Measurement of the Ratio of Top-quark Branching Fractions $\mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$ in the Lepton+Jets and Dilepton Channels at the Collider Detector at Fermilab

by

Dmitri Smirnov

M.S., Physics, Moscow State University, 2001

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Physics

The University of New Mexico

Albuquerque, New Mexico

October, 2005

Dedication

To my parents and sister for their support, encouragement and love.

Acknowledgments

I would like to sincerely thank my advisor, Prof. Michael Gold, for his guidance, support, and encourangement in my research work. It was my pleasure to work with Michael, a remarkable scientist with a wide spectrum of knowledge, ranging from the profound details of experimental techniques to the deep theoretical concepts of modern particle physics. I am extremely gratefull to Dr. John Strologas who helped me with this analysis in many ways, especially I thank him for the interesting disscussions and valuable comments from which I benefited substantially. I express my gratitude to Prof. Ken Bloom who has been the Top Group convener at the time when this analysis was accepted by the CDF Collaboration, and who later led the analysis towards its successful publication. I deeply appreciate the warm welcome I always received from Prof. Igor Gorelov since I started my study at UNM as a graduate student. I thank him for his unceasing interest in my work and willingness to help with virtually any problem. I would also like to thank the former member of the UNM group, Dr. Volker Drollinger, who helped me to make a good start and gave me a good example of how to work hard and get the job done.

Measurement of the Ratio of Top-quark Branching Fractions $\mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$ in the Lepton+Jets and Dilepton Channels at the Collider Detector at Fermilab

by

Dmitri Smirnov

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Physics

The University of New Mexico

Albuquerque, New Mexico

October, 2005

Measurement of the Ratio of Top-quark Branching Fractions $\mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$ in the Lepton+Jets and Dilepton Channels at the Collider Detector at Fermilab

by

Dmitri Smirnov

M.S., Physics, Moscow State University, 2001

Ph.D., Physics, University of New Mexico, 2005

Abstract

According to the Standard Model, the top quark decays to a W boson and a b quark virtually 100% of the time. The measurements of $t\bar{t}$ production cross section depend strongly on that assumption. We test this hypothesis with a measurement of $R = \mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$, using a combination of event kinematics and b-tagging techniques. The measurement is carried out using a data sample produced in $p\bar{p}$ collisions at 1.96 TeV and collected at the Collider Detector at Fermilab between March 2002 and September 2003 with an integrated luminosity of ~ 162 pb^{-1}.

The branching ratio R is determined from the relative $t\bar{t}$ tagging rates making the measurement independent of any assumption on the $t\bar{t}$ cross section. Any two tagging rates are sufficient to determine the R but the problem is overconstrained if more than two tagged subsamples are used. The $t\bar{t}$ events are classified by the number

of leptons in the final state. In lepton-plus-jets channel only one of the W bosons decays leptonically, whereas in dilepton channel both W bosons decay leptonically. The measurement of R is performed in both lepton-plus-jets and dilepton samples.

In the lepton-plus-jets channel the background is estimated using the artificial neural network (ANN) technique. The ANN approach allows us to measure the signal fraction in samples with any number of tags. By applying this method alone the branching ratio was measured to be $R = 1.06^{+0.27}_{-0.24}(\text{stat.}) \pm 0.16(\text{syst.})$.

Alternatively, the tagged background contamination in lepton-plus-jets channel is determined from a traditional *a priori* method using data driven and Monte Carlo based techniques. A similar approach is used to determine the $t\bar{t}$ content in the dilepton sample. The combination of ANN background measurement in lepton-plus-jets data sample with the *a priori* lepton-plus-jets and dilepton estimations leads to improved sensitivity in the final value of $R = 1.12^{+0.21}_{-0.19}(\text{stat.})^{+0.17}_{-0.13}(\text{syst.})$.

Finally, we construct the confidence level bands based on the method proposed by Feldman and Cousins. From these bands and our final measurement of R we set the lower limit of $R > 0.61$ at 95% confidence level.

Contents

List of Figures

List of Tables

Chapter 1

Top Quark

1.1 Top Quark in the Standard Model

The top quark was discovered at Fermilab Tevatron in 1995 by both CDF and DØ collaborations [1, 2, 3], thus completing the third generation of fermions in the Standard Model (SM). Among other known fundamental particles the top quark stands out by its surprisingly large mass $m_t = 172.7 \pm 4.3$ GeV/c^2 [4]. The study of the top quark properties is important for particle physics, since deviations from the Standard Model predictions believed to be more pronounced in the behaviour of a heavy particle. The large value of top mass also suggests that the top quark may play a role in the breaking of electroweak symmetry and hence in the origin of fermion masses.

In this Chapter we will briefly discuss the theory developed in order to describe all known properties of the fundamental particles. In the presented theoretical framework the role of top quark in the spectrum of all particles is fixed by its properties. Whether the properties are predicted or not by the theory they need to be verified by experiment. This analysis is dedicated to an important measurement of one of the predicted top quark quantities: the probability of top quark decay into a particular

final state. Based on obtained result one can make conclusions about the consistency between the theory and observations. The reader will find more about the stated problem and proposed solution in the last section of this Chapter.

1.1.1 Standard Model of Elementary Particles

In modern physics the interaction and behaviour of elementary particles are described by a special kind of quantum mechanical theory—gauge field theory. Based on its formulae particle physicists have developed a number of competing models predicting and describing the actual outcomes from all known experiments. Since 1970's many theories were rejected by inconsistency with measured quantities but the one which was capable of surviving all the tests and in a good agreement with experimental results became a Standard Model.

Currently experimental resolution allows us to verify particle physics phenomena down to scales of 10^{-19} m. The particles are considered to be elementary (or fundamental) if they do not reveal any inner structure in the experiments. Within given scope the matter is known to consist of 12 elementary particles including six flavours of leptons and six flavours of quarks. The matter particles are divided into three generations (families) of two leptons and two quarks each as shown in Table 1.1. Each generation consists of one lepton, one neutrino, and two quarks. The charged leptons are electron e, muon μ, and tau τ with $Q = -1$; each of which is associated with

Table 1.1: Three generations of Standard Model elementary particles.

Generations	**1**st		**2**nd		**3**rd	
		Mass, MeV/c^2		**Mass**, MeV/c^2		**Mass**, MeV/c^2
Leptons	$\boldsymbol{e}$	0.511	$\boldsymbol{\mu}$	105.66	$\boldsymbol{\tau}$	1777
	$\boldsymbol{\nu_e}$	$< 3 \times 10^{-6}$	$\boldsymbol{\nu_\mu}$	< 0.19	$\boldsymbol{\nu_\tau}$	< 18.2
Quarks	$\boldsymbol{u}$	1.5-4.5	$\boldsymbol{c}$	1-1.4 $\times 10^3$	$\boldsymbol{t}$	175×10^3
	$\boldsymbol{d}$	5 - 8.5	$\boldsymbol{s}$	80-155	$\boldsymbol{b}$	4-4.5 $\times 10^3$

a corresponding neutral neutrino ν_e, ν_μ, and ν_τ. The quarks have fractional electric charges: $Q = +\frac{2}{3}$ for the "up" type quarks: Up quark u, charm c, and top t; and $Q = -\frac{1}{3}$ for the "down" type quarks: Down quark d, strange quark s, and bottom b. In SM both quarks and leptons are fermions with spin $\sigma = 1/2$. Each particle is described by the mass and several intrinsic quantum numbers assigned to it. The antiparticles have the same mass and spin but opposite sign of quantum numbers like, for example, charge, leptonic and baryonic numbers. Each lepton and quark has a corresponding antiparticle, but for neutrinos it is not known whether they are their own antiparticles or not.

The quarks differ from other fermions in a way that they can not be observed in a free state. This restriction on quark isolation is called the *quark confinement* and it arises from the behaviour of the strong force induced and mediated by the colour-charged particles. A quark is assigned to have one of the three colour charges, conventionally named as: red R, green G, and blue B; or anti-colour charges: $\bar{R}$, $\bar{G}$, and $\bar{B}$. Only colourless combinations of quarks confined inside *hadrons* can be detected in experiments. The *baryons* are the hadrons formed of three valence quarks with different colours and the *mesons* are the hadrons consisting of valence quark and antiquark having opposite colours: C and $\bar{C}$. The strength of the strong interaction converges to zero at very short distances between the quarks, the phenomena known as *asymptotic freedom*, and rapidly increases when particles are pulled apart. A speculative formulation of the quark confinement is usually posed as follows. If the energy stored in the field between separated quarks is high enough than a new quark-antiquark pair can be produced out of the vacuum so that produced final states being again colour-neutral. Thus, the properties of quarks must be inferred from measurable properties of the composite particles which are made up of quarks. The top quark is an exception for it is so heavy that it decays before any colourless hadron can be formed.

Table 1.2: Force carriers of known interactions and their basic properties.

Interaction	**Participants**	**Carriers**	**Mass**, GeV/c^2	**Charge**	**Time**[1], s	**Range**[1], m
Strong	Quarks	8 gluons, g_α	0	0	10^{-23}	10^{-15}
Electromagnetic	All charged	Photon, γ	0	0	10^{-18}	∞
Weak	All particles	$W^\pm$ boson	80.4	±1	10^{-12}	10^{-18}
		Z^0 boson	91.2	0		
Gravitational	All massive particles	Graviton (?)	-	0	-	∞

Ordinary matter consists only of particles from the 1$^{\text{st}}$ generation. The particles in the second and third generations have the identical quantum numbers as the first generation particles but they are heavier. The higher generation particles decay quickly into the first generation ones and can only exist for a short period of time in high-energy experiments or cosmic rays.

The Standard Model also describes the forces through which the elementary particles of matter interact with each other. The interactions in SM are mediated by *gauge bosons* having an integer spin $\sigma = 1$. Traditionally, we distinguish four fundamental interactions in nature: strong (1), electromagnetic (10^{-2}), weak (10^{-6}), and gravitational (10^{-38}), where the number in parenthesis shows the relative magnitude of the force. The fundamental interactions are also listed in Table 1.2 along with their mediator particles and characteristic properties. Only charged particles participate in electromagnetic interactions by exchanging chargeless and massless photons γ. The positively and negatively charged $W^\pm$ bosons and neutral Z^0 boson carry the weak interactions between all the particles. The strong interactions among quarks are carried out by eight massless, electrically neutral, but colourful gluons. The total number of gluons arises from the possibility of constructing eight independent colour-anticolour combinations out of three available colour charges. The SM unifies only strong, electromagnetic, and weak interactions, while gravity remains outside the SM.

[1]**Time** and **Range** are the characteristic time and range of the relevant interaction.

The theory which would unify and explain all the forces as manifestations of a single force is being sought.

In theoretical particle physics the concept of symmetry plays a crucial role. Generally speaking, the symmetry S is implied when a physical system subject to some transformation S remains invariant. Indeed, in practice we see a number of physical quantities attributed to the system which do not change while the system is being translated from initial to final state. Such phenomena are declared as conservation laws and it can be shown that each symmetry is directly related to a conservation law ([5], Noether's theorem). In quantum mechanics the conservation laws are expressed as commutation relations of operator S with the Hamiltonian H of the system, so that H remains invariant:

$$SHS^{+} = H. \tag{1.1}$$

The set of such transformations forms a symmetry group G. The symmetries are classified as *discrete* if the relevant parameters can take only discrete values as in, for example, particle-to-antiparticle transformation by charge conjugation C, parity flipping P, and time reversal T. There are also *continuous symmetries* in SM. The rotation and translation are the typical examples of the continuous space-time symmetries. The symmetries acting on the internal quantum numbers, *e.g.* weak isospin symmetry, colour symmetry, are called the *internal symmetries.* If the latter transformations are functions of space-time coordinates then the symmetries are referred as *local gauge symmetries.*

The interactions of particles forming the content of SM are entirely based on the concept of local gauge invariance. Let a system of particles or a single particle be represented as a complex field Ψ and L is a Lagrangian (or Hamiltonian) describing the dynamics of Ψ. The preservation of invariance of L under the local gauge transformations from a symmetry group G forces the introduction of additional gauge fields that interact with Ψ. The number of gauge fields is equal to the number of generators

of G. In the simplest case of $U(1)$ symmetry group, the transformation $\Psi \rightarrow e^{i\alpha(x)\Psi}$ is associated with only one gauge boson field.

The Standard Model is the gauge theory based on the total gauge symmetry of the fundamental interactions in particle physics, $SU(3)_C \times SU(2)_L \times U(1)_Y$. The $SU(3)_C$ symmetry group is used to describe the gauge theory known as Quantum Chromodynamics (QCD). The composed group $SU(2)_L \times U(1)_Y$ describes the gauge theory called Electroweak Theory (EWT). The EWT unifies the weak and electromagnetic interactions with the latter described by a subgroup $U(1)_{\rm em}$, the base for Quantum Electodynamics (QED), also a gauge theory. The subscripts C, L, and Y refer to the entities taken to define the corresponding local symmetries: colour charge, weak isospin, and hypercharge. The number of gauge fields produced by the special unitary group $SU(N)$ is N^2-1, thus one obtain eight gluons, three weak bosons, and a photon.

The development of any theory is gradual and new concepts are introduced while fresh experimental results become available. The detailed description of Standard Model can be found elsewhere [6, 7]. Here we give an overview of the most essential theoretical excerpts. The top quark is a heavy fermion with electric and colour charges which can participate in all interactions.

Quantum Electodynamics The QED is the most successful gauge theory found to be in excellent agreement with experimental results. The QED Lagrangian is obtained from the one describing fermions with mass m and electric charge Q as Dirac fields $\Psi(x)$:

$$\mathcal{L} = \bar{\Psi}(x)(i\partial\!\!\!/ - m)\Psi(x), \tag{1.2}$$

with $\partial\!\!\!/ \equiv \partial_\mu \gamma^\mu$ and corresponding equation of motion (Dirac equation) given by $(i\partial\!\!\!/ - m)\Psi = 0$. One can show that under the $U(1)$ gauge transformation $\Psi \rightarrow e^{iQ\alpha(x)}\Psi$

the Lagrangian in (1.2) can be written in gauge invariant form:

$$\mathcal{L}_{\mathrm{QED}} = \bar{\Psi}(x)(i\not{D} - m)\Psi(x) - \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}, \tag{1.3}$$

where the normal derivative ∂_μ is replaced by covariant derivative D_μ,

$$D_\mu\Psi \equiv (\partial_\mu - iQA_\mu(x))\,\Psi. \tag{1.4}$$

There is a hidden term inside D_μ with a gauge vector boson field $A_\mu(x)$, a photon field, inside the covariant derivative. This term is responsible for interactions of A_μ and Ψ fields, *i.e.* a photon and a charged particle.

The last term in (1.3) is a so-called photon kinetic energy term which constitutes the propagation of the photon. It is given in terms of the field strength tensor $F_{\mu\nu}$ defined as:

$$F_{\mu\nu} \equiv \partial_\mu A^\nu - \partial_\nu A^\mu. \tag{1.5}$$

In case of QED theory the gauge invariance of $\mathcal{L}_{\mathrm{QED}}$ leads to the conservation of electromagnetic current, J_μ, and charge, Q:

$$J_\mu = \bar{\Psi}\gamma_\mu Q\Psi, \quad \partial_\mu J^\mu = 0, \quad Q = \int J_0(x)d^3x. \tag{1.6}$$

Quantum Chromodynamics As it was noted earlier quarks beside having an electrical charge also possess a colour charge which stipulate the participation in strong interactions. The QCD is a gauge theory of strong interactions. Since there are three possible colour states, the QCD is based on the gauge symmetry group $SU(3)$ of colour transformations. In this representation the particle field is given as three-vector:

$$q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}, \tag{1.7}$$

where q_i are the quark fields. Also, there are eight gauge bosons, gluons $g_\alpha,\ \alpha = 1,\ldots,8$, mediating the strong interactions among the quarks. The QCD Lagrangian can be written in the form similar to the one in QED:

$$\mathcal{L}_{\text{QCD}} = \sum_q \bar{q}(x)(i\not{D} - m_q)q(x) - \frac{1}{4}F^\alpha_{\mu\nu}(x)F^{\mu\nu}_\alpha. \tag{1.8}$$

The covariant derivative in the QCD case is defined as

$$D_\mu q \equiv \left(\partial_\mu - ig_s\left(\frac{\lambda_\alpha}{2}\right)A^\alpha_\mu(x)\right)q, \tag{1.9}$$

where g_s is the strong coupling constant, $\frac{\lambda_\alpha}{2}$ are $SU(3)$ generators given by 3×3 traceless hermitian matrices, and $A^\alpha_\mu,\ \alpha = 1,\ldots,8$ are gluon fields. If D_μ is substituted into Lagrangian one will see that the sum will contain terms for kinematic energy of the quarks and terms for quarks-gluons interactions.

Similarly to the QED Lagrangian the last term in (1.8) defines gluon field strength with tensor $F^\alpha_{\mu\nu}$ given by

$$F^\alpha_{\mu\nu} = \partial_\mu A^\alpha_\nu(x) - \partial_\nu A^\alpha_\mu(x) + g_s f^{\alpha\beta\gamma}A_{\mu\beta}A_{\nu\gamma}, \tag{1.10}$$

where an extra bilinear term with structure constants $f^{\alpha\beta\gamma},\ \alpha,\ \beta,\ \gamma = 1,\ldots,8$ corresponds to gluon self-interaction.

Electroweak Theory The EWT rose from an attempt to produce a theory describing weak interactions analogous to QED. The experimental data showed that the fermions with opposite helicities couple differently with respect to weak interactions. In order to describe such behavior mathematically we introduce the left-handed (negative helicity) and right-handed (positive helicity) fermions. The handedness of the (anti)particles is achieved by projecting the massless fermionic Dirac field Ψ ($\bar{\Psi}$) into left and right handed components by using chirality operator γ_5:

$$\Psi_L = \frac{1}{2}(1-\gamma_5)\Psi, \qquad \bar{\Psi}_L = \bar{\Psi}\frac{1}{2}(1+\gamma_5), \tag{1.11}$$

$$\Psi_R = \frac{1}{2}(1+\gamma_5)\Psi, \qquad \bar{\Psi}_R = \bar{\Psi}\frac{1}{2}(1-\gamma_5). \tag{1.12}$$

The particles with opposite handedness are assigned different values of some quantum number. This quantum number is called weak isospin, $\vec{T}$, and it is conserved in the weak interactions. The corresponding symmetry is based on the $SU(2)$ group. Therefore, the optimal choice of fermionic fields consistent with the experiment is the combination of fermions into the following left doublets, f_L, and right singlets, f_R:

$$\begin{array}{lccccc} f_L: & \begin{pmatrix} e \\ \nu_e \end{pmatrix}_L & \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_L & \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}_L & \begin{pmatrix} u \\ d \end{pmatrix}_L & \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L . \\ f_R: & e_R & \mu_R & \tau_R & u_R, d_R & c_R, s_R \quad t_R, b_R \end{array} \tag{1.13}$$

The right-handed neutrinos were not observed and thus, they are not incorporated into the SM.

It was also noticed that only left-handed particles participate in weak couplings accompanied by change in electric charge, so called weak charged currents. In the reactions where electric charge is unchanged, the weak neutral currents, the coupling can take place between particles with both chiralities just like it happens in pure electromagnetic interactions of massless particles. The idea of unification of weak and electromagnetic theories was proposed by Glashow. He proposed to add to $SU(2)_L$ a new $U(1)$ group but with generators different than those in QED $U(1)_{\text{em}}$. The suggested group $U(1)_Y$ is based on one generator, the *weak hypercharge* $\frac{Y}{2}$, which commutes with the weak isospin $\vec{T}$. Hence, the combined electroweak group $SU(2)_L \times U(1)_Y$ has four generators, three of which are the $SU(2)_L$ generators, $T_i = \frac{\sigma_i}{2},\ i = 1, 2, 3$ and the fourth one is the $U(1)_Y$ generator, $\frac{Y}{2}$. The relation among corresponding quantum numbers is given by

$$Q = T_3 + \frac{Y}{2}, \tag{1.14}$$

where Q is the electric charge, T_3 is the third component of the weak isospin, and Y is the hypercharge. All generations of fermions have the same electroweak quantum numbers which are shown in Table 1.3 for the first generation only.

The number of associated gauge bosons is equal to the total number of generators,

Table 1.3: Lepton and quark quantum numbers.

	Leptons			**Quarks**			
	$\boldsymbol{\nu_L}$	$\boldsymbol{e_L}$	$\boldsymbol{e_R}$	$\boldsymbol{u_L}$	$\boldsymbol{d_L}$	$\boldsymbol{u_R}$	$\boldsymbol{d_R}$
$\boldsymbol{T}$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
$\boldsymbol{T_3}$	$\frac{1}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	$\frac{1}{2}$	0	0
$\boldsymbol{Q}$	0	-1	-1	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{2}{3}$	$-\frac{1}{3}$
$\boldsymbol{Y}$	-1	-1	-2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{4}{3}$	$-\frac{2}{3}$

thus EWT includes: three weak bosons W^i_μ, $i = 1, 2, 3$ of $SU(2)_L$ group and one hypercharge boson $U(1)_Y$ group.

The Lagrangian of electroweak theory needs to be gauge invariant under the transformations corresponding to $SU(2)_L$ and $U(1)_Y$:

$$\begin{aligned} SU(2)_L: &\quad f_L \to e^{i\vec{T}\vec{\Theta}(x)} f_L, \quad f_R \to f_R, \\ U(1)_Y: &\qquad f_{L,R} \to e^{i\frac{Y}{2}\alpha(x)} f_{L,R}, \end{aligned} \tag{1.15}$$

where under $SU(2)_L$ the left-handed fermions transform as doublets and right-handed fermions transform as singlets. The Lagrangian includes the following major interaction terms:

$$\mathcal{L}_{\text{EWT}} = \sum_f \bar{f}(x) i \not{D} f(x) - \frac{1}{4} W^i_{\mu\nu} W^{\mu\nu}_i - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}. \tag{1.16}$$

Similarly to QED and QCD Lagrangians the sum over all fermions in the above expression represents the energy of fermions and the last terms include kinetic energy and the self-interaction of W^i_μ gauge bosons which are written in terms of the field strength tensors:

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{ijk} W^j_\mu W^k_\nu, \tag{1.17}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{1.18}$$

The covariant derivative in EWT is given by

$$D_\mu f = \left(\partial_\mu - ig\vec{T} \cdot \vec{W}_\mu - ig' \frac{Y}{2} B_\mu \right) f, \tag{1.19}$$

where g and g' are coupling constants corresponding to $SU(2)_L$ and $U(1)_L$.

There are two problems with Lagrangian in (1.16): First we did not include the mass terms of the form $\bar{f}mf$ because they are not invariant under $SU(2)_L$ transformation, and second the gauge fields W^i_μ and B_μ do not correspond to observed physical bosons: photon A_μ, $W^\pm_\mu$ and Z^0_μ bosons. The former difficulty will be solved later by introducing mass generation mechanism named after Higgs. The latter problem is solved by rotating the electroweak eigenstates by some angle Θ_W. Therefore, the neutral fields, A_μ and Z^0_μ, can be expressed as linear combination of B_μ and the third component of W^i_μ in the form:

$$Z^0_\mu = W^3_\mu \cos\Theta_W - B_\mu \sin\Theta_W, \tag{1.20}$$

$$A_\mu = W^3_\mu \sin\Theta_W + B_\mu \cos\Theta_W, \tag{1.21}$$

and the charged boson fields are defined in terms of first two components of W^i_μ field as

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu). \tag{1.22}$$

The angle Θ_W is called the *weak mixing angle*. This is one of the 19 free parameters of the SM. Its value is measured experimentally

$$\sin^2\Theta_W = 0.2255 \pm 0.0021. \tag{1.23}$$

Higgs Mechanism In order to get the mass spectra of particles the procedure known as *Spontaneous Symmetry Breaking* (SSB) is applied to electroweak sector of Standard Model. SSB is a general concept widely used in many other theories where the system of interest possesses a given symmetry S, with transformation operator U, except when the system is in its ground state, *i.e.* $U|0\rangle \neq |0\rangle$. As it was pointed out previously the Lagrangian of electroweak interactions is gauge invariant only if the mass terms are absent in the equation, *i.e.* the particles are massless. In SM in order to make the EWT Lagrangian gauge invariant and generate the masses, the

Higgs mechanism based on the SSB principles is introduced. The symmetry breaking in the ground state is equivalent to having some scalar field Φ with non-vanishing expectation values in the ground state: $\langle 0|\Phi|0\rangle \neq 0$. The field Φ is described as fundamental complex doublet with hypercharge $Y = 1$:

$$\Phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix}. \tag{1.24}$$

The simplest possible SSB Lagrangian for the Φ field can be written in the form:

$$\mathcal{L}_{\text{SBS}} = (D_\mu \Phi)^+ (D^\mu \Phi) - V(\Phi). \tag{1.25}$$

The L_{SBS} containing the kinetic and self-interaction parts is added to L_{EWT} given by (1.16). The scalar potential is given by the simplest renormalisable form:

$$V(\Phi) = -\mu^2 \Phi^+ \Phi + \lambda (\Phi^+ \Phi)^2, \tag{1.26}$$

where the mass parameter $-\mu^2 < 0$ and $\lambda > 0$. The minimum of the potential V is achieved at $|\Phi|^2 = -\frac{1}{2}\mu^2/\lambda \equiv \frac{v^2}{2}$, so that the vacuum expectation value is

$$\langle 0|\Phi|0\rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}. \tag{1.27}$$

By choosing a particular value for the vacuum we define a preferred direction in hypercharge and weak isospin space which spontaneously breaks $SU(2)_L \times U(1)_Y$.

The next step in Higgs mechanism involves the construction of physical spectrum by performing "small oscillations" around the vacuum. The scalar field Φ can be parameterized in terms of small fields, $\xi(x)$ and $H(x)$, in the following form:

$$\Phi(x) = \exp\left(\frac{i\vec{\xi}(x)\cdot\vec{\sigma}}{2v}\right)\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{1.28}$$

where the $\vec{\xi}(x)$ are the three scalar fields, the Goldstone bosons, and $H(x)$ is a massive Higgs field with zero expectation value. The Goldstone fields are eliminated by the $SU(2)$ gauge transformation

$$U(\xi) = \exp\left(-i\frac{\vec{\xi}\cdot\vec{\sigma}}{v}\right) \tag{1.29}$$

performed on every field entering the SM Lagrangian: Φ, f_L, f_R, W_μ, B_μ. After rotating the weak eigenstates to the mass eigenstates as in equations (1.20), (1.21) and (1.22), we substitute the result expressions in (1.16) and (1.25), obtaining for the bosonic mass-related terms:

$$\mathcal{L}_{\text{boson mass}} = \frac{1}{2}\left(\frac{g^2v^2}{4}\right)W^+_\mu W^{-,\mu} + \frac{1}{2}\left(\frac{(g^2+g'^2)v^2}{4}\right)Z^0_\mu Z^{0,\mu} + \ldots + \frac{1}{2}(2\mu^2)H^2 + \ldots. \tag{1.30}$$

Therefore, predictions for the corresponding masses are:

$$M_W = \frac{gv}{2}, \qquad M_Z = \frac{\sqrt{g^2+g'^2}v}{2}, \qquad M_H = \sqrt{2}\mu. \tag{1.31}$$

The masses of leptons and quarks are established through their Yukawa couplings to the Higgs doublet. The corresponding part of the SM Lagrangian before the spontaneous symmetry breaking can be written as

$$\mathcal{L}_{\text{Yukawa}} = \lambda_e \bar{l}_L \Phi e_R + \lambda_u \bar{q}_L \widetilde{\Phi} u_R + \lambda_d \bar{q}_L \Phi d_R + \text{h.c.} + \left[2^{\text{nd}} \text{ and } 3^{\text{rd}} \text{ generations}\right], \tag{1.32}$$

where the first three terms are written explicitly for the first generation of fermions with $l_L = \binom{e}{\nu_e}_L$, and $q_L = \binom{u}{d}_L$. $\widetilde{\Phi} = \binom{\phi^{0,*}}{\phi^-}$ is the charge conjugated scalar Higgs doublet Φ and the "h.c." stands for the sum of the same terms but *hermitian conjugated*. The similar terms are understood for the second and third generation particles. Following the Higgs Mechanism procedure established for the bosons and substituting the modified fields into equation (1.32), one obtain for the fermionic mass terms:

$$\mathcal{L}_{\text{Yukawa}} = \frac{\lambda_e v}{\sqrt{2}}\bar{e}_L e_R + \frac{\lambda_u v}{\sqrt{2}}\bar{u}_L u_R + \frac{\lambda_d v}{\sqrt{2}}\bar{d}_L d_R + \ldots, \tag{1.33}$$

from where it is straight forward to write the expressions for mass:

$$m_e = \lambda_e \frac{v}{\sqrt{2}}, \qquad m_u = \lambda_u \frac{v}{\sqrt{2}}, \qquad m_d = \lambda_d \frac{v}{\sqrt{2}}, \qquad \ldots. \tag{1.34}$$

From the above examples it is clear that the Higgs Mechanism generates the masses of the particles given in terms of a unique mass parameter v and the couplings g, g',

λ, λ_e, *etc.* The vacuum expectation value v is determined experimentally from μ decays and its value is $v = \sqrt{\frac{\mu^2}{\lambda}} = (\sqrt{2}G_F)^{\frac{1}{2}} = \frac{2M_W}{g} = 246$ GeV, where G_F is the Fermi constant. Thus, the masses of $W^{\pm}$ and Z^0 were anticipated before they were measured in experiment. In contrast to the gauge boson sector the Higgs boson self-coupling λ and the fermion masses are completely undetermined in the SM. So far the Higgs boson remains the only undiscovered particle predicted by the Standard Model.

CKM Mixing Matrix It was found that charged current exchanges via $W^{\pm}$ do not only couple different (anti-)particle states within the same generation but also states from different generations. The underlying occurrence of such behaviour is that the mass eigenstates and weak eigenstates are not identical. In order to find the relation between those bases, let us rewrite Equation 1.32 in a matrix form:

$$\mathcal{L}_{\text{Yukawa}} = \bar{L}'_L M'_E \Phi E'_R + \bar{Q}'_L M'_U \widetilde{\Phi} U'_R + \bar{Q}'_L M'_D \Phi D'_R + \text{h.c.}, \tag{1.35}$$

where the three generations are combined into right-handed columns:

$$E'_R = \begin{pmatrix} e' \\ \mu' \\ \tau' \end{pmatrix}_R , \qquad U'_R = \begin{pmatrix} u' \\ c' \\ t' \end{pmatrix}_R , \qquad D'_R = \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_R ; \tag{1.36}$$

and the left-handed Dirac-conjugated fields are given by the strings:

$$\bar{L}'_L = \begin{pmatrix} \bar{e}' & \bar{\mu}' & \bar{\tau}' & \bar{\nu}'_e & \bar{\nu}'_\mu & \bar{\nu}'_\tau \end{pmatrix}_L , \qquad \bar{Q}'_L = \begin{pmatrix} \bar{u}' & \bar{c}' & \bar{t}' & \bar{d}' & \bar{s}' & \bar{b}' \end{pmatrix}_L . \tag{1.37}$$

Here we will use primed notation for the weak quark and lepton eigenstates. We have no *a priori* knowledge about the 3×3 complex matrices $M'_{E,U,D}$.

When we substitute the Higgs field and its conjugate into Equation (1.35), we obtain the generalized mass terms:

$$\mathcal{L}_{\text{Yukawa}} = \bar{E}'_L M_E E'_R + \bar{U}'_L M_U U'_R + \bar{D}'_L M_D D'_R + \text{h.c.}, \tag{1.38}$$

where $M_{E,U,D}$ are the mass matrices equivalent to their primed counterparts rescaled by the Higgs vacuum expectation value. In general, these matrices are not diagonal but they can be diagonalized by means of bi-unitary transformations:

$$M_E = V_{E,L}^{\dagger}\mathbf{M}_E V_{E,R}, \qquad M_U = V_{U,L}^{\dagger}\mathbf{M}_U V_{U,R}, \qquad M_D = V_{D,L}^{\dagger}\mathbf{M}_D V_{D,R}, \tag{1.39}$$

where $\mathbf{M}_{E,U,D}$ are diagonal 3×3 matrices and $V_{E,L/R}$, $V_{U,L/R}$, $V_{D,L/R}$ are unitary matrices. Using the above expressions and plugging them into Equation (1.38) we obtain the standard form of the mass terms in the Lagrangian:

$$\mathcal{L}_{\text{Yukawa}} = \bar{E}_L \mathbf{M}_E E_R + \bar{U}_L \mathbf{M}_U U_R + \bar{D}_L \mathbf{M}_D D_R + \text{h.c.}, \tag{1.40}$$

where we defined the mass eigenstates as linear combinations of corresponding weak eigenstates:

$$E_L = V_{E,L} E'_L \qquad E_R = V_{E,R} E'_R \tag{1.41}$$

$$U_L = V_{U,L} U'_L \qquad U_R = V_{U,R} U'_R \tag{1.42}$$

$$D_L = V_{D,L} D'_L \qquad D_R = V_{D,R} D'_R \tag{1.43}$$

As for leptons the weak and mass eigenstates should be very close to each other since there are no right-handed neutrinos in the SM and the left-handed neutrinos have very small masses. On the contrary, the quark states transformations have far-reaching consequences. Using Equations 1.43 one can write the bi-linear terms of the charged-current weak interactions equivalently in both bases as

$$\begin{pmatrix} \bar{u}' & \bar{c}' & \bar{t}' \end{pmatrix}_L \gamma_\mu \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_L = \begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}_L V_{U,L}^{\dagger} V_{D,L} \gamma_\mu \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L , \tag{1.44}$$

where $\mathbf{V} = V_{U,L}^{\dagger} V_{D,L}$ is the generation mixing matrix acting, by convention, on the down-like quarks D. The matrix $\mathbf{V}$ is referred to as the CKM matrix for the explicit parametrization was provided by Cabbibo-Kobayashi-Maskawa [8, 9, 10].

The CKM matrix is complex and unitary by construction satisfying the relation:

$$\mathbf{V}^{\dagger} = \mathbf{V}^{-1}, \tag{1.45}$$

where $\mathbf{V}^{\dagger}$ and $\mathbf{V}^{-1}$ are the conjugate transpose and inverse matrices of $\mathbf{V}$. In general, the dimension of CKM matrix, n, depends on number of generations and in order to define a $n \times n$ complex unitary matrix one need to introduce $(n-1)^2$ observable parameters. The following is considered to be the "standard" parametrization of CKM matrix for three generations:

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13} \end{pmatrix}, \tag{1.46}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$ with the "generation" labels $i, j = 1, 2, 3$. The four parameters are: three angles θ_{12}, θ_{23}, θ_{13}, and the phase δ_{13}. Since only four parameters are used to determine nine matrix elements and from the CKM matrix unitarity[2], it follows that the elements are mutually constrained. In other words, by measuring some of them one is able to predict the magnitude of others. The elements from the first two rows are well measured from various weak decays and their values are reported elsewhere [11]. Therefore, the value of V_{tb} element is expected to be very close to unity: $0.9990 \leq V_{tb} \leq 0.9992$. It is important to note that the expected values of unknown CKM matrix elements will change if the CKM matrix is expanded to accommodate more generations, nevertheless, the matrix itself will remain unitary by construction.

[2]Multiplying both sides of Equation (1.45) from left by $(V^{\dagger})^{-1}$ (from right by V) and using $V^{-1}V = 1$ identity, we obtain $V^{\dagger}V = 1$ ($VV^{\dagger} = 1$). These expressions imply that the sum of squared elements in each row and column in $\mathbf{V}$ is equal to 1.

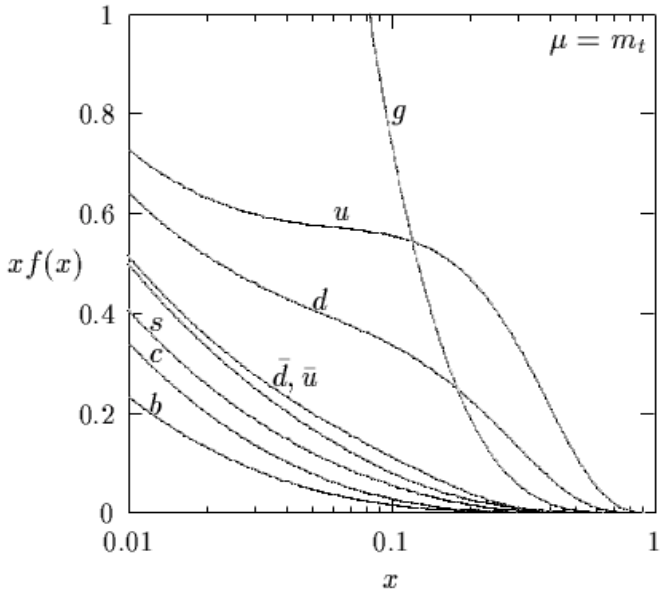


Figure 1.1: Parton distribution functions for different parton species inside the proton at the scale $\mu = m_t$.

1.1.2 Top Quark Production

It was noted in Section 1.1.1 that in the Standard Model the top quark can participate in electromagnetic, weak, and strong interactions via coupling to a photon, $W^{\pm}$ and Z^0 bosons, and a gluon correspondingly. Therefore, the top quark can be produced in many ways if the energy of interacting particles is sufficient to produce such a massive quark. To date, only Fermilab Tevatron is capable of producing top quark in proton-antiproton collisions. The (anti)proton is regarded as a collection of valence and sea partons including quarks, antiquarks, and gluons, each carrying some fraction x of the proton's four-momentum. The chances of specific interaction to occur depend on the probability of finding a parton inside the proton with a specific momentum fraction. These probabilities for different types of partons are given by the *parton distribution functions* (PDF) as shown in Figure 1.1 [12]. More specifically, in $p\bar{p}$ collisions at Tevatron the top quark preferably produced as a result of three distinct

types of interactions at the parton level: quark-antiquark ($q\bar{q}$), qluon-(anti)quark (gq, $g\bar{q}$), and gluon-gluon (gg) interactions.

The inclusive cross section of top quark production in hadronic interactions at the center-of-mass energy $\sqrt{s}$ is given by the formula:

$$\sigma(s) = \sum_{i,j} \int \hat{\sigma}_{ij}(x_i x_j s, m_t^2, \mu) F_i^p(x_i, \mu) F_j^{\bar{p}}(x_j, \mu) dx_i dx_j, \tag{1.47}$$

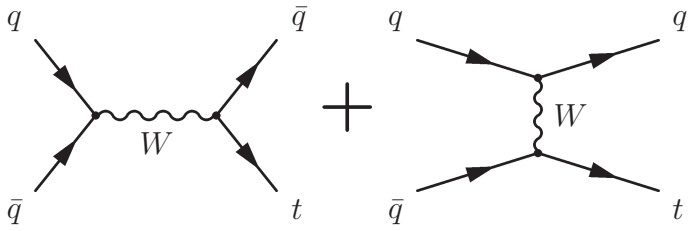
where the summation indexes i and j run over all the partons contained in the initial proton and antiproton and carrying the momentum fractions x_i and x_j. $F_i^{p(\bar{p})}(x_i, \mu)$ is the PDF of parton i in proton (antiproton) evaluated at a scale μ, and $\hat{\sigma}_{ij}(\hat{s}, m_t^2, \mu)$ represents the parton-level cross section of interaction of hadron constituents as a function of top mass m_t and the square of the partonic center-of-mass energy $\hat{s} = x_i x_j s$. The total short distance cross section $\hat{\sigma}$ can be parametrized [13] as

$$\hat{\sigma}_{ij} = \frac{\alpha_S^2(\mu)}{m_t^2} f_{ij}\left(\frac{4m_t^2}{s}, \frac{\mu^2}{m_t^2}\right), \tag{1.48}$$

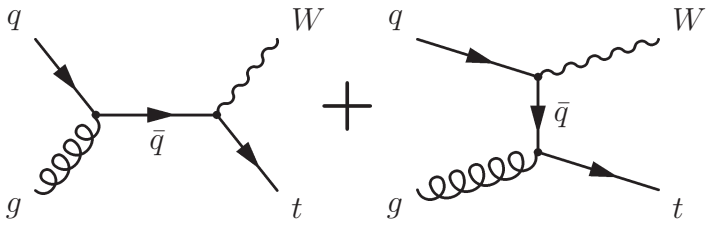
where $\alpha_S(\mu)$ is the QCD running coupling and dimensionless function f_{ij} includes contributions from the higher perturbative expansion terms in α_S. In QCD calculations the higher order terms can contribute significantly. The full cross section is independent of the choice of renormalization scale μ but for calculations with finite number of terms the impact of high order terms may be reduced by choosing $\mu = m_t$. The dependence of α_S on the scale μ is logarithmic and uncertainties are estimated by varying the μ by factor of 2 around the top mass. This way the uncertainty on top quark cross section will reflect the absence of the higher order terms.

The most recent prediction of the top pair cross section with inclusion of next-to-leading order corrections is $\sigma_{t\bar{t}} = 6.7^{+0.7}_{-0.9}$ pb^{-1} at $\sqrt{s} = 1.96$ GeV and $m_t = 175$ GeV [14, 15]. The single top cross section is smaller by a about a factor of two.

Mostly for experimental identification of the top quark production processes, they are distinguished by the number of produced top quarks. Single top quark is produced in weak couplings to $W^{\pm}$ boson. The leading order Feynman diagrams of such



(a) $q\bar{q}$ annihilation.



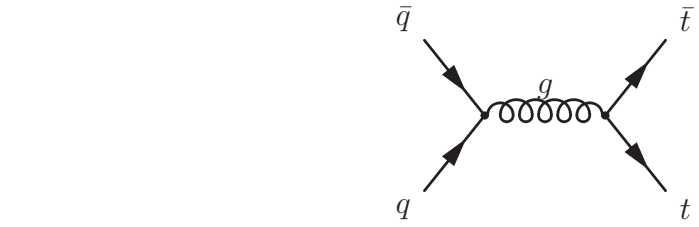
(b) qg interaction.

Figure 1.2: The leading-order diagrams of single top quark electroweak production in t- (left) and s- (right) channels.

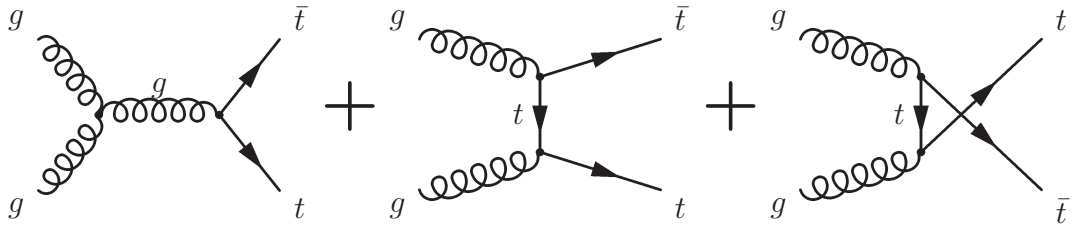
production mechanism are shown in Figure 1.2. At the Tevatron energies the dominant diagrams are the t-channel $W^{\pm}$ exchange and the s-channel $q\bar{q}$ annihilation. At the LHC the $gb \to tW$ diagram will contribute significantly [16, 17]. Although, predicted total single top cross section is not very small, the background reduction is much more complicated problem comparing to $t\bar{t}$-pair production. At the moment the single top signal was not observed. In this analysis we are focused on the strong production of the $t\bar{t}$-pair. In Figure 1.3 the leading order diagrams are shown for the processes dominated at Tevatron: the annihilation of quark and antiquark and gluon-gluon fusion. In order to produce two top quarks the total center-of-mass energy of interacting partons must be no less than two top quark masses: $\hat{s} \geqslant 2m_t$. From where for Tevatron energy $\sqrt{s} = 1.96$ GeV, it follows that typical values for momentum fractions x are:

$$x \sim \frac{2m_t}{\sqrt{s}} \approx 0.18. \tag{1.49}$$

From Figure 1.1 one can see that at values $x \sim 0.2$ the probability of finding up quark



(a) $q\bar{q}$ annihilation.



(b) gg fusion.

Figure 1.3: The leading-order QCD diagrams of top quark pair production as the result of quark-antiquark annihilation and gluon-gluon fusion.

in proton is approximately twice as much of finding down quark or a gluon. This roughly suggests that $q\bar{q}$ interaction should be at least five times more frequent than gg interaction. Indeed, more careful calculations predict the relative contributions to the top pair production from $q\bar{q}$ and gg to be about 85% and 15% respectively.

1.1.3 Top Quark Decay

The understanding of top quark decay is essential for determination of its properties. As any other coloured particle the top quark shortly after its production is allowed to form a baryonic or mesonic bound state. In QCD the time required for formation of colourless states is about 10^{-23} seconds. However, the top quark does not form any hadronic states simply because its lifetime is shorter (10^{-25} s) than hadronization time. The weak decay of the top quark is considered in details in Appendix A. The bottomline is that the partial width (the inverse of lifetime) of the top decay to any

other bottom-type quark q can be written as

$$\Gamma(t \to Wq) = |V_{tq}|^2 \Phi(m_q), \tag{1.50}$$

where the phase space factors $\Phi(m_q)$ for $q = d, s, b$ equal to each other with a good precision due to the fact that the masses of d, s, and b are much smaller than top mass. Also the CKM matrix element $|V_{tb}|$ is expected to be close to unity while $|V_{td}|$ and $|V_{ts}|$ values are close to zero suppressing the decays $t \to Wd$ and $t \to Ws$. Therefore the top quark is expected to decay to W and b nearly 100% of the time. Another consequence of the large top quark mass is that the W boson is produced on-shell, *i.e.* as a real particle.

Let us consider decay of top pair more closely. As the result of decay of each top quark in $t\bar{t}$ event, two oppositely charged W bosons are produced. In their turn the bosons will decay either hadronically ($W \to q\bar{q}$) or leptonically ($W \to l\bar{\nu}$). With a good approximation we can assume that the branching fraction of each individual channel is close to 11%. The hadronic decay channel includes six possible final states, since the W can decay only into quark pairs from the first and the second generations and each pair can appear in three different colour combinations ($R\bar{R}$, $G\bar{G}$, $B\bar{B}$). The leptonic decay has three final states corresponding to different lepton types (e, μ, τ). Hence, each W boson has nine final state degrees of freedom and the total number of final states for a pair of W bosons is $81 = 9 \times 9$.

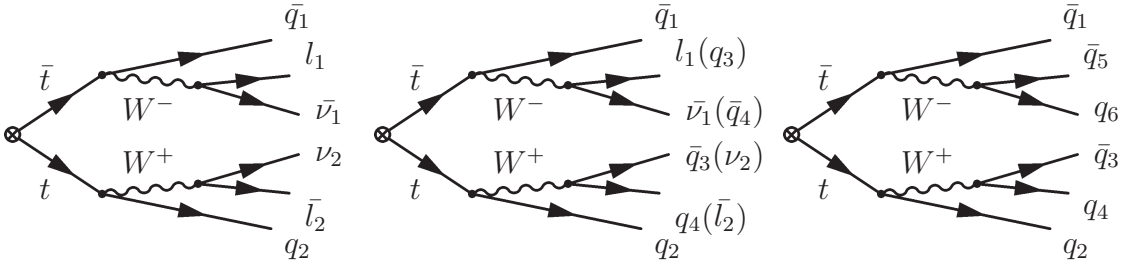
A special comment needs to be made on final states where one of or only lepton produced is τ. Unlike other leptons, tau decays in the detector into an electron or muon in about 36% of the time and in other cases tau decays hadronically forming τ-jets. The difference between the nominal $W \to l\bar{\nu}$ signature and the subsequent tau decay $W \to \tau\bar{\nu}_\tau \to l\bar{\nu}_l\nu_\tau\bar{\nu}_\tau$ will be indicated by larger missing E_T and a softer lepton. Also an additional effort is required to identify τ-jets, otherwise, they may be incorrectly assigned to quarks. Taking all this into account, it is advisable to treat the events with taus in intermediate states separately. Through the rest of this analysis

we will refer only to electrons and muons as leptons.

From experimental point of view it is convenient to combine the final states, in which $t\bar{t}$ pair can be observed, in the following three channels:

Dilepton channel In this channel both W bosons decay leptonically as shown in Figure 1.4a. These final states include lepton pairs e^+e^- (1/81), $\mu^+\mu^-$ (1/81), $e^\pm\mu^\mp$ (2/81), plus contributions from leptonic τ decays. Altogether the dilepton channel contributes in approximately 7% of all signal events. The final $t\bar{t}$ signature is recognized in events having highly energetic leptons, large imbalance in transverse energy due to undetectable neutrinos, and two jets presumably originated by b quarks. This decay mode is the cleanest one since the backgrounds are small and events with two energetic leptons can be triggered efficiently.

Lepton-plus-jets channel In this channel only one W boson decays leptonically whereas another one decays hadronically as shown in Figure 1.4b. The fraction of events with a single lepton is expected to be approximately 30% with equal contributions from events with electron (12/81) and muon (12/81) in the final state. Events with tau may contribute additional $\sim 8\%$ if τ-jets are not separated. Lepton-plus-jets events are characterized by having one high-P_T lepton, extensive missing E_T, and four energetic jets including at least two heavy-flavour



(a) Dilepton channel (b) Lepton-plus-jets channel (c) All-jets channel

Figure 1.4: Different final states of $t\bar{t}$-pair production and its subsequent decay. According to the final signature the events are divided into three classes.

jets. The signal events in this decay mode are more difficult to separate from the background than in dilepton mode, but overall more lepton-plus-jets events will be produced.

All-jets channel In this channel both W bosons decay hadronically as shown in Figure 1.4c. With only quarks produced in the final state we can have 36—out of 81 total—combinations distinguished by quark flavour and colour. The fraction of such events will be $\sim 44\%$, with additional contributions from hadronically decayed taus for up to $\sim 10\%$. The events in all-jets channel have 6 possibly separable jets in the final state. Such signature is not easily separable from multi-jet QCD backgrounds which dominate 2–3 orders of magnitude over the signal processes.

In this analysis we utilize the $t\bar{t}$ events in lepton-plus-jets and dilepton channels.

1.2 Analysis Problem and Strategy

This thesis presents a new measurement of the ratio R of top quark branching fractions $\mathcal{B}$ defined as

$$R \equiv \frac{\mathcal{B}(t \to Wb)}{\mathcal{B}(t \to Wq)} = \frac{\Gamma(t \to Wb)}{\sum_q \Gamma(t \to Wq)}, \tag{1.51}$$

where $\Gamma(\cdot)$ is the partial width of an appropriate top quark decay, and q can be any quark. If we assume that there are only three generations of quarks ($q = b, s, d$) then Equation (1.51) can be rewritten in terms of CKM matrix elements in the form

$$R = \frac{|V_{tb}|^2}{|V_{tb}|^2 + |V_{ts}|^2 + |V_{tb}|^2}, \tag{1.52}$$

where the phase space factors were canceled due to very small relative difference in their values (m_d, m_s, and $m_b \ll m_t$, see Appendix A). Further, from CKM matrix unitarity (see footnote on page 16) one can construct two identities:

$$|V_{tb}|^2 + |V_{ts}|^2 + |V_{tb}|^2 = 1, \tag{1.53}$$

$$|V_{ub}|^2 + |V_{cb}|^2 + |V_{tb}|^2 = 1. \tag{1.54}$$

The first equation suggests that the denominator in Equation (1.52) is just unity and from the second equation knowing the measured values of $|V_{ub}|$ and $|V_{cb}|$ one can predict a value for the numerator in the same Equation (1.52). The $|V_{ub}|$ and $|V_{cb}|$ are small and $|V_{tb}|$ is predicted to be close to unity: $0.9990 \leq |V_{tb}| \leq 0.9992$ at 90% confidence level (CL). Therefore, with the above assumptions the Standard Model top quark is expected to decay to a W and bottom quark at least 99.8% of the time at 90% confidence level.

In order to determine the ratio R experimentally, one need to measure how often the top quark decays to Wb versus Wq. Luckily there are a number of techniques allowing us to identify (tag) the b quarks in the final state. For this measurement one needs to count the number of $t\bar{t}$ events N_i^{obs} with different tag multiplicities i. Ideally,

the number of tags i can be 0, 1, or 2. On the other hand, the number of signal events with different tag rates can be predicted as a function of unknown quantity R:

$$N_i^{\rm exp} = \epsilon_i(R) \cdot N_{t\bar{t}}, \tag{1.55}$$

where $\epsilon_i(R)$ is the efficiency to have $i = 0$, 1, or 2 b quarks per event and $N_{t\bar{t}}$ is the total number of signal events.[3] Once $N_i^{\rm obs}$ and $N_i^{\rm exp}$ are determined, one can compare them through the use of likelihood function:

$$L(R) = L(N_i^{\rm obs}, N_i^{\rm exp}(\epsilon_i)). \tag{1.56}$$

The value of R where the likelihood reaches its maximum corresponds to the best fit between the experimental data and the theoretical prediction.

A first direct measurement of the ratio $\mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$ is reported in [18]. The measurement was performed in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ GeV using the Run I data collected by Collider Detector at Fermilab. During Run I (August 1992 to July 1995) CDF collected about 109 pb^{-1} of integrated luminosity. The value of R was determined by from both lepton-plus-jets and dilepton data sets with 163 and 9 events respectively. The reported central value for R is $0.94^{+0.26}_{-0.21}$(stat) $^{+0.17}_{-0.12}$(syst) and the lower limit is $R > 0.61$ (0.56) at 90% (95%) confidence level. The lower limit on R was obtained by a numerical integration of the likelihood function used to compare the number of observed and predicted tagged events.

Like the previous measurement of R the present one utilizes the lepton-plus-jets and dilepton data sets collected during Run II of CDF with the total integrated luminosity is ~ 162 pb^{-1}. Run II of the Tevatron provided $p\bar{p}$ collisions at higher energy ($\sqrt{s} = 1.96$ TeV) and luminosity. The cross section of top quark pair production increases in Run II by $\sim 30\%$ relative to the Run I. As a result of increased statistics and enhancements in the upgraded CDF detector, this measurement is expected

[3] In reality, the event tagging efficiencies ϵ_i are functions of R as well as other quantities. Here we do not show an explicit form of ϵ_i for the sake of brevity.

to yield better result than the previous one. Additionally, the systematic errors are reduced by more sensitive methods developed in this thesis.

This measurement of R is proceed in three basic steps:

1. We apply appropriate requirements on the data sample to identify the $t\bar{t}$ enriched events and estimate the background level in the samples as a function of the number of b-tagged jets in the event. Most importantly, in the lepton-plus-jets channel we utilize an Artificial Neural Network technique to obtain the background estimation.

2. We predict the expected tag rates in the $t\bar{t}$ events as a function of R. In our model we take into account the efficiencies to mistag the jets originated by quarks other than bottom quark. These efficiencies as well as event acceptances are determined from the Monte Carlo samples. The scale factor accounts for the difference in b-tagging efficiency measured in MC and data.

3. We compare the observed $t\bar{t}$ tag rates to the expectations, find the most likely value of R, and, finally, set a lower limit on R. Unlike Run I measurement we apply the Feldman-Cousins technique in order to set a lower limit on R.

A measurement of the ratio R significantly less than unity would contradict our theoretical assumptions, implying either a fourth generation of quarks, non-SM top decay, or a non-SM background to top pair production.

In the next chapter we present an overview of the Tevatron accelerator complex and CDF experiment. Event reconstruction and selection of signal enriched events is discussed in Chapter 3. Chapter 4 contains a brief introduction to the field of Artificial Neural Networks and their application to high energy physics experiment. The measurement of the ratio of top quark branching fractions is presented in Chapter 5. Conclusions about the results and future perspectives are made in the last

chapter. There are two additional appendixes for easy references from the main text. Appendix A contains formulae on top quark decay and Appendix B includes information related to likelihood method.

Chapter 2

Experimental Apparatus

Currently, the Fermilab Tevatron is the only accelerator capable of producing top quarks. There are two large, general-purpose detectors independently designed, constructed and installed at two interaction points on the Tevatron ring. In this analysis we use the data collected at the CDF experiment during Run II. This run started in 2001 after both detectors underwent significant improvements. The accelerator complex was also significantly upgraded, primarily for higher instantaneous luminosity but also for a small increase in the center of mass energy ($\sqrt{s}$) from 1.8 TeV to 1.96 TeV which increases the $t\bar{t}$ cross section by about 30%. The Tevatron will remain the highest energy collider until Large Hadron Collider at CERN starts operating in 2007 at center-of-mass energy $\sqrt{s} = 14$ TeV. In this chapter we will give an overview of the various accelerating machinery and a description of the CDF components.

2.1 Fermilab Accelerating Facilities

The Fermilab accelerating complex consists of several particle accelerators schematically shown in Figure 2.1. Various installations are used at different stages of the entire acceleration course, the purpose of which is the production of intense high energy proton and antiproton beams accelerated to the 980 GeV each.

There are two common types of accelerators used at Fermilab: Synchrotrons and linear accelerators. A synchrotron is a cyclic charged particle accelerator in which the magnetic field turns the particles so they circulate and the electric field accelerates the particles. Both fields are carefully synchronized with the travelling particle beams. By appropriately increasing the strength of the electric component the particles gain energy while travelling thousands of times through the accelerator ring. The magnetic field is adjusted in order to maintain a constant-radius path as the particles are accelerated. A detailed description of operation of synchrotron can be found in literature [19]. In linear accelerators the magnets bending the particle's path are not used, hence the paths of the accelerated particles are essentially straight lines rather than circles.

Conventionally, the whole acceleration chain at Fermilab can be split into the following three stages: (1) creation and preacceleration of protons, (2) creation and preacceleration of antiprotons, and (3) injection of the particle beams for the final acceleration in Tevatron.

Proton Source The creation of protons begins in the Cockroft-Walton chamber using hydrogen gas H_2. Negatively charged ions H^- are produced by electrical discharges in the volume of the chamber. The ions are extracted from the chamber by applying a positive voltage of 750 kV on one side of the dome. The preaccelerated ions are fed into the linear accelerator, Linac [20], where the particles are accelerated

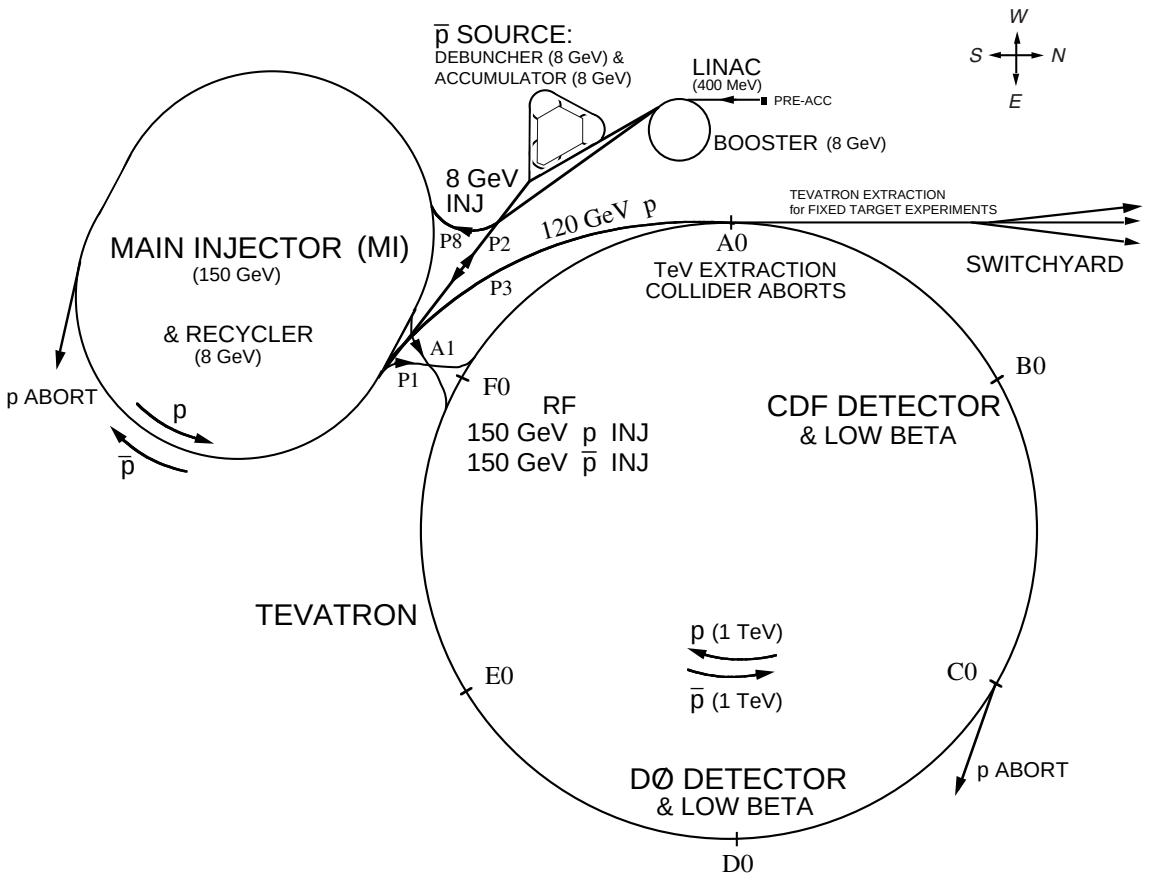


Figure 2.1: Schematic view of the Fermilab accelerator complex.

in RF cavities to 400 MeV. The Linac is 150 m long and it converts a continuous stream of particles into discrete packets (bunches). The H^- ions are transfered to the next accelerator called the Booster. The Booster is a synchrotron with a radius of 75 m. The Linac pulse length is about 10 times longer than the period of time needed for the protons to complete one circle in the Booster. In order not to disturb the protons that are already in the Booster, the ions are injected using one of the Booster's bending magnets, passing which the H^- ions and protons are merged in one beam. Then the particles are passed through a thin carbon foil, where the electrons are stripped off from the ions leaving only the protons. The proton bunches repeatedly pass through the circular array of RF cavities where they are accelerated

by electric component of the oscillating electromagnetic field. After several thousand acceleration cycles the protons acquire energy of 8 GeV. At the next stage the proton bunches are transfered into the Main Injector which is also a synchrotron, but which is seven times the circumference of the Booster. In the Main Injector the protons can be accelerated either to 150 GeV for the next final acceleration in the Tevatron or to 120 GeV for the antiproton production. In the former case the Linac and Booster provide the beam with typically 6×10^{10} protons per bunch, in the latter case they deliver a beam with up to 5×10^{12} protons per bunch.

Antiproton Source Two small synchrotrons, the Debuncher and the Accumulator, are referred to as the antiproton source. Protons with energy of 120 GeV are extracted from the Main Injector and transferred to the target area where they are used to bombard a nickel target. As the result of proton scattering on the target, a spray of various particles is produced at many different angles. These particles are focused with the Lithium lens and then, the antiprotons are separated from other particles by a magnetic charge-mass spectrometer. These antiprotons are moved to the Debuncher, where the spread in the particle momenta is reduced by stochastic cooling process. The Debuncher has a triangle shape with rounded corners. The beam turning magnets are installed in the corners on a circle with a radius of 90 m. The 8 GeV, almost continuous in time (debunched), beam of antiprotons is injected into the Accumulator whose purpose is to accumulate the antiprotons. The Accumulator is also a triangular-shaped synchrotron with 75-meter radius, and is housed in the same tunnel as the Debuncher. Here the beam can be stored for many hours, collecting antiprotons from the Debuncher in a process called *stacking*. Also, a much more thorough beam cooling is performed. The Accumulator can store up to 140×10^{10} antiprotons and whether typical stacking rates vary from 6×10^{10} to 11×10^{10} protons per hour. When the intensity of antiproton beam reaches its maximum, the beam is sent back to the Main Injector.

Main Injector From the above discussion and the name itself it is clear that the main purpose of the Main Injector is to prepare the protons and antiprotons for their injection into the Tevatron for the final acceleration. The Main Injector was designed and built specifically for Run II. Before that the role of injector had been played by Main Ring, which was the original 400 GeV proton synchrotron built during the early 1970's. More specifically, the Main Injector is used for the following tasks:

- accept the 8 GeV protons from the Booster and antiprotons from the Accumulator;
- produce 120 GeV protons used for the antiprotons production;
- accelerate protons and antiprotons to 150 GeV and inject them into the Tevatron.

Tevatron The largest accelerator at Fermilab is the Tevatron, a synchrotron collider. After accepting 150 GeV protons and antiprotons from the Main Injector, it is capable of accelerating particles to the energy of 980 GeV. Thirty-six bunches each of protons and antiprotons circulate in opposite directions in the Tevatron. There are two points along the ring designated as B0 and D0 (see Figure 2.1), where the beams are focused and brought into collision. The results of such collisions are observed by two multi-purpose detectors installed at these points: The Collider Detector at Fermilab (CDF) and the DØ Detector. The Tevatron is installed inside a tunnel of exactly 2 km in diameter. A very strong magnetic field of 4.2 T is required to turn particles with such high energies. Superconducting magnets are used in order to create the magnetic field. The magnets are made of niobium-titanium alloy and cryogenically cooled to about 4.6 K with liquid Helium.

The process of loading protons and antiprotons in the Tevatron is called the *shot setup* and it begins when the size of the antiproton stack in the Accumulator is large

enough. The proton bunches are injected first by the transfer system with a fast kicker magnet having 396 ns rise time. The 36 bunches are injected in 3 *trains* of 12 coalesced bunches with an abort gap between each train of 2.617 μs. In fact, a single proton bunch in the Tevatron is a product of merging of several bunches prepared in the Booster providing an increased intensity of $\sim 30 \times 10^{10}$ protons per bunch. The antiproton bunches are injected similarly to the protons when either of the three abort gaps in the proton beam passes the antiproton kicker magnet. Like with protons, each bunch is a merging of several antiproton bunches prepared in the Accumulator resulting in $\sim 30 \times 10^{9}$ antiparticles per bunch.

Once the Tevatron loading is complete, beams are accelerated to the maximum energy and a stable configuration is reached the data taking period (the *store*) begins. During this period the luminosity decreases due to loss of particles in $p\bar{p}$ collisions and beam-gas interactions. Also, the beams start to heat up and their effective width increases due to long range beam-beam interactions. The instantaneous luminosity is given by the formula

$$N = \sigma\mathcal{L} = \sigma\frac{f N_b N_p N_{\bar{p}}}{2\pi\sqrt{s_p^2 + s_{\bar{p}}^2}}, \tag{2.1}$$

where σ is effective cross section of a particular process, f is the bunches revolution frequency, N_b is the number of bunches in the beam, N_p and $N_{\bar{p}}$ are the number of protons and antiprotons in each bunch respectively, s_p and $s_{\bar{p}}$ are the transverse sizes of the proton and antiproton beams at the interaction point respectively. From the above relation one can see that a desired gain in the luminosity is achievable either by increasing the number of (anti)protons or by reducing the effective beam sizes. If the cross section of specific interaction is σ then the interaction rate $\mathcal{R}$ is proportional to the instantaneous luminosity $\mathcal{L}$

$$\mathcal{R} = \sigma\mathcal{L}. \tag{2.2}$$

While detectors are collecting data, a new stacking period begins in the Accumulator preparing antiprotons for the next store. A typical store lasts for 15 hours. After

that time it is more efficient to abort the current store and start a new cycle with fresh beams. The period of time between two stores is about 2 hours and it can be used for detector calibrations.

2.2 Collider Detector at Fermilab

2.2.1 Overview and Coordinate System

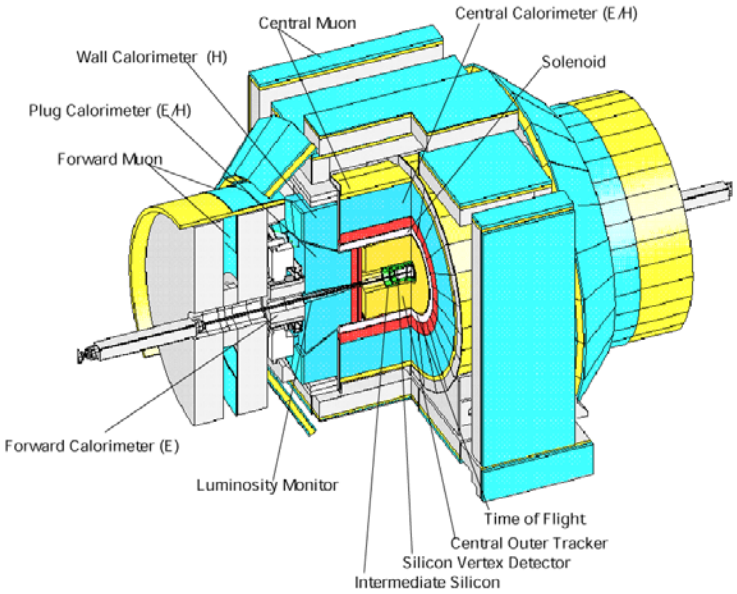
CDF is a magnetic detector designed to observe a wide range of physics phenomena in high-energy $p\bar{p}$ collisions. A general view of CDF is shown in Figure 2.2a and Figure 2.2b provides a more detailed cross section view. The detector has undergone significant upgrades for Run II [21, 22]. The upgrade was necessitated by the increased Tevatron luminosity and shorter bunch-crossing time. For improved momentum resolution, b-tagging, electron and muon identification in Run II some old subdetectors were replaced and new hardware was added to the CDF.

This analysis uses all of the major detector components:

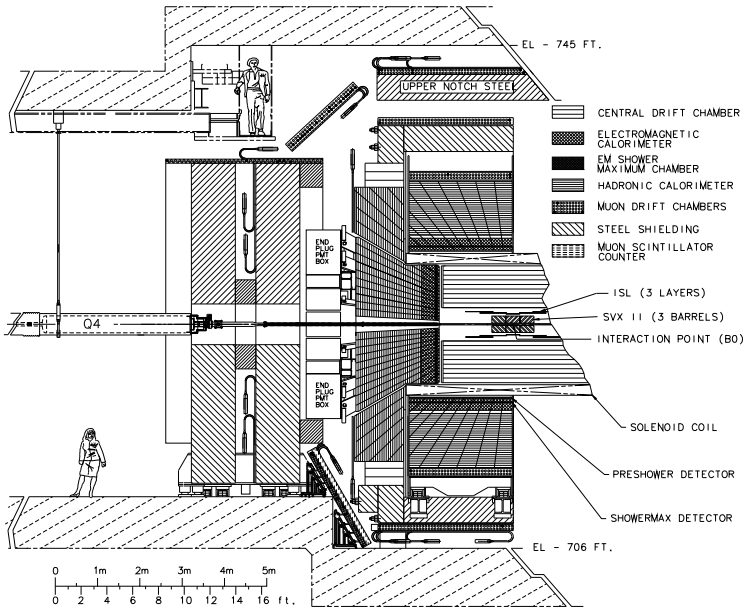
The tracking system is the innermost part of CDF used for particles charge and momentum measurements. This system is placed inside the superconducting solenoid of 4.8 m long and 1.5 m radius which produces a 1.4 T magnetic field coaxial with the beam direction.

The calorimeter system is composed of electromagnetic (EM) and hadronic (HAD) scintillator-based sampling calorimeters surrounding the solenoid in a projective-tower geometry. The EM calorimeter features an embedded proportional chamber at shower maximum for precise electron identification and reconstruction.

The muon system is located on the exterior of CDF. Only high energy muons are



(a) Isometric view of the CDF with cut-away quarter.



(b) Elevation view of one half of the CDF.

Figure 2.2: Collider Detector at Fermilab. The detector has both azimuthal symmetry with respect to the beamline and forward-backward symmetry with respect to the nominal collision point.

able to pass the absorption layers of steel beyond which the muon detectors are located.

A Cartesian coordinate system is chosen with origin at the geometrical center of the detector. The x axis is defined in the direction radially outward from the accelerator center, the y axis is defined perpendicular to the accelerator plane pointing vertically upward, and the z axis is along the antiproton beam direction. Because of the cylindrical shape of the detector it is more convenient to identify particle directions and other locations within the detector in a cylindrical-like coordinate system. The three coordinates are (z, ϕ, η), where ϕ is the regular azimuthal angle in x–y plane and η is the pseudorapidity expressed in terms of the polar angle θ as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{2.3}$$

The pseudorapidity is chosen as a coordinate because it is closely related to the true rapidity y defined as

$$y = \frac{1}{2}\ln\left(\frac{E - p_z}{E + p_z}\right), \tag{2.4}$$

where E is energy and p_z is the z component of the momentum of the particle. The rapidity shows how much the particle is boosted along the z axis. The differences in rapidity are Lorentz invariant, because under Lorentz transformation to a frame moving along z direction with velocity β relative to the lab frame, the rapidity modifies as an additive quantity:

$$y \to y - \tanh^{-1}\beta, \tag{2.5}$$

For highly relativistic particles ($|\vec{p}| \gg m$) pseudorapidity is a good approximation to rapidity y: $\eta \approx y$. The consequence of $\Delta\eta$ invariance under Lorentz transformation along z axis is that the shape of the pseudorapidity distribution $\frac{dN}{d\eta}$ also remains invariant. This is the main reason why the CDF calorimeter is segmented in equal units of pseudo-rapidity. Another important Lorentz invariant quantity is the measure

of an opening angle between two particles given by

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{2.6}$$

2.2.2 Tracking System

The ionization signal due to a charged particle traversing the tracking medium can be collected on a localized electrode, thereby providing position information, is called a *hit*. By detecting many discrete hits along the path the particle's trajectory can be reconstructed in a process know as *tracking*. The CDF tracking system consists of two major detectors: the silicon detector at the small radii, and an open cell drift chamber (COT) at larger radii surrounding the silicon detector. Both detectors are submerged into a uniform magnetic field $B = 1.4$ Tesla along z axis created by the solenoid.

The trajectories of charged particles inside the solenoid are helices, which are described by the five parameters:

z_0 – z coordinate of the point of the closest approach to the z axis;

$d_0 = q(\sqrt{x_c^2 + y_c^2} - R)$ – distance from the point of closest approach to the z axis, impact parameter. q is the charge of the particle and (x_c, y_c) is the center of the helix projected onto the x–y plane;

ϕ_0 – azimuthal angle of particles transverse momentum at the point of closest approach to the z axis;

$\cot\theta$ – ratio of helix step to its diameter, helix pitch;

$C = \frac{q}{2R}$ – helix curvature.

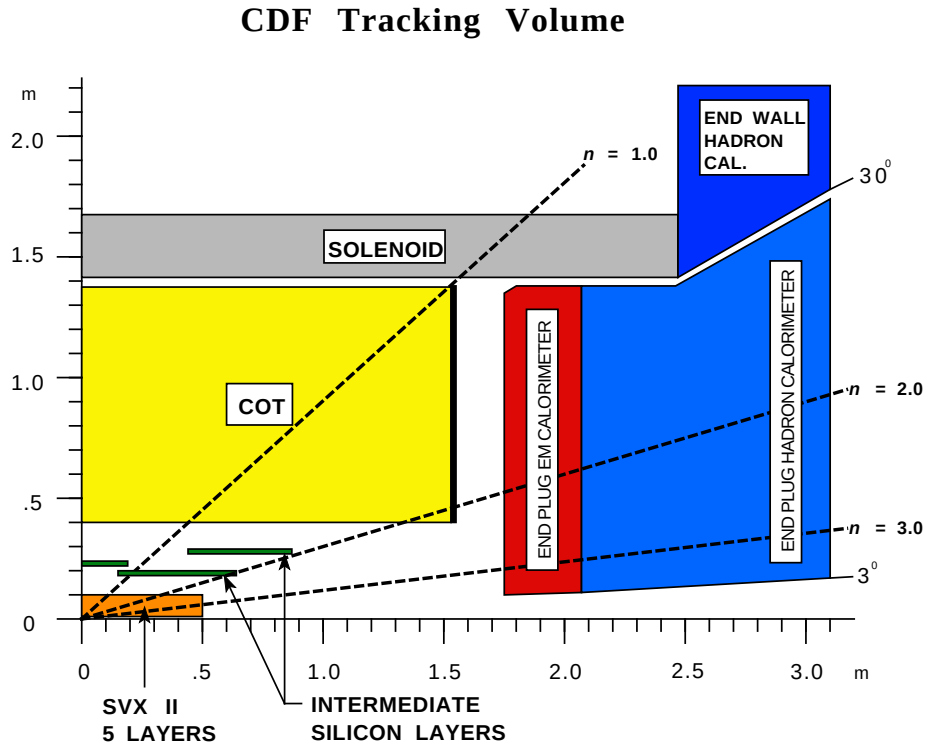


Figure 2.3: Cutaway view of one quadrant of the CDF tracking system showing the COT and silicon detectors surrounded by solenoid and end-cap calorimeters.

The above parameters can be used to calculate the particle's transverse and longitudinal momenta:

$$p_T = \frac{cB}{2|C|}, \tag{2.7}$$

$$p_z = p_T \cot\theta. \tag{2.8}$$

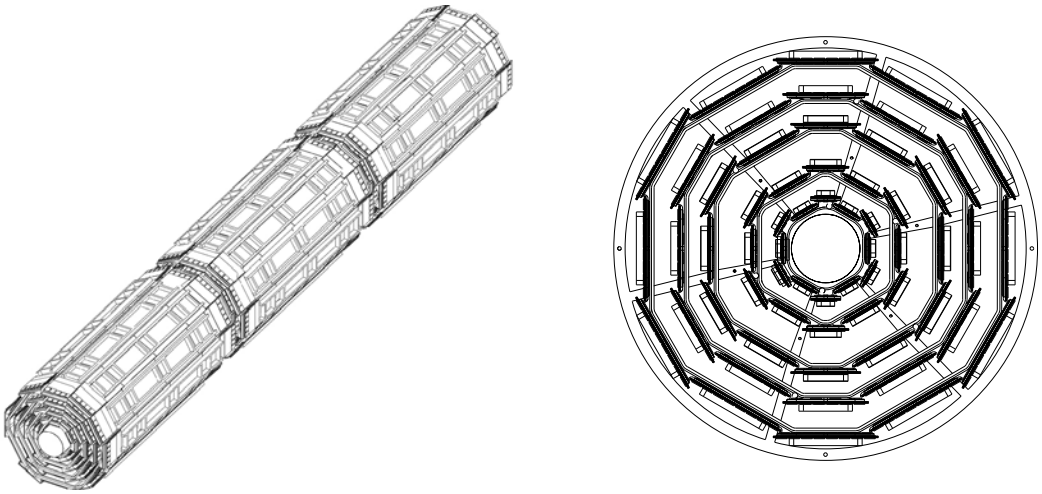
The track parameters are determined by the helical track fit. The input to this fitter is a set of spacial coordinates of the measured hits which specially preselected with the pattern recognition algorithm. The primary vertex of event is determined from the reconstructed tracks in the event by fitting them to the same space point and using the beamline position as a constrain. Tracks with hits in the silicon detector are reconstructed with high-precision and can be used to reconstruct secondary verticies. The schematic side view of one quarter of tracking system is shown in Figure 2.3. The parameters of each component of tracking system will be briefly described below.

Silicon Detector

Many analyses at CDF require a precise measurement of the decay length of long-lived B hadrons. The silicon detector is comprised of eight concentric cylindrical layers of silicon microstrip sensors surrounding the beam-pipe. The layers span radii from 1.35 cm to 28 cm and extend from 90 cm to almost two meters along the Z direction. The detector is divided into three sub-systems: Layer-00, SVXII, and ISL. Each component has a cylindrical form with an axis coinciding with the beam direction.

The Silicon Vertex Detector (SVXII) is the primary component of the silicon tracking system, consisting of 5 layers of double-sided microstrip detectors. The most inner and outer layers have the radii of 2.5 cm and 10.7 cm respectively. Along the beam pipeline the SVXII covers 2.5σ of the $p\bar{p}$ luminous region, corresponding to the length of 96 cm, and in pseudorapidity it covers $\eta < 2.5$ (see Figure 2.3). In the axial direction, the SVXII is divided into three barrels each 29 cm long with beryllium bulkheads at each end. The bulkheads support the five silicon layers and the cooling system for readout electronics. In the plane transverse to the beamline, each layer is segmented in 12 azimuthal wedges as shown in Figure 2.4.

In order to reduce residual interactions, the silicon sensors are supported by light-weight substrates forming a basic structural unit called *ladder*. On one side of the silicon sensors the finely spaced p-type silicon strips run in the axial direction allowing the measurement of the hits in r–ϕ plane. On the other side of silicon sensors, n-type silicon strips are rotated forming either 90° (0^{th}, 1^{st}, and 3^{rd} layers) or $\pm 1.2^\circ$ (2^{nd} and 4^{th} layers) stereo angles allowing the measurement of r–z. The number of strips per ladder in each layer along with other mechanical parameters are shown in Table 2.1. The signal is detected on a small cluster of non-stereo (r–ϕ) and stereo strips and the hit position is determined by the appropriate strip coordinates. The accuracy of



(a) Isometric view of the three-barrel structure of SVX.

(b) End view of SVXII bulkhead supporting 5 layers divided into 12 wedges.

Figure 2.4: CDF Silicon Vertex Detector.

individual hit position measurement is about 12 μm.

The readout units are mounted directly on silicon surface on both ends of each ladder. Such construction helps to improve readout speed and minimize electronic noise. Each wedge has 44 readout chip sets and each chip has 128 channels with charge-sensitive amplifiers. The total number of readout channels for six barrel ends is 405,504. The signals are converted by special port cards from electrical to optical

Table 2.1: SVX layers technical summary.

Layer	**Radius**, cm		**# of strips**		**Strip pitch**, μm		**Stereo angle**	**Ladder active width**, mm
	Stereo	***r*–ϕ**	**Stereo**	***r*–ϕ**	**Stereo**	***r*–ϕ**		
0	2.55	3.00	256	256	60	141	90°	15.30
1	4.12	4.57	576	384	62	125.5	90°	23.75
2	5.52	7.02	640	640	60	60	+1.2°	38.34
3	8.22	2.72	512	768	60	141	90°	46.02
4	10.10	10.65	896	896	65	65	−1.2°	58.18

form, and transmitted to the external electronics at 53 MHz rate. The highly parallel readout permits the entire detector to be read in approximately 10 μm.

In order to further improve the impact parameter resolution, Layer-00 ("Layer zero zero") is mounted directly on the beam pipe at a radius of 1.6 cm. The length of Layer-00 is 80 cm and its azimuthal structure is similar to the SVXII (Figure 2.5). Layer-00 utilizes radiation-tolerant single-sided microstrip sensors. The narrow sensors have 128 channel and wider sensors have 256 channels with a strip pitch of 50 μm. To

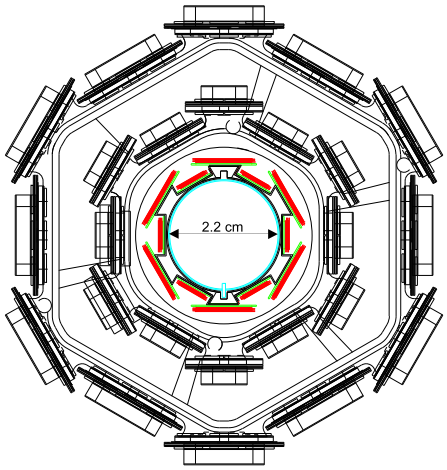


Figure 2.5: Transverse view of Layer 00 (innermost) inside the first two inner layers of SVXII.

minimize the radiation exposure of the electronics, Layer-00 is readout via finely pitched kapton cables connected to the readout chips mounted outside the tracking volume at $|z| > 40$ cm.

The ISL is placed outside the SVXII as can be seen in Figure 2.3. There is one cylindrical layer installed in the central region $|\eta| < 1.0$ at radius of 22 cm. Two other layers are installed at radii of 20 cm and 29 cm, covering the pseudorapidity region $1.0 < |\eta| < 2.0$. Double-sided silicon sensors are used in ISL ladders. On axial side, the silicon microstrips are spaced by 55 μm and on the stereo side by 73 μm with 1.2° stereo angle. In order to reduce the total number of channels (268,800) only every

other strip is read out in ISL ladders. Hence, the single hit resolution is somewhat worse than in SVXII, achieving 16 and 23 μm on axial and stereo sides respectively. In general the design of ISL ladders is similar to the one of the SVXII ladders. Extra hits provided by ISL are very important for track reconstruction, providing intermediate hits between the SVXII and the Central Outer Tracker (COT, described below).

Overall, the silicon system measures the impact parameter of high p_T tracks with a resolution of about 40 μm, including 25–30 μm contribution from the beam width. The long-term performance of the device depends on its resistance to radiation damage. The innermost layer of SVX is the most vulnerable component of the silicon system and it is expected to last at least 7.4 fb^{-1}. More information on CDF silicon detector can be found in [23, 24, 25].

Central Drift Chamber

The Central Outer Tracker (COT) is a cylindrical, multi-wire open-cell drift chamber designed to find tracks of charged particles in the central region $\eta < 1.0$. As shown in Figure 2.3 the COT is located outside the silicon detectors and inside the solenoid, which provides the magnetic field of 1.4 T. Along z axis the tracker covers the region of $|z| < 155$ cm and in radial direction its closed internal volume is between radii of 44 cm and 132 cm. The COT is hermetically sealed and filled with fast gas mixture of 50:50 Argonne-Ethane, which gives a drift velocity of about 50 μm/ns. Small drift times are essential in order to prevent the subsequent events to pile-up at high collision rates.

As charged particles pass through the COT they leave a trail of ionization in the active medium. The electrons drift to sense wires strung between the COT end plates. The sense wires are placed in electrostatic drift field created by field-shaping wires and cathode field panels, both running parallel to the sense wires. The COT cell

consists of 12 sense wires alternating with 13 potential wires as shown in Figure 2.6. Two adjacent cells share the same field panel; and four shaper wires (two at each

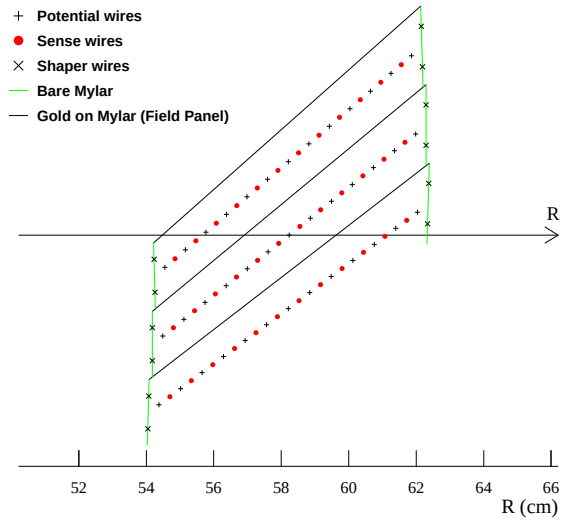


Figure 2.6: Cross section of three COT cells in second superlayer. The R-arrow shows the radial direction. Electrostatic field is perpendicular to the field panels and drift velocity is perpendicular to the radius.

side) serve to close the cell electrostatically. Because of the magnetic field in the z direction, the ionization electrons do not move along the electric field lines but rather cross them at an angle α known as the Lorentz angle. The value of α depends on the strengths of electric and magnetic fields, and for COT it is $\simeq 35°$. Therefore, the COT cells are rotated by α with respect to the radial direction in order to reduce the drift length.

The field panels are kept at ground potential, while the potential and sense wires are operated under potentials of approximately 2 and 3 kV respectively. Within a cell the actual voltages are slightly varied from wire to wire in order to maintain a uniform drift field.

There are total 30,240 sense wires in COT. They are arranged in eight radially spaced superlayers as shown in Figure 2.7. Since there are 12 sense wires in one cell

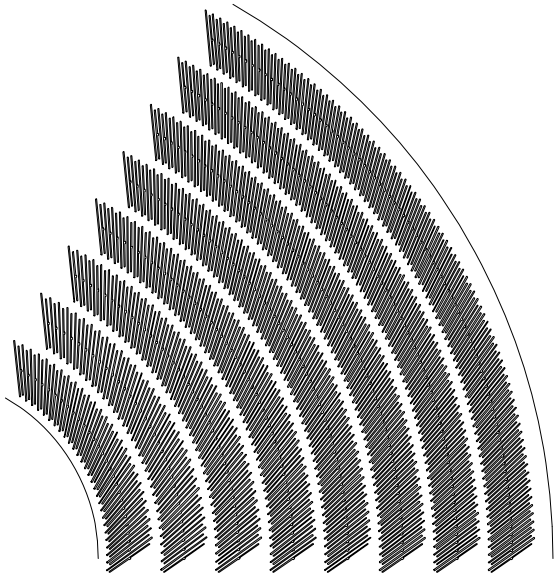


Figure 2.7: End view of a one sixth of COT endplate. The wire-plane slots are grouped into eight concentric super-layers. The slots are tilted by 35° with respect to radial direction.

the total number of possible measurements per track traversing the COT radially is 96. The even layers are non-stereo, *i.e.* wire plane is parallel to the z axis, and odd layers have small stereo angle of $\pm 3°$ which is achieved by connecting the wire slots with eight slot offset on opposite endplates. The number of cells and their radii are summarized in Table 2.2.

Table 2.2: COT layers summary.

Superlayer	**Radius at center**, cm	**# of cells**	**Stereo angle**
1	46	168	$+3°$
2	58	192	$0°$
3	70	240	$-3°$
4	82	288	$0°$
5	94	336	$+3°$
6	106	384	$0°$
7	119	432	$-3°$
8	131	480	$0°$

The signals from sense wires are readout by electronics directly placed on the chamber endcaps. Specially designed pattern recognition logic in the Level-1 trigger is used to recognize multiple hits from single wire. The resolution of a single hit is measured to be about 140 μm, which corresponds to the transverse momentum resolution of $\delta p_T/p_T^2 \sim 0.15\%$ $(\mathrm{GeV}/c)^{-1}$. In absolute magnitude the tracks can be measured down to 300 MeV/c. If information from COT is combined with the one from silicon detectors, then the transverse momentum resolution is: $\delta p_T/p_T^2 \lesssim 0.10\%$ $(\mathrm{GeV}/c)^{-1}$. More information on CDF silicon detector can be found in [26].

2.2.3 Calorimeter System

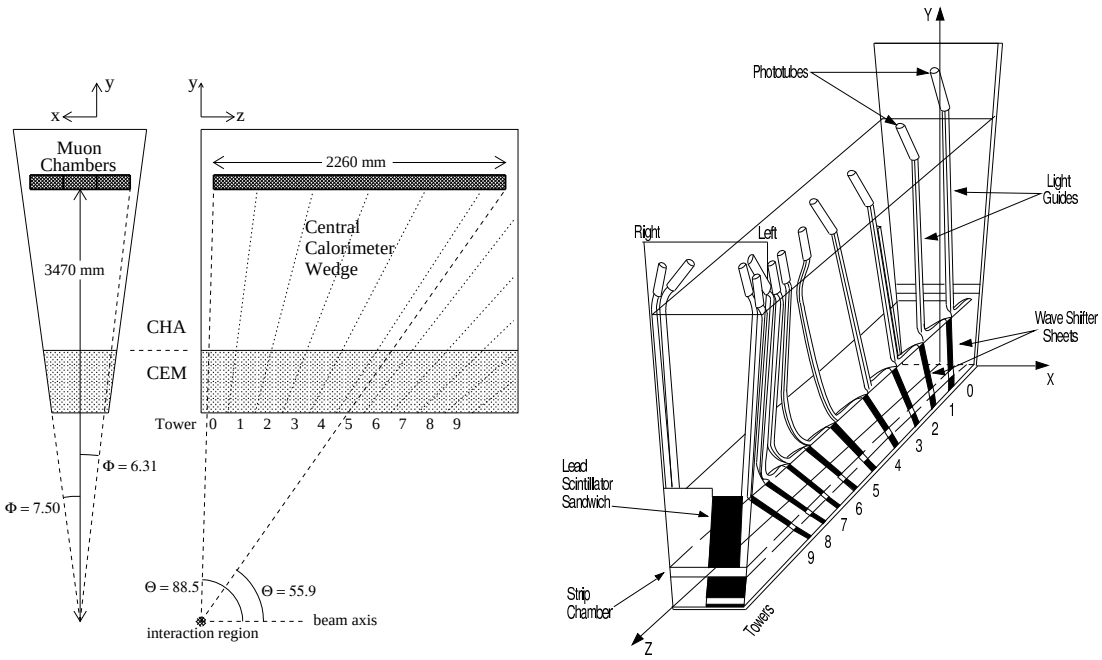
The purpose of the CDF calorimeter system is to measure the energies of the charged and neutral particles produced in $p\bar{p}$ collisions. The absorption and scintillation materials are the main components of each calorimeter. As particle traverse the calorimeter it looses its energy by interacting with the dense absorption material and producing a shower of secondary particles [27]. The number of charged secondary particles is proportional to the energy of incident particle. A scintillator material is used to collect these charged particles and release an equivalent amount of energy by means of photons. The light from scintillators is read out by light guides and converted into electric signal using photomultipliers (PMT) and amplifiers. The strength and position of the signals are used in event reconstruction. This information combined with tracking information leads to better identification of the original particle.

The resolution on energy measurement by a calorimeter, $\frac{\sigma_E}{E}$, is dominated by statistical sampling fluctuations and the photo-statistics of the PMTs, which is inversely proportional to the square root of the incident energy. Other contributions to the total resolution come from the non-uniform response of the calorimeter, calibration errors, and electronics noise. These factors are taken into account by adding smaller

constant term in quadrature.

The calorimeter detectors are installed just outside the solenoid providing 2π azimuthal coverage out to $|\eta| < 3.6$ (Figure 2.2a). The classification of different calorimeter portions is based on the coverage in pseudorapidity regions: The central calorimeter provides coverage in the region where $|\eta| < 1.1$ and the plug calorimeter covers the pseudorapidity over $1.1 < |\eta| < 3.6$ corresponding to polar angles between 37° and 3°. The gap between central and plug regions is closed by endwall calorimeter. Both central and plug calorimeters consist of two sections: Electromagnetic (EM) and Hadronic (HAD), while the endwall calorimeter has only hadronic section. The EM section is located on the inner side of the calorimeter, closer to the global detector center, and HAD section is placed further in radial direction. Such placement is conditioned by the different penetration lengths for the electrons/photons and hadrons. The high energy electrons lose their energy through the bremsstrahlung processes or photon radiation, whereas the photons' primary source of energy loss is the production of electon-positron pairs. In either case the showers are produced via electromagnetic processes. In case of hadrons their interaction with matter is more complicated, since they can undergo electroweak and strong interactions with the latter prevailing. The hadrons at CDF lose their energy mostly through inelastic collisions with the nuclei of the absorbing medium. The cross section of nuclear interactions is smaller than electromagnetic, therefore hadrons travel greater distances before losing all of their energy. It is convenient to define characteristic penetration depth for EM and HAD segments independent of specific material they are made of. For electrons and photons we define the *radiation length* X_0 as the distance at which particle loses on average $1 - e^{-1}$ of its energy. The nuclear *interaction length* λ_I is defined as a mean free path a hadron can go before inelastic scattering happens.

The central calorimeters consist of 24 azimuthal wedges just like the one shown in Figure 2.8. Each wedge covers an azimuthal angle of 15° and 250 cm along the beam

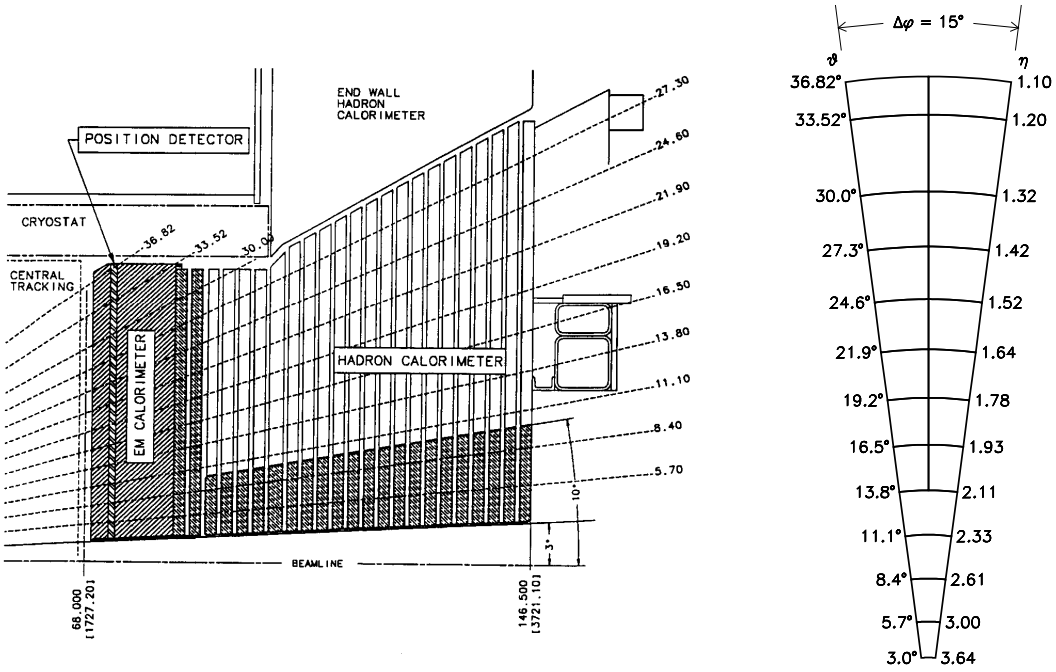


(a) Front and end view with designated dimensions.

(b) Isometric view.

Figure 2.8: Central calorimeter wedge with EM and HAD sections. 10 towers in pseudorapidity are shown.

axis in both positive and negative directions. In radial direction the central electromagnetic calorimeter (CEM) section starts at 173 cm and is 35 cm thick, which corresponds to 18 radiation lengths. In the outer part of the wedge the central hadronic calorimeter (CHA) is placed next to the CEM. Each wedge is divided along pseudorapidity forming projective towers pointing back to the nominal interaction point. A tower covers $\Delta\eta = 0.11$. The CEM calorimeter is made of 0.5 cm thick plastic scintillator plates interspersed with 0.32 cm thick lead layers as an absorber. Each scintillator layer of each wedge is readout by two PMTs. The resolution on energy measurement of a single electron or photon at normal incidence is $\frac{\sigma_E}{E} = 13.5\%/\sqrt{E_T} \oplus 2\%$, with the constant term added in quadrature. In the CHA calorimeter the scintillator layers are 1 cm thick and the absorption layers made of steal are 2.5 cm thick. The wall hadronic calorimeter (WHA) has similar to CHA active layers but the steal absorber



(a) Cross section of an upper part of the plug calorimeter.

(b) Tower segmentation.

Figure 2.9: Plug calorimeter illustration.

is even thicker (5 cm), which corresponds to the fact that the total energy deposited in this calorimeter is on average a factor $\sqrt{2}$ larger than in the central one for the particles with the same transverse energy. The width of CHA and WHA in terms of λ_I and their resolutions are given in Table 2.3.

The plug calorimeter is divided into concentric η regions, which in turn segmented in 24 for $|\eta| < 2.11$ or 12 for $|\eta| < 2.11$ projective towers (see Figure 2.9). The electromagnetic section of the plug calorimeter (PEM) consists of 23 absorber-scintillator layers. A calcium-tin-lead alloy enclosed between steel plates serves as an absorber. The plug hadronic calorimeter (PHA) follows the PEM section and is comprised of 23 unit layers of 5 cm iron and 0.6 cm scintillator. The thickness and the resolution of the plug calorimeter sections is presented in Table 2.3.

Table 2.3: Summary of the CDF calorimeter subsystem properties. The energy resolution is given for a single incident electron/photon for EM section and pion for HAD section. See text for radiation (X_0) and interaction (λ_I) lengths.

Calorimeter Subsystem	**η coverage**	**Depth**	**Resolution $\frac{\sigma_E}{E}$**
CEM	$\lvert\eta\rvert < 1.1$	$18X_0$	$13.5\%/\sqrt{E_T} \oplus 2\%$
PEM	$1.1 < \lvert\eta\rvert < 3.6$	$21X_0$	$16\%/\sqrt{E_T} \oplus 1\%$
CHA	$\lvert\eta\rvert < 0.9$	$4.5\lambda_I$	$50\%/\sqrt{E_T} \oplus 3\%$
WHA	$0.7 < \lvert\eta\rvert < 1.3$	$4.5\lambda_I$	$50\%/\sqrt{E_T} \oplus 3\%$
PHA	$1.2 < \lvert\eta\rvert < 3.6$	$7\lambda_I$	$80\%/\sqrt{E_T} \oplus 5\%$

A detailed description of EM and HAD calorimeters and changes related to Run II upgrade can be found in [28, 29].

2.2.4 Muon System

The muons produced in $p\bar{p}$ collisions and escaped the magnetic field do not interact heavily with the calorimeters. At such high energies the muons undergo multiple Coulomb scatterings in the calorimeter without developing a shower. The energy of a muon must be higher by a factor of $\sim \left(\frac{m_\mu}{m_e}\right)^2$, where m_e and m_μ are the muon and the electron masses respectively, for the radiation energy losses become dominant. The detection of muons is performed with scintillator and proportional drift chambers installed on the outermost part of the detector. The muon system is composed of four subdetectors covering the following regions in pseudorapidity: The Central Muon (CMU) and the Central Muon Upgrade (CMP) chambers cover the region where $|\eta| < 0.6$, the Central Muon Extension (CMX) chambers cover the pseudorapidity in the range of $0.6 < |\eta| < 1.0$, and the Intermediate Muon (IMU) detector extend the coverage further in the forward and backward regions for up to $|\eta| \approx 2$. The muon detectors do not provide full 2π coverage in ϕ. The CMU and CMP cover 84% and 63% of the solid angle respectively. Due to CMU and CMP overlap, 53% of the solid angle is covered by both of them. The complete coverage in ϕ–η of muon system is

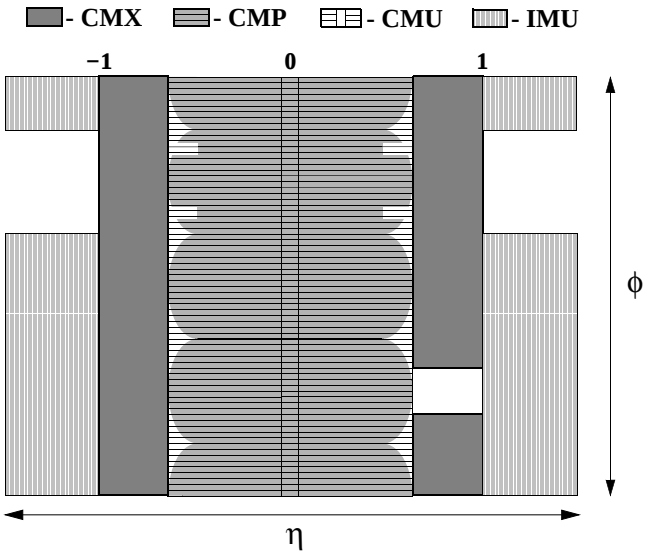


Figure 2.10: Coverage map of four muon subsystems in ϕ–η plane.

shown in Figure 2.10.

The CMU detectors are located within the central calorimeter wedges, directly behind the CHA at the radius of 347 cm (see Figure 2.8). The material in the return yoke of magnet and in the calorimeter serve as an absorber for hadrons protecting the muon detectors. The amount of the material from the detector's center to the CMU is equivalent to a 5.5 hadron interaction lengths, causing the muon rangeout threshold to be about 1.4 GeV/c. Like CEM and CHA the CMU is spit into 24 azimuthal wedges but the instrumented part of the wedge is only 12.6° leaving 2.4° gaps between chambers. Each wedge contains three 4.2° towers with four radial layers of four rectangular drift cells forming a CMU chamber. There is a sense wire at the center of each drift cell running from one end-plate of the cell to the opposite one. The length of the cells and the wires is 226 cm. The wires in the cells of the first and the third layers lie in the plane passing through the z axis, while in the second and fourth layers the wires are shifted by 2 mm with respect to center of the cells in

order to provide unambiguous ϕ measurement of the muon track. Single hit position can be measured with resolution of 250 μm in the r–ϕ plane and 1.2 mm in the r–z plane. The z coordinate is determined by comparing the pulse heights at each end of the sense wires. If the muon has hits on at least 3 of the 4 layers, a muon *stub* is reconstructed.

The CMP chambers form a rectangular "box" around the central part of the CDF. In order to further reduce the number of hadrons entering the CMP detector, an additional layer of 60 cm of steel is placed in front of the CMP providing extra $2.3\lambda_I$ and raising the muon momentum threshold to 2.2 GeV/c. The drift cells composing the CMP are similar to the ones used in the CMU and they are also arranged in four layers with alternate half-cell staggering. The inner and outer surfaces of the CMP is lined with scintillator counters (CSP), which are used to provide timing information. Because of the rectangular shape CMP pseudorapidity coverage varies as a function of ϕ, nevertheless at some values of ϕ the CMP compensates for CMU gaps.

The muon coverage is extended by additional muon chambers composing the CMX and IMU detectors. The CMX is comprised of two conical arches of drift tubes with layers of scintillators (CSX) installed at each end of the central part of the detector. No additional absorption material is added before CMX, since at these polar angles (45°–55°) a particle will traverse through enough detector material. The IMU consists of a barrel of drift chambers and scintillator counters placed on both sides of toroid magnets. The IMU drift chambers, scintillators, and read out electronics are identical to the ones used in CMU. The information from IMU is not used in this analysis because there is no trigger associated with it.

More information about muon detectors can be found in [30, 31].

2.2.5 Shower Detectors

The electromagnetic showers are very narrow in transverse direction. The Moliere radius, within which approximately 90% of the total energy of a EM shower is deposited, has typical values of ~ 2 cm in the central calorimeter. In order to enable a precise measurement of the shower position, highly segmented detectors are embedded in each tower of central (CES) and plug (PES) electromagnetic calorimeters [32]. These detectors are placed at a depth of about 6 radiation lengths from the front face of the CEM and PEM detectors corresponding to the average shower (particle multiplicity) maximum.

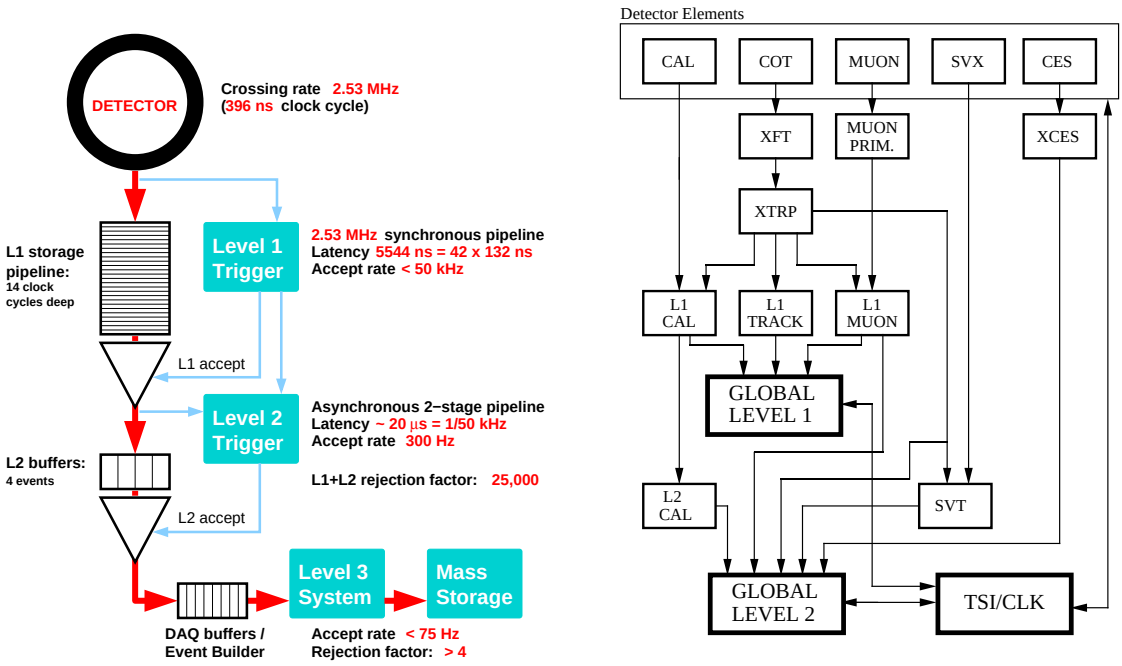
The CES detector consists of proportional strip and wire chambers providing the position and transverse shape of the electromagnetic showers in both z and r–ϕ directions. The position resolution of CES measurements is 0.2 cm electrons with energy of 50 GeV. The PES are made of 0.5 cm pitch scintillator strips and arranged in two layers. The direction of strips in these layers are rotated at $+22.5^\circ$ and -22.5° with respect to the radial direction, to provide a two dimensional position measurement. The PES position resolution is about 0.1 cm.

There is also a set of multiwire proportional chambers in front of the CEM called Central Preshower (CPR) chambers. These chambers detect electromagnetic showers that begin in the solenoid magnet material, providing enhanced photon and soft electron identification. There is no dedicated preshower detector in front of PEM, instead the first scintillator layer of PEM is made thicker (1 cm) to yield more light and is read out separately. This layer is called PPR and used to statistically separate prompt γs from the two-photons coming from the decay of π^0s.

2.2.6 Event Triggers

The output signals produced by various subdetectors of CDF must be stored on an event-by-event basis for later analysis. At typical instantaneous luminosities of $\lesssim 3 \times 10^{31}$ cm^{-2}s^{-1} it is expected to have 1.4 proton-antiproton interactions on average per bunch crossing which occurs at rate of $\sim$ 2.5 MHz for 396 ns bunch spacing. The time required to read out a single event from the entire detector is of order of 2 ms which is about 5000 times larger than the time between two consecutive bunch crossings. Another constraint on event acceptance rate comes from the speed at which information can be recorded on a permanent storage device, which is currently 10–20 ms for a typical event size of 100–200 kB. Therefore, not all the events can be recorded. However, most of the inelastic scattering events produced in $p\bar{p}$ collisions are the result of low-momentum transfer "minimum bias" events, with no significant energy flow in the directions transverse to incoming partons. The rate of producing the more interesting large momentum-transfer events probing short-distance physics typically have cross sections at least of factor of 10^5 less than the minimum bias cross section ($\sim$ 60 mb). For example, $t\bar{t}$ cross section is 9 orders of magnitude less than minimum bias events cross section. At CDF the trigger system is designed to make an efficient on-the-fly selection of events with interesting physics signatures and discard those with low p_T particles.

Before being written to tape the data from the detector passes three consecutive levels of the CDF trigger system, the block diagram of which is shown in Figure 2.11a. The more detailed block diagram showing the selection logic for the first two levels is presented in Figure 2.11b. The Level-1 trigger has an acceptance rate of no more than 50 kHz. The events satisfied the Level-1 requirements are more accurately analysed by Level-2 trigger with 300 Hz acceptance rate. Finally, the Level-3 trigger will decide whether to accept the events for offline storage at the maximum possible writing speed of 75 Hz. Since Level-1 and Level-2 systems have to make fast decisions



(a) Data flow block diagram.

(b) Level-1 and Level-2 triggers system block diagram.

Figure 2.11: CDF data acquisition system and triggers in Run II.

on further event processing they are projected to use only part of the entire event data. At Level-1 the decisions are made simply by counting the number of physics objects. These systems equipped with a custom hardware to do a limited event reconstruction, whereas Level-3 performs almost full event reconstruction using an appropriate software running on computing farm.

As it shown in Figure 2.11b the inputs for Level-1 are the signals coming from the calorimeters, the COT, and the muon detectors. Up to 42 events (at 132 ns clock circle) can be stored in pipelined buffers where the data is stored for 5.5 μs and analysed by the three parallel synchronous streams. The first stream collects the information from the object and global calorimeter triggers. The object triggers fire off when the transverse energy deposited in individual calorimeter towers is above a certain threshold. The electron/photon trigger make decision based on the EM

calorimeter energy, while jet trigger sets a threshold on the sum of EM and HAD energies in a tower. The global trigger compares the total energy from all towers and compares it to the reference value. All the thresholds and requirements on the number of objects can be preset to a desired value in the hardware for each detector regions. The second stream is responsible for finding stubs in the muon chambers. A novel part of the Level-1 trigger upgrade in Run II is the installation of the Extremely Fast Tracker (XFT, [33]) forming the third Level-1 trigger stream. The XFT is able to reconstruct tracks in the transverse plane of the COT. Together with the extrapolation unit (XTRP) the tracks can be matched to clusters in EM calorimeter and muon chambers forming electron and muon primitives. The decisions from above streams are treated with logical AND and OR gates and final decision is formed on event acceptance at Level-1.

In case an event is accepted by Level-1 trigger all the information about this event is passed into one of four asynchronous event buffers of the Level-2 trigger system. In addition the information from shower maximum strip chambers (CES) and from the r–ϕ strips of the SVX detector is transmitted to the Level-2 system. The Level-2 buffers are not able to accept more than four events passed the Level-1 requirements. On average the time required to analyze an event from the buffer is 20 μs during which the trigger incurs dead time. There are three hardware subsystems helping to make the final decision at the Level-2: L2CAL, XSEC, and SVT. The L2CAL is the cluster finder which combines the energies collected by individual towers. Thus L2CAL provides an approximate measurement of jet E_T and direction. The XSEC system discriminates for a certain threshold on the signals from the strips of CES detector. This information combined with tracks extrapolated by XTRP is used for better track-to-cluster matching and therefore electron identification. The Silicon Vertex Tracker (SVT) reconstructs tracks in silicon vertex detector and provides the measurement of the track impact parameter. If all the triggers of Level-2 are satisfied then the full detector readout is initiated.

After Level-1 and Level-2 triggers have made their decisions the events are transfered to the data acquisition system (DAQ) buffers. The system controlling the DAQ buffers, called the Event Builder, orders event fragments and sends them to a farm of conventional PCs serving as the Level-3 trigger. The event fragments are assembled in a block of data, *event record*, suitable for analysis by CDF reconstruction software. The reconstruction of an event takes advantage of the full detector information and improved resolution not available at the lower trigger levels. If an event satisfies Level-3 requirements, the corresponding event record is sent for storage on tape. Each event is marked with a run number corresponding to a continuous period of time for which the detector was taking data. During each run the configuration of the detector including calibration constants is stable and later it can be retrieved for each event.

More information on CDF trigger system can be found elsewhere [34].

Chapter 3

Event Reconstruction and Selection

If an event satisfies the high-pt single e or μ triggers, the information from the detector is digitized and saved on permanently storage device, where it can be analyzed. Before any physics analysis can be done on the data, the physics objects such as jets, electrons, muons, and missing transverse energy must be made from the raw data by a sequence of sophisticated pattern recognition and fitting algorithms called reconstruction. In this Chapter we shall discuss the algorithms used to reconstruct the high level objects and the selection requirements used to create the final data set. The reconstruction can be divided into three general steps. First, the hits from the various detector components are combined to form energy clusters and tracks. Second, the cluster and tracking information is identified as a particle "candidate" based on the particle's signature in the detector as determined by a number of pattern recognition variables. These variables may be used later in the event selection to improve or "tighten" the particle identification. A first pass or "raw" energy and momentum is calculated for each particle candidate. Finally, in the third step, the raw energies and momenta are corrected to achieve the best estimate of the particle's

energy-momentum four-vector.

3.1 Jets Reconstruction

According to the parton model of hadron structure as developed by Bjorken and Feynman (and later justified by the asymptotic freedom property of QCD), at high energies the colliding hadrons can be considered to be composed of free partons (quarks and gluons), each carrying a certain momentum fraction of the parent hadron. The partons undergo hard-scattering, and the total cross section is the incoherent sum of the parton sub-process cross sections. Due to confinement the coloured partons are transformed into colourless hadrons in the process called hadronization or fragmentation[1]. The typical scale for fragmentation is $\sim$ 300 MeV—harder for heavy quark (b,c) fragmentation, which can be parameterized by, for example, the Peterson function [35]. In the string fragmentation model the energy of interaction between a scattered parton q and a parton in the hadron remnant $\bar{q}$ increases as they move apart. If the potential energy of such interaction is high enough a new pair of q' and $\bar{q}'$ quarks will be produced. These quarks along with the initial quarks form colour-singlet states $q\bar{q}'$ and $q'\bar{q}$. The invariant mass of the colourless systems can be still high enough to cause a new splitting followed by another formation of colorless states. The chain of such interactions will continue until the potential energy between the secondary partons will be enough to escape the coupling. As a result, a spray of collimated hadrons is produced in the the direction of the fragmenting parton. The final particles share the total energy of the shower-initiator, which, if deposited and measured in the calorimeter, can be used to estimate the energy of the initial parton.

In the detector the jets of particles are recognized as clusters of energy deposited

[1]The fragmentation mechanism is not well understood yet from first principles. The hadronization also assumes the subsequent decay of unstable particles produced during the fragmentation.

in adjacent calorimeter towers. Since many jets can be produced in event, the jet clustering algorithm— combining and rejecting certain towers—is required to reconstruct the correct topology of the event. In this analysis we use the standard CDF cone algorithm [36] for jet reconstruction. First, the *seed* towers with deposited transverse energy higher than 1 GeV are chosen. Then, these towers are used as initial centroids around which all the towers within a particular open angle $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ grouped into a cluster. The towers belonging to the cluster and having transverse energy above 100 MeV are used to calculate a new E_T-weighted centroid:

$$\phi_{\text{centroid}} = \frac{\sum_{i\subset\Delta R} E_{T,i}\phi_i}{\sum_{i\subset\Delta R} E_{T,i}}, \tag{3.1}$$

$$\eta_{\text{centroid}} = \frac{\sum_{i\subset\Delta R} E_{T,i}\eta_i}{\sum_{i\subset\Delta R} E_{T,i}}, \tag{3.2}$$

where angular coordinates (ϕ_i, η_i) define the direction to the center of i^{th} calorimeter tower with transverse energy deposited in the tower $E_{T,i}$. The sum in Equations 3.1 and 3.2 is taken over all towers with centers inside the cone with $\Delta R = 0.4$. The new cone is drawn around the recalculated centroid and a new iteration is performed. The whole process is repeated until there is no change in the centroid's angular coordinates. In the case if two clusters are overlapping, they may be combined or left intact. The former happens if the energy in the common towers is equal or greater than 75% of the total energy in the smaller cluster, otherwise, the overlap towers are assigned to the cluster with the nearest centroid. In either case, the new centroids have to be recalculated.

Once all stable cones are identified, the following sums over the towers inside each cluster are used to calculate the raw four-momentum of the jets:

$$\begin{aligned} p_{\text{jet}}^{\text{raw}} &\equiv (E; p_x, p_y, p_z) \\ &= \left(\sum_{i\subset\Delta R} E_i; \sum_{i\subset\Delta R} E_i \sin\Theta_i \cos\phi_i, \sum_{i\subset\Delta R} E_i \sin\Theta_i \sin\phi_i, \sum_{i\subset\Delta R} E_i \cos\Theta_i \right), \end{aligned} \tag{3.3}$$

where E_i is the energy deposited in the i^{th} tower and Θ_i and ϕ_i are the polar and azimuthal angles defining the direction to the center of this tower.

It must be noted that the cone algorithm described above showed a satisfactory performance in the Monte Carlo studies [37], although it suffers from a few drawbacks. Two "real" jets can be mistakenly merged into one cluster if the soft emission between them is above the seed threshold. On the contrary, the jets may not be reconstructed at all if their energy is evenly distributed among several towers where deposited E_T is less than the seed threshold. Thus, the number of reconstructed jets depends on empirically determined parameters such as jet cone size and seed threshold.

Before the raw energy of a jet cluster can be associated with the energy of a parton produced during the hard scattering, it must be corrected. Corrections take care of differences occurring on average between the measured energy in the calorimeters and the actual energy of the initial parton. The differences can occur either due to detector's imperfections or due to physics effects. Examples of the first kind are the cracks between towers and calorimeter response as a function of rapidity. The physical reasons can be related to multiple interactions and particles depositing their energy outside the cone. Seven additive and multiplicative corrections are applied to the raw energy in the following way:

$$E_T = \left(E_T^{\text{raw}} \times f_{\text{rel}} \times f_{\text{time}} \times f_{\text{scale}} - E_T^{\text{MI}}\right) \times f_{\text{abs}} - E_T^{\text{UE}} + E_T^{\text{OOC}}. \tag{3.4}$$

The description of each factor is presented below.

- The **relative correction (f_{rel})** is responsible for adjusting the responses from the calorimeter towers at different pseudorapidities making them uniform function of η. The scale factor is determined from the disagreement in transverse energy balance in dijet events where one jet is measured in well-calibrated central region far from the detector cracks and the other is lying outside the $0.2 < |\eta| < 0.6$ region.

- The **time-dependent correction ($f_{\mathbf{time}}$)** is responsible for justifying calorimeters responses due to their deterioration in time, *e.g.* due to scintillator aging.

- The **scale correction ($f_{\mathbf{scale}}$)** accounts for the differences in Run I and Run II jet energy measurements,which are due to the fact that there is less inactive material in the crack regions. The photon energy deposited in the electromagnetic calorimeter can be measured with high precision, and the scale factor is determined from the photon-jet balancing:

$$f_{\text{scale}} = \frac{1 + \langle f^{\text{Run I}} \rangle}{1 + \langle f^{\text{Run II}} \rangle}, \tag{3.5}$$

  where $f^{\text{Run I,II}} = \frac{P_T^{\text{jet}} - P_T^{\gamma}}{P_T^{\gamma}}$.

- The **correction for multiple interactions ($E_T^{\mathbf{MI}}$)** is required if more than one interaction happens per bunch crossing. The $t\bar{t}$ events can have extra, overlapping minimum bias events, the Poisson probability of which increases with instantaneous luminosity. The energy from these events may fall into the jet clustering cone and must be subtracted from the raw value. The number of additional reconstructed vertices is used as measure for the number of soft $p\bar{p}$ scatterings. The additional energy per interaction E_T^{MI} in the jet cone of a given size $\Delta R = 0.4$ is estimated by measuring transverse energy of a random cone as a function of reconstructed primary vertices in minimum bias events.

- The **absolute energy correction ($f_{\mathbf{abs}}$)** scales the measured jet energy to match the energy of the original parton as it defined by Monte Carlo simulation. The true parton energy is usually underestimated because of nuclear absorption, energy loss in un-instrumented calorimeter regions, and non-linear response to neutral and charged pions. The multiplicative factor f_{abs} is determined from the detector simulation by comparing the sum of P_T's of the particles within the cone of same size around the parton direction which matched the jet direction

with $\Delta R < 0.4$:

$$f_{\text{abs}} = \left\langle \frac{\sum\limits_{i \subset \Delta R} P_{T,i}^{\text{MC particle}}}{E_T^{\text{jet}}} \right\rangle . \tag{3.6}$$

- The **underlying event correction ($E^{\mathbf{UE}}$)** is defined by activities that are not directly related to the physical process of interest. The sources of such parasitic activities include the interactions, also referred as underlying events, of spectator partons and proton-antiproton remnants. Due to this effect, an extra small amount of energy produced in underlying events needs to be subtracted from each reconstructed jet. The value of E^{UE} is measured from the minimum bias data requiring events with only one vertex.

- The **out-of-cone correction ($E^{\mathbf{OOC}}$)** is conditioned by the fact that not all particles produced by the initial parton lie inside the jet cone. The usual cause of such energy leakage is the soft gluons radiation. In order to account for the escaped particles, the certain amount of energy needs to be added to the energy of the jet cluster. An exact value of E^{OOC} is determined from the Monte Carlo simulation by calculating the difference in the actual parton transverse momentum and the sum of transverse momenta of all daughter particles lying inside the jet cone:

$$E^{\text{OOC}} = \left\langle P_T^{\text{parton}} - \sum_{i \subset \Delta R} P_{T,i}^{\text{MC particle}} \right\rangle . \tag{3.7}$$

The application of the corrections to the raw jet energy helps to estimate the true energy of the parton as accurate as possible. The total systematic uncertainty on the jet E_T is determined from the individual uncertainties assigned to each correction level. The overall uncertainty depends on jet E_T value and varies from 5% to 12%. The central jets are measured slightly better than the jets in the plug region.

Before applying the final physical cuts we correct the jet energy by applying the relative f_{rel}, time-dependent f_{time}, scale f_{scale}, and multiple interaction E_T^{MI} correc-

tions. Other corrections are applied in the analyses, where the full energy of initial partons is required, for example, in a top quark mass reconstruction. In our case, we measure signal rates by comparing data to MC simulation. We consider only those jets that satisfy the physical cuts listed in Table 3.1.

Table 3.1: Summary on jet identification cuts. To be accepted a jet must satisfy the listed requirements.

Variable	**Selection Requirement**
E_T	≥ 15 GeV
$\|\eta\|$	≤ 2

3.2 Lepton Identification

The lepton identification is crucial for this analysis. One or both of the W bosons produced in $t\bar{t}$ events decay leptonically producing—with the equal probability—a highly energetic $e^{\pm}$, $\mu^{\pm}$, or $\tau^{\pm}$. Unlike electrons, muons and taus are unstable particles with an average lifetimes $\tau \approx 2.2\times10^{-6}$ and $\tau \approx 291\times10^{-15}$ seconds respectively. The relativistic muons generated in signal events pass through the detector undecayed due to their relatively long lifetime. In contrast, the taus have much shorter lifetime and decay soon after they were produced. With almost 40%-probability the product of tau decay will include a lighter lepton ($e^{\pm}$ or $\mu^{\pm}$) and two neutrinos. Events with such signature will be classified as signal and contribute to the lepton-plus-jets channel. However, in about 60% of the time the taus will decay hadronically producing particle jets. The identification of tau-jets is possible, although not trivial. In this study we will constrain ourselves to those events where one energetic isolated electron or muon is observed. In the following subsections we present the identification cuts accepted by the CDF Top Group, other groups may use different definitions [38, 39].

3.2.1 Electron

The electron is the lightest electrically-charged lepton. Electron showers differ from the hadronic showers in jets by its length, shape, and number of tracks. Unlike hadronic jets, electron jets can be identified by the substantial amount of energy deposited in electromagnetic calorimeter and a high-P_T track reconstructed in the drift chamber. Typically, electron jets deposit most of their energy in a single calorimeter tower, while parton jets are much broader and continue into hadronic calorimeter towers. The shape and position of electromagnetic shower is measured to high precision by shower maximum detectors. This information is also used in electron identification. Only electrons detected in the central calorimeter (CEM, $|\eta| < 1.1$) are considered for this analysis. The full list of all variables and requirements on their values is presented in Table 3.2 followed by detailed explanations.

Table 3.2: Summary of central electron identification cuts. To be accepted an electron must satisfy the listed requirements.

Variable	**Selection Requirement**
E_T	≥ 20 GeV
p_T	≥ 10 GeV/c
E/p (if $p_T \leq 50$ GeV/c)	≤ 2
$E_{\rm Had}/E_{\rm EM}$	$\leq 0.055 + 0.00045E$
$\lvert L_{\rm shr}\rvert$	< 0.2
$\lvert\Delta z\rvert$	≤ 3.0 cm
$Q \cdot \Delta x$	≤ 1.5 and ≥ -3.0 cm
$\chi^2_{\rm strip}$	< 10
$\lvert z_0\rvert$	≤ 60 cm
number of axial segments	≥ 3
number of stereo segments	≥ 3
Isolation	≤ 0.1
Conversion check	$\Delta\Lambda > 0.04$ and $\Delta r > 0.2$ cm
Fiduciality	required

$\boldsymbol{E_T \equiv E \sin\Theta \geq 20}$ **GeV** The transverse energy of EM cluster E_T is defined as a product of total energy deposited in the EM cluster and $\sin\Theta$, where Θ is the angle of a COT track pointing to that cluster. Usually, EM cluster has only one seed tower in longitudinal and azimuth directions. In order to reject most of the background events, where electron is not produced from W, we select events only with high values of E_T.

$\boldsymbol{p_T \geq 10}$ **GeV/**$\boldsymbol{c}$ An electron in magnetic field moves along a helix whose curvature measures the electron's transverse momentum p_T. The COT's response to electrons and positrons has been found to be slightly different, and the curvature of charged tracks is corrected by a small additive term:

$$\frac{Q}{p_T} = \frac{Q}{p_T^{\text{raw}}} - 0.00037 - 0.0011 \times \sin(\phi + 0.28), \tag{3.8}$$

where Q is the charge of particle and ϕ is track's azimuthal angle.

$\boldsymbol{E/p \leq 2}$ The absolute value of electron's momentum as measured by COT track is required to be no less than 50% of the full E measured in EM calorimeter. This requirement is due to external bremsstrahlung in the material preceding the EM calorimeter. An electron loses a portion of its energy by emitting a photon while passing through the material at the inner COT radius. For electrons with $p_T >$ 50 GeV/c the radiated photon will almost certainly be collinear with the electron and contribute to the same calorimetric energy. For less energetic electrons, the photon can end up in an adjacent EM tower. For the above reasons we use EM energy for the electron's four-momentum calculation.

$\boldsymbol{E_{\text{Had}}/E_{\text{EM}} \leq 0.055 + 0.00045E}$ In order to avoid misidentification of hadronic jet as an electron one, the fraction of energy E_{Had} deposited in hadronic calorimeter is required to be a small fraction of energy E_{EM} deposited in electromagnetic calorimeter.

The last term compensates for leakage at high energies into the hadronic part of the calorimeter.

$L_{\rm shr} < 0.2$ The energy of the shower deposited in towers adjacent to the seed tower is expected to match the data from test beam electrons. This lateral shower profile is characterized by the variable $L_{\rm shr}$ calculated from the measured energy $E^{\rm meas}$ in adjacent towers and the expected energy $E^{\rm exp}$:

$$L_{\rm shr} = 0.14 \sum_{\rm towers} \frac{E_i^{\rm meas} - E_i^{\rm exp}}{\sqrt{(0.14\sqrt{E})^2 + (\Delta E^{\rm exp})^2}}, \tag{3.9}$$

where $0.14\sqrt{E}$ and $\Delta E^{\rm exp}$ are the uncertainties on the measured and expected energies. Typically, the sum is taken over two nearby towers in a cluster.

$|\Delta z| < 3.0$ cm, $-3.0 \leq Q \cdot \Delta x \leq 1.5$ cm Using CES detector, the shower is reconstructed as clusters in x and z projections corresponding to "strip" (r–z) and "wire" (r–ϕ) views. The electron identification algorithm is searching for the best COT track matching the CES clusters within ranges given by above $Q \cdot \Delta x$ and Δz cuts, where Q is ± 1 for $e^{\pm}$. The former cut is asymmetric in order to account for an offset between the track and the CES cluster due to bremsstrahlung from the electron or positron. These proposed requirements are efficient in rejecting electron candidates faked by simultaneous penetration of π^0 and $\pi^{\pm}$ into the same tower.

$\chi^2_{\rm strip} < 10$ This quantity compares the CES shower profile in r–z view to the profile measured with test beam electrons. The χ^2-fit is performed on the distribution of energy deposited on each of the 11 strips in the CES shower. The purpose of this requirement is similar to the previous cut, providing discrimination between electrons and the wider clusters from fully or partially merged showers from π^0 decays.

$|z_0| \leq 60$ cm The distance along beam direction between the nominal interaction point $z = 0$ and the point where the reconstructed track intersects the beam axis. In order to be well reconstructed, the events are required to be near the center of the detector. The distribution of event vertices in z due to the long beam bunch size is well described by a Gaussian distribution with $\sigma = 26$ cm.

Number of axial and stereo segments ≥ 3 each In order to ensure a good quality of the track associated with an electron candidate, it is required to pass through at least three axial and three stereo superlayers of the COT.

Isolation ≤ 0.1 This variable defines how much energy is left by other particles outside the EM cluster lying inside the cone $\Delta R \leq 0.4$ and it is given by

$$\text{Isolation} \equiv \frac{E_T^{\text{iso}}}{E_T^{\text{EM cluster}}} = \frac{E_T^{\text{cone}} - E_T^{\text{EM cluster}}}{E_T^{\text{EM cluster}}}, \tag{3.10}$$

where E_T^{cone} and $E_T^{\text{EM cluster}}$ are the transverse energies corresponding to the cone and EM cluster. This isolation variable is then corrected for two effects [40]: The leakage of energy into the neighbouring ϕ wedge, E_T^{leak}, and the extra energy in the cone from an underlying events E_T^{UE}. Thus,

$$E_T^{\text{corrected iso}} = E_T^{\text{iso}} + E_T^{\text{leak}} - E_T^{\text{UE}}. \tag{3.11}$$

The first correction was determined from Monte Carlo studies and parametrized as a function of x coordinate in CES:

$$E_T^{\text{leak}} = E_T^{\text{EM cluster}} \cdot p_0 \cdot \exp p_1 \cdot (|x_{\text{CES}}| - 21), \tag{3.12}$$

where $p_0 = 0.0511 \pm 0.0075$ and $p_1 = 0.33 \pm 0.061$. The second correction is determined similarly to hadronic jet energy correction described in Section 3.1. In fact, the Isolation is not an electron identification cut but rather a discriminator between non-isolated electrons produced in semi-leptonic decays inside hadronic jets and isolated

electrons coming from W and Z boson decays. This cut has been applied in the final data set selection, unless otherwise stated.

Conversion removal: $\Delta\Lambda > 0.04$ and $\Delta r > 0.2$ cm Non-prompt electrons can be produced as a result of photons converting into electron-positron pairs in the detector material. These electrons are identified matching the electron track to an oppositely charged track that reconstructs a conversion point with the electron. At this point, the two helicies are tangent, as determined by a near-zero separation of the track circles in the r–ϕ plane, $\Delta r < 2$ mm, and a near-zero difference in polar angles $\Delta\Lambda \equiv \Delta \cot\Theta \leq 0.04$. If the high-$p_T$ electron is identified as conversion, the whole event is rejected.

Fiduciality The electrons are required to be in the fiducial (well instrumented) regions of the detector. The following regions are non-fiducial and excluded:

- **$0.77 < \eta < 1.0$ and $75^\circ < \phi < 90^\circ$** The chimney with supporting cables and cryogenic suppliers.
- **$1.05 < |\eta| < 1.10$** The region with small depth of the EM calorimeter.
- **$|z_{\mathrm{CES}}| < 9$ cm and $|z_{\mathrm{CES}}| > 230$ cm** The parts of CES detector at the inner and outer edges of the central EM calorimeter.

3.2.2 Muon

The muons produced in $t\bar{t}$ processes are highly relativistic and, therefore, traverse the detector undecayed. In the detector the muons are identified as a near-minimum ionizing charged track matched to a muon stub. The reconstructed muons are identified according to where the muons stubs are located, and are called CMU, CMP, and

CMX accordingly. Muons with both CMU and CMP stubs are called CMUP muons. In this analysis we use CMUP and CMX muons.

The full list of the muon identification variables and requirements on their values is presented in Table 3.3 followed by detailed explanations.

Table 3.3: Summary of CMU, CMP, and CMX muon identification cuts. To be accepted a muon must satisfy the listed requirements.

Variable	**Selection Requirement**
p_T	≥ 20 GeV/c
$E_{\rm EM}$	$\leq \max(2, 2 + 0.0115 \cdot (p - 100))$ GeV
$E_{\rm Had}$	$\leq \max(6, 6 + 0.0280 \cdot (p - 100))$ GeV
$\lvert\Delta x\rvert_{\rm CMU}$	≤ 3 cm
$\lvert\Delta x\rvert_{\rm CMP}$	≤ 5 cm
$\lvert\Delta x\rvert_{\rm CMX}$	≤ 6 cm
$\rho(\eta, z_0)_{\rm CMX}$	≥ 140 cm
$\lvert z_0\rvert$	≤ 60 cm
d_0	≥ 0.02 (≥ 0.2) cm with (without) SVX
number of axial segments	≥ 3
number of stereo segments	≥ 3
Isolation	≤ 0.1
Non-cosmic	required

$\boldsymbol{p_T \geq 20}$ **GeV/**$\boldsymbol{c}$ The transverse momentum of a muon is measured in the same way as of electron. The curvature correction is also applied for positively and negatively charged tracks.

$\boldsymbol{E_{\rm EM} \leq \max(2, 2+0.0115\cdot(p-100))}$ **GeV** This construction means that muons with momentum p less than 100 GeV/c are not expected to deposit energy more that 2 GeV in electromagnetic calorimeter. For muons with $p < 100$ GeV/c the cut is corrected in order to make it more efficient.

$\boldsymbol{E_{\mathrm{Had}} \leq \max(6, 6 + 0.0280 \cdot (p - 100))\ \mathrm{GeV}}$ This is similar to the E_{EM} requirement, but for energy deposited by the muon in hadronic calorimeter E_{Had}.

$\boldsymbol{|\Delta x|_{\mathrm{CMU}} \leq 3\ \mathrm{cm},\ |\Delta x|_{\mathrm{CMP}} \leq 5\ \mathrm{cm},\ |\Delta x|_{\mathrm{CMX}} \leq 6\ \mathrm{cm}}$ These constraints are imposed on the distance in r–ϕ plane between the muon track extrapolated in the relevant detector chambers and the muon stub segment. The CMU and CMP type cuts are related to all muons having hits in either CMU, CMP, or both detector chambers. The CMP and CMX cuts are looser than for CMU because of additional multi-coulomb scatterings in the material and the deflection angle is projected over a longer distance to the chamber surface.

$\boldsymbol{\rho(\eta, z_0)_{\mathrm{CMX}} \geq 140\ \mathrm{cm}}$ The level 1 CMX18 trigger requires that a track has hits in all 4 axial superlayers of the COT. However, if track has η close to 1 and $|z_0|$ far from COT center, the high-p_T muon will not traverse these layers. In order to ensure that the track has a minimum number of hits, the cut on COT exit radius ρ is imposed only for CMX muons. The ρ is defined as

$$\rho(\eta, z_0) = \frac{\mathrm{sign}(\eta) \cdot z_{\mathrm{COT}} - z_0}{\cot\Theta}, \tag{3.13}$$

where η and z_0 are the pseudorapidity and z coordinate of the event vertex, $z_{\mathrm{COT}} = 155$ cm is the length of the COT, and Θ is the polar angle.

$\boldsymbol{|z_0| \leq 60\ \mathrm{cm}}$ As for electrons the z coordinate of the event vertex, as reconstructed from the muon track, must lie within 60 centimeters of the detector's geometric center.

$\boldsymbol{d_0 \geq 0.02\ (0.2)\ \mathrm{cm\ with\ (without)\ SVX}}$ The impact parameter d_0 is the distance between the reconstructed muon track and the beam axis in the r–ϕ plane. This requirement selects muons from the primary vertex, removing cosmic muons and muons from kaon and pion decays.

Number of axial and stereo segments $\geq$ 3 each This is identical to the same requirement on the quality of electron track reconstructed in the COT.

Isolation $\leq$ 0.1 Again, this is similar to the electron isolation cut: The extra energy inside the cone with $\Delta R = 0.4$ is required to be less than 10% of the calorimeter tower associated with the lepton.

Cosmic ray removal Usually, muons in cosmic showers are produced in $\mu^+\mu^-$ pairs with a very high energy. If passed through the detector's tracking system such pairs are recognized as nearly straight tracks. In that case, the cut on impact parameter d_0 is proven to be very efficient. Cosmic muons can pass through the calorimeter towers only. To identify such muons the time delay between hits in two adjacent towers is measured by the Time of Flight detector (TOF). If muon is identified as cosmic then the whole event is rejected.

3.3 Missing Energy

Neutrinos are electrically-neutral leptons participating only in weak interactions with other particles. Thus the presence of the neutrino in an event can only be indirectly deduced from momentum conservation. Since only the total transverse momentum is detected in the $p\bar{p}$ collision, only the transverse momentum of the neutrino can be measured. The initial transverse momentum of the system of interacting partons is negligible and so must be the vector sum of all final particles momenta projected on the transverse plane. If the latter is not consistent with zero, the missing transverse energy ($\not\!\!E_T$) is introduced to restore the balance in accordance with the total momentum conservation law:

$$\vec{\not\!\!E}_T = -\sum_i \vec{P}_{T,i}. \tag{3.14}$$

Since the CDF is sealed around the z axis the sum on the right sum in Equation 3.14 can be calculated over all detected particles, and thus the missing transverse energy can be associated with the neutrino's transverse momentum. Unlike the transverse momentum we cannot make any assumptions about the initial longitudinal components of the interacting partons. Moreover, many particles with very high pseudorapidity are not detected and, therefore, neutrino's z-component cannot be measured.

The raw value of $\vec{\not\!\!E}_T$ is estimated very fast at the trigger level by summing the transverse energies of all calorimeter towers:

$$\vec{\not\!\!E}_T^{\text{raw}} = -\sum_i (E_i \sin\Theta_i)\vec{n}_i \equiv -\sum_i E_{T,i}\vec{n}_i, \tag{3.15}$$

where E_i is the energy deposited in the i^{th} tower, Θ_i is the polar angle of the vector pointing from the event vertex to the center of the i^{th} tower, and $\vec{n}_i$ is the component of that vector lying in the plane perpendicular to the beam direction. If the azimuthal angle of $\vec{\not\!\!E}_T^{\text{raw}}$ is $\phi_{\not\!\!E_T}$ then the x and y components are given by

$$\not\!\!E_x^{\text{raw}} = \not\!\!E_T^{\text{raw}} \cdot \cos\phi_{\not\!\!E_T} \text{ and } \not\!\!E_y^{\text{raw}} = \not\!\!E_T^{\text{raw}} \cdot \sin\phi_{\not\!\!E_T}. \tag{3.16}$$

All towers with pseudorapidity $|\eta_i| < 3.6$ and deposited energy $E_i > 100$ MeV are taken into account in Equation 3.15.

Before missing energy can be linked to neutrino's momentum, its raw value $\not\!\!E_T^{\text{raw}}$ should be corrected for mismeasurements due to the detector imperfections. In the definition of the $\vec{\not\!\!E}_T^{\text{raw}}$, given by Equation 3.15, the right part initially contains the information used in determination of uncorrected jet energies and unclustered energy outside the reconstructed jets:

$$\vec{\not\!\!E}_T^{\text{raw}} = -\left(\sum_{i\subset\text{jets}} E_{T,i}\vec{n}_i + \sum_{i\not\subset\text{jets}} E_{T,i}\vec{n}_i\right) = -\sum_{j=1}^{N_{\text{jets}}} \vec{E}_{T,j}^{\text{uncorr}} + \vec{E}_T^{\text{unclust}}. \tag{3.17}$$

From where it follows that, once the jets energies are corrected, the missing transverse energy has to be recalculated by substituting uncorrected jet energies $\vec{E}_{T,j}^{\text{uncorr}}$ with the

corrected ones $\vec{E}_{T,j}^{\text{corr}}$, *i.e.* in the vector notation we add the former and subtracting the latter.

Another correction of $\vec{\not\!\!E}_T^{\text{raw}}$ is necessary in order to resolve the possible difference in muons energies measured by calorimeters and COT. Typically a muon will not deposit all of its energy in the calorimeter, therefore, its momentum has to be determined from the track curvature measured in COT. In the case of an energetic isolated muon coming from the primary vertex, we have to add the muon $\vec{P}_T$ with a negative sign and subtract the calorimeter energy $\vec{E}_T$ deposited by the muon since it was already counted in $\vec{\not\!\!E}_T^{\text{raw}}$.

Finally, the correction of the raw transverse missing energy can be written as

$$\vec{\not\!\!E}_T = \vec{\not\!\!E}_T^{\text{raw}} - \left(\sum_{j=1}^{N_{\text{jets}}} \vec{E}_{T,j}^{\text{corr}} - \sum_{j=1}^{N_{\text{jets}}} \vec{E}_{T,j}^{\text{uncorr}} \right) - \left(\sum_{\mu=1}^{N_{\text{muons}}} \vec{P}_{T,\mu} - \sum_{\mu=1}^{N_{\text{muons}}} \vec{E}_{T,\mu} \right). \quad (3.18)$$

The uncertainty on missing transverse energy primarily comes from the uncertainties on jet energy and muon momentum measurements.

3.4 Secondary Vertex Tagging

The identification of jets originating from b quarks is called b-tagging. Tagging is very important for background reduction in top analyses, because most of the $t\bar{t}$ backgrounds do not contain heavy quarks. Moreover, in this analysis we measure the rate of top decay to a Wb final state. Therefore, tagging is an essential part of this analysis.

There are several methods available to tag the b-jets. The most powerful techniques are *secondary vertex* (SecVtx) and *soft lepton* (SLT) tagging. The first method is based on measurable distance traveled by highly relativistic B-mesons before they decay (several millimeters) which can be measured by the silicon vertex detector. A

secondary vertex can be reconstructed from the charged decay products. The second method identifies the low p_T (soft) leptons from semileptonic B-hadron decays, occurring with approximately 10% probability. In this measurement we use the higher efficiency SecVtx algorithm to identify b-jets.

The SecVtx algorithm was used in Run I to discover the top quark. In Run II the algorithm is essentially the same as it was implemented and used in Run I measurements [41]. The description of this method is given below.

The procedure starts with the reconstruction of the primary vertex on event-by-event basis. All well-measured tracks in the event are fitted to a common point of origin. The situation when several collisions occur per one bunch crossing is typical for high luminosities. The fitter is designed to find several primary vertices, if they are separated in the z coordinate. In the data used in this analysis, the average number of reconstructed vertices per event is 1.4. We proceed by considering only the primary vertex closest to the track associated with the high-p_T lepton. The precise location of that primary vertex ($x_{\rm pv}$, $y_{\rm pv}$, $z_{\rm pv}$) is determined by refitting the tracks in ± 1 cm window around first iteration vertex. Only tracks consistent with the primary vertex are used in the fit by requiring an impact parameter significance $|d_0/\sigma_{d_0}| < 3$, where the *impact parameter* (d_0) is the distance of closest approach of the track to the beamline, and the error (σ_{d_0}) includes both the uncertainties of on the track and the beamline.

Once the primary vertex is found, the search for secondary vertex is performed on a per-jet basis. The algorithm considers all the jets with at least two good tracks inside a given jet cone. A track is regarded as good if it satisfies a set of requirements on a number of hits in silicon and COT detectors, the quality of the track fit, and transverse momentum value. A jet is referred as "taggable" if it has at least two good tracks. The SecVtx algorithm first tries to find at least three tracks with $|d_0/\sigma_{d_0}| > 2.5$ including two tracks having $p_T > 0.5$ GeV and one track having $p_T > 1.0$ GeV. If

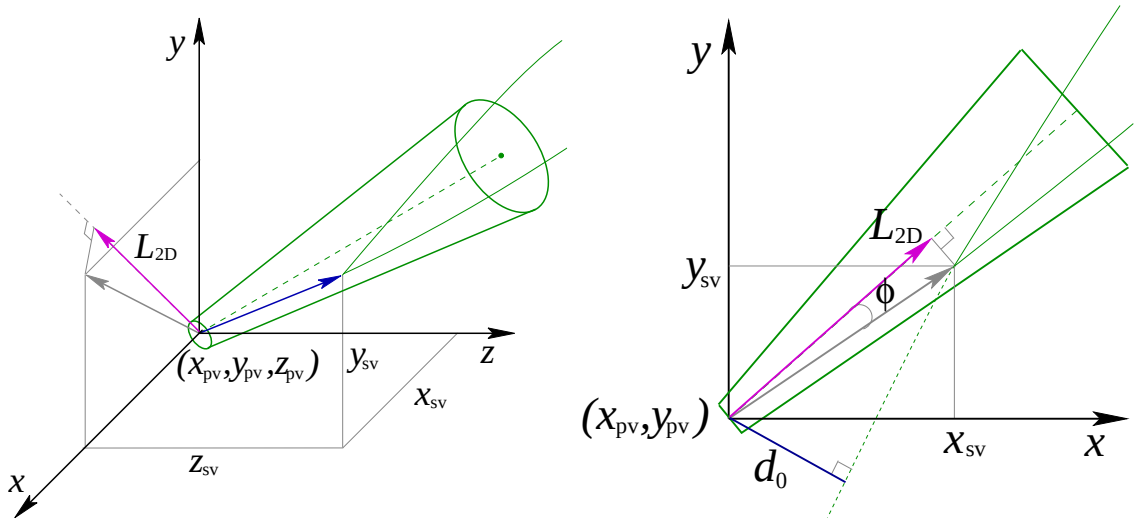


Figure 3.1: These pictures illustrate the reconstruction of a secondary vertex inside a b-jet.

this fails, then an attempt is made to find two tracks with g $|d_0/\sigma_{d_0}| > 3.0$, one with $p_T > 1.0$ GeV and the other with $p_T > 1.5$ GeV. The selected tracks are fitted to a point of relative closest approach, determining the location of the secondary vertex $(x_{\rm sv}, y_{\rm sv}, z_{\rm sv})$. Knowing the coordinates of both primary and secondary vertices one can calculate the two-dimensional decay length $L_{\rm 2D}$ as shown in Figure 3.1. The transverse component in xy plane of the vector pointing from the primary to the secondary vertex is projected onto the jet axis giving the decay length $L_{\rm 2D}$. If the absolute value of angle ϕ is less than 90° then $L_{\rm 2D}$ is said to be positive, otherwise, the $L_{\rm 2D}$ is negative. It is expected that secondary vertices from the decay of b and c hadrons will have larger positive values of $L_{\rm 2D}$ than that from occasionally mis-measured tracks. The jet is positively tagged if $L_{\rm 2D}/\sigma_{L_{\rm 2D}} > 3$ and negatively tagged if $L_{\rm 2D}/\sigma_{L_{\rm 2D}} < -3$, where $\sigma_{L_{\rm 2D}}$ is the total estimated uncertainty on $L_{\rm 2D}$.

In this analysis we are heavily dependent on the knowledge of the tagging efficiency, *i.e.* the probability to positively tag a jet coming from a b quark. The b-tagging efficiency is measured in data and Monte Carlo samples. A special data

sample with low-p_T non-isolated electrons is prepared. The electrons inside the jets are believed to be a product of semileptonic decays of bottom and charm hadrons. The fraction of each kind of heavy flavour events was estimated and a matching MC sample was generated using the HERWIG package. In both samples the efficiency to tag a heavy flavour jet was calculated in [23] yielding their averaged ratio of $\frac{\epsilon_{\rm DATA}}{\epsilon_{\rm MC}} = 0.82 \pm 0.06$ (stat. + syst.). This ratio is called the tagging scale factor (SF) and it is used to scale the tagging efficiencies in $t\bar{t}$ Monte Carlo samples.

3.5 Event Selection

The data sample considered in this analysis was collected at CDF during Run II period lasting from March 2002 to September 2003. The measured integrated luminosity corresponding to this period and including only those runs where all the detector parts except for the silicon detector functioned properly is 193.5 ± 11.2 pb^{-1}. For the same period of time but with additional requirement on silicon detector performing correctly, the integrated luminosity is measured to be 162.1 ± 9.4 pb^{-1}. The quoted uncertainties were studied by the CDF luminosity group ([42, 43]). There are two sources contributing to the total 5.8% uncertainty. First one comes from the measurement of average number of primary interactions using CLC detector, and its value is 4.2%. The second contribution of 4.0% comes from the measurement of $p\bar{p}$ inelastic cross section $\sigma_{p\bar{p}} = 59.3 \pm 2.4$ mb [44].

After full event reconstruction, the particles are identified using the various cuts considered in Sections 3.1 and 3.2. In order to select a signal-rich dataset we apply further selection criteria on the number of reconstructed objects consistent with the expected final signature of the signal events. All the selection requirements were studied and optimized by the Top Group at CDF and used in the $t\bar{t}$ cross section measurement [45]. In lepton-plus-jets channel we select events with exactly one high-

p_T lepton and at least 4 hadronic jets (Table 3.4).

Table 3.4: List of global event cuts in lepton-plus-jets channel.

Variable	**Selection Requirement**
$\not\!\!E_T$	≥ 20 GeV
number of leptons	$= 1$
number of jets	≥ 4

In the remainder of the thesis we will refer to the selected inclusive data sample selected without any b-tag requirement as *pretag sample*. The pretag sample is split into three orthogonal subsamples according to the number of b-tags observed in an event: the *0-tag* subsample, the *1-tag* subsample, and the *2-tag* subsample which includes events with two and more b-tags. In fact, we do not observe events with more than 2 b-tags. The summary on total event yields are shown in Table 3.5 as a function of different tag multiplicities.

Table 3.5: Number of total selected data events (pretag) and subsamples with different tag multiplicities in 162.1 pb^{-1}.

pretag	**0-tag**	**1-tag**	**2-tag**
107	79	23	5

Chapter 4

Artificial Neural Networks

The method used in this analysis is heavily based on the application of Artificial Neural Networks (ANN). This work would not be complete without a detailed description of this technique. Many physics problems require us to establish a functional dependency between two sets of measured or *a priori* known quantities. For example, in the simplest case of two single variables the scientist would fit the data points in order to get an analytical function. Unfortunately, the analytical solution is not always possible or easy to find, especially if the problem is formulated in many dimensions. The ANN technique provides a powerful method which, if used correctly, can help to construct a computational solution to such problems. In this chapter we would like to describe the fundamental features of the ANN algorithm. Then in Chapter 5.1 we report in details how this method is applied to the top branching ratio calculation.

4.1 Introduction to Artificial Neural Networks

The computational scheme referred as the Artificial Neural Networks was inspired by the work done in research fields studying the properties of the human brain. By the

middle of the twentieth century it was well known that the brain consisted of a large number of interconnecting cells called *neurons*. The way individual neuron functions was qualitatively described by neuroscientists and in 1943 McCulloch and Pitts [46] introduced an oversimplified model of a neuron as a threshold logic unit. Later, in 1958, Rosenblatt explored [47] the properties of several reciprocally connected threshold neurons forming a simple artificial network known as *perceptron*. Initially, such configuration seemed to be a powerful tool being able to reproduce binary logical operations such as AND, OR, and even exclusive OR (XOR), although the latter required grouping neurons in more than 2 layers. However, the method used to adjust the inner neuron parameters could not be applied effectively towards the solution of many practical problems. Moreover, whenever this method worked it required too large computational times on contemporary machines. These problems were pointed out by Minsky and Papert in 1969 [48] and, as a result, the interest to ANN has faded until 1980, when new algorithms were developed and the computer machinery was improved significantly.

Since its rediscovery, the ANN technique was extensively used in many fields including Statistical and High Energy Physics. Today the ANN has proven its ability to solve many practical problems by providing successful classification models. In order to understand the main idea behind the ANN technique it is advisable to consider artificial neuron properties inherited from its biological prototype.

4.1.1 Biological Neuron and Its Mathematical Model

One should be very careful when discussing parallels between a real neuron and its mathematical concept used in ANN theory. An analytical description of the biological neuron is indeed meant to simulate its principal functionality, nevertheless, it should be always kept in mind that a real neuron is much more complicated and still

ambiguous object. Having said this, it is still worth noticing that systems based on such oversimplified model preserve the main features of their complex originals.

The neuroscientists estimate the total number of nerve cells in the human brain to be of the order of 10^{11}, whereas, for instance, the nematode worm has only 302 neurons. The neurons are connected with each other in a complex way. They are capable in exchanging tiny electrical (or chemical) signals with the neighbouring neurons. A typical biological neuron is shown in Figure 4.1 [49]. It consists of the three major parts called *soma*, *dendrites*, and *axon*, each having its own purpose:

Soma is the body of the neural cell. The vital functionality of the cell as a whole depends on the soma.

Dendrites are the short tree-like extensions of the soma. The dendrites are responsible for receiving signals from other neurons and transmitting them to the cell's body.

Axon is a much longer projection of the soma transmitting the signals away from the cell's body. Each neuron has only one axon, but this axon may branch into many supplementary terminals connecting with other neuron's dendrites.

Neurons can have thousands of dendrites making it possible for them to connect with tens of thousands of other cells. The impulse traveling from one neuron to another is bifurcated at the axon terminal joints affecting more than one other cell. At the junctions between axon and dendrite tips, the signal propagates through the special formations called *synapses*. From a simplified point of view, the synapses act like signal modulators or resistors altering the strength of the signal. The modified signals from many cells propagate through the dendrites where their strength is summed. If the amplitude of the arrived impulse is higher than a specific threshold level (or activation time is long enough) then the neuron is able to fire its own signal

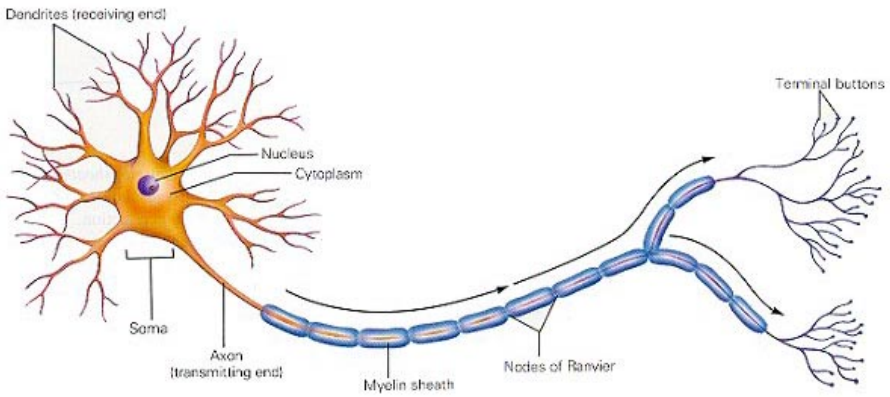
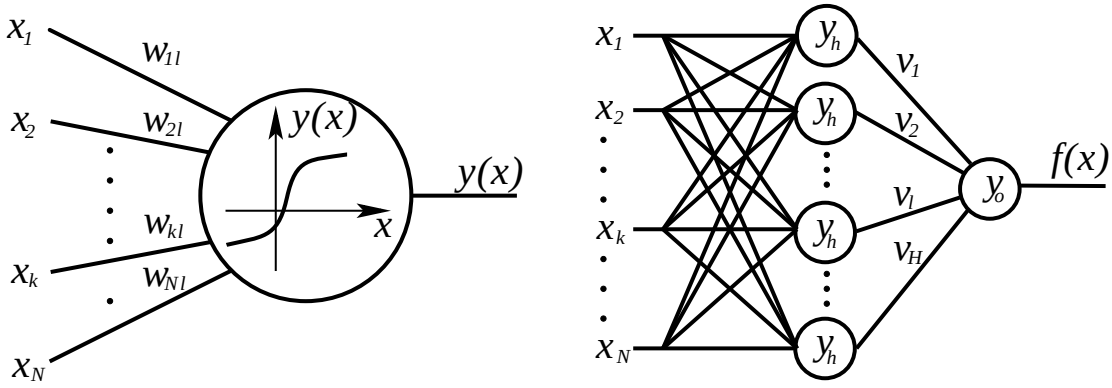


Figure 4.1: Artistic interpretation of the biological neuron. The arrows show the direction in which small electric impulses propagate through the neuron.

and send it to the axon. The generated signal affects other neurons in the similar way.

Based on the biological neuron functionality described above, one can build its mathematical model. In our model the neuron will have N number of inputs (dendrites) and one output (axon). The sketch of a neuron is shown in Figure 4.2a. The strength of the signal received by the i^{th} input we denote as x_i. For simplicity we



(a) Mathematical model of neuron.

(b) Artificial neural network.

Figure 4.2: Graphical representation of the mathematical model of a neuron (a) and a simple artificial neural network with N inputs, H hidden units, and one output (b).

assume that the input signals are linearly scaled by some weight w_i (synapse), so that the neuron receives a set of modified signals $(w_1x_1, w_2x_2, \ldots, w_Nx_N)$. In addition to N neuron inputs we present a so-called *bias* or *threshold* term w_b which can be regarded as an extra input with a constant unit signal $x_{N+1} = 1$. The meaning of the bias term will be cleared out later. The weighted input signals are summed within the neuron:

$$X = w_1x_1 + w_2x_2 + \ldots + w_Nx_N + w_b, \tag{4.1}$$

and then some fixed function $y(\cdot)$ takes the sum as an argument to form the output of the neuron:

$$Y = y(X) = y(w_1x_1 + w_2x_2 + \ldots + w_Nx_N + w_b). \tag{4.2}$$

The function $y(\cdot)$ is called the *neuron activation function* (sometimes it is referred as the *transfer function*). In its simplest form proposed by McCulloch and Pitts (1943) it is defined as a step function with some threshold T:

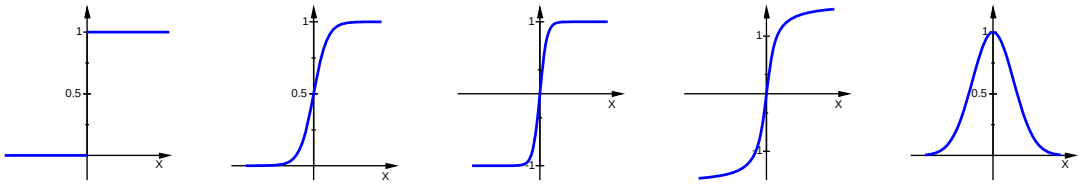
$$Y = \begin{cases} 1, & \text{if } \left(\sum_{i=1}^{N} w_ix_i + w_b\right) \geq T, \\ 0, & \text{if } \left(\sum_{i=1}^{N} w_ix_i + w_b\right) < T. \end{cases} \tag{4.3}$$

Many other activation functions were studied since then. The form of activation function is important for it is responsible for the general ability of neural network to solve the problem. Several questions related to the linearity of activation function are considered in the next subsection. Another trivial case of a linear function is the *identity function* when the input sum $\sum_{i=1}^{N} w_ix_i + w_b$ remains unchanged. In this analysis we studied the most commonly used of non-linear functions:

Sigmoid (logistic) function is defined as

$$y(X) = \sigma(X) \equiv \frac{1}{1 + e^{-X}}. \tag{4.4}$$

The sigmoid function is very popular among ANN users since it complies with all the significant requirements peculiar to both biological prototype and numerical



(a) Step (b) Sigmoida (c) tanh (d) arctan (e) Gaussian

Figure 4.3: Some neuron activation functions.

computations. Unlike the step function sigmoid is differentiable, making it easier to adapt small change in the input to an appropriate change in the output. The confinement of this function is also advantageous for numerical calculations.

Hyperbolic tangent is given by

$$y(X) = \tanh(X) \equiv \frac{e^X - e^{-X}}{e^X + e^{-X}}. \tag{4.5}$$

This function resembles all the features of the sigmoid function, except that it can produce negative output values as well as positive ones.

Less common examples of activation functions are not considered here but include arctan and Gaussian functions. The shapes of the different activation functions can be compared in Figure 4.3. It will be shown that, from the theoretical point of view one can always find an appropriate continuous function that can be used as an activation function.

4.1.2 Artificial Neural Networks

Formally an Artificial Neural Network can be defined as a collection of neurons reciprocally connected with each other in a certain way. As an example, let us consider one of the most widely used ANN structures where the neurons are connected in a way shown in Figure 4.2b. This ANN has a three-layer structure. The first, from the

left, layer is called the *input layer*. The input signals are received by the input layer and distributed farther among the neurons located in the next layer called the *hidden layer*. The last layer where we typically read the output from the neural network is called the *output layer*. Depending on the posed problem the number of hidden layers and the number of neurons in the output layer can be more than one.

From a practical point of view, it rarely makes sense to assign different activation functions to individual neurons belonging to the same layer. Without a loos of generality we will assume that the hidden and output neurons are characterized by activation functions $y_h(\cdot)$ and $y_o(\cdot)$ respectively. Let us consider the ANN with only one output, H hidden neurons, and N inputs. Thus, the number of inputs of each neuron in hidden layer is N and the number of inputs of output neuron is H. We will denote as $(w_{j1}, \ldots, w_{jN}, w_{jb})$ the weights of the j^{th} hidden neuron and the weights of output neuron as $(v_1, \ldots, v_H, v_b)$. If the signals received by ANN are $(x_1, \ldots, x_N)$ then according to Equation 4.2 one can write for the output of hidden neurons:

$$Y_j = y_h \left(\sum_{i=1}^{N} w_{ji} x_i + w_{jb} \right), \tag{4.6}$$

and for the ANN output we obtain:

$$Y = y_o \left(\sum_{j=1}^{H} v_j Y_j + v_b \right) = y_o \left(\sum_{j=1}^{H} v_j \cdot y_h \left(\sum_{i=1}^{N} w_{ji} x_i + w_{jb} \right) + v_b \right). \tag{4.7}$$

The derived equation represents a multivariate function of N arguments with $(N + 2) \times H + 1$ unknown parameters: the neuron weights. Obviously, the output returned by ANN will depend on input values as well as weights.

The first ANN with simple threshold neurons was constructed by Rosenblatt in the late 1950's. Rosenblatt's ANN did not have a hidden layer and it was capable of solving only problems with linearly separable patterns. This class of ANNs was called *perceptrons* as they could recognize the images of letters. It was noticed later that by introducing the hidden layer with extra neurons some non-linear problems could be

solved. For example, the initially non-linear relation between inputs and outputs in XOR problem can be represented by ANN inner structure in three dimensions where the problem is linearly solvable. The perceptrons with one and more hidden layers are called *multilayer perceptrons* (MLP).

Depending on the problem addressed, the ANN function can be applied for solving, among others, classification problems. Usually the problem has to be given numerically by a set of known input variables and matching outputs. By modifying the weights, the ANN function is required to reproduce the desired output with a tolerant precision for given inputs. The procedure of adjusting the weights is referred to as *training* and the ANN is said to *learn* the training sets. After successful training, one can present input data alone to the ANN and the ANN will compute an output that approximates the desired outputs as learned by training. In classification problems the desired outputs are usually discrete with values corresponding to two or more different input classes. The ANN function can be used to interpolate or even extrapolate the data after it was used to fit the training data. From a slightly different perspective, the trained ANN can be viewed as a memory device where the information about the training patterns is memorized among all the weights. Such outlook is very much favoured by cognitive scientists, who believe that a large degree of parallelism and redundancy in connections among the neurons in the brain are used to store memory. The practical side of the ANN method will be considered in details in Section 4.2. In this analysis we use ANN as a pattern classifier.

There is an important question that remains unanswered: Can one be sure that MLP will approximate any complex enough function? Amazingly, the very same question but in more strict mathematical wording was asked by David Hilbert, the father of functional analysis, almost 90 years before the world learned about MLP.

4.1.3 Thirteenth Hilbert's Problem

In order to understand the concepts of ANN more deeply we should consider the following mathematical formulae. In 1900 on the International Congress of Mathematicians at Paris the great German mathematician David Hilbert formulated 23 unsolved problems [50] which he claimed to be the most important at that time. Although not all of the "Hilbert problems" are solved, their influence on the development of mathematics during the twentieth century was significant. During the last decades the theory of neural networks was evolving rapidly, and it became clear that one of the Hilbert's problems had a strong connection with this technique. In his 13^{th} problem Hilbert questions: "Is there an analytical function of three arguments x, y, z which cannot be obtained by a finite chain of functions of only two arguments?" In this statement the finite chain of functions implies the finite number of insertions of functions, *i.e.* function $f(x, y, z) = xy + xz$ can be written as $f(x, y, z) = g(h(x, y), h(x, z))$, where $g(a, b) = a + b$ and $h(a, b) = ab$.

The 13^{th} Hilbert problem was solved by Vladimir Arnold in 1957 [51], when he was a student at the mathematical department of Moscow State University. He showed that any continuous function of three variables can be represented in terms of continuous functions of only two arguments. Shortly after this work the soviet mathematician Andrey Kolmogorov proved a much stronger theorem [52, 53]. Due to the importance of this theorem for the ANN theory we will formulate it here in its original form.

Kolmogorov theorem: For every integer dimension $N \geqslant 2$, there exist continuous real functions $h_{ij}(x)$ defined on the initial interval $U = [0, 1]$, such that, for every continuous real function $f(x_1, \ldots, x_N)$ defined on the N-dimensional unit hypercube

U^N, there exist real continuous functions $g_i(x)$ such that

$$f(x_1, \ldots, x_N) = \sum_{i=1}^{2N+1} g_i \left\{ \sum_{j=1}^{N} h_{ij}(x_j) \right\}. \tag{4.8}$$

Unfortunately, Kolmogorov theorem is not much of a practical use because it does not provide any clue about neither the specific form of h and g functions nor how to choose them. Nevertheless, this theorem shows in principle that any continuous function of several variables can be represented by a superposition of some other continuous functions of only one variable. One can note that Equation 4.8 has the same structure as the expression for the MLP output in Equation 4.7. In order to get more practical conclusions for ANN theory, it is possible to relax some of the requirements in the Kolmogorov theorem. For example, the function $f(x_1, \ldots, x_N)$ does not need to be expressed exactly as in Equation 4.8 but rather be approximated with an expression of similar form as the right part of Equation 4.8. Also, the upper limit of the first sum $(2N + 1)$ can be independent of N, *i.e.* we can increase or decrease the number of neurons in the hidden layer if necessary. This said, in 1989 several authors independently formulated [54, 55, 56] a new version of the Kolmogorov theorem specifically suited for ANN applications. Here we provide the formulation of the theorem as it appears in one of the articles.

Universal approximation theorem: For every continuous real function $f(x_1, \ldots, x_N)$ of N variables and any arbitrary small number $\epsilon > 0$, there exist an integer number H and sets of real numbers w_{ji}, w_{jb} and v_j, v_b, such that

$$|f(x_1, \ldots, x_N) - Y(x_1, \ldots, x_N)| < \epsilon \tag{4.9}$$

everywhere on $f(x_1, \ldots, x_N)$ definition interval, and $Y(x_1, \ldots, x_N)$ is given by

$$Y(x_1, \ldots, x_N) = \sum_{j=1}^{H} v_j \cdot \sigma \left(\sum_{i=1}^{N} w_{ji} x_i + w_{jb} \right) + v_b, \tag{4.10}$$

where σ is a sigmoid function.

An important conclusion can be made based on the above theorem and from the comparison of Equations 4.7 and 4.10: **a MLP with sigmoid and identity activation functions in the hidden and output layers correspondingly can reproduce any continuous function.** In practice, other activation functions can be used in order to achieve better approximation or faster training times. A particular choice of activation functions can also depend on normalization of inputs or a desired nonlinear transformation of the ANN output. In any case the functions with sigmoid-like shapes provide better performance of a MLP (*e.g.* see [57]).

4.2 Practical Issues

In the previous section we presented the basic concepts of ANN technique mostly from the theoretical point of view. A specific implementation of ANN heavily depends on the particular problem in question and, therefore, theoretical derivations can only serve as very general guidelines. In this section we will limit ourselves to the consideration of some practical issues related to the solution of the classification problem using MLP. In our analysis we prepare two classes of events: signal and background. These events are described by several variables which are used as inputs for the ANN. Each event is assigned a desired ANN output (*target value*): 0, if event belongs to background class, and 1, if event belongs to signal class. The input variable along with the corresponding targets form a *itraining pattern*. The set of all training patterns is called the *training sample*.

The ANN we use is implemented as a set of software routines in a package called Stuttgart Neural Network Simulator (SNNS, [58]). The SNNS is used by many scientists in various disciplines and was proven to provide a flexible environment with great functionality. The package also provides numerous features for ANN implemen-

tation. The software code is written in standard ANSI-C and can be compiled as a software library. In order to naturally integrate the SNNS functionality into the CDF software environment we developed an interface coupling the SNNS and ROOT ([59]) libraries.

4.2.1 Training

In order for the ANN to solve the classification problem one needs to adjust the weights so that the ANN output value is close enough to the desired one for a given vector of inputs. Since the nominal target value is known *a priori*, the procedure of adjustment of weights is called *supervised learning*. The most common supervised learning algorithm used to train an MLP is based on backward propagation of errors and referred to as *backpropagation*. The error of ANN represents a measure of the discrepancy between the ANN output and the target value, and it is defined as[1]

$$E_i = \frac{1}{2}(t_i - Y_i)^2, \tag{4.11}$$

where t_i and Y_i are the target and output values correspondingly for the i^{th} training pattern. As the algorithm's name suggests, the error E_i propagates backwards from the output layer to the input layer. First, the weights of output neuron are updated towards E_i reduction, and then, based on those modified weights, the algorithm calculates the correction factors for the weights in the previous layer, and so on until the input layer is reached. The weight changes can be applied after each training pattern is shown to the ANN or, alternatively, the weights can be updated after ANN outputs are calculated for all the patterns in the training sample. In the latter case the cumulative error function is used as a measure of current ANN performance:

$$E = \frac{\sum_{i}^{N_{\text{train}}} (t_i - Y_i)^2}{N_{\text{train}}}, \tag{4.12}$$

[1]Note that Equation 4.11 is written for a MLP with a single output, if a MLP has more outputs the errors for different outputs are summed.

where N_{train} is the number of training samples. The full cycle including cumulative error calculation and update of the weights is called *epoch*. Obviously, the goal of ANN training is to find such configuration of the weights that will make E as small as possible. Usually several hundreds of epochs are needed to achieve stable values of error function.

Technically speaking, the backpropagation is used to calculate the gradient of the error of the network with respect to the network's modifiable weights. If one considers a unit j with weight w_{ij} on i^{th} input, then according to the gradient descent the weights should be updated by

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = -\eta \frac{\partial E}{\partial \Sigma_j} \cdot \frac{\partial \Sigma_j}{\partial w_{ij}} = -\eta \delta_j x_{ij}, \tag{4.13}$$

where $\Sigma_j = \sum_i^{N_{\text{inputs}}} w_{ij} x_{ij}$ is the weighted sum of all inputs to unit j, η is a constant learning factor, δ_j is the difference between the real and teaching outputs of j^{th} unit, which specific form depends on whether unit j belongs to output of hidden layer, and x_{ij} is the i^{th} input to the j^{th} unit. For a particular case of MLP network with one hidden layer and y_o and y_h activation functions in output and hidden layers respectively, a single training epoch will include the following steps:

1. **Forward propagation phase:** Calculate the output value of the network by presenting the input pattern to the network, and consequently calculating the activations of hidden and output units.

2. **Backward propagation phase:** Calculate the errors in output of unit j:

$$\delta_j = \begin{cases} \frac{\partial y_o(\Sigma_j)}{\partial \Sigma_j}(t_j - Y_j), & \text{if } j \text{ unit is an output unit,} \\ \frac{\partial y_h(\Sigma_j)}{\partial \Sigma_j} \sum_k \delta_k w_{jk}, & \text{if } j \text{ unit is a hidden unit,} \end{cases} \tag{4.14}$$

   where index k refers to the k^{th} successor unit relative to the current j unit. Then, according to Equation 4.13 calculate the weight changes Δw_{ij}.

3. Repeat forward and backward phases for all training samples and apply the sum of all changes Δw_{ij} to the corresponding weights.

From the geometrical point of view the error function is a surface in the weight space. The backpropagation can not guarantee that the minimum of E found for a given training data set is the global rather than a local one. In practice the simplest suggested way in avoiding local minimum is to start the minimization from different starting points by initializing the weights with small random numbers.

As the Equation 4.13 suggests the step size along the error surface is proportional to the magnitude of the gradient. Such behaviour is not always beneficial since, for example, flat spots on the error surface are better be traversed rapidly with a few big steps, while the smaller step sizes are needed in the regions close to a local minimum. The SNNS package provides an enhanced version of backpropagation algorithm with an extra regularization term. This term is called the *momentum term* and it uses the old weight change as a parameter for the computation of the new weight change. The new weight change is computed as

$$\Delta w_{ij}(t+1) = -\eta \delta_j x_{ij} + \alpha \Delta w_{ij}(t), \tag{4.15}$$

where α $(0 \leq \alpha \leq 1)$ is a constant specifying the influence of the momentum. This adaption of the step size can increase the learning speed. We find that, for our data, the ANN learns fast and accurate enough if $\eta = 0.01$ and $\alpha = 0.5$ which are within typical intervals suggested by SNNS authors.

4.2.2 Choosing Optimal Structure and Variables

The number of inputs, number of hidden layers, and number of units in the hidden layers define the structure of the ANN. As it was shown from theoretical point of view one hidden layer with neurons possessing sigmoid-like activation function is sufficient

to approximate any function. The choice of nonlinear activation function in hidden units is essential in our analysis. Without nonlinearity, hidden units would not make ANN any more powerful than just a plain perceptron. This is because a linear function of linear functions is again a linear function. In our classification problem the linear transformation of input signals would be equivalent to a set of rectangular cuts in the input space. Therefore, we constrain ourselves to MLP with one hidden layer and nonlinear activation functions.

Except for the number of hidden layers the theory does not provide any clue about the number of inputs and number of hidden units. There are several "rules of thumb" relating the number of training cases to the total number of weights. For instance, a typical recommendation is to use at least 30 times more training templates than weights. Such recommendations should be considered as a crude approximation for a proper number of hidden units. The ideal number of hidden units may depend in a complex way on the number of inputs and outputs, the training algorithm, the type of activation functions, and the complexity of classification function to be learned. The best way to determine the number of hidden units is to train several networks and compare their errors.

It is not feasible to use every imaginable variable as input to ANN, since the large number of inputs will lead to exponential growth of the hypervolume of the input space. In order to evenly represent every part of the high-dimension hypervolume without significant loss in input space coverage, one would need to increase accordingly both the number of training samples and the number of weights. Hence, it is advisable to select only the most relevant input variables and prevent the waste of ANN resources representing unimportant portions of input space. In our work we follow the approach where the input variables are selected according to their separation power in one dimension and then the best of them are combined by the network.

The reprocess of the input variables can also help to find an optimal ANN solution.

In particular, we consider the normalization of the variables before they are shown to the network. The reason for data rescaling is based on the fact that different variables may be defined on different ranges, *e.g.* the H_T variable has typical values from 100 to 500, whereas sphericity of the event is defined on interval $[0, 1]$. It is not a problem in itself since the inputs can be rescaled by appropriate changes in corresponding weights, but large weights can lead to a greater risk of weight saturation while training. Normalized inputs remove the problem of scale dependence of initial weights. The deeper reasoning for the input data normalization comes from the interpretation of each hidden (or output) unit as a separating hyperplane in the N dimensional space, where N is the number of inputs to that unit. The position and orientation of such hyperplane is determined by the weights, and the bias determines the distance of the hyperplane from the origin. Initially, all the hyperplanes are oriented in random directions, and pass close to the origin. If the data points are not centered at the origin, the hyperplanes may fail to pass through the data cloud and the chance of getting stuck in local minimum increase. In our analysis we normalize the input variables by subtracting the mean, and dividing by the standard deviation, thereby obtaining a normal variables with mean 0 and standard deviation 1.

The choice of optimal structure of an ANN is closely related to ANN ability to perform accurately on data which is not in the training set.

4.2.3 Generalization

During learning, the outputs of a supervised neural network come to approximate the target values given the inputs in the training set. This ability may be useful in itself, but more often the purpose of using a neural network is to generalize, *i.e.* to have the outputs of the network approximate target values given inputs that are not in the training set.

Since our goal is to find the network having the best performance on new data, the simplest approach to the comparison of different networks is to evaluate the error function using data which is independent of that used for training. Various networks are trained by minimization of an appropriate error function defined with respect to a training data set. The performance of the networks is then compared by evaluating the error function using an independent *validation set*, and the network having the smallest error with respect to the validation set is selected. This approach is called the *hold out method.* On the other hand, the validation set is used to control overfitting of ANN and, therefore, the hold out procedure can itself lead to a biased decision with respect to the validation set. The performance of the selected network should be confirmed on a third independent set of data called a *test set.* It is worth noticing that the error function for the training set, given by Equation 4.12, can be calculated also for both the validation and test sets at the end of each epoch. These errors are referred to as *validation error* and *generalization error* respectively.

Like many other nonlinear fitting techniques the ANNs can suffer from either underfitting or overfitting of the data. The underfitting usually happens when a network does not have enough inner parameters to fully represent the true classification function. In contrary, the overfitting happens when a network is too complex and it can learn undesired particulars of the finite training sample, *e.g.* noise. Apparently, the overtrained ANN should not be expected to provide a good generalization. In order to avoid underfitting and overfitting we utilize the *early stopping* approach. The idea of this method is based on the calculation of the validation error during the training, for instance, every time the weight are being updated. Since the validation set is independent of the training one, the validation error rate is expected "to go up" at the point where further training will cause an overfitting. The validation error can go up and down several times during the training, hence we train the network to convergence and then find an epoch where the network produced the lowest validation error.

Chapter 5

Measurement of the Branching Ratio $R = \frac{\mathcal{B}(t \to Wb)}{\mathcal{B}(t \to Wq)}$

In Chapter 3 we described the selection of signal-rich dataset collected at CDF. The identification of jets produced by b quarks is important for this analysis and was presented in details. We use this information to split the selected events in three groups based on the number of b-tagged jets. In Chapter 4 we discussed the ANN technique and its application to classification problems. In this Chapter we first describe the application of ANN to the determination of the number of $t\bar{t}$ events in lepton-plus-jets channel. We then discuss the sources of systematic uncertainties affecting the ANN output shape and, hence, the measurement of signal and background contributions. The background contamination is also determined using a direct measurement in lepton-plus-jets and dilepton channels. Subsequently, we present the measurement of R using the above information. We finally, establish the lower confidence limit on the branching ratio using the Feldman-Cousins method.

5.1 Top Quark Signal Determination in the Lepton-plus-jets Channel

In the lepton-plus-jets analysis we use ANN to estimate the number of signal and background events in each tagged subsample.

5.1.1 Signal and Background Modeling

For the proper modeling of the data events passed the lepton-plus-jets selection requirements, we use a combination of Monte Carlo simulation and data. The Monte Carlo samples used in this analysis undergo complete detector simulation[1]and standard event reconstruction as if it was real data. The main background to $t\bar{t}$ production in the lepton-plus-jets channel comes from W production in association with jets. The background from generic QCD multi-jet processes is the second largest contributor. Much smaller contributions come from other electroweak processes resulting in at least one high-p_t lepton and jets. We will further describe the $t\bar{t}$ signal and W+jets Monte Carlo generation since these samples are also used for the ANN training. The modeling and estimation of QCD multi-jet background is performed using an independent data sample.

The $t\bar{t}$ events in our default signal sample are generated with HERWIG Monte Carlo program [62], which has a leading order matrix element for the parton hard-scattering convoluted with CTEQ5L parton distribution functions [63]. The top quark mass is assumed to be 175 GeV/c^2 and the default $W \to l\nu$ branching fraction is 11.1% for each lepton flavour l. HERWIG includes initial and final state radiation, which cannot be turned off, and this appears to be a primary concern in systematic

[1]The detector simulation is performed with the CDFSim package [60] developed by the CDF collaboration. The CDFSim is based on GEANT program [61] simulating the passage of particles through the matter.

uncertainty associated with our use of HERWIG. We use PYTHIA generator [62] to study the systematic effects described in Section 5.1.5.

We use the ALPGEN matrix element generator [64], convoluted with the CTEQ5L parton distribution functions, to generate $W + 4$ partons ($W + 4$p) in the final state. We require parton $|\eta| \leq 3.0$, $p_T \geq 8$ GeV/c, and a minimum separation $\Delta R \geq 0.2$ between u, d, s, and g partons at the generator level. Our default momentum transfer scale for the PDFs is $Q^2 = M_W^2 + \sum_i p_{T,i}^2$, where $p_{T,i}$ is the transverse momentum of the i^{th} parton. For parton shower evolution to the colorless hadrons, the ALPGEN information is transfered to the HERWIG where the hadron jets are adequately modeled with parton shower algorithms. It is assumed that the gluon radiation can correctly model extra jets in the final state. Our generated sample contains W associative production with all parton flavours, but only a small fraction of W bosons is produced in association with heavy flavour partons. The lack of statistics becomes sufficient if the b-tagging is required. Thus, we generate $W + b\bar{b} + 2$ partons MC samples to study the $W +$ jets background in the tagged subsamples.

Theoretical predictions for the total rate of the generic QCD processes can not be easily obtained. A common solution is to model the kinematics of the QCD multi-jet background with data events that pass all of our standard selection requirements except for the requirement on lepton isolation, I, which in case of QCD is inverted $I > 0.2$. This requirement is consistent with the typical signature of generic QCD events, *i.e.* large amount of additional energy deposited in the cone around the lepton and low missing transverse energy. The fraction of QCD multi-jet events, in the sample of events passing the default cuts, can be estimated by assuming that the lepton isolation and the missing transverse energy $\not\!\!E_T$ are uncorrelated. By dropping off the lepton isolation and $\not\!\!E_T$ cuts from the list of our default selection requirements we parameterize the events in two dimensional space: Isolation versus $\not\!\!E_T$. Then we define the following regions:

- region A: $I > 0.2$ and $\not\!\!E_T < 10$ GeV,
- region B: $I < 0.1$ and $\not\!\!E_T < 10$ GeV,
- region C: $I > 0.2$ and $\not\!\!E_T > 20$ GeV,
- region D: $I < 0.1$ and $\not\!\!E_T > 20$ GeV.

The fraction $f_{\rm QCD}$ of QCD multi-jet events in the signal region D is deduced from the ratio of event populations found in the above control regions:

$$f_{\rm QCD} = \frac{N_{\rm B} \times N_{\rm C}}{N_{\rm A} \times N_{\rm D}}. \tag{5.1}$$

We use the above procedure to calculate the QCD background fractions for all b-tag subsamples. The results are shown in Table 5.1. We assume a $\pm 50\%$ systematic uncertainty in the above estimation which is translated to a $\pm 50\%$ variation in the normalization of the ANN QCD output shape (see Section 5.1.5). We also correct the QCD background ANN shape for the expected contributions from $t\bar{t}$ events in the non-isolated region.

Table 5.1: Number and the fraction of QCD multi-jet events in pretag, 0-tag, 1-tag, and 2-tag data subsets.

	Pretag	**0-tag**	**1-tag**	**2-tag**
Number of QCD events	10.7	8.9	1.0	0.1
QCD fraction (%)	10.0	11.3	4.4	2.0

5.1.2 Kinematic Discriminators

The selection of appropriate ANN input variables is essential for using the ANN technique. Our choice of ANN inputs is based on kinematic differences between signal and background processes. One of the most essential characteristics of signal events is

the production of two massive particles. The top quarks are produced almost at rest, and, therefore, their decay products are likely to be distributed uniformly in space, causing final state particles to be distributed more spherically that the background ones. Also, for the same reason, the signal events are expected to have jets with higher transverse energy on average. Although, the W + jets background can well mimic the signal highly energetic lepton and missing energy, the jets with the highest energies typically come from other QCD processes, and their reconstructed invariant mass can be used for discriminating from $t\bar{t}$ events.

The complete list of the initially considered kinematic variables and their short definitions are presented in Table 5.2. Not all of these variables will be used as inputs in the final network. The most discriminating inputs are determined at the training procedure, as described in the next section. Since the training of ANN is performed only on $t\bar{t}$ and W + jets samples, we compare their distribution shapes in Figures 5.1–5.5.

Table 5.2: The list and description of all the kinematic variables considered as inputs to ANN in this analysis. The reader may refer to the text for detailed definitions of these variables.

No.	**Variable**	**Description**[2]
0	$E_T(j_{\rm 3rd})$	E_T of the third jet.
1	$M^{\rm min}(j_k, j_l)$	Minimum invariant mass of a jet pair among the four first jets.
2	$E_T(j_{\rm 2nd})$	E_T of the second jet.
3	H_T	Sum of the transverse energies of the first four jets, the lepton and the missing energy.
4	$E_T(j_{\rm 1st})$	E_T of the first jet.
5	$E_T(j_{\rm 4th})$	E_T of the fourth jet.
6	$M_T^W(j_k, j_l)$	Transverse mass of the jet pair closest to the W mass.

[2]The jets in event are ordered by E_T in descending order.

Table 5.2: (continued)

No.	Variable	Description
7	$\frac{\sum P_z}{\sum P_T}$	Sum of the z momentum-components of the first four jets, the lepton and the missing energy divided by the sum of corresponding transverse components.
8	Aplanarity	Aplanarity of the event.
9	$\langle\eta^2\rangle$	Arithmetic average of squared pseudo-rapidities of the first four jets.
10	$\frac{\Delta R^{\min}(j_k,j_l)E_T^{\min}}{E_T(l)}$	Separation in η–ϕ space.
11	$M^W(j_k,j_l)$	Invariant mass closest to the W mass of a jet pair among all jets in an event.
12	$\Delta R^{\min}(j_k,j_l)$	Minimum opening angle, $\Delta R = \sqrt{(\phi_{j_k}-\phi_{j_l})^2+(\eta_{j_k}-\eta_{j_l})^2}$, between a jet pair among all jets in event.
13	$\not\!\!E_T$	Missing transverse energy in the event.
14	$\lvert\eta\rvert_{\max}$	Maximum absolute value of pseudo-rapidity among all jets in the event.
15	Sphericity	Sphericity of the event.
16	$M_T^W(l,\nu)$	Transverse mass of the high-P_T lepton and neutrino ($\not\!\!E_T$)

Some of the chosen variables are the simple variables like the transverse energies of the jets, E_T(jet), event missing transverse energy, $\not\!\!E_T$, and jet pseudo-rapidities, η. However, additional information about an event can be obtained by either constructing composite variables or by calculating variables reflecting the features of the event as a whole. For example, the *invariant mass*, M, can be calculated for a pair of particles, for which the full four-momentum, p, is known. For the construction of variables $M^{\min}(j_k,j_l)$ and $M^W(j_k,j_l)$, we calculate the invariant mass for a pair of jets (k^{th}, l^{th}), using the formula:

$$M = \sqrt{(p_k+p_l)^2} = \sqrt{(E_k+E_l)^2-(\vec{p}_k+\vec{p}_l)^2}. \tag{5.2}$$

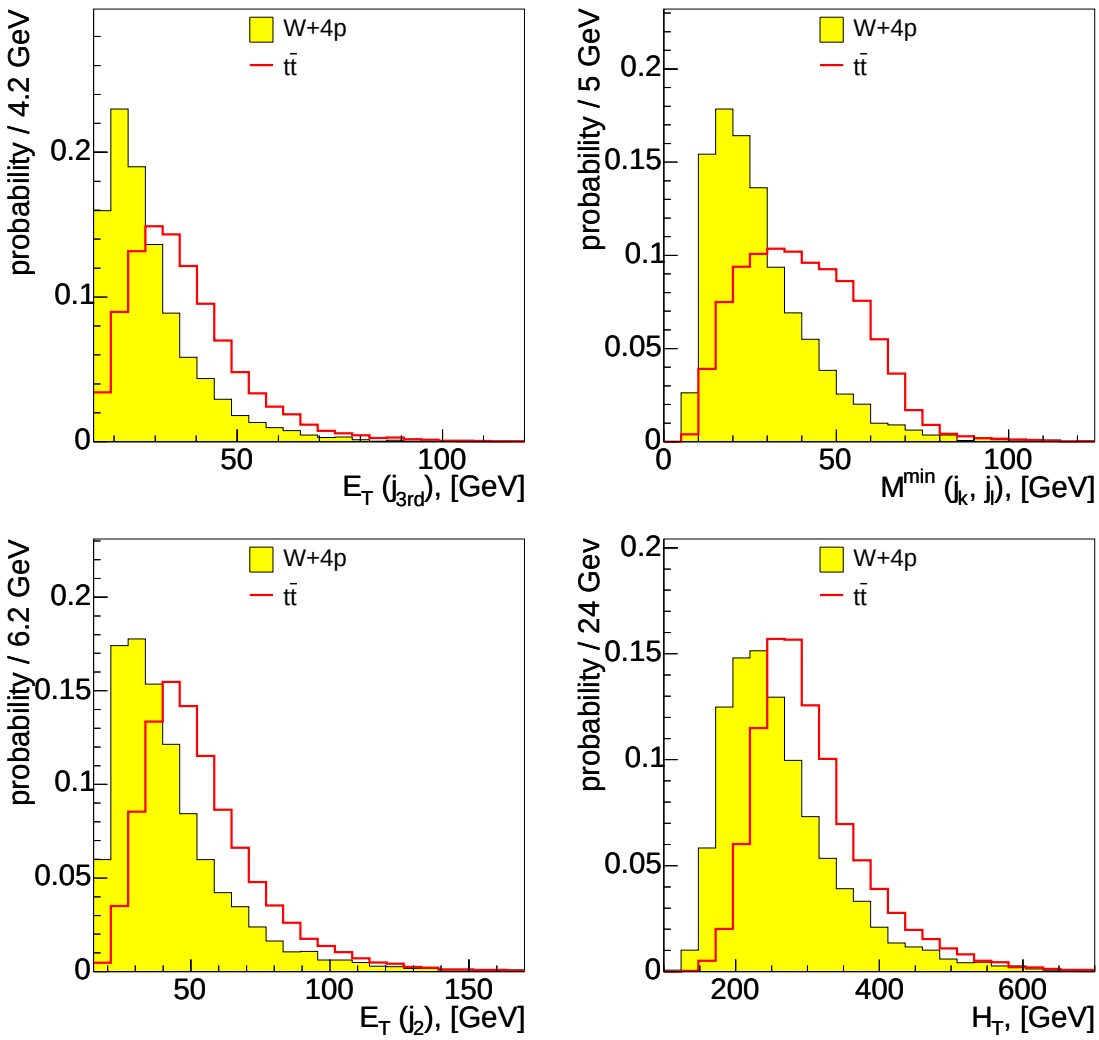


Figure 5.1: Distributions of the considered ANN input variables $E_T(j_{3\mathrm{rd}})$, $M^{\mathrm{min}}(j_k, j_l)$, $E_T(j_{2\mathrm{nd}})$, and H_T. The filled and the empty histograms represent $W + 4\mathrm{p}$ and $t\bar{t}$ Monte Carlo samples correspondingly. The distributions are normalized to the unit area. These variables will be used as inputs to the ANN, as discussed in Section 5.1.3.

The calculation of the *transverse invariant mass*, M_T does not require the knowledge of the z component of particles momenta. For a pair (k^{th}, l^{th}) of jets with transverse momenta $p_{T,i}$ and azimuthal angle $\phi_i = \arctan(\frac{p_{y,i}}{p_{x,i}})$, the transverse mass $M_T(j_k, j_l)$

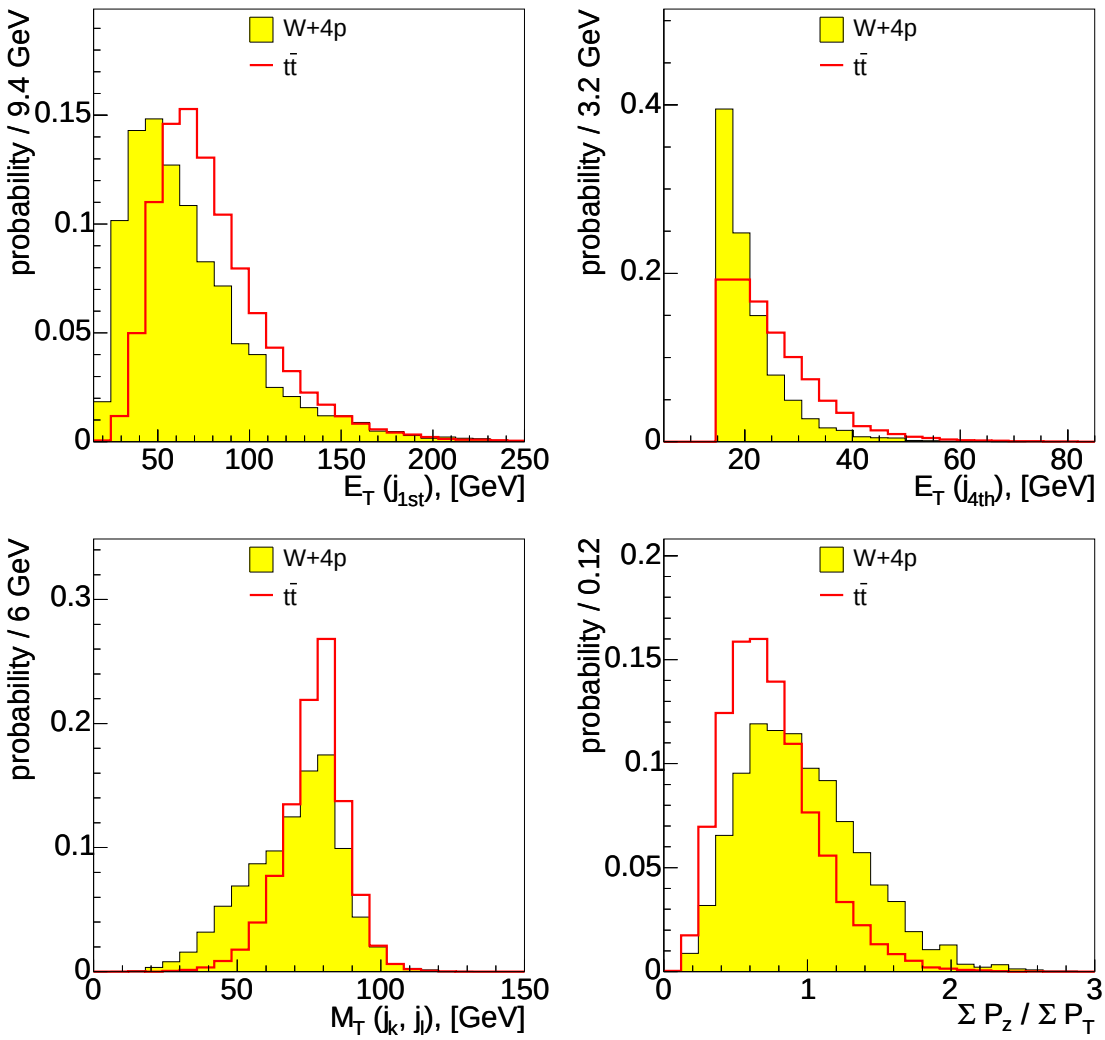


Figure 5.2: Distributions of the considered ANN input variables $E_T(j_{1\text{st}})$, $E_T(j_{4\text{th}})$, $M_T(j_k, j_l)$, and $\sum P_z / \sum P_T$. The filled and the empty histograms represent $W + 4\text{p}$ and $t\bar{t}$ Monte Carlo samples correspondingly. The distributions are normalized to the unit area. These variables will be used as inputs to the ANN, as discussed in Section 5.1.3.

is given by

$$M_T = \sqrt{2 \cdot p_{T,k} \cdot p_{T,l} \cdot (1 - \cos(\phi_k - \phi_l))}. \tag{5.3}$$

The transverse mass of a lepton and the neutrino ($\not\!\!E_T$), $M_T(l, \nu)$, is calculated in a similar manner.

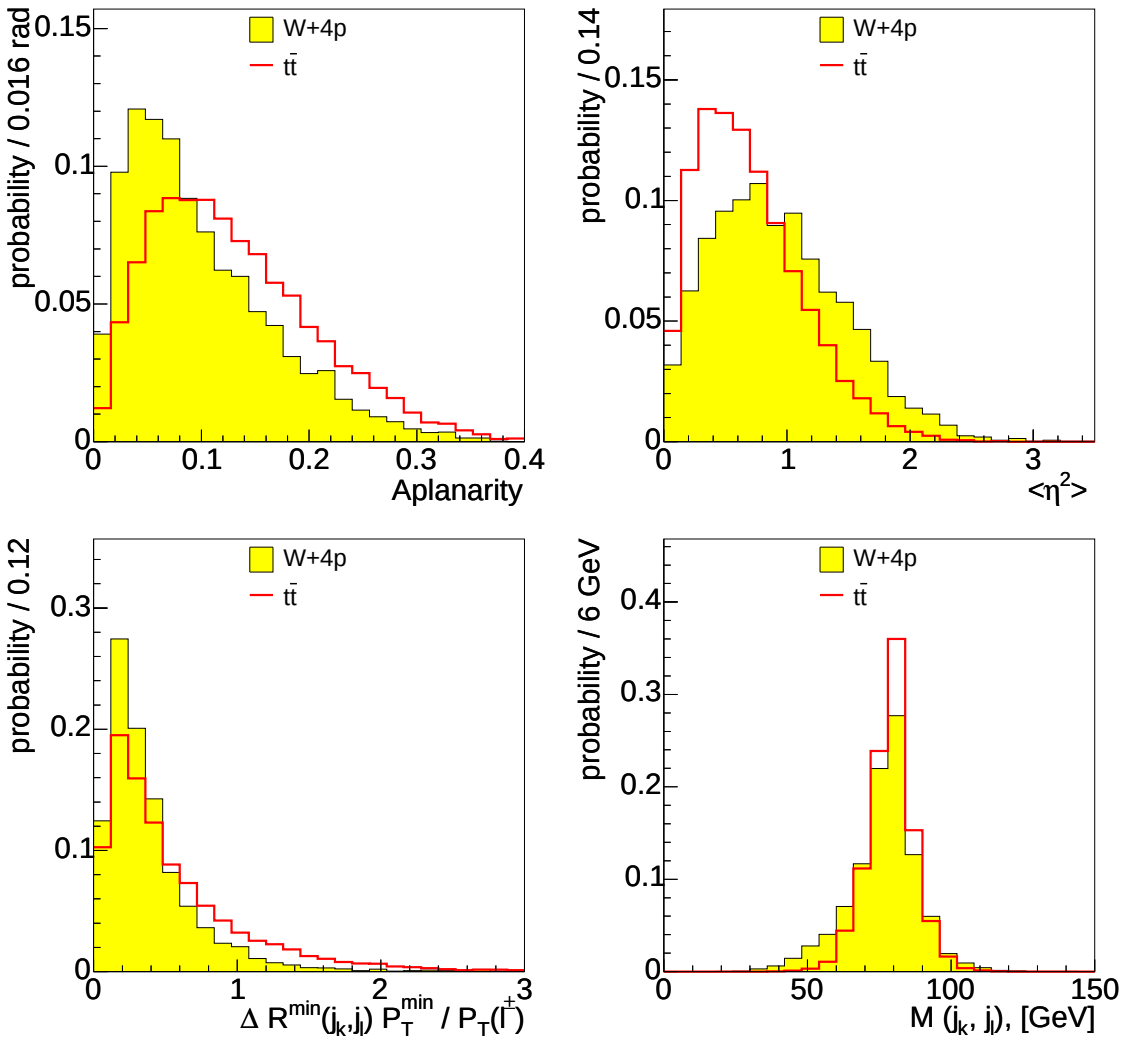


Figure 5.3: Distributions of the considered ANN input variables Aplanarity, $\langle\eta^2\rangle$, $\frac{\Delta R^{\min}(j_k,j_l)E_T^{\min}}{E_T(l)}$, and $M(j_k,j_l)$. The filled and the empty histograms represent $W+4$p and $t\bar{t}$ Monte Carlo samples correspondingly. The distributions are normalized to the unit area. The Aplanarity will be used as input to the ANN, as discussed in Section 5.1.3.

The *summed transverse energy*, H_T, tends to be higher for $t\bar{t}$ events than $W+$ jets events even after the application of all the cuts. For the calculation of this quantity we consider the transverse momenta of lepton, missing energy, and first four

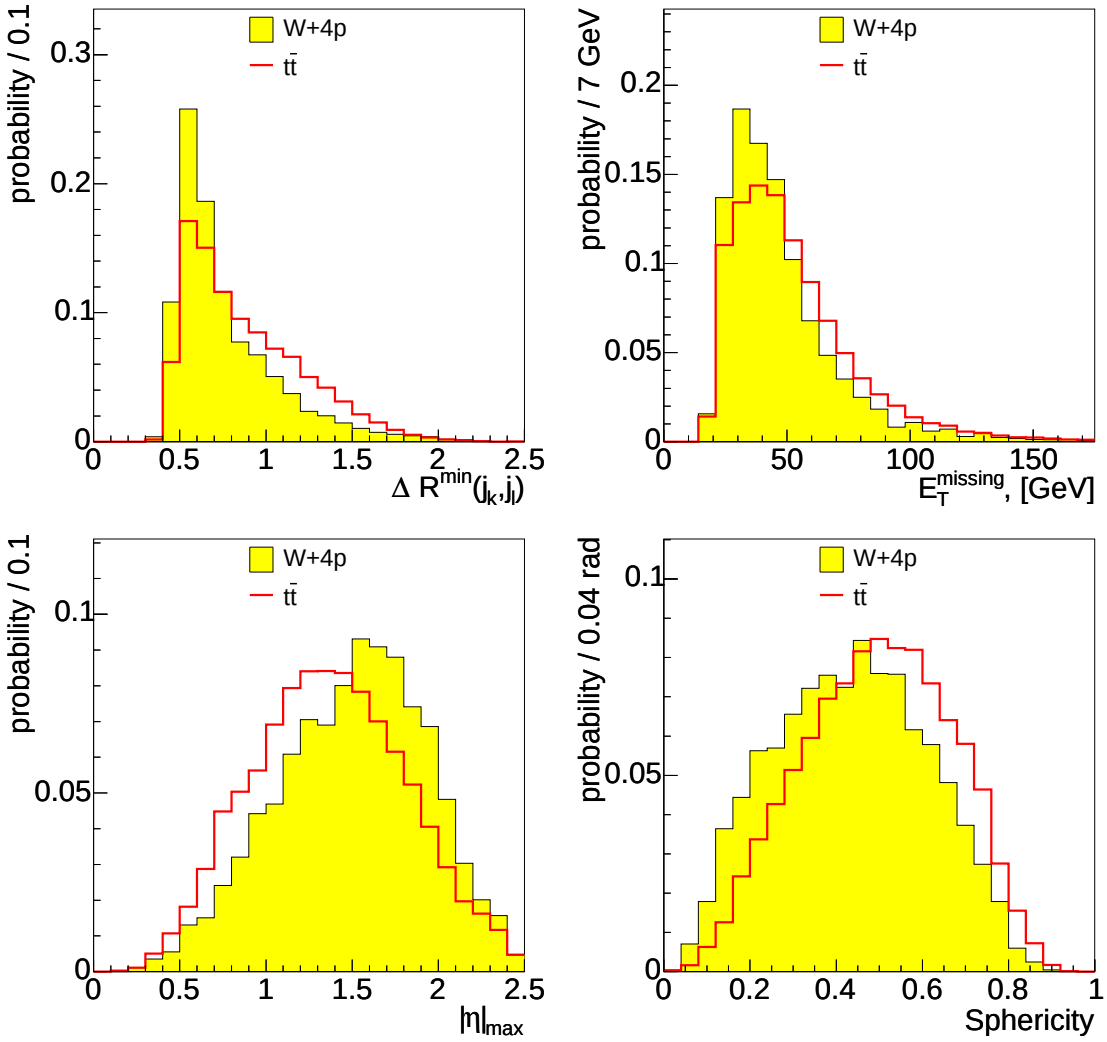


Figure 5.4: Distributions of the considered ANN input variables $\Delta R^{\min}(j_k, j_l)$, $\not\!\!E_T$, $|\eta|_{\max}$, and sphericity. The filled and the empty histograms represent $W + 4\text{p}$ and $t\bar{t}$ Monte Carlo samples correspondingly. The distributions are normalized to the unit area.

jets:

$$H_T = E_T(l) + \not\!\!E_T + \sum_{i=1}^{4} E_T(j_i). \tag{5.4}$$

The separation $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ of a pair of jets in η–ϕ space can depend

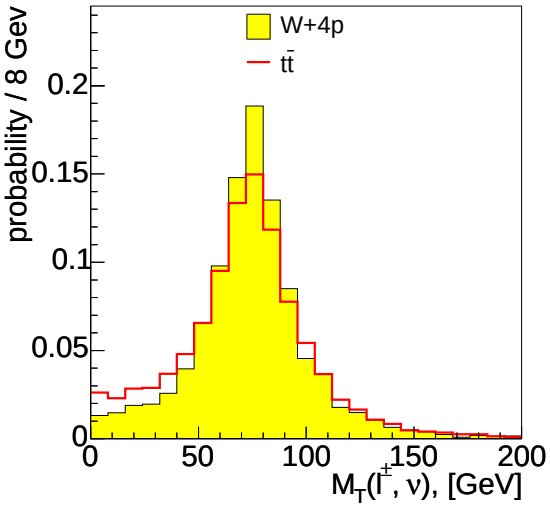


Figure 5.5: Distributions of the considered ANN input variable $M_T(l^{\pm}, \nu)$. The filled and the empty histograms represent $W + 4\text{p}$ and $t\bar{t}$ Monte Carlo samples correspondingly. The distributions are normalized to the unit area.

on their production process. Two jets produced through gluon radiation usually have smaller separation. We calculate the minimum separation, $\Delta R^{\min}(j_k, j_l)$, between any two jets passing our cuts. The same quantity is also used in the compound variable $\Delta R^{\min} \frac{E_T^{\min}(j)}{E_T(l)}$, where the separation $\Delta R^{\min}$ is scaled by the ratio of the lower jet E_T in the pair to lepton E_T.

The topology of signal and background events can be described by two extra variables: *Aplanarity* and *Sphericity*. The definition of both variables is based on the calculation of the normalized momentum tensor:

$$T_{ab} = \frac{\sum\limits_{q=j_{1,\ldots,5},l,\nu} p_a(q) \cdot p_b(q)}{\sum\limits_{q=j_{1,\ldots,5},l,\nu} |\vec{p}(q)|^2}, \tag{5.5}$$

where a and b run over the three spacial dimensions and $p(q)$ is the momentum of particle q. In our case the sum over q includes the five highest jets, the lepton, and the neutrino. T_{ab} is a 3×3 symmetric matrix which can be diagonalized resulting in three eigenvalues $Q_1 \leq Q_2 \leq Q_3$. The Sphericity and Aplanarity are defined in terms

of the two lowest eigenvalues as:

$$\text{Aplanarity} = \frac{3}{2}Q_1, \tag{5.6}$$

$$\text{Sphericity} = \frac{3}{2}(Q_1 + Q_2). \tag{5.7}$$

Small values of Aplanarity correspond to coplanar or collinear shape of event. The Aplanarity can be considered as the fraction of the total momentum out of the highest p_T plane. The Spericity values lie in a range of 0 to 1, with larger values corresponding to more spherical event shapes and smaller values corresponding to cigar-like shapes. Similarly, the Sphericity can be interpreted as the fraction of the total momentum out of the lowest p_T plane.

Several of our kinematic variables require the knowledge of the z component of the neutrino. In general, the problem with measuring the missing longitudinal energy, $\not\!\!E_z$, is that the total particle momentum along the beam direction is not known due to particles lost along the beam-pipe. However, in $t\bar{t}$ events, we can assume that most of the $\not\!\!E$ is due to the neutrino from the leptonically decaying W boson. Since in lepton-plus-jets channel only one neutrino is present, we can reconstruct the invariant mass of W boson which is known for W produced on shell. Therefore, we can deduct the third component of the neutrino's momentum from the equation:

$$M_W^2 = (p_l + p_\nu)^2, \tag{5.8}$$

where p_l and p_ν are the four-momenta of the lepton and neutrino respectively. The quadratic equation leads to two possible solutions for $\not\!\!E_z$. Once real solutions are produced, the smaller solution is chosen for the use in our kinematic variables, since this one is the actual solution in 70% of the times [65].

A number of statistical tests are performed on the kinematic variables in Table 5.2 to help us decide which ones have the highest separation power. There are many universal statistical methods that allow us to verify whether two data sets are drawn

from the same underlying distribution function or not. The most generally accepted test for continuous data is the *Kolmogorov-Smirnov (KS) test* and for binned data it is the *chi-square* test. We can also determine if two data sets are drawn from functions with the same mean or variance by applying the *student's t-test* and *f-test* respectively [66]. In this thesis, we use the KS test applied on signal $t\bar{t}$ and background $W+$jets distributions to determine the discriminating power or the proposed kinematic variables. We actually have conducted all of the mentioned statistical tests on our variables and found the results consistent with each other. The KS test is based on a construction from the data points of an unbiased estimator for the cumulative distribution function. Then the two cumulative distribution can be compared by introducing some measure. In the case of KS statistics, the measure is defined as the maximum value of the absolute difference between two cumulative distributions. The detailed description of KS test can be found in [66]. We sort the input variables in descending order by their discrimination power as determined by the unbinned KS test (Table 5.3). In the next section we explain how the compiled list of variables is used for ANN architecture optimization.

Table 5.3: List of the kinematic variables sorted by the maximum distance (highest discriminating power) as calculated from Kolmogorov-Smirnov statistical test.

No.	**Variable**	**KS test $\mathbf{d_{max}}$**
0	$E_T(j_{\mathrm{3rd}})$	0.358
1	$M^{\mathrm{min}}(j_k, j_l)$	0.342
2	$E_T(j_{\mathrm{2nd}})$	0.301
3	H_T	0.300
4	$E_T(j_{\mathrm{1st}})$	0.244
5	$E_T(j_{\mathrm{4th}})$	0.268
6	$M_T^W(j_k, j_l)$	0.219
7	$\frac{\sum P_z}{\sum P_T}$	0.239
8	Aplanarity	0.190
9	$\langle \eta^2 \rangle$	0.198
10	$\frac{\Delta R^{\mathrm{min}}(j_k,j_l) P_T^{\mathrm{min}}}{P_T(l^{\pm})}$	0.163
11	$M^W(j_k, j_l)$	0.109
12	$\Delta R^{\mathrm{min}}(j_k, j_l)$	0.186
13	$\not{E}_T$	0.125
14	$\lvert\eta\rvert_{\mathrm{max}}$	0.154
15	Sphericity	0.124
16	$M_T^W(l, \nu)$	0.076

5.1.3 Training of the ANN

The purpose of the ANN is to distinguish signal events from background events. Although there are several different backgrounds to $t\bar{t}$ production, the $W+$jets background is the dominant one. Thus, the ANN is trained on $t\bar{t}$ and $W+4$p Monte Carlo samples. Based on the information in Chapter 4, we choose the basic network architecture to have one hidden layer and one output providing continuous values between 0 and 1. For the ANN implementation, we use the SNNS [58] software libraries. We developed a software that interfaces the SNNS with ROOT [59] data analysis tool. For ANN learning, we use the backpropagation algorithm which is widely accepted and well-justified for training a network as a classificator.

The training of ANN is performed on 5000 MC events consisting of equal number of events from the $t\bar{t}$ and $W+4$p samples. We also compile a test and validation sets, 1000 events each, consisting of equal number of $t\bar{t}$ and $W+4$p events. The training, validation, and test samples are constructed with no common events to avoid biases. Signal events are trained for target 1, while background events are trained for target 0 (see Table 5.4).

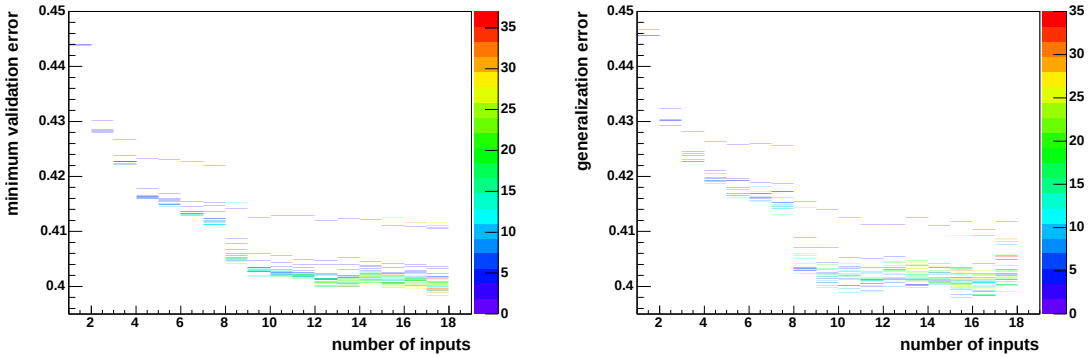
Table 5.4: Split of $t\bar{t}$ and $W+$jets MC samples into three statistically independent sets: training, validation, and test.

Sample	**Target**	**Training**	**Validation**	**Test**
$t\bar{t}$	1	2500	1000	1000
$W+4p$	0	2500	1000	1000

We choose not to use any automatic optimization algorithm for the determination of the best number of network inputs and hidden units. Instead, we train many independent networks and then rely on their generalization errors for choosing the one with the best performance. The networks are built up by sequentially increasing the dimension of input space while varying the number of hidden units. Thus, we

start with networks having only one input corresponding to the first variable in the KS-ordered list (Table 5.3), then we consider networks with two inputs corresponding to the first two variables from the KS-ordered list, and so on, until, finally we train networks with as many inputs as there are available variables. For each number of inputs N the number of hidden units is varied from 1 to $2N + 1$ resulting in $2N + 1$ separate networks[3]. The weights of every ANN are initialized by small random numbers before training procedure begins. The training, validation, and test distributions are standardized in advance by subtracting the mean and dividing by the standard deviation of the training set. The networks are trained until the moving average validation error starts to increase significantly relative to its minimum value. Typically, this convergence occurs after about 300 training epochs. The training samples are randomly shuffled before each epoch to mix signal and background events and avoid memorizing of particular order of the patterns by ANN. At every epoch the error function is calculated on all three samples. The value of the training error function can not be used to make any decision on the ANN training. We use the validation error to determine an epoch at which trained ANN started to loose the generalization ability, *i.e.* when the validation error function starts to go up. At this point the ANN weight configuration is saved along with the error values. The calculation of the error function on the independent test sample is essential since the validation sample was already used to make a decision on proper ANN training. We can choose the best network architecture by plotting the validation and generalization errors as functions of number of inputs and hidden units as presented in Figure 5.6. The plots show the desired decrease in both errors when new input variables are added to the network. An increase in ANNs performance, *i.e.* the decrease in both validation and generalization errors, is pronounced for the networks with up to 8 input variables. The networks with more than 9 inputs do not indicate significant improvement in both validation and generalization errors. One also can notice that for each given number

[3]The maximum of $2N + 1$ is inspired by the Kolmogorov theorem, *v.s.* Section 4.1.3



(a) Validation errors. (b) Generalization errors.

Figure 5.6: Each color line corresponds to a trained network with a color representing the number of hidden units. The validation and generalization errors are calculated at the epoch when ANN overtraining is indicated from the increase in validation error function.

of inputs the worst performance is shown by ANNs with very small number of hidden units. Moreover, for 9 inputs, the best performance is observed for the network with more than 10 hidden units. Therefore, as a reasonable trade-off between ANN complexity and its good performance, we decide to chose the network with 9 inputs and 10 hidden units.

Figure 5.7 illustrates the development of the training and validation errors for the chosen 9-10-1 network. The weight configuration of this ANN is saved at the 60^{th} epoch when the validation error has the lowest value. Although the training error is decreasing, further training would lead to overfitting of the data.

5.1.4 Measurement of the Top Signal Fraction

The ANN described in Section 5.1.3 is a very useful tool that produces a one-dimensional distribution for any simulated or measured data set. Now we create ANN output distributions for the CDF data and all the signal and backgrounds MC

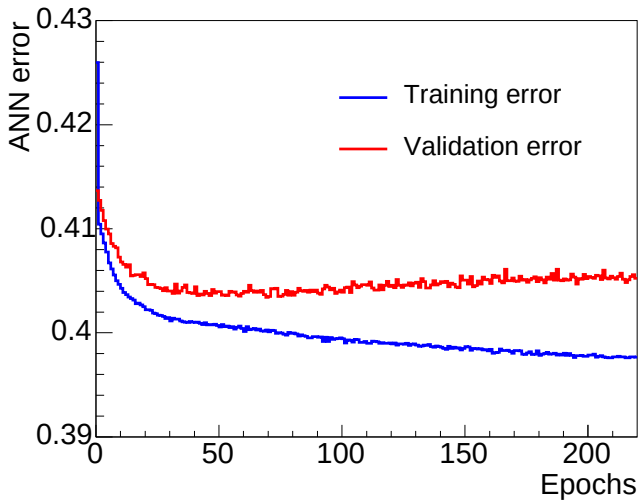


Figure 5.7: Change in training and validation errors for 9-10-1 neural network. The minimum of validation error indicates the best configuration of ANN weights.

sets by simply propagating the corresponding samples through the trained ANN.

There are at least two available methods to estimate the number of signal events in the data sample using the data and MC ANN outputs. The first approach is to set a classical cut on the ANN output by choosing a single value where the significance and the purity for such value (as determined from the contributions of signal and background MC ANN outputs) would be as small as possible in accordance with the Neyman-Pearson theorem (see *e.g.* [67]). Once the optimal ANN output value is determined, the cut at this value can be applied to the data ANN output and thereby define a number of signal events in this data sample. One of the drawbacks of this method is that we need to know the normalization for each MC output templates, *i.e.* cross section for each component in the model. Another disadvantage is that we would not use the information about the shapes of the ANN output. The better way to exploit the shape information is to fit the data shape and let the fitter decide the appropriate weights for the signal and background.

According to our hypothesis we model the data with three major constituents:

$t\bar{t}$ signal, W+jets and QCD multi-jet backgrounds. The ANN output template for QCD multi-jet background is modeled with the non-isolated data ($I > 0.2$, *v.s.* Section 5.1.1), whereas the $t\bar{t}$ and W + jets templates are determined by propagating an independent test sample through the network. Following our CDF data categorization, we create three separate $t\bar{t}$ and W + jets subsets with 0, 1, and 2 b-tags per event respectively. We subsequently propagate them through the ANN to acquire the output shapes shown in Figures 5.8 and 5.9 for W + jets and $t\bar{t}$ respectively. In these figures we compare the 0-tag, 1-tag, and 2-tag ANN output distributions with the respective pretagged ones. One can conclude that these shapes do not change considerably as we move from the pretagged to tagged samples, however, in the case of our default W + jets sample, we run out of statistics in the 1-tag, and 2-tag cases. At the same time most of the background in those bins comes from the production of W in association with heavy flavor ($W + b\bar{b} + 2\text{p}$). Figure 5.10 shows the distribution of the ANN output for $W + b\bar{b} + 2\text{p}$ for the 0-tag, 1-tag and 2-tag subsamples, compared to the pretag shape of the W + jets. From the poor agreement between these shapes, we realize that it is not sufficient to use the W + jets shape to model the heavy flavor in the 1-tag and 2-tag bins, and we thus use the $W + b\bar{b} + 2\text{p}$ shapes. The similar comparisons of pretagged and tagged ANN output shapes are done for the QCD multi-jet samples shown in Figure 5.11. Since there is no noticeable discrepancy, we use the pretagged QCD sample ANN output shape as the shape of this background in the tagged sets, for increased statistics. Summarizing, for the fitting of the data in the 0-tag case, we use the pretag shapes of $t\bar{t}$ and W + jets, since they are almost identical. For the 1-tag and 2-tag cases, we fit the data using the $W + b\bar{b} + 2\text{p}$ (instead of the and W + jets MC ANN output shapes), and also using the 1-tag and 2-tag $t\bar{t}$ MC ANN output shapes respectively, since our MC statistics allows us. The pretagged QCD shape is used across all the subsamples.

To fit the ANN output distributions we use the binned maximum likelihood method. The fit is performed in every subsample including pretag sample. Since

the statistics of the binned data outputs is low, we choose the Poisson distribution as a probability distribution function (p.d.f.) for the number of events in each bin, and the likelihood function takes the form:

$$L = \prod_{b=1}^{N_{\rm bins}} \frac{(N_b^{\rm exp})^{D_b} \cdot e^{-N_b^{\rm exp}}}{D_b!}, \tag{5.9}$$

where D_b is $b^{\rm th}$ bin content in the data histogram and $N_b^{\rm exp} = f_{t\bar{t}} S_b^{t\bar{t}} + f_{W+{\rm jets}} B_b^{W+{\rm jets}} + f_{\rm QCD} B_b^{\rm QCD}$ is the three-component sum of contributions from $t\bar{t}$, $W+{\rm jets}$, and QCD histograms in the same bin weighted by fit parameters f_α. It is convenient to minimize the negative logarithm of the likelihood L instead of maximizing L itself. Equation (5.9) is then transformed into

$$\begin{aligned} -\ln L^{\rm fit} = & \sum\nolimits_{b=1}^{N_{\rm bins}} \left(f_{t\bar{t}} S_b^{t\bar{t}} + f_{W+{\rm jets}} B_b^{W+{\rm jets}} + f_{\rm QCD} B_b^{\rm QCD} \right) \\ & - \sum\nolimits_{b=1}^{N_{\rm bins}} \left(D_b \ln(f_{t\bar{t}} S_b^{t\bar{t}} + f_{W+{\rm jets}} B_b^{W+{\rm jets}} + f_{\rm QCD} B_b^{\rm QCD}) \right), \end{aligned} \tag{5.10}$$

where we dropped a constant factorial term. The minimization of the negative logarithmic likelihood is performed using the MINUIT [68] library which is a part of the ROOT package. During the minimization, the QCD fraction $f_{\rm QCD}$ is kept constant at the estimated QCD multi-jet background rate (Table 5.1). Since the total number of events is fixed by the observed event yields in each tagged bin (Table 3.5), the $W+{\rm jets}$ background fraction can be constrained as $f_{W+{\rm jets}} = 1 - f_{t\bar{t}} - f_{\rm QCD}$ leaving $f_{t\bar{t}}$ as the only one free parameter in the fit.

Figures 5.12 to 5.15 show the results of the binned likelihood fit along with the shape of the negative logarithm likelihood. The summary of final results for the measurement of the number of signal events by fitting the ANN output shapes are shown in Table 5.5. The signal rate in pretag subsample is consistent with the similar result obtained for the cross section measurement performed at CDF [45].

Table 5.5: Number and the fraction of signal events as measured with 3-component binned likelihood fit in pretag, 0-tag, 1-tag, and 2-tag data subsets. Statistical uncertainty is shown.

	Pretag	**0-tag**	**1-tag**	**2-tag**
Number of signal events	$43.9^{+10.0}_{-9.8}$	$16.6^{+8.1}_{-7.2}$	$17.2^{+4.7}_{-5.1}$	$4.9^{+0.0}_{-0.8}$
Signal fraction (%)	$41.1^{+9.3}_{-9.2}$	$21.0^{+10.3}_{-9.1}$	$74.7^{+20.5}_{-22.3}$	$98.0^{+0.0}_{-16.5}$

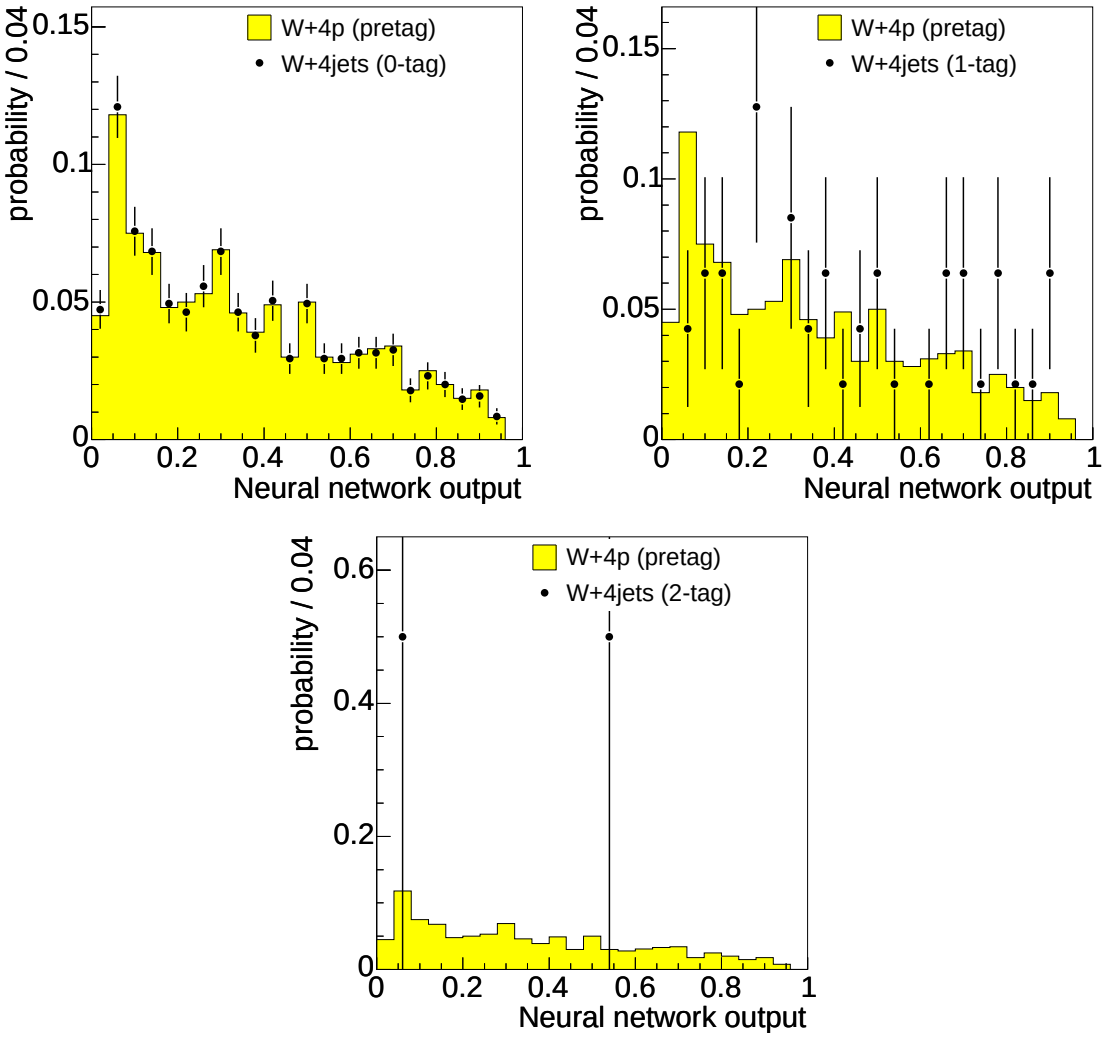


Figure 5.8: Comparison of the ANN output shapes of inclusive $W + 4$p MC sample (solid histogram) with the $W + 4$p 0-tag, 1-tag, and 2-tag subsamples (points with errors).

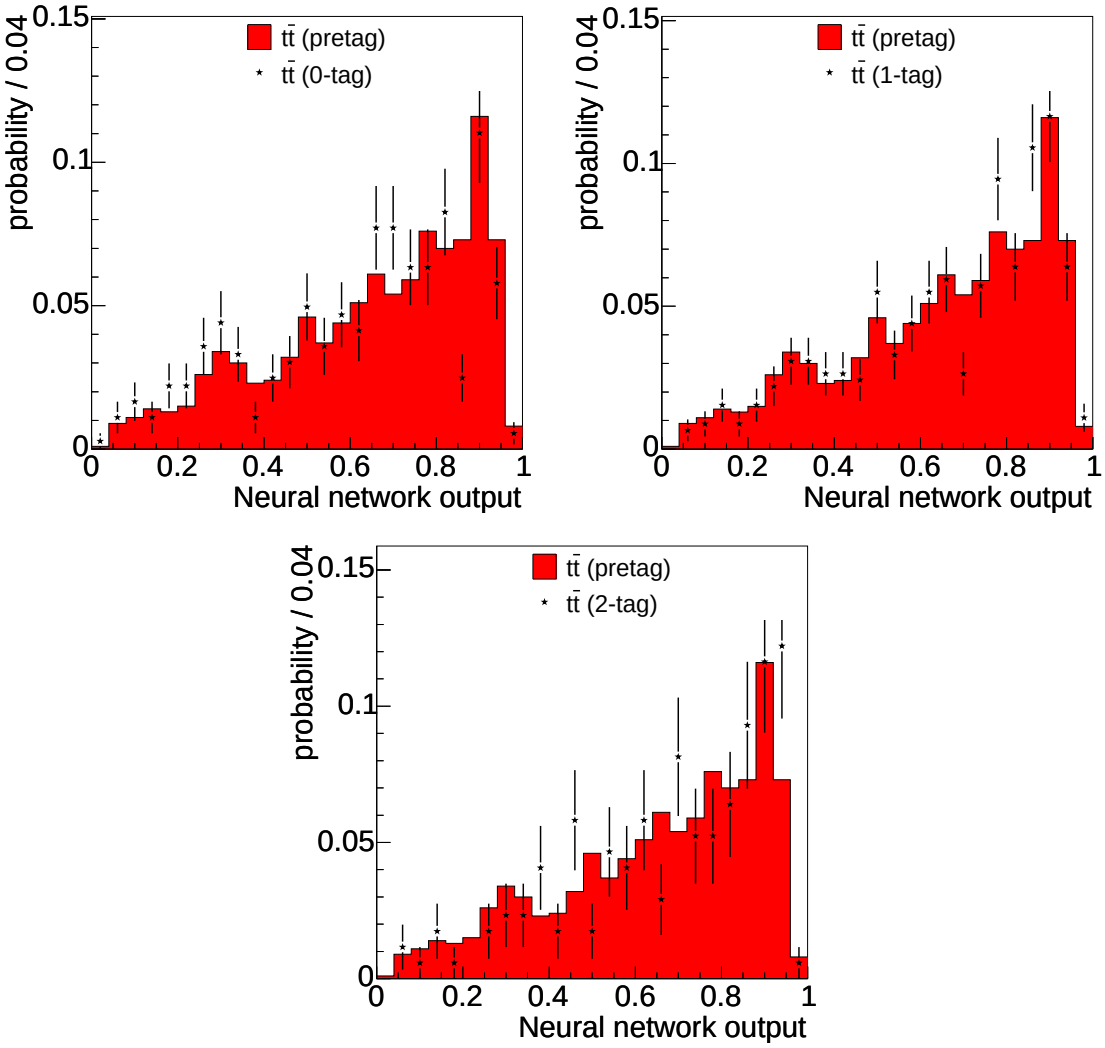


Figure 5.9: Comparison of the ANN output shapes of inclusive $t\bar{t}$ MC sample (solid histogram) with the $t\bar{t}$ 0-tag, 1-tag, and 2-tag subsamples (points with errors).

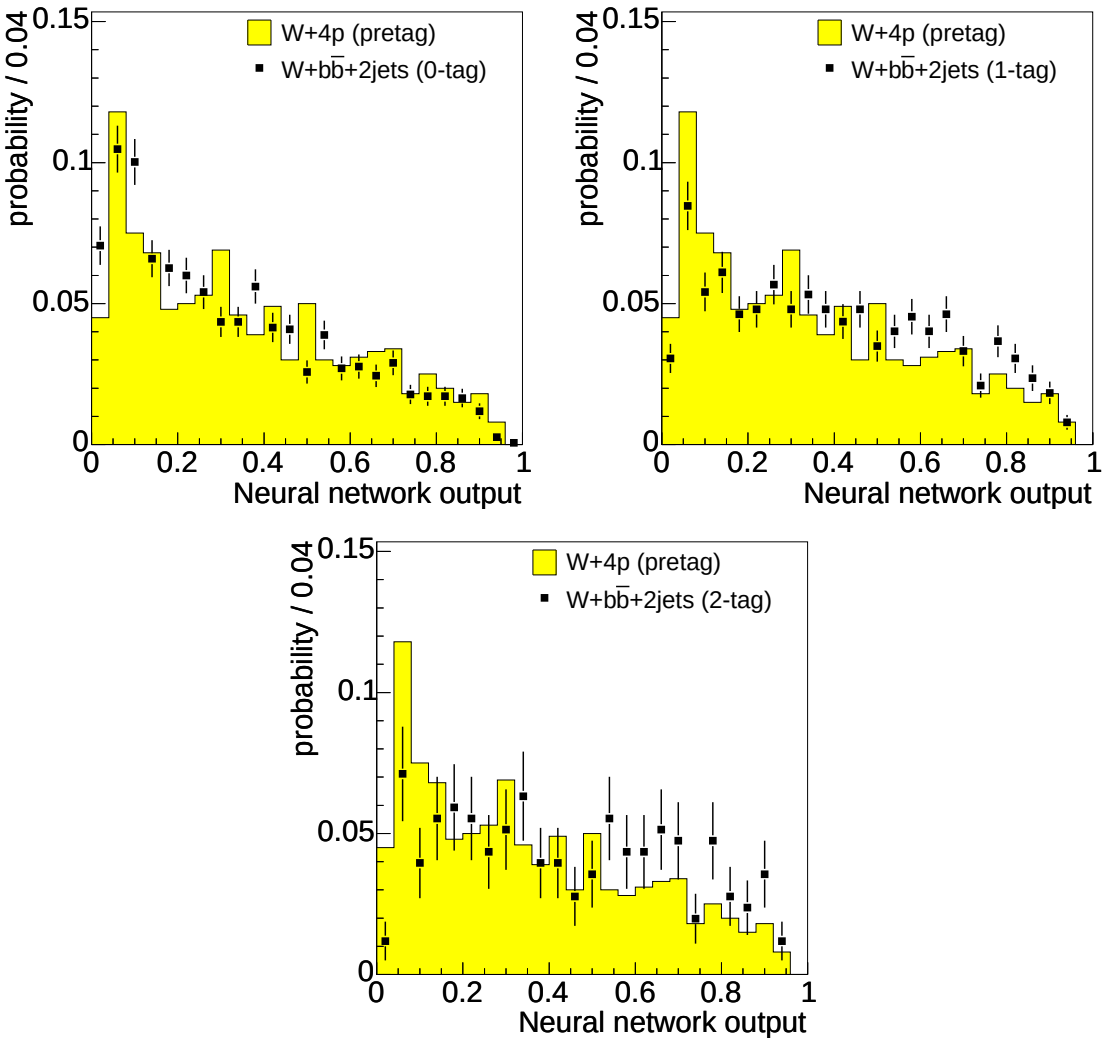


Figure 5.10: Comparison of the ANN output shapes of inclusive $W + 4$p MC sample (solid histogram) with the $W + b\bar{b} + 2$p MC 0-tag, 1-tag, and 2-tag subsamples (points with errors).

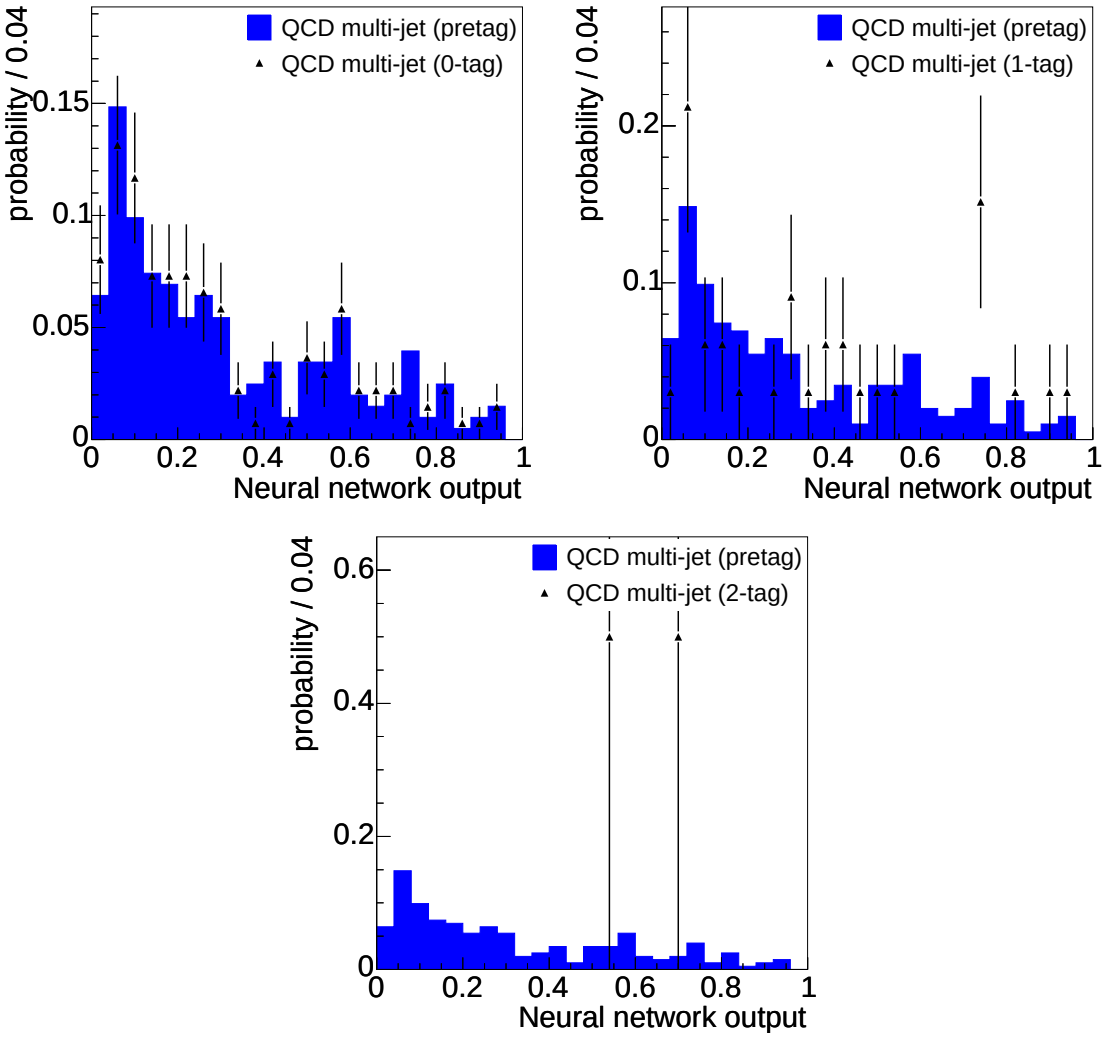


Figure 5.11: Comparison of the ANN output shapes of inclusive QCD multi-jet data sample (solid histogram) with the QCD multi-jet 0-tag, 1-tag, and 2-tag subsamples (points with errors).

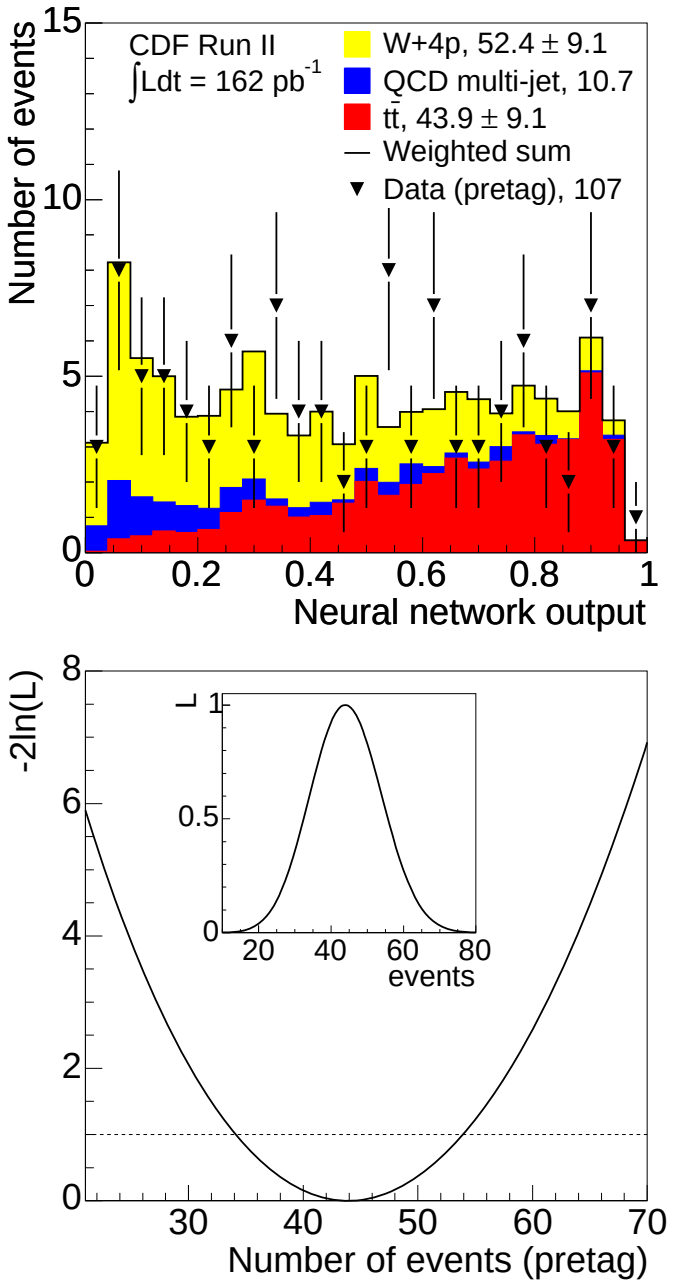


Figure 5.12: The top plot shows a three-component fit to the ANN output distribution of the pretagged lepton-plus-jets data subset (inverted triangles). The fraction of QCD multi-jet background (middle histogram) is kept constant in the fit, while the only free parameter (the fraction of $t\bar{t}$ signal events) defines the weights for both $W+4$p background (upper histogram) and $t\bar{t}$ signal (lower histogram) MC templates. The bottom plot shows the shape of the binned maximum likelihood as a function of the free parameter.

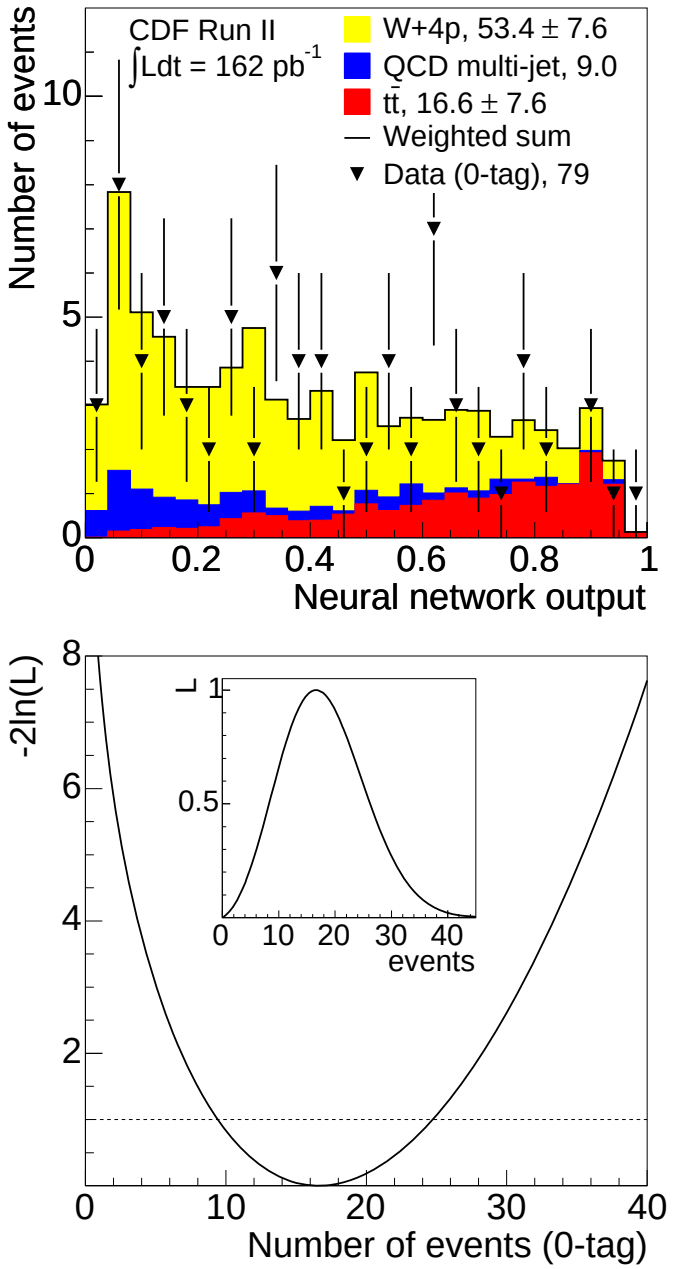


Figure 5.13: The top plot shows a three-component fit to the ANN output distribution of the 0-tag lepton-plus-jets data subset (inverted triangles). The fraction of QCD multi-jet background (middle histogram) is kept constant in the fit, while the only free parameter (the fraction of $t\bar{t}$ signal events) defines the weights for both W + 4p background (upper histogram) and $t\bar{t}$ signal (lower histogram) MC templates. The bottom plot shows the shape of the binned maximum likelihood as a function of the free parameter.

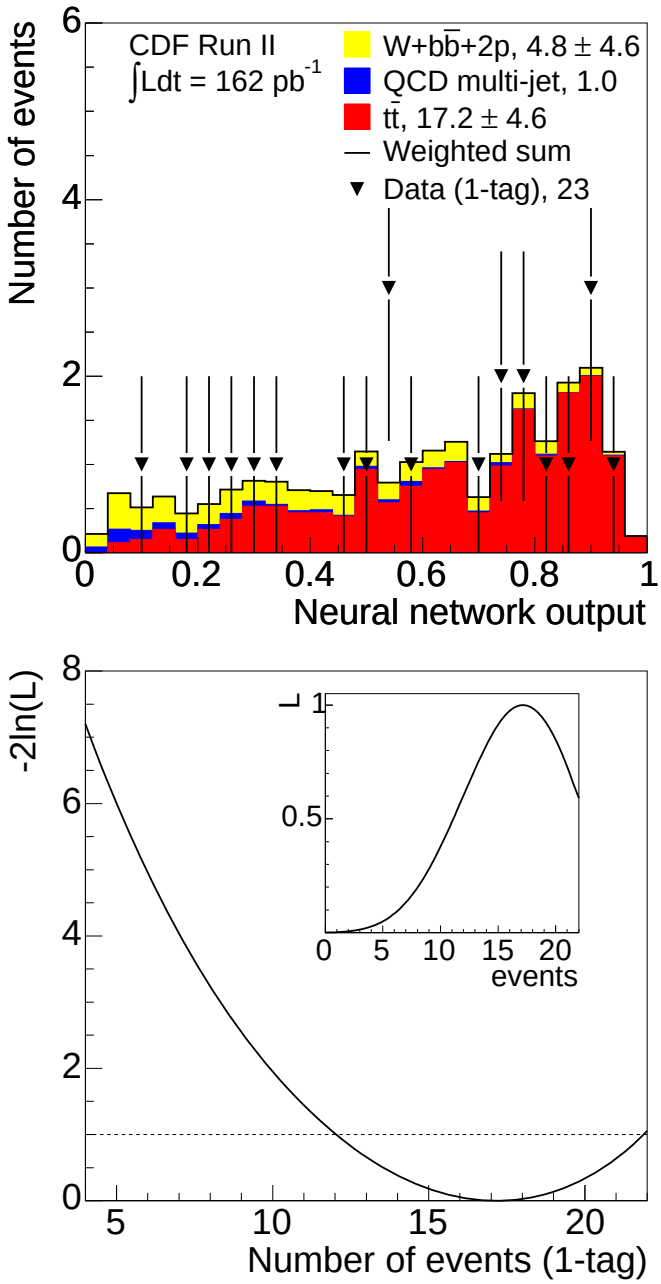


Figure 5.14: The top plot shows a three-component fit to the ANN output distribution of the 1-tag lepton-plus-jets data subset (inverted triangles). The fraction of QCD multi-jet background (middle histogram) is kept constant in the fit, while the only free parameter (the fraction of $t\bar{t}$ signal events) defines the weights for both W + 4p background (upper histogram) and $t\bar{t}$ signal (lower histogram) MC templates. The bottom plot shows the shape of the binned maximum likelihood as a function of the free parameter.

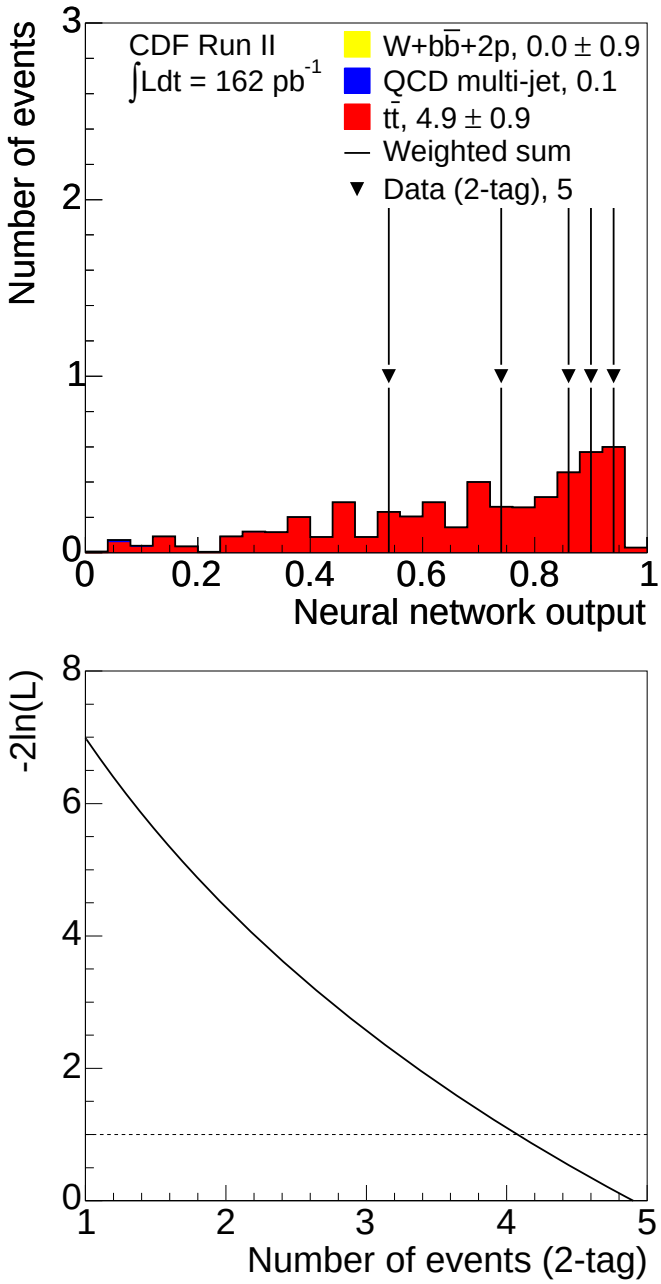


Figure 5.15: The top plot shows a three-component fit to the ANN output distribution of the 2-tag lepton-plus-jets data subset (inverted triangles). The fraction of QCD multi-jet background (middle histogram) is kept constant in the fit, while the only free parameter (the fraction of $t\bar{t}$ signal events) defines the weights for both W + 4p background (upper histogram) and $t\bar{t}$ signal (lower histogram) MC templates. The bottom plot shows the shape of the binned maximum likelihood as a function of free parameter.

5.1.5 Systematic Uncertainties on the Measurement of Top Signal Fraction

Our measurement of the signal fraction is sensitive to systematic effects that affect the shape and normalization of ANN output distribution. These systematics also affect the measurement of the branching ratio. To estimate these systematic effects, we use the trained ANN to create alternative systematic-shifted ANN output shapes for each systematic source. In order to create the shifted ANN output shapes, we use 2,000 simulated experiments (pseudo-experiments) based on a model with a particular systematic effect applied. We generate another 2,000 pseudo-experiments based on our default signal and background models. The total number of events in a pseudo-experiment equals on average to the observed data yields listed in Table 3.5. Each simulated experiment contains $N_{t\bar{t}}$ events drawn from a Poisson distribution with mean given by Table 5.5, N_{QCD} background events drawn from a Poisson distribution with mean given by Table 5.1, and $N_{W+\mathrm{jets}}$ background events drawn from a Poisson distribution with mean equal to the remainder. In every simulated experiment, we perform the binned maximum likelihood fit of ANN output to the signal and background ANN output distributions from our default model. The average difference in the fitted number of signal events—as calculated in 2,000 pseudo-experiments for shifted and nominal ANN output shapes—defines the systematic uncertainty for the particular source. We consider the following systematic sources:

Jet Energy Scale The systematic effect due to jet energy scale variations is the largest contributor to the total uncertainty. The jet energy corrections with respect to the measurement of the raw calorimeter was described in Chapter 3. There are several levels of correction intended to bring the energy of the jet closer to the true energy of original parton. There are two ways to calculate the overall impact on the ANN output shapes from the jet energy scale variations. One way is to vary the jet

energy corrections by varying one particular level of correction at a time and then add the effects of all levels in quadrature. The other way is to vary jet energy corrections taking into account all levels of corrections and calculate only one systematic number. We find similar results using either method.

W+jets background (Q^2) The shape of $W+$jets background is affected by different values of the scale of momentum transfer Q^2 in the hard scattering process. The value of Q^2 affects the relative weights of diagrams in the leading order matrix element. Our default ALPGEN MC is generated with $Q = M_W^2 + \sum_i p_{T,i}^2$, where M_W is the W boson mass and the sum is taken over all partons i with transverse momenta $p_{T,i}$. For the estimation of the Q^2 systematic we use two $W+$jets MC samples where Q^2 is $Q^2 = 4M_W$ and $Q^2 = \frac{1}{4}M_W$ respectively.

QCD multi-jet background There are two kinds of systematic effects related to this background component: the QCD-modeling and the absolute normalization of the sample. As an alternative to our default model with non-isolated leptons we use the sample with identified conversion electrons. A large fraction of QCD background in the electron channel is expected to come from photon conversions. For the uncertainty on the multi-jet background normalization, we vary the contributions by $\pm 50\%$ around the central value [45].

Monte Carlo Generators This uncertainty reflects the choice of $t\bar{t}$ event generator. We compare our default generator, HERWIG, with PYTHIA.

Top quark mass, m_t The systematic effect due to choice of the top mass is accounted by using two MC samples with $m_t = 170$ GeV and $m_t = 180$ GeV and comparing with our default model with $m_t = 175$ GeV.

Initial and Final State Gluon Radiation (ISR/FSR) We produce four PYTHIA Monte Carlo samples with different values for the transverse momentum scale (defined by $\Lambda_{\rm QCD}$ and K-factor), tuned to produce either less or more decay products as the result of ISR and FSR. The number of detected jets directly depends on the amount of initial or final state gluon radiation and, hence, the event kinematics may be affected.

Parton Distribution Functions and α_s For the calculation of systematic uncertainty caused by uncertainties in the proton and antiproton PDFs we follow the method described in [69]. The variations in PDF affect the relative contributions of the $q\bar{q} \to t\bar{t}$ and $gg \to t\bar{t}$ processes as well as the momentum of the $t\bar{t}$ system. The basic idea of the method is to re-weight the events in the default MC sample using the shifted PDF, momentum transfer Q, and momentum fractions x_i of incoming particles:

$$w_{\rm event} = \frac{F^{p}_{i,\rm shifted}(x_i,Q)\cdot F^{\bar{p}}_{j,\rm shifted}(x_j,Q)}{F^{p}_{i,\rm default}(x_i,Q)\cdot F^{\bar{p}}_{j,\rm default}(x_j,Q)}. \tag{5.11}$$

Our default signal sample is generated using CTEQ5L PDF library. We compare the results from the default model to the CTEQ6M version of PDFs. The CTEQ6M PDF is defined by 20 different eigenvalues which are varied up and down within a 90% confidence interval around their default values. The individual systematic contributions from 40 CTEQ6M PDFs are added in quadrature. Alternatively 2 MRST PDF sets [70] are utilized returning consistent results. The MRST PDFs are also used to estimate the α_s-systematic uncertainty for two values $\alpha_s = 0.1125$ and $\alpha_s = 0.1175$ (the default value is $\alpha_s = 0.115$).

The systematic uncertainties are summarized in Table 5.6. For those systematics where only one variation of the shifted source is available, we quote one number and use its absolute value to obtain both negative and positive variations on the signal fraction. For the systematic effects where shifted variations of the source are

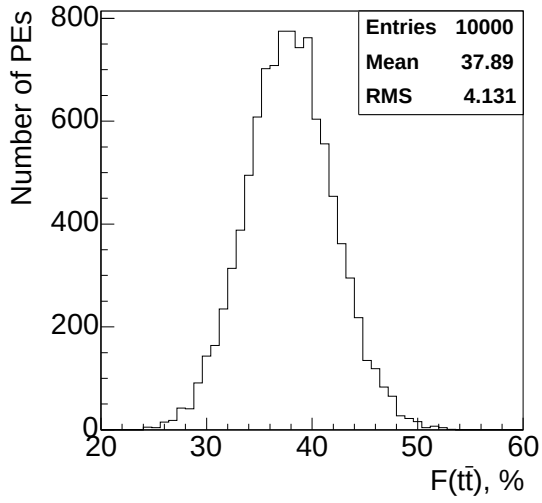
Table 5.6: Systematic uncertainties on the measured signal fraction in pretag, 0-tag, 1-tag, and 2-tag data subsets. The overall uncertainty is obtained by adding contributions from each source in quadrature.

Source	**pretag (%)**	**0-tag (%)**	**1-tag (%)**	**2-tags (%)**
central value	41.1	21.0	74.7	98.0
Jet E_T	$+5.82$	$+4.38$	$+2.62$	$+2.54$
	-1.68	-0.45	-9.63	-4.72
Q^2	$+5.05$	$+6.13$	$+0.322$	$+1.59$
	$+2.23$	$+2.09$	$+3.90$	-0.14
QCD (shape)	-3.65	-3.13	-2.44	-0.94
QCD (norm.)	-0.09	$+0.12$	-1.06	-1.15
$\pm 50\%$	$+0.06$	$+0.04$	-0.29	$+2.07$
MC generator	$+0.79$	$+0.44$	-4.78	$+0.17$
m_t	$+2.63$	$+1.79$	-0.47	$+2.22$
	-0.99	-0.37	-6.54	-4.15
FSR	$+0.44$	$+0.26$	$+1.93$	-4.68
	-0.21	-0.07	$+2.21$	-2.87
ISR	$+0.78$	$+0.41$	$+3.45$	-3.05
	$+0.37$	$+0.28$	$+1.48$	-5.06
PDF	$+0.14$	$+0.17$	$+1.81$	$+2.80$
	-0.57	-0.41	-1.94	-4.86
α_s	$+0.04$	-0.01	-1.15	-0.29
Total	± 6.98	± 6.02	± 10.62	± 8.95

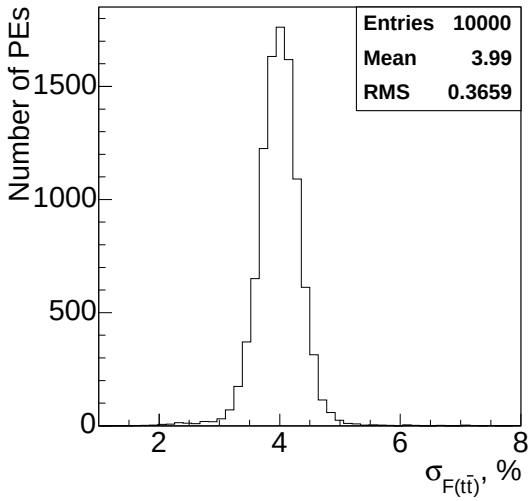
available below and above our default value, we quote two numbers accordingly. For the calculation of the total symmetric uncertainty we use a symmetric uncertainty $\pm\sigma$ for every source, where σ is either the average of the two systematic uncertainties, if they are of the same sign, or the largest absolute value, if they are of opposite sign. These symmetric uncertainties are then summed in quadrature, to give the overall symmetric systematic uncertainty on the signal fraction.

In order to test the overall performance of our fitter, we also generate pseudo-

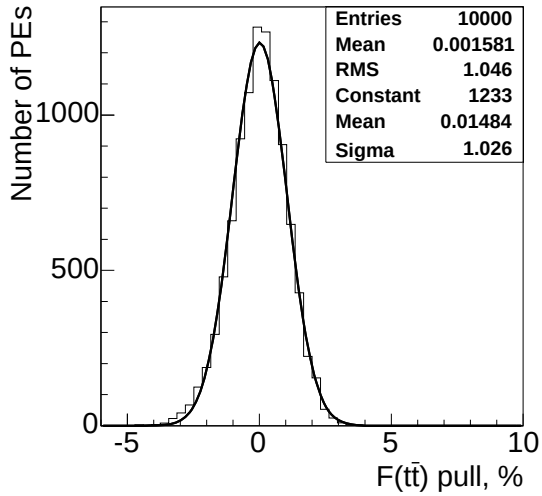
experiments following the same pseudo-experiment procedure as the one used for the systematic uncertainties estimation, using the default ANN output shapes as templates. We use the pretagged signal and background templates which are statistically fluctuated using a Poisson distribution with mean corresponding to the expected number of events for $R = 1$. The calculation of expected signal events in the three tagged subsamples is described in Section 5.3.1. We predict the fraction of $\approx 38\%$ signal events in the pretagged sample with total 107 events. The results of the measurement of signal fraction for the 10,000 pseudo-experiments are shown in Figure 5.16, along with the parabolic error and pull distribution. We conclude that the performance of our fitter is satisfying.



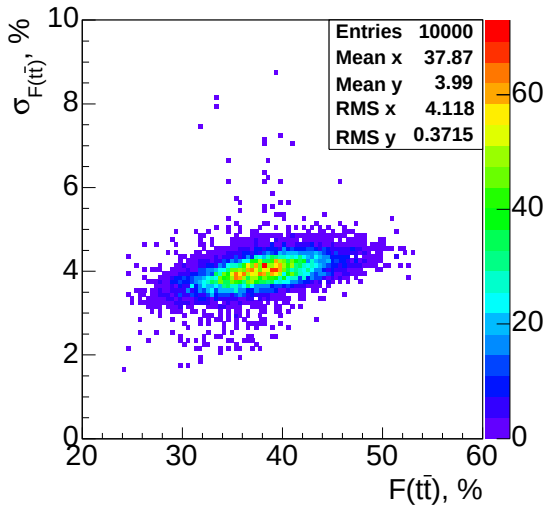
(a) Measured signal fraction $F_{t\bar{t}}$.



(b) Parabolic error $\sigma_{F_{t\bar{t}}}$ as returned by the fitter.



(c) Pull distribution $P = \frac{F_{t\bar{t}} - F_{t\bar{t}}^{\text{exp}}}{\sigma_{F_{t\bar{t}}}}$.



(d) $F_{t\bar{t}}$ versus $\sigma_{F_{t\bar{t}}}$.

Figure 5.16: Performance of the ANN output fitter as verified by 10,000 pseudo-experiments.

5.1.6 Alternative Background Estimation

It was shown that the signal and background fractions in the pretagged and tagged data subsamples can be measured using the ANN framework. There is an alternative method for the background determination. This technique is widely used by the CDF collaboration and it can be applied in both lepton-plus-jets and dilepton channels. Despite of small variations in the actual implementation of this method in the two channels, the general idea is to calculate the contribution of each possible background component using a combination of data driven and Monte Carlo based techniques. In the remainder of the thesis we will refer to this method as *a priori* or direct background estimations[4]. Below we give a short overview of the background estimation in lepton-plus-jets.

In the lepton-plus-jets samples, the dominant background to $t\bar{t}$ is from W production in association with QCD radiation. Once a tagged b-jet is required, *i.e.* in the 1-tag and 2-tag subsamples, these backgrounds are greatly reduced, and the residual backgrounds include: W production in association with heavy-flavour jets ($Wb\bar{b}$, $Wc\bar{c}$, Wc); W production in association with light-flavour jets that are mistakenly tagged (*mistags*); non-W (generic QCD) events containing fake leptons or incorrectly-measured missing energy; dibosons (WW, WZ) and single-top quark production. In the tagged samples, we can make a direct estimate of the background rates. $W+$ heavy-flavour production backgrounds are estimated using the predicted fraction of inclusive $W+$jets events containing heavy-flavour jets and the efficiency to tag those jets, normalized to the observed number of $W+$jets events. Mistags are estimated using determined parameterizations of the rates of incorrect tags measured in complementary data samples. Non-W rates are estimated in a similar way as it was done for the ANN QCD background, *i.e.* we extrapolate the number of events with isolated leptons and small missing energy into the signal region of large missing energy. The

[4]By construction, the *a priori* method applies only to the 1-tag and 2-tag subsamples.

Table 5.7: Background estimates in the lepton-plus-jets sample, based on the *a priori* estimate.

	Pretag	**0-tag**	**1-tag**	**2-tag**
Estimated background	undetermined	undetermined	4.2 ± 0.7	0.2 ± 0.1

remaining backgrounds are estimated using Monte-Carlo-simulated processes, with rates normalized to their Standard-Model cross sections. In total, these background rates can be measured to a precision of about 15%, but these techniques cannot be used to estimate the background in the 0-tag sample, where the absolute rate of W plus jets production is difficult to predict.

The background estimates in the lepton-plus-jets samples from the *a priori* estimation method are given in Table 5.7. These results are consistent with the ANN results in the tagged subsamples where both techniques can be used. For 0-tag sample the ANN gives a background measurement whereas the *a priori* method cannot. At the same time, the *a priori* estimate provides more precise estimates in 1-tag and 2-tag subsamples. More details on the *a priori* background estimation in lepton-plus-jets channel can be found in [23, 71].

5.2 Top Quark Signal Determination in the Dilepton Channel

The dominant backgrounds in the dilepton sample are Drell-Yan production (for ee and $\mu\mu$ events only), fake leptons, and diboson (WW , WZ, ZZ) production. The Drell-Yan rate is estimated using simulations that are normalized to the observed rate of Z events. Fake-lepton background rates are based on measurements in complementary, lepton-poor data samples, and diboson rates are estimated from simulations, normalized to the Standard-Model cross sections. However, these es-

Table 5.8: Background estimates in the dilepton sample, based on the *a priori* estimate.

	Pretag	**0-tag**	**1-tag**	**2-tag**
Estimated background	2.2 ± 0.6	2.0 ± 0.6	0.2 ± 0.1	< 0.01

timates do not address how the backgrounds are distributed across the subsamples with different numbers of tagged jets. Most of the jets in the background events arise from generic QCD radiation. To estimate the dibosons amount of background in the tagged subsamples, we parameterize the probability to tag a generic QCD jet using complementary jet-data samples, and apply this parameterization to the jets in dilepton candidate events. We correct this estimate for the fact that most of the events in this sample are in fact $t\bar{t}$ events and not background. The resulting background estimates are given in Table 5.8. More details on the *a priori* background estimation in dilepton channel can be found in [72].

5.3 Measurement of $R = \frac{\mathcal{B}(t\to Wb)}{\mathcal{B}(t\to Wq)}$

The idea behind the measurement of the ratio of branching fractions R is to describe the number of top-signal events we expect in our data as a function of R. By comparing our predictions to the number of observed top-signal events, we extract R. This procedure is performed simultaneously in each of the tagged subsamples in order to further constrain the measurement. We start by discussing the determination of the theoretically expected number of top-signal events with different b-tag multiplicities. The comparison between the number of observed and predicted events is accomplished by constructing a proper likelihood function.

5.3.1 Prediction of Number of Signal Events

The ratio R is sensitive to the number of $t\bar{t}$ signal events in the 0-tag, 1-tag and 2-tag bins. This sensitivity is due to the fact that the number of tagged jets depends on both our efficiency to tag a jet and the fractions of top quarks that decay to a b quark. Tagged jets in $t\bar{t}$ events arise from a variety of sources. By far the most likely source is real b quarks that come from the decay $t \to Wb$, and the efficiency to tag such a b-type jet is ϵ_b. If $R \neq 1$, then the decay $t \to Wq$ can also occur. If ϵ_q is the efficiency to tag a light quark or gluon, then the efficiency to tag a jet from a top decay is $E_b(R) = R\epsilon_b + (1 - R)\epsilon_q$. Along with the b-jet and q-jet tagging efficiencies, we take into account the efficiency ϵ_c to tag c-jets that come from c quarks. All the tagging efficiency values apply to jets that are within our fiducial acceptance, *i.e.* the jets must be *taggable*[5]. We use a sample of simulated Monte Carlo $t\bar{t}$ events, where we can precisely determine the flavour content of jets, to estimate the efficiencies to tag the b, c, and q-jets. Note, that the SecVtx tagging algorithm performs differently in data and Monte Carlo. Therefore, we account for this difference by introducing a scale factor SF [45], which weights the MC-determined tagging efficiencies. Also, from the same simulation sample we determine the fractions of events that have different combinations of taggable jets from different sources, *i.e.* the fraction of events with i taggable q-jets, j taggable c-jets, and k taggable b-jets defines the acceptance F_{ijk}.

By using appropriate event combinatorics, applying the proper efficiency values according to the types of jets that are present in a given event, and weighting by the acceptances F_{ijk}, we predict the efficiencies ϵ_0, ϵ_1, and ϵ_2 to observe zero, one and two tags correspondingly in a $t\bar{t}$ event, as a function of R. Because of imperfect tagging efficiencies, we may observe either less or more b-tagged jets than the expected two b-jets in a top-signal event. Since we do not observe events with more that 3 tagged

[5]A jet is taggable if it has at least two tracks passed quality cuts based on the number of hits found in CDF tracking systems. Two tracks are needed to reconstruct a secondary vertex.

jets, the efficiencies ϵ_i, where $i \geq 3$, are not considered here. The explicit expressions for the event tagging efficiencies ϵ_i are given below:

$$\epsilon_0(R) = \sum_{i,j,k} F_{ijk}(1-\epsilon_q)^i(1-\epsilon_c)^j(1-E_b(R))^k, \tag{5.12a}$$

$$\begin{aligned}\epsilon_1(R) = &\sum_{i\geq 1,j,k} F_{ijk}\binom{i}{1}\epsilon_q(1-\epsilon_q)^{i-1}(1-\epsilon_c)^j(1-E_b(R))^k \\ &+ \sum_{i,j\geq 1,k} F_{ijk}\binom{j}{1}(1-\epsilon_q)^i\epsilon_c(1-\epsilon_c)^{j-1}(1-E_b(R))^k \\ &+ \sum_{i,j,k\geq 1} F_{ijk}\binom{k}{1}(1-\epsilon_q)^i(1-\epsilon_c)^j E_b(R)(1-E_b(R))^{k-1},\end{aligned} \tag{5.12b}$$

$$\begin{aligned}\epsilon_2(R) = &\sum_{i\geq 2,j,k} F_{ijk}\binom{i}{2}\epsilon_q^2(1-\epsilon_q)^{i-2}(1-\epsilon_c)^j(1-E_b(R))^k \\ &+ \sum_{i,j\geq 2,k} F_{ijk}\binom{j}{2}(1-\epsilon_q)^i\epsilon_c^2(1-\epsilon_c)^{j-2}(1-E_b(R))^k \\ &+ \sum_{i,j,k\geq 2} F_{ijk}\binom{k}{2}(1-\epsilon_q)^i(1-\epsilon_c)^j E_b^2(R)(1-E_b(R))^{k-2} \\ &+ \sum_{i\geq 1,j\geq 1,k} F_{ijk}\binom{i}{1}\binom{j}{1}\epsilon_q(1-\epsilon_q)^{i-1}\epsilon_c(1-\epsilon_c)^{j-1}(1-E_b(R))^k \\ &+ \sum_{i\geq 1,j,k\geq 1} F_{ijk}\binom{i}{1}\binom{k}{1}\epsilon_q(1-\epsilon_q)^{i-1}(1-\epsilon_c)^j E_b(R)(1-E_b(R))^{k-1} \\ &+ \sum_{i,j\geq 1,k\geq 1} F_{ijk}\binom{j}{1}\binom{k}{1}(1-\epsilon_q)^i\epsilon_c(1-\epsilon_c)^{j-1}E_b(R)(1-E_b(R))^{k-1}.\end{aligned} \tag{5.12c}$$

The sums in Equations 5.12 are calculated over all taggable jets in event and the summation indexes i, j, and k correspond to the number of taggable q-jets, c-jets, and b-jets. All additive terms are multiplied by the acceptances F_{ijk} presented in Table 5.9. The values of jet tagging efficiencies and the scale factor are given in Table 5.10. The terms in Equations 5.12 are constructed as a product of the binomial probability distributions $\binom{n}{k}(\epsilon_\alpha)^k(1-\epsilon_\alpha)^{n-k}$, where $\binom{n}{k}$ is a binomial coefficient and ϵ_α and $(1-\epsilon_\alpha)$ are the probabilities to tag and not to tag a jet coming from a

Table 5.9: Most significant acceptances F_{ijk} of fraction of events with i taggable q-jets, j taggable c-jets, and k taggable b-jets as calculated from $t\bar{t}$ Monte Carlo sample.

F_{ijk}						
i	j	k				
		0	1	2	3	4
0	0	0.0357	0.0055	0.0015	0.0003	0.0001
0	1	0.0237	0.0182	0.0082	0.0018	0.0003
0	2	0.0301	0.0308	0.0114	0.0012	0.0002
1	0	0.0148	0.0091	0.0035	0.0007	0.0000
1	1	0.0658	0.0558	0.0172	0.0027	0.0004
1	2	0.1267	0.0875	0.0151	0.0013	0.0003
2	0	0.0136	0.0104	0.0022	0.0002	0.0000
2	1	0.0984	0.0484	0.0070	0.0007	0.0002
2	2	0.1970	0.0457	0.0055	0.0007	0.0001

Table 5.10: Jet tagging efficiencies and the scale factor.

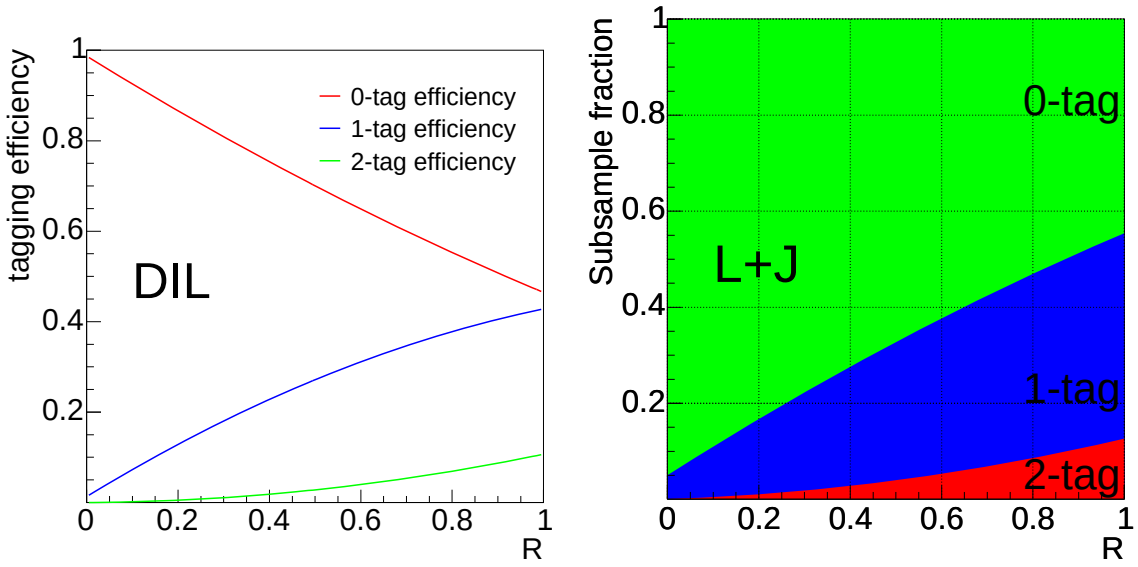
Parameter	Central Value	Uncertainty
ϵ_b	0.532	0.039
ϵ_c	0.106	0.015
ϵ_q	0.005	0.001
SF	0.820	0.060

partonic source α, correspondingly. The event tagging efficiencies for lepton-plus-jets and dilepton samples are nearly identical for all values of R. For $R = 1$ their values are shown in Table 5.11. In Figure 5.17 the dilepton efficiencies are drawn as functions of R but in case of lepton-plus-jets the "stacked" curves are drawn instead, so that the areas between the curves show the fraction of the $t\bar{t}$ MC sample that has zero, one, or two tags. This helps to see that at $R = 0$, the sample is almost completely composed of 0-tag events; even though no b quarks are produced in that case, there

Table 5.11: Event tagging efficiencies for $R = 1$.

	ϵ_0	ϵ_1	ϵ_2
Lepton-plus-jets	0.45 ± 0.03	0.43 ± 0.02	0.12 ± 0.02
Dilepton	0.47 ± 0.03	0.43 ± 0.02	0.10 ± 0.02

is still a small chance to tag other jets in the event. At $R = 1$, the fraction of a the sample with two tags is at a maximum. The fraction of the sample with only one tag stays fairly constant in the region of interest at $R > 0.7$. The measured value of R will be most sensitive to the ratio of 0-tag to 2-tag event counts, as that ratio changes more rapidly with R compared to other ratios of tagged events.



(a) Event tagging efficiencies in dilepton sample.

(b) "Stacked" tagging efficiencies in lepton-plus-jets sample.

Figure 5.17: The event tagging efficiencies expressed in a unity fractions as functions of R (*c.f.* Equations 5.12).

Finally, the number of expected signal events with i tags can be calculated as

$$N^{t\bar{t}}_{\text{exp},i}(R) = N^{t\bar{t}}_{\text{inc}} \cdot \epsilon_i(R), \tag{5.13}$$

where $N^{t\bar{t}}_{\text{inc}}$ is the inclusive number of $t\bar{t}$ events before any tagging requirement, as determined from $N^{t\bar{t}}_{\text{inc}} = \sum_{i=0}^{2}(N_{\text{obs},i} - N^{\text{bkg}}_{\text{obs},i})$, where the $N_{\text{obs},i}$ and $N^{\text{bkg}}_{\text{obs},i}$ are the numbers of total observed and measured background events with i b-tags.

5.3.2 ANN Measurement of R in Lepton-plus-jets Channel

In this Section we present the measurement of R using only the ANN measured number of $t\bar{t}$events in the three tagged subsamples. A relatively simple likelihood function can be constructed in this case. Since fitting the ANN data output to the signal and background templates returns the 100% anti-correlated number of background and signal events, we can use only the signal measurement to extract R. The following likelihood is constructed:

$$\begin{aligned} L = & P(\widetilde{N}_0^{t\bar{t}}; N_{\rm exp,0}^{t\bar{t}}(R)) \times P(\widetilde{N}_1^{t\bar{t}}; N_{\rm exp,1}^{t\bar{t}}(R)) \times P(\widetilde{N}_2^{t\bar{t}}; N_{\rm exp,2}^{t\bar{t}}(R)) \\ & \times G(\widetilde{\epsilon}_b; \epsilon_b, \sigma_{\epsilon_b}) \times G(\widetilde{\epsilon}_c; \epsilon_c, \sigma_{\epsilon_c}) \times G(\widetilde{\epsilon}_q; \epsilon_q, \sigma_{\epsilon_q}) \\ & \times G(\widetilde{N}_0^{t\bar{t}}; N_{\rm obs,0}^{t\bar{t}}, \sigma_{N_{\rm obs,0}^{t\bar{t}}}) \times G(\widetilde{N}_1^{t\bar{t}}; N_{\rm obs,1}^{t\bar{t}}, \sigma_{N_{\rm obs,1}^{t\bar{t}}}) \times G(\widetilde{N}_2^{t\bar{t}}; N_{\rm obs,2}^{t\bar{t}}, \sigma_{N_{\rm obs,2}^{t\bar{t}}}) \end{aligned} \quad (5.14)$$

where $P(\widetilde{N}_i^{t\bar{t}}; N_{{\rm exp},i}^{t\bar{t}})$ are Poisson probabilities to observe $\widetilde{N}_i^{t\bar{t}}$ events when the expected mean values are $N_{{\rm exp},i}^{t\bar{t}}$. $G(\widetilde{\epsilon}_\alpha; \epsilon_\alpha, \sigma_{\epsilon_\alpha})$ are gaussian smearings of the efficiencies $\widetilde{\epsilon}_\alpha$ based on the mean values ϵ_α and uncertainties σ_{ϵ_α}, and $G(\widetilde{N}_i^{t\bar{t}}; N_{{\rm obs},i}^{t\bar{t}}, \sigma_{N_{{\rm obs},i}^{t\bar{t}}})$ are gaussian smearings of the number of observed signal events $\widetilde{N}_i^{t\bar{t}}$ with a mean equal to the measured values $N_{{\rm obs},i}^{t\bar{t}}$ and a standard variation equal to the measured errors $\sigma_{N_{{\rm obs},i}^{t\bar{t}}}$, as presented in Table 5.5. For the likelihood maximization, we calculate the number of inclusive $t\bar{t}$ events as

$$N_{\rm inc}^{t\bar{t}} = \frac{\widetilde{N}_0^{t\bar{t}} + \widetilde{N}_1^{t\bar{t}} + \widetilde{N}_2^{t\bar{t}}}{\epsilon_0 + \epsilon_1 + \epsilon_2}, \quad (5.15)$$

where $\widetilde{N}_i^{t\bar{t}}$ is the floating number of signal events. Note that the $\widetilde{\epsilon}_i$ are also varied during the minimization procedure, since they depend on R. The minimization of the negative logarithm of the L is shown in Figure 5.18 and it yields the central value of $R = 1.06^{+0.27}_{-0.24}$, where the uncertainty is statistical only. The systematic uncertainty on R is calculated by generating pseudo-experiments in a way described in Section 5.1.5. Every time we extract the number of signal events for the systematically shifted pseudo-data, we find a new shifted value of R by minimizing the likelihood. The Gaussian terms are removed in order to measure the systematic effect only. We obtain

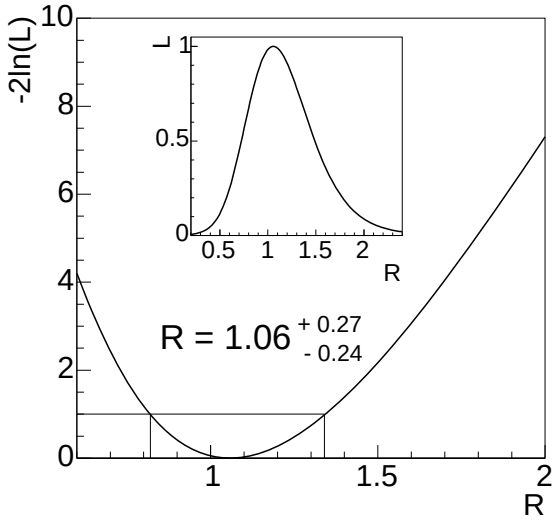


Figure 5.18: Minimization of the negative logarithmic likelihood constructed using only ANN information in lepton-plus-jets channel. The inset shows the likelihood. The intersections of the horizontal line $\ln L = 1$ with the likelihood define the statistical 1σ errors on R.

asymmetric shifts of the average systematically shifted R from the central value to be $^{+0.07}_{-0.16}$. If we conservatively assign the highest shift on one systematic direction as a symmetric uncertainty, our systematic uncertainty on R is ± 0.16.

5.3.3 Combined Measurement of R in Lepton-plus-jets and Dilepton Channels

To improve the statistical and systematic uncertainties of our measurement, we measure R using both lepton-plus-jets and dilepton samples. For this combined R measurement we follow a slightly different approach in the creation of the likelihood function. As the *a priori* estimates have smaller uncertainties in the 1-tag and 2-tag subsamples, the ANN-determined background level is used only for the 0-tag subsample. The 1-tag and 2-tag statistical uncertainties in lepton-plus-jets background

Table 5.12: Summary of observed number of events with i tags in the lepton-plus-jets and dilepton samples, with estimates of background levels and expected event yields. The *a priori* backgrounds estimates are given for $R = 1$. Equations 5.12 and 5.19 are used for the calculation of the expected total number of events. The statistical and systematic uncertainties have been combined.

Lepton-plus-jets (LJ)	**0-tag**	**1-tag**	**2-tag**
ANN background, N_i^{bkg}	62.4 ± 9.0	5.8 ± 5.2	$0.1^{+1.0}_{-0.1}$
A priori Background, N_i^{bkg}	undetermined	4.2 ± 0.7	0.2 ± 0.1
Total expected, $N_{\text{exp},i}$	80.4 ± 5.2	21.5 ± 4.1	5.0 ± 1.4
Observed, $N_{\text{obs},i}$	79	23	5
Dileptons (DIL)	**0-tag**	**1-tag**	**2-tag**
Background, N_i^{bkg}	2.0 ± 0.6	0.2 ± 0.1	< 0.01
Total expected, $N_{\text{exp},i}$	6.1 ± 0.4	4.0 ± 0.2	0.9 ± 0.2
Observed, $N_{\text{obs},i}$	5	4	2

measured with the ANN are larger than but consistent with the *a priori* estimates shown in Table 5.12. The expected number of total events for $R = 1$ are also shown in the same table. Since the fraction of signal events is not determined directly in tagged subsamples using *a priori* method, the total number of observed events is fitted to the corresponding predictions in all three subsamples.

The likelihood function consists of two independent parts describing the lepton-plus-jets (LJ) and dilepton (DIL) samples and has the following form:

$$L = L_{\text{LJ}} \times L_{\text{DIL}} \times G(\widetilde{SF}; SF, \sigma_{SF}), \tag{5.16}$$

where each individual likelihood is a product of Poisson and Gaussian probability functions:

$$L_{\text{LJ}} = \prod_{i=0}^{2} P(N_{\text{obs},i}; N_{\text{exp},i}(R)) \times G(\widetilde{N}_0^{\text{bkg}}; N_{\text{obs},0}^{\text{bkg}}, \sigma_{N_{\text{obs},0}^{\text{bkg}}}) \times G(\widetilde{C}_1; C_1, \sigma_{C_1}), \tag{5.17}$$

and

$$L_{\text{DIL}} = \prod_{i=0}^{2} P(N_{\text{obs},i}; N_{\text{exp},i}(R)) \times G(\widetilde{C}_2; C_2, \sigma_{C_2}) \times G(\widetilde{C}_3; C_3, \sigma_{C_3}). \tag{5.18}$$

The Poisson terms $P(N_{\mathrm{obs},i}; N_{\mathrm{exp},i}(R))$ are responsible for constraining the number of observed events $N_{\mathrm{obs},i}$ with i tags to an expected number of events $N_{\mathrm{exp},i}$ within statistical uncertainty. Here,

$$N_{\mathrm{exp},i} = N_{\mathrm{inc}}^{t\bar{t}} \times \epsilon_i + N_{\mathrm{obs},i}^{\mathrm{bkg}}, \tag{5.19}$$

where $N_{\mathrm{inc}}^{t\bar{t}}$ is the number of events in the $t\bar{t}$ sample summed over all numbers of tags. This quantity is inferred from the populations of the subsamples, and depends on R through the event tagging efficiencies:

$$N_{\mathrm{inc}}^{t\bar{t}} = \frac{\sum_{i=0}^{2}(N_{\mathrm{obs},i} - N_{\mathrm{obs},i}^{\mathrm{bkg}})}{\sum_{i=0}^{2}\epsilon_i}. \tag{5.20}$$

Since $\sum_{i=0}^{2}\epsilon_i$ is nearly unity for all values of R, $N_{\mathrm{inc}}^{t\bar{t}}$ varies very little with R, due to the dependence of the *a priori* estimated background on R, and, thus, providing a strong constraint on the measurement. Also, since we do not use the $t\bar{t}$cross section for the determination of $N_{\mathrm{inc}}^{t\bar{t}}$, our measurement is independent of $t\bar{t}$cross section.

Systematic uncertainties in the measurement are folded into the likelihood by multiplying the Poisson functions by a set of Gaussian functions that constrain various parameters to their expected values with some uncertainty. We assume the uncertainties on the jet tagging efficiencies ϵ_α by including an extra gaussian constrain on the scale factor SF, and, therefore, the uncertainties are correlated across all six subsamples. This constrain is incorporated as a common term in Equation 5.16, since the SF is the same for both lepton-plus-jets and dilepton samples. The *a priori* uncertainty on the lepton-plus-jets 2-tag background is anti-correlated with the uncertainty on the corresponding 1-tag background, as the 1-tag value was obtained by subtracting the 2-tag estimates from the $\geq$ 1-tag estimates. The Gaussian term with anti-correlation coefficient C_1 is accounting for this uncertainty. The uncertainty on the lepton-plus-jets 0-tag background, obtained from the ANN technique, is uncorrelated with all the other uncertainties. In the dilepton sample, the overall background normalization

Table 5.13: Correlation coefficients as set by *a priori* background estimation.

	$C \pm \sigma_C$
C_1	0.00 ± 0.20
C_2	0.00 ± 0.20
C_2	0.00 ± 0.26

uncertainty is correlated across all the tagged subsamples (C_2), whereas the uncertainties on the 0-tag and 1-tag backgrounds are anti-correlated (C_3). The values of various correlation coefficients, coming from the *a priori* method of background estimation, are given in Table 5.13.

The minimization of the negative logarithm of the combined likelihood function is shown in Figure 5.19. We find the central value $R = 1.12^{+0.27}_{-0.23}$, where the uncertainty includes statistical and systematic contributions. The systematic uncertainty is determined by minimization of L with Gaussian components removed. The final result can be written as $R = 1.12^{+0.21}_{-0.19}(\text{stat})\ ^{+0.21}_{-0.19}(\text{syst})$. The dominant systematic uncertainties arise from the uncertainty on the background measurement in the 0tag lepton-plus-jets sample ($^{+0.14}_{-0.11}$) and from the overall normalization of the tagging efficiencies ($^{+0.09}_{-0.06}$). We also perform a separate minimizations of lepton-plus-jets and dilepton likelihoods and find consistent results: The lepton-plus-jets sample alone yields $R = 1.02^{+0.23}_{-0.20}$ (stat) $^{+0.21}_{-0.13}$ (syst), and the dilepton sample alone yields $R = 1.41^{+0.46}_{-0.40}$ (stat) $^{+0.17}_{-0.13}$ (syst). These results are consistent with SM expectations.

We also validate the performance of the branching likelihood fitter by generating 10,000 pseudo-experiments. Each pseudo-experiment is defined by a random smearing of the default likelihood inputs around their expected values. For example, the number of observed events $N_{\text{obs},i}$ for each pseudo-experiment is determined by calculating the number of expected signal events for $R_{\text{true}} = 1$ and by Gaussian smearing the measured background rates. Since the total expected rates depend on other quantities such as total integrated luminosity and the scale factor, we Gaussian-vary these

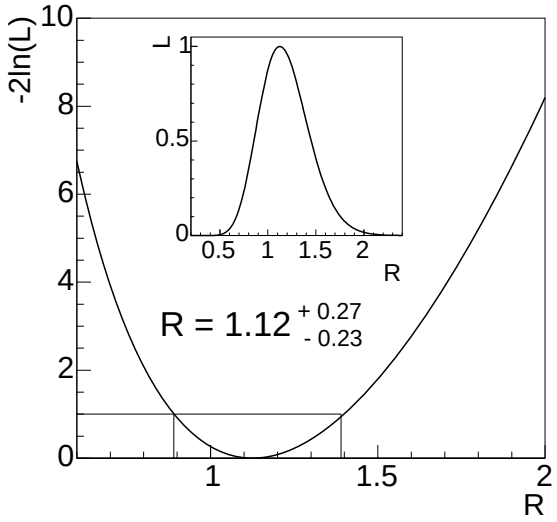
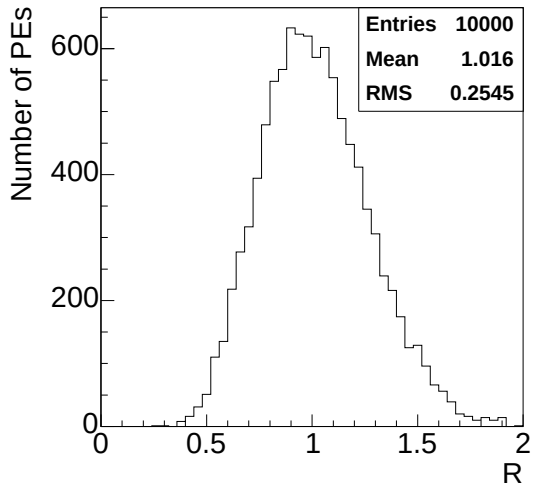
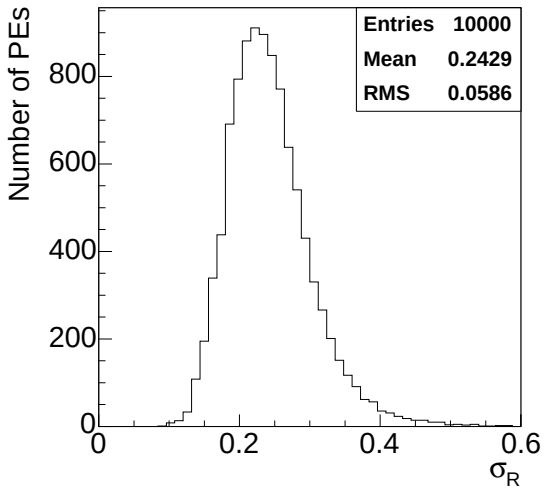


Figure 5.19: Combined likelihood for lepton-plus-jets and dilepton channels (inset) and its doubled negative logarithm. The intersections of the horizontal line $\ln L = 1$ with the likelihood define the total 1σ errors on R.

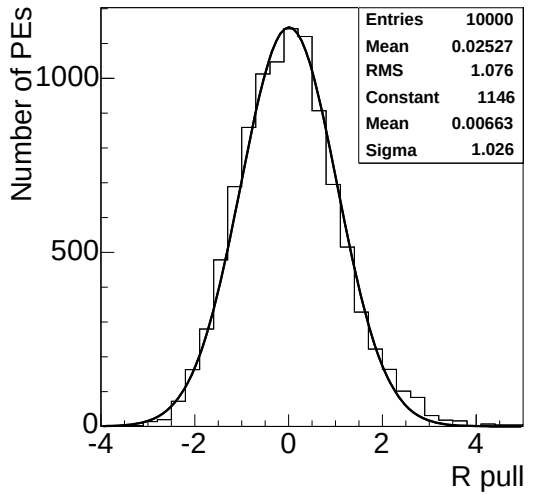
quantities as well. The detailed description of pseudo-experiments generation will be provided in Section 5.4. The results of R measurement in 10,000 pseudo-experiments for $R_{\text{true}} = 1$ are presented in Figure 5.20. The pull distribution $P = \frac{R-R_{\text{true}}}{\sigma_R}$ is consistent with a Gaussian distribution having mean centered at 0 and unit standard deviation. The correlation between the measured R and the corresponding uncertainty is due to a significant contribution from the ANN measurement in the 0-tag subsample. For those experiments where $N_{\text{obs},0}$ is large, the returned values of R are small. At the same time, the likelihood function tends to have a smaller width due to significant increase in the denominator of the corresponding Poisson term.



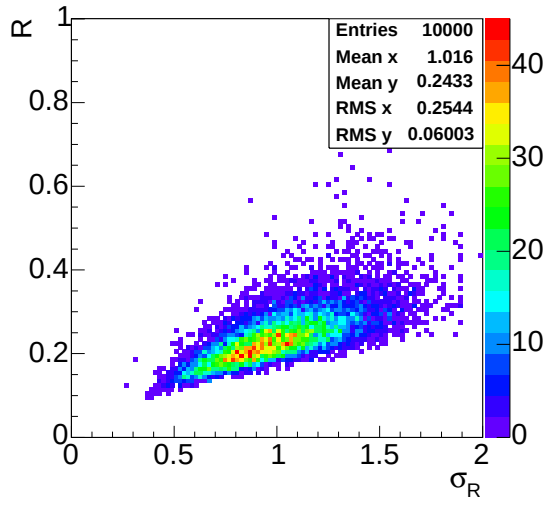
(a) Measured R.



(b) Parabolic error σ_R as returned by the likelihood minimization.



(c) Pull distribution $P = \frac{R-R_{\text{true}}}{\sigma_R}$.



(d) R versus σ_R.

Figure 5.20: Performance of the branching likelihood fitter as verified by measuring R in 10,000 pseudo-experiments with $R_{\text{true}} = 1$.

5.4 Confidence Level Limits on R

In the previous Chapter we presented the measurement of R and the estimation of its statistical and systematic uncertainties. In most measurements these uncertainties are considered sufficient to describe the range of values the measured quantity can take. However, in special cases where, for example, the p.d.f. of the estimator is not gaussian or the parameter's values must lie within physical boundaries, one can quote an interval covering the actual true value of the measured quantity with a given probability. In this final part of our analysis we describe the procedure aiming to set such boundaries.

The classical way of constructing the confidence level intervals is based on Neyman's procedure [73]. This technique utilizes the frequentist approach in defining the probability of a given outcome to occur, *i.e.* the probability distribution function of the measured quantity is defined after a large number of experiments has been performed. According to Neyman each measurement of quantity x has to be repeated for a given value of unknown parameter Θ for which we want to construct a confidence interval. All possible outcomes x will be described by the p.d.f. $P(x|\Theta)$, where, generally speaking, observable x does not have to be a function of Θ. In our case we would like to determine the probability $P(R|R_{\text{true}})$ that we measure R when the actual value is R_{true}. Since we perform our experiment only once, we cannot construct $P(R|R_{\text{true}})$ by repeating the measurement, instead, we generate many pseudo-experiments for a specific R_{true} and measure R for each one of them, using our analysis machinery.

After the p.d.f.s for different R_{true} values are generated in such way, a number of methods can be used to proceed with the construction of the confidence intervals. In this thesis we use a method based on the likelihood ratio ordering principle. This method is described by G. Feldman and R. Cousins in [74], where the main advantage of the method is pointed: it always providing physical and non-empty[6]intervals.

Below we present the pseudo-experiments generation and the confidence belts construction.

During the pseudo-experiment generation process, it is essential to vary all parameters which determine the signal and background yields in the three tagged subsamples in both lepton-plus-jets and dilepton channels. Based on the expected signal and background yields we calculate the total pseudo-experiment yields in each subsample and use them as input in our combined likelihood function. The block diagram of a single pseudo-experiment simulation is shown in Figure 5.21. We start by selecting the true value of R_{true} from interval of physical values between 0 and 1. We smear the luminosity, $\mathcal{L} = 193 \pm 11.2$ pb^{-1}, and theoretical prediction for $t\bar{t}$ cross section, $\sigma_{t\bar{t}} = 5.5 \pm 0.8$ pb, according to their uncertainties, and based on these values, we calculate the expected mean number of signal events in both channels, which in turn is smeared by Poisson distribution. The fraction of 0-tag, 1-tag, and 2-tag events are obtained by utilizing the event tagging efficiencies, $\epsilon_{0,1,2}$, computed for the chosen value R_{true}, and smeared by the scale factor SF. The same values of $\epsilon_{0,1,2}$ and SF are used to calculate the *a priori* background expectations in the lepton-plus-jets tagged samples. The background in the lepton-plus-jets 0-tag and dilepton samples are randomly chosen within their nominal measured values smeared by Poisson probability. Finally, we add the computed signal and background events to get the number of pseudo-experiment data events with different tag multiplicities. These values are fed into the combined branching likelihood function presented in Section 5.3.3. By minimizing the negative likelihood, we measure a value R for a given initial R_{true}. In total 10,000 pseudo-experiments are generated, for each 100 equidistant values of R_{true} from 0 to 1.

Figure 5.22 shows the further steps we perform in order to construct the confidence level belts. The results of pseudo-experiments generation are shown in Figure 5.22a,

[6]Non-empty interval is the interval which can be constructed in principle for a given R_{true}.

where the 2-dimensional histogram is filled with the number of pseudo-measurements of a given R_{true}. For every R_{true} we normalize the entries across the R bins and fit these distributions with a double bifurcated Gaussian. This way, we create a set of likelihood distributions $P(R|R_{\text{true}})$ to measure R when the actual mean value is R_{true}. The set of likelihoods is presented in Figure 5.22b and as an illustration a single likelihood for $R_{\text{true}} = 0.61$ is shown in Figure 5.22c. Next, according to the Feldman-Cousins ordering principle we calculate the likelihood ratio

$$r(R) = \frac{P(R|R_{\text{true}})}{P(R|R_{\text{best}})}, \tag{5.21}$$

where R_{best} is the value of R that maximizes the likelihood $P(R|R_{\text{true}})$ for a specific R_{true}. The likelihood ratios for different values of R_{true} are plotted in Figure 5.22d and Figure 5.22e shows a single likelihood ratio for $R_{\text{true}} = 0.61$. The likelihood ratio functions $r(R)$ are used to find intervals $[R_{\text{min}}, R_{\text{max}}]$ for each R_{true}, satisfying the following two conditions: First, $r(R_{\text{min}}) = r(R_{\text{max}})$, and second, the area under the corresponding likelihood function $P(R|R_{\text{true}})$ within $[R_{\text{min}}, R_{\text{max}}]$ must be equal to some desired probability. For example, the vertical lines in Figure 5.22e define the interval $[R_{\text{min}}, R_{\text{max}}]$, such that the 95% of the area under the likelihood function is constrained within this interval. The R-intervals for all values of R_{true} result to 2-dimensional bands shown in Figure 5.22f, where the 68% and 90% acceptance intervals are drawn on top of the 95% intervals. Finally, Figure 5.23 can be used to set the desired confidence intervals on any measured values of R by drawing a vertical line at the central value of R and finding the points where this line intersects the bands' boundaries. If the line intersects the boundary of a coloured band only once, then only a lower or upper limit can be set, whereas if the line has two intersections with a band's boundaries, then a two-sided confidence level interval can be set. For our case where $R = 1.12$, we find lower limits $R > 0.61$, $R > 0.69$, and $R > 0.84$ at 95%, 90%, and 68% CL respectively.

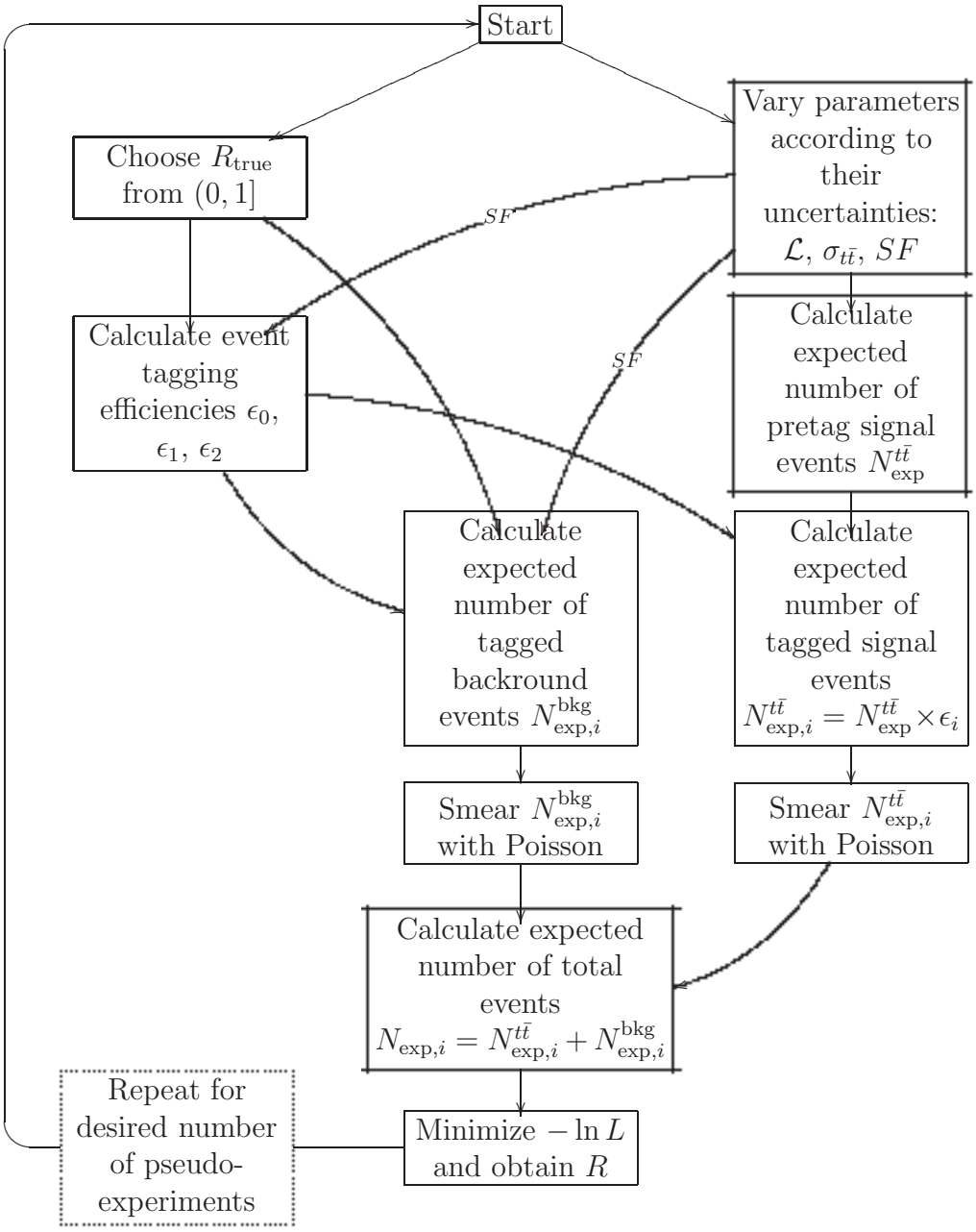
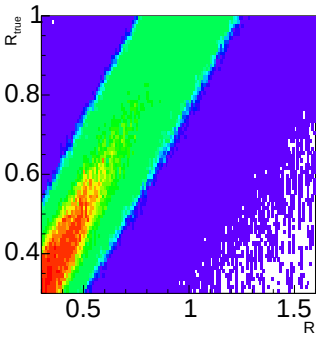
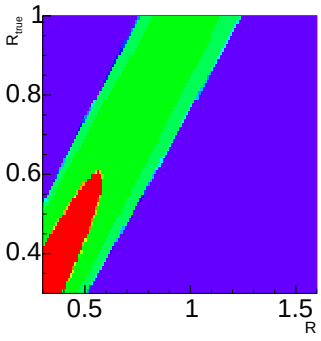


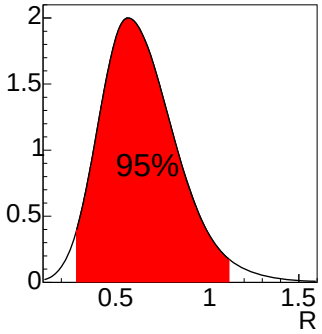
Figure 5.21: Block-diagram showing the general procedure of pseudo-experiment generation. See text for details.



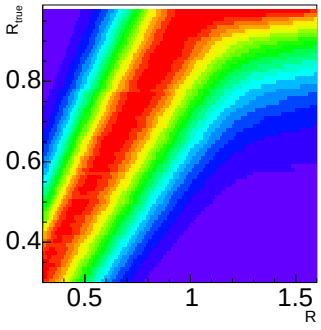
(a) Result of pseudo-experiments generation.



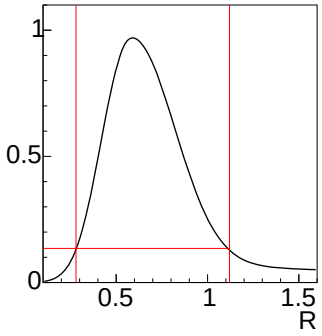
(b) Likelihoods $P(R|R_{\text{true}})$.



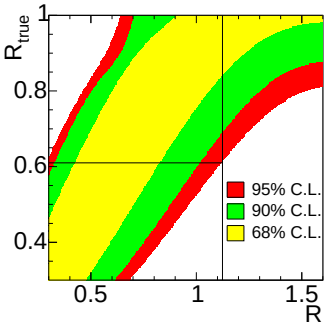
(c) Likelihood $P(R|R_{\text{true}})$ for $R_{\text{true}} = 0.61$.



(d) Likelihood ratios $\frac{P(R|R_{\text{true}})}{P(R|R_{\text{best}})}$



(e) Likelihood ratio $\frac{P(R|R_{\text{true}})}{P(R|R_{\text{best}})}$ for $R_{\text{true}} = 0.61$.



(f) Confidence level bands.

Figure 5.22: Step-by-step explanation of the construction of the Feldman-Cousins confidence bands. See text for details.

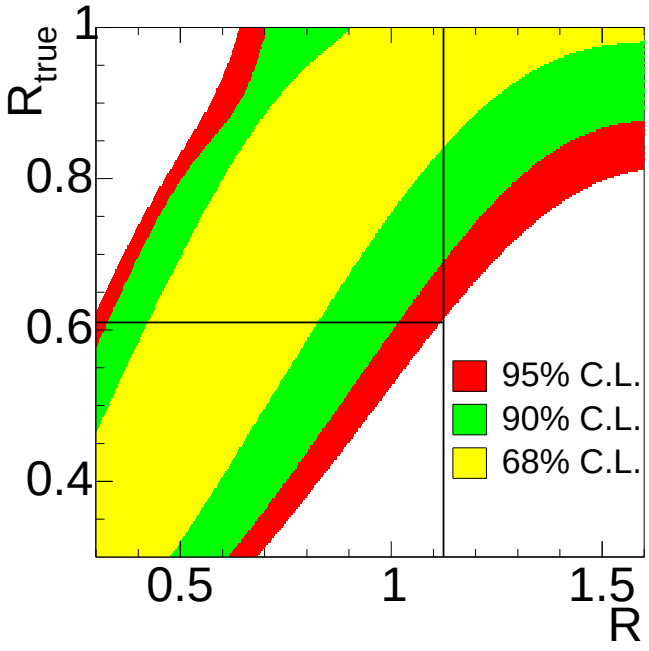


Figure 5.23: Confidence level bands constructed following the Feldman-Cousins technique. The innermost, middle, and outermost bands can be used to set the 68%, 90%, and 95% CL intervals or limits correspondingly.

Chapter 6

Summary and Conclusions

We have presented the measurement of the ratio of top quark branching fractions $R = \frac{\mathcal{B}(t\to Wb)}{\mathcal{B}(t\to Wq)}$, which shows how often the top quark decays to a bottom quark compared to any possible quark-channel decay. According to the Standard Model, with the assumption of three fundamental particle generations, and given CKM matrix unitarity, the value of R is expected to be very close to unity: $R > 0.998$ at 90% CL. Any deviation of the measured R from unity would indicate that one or more of our SM assumptions are not correct. For example, it would suggest a fourth generation of quarks, or a non-SM top decay, or a non-SM background to top pair production.

For this measurement we used a part of the data sample collected at CDF during the Run II period. Only events when all of the CDF subdetectors including SVX functioned properly are considered here, which corresponds to a total integrated luminosity of 162.1 pb^{-1}. The CDF-detected events are produced from the collisions of proton and antiproton beams accelerated by the Tevatron accelerator at center-of-mass energy $\sqrt{s} = 1.96$ TeV. At these energies top quarks are produced mainly in $t\bar{t}$ pairs via strong interactions. The top quark has an extremely short lifetime and decays weakly to a W boson and a quark q. Depending on whether the produced

W bosons decay hadronically or leptonically, the signal events can be arranged in three categories with respect to the expected final signatures: dilepton, lepton-plus-jets, and all-jets. In this analysis we performed the measurement of R using the lepton-plus-jets and dilepton channels.

The probability of the top quark to decay into bottom quark is measured from the relative rates of $t\bar{t}$ events with different number of b-jets. To identify the jets that originated from b quarks, we use the secondary vertex algorithm based on the long lifetime of B hadrons. In the data we observe events with 0, 1, or 2 b-tagged jets and we measure the signal fraction in these subsets. In principle, any two tagging rates could be used to determine the R uniquely but if all three values are used simultaneously the sensetivity of the measurement increases. The expected signal event tagging rates are predicted as a function of R, background rates, and estimation of jet b-tagging efficiencies from pure signal MC samples. The predicted number of signal events is compared with the observed ones by means of a likelihood function in which the unknown ratio R is allowed to vary. The maximum value of the likelihood corresponds to the best match between observed and predicted number of events and, therefore, to the most probable value of R.

In the lepton-plus-jets channel the background contamination in the selected data events is determined by using the Artificial Neural Network technique. This method allows us to combine multiple kinematic variables in order to enhance signal-backround separation. Only the best kinematic variables reflecting the difference in energy distribution and shape between signal and background events are selected. After training and selection of the most optimal ANN, a binned likelihood fit to the ANN output distribution is performed in all subsamples yielding the fraction of signal and background events. The achieved level of precision is comparable to the traditional background reduction performed on a sample with at least one secondary vertex. A more precise background estimation in the tagged samples can be achieved

with an *a priori* method utilizing the data-driven and simulation-based techniques. The dominant backgrounds to $t\bar{t}$ events are W production in association with heavy-flavour jets ($Wb\bar{b}$, $Wc\bar{c}$, Wc) and non-W events containing fake leptons or incorrectly measured missing energy. Smaller contributions to the background come from diboson (WW, WZ), single top quark production, and associative production of W with light quark that are mistakenly tagged ("mistags"). While the *a priori* method is good for estimating the background in the tagged samples, it cannot be applied to the 0-tag sample where the total rate of W + jets production is difficult to predict.

The dominant backgrounds in the dilepton sample are Drell-Yan production, fake leptons, and diboson (WW, WZ, ZZ) events. The background rates are estimated either from using complementary data samples or from simulations normalized to the standard-model cross sections. The fraction of tagged events is based on the probability to tag a generic QCD jet in the complementary data sample, since most of the jets in dilepton background events are arise from generic QCD radiation.

The likelihood function used to determine the most consistent with the observations value of R is a product of Poisson and Gaussian terms. The Poisson returns the probability to observe the measured number of events given an expected number of events, and the Gaussian reflects the uncertainties on measured quantities. In the lepton-plus-jets channel we can use information from ANN alone, which results in the following value:

$$R = 1.06^{+0.27}_{-0.24}\ (\text{stat}) \pm 0.16(\text{syst}). \tag{6.1}$$

Replacing the tagged ANN background estimations with the *a priori* ones, we get:

$$R = 1.02^{+0.23}_{-0.20}\ (\text{stat})\ ^{+0.21}_{-0.13}\ (\text{syst}). \tag{6.2}$$

Dilepton sample alone yields:

$$R = 1.41^{+0.46}_{-0.40}\ (\text{stat})\ ^{+0.17}_{-0.13}\ (\text{syst}). \tag{6.3}$$

Finally, combining lepton-plus-jets and dilepton information, and using the *a priori* method for the 1-tag and 2-tag letpon-plus-jets samples, we measure:

$$R = 1.12^{+0.21}_{-0.19}\ (\text{stat})\ ^{+0.17}_{-0.13}\ (\text{syst}). \tag{6.4}$$

The ratio R can only take on physical values between zero and unity. Using the combined measurement (Equation 6.4), we set a limit on R following the prescription proposed by Feldman and Cousins. After generating ensemles of pseudo-experiments for different values of R_{true} and using the likelihood-ratio ordering principle, we find that

$$R > 0.61 \text{ at the } 95\% \text{ CL} \tag{6.5}$$

From the last relation a lower limit on V_{tb} can be also set, using the CKM matrix unitarity and assuming only three quark generations. The result is

$$V_{tb} > 0.78 \text{ at the } 95\% \text{ CL} \tag{6.6}$$

The results presented in Equations 6.4, 6.5, and 6.6 can be compared with the corresponding measurements accomplished at CDF Run I, which are:

$$R = 0.94^{+0.26}_{-0.21}\ (\text{stat.})\ ^{+0.17}_{-0.12}\ (\text{syst.}),$$

$$R > 0.56 \text{ at } 95\% \text{ CL}$$

$$V_{tb} > 0.75 \text{ at } 95\% \text{ CL}$$

All of our current measurements are consistent with the Standard Model predictions. The lower limits on R and V_{tb} are the strongest constraints on these quantities placed to date. At the moment the results are dominated by statistical uncertainties. The systematic uncertainties are expected to reduce further with future improvements in the jet energy calibration in the data and MC background simulations. In Appendix C the reader can find a study on the prediction of future improvements of these lower limits as functions of the integrated luminosity and two extreme scenarios of systematic uncertainty development.

Appendices

Appendix A

Width of the Top Quark Weak Decay

Let us consider the weak decay of the top quark t with mass m_t and momentum $\mathbf{p}_t$ into a W boson and a quark q with masses and momenta correspondingly equal to m_W, m_q and $\mathbf{p}_W$ and $\mathbf{p}_q$ respectively. The simplest possible Feynman diagram for such decay is shown in Figure A.1. Due to the conservation of the total charge, the q quark has to be any "down" type quark ($q \equiv d,\ s,\ b$) with a charge $-e/3$. This interaction is represented by three terms in the Standard Model Lagrangian of the

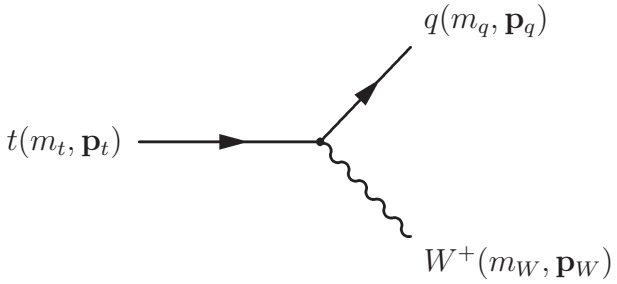


Figure A.1: Feynman diagram of the top quark weak decay.

form

$$\mathcal{L} \supset \frac{ig}{2\sqrt{2}} V_{mn} \bar{u}_m W^+_\mu \gamma^\mu (1+\gamma^5) v_n + h.c., \tag{A.1}$$

where indexes $m = t$ and $n = q = d, s, b$, and V_{mn} is an element of the CKM quark-mixing matrix.

In order to estimate the width Γ of the top quark decay, we can use the standard decay formula:

$$\Gamma(t \to W^+ q) = \frac{(2\pi)^4}{2m_t} \int \overline{|\mathcal{M}|^2} d\Phi, \tag{A.2}$$

where $\overline{|\mathcal{M}|^2}$ is the squared amplitude of the matrix element corresponding to $t \to W^+ b$ decay and averaged over the initial and final states polarizations. The integration is carried out over a 6-dimensional phase space with the $d\Phi$ element given by

$$d\Phi = \delta^4(p_t - p_W - p_q) \frac{d^3\mathbf{p}_W d^3\mathbf{p}_q}{(2\pi)^6 \cdot 2E_W \cdot 2E_q}, \tag{A.3}$$

where $p_t \equiv (-E_t; \mathbf{p}_t)$, $p_W \equiv (-E_W; \mathbf{p}_W)$, and $p_q \equiv (-E_q; \mathbf{p}_q)$ are the four-vectors of the corresponding particles. Applying the Feynman rules for the diagram shown in Figure A.1 we can write the matrix element as:

$$\mathcal{M} = -e_W \bar{u}(\mathbf{p}_q, \sigma_q) \epsilon^*_\mu(p_W, \lambda_W) V_{tq} \left[\gamma^\mu (1+\gamma_5)\right] v(\mathbf{p}_t, \sigma_t), \tag{A.4}$$

where spinors u and v describe the fermions with momentum $\mathbf{p}$ and spin σ, and ϵ_μ is a four-vector corresponding to a massive spin-one W boson with polarization λ_W. Now we have to square the matrix element and sum over the initial and final polarizations since we do not measure spins in the detector:

$$\begin{aligned} \overline{|\mathcal{M}|^2} &= \sum_{\lambda_W=-1}^{+1} \sum_{\sigma_t=-\frac{1}{2}}^{+\frac{1}{2}} \sum_{\sigma_q=-\frac{1}{2}}^{+\frac{1}{2}} |\mathcal{M}|^2 \\ &= e_W^2 |V_{tq}|^2 \sum_{\lambda_W} \epsilon_\mu \epsilon^*_\nu \sum_{\sigma_t} \sum_{\sigma_q} \left[\bar{u}\gamma^\mu(1+\gamma_5)v\right] \left[\bar{u}\gamma^\mu(1+\gamma_5)v\right]^*. \end{aligned} \tag{A.5}$$

The first sum can be calculated using the completeness relation for polarization vectors:

$$\sum_{\lambda_W} \epsilon_\mu \epsilon^*_\nu = \eta_{\mu\nu} + \frac{p_{W\mu} p_{W\nu}}{m_W^2}, \tag{A.6}$$

where in the last expression we used the usual Lorentz metric $\eta_{\mu\nu} = \mathrm{Diag}[-+++]$. The summation over the fermion spinors can be evaluated as a trace over the Dirac matrices using the following relations:

$$\bar{u}\mathcal{G}v = \sum_{ij} \bar{u}_i \mathcal{G}_{ij} v_j = \mathrm{Tr}\left[\mathcal{G}(u\bar{v})\right], \tag{A.7}$$

where $\mathcal{G}$ is a set of γ-matrices and $(u\bar{v})$ stands for the two-component matrix which elements are given by $(u\bar{v})_{ij} = u_i\bar{v}_j$. Thus, we get:

$$\begin{aligned}
&\sum_{\sigma_t}\sum_{\sigma_q} \left[\bar{u}\gamma^\mu(1+\gamma_5)v\right]\left[\bar{u}\gamma^\nu(1+\gamma_5)v\right]^* \\
&\qquad = -\sum_{\sigma_t,\sigma_q} \left[\bar{u}\gamma^\mu(1+\gamma_5)v\right]\left[\bar{v}\gamma^\nu(1+\gamma_5)u\right] \\
&\qquad = -\sum_{\sigma_t,\sigma_q} \mathrm{Tr}\left[\gamma^\mu(1+\gamma_5)v\bar{v}\gamma^\nu(1+\gamma_5)u\bar{u}\right] \\
&\qquad = \mathrm{Tr}\left[\gamma^\mu(1+\gamma_5)(m_t - \not{p}_t)\gamma^\nu(1+\gamma_5)(m_q - \not{p}_q)\right].
\end{aligned} \tag{A.8}$$

The last line in this expression is obtained by substituting the sums over the fermion spins with the identity:

$$\sum_{\sigma=-\frac{1}{2}}^{+\frac{1}{2}} u(\mathbf{p},\sigma)\bar{u}(\mathbf{p},\sigma) = m - \gamma^\mu p_\mu \equiv m - \not{p}. \tag{A.9}$$

In order to calculate the trace in Equation A.8 we use some common relations involving the Dirac matrices:

1. $\gamma_5\gamma^\mu = -\gamma^\mu\gamma_5$;
2. $\mathrm{Tr}[\gamma^{\mu_1}\ldots\gamma^{\mu_n}] = 0$ and $\mathrm{Tr}[\gamma_5\gamma^{\mu_1}\ldots\gamma^{\mu_n}] = 0$, if n is odd;
3. $\mathrm{Tr}[\gamma^\mu\gamma^\nu\gamma^\lambda\gamma^\rho] = 4(\eta^{\mu\nu}\eta^{\lambda\rho} - \eta^{\mu\lambda}\eta^{\nu\rho} + \eta^{\mu\rho}\eta^{\nu\lambda})$;
4. $\mathrm{Tr}[\gamma_5\gamma^\mu\gamma^\nu\gamma^\lambda\gamma^\rho] = 4i\epsilon^{\mu\nu\lambda\rho}$.

Now with the rules listed above in hand, we transform the trace into the following expression:

$$\begin{aligned}
&\mathrm{Tr}\left[\gamma^\mu(1+\gamma_5)(m_t-\not{p}_t)\gamma^\nu(1+\gamma_5)(m_q-\not{p}_q)\right] \\
&\quad=\mathrm{Tr}\left[\gamma^\mu(1+\gamma_5)\gamma^\nu(1+\gamma_5)m_tm_q-\gamma^\mu(1+\gamma_5)\not{p}_t\gamma^\nu(1+\gamma_5)\not{p}_q\right] \\
&\quad=\mathrm{Tr}\left[\gamma^\mu(1-\gamma_5^2)\gamma^\nu m_tm_q-\gamma^\mu(1+\gamma_5)^2\not{p}_t\gamma^\nu\not{p}_q\right] \\
&\quad=8(\eta^{\mu\lambda}\eta^{\nu\rho}-\eta^{\mu\nu}\eta^{\lambda\rho}+\eta^{\mu\rho}\eta^{\lambda\nu})p_{t\lambda}p_{q\rho}+8i\epsilon^{\mu\nu\lambda\rho}p_{t\lambda}p_{q\rho}. \qquad \text{(A.10)}
\end{aligned}$$

Finally, combining Equations A.8 and A.10, and substituting the resulting expression along with Equation A.6 into the expression for the squared matrix element (Equation A.5), we obtain:

$$\begin{aligned}
\overline{|\mathcal{M}|^2}&=8e_W^2|V_{tq}|^2\left(\eta_{\mu\nu}+\frac{p_{W\mu}p_{W\nu}}{m_W^2}\right)\left(p_t^\mu p_q^\nu+p_t^\nu p_q^\mu-\eta^{\mu\nu}p_t\cdot p_q\right) \\
&=8e_W^2|V_{tq}|^2\left(-p_q\cdot p_t+2\frac{(p_q\cdot p_W)(p_t\cdot p_W)}{m_W^2}\right) \qquad \text{(A.11)} \\
&=8e_W^2|V_{tq}|^2\left(E_qm_t+2\frac{E_Wm_t(E_qE_W+|\mathbf{p}_q|^2)}{m_W^2}\right). \qquad \text{(A.12)}
\end{aligned}$$

In the last line of Equation A.12 we assumed that the system is at rest and the four-vector $p_t=(m_t;0)$. The event kinematics in that case is defined by the following relations:

$$m_t=E_q+E_w=\sqrt{m_q^2+|\mathbf{p}_q|^2}+\sqrt{m_W^2+|\mathbf{p}_W|^2}, \qquad \text{(A.13)}$$

$$|\mathbf{p}_q|=-|\mathbf{p}_W|\equiv|\mathbf{p}|. \qquad \text{(A.14)}$$

Solving the system of the above equations we obtain for E_q, E_W, and $|\mathbf{p}|$:

$$E_q=\frac{m_t^2+m_q^2-m_W^2}{2m_t}, \qquad \text{(A.15)}$$

$$E_W=\frac{m_t^2-m_q^2+m_W^2}{2m_t}, \qquad \text{(A.16)}$$

$$|\mathbf{p}|^2=\frac{[(m_q+m_W)^2-m_t^2]\,[(m_q-m_W)^2-m_t^2]}{4m_t^2}. \qquad \text{(A.17)}$$

The expression for the top quark decay width given by Equation A.2 now can be written in explicit form:

$$\begin{aligned}
\Gamma &= \frac{(2\pi)^4}{2m_t}\int \overline{|\mathcal{M}|^2}\delta(m_t - E_q - E_W)\delta^3(\mathbf{p}_W + \mathbf{p}_q)\frac{d^3\mathbf{p}_W d^3\mathbf{p}_q}{(2\pi)^6\cdot 4E_W E_q} \\
&= \frac{(2\pi)^4}{2m_t}\int \overline{|\mathcal{M}|^2}\delta(m_t - E_q - E_W)\frac{4\pi|\mathbf{p}|^2 d|\mathbf{p}|}{(2\pi)^6\cdot 4E_W E_q} \\
&= \frac{(2\pi)^4}{2m_t}\int \overline{|\mathcal{M}|^2}\delta(|\mathbf{p}'|^2 - |\mathbf{p}|^2)\frac{E_W E_q}{(E_W + E_q)}\frac{4\pi|\mathbf{p}'|d|\mathbf{p}'|^2}{(2\pi)^6\cdot 2\cdot 4E_W E_q} \\
&= \frac{(2\pi)^4}{2m_t}\left\{8e_W^2|V_{tq}|^2\left(E_q m_t + 2\frac{E_W m_t(E_q E_W + |\mathbf{p}|^2)}{m_W^2}\right)\right\}\frac{4\pi|\mathbf{p}|}{(2\pi)^6\cdot 8m_t} \qquad \text{(A.18)} \\
&\equiv |V_{tq}|^2\Phi(m_q). \qquad \text{(A.19)}
\end{aligned}$$

The second line in this chain of identities was derived by taking the integral over the 3-dimensional delta function corresponding to the condition in Equation A.14. In going from the second to the third lines we used the fact that E is a function of $|\mathbf{p}|$, and, therefore, applied the property of the Dirac's delta function of a function: $\delta(g(|\mathbf{p}'|^2)) = \frac{\delta(|\mathbf{p}'|^2 - |\mathbf{p}|^2)}{g'(|\mathbf{p}|)}$. The line before the last one simply employs the fact that $\int f(x)\delta(x-a)dx = f(a)$. One can note that the contributions from the partial top quark decay widths are proportional to the $|V_{tq}|^2$ with some invariant coefficient $\Phi(m_q)$.

Let us assume the following numerical values for the quantities in Equation A.19: $m_W = 80.4$ GeV, $m_t = 172.7$ GeV, $m_b = 4.2$ GeV, $m_s = 0.1$ GeV, $m_d = 0.006$ GeV, and $e_W^2 = \frac{g_2^2}{(2\sqrt{2})^2} = \frac{\pi\alpha_2}{2} = 0.053$. It is natural to neglect the mass of the final state quark due to its small value comparing to the W boson and top quark masses. Now, calculating the identities in Equation A.17 and substituting them into the final expression for the top partial decay width we obtain:

$$\Gamma(t \to W^+ q) = |V_{tq}|^2\Phi(m_q), \qquad \text{(A.20)}$$

$$\text{where } \Phi(m_b) \approx \Phi(m_s) \approx \Phi(m_d) \approx 1.52 \text{ GeV}. \qquad \text{(A.21)}$$

An explicit calculation, *i.e.* without neglecting the mass of the light quark q, of the

top decay width gives the difference in $\Phi(m_b)$ and $\Phi(m_{d,s})$ values less than 0.2%.

Under the restriction that only 3 generations exist, the branching ratio of the top quark decaying into the bottom quark is:

$$R(t \to W^+ b) = \frac{\Gamma(t \to W^+ b)}{\sum\limits_{q=b,s,d} \Gamma(t \to W^+ q)} = \frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2}. \tag{A.22}$$

The denominator in this expression equals unity, as follows from the unitarity of the CKM matrix V [75].

Up to this moment, we did not assume any particular value for $|V_{tq}|$. These values can be estimated from the unitarity of the CKM matrix, using the identity $V^\dagger V = 1$. The latter is equivalent to the system of nine equations, where the relevant three are:

$$|V_{ud}|^2 + |V_{cd}|^2 + |V_{td}|^2 = 1, \tag{A.23}$$

$$|V_{us}|^2 + |V_{cs}|^2 + |V_{ts}|^2 = 1, \tag{A.24}$$

$$|V_{ub}|^2 + |V_{cb}|^2 + |V_{tb}|^2 = 1. \tag{A.25}$$

The elements $|V_{uq}|$ and $|V_{cq}|$ are measured with a good precision [10], therefore, the above conditions give $|V_{td}| = 0.009$, $|V_{ts}| = 0.040$, and $|V_{tb}| = 0.9992$. Since the partial width decay is proportional to $|V_{tq}|^2$, we should expect top quark decay into the bottom quark almost 100% of the time.

Another interesting property of the top quark can be concluded from the expected value of $|V_{tb}|^2 \approx 1$. In that case, the width $\Gamma(t \to W^+ b) \approx 1.5$ GeV converts into an extremely short lifetime

$$\tau = \frac{\hbar}{\Gamma} \approx 4.1 \times 10^{-25} \text{ sec.} \tag{A.26}$$

Such a short lifetime leads to almost immediate decay of the top quark before any top-flavored hadrons can be formed.

Appendix B

Method of Maximum Likelihood

The maximum likelihood method was first introduced by R. A. Fisher in 1922. This method is commonly used in experimental science where one needs to estimate either one or a set of several unknown parameters describing a given observation. Let us denote a set of N unknown parameters defining an expected observation as $\mathbf{\Theta} \equiv (\Theta_1, \ldots, \Theta_N)$ and a set of K independent measurements of quantity x as $\mathbf{x} \equiv (x_1, \ldots, x_K)$. Further, let us assume that the actually measured values $\mathbf{x}$ are coming from an underlying distributions with a known probability density function (p.d.f.) $f = f(x|\mathbf{\Theta})$. The joint p.d.f. of all independent measurements is called the *likelihood function* and can be written as

$$L(\mathbf{\Theta}) = \prod_{i=1}^{K} f(x_i|\mathbf{\Theta}). \tag{B.1}$$

In this form the likelihood function $L(\mathbf{\Theta})$ is understood to be a function of unknown parameters $\mathbf{\Theta}$ which values $\hat{\mathbf{\Theta}} \equiv (\hat{\Theta}_1, \ldots, \hat{\Theta}_N)$ are called the *best maximum likelihood estimators* if they maximize the likelihood: $L_{\max} = L\big|_{\mathbf{\Theta}=\hat{\mathbf{\Theta}}}$.

Sometimes, it is the case that an observable phenomena can require measurements of several independent quantities $x^{(1)}, x^{(2)}, \ldots, x^{(M)}$ in which case the final likelihood

function is just a product of M independent likelihoods,

$$L = L^{(1)} \cdot L^{(2)} \cdot \ldots \cdot L^{(M)}. \tag{B.2}$$

The maximum likelihood estimators $\hat{\mathbf{\Theta}}$ can be found by solving the system of *likelihood equations*

$$\frac{\partial L}{\partial \Theta_i} = 0, \qquad i = 1, \ldots, N \tag{B.3}$$

or, most commonly used in practice, by finding the maximum likelihood value $L_{\max}$ numerically. Instead of locating maximum of L one can equally search for a minimum of a negative logarithm of L and obtain the same parameter values $\hat{\mathbf{\Theta}}$, i.e. $L'_{\min} = -\ln L_{\max} = -\ln L\big|_{\mathbf{\Theta}=\hat{\mathbf{\Theta}}}$. This transformation is especially useful when p.d.f.'s are of exponential form as, for instance, in the Poisson and Gaussian distributions. In general case, the probability density functions must be normalizable for all possible values of free parameters. If the observable x is in the range $x_{\min} \leq x \leq x_{\max}$ then

$$\int_{x_{\min}}^{x_{\max}} f(x|\mathbf{\Theta})dx = A(\mathbf{\Theta}), \tag{B.4}$$

where $A(\mathbf{\Theta})$ is required to take only finite values in order for a numerical minimization procedure to converge, otherwise free parameters may run away to infinity. Actually, any normalization factor in p.d.f. can be dropped completely if it depends only on measured values and constants but not on the variable parameters Θ_i. This can be also seen from the fact that taking logarithm of L transforms any multiplicative term into an additive one, causing shift in $L'_{\min}$ absolute value and not affecting the location of the minimum. The consequence of such arbitrary normalization is the meaningless of a likelihood value at the solution point. In contrary, the difference in L' values between two points in parameter space is important, especially for the estimation of the parameter's errors.

B.1 Parameter's Uncertainty

The parameters' errors can be estimated by calculating the *error matrix* which is also called the *covariant matrix*. The covariant matrix V is a $N \times N$ matrix which elements are defined for each pair of all parameters $\mathbf{\Theta}$ as

$$V_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\Theta_i - \bar{\Theta}_i)(\Theta_j - \bar{\Theta}_j) L(\Theta_i, \Theta_j) d\Theta_i d\Theta_j. \tag{B.5}$$

The Equation B.5 is a standard definition of the first central moment of parameters Θ_i and Θ_j about their mean values $\bar{\Theta}_i$ and $\bar{\Theta}_j$ and a joint probability function $L(\Theta_i, \Theta_j)$. In the special case where $j = i$, the covariance V_{ii} is equal to the usual variance $\sigma_{ii} \equiv \sigma_i^2$ of variable parameter Θ_i.

A reasonable approximation of a general case can be made if marginal p.d.f.'s for each Θ have a gaussian form. In that case, the shape of the likelihood function L is also very close to a Gaussian and $\ln L$ is a hyper-parabola. Such an approximation makes an estimation of the covariant matrix elements quite simple. In case of only one parameter we have:

$$L = c \exp\left(-\frac{1}{2}\frac{(\Theta - \bar{\Theta})^2}{\sigma^2}\right) \tag{B.6}$$

where c is some constant which does not depend on Θ. From this formula we can write for the variance:

$$(\sigma^2)^{-1} = -\frac{\partial^2 \ln L}{\partial^2 \Theta}. \tag{B.7}$$

The last two equations can be easily generalized to represent a multidimensional case where L depends on a vector of N parameters $\mathbf{\Theta}$ and "multidimensional variance" is represented by the covariant matrix V:

$$L = c \exp\left(-\frac{1}{2}(\mathbf{\Theta} - \bar{\mathbf{\Theta}})^{\mathrm{T}} V^{-1} (\mathbf{\Theta} - \bar{\mathbf{\Theta}})\right), \tag{B.8}$$

from where

$$V_{ij}^{-1} = \frac{\partial \ln L}{\partial \Theta_i \partial \Theta_j}. \tag{B.9}$$

One should not forget that the last equality is an approximation, since, in general, L might not have a gaussian form.

Let us consider one dimensional case again where $-\ln L(\Theta) = -\frac{1}{2}\frac{(\Theta-\bar{\Theta})^2}{\sigma^2}$ is a simple parabola, and let us assume that the $-\ln L(\Theta)$ was minimized. The best estimators for the true parameter's value $\bar{\Theta}$ was found to be $\hat{\Theta}$ such that $-\ln L_{\max} = -\ln L(\Theta)\big|_{\Theta=\hat{\Theta}}$. We are interested in determining the uncertainty $\pm\Delta\Theta$ on the measured parameter's value $\hat{\Theta}$ numerically. The former can be expressed in terms of s standard deviation errors from the mean value:

$$\pm\Delta\Theta = \pm s\sigma = \Theta - \hat{\Theta}. \tag{B.10}$$

By varying Θ in each direction by $\pm\Delta\Theta$ from the minimum, the variation of $\ln L$ can be written in the following form:

$$\ln L\big|_{\Theta=\hat{\Theta}\pm\Delta\Theta} - \ln L\big|_{\Theta=\hat{\Theta}} = \ln L(\Theta) - \ln L_{\max} = -\frac{s^2}{2}. \tag{B.11}$$

Therefore, determining the uncertainties on $\hat{\Theta}$ is equivalent to finding the intersections of $-\ln L$ with a straight line shifted up by $\frac{s^2}{2}$ from the minimum.

In multidimensional case the situation is similar except that $-\ln L(\boldsymbol{\Theta})$ is hyperparabola and parameters' uncertainties are defined by the N-dimensional ellipsoids $(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}})^T V^{-1} (\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}}) = s^2$.

B.2 Likelihood for Normal Distribution

The normal distribution is also commonly known as the Gaussian distribution. It is interpreted as a probability to measure a true value μ of some random variable x with a known variance σ^2. The probability function is given by the formula:

$$G(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{B.12}$$

Suppose we perform K independent measurements of x, $(x_1, \ldots, x_K)$, then it can be shown [76] that unbiased estimators for μ and σ are defined as

$$\hat{\mu} = \frac{\sum_{i=1}^{K} x_i}{K}, \tag{B.13}$$

$$\hat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{K}(x_i - \mu)^2}. \tag{B.14}$$

The true values μ and σ can be also estimated using the method of maximum likelihood. The joint likelihood function of K measurements has the form:

$$L = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdot \ldots \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_K-\mu)^2}{2\sigma^2}} = \frac{1}{(\sigma\sqrt{2\pi})^K} e^{-\sum_{i=1}^{K}\frac{(x_i-\mu)^2}{2\sigma^2}}, \tag{B.15}$$

and taking logarithm of L yields, it is transformed to:

$$\ln L = -\frac{K}{2}\ln(2\pi) - K\ln\sigma - \sum_{i=1}^{K}\frac{(x_i-\mu)^2}{2\sigma^2}. \tag{B.16}$$

In order to calculate the maximum likelihood estimators we take partial derivatives of $\ln L$ with respect to each parameter and set the equations equal to 0:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{K}(x_i - \mu) = 0, \tag{B.17}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{K}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{K}(x_i - \mu) = 0, \tag{B.18}$$

from where

$$\hat{\mu} = \frac{\sum_{i=1}^{K} x_i}{K}, \tag{B.19}$$

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_{i=1}^{K}(x_i - \mu)^2}. \tag{B.20}$$

Note that analytical determination of the maximum of L returns the same estimate for μ as in Equation B.13 but the expression for $\hat{\sigma}$ differs from the unbiased one given by Equation B.14. This difference is negligible if K is a large number.

There is at least one more type of practical cases where the Gaussian distribution can appear in a likelihood function. In such cases the normal distribution with known mean μ and variance σ is used as a constraint on a corresponding model parameter. This is opposite to the above example where we estimated the mean and the variance assuming that the measurements are distributed according to the normal probability.

B.3 Likelihood for Poisson Distribution

The Poisson distribution is useful in those cases where we need to know the probability of selecting n rare events while knowing the prediction for the expected mean number of selected events. The number of such expected events can be written as $\lambda \equiv Np$, where N is a large number of total trials and p is a probability of choosing an event, which value is small. Mathematically the Poisson probability is given by the limit of binomial distribution where $N \to \infty$ and $p = \frac{\lambda}{N} \to 0$:

$$\begin{aligned} P(n;\lambda) &= \lim_{\substack{N\to\infty \\ p=\frac{\lambda}{N}}} \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \\ &= \lim_{N\to\infty} \frac{N!}{n!(N-n)!} \left(\frac{\lambda}{N}\right)^n \left(1-\frac{\lambda}{N}\right)^{N-n} \\ &= \lim_{N\to\infty} \underbrace{\frac{N(N-1)\cdot\ldots\cdot(N-n+1)}{N^n}}_{1} \frac{\lambda^n}{n!} \underbrace{\left(1-\frac{\lambda}{N}\right)^N}_{e^{-\lambda}} \underbrace{\left(1-\frac{\lambda}{N}\right)^{-n}}_{1} \\ &= \frac{\lambda^n e^{-\lambda}}{n!}. \end{aligned} \tag{B.21}$$

In those experiments where we do not know the *a priori* value of λ, it can be estimated by solving the likelihood Equations B.3. Assuming that the measurements of x are distributed according to the Poisson probability, the joint likelihood function is given by

$$L = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \ldots \cdot \frac{e^{-\lambda}\lambda^{x_K}}{x_K!} = \frac{e^{-K\lambda}\lambda^{\sum_i x_i}}{x_1! \cdot \ldots \cdot x_K!}. \tag{B.22}$$

It is easier to take logarithm of L and calculate an extremum for $\ln L$, in order to obtain the best estimator for λ,

$$\ln L = -n\lambda + \ln\lambda \sum_{i=1}^{K} x_i - \ln\left(\prod_{i=1}^{K} x_i!\right). \tag{B.23}$$

Finally, solving for $\hat{\lambda}$, the equation

$$\frac{\partial \ln L}{\partial \lambda} = -n + \frac{\sum_{i=1}^{K} x_i}{\ln\lambda} = 0, \tag{B.24}$$

returns for $\hat{\lambda}$

$$\hat{\lambda} = \frac{\sum_{i=1}^{K} x_i}{n}. \tag{B.25}$$

This result indicates that the set of given measurements $(x_1, \ldots, x_K)$ will be best described by the Poisson distribution with a λ equal to their mean value which, by the way, does not have to be integer.

In conclusion, an important note can be made on the Poisson distribution. Although, it is primarily defined for processes involving integer number of events n, we can calculate the Poisson probability for any positive real number. In that case the factorial in the denominator is replaces by the gamma function extending the Poisson distribution to the class of non-integer arguments:

$$n! \equiv \Gamma(n+1) \equiv \int_0^\infty e^{-t} t^n dt. \tag{B.26}$$

B.4 Binned Maximum Likelihood Fit

Let us again consider the measurements $(x_1, \ldots, x_K)$ of some quantity x distributed among M bins with a fixed width $w = \frac{x_{\max} - x_{\min}}{M}$. A relevant histogram will be filled with exactly K events in total and the number of events in each bin will be $(n_1, \ldots, n_M)$, where n_b for the b^{th} bin is given by

$$n_b = \sum_{i=1}^{K} \int_{x_{\min}+w(b-1)}^{x_{\min}+wb} \delta(x - x_i) dx. \tag{B.27}$$

One can think of each individual measurement as a successfully selected event if its value falls inside the bin b. In that case, the probability of measuring n_b events given an expected number of events f_b in the same bin is defined by the *binomial distribution*:

$$P(n_b|K) = C_K^{n_b} \left(\frac{f_b}{K}\right)^{n_b} \left(1 - \frac{f_b}{K}\right)^{K-n_b}, \tag{B.28}$$

where $C_K^{n_b} = \frac{K!}{n_b!(K-n_b)!}$ is a *binomial coefficient*. According to our hypothesis all measurements come from the expected distribution described by a p.d.f. $f(x|\mathbf{\Theta})$, thus the number of expected events f_b can be estimated by the value of the p.d.f. at the center of a b^{th} bin which is given by $f_b \equiv f_b(\mathbf{\Theta}) = f(x_{\min} + w(b - \frac{1}{2})|\mathbf{\Theta})$. As it follows from its definition, this value does not at all need to be an integer.

In the cases where the registered events are rare, i.e. the probability of successes is small, the conditional probability given by Equation (B.28) can be approximated with a good precision by a *Poisson distribution*:

$$P(n_b) = \frac{f_b^{n_b} e^{-f_b}}{n_b!}. \tag{B.29}$$

The likelihood function is given by the joint probability calculated over all bins in the histogram as:

$$L(\mathbf{\Theta}) = \prod_{b=1}^{M} \frac{f_b^{n_b}(\mathbf{\Theta}) e^{-f_b(\mathbf{\Theta})}}{n_b!}. \tag{B.30}$$

Taking logarithm of this product and multiplying both sides of this equation by -1 we obtain the minimization function:

$$\Lambda \equiv -\ln L(\mathbf{\Theta}) = -\sum_{b=1}^{M} (n_b \ln f_b(\mathbf{\Theta}) - f_b(\mathbf{\Theta})) + \sum_{b=1}^{M} \ln(n_b!), \tag{B.31}$$

where the last sum can be dropped because it does not depend on $\mathbf{\Theta}$.

An interesting conclusion can be made on Λ: If we substitute the n_b and return

back to the exact expression for $f(x|\mathbf{\Theta})$ instead of f_b in the last equation, we obtain:

$$\begin{aligned}
\Lambda &\approx -\sum_{b=1}^{M}\left(\sum_{i=1}^{K}\int_{x_{\min}+w(b-1)}^{x_{\min}+wb}\delta(x-x_i)dx\ln f_b(\mathbf{\Theta})\right)-\sum_{b=1}^{M}f_b(\mathbf{\Theta})\\
&=-\sum_{i=1}^{K}\left(\sum_{b=1}^{M}\int_{x_{\min}+w(b-1)}^{x_{\min}+wb}\delta(x-x_i)\ln f(x|\mathbf{\Theta})dx\right)-N(\mathbf{\Theta})\\
&=-\sum_{i=1}^{K}\int_{x_{\min}}^{x_{\max}}\ln f(x|\mathbf{\Theta})\delta(x-x_i)dx-N(\mathbf{\Theta})\\
&=-\sum_{i=1}^{K}\ln f(x_i|\mathbf{\Theta})-N(\mathbf{\Theta}),
\end{aligned}\tag{B.32}$$

where $N(\Theta)$ can be considered as a normalization of $f(x|\mathbf{\Theta})$. Introducing normalized likelihood function $f'(x|\mathbf{\Theta})=\frac{f(x|\mathbf{\Theta})}{N(\mathbf{\Theta})}$, Equation B.32 can be rewritten as:

$$\Lambda(\mathbf{\Theta})\approx\sum_{i=1}^{K}\ln f'(x_i|\mathbf{\Theta}),\tag{B.33}$$

which is equivalent to the definition of unbinned likelihood function given by Equation B.1. From the comparison of the binned and unbinned likelihood functions one may conclude that both approaches will give us consistent results. Decrease in the bins' width will lead to a better agreement between binned and unbinned likelihood functions.

Appendix C

Predictions for R and V_{tb} Lower Limits at Higher Luminosities

An interesting prediction can be made on lower limits of R, and hence V_{tb}, as functions of luminosity. A study was performed using the same likelihood function as for the combined lepton-plus-jets and dilepton measurement of R (see Section 5.3). This investigation is not intented to provide a scrutinized examination of the problem but rather give a general idea of what future measurements of the top branching fractions may look like. Therefore, a number of assumptions are made. We begin with the assumption that the number of signal and background events scales linearly with the integrated luminosity. Also, we assume that the standard deviations of the established lower limits on R are solely functions of the uncertainties of the measured value of R itself.

Ideally, one would need to generate a large set of pseudo-experiments for different values of R_{true} and a given luminosity value using the same procedure as in the construction of the Feldman-Cousins CL intervals (see Section 5.4). Given the level of assumptions, we limit ourselves to only three "ideal" pseudo-experiments for every

luminosity value. Here "ideal" means that calculated number of expected signal and background events are used directly as number of "observed" events, *i.e.* without being fluctuated statistically. The selected luminosity values are increased with a step of $\sqrt{2}$ between 140 pb^{-1} and 20 fb^{-1}. The three pseudo-experiments are implemented for three different values of R_{true}, where one of them represents the expected central value of $R_{\text{true}} = 1$ and the other two are calculated at R_{true} plus or minus one standard deviation of the central measurement. In order to calculate the central lower limit on R along with the standard deviations on it, we proceed in the following steps:

1. For a given luminosity and a nominal value of $R_{\text{true}} = 1$ we calculate the expected number of events with i b-tags, $N_{\text{exp},i}$.

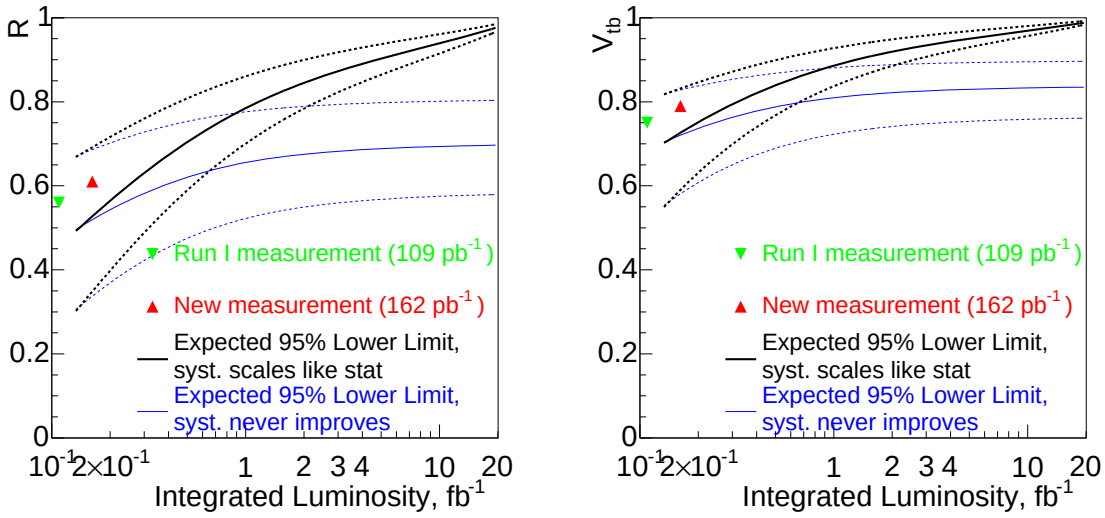
2. The pseudo-experiment is simulated by simply assigning $N_{\text{exp},i}$ to the number of pseudo-measured events $N_{\text{obs},i}$.

3. The numbers of observed events $N_{\text{obs},i}$ along with the other parameters are used as input in our combined likelihood function and the first pseudo-measurement of R is performed by maximizing it. We call the measured value R_c and the uncertainties $\pm\sigma_{R_c}$ on R_c are calculated from the shape of the likelihood function as usual.

4. Two more pseudo-experiments are simulated for the same luminosity but using $R_{\pm} = R_c \pm \sigma_{R_c}$ as R_{true}.

5. The three separate likelihoods corresponding to the three different $R_{\text{true}} = R_c, R_{\pm}$ are integrated to find 95% Bayesean lower credibility interval for the given luminosity. The likelihood with nominal R_{true} value provides the central lower limit on R, whereas the other two correspond to one standard deviation from the central value.

The lower limit on R and its standard deviations are calculated for all desired



(a) Ratio R of top branching fractions.

(b) CKM matrix element $V_{tb} = \sqrt{R}$.

Figure C.1: Future prediction for the measurement of the lower limit on R and V_{tb} as functions of integrated luminosity. The solid lines show the expected central lower limit and dotted lines correspond to the $\pm 1\sigma$ deviations for this value. The lower three lines represent the case when the systematic uncertainty remains at the current level and the upper three lines reflect the case when the systematic uncertainty decreases along with the statistical one.

values of luminosity which can be viewed as functions of σ_{R_c}. This parametrization is essential if one wants to see the variation of the lower limits under either of the two following cases. First, when the total error includes the systematic uncertainty on the real measurement $\sigma_R^{\text{syst}} = {}^{+0.17}_{-0.13}$ which is added to the σ_{R_c} in quadrature as a constant term, and, second, when the systematic uncertainty is scaled down by the ratio $\frac{\sigma_{R_c}}{\sigma_R^{\text{stat}}}$, where $\sigma_R^{\text{stat}} = {}^{+0.21}_{-0.19}$ is the statistical uncertainty on the current real measurement. Given this, we plot three curves for each case as shown in Figure C.1a. Further, assuming existence of only three quark generations, we plot lower limits on V_{tb} as shown in Figure C.1b by merely taking the square root of the R curves.

References

[1] F. Abe et al. Evidence for top quark production in anti-p p collisions at s**(1/2) = 1.8-tev. *Phys. Rev.*, D50:2966–3026, 1994.

[2] F. Abe et al. Observation of top quark production in anti-p p collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.

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

[4] the Tevatron Electroweak Working Group. Combination of cdf and d0 results on the top-quark mass. 2005.

[5] E. Noether. Invariante Varlationsprobleme. *Nachr. d. Konig. Gesellsch. d. Wiss. zu Gottingen, Math-phys. Klasse*, pages 235–257, 1918. English translation M. A. Travel, Transport Theory and Statistical Physics 1(3) 1971,183-207.

[6] A. D. Martin F. Halzen. *Lepton and Quarks: An Introductory Course in Modern Particle Physics*. John Wiley and Sons, Inc, 1984.

[7] S. F. Novaes. Standard model: An introduction. 1999.

[8] N. Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–532, 1963.

[9] M. Kobayashi and T. Maskawa. Cp violation in the renormalizable theory of weak interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[10] K. Hagiwara et al. Review of particle physics. *Phys. Rev.*, D66:010001, 2002.

[11] S. Eidelman et al. Review of particle physics. *Phys. Lett.*, B592:1, 2004.

[12] Parton distribution functions.
See also `http://durpdg.dur.ac.uk/HEPDATA/PDF`.

[13] P. Nason, S. Dawson, and R. K. Ellis. The total cross-section for the production of heavy quarks in hadronic collisions. *Nucl. Phys.*, B303:607, 1988.

[14] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi. The t anti-t cross-section at 1.8-tev and 1.96-tev: A study of the systematics due to parton densities and scale dependence. *JHEP*, 04:068, 2004.

[15] Nikolaos Kidonakis and Ramona Vogt. Next-to-next-to-leading order soft-gluon corrections in top quark hadroproduction. *Phys. Rev.*, D68:114014, 2003.

[16] P. K. Sinervo. Top quark studies at hadron colliders. 1995. hep-ex/9608005.

[17] M. Beneke et al. Top quark physics. 2000. hep-ph/0003033.

[18] T. Affolder et al. First measurement of the ratio $B(t \to Wb)/B(t \to Wq)$ and associated limit on the CKM element $|V_{tb}|$. *Phys. Rev. Lett.*, 86:3233–3238, 2001.

[19] P. J. Bryant. *The principles of circular accelerators and storage rings*. Cambridge University Press, 1993.

[20] Fermilab Beam Division. Fermilab linac upgrade conceptual design revision 4a. FERMILAB-LU-CONCEPTUAL-DESIGN.

[21] R. Blair et al. The cdf-ii detector: Technical design report. FERMILAB-PUB-96-390-E.

[22] D. Acosta et al. Measurement of the j/psi meson and b-hadron production cross sections in p anti-p collisions at s**(1/2) = 1960- gev. 2004.

[23] D. Acosta et al. Measurement of the t anti-t production cross section in p anti-p collisions at $\sqrt{s} = 1.96$ tev using lepton + jets events with secondary vertex b-tagging. 2004.

[24] M. Aoki et al. The CDF Run IIb silicon detector. *Nucl. Instrum. Meth.*, A518:270–276, 2004.

[25] P. Merkel. The CDF silicon detector upgrade and performance. *Nucl. Instrum. Meth.*, A501:1–6, 2003.

[26] T. Affolder et al. CDF central outer tracker. *Nucl. Instrum. Meth.*, A526:249–299, 2004.

[27] R. Wigmans. *Calorimetry Energy Measurement in Particle Physics*. Claredon Press, Oxford, 2000.

[28] L. Balka et al. The cdf central electromagnetic calorimeter. *Nucl. Instrum. Meth.*, A267:272, 1988.

[29] S. Kuhlmann et al. The cdf calorimeter upgrade for run iib. *Nucl. Instrum. Meth.*, A518:39–41, 2004.

[30] G. Ascoli et al. Cdf central muon detector. *Nucl. Instrum. Meth.*, A268:33, 1988.

[31] T. Dorigo. The muon system upgrade for the cdf ii experiment. *Nucl. Instrum. Meth.*, A461:560–562, 2001.

[32] G. Apollinari, Konstantin Goulianos, P. Melese, and M. Lindgren. Shower maximum detector for the cdf plug upgrade calorimeter. *Nucl. Instrum. Meth.*, A412:515–526, 1998.

[33] E. J. Thomson et al. Online track processor for the cdf upgrade. *IEEE Trans. Nucl. Sci.*, 49:1063–1070, 2002.

[34] B. L. Winer. Status of the cdf ii online trigger. *Int. J. Mod. Phys.*, A16S1C:1169–1171, 2001.

[35] C. Peterson, D. Schlatter, I. Schmitt, and Peter M. Zerwas. Scaling violations in inclusive e+ e- annihilation spectra. *Phys. Rev.*, D27:105, 1983.

[36] J. Mueller B. Flaugher. A guide to JETCLU: the cdf jet cluster algorithm. CDF/DOC/JET/CDFR//1814, August 21, 1992.

[37] Gerald C. Blazey et al. Run ii jet physics. 2000.

[38] E. Thomson et al. Top and Electroweak Event Classification Module for CDF Run II. CDF/DOC/TOP/PUBLIC/5947, October 22, 2003.

[39] E. Thomson et al. TopNtuple: a Public Analysis Module for the Top Group. CDF/DOC/TOP/CDFR/6737, October 22, 2003.

[40] A. Wyatt and Beate Heinemann. Correction for Leakage Energy in the Central and Plug Calorimeters in Run II. CDF/ANAL/ELECTRON/CDFR/6167, December 7, 2002.

[41] T. Affolder et al. Measurement of the t anti-t production cross section in p anti-p collisions at s**(1/2) = 1.8-tev. *Phys. Rev.*, D64:032002, 2001.

[42] D. Acosta et al. A first look at the CLC Luminosity Measurements. CDF/ANA/6052, August 28, 2002.

[43] S. Jindariani et al. Luminosity Uncertainty for Run 2 up until August 2004. CDF/ANA/7446, February 3, 2005.

[44] S. Klimenko, J. Konigsberg, and T. M. Liss. Averaging of the inelastic cross sections measured by the cdf and the e811 experiments. FERMILAB-FN-0741.

[45] D. Acosta et al. Measurement of the cross section for tt-bar production in pp-bar collisions using the kinematics of lepton + jets events. *Phys. Rev.*, D72:052003, 2005.

[46] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[47] F. Rosenblatt. Principles of neurodynamics. *Washington D.C.: Spartan Press*, 1961.

[48] M. Minsky and S. Papert. An introduction to computational geometry. *The MIT Press*, 1969.

[49] P. G. Zimbardo R. J. Gerrig. *Psychology and Life*. Allyn & Bacon, May 14, 2004.

[50] David Hilbert. Mathematical problems. *Bull. Am. Math. Soc., New Ser.*, 37(4):407–436, 2000. Originally published as *Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematike-Congress zu Paris 1900*, Gött. Nachr. 1900, 253-297, Vandenhoeck & Ruprecht, Göttingen. Translated for the *Bulletin*, with the author's permission, by Dr. Mary Winston Newson, 1902.

[51] V. I. Arnold. Representation of continuous functions of three variables by the superposition of continuous functions of two variables. *Transl., Ser. 2, Am. Math. Soc.*, 28:61–147, 1963. Translation from Mat. Sb., N. Ser. 48(90), 3-74 (1959).

[52] A. N. Kolmogorov. On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. *Am. Math. Soc., Transl., II. Ser.*, 17:369–373, 1956. Translation from Dokl. Akad. Nauk SSSR 108, 179-182 (1956).

[53] A. N. Kolmogorov. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Transl., Ser. 2, Am. Math. Soc.*, 28:55–59, 1963. Translation from Dokl. Akad. Nauk SSSR 114, 953-956 (1957).

[54] White Hornick, Stinchcombe. Multilayer feedforward networks are universal approximators. *Neural Networks*, v. 2, #5, 1989.

[55] Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematical Control Signals Systems*, 2, 1989.

[56] Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, v. 2, #3, 1989.

[57] Maintainer: saswss@unx.sas.com (W. S. Sarle). Neural Network FAQ.
See also `ftp://ftp.sas.com/pub/neural/FAQ.html`.

[58] A. Zell et al. *SNNS: Stuttgart Neural Network Simulator*. User Manual, V4.2.
See also `http://www-ra.informatik.uni-tuebingen.de/SNNS/`.

[59] R. Brun and F. Rademakers. *ROOT — An Object Oriented Data Analysis Framework*. Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86.
See also `http://root.cern.ch`.

[60] Cdfsim software description.
See also `http://www-cdf.fnal.gov/cdfsim_cdfsim_main.html`.

[61] R. Brun and F. Carminati. Geant, detector description and simulation tool. *CERN Programming Library Long Writeup*, W5013, 1993.

[62] G. Corcella et al. Herwig 6.5: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 01:010, 2001. arXiv:hep-ph/0011363.

[63] J. Pumplin et al. New generation of parton distributions with uncertainties from global qcd analysis. *JHEP*, 07:012, 2002.

[64] M. L. Mangano et al. Alpgen, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003. arXiv:hep-ph/0206293.

[65] J. Strologas. *Measurement of the differential angular distribution of the W boson produced in association with jets in proton-antiproton collisions at sqrt(s)=1.8 TeV*. PhD thesis, University of Illinois at Urbana-Champaign, 2002.

[66] W. H. Press et al. *Numerical recipes in C: the art of scientific computing*. Press Syndicate of the University of Cambridge, 2nd edition, 1992.
See also `http://www.numerical-recipes.com`.

[67] Pekka K. Sinervo. Signal Significance in Particle Physics. CDF/PUB/STATISTICS/PUBLIC/6031, July 6, 2002.

[68] F. James. *MINUIT: Function Minimization and Error Analysis*. CERN, March 1994. CERN Program Library Long Writeup D506.

[69] Carsten Rott Oscar Gonzalez. Uncertainties due to the PDFs for the gluino-sbottom search. CDF/PHYS/EXOTIC/CDFR/7051, June 7, 2004.

[70] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne. Physical gluons and high-e(t) jets. *Phys. Lett.*, B604:61–68, 2004.

[71] D. Amidei J. S. Miller. Method 2 Background Estimation for the Secvtx Tagged Lepton+Jets Sample and Measurement of the $t\bar{t}$ Production Cross Section. CDF/ANAL/TOP/CDFR/6907, March 23, 2004.

[72] D. Acosta et al. *Phys. Rev. Lett.*, 93:142001, 2004. We restrict the 193 pb^{-1} data set used here to runs in which the silicon detector was included.

[73] J. Neyman. *Philos. Trans. R. Soc.*, A236:333, 1937.

[74] G. J. Feldman and R. D. Cousins. A unified approach to the classical statistical analysis of small signals. *Phys. Rev.*, D57:3873–3889, 1998.

[75] Ling-Lie Chau and Wai-Yee Keung. Comments on the parametrization of the kobayashi-maskawa matrix. *Phys. Rev. Lett.*, 53:1802, 1984.

[76] Eric W. Weisstein. Mathworld—a wolfram web resource.
See also `http://mathworld.wolfram.com/`.