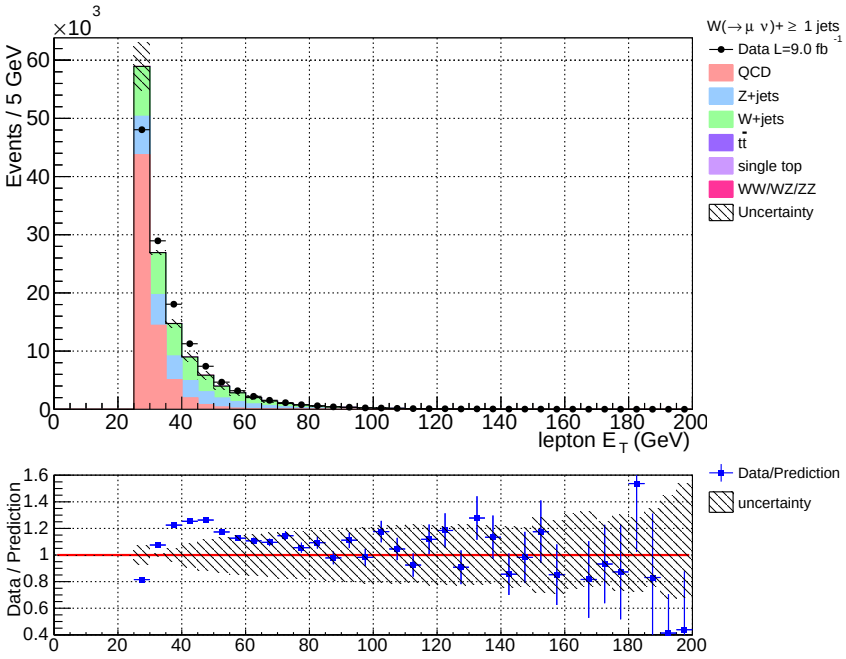


(a)

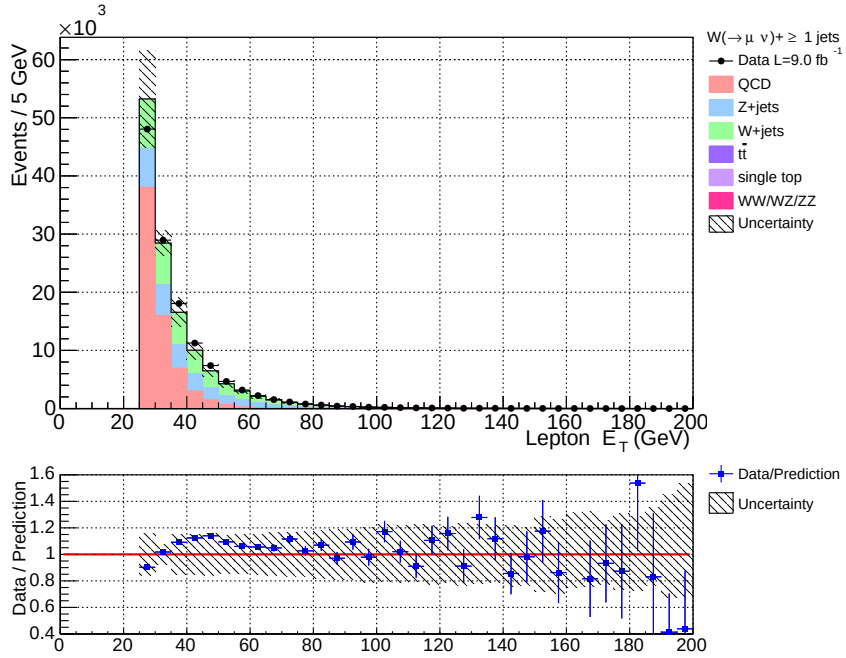


(b)

Figure 4.36: The leading jet E_T distributions for $W(\rightarrow \mu\nu)+ \geq 1$ jet in the control region. The multi-jet background is modeled with (a) *non-isolated muons* (b) *almost-isolated muons*. In (b) systematic uncertainties include those relative to the choice of the isolation cut.

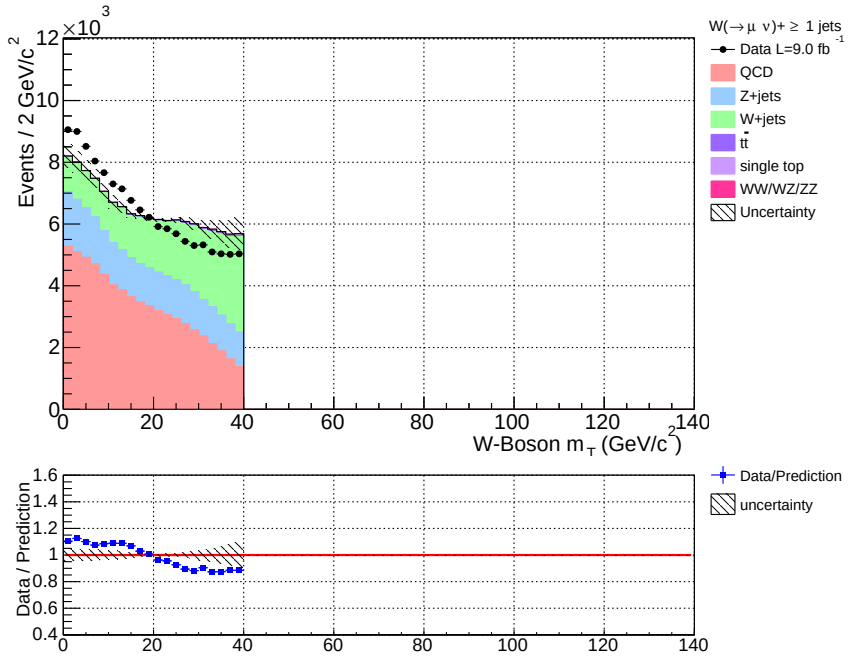


(a)

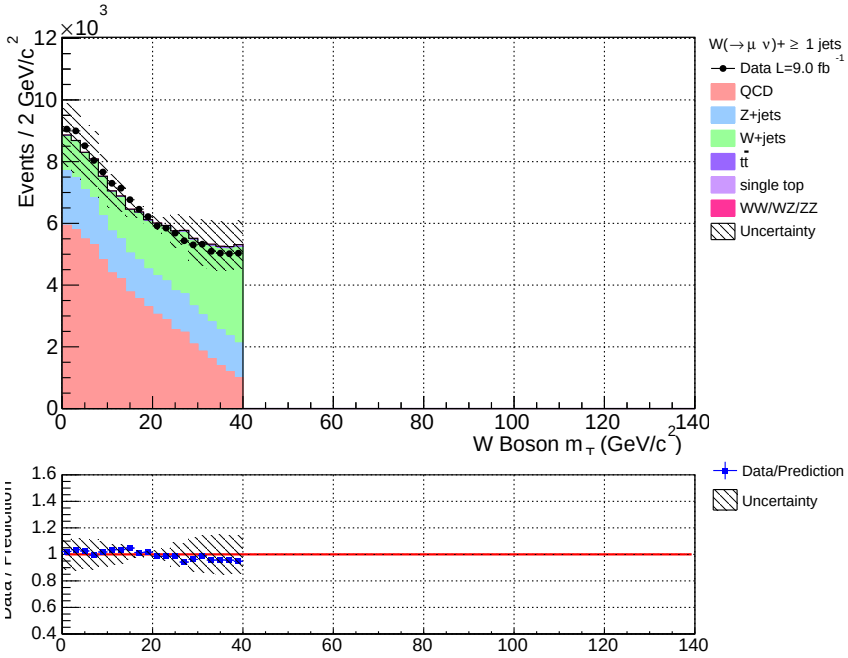


(b)

Figure 4.37: The transverse energy distribution of the muon for the $W(\rightarrow \mu\nu)+ \geq 1$ jets sample in the control region in which the QCD multi-jet background is modeled with (a) *non-isolated muons* (b) *almost-isolated muons*. Lower pads illustrate the ratio between Data and Prediction.

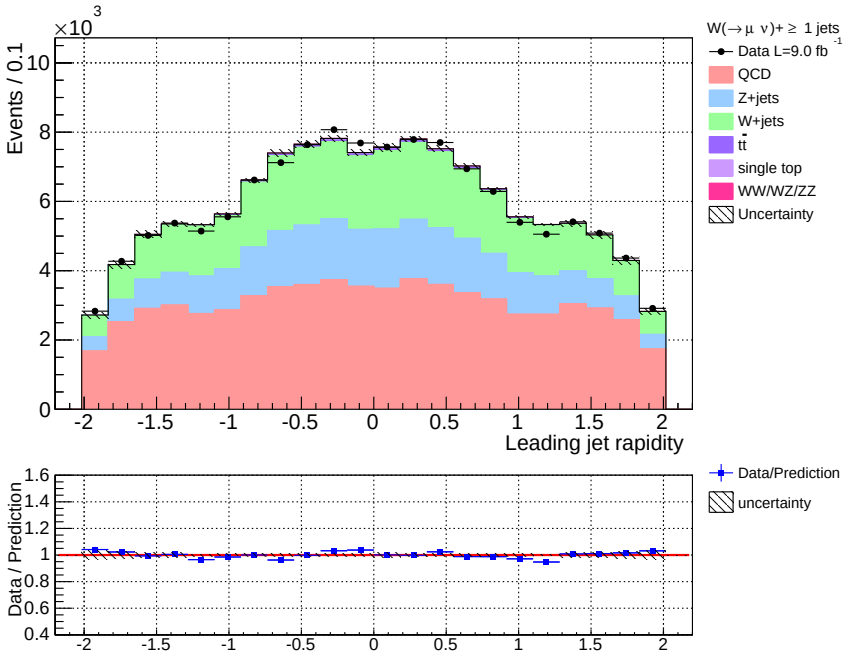


(a)

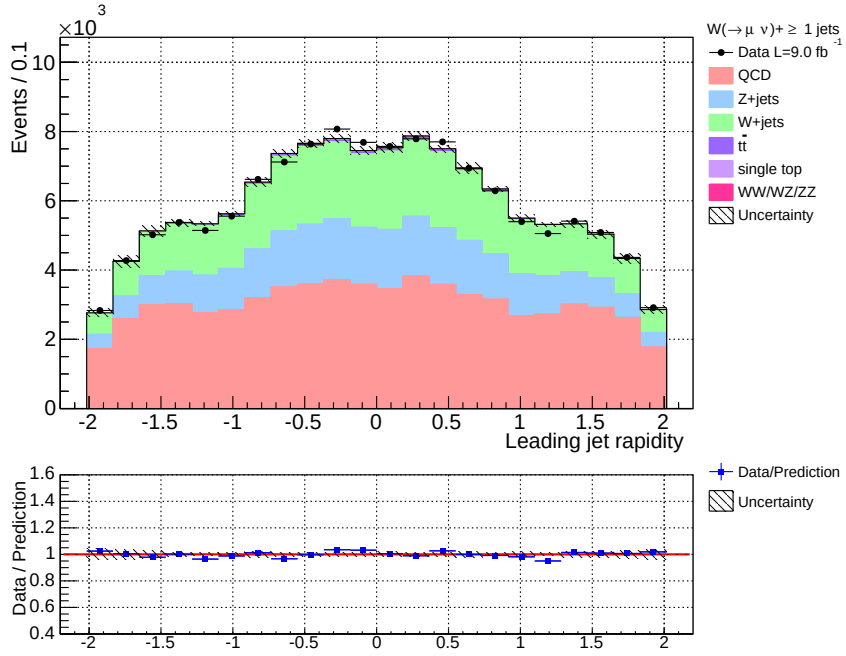


(b)

Figure 4.38: The W-boson transverse mass distribution for the $W(\rightarrow \mu\nu)+ \geq 1$ jets sample in the control region in which the QCD multi-jet background is modeled with (a) *non-isolated muons* (b) *almost-isolated muons*. Lower pads illustrate the ratio between Data and Prediction.

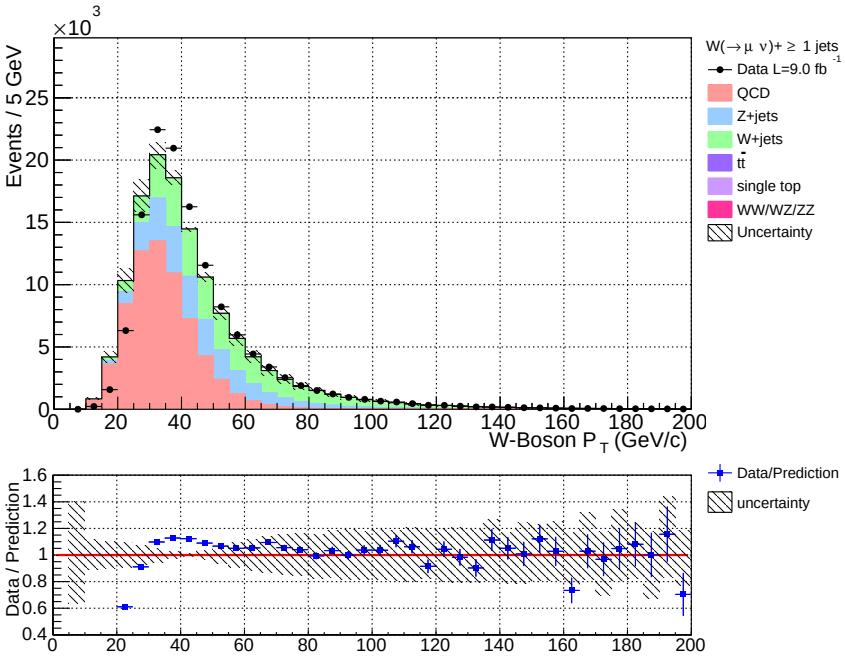


(a)

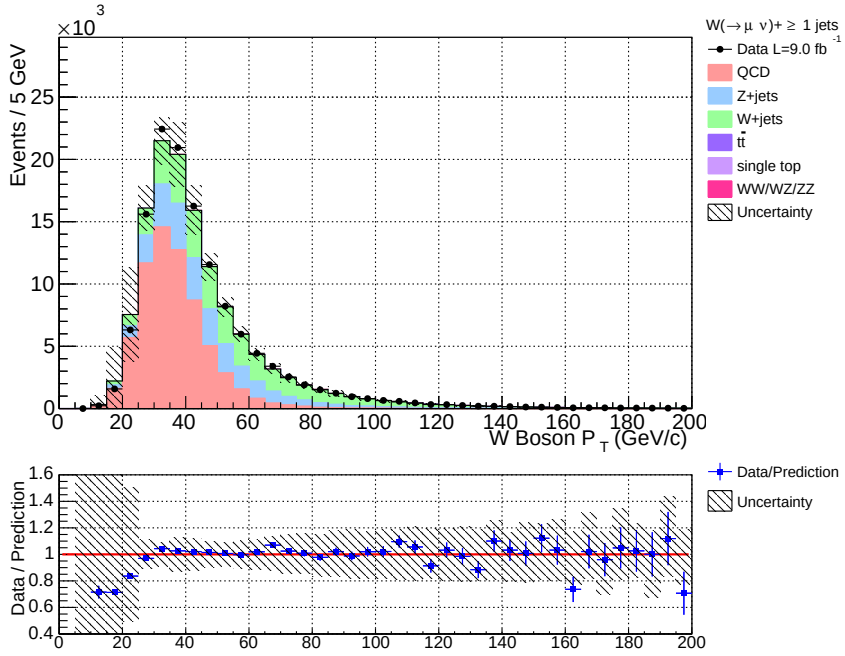


(b)

Figure 4.39: The leading jet rapidity distribution for the $W(\rightarrow \mu\nu)+ \geq 1$ jets sample in the control region in which the QCD multi-jet background is modeled with (a) *non-isolated muons* (b) *almost-isolated muons*. Lower pads illustrate the ratio between Data and Prediction.

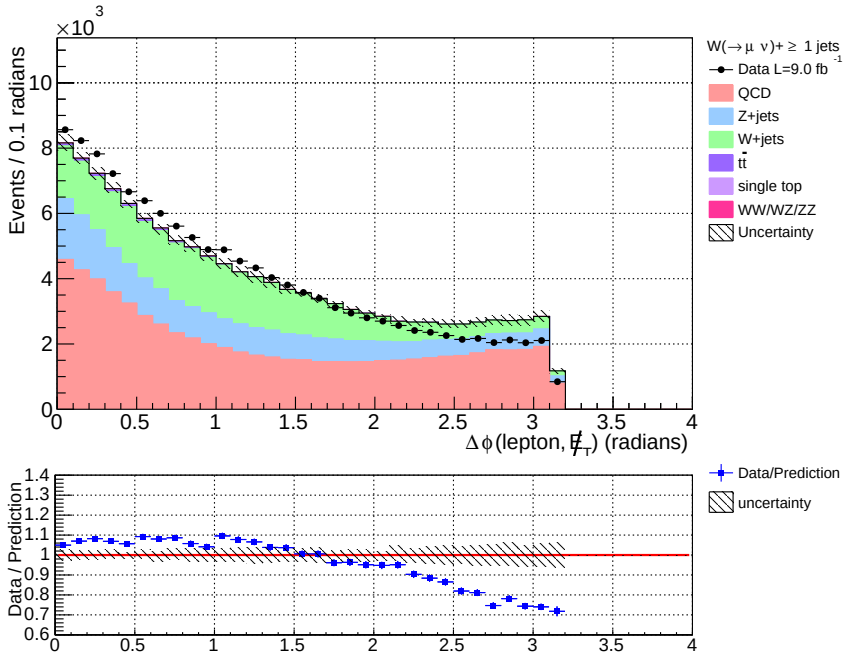


(a)

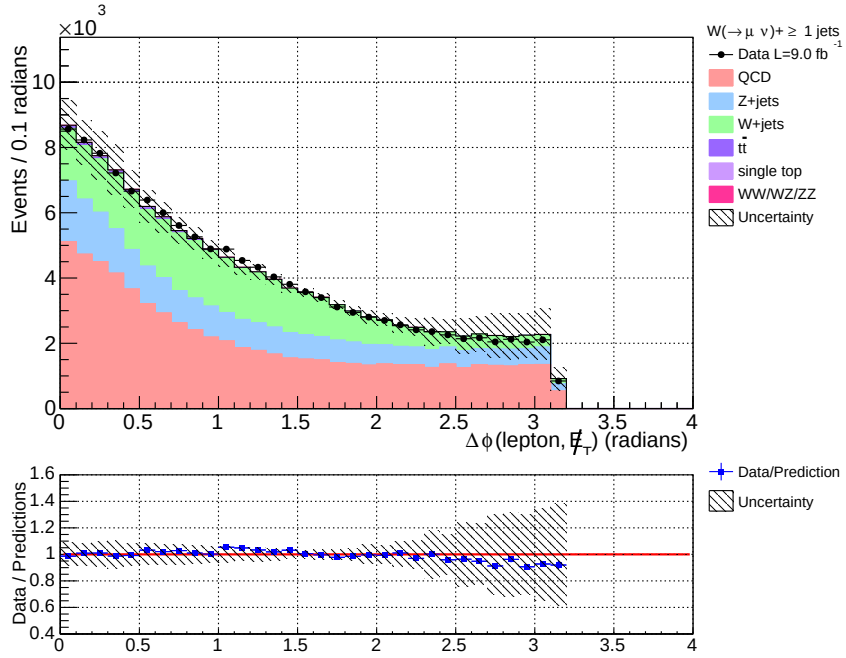


(b)

Figure 4.40: The W-boson transverse momentum distribution for the $W(\rightarrow \mu\nu)+ \geq 1$ jets sample in the control region in which the QCD multi-jet background is modeled with (a) *non-isolated muons* (b) *almost-isolated muons*. Lower pads illustrate the ratio between Data and Prediction.

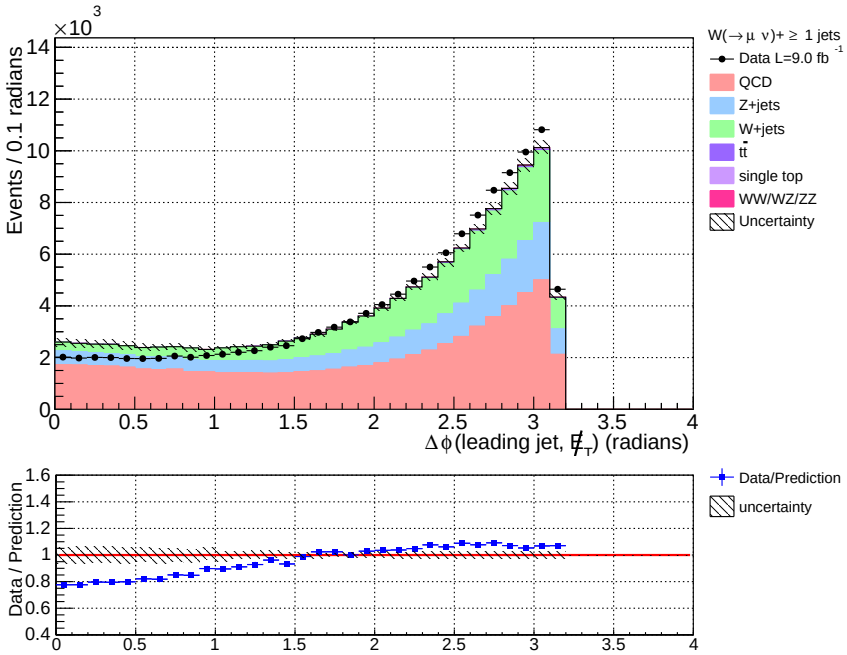


(a)

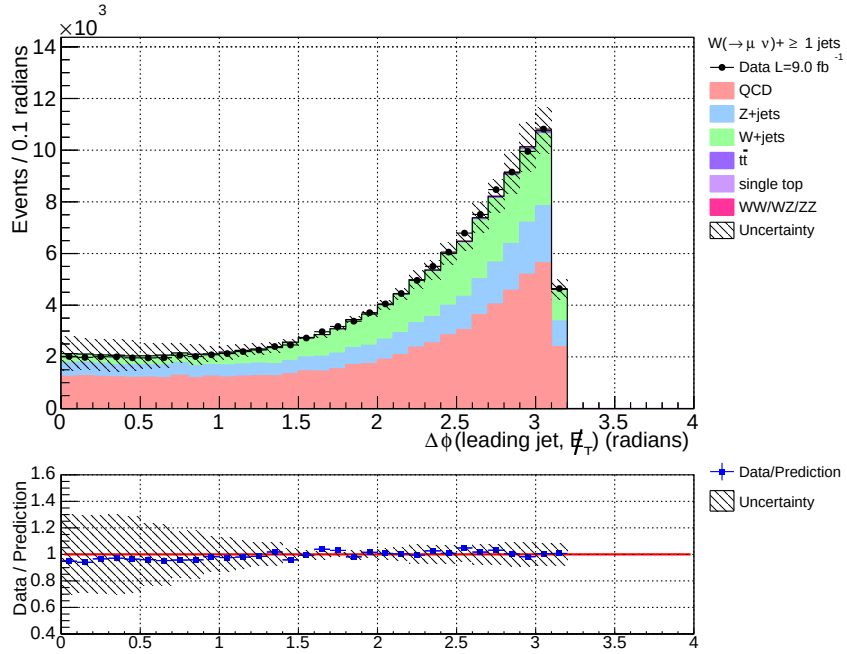


(b)

Figure 4.41: The angle between the muon and the $\not{E}_T$ distribution for the $W(\rightarrow \mu\nu)+ \geq 1$ jets sample in the control region in which the QCD multi-jet background is modeled with (a) *non-isolated muons* (b) *almost-isolated muons*. Lower pads illustrate the ratio between Data and Prediction.

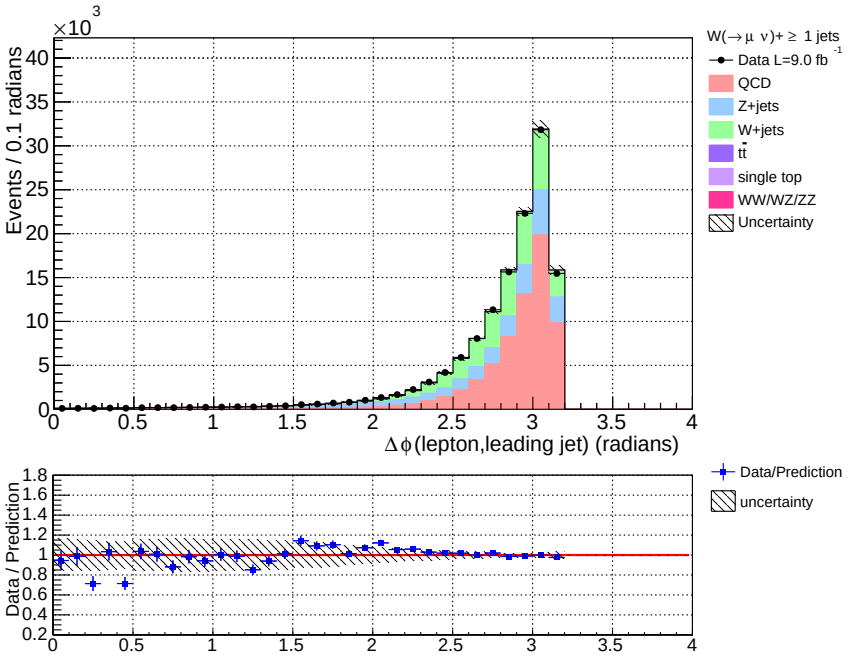


(a)

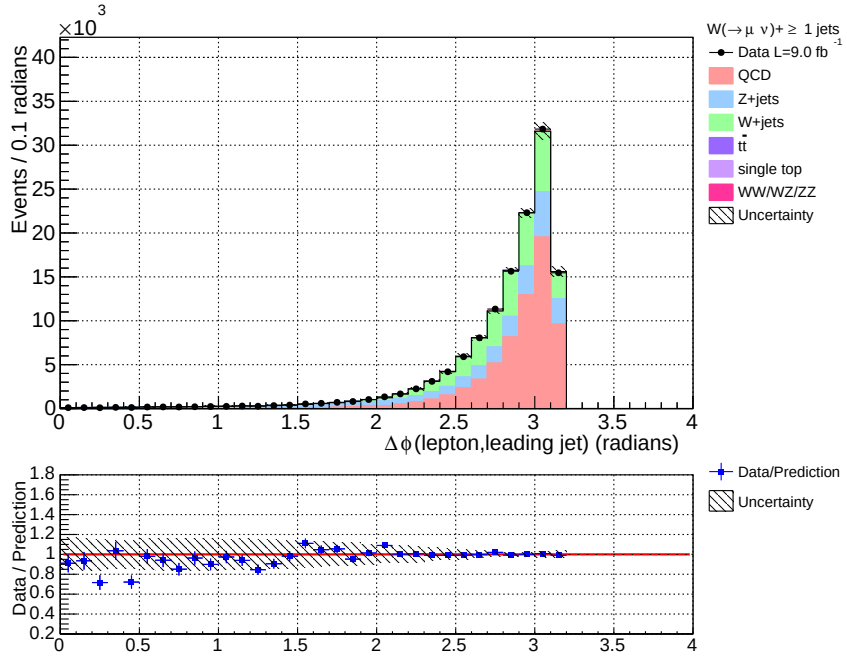


(b)

Figure 4.42: The angle between the leading jet and the $\not{E}_T$ distribution for the $W(\rightarrow \mu\nu)+ \geq 1$ jets sample in the control region in which the QCD multi-jet background is modeled with (a) *non-isolated muons* (b) *almost-isolated muons*. Lower pads illustrate the ratio between Data and Prediction.

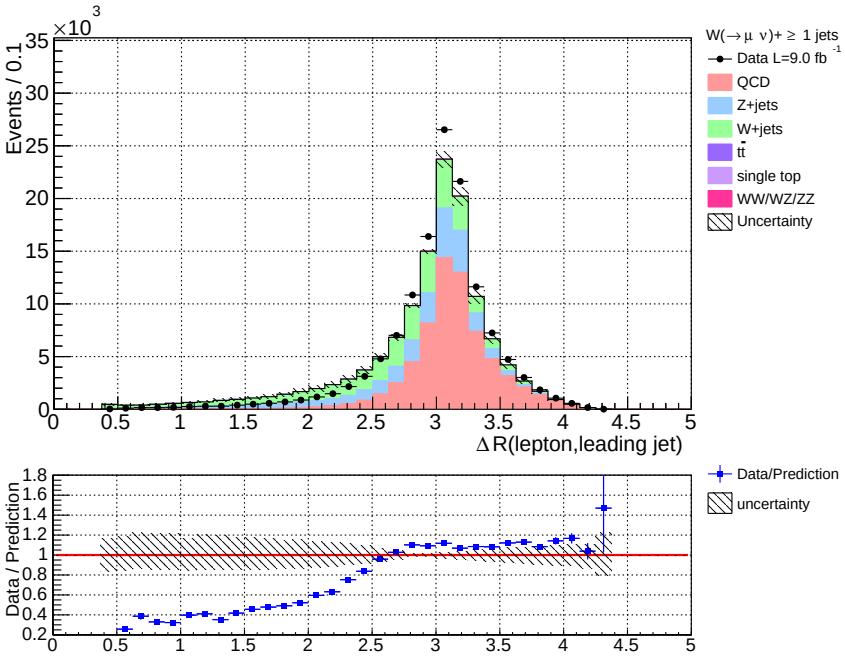


(a)

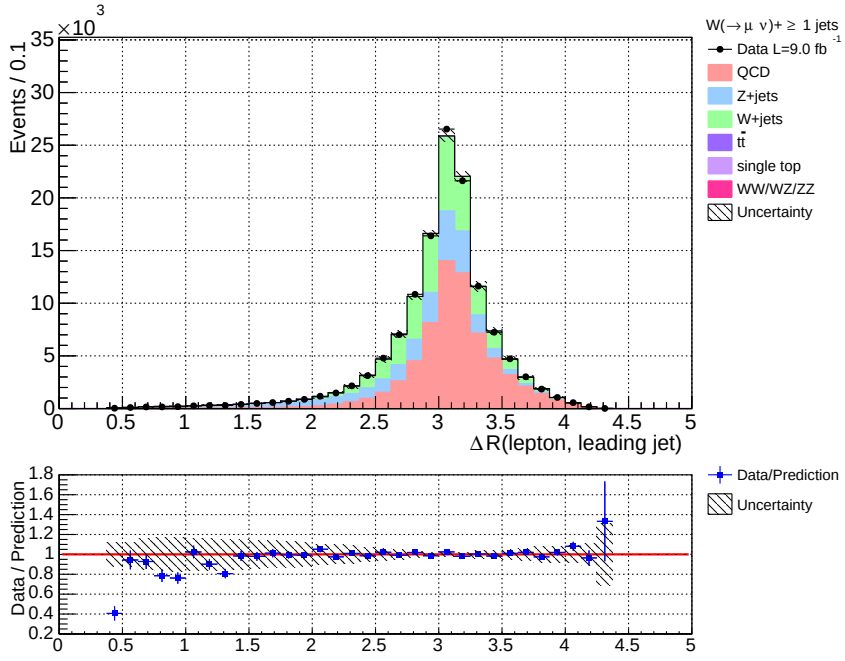


(b)

Figure 4.43: The angle between the leading jet and the muon distribution for the $W(\rightarrow \mu\nu)+ \geq 1$ jets sample in the control region in which the QCD multi-jet background is modeled with (a) *non-isolated muons* (b) *almost-isolated muons*. Lower pads illustrate the ratio between Data and Prediction.



(a)



(b)

Figure 4.44: The ΔR between the leading jet and the $\not{E}_T$ distribution distribution for the $W(\rightarrow \mu\nu)+ \geq 1$ jets sample in the control region in which the QCD multi-jet background is modeled with (a) *non-isolated muons* (b) *almost-isolated muons*. Lower pads illustrate the ratio between Data and Prediction.

Modeling of the distribution clearly benefits from the use of the *almost-isolated muons*. This indicates that the isolation is highly correlated with the kinematical variables of the events. In particular isolation is correlated with the amount of transverse energy carried by the jets, and this is confirmed by the fact that the systematic uncertainty due to the isolation selection is very large. In any case, since the contribution of the QCD multi-jet background in the signal region is expected to be very small, the impact of the systematic uncertainty on the final results should by small.

4.6 Cross Section Measurement

The measured differential cross section per bin of a variable α, which in this thesis is either the number of jets or the E_T of the leading jet, is defined as:

$$\frac{\Delta\sigma^{meas}_{W(\to\ell\nu)+jets}}{\Delta\alpha} = \frac{1}{\Delta\alpha}\, U(\alpha)\, \frac{N_{DATA}(\alpha) - N_{BKG}(\alpha)}{L} \tag{4.6.1}$$

where $\Delta\alpha$ is the width of the bin in which the cross section is measured and:

$N_{DATA}(\alpha) - N_{BKG}(\alpha)$**:** in obtain by subtracting the estimated background events $N_{BKG}(\alpha)$ from the $W+$jets data, $N_{DATA}(\alpha)$. Details in section 4.6.1.

$U(\alpha)$ is the unfolding factor which is applied to remove all the detector effects, *e.g.*, offline reconstruction and trigger efficiencies ($A\epsilon$) and unfold the measurement back to particle level in order to facilitate comparison with theoretical predictions. The selection phase space of the unfolded measurement reflects the kinematic cuts made in the event selection and is summarized in table 4.13. Details in section 4.6.2.

	Kinematical requirements
Leptons	$E_T > 25\,\text{GeV}$ and $\lvert\eta\rvert < 1.0$
Jets	$E_T > 25\,\text{GeV}$ and $\lvert\eta\rvert < 2.0$
W boson	$M_T^W > 40\,\text{GeV}$

Table 4.13: Definition of the selection phase space of the measurement

L is the integrated luminosity.

4.6.1 Background Subtraction

Tables 4.14 and 4.15 report the estimated number of background events in $9.0\,\text{fb}^{-1}$ for each background separately for the electron and the muon channel, respectively. The last line of each table reports the data yields.

CDF Run II Preliminary

Estimated and Observed Number of Events in 9.0 fb^{-1} $W(\rightarrow e\nu)$+ jets

Backgrounds	≥ 1 jet	≥ 2 jets	≥ 3 jets	≥ 4 jets
Multijet	142382 ± 23500	21706 ± 2783	3022 ± 427	470 ± 68
Z+jets	21959 ± 4392	2771 ± 554	348 ± 70	41 ± 8
$W(\rightarrow \tau\nu)$+jets	8717 ± 3487	948 ± 379	99 ± 40	10 ± 4
WW, WZ, ZZ	3759 ± 229	1693 ± 103	231 ± 14	31 ± 2
$t\bar{t}$	2769 ± 83	2535 ± 76	1756 ± 53	736 ± 22
Single Top	2038 ± 228	1187 ± 133	243 ± 27	33 ± 4
Total Backgrounds	181624 ± 24162	30840 ± 2869	5699 ± 438	1321 ± 72
Data	477665	65029	9483	1642

Table 4.14: Estimated background events in 9.0 fb^{-1} for $W(\rightarrow e\nu)+ \geqslant 1$, 2, 3 and 4 jets and data yields.

CDF Run II Preliminary

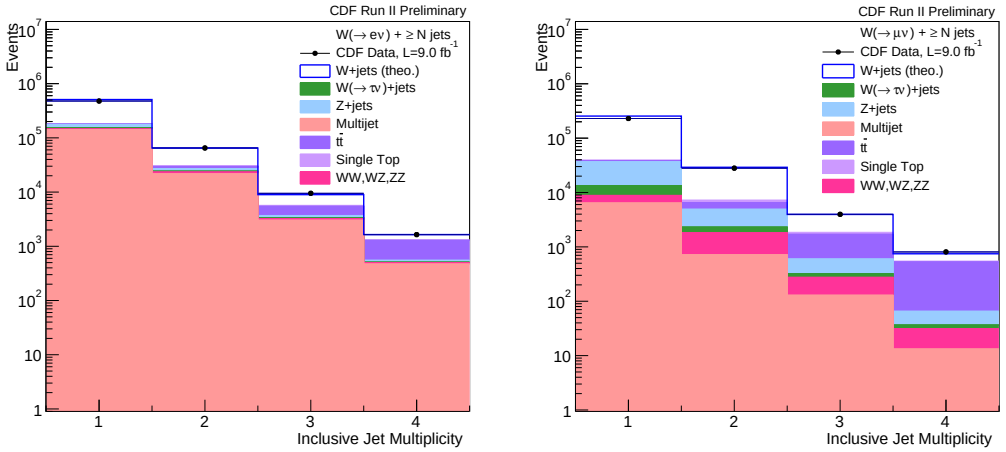
Estimated and Observed Number of Events in 9.0 fb^{-1} $W(\rightarrow \mu\nu)$+ jets

Backgrounds	≥ 1 jet	≥ 2 jets	≥ 3 jets	≥ 4 jets
Z+jets	23287 ± 4657	2600 ± 520	275 ± 55	29 ± 6
$W(\rightarrow \tau\nu)$+jets	4571 ± 1828	496 ± 199	48 ± 19	5 ± 2
Multijet	6407 ± 990	716 ± 291	129 ± 80	13 ± 13
WW, WZ, ZZ	2430 ± 146	1102 ± 66	146 ± 9	18 ± 1
$t\bar{t}$	1739 ± 52	1593 ± 48	1101 ± 33	462 ± 14
Single Top	1317 ± 149	767 ± 85	156 ± 18	22 ± 3
Total Backgrounds	39751 ± 5104	7274 ± 639	1855 ± 106	549 ± 20
Data	229823	28038	3967	807

Table 4.15: Estimated background events in 9.0 fb^{-1} for $W(\rightarrow \mu\nu)+ \geqslant 1$, 2, 3 and 4 jets and data yields.

The total background fraction increases with the jet multiplicity. As anticipated the multi-jet process is the main background for the electron channel but not for the muon channel. In both channels, the estimated fraction of multi-jet events decreases with increasing jet multiplicity. For $\geqslant 3$ and $\geqslant 4$ jets bins the $t\bar{t}$ background is significant, as expected. The contribution of this background can be reduced by using a veto based on b-tagging information. However, the event selection for this analysis needs to be kept as inclusive as possible to allow better comparison with theory. Therefore the b-tag veto has been considered but not applied. The contribution of the WW, WZ, ZZ and the single-top backgrounds grows in the $\geqslant 2, 3, 4$ jet multiplicities but is almost negligible in both channels. These observations are evident in figures 4.45a and 4.45b in which the comparison between the inclusive number of jets in the data and the theory (theoretical prediction generated with ALPGEN+PYTHIA Monte Carlo calculation and normalized to LO×K-factor)

plus background for $W(\to e\nu)+ \geqslant N$ jets and $W(\to \mu\nu)+ \geqslant N$ jets, are shown.



(a) (b)

Figure 4.45: Data and background for each inclusive jets multiplicity in the (a) $W \to e\nu$ channel and (b) $W \to \mu\nu$ channel. The ALPGEN+PYTHIA normalized to the LO×K-factor theoretical prediction is also shown.

4.6.2 Unfolding

In order to facilitate comparison with theoretical predictions, the particle level cross section is needed. The cross section at particle level can be evaluated by correcting the measurement for all the detector effects, such as finite resolution and limited acceptance, that can distort the distributions under study. The procedure used for this evaluation is known as "unfolding". This unfolding procedure takes into account migration of events from one bin to another due to the limited resolution and acceptance of the detector. The unfolding can be accomplish by means of a response matrix ($\mathcal{R}_{ij}$) that maps the particle level distribution P_j to the detector reconstructed distribution D_i:

$$D_i = \sum_j \mathcal{R}_{ij}\ P_j. \tag{4.6.2}$$

Ideally, the unfolded distribution can be obtained by simply applying the inverted matrix $\mathcal{R}^{-1}$ to the measured data (background subtracted data):

$$P_j = \sum_i \mathcal{R}^{-1}_{ji}\ D_i. \tag{4.6.3}$$

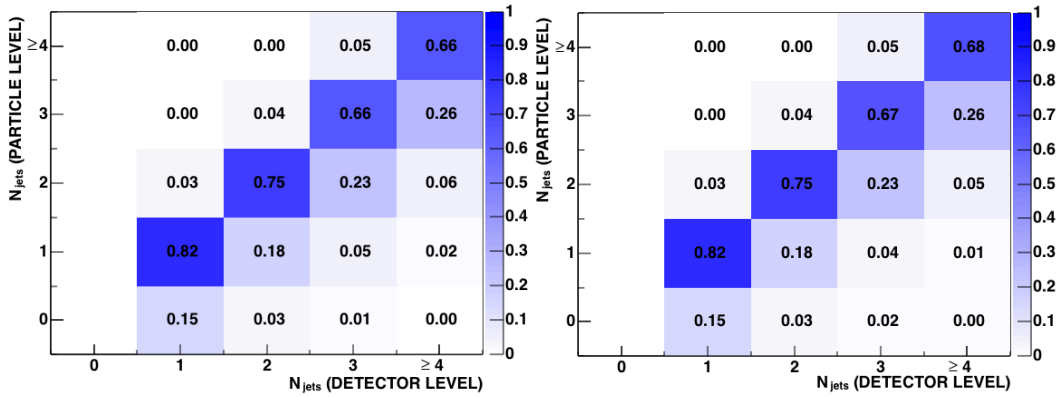
The response matrix $\mathcal{R}$ can be factorized, into the diagonal matrix $\mathcal{A}$ that accounts for the acceptance and the non-diagonal migration matrix $\mathcal{M}$. Inversion of the diagonal matrix $\mathcal{A}$ is trivial, on the other hand inversion of the non-diagonal migration matrix $\mathcal{M}$ requires special care since numerical instabilities can be introduced by low-significance (low statistics) bins resulting in unphysical oscillating fluctuations [100].

Migration and Acceptance Matrices

The migration matrix $\mathcal{M}$ is derived from the W+jets ALPGEN+PYTHIA Monte Carlo. Each entry (i, j) of the matrix corresponds to the number of reconstructed events ($x_i^{MC\ detector}$) in bin i that are generated in bin j at particle level ($x_j^{MC\ particle}$):

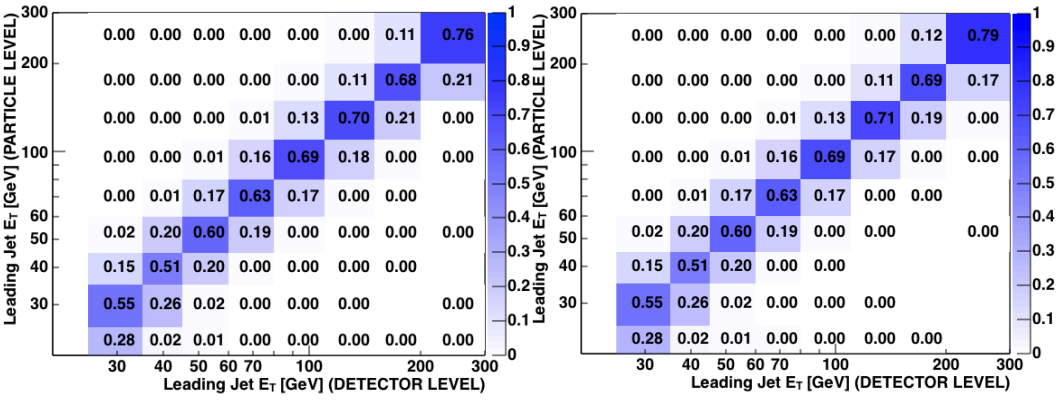
$$x^{MC\ detector} = \mathcal{M}\ x^{MC\ particle} \tag{4.6.4}$$

Figures 4.46 and 4.47 show the migration matrices used to unfold the observables measured in this analysis: the number of jets in the electron channel (4.46a) and in the muon channel (4.46b), the transverse energy of the leading jet in the electron channel (4.47a) and in the muon channel (4.47b). The matrices are normalized to the number of reconstructed events in each bin of jet multiplicity or leading jet E_T, so that $\sum_i \mathcal{M}_{ij} = 1$.



(a) (b)

Figure 4.46: Migration matrices to unfold the exclusive number of jets cross section (a) for $W(\to e\nu)$+ jets and (b) for $W(\to \mu\nu)$+ jets.



(a) (b)

Figure 4.47: Migration matrices to unfold the E_T of the leading jet cross section (a) for $W(\to e\nu)$+ jets and (b) for $W(\to \mu\nu)$+ jets.

The particle level jets are reconstructed in the Monte Carlo sample by applying the JETCLU algorithm to the stable particles of the HEPG bank in the final states. The same requirements applied to jets and the lepton at detector level are applied at particle level. Moreover, in the definition of the particle level used in this analysis photons which are radiated are recombined with the leptons using a scheme similar to that described in [101]. Particle level requirements are summarized in table 4.13.

The measured distributions are unfolded with the Single Value Decomposition (SVD) technique [102] within the ROOT software package [103]. The SVD unfolding algorithm allows for a better understanding of, and correction for, the causes of the oscillations due to statistical fluctuations in the data. SVD consists in factorizing the the $(m \times n)$-matrix $\mathcal{M}$ in the form $\mathcal{M} = \mathcal{U}\mathcal{S}\mathcal{V}^T$ where $\mathcal{U}$ is an orthogonal $(m \times m)$-matrix and $\mathcal{V}$ is an orthogonal $(n \times n)$-matrix, where $m \geq n$. $\mathcal{S}$ is a $(m \times n)$ diagonal matrix ($s_{ij} = 0$ for $i \neq j$) and the non-negative entries along the diagonal, s_i, called the singular values, are arranged in descending order. Defining $z = \mathcal{V}^T x^{particle}$ and $d = \mathcal{U}^T x^{detector}$ we have $\mathcal{S}z = d$ and $s_i z_i = d_i$ (diagonal system of equations). When applying the inverted matrix to the data, the exact solution of the system can then be expressed as:

$$x^{particle} = \mathcal{V}z = \mathcal{V}\mathcal{S}^{-1}d = \mathcal{V}\mathcal{S}^{-1}\mathcal{U}^T x^{detector}. \tag{4.6.5}$$

Because $z_i = d_i/s_i$ problems arise if one of the singular values s_i is too small. In this case errors in d_i are exaggerated and the values returned for z_i, which are related to $x_i^{particle}$ are unrealistic.

In order to mitigate these problems, one rescales equation 4.6.4 by multiplying $\mathcal{M}_{ij}$ by the Monte Carlo simulated particle level distribution $x^{MC,\ particle}$ and transforms the unknowns into

$$w_i = \frac{x_i^{particle}}{x_i^{MC,\ particle}} \tag{4.6.6}$$

the fractional deviation of the measured particle level distribution from the Monte Carlo simulation. The matrix is also rescaled by dividing each equation by the error $(\Delta x_i^{detector})$ corresponding to the measured value $x_i^{detector}$ so as to have equally weighted equations when one comes to minimizing the residuals. The result of the rescaling leads to the following system:

$$\widetilde{\mathcal{M}}_{ij} w_j = \widetilde{x}_i^{detector} \tag{4.6.7}$$

which, after SVD factorization, obtains $z = \mathcal{V}^T w$ and $d = \mathcal{U}^T \widetilde{x}_i^{detector}$. Again $\mathcal{S}z = d$ and $s_i z_i = d_i$ and it is still possible that large errors in $x^{detector}$ lead to insignificant d_i and correspondingly small s_i which give large (but insignificant) z_i. Oscillations in the solution of 4.6.7 are therefore still possible. Because $m \geq n$, the system is overdetermined and the solution is obtained by minimizing the square of the residuals:

$$\chi^2 \equiv (\widetilde{\mathcal{M}}w - \widetilde{x}_i^{detector})^T (\widetilde{\mathcal{M}}w - \widetilde{x}_i^{detector}) = min \tag{4.6.8}$$

The spurious oscillatory component can be suppressed by adding a regularization term to 4.6.8:

$$\chi^2 \equiv (\widetilde{\mathcal{M}}w - \widetilde{x}_i^{detector})^T (\widetilde{\mathcal{M}}w - \widetilde{x}_i^{detector}) + \tau(Cw)^T(Cw) = min \tag{4.6.9}$$

which has the effect of smoothing the solution. One now obtains:

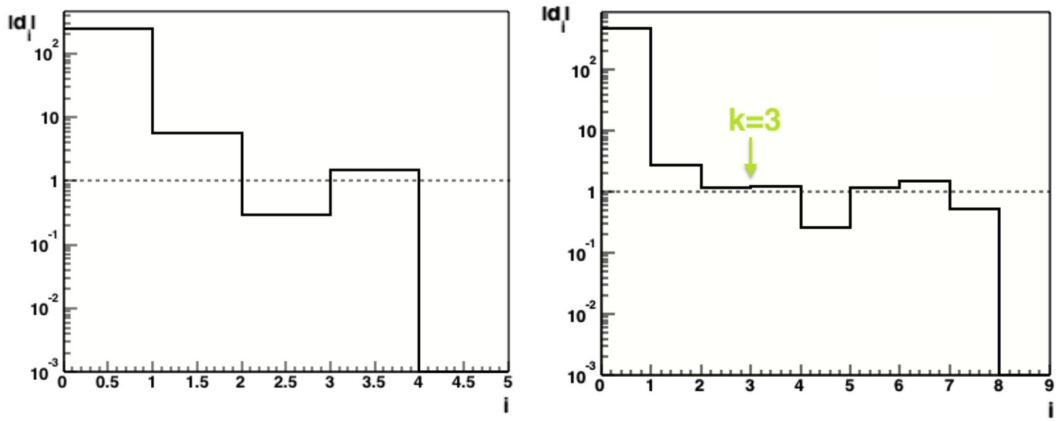
$$z_i = \frac{d_i s_i}{s_i^2 + \tau}. \tag{4.6.10}$$

and oscillations are eliminated by setting:

$$\tau = s_k^2 \tag{4.6.11}$$

where k is the index of the last significant d. The optimal value of k is problem-dependent and its determination is an important part of the procedure. Since we rescaled the equations by the corresponding errors in $x_i^{detector}$, d_i remains significant for values > 1. d_i is expected to decrease with increasing i and the optimal value of k therefore corresponds to the value of i for which d_i assumes a value around 1.

Figure 4.48, shows the plot of $\log d_i$ *vs* i (i runs over the number of bins) for the matrices used in the analysis.



(a) (b)

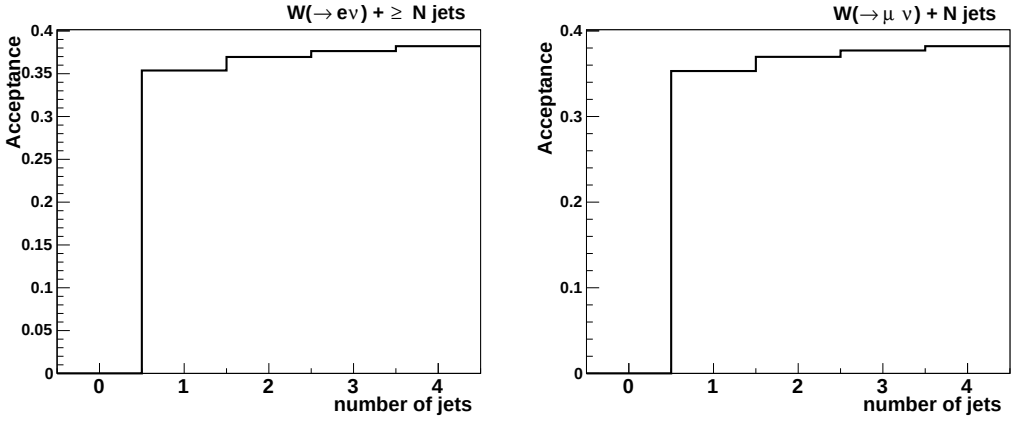
Figure 4.48: Plot of $\log d_i$ *vs* i for (a) the exclusive number of jets distributions and (b) the leading jet E_T distributions in which the arrow indicates the boundary between the significant and non-significant equations.

From figure 4.48b, it can be inferred that the last significant d is for $i = 3$, for this reason the optimal choice of the regularization parameter for unfolding the leading jet E_T distributions is $k = 3$. In figure 4.48a d_i starts oscillating around 1 after $i = 3$, which means that $k = 3$ and $k = 4$ are both adequate choices.

The diagonal acceptance matrices ($\mathcal{A}$) are also derived using the same W+jets ALPGEN+PYTHIA MC simulation. The correction carried by this matrix are required by the fact that the experimental resolution affects the matching between the detector level and the particle level so that there are some events reconstructed at detector level that do not satisfy the selection criteria at particle level.

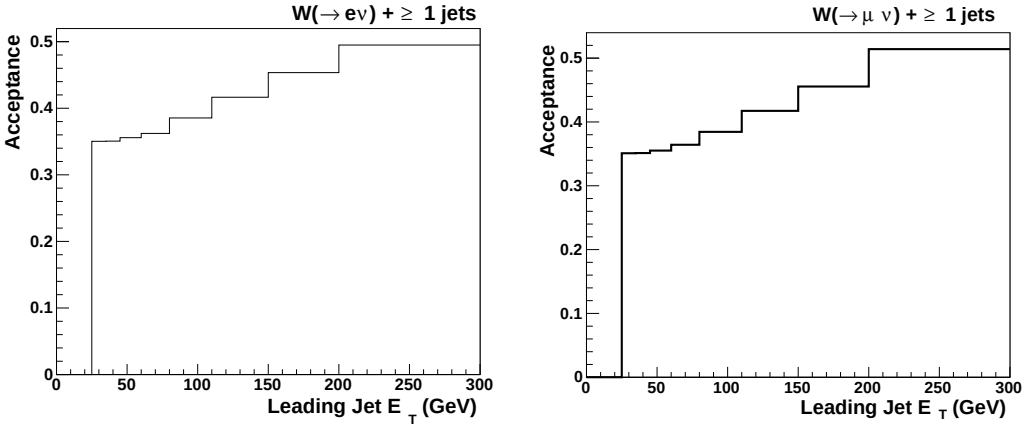
Figures 4.49 and 4.50 show the acceptance matrices used to correct the observables measured in this analysis: the number of jets in the electron channel (4.49a)

and in the muon channel (4.49b), the transverse energy of the leading jet in the electron channel (4.50a) and in the muon channel (4.50b).



(a) (b)

Figure 4.49: Acceptance matrices used to unfold the exclusive number of jet distribution (a) for $W(\rightarrow e\nu)+$ jets and (b) for $W(\rightarrow \mu\nu)+$ jets.



(a) (b)

Figure 4.50: Acceptance matrices used to unfold the E_T of the leading jet distribution (a) for $W(\rightarrow e\nu)+$ jets and (b) for $W(\rightarrow \mu\nu)+$ jets.

Validation of the unfolding procedure

Before unfolding the data, two different checks are performed to validate the procedure:

1. *Identity test:* perform the unfolding on the same signal MC distribution as the one used to determine the migration matrix;

2. *Closure test:* perform the unfolding on signal MC distributions that differ from the ones used to determine the migration matrix. Four MC test samples are built by varying the cross-section at parton level so has to mimic the data distributions of interest at particle and reconstruction levels. The variation used to build the four test samples are shown in figure 4.51.

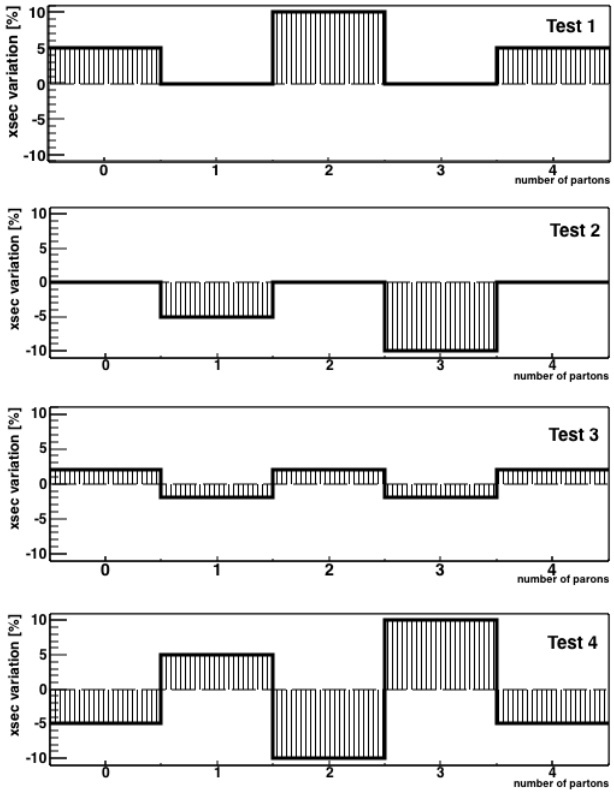


Figure 4.51: Percentage variation of the cross-section at parton level used to build the 4 samples for performing the closure test.

Both tests are performed with a set of different regularization parameters to find the optimal value of k.

For each test distribution, an ensemble of one thousand distributions from Pseudo-Experiments (PEs) is built. In each pseudo-experiment the number of events in each bin is randomly changed using a Poisson distribution. Only the distributions with, on average, the same statistics and fluctuations (in each individual bin and in the total number of events) as those of the measured distributions are collected in the ensemble. Each of the distributions of an ensemble is unfolded using the same procedure and, for each bin, the residual:

$$residual = \frac{\text{unfolded result} - \text{particle level value}}{\text{particle level value}} \tag{4.6.12}$$

is calculated. In a given bin the distribution of the residual is a Gaussian distribution with the mean and sigma representing the bias and the statistical uncertainty of the unfolding procedure, respectively. The bias and the uncertainty of the unfolding procedure for different values of k are shown in figures 4.52, 4.53 and 4.54. We choose as optimal regularization parameter k that obtains no bias and the lowest uncertainty in the closure test shown in figure 4.53 and 4.54.

The results for the identity test are shown is figure 4.52. As expected, no bias is present. The identity test is only used to check if the migration matrix is correctly built.

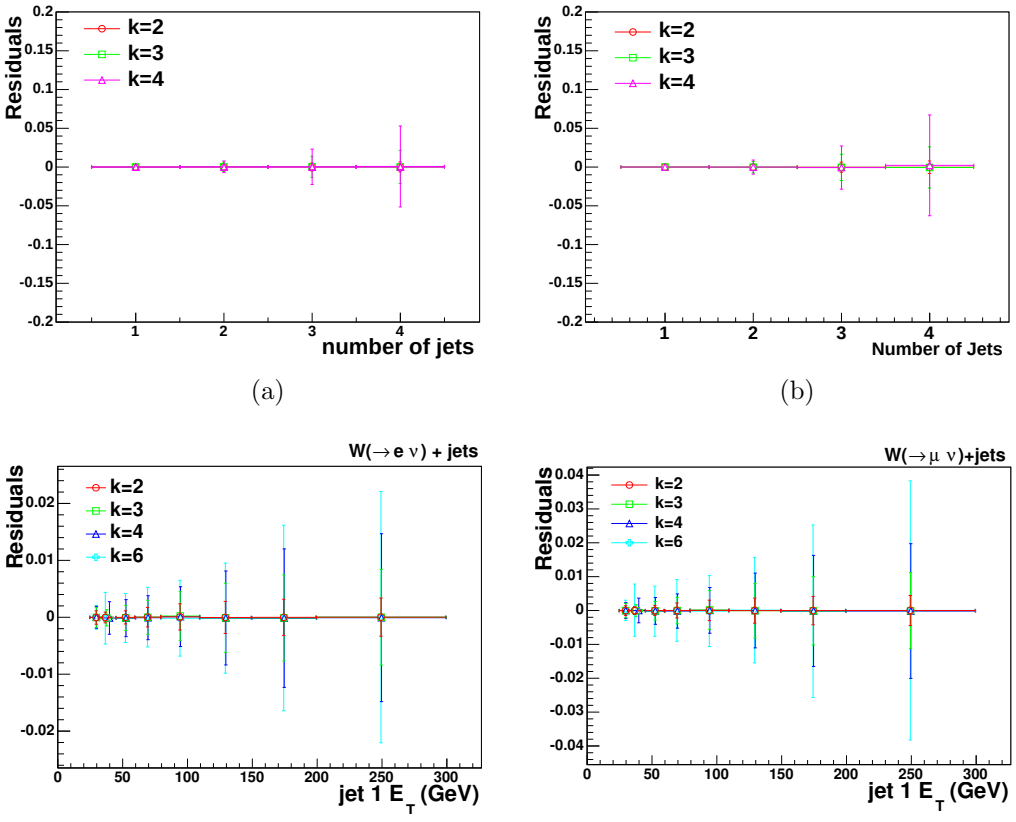


Figure 4.52: Results of the identity test for number of jets in the electron channel (a) and in the muon channel (b), and for the transverse energy of the leading jet in the electron channel (c) and in the muon channel (d).

Figure 4.53, shows the result of the 4 closured tests for the distribution of the number of jets in the electron channel. Tests suggest that $k = 4$ is the optimal choice in so far as it minimizes the bias and the errors do to fluctuations.

For the leading jet E_T distribution in the electron (fig. 4.54a) and muon (fig. 4.54b) channels we only show test 4 because it is the most stringent test insofar as it includes the largest variation on the cross-section.

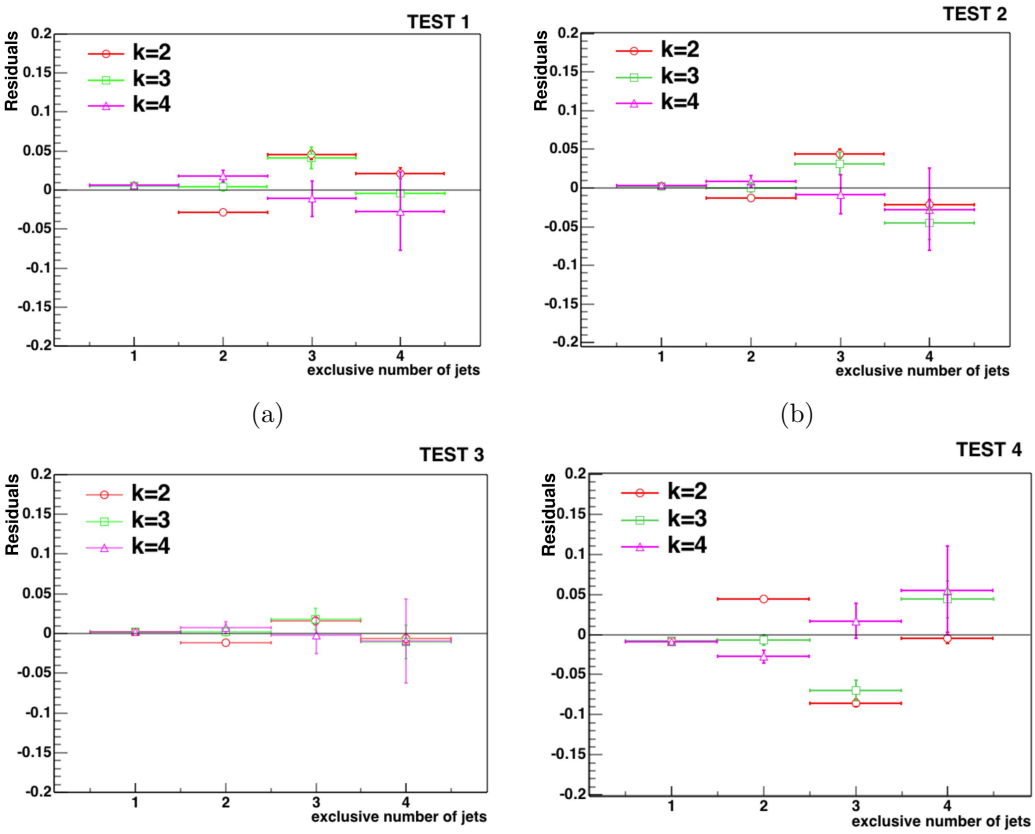


Figure 4.53: The bias and the uncertainty of the unfolding procedure for each bin in the distribution of the number of jets in the electron channel from closure tests (a) 1, (b) 2, (c) 3 and (d) 4.

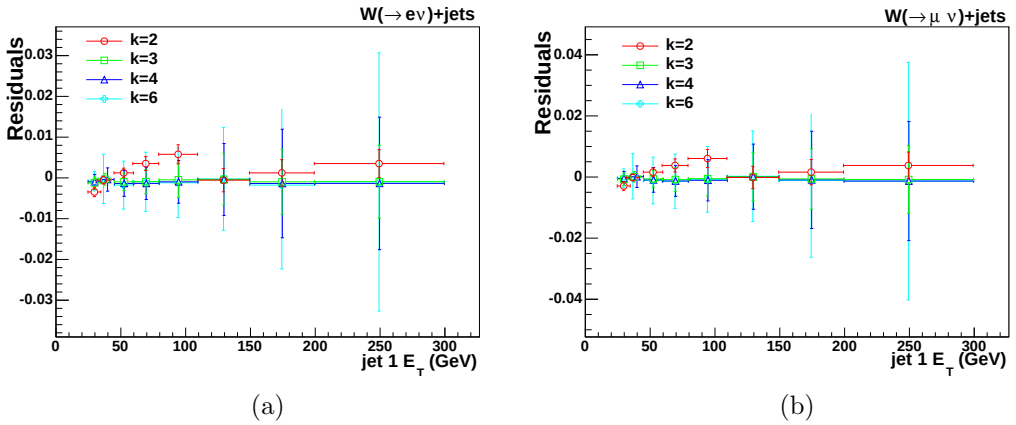


Figure 4.54: Bias and the uncertainty of the unfolding procedure for each bin of the jet1 E_T distribution in the electron channel (a) and in the muon channel (b), from closure test 4.

Figure 4.54 confirms that the optimal choice for the regularization parameter is $k = 3$ because it guarantees the lowest bias and fluctuations in both distributions.

4.7 Systematic Uncertainties

In this analysis, we considered the uncertainties in the background estimation and those in the unfolding procedure as systematic uncertainties. The unfolding procedure is repeated for each systematic variation, and the difference between unfolded data and nominal result is taken as uncertainty.
The uncertainties in the background estimation include:

- **Lepton Acceptance:** we quote 2.2% as the systematic uncertainty due to the trigger efficiency. This systematic is not applied to the multi-jet background since is data-driven.
- **Monte Carlo (MC) Backgrounds:** is the uncertainty on the production cross section assigned to the Monte Carlo. Details regarding these cross sections are reported in section 4.4.1.
- **Multijet (MJ) Backgrounds:** uncertainty in the estimates of the multijet background. All sources of uncertainties discussed in section 4.4.2 are accounted for.
- **Jet Energy Scale (JES)**: uncertainty in the modeling of the jet energy scale. Since an additional correction, which takes into account the difference between gluon and quark jets, has been added to the standard L5 correction, we assumed a 2.7% uncertainty for quark jets and 4.4%uncertainty for gluon jets with 100% anti-correlation.

The sources of systematic uncertainties which have been considered for differential cross section measurements listed above are summarized in table 4.16.

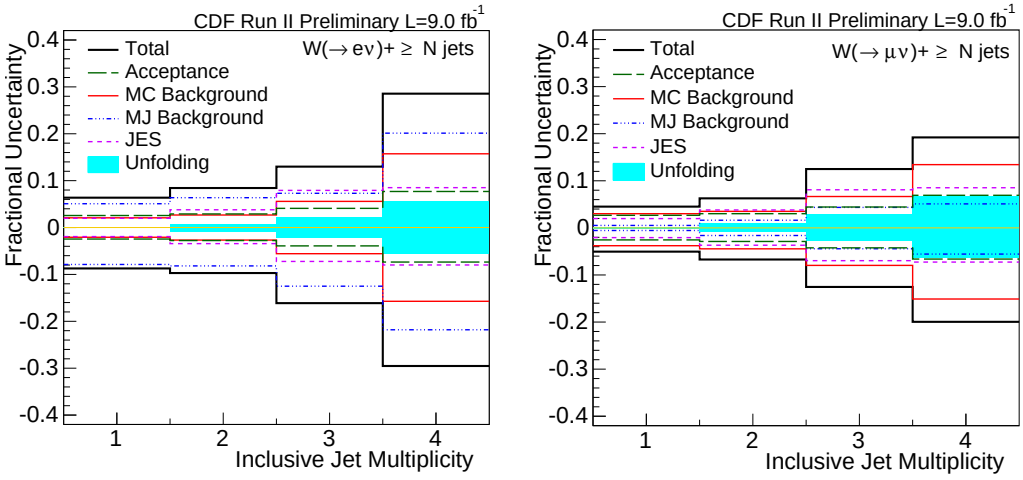
Systematic	Wjets	Zjets	$t\bar{t}$	sing.-t	VV	W($\tau\nu$)jets	QCD
Acceptance							
Lepton Acc.	2.2%	2.2%	2.2%	2.2%	2.2%	2.2%	
Backgrounds							
Monte Carlo		20%	3%	11%	6.1%	40%	
Multijet (ele)							13-18%
Multijet (muo)							15-95%
JES	$\pm1\sigma$	$\pm1\sigma$	$\pm1\sigma$	$\pm1\sigma$	$\pm1\sigma$	$\pm1\sigma$	
Luminosity		6%	6%	6%	6%	6%	

Table 4.16: Summary of the systematic uncertainties applied to the measured cross section.

The uncertainty in the unfolding procedure (width of the residual distribution) is 0.1% - 6%. We also account for the acceptance and resolution uncertainties in the

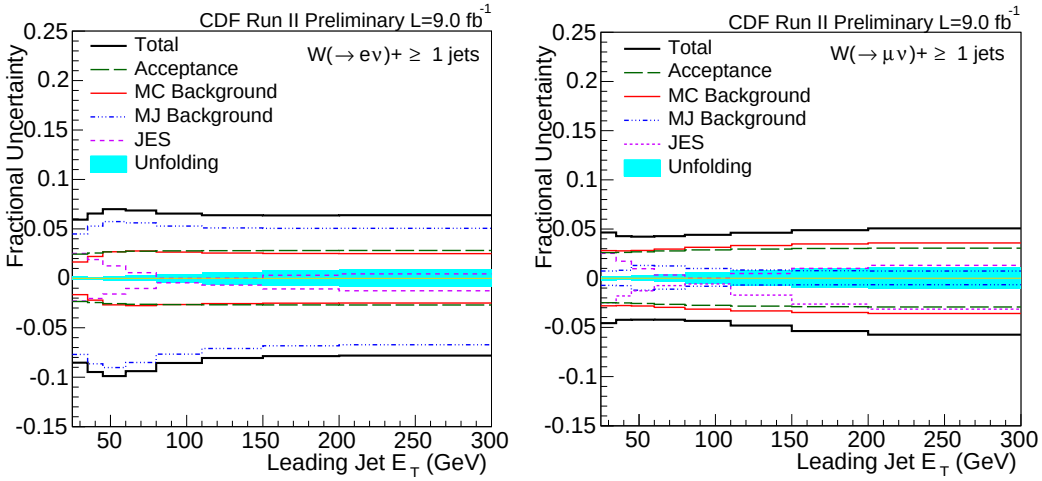
$W+$jets MC (reported in the first column of table 4.16) used to build the unfolding matrix. All correlations are properly accounted for.

Figures 4.55 and 4.56 show the systematic uncertainties as a function of inclusive jet multiplicity and first jet transverse energy in both electron and muon channels.



(a) (b)

Figure 4.55: Total systematic uncertainties vs inclusive jet multiplicity (a) electron channel (b) muon channel.



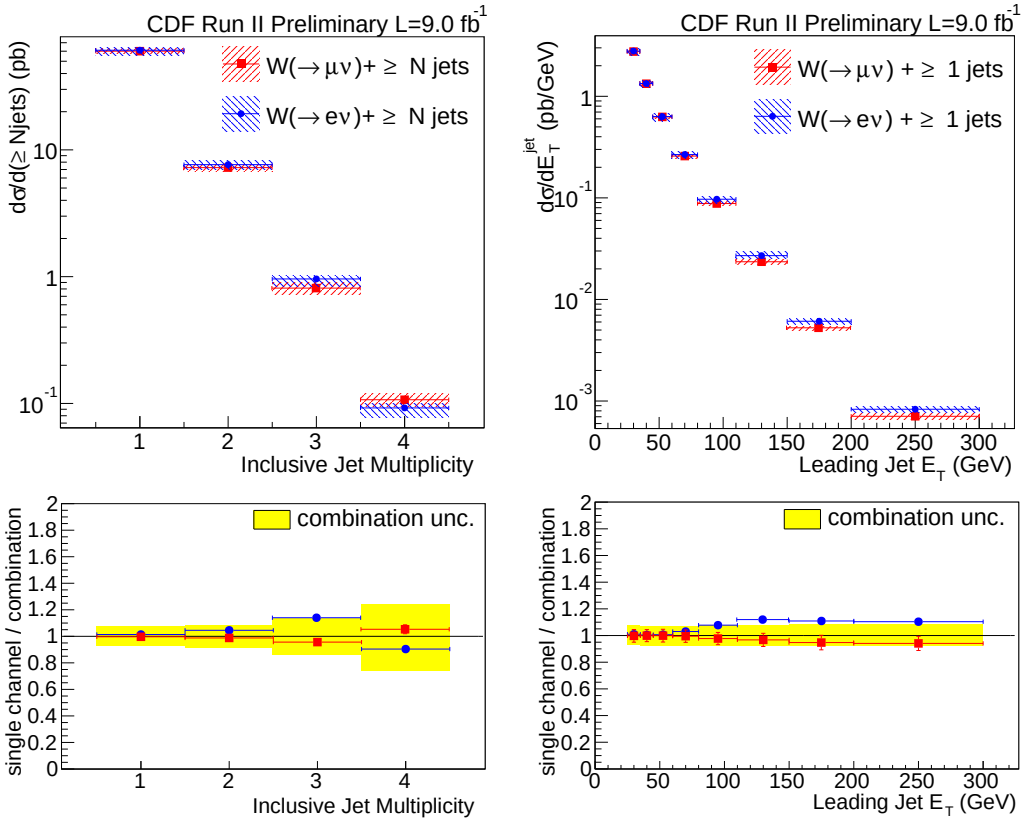
(a) (b)

Figure 4.56: Total systematic uncertainties vs leading jet E_T (a) electron channel (b) muon channel.

4.8 Electron and Muon decay channel combination

The $W+$jets cross sections are measured independently in the electron and muon decay channels. In order to combine them the Best Linear Unbiased Estimate

method (BLUE) has been used [104]. The algorithm takes into account possible correlations of the uncertainties in the two channels and the result is a weighted average of the them. Moreover, in order to process asymmetric uncertainties we used an algorithm that was developed at CDF for the single top production measurement called Asymmetric Iterative BLUE (AIB) [105]. AIB is a combination of three BLUE outputs in which the first has average uncertainties as input and the others upper and lower, respectively. Systematic uncertainties related to JES, MC background and luminosity are considered 100% correlated. On the other hand, statistical, trigger efficiency and multi-jet background uncertainties are considered uncorrelated. The measured cross sections as a function of inclusive jet multiplicity and leading jet E_T for $W(\to e\nu)$ and $W(\to \mu\nu)$ are shown in figure 4.57. The lower plots show the comparison between the single channel cross sections and the combined $W(\to \ell\nu)$ cross sections with the total systematic uncertainties which account for all correlations.



(a) (b)

Figure 4.57: Measured cross section in the $W(\to e\nu)+$jets and $W(\to \mu\nu)+$jets decay channels as a function of (a) inclusive number of jets and (b) inclusive leading jet E_T in events with ≥ 1 jet. Lower figures show the ratio between measured cross sections and combination with the total systematic uncertainty (all correlations are properly accounted for).

Chapter 5

Results and Discussion

The previous chapter contains the detailed description of the analysis and of the measurements. In this chapter the measured cross sections resulting from the channel combination are summarized and discussed.

Section 5.1 presents the measured cross sections for the production of a W-boson in association with inclusive jets. In section 5.2, the differential cross-sections as a function of jet multiplicity and leading jet E_T are compared with theory. Theoretical predictions are simulated with W+jets leading order Monte Carlo generators and their uncertainties are evaluated.

5.1 W+jets Production Cross Sections

We present cross sections of W-bosons, identified through their decay to electrons and muons, in association with an inclusive number of jets, $\sigma_{(p\bar{p}\to W(\to\ell\nu)+\geqslant N\,\text{jets})}$. We considered jet multiplicities up to $N = 4$. Measurements of the combined differential cross sections are reported in tables 5.1 and 5.2. The uncertainties quoted are factorized into statistical, systematic and luminosity sources. The uncertainties are dominated by systematics mainly because of uncertainties in the background modeling, as detailed in section 4.7.

CDF Run II Preliminary, L=9.0 fb^{-1}

Jet Multiplicity	**Inclusive σ [pb]**
≥ 1 jet	$60.3 \pm 0.1\,(\text{stat.})^{+2.7}_{-2.6}(\text{syst.}) \pm 3.6\,(\text{lum.})$
≥ 2 jets	$7.31 \pm 0.04\,(\text{stat.})^{+0.43}_{-0.44}(\text{syst.}) \pm 0.44\,(\text{lum.})$
≥ 3 jets	$0.84 \pm 0.02\,(\text{stat.})^{+0.10}_{-0.11}(\text{syst.}) \pm 0.05\,(\text{lum.})$
≥ 4 jets	$0.102 \pm 0.007\,(\text{stat.})^{+0.024}_{-0.026}(\text{syst.}) \pm 0.006\,(\text{lum.})$

Table 5.1: The unfolded $W(\to \ell\nu)$ +jets cross section measurements for inclusive jet multiplicities of up to 4 jets, together with statistical, systematic and luminosity uncertainties. The kinematic region to which the measurements are restricted is $E_T^\ell > 25\,\text{GeV}$, $|\eta^\ell| < 1$, $E_T^{jet} > 25\,\text{GeV}$, $|\eta^{jet}| < 2$ and $M_T^W > 40\,\text{GeV/c}^2$.

The measurements refer to particle level cross sections so that comparison with theoretical prediction can be performed. The kinematic region investigated is:

$$E_T^\ell > 25\,\text{GeV},\ |\eta^\ell| < 1,\ E_T^{jet} > 25\,\text{GeV},\ |\eta^{jet}| < 2 \text{ and } M_T^W > 40\,\text{GeV/c}^2$$

where E_T^ℓ and $|\eta^\ell|$ are the transverse energy and the rapidity of the lepton, whereas E_T^{jet} and $|\eta^{jet}|$ are the same quantities referred to the jet and M_T^W is the transverse mass of the W-boson.

5.2 Comparison with Theoretical Predictions

The cross sections measured are compared with the LO ALPGEN+PYTHIA prediction scaled up by a K-factor of 1.4 to account for the next to leading order effects. It is the same MC prediction as was used to unfold the measurement as described in section 4.6.2. The same prediction was also used at CDF to estimate the W+jets

$\mathbf{E_T}$ **Range** [GeV]	$\mathbf{d\sigma/dE_T^{1st\,jet}}$ **[pb/GeV]**
$25-35$	$2.76 \pm 0.01\,(\text{stat.})^{+0.11}_{-0.11}(\text{syst.}) \pm 0.17\,(\text{lum.})$
$35-45$	$1.331 \pm 0.004\,(\text{stat.})^{+0.053}_{-0.054}(\text{syst.}) \pm 0.080\,(\text{lum.})$
$45-60$	$0.628 \pm 0.003\,(\text{stat.})^{+0.027}_{-0.026}(\text{syst.}) \pm 0.038\,(\text{lum.})$
$60-80$	$0.261 \pm 0.002\,(\text{stat.})^{+0.012}_{-0.011}(\text{syst.}) \pm 0.016\,(\text{lum.})$
$80-110$	$0.090 \pm 0.001\,(\text{stat.})^{+0.004}_{-0.004}(\text{syst.}) \pm 0.005\,(\text{lum.})$
$110-150$	$0.0243 \pm 0.0004\,(\text{stat.})^{+0.0012}_{-0.0011}(\text{syst.}) \pm 0.0015\,(\text{lum.})$
$150-200$	$0.0055 \pm 0.0001\,(\text{stat.})^{+0.0003}_{-0.0003}(\text{syst.}) \pm 0.0003\,(\text{lum.})$
$200-300$	$0.00075 \pm 0.00002\,(\text{stat.})^{+0.00004}_{-0.00004}(\text{syst.}) \pm 0.00005\,(\text{lum.})$

Table 5.2: The unfolded leading jet E_T differential cross section for $W(\to \ell\nu) + \geqslant 1$ jet. Statistical, systematic and luminosity uncertainties are reported separately. The kinematic region to which the measurements are restricted is $E_T^\ell > 25\,\text{GeV}$, $|\eta^\ell| < 1$, $E_T^{jet} > 25\,\text{GeV}$, $|\eta^{jet}| < 2$ and $M_T^W > 40\,\text{GeV/c}^2$.

background in various analyses such as Higgs searches and diboson production cross-section measurements. Its validation is therefore of fundamental importance.

The Parton Distribution Functions, PDFs, used are from the leading order CTEQ5L set. The PDF uncertainties reported, account for the the uncertainties associated with the global fit of the experimental results and are calculated using the Hessian method [106]. The method consists in evaluating both the positive and negative deviations from the nominal global fit for each of the 20 free parameters considered by the CTEQ PDFs. The positive and negative variations are summed separately in quadrature, so that the asymmetric uncertainties represent the maximal deviation in each direction. In the case of variations with the same sign, only the largest is considered.

The nominal choice of the renormalization and factorization scale is

$$\mu_0 = \sqrt{M_W^2 + P_T^2} \tag{5.2.1}$$

where P_T^2 is the squared sum of transverse energies of all final state partons. To illustrate the dependence of the prediction on the renormalization and factorization scale the LO ALPGEN+PYTHIA predictions with $\mu_0/2$ and $2\mu_0$ has been evaluated. This uncertainty in the theoretical cross section is large since the leading order predictions suffer from dependence on the choice of the renormalization scale.

The comparison between the cross sections measured as a function of jet multiplicity reported in table 5.1 with the theoretical predictions are shown in figure

5.1. The systematic uncertainties on the measured cross sections are mainly due to the background modeling and are of the order of 5% for $N \geq 1$ jet up to 25% for $N \geq 4$ jets.

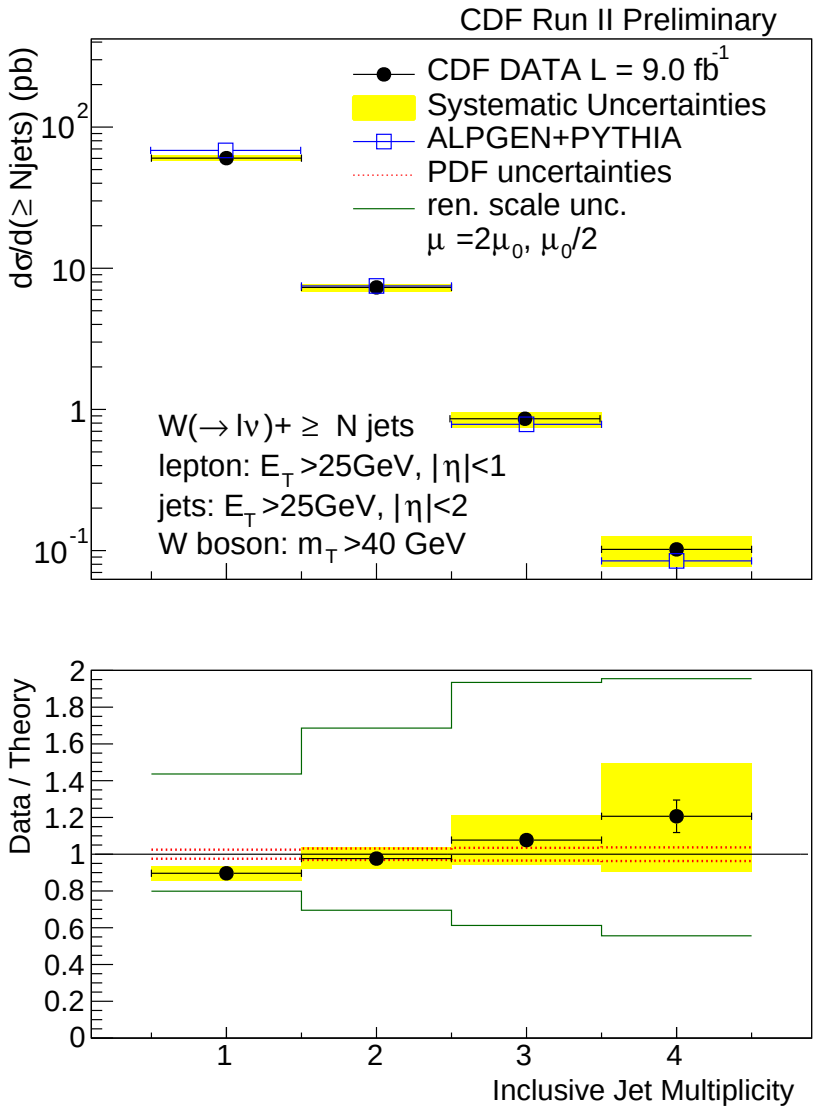


Figure 5.1: Measured inclusive differential cross sections (black dots) as a function of the inclusive jet multiplicity compared to LO calculations corrected by K-factor pQCD predictions (open squares). The shaded bands show the total systematic uncertainty, with the exclusion of the 6% luminosity uncertainty. The dashed indicate the PDF uncertainty whereas the solid lines the variation of the renormalization scale μ_0 as described in the text.

Figure 5.2 presents the measured differential cross section as a function of the leading jet E_T in $W(\to \ell\nu)+ \geq 1$ jets events. The differential cross section is measured in eight bins of leading jet E_T in the range 25 GeV to 300 GeV. The cross

sections decrease by approximately three orders of magnitude as the leading jet E_T increases. The systematic uncertainty on the measured cross sections is about 5% and is mainly due to the background modeling.

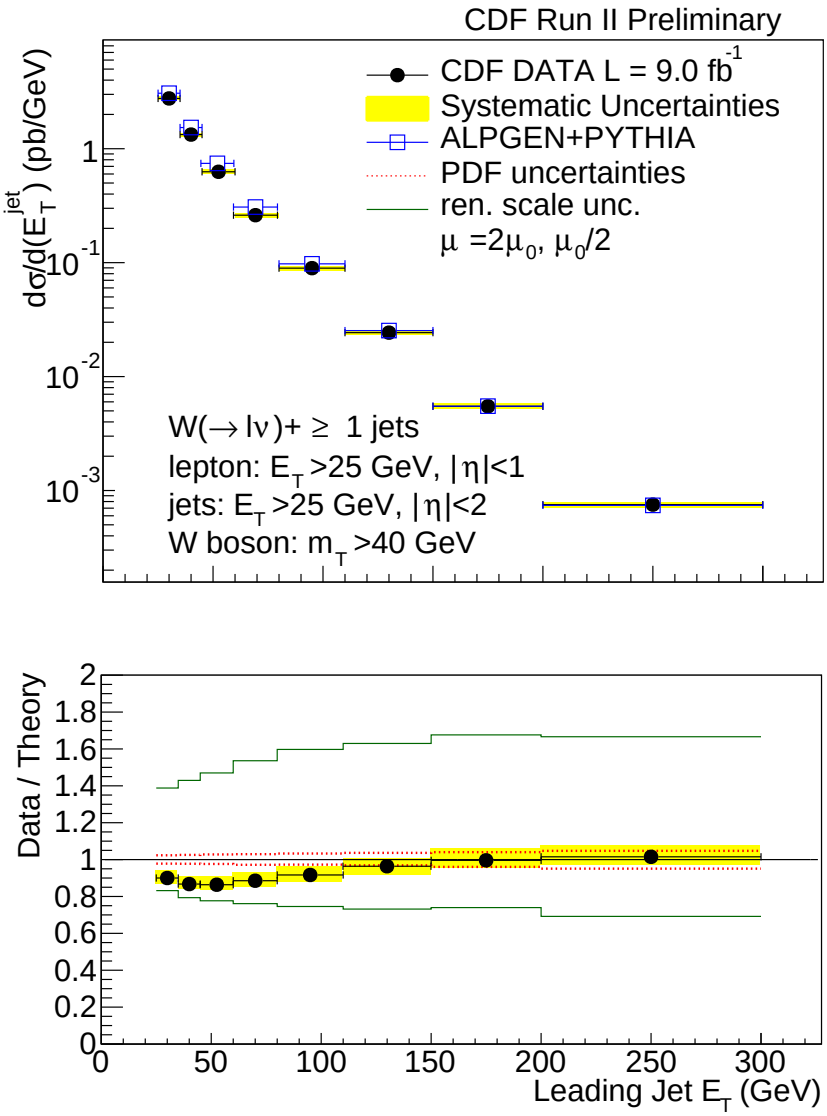


Figure 5.2: Measured inclusive jet differential cross sections (black dots) as a function of the leading jet E_T compared to LO calculations corrected by K-factor pQCD predictions (open squares). The shaded bands show the total systematic uncertainty, with the exclusion of the 6% luminosity uncertainty. The dashed indicate the PDF uncertainty whereas the solid lines the variation of the renormalization scale μ_0 as described in the text.

Whereas the choice of the LO-to-NLO K-factor appears to be appropriate the significance of this comparison is limited as shown by the uncertainties on the normalization scale (calculated as described above). This scale uncertainty reflects the

uncertainty connected with missing higher order corrections. Given the superior accuracy of these measurements, comparison with higher order calculations is very desirable.

Chapter 6

Conclusions

This thesis presents cross section measurements for the production of a W-boson in association with an inclusive number of jets. The entire CDF Run II dataset (integrated luminosity of 9.0 fb^{-1}) has been analyzed considering the events collected by the hight-p_T electron and muon triggers so that the W-bosons could be identified using both the electron and muon decay channels.

In order to optimize the measurement, the substantial backgrounds that enter the dataset under analysis were carefully studied. In particular, modeling of the multi-jet background has been the central focus of this work. For the high-p_T electron sample, we verified that the data-driven "nonelectron" model introduced for the specific study of the W+2 jets channel can describe the event kinematics of the whole W+jets sample provided all possible sources of systematic uncertainties related to the generation of the "nonelectron" data sample are considered. For the high-p_T muon sample, good modeling is achieved by introducing the "almost-isolated" data-driven sample together with the appropriate systematic uncertainties. Moreover, for a better description of the energy of the jets in the simulated backgrounds, the CDF standard energy scale corrections have been supplemented with a correction that accounts for the difference between quark-jets and gluon-jets.

The cross sections of both channels are unfolded to particle level for comparison with theory and then combined. The unfolding procedure, which relies on Monte Carlo simulations, corrects for all the detector effects, including bin migration due to limited detector resolution.

Perturbative QCD predictions, generated with ALPGEN interfaced with PYTHIA are compared with the measured cross-section as a function of the inclusive jet multiplicity and the leading jet E_T. The LO×K-factor ALPGEN+PYTHIA predictions are consistent with the measured cross sections, but the significance of this comparison is limited by the large uncertainty associated with the renormalization and factorization scale. It would be more significant to compare these results with next-to-leading order predictions. Comparison with NLO fixed order calculation is complicated by the infrared unsafe characteristic of the jet algorithm in use in this analysis [35]. An infrared and collinear safe jet algorithm would be preferable for such comparisons but the corrections necessary for good modeling of $V+$jets data (*e.g.*, quark/gluon jets corrections) were available only for the JETCLU algorithm. Studies to overcome this problem are contemplated as extension of this work.

Bibliography

Chapter 1

[1] CDF collaboration, T. Aaltonen et al., *Invariant-mass distribution of jet pairs produced in association with a W boson in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,TeV$ using the full CDF Run II data set*, *Phys. Rev.* **D 89**, 092001 (2014) [arXiv:1402.7044] [INSPIRE].

[2] S.L. Glashow, *Partial symmetries of weak interactions*, *Nucl. Phys.* **22**, 579 (1961) [INSPIRE].

[3] S. Weinberg, *A model of leptons*, *Phys. Rev. Lett.* **19**, 1264 (1967) [INSPIRE].

[4] A. Salam, *Weak and electromagnetic interactions*, originally printed in *Svartholm: Elementary Particle Theory, Proceedings Of The Nobel Symposium Held* 1968 *At Lerum, Sweden*, Stockholm Sweden (1968), pg. 367 [INSPIRE].

[5] G. 't Hooft, *Renormalizable Lagrangians for massive Yang-Mills fields*, *Nucl. Phys.* **B 35**, 167 (1971) [INSPIRE].

[6] Wikimedia Commons.

[7] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* **13**, 321 (1964) [INSPIRE].

[8] P.W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* **13**, 508 (1964) [INSPIRE].

[9] M.B. Robinson, K.R. Bland, G.B. Cleaver and J.R. Dittmann, *A simple introduction to particle physics. Part I — foundations and the Standard Model*, arXiv:0810.3328 [INSPIRE].

[10] Particle Data Group collaboration, K.A. Olive et al., *Review of particle physics*, *Chin. Phys.* **C 38**, 090001 (2014) [INSPIRE].

[11] S. Dürr et al., *Ab-initio determination of light hadron masses*, *Science* **322** (2008) 1224 [arXiv:0906.3599] [INSPIRE].

[12] J.D. Bjorken and E.A. Paschos, *Inelastic electron proton and gamma proton scattering and the structure of the nucleon*, *Phys. Rev.* **185**, 1975 (1969) [INSPIRE].

[13] R.P. Feynman, *The behavior of hadron collisions at extreme energies*, in *High Energy Collisions: Third International Conference at Stony Brook, N.Y.*, Gordon & Breach, U.S.A. (1969) [INSPIRE].

[14] V.N. Gribov and L.N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15**, 438 (1972) [*Yad. Fiz.* **15**, 781 (1972)] [INSPIRE];
G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, *Nucl. Phys.* **B 126**, 298 (1977) [INSPIRE];
Y.L. Dokshitzer, *Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics.*, *Sov. Phys. JETP* **46**, 641 (1977) [*Zh. Eksp. Teor. Fiz.* **73**, 1216 (1977)] [INSPIRE].

[15] R.K. Ellis, W.J. Stirling and B.R. Webber, *QCD and collider physics*, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **8**, 1 (1996) [INSPIRE].

[16] R. Placakyte, *Parton distribution functions*, in *Proceedings, 31st International Conference on Physics in collisions (PIC 2011)*, Vancouver Canada (2011) [arXiv:1111.5452] [INSPIRE].

[17] G.P. Salam, *Elements of QCD for hadron colliders*, lectures given at *Conference, 2009 European School of High-Energy Physics (ESHEP 2009)*, Bautzen Germany (2009) [arXiv:1011.5131] [INSPIRE].

[18] M.A. Dobbs et al., *Les Houches guidebook to Monte Carlo generators for hadron collider physics*, hep-ph/0403045 [INSPIRE].

LHAPDF website.

[19] G. Watt, *MSTW PDFs and impact of PDFs on cross sections at Tevatron and LHC*, *Nucl. Phys. Proc. Suppl.* **222-224**, 61 (2012) [arXiv:1201.1295] [INSPIRE].

[20] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07**, 012 (2002) [hep-ph/0201195] [INSPIRE].

[21] NNPDF collaboration, R.D. Ball et al., *A determination of parton distributions with faithful uncertainty estimation*, *Nucl. Phys.* **B 809**, 1 (2009) [*Erratum ibid.* **B 816**, 293 (2009)] [arXiv:0808.1231] [INSPIRE].

[22] N. Kobayashi, *Study of fragmentation from a quark to hadrons*, *J. Phys. Conf. Ser.* **302**, 012056 (2011) [INSPIRE].

[23] T. Sjöstrand, *Monte Carlo generators*, hep-ph/0611247 [CERN-LCGAPP-2006-06] [INSPIRE].

[24] Y.L. Dokshitzer, V.A. Khoze and S.I. Troian, *On the concept of local parton hadron duality*, *J. Phys.* **G 17**, 1585 (1991) [INSPIRE].

[25] J.E. Hutcj et al., *Toward a standardization of jet definitions*, in *Snowmass summer study*, World Scientific, Singapore (1990), pg. 134 [FERMILAB-CONF-90-249].

[26] G.C. Blazey et al., *Run II jet physics*, hep-ex/0005012 [INSPIRE].

[27] G.P. Salam, *Towards jetography*, *Eur. Phys. J.* **C 67**, 637 (2010) [arXiv:0906.1833] [INSPIRE].

[28] S.D. Ellis, J. Huston, K. Hatakeyama, P. Loch and M. Tonnesmann, *Jets in hadron-hadron collisions*, *Prog. Part. Nucl. Phys.* **60**, 484 (2008) [arXiv:0712.2447] [INSPIRE].

Chapter 2

[29] U. Baur and D. Wackeroth, *Electroweak radiative corrections to* $p\bar{p} \to W^{\pm} \to \ell^{\pm}\nu$ *beyond the pole approximation*, *Phys. Rev.* **D 70**, 073015 (2004) [hep-ph/0405191] [INSPIRE].

[30] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A.D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07**, 001 (2003) [hep-ph/0206293] [INSPIRE].

[31] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA* 6.4 *physics and manual*, *JHEP* **05**, 026 (2006) [hep-ph/0603175] [INSPIRE].

[32] F. Caravaglios and M. Moretti, *An algorithm to compute Born scattering amplitudes without Feynman graphs*, *Phys. Lett.* **B 358**, 332 (1995) [hep-ph/9507237] [INSPIRE].

[33] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A.D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07**, 001 (2003) [hep-ph/0206293] [INSPIRE].
M.L. Mangano, M. Moretti and R. Pittau, *Multijet matrix elements and shower evolution in hadronic collisions:* $Wb\bar{b} + n$ *jets as a case study*, *Nucl. Phys.* **B 632**, 343 (2002) [hep-ph/0108069] [INSPIRE].

[34] S. Catani, F. Krauss, R. Kuhn and B.R. Webber, *QCD matrix elements + parton showers*, *JHEP* **11**, 063 (2001) [hep-ph/0109231] [INSPIRE].

[35] R.K. Ellis, K. Melnikov and G. Zanderighi, W + 3 *jet production at the Tevatron*, *Phys. Rev.* **D 80**, 094002 (2009) [Fermilab-PUB-09-293-T] [arXiv:0906.1445] [INSPIRE].

[36] J.M. Campbell and R.K. Ellis, *MCFM for the Tevatron and the LHC*, *Nucl. Phys. Proc. Suppl.* **205-206**, 10 (2010) [FERMILAB-CONF-10-244] [arXiv:1007.3492] [INSPIRE].

[37] S. Frixione, F. Stoeckli, P. Torrielli, B.R. Webber and C.D. White, *The MC@NLO* 4.0 *event generator*, arXiv:1010.0819 [INSPIRE].

[38] CDF collaboration, T. Aaltonen et al., *Measurement of the cross section for* W^- *boson production in association with jets in* $p\bar{p}$ *collisions at* $\sqrt{s} =$ 1.96 *TeV*, *Phys. Rev.* **D 77**, 011108 (2008) [arXiv:0711.4044] [INSPIRE].

[39] D0 collaboration, V.M. Abazov et al., *Measurements of inclusive W+jets production rates as a function of jet transverse momentum in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Lett.* **B 705**, 200 (2011) [arXiv:1106.1457] [INSPIRE].

[40] CMS collaboration, V. Khachatryan et al., *Differential cross section measurements for the production of a W boson in association with jets in proton-proton collisions at $\sqrt{s} = 7$ TeV*, *Phys. Lett.* **B 741**, 12 (2015) [arXiv:1406.7533] [INSPIRE].

[41] ATLAS collaboration, G. Aad et al., *Measurements of the W production cross sections in association with jets with the ATLAS detector*, *Eur. Phys. J.* **C 75**, 82 (2015) [arXiv:1409.8639] [INSPIRE].

Chapter 3

[42] Link to Fermilab website.

[43] Link to Fermilab Beam Division (Concepts).

[44] Link to Fermilab Beam Division (LINAC).

[45] Link to Fermilab Beam Division (Booster).

[46] Link to Fermilab Beam Division (Main Injector).

[47] Link to Fermilab Beam Division (Recycler Ring).

[48] Link to Fermilab Beam Division (Tevatron).

[49] Link to Fermilab Sync Lite system website.

[50] Link to CDF website.

[51] CDF collaboration, D. Acosta et al., *Measurement of the J/ψ meson and b-hadron production cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1960$ GeV*, *Phys. Rev.* **D 71**, 032001 (2005) [hep-ex/0412071] [INSPIRE].

[52] CDF collaboration, T. Aaltonen et al., *Operational experience, improvements and performance of the CDF Run II silicon vertex detector*, *Nucl. Instrum. Meth.* **A 729**, 153 (2013) [FERMILAB-PUB-13-015-E] [arXiv:1301.3180] [INSPIRE].

[53] Link to CDF ToF webpage.

[54] CDF collaboration, A.A. Affolder et al., *CDF central outer tracker*, *Nucl. Instrum. Meth.* **A 526**, 249 (2004) [INSPIRE].

[55] J. Elias et al., *Luminosity monitor based on Cherenkov counters for $p\bar{p}$ colliders*, *Nucl. Instrum. Meth.* **A 441**, 366 (2000) [INSPIRE].

[56] CDF collaboration, A. Abulencia et al., *Measurements of inclusive W and Z cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *J. Phys.* **G 34**, 2457 (2007) [hep-ex/0508029] [INSPIRE].

[57] Link to CDF L2CAL UPGRADE website.

[58] CDF collaboration, J.A. Adelman et al., *The silicon vertex trigger upgrade at CDF*, *Nucl. Instrum. Meth.* **A 572**, 361 (2007) [INSPIRE].

[59] CDF collaboration, L. Balka et al., *The CDF central electromagnetic calorimeter*, *Nucl. Instrum. Meth.* **A 267**, 272 (1988) [INSPIRE].

[60] CDF collaboration, S. Bertolucci et al., *The CDF central and endwall hadron calorimeter*, *Nucl. Instrum. Meth.* **A 267**, 301 (1988) [INSPIRE].

[61] CDF collaboration, M.G. Albrow et al., *The CDF plug upgrade electromagnetic calorimeter: test beam results*, *Nucl. Instrum. Meth.* **A 480**, 524 (2002) [INSPIRE].

[62] CDF collaboration, M. Gallinaro et al., *A new scintillator tile/fiber preshower detector for the CDF central calorimeter*, *IEEE Trans. Nucl. Sci.* **52**, 879 (2005) [physics/0411056] [INSPIRE].

[63] A. Artikov et al., *CDF central preshower and crack detector upgrade*, FNAL-PUB-07-023-E, Fermilab, Batavia U.S.A. (2007) [arXiv:0706.3922] [INSPIRE].

[64] G. Apollinari, K.A. Goulianos, P. Melese and M. Lindgren, *Shower maximum detector for the CDF plug upgrade calorimeter*, *Nucl. Instrum. Meth.* **A 412**, 515 (1998) [INSPIRE].

[65] M.G. Albrow et al., *A preshower detector for the CDF plug upgrade: test beam results*, *Nucl. Instrum. Meth.* **A 431**, 104 (1999) [INSPIRE].

[66] CDF collaboration, C.M. Ginsburg, *CDF Run 2 muon system*, *Eur. Phys. J.* **C 33** (2004) S1002 [INSPIRE].

[67] CDF collaboration, T. Dorigo, *The muon system upgrade for the CDF II experiment*, *Nucl. Instrum. Meth.* **A 461**, 560 (2001) [INSPIRE].

[68] R. Downing et al., *Track extrapolation and distribution for the CDF-II trigger system*, *Nucl. Instrum. Meth.* **A 570**, 36 (2007) [physics/0606247] [INSPIRE].

[69] CDF collaboration, A. Canepa et al., *Level-3 calorimetric resolution available for the Level-1 and Level-2 CDF triggers*, arXiv:0810.3738 [INSPIRE].

[70] CDF collaboration, A. Lister, *The CDF L2 XFT trigger upgrade*, in *Proceeding of the 34th International Conference on High Energy Physics*, Philadelphia U.S.A. (2008) [arXiv:0810.3348] [INSPIRE].

[71] Trigger and Datasets Working Group collaboration, *Run II trigger table and datasets plan*, CDF Note 4718, Fermilab, Batavia U.S.A. (1998).

[72] CDF collaboration, D. Acosta 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 secondary vertex b-tagging*, *Phys. Rev.* **D 71**, 052003 (2005) [hep-ex/0410041] [INSPIRE].

[73] M. Trovato, *Search for WZ/ZZ production in the lepton(s) + MET + jets channel with the CDF experiment at the Tevatron collider*, Ph.D. thesis, Scuola Normale Superiore, Pisa Italy (2014) [FERMILAB-THESIS-2014-02] [INSPIRE].

[74] CDF collaboration, A. Bhatti et al., *Determination of the jet energy scale at the collider detector at Fermilab*, *Nucl. Instrum. Meth.* **A 566**, 375 (2006) [hep-ex/0510047] [INSPIRE].

[75] CDF collaboration, T. Aaltonen et al., *Precision top-quark mass measurements at CDF*, *Phys. Rev. Lett.* **109**, 152003 (2012) [arXiv:1207.6758] [INSPIRE].

[76] W.R. Ketchum, *Search for $ZW/ZZ \to \ell^+\ell^-$ + jets production in $p\bar{p}$ collisions at CDF*, Ph.D. thesis, University of Chicago, Chicago U.S.A. (2012) [FERMILAB-THESIS-2012-36] [INSPIRE].

[77] CDF collaboration, T.A. Aaltonen et al., *Search for the production of ZW and ZZ boson pairs decaying into charged leptons and jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev.* **D 88**, 092002 (2013) [arXiv:1310.0086] [INSPIRE].

Chapter 4

[78] B. Cooper and A. Messina, *Estimation of the background to $W^{\pm} \to e^{\pm}\nu + n$ jet events*, CDF Note 7760, Fermilab, Batavia U.S.A. (2005).

[79] CDF collaboration, T. Aaltonen et al., *Invariant mass distribution of jet pairs produced in association with a W boson in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev. Lett.* **106**, 171801 (2011) [arXiv:1104.0699] [INSPIRE].

[80] G. Bellettini et al., *Search for WZ/ZZ production in events with lepton(s) plus jets plus missing transverse energy*, CDF Note 10837, Fermilab, Batavia U.S.A.

[81] V. Boisvert, *Trigger efficiencies for the high E_T central electrons in the Gen6 data*, CDF Note 7939, Fermilab, Batavia U.S.A. (2006).

[82] U. Grundler, A. Taffard and X. Zhang, *High-p_T muons recommended cuts and efficiencies for Summer* 2006, CDF Note 8262, Fermilab, Batavia U.S.A. (2006).

[83] M. Griffiths, B. Heinemann and G. Manca, *Fake rate for low-p_T leptons*, CDF Note 7470, Fermilab, Batavia U.S.A. (2005).

[84] CDF collaboration, T. Aaltonen et al., *Invariant-mass distribution of jet pairs produced in association with a W boson in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\ TeV$ using the full CDF Run II data set*, *Phys. Rev.* **D 89**, 092001 (2014) [arXiv:1402.7044] [INSPIRE].

[85] R. Brun, F. Carminati and S. Giani, *GEANT detector description and simulation tool*, CERN Program Library Long Writeup, CERN, Geneva Switzerland (1994).

[86] F. Maltoni and T. Stelzer, *MadEvent: automatic event generation with MadGraph*, *JHEP* **02**, 027 (2003) [hep-ph/0208156] [INSPIRE].

[87] J. Alwall et al., *MadGraph/MadEvent v4: the new web generation*, *JHEP* **09**, 028 (2007) [arXiv:0706.2334] [INSPIRE].

[88] CTEQ collaboration, H.L. Lai et al., *Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions*, *Eur. Phys. J.* **C 12**, 375 (2000) [hep-ph/9903282] [INSPIRE].

[89] G. Grindhammer, M. Rudowicz and S. Peters, *The fast simulation of electromagnetic and hadronic showers*, *Nucl. Instrum. Meth.* **A 290**, 469 (1990) [INSPIRE].

[90] E. Gerchtein and M. Paulini, *CDF detector simulation framework and performance*, talk from the 2003 *Computing in High Energy and Nuclear Physics (CHEP03)*, La Jolla U.S.A. (2003) [*eConf* **C 0303241** (2003) TUMT005] [physics/0306031] [INSPIRE].

[91] J.M. Campbell and R.K. Ellis, *An update on vector boson pair production at hadron colliders*, *Phys. Rev.* **D 60**, 113006 (1999) [hep-ph/9905386] [INSPIRE].

[92] M. Czakon, P. Fiedler and A. Mitov, *Total top-quark pair-production cross section at hadron colliders through $O(\alpha_S^4)$*, *Phys. Rev. Lett.* **110**, 252004 (2013) [arXiv:1303.6254] [INSPIRE].

[93] N. Kidonakis, *Single top quark production cross section at hadron colliders*, PoS(DIS 2010)196 [arXiv:1005.3330] [INSPIRE];
N. Kidonakis, *NNLL resummation for s-channel single top quark production*, *Phys. Rev.* **D 81**, 054028 (2010) [arXiv:1001.5034] [INSPIRE].

[94] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, *Phys. Rev.* **D 83**, 091503 (2011) [arXiv:1103.2792] [INSPIRE].

[95] CDF collaboration, T. Aaltonen et al., *Measurement of cross sections for b jet production in events with a Z boson in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\ TeV$*, *Phys. Rev.* **D 79**, 052008 (2009) [arXiv:0812.4458] [INSPIRE].

[96] CDF collaboration, T. Aaltonen et al., *Observation of single top quark production and measurement of $|V_{tb}|$ with CDF*, *Phys. Rev.* **D 82**, 112005 (2010) [arXiv:1004.1181] [INSPIRE].

[97] CDF collaboration, T.A. Aaltonen et al., *Measurement of differential production cross section for Z/γ^* bosons in association with jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,TeV$*, *Phys. Rev.* **D 91**, 012002 (2015) [arXiv:1409.4359] [INSPIRE].

[98] B. Cooper, A. Messina and D. Waters, *Measurement of the $W(\to e\nu)+ \geqslant n$ jet differential cross-section using $320\,\mathrm{pb}^{-1}$ of CDF Run II data*, CDF Note 8046, Fermilab, Batavia U.S.A.

[99] W. Ketchum, V. Rusu and M. Trovato, *New jet energy scale corrections for quarks and gluons*, CDF Note 10829, Fermilab, Batavia U.S.A. (2012).

[100] O. Behnke, K. Kroninger, G. Schott and T. Schorner-Sadenius, *Data analysis in high energy physics: a practical guide to statistical methods*, first edition, Wiley-VCH, Germany (2013).

[101] A. Denner, S. Dittmaier, T. Kasprzik and A. Mück, *Electroweak corrections to $W+$ jet hadroproduction including leptonic W-boson decays*, *JHEP* **08**, 075 (2009) [arXiv:0906.1656] [INSPIRE].

[102] A. Höcker and V. Kartvelishvili, *SVD approach to data unfolding*, *Nucl. Instrum. Meth.* **A 372**, 469 (1996) [hep-ph/9509307] [INSPIRE].

[103] R. Brun and F. Rademakers, *ROOT: an object oriented data analysis framework*, *Nucl. Instrum. Meth.* **A 389** (1997) 81 [INSPIRE];
ROOT webpage, http://root.cern.ch/.

[104] L. Lyons, D. Gibaut and P. Clifford, *How to combine correlated estimates of a single physical quantity*, *Nucl. Instrum. Meth.* **A 270**, 110 (1988) [INSPIRE].

[105] CDF collaboration, R.C. Group, C.I. Ciobanu, K. Lannon and C. Plager, *Combination of single top quark production results from CDF*, in *Proceedings of the 34^{th} International Conference in High Energy Physics (ICHEP08)*, Philadelphia U.S.A. (2008) [arXiv:0809.4670] [INSPIRE].

Chapter 5

[106] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07**, 012 (2002) [hep-ph/0201195] [INSPIRE].

Acknowledgements

I must begin by expressing my sincere gratitude to my advisor Prof. Giovanni Pauletta, without whom this work would not have been possible. He guided me over the years and helped me develop and present this thesis.

My special thanks to Marco Trovato and Vadim Rusu for guiding me through the analysis and for always having suggestions to understand and solve the problems I encountered. They always found time for a meeting, no matter in what country they were.

I am thankful to all members of the CDF collaboration, in particular to the Spokepersons Luciano Ristori, Costas Vellidis, David Toback and Giorgio Bellettini who followed my progress during the CDF meetings and gave me a lot of suggestions.

I would like to thank the Professors of my institution, Alessandro De Angelis, Barbara De Lotto, Massimo Persic, Stefano Ansoldi, Diego Cauz and Lorenzo Santi from whom I learned many things about high energy physics, astrophysics, general relativity and more. A double thanks to Massimo Persic, his enthusiasm for physics inspired me.

I want to thank the other students that walked through this path with me: Barbara, Carlo, Michele, Michele who shared the long days at the university, Emanuele for all the discussions about mass (theoretical vs. experimental point of view) and Alessandra to be my "jet sister". I cannot forget Luigi, Manuel, Giulia and Jon who made Fermilab feel like home.

I would also like to extend my most heartfelt thanks to these friends for their strong support and timely encouragement, Ilaria and Maurizio, Giulia and Christian, Gianpaola and Gaia, Marco, Maura, Amedeo, Paola, my brother Nicola, and everybody else with whom I shared good times through these years.

I am extremely thankful to Barbara and Peter for all they did for me and for always being my biggest fans.

The biggest thanks goes to my parents for their continuous love and support.

Finally, thank you Jeff for all your patience and for staying next to me "no matter what".