Università degli Studi di Pisa

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI

Corso di Laurea Specialistica in Scienze Fisiche

Tesi di laurea specialistica

Improving the Acceptance in the Single Top Quark Analysis using 7.5 fb^{-1} of CDF data

Candidato:
Manfredi Ronzani

Relatore:
Prof. Giorgio Chiarelli

Anno Accademico 2011-2012

Contents

Introduction

The Top quark was observed at the Fermilab Tevatron by the CDF and D0 collaboration in 1995 via strong production of top quarks pairs at center-of-mass energy of $\sqrt{s}$=1.8 Tev, using 60 pb^{-1} of data [1]. Standard Model of elementary particles predicts top quark to be produced individually via electroweak interactions in three different channels (s, t and Wt). This is defined as "single top" (from now on also ST). ST was observed by CDF collaboration in 2009 at center-of-mass energy of $\sqrt{s}$=1.96 Tev using 3.2 fb^{-1} of $p\bar{p}$ collected data [2]. This thesis concerns the upgrade of this measurement using 7.5 fb^{-1} of data.

There are several reasons to search for Single Top quarks production, besides the interest of the process itself. The ST cross-section measurement provides a direct determination of the $|V_{tb}|$ element of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. SM predicts a value of $|V_{tb}| \approx 1$, in order to mantain the unitarity of 3x3 CKM matrix. A significant deviation from unity may imply a fourth generation of fermions and open the road to new physics.

Another important reason to study this process, even in absence of new phenomena, is its connection with one of the most important topic of modern HEP research, that is to directly observe the Higgs boson. Single Top production is one of the background processes of $WH \rightarrow l\nu b\bar{b}$ channel, as they share very similar final states. Also, Single Top and WH processes share the same backgrounds. Therefore, understanding the background and perform a direct measurement of Single Top production may help to optimize the Higgs boson searches in the WH channel.

First observation by CDF [2] measured the cross section $\sigma_{s+t} = 2.3^{+0.6}_{-0.5}$ pb (stat+syst) for the s- and t- combined channels while measured the individual cross sections $\sigma_s = 1.8^{+0.7}_{-0.5}$ pb and $\sigma_t = 0.8^{+0.4}_{-0.4}$ pb for s-channel and t-channel respectively. These results show a tension with the SM expected value at NLO of $\sigma_s = 0.88 \pm 0.11$ pb and $\sigma_t = 1.98 \pm 0.25$ pb [3] [4] (Fig. 1). A value of $|V_{tb}| = 0.91 \pm 0.11$(stat.+syst)$\pm$0.07(theory) and a 95% c.l. limit of $|V_{tb}| > 0.71$ were obtained using this result.

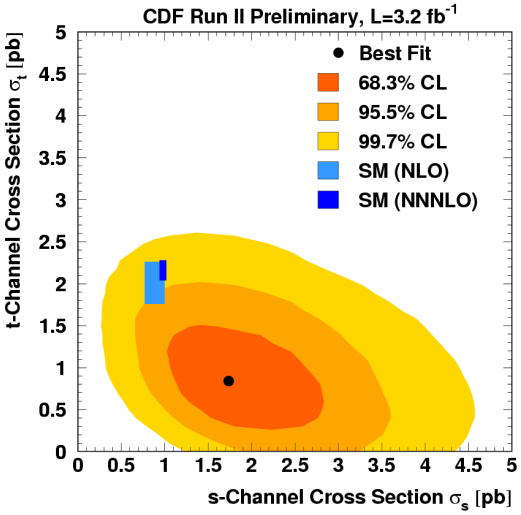


Figure 1: The results of the two-dimensional fit for σ_s and σ_t for the 3.2 fb^{-1} analysis.

This thesis was centered on the event selection and background estimate. The analysis is based on the WHAM (WH Analysis Module) package which is the new platform with common selection tools and same lepton definitions for all WH searches in the WH-SingleTop group at CDF (see Sec. 3.3.1 and App. C). I adapted WHAM to be able to select ST events, by exploiting some specific kinematics (more jets in the forward region). First, the event selection of this new analysis was validated reproducing exactly the 3.2 fb^{-1} acceptance (App. D). Then, we implemented a different selection from the one used in [2]. The other important change with respect to the previous analysis is that I included a new lepton cathegory (EMC) in order to recover muons lost in detector cracks.

The measurement of Single Top quark cross section presents substantial experimental challenges compared to the $t\bar{t}$ case because of the lower SM production rate and the larger kinematically similar backgrounds. A simple counting events analysis will not provide a precise measurement of the single top cross section, no matter how much data are accumulated. Further seperation of the signal from the background is required. In this analysis, we use the artificial neural networks NeuroBayes® package, which exploits some important input variables to separate signal from background. For the sake of completeness we will present the final results of the whole NN analysis in the last chapter (Cap. 8), even though the NN part was not the specific task of my work and it is essentially the same as for the original analysis.

The final results obtained by CDF for the single top quark cross section is $\sigma_{s+t} = 3.04^{+0.57}_{-0.53}$ pb. Thanks to the larger statistics we were also able to separate s- and t- channel and obtain the values $\sigma_s = 1.81^{+0.63}_{-0.58}$ pb and $\sigma_t = 1.49^{+0.47}_{-0.42}$ pb. We also extract a value of $|V_{tb}| = 0.92^{+0.10}_{-0.08}(\text{stat} + \text{sys}) \pm 0.05(\text{theory})$ and a 95% confidence level lower limit of $|V_{tb}| > 0.78$.

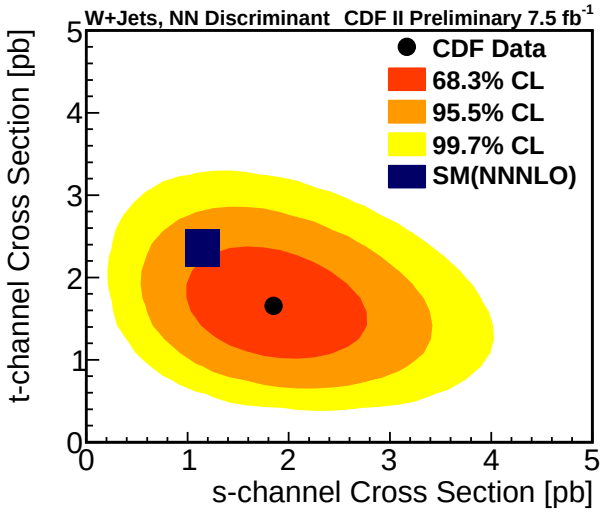


Figure 2: The results of the two-dimensional fit for σ_s and σ_t. The black point shows the best fit value, and the 68.3%, 95.5%, and 99.7% credibility regions are shown as shaded areas. The SM predictions are also indicated with their theoretical uncertainties.

Chapter 1

Standard Model and Single-Top production

At the present, in Nature we know four fondamental forces: gravitational, electromagnetic, weak and strong nuclear forces. Most of the efforts in theoretical physics are directed towards the unification of all interactions under a complete theory. Today, three out of four (electromagnetic, weak and strong nuclear force) are included in a larger frame called *Standard Model* (SM) of the fondamental interactions. As it does not include the gravitation in a complete ultimate theory, the search for such a universal model is still ongoing.

The SM is an elegant and calculable theory that describes most of the elementary particle physics processes. It successfully predicted the existence of new particles afterwards discovered in high energy physics experiments, like the W and Z bosons.

The evidence for a signal consistent with the SM Higgs boson was reported at LHC in 2012 by ATLAS [82] and CMS [83] collaborations. Its observation would be a crucial verification of SM. On the other hand, even if different models were proposed to replace or extend the SM[1], there is still no experimental evidence of physics processes with respect to large discrepancy with SM expectations.

1.1 Particles and Fields: a Gauge Theory

The SM is a quantum field theory in which particles are treated as quantum local fields interacting via the exchange of some force-mediators (each type of force has its own mediators), that are introducted in order to conserve local symmetries i.e. the theory has to be *gauge invariant* under local trasformations [5]. The SM

[1] the so-called *Physiscs Beyond Standard Model*

is based on gauge symmetry group $SU(3) \otimes SU(2) \otimes U(1)$: SU(3) describes QCD (Quantum-Chromo-Dynamics) theory of strong interactions, SU(2) describes weak theory and U(1) the QED (Quantum-Electro-Dynamics) theory of electomagnetims.

As summarized in table 1.1, elementary particles [2] are divided in two fondamental groups by their spin: leptons and quarks, with half integer spin, and bosons, with integer spin. Quarks and leptons follow Fermi-Dirac statistics, therefore are called *fermions* while bosons follow Bose-Einstein statistics.

	Name	Symbol	Spin($\hbar$)	Charge(e)	Mass(MeV/c^2)
Leptons	e neutrino	ν_e	1/2	0	$< 2 \cdot 10^{-6}$
	electron	e	1/2	-1	$0.510998910 \pm 0.000000013$
	μ neutrino	ν_μ	1/2	0	< 0.19
	muon	μ	1/2	-1	$105.6583668 \pm 0.0000038$
	τ neutrino	ν_τ	1/2	0	< 18.2
	tau	τ	1/2	-1	1776.82 ± 0.16
Quarks	up	u	1/2	2/3	$2.49\pm^{0.81}_{0.79}$
	down	d	1/2	-1/3	$5.5\pm^{0.75}_{0.95}$
	charm	c	1/2	2/3	$(1.27\pm^{0.07}_{0.09}) \cdot 10^3$
	strange	s	1/2	-1/3	$101\pm^{29}_{31}$
	beauty	b	1/2	2/3	$(4.19\pm^{0.18}_{0.06}) \cdot 10^3$
	top	t	1/2	-1/3	$(172.0\pm^{0.9}_{1.3}) \cdot 10^3$
Gauge Bosons	photon	γ	1	0	0
	W boson	$W^\pm$	1	± 1	$(80.399 \pm 0.023) \cdot 10^3$
	Z boson	Z^0	1	0	$(91.1876 \pm 0.0021) \cdot 10^3$
	gluon	g	1	0	0

Table 1.1: Properties of elementary particles in SM (numbers from PDG [6]).

1.1.1 Fermions: Lepton and Quarks

The matter, as we know so far[3], is made of fermions that are divided in two classes: leptons and quarks. For each fermion exists a corresponding anti-particle with same mass and opposite quantum numbers. Each class can be ordered in three generations - also called *families* - (see table 1.1).Only first generation particles can form stable matter, while second and third generations fermions are created high

[2] A particle is defined as *elementary* if is not possible to describe its quantum state as a combination of other particles.

[3] The so-called *Dark Matter*, supposed to be made of non-barionic particles is not included in our description.

energy processes and subsequently decay into first generation states[4].

Quarks exist in six differents types, called flavors, as shown in table 1.1, with a fractional electric charge of 2/3 and -1/3. They interact through EM, weak and strong interaction. Quarks can be arranged by strong interaction in bound states, called hadrons, of $q\bar{q}$ pairs (mesons) or three quarks (barions). As it is forbidden to have fermions in the same state, due to Pauli's principle, quarks must have a new quantum number, that is called *color* and comes in three different types: red, yellow, blue. Bound states of quarks are color-neutral. Gluons, the strong-force mediators, carry the color number itself, hence are self-coupled. This cause an increase of the force strength with the distance, so that quarks are tightly forced to be bound in hadrons (*confinement*) and they cannot be observed as free particles.

Leptons interact via weak force and, if they carry electric charge, also via EM. They carry a *leptonic* number that characterizes each family. Neutrinos were originally assumed as massless while several measurement showed the presence of non-zero mass[7, 8].So far no direct measurement of ν mass exists [6].

1.1.2 Gauge Theory: Bosons and interactions

Interactions between particles in SM are mediated by spin-1 particles called bosons: photon (γ) mediates electromagnetic interaction, the $W^{\pm}$ and Z^0 mediates weak interaction and color-carrying gluons (g) mediates strong interactions. Those particles are needed in the SM in order to mantain local invariance of the Lagrangian density function of the physical system. In fact, local phase transformations of the free Lagrangian density produce the need to introduce gauge fields (i.e. the gauge bosons) to preserve local-gauge invariance. The invariance can be seen as a symmetry of the function and, according to Noether's theorem, when a symmetry is introducted in the system, it must exist a conserved current verifying continuity equation:

$$\partial_\mu J^\mu = 0 \tag{1.1}$$

Since the conservation of a current is associated with conservation of a charge (the time-component J^0 of 4-vector J^μ integrated over the space), it means that each interaction force has to conserve a quantum number: the electric charge Q in EM, the weak isospin charge I_3^W (and the associated weak hypercharge $Y = Q/e - I_3^W$) in weak interaction and color charge C_q in strong interaction [9]. Coupling with those

[4] the existence of a fourth generation, in principle not forbidden by the theory, involves directly single-top measurement, since $\sigma_{ST} \sim |V_{tb}|$.

charges, the gauge fields generates the interactions with strengths characterized by coupling constants with approximate relative magnitudes:

$$\alpha_{strong} : \alpha_{em} : \alpha_{weak} \approx \frac{1}{10} : \frac{1}{100} : \frac{1}{10000} \tag{1.2}$$

Typical lifetimes of processes belonging to those interactions are (in seconds):

$$\tau_{strong} \approx 10^{-23}, \quad \tau_{em} \approx 10^{-20} - \ 10^{-16}, \quad \tau_{weak} \approx 10^{-12} \tag{1.3}$$

QED

The quantum field theory of electromagnetism is quantum-electro-dynamics (QED). The Lagrangian density function in QED for a free-fermion is:

$$\mathscr{L}_0 = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) \tag{1.4}$$

where ψ is the Dirac field of mass m. Although 1.4 is already invariant for *global* U(1) transormation $\psi \to e^{-iqf}\psi$, it is not under *local* U(1) tranformation (i.e. with a space-time dependent differentiable function f(x)):

$$\psi \to e^{-iqf(x)}\psi \tag{1.5}$$

Through the minimal substitution of the derivative with covariant derivative:

$$\partial_\mu \to D_\mu \equiv (\partial_\mu + iqA_\mu(x)) \tag{1.6}$$

where A_μ is the real electromagnetic field transforming as

$$A_\mu \to A_\mu \ - \frac{1}{q}\partial_\mu f(x) \tag{1.7}$$

the Lagrangian becomes invariant in the final form

$$\mathscr{L}_{QED} = \mathscr{L}_0 - q\bar{\psi}(x)\gamma^\mu\psi(x)A_\mu(x) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.8}$$

The second term in 1.8 represents the interaction between charged particles through the gauge field A_μ, i.e. the exchange of a photon. Local invariance also requires photon to be massless (otherwise, a term as $\frac{1}{2}mA_\mu A^\mu$ would emerge from the calculation), as experimentally verified. Finally, the masslessness of the photon also implies that electromagnetic interaction has infinite range.

Weak interaction

The first theory for weak processes was proposed by E.Fermi in 1934 [10]. In order to explain β-decay $n \to pe^-\bar{\nu}_e$, he introduced a current-current amplitude of the form:

$$\mathscr{M} = G_F(\bar{u}_p\gamma^\mu u_n)(\bar{u}_e\gamma_\mu u_{\nu_e}) \tag{1.9}$$

with the Fermi coupling constant $G_F/(\hbar c)^3 = 1.166 \cdot 10^{-5}$ GeV^{-2}. In 1957 C. S. Wu's experiment showed parity violation of ^{60}Co decay in ^{60}Ni that was not explained by Fermi's theory, therefore a new form of interaction was suggested by replacing γ^μ in Eq. 1.9 with $\gamma^\mu(1-\gamma^5)$. This is the so-called V-A structure of weak interactions that couples differently with left-handed and right-handed components of the spinors $\psi = \psi_L + \psi_R$ (i.e. the fermions).

In high energy approximation[5] leptonic currents involve only left-handed lepton fields. Limited to the case of leptonic interactions[6] (i.e. for e, μ, τ and their neutrinos), for any Dirac spinor $\psi(x)$ we can define the weak *isospin* doublet:

$$\Psi_l^L(x) \equiv \begin{pmatrix} \psi_{\nu_l}^L(x) \\ \psi_l^L(x) \end{pmatrix} \tag{1.10}$$

with

$$\psi_{l,\nu_l}^L(x) = \frac{1}{2}(1-\gamma_5)\psi_{l,\nu_l}(x) \tag{1.11}$$

The formulation of weak interaction as gauge theory relies on the invariance under SU(2) local phase transformations[7]:

$$\Psi_l^L(x) \to e^{ig\alpha_j(x)\tau_j/2}\Psi_l^L(x), \qquad \bar{\Psi}_l^L(x) \to \bar{\Psi}_l^L(x)e^{-ig\alpha_j(x)\tau_j/2} \tag{1.12}$$

where g is the weak coupling constant, τ_j are Pauli spin matrices and $\alpha_j(x)$ three arbitrary real differentiable functions of x. As it can be shown, we can obtain from this invariance three conserved weak currents:

$$J_j^\alpha(x) = \frac{1}{2}\bar{\Psi}_l^L(x)\gamma^\alpha\tau_j\Psi_l^L(x), \quad j = 1,2,3 \tag{1.13}$$

and finally the corresponding conserved charges, called *weak isospin charges*:

[5] this weak theory is gauge-invariant only if leptons and bosons are considered massless. For high energy approximation we refer to E≫m.

[6] quarks weak interactions are similarly described

[7] SU(2) transformation operators $U(\alpha) \equiv e^{i\alpha_j\tau_j}/2$ forms a non -Abelian symmetry group

$$I_j^W = \int d^3x J_j^0(x), \quad j = 1, 2, 3 \tag{1.14}$$

The third current J_3^α is called *neutral* current, because it couples also to electrically neutral leptons (i.e. neutrinos). As said above, we can define the *weak hypercharge* from the electric and weak charge:

$$\frac{Y}{2} = Q/e - I_3^W \tag{1.15}$$

so that left-handed ν_l has $I_3^W = \frac{1}{2}$,$Y = -1$ and the left-handed charged lepton has $I_3^W = -\frac{1}{2}$,$Y = -1$.

Flavor mixing: CKM matrix

Leptons form SU(2) doublets under the weak interaction:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}, \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{1.16}$$

For quarks[8], it must be took into account the experimentally verified flavor change via W boson exchange[9], so that hadronic coupling is possible using the CKM rotation on quarks d,s,b. In this case, the flavor states are different from the mass states:

$$\begin{pmatrix} u \\ d' \end{pmatrix}, \quad \begin{pmatrix} c \\ s' \end{pmatrix}, \quad \begin{pmatrix} t \\ b' \end{pmatrix} \tag{1.17}$$

where Cabibbo-Kobayashi-Maskawa (CKM) matrix (an unitary 3X3 matrix) operates on mass states d,s,b:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.18}$$

Electro-Weak Unification

In 1961, Glashow [11] proposed an unified gauge theory for QED and weak interactions, based on $SU(2) \otimes U(1)$ group symmetry. Similar to QED, even weak lagrangian density can be made invariant under local gauge transformations (Eq. 1.12) through replacement of the derivative and the introduction of gauge fields. If we make both replacements valid for U(1) and SU(2) at same time:

[8] leptons don't mix the families in SM, cause neutrinos are massless.

[9] g.e. in $K^+ \to \mu^+ \nu_\mu$ decay, K meson is made by $u\bar{s}$ quarks.

$$\partial^\mu \to D^\mu = (\partial^\mu + ig\tau_j W_j^\mu(x)/2 - ig' B^\mu(x)/2) \tag{1.19}$$

where $\bar{W}^\mu$ is the real vector gauge field for weak interactions (according to SU(2) symmetry) and B^μ is the real gauge field for QED (according to U(1)), we obtain the invariant leptonic electro-weak Lagrangian density in the form $\mathscr{L}^L = \mathscr{L}_0 + \mathscr{L}_I$. In order to have a version of the function in the form

$$\mathscr{L}_I = \mathscr{L}_{CC} + \mathscr{L}_{NC} \tag{1.20}$$

where CC and NC indicate charged and neutral currents. It is usually used a linear combination of W_1^μ, W_2^μ for weak charged current while we write photon and Z^0 boson as linear combinations of B^μ and W_3^μ

$$W_\mu^{(\dagger)} = \frac{1}{\sqrt{2}}(W_{1\mu} \mp W_{2\mu}) \tag{1.21}$$

$$A_\mu = B_\mu \cos\vartheta_W + W_\mu^3 \sin\vartheta_W, \quad Z_\mu = -B_\mu \sin\vartheta_W + W_\mu^3 \sin\vartheta_W \tag{1.22}$$

$$g \sin\vartheta_W = g' \cos\vartheta_W = e \tag{1.23}$$

where ϑ_W is the Weinberg angle with a measured value of $\sin^2\vartheta_W = 0.2312 \pm 0.0006$ [6].

As we can see, from previuos formula it is straightforward to interpret the quanta of gauge fields as the EM and weak force mediators $\gamma, W^\pm, Z^0$.

So far, we have considered the fields as massless, as in Glashow's first theory, because of the gauge symmetry of the system. The Higgs mechanism was introduced to solve this problem adding the mass term for lepton and boson field, preserving at the same time gauge invariance.

1.1.3 Higgs mechanism: spontaneous symmetry breaking

The Higgs mechanims was proposed at the beginning of the '60s by several authors (Higgs, Englert, Guralnik et al.) [12, 13, 14] and was fully incorporated into the SM by Weinberg and Salam [15, 16]. It relies on the idea to have a Lagrangian density invariant under a symmetry group of transformations that produces degenerate asymmetric states. Arbitrary selection of one of these states produces a *spontaneous symmetry breaking*.

The core of the mechanism is to find a non-unique ground state (i.e. the *vacuum*) that implies a non-vanishing quantity in the system. This quantity will be assumed as the vacuum expectation value of quantized field.

In its simplest configuration the Higgs mechanism can be shown in a lagrangian density of scalar electrodynamics:

$$\mathscr{L}(x) = (D^{\mu}\varphi(x))^{*}(D_{\mu}\varphi(x)) - \mu^{2}\varphi^{*}(x)\varphi(x) - \lambda(\varphi^{*}(x)\varphi(x))^{2} - \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) \tag{1.24}$$

where D^{μ} is the covariant derivative 1.6 so that $\mathscr{L}$ is invariant under U(1) gauge transformations 1.5, λ and μ^2 are real parameters and $\varphi(x)$ is a complex scalar field that interact with the A_{μ} gauge field defined as usual by $F_{\mu\nu} = \partial_{\nu}A_{\mu} - \partial_{\mu}A_{\nu}$.

The potential part of the field $\mathscr{V}(\varphi) = \mu^{2}\varphi^{*}(x)\varphi(x) + \lambda(\varphi^{*}(x)\varphi(x))^{2}$ depends on λ, μ^2 values. If we take $\lambda > 0$, two situations occour for the bounded potential: for $\mu^{2} > 0$ $\mathscr{V}(\varphi)$ has an absolute miminum at $\varphi(x) = 0$, while for $\mu^{2} < 0$ $\mathscr{V}(\varphi)$ possesses a local minimum at $\varphi(x) = 0$ and a whole circle of absolute minima, leading to a set of degenerate states i.e. different *vacuum states* (see fig. 1.1):

$$\varphi(x) = \varphi_{0} = \left(\frac{-\mu^{2}}{2\lambda}\right)^{\frac{1}{2}} e^{i\vartheta}, \quad 0 \leq \vartheta < 2\pi \tag{1.25}$$

where the phase angle ϑ defines a direction in the complex φ-plane. Thanks to the invariance, gauge freedom allows to choose $\vartheta = 0$, leading to the *real* value $\varphi_{0} = \left(\frac{-\mu^{2}}{2\lambda}\right)^{\frac{1}{2}} \equiv \frac{1}{\sqrt{2}}\, v \;\; (> 0)$.

At the end, the original scalar field $\varphi(x)$ can be expressed as a real field[10] in a perturbative expansion of the chosen minumum:

$$\varphi(x) = \frac{1}{\sqrt{2}}[v + \sigma(x)] \tag{1.26}$$

At this point we can say that symmetry of $\mathscr{V}(\varphi)$ has been removed and ground states has *broken* the symmetry. Substituting 1.26 in 1.24 gives a form like $\mathscr{L}(x) = \mathscr{L}_{0}(x) + \mathscr{L}_{I}(x)$:

$$\begin{aligned} \mathscr{L}(x) \;\; = \;\; & \frac{1}{2}(\partial_{\mu}\sigma(x))(\partial^{\mu}\sigma(x)) - \frac{1}{2}(2\lambda v^{2})\sigma^{2}(x) \\ & -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \frac{1}{2}(qv^{2})A_{\mu}(x)A^{\mu}(x) \\ & -\lambda v\sigma^{3}(x) - \frac{1}{4}\lambda\sigma^{4}(x) \\ & +\frac{1}{2}q^{2}A_{\mu}(x)A^{\mu}(x)[2v\sigma(x) + \sigma^{2}(x)] \end{aligned} \tag{1.27}$$

[10] since $\varphi(x)$ is originally complex we should write a form as $\varphi(x) = \frac{1}{\sqrt{2}}[v + \sigma(x) + i\eta(x)]$, but it can be shown that this leads to the presence of un-physical terms in the $\mathscr{L}$, so that we can perform a U(1) tranformation on $\varphi(x)$ an obtain the real form as in 1.26

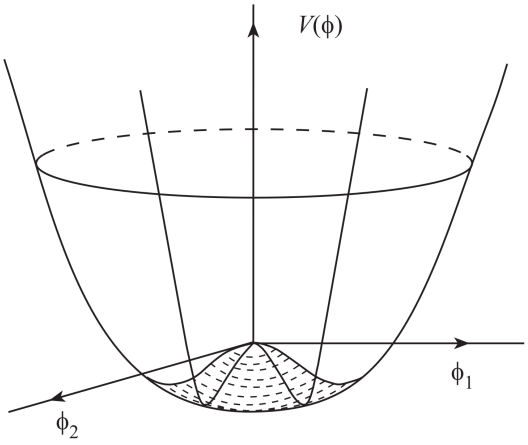


Figure 1.1: Symmetry breaking for a complex scalar field $\varphi(x) = 1/\sqrt{2}[\varphi_1(x) + i\varphi_2(x)]$ choosing $\mu^2 > 0$ in $\mathscr{V}(\varphi)$

The interpretation of the 1.27 terms brings to some crucial considerations. First two lines are $\mathscr{L}_0$, where we have a free scalar field $\sigma(x)$ with a mass $M_H = \sqrt{2\lambda v^2}$ and a *massive* vector field $A_\mu(x)$ of mass $|qv|$ replacing the initial massless gauge field (!). Second two lines represents the $\mathscr{L}_I$, with self-interaction of the scalar field and the interaction with the vector field with coupling strengths of $q^2 v$ and $\frac{1}{2}q^2$.

What we got is just the *spontaneous symmetry breaking* coming from the assigment of one initial degree of freedom of φ to the vector field A_μ, which then aquires mass. The massive spin 0 boson associated with the field $\sigma(x)$ is called *Higgs boson.*

Symmetry breaking in SM

Since electro-weak theory is $SU(2) \otimes U(1)$ gauge invariant, the Higgs mechanism must be adapted to such a symmetry group in order to produce the lepton and boson masses and finally arrive to the Weinberg and Salam formulation [15]. Here we give a brief description of that procedure, while the details can be found in [9].

The basic idea is to introduce again an *Higgs field* that can break U(1) as well as SU(2) symmetry, so we use a weak isospin doublet made by two scalar fields:

$$\Phi(x) = \begin{pmatrix} \varphi_a(x) \\ \varphi_b(x) \end{pmatrix} \tag{1.28}$$

The 1.28 transforms in the same way that 1.10 under $SU(2) \otimes U(1)$ gauge transformations. Including the Higgs field and its interactions with boson and lepton fields in the electro-weak Lagrangian, we obtain a generalized form as:

$$\mathscr{L} = \mathscr{L}^L + \mathscr{L}^B + \mathscr{L}^H + \mathscr{L}^{HL} \tag{1.29}$$

Next steps follows Higgs model searching for the vacuum state energy. The particular value:

$$\Phi_0 = \begin{pmatrix} \varphi_a^0 \\ \varphi_b^0 \end{pmatrix} = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \tag{1.30}$$

where $v = (-\mu^2/\lambda)^{1/2}$, is chosen for the ground state, so that we derive the Higgs field as $\Phi(x) = \frac{1}{\sqrt{2}}\binom{0}{v+\sigma(x)}$, as in 1.26. Since the process began with a doublet of complex scalar fields and finished with a real doublet, we can say that in the Lagrangian three degrees of freedom are absorbed by the $W^{\pm}, Z$ bosons to aquire mass, while the photon remain massless as desired and the scalar massive Higgs boson comes from $\sigma(x)$, with a mass $M_H = \sqrt{2\lambda v^2}$ that is a free parameter to be measured experimentally.

Lepton and quark masses arise from Yukawa interactions with the scalar Higgs. They are directly proportional to Yukawa coupling constants $Y_{q,l}$ and are expressed, in lepton case, in $\mathscr{L}^{LH}$ part:

$$\begin{aligned} \mathscr{L}^{LH}(x) \quad = \quad & -Y_l[\bar{\Psi}_l^L(x)\psi_l^R(x)\Phi(x) + \Phi^\dagger\bar{\psi}_l^R(x)\Psi_l^L(x)] \\ & -Y_{\nu_l}[\bar{\Psi}_l^L(x)\psi_{\nu_l}^R(x)\tilde{\Phi}(x) + \tilde{\Phi}^\dagger\bar{\psi}_{\nu_l}^R(x)\Psi_l^L(x)] \end{aligned} \tag{1.31}$$

where $\tilde{\Phi}(x) = -i[\Phi^\dagger(x)\tau_2]^T$. The masses are introducted as parameters to be experimentally measured, derived from coupling constants and the Higgs field:

$$m_l = Y_l\frac{v}{\sqrt{2}} \tag{1.32}$$

In quarks case, we must take in account that also upper member of quark doublet must have mass. To aquire this feature the Higgs doublet is re-constructed as $\Phi_c = -i\tau_2\Phi^*(x)$ and an hermitian conjugate member is added to the lagrangian. Since the vacuum expectation value of the Higgs field is $v = 246$ GeV, the Top quark Yukawa coupling is:

$$Y_t = \frac{m_{\text{top}}\sqrt{2}}{v} \sim 1 \tag{1.33}$$

1.2 Top and Single-Top in SM: Top physics at Tevatron

Top quark is the heaviest fondamental particle. First observation was obtained by CDF in 1994 [1] and later confirmed by DØ [17]

Most recent Tevatron combination of CDF and DØ results yields a top-quark mass of $m_t = 173.2 \pm 0.9$ Gev [18]. Top quark belongs to the third generation of fermions - being the upper member of $\binom{t}{b}$ quark doublet - and carries both electroweak and strong couplings, so that it can be produced by both interactions.

It decays via electroweak interaction almost exclusively into a W boson and a b quarks. Decays in first and second generation quarks are strongly soppressed by small CKM matrix elements $|V_{ts}|$ and $|V_{td}|$, while $|V_{tb}| \approx 1$.

The top-quark lifetime is $\tau \approx 0.5 \cdot 10^{-24}$ s, one order of magnitude smaller than time scale for hadronization, therefore no top-flavored hadrons exists. The small lifetime also implies that top quark passes all the information, including the spin, to the decay products. It provides the unique opportunity to study a bare quark.

1.2.1 $t\bar{t}$ strong production

The dominant process for top quark production in $p\bar{p}$ collisions at Tevatron is top quark pair production via strong interaction.

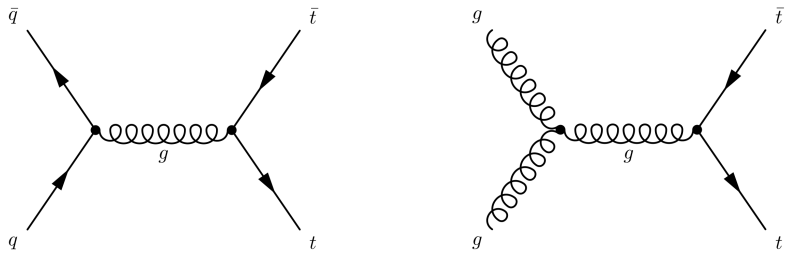


Figure 1.2: Feynman diagrams for top pair production: quark-antiquark annihilation (a) and gluon fusion (b). At the Tevatron diagram (a) is the most relevant

To produce $t\bar{t}$ pairs the kinematic threshold is $\hat{s}_{min} = (P_t + P_{\bar{t}})^2 = 4m_t^2$, while the square of center of mass energy of the partons is:

$$\hat{s} = (x_1 P_p + x_2 P_{\bar{p}})^2 = x_1 x_2 s \tag{1.34}$$

where s is the Tevatron colliding cm energy $\sqrt{s} = 1.96$ Gev. Thus, the momentum fractions of the partons verify the relation:

$$x_1 x_2 \geq \frac{\hat{s}_{min}}{s} \approx (0.18)^2 \tag{1.35}$$

At the Tevatron about 85% of cross section is due to $q\bar{q}$ annihilation and only about 15% due to gluon fusion.

1.2.2 Electroweak Single-Top production

An alternative process to produce top in Tevatron collisions is top electroweak production from a Wtb vertex. Within SM three electroweak processes result in the final state with a single top: t-channel, s-channel and Wt production. The first two processes are dominating at the Tevatron, while the third has much smaller predicted cross section. All three production modes are distinguished by virtuality of $Q^2 = -q$, where q is the 4-momentum of the charged W boson which mediates the interaction.

t-channel

In the t-channel production mode a b sea quark from the proton or the antiproton scatters off a light quark in the other baryon by exchanging a virtual spacelike W boson ($q^2 = \hat{t} = (p_b + p_t)^2 < 0$, where $\hat{t}$ is the Mandelstan variable). Most important NLO correction to t-channel leading order is the $2 \to 3$ process, known as W-gluon fusion, where an initial gluon splits into a $b\bar{b}$ pair. At the Tevatron the predicted cross-section of the t-channel at the Tevatron is $\sigma_{\text{t-ch}}^{\text{NNNLOapprox}} = 2.10 \pm 0.19$ pb.

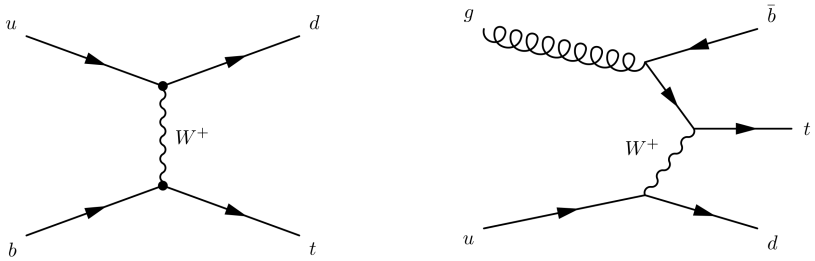


Figure 1.3: Feynman diagrams for t-channel single-top production: leading order (a) and next-to-leading order with initial gluon splitting (b). At the Tevatron

s-channel

In s-channel production mode the annihilation of two quarks q and $\bar{q}'$, one from proton and the other from antiproton, produces a timelike W boson ($q^2 = \hat{s} \leq (m_b + m_t)^2$, where $\hat{s}$ is the Mandelstan $\hat{s} = (p_q + p_{\bar{q}'})^2$). The coupling of the (u, d) pair is by far the dominant process; diagrams with an s or c quarks in the initial state contribute only for about 2%. The NLO most important contributions to the s-channel lead to additional soft light quarks, since is much less likely for gluon to split into heavy quarks pair. The NLO initial state gluon splitting process for

s-channel in fig. 1.4(b) can lead to the same final state as the W-gluon fusion NLO process for t-channel in fig. 1.3(b). They could give rise to an interference in the NLO cross section calculation. This is avoided since in the t-channel process the tb pair forms a color-octet cause it comes from a gluon, while in the s-channel the t and b quarks form a color singlet since both quarks in the final state come from the W boson, therefore the separation between the processes is manteined at NLO. At the Tevatron the predicted cross-section of the t-channel at the Tevatron is $\sigma_{\text{s-ch}}^{\text{NNNLOapprox}} = 1.05 \pm 0.07$ pb.

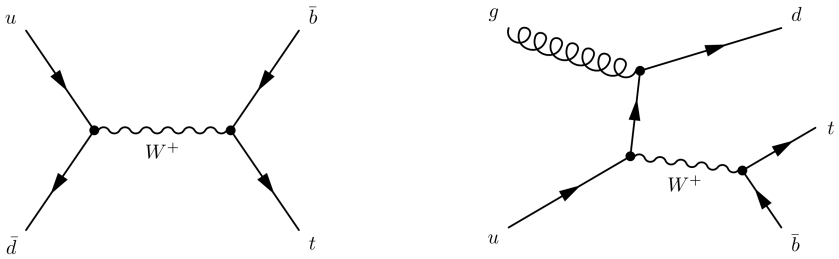


Figure 1.4: Feynman diagrams for s-channel single-top production: leading order (a) and next-to-leading order with initial gluon-splitting (b)

Wt-channel

The Wt channel is also dubbed as associated production, in which a on-shell W boson ($q^2 = m_W^2$) is produced in association with a top quark. This process has a cross section much smaller than the other production modes: NNNLO order calculations predict a cross section at the Tevatron of $\sigma_{\text{Wt-ch}}^{\text{NNNLO}_{\text{approx}}} = 0.22 \pm 0.08$ pb, with $m_{\text{top}} = 172.5$ GeV [19].

Wt-channel was considered negligible in the previuos analysis. Taking advantage of a new calculation of MC recently made available at CDF, we include this process in the signal sample, altough its contribution is negligible.

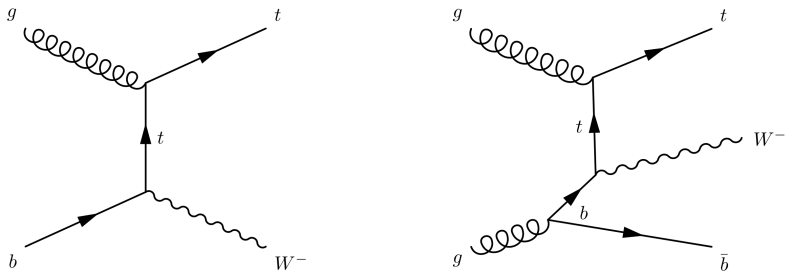


Figure 1.5: Feynman diagrams for Wt-channel single-top production: leading order (a) and next-to-leading order with initial gluon-splitting (b)

Chapter 2

Single-Top Experimental Searches

This work concerns the new Single Top analysis performed at CDF using 7.5 fb^{-1} of $p\bar{p}$ data collected until march 2011. This analysis started as an update of the previous Single Top search, which used 3.2 fb^{-1} of data. The 3.2 fb^{-1}analysis combined four parallel W+jets analyses which used different multivariate methods for the background discrimination [2]. Here, we provide the update of the artificial neural networks (NN) Single Top analysis. As the previous analyses were statistically limited, we increased the lepton acceptance by defining new lepton categories. We then adjusted and retrained the old Neural Networks for the new available integrated luminosity of 7.5 fb^{-1}.

In the following, we will briefly present the structure and results of the previous 3.2 fb^{-1}analysis and provide an introductory overview of the new 7.5 fb^{-1}analysis, pointing out the work done in this thesis. Then, we will review the most important Single Top searches and results obtained by the other experiments.

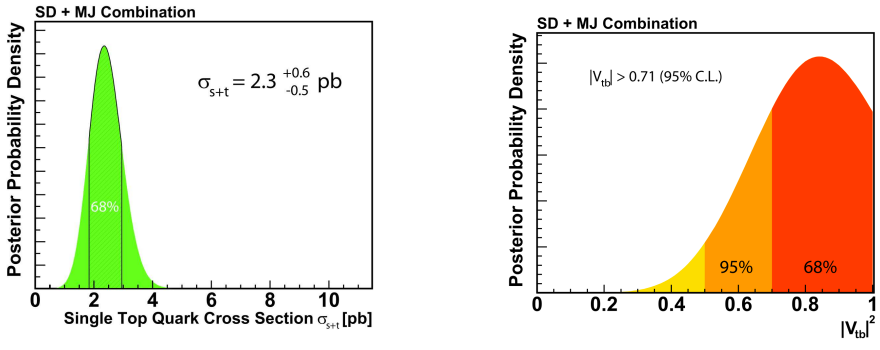
2.1 Single-Top at CDF

2.1.1 3.2 fb^{-1} Single-Top Analysis

The CDF Collaboration reported evidence for the production of single top quarks in 2.2 $fb^{-1}$of data [20] and observation of the process in 3.2 $fb^{-1}$of data [2]. These results were obtained by a combination of four analyses: multivariate likelihood functions (LF), artificial neural networks (NN), matrix elements (ME), and boosted decision trees (BDT). These analyses select events with a high-p_T charged lepton,

large missing transverse energy E_T and two or more jets, at least one of which is b-tagged. Each analysis separately measures the single top quark production cross section and calculates the significance of the observed excess. Because the selection was in common between the four analyses -with a 100% overlap in the data and Monte Carlo events selected-, the output discriminant of each one was used as input to a *super discriminant* (SD) function evaluated for each event, in order to combine the results from the 4 different techniques.

A separate analysis was conducted on events without an identified charged lepton, in a data sample which corresponds to 2.1 fb^{-1}of data. Missing transverse energy plus jets, one of which is b-tagged, is the signature used for this (fifth) analysis (MET+jets or MJ), which is described in detail in [21]. There is no overlap of events selected by the MJ analysis and the W+jets analyses as they were collected by different triggers and had a completely orthogonal event selection, therefore they can be statistically combined to maximize the total sensitivity. The result of the MJ analysis was combined with the result of the SD from the W+jets analysis to obtain the final measurement. In the end, it was reported the observation of electroweak single top quark production with a significance of 5.0 standard deviations. The measured value of the combined s- and t-channel cross section is $\sigma_{s-t} = 2.3^{+0.6}_{-0.5}$(stat.+syst.)pb (fig. 2.1(a)) assuming the top quark mass is 175 GeV/c^2 and using the SM value of σ_s/σ_t. A value for the CKM matrix element $|V_{tb}| = 0.91 \pm 0.11$(stat + sys) ± 0.07(theory) was extracted, with a lower limit $|V_{tb}| > 0.71$ at the 95% c.l.(Fig. 2.1(b)).



(a) (b)

Figure 2.1: cross section measurement histogram (a), posterior curve for the $|V_{tb}|$ calculation (b)

NN Discriminant Analysis

Our goal is to update the Neural Network branch of the 3.2 fb^{-1}analysis. This method uses artificial neural networks created with the `NeuroBayes` ® package in order to combine variables capable to separate single top quark signal from background events. The NN analysis is structured in this way:

1. The neural networks are **trained** in the different selected data samples (TLC or EMC events with 2-3 jets with 1-2 b-tag[1]). The t- and s-channel Montecarlo are used as training signal sample while as background training sample is used a mix of SM processes in the ratios of the estimated yields provided by the basckground estimation (see Chap. 7). The training procedure uses a set of variables listed in Tab. 2.1, whose precise description can be found in [22].

2. For some of those variables the **reconstruction** of the top quark is necessary. The top quark four-momentum is built out of the reconstructed W boson and the btagged jet. The W boson is reconstructed from the measured tight lepton and the reconstructed neutrino. If there is more than one jet with a reconstructed secondary vertex, the tagged jet with the largest $Q_l \cdot \eta_j$, is assigned to belong to the top-quark decay.

3. In each data sample a **neural network** is constructed with the input variables described above. The outputs lie in the range [-1,+1], where -1 is background-like and +1 is signal-like. The sum of predictions normalized to our signal and background models -given by the background estimation- are compared with data. As example, the predicted distributions (Fig. 2.2(b)) and the corresponding template (Fig. 2.2(a)) in the 2Jet-1Tag channel are shown below.

4. The observed and predicted distributions of some of the variables that provide the best sensitivity are compared. The Monte Carlo **modeling** of the output NN discriminant distributions are validated searching an agreement between the predictions and the observations in both the input and the output variables.

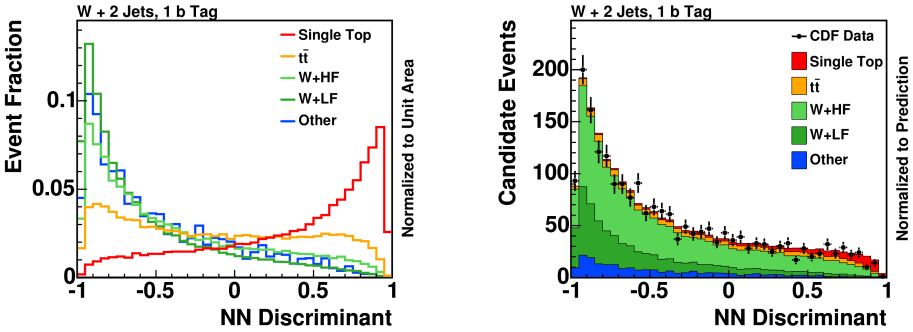
5. The last step of the NN procedure is to increase the quality of the signal contribution in the highly signal-enriched region by requiring a **NN discriminant** output above 0.4 in the event sample with 2 jets and 1 b tag comparing the expectations of the signal and background processes to the observed data of this

[1]In the following we will define tight leptons as TLC. EMC are the Extended Muon Category, that is an inclusive definition for muons collected with loose requirements (Chap. 6)

Variable/Channel	2J1T	2J2T	3J1T	3J2T
$M_{l\nu b}$	✓	✓	✓	
$M_{l\nu bb}$		✓		✓
$M_T^{l\nu b}$	✓	✓	✓	✓
M_{jj}	✓	✓	✓	✓
M_T^W	✓	✓		
$E_T^{b_{top}}$		✓	✓	
$E_T^{b_{other}}$				✓
$\sum E_T^{jj}$			✓	✓
E_T^{light}	✓			✓
p_T^l	✓			
$p_T^{l\nu jj}$			✓	✓
H_T	✓		✓	
$\not{E}_T$		✓		
$\not{E}_{T\,\mathrm{sig}}$			✓	
$cos\vartheta_{lj}$	✓		✓	✓
$cos\vartheta_{lW}^W$	✓			
$cos\vartheta_{lW}^t$	✓			
$cos\vartheta_{jj}^t$		✓		✓
$Q_{lep} \cdot \eta_{j_1}$	✓		✓	✓
η_l		✓		
η_W	✓	✓		
$\sum \eta_j$	✓		✓	
$\Delta\eta_{jj}$			✓	✓
$\Delta\eta_{t,light}$			✓	
$\sqrt{\hat{s}}$				✓
Centrality				✓
Jet flavor sep.	✓	✓	✓	

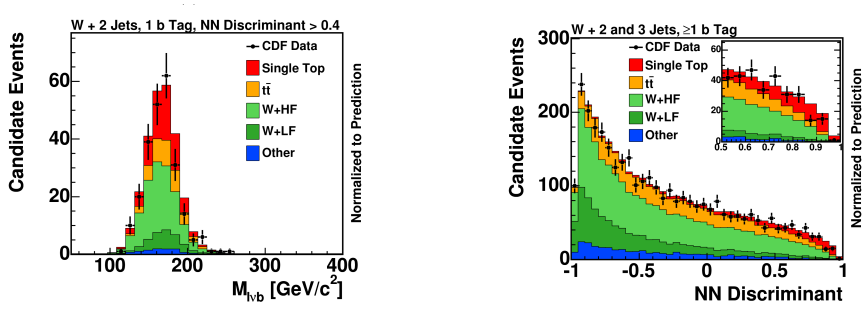
Table 2.1: Summary of variables used in the different neural networks.

subsample in various highly discriminating variables. In Fig.2.3(a) the single top invariant mass plot from NN analysis is shown, with the characteristic observation peak, for events in the NN discriminant optimized region.



(a) (b)

Figure 2.2: Comparison of the predictions and the data for $M_{l\nu b}$ (a) for events in the 2Jet-1Tag channel with an output above 0.4 of the NN discriminant (b)



(a) (b)

Figure 2.3: Comparison of the predictions and the data for $M_{l\nu b}$ (a) for events in the 2J1T channel with an output above 0.4 of the NN discriminant (b)

The NN method for the Single Top analysis was developed and succesfully used by the KIT (Karlsruhe Insitute of Technology) group at CDF. A full detailed description of this method can be found in [23] and [22].

2.1.2 7.5 fb^{-1} Single-Top Analysis

In the first observation of Single Top, several research groups at CDF worked on the development of different techniques in order to observe such a rare process and discriminate the few signal events from the background. Moreover the different analyses were based on different data selection softwares. Therefore it was decided to integrate this update into a more general effort that CDF was planning, the development if a common selection tool named WH Analysis Module, or WHAM (App. C).

The strategy followed for my analysis was to:

1. re-write the whole selection with WHAM and improve it;

2. interface this to the NN package used to obtain the cross section;

3. obtain the cross section;

Besides increasing the total integrated luminosity to 7.5 fb^{-1}, the current work presents several improvements in the data analysis techniques:

- The signal MonteCarlo generator POWHEG [24] is different from the one used before. It fully exploits the NLO level of the ST processes and well modelize Wt associated production, which is now available to be included in the cross section measurement for the first time. It will be introduced in Sec. 5.1.1.

- The selection was changed to include the new ISOTRK muon events in the inclusive EMC lepton category (Sec. 6.2).Moreover, we changed the QCD rejection procedure after some studies on the QCD-oriented cuts (Sec. 6.4.2).

- We also improved the separation of QCD events faking our signal.(Sec. 7.4.1).

- From the code point of view, the part of my work was to adapt the outputs of the selection and background estimates provided by the WHAM framework in a fully complatible format for the old software developed by the KIT group. This was needed as the old package provided the reconstruction for the physics objects used in the NN package.

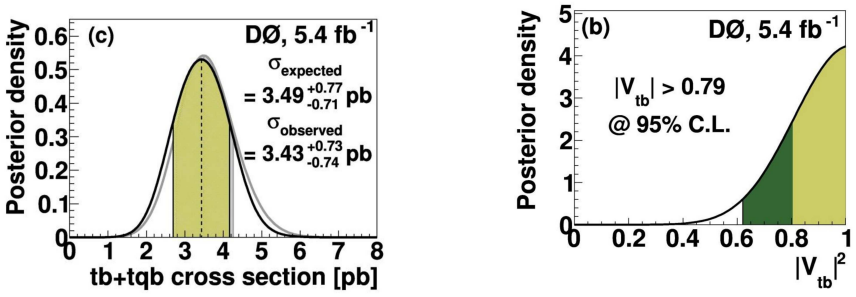
2.2 Other Experimental results

2.2.1 DØ

The DØ Collaboration reported evidence for the production of single top quarks in 0.9 fb^{-1}of data [25], and observation of the process in 2.3 fb^{-1}. Three multivariate analyses in Lepton+jets channel were performed and succesfully combined [26].

In 2010, DØ measured the single top production cross section in the τ+jets final state using 4.8 fb^{-1}of data [27].

In 2011, DØ updated and combined the three multivariate analysis used for the observation search in Lep+jets using 5.4 fb^{-1}of data[28]. Their latest results in the cross section measurement and $|V_{tb}|$ value extraction are shown in Fig. 2.4(a) and 2.4(b).



(a) (b)

Figure 2.4: DØ latest results for the cross section measurement histogram (a), posterior curve for the $|V_{tb}|$ calculation (b)

2.2.2 LHC experiments: CMS and ATLAS

The start of LHC produced new data on Single Top. ATLAS and CMS studied the ST production in their data.

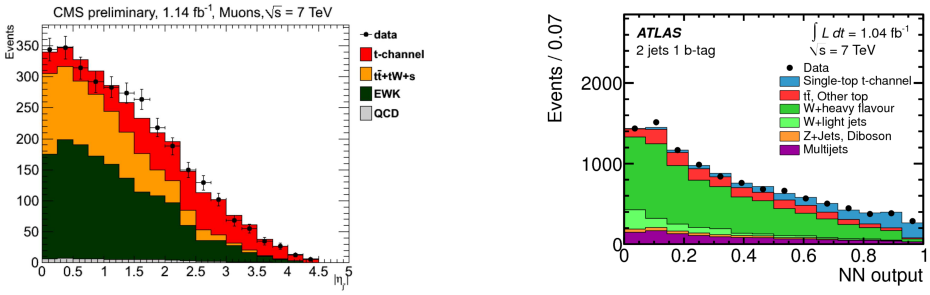
Cross sections at the LHC are 4:30 times larger than at the Tevatron. Especially the Wt-channel has a significantly large cross section at the LHC as the available parton energy and the gluon flux are larger than at Tevatron. The search in the s-channel is indeed more challenging due to the small cross section of this process. The latest numbers for Single Top cross section are showed in Tab. 2.2.

$\sqrt{s}$	t-ch	Wt-ch	s-ch
7 TeV	64.2 ± 2.6	15.6 ± 1.3	4.6 ± 0.2
1.96 Tev	2.1 ± 0.1	0.25 ± 0.03	1.05 ± 0.05

Table 2.2: Cross sections at NNLL for Single Top production channels at LHC and Tevatron, with an assumed top quark mass of 173 GeV [29].

So far, the t-channel single top quark production has been measured by the ATLAS and CMS collaborations. First measurements of the t-channel cross section were performed in the first 35 pb^{-1}(CMS) [30] and 36 pb^{-1}(ATLAS) [31]. The latest update in the t-channel by CMS [32] measured a value of $\sigma_{t-ch} = 70.2 \pm 5.2(\text{stat.}) \pm 10.4(\text{syst.}) \pm 3.4(\text{lumi.})$pb (Fig. 2.5(a)). ATLAS latest results with 700 pb^{-1}[33] is $\sigma_{t-ch} = 83 \pm 4(\text{stat.})^{+20}_{19}(\text{syst.})$m pb (Fig. 2.5(b)).

In the most promising channel at the LHC, the single top production channel with an associated W boson (Wt-channel), CMS performed an analysis on 2.1 fb^{-1}[34] and measured a cross section of $\sigma_{Wt-ch} = 22^{+9}_{-7}$ pb with 16 ± 1 pb expected.



(a) (b)

Figure 2.5: LHC latest results for the single top t-channel production mode by CMS (a) and ATLAS (b)

The result is still compatible with the background-only hypothesis at the 2.7 σ level. ATLAS recently found the first evindence of Wt single top associated production in 2.05 fb^{-1}with a signifincance level of 3.3 σ and measured a cross section value $\sigma_{Wt-ch} = 16.8 \pm 2.9(\text{stat.}) \pm 4.9(\text{syst.})$ [35].

Chapter 3

The Fermilab Tevatron Collider and the CDF II Detector

This chapter provides a brief description of the Tevatron accelerator complex and of the CDF II detector. We will describe the $p\bar{p}$ beam production and acceleration process, the structure of CDF II detector - specially focusing on the tracking system - and the trigger and data acquisition system.

3.1 The Tevatron

The Tevatron is a proton-antiproton collider located at Fermi National Accelerator Laboratory (Fermilab) and was the particle accelerator with the worldwide highest center-of-mass energy ($\sqrt{s}$=1.96 TeV) until March 2010, when the Large Hadron Collider (LHC) at CERN started to run at $\sqrt{s}$=7 TeV. The Fermilab is located at Batavia, Illinois (USA) about 50 km west of Chicago.

The Tevatron was completed in 1983 and started operations in 1985, delivering data until 1996 at $\sqrt{s}$=1.8 TeV, in the so-called Run I. Afterwards, the accelerator complex was upgraded in order to achieve a higher center-of-mass energy as well as a much higher luminosity. RunII, with a c.m. energy of 1.96 TeV, started in 2002 and lasted until September 30, 2011.

As we can see in Fig. 3.1 the accelerator chain is a complex process involving several machines, and it can be summarized in: proton and antiproton beams production and accelaration, injection and collision in the Tevatron.

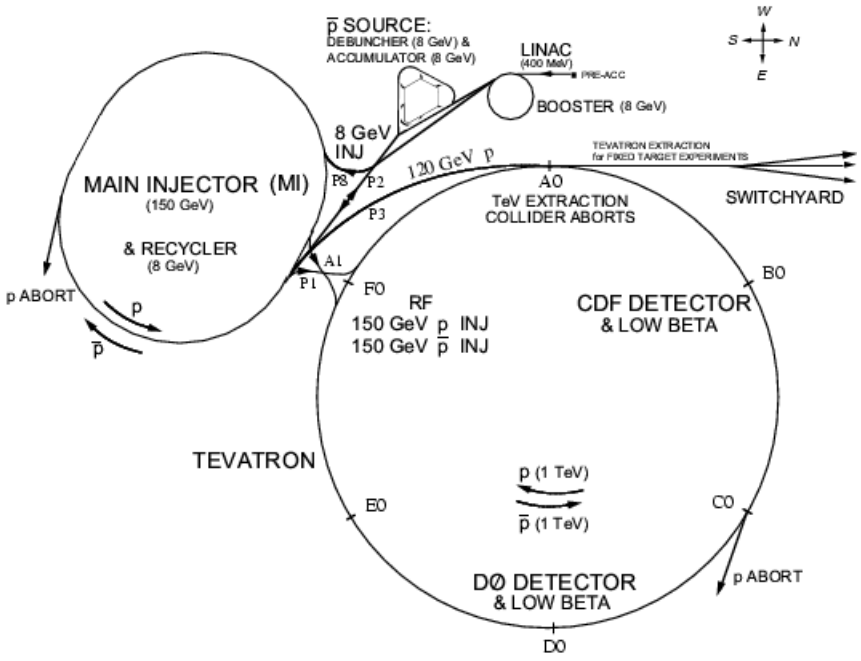


Figure 3.1: The Fermilab accelerator complex.

3.1.1 Proton Production

As first step protons are produced from the Proton Source (gaseus H_2 in a negative electrically charged enclosure with -750 kV potential) by ionizing the H_2 into electrons and H^+ ions. The protons H^+ are collected and brought to strike a cathode surface made by negative charged coated Cs from which they can absorb two electrons and convert into H^-, so that they are then accelerated to an energy of 750 KeV by a Cockroft-Walton accelerator at 15 Hz rate (that is the Preaccelerator stage). The H^- are then passed to the Linear Accelerator (*Linac*).

The Linac is made of two sections: a low energy one consisting in five drift tubes which accelerates the H^- to about 116 MeV, and an high energy section with a series of side coupled cavities. At the end of the Linac (130 m of overall length) the negative ions have an energy of 400 MeV and are collected into bunches, so that they have the characteristic RF bunch structure. The beam pulse is typically of 20 ms and from the Linac it's injected in the first synchrotron of the accelerator chain, the *Booster*.

The Booster is a circular proton accelerator of 150 diameter that works at same duty cycle of the Linac, 15 Hz. Passing through carbon foils, electrons are stripped off the H^- and resulting protons are accelerated in the Booster up to 8 GeV. The injecton of H^- instead of protons is motivated by the need of injecting more

particles at the same time in which protons are already running in, avoiding the loss of protons by kick out from the bending magnets.

After the booster, the proton bunches are trasferred to the *Main Injector*[1], a large oval synchrotron with a mean radius of 0.5 Km. This accelerator has a collection of accelerating and focusing magnets, leading to full energy acceleration in a cycle time of 2 s. The main function of the Main Injector is to accelerate protons up to 120 GeV to be sent in AntiProton Source, and to accelerate protons and antiprotons up to 150 Gev to be sent in the Tevatron collider.

3.1.2 Antiproton production and accumulation

Antiprotons must be first produced and stored. The production involves the protons bunches extracted from the Main Injector at an energy of 120 GeV. Those protons are sent to the *Antiproton Source*, where they hit a 7 cm nickel target, so that new protons, antiprotons and many other particles are produced by the collision. Secondary particles are focused by a lithium lens and analyzed in a charge-mass spectrometer selecting negatively charged particles at an energy of 8 GeV. This system of production has an efficency of $2 \cdot 10^{-6}$ antiprotons per proton hitting the target. Since antiprotons exit from the source with a large spatial and momentum spread, they are then sent to the *Debuncher*, a rounded triangular shaped synchrotron with a mean radius of 90 m. In this machine, thanks to RF manipulation called "bunch rotation", they loose their RF bunch structure turning into a continuous nearly monochromatic 8 GeV beam, stabilized by stochastic cooling [36].

The antiproton debunched beam is then transferred to the *Accumulator Ring*, a sinchrotron housed in the same tunnel of the Debuncher, where the $\bar{p}$ are stacked at energy of 8 GeV and further cooled down using again RF and stochastic cooling system. Once reached the optimal capacity for the antiproton stack in the Accumulator, the bunches are trasferred in the *Recycler*. The Recycler is a constant 8 GeV-energy storage-ring placed in the Main Injector tunnel, that uses permanent magnets (magnetized strontium ferrite). It is used [2] to store antiprotons from

[1] The Main Inejctor synchrotron was built during the upgrade of the Tevatron accelerator complex that took place between the end of RunI (1996) and the start of RunII (2002). The Main Injector replaced the old Main Ring, that was placed in the Tevatron tunnel, and it allowed to have more current with faster cycle time, leading to higher luminosity of the Collider.

[2] Its original purpose was to recycle -from which the name- the $\bar{p}$ left over from the previuos Tevatron store and merging them with the $\bar{p}$ coming from the Antiproton source, but those plans were abandoned in early Run II cause of efficency and cost problems.

the Accumulator before the injection in the Tevatron main machine. Since 2008 relativistic electron cooling was successfully implemented into the Recycler, significatively improving the Tevatron performancies.

3.1.3 Tevatron Collider

The *Tevatron* is the largest accelator at Fermilab: it is a circular synchrotron with a radius of about 1 km that employs about 1000 dipole bending magnets with niobium-titanium superconducting coils in order to obtain the 4.2 T magnetic field needed to bend the particles. The coils are kept at a temperature of 4 K to reach the superconducting state and they are cooled by liquid helium.

The Tevatron stores and accelerates protons coming from the Main Injector and the antiprotons extracted from the Recycler and brings them to collide in several specific high luminosity intersection points. Two of them are used by the CDF and DØ .

During the standard operations, protons and antiprotons circulate in opposite direction and are kept inside the beampipe by the magnets. Elecrostatic separators make the beams spiraling around each other (producing the so-called *helix*) during the revolution in the ring.

A Tevatron *store* starts from the injecton of 36 proton bunches ($\sim 2.9 \cdot 10^{11}$ p per bunch) and 36 antiproton bunches ($\sim 0.9 \cdot 10^{11}$ $\bar{p}$ per bunch) at an energy of 150 GeV, while the p and $\bar{p}$ beams are split in three bunch trains, each containing 12 bunches separated in time by 0.396 μs. The beams are then accelerated up to 980 GeV each within 90 s by eight RF accelerating cavities, leading to final bunch revolution time of $\sim$ 21 μs (0.9999996 c). Finally proton and antiprotons beams in the Tevatron are brought to collision at BØ and DØ after they are focused to minimal transvers size in order to maximize the collision rate [3]. Near the interstection point, the polarity of the separators are inverted, allowing head-on collisions of the opposite incoming bunches, with a final center-of-mass energy of $\sqrt{s}$=1.96 TeV.

3.1.4 Tevatron Performance

The two most important parameters of a particle accelerator are the energy E and the instantaneous luminosity $\mathcal{L}$, since the rate of a physical process with cross

[3] To focus the beams are used some special high-power quadrupole magnets ("low-β squeezers"), installed on the beam pipe at either side of the detectors, so that the resulting transverse spatial distribution of the luminous region is approximately a two-dimensional Gaussian, with $\sigma_{x,y} \approx 30\mu$m.

section σ (which depends on energy) is:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma \tag{3.1}$$

The instantaneus luminosity, usually measured in $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, is defined by:

$$\mathcal{L} = n \cdot f \cdot \frac{N_p N_{\bar{p}}}{4\pi\sigma_x\sigma_y} \tag{3.2}$$

where n is the number of bunches inside the accelerator, f is the bunch crossing frequency, N_p and $N_{\bar{p}}$ are the number of protons and antiprotons per bunch, σ_x and σ_y are the transverse dimensions of the beam with respect to the interaction point.

At the end of Run I the typical instantaneous luminosity at the beginning of each store was $\mathcal{L} = 2.5 \cdot 10^{31}\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

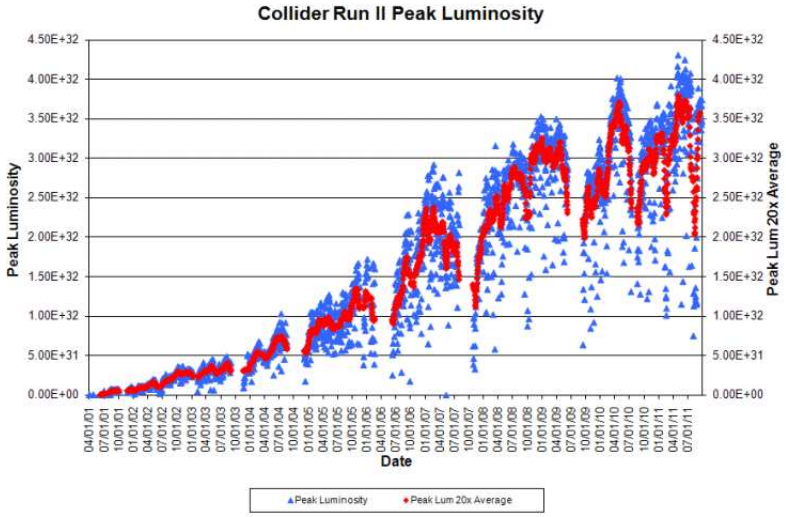


Figure 3.2: The Tevatron initial luminosity in the Run II: the blue triangles are the peak luminosity at beginning of each store, the reds are an average on last 20 peaks.

Many improvements were realized during the 1996-2001 upgrade such as the increase of *n*, *f*, and the number of particles per bunch. In Run II the increase of total number of $\bar{p}$ in the beam, made possible to obtain a (typical) value $\mathcal{L} \geq 2 \cdot 10^{32}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. At the end of its life, the accelerator provided a value up to $\mathcal{L} = 4.03 \cdot 10^{32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, the highest one for a proton-antiproton collider [4](as in Fig. 3.2).

The amount of collected data is measured by the integrated luminosity $L = \int \mathcal{L}\, dt$, usually measured in fb^{-1}($1\ \mathrm{fb}^{-1} = 10^{-39}\mathrm{cm}^2$). Up to September 30, 2011,

[4]in april 25th, 2011 the LHC has produced an initial luminosity of $4.3 \cdot 10^{32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, achieving the new world record for an hadron collider.

the Tevatron delivered about 12 fb^{-1}(Fig. 3.4). Since the CDF recording efficiency was $\approx 80\%$, about 10 fb^{-1}of data represent the final complete CDF dataset.

The analysis we present works on 7.5 fb^{-1}of data.

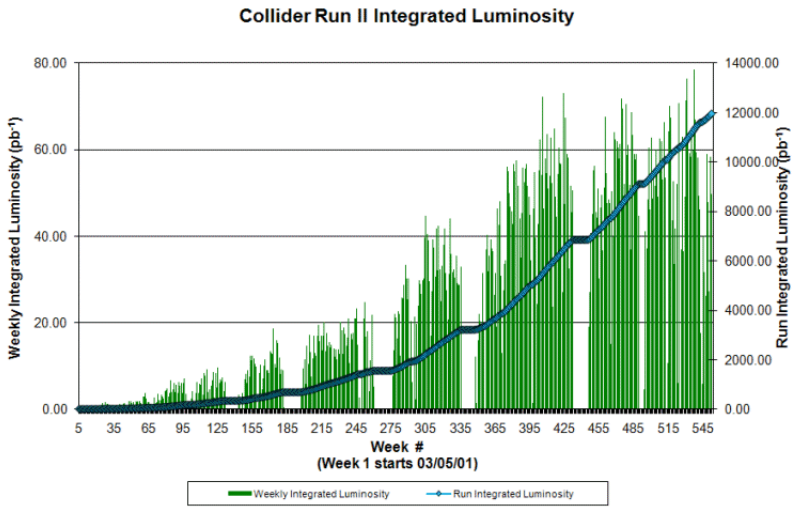


Figure 3.3: The Tevatron integrated luminosity in the Run II. Empty periods are due to Tevatron shutdown

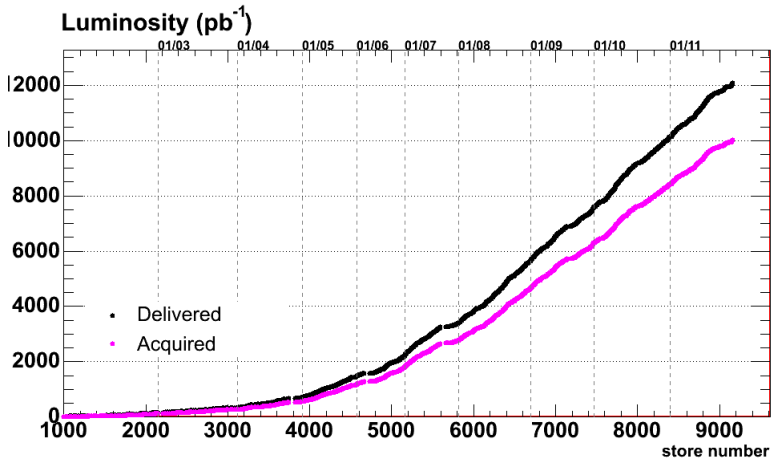


Figure 3.4: The delivered and aquired luminosity in the Run II by the CDF detector. The final integrated luminosity is about 10 fb^{-1}.

3.2 CDF II Detector

The Collider Detector at Fermilab (CDF II) is a general purpose solenoidal detector with azimuthal and forward-backward symmetry, located at the BØ intersection point of the Tevatron accelerator ring. With the combination of precise charged particle tracking, fast projective calorimetry and fine grained muon detection, CDF II is able to detect energy, momentum and charge of the particles resulting from the 1.96 TeV proton-antiproton collisions.

The detector started the operations in 1985 and ended its data taking in September 30, 2011, becoming the longest lasting HEP experiment ever. In 25 years the detector was upgraded several times: the most important one took place in 1996-2001, in order to deal with the increase in instantaeous luminosity. Currently CDF II experiment is operated and mantained by CDF collaboration, which involves more than 60 universities and research institutions with hundreds of physicists.

The detector is composed by several coaxial subdetectors, as showed in Fig. 3.5.

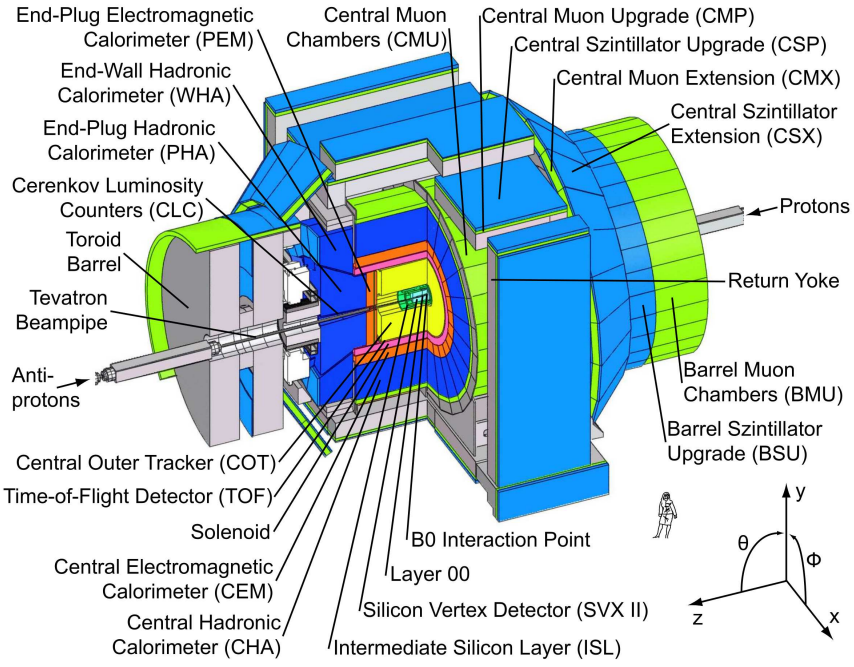


Figure 3.5: Isometric view of main components of CDF II detector and coordinate system

IN CDF we ususally use cylindrical coordinate system (appropriate for its symmetry), with the origin of the axes in BØ . The z axis is directed along the proton beam, the azimuthal angle φ is defined with respect to outgoing radial direction and the polar angle ϑ is defined with respect to the proton beam direction. Using these

coordinates, *longitudinal* means along proton direction (the z-axis) and *transverse* means placed in the plane perpendicular to the beam, i.e. (r, φ) plane. With these definition it is useful to define the transverse momentum p_T and the transverse energy E_T of a particle:

$$p_T \equiv p \cdot \sin\vartheta, \quad E_T \equiv E \cdot \sin\vartheta \tag{3.3}$$

In collider physics it is sometimes more convenient to use another variable that is invariant under $\hat{z}$-boosts. One can replace ϑ by the *pseudorapidity* η, defined as:

$$\eta \equiv -\ln(\tan(\frac{\vartheta}{2})) \tag{3.4}$$

The pseudorapidity is the ultra-relativistic approximation of the *rapidity* variable Y. If $P_\mu = (E, \vec{p})$ is the 4-momentum of the particle, then Y is defined as:

$$Y \equiv \frac{1}{2}\ln\left(\frac{E + p\cos\vartheta}{E - p\cos\vartheta}\right) \tag{3.5}$$

Therefore, a measurement of rapidity requires a detector with full particle identification capability because of the mass term entering E. If we refer to an ultra-relativistic particle ($p \gg m$), we can approximate Y as:

$$Y \overset{p\gg m}{\longrightarrow} \eta + O(m^2/p^2) \tag{3.6}$$

It can be shown that the difference in Y (and, in u.r. limit, in η) is invariant under Lorentz $\hat{z}$-boosts.

The detector can be described following its radial development (Fig. 3.5). It can be divided in three coaxial parts, one inner and two outer with respect to the 1.4 T superconducting solenoid: the inner part is made by the silicon tracking system, which consists of 3 subdetectors (*Layer00, Silicon Vertex Detector II, Intermediate Sylicon Layers*), and an open-cell drift chamber (Central Outer Tracker). Outside the solenoid is located the complex calorimetry system, composed by projective towers each one divided into electromagnegnetic and hadronic part. Muon detectors are the third and outermost part of the detector.

In the following sections there will be a short description of the detector, while a complete description of CDF II can be found in [37].

3.2.1 Tracking system

The CDF tracking system is designed and installed to provide a precise transverse momentum measurement of three-dimensional charged particles. It is surrounded by a solenoid made by NbTi superconducting helium-cooled coils. The

solenoid generates a 1.4 T magnetic fied parallel to beam axes, bending the tracks of charged particles to be helicoidal in the tracking volume ($|r| \leq 150$, $|z| \leq 2.4$). The radial thickness of the solenoid is 0.85 X_0 and particles with $p_T < 0.3$ GeV/c can spiral inside.

As mentioned, the silicon tracking system is divided into three subdetectors interacting with the particles exiting from the primary interaction vertex. With the precise recontruction of the tracks it permits to identify secondary vertices associated with long-living b and c hadrons.

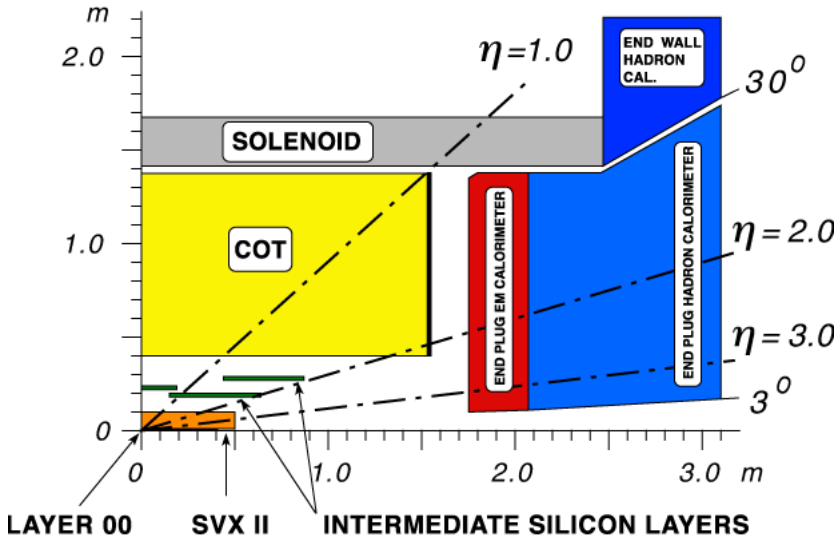


Figure 3.6: Elevation view of the CDF II tracking system, the surrounding solenoid and the forward calorimeters

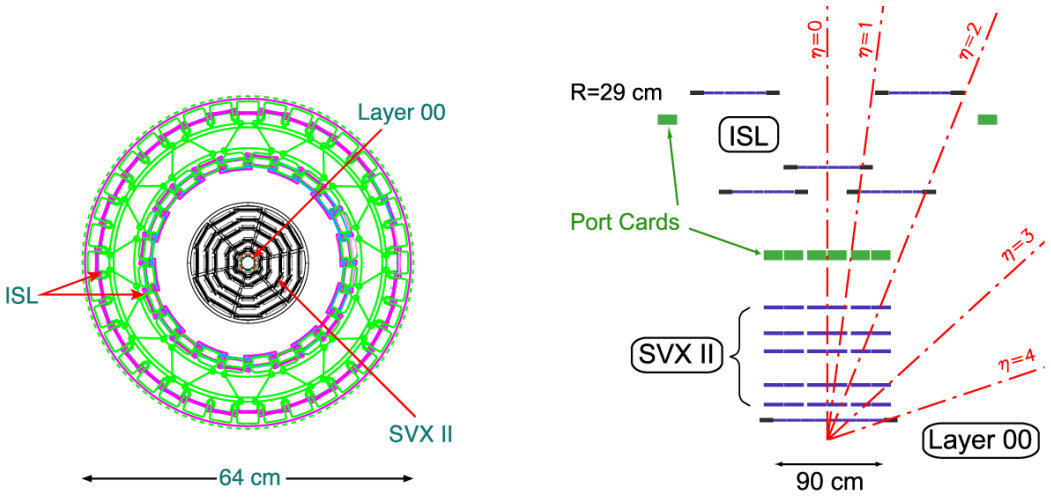
Silicon System

Installed directly on the Tevatron beam pipe, *Layer00* is a single-sided silicon microstrip detector. It consists of 12 sensors along the beam line, for a total length of 94 cm ($|z| \leq 47$), arranged in two overlapping sub-layers at radii of $r_1 = 0.35$ cm and $r_2 = 0.62$ cm, covering the region $|\eta| \leq 4.0$. Layer00 is used to recover the degradation in resolution of the reconstructed vertex position due to multiple scattering, which is particularly significant for tracks passing through the SVXII read-out electronics and cooling system that are installed within the tracking volume.

The most important component of the silicon system is the Silicon VerteX detector (SVXII), built in three cylindrical barrels each one divided in twelve azimuthal sectors (*wedges*). Each sector is arranged in five concentric layers with radii from

$2.5 < r < 10.7$ cm ((see Fig. 3.7(a))), supporting double-sided silicon strip detectors. The total length of SVXII is 96 cm along z, covering the region $|\eta| \leq 2$.

The sensors of SVXII layers have strips parallel to z axis on one side, so that they all measure the coordinates in the (r, φ) plane. Three of five layers have backside strips rotated by $\pi/2$ with respect to axial parallel strips, in order to combine (r, φ) measurament with precise z coordinate determination. The other two layers has backside strips twisted by 0.021 rad with respect to axial parallel strips, combining (r, φ) with small angle stereo (SAS) measurement in plane (r,z).



(a) (b)

Figure 3.7: Frontal and side views of silicon microstrip system

Between the SVXII and the COT there is the Intermediate Silicon Layers (ISL), placed at $20 \leq r \leq 29$ and covering the region of $|\eta| \leq 2$. The ISL is divided into three barrels segmented into φ wedges: the central barrel ($|\eta| \leq 1$) has a single layer of silicon at radius r= 22 cm, the other two barrels at $1 \leq |\eta| \leq 2$ are divided in 2 layers at radii $r_1 = 20$ cm and $r_2 = 29$ cm (Fig. 3.7(b)). The total length of ISL along z is 174 cm. Its purpose is to provide an efficent linking between tracks reconstructed in the chamber and hits detected in the SVXII, extending the track recontruction in the region $1 \leq |\eta| \leq 2$, not fully covered by the outer chamber.

The entire silicon tracking system has a resolution on the impact parameter of 40 μm and a resolution on z_0 of 70 μm.

Central Outer Tracker

The COT is the main tracker at CDF. It is a cylindrical multi-wire drift chamber surrounding the silicon system, with an active volume covering the region $43.4 \leq$

$r \leq 132.3$ cm, $|z| \leq 155$, $|\eta| \leq 1$. The chamber is filled with a 50:50 mixture of argon and ethane gas. The COT contains 96 sense wire layers, which are radially grouped into 8 superlayers, as inferred from the end plate slot structure shown in Fig. 2.7. Each superlayer is divided in φ into *supercells*, and each supercell has 12 alternated sense and field shaping wires. So within the supercell width the trajectory of a particle is sampled 12 times. The maximum drift distance is approximately the same for all superlayers. Therefore, the number of supercells in a given superlayer scales approximately with the radius of the superlayer. The entire COT contains 30,240 sense wires. Approximately half the wires run along the z direction (*axial*), the other half are strung at a small stereo angle (2°) with respect to the z direction (*stereo*). The combination of the axial and stereo information allows us to measure the z positions. Particles originated from the interaction point, which have $|\eta| < 1$ pass through all the 8 COT superlayers. In this η range the tracking effincency is nearly 100%.

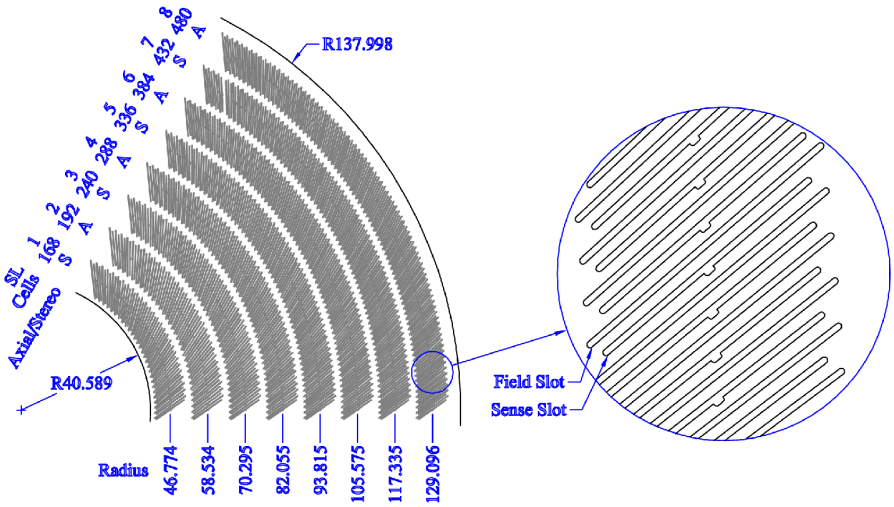


Figure 3.8: A 1/6 section of the COT end-plate with the eight super-layers structure and cell slots: for each superlayers there is the total number of cells, the wire orientation (A=axial, S=stereo) and the average radius.

Thanks to the 8x12 sampling planes structure of the chamber (see Fig. 3.8), for each particle fully crossing the COT 96 hits are measured. The layers with axial wires provide meaurements on the (r, φ) plane, while the other four with stereo wires measure (r, z) coordinates.

When a charged particle passes through, the gas is ionized. Electrons are attracted toward the sense wires. However, due to the magnetic field that the COT is

immersed in, electrons drift at a Lorentz angle of 35°. In order to optimize momentum measurement, the supercell is tilted by 35° with respect to the radial direction to compensate for this effect and make the drift path perpendicular to the radial direction.

The single hit position resolution is 140 μm and the resolution on p_T is $\Delta p_T/p_T \sim 0.01\% \cdot p_T$, that slowly degrades with increasing $|\eta|$. In addition, the COT provides some dE/dx information for the tracks.

3.2.2 Calorimetry

The scintillator-based CDF calorimetry system is located ouside the solenoid. It covers most of the 4π solid angle around the interaction point (up to $|\eta \leq 3.64$). The role of the system is to provide a precise measurement of the energy flow of electrons, photons and jets by fully absorption of those particles. Calorimeters are segmented into projective towers, each one covering a small range in pseudorapidity and azimuth ($\Delta\varphi \sim 15°$, $\Delta\eta \sim 0.1$), and each tower is an independent read-out unit. As can be seen in Fig. 3.5 the system have two pairs of a lead-scintillator electromagnetic calorimeters and iron scintillator hadronic calorimeters, where the latters are radially outer and thicker with respect to the formers. In Tab. 3.1 we summarize the main characteristics of the subsesctions. In central and plug region calorimeters have an EM section backed by an HAD one. In the transition region between central and forward there is an additional HAD system (WHA).

Calorimeter	η range	Depth	Energy Resolution	Absorber
CEM	$\|\eta\| < 1.1$	$19X_0$	$13.5\%/\sqrt{E} \oplus 1.5\%$	lead
PEM	$1.1 < \|\eta\| < 3.64$	$23.2X_0$	$16\%/\sqrt{E} \oplus 1\%$	lead
CHA	$\|\eta\| < 0.9$	$4.7\lambda_I$	$50\%/\sqrt{E} \oplus 3\%$	steel
WHA	$0.9 < \|\eta\| < 1.3$	$4.7\lambda_I$	$75\%/\sqrt{E} \oplus 4\%$	steel
PHA	$1.3 < \|\eta\| < 3.64$	$6.8\lambda_I$	$80\%/\sqrt{E} \oplus 5\%$	steel

Table 3.1: Overview of main technical details of the calorimetry system. The depth is quoted in radiation lenghts X_0 and hadronic interaction lenghs λ_I.

The central region $|\eta| < 1.1$ is covered by the Central Electromagnetic Calorimeter (CEM) and central and end-wall hadronic calorimeters (CHA in $|\eta| < 0.9$ and WHA up to $|\eta| \simeq 1.3$). CEM, located directly outside the solenoid in the radial region $173 < |r| < 208$, is a sampling calorimeter consisting of 31 layers of 5mm thick

polystyrene scintillator as active medium, radially alternated with 3.2mm thick layers of aluminum-clad lead, for a total thickness of 19 radiation lenghts (X_0) . As result of energy loss in the medium, the blue-light emitted by the scintillator is collected by a wave shifters (light to green) placed on the sides of each tower and then is brought to two photo-multipliers for the read-out (Fig. 3.9). CHA has similar structure as the CEM, but with absorber medium made by 32 layers of acrylic scintillator alternated with 2.5cm thick steel absorber (4.7 X_0). For charged pions that do not interact in the CEM [5] the resolution is $50\%/\sqrt{E} \oplus 3\%$. Finally, the WHA is made by 15 layers of acrylic scintillator alternated by 5.1 cm thick steel and extends the coverage of CHA up to $|\eta| \simeq 1.3$ with six additional towers, three overapping the CHA and three out of its coverage.

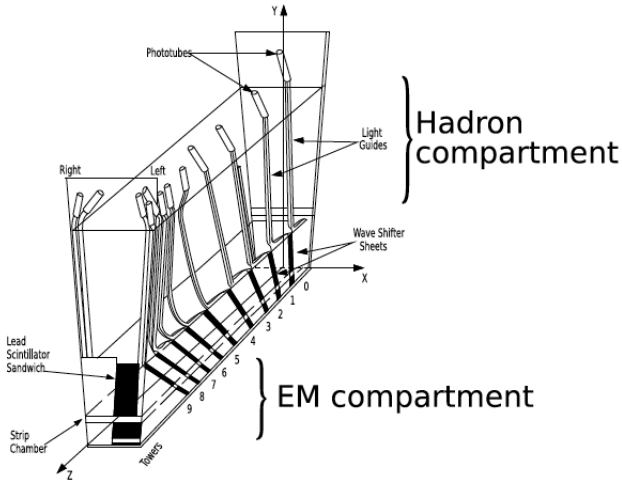


Figure 3.9: Scheme of a wedge of the central calorimeter

The forward region $1.1 < |\eta| < 3.6$ is covered on each symmetric side with respect to the origin by plug-calorimeters: the end-plug electromagnetic PEM and the end-plug hadronic PHA (see Fig. 3.10).

The PEM has a structure similar to CEM: 22 layers of 4.5 mm thick lead absorber alternated by 22 layers of 4 mm thick scintillator tiles. Each scintillator tile is coupled to a different PMT, except for the first layer which is a 10 mm thick plane of scintillator bars read by a multi-anode PMT, that act as preshower detector (the subdetector PPR[6]).

Both the central CEM and the forward PEM calorimeters have two dimensional

[5] 19 X_0 is $1\lambda_{int}$ for charged pions

[6] Plug PReshower: the pre-shower measurement allow discrimination between electrons and pions and improve phton identification

redout strip detectors located at e.m. shower maximum (at $\sim 6X_0$ into the calorimeter), respectively dubbed CES and PES [7]. They are finely segmented to provide a good position resolution for e.m. clusters. Shower maximum detectors contribute to identificate electrons and photons, and help separate them from γ coming from π^0 decays.

The hadronic section of the forward calorimetry is the Plug Hadron Calorimeter PHA, located behind the PEM and similar in structure to the CHA: 23 layers of 2 cm thick steel absorber alternated to 6 mm thick scintillator for a total thickness of $6.8X_0$.

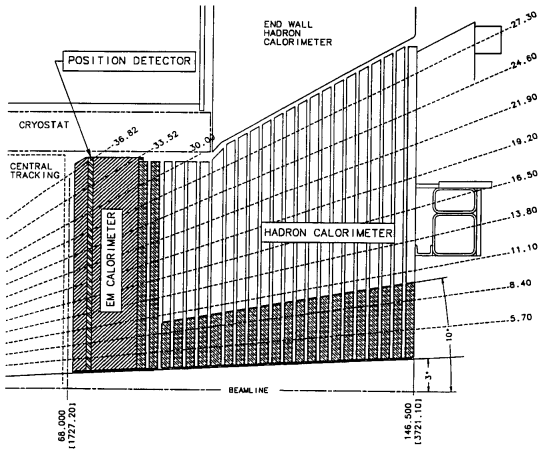


Figure 3.10: Elevation view of a quarter of plug calorimeter

3.2.3 Muon System

The muon detecting system is located outside the tracking and the calorimeter detectors, and is the outermost structure of CDF. High energy muons are m.i.p. losing very little energy while traversing the inner detectors and the calorimenters. The muon system covers the region $|\eta| \leq 1.5$ and it is arranged in four separated subdetectors made of proportional chambers [8] and scintillators (Fig. 3.11).

Outside the CHA, in the region $|\eta| < 0.6$, after a thickness of 5 absorption lenghts, there is the Central Muon Chambers (CMU). Made of four layers of planar multi-wire drift chambers it detects muons with a $p_T \geq 1.4$ GeV. Central Muon uPgrade (CMP) is a second set of four layers of drift chambers, placed behind the magnet return yoke. This is a box made by 60 cm of steel surrounding the central

[7] Central Electromagnetic Shower maximum, Plug Electromagnetic Shower maximum

[8] filled with a mixture of 50:50 arghon and ethane

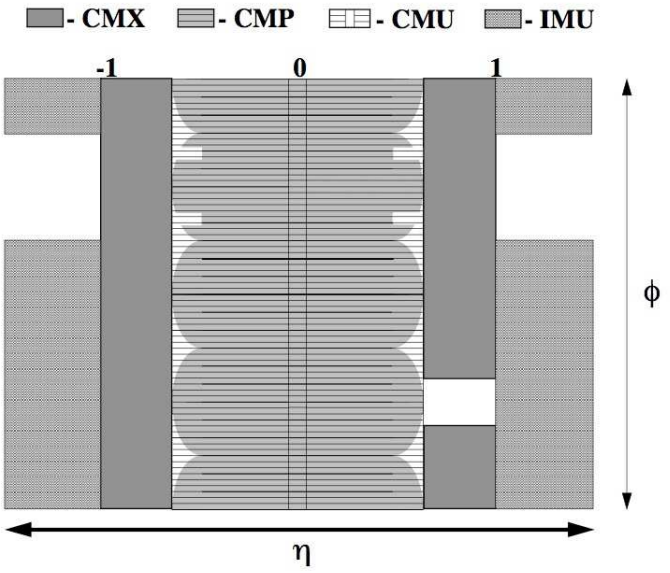


Figure 3.11: Diagram in the (η, φ) plane of the muon systems at CDF: CMU, CMP, CMX and BMU muon detectors. The BMU detector is referred in this diagram as IMU.

detector, which also has the function of absorbing hadrons escaping from central hadronic calorimeter. The CMP detects muons with $p_T > 2.2$ GeV in the same region covered by CMU.

The region $0.6 < |\eta| < 1.0$ is covered by the Central Muon Extension (CMX), a structure of conical sections of drift tubes and scintillator counters located at the west and east ends of central detector. Thanks to the CMX presence, the fully coverage of the COT region is complete. The fourth subsytem is the Barrel Muon Chambes (BMU), a new detector for Run II, which covers the interval $1.0 < |\eta| < 1.5$ and provides the momentum and timing with drift chambers and scintillators.

	η coverage	φ coverage	drift time	p_T threshold (GeV/c)
CMU	$\|\eta\| < 0.6$	360°	800	1.4
CMP	$\|\eta\| < 0.6$	360°	1500	2.0
CMX	$0.6 < \|\eta\| < 1.0$	360°	1600	1.4
BMU	$1.0 < \|\eta\| < 1.5$	270°	800	1.4-2.0

Table 3.2: Design parameters of muon system.

The resolution in the (η, φ) plane for a single hit is ∼25 mm, obtained measuring the difference in time of the drifting path between neighbours cells. The z coordinate is measured with a resolution of ∼1.2 mm.

A muon candidate is identified when a short-track segment (*stub*) resulting from three matching radial-layers matches to the outward extrapolation of a COT track.

3.2.4 Luminosity counters

The instantaneuos luminosity $\mathcal{L}$ of the Tevatron beam is determined by the Cerenkov Luminosity Counters (CLC), two multi-cell gaseous Cherenkov modules placed in both end-plug regions at pseudorapidity $3.7 < |\eta| < 4.7$. Each part is made by 48 conical multicell counters filled with isobutane and arranged in three concentric layers splitted in azimuthal sectors pointing to BØ . As a charged particle, outgoing from the interaction point with a small angle passes through the gas in the CLC, it radiates Cherenkov light if its momentum is above the threshold[9]. Cherenkov light is then collected by the conical mirrors at the base of each cone and sent to the PMT to be read out. The detector is designed to have large signal by particles from $p\bar{p}$ collisions and a small signal from beam halo and secondary interactions.

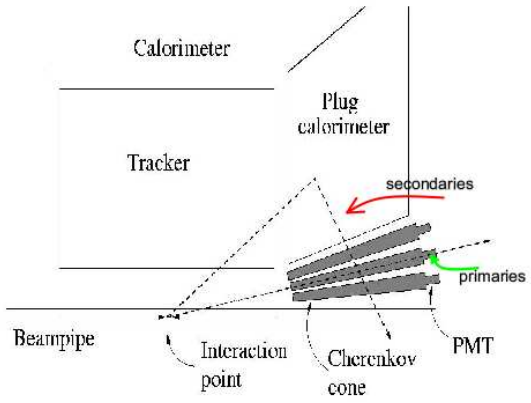


Figure 3.12: Schematic view of the Cherenkov Luminosity Counters

The method used to measure $\mathcal{L}$ is through determination of the rate (the average number) of inelastic $p\bar{p}$ processes with well-known cross sections σ_{in}. The number of $p\bar{p}$ interactions in a bunch crossing follows a Poisson distribution with mean μ:

$$P(n) = \frac{e^{-\mu}\mu^n}{n!} \tag{3.7}$$

If we use the frequency of bunch-crossing $f_{b.c.}$, which is precisely known from Tevatron RF cavities, we obtain the instantaneous luminosity as:

$$\mathcal{L} = \frac{\mu \cdot f_{b.c.}}{\sigma_{in} \cdot \varepsilon} \tag{3.8}$$

[9]With a Cherenkov angle $\vartheta = 3.4°$, the momentum thresholds for light emission are 9.3 MeV/c for electrons and 2.6 GeV/c for charged pions

where ε is the CLC efficency for inelastic scattering. The cross section $\sigma_{in} = 60.7 \pm 2.4$ mb is the inelastic $p\bar{p}$ cross section resulting from the averaged CDF and E811 luminosity-independent measurements at $\sqrt{s} = 1.80$ TeV and extrapolated to $\sqrt{s} = 1.96$ TeV [38].

The total uncertainty on instantaneous luminosity is $\simeq 6\%$ of which about 4.4% comes from measurement of $\bar{N}$ and the efficency and about 4% comes from uncertainty of σ_{in}.

3.2.5 Trigger and Data Acquisition System

For equally spaced 36 on 36 bunches the average interaction rate at the Tevatron would be 2.5 MHz. The actual interaction rate is higher because the beams circulate in three trains of 12 bunches spaced 396 ns, which leads to a crossing rate of 2.53 MHz. This interaction rate is orders of magnitude higher than the maximum rate that the CDF data acquisition system can handle. Furthermore, most of the interesting processes have cross sections much smaller than those for $p\bar{p}$ inelastic collision[10], making pointless trying to record this huge amount of data. An elaborate system of triggers has been designed and put in operation at CDF in order to provide online data reduction and decide on line which event is worth to be recorded.

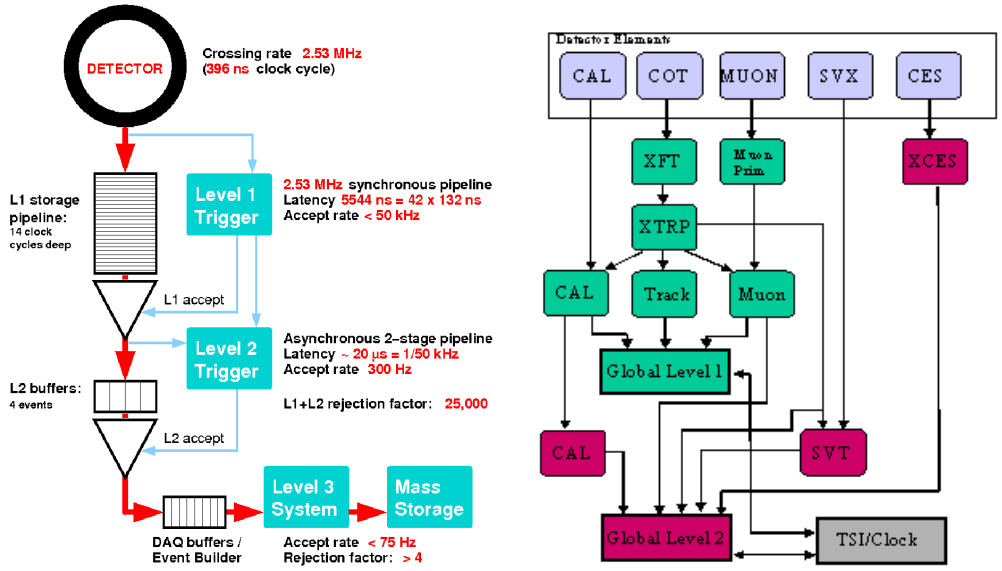
The CDF Data Acquisition System (DAQ) is a three-level trigger structured in two special-purpose hardware levels and one software trigger, as it shown in Fig. 3.13. *Level 1* is composed by three parallel streams which examine each event synchronously with a clock cycle and send inputs to Global Level1 decision unit. The first stream finds calorimeter-based objects, the second one collects information from muon chambers to identify muon stubs and the last stream (the Extremely Fast Trigger, XFT) reconstructs tracks in COT and matches them with energy deposition in the calorimeter towers or muon chambers. From an effective beam-crossing rate of 1.7 MHz, Level 1 reduces event rate down to $\sim$ 15 kHz, storing events during trigger decision time in 42 pipelined buffers synchronized with Tevatron clock cycle of 132 ns. Latency time is 132 ns $\cdot 42 \simeq 5.5\mu$s.

Level2 processes the time-ordered events accepted by Level1 adding information from CES and SVXII detectors. By exploiting the azimuthal position at shower maximum, it provides a better electron recognition and it improves jet identification using a crude reclustering algorithms to merge energies from adjacent towers.

[10] The single top theoretical cross section is $\sigma = 0.88^{+0.12}_{-0.11}$ pb for s-channel and $\sigma = 1.98^{+0.28}_{-0.22}$ pb, several orders of magnitude below the inelastic cross section ($\sigma_{in} = 61.7mb$).

Informations on (r, φ) coordinates read-out by SVX II are combined with Level1 tracks: using SVT data on impact parameter the resolution on tracks is similar to off-line one. SVT allows to trigger on secondary tracks displaced from the interaction point and mostly coming from B and C hadrons decay. Since there are 4 memory buffers the latency time of L2 is $4 \cdot 5.5 \simeq 21\mu$s. At this level the trigger rate is further reduced down to 300 kHz, which is the output rate for the next level.

Level3 is the software level of CDF trigger system: digitalized outputs of L2 accepted events are collected by a modular structure of 16 parallel processors (a LINUX running processor farm). Accessing to the whole event record[11], the software performs event recontruction based on the full detector information, making furthermore a selection based upon a semplified version of the off-line selection criteria. Accepted events are finally transferred to storage mass to be further processed with offline algorithms for data analysis.



(a) (b)

Figure 3.13: Diagrams of the CDF II entire trigger system (a), L1 and L2 hardware triggers (b)

[11] The Event Builder EVB assembles the fragmented output in a logical block of data univocally corresponding to the same bunch crossing, i.e. the *event record*.

3.3 CDF Offline Reconstruction and Data Analysis

The CDF experiment uses a custom object-oriented software to manage data acquisition, offline event reconstruction and most of the analysis. The C++ language is used to define objects, i.e self-consisting structures containing properties and definition common to all entries of the same kind. Usually an *object* is a physics interesting quantity. It can be simple, as a track reconstructed with a particular algorithm or the energy deposit in the calorimeters, or more complex, like a "jet object", that will contain links to tracks, calorimeter energy deposits, vertex information and much more. All the objects characterizing a reconstructed event are stored into large arrays ("n-tuples") written in `ROOT` format [39] and suitable for high level analysis. Different types of those n-tuples were developed at CDF, according to the different types of information needed in the various analysis. In this work we use the so-called `TopNtuples` [40], developed by the Top quark studies group at CDF and widely used in high-p_t analyses.

All analysis streams start from a common stable framework, the Reconstruction package. When a large set of new features is available, a new stable "generation" (or "Gen") of the CDF software is released. Since several years the CDF collaboration is using Gen6 version. The release of a new generation improves several aspects of CDF analysis because, as time goes on, new algorithms are implemented and a better knowledge of the detector is obtained. Unfortunately, all the data acquired before the date of the new release as well as all the Monte Carlo events need to be reprocessed each time with the new software. Sometimes older analysis may need adjustments because of backward compatibility problems. This is called "production" and is a time and CPU consuming task. At the moment of this thesis is submitted the Gen6 is the latest version of CDF software.

3.3.1 Single Top Specific Software Packages

The Single Top quark analysis is performed using several packages on TopNtuples. The TopNtuples are of the same format for data and simulated events of signal and background processes. Those ntuples are the input for the WHAM softare package, that is presented and fully described in App. C.

WHAM performs the selection in order to accept only the event informations that are useful for the single top quark analysis, as described in Chap. 6. As output of the selection, we have a new and size-reduced `ROOT` file named `CatoTree` (Single Top specific file). A complete run on all the W+jets data and MonteCarlo samples

is dubbed as "skim", composed by a dataset of `CatoTrees`. After a complete "skim" the output is used by a specific routine inside WHAM that calculates the expected rate for all the processes as described in Chap. 7.

On the other hand the tree was built (see App. D) in order to be taken as compatible input also for the SingleTop CATO analysis package. In CATO a further program is then run on `CatoTrees`, the `SingleTopReconstruction`. This package reconstructs for event candidates the four-vectors of the top quark, the W boson and the neutrino as four-vectors. Since there are several possible hypotheses leading to the top quark reconstruction, a container class, dubbed as `SingleTopInterpretation`, stores the different hypotheses, so that all of them can later be accessed in the analysis. The resulting `RecoTrees` are then taken as input by the external `NeuroBayes`® package the further Neural Network analysis.

Chapter 4

Identification and Reconstruction of Physics Objects

This analysis deals with the Single Top production processes, which have a final state with a charged and isolated lepton and a neutrino (corresponding to W decay) and two or three energetic jets, at least one of them identified as coming from a b-quark fragmentation.

Each of those elements in final state are physics objects defined after a reconstruction process.

At CDF both real and simulated data go through the same reconstruction path. This is done in two steps. First of all, information from the subdetectors (real or simulated) are collected and combined to form high-level detector objects, as tracks in the tracking system (COT and silicon trackers) or energy clusters in the calorimeters. Then, objects passing given requirements are defined as candidates of physical objects: leptons (electrons or muons), neutrinos or jets. In this chapter we will give a brief explanation of the most relevant physics objects.

4.1 Track reconstruction

The detection and reconstruction of charged particles is one of the most important and essential part of high energy collisions data analysis at CDF II. Tracking detectors provides the measurement of the position of charged particles by identifying the hits in the detector along particle trajectories, so that track parameters can be reconstructed.

Due to the presence of an uniform axial magnetic field in the CDF tracking system (see 3.2.1) charged particles have an helicoidal trajectory[1], with the curvature depending on the momentum and charge of the particle. Therefore, the helix direction points back to the origin of the track. First hits are reconstructed at detector level, then are collected in candidate tracks by the pattern recongnition programs. The reconstruction of a charged-particle trajectory consists in determining the helix parameters through a χ^2-based fit of a set of the spatial measurements of a candidate track. The helical fit takes into account field non-uniformities and scattering in the detector material.

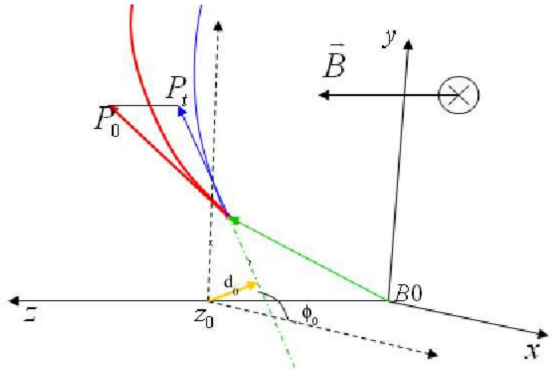


Figure 4.1: Helix parameters of the track in the magnetic field and coordinate system

Five parameters are used to describe the helix trajectory (see fig. 4.1):

- $C \equiv 1/2qr$: the half-curvature of the trajectory, where r is the helix radius and q the charge of the particle;
- $d_0 \equiv q\sqrt{x_0^2 + y_0^2} - r$: the impact parameter, i.e. the distance of closest approach in the transverse plane between the helix and the origin, where x_0 and y_0 are the coordinates of the center obtained by the projection of the helix on the transverse plane;
- λ: the helix pitch i.e. the cotangent of the polar angle between the track and the z-axis;
- z_0: the z position of the track vertex ;
- φ_0: the azimuthal angle of the track at its vertex.

[1]Neglecting at first order of approximation the energy loss by ionization in the detector material

4.1.1 Tracking Algorithms

The five parameters described above are derived from the hits detected by the tracking system using a track pattern recognition algorithm. The experiment exploits several tracking algorithms [41], each optimized for the information available in different detector region.
At the beginning, tracks are reconstructed in the COT using two different algorithms. The first ("segment linking") reconstructs and connects segments in each of the 8 COT superlayers and fits them to a circle with the minimum χ^2 method in order to find the trajectory. The second ("histogram linking") operates at single hit level starting from the outermost superlayer and fill the histogram with the distance from the center of the circle of all hits in 1cm wide road, adding hits to the track if belonging to a bin with more than 10 hits. Duplicates reconstructed by both algorithms are removed. Alltogheter these algorithms are able to reconstruct tracks of charged particles passing the entire volume of the COT. Due to the detector geometry, those tracks are limited to the range of $|\eta| \leq 1$ and $p_T > 0.5$ GeV/c, and the efficency of the algorithms quickly falls for larger η or smaller p_T.

The silicon track reconstruction is performed by three different algorithms.

The first one ("Outside-In") uses tracks found in the COT which are then extrapolated into the silicon system. Axial and stereo silicon hits are added to the track if they lie on a predefined road defined using the uncertainties on COT track parameters. The impact parameter resolution of those COT+SVX tracks is found to be $\sigma_{d_0} \sim 20\mu m$.

The second algorithm ("Silicon-stand-alone") consists of stand-alone pattern recognition in the silicon sub-detectors. Since hits already assigned to another track are not considered, this method is mostly used to find tracks in the forward region up to $|\eta| < 2$ and, with lower efficency, up to $|\eta| \lesssim 2.8$.

The third algorithm ("Inside-Out") is built on the purpose to recover efficency and p_T resolution in the region $1.2 \leq |\eta| \leq 1.8$ where the COT coverage is limited. This is done matching tracks extrapolated from the previuos Silicon-stand-alone alghoritms with COT tracks in order to identify particles that did not pass the entire COT volume.

4.1.2 Primary Vertex Reconstruction

The determination of the primary interaction point of the particle collisions is essential in the analysis because it affects crucial parameteres of the kinematic of

the event. Moreover, it is needed to identify displaced secondary vertices.

Primary vertex position is reconstructed by two alghoritms, `PrimVtx` and `ZVertex`. `PrimVtx` uses as seed the average event z position measured during collisions. All tracks with $|z_{\rm trk} - z_{\rm vtx}| < 1$ cm, $|d_0| < 1$ cm and $|d_0/\sigma_{d_0}| < 3$ are selected, ordered in decreasing p_T and fitted to a common vertex. The tracks with χ^2 above a given threshold are removed and the procedure is iterated until all tracks have fullfilled the requirements. `ZVertex` selects vertices from tracks passing minimal quality requirements and removes those without a minimum associated number of tracks with $p_T > 300$ MeV. A track is associated to a vertex if its distance from the vertex is less than 1 cm (silicon vertex) or 5 cm (COT vertex). The z position of each vertex is the mean z_0 of its associated tracks, weighted on the respective uncertainty σ_{z_0} .

4.2 Electron Identification

An electron typically deposits most of its energy in the electromagnetic calorimeters. The basic idea for the identification of an electron candidate is to reconstruct it as physics object from an high-p_T isolated track in the COT matched to an energy deposit (also called *cluster*) in the electromagnetic calorimeters. For central and forward (plug) electrons the clusters must be in the CEM and the PEM, respectively. In the following we will dscribe the definition used for high-p_T electrons.

4.2.1 CDF EM Clustering Algorithm

In central region electron identification relies on matching a track in the COT with a cluster in the CEM central calorimeter.

The CDF electromagnetic clustering algorithm [42] maps in the (η, φ) plane the calorimeter towers. Then, it orders the towers in decreasing observed energy. It takes the most "energetic" towers (with $E_T > 2$GeV), also called *seed* towers, and creates a cluster in the plane adding the neighbours towers in order to form a 2x2 (for PEM region) or 3x3 (for CEM region) clusters [2]. A cluster is not allowed to cross the boundary between different subdetectors (i.e. between PEM and CEM). The energies measured in the shower max (PES) and preshower (PPR) detectors (see 3.2.2) are added to the reconstructed energy. PES is also used to compare the shower profile of electrons or photons to the profile measured in an electron test beam (to distinguish electrons from photons) and to measure the spatial position

[2] Different clustering for CEM and PEM are used in order to reduce the probability of overlapping between clusters of two different electrons.

of the EM shower centroid in the (η, φ) plane.

After the cluster energy is reconstructed, the used towers are removed from the lists and the algorithm iterates the process by selecting the next seed tower. The algorithm ends when all seed towers are used.

4.2.2 Tracks Matching and Electron Reconstruction

The next step is to search for a track to be associate to the cluster. From all the tracks reconstructed in the COT which point to an instrumented region of the CEM detector, the one with the highest p_T that best matches the cluster is chosen as the geometrical center of the main cluster.

The selected track and EM cluster must pass several requirements to be matched. For the CEM electrons, identified in the central calorimeter CEM within a range $|\eta| < 1.1$, the requests are on *kinematic* and *identification* variables. First we consider the *kinematic* selection variables:

- $E_T > 20$ GeV;
- $|z_0^{\text{track}}| < 60.0$ cm;
- $N^{\text{SL}}_{\text{Axial}}(\text{with} \geq 5 \text{ hits}) \geq 3$;
- $N^{\text{SL}}_{\text{Stereo}}(\text{with} \geq 5 \text{ hits}) \geq 2$;
- $p_T^{\text{track}} > \begin{cases} 50.\ \text{GeV/c} & \text{if } E_T \geq 100.\ \text{GeV} \\ 10.\ \text{GeV/c} \quad \text{and } E/p < 2 & \text{if } E_T < 100.\ \text{GeV} \end{cases}$
- IsoRel < 0.1.

which correspond to the following definitions (E_T and z_0^{track} are trivial):

- $N^{\text{SL}}_{\text{Axial(Stereo)}}$ are the number of COT SuperLayers, with at least five hits in each superlayer, where a track is registered.
- $E^{\text{cluster}}/p^{\text{track}}$ is defined as:

$$\frac{E^{\text{cluster}}}{p^{\text{track}}} = \sqrt{1 + \frac{m}{p}} \tag{4.1}$$

  Its distribution is peaked at 1 because the electrons have much smaller mass than their momentum. However, it has a large tail beyond 1 because electrons can radiate collinear photons in the passage through matter in the tracking volume.

- Electron isolation, or *IsoRel*, is an important quantity derived from calorimeter informations. It is defined for all electrons as:

$$\text{IsoRel} \equiv \frac{E_T^{0.4} - E_T^{cluster}}{E_T^{cluster}} < 0.1 \tag{4.2}$$

  where $E_T^{0.4}$ is the energy collected by the calorimeter within a radius $\Delta R < 0.4$ from the centroid of the EM cluster. Studies performed with clean $Z_0 \to e^+e^-$ events [43] show that electrons detected in the central or in the plug region have a little deposit in the hadronic part of the calorimeter (see fig. 4.2). Isolation is a topological variable used in analyses involving a $W^{\pm}$ or Z boson since leptons coming from the bosons decay are usually far from jets or other particles.

Now we can consider the requirements on the *identification* variables:

- $|\Delta z| \leq 3$ cm;
- $-3 \leq Q \cdot \Delta x \leq 1.5$;
- $\chi^2 < 10$;
- $L_{\text{shr}} \leq 0.2$;
- $E_{\text{HAD}}/E_{\text{EM}} \leq (0.055 + (0.00045 \cdot E))$;

which correspond to the following definitions:

- Δz and Δx are the difference in the distances between the CES cluster position and its associated extrapolated track along z and x (x is a local variable).
- χ^2 is the result from a fit of the measured shower profile to an electron baseline shower profile acquired during test beam.
- L_{shr} is the lateral shower sharing variable, defined as:

$$L_{\text{shr}} = \frac{0.14 \sum_i (E_i - E_i^{\text{exp}})}{\sqrt{(0.14\sqrt{E_i})^2 + \sum_i (\Delta E_i^{\text{exp}})^2}} \tag{4.3}$$

  where E_i is the measured energy in the i-th tower, E_i is the expected value for the energy in the same tower according to a parametrization based on test-beam data. The index i runs over all the towers adjacent to the seed one.
- $E_{\text{HAD}}/E_{\text{EM}}$: The ratio of energy measured in the hadronic calorimeter divided by the energy measured by the electromagnetic calorimeter. Electrons deposit the majority of energy in the electromagnetic calorimeter, so this ratio should be small for real electrons.

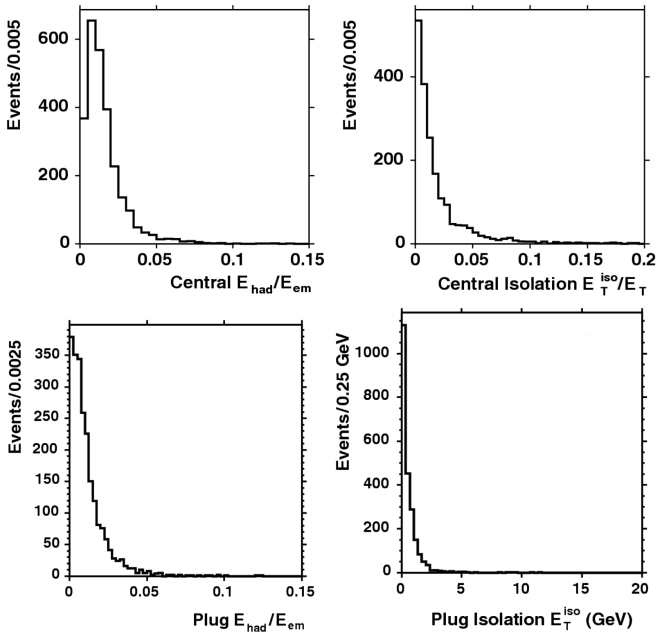


Figure 4.2: E_{Had}/E_{EM} (left) and isolation (right) distribution of central (top) and plug (bottom) calorimeter electron selection from unbiased, second legs of $Z_0 \to e^+e^-$ candidate events in data

Phoenix Algorithm for Plug Electrons

The forward (plug) region ($1.2 \leq |\eta| \leq 2.0$) is not covered by the COT, making the identification of forward electrons more complicated. d. At CDF a special alghoritm -"Phoenix"[44]- was developed in this sense. It combines calorimetric information, the energy cluster in the PEM, and stand-alone silicon hits in the SVXII in order to identify forward electrons that are not tracked by the COT. The position of the cluster and the primary vertex are used to construct the *seeds*. For each seed, two hypotheses about the charge of the particle are considered by computing the curvature for both an electron and a positron corresponding to the deposited energy. If the hits in the silicon match the seed, a PHX (from "Phoenix") electron is found.

Photon Discrimination

At CDF many photons convert in e^+e^- pairs. Therefore, photons can be misidentified as an electron if the electron track is not reconstructed or if the photon converts to an electron-positron pairs ($\gamma \to e^-e^+$) as it transverses the tracker. Alternatively, a track can be mistakenly associated to the EM cluster of a photon [45]. Electrons from photon conversion are identified and rejected if another

opposite-sign COT track close in space is found.

4.3 Muon Identification

Muons plays a crucial role in the Single-Top analysis and in any other analysis using W+jets sample, as W bosons are likely to decay in $\mu\nu$ as well as in $e\nu$. The basic idea of muon identification is that high energy muons pass through the calorimetric system as Minimum Ionizing Particles (MIP). They are identified as a track inside the COT matching an energy deposit consistent with a MIP inside EM and HAD calorimeters and a track segment ("stub") in the outer muon chambers. Muons can be faked by cosmic rays, by π, k leptonic decays in the tracker and by hadrons deeply showering or not showering at all inside the calorimeters ("punch-through").

The track of a candidate muon must satisfy the following requirements:

- $p_T > 20$ GeV/c;
- $E_T/p_T < 0.1$;
- $E_{\rm EM} < 2 + \max(0, (p - 100) \cdot 0.0115)$ GeV;
- $E_{\rm HAD} < 6 + \max(0, (p - 100) \cdot 0.028)$ GeV;
- $N^{\rm SL}_{\rm Axial}$ with ≥ 5 hits ≥ 3;
- $N^{\rm SL}_{\rm Stereo}$ with ≥ 5 hits ≥ 2;
- $|z_0| < 60.$ cm;
- $|d_0| < \begin{cases} 0.2 \text{ cm} & \text{for tracks w/o silicon hits} \\ 0.02 \text{ cm} & \text{for tracks with silicon hits} \end{cases}$

Tight central muons CMUP and CMX are reconstructed in the corresponding muon chambers if belonging to their respective fiducial regions (CMU and CMP for the firsts, CMX region for the others). Additional requirements have to be fullfilled for the reconstruction. For the CMUP:

- $|\Delta x_{\rm CMU}| < 3$ cm;
- $|\Delta x_{\rm CMP}| < 5$ cm;

while for the CMX it is required:

- $|\Delta x_{\rm CMX}| < 6$ cm;

- $\rho_{\mathrm{COT}} > 140.$ cm;

where $\Delta x_{\mathrm{CMX/CMU/CMP}}$ is the distance between the extrapolated COT track and the matched stub; ρ_{COT} is the radial distance from the beam pipe at which the reconstructed track crosses the end plane of the COT. The latter requirement ensures that the muon is within the range of the CMX trigger.

Loose muons are included in the inclusive EMC category which includes several mutually esclusive subcategories depending by the muon sub-detector in which they are identified [46]. First two loose categories are CMU-only and CMP-only which must have a stub and lie in their namesake subdetectors fiducial region, while no stubs in the other subdetector has to be found and $|\Delta x_{\mathrm{CMU(CMP)}}| > 7(5)$ cm is required. CMIO muons are events with no stubs in CMU,CMP,CMX,BMU detectors and not belonging to their fiducial region, but passing a minimum energy cut of $E_{\mathrm{EM}} + E_{\mathrm{HAD}} > 0.1$ GeV. The SCMIO muons have a stub in some muon detector but do not meet any other requirement but the $E_{EM} + E_{HAD} > 0.1$ GeV cut. The reconstruction of BMU muons requires a BMU stub and a track fiducial to the BMU chamber with $|\Delta x_{\mathrm{BMU}} < 9|$ cm. Additional requirement and an exahustive description of this category can be found in [47]. Finally, CMXNT have the same CMX requirements but they extend their fiducial region to the most forward part of CMX detector ($|\eta| \sim 1$) that is not usable for the trigger.

4.4 Jet definition and reconstruction

Emerging quarks manifest themselves as "jets" of particles as they cannot exist as free particles and must form colorless hadrons. The *hadronization*, or *showering*, is the process involving partons resulting from the interactions. The result of such a process is a collimated cluster of stable particles named "jet", which approximately retains the total momentum and direction of the initial parton (fig. 4.3). We define the jets from the characteristic energy deposition in calorimeters, as result of the particles flying away from the collision point.

A universal definition of a jet does not exist in HEP. However, from an experimental point of view, a jet is defined as a large energy deposit in a localized area of the detector, in our case measured by calorimenters[3].

The main purpose of jet analysis is to recover information from the detector about initial energy, momentum and, possibly, the nature of the parton produced in the original interaction. A "jet algorithm" is a tool to reconstruct such information.

[3]Typically jet energy is found for $\sim 40\%$ in EM and for $\sim 60\%$ in HAD calorimeter

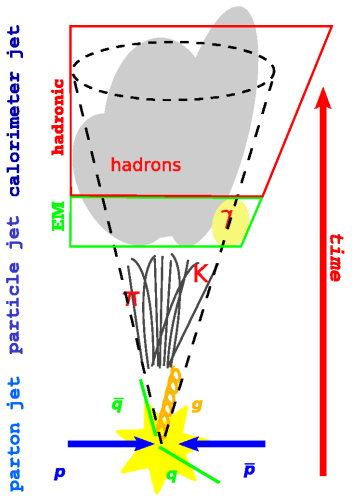


Figure 4.3: Schematic picture representing the hadronization process from the initial parton to the final cluster, called jet.

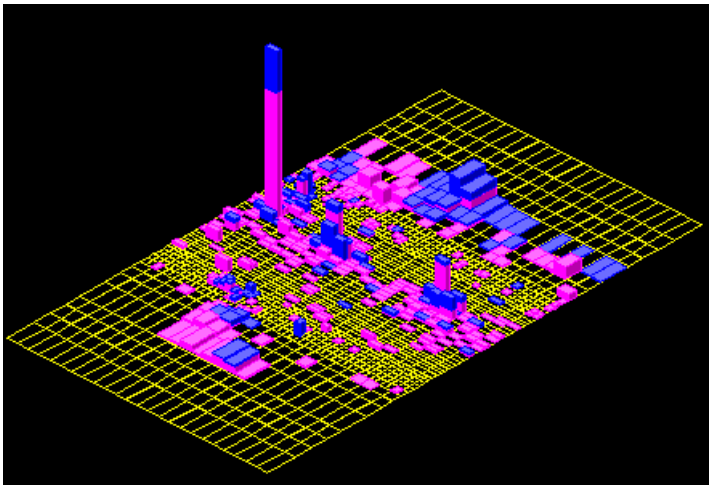


Figure 4.4: Calorimetric deposit in the (η, φ) plane as represented in the CDF event display of a typical event. EM deposits are red while HAD deposits are blue.

The jet energy is calculated from the energy deposited in the electromagnetic and hadron calorimeter towers using a jet clustering algorithm [48]. The official jet algorithm at CDF is dubbed `JETCLU` [49], an iterative fixed cone jet reconstruction algorithm based only on calorimetric information.

4.4.1 `JETCLU` Cone Alghoritm

The basic idea of the cone algorithm is to cluster the jet with a fixed cone size in which the center of the jet is defined as $(\eta_{jet}, \varphi_{jet})$ and the size of the cone is $R = \sqrt{(\eta_{\text{tower}} - \eta_{\text{jet}})^2 + (\varphi_{\text{tower}} - \varphi_{\text{jet}})^2} \leq R_{cone}$. In our analysis we will use jets

with $R_{cone} \leq 0.4$, but at CDF there are analyses that use cone size of 0.7 and 1.

The algorithm begins by creating a list of *seed* towers, sorted by decreasing E_{T_i} above the threshold of 1 GeV, where $E_{T_i} = E_i \cdot \sin\vartheta$ is the transverse energy deposited in tower i^{th} with respect to the primary vertex z position, and E_i is the sum of energies measured in the electromagnetic and hadronic components of that tower. For each seed tower, starting with the highest-E_T one, a precluster is formed by combining together all adjacent seed towers within a cone of given radius R. The procedure is iterated for the next tower until the list is over. Now, for each precluster, the E_T-weighted centroid is calculated and a new cone of radius R, centered on the new centroid, is defined. New towers with $E_T > 100$ MeV are added to the cluster in that radius and a new centroid is formed. The process is iterated until the centroid of the energy deposition within the cone is aligned with the geometric axis of the cone (that is called "stable solution"). Jets are merged if they overlap by more than 50%; otherwise, each tower in the overlap region is assigned to the nearest jet in (η, φ) plane. The final jet energy is computed from the final list of N_{tow} towers:

$$E_{\text{jet}} = \sum_{i=0}^{N_{\text{tow}}} (E_i^{EM} + E_i^{HAD}) \tag{4.4}$$

4.4.2 Jet Corrections

The ultimate purpose of the jet reconstruction algorithm is to obtain the best estimate of the energy of the outgoing partons produced by the initial hard scattering. The raw energy reconstructed by the cone alghoritm can be quite different from the parton energy at the beginning of the hadronization, because of several factors depending from the event as well as from the response of the detector.

CDF developed a set of generic jet energy corrections depending raw of η, E_T and R of the jet reconstructed by `JETCLU` algorithm. The corrections are divided in several sub-levels (L1,L4,L5,L6,L7[4]) and separately stored, so that a subset of them can be applied according to the different analyses [50] [51].

Level-1: η dependent Correction

Due to the geometry of the CDF calorimeter, its response is not uniform in pseudorapidity. L1 correction takes into account the different performances of calorime-

[4]The actual naming skips a Run I step named L2, because it is absorbed in L1, and L3, as it was introduced as a temporary MC calibration in Run II.

ter in central and forward region and other hardware non-uniformities due, for example, to the presence of cracks between sections of the calorimeter.

The final L1 correction is defined as:

$$f_{L1}(\eta, E_T^{raw}, R) = \beta^{-1} \tag{4.5}$$

where β is a factor depending on the average momentum balancing fraction between jets selected only in the central part of the calorimeter ($0.2 < |\eta| < 0.6$) and jets in the wide η spectrum ($|\eta| < 3.6$), measured using a large di-jet sample (Fig. 4.4).

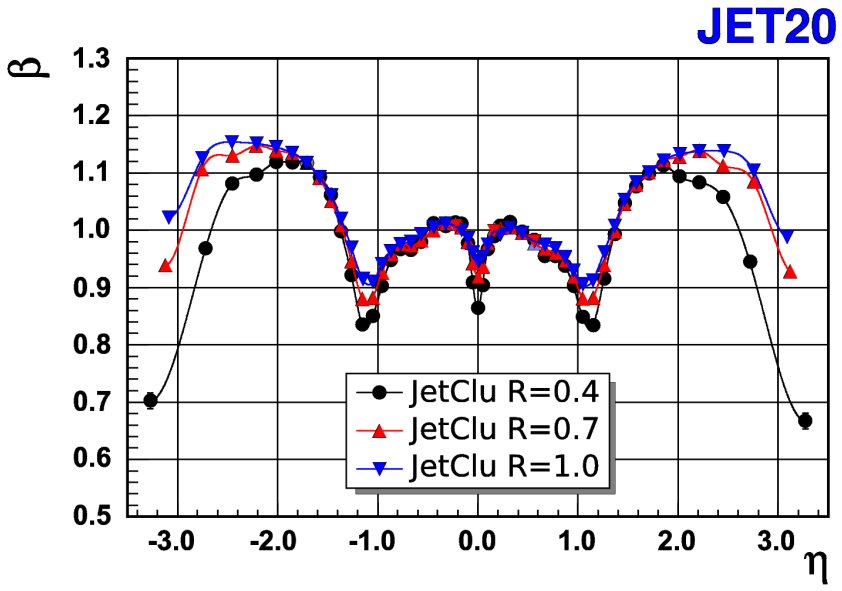


Figure 4.5: η-dependence of β factor for different cone radii measured in the di-jet component of jet20 sample

Level-4: Multiple $p\bar{p}$ Interactions Correction

At higher luminosities, more than one $p\bar{p}$ interaction occurs at the same bunch crossing. The number of interactions per bunch crossing varies and is Poisson-distributed with the mean value depending almost linearly from the luminosity. The result of these additional minimum-bias events can be an amount of extra energy deposited in the calorimeter, which therefore needs to be subtracted from the jet energy.

Level-5: Absolute Jet Energy Scale Correction

The L5 correction deals with the difference between the final jet energy measurement of the event in the calorimeter with respect to the parton level physics.

The study is an MC driven comparison between simulated jet events that are generated by MC and then go through the full CDF detector simulation. Then, after simulation, jets are reconstructed both at calorimeter and hadron generation levels (HEPG) using the of same clustering algorithm. A calorimeter jet (C) is associated to the corresponding hadron jet (H) if $\Delta R < 0.1$. For both HEPG and detector jets the transverse momenta p_T^H and p_T^C are calculated. The jet energy correction is defined as $\mathcal{P}(p_T^C/p_T^H)$, the probability to measure p_T^C, given p_T^H. Figure 4.6 shows the correction factor f_{L5} for different cone sizes as function of the different jet E_T. The total uncertainty is about 3% and arises mainly from the determination of calorimetric response to single particles and MC fragmentation modeling.

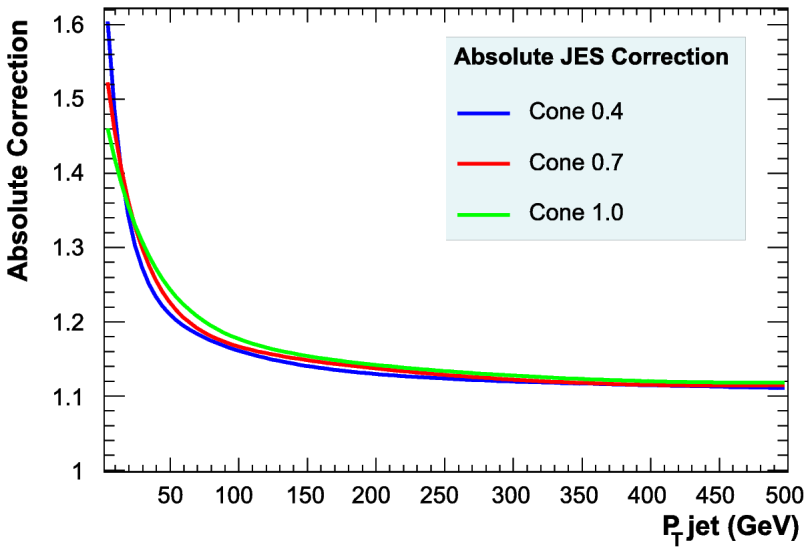


Figure 4.6: Absolute jet energy scale correction (f_{L5}) for different cone sizes.

This is the most important correction and the L5-corrected jet energy will be used everywhere in our analysis, unless otherwise specified.

Level-6: Underlying Event Correction

In order to determine the energy of the original parton rather than the corresponding jet energy, a correction taking in account the interaction that can occour between spectator partons or originating from initial-state gluon radiation (ISR) is introduced. The energy contribution coming from these "underlying" event must be subtracted to the total jet energy. It is usually parametrized with a value that scales with the cone radius and that was determined with minimum-bias studies done during Run I.

Level 7: Out of the Cone Correction

Level 7 correction considers the fraction of particles coming form the original parton, usually final-state gluon radiation (FSR), that falls outside the jet clustering cone up to $R = 1.3$, because they are radiated at large angles with respect to the parent parton. This energy contribution must be added to the total jet energy and is given by corrections representing the cone-size dependance of the energy depositions determined with the same jet-to-parton matching method used for Level5.

Corrected Jet energy

By construction, the L1 and L5 correction are moltiplicative factors (f_{L1} and f_{L5}) applied on the raw E_T of the jet, while L4, L6 and L7 corrections are additive constants (A_{L4}, A_{L6} and A_{L7}). The general equation used to apply all the correction is:

$$E_T^{corr}(\eta, E_T^{raw}, R) = (E_T^{raw} f_{L1} - A_{L4}) f_{L5} - A_{L6} + A_{L7} \tag{4.6}$$

4.5 Neutrino Identification

CDF detects neutrinos indirectly, as they pass through the detector without interacting. Their presence is revealed exploiting momentum conservation.

In principle, they manifest as missing energy in the overall three-dimensional energetic balance of the event. While is not possible to know the exact longitudinal momentum of the colliding partons, the trasverse component p_T of the interacting partons is assumed to be zero. By eploiting the conservation of the transverse energy we can measure the neutrino momentum as the $\not\!\!E_T$:

$$\vec{\not\!\!E_T} \equiv -\sum_i E_T^i \hat{n}_i \tag{4.7}$$

where i denotes the calorimeter tower number with $|\eta| < 3.6$ and total energy deposit above 0.1 GeV, $\hat{n}_i$ is a unit vector which points from the interaction vertex to the center of the i^{th} calorimeter tower.

At offline level, the algorithm corrects for the position of the reconstructed event vertex and for any identified muon in the tracker. We discuss the main corrections to $\not\!\!E_T$ in the section 4.6.

4.6 Corrections to the Raw Measured $\not\!\!E_T$

The raw missing transverse energy for any event is defined as vector by (4.7). Due to the presence of the muons, which do not shower in the calorimeter, and the variuos jet energy corrections, the $\not\!\!E_T$ has to be corrected for such effects.

For the muons, the correction is performed in two steps. First of all, the muon track is extrapolated to the calorimeters and an energy corresponding to a MIP ($\sim$ 350 MeV in CEM and $\sim$ 1.6 GeV in CHA) is removed from the towers traversed and from the total p_T balance. Then, the muon tranverse momentum is included in the total p_T balance.

For the jet energy corrections, the procedure is similar to the muon case: the contribution given by the raw measured energy of the identified jets is removed from the total P_T balance. Then, each jet is corrected at the level used in the analysis (L5 in our case) and the p_T of all jets, corrected at the desired level, is put back in the total P_T balance. In order to avoid underestimates of the total activity in the calorimeters because of the minimum bias energy subtraction performed by L4 correction, the jets are corrected in this case excluding L4 correction if the applied level is superior (as in our case, Level 5).

In the end, the formula for the missing transverse energy changes from (4.7) to the following (4.8):

$$\vec{\not\!\!E_T}^{\,L5} = -\sum_{i=1}^{N_{tow}} E_T^i \cdot \hat{n}_i + E_T^{MIP} \cdot \hat{n}_\mu - \vec{P}_{T_\mu} + \sum_{j=1}^{N_{jets}} P_{T_j}^{RAW} - \sum_{j=1}^{N_{jets}} P_{T_j}^{(L5-L4)} \qquad (4.8)$$

4.7 Secondary Vertex Tagging

A fondamental tool in collider physics is the identification of jets coming from heavy flavor, i.e. botton or charm hadrons. This tool is strongly required in top analysis and is powerfully used to reject background processes with only light flavor final state.

Heavy flavor jets can be identified by an observable secondary vertex inside a jet cone. The top quark, after its production, decays instantaneously (due to its mass of $\sim$ 172.5 GeV/c^2) into a W boson and a b quark. The latter, independently of its production mode, hadronizes in B hadron, that can be a b-meson as $B^0, B^\pm$ or B_s, or a b-baryon as the Λ_B and usually carries most of the transverse momentum of the original b quark.

B hadrons decay with a life time of ~ 1.6 ps which, together with a large relativistic boost, make the the decay length of the order of some millimetres, that is enough displaced from the primary vertex to be appreciable for the detector[5]. The further decay of B-hadrons produces tracks with large impact parameter d_0 i.e. with low probability of coming from primary vertex. The impact parameter is reconstructed by the silicon detector with a precision of $\approx 50\mu m$, making possible to separate displaced tracks from prompt tracks coming from the primary interaction, as shown in fig. 4.7 for a W+jets event with two displaced secondary vertices. Those are identified by the Secondary Vertex Tagger algorithm (`SecVtx`)[52], that is one of the main b-taggers used at CDF and the mostly used in this analysis.

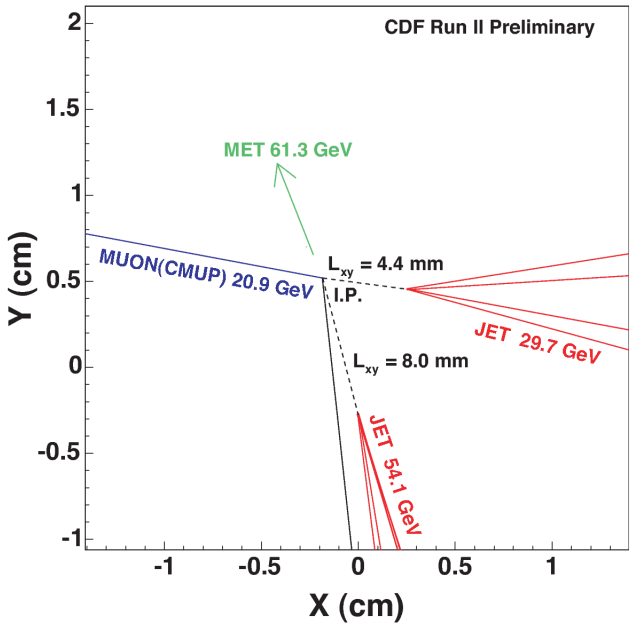


Figure 4.7: W +jets candidate event with two secondary vertices tagged by SecVtx (run 166063, event 279746). The ET direction, a muon track, a prompt track and tracks from the secondary vertices are shown

4.7.1 `SecVtx` algorithm

The `SecVtx` algorithm searches for secondary vertices using the tracks within the jet cone (i.e. with $\Delta R_{jet,trk} < \Delta R$) of radius $\Delta R = 0.4$, selecting tracks

[5] as g.e. the mean transverse momentum of a b-hadron coming from a single-top s-channel process is ~ 53.4 GeV/c. Since B-hadrons has a mass of ~ 5.28 GeV/c^2, it undergoes a boost of $\beta\gamma \sim 10.12$ with a decay length of about 4.8 mm.

with large d_0 with respect to the primary vertex, in a fiducial region of $|d_0| < 1.5$ mm, and trying to merge them into a common vertex. Tracks passing certain requirements ($p_T > 0.5$ GeV/c, $d_0/\sigma_0 > 2.0$) are defined as "usable", while a jet containing at least 2 usable tracks is defined as "taggable". The algorithm uses a two-pass approach to find secondary vertices: the first pass attempts to reconstruct a secondary vertex including at least three tracks with $d_0/\sigma_0 > 2.5$ and at least one with $p_T > 1.0$ GeV/c; the second is performed in case of negative pass-1 result and requires only two tracks but with $d_0/\sigma_0 > 3.5$, one track with $p_T > 1.5$ GeV/c and the other one with $p_T > 1.0$ GeV/c. If a secondary vertex is found in a jet, the jet is defined as *tagged*. The two-dimensional decay length of the secondary vertex L_{xy} is calculated as the projection into the jet axis, in the (r, φ) plane, of the vector pointing from the primary vertex to the secondary vertex. The sign of L_{xy} is defined by the angle α between the jet axis and the SecVtx vector, being positive if $\alpha < \pi/2$ and negative if $\alpha > \pi/2$, as shown in fig. 4.8:

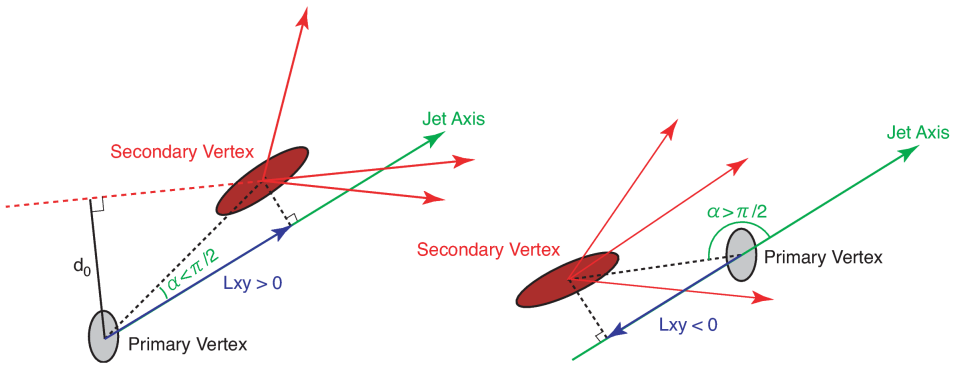


Figure 4.8: Left: true reconstructed secondary vertex. Right: negative SecVtx tag, falsely reconstructed negative secondary vertex

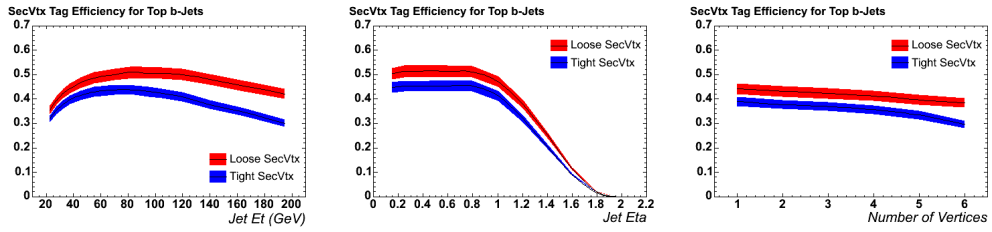
Events containing heavy flavour hadrons, i.e. with tagged jets, are expected to have secondary vertices with large positive L_{xy}. To reduce background from mismeasured tracks, a cut on $|L_{xy}/\sigma_{xy}| > 7.5$ is required. Negative tags are due to the finite tracking resolution of the CDF tracking system, but are not cut away since they are useful to calculate the false positive tag rate (mistag) background, as we will explain in Sec. 7.7.

Due to the difficulty of the b-tagging to reach optimal efficency in the wide variety of b-hadrons decays, CDF developed other tagger algorithms beside `SecVtx`. In particular, the "soft lepton tagger" alghoritm looks for semileptonic heavy flavor decays (the difficulty is to discriminate low energy leptons inside high energy jets), while the "jet probability" alghoritm [53] assigns to each track the probability to

come from the primary interaction vertex, setting different thresholds for heavy flavor tagging possibility. New b-taggers that use multivariate techniques were tried and validated such as the "Roma b-tag"[54], the "NeuroBayes®b-tag" [55] or the newest "B-ness" tagger. However, in my analysis I will use as btagger only SecVtx. The NeuroBayes®b-tag (NeuroTag) value is commonly stored for each event and is used as input variable in the NN training part (Sec. 8.1.1).

4.7.2 Tagging Performancies and Scale Factors

The performances of a b-tagger are evaluated on its efficiency, i.e the rate of correctly identified b-hadrons over all the produced b-hadrons, and on its purity, i.e the rate of falsely identified b-hadrons in a sample with no true b-hadrons (mistags). CDF uses $t\bar{t}$ MonteCarlo to evaluate `SecVtx` efficiency relying on detector and physical processes simulation. In fig. 4.9 the b-tagging efficiency is plotted as a function of jet η, E_T and the number of vertices for two SecVtx modes *tight* and *loose*. Tagging efficiency drops essentially because of lower track reconstruction efficency.



(a) (b) (c)

Figure 4.9: b-tagging efficiency for the two operation modes of the SecVtx algorithm vs η(a),E_T(b),nVtx(c).

As MonteCarlo simulation does not reproduce the exact b-tagging efficiency of SecVtx, a "Scale Factor" (SF or Φ) [56] is introduced to account data/MC difference in the form of the ratio $\Phi \equiv \varepsilon_{data}/\varepsilon_{MC}$. The tagging efficency is calculated multiplying the tag rate of jets matched with heavy flavour quarks with the data /MC scale factors, that are listed for the three modes in Tab. 4.1.

mode	Φ	$\sigma_\Phi(stat)$	$\sigma_\Phi(syst)$
loose	0.96	0.01	0.05
tight	0.96	0.01	0.04

Table 4.1

Chapter 5

Monte Carlo Simulated Samples

In this chapter we describe the Monte Carlo simulation of events used to model the various processes of signal and background. Beyond the Single Top signal samples, we identify the following processes as source of background events: top-quark pair, diboson production (WW,WZ,ZZ), W+jets and Z+jets production, QCD multijet events. We use Monte Carlo generators to simulate signal and background processes, except for the QCD multijet background (also dubbed "non-W"), which is estimated using a mix of simulated and data set (see Sec. 7.4).
The method of classification of the various backgrounds and their normalization estimate will be described in Chap. 7.

5.1 Monte Carlo Event Simulation

Quantum field theory calculates the kinematic properties of leading-order Feynman diagrams of processes generated from $p\bar{p}$ collision in the CDF detector at quarks (parton) level. On the other hand, as quarks do not exist as free particles, but materialize as showers of hadrons, we must use MonteCarlo programs to simulate the parton [1] showering.
A MonteCarlo simulation (MC) starts from the leading order diagram of the process of interest and then identifies the total phase space available for the interacting particles through statistical techniques, yielding as output individual events that approximate the real physics of the process at higher perturbative level. Those particles are then simulated in their interactions with the detector.

[1] In this context, parton is a generic term used for a constituent of particle and represents physically either a quark or a gluon.

5.1.1 Monte Carlo Events Generation

Signal Events

The signal (single top s-channel and t-channel) MonteCarlo simulated events are generated with POWHEG [24] as described in [57]. In the previous analysis CDF used MADEVENT [58] program.

The main improvement obtained by using POWHEG generator is the implementation of NLO process calculation. POWHEG package interfaces NLO calculations with parton shower generators and produces parton level events with positive (constant) weight at NLO accuracy. This is most important in the case of t-channel and Wt-channel associated production, since leading order parton shower MC does not model well the momentum distribution (expecially in the low-p_T portion)of the $\bar{b}$ quark originated by the initial state gluon splitting in $b\bar{b}$ pairs (as in Fig. 6.1(a)). POWHEG avoids the complicated procedure of LO and NLO t-channel matching (see Ref. [2]). Moreover, POWHEG can produce Wt-channel signal sample, as it takes into account the interference between the Wt diagram and the $t\bar{t}$ production processes.

Background Events

The $t\bar{t}$ pairs production and diboson (WW,WZ,ZZ) production are simulated with PYTHIA [59], a general-purpose event generator. PYTHIA uses the parton distribution functions (PDF) provided by CTEQ6.6 version [60]. Since PYTHIA is also a showering and hadronization program, for such processes the entire MonteCarlo simulation is provided by the same program.

The W +jets and Z+jets background events are simulated using ALPGEN [61].It is an event generator specialized in electroweak bosons (W and Z bosons) produced in association with a desired number of jets coming from either quarks or gluons.

5.1.2 Parton Showering and Hadronization

After generation, all MonteCarlo events pass through PYTHIA package that models parton showering and hadronization. The **parton shower** simulates the higher-order effects of initial and final state gluon radiation and the splitting of those gluons into quark pairs, increasing the number of jet that can be detected by CDF. All the resultant partons are then grouped together and passed to the **hadronization** stage where they are hadronized into color-singlet hadrons and decay resonances. Since the hadronization of quarks and gluons, which describes the formation of jets,

takes place at low Q^2 and large α_s perturbation theory cannot be applied. The final Monte Carlo samples include contributions from initial-state sea quarks via the proton PDFs and the beam remnants (mostly forward jets).

The event generators corresponding to the differents processes used in the current analysis are summarized in Tab. 5.1.

Process	Event Generator	Parton Showering
Single Top	POWHEG	PYTHIA
$t\bar{t}$	PYTHIA	PYTHIA
diboson	PYTHIA	PYTHIA
W+jets	ALPGEN	PYTHIA
Z+jets	ALPGEN	PYTHIA

Table 5.1: Monte Carlo event generators and parton showering software programs used for Monte Carlo simulated events for signal and background processes used in this analysis.

5.1.3 Detector Simulation

After the hadronization, a list of "stable" particles is passed to a simulation of the CDF II detector system. They are propagated through the detector (simulated using the GEANT3 [62] package). The output is a list of hits and signals in the detector which is identical to the one obtained in real data.

After this, simulated events go through the same reconstruction and selection as real data. A detailed description of the CDF simulation can be seen in Ref. [63].

5.2 Monte Carlo Samples

5.2.1 Single Top Signal Production

In this analysis the signal process is the electroweak Single Top quark production. It can result in the final state of three different production modes: s-channel. t-channel and Wt-channel. The final state signature considered in this analysis is made of a charged lepton plus two or three jets and $\not\!\!E_T$. Due to the kinematic of the various production modes, the relevant processes are s-channel at leading order, t- and Wt-channel at NLO (Fig. 5.1). The Monte Carlo signal samples are generated with POWHEG generator described above.

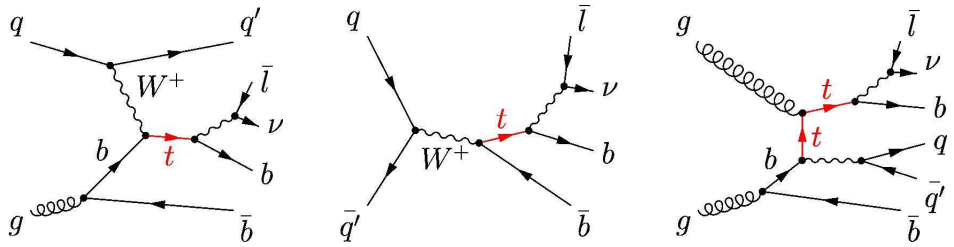


Figure 5.1: Feynman diagrams for NLO t-channel (a), LO s-channel (b) and NLO Wt-channel (c) single-top production with successive $t \to Wb$ and $W \to l^+\nu_l$ leptonic decays.

5.2.2 Top-Quark Pair Production

$t\bar{t}$ events are classified by the decay mode of the W bosons coming from the t and $\bar{t}$ decays (BR($t \to W\bar{b} \approx 100\%$)). If both W-bosons decay leptonically the event is called *dilepton*(Fig.5.2(a)), while if on W decays leptonically and the other W decays in jets the event is dubbed as lep+jets (Fig.5.2(a)). Events with both W-bosons decaying hadronically are named *all-hadronic* and are not included in our selection. For Single Top analysis events from top-pair production are expected pass the data selection for various reasons: either one lepton is lost in dilepton events or several jets are lost or misidentified in lep+jets case.

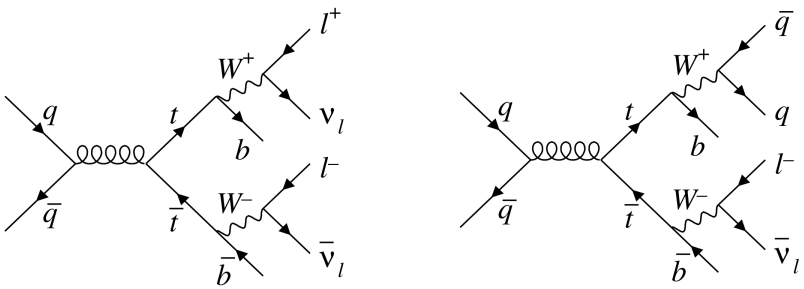
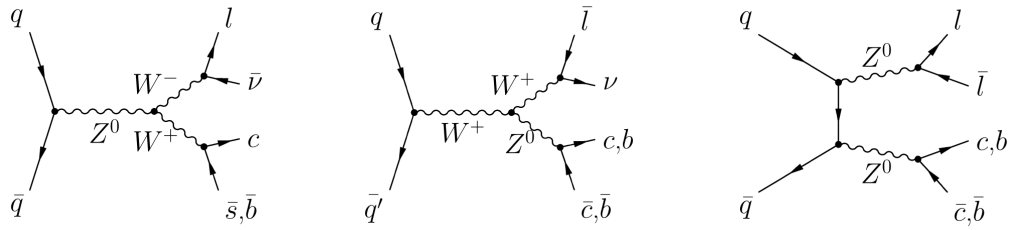


Figure 5.2: LO Feynman diagrams for the top quark pair production and decay, reducible background processes for our SingleTop analysis.

5.2.3 Diboson Production

The diboson production modes WW and WZ contribute to selected data sample since their final states include a lepton a neutrino and heavy-flavor jets (see Fig. 5.3(a) and Fig. 5.3(b)). ZZ events mimic a lepton+jets signature only if one

of the two leptons is lost, which coincidently fakes a neutrino through the resulting missing transverse lepton energy (see Fig. 5.3(c)). It is remarkable that the non-resonant production of diboson processes predicted by the Standard Model were observed at the Tevatron by CDF in various dacay modes [64] [65] [66].



(a) (b) (c)

Figure 5.3: LO Feynman diagrams for the WW (a), WZ (b) or ZZ (c).

5.2.4 Z+jets Production

The Z+jets background principaly comes either from the associated production of a Z boson, which decays in leptons, and a gluon, which decays in a $b\bar{b}$ or $c\bar{c}$ quarks pair, or from $Z \to \tau\bar{\tau}$ process, which has light flavor quarks in final state (also called $Z + lf$). In such events, the mismeasured jet energy produce a fake $\not\!\!E_T$. The $Z \to b\bar{b}$, $Z \to c\bar{c}$ and $Z + lf$ samples are not generated togheter, but in different MonteCarlo samples with different theoretical cross section for each process. Since the samples are then combined togheter, we weight the cross sections as follows:

$$\frac{N^{\text{sel}}_{\text{Z+jets}}}{N^{\text{tot}}_{\text{Z+jets}}} = \sigma^{-1}_{\text{total}} \sum_j \sigma_j \frac{N^{\text{sel}}_j}{N^{\text{tot}}_j} \tag{5.1}$$

where $\sigma_{\text{total}} = \sum \sigma_j$ and $j == b\bar{b}, c\bar{c}, lf$. Eq. 5.1 is then inserted into Eq. 7.1 to calculate the number of Z + jets events.

5.2.5 W+Jets

The W+jets process is by far the most important source of background for our analysis, as we will show later (Cap. 7). There are several W +jets processes. Just as in the case of the signal, $W + b\bar{b}$ and $W + c\bar{c}$ (Fig. 5.5(a)) present a real W bosons and two jets. Wc events feature the lepton+jets characteristics by additional jet production through higher order effects which is modeled by showering procedures, resulting as the $W + cj$ sample (Fig. 5.5(b)).

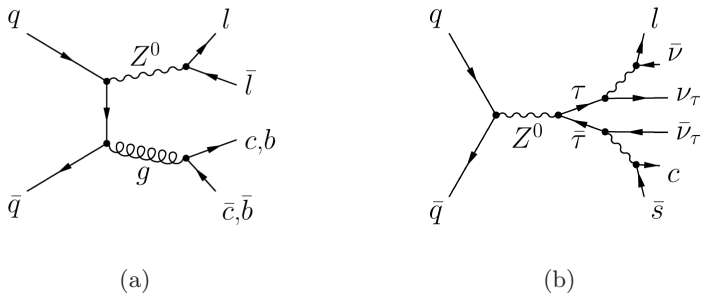


Figure 5.4: LO Feynman diagrams for the Z+jets in heavy flavor (a) or light flavor (b).

The jets may also originate from light flavour partons (up, down, strange quarks and any type of gluons). As in signal events we require at least one b-jet identified by b-tagging algorithms. While in W+light events there are no b quarks in final states, sometimes jets are (mis-)tagged as well. The mistag events sample is generated by ALPGEN , showered by PYTHIA and named $W + lf$ (light flavor) (Fig. 5.5(c)).

As for Z+jets, W+jets are produced in different subsets (w+0, w+1, w+2, etc.) with respect to the number of partons considered. The subsets are then put togheter and weighted by the appropriate cross-sections.

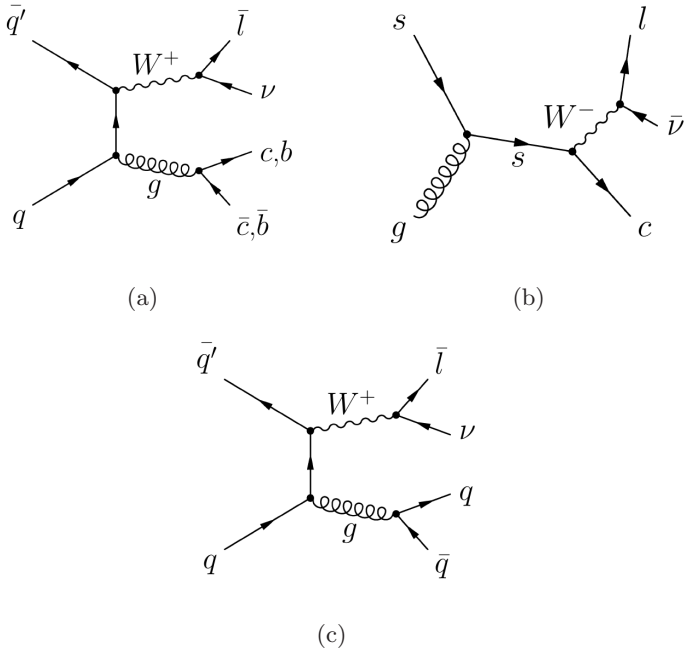


Figure 5.5: Feynman diagrams for several W + jets associated production and decay modes: W+HF (a), W+cj and W+lf.

Chapter 6

Event Selection and Data Reduction

6.1 Introduction

In our analysis we concentrate on events in which the W from top decays leptonically. As shown in Fig. 6.1(a) and Fig. 6.1(c), the s and t production modes for single top quark events show a final state topology containing two leptons (charged and neutral) coming from W boson decay and two or three jets. Therefore, our selection requires two or three energetic jets in the final state. The third jet can also originate from an additional gluon radiated in final state. Furthermore, we require one of those jets to be *b-tagged*, i.e. to be identified as coming from an hadronized b-quark. Our requirement of the W boson to decay in a charged lepton (e or μ) and its neutrino (ν_e or ν_μ) reduces QCD background from multijet production.

Due to small τ-efficency, a small fraction of events in which W boson decay in τ leptons and the τ in turns decays leptonically, is included in this selection even if τ is not directly identified.

6.2 Data Samples and Trigger Requirements

Single top analysis selects W+jets events requiring the signature of a charged lepton, large missing transverse energy $\not\!\!E_T$ and two or three jets.

Tight leptons samples (CEM, PHX, CMUP, CMX triggers) are collected using appropriate high-p_T lepton triggers. $\not\!\!E_T$ triggers are used to collect loose charged leptons (or EMC i.e. Extended Muon Category). $\not\!\!E_T$trigger requires large $\not\!\!E_T$ plus at least two additional high-p_T jets.

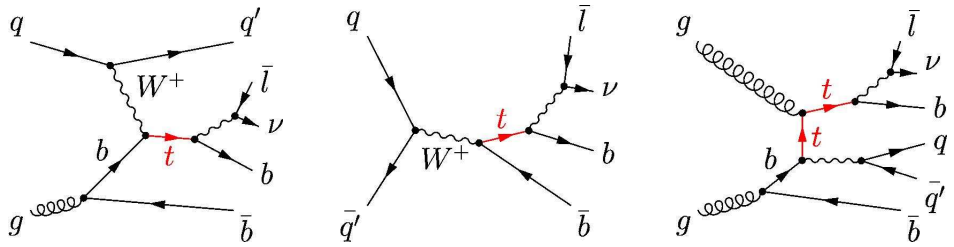


Figure 6.1: Feynman diagrams for NLO t-channel (a), LO s-channel (b) and NLO Wt-channel (c) single-top production with successive $t \to Wb$ and $W \to l^+\nu_l$ leptonic decays.

6.2.1 Central Electrons and Muons Triggers

Tight central electrons (CEM) are collected by the ELECTRON_CENTRAL_18 trigger. It requires a COT track with $p_T > 9$ GeV/c matched to an energy cluster in the CEM calorimeter with $E_T > 18$GeV. The ratio between the energy deposited ithe hadronic calorimeter to that in the EM calorimeter must be $E_{\rm HAD}/E_{\rm EM} < 0.125$. The shower profile of this cluster as measured by shower-maximum detector (see 3.2.2) is required to be consistent with expectations obtained using test-beam electrons.

Tight central muons (CMUP and CMX), identified in the muon detectors, are required to pass the high p_T MUON_CMUP18 or MUON_CMX18 trigger respectively. For CMUP candidates a track in the COT with $p_T > 18$ GeV/c has to match track segments in both central muon chambers CMU and CMP.

Forward muons candidates CMX ($0.65 < |\eta| < 1.0$) must have a COT track segment matched to hits in the CMX muon chambers.

6.2.2 Forward Electrons

Forward electron candidates ($1.2 < |\eta| < 2.0$), also named plug electrons (see 4.1.1), are collected by the MET_PEM trigger which requires a cluster with an energy deposition of at least 20 GeV in the PEM and $E_{HAD}/E_{EM} < 0.0125$, while no tracking is available for this class of electron.

6.2.3 EMC muons and ISOTRK

Extended Muon Coverage is an inclusive category of muon candidates used to increase acceptance for events that do not fire a μ trigger. Therefore, EMC comprises muons collected by the $\not\!\!E_T$+jets trigger path. They are ordered in six mutually

exclusive sub-categories, namely CMU,CMP, BMU,CMIO,SCMIO,CMXNT. Those sub-categories are then collected in a preliminary *loose* muons category.

Single Top observation analysis [2] used just one trigger (MET2J) for EMC events. In this analysis update we added a new category (ISOTRK) of candidate muons to the EMC.

We used a novel trigger combination among MET2J, MET45 and METDI triggers [67] for the EMC events. In the following we describe the triggers used to collect EMC and ISOTRK.

$\not\!\!E_T$-based triggers

MET2J is a combination of different versions of a trigger that cover in sequence all the periods of CDF RunII evolving with the increasing instantanous luminosity. MET2J is fired if there is a large MET ($\not\!\!E_T > 35$ GeV) and two energetic jets. Over time, more requirements were added. Also, the trigger is prescaled above a certain value of the instantaneous luminosity. The jet selection requires two jets with $E_T > 25$ GeV, $\Delta R > 1.0$ [1] between two jets and one central jet ($|\eta_{det}(jet)| < 0.9$). Latter requirement was added only after period 14.

MET45 trigger comes in two different version (MET45/MET40) according to different running periods due to high intantaneous luminosity. It requires either 40 or 45 GeVof $\not\!\!E_T$, higher than MET2J (as suggested by the name).

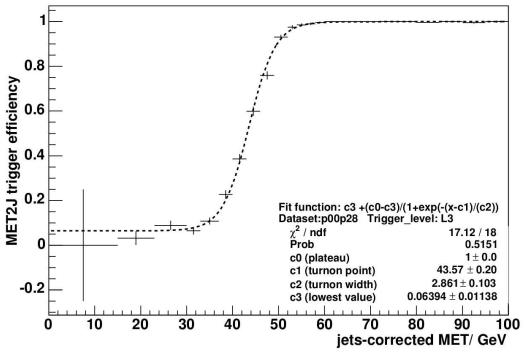
METDI trigger is very similar to MET2J. It requires exactly two jets, which must have energy greater than 40 GeV and 25 GeV respectively[2].

In order to parametrize the efficency of these triggers (MET2J, MET45, METDI) we use sigmoid turn-on curves as function of only one variable common for all of them, that is the missing transverse energy $\not\!\!E_T$(Fig. 6.2). After correcting for the z position of the interaction vertex, if we do not correct for the muon energy we can take it as a rough estimate of the transverse energy of the W boson[3]. When in one event more than one trigger is fired, we compute, for each trigger fired, the combined probability (called "trigger weight") to pass the specific trigger (L1,L2,L3) taking also into account the prescale factor of each level. We define that the event belongs to the trigger with the largest "trigger weight".

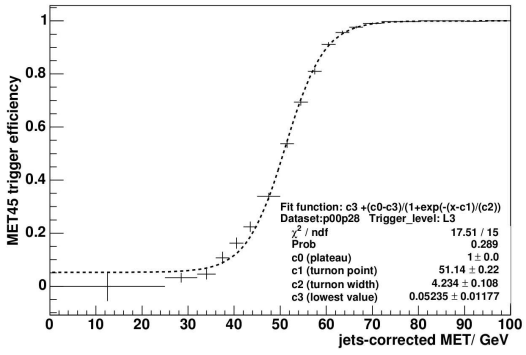
[1] The definition of ΔR quantity is $\Delta R \equiv \sqrt{(\Delta\eta_{det})^2 + (\Delta\varphi)^2}$

[2] Jets are ordered in decreasing E_T.

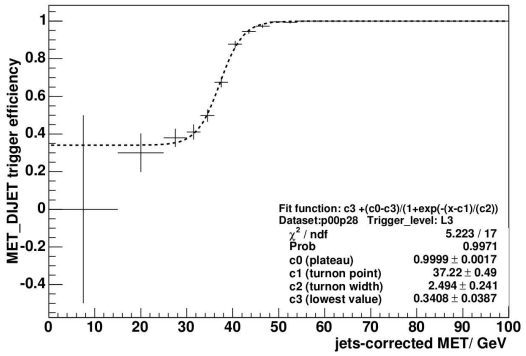
[3] Otherwise it would be the momentum of the neutrino from W boson decay.



(a)



(b)



(c)

Figure 6.2: MET2J (a), MET45 (b), METDI (c) triggers turn-on curves, parametrized as function of MET, calcucated at third level of the trigger which is the closest to the offline analysis. Figure taken from [67].

Data Streams

Data used for this analysis are taken from high-p_T central eletron data stream (*bhel*), high-p_T central muon data stream (*bhmu*), high-p_T plug electrons (*bpel*)

and high-$\not\!\!E_T$ data stream (*emet*), collected up to March 2011, for a total integrated luminosity of 7.5 fb^{-1}. From these datasets events are selected for the subsequent analysis if containing an energetic charged lepton with $E_T(p_T) > 20$ GeV(GeV/c).

6.3 Signal Selection

For each lepton category, selected events must be consistent with a topology of W-boson semileptonic decay and two energetic jets.

W-boson events are required to have a single charged lepton with $p_T > 20$GeV/c and $|\eta| < 1.6$. As previuosly described, a lepton candidate can be categorized as tight or loose: the main difference is that tight leptons are required to be isolated, meaning that the E_T not assigned to the lepton inside a cone defined by $R < 0.4$ and centered around the lepton is less than 10% of the lepton E_T (4.2), while loose leptons do not have this requirement.

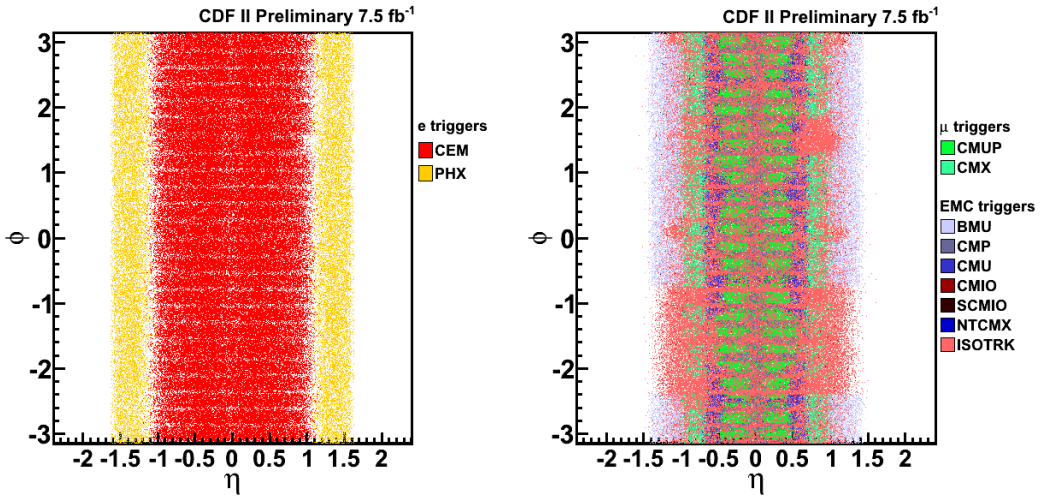
The presence of ν is estimated from the MET. We require $\not\!\!E_T$ greater than 25 GeV, corrected for the presence of the muons and for the energy of the jets. Different p_T thresholds (above 20 or 25 GeV) are applied for the different lepton types.

6.3.1 Lepton Identification

In order to clean our sample from multijet contamination, a number of offline requirements are applied to lepton candidates. As for the trigger case, we distinguish between tight electrons (CEM,PHX), high quality trigger muons (CMUP,CMX), loose muons and isolated traks. Last two are included in EMC leptons as they are collected with the same $\not\!\!E_T$ triggers. The aim is to gain as much acceptance as possible. This is of the outmost importance for muons, as we can recover in this way the losses in the cracks of the μ coverage. In Fig. 6.3 is shown the overall geometrical region covered by different lepton types, respectively for electrons (Fig. 6.3(a)) and muons (Fig. 6.3(b)).

Tight Leptons

CEM electrons have a Central ElectoMagnetic isolated object with $|\eta| < 1.1$ and with energy deposition in the CEM > 20. GeV. The ratio of energy left in hadronic calorimeter and in em section of the calorimeter is requested to be $E_{HAD}/E_{EM} \leq 0.055 + (0.00045 \cdot E)$ GeV. The ratio between the energy of the cluster and the momentum of the associated track in the COT has to be $E/p < 2.0$ for track momenta ≤ 50 GeV/c.



(a) (b)

Figure 6.3: The leptonic acceptance distribution in (η, φ) plane, showing the coverage of the detector provided by each lepton category. It can be seen how much the new ISOTRK type helps in covering the fiducial region of the detector

PHX forward electrons are detected by PEM calorimeter if a cluster with $E_T >$ 20 GeV is identified. The selection take in accout only events with $1.1 < |\eta| < 1.6$ and $E_{HAD}/E_{EM} < 0.05$. Cluster and primary vertex positions are combined to form a trajectory in the silicon tracker. If at least 3 silicon hits are found by the pattern recognition, a PHX electron is identified.

If an additional high-p_T track is found forming a common vertex with track of the electron but with opposite sign of curvature, the event is rejected as probably coming from photon-conversion.

CMUP and CMX muon selection requires a COT track segment with $p_T > 20$ GeV matched to a stub in the muon detectors, in the region of $|\eta| \leq 0.6$ and $0.6 \leq |\eta| \leq 1.0$ respectively.

EMC muons

EMC muon candidates come from the $\not\!\!E_T$ triggers described in Sec. 6.2.3), divided in six different categories (as in 6.2.3) that represent different selection criteria. CMU and CMP muons have tracks in their respective muon chamber and in the COT but not overlapped one with each other, while BMU have track segment in barrel muon chambers, geometrically forward ($1.0 \leq |\eta| \leq 1.5$). CMIO and SCMIO come from tracks matched to calorimetric deposits identified as minimum ioniza-

tion particles but not fiducial to any muon segment -so that they do not point towards a muon detector- (CMIO) or matched to a muon segment not fiducial to any muon detector - so that the track points to a non-fiducial part of any muon detector-(SCMIO). CMXNT represents tracks segment in CMX but in the most forward part of that detector which is not usable for triggers as it extends beyond COT region of full efficency.

ISOTRK charged leptons

In order to gain acceptance and recover events that are not included in any of the previuos categories, a new category was introducted named ISOTRK leptons. If the event is reconstructed as containing a high quality, high-P_T isolated track, with $|\eta| < 1.2$, it can be accepted as an "`isotrk`" lepton. First of all we estabilish if the track has nearby activity using the track isolation variable defined as:

$$\text{TrkIsol} = \frac{p_T(\text{candidate})}{p_T(\text{candidate}) + \sum p_T(\text{trk})} \tag{6.1}$$

where $\sum p_T(trk)$ is the sum of the p_T tracks in a cone of radius $\Delta R = 0.4$ around the candidate passing the following requirements:

- $p_T > 0.5$ GeV/c;
- ΔZ(trk,candidate) < 5 cm;
- Number of COT axial hits ≥ 24;
- Number of COT stereo hits ≥ 20;

so that with this definition a track with no surrounding activity has an isolation of 1.0. The final candidate must pass these following requirements

- $p_T > 20$ GeV/c;
- $|z_0| < 60$ cm;
- $|d_0|_{\text{corr}} < 0.2$ cm;
- $|d_0|_{\text{corr}}$(w/si) < 0.02 cm[4];
- TrkIsol> 0.9, that is 90% of the momentum in the cone must come from the candidate track p_T;

[4](w/SI) means with an hit in the silicon tracker, so that a larger impact parameter d_0 would lead to have a jet from some long-lived particles as K_S^0 or K_L^0 decay instead of a letpon.

- number of silicon hits ≥ 3.

Additional vetoes are applied to ensure that ISOTRK events originate from W events and they do not overlap with other lepton types. Furthermore, we apply requirements aimed to enrich the sample in W initiated events ensuring that ISOTRK category remains ortoghonal to any other lepton type used in the analysis. Below those vetoes and requirements are listed:

- ***Tight Jet Veto*** requires an angular separation of $\Delta R > 0.4$ from any tight jet in the event.
- ***Two Track Veto*** counts if there are two or more isolated tracks in the event before applying the tight jet veto, and in case the event is vetoed.
- ***Tight Lepton Veto*** check if there is any tight lepton in the event, so that the candidate should not be processed as ISOTRK. Finally, $\Delta\varphi$(trk,Jet1) must be greater than 0.05.

Loose muons and the ISOTRK are collected in the same EMC category and, as for tight leptons, we require a corrected $\not\!\!E_T > 25$ GeV.[5].

Fig. 6.4 shows the increased number of accepted events by including ISOTRK in the EMC category as a function of run number. The number of total selected events in the EMC category increases by about 20% (as shown in Fig. 6.4).

The number of selected events in EMC signal region (2J1T, 3J1T, 2J2T, 3J2T) is increased by 15%. This increase can be seen in Fig. 6.3 obtained after the full event selection before applying the SecVtx tag. In this scatter plot we can see how ISOTRK charged lepton candidates fill the gaps in the calorimentes. This increases the number of selected events containing charged leptons and thus the signal event prediction. Unlike the case of muon candidates, the distribution for electron candidates (Fig. 6.3(a)) is smoother. This happens as ISOTRK candidates are mostly muon candidates and increase the EMC selected events, as in Fig. 6.4. A detailed study [68] identified that ISOTRK candidates are muon candidates in 85% of cases, electron candidates in 6% of cases and tau candidates in 7% of cases.

6.3.2 Jet Identification

Hadronic jets are reconstructed by a fixed cone (JetClu) algorithm summing the transverse calorimeter energy in a cone of radius $R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} < 0.4$ in

[5]The usual correction to the MET due to muon presence has been modified to include ISOTRK events: ISOTRK category definition avoids to mismatch a tight muon but it is possible to match a BMU, so in such a case $\not\!\!E_T$ is not corrected for ISOTRK since it is already done for BMU.

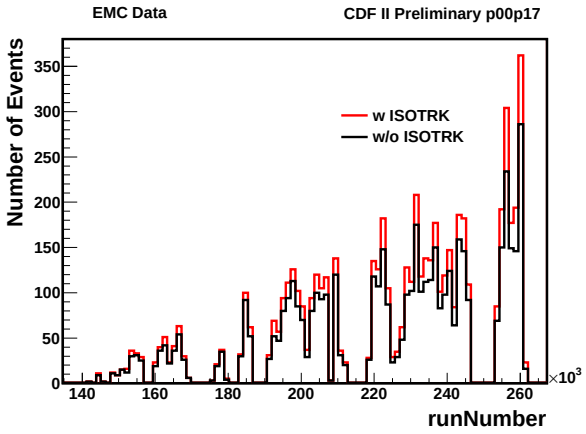
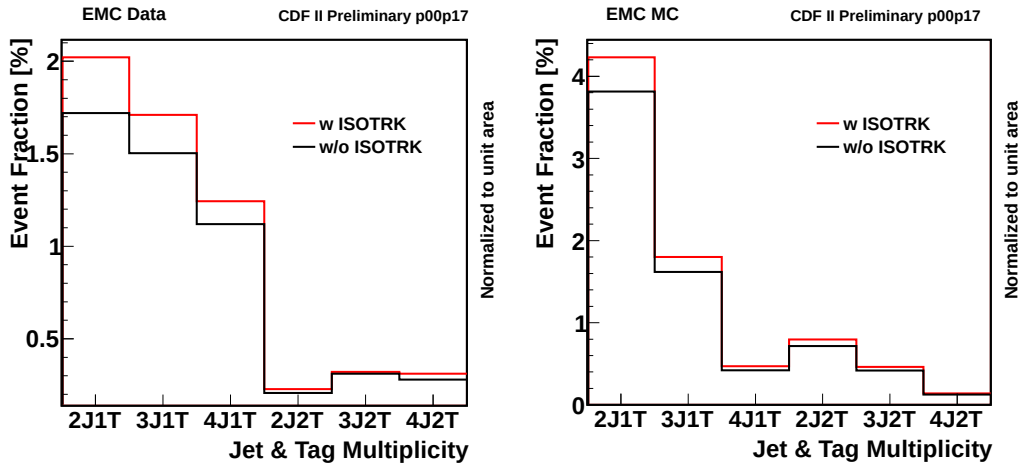


Figure 6.4: Increase in acceptance of EMC events due to ISOTRK inclusion. The study is performed on the first 2.6 fb^{-1} of integrated luminosity sample corresponding to data collected up to period 17. Afterwards the events were reprocessed and the trigger changed. Those data were also used in the previuos analysis and belongs to the common dataset among the 3.2 and 7.5 fb^{-1}searches.



(a) (b)

Figure 6.5: Jet and b-tag multiplicity of EMC events with and without ISOTRK. The study is performed as in fig. 6.4.

the (η, φ) space and removing -if present- the energy deposition due to the lepton candidate. Unlike CDF standard lepton+jets selection (used also in WH search, which shares most of the selection tools with our Single Top analysis), we require jets to have $|\eta_{det}| < 2.8$, while the usual requirement is $|\eta_{det}| < 2.0$. This difference originates from the kinematic features of single top signal events, which are likely to contain forward jets.

Jet candidates are required to have $E_T > 20$ GeV corrected up to Level-5 jet

energy correction[6]. Most of the signal events have two or three jets, so that we apply this requirement in our selection.

In order to reduce the background from light flavor events, at least one jet is required to be b-tagged by the secondary vertex alghoritm (SecVtx) because, as in fig. 6.1, at least one of the 2 b jets is present in signal events.

6.3.3 Missing Transverse Energy

The neutrino coming from W-boson decay is identified from the missing-trasverse-energy ($\not\!\!E_T$) present in the event.

$$\vec{\not\!\!E_T} \equiv -\sum_i \vec{E}_T^i \cdot \hat{n}_i \tag{6.2}$$

where i runs over the calorimeter tower number with $|\eta| < 3.6$. In the following, $\not\!\!E_T = |\vec{\not\!\!E_T}|$.

We use the level 5 corrected $\not\!\!E_T$ as defined in Eq. 4.8. We require $\not\!\!E_T >$ 25 GeV for all lepton categories, after correcting for the presence of muons (including ISOTRK) and jet energies.

6.4 Background and Event Vetoes

Several physics processes predicted in $p\bar{p}$ collisions can be mistaken as ST production. Some of them have exactly the same final state as single top, while some others are likely to mimic the single top signature because of misreconstruction or loss of components of the expected final state. All of the events coming from such processes are called *background* events. Sometimes the first type of background (processes with same signature as signal) is called "irreducible" and the second (different signature) is called "reducible", depending from the fact that for the first type we can search for some improvements in the selection techniques while for the latter this is not possible.

We consider as background sources the following processes: top-quark pair production, diboson production (WW,WZ,ZZ), W+jets and Z+jets production, QCD multijet events.

In order to reduce known background presence, some vetos are included in the event selection . A special treatment must be reserved to QCD background because of the absence of an exact modeling and simulation.

[6]defined in 4.4.2

6.4.1 Event Vetoes

Dilepton veto requires that an event must contain exactly one lepton and those which contain more leptons are rejected. This veto rejects events in which $t\bar{t}$ decays into leptons as well as diboson events.

Z boson veto is implemented to avoid events containing a Z boson. We check that lepton, paired with another high-p_T track, does not form an invariant mass $M_{\text{l,track}}$ consistent with $76 \leq M_Z \leq 106$ GeV/c^2.

z Vertex cut requires the reconstructed primary z vertex of event to be within a range of ± 60 cm with respect to the center of the detector.

cosmic veto remove event tagged as cosmic rays taking advantage of their characteristic timing and topology, going from outer space to the inside of the detector. This veto is more important for muons, since highly energetic muons can be created in the upper atmosphere and then penetrate the detector. It is not applied to MC events as we do not simulate cosmic rays.

6.4.2 Single Top QCD Veto Studies

An important fraction of the background for W+jets final state processes comes from the QCD-multijet events which do not contain a W boson (so that they are also called "non-W" events), but fake a W-boson-like signature with one jet mistaken as a lepton and $\not\!\!E_T$ generated from jet energy mis-measurement. In such a case, the $\vec{\not\!\!E}_T$ often points to the lepton candidate direction.

In general we expect this background to be more important for electron candidates than for muons as it is easier for a generic jet to mimic an electron than an high-p_T muon. The so-called "Single-Top QCD Veto" (STQCDVeto) is a procedure designed to reject multijets events. It relies on the study of the W boson transverse mass M_T^W, which is reconstructed from the final state kinematics, and the special variable MET-significance.

$\mathrm{M_T^W}$ requirements

Defining the trasverse mass (for W boson in this case) as:

$$M_T^W = \sqrt{2[p_T^{\text{lep}} \not\!\!E_T - p_x^{\text{lep}} \not\!\!E_{T,x} - p_y^{\text{lep}} \not\!\!E_{T,y}]} \simeq \sqrt{2p_T^{\text{lep}} \not\!\!E_T [1 - cos\Delta(p_T^{\text{lep}}, \not\!\!E_T)]} \quad (6.3)$$

we expect multijet events to accumulate at low M_T^W region, because of their $\vec{\not\!\!E}_T$ direction. After studying the different leptons, we require for electron candidates

$M_T^W > 20$ Gev while for tight muons $M_T^W > 10$ GeV. In the case of EMC $M_T^W > 20$ GeV is required because of larger QCD contamination (shown in fig. 6.7). In this case the current analysis differs from the previuos one which used a cut at 10 GeV (but for SCMIO leptons a $M_T^W > 20$ GeV cut was applied) without including ISOTRK events. Fig. 6.6 shows the effect of this requirement on EMC muons.

$\not\!\!E_{T,\mathrm{sig}}$ requirements

Another tool to further reduce the non-W events stats by defining a variable called MET-significance (or $\not\!\!E_{T,\mathrm{sig}}$):

$$\not\!\!E_{T,\mathrm{sig}} = \frac{\not\!\!E_T}{\sqrt{\sum_{\mathrm{jets}} C_{\mathrm{jets}}^2 cos^2(\Delta\varphi(\mathrm{jet}, \vec{\not\!\!E}_T)) E_{T,\mathrm{jet}}^{\mathrm{raw}} + cos^2(\Delta\varphi(\not\!\!E_T^{\mathrm{raw}}, \not\!\!E_T^{\mathrm{corr}}))(\not\!\!E_T - \sum_{\mathrm{jets}} E_T)}} \tag{6.4}$$

where C_{jes} is the jet energy correction factor, E_T^{raw} and E_T^{corr} are the uncorrected and corrected missing transverse energy. The physical meaning of $\not\!\!E_{T,\mathrm{sig}}$ variable is the dispersion of measured $\not\!\!E_T$, approximated by the quantity at the denominator, in events without true $\not\!\!E_T$. Some lepton categories are required to pass requirements on $\not\!\!E_{T,\mathrm{sig}}$.

- CEM events are required to pass the triangular cut

$$\not\!\!E_{T,\mathrm{sig}} > 3.5 - 0.05 \cdot M_T^W, \tag{6.5}$$

  which is motivated by Fig. 6.7(c). In addition, CEM events must fullfill (as shown in Fig. 6.8(c) and Fig. 6.8(f))

$$\not\!\!E_{T,\mathrm{sig}} > \begin{cases} -7.6 + 3.2 \cdot |\Delta\varphi(l, \mathrm{jet})| & \text{1-jet bin} \\ 2.5 - 3.125 \cdot |\Delta\varphi(\not\!\!E_T, \mathrm{jet}_2)| & \text{2-jets, 3-jets bin} \end{cases} \tag{6.6}$$

  where jet_2 is the second most energetic jet since they are ordered in decreasing energy.

- PHX candidate events must satisfy:

$$\not\!\!E_{T,\mathrm{sig}} > 2; \quad \not\!\!E_T > 45 - 30 \cdot |\Delta\varphi(\not\!\!E_T, \mathrm{jet})| \tag{6.7}$$

  the last requirement has to be passed by all jets in event.

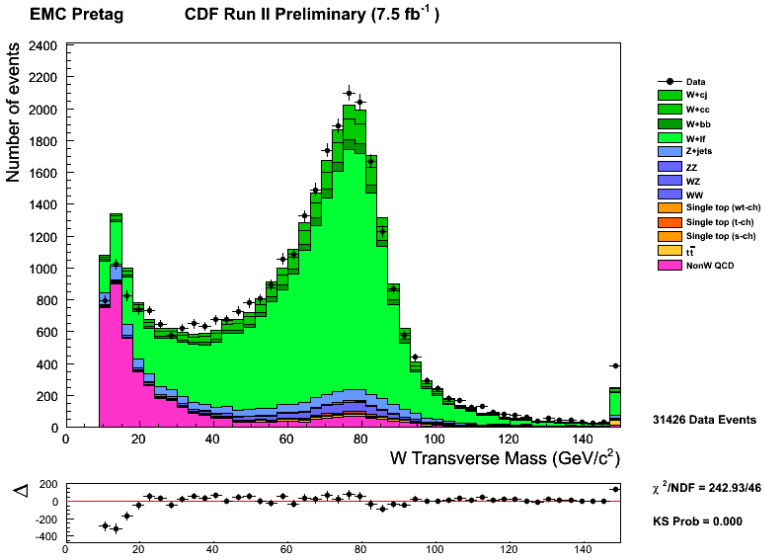
- CMUP muons are required to have:

$$\not\!\!E_T > -145 + 60 \cdot |\Delta\varphi(l, \mathrm{jet})| \tag{6.8}$$

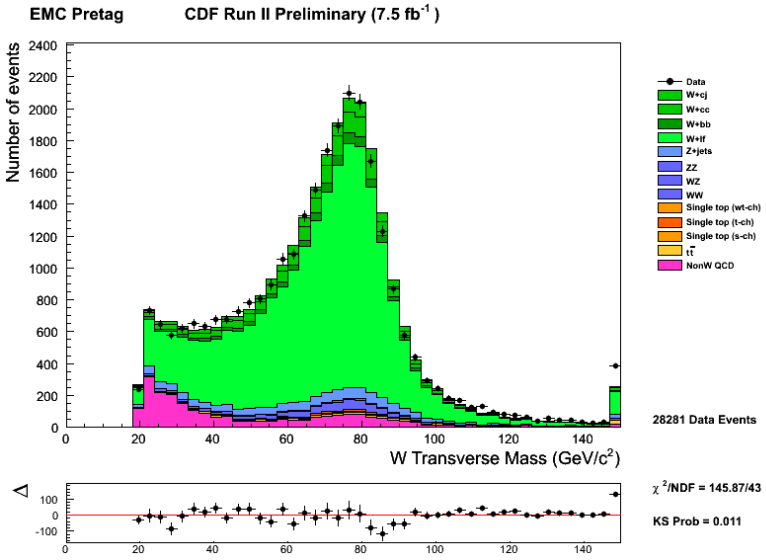
  for the single jet in the 1-jet bin.

There are no requirements on $\not\!\!E_T(\not\!\!E_{T.sig})$ for CMX and EMC events.

All these cuts on M_T^W and $\not\!\!E_{T,\text{sig}}$ were determined and then applied after studies on QCD contamination on the different TLC and EMC lepton categories, based on the difference between data and simulated events as shown in plots (6.7(c), 6.8(c),6.8(f)) for CEM, (6.7(f),6.9(f)) for CMUP and 6.7(i) for EMC.



(a)



(b)

Figure 6.6: Difference between background calculation before (a) and after (b) the cut on $M_T^W > 20$ GeV for EMC leptons.

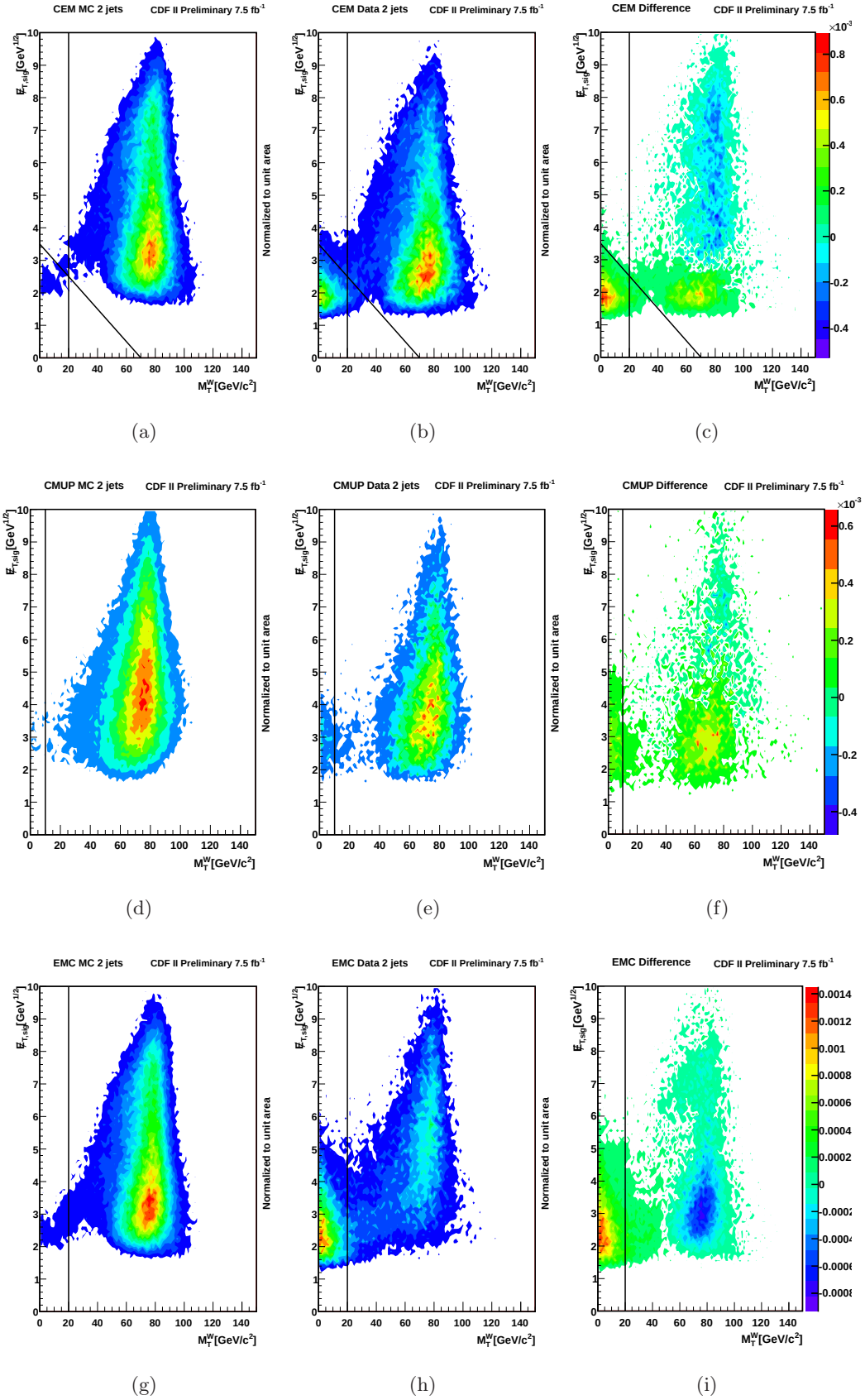


Figure 6.7: Scatter plot representing Single Top QCD Veto variables $\not\!\!E_{T,sign}$ vs M_T^W for CEM (a,b,c), CMUP (d,e,f) and EMC (g,h,i) pretag events in their 2 jets bin. In (a,d,g) simulated events from a W+jets MonteCarlo sample are shown. In (b,e,h) there are events from selected data. In (c,f,i) the MonteCarlo and Data distributions are subtracted showing the presence of not modelized QCD background in the low W mass region which is rejected by the cuts represented by the lines.

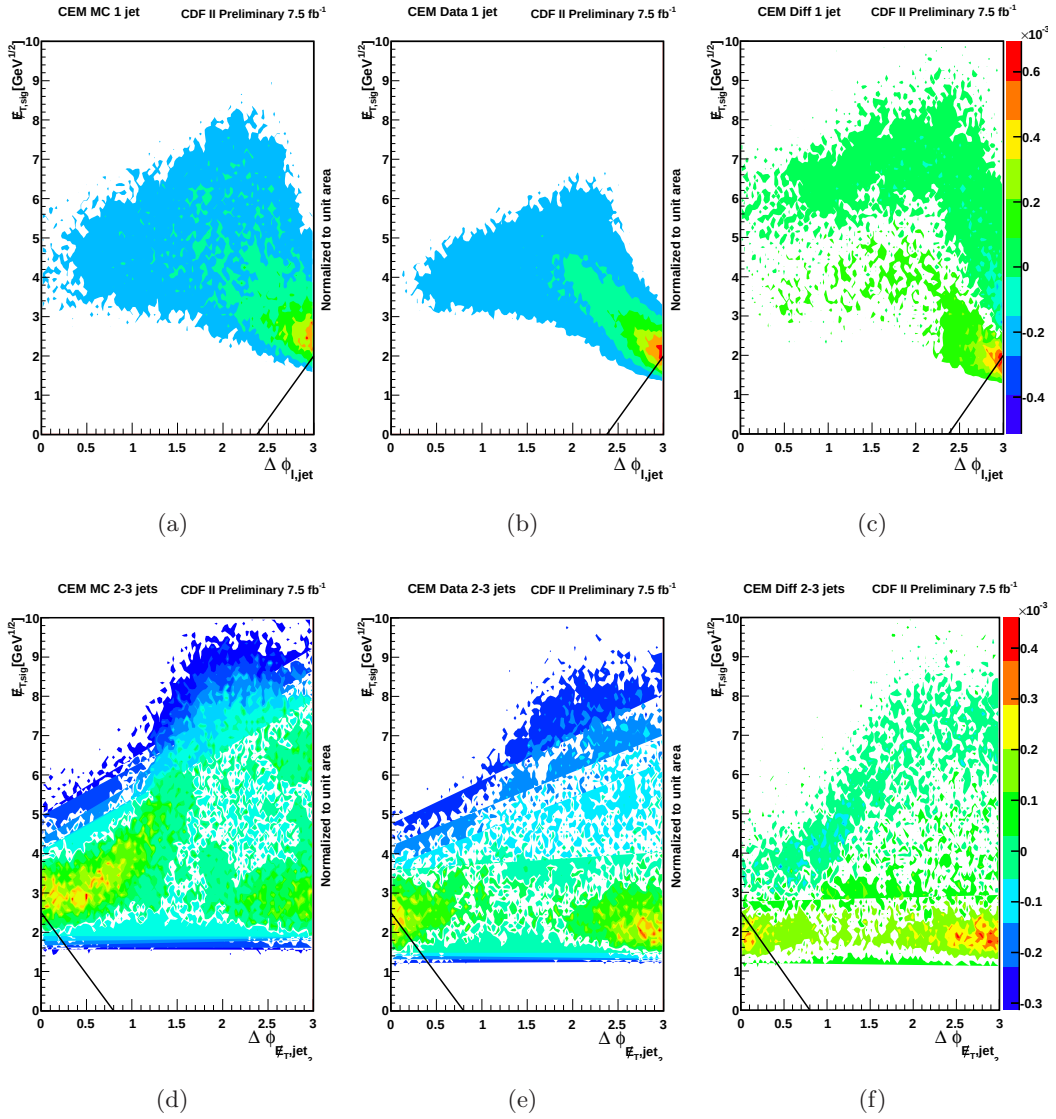
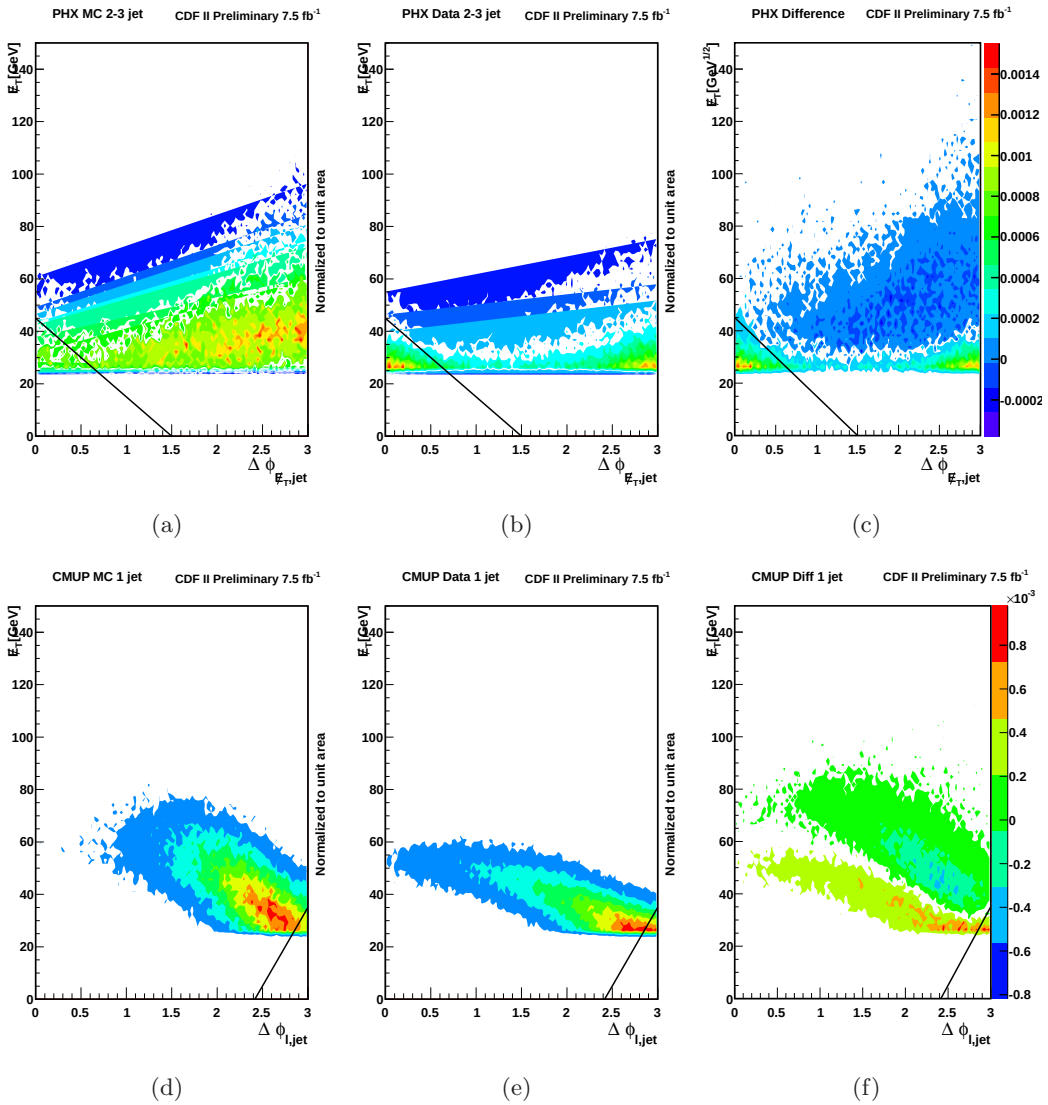


Figure 6.8: Scatter plot representing Single Top QCD Veto variables $\not\!\!E_{T,sign}$ vs $|\Delta\varphi(l,\text{jet})|$ in 1-jet bin (a,b,c) and vs $|\Delta\varphi(\not\!\!E_T,\text{jet}_2)|$ in 2-3 jets bin (d,e,f) for CEM pretag events. In (a,d) simulated events from a W+jets MonteCarlo sample are shown. In (b,e) there are events from selected data. In (c,f) the MonteCarlo and Data distributions are subtracted showing the presence of not modelized QCD background in the low W mass region which is rejected by the cuts represented by the lines.



(a) (b) (c)

(d) (e) (f)

Figure 6.9: Scatter plot representing Single Top QCD Veto variables $\not\!\!E_T$ vs $|\Delta\varphi(\not\!\!E_T, \text{jet})|$ in 2-3 jets bin for PHX pretag events (a,b,c) and vs $|\Delta\varphi(l, \text{jet})|$ for CMUP pretag events 1-jet bin (d,e,f). In (a,d) simulated events from a W+jets MonteCarlo sample are shown. In (b,e) there are events from selected data. In (c,f) the MonteCarlo and Data distributions are subtracted showing the presence of not modelized QCD background in the low W mass region which is rejected by the cuts represented by the lines.

Chapter 7

Signal and Background Yield Estimate

7.1 Introduction

In order to calculate the amount of signal in our data, we first need to understand all backgrounds. The samples, both simulated[1] and observed data, consist of events that pass the selection criteria described in Cap. 6. As already mentioned in Sec. 6.4, this lepton+jets signature is common for Single Top and several other physical processes produced by $p\bar{p}$ high energy collision. Those processes can have exactly the same Single Top final state, as for $Wb\bar{b}$ events, or similar final states, as for $Wc\bar{c}$ or $t\bar{t}$ events, which mimic the Single Top signature because of misreconstruction or loss of some components of their real final state.

In this chapter, we explain how we model the bckground processes and how we compute the normalization in the pretag and tagged lepton+jets samples.

7.2 Method II Background Estimation Tool

The procedure used in this analysis for the background estimation is commonly known at CDF as `MethodII` [69] and was successfully used in the $t\bar{t}$ cross section measurement [52] and the single top observation [2].

`MethodII` relies on the assumption that we know all of processes that contribute to form the lepton+jets sample.

In the Single Top search these are single-top quark in s-channel and t-channel, top quark pair, diboson production (WW,WZ,ZZ), W+jets and Z+jets production,

[1] see Sec. 5

QCD-multijet processes (also called "non-W"). The event yield for the latter process is obtained using both MC simulations and data. It is described in more details in Sec. 7.5.

`MethodII` procedure is performed separately for each channel in which we divide the samples. The channels are the two or three jets bins each one with one or two tags (2J1T, 2J2T, 3J1T, 3J2T), divided by the various lepton categories.

In the following we will describe each step and show the results obtained in our analysis.

7.3 Simulated Based Event Yield Estimate

For each physical process, the number of expected events is given by:

$$N_i = \varepsilon_{\text{evt}} \cdot \varepsilon_{\text{tag}} \cdot \sigma_i \cdot \int \mathcal{L} \cdot dt \tag{7.1}$$

EW processes	Cross Section σ pb	Branching Ratio BR
$t\bar{t}$	7.04 ± 0.44	1.0
Single Top (s-ch)	1.06 ± 0.06	0.324
Single Top (t-ch)	2.12 ± 0.22	0.324
Single Top (Wt-ch)	0.22 ± 0.08	1.0
WW	11.60 ± 0.7	1.0
WZ	3.46 ± 0.3	1.0
ZZ	1.51 ± 0.2	1.0
Z+jets	787.4 ± 85.0	1.0

Table 7.1: The theoretical cross sections and branching ratios used for MC-based calculation. Cross sections are calculated at NLO assuming a top-quark mass of 172.5 GeV/c^2.

Here, σ_i is the theoretically predicted cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. We use cross sections and branching ratios specific for each simulated process, as listed in Tab. 7.1. $\mathcal{L}$ is the instantaeous luminosity integrated over the run time corresponding to a value of $7.50 \pm 0.45\ fb^{-1}$.

ε_{evt} is the event detection efficency, which is given by several factors:

$$\varepsilon_{\text{evt}} = \varepsilon_{\text{evt}}^{\text{MC}} \cdot \varepsilon_{\text{BR}} \cdot \varepsilon_{z_0} \cdot \varepsilon_{\text{trigger}} \cdot \varepsilon_{\text{leptonID}} \tag{7.2}$$

$\varepsilon_{\text{evt}}^{\text{MC}}$ is the event detection efficency obtained for simulated events.

$\varepsilon_{\rm BR} = \sum_{l=e,\mu,\tau} BR(W \to l\nu)$ is the branching ratio of W bosons decaying into leptons.

$\varepsilon_{\rm z_0}$ is the efficiency of the cut $|{\rm z_0}| < 60$ cm.

$\varepsilon_{\rm trigger}$ is the efficiency of the requirement for the event to fire its respective trigger and it is measured in data for each trigger and lepton category.

$\varepsilon_{\rm leptonID} = \frac{\varepsilon^{\rm data}_{\rm lepID}}{\varepsilon^{\rm MC}_{\rm lepID}}$ is the ratio of the lepton identification efficiencies for data and simulated signal events[2]. Tab. 7.2 shows the efficency and scale factors used in the current analysis for the various lepton category.

$\varepsilon_{\rm tag}$ is the b-tagging scale factor. It takes in account differences between data and simulation efficency in b-quarks jet identification for the SecVtx tagger (see Sec. 4.7.2).

Lepton Type	$\varepsilon_{\rm z_0}$	$\varepsilon_{\rm trig}$	$\varepsilon_{\rm lepID}$
CEM	0.972 ± 0.001	0.961 ± 0.001	0.976 ± 0.001
PHX	0.972 ± 0.001	0.946 ± 0.003	0.919 ± 0.002
CMUP	0.972 ± 0.001	0.877 ± 0.002	0.892 ± 0.002
CMX	0.972 ± 0.001	0.902 ± 0.002	0.984 ± 0.002
EMC	0.972 ± 0.001	1.0 ± 0.0	0.942 ± 0.012

Table 7.2: The efficencies and scale factors used in the current analysis. These numbers are the official released from the Joint Physics Group at CDF collaboration for periods p00-p34 of run. The EMC trigger efficency is set to unity because they are not-trigggered events (see Sec. 6.2.3).

7.3.1 Signal: Single Top-Quark Production

For the signal samples, the event detecton efficency in the three different channels is: $\varepsilon_{\rm evt}({\rm t-ch}) = (1.9 \pm 0.3)\%$, $\varepsilon_{\rm evt}({\rm s-ch}) = (2.7 \pm 0.3)\%$ and $\varepsilon_{\rm evt}({\rm Wt-ch}) = (2.7 \pm 1.2)\%$. Using the theoretical cross sections listed in Tab. 7.1 [19] we find, through Eq. 7.1, that the expected total number of signal events are 297.4 ± 39.3 in the t-channel, 214.3 ± 21.1 in the s-channel and 44.9 ± 11.9 in the Wt-channel. The uncertainties in the values take in account the systematics uncertainties on theoretical cross section, luminosity, trigger efficency, lepton ID efficency and b-tagging correction scale factor.

For the Single-Top cross section measurement we will take in account several other source of systematic uncertainties, as those coming from JES correction and the initial and final state and gluon radiation.

[2] Also called *lepton identification scale factor*

7.4 Non-W (QCD) Background Modeling

The QCD background originates from a variety of QCD processes that produce a huge amount of multijet events lacking a leptonic W boson decay to be selected. For this reason, the QCD background is also dubbed as "non-W". For the events passing the selection requirements, there must be a fake lepton (a jet producing a track in the COT associated with an EM energy deposit) or a real lepton from a heavy flavor quark decay[3]. In addition to that, the $\not\!\!E_T$ must be mismeasured [70].

The QCD background is one of the more difficult to estimate as it is poorly predicted and it is the less understood. Besides, we cannot generate MC samples properly describing mismeasurements in the detector. In absence of trustable models we use a modeling method that is based on data: we create several non-W templates from fake leptons. Usually this means selecting data samples with less stringent requirements than for the signal case. The resulting samples are enriched of non-W events with similar kinematic distributions as the non-W contribution to the signal sample.

In the following sections, we describe the different non-W templates used for each lepton type are listed.

7.4.1 Improved Central Antielectron for CEM, CMUP, CMX fakes

All central tight fake leptons, either electrons or muons, are modeled from a template called "anti-electron", taking advantage of the fact that fake leptons from non-W events have difficulty passing the lepton identification requirements.

We require the electron trigger to be fired by all but two of the five eletron identification requirements that are listed below [4]:

1. $E_{\mathrm{HAD}}/E_{\mathrm{EM}} \leq 0.055 + (0.00045 \cdot E)$ GeV;
2. $\chi^2 < 10.0$;
3. $\mathrm{L_{shr}} < 0.2$;
4. $|\Delta Z| < 3.0$ cm;
5. $-3.0 < Q \cdot \Delta x < 1.5$;

[3] Each charged lepton category is susceptible to different kinds of faking due to the different reconstruction criteria.

[4] see Sec. 4.2.

The electron isolation requirement IsoRel < 0.1 (Eq. 4.2) is retained.

The same sample also simulates the kinematics of events with misidentified central triggered muon candidates (CMUP, CMX).

For this analysis we improved CEM anti-electron modeling by correcting the $\not\!\!E_T$ for a fraction of 5% of the L5 jet transverse energy correction associated with the jet misidentified as the lepton. The new improved QCD model produces a better description of the $\not\!\!E_T$ and its correlations with the other variables.

7.4.2 Antielectron for PHX fakes

The QCD template used for fake PHX events is built with a similar procedure, except for the trigger requirements that are the same for PHX events[5].

7.4.3 Non-Isolated Loose Muons for EMC fakes

The kinematic distributions of the reconstructed objects in the EMC sample are different from those in the TLC samples due to the trigger requirement.

A separate sample must be used to model the non-W background in the EMC data. It consists of events that are collected with the $\not\!\!E_T$+jets trigger path and that have a muon candidate passing all selection requirements except for the isolation requirement (Isolation> 0.2). It is dubbed the "non-isolated" sample.

The choice of this modeling is based on the physics of non-isolated leptons. They are typically leptons contained in jets, and jets that contain energetic leptons are more likely to pass lepton identification cuts. This sample has the advantage of a large statistics.

7.5 Normalization of QCD background

The normalization of the non-W prediction is separately determined by fitting with a binned likelihood fit the templates of the $\not\!\!E_T$ distribution of W+jets and non-W samples to the data.

First of all, we use the pretag data sample where the Single Top (N_{ST}), electroweak (N_{EW}) and $t\bar{t}$ normalization values are fixed by the theoretical cross sections while the normalization of non-W (N_{QCD}) and W+jets are free to float.

Those fits are performed in the $0 < \not\!\!E_T < 120$ GeV range separately for each lepton category (CEM, PHX, CMUP, CMX, and EMC) as the instrumental fake lepton fractions are different for electrons and muons.

[5] see Sec. 6.2.2.

The QCD fraction is obtained using the fit results after requiring $\not\!\!E_T > 25$.

For the PHX electron sample, we require $\not\!\!E_T > 20$ GeV in order to minimize sensitivity to the trigger. The results of $\not\!\!E_T$ fits in the pretag samples for this analysis are shown in Fig. 7.1 and Fig. 7.4 for the 2-Jets and 3-Jets bins.

7.5.1 Pretag Sample

The total number of QCD events in the pretag sample ($N_{\rm QCD}^{\rm pretag}$) is then given by:

$$N_{\rm QCD}^{\rm pretag} = F_{\rm QCD}^{\rm pretag} \cdot N^{\rm pretag} \tag{7.3}$$

where $F_{\rm QCD}^{\rm pretag}$ is the fraction of non-W events obtained from the fit. Then, we find the normalization of W+jets samples as:

$$N_{\rm W+jets}^{\rm pretag} = N^{\rm pretag} \cdot (1 - F_{\rm QCD}^{\rm pretag}) - N_{\rm top+EW}^{\rm pretag} - N_{\rm ST}^{\rm pretag} \tag{7.4}$$

7.5.2 Tagged Sample

The tagged QCD fraction is evaluated after we obtain the W+jets (W+bb, W+cc, W+c, W+lf) normalizations: non-W QCD component is fitted to the tagged data against a template composed by all the other backgrounds; the two normalizations are free to float and the QCD one is extracted. For the QCD template in the tagged fraction, we use the same pretag non-W templates rejecting the events without taggable jets[6]. We use a separate template for the tagged W+jets template, which is composed of the (properly normalized) contribution of W+bb, W+cc, W+cj, W+lf plus the electroweak content weighted by the tagging rates. The non-W normalization is then obtained by the formula:

$$N_{\rm QCD}^{\rm tag} = F_{\rm QCD}^{\rm pretag} \cdot N^{\rm tag} \tag{7.5}$$

The results of the $\not\!\!E_T$ fits in the single SecVtx (SVT) are shown in Fig. 7.2 and Fig. 7.5 for the 2J1T and 3J1T bins, while double SecVtx (SVTSVT) tagged samples are shown in Fig. 7.3 and Fig. 7.6 for the 2J2T and 3J2T bins. Fits in the double tagged region suffer from low statistics.

[6] A taggable jet is one that is within the acceptance of the silicon tracking detector and which has at least two tracks in it.

7.6 W + Heavy Flavor Event Yield Estimate

The number of W+jets expected events in the tagged sample comes from the combination of W + heavy flavor (W+HF) and W + light flavor (W+LF) MC event selection and their subsequent normalization. The major sources of W+HF background events are $W + b\bar{b}$, $W + c\bar{c}$ and $W + cj$ processes[7]. The contribution of W+HF ($N_{\mathrm{W+HF}}$) is computed starting from the normalization of W+jets in the pretag sample (Eq. 7.5.1). Then we need to moltiplicate $N^{\mathrm{pretag}}_{\mathrm{W+jets}}$ for its heavy flavor fraction f_{HF}, the b-tagging efficency $\varepsilon_{\mathrm{tag}}$ and the K_{HF} scale factor, leading to the following formula:

$$N_{\mathrm{W+HF}} = N^{\mathrm{pretag}}_{\mathrm{W+jets}} \cdot f_{\mathrm{HF}} \cdot \varepsilon_{\mathrm{tag}} \cdot K_{\mathrm{HF}} \tag{7.6}$$

where f_{HF} is the fraction of pretag events with jets matched to heavy flavour quarks[8], in the total W+jets MC:

$$f_{\mathrm{HF}} = \frac{N_{\mathrm{W+b\bar{b}}} + N_{\mathrm{W+cj/c\bar{c}}} + N_{\mathrm{W+lf}}}{\sum N_{\mathrm{W+jets}}} \tag{7.7}$$

A discrepancy between MC and data prediction is solved correcting the expected number of events by the scale factor $K_{\mathrm{HF}} = 1.4 \pm 0.4$ [72], that is the fraction of MC and data prediction in the b-tagged sample:

$$K_{\mathrm{HF}} = \frac{f^{HF}_{\mathrm{data}}}{f^{HF}_{\mathrm{MC}}} \tag{7.8}$$

The K-factor is measured in the W+1jets sample, that is a "sideband" and not a signal region. With neglibile signal contamination and the largest statistics of all the jet bins, it is the ideal region to measure heavy flavor content. The same value is used for the three W+HF processes considered ($W + b\bar{b}$, $W + c\bar{c}$ and $W + cj$).

7.7 W +Light Flavor (Mistags) Event Yield Estimate

The last background to be determined in the tagged sample is the one containing jets originating from light-flavoured quarks which have a misreconstructed secondary vertex (mistags). Mistags are caused mostly by random overlap of tracks which are displaced from the primary vertex due to the finite tracking resolution. There are also contributions to the mistag sample from long-living particles and nuclear interactions with the detector.

[7] see Sec. 5.2.5.

[8] There and in the following *matching* refers to the process of matching detector-level jets to quarks from the observed particle branch which only exists in MC samples and identifies the particles that are observed in the detector [71].

The number of expected W+lf events is computed analougsly to the previous case:

$$N_{\mathrm{W+LF}} = N^{\mathrm{pretag}}_{\mathrm{W+jets}} \cdot (1 - f_{\mathrm{HF}} \cdot K_{\mathrm{HF}}) \cdot \varepsilon_{\mathrm{mistag}} \tag{7.9}$$

Here, we multiply the $N^{\mathrm{pretag}}_{\mathrm{W+jets}}$ for the number of pretag events that do *not* match heavy flavor and for the overall fake tag rate.

The tracking-related mistag rate ε_{mistag} is parametrized in data from the rate of **negative** secondary vertex tags[9]. They are jets which show the diplaced vertices as originating from *behind* the primary vertex relatively to the momentum of the jet, meaning that their transverse decay-leght significance is negative $|L_{\mathrm{xy}}/\sigma_{\mathrm{xy}}| < 0$[10] This is unphysical and due to errors of the tracking and the algorithm. Since the SecVtx alghoritm is symmetric in L_{xy}, the tracking-related mistags should occour at the same rate for $L_{\mathrm{xy}} > 0$ and $L_{\mathrm{xy}} < 0$. The assumption is that a good estimate of the positive mistag rate can be obtained from the negative tag rate [52].

The mistag rate derived from negative tags is corrected for the effects of HF in the jet sample and the other contributions to the mistag sample [73].

In the end, CDF uses a function that calculate the mistag probability for each jet in the pretag event. Such a function is named *mistag matrix*. For SecVtx the mistag matrix is a function of η and E_{T} of the jets, number of interaction vertices in the event, track multiplicity and $\sum_i E_{\mathrm{T}^{\mathrm{i}}}$ for the objects in the event. The output probability for each jet is used to calculate the overall probability of the event to be single or double tagged. Finally, the sum of tag probabilities for all pretag events provides the total mistag rate estimate. The mistag matrix used in our analysis was validated for the entire 7.5 fb^{-1}dataset.

7.8 Background Fits and Event Results

We described the contributions of the different background sources to the final background estimate. From Fig. 7.1 to Fig. 7.6 we show the results in the $\not\!\!E_T$ fit for Pretag and Tag in the 2 or 3 jets bin (the ST signal region) separately for the CEM, PHX, CMUP, CMX, EMC charged lepton categories. Then, we present the numbers of observed and expected events for all lepton categories in the 2J1T-3J1T in Tab. 7.3 and 2J2T-3J2T bins in Tab. 7.4.

The total event yield estimate is shown in terms of jet-multiplicity in Fig. 7.7 and summarized in the Tab. 7.5 numbers.

[9] see Sec. 4.7.1

[10] while it should be $L_{\mathrm{xy}}/\sigma_{\mathrm{xy}} > 7.5$ for a positive tag

As shown in Tab. 7.4 and 7.5, the total number of predicted events in the signal region is 9196.9 ± 1287.8 and the number of predicted signal events is 556.6 ± 72.3. Therefore, the signal expectation is 6% of the background, while the previuos analysis accounted of a 5%. That means that our signal estimation rate has an increase of 20%. Nevertheless, there is still about 40% of the overall rate uncertainty on all the background processes, making impossible any significant result based on a simple counting experiment.

In order to solve this problem, this analysis uses a multivariate technique (Neural Network using NeuroBayes® package) to separate signal and background and get a significant result.

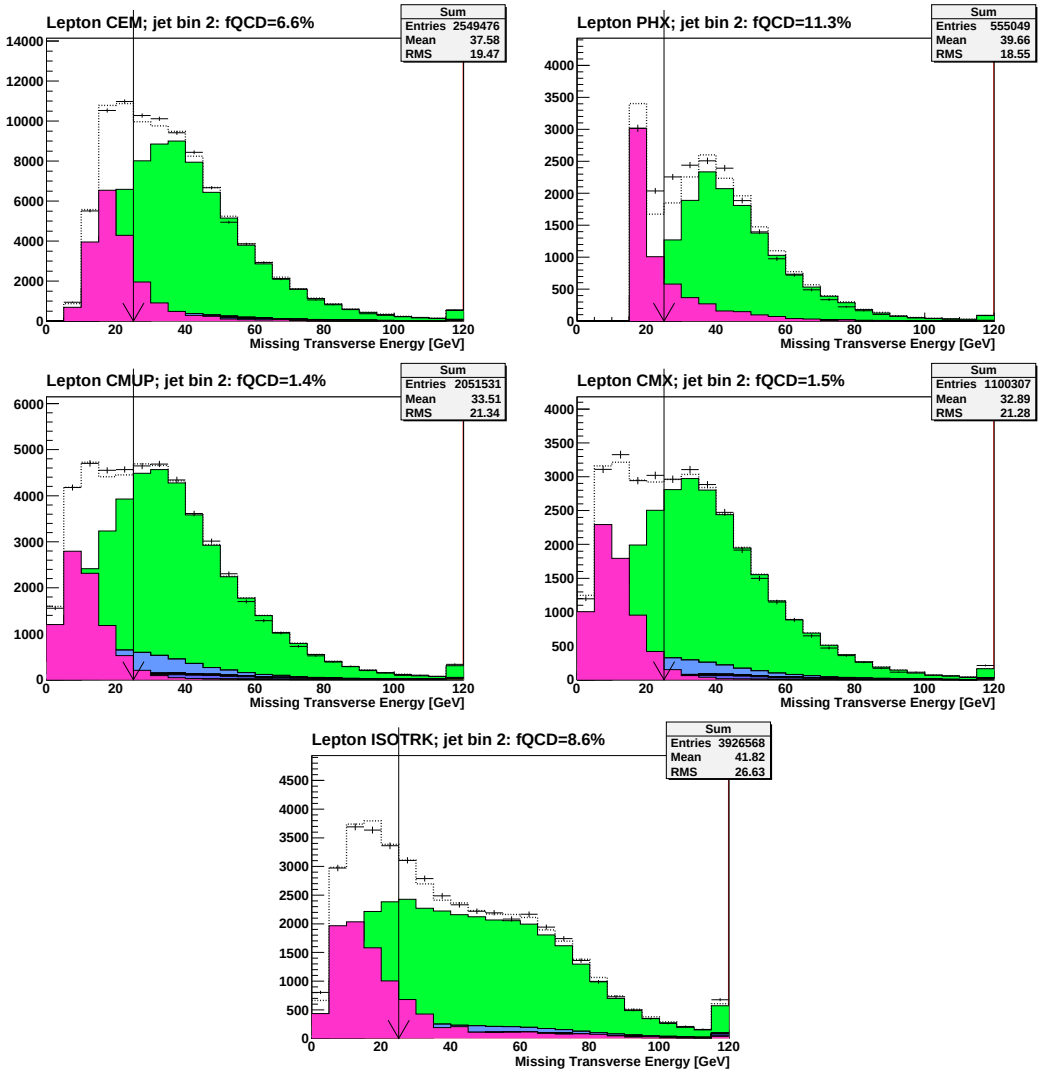


Figure 7.1: QCD fraction estimate for the 2-Jets bin Pretag sample. The horizontal axis represents the fully corrected MET. The QCD background is represented in pink, the remainder of backgrounds in green. The dashed line represents the sum of all the backgrounds and the points represent the data. The figures represent (left to right and top to bottom) the CEM, PHX, CMUP, CMX and EMC (named as ISOTRK) charged lepton categories.

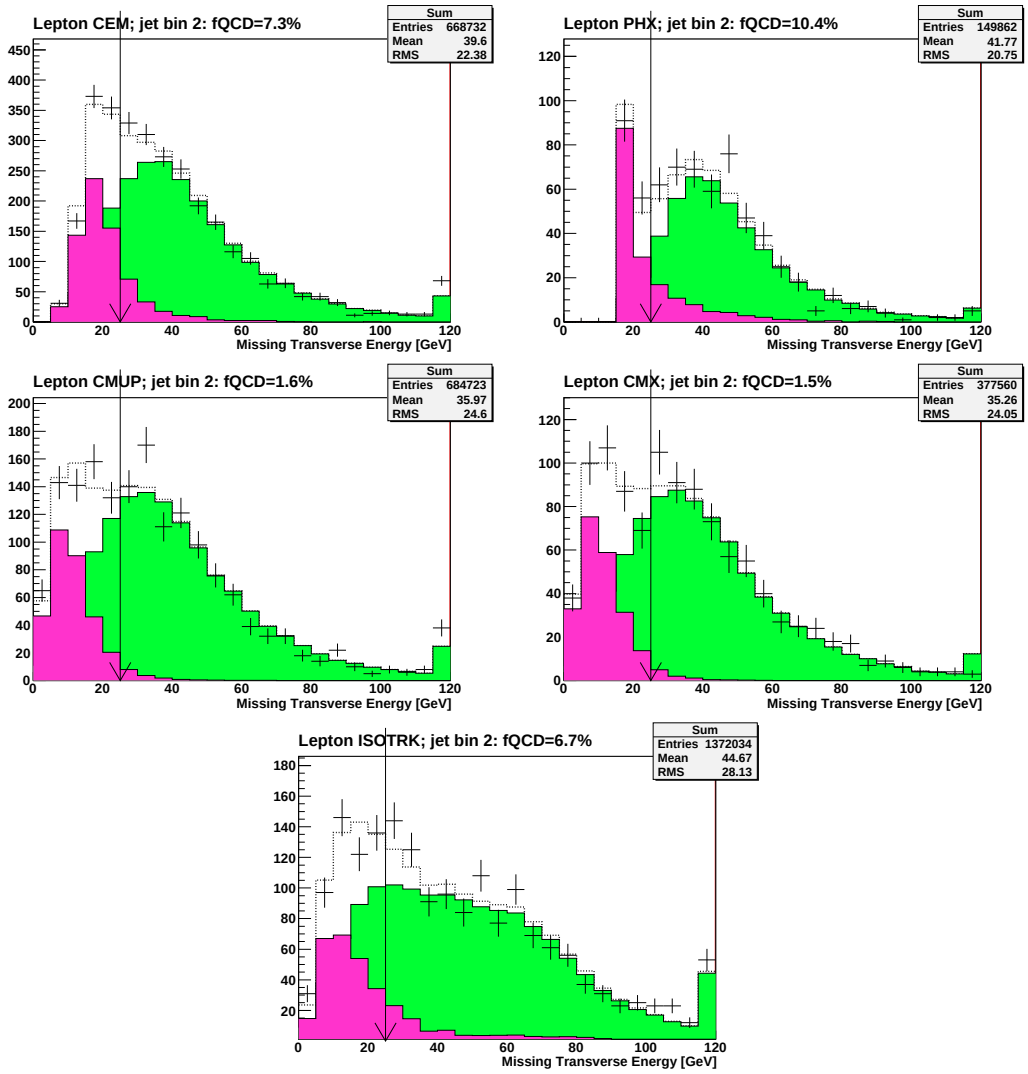


Figure 7.2: QCD fraction estimate for the 2-Jets bin SVT sample. The horizontal axis represents the fully corrected MET. The QCD background is represented in pink, the remainder of backgrounds in green. The dashed line represents the sum of all the backgrounds and the points represent the data. The figures represent (left to right and top to bottom) the CEM, PHX, CMUP, CMX and EMC (named as ISOTRK) charged lepton categories.

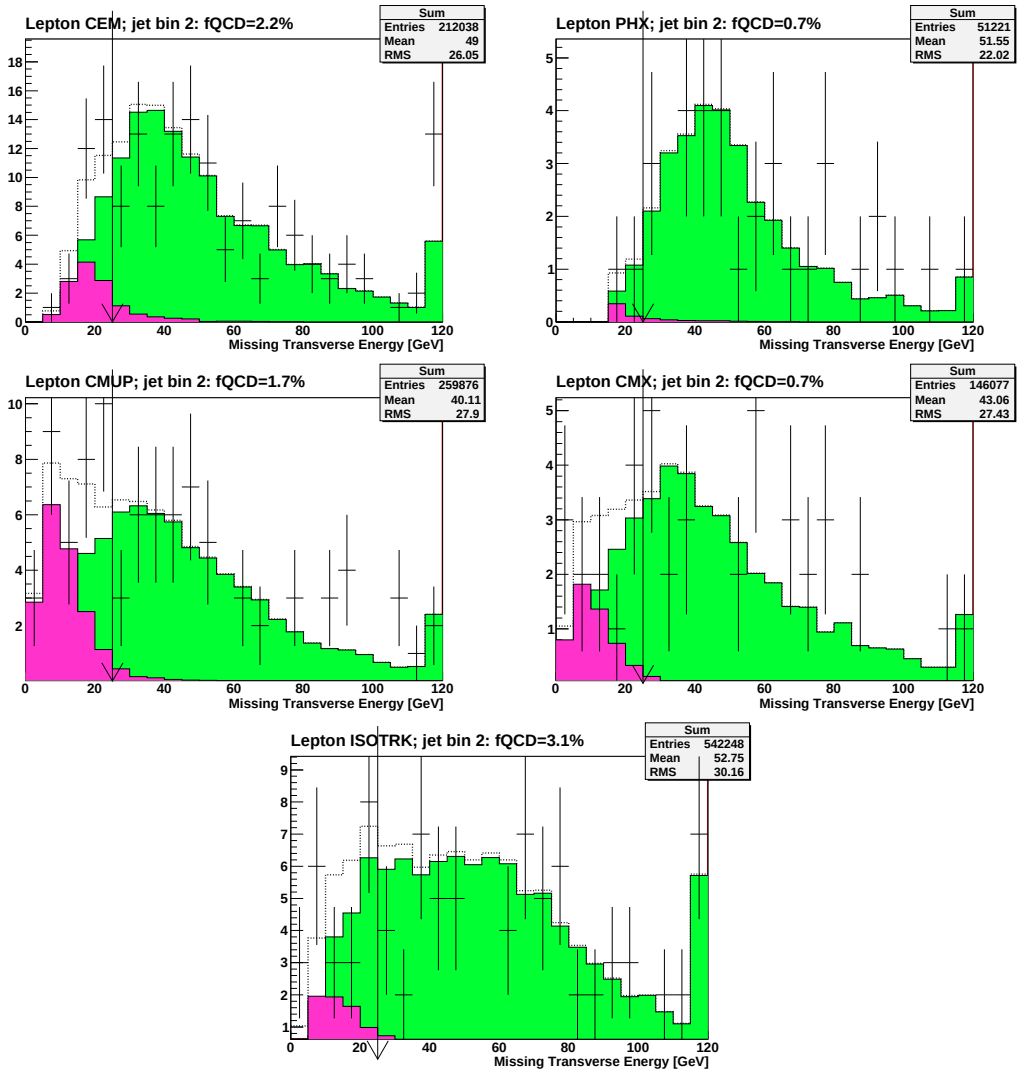


Figure 7.3: QCD fraction estimate for the 2-Jets bin SVTSVT sample. The horizontal axis represents the fully corrected MET. The QCD background is represented in pink, the remainder of backgrounds in green. The dashed line represents the sum of all the backgrounds and the points represent the data. The figures represent (left to right and top to bottom) the CEM, PHX, CMUP, CMX and EMC (named as ISOTRK) charged lepton categories.

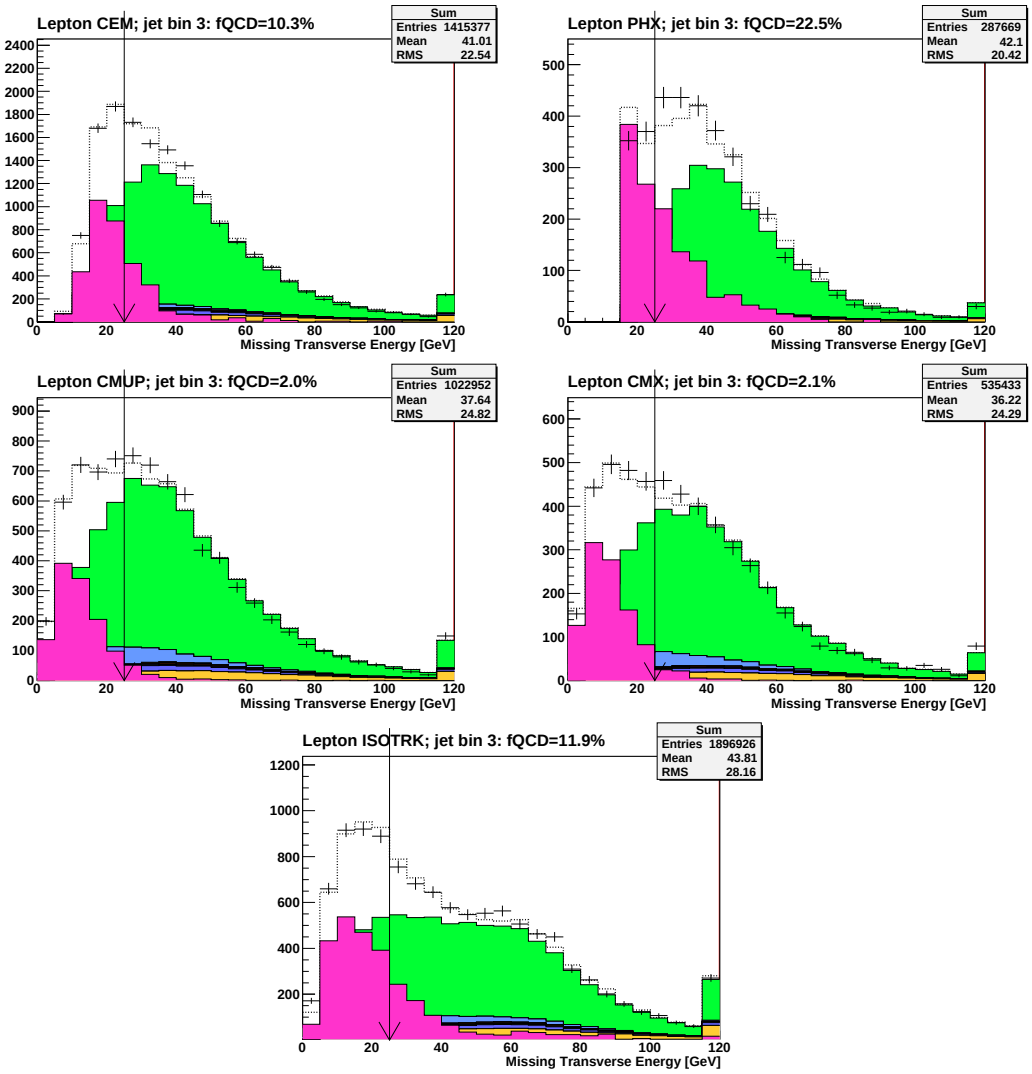


Figure 7.4: QCD fraction estimate for the 3-Jets bin Pretag sample. The horizontal axis represents the fully corrected MET. The QCD background is represented in pink, the remainder of backgrounds in green. The dashed line represents the sum of all the backgrounds and the points represent the data. The figures represent (left to right and top to bottom) the CEM, PHX, CMUP, CMX and EMC (named as ISOTRK) charged lepton categories.

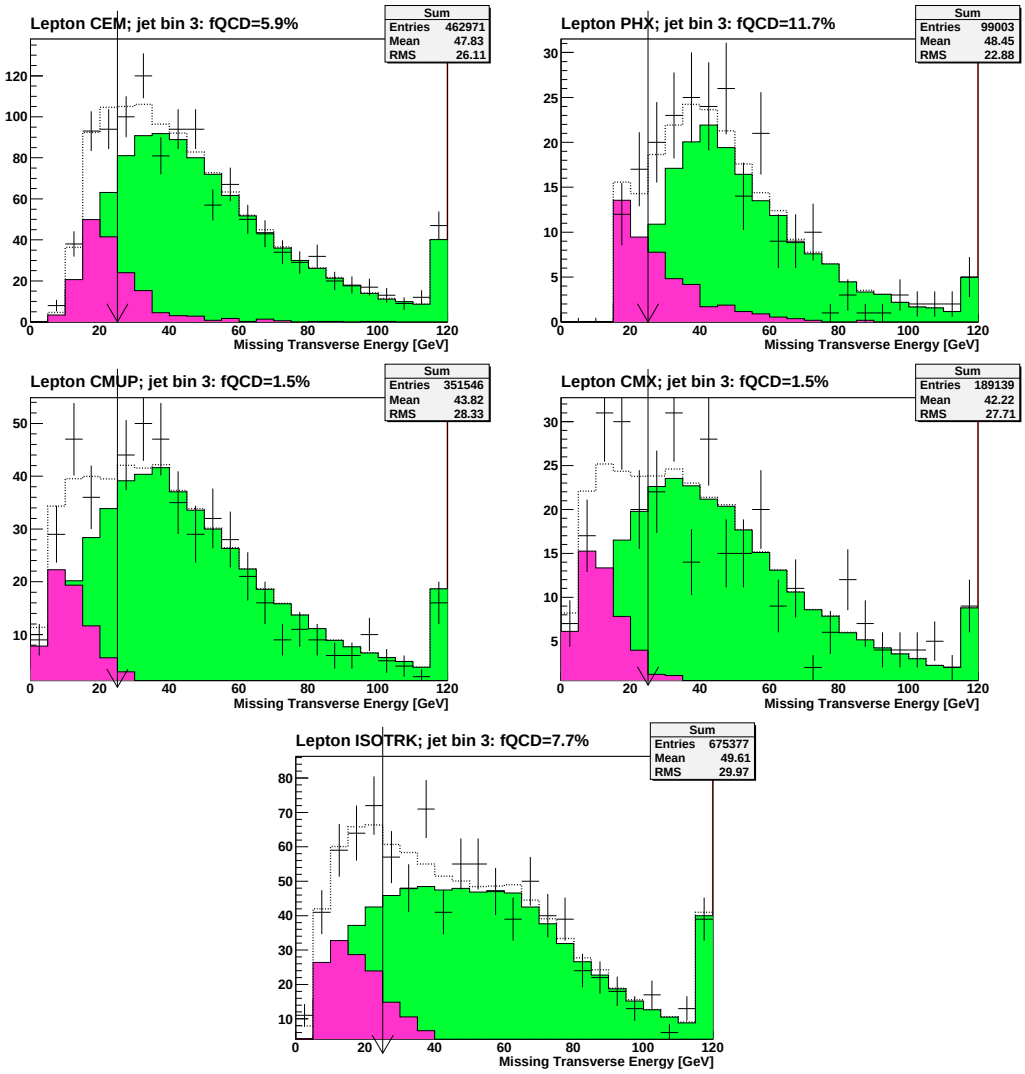


Figure 7.5: QCD fraction estimate for the 3-Jets bin SVT sample. The horizontal axis represents the fully corrected MET. The QCD background is represented in pink, the remainder of backgrounds in green. The dashed line represents the sum of all the backgrounds and the points represent the data. The figures represent (left to right and top to bottom) the CEM, PHX, CMUP, CMX and EMC (named as ISOTRK) charged lepton categories.

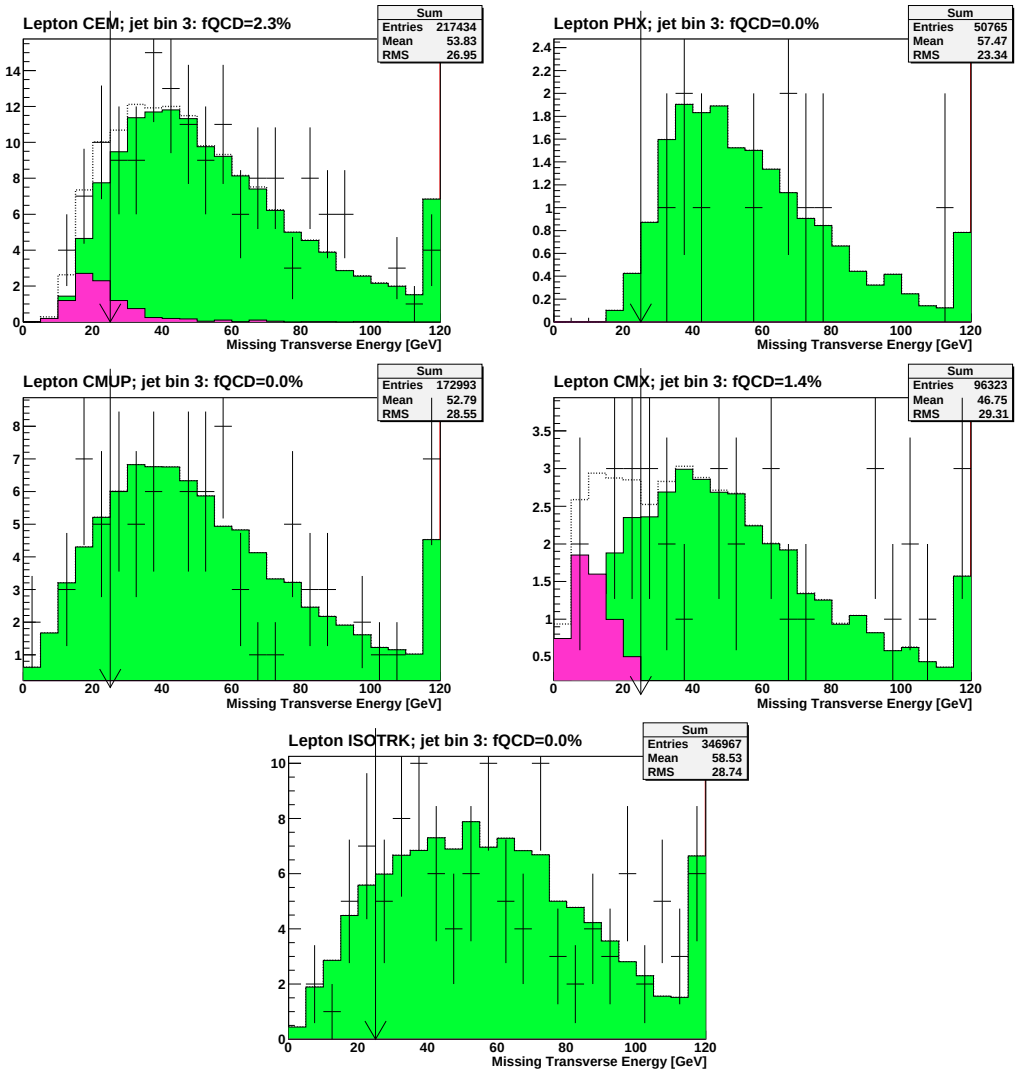


Figure 7.6: QCD fraction estimate for the 3-Jets bin SVTSVT sample. The horizontal axis represents the fully corrected MET. The QCD background is represented in pink, the remainder of backgrounds in green. The dashed line represents the sum of all the backgrounds and the points represent the data. The figures represent (left to right and top to bottom) the CEM, PHX, CMUP, CMX and EMC (named as ISOTRK) charged lepton categories.

Process	W+2jets	W+3jets
$t\bar{t}$	473.9 ± 49.1	1067.2 ± 108.6
WW	147.7 ± 20.6	48.3 ± 6.6
WZ	52.9 ± 5.8	14.4 ± 1.6
ZZ	1.7 ± 0.2	0.7 ± 0.1
Zjets	118.3 ± 15.5	45.6 ± 5.9
Wbb	1452.1 ± 436.9	434.1 ± 131.1
Wcc	766.1 ± 233.0	254.5 ± 77.4
Wcj	583.2 ± 177.4	127.7 ± 38.8
W+Mistags	1459.3 ± 148.5	432.6 ± 47.0
Non-W	315.7 ± 126.3	141.5 ± 56.6
t-channel	192.8 ± 25.3	84.0 ± 10.6
s-channel	127.6 ± 11.3	42.8 ± 3.8
Wt-channel	16.2 ± 4.3	25.7 ± 6.8
Total Prediction	5707.4 ± 876.6	2719.1 ± 292.9
Observed	5533.0	2432.0

Table 7.3: Summary of the predicted numbers of signal and background events with exactly one $b-$tag in the 2-Jets and 3-Jets samples, with systematic uncertainties on the cross section and Monte Carlo efficiencies included.

Process	W+2jets	W+3jets
$t\bar{t}$	98.3 ± 14.5	284.2 ± 41.8
WW	1.1 ± 0.3	1.2 ± 0.3
WZ	8.8 ± 1.3	2.4 ± 0.4
ZZ	0.3 ± 0.0	0.1 ± 0.0
Zjets	4.8 ± 0.7	2.7 ± 0.4
Wbb	182.9 ± 56.1	64.7 ± 19.8
Wcc	10.2 ± 3.2	7.0 ± 2.2
Wcj	7.8 ± 2.4	3.5 ± 1.1
W+Mistags	7.4 ± 1.5	5.4 ± 1.1
Non$-W$	6.8 ± 3.5	3.4 ± 3.2
t-channel	5.9 ± 1.0	14.7 ± 2.4
s-channel	32.3 ± 4.4	11.6 ± 1.6
Wt-channel	0.7 ± 0.2	2.3 ± 0.6
Total Prediction	367.3 ± 65.7	403.1 ± 52.6
Observed	335.0	355.0

Table 7.4: Summary of the predicted numbers of signal and background events with two or more b-tags in the 2-Jets and 3-Jets samples, with systematic uncer- tainties on the cross section and Monte Carlo efficiencies included.

W+jets, ≥1 b Tag	W+2jets	W+3jets	TOTAL
Total Single Top	375.5 ± 46.5	180.96 ± 25.80	556.46 ± 72.30
Total Prediction	6074.7 ± 942.3	3122.2 ± 345.5	9196.9 ± 1287.8
Total Observed Data	5868	2787	8655

Table 7.5: Summary of the total predicted numbers of the signal and the total event yield estimate, compared with the total observed data. The numbers refer to the signal region composed by events with 2 or 3 jets with one or more b-tags.

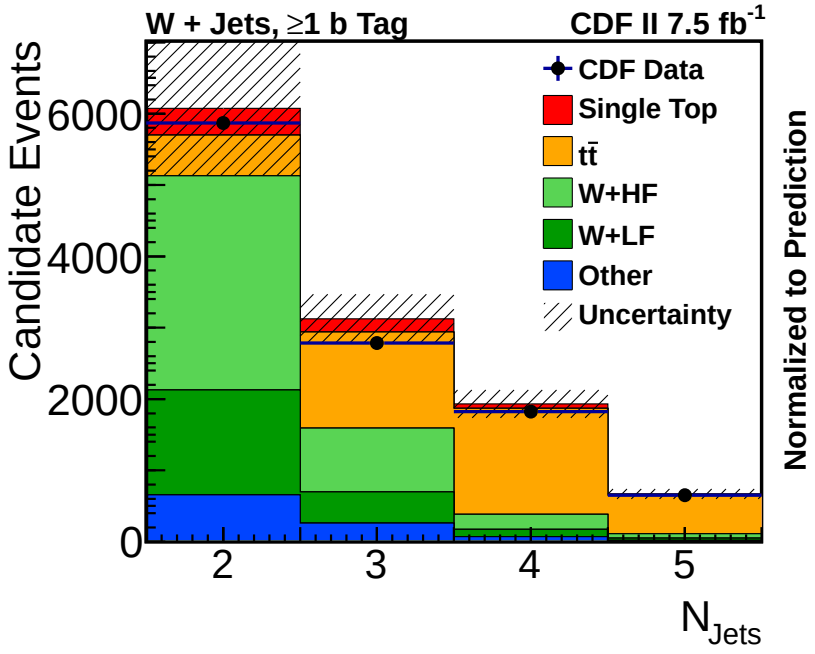


Figure 7.7: Jet multiplicity of the candidate event yield estimate compared to the number of observed events in collision data.

Chapter 8

Neural Networks Analysis and Results

My thesis work focused on the event selection and background estimation. Moreover, we also provided the input variables reconstructed as required for the NN analysis. The NN analsysis and the final cross section measurement was not a specific task for this thesis. However, by now the final analysis results were already achieved and publically presented by the CDF group at the 2012 winter conferences. For sake of completeness, we present the final part of the analysis and the results.

8.1 Neural Network discriminant

The Neural Network part follows the structure of the previous analysis described in Sec. 2.1.1.

8.1.1 NN Training

For the training of the neural network we arrange the composition of the training sample in different way for each channel in such a way that the relative signal process contributes 50% to the total number of events. [1] The relative fractions of all background processes in Tab. 8.1 are obtained by the prediction shown in Tab. 7.3 and Tab. 7.4 [2]. The input variables used for the training are the same of

[1] The s- and t-channel can be treated alternatively as signal or background due to their different topology, in order to produce a further separation in our NN.

[2] except for QCD events, whose event properties and kinematics are hard to model. Our studies showed that even if we trained without QCD processes, the final neural network can distinguish the QCD events as background like (see [74]).

the previous analysis listed in Tab. 2.1. The most discriminating variables will be shown in App. E.

Category	2 jets 1 tag	2 jets 2 tags	3 jets 1 tag	3 jets 2 tags
t-channel	50.0%	—	50.0%	50.0%
s-channel	—	50.0%	—	—
$t\bar{t}$	5.1%	15.9%	22.0%	37.0%
$Wb\bar{b}$	13.7%	27.2%	7.9%	11.2%
$Wc\bar{c}/Wc$	14.0%	3.2%	7.8%	1.8%
$Wq\bar{q}$	14.0%	—	10.0%	—
Diboson	2.4%	2.2%	1.6%	—
Z+jets	0.8%	0.6%	0.7%	—

Table 8.1: Composition of the training samples used to train the neural networks for the combined search.

8.1.2 NN Outputs

After the training procedure, the output distributions of s–, t–and w–channel events are combined into one signal distribution, where the ratio among them is as predicted by the SM. We combine the background processes whose output distributions look very similar and are hence difficult to distinguish, reusulting in six remaining background templates: $t\bar{t}$, W + HeavyFlavor, W + LightFlavor, Diboson, Z+jets, and QCD. The templates of the neural networks in the different channels are shown below. In particular, Fig. 8.1 shows the distribution for the signal and background processes in all channels. Single Top quark events are predominantly found on the right-hand sides (in the zoomed region).

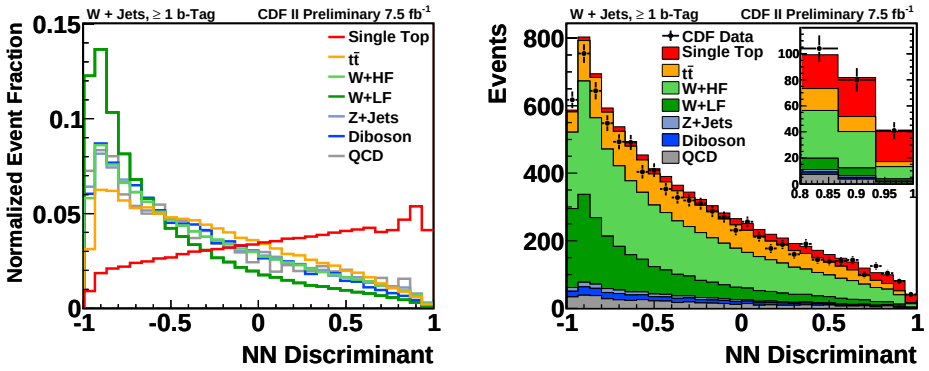


Figure 8.1: Signal and background templates in the two-and three-jet signal region

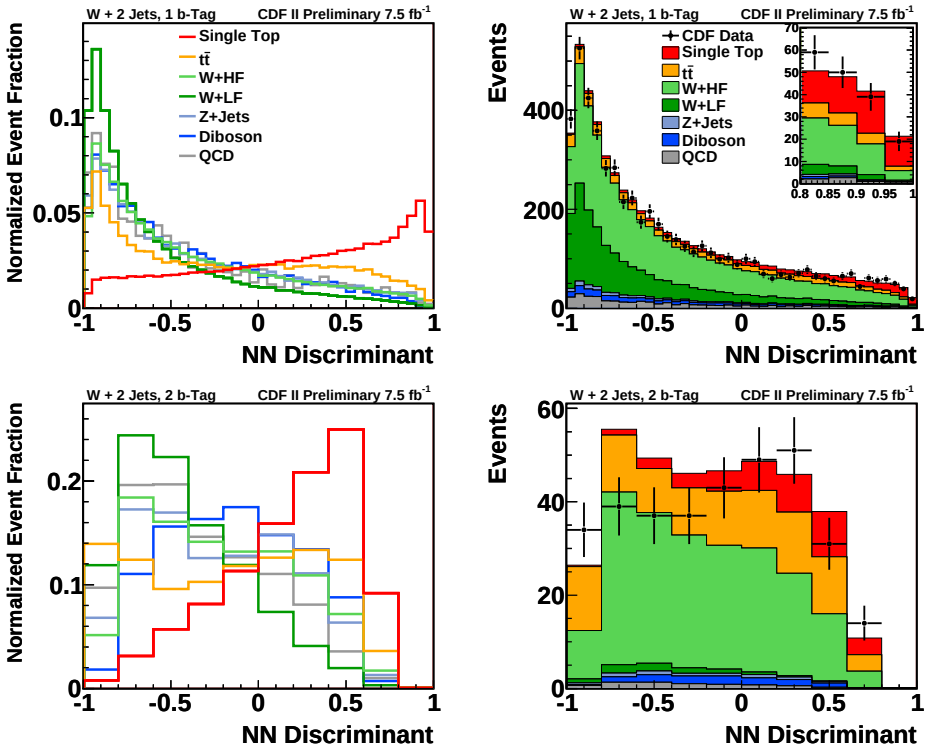


Figure 8.2: Signal and background templates in the two-jet signal region

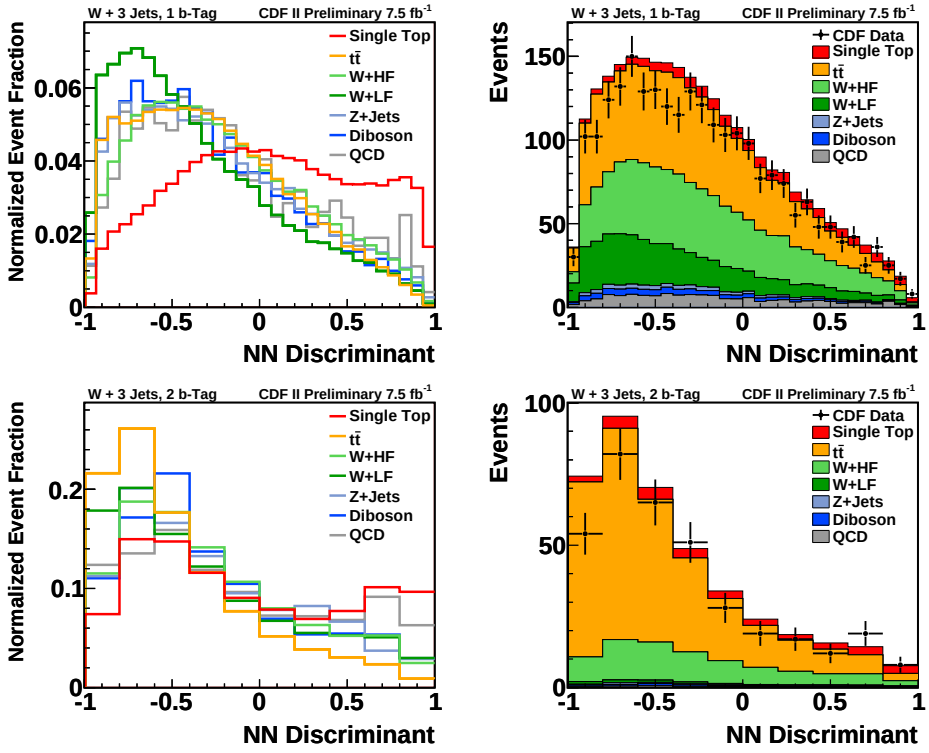


Figure 8.3: Signal and background templates in the three-jet signal region

8.2 Systematic Uncertainties

The measurement of the single top cross section requires substantial input from theorical models, Monte Carlo simulations and extrapolations from control samples in data. We assign systematic uncertainties to our predictions and include the effects of those uncertainties on the measured cross section. The uncertainties we consider are mostly the same as the previous analysis [2]. We consider three categories: uncertainty in the predicted rates of the signal and background processes, uncertainty in the shapes of the distribution of the discriminant variables, and uncertainty arising from the limited number of Monte Carlo events used to predict signal and background expectations in each bin of each discriminant distribution.

Rate uncertainties affects the expected contributions of the signal and background samples. Some sources have asymmetric uncertainties. In Tab. 8.2 the sources of systematic uncertainty considered in this analysis are listed. A detailed description of each source can be found in [74].

Source of Uncertainty	Rate	Shape	Processes affected
Jet energy scale	0–8%	X	all
Initial and final state radiation	0–6%	X	single top, $t\bar{t}$
Parton distribution functions	0–1%	X	single top, $t\bar{t}$
Acceptance and efficiency scale	1–7%		single top, $t\bar{t}$, diboson, Z/γ^*+jets
Luminosity	6%		single top, $t\bar{t}$, diboson, Z/γ^*+jets
Jet flavor separator		X	all
Mistag model		X	W+light
Non-W model		X	Non-W
Factorization and renormalization		X	$Wb\bar{b}$
Jet η and ΔR distribution		X	W+light
Non-W normalization	40%		Non-W
$Wb\bar{b}$ and $Wc\bar{c}$ norm	30%		$Wb\bar{b}$, $Wc\bar{c}$
Wc normalization	30%		Wc
Mistag normalization	10–20%		W+light
$t\bar{t}$ normalization	8%		$t\bar{t}$
Monte Carlo generator	3–7%		single top, $t\bar{t}$
Single top normalization	7%		single top
Top mass	2-12%	X	single top, $t\bar{t}$

Table 8.2: Sources of systematic uncertainty considered in this analysis. X indicates the sources of uncertainty from shape variation. Sources listed below double line are used only in $|V_{tb}|$ measurement

8.3 Analysis Results

8.3.1 Cross Section Measurement

The measurement of single top–quark cross section is performed using the CDF MClimit package, which uses Bayesian handling of the systematic uncertainties. Since the single top and the $t\bar{t}$ background are functions of m_t, the single top quark and $t\bar{t}$ cross section are quoted assuming a top quark mass $m_t = 172.5$ GeV/c^2. For this reason, the uncertainty on the top quark mass is not included in the cross section measurement.

We measured the total cross section of single top quark production σ_{st}, assuming the SM ratio between s-channel and t-channel production: $\beta_s = \beta_t = \beta_w t \equiv \beta$. The posterior distribution is shown in Figure 8.4. From this distribution, we obtain a single top quark cross section measurement of $\sigma_{s+t} = 3.04^{+0.57}_{-0.53}$ pb, assuming a top quark mass of 172.5 GeV/c^2.

8.3.2 Extraction of Bounds on $|V_{tb}|$

To extract $|V_{tb}|$ from the combined measurement, we take advantage of the fact that the production cross section σ_{s+t} is directly proportional to $|V_{tb}|^2$. We use the relation:

$$|V_{tb}|^2_{\text{measured}} = \sigma^{\text{measured}}_{s+t} \frac{|V_{tb}|^2_{\text{SM}}}{\sigma^{\text{SM}}_{s+t}} \tag{8.1}$$

where $|V_{tb}|^2_{\text{SM}} \approx 1$ and $\sigma^{\text{SM}}_{st} = 3.35 \pm 0.34$ [19] Equation 8.1 further assumes that $|V_{tb}|^2 \gg |V_{ts}|^2 + |V_{td}|^2$, because we are assuming that the top quark decays to Wb 100% of the time, and because we assume that the production cross section scales with $|V_{tb}|^2$, while the other CKM matrix elements may contribute as well if they were not very small. Fig. 8.5 shows the joint posterior distribution of all of our independent channels as a function of $|V_{tb}|^2$ (which includes the theoretical uncertainty on the predicted production rate, which is not part of the cross section posterior), from which we obtain a 95% confidence level lower limit of $|V_{tb}| > 0.78$ and extracted $|V_{tb}| = 0.92^{+0.10}_{-0.08}(\text{stat} + \text{sys}) \pm 0.05(\text{theory})$.

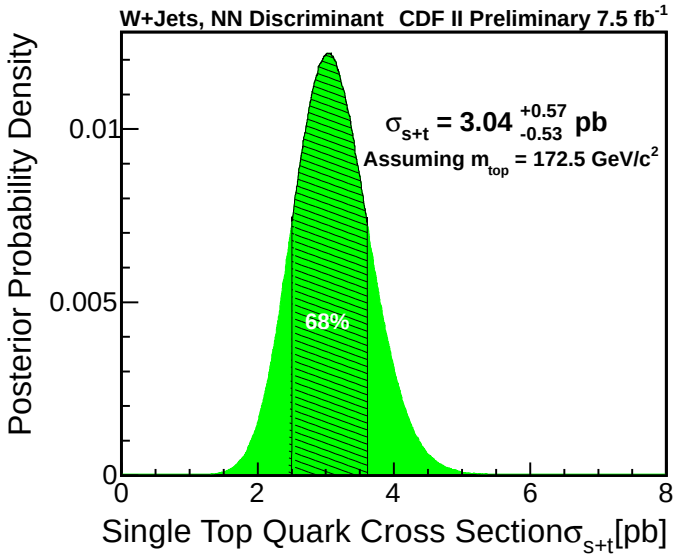


Figure 8.4: The posterior curve of the cross section measurement calculated with the super discriminant histograms as inputs

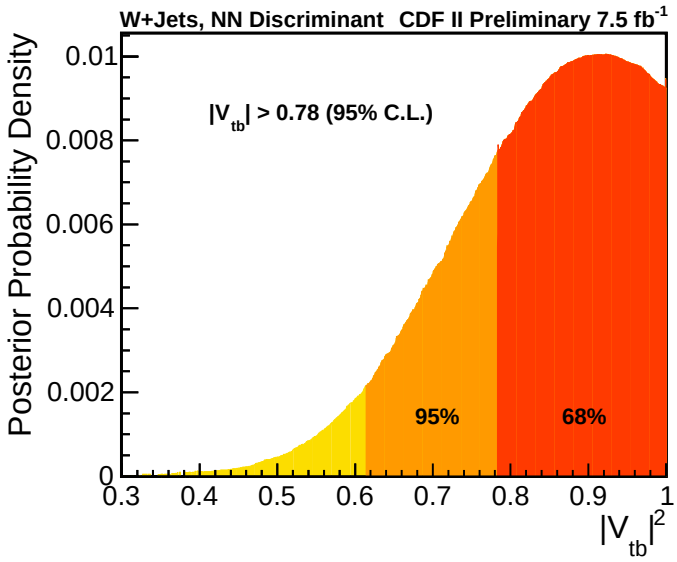


Figure 8.5: The posterior curve for the $|V_{tb}|$ calculation (b).

8.3.3 Two-Dimensional Fit Results

The extraction of the combined signal cross section σ_{s+t} proceeds by constructing a one-dimensional Bayesian posterior with a uniform prior in the cross section to be measured. An extension of this is to form the posterior in the two-dimensional plane, σ_s vs. σ_{t+wt}, and to extract the s-channel and the t-and wt-channel cross sections separately. Here we combined the Wt-channel with t-channel due to the small predicted cross section of Wt-channel at Tevatron and the similar final state signature with t-channel. Moreover, our studies show that the Wt-channel contributes negligible effects in this two-dimensional fit. Thus we neglect the $Wt-$channel contribution in the final result.

The best-fit cross section is the one for which the posterior is maximized, and corresponds to $\sigma_s = 1.81^{+0.63}_{-0.58}$ pb and $\sigma_t = 1.49^{+0.47}_{-0.42}$ pb. The uncertainties on the measurements of σ_s and σ_t are correlated with each other because s-channel and t-channel signals both populate the signal-like bins of each of our discriminant variables. Regions of 68.3%, 95.5%, and 99.7% c.l. are derived from the distribution of the posterior by evaluating the smallest region in area that contains 68.3%, 95.5% or 99.7% of the integral of the posterior. The best-fit values, the credibility regions, and the SM predictions of σ_s and σ_t are shown in Fig. 8.6. We compare these with the NNNLO predictions of $\sigma_{t+wt} = 2.32 \pm 0.27$ pb and $\sigma_s = 1.05 \pm 0.07$ pb [19].

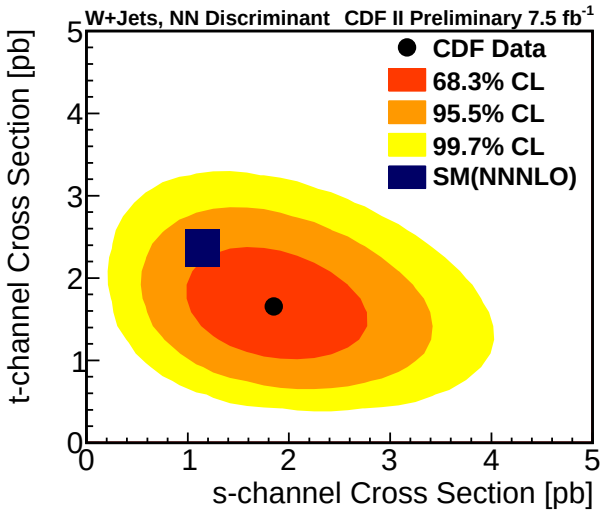


Figure 8.6: The results of the two-dimensional fit for σ_s and σ_t. The black point shows the best fit value, and the 68.3%, 95.5%, and 99.7% credibility regions are shown as shaded areas. The SM predictions are also indicated with their theoretical uncertainties.

Chapter 9

Conclusions

In this dissertation I have presented the Single Top in W+jets analysis at CDF II using 7.5 fb^{-1} of $p\bar{p}$ data.

The main reason for starting a new analysis from the previous one in 3.2 $fb^{-1}$of data, which reported the first observation of the process, was to further investigate with a much larger set of data the open issue of the deviation of the two dimensional fit on s- and t-channel cross section from Standard Model prediction. Single Top cross section measurement also offers the possibility of a direct measurement of the $|V_{tb}|$ element of CKM matrix.

Moreover, due to the different physics conditions between the Tevatron and LHC, the s-channel is completely accessible in CDF while it suffers of a small cross section and a relatively large background at LHC experiments for a significant measurement.

In this thesis I updated the "Single Top analysis using Neural Network", which was one of the 4 analysis involved in the first ST observation in 2009.

In my work I improved the signal acceptance by about 20%, by including more lepton categories. I also reduced some backgrounds with dedicated selection.

The final outcome of the whole analysis was the new CDF measurement of $\sigma_{s+t} = 3.04^{+0.57}_{-0.53}$ pb, $\sigma_s = 1.81^{+0.63}_{-0.58}$ pb and $\sigma_t = 1.49^{+0.47}_{-0.42}$ pb, assuming a top quark mass of 172.5 GeV/c^2. Those results show a better agreement with the SM expectations.

Appendices

Appendix A

KIT Flavor Separator

One of the most significative variables for the Single Top signal events discrimination is the jet flavor separator, also called *KIT flavor separator* [1][75].

The KIT flavor separator is a neural network based tool to improve the purity of the b-quark identification in high-p_T jets. It gives an additional handle to reduce the large background components where no real b-quarks are contained but c-quarks and light jets, as they partecipate for about 50% in the dominant two jet data sample even after imposing the requirement that one jet is SecVtx tagged. The SecVtx algorithm merely provides a binary decision whether a jet contains a reconstructable secondary vertex.

The KIT flavor separator is a neural network trained on simulated SecVtx tagged jet samples. It exploits many different informations as inputs, which are combined by the network. Besides the lifetime based informations, like the impact parameter d_0, and the decay length significance L_{xy}/σ_{xy}, the KIT flavor separator makes use of further variables like the reconstructed SecVtx vertex mass and its decay multiplicity. Finally, the KIT flavor separator output can be treated as a measure for the probability of a true b quark being present within the SecVtx tagged jet.

The latest validation -which means training and comparision to the previous results- of the KIT flavor separator was made on the CDF dataset up to p28, which corresponds to an integrated luminosity of 5.6 ± 0.3 fb^{-1}[76]. We trust on this result also for our 7.5 fb^{-1}dataset. In Fig. A.1 is shown the output of the KIT flavor separator in the 2J1T signal channel for our analysis. For jets containing a b quark, the output of the network accumulates at +1, whereas jets without any heavy quark produce an output close to -1. It is also possible to distinguish jets with c but no b quarks, their output distribution lies in-between the b respectively

[1] KIT refers to the Karlsruhe Institue of Technology, where this method was develeoped.

light jets.

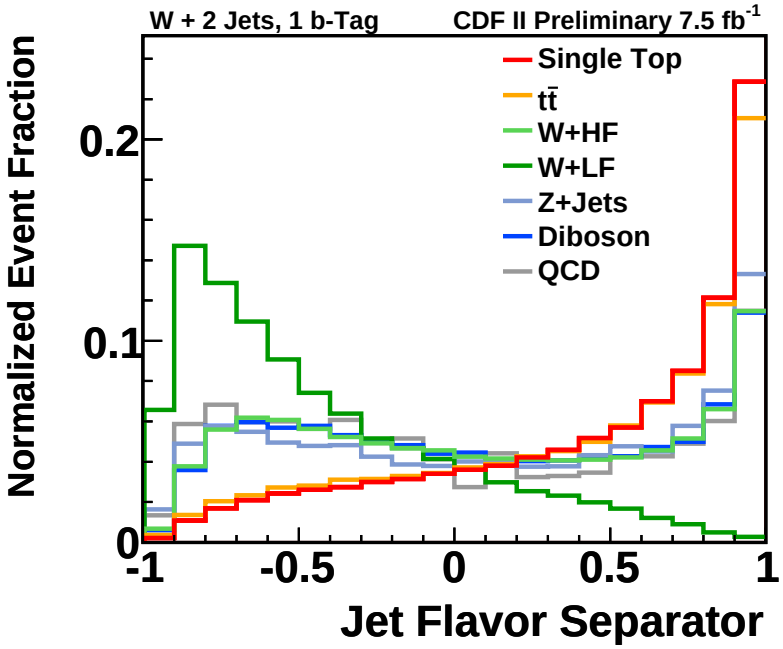


Figure A.1: The results of the two-dimensional fit for σ_s and σ_t. The black point shows the best fit value, and the 68.3%, 95.5%, and 99.7% credibility regions are shown as shaded areas. The SM predictions are also indicated with their theoretical uncertainties.

Appendix B

Top Quark Reconstruction

Some variables used for the training of the neural networks necessitate a reconstruction of the top quark from the final state particles in the event. The top quark decays into a W boson and a b quark. The event selection has been chosen in a way that only leptonically decaying W bosons are considered. It is thus assumed that the lepton and the missing transverse energy $\not{E}_T$ in the event originate from the W boson decay. In a first step this information is used to reconstruct the z component of the 4-momentum of the neutrino. The following relation exists between the 4-momentum of the neutrino p_ν, the 4-momentum of the lepton p_l and the W boson mass m_W^2:

$$(p_l + p_\nu)^2 = (p_W)^2 = m_W^2 \approxeq 80.4\text{GeV/c}^2 \tag{B.1}$$

This quadratic equation can be solved for the z-component of the 4-momentum of the neutrino. There are two solutions:

$$\begin{aligned} p_z^\nu = & \frac{k p_z^l}{E_l^2 - (p_z^l)^2} \pm \frac{1}{2(E_l^2 - (p_z^l)^2)} \\ & \cdot \sqrt{(2kp_z^l)^2 - 4(E_l^2 (p_T^\nu)^2 - k^2) \cdot (E_l^2 - (p_z^l)^2)} \end{aligned} \tag{B.2}$$

with:

$$k = \frac{1}{2}(m_W^2 - m_l^2) + cos(\varphi_l - \varphi_\nu) \cdot p_T^l p_T^\nu \tag{B.3}$$

where the mass values can be found in Tab. 1.1. If both solutions are real, the one with smallest absolute value is taken, since neutrinos produced in top quark decays are rather central. It can also happen, that the p_z^ν solutions become complex if $\not{E}_T$ has been mismeasured. If one would take only the real part of the solution,

Eq. B.1 would no longer be fulfilled. Nevertheless, a physically reasonable solution can be found doing a minimization that corrects $\not{E}_T$ and leads a real p_z^ν solution, as shown in [77].

The energy of the neutrino can then be calculated with:

$$E_\nu^2 = \not{E}_T^2 + (p_z^\nu)^2 \tag{B.4}$$

and the 4-momentum of the W boson is reconstructed to $p_W = p_l + p_\nu$.

A choice has to be made concerning the identification of one of the jets in the event as the jet originating from the b quark of the top quark decay, as required by the event signature, which prescrives at least one jet to be tagged. In the previous analysis, studies were performed on the top quark reconstruction in order to optimize the b-tagged jet assignement procedure (*best hypothesys*) [78].

For the reconstruction of t-channel events with exactly one tagged jet, this jet is attributed to the b quark emanating from the top quark. In case there are two jets with one b-tag in the t-channel MC sample, this is the best possible choice for 75% of the events. In case of two tags, the tagged jet with the largest product of its pseudorapidity times the lepton charge $Q_l \cdot \eta_j$ is taken.

For s-channel events the jet in the event with the highest $Q_L \cdot \eta_j$ is chosen as being the jet coming from the top quark decay. In the s-channel MC sample this is the best possible choice for 60% of the events with two jets and one b-tag and for 40% of the events with two jets and two b-tags. In this analysis we use the same procedure referring to the previuos studies. The Wt-channel was treated as the t-channel, since its contribution to the final signal sample is almost irrelevant.

Finally, the mass of the b-quark jet is set to $m_b = 5$ GeV/c^2 and its energy is calculated by $E_b 2 = m_b^2 + |\vec{p}|^2$. The top quark 4-vector can then be reconstructed: $p_t = p_b + p_W$.

Appendix C

WHAM Analysis Framework at CDF

C.1 Introduction

Associated WH production search Analysis Modules (WHAM) is a new data analysis framework developed in `C++` language for the CDF collaboration. It derives several tools from a previous framework developed by the top quark study group inside CDF and adds functionality through its improved modular structure. WHAM performs almost all the analysis stages from the selection on data and Monte Carlo ntuples produced by the CDF production group, through background estimation and systematics calculation, up to the final measurements such as limits or cross sections. It also contains several applications to easily perform studies and checks on the intermediate and final results of the various analysis stages. I am one of the contributors to WHAM code development and validation. I used WHAM to produce the most of the results presented in this thesis.

C.2 "History" and Motivations

The original idea behind the WHAM package was to perform a combination analysis between two WH searches inside CDF: the one using as a discriminant an artificial neural network (WHNN) and the one using a boosted decision tree and matrix elements (WHME).

The combination between those different analyses can be successfully executed once a trustable common event selection and a common physical objetcs reconstruction, done event by event, are achieved. Such a process is more easily practicable

if made by a common data analysis framework instead of using different packages and afterwords combining the results as done before[1].

For the purpose of this work, it can be noticed that since the WHAM package is constructed in order to perform WH data analysis, it can be therefore adapted to perform other searches that have the same or very close signature as WH one, such as technicolour and $t\bar{t}H$ searches as well as WZ and, of course, Single Top measurements. Particularly, our work focused on building a "Single Top-oriented" branch of WHAM package which could provide a trustable output and share the tools inside the common framework.
The realization and the following usage of a same framework in different analysis allows that a given improvement from a single collaborator, after being validated and integrated in the framework, can be immediatly used by the other developers and users of WHAM in their own analysis. Moreover, the gain in having a common and stable framework would consist in allowing quick iterations and new studies during the analysis process, particularly under the blessing procedures. Important goals of the computational development work also relie in trying to avoid code-redondacies and unnecessary use of memory-space or processing power. Such an open sharing beheaviour (and totally free, in the internal users) stimolates and enriches the scientific process of finding new solutions to new problems, that is, at the end, the physics research.

C.3 A Brief Overview of the WHAM Package Structure

WHAM code is composed by a collection of variuos folders and subfolders whose name describes what is inside. The Main directory contains the 5 most important folders for the package: Setup, Inputs, Commands, Modules and Results. Here below a brief description of each folder is given:

Setup allows an easy setup of the entire analysis framework. It is used mostly in the first installation of the package.

Inputs contains all the preliminary inputs needed for the analysis, such as the list of files with events to be processed, either from data or MC or background simulations, the lists of runs with good data (updated for the dataset used in the analysis). Then, it contains the text files with information for the analysis work, such as the cuts to be applied, tasks to be performed, scale factor to be

[1]Different frameworks can have several internal different definitions which cause a long time consuming process of comparision and validation.

used etc. Since those information are linked in the compiled code, this allows a quick change of the analysis parameters without recompiling everything, that is a long time-consuming process. Also, those parameters are saved with the results in the output, allowing for an immediate *a posteriori* checks.

Commands contains various tools such as submitting computing jobs to the CDF's Central Analysis Farm (CAF) and the limit calculation code. Moreover, it contains several `ROOT` macros for making studies on the variuos analysis output, as for g.e. the QCDVeto studies presented in Sec. 6.4.2.

Modules is the most important folder, containg most of the code in WHAM. It is built in a vertical and modular structure:

- The folder "dep" contains the dependencies files produced in the compilation of code;
- The folder "shlib" contains the shared objects produced during the compilation of code;
- The folder "inc" is the include directory containig symbolic links to all the packages inside the "modules" folder;
- The folder "external" contains all the packages already existing in the CDF software archive and used by WHAM, such as high level object reconstruction code, b-tagging algorithm and their mistag matrices, background calculation methods (`Method II`), limit calculation method and ABCDF[2]
- The folder "Native" contains all the packages exclusively produced for WHAM: the complete event reconstruction, the construction of the analysis trees storing the desired informations of the events selected from original data or MC simulations, the production of histogram root files used for background estimation and several others tasks. Commonly with the previuos "external" folder, "Native" is built in modular structure, meaning that each subfolder has the name of its main containing, and all the subfolders have the same internal structure, with the header files in a namesake sub-subfolder and the source file in the `src` sub-subfolder. This allows that they can be compiled individually instead of

[2] ABCDF is the software package developed in order to model the trigger simulation for MET+jets triggers and combine them for the inclusion of the ISOTRK lepton category (see Sec. 6.2.3). It is placed in the external package as it can be used by other CDF frameworks and analyses as well.

re-compile everything after a little change in some internal file. The idea is to have a saving-time structure with independent subfolders connected by symbolic liks.

Results contains all the results of tasks from "modules". Here we store the `ROOT` trees produced with WHAM using CAF that are used then for background estimation and limit calculation, the limit results, text files with enumerations of events that pass our event selection and their kinematic properties (event dumps), histogram files used for background calculation, various smaller trees for various studies such as signal acceptance improvement or jet energy resolution improvement.

C.4 Future Plans for WHAM

Since the introduction of WHAM package, in late 2009, the WHAM development team worked on the improvement and optimization of the framework in order to make it as much stable as possible and give it at the same time the correct flexibility to be used in different analysis in W+jets dataset. At the moment that this thesis is submitted, the analyses performed and validated with WHAM were the WH search using Neural Networks [79], WW/WZ cross section measurement [80], $R(t\bar{t})$ measurement [81] and several other analsysis and studies. Our Single Top analysis is currently under blessing procedure and used WHAM for the most part of the work, i.e. event recontruction and selection, background calculcation and final cross section measurement [74]. Several other analyses are currently in progress performed by WHAM in the CDF collaboration and others more. Either new or updates analyses are in program to be assigned to students, post-doctoral physicists or whoever else could afford them. The final goal of WHAM would be to offer a stable framework to let the researchers easily work on the final full CDF dataset in the last years of the collaboration.

Appendix D

Validation of the Selection in the 7.5 fb^{-1} ST analysis at CDF

D.1 Introduction

One of the most challeging task to be performed during the analysis is to validate the event selection. There are several reason for putting particular attection to this point: the applied selection cuts must not give a worst result than other fiducial selection to be compared with; if one is using a new data analysis framework, as in our case (see App. C), the output of the selection must be trustable, because errors in this preliminary part would terribly propagate to the following steps of the analysis up to the final results.

There are several different ways one can adopt to validate his own selection. Since this work originates as a new run of the previuos 3.2 fb^{-1} analysis, we decided that a good strategy would have been to compare the output of the current selection made using the new WHAM framework with the previous one made using the old SingleTop CATO package. Therefore, we used exactly the same cuts and the same technical settings of the previous selection (without selecting on the ISOTRK leptons) on the first 3.2 fb^{-1} of CDF data[1], trying to reproduce the old validated output as more precisely as possible. The comparision was made on the previuosly used ST signal MonteCarlo samples as well[2].

The results of the validation were completely satisfying. The comparision shows an event by event agreement very close to the 100% and the differences (just 2 events

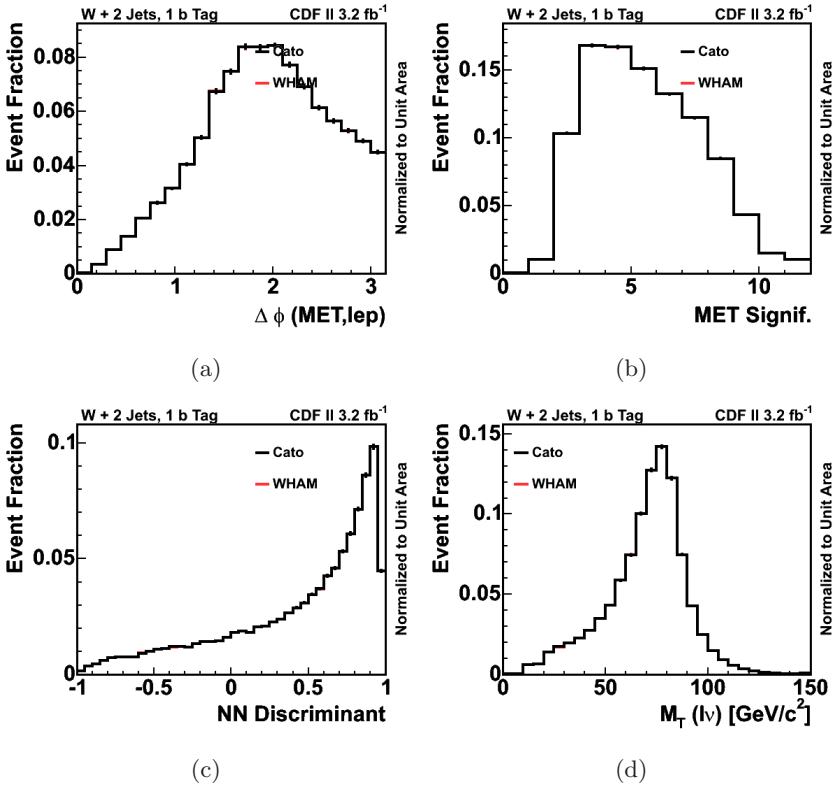
[1] The 3.2 fb^{-1} dataset corresponds to data collected up to Period 19.

[2] The previuos analysis used ST MonteCarlo samples generated with $m_{top} = 175$ GeV, while in the current work new MonteCarlo samples with $m_{top} = 172.5$ were used, as in Sec. 5.1.1.

on about $2 \cdot 10^6$ in the MonteCarlo ST samples) were well understood as coming by different lepton definitions among the two packages. It is remarkable to notice that such a perfect agreement is very different to reach and rarely showed, as the analysis softwares pass through many variation in time. In the following we present some of the control plots used for the validation. They belong to the 2J1T bin, that is the signal bin with the biggest statistics. For the MonteCarlo comparision we show some important variables for the analysis. For the data we show the output of the Neural Network discrmininant for each lepton category. Since this analysis uses the previuos Neural Network package, a perfect agreement of this variable with the previous results is a optimal starting point for the following NN analysis.

D.2 Comparision with the 3.2 fb^{-1} analysis

D.2.1 ST signal MonteCarlo: 2J1T



(a) (b)

(c) (d)

Figure D.1: Comparision between CATO and WHAM packages selection on the MonteCarlo samples for $\Delta\varphi(\not\!\!E_{\rm T}, \text{lep})$ (a), $\not\!\!E_{\rm T,sig}$ (b), NNoutput (c), $M_T^{\rm top}$ (d).

D.2.2 3.2 fb^{-1} Data 2J1T: CEM, PHX, CMUP, CMX, EMC

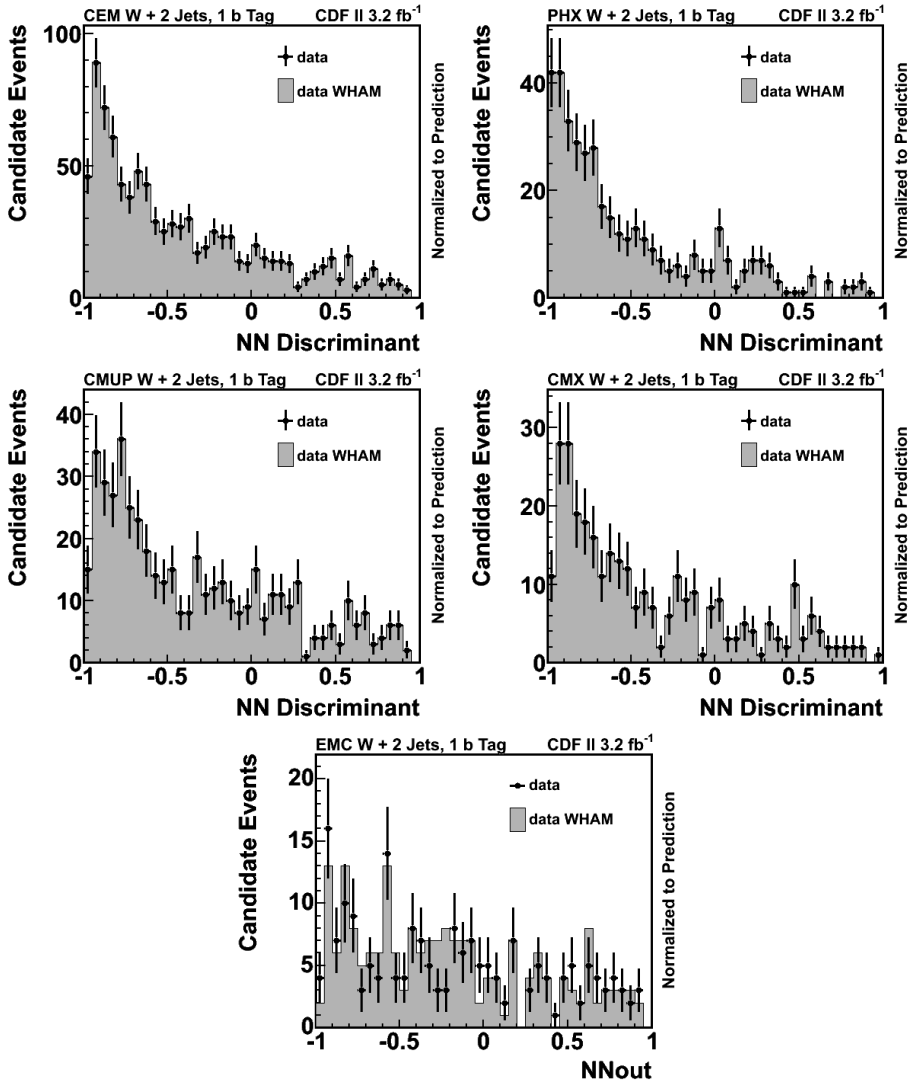
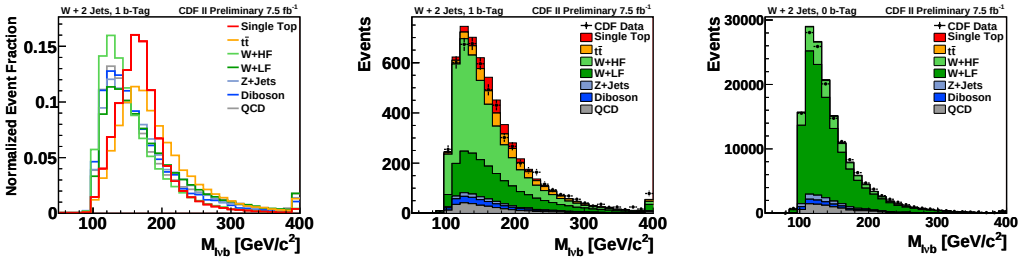


Figure D.2: Comparision between CATO and WHAM packages selection on 3.2 fb^{-1} Data: NNdiscriminant and NNoutput variable (which are both the NN btagger output) for CEM, PHX, CMUP, CMX, EMC lepton categories.

Appendix E

Input Variables for the NN Discriminant

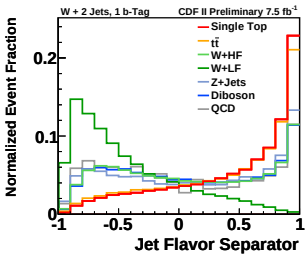
We show some of the most significant variables reconstructed as described in App. B and then used as input for the training of the neural networks. Events come from TLC sample in the 2J1T channel.



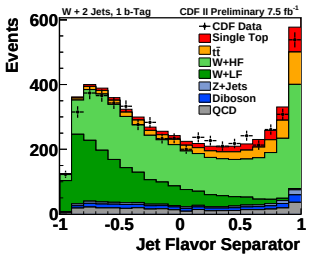
(a)

(b)

(c)

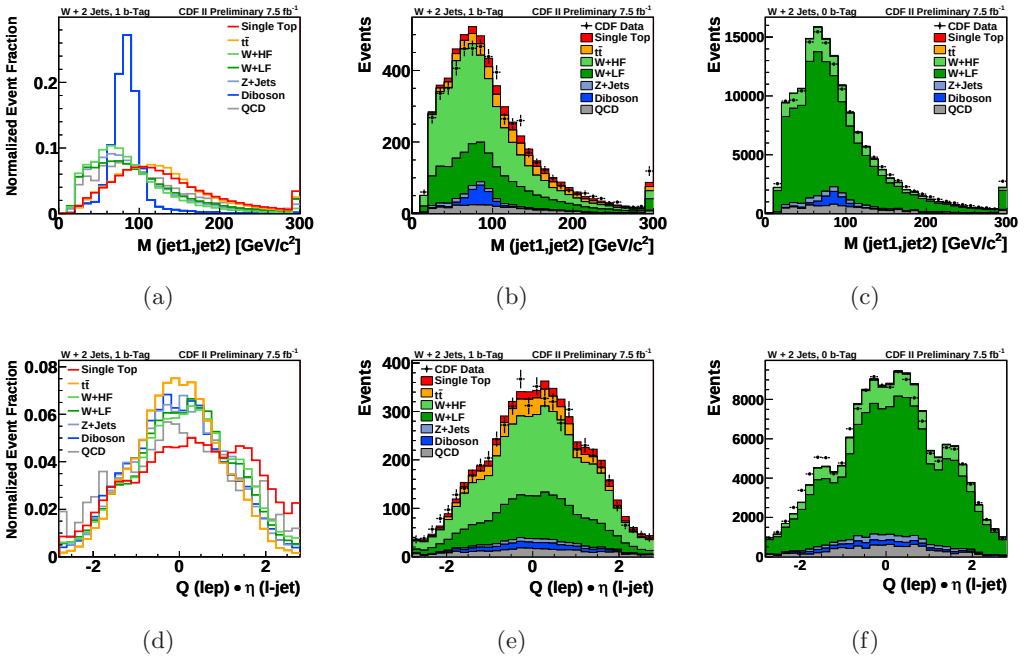


(d)



(e)

Figure E.1: Shape comparison (first column) and MC modeling validation (second column) in the two-jet one-tag signal region, and MC modeling validation (third column) in the untagged two-jet sideband of the discriminating input-variables for TLC events.



(a) (b) (c)

(d) (e) (f)

Figure E.2: Shape comparison (first column) and MC modeling validation (second column) in the two-jet one-tag signal region, and MC modeling validation (third column) in the untagged two-jet sideband of the discriminating input-variables for TLC events.

Bibliography

[1] Abe F. et al. Observation of top quark production in $p\bar{p}$ collisions with the collider detector at fermilab. *Phys. Rev. Lett.*, **74**:2626–2631, 1995. v, 10

[2] Aaltonen et al. Observation of electroweak single top-quark production. *Phys. Rev. Lett.*, **103**:092002, 2009. v, vi, 15, 64, 71, 85, 107

[3] B. W. Harris, Eric Laenen, L. Phaf, Z. Sullivan, and S. Weinzierl. The Fully differential single top quark cross-section in next to leading order QCD. *Phys. Rev.*, **D66**:054024, 2002. v

[4] Zack Sullivan. Understanding single-top-quark production and jets at hadron colliders. *Phys.Rev.*, **D70**:114012, 2004. v

[5] I. J. R. Aitchison and A. J. G. Hey. *Gauge Theories in Particle Physics: A Practical Introduction*. Adam Hilger, 1989. 1

[6] K. et al. Nakamura. Review of particle physics. *J.Phys.G*, **G37**:075021, 2010. 2, 3, 7

[7] Ahmad et al. Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. *Phys.Rev.Lett.*, **89**:011301, 2002. 3

[8] Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, **81**:1562–1567, 1998. 3

[9] G. Shaw F. Mandl. *Quantum field Theory*. John Wiley and Sons, 1984. 3, 9

[10] E.Fermi. *Nuovo Cim.*, **1**(11), 1934. 5

[11] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, **22**:579–588, 1961. 6

[12] P.W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, **12**:132–133, 1964. 7

[13] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, **13**:321–323, 1964. 7

[14] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, **13**:585–587, 1964. 7

[15] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, **19**:1264–1266, 1967. 7, 9

[16] A. Salam. Elementary particle theory. Stockholm, Sweden, 1968. Proc. of the 8th Nobel Symposium. 7

[17] Abachi S. et al. Observation of the top quark. *Phys.Rev.Lett.*, **74**:2632–2637, 1995. 10

[18] Tevatron Electroweak Working Group. Combination of CDF and D0 results on the mass of the top quark using up to 5.8 fb-1 of data. 2011. 11

[19] Nikolaos Kidonakis. Higher-order corrections to top-antitop pair and single top quark production. Technical Report arXiv:**0909.0037**, 2009. proceedings of DPF-2009. 13, 87, 108, 110

[20] Aaltonen et al. Measurement of the Single Top Quark Production Cross Section at CDF. *Phys.Rev.Lett.*, **101**:252001, 2008. 15

[21] Aaltonen et al. Search for single top quark production in pbar p collisions at sqrts=1.96 TeV in the missing transverse energy plus jets topology. *Phys.Rev.*, **D81**:072003, 2010. 16

[22] Jan Lueck. *Observation of Electroweak Single Top-Quark Production with the CDF II Experiment.* PhD thesis, Universität Karlsruhe, 2009. 17, 19

[23] Irja Schall. *Development and Application of a Software Framework for the Single Top Quark Analysis using Neural Networks with the CDF II Experiment.* PhD thesis, Universität Karlsruhe, 2008. 19

[24] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, **11**:070, 2007. 20, 64

[25] Abazov et al. Evidence for production of single top quarks and first direct measurement of V(tb). *Phys. Rev. Lett.*, **98**:181802, 2007. 20

[26] Abazov et al. Observation of single top-quark production. *Phys. Rev. Lett.*, **103**:092001, 2009. 20

[27] Abazov et al. Search for single top quarks in the tau+jets channel using 4.8 fb^{-1} of $p\bar{p}$ collision data. *Phys. Lett.*, **B690**:5–14, 2010. 20

[28] Victor Mukhamedovich Abazov et al. Measurements of single top quark production cross sections and $|V_{tb}|$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys.Rev.*, **D84**:112001, 2011. 20

[29] Nikolaos Kidonakis. Top quark cross sections and differential distributions. *Phys.Rev.*., **D83**, 2011. 21

[30] Serguei Chatrchyan et al. Measurement of the t-channel single top quark production cross section in pp collisions at sqrt(s) = 7 TeV. *Phys.Rev.Lett.*, **107**:091802, 2011. 21

[31] The ATLAS collaboration. Searches for single top-quark production with the atlas detector in pp collisions at sqrt(s) = 7tev. ATLAS-CONF-2011-027, 2011. 21

[32] The CMS collaboration. Measurement of the single top t-channel cross section in pp collisions at sqrt(s)=7 tev. CMS-PAS-TOP-11-021, 2012. 21

[33] The ATLAS collaboration. Measurement of the t-channel single top-quark production cross section in 0.70 fb-1 of pp collisions at sqrt(s) = 7 tev with the atlas detector. ATLAS-CONF-2011-101, 2011. 21

[34] The CMS collaboration. Search for single top tw associated production in the dilepton decay channel in pp collisions at sqrt s =7 tev. CMS-PAS-TOP-11-022, 2011. 21

[35] Georges Aad et al. Evidence for the associated production of a W boson and a top quark in ATLAS at sqrt(s) = 7 TeV. 2012. 22

[36] Simon Van Der Meer. Stochastic Cooling and the Accumulation of Anti-Protons. *Rev. Mod. Phys.*, **57**:092002, 1985. 25

[37] R. et al. Blair. The CDF-II detector: Technical design report. 1996. FERMILAB-DESIGN-1996-**01**. 30

[38] E-811 Collaboration and Avila et al. A measurement of the proton-antiproton total cross section at $\sqrt{s} = 1.8$ TeV. *Physics Letters B*, **445**, 1999. 39

[39] R. Brun. ROOT — An object oriented data analysis framework. *Nucl.Instrum.Meth*, **389**(1-2), 1997. 41

[40] E.Thomson et al. Topntuple: a public analysis module for the top group. CDF Internal Note **6737**, 2003. 41

[41] F. D. Snider. Tracking at CDF: Algorithms and experience from Run I and Run II. *Nucl. Instrum. Meth.*, **A566**:133 – 141, 2006. 45

[42] J. Goldstein at al. Silicon tracking for plug electrons. CDF Internal Note **5970**, 2002. 46

[43] A. Abulencia et al. Measurements of inclusive W and Z cross sections in p anti-p collisions at $\sqrt{s} = 1.96$ TeV. *J.Phys.G*, **G34**:2457–2544, 2007. 48

[44] S. Harper et al. Reconstructing the plug electron energy in 5.3.3. CDF Internal Note **7687**, 2005. 49

[45] D. Acosta et al. Direct photon cross section with conversions at CDF. *Phys.Rev.*, **D70**:074008, 2004. 49

[46] Lubomir Lovas Simona Rolli. Muon id, reconstruction and trigger efficiencies and scale factors for perdiod 9-12 data. CDF Internal Note **9085**, 2007. 51

[47] Salvador Carrillo et al. High pt intermediate muon reconstruction and identification with 5.3.3/6.1.2 data vs 5.3.3 mc. CDF Internal Note **8088**, 2006. 51

[48] C. Gerald et al. Run ii jet physics. CDF Internal Note **5293**, 2000. 52

[49] Abe et al. Topology of three-jet events in $\overline{p}p$ collisions at $\sqrt{s} = 1.8$ tev. *Phys. Rev. D*, **45**:1448–1458, 1992. 52

[50] A. Bhatti et al. Determination of the jet energy scale at the collider detector at Fermilab. *Nucl.Instrum.Meth.*, **A566**:375–412, 2006. 53

[51] J. Adelman et al. Generic jet scale correction for run ii. CDF Internal Note **7358**, 2005. 53

[52] 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. 58, 85, 92

[53] S. Acosta et al. Introduction to run ii jet probability heavy flavor tagger. CDF Internal Note **6315**, 2003. 59

[54] C. Ferrazza and D. Jeans. A new b & c jet identification algorithm. CDF Internal Note **8451**, 2006. 60

[55] Michael Feindt et al. A neural network b tagger for single-top analyses. CDF Internal Note **7816**, 2006. 60

[56] John Freeman. Summer 2009 secvtx scale factors calculated using the electron method through period 22. CDF Internal Note **9848**, 2009. 60

[57] Jan Lueck Zhenbin Wu. Powheg signal samples used in single top analysis. CDF Internal Note **10325**, 2010. 64

[58] F. Maltoni and T. Stelzer. Madevent: Automatic event generation with madgraph. *J. High Energy Phys.*, **02**(027):027, 2003. 64

[59] T. Sjostrand et al. High-energy-physics event generation with pythia 6.1. *Comput. Phys. Commun.*, **135**(238):238–259, 2001. 64

[60] Pavel M. Nadolsky et al. Implications of cteq global analysis for collider observables. *Phys. Rev.*, **D78**:013004, 2008. 64

[61] M.L. Mangano et al. Alpgen, a generator for hard multiparton processes in hadronic collisions. *JHEP*, **07**(001):001, 2003. 64

[62] R. Brun and C. Carminati. Geant detector description and simulation tool. W5013, CERN Program Library Writeup. 65

[63] E.Gerchtein and M.Paulini. Cdf detector simulation framework and performance. La Jolla, CA, USA, 2003. Talk from the 2003 Computing in High Energy and Nuclear Physics (CHEP03). 65

[64] A. Abulencia et al. Observation of WZ Production. *Phys. Rev. Lett.*, **98**:161801, 2007. 67

[65] T. Aaltonen et al. Measurement of the WW+WZ Production Cross Section Using the Lepton+Jets Final State at CDF II. *Phys.Rev.Lett.*, **104**:101801. 67

[66] T. Aaltonen et al. Measurement of the $WW + WZ$ Production Cross Section Using a Matrix Element Technique in Lepton + Jets Events. *Phys.Rev.*, **D82**:112001. 67

[67] Adrian Buzatu et al. WH Neural Network Analysis in the WHAM framework with Improved Triggers for Isolated Track Channel. CDF Internal Note **10440**, 2011. 71, 72

[68] J. M. Slaunwhite. PhD thesis, FERMILAB-THESIS, 2009-01. 76

[69] J. Adelman et al. Method ii for you. CDF Internal Note **9185**, 2008. 85

[70] B. Cooper and A. Messina. "estimation of the background to $w^{\pm} \to e^{\pm}\nu$+ n-jet events". CDF Internal Note **7760**, 2005. 88

[71] Franklin et al. Heavy-flavor content of the w+jets sample. CDF Internal Note **8765**, 2007. 91

[72] C. Cully et al. Calibration of Heavy-Flavor Production in W + 1 Jet Data. CDF Internal Note **9187**, 2008. 91

[73] John Freeman et al. SecVtx Mistag Asymmetry Corrections through Period 22. CDF Internal Note **9847**, 2009. 92

[74] M.Ronzani et al. Measurement of single top quark production in 7.5 fb^{-1} of cdf data using neural networks. CDF Internal Note **10703**, 2011. 103, 107, 122

[75] Svenja Richter. *Search for electroweak single top-quark production with the CDF II experiment.* PhD thesis, Universität Karlsruhe, 2007. 115

[76] P. Butti et al. Kit flavor separator validation with high luminosity samples. CDF Internal Note **10475**, 2011. 115

[77] Antonios Papaikonomou. *Search for single top-quark production via flavor-changing neutral currents with the CDF II experiment.* PhD thesis, 2009. FERMILAB-THESIS-2009-21. 118

[78] Thorsten Walter. *Search for electroweak single top quark production with cdf in proton - anti-proton collisions at s**(1/2) = 1.96-TeV.* PhD thesis, 2005. FERMILAB-THESIS-2005-28. 118

[79] A. Buzatu et al. Search for the standard model higgs boson production in association with a w boson using 7.6/fb. CDF Internal Note **10571**, 2011. 122

[80] F.Sforza et al. Search for $ww/wz \to l\nu$+heavy flavors vector boson production in 7.6 fb-1 of cdf data. CDF Public Note **10598**, 2011. 122

[81] P.Butti et al. Measurement of R(t $\to$ Wb/t $\to$ Wq). CDF Internal Note **10723**, 2012. 122

[82] The ATLAS collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. CERN-PH-EP-2012-218, 2012.

[83] The CMS collaboration. Observation of a new boson with a mass near 125 gev. CMS-PAS-HIG-12-020, 2012.

Acknowledgements

The end of this thesis has a special meaning for me and all the people who supported me during this long and laborious work. I would like to individually thank some of them, including just the people directly involved in this particular thesis work.

Many thanks to my supervisor, Prof. Giorgio Chiarelli, who guided me in this travel and let me discover the wonderful world of Fermilab, giving me the possibility to spend lot of time there. He was always very helpful and willing to solve the problems i faced, which were a lot.

Thanks to Federico Sforza, the great Phd student who introduced me to the magic and terrible world of the CDF code (and WHAM, of course) and always always always helped me. They probably should give him at least two or three Phd's at the end of his studies for all the work he did.

Thanks to Dr.Sandra Leone, who kindly answered my questions and helped me everytime i needed.

A special thanks to Zhenbin Wu, from Baylor University. He shared with me this Single Top analysis work and drove it to the end, dealing with every particular issue and overcoming it. We managed to stay in touch for all the work period, no matter if we were in Italy, Usa or China.

Thanks to Dr. Jan Lück, from KIT. He was one of the most contributors of the previuos ST analysis and he helped us a lot in this update.

Thanks to Homer Wolfe, Adrian Buzatu, Azedine Kasmi and the other WHAM guys (the code group, of course!).

Thanks to Angelo, Fabrizio, Matteo, Paola, Patrizia, Pierfrancesco, Sabatino, Viviana and all the other italian - or not! - people i met in CDF experiment which helped me in some way in Pisa or at Fermilab.

There would be more and more people deserving personal acknowledgements, as they supported or inspired me in these years. To all of them I express all of my gratitude.

In the end, but at the first place for importance, the greatest thank goes to my family.