A Search for the Top Quark Using Artificial Neural Networks

A Dissertation Presented

by

Erfan Amidi

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Physics

Northeastern University
Boston, Massachusetts

February 1996

NORTHEASTERN UNIVERSITY
Graduate School of Arts and Sciences

Dissertation Title: A Search for The Top Quark Using Neural Networks

Author: Erfan Amidi

Department: Physics

Approved for Dissertation Requirements of Doctor of Philosophy Degree

S. Neuchuh — Febr. 16, 1996

2/16/96

2/16/96

Thesis Committee — **Date**

Paul M. Chapin — 2/16/96

Head of Department — **Date**

Graduate School Notified of Acceptance

Kay Onan (mc) — 16 Feb 1996

Dean — **Date**

A Search for the Top Quark Using Artificial Neural Networks

by

Erfan Amidi

ABSTRACT OF DISSERTATION

The neural networks method has been applied to 75 pb^{-1} of data collected by the DØ detector at Fermilab during the 1993-1995 $p\bar{p}$ collider run at $\sqrt{s} = 1.8$ TeV, to isolate the top quark in the e+jets+$\not\!\!E_T$ channel. It is shown that the neural networks method provides a higher signal-to-background result compared to conventional methods. The number of observed top events and the cross section are presented.

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate School of Arts and Sciences of
Northeastern University, February 1996

Acknowledgements

I would like to thank Professor Steve Reucroft for his continuous support and encouragement throughout my student years at Northeastern. It is hard to imagine an advisor more supportive than him. I consider myself very lucky for having a chance to be his student. I would also like to thank the rest of our high energy group at DØ , Takahiro Yasuda, Judd Wilcox, Regina Demina, Hossein Johari, Mike Glaubman, Ssumin Chang and newly joined Darien Wood for their support and kindness. It is a great pleasure and fun to work with the high energy group at Northeastern. I thank Professors John Swain and Alain Karma in my thesis committee for their time and valuable comments.

I sincerely thank Meenakshi Narain and Sailesh Chopra for their support, help, and detailed discussions about e+jets analysis. Their contributions to Top analyses in DØ is immeasurable. I would also like to thank the conveners of the Top group at DØ, Boaz Klima and Nick Hadley for their wise leadership at tough times.

Above all, I would like to thank Gene Alvarez, my coach at the Gym, my captain in our soccer team Pererna, and a friend with a great soul and honesty. I also thank him for reading my thesis and for his valuable comments. Special thanks to my teammates in Pererna who shared their fun time with me. I would also like to thank Paul Russo who has always been willing to go out of his way to help everyone with his computer related problem.

Special thank goes to all of my teachers who taught me Physics, Math and most importantly, about life. Saving the best for last, I would like to thank my first and

best teachers, my parents, to whom I dedicate this thesis.

Erfan Amidi

February 1996

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The present understanding of particle physics provides a beautiful and simple picture of the fundamental constituents of matter and the interactions among them. All matter appears to be composed of quarks and leptons which are structureless, pointlike, spin-$\frac{1}{2}$ particles. Leaving aside gravitation as a negligible perturbation at the energy scales usually considered, the interactions among these particles are of three types: weak, electromagnetic, and strong. All three of these interactions are mediated by spin-1 gauge bosons, as described by gauge theories. The interaction Lagrangians are determined assuming that the states of the particles are invariant under local phase transformations which are a consequence of local charge conservation.

The weak and electromagnetic interactions are successfully unified in the electroweak theory, based on the gauge group $SU(2)_L \otimes U(1)_Y$. In this theory, local gauge invariance is spontaneously broken by the Higgs mechanism. This causes the intermediate bosons (W^+ , W^-, and Z) of the weak interaction to acquire mass, while leaving the photon massless. The theory of the strong interaction is known as quan-

tum chromodynamics (QCD) and is based on the exact local gauge symmetry $SU(3)_c$, which results in conservation of color (the strong charge). A distinctive property of non-Abelian gauge theory is the tendency of the coupling strength to decrease at short distances; this is known as asymptotic freedom. This behavior explains why quarks behave as free particles within hadrons, but cannot be liberated.

The phenomenology of the interactions of the elementary particles is incorporated in the so-called "Standard Model" of QCD plus the $SU(2)_L \otimes U(1)_Y$ electroweak theory. The Standard Model suggests grouping the six known leptons into three families:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}.$$

The quarks are similarly grouped in doublets:

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

where the primes denote Cabbibo-Kobayashi-Maskawa mixing among the charge $-\frac{1}{3}$ quarks. Symmetry considerations and b-quark decay features motivated the search for the third quark with charge $+\frac{2}{3}$, the top quark or simply t-quark. The pursuit of this search finally resulted in the simultaneous announcement of the discovery of the t-quark in 1995 by both CDF and DØ collaborations at Fermilab. At the Tevatron, the t-quark is mostly produced in association with a tbar-quark. The quark-antiquark annihilation contribution to $t\bar{t}$ production is highly dominant as compared to the gluon fusion contribution. The DØ experiment has good electron identification and

jet energy measurement. Because of this and the high branching ratio and relatively high signal-to-noise ratio, $t\bar{t} \rightarrow e + \nu_e +$ jets yields an attractive signature channel to search for the t-quark at DØ. The small ttbar production cross-section makes the t-events rare.

The resemblance of background noise to top signal persuades analysts to use several features of top events to separate signal from background. This brings about the need to use classification methods which are capable of constructing a decision boundary between signal and background in a multi-dimensional configuration space. Several advanced classification methods of this sort have emerged in recent years. These techniques are of two types: parametric and non-parametric. In parametric techniques one assumes a particular parametrized functional form for the discriminator function which determines the boundary between signal and background. In non-parametric techniques, no preassumption is made about the functional form of the discriminator. However, non-parametric techniques require higher statistics to approximate the discriminator function as compared to parametric techniques.

One of the components of the background to the top signal is that due to the misidentification of a jet as an electron. The samples which are used to study the characteristics of this background are derived from data. This limits the statistics of the samples employed to construct the discriminator function. Due to these limited statistics, the non-parametric techniques fail to adequately approximate the discriminator function in the region with low signal efficiency. Since this region is of great interest due to its high signal-to-background ratio, parametric techniques are prefered for discriminator function approximation. It is crucial to use a parametric technique

which promises the most general parametrized functional form for the discriminator function. It has been proven mathematically (see Chapter 5) that the neural network method provides the most general functional form for function approximation. This makes the neural network technique a powerful enough method that it is frequently referred to as "The Universal Approximator". It is worth mentioning that neural networks approximate functions by using a close-to-minimal number of parameters. This is important when sample statistics are limited and a decision boundary is sought based on general features of the signal and background, rather than on idiosyncrasies in the samples.

This thesis is based on a search for the top quark in the $t\bar{t} \rightarrow e + \nu_e +$ jets channel using neural network techniques. We use data collected with the DØ experiment at Fermi National Accelerator Laboratory (Fermilab). The thesis is arranged as follows. In Chapter 2 we have a short overview of the Standard Model, including a discussion of the fundamental building blocks of matter, and the jets which play a major role in the top search. We also see how the existence of the top quark is essential for the integrity and simplicity of the Standard Model. In Chapter 3, those components of the experimental apparatus at DØ which are crucial to the t-quark search are discussed. Chapter 4 explains how physical particles and jets are identified and their parameters derived from the digitized data provided by the experimental apparatus. Chapter 5 covers the basic concepts and theorems of neural networks which are important to this thesis. In Chapter 6 we estimate the amounts of signal and background in the data. In Chapter 7, we study features of signal and background events which form the basis of our classification of the data events discussed in Chapter 8. In conclusion,

the neural network results have been compared with DØ published results [44] in which conventional cuts are applied for event classification. This comparison strongly indicates that the neural networks method is far superior to conventional method of event classification. Natural units are used throughout the thesis *i.e.* $\hbar = c = 1$ and energies and momenta are in GeV unless stated otherwise. The unit of electric charge is $|e|$ where e is the charge of the electron.

Chapter 2

Theoretical Overview

2.1 The Standard Model

The goal of particle physics is to identify the structureless units of matter, and understand the nature of the forces acting between them. The matter/force nature of this challenge is beautifully illustrated by Maxwell's theory of the electromagnetic field. The electron is considered both as a source of the electromagnetic field and also as a constituent of matter. During the past hundred years, the deeper study of matter has led to the discovery of two new forces, the weak and strong forces. The *Standard Model (SM)* generalizes the beautiful electron/electromagnetic field relationship to the weak and strong forces.

According to the Standard Model there are two constituents of matter, *leptons* and *quarks*. Both are structureless at the smallest distances currently attainable by the highest-energy accelerators and both have spin $\frac{1}{2}$. Charged leptons participate in both electromagnetic and weak interactions, whereas neutral leptons only interact weakly. Quarks interact via all three interactions, strong, electromagnetic and weak.

Force	gauge boson	symbol	charge(e)	spin($\hbar$)	mass(GeV)
Strong	gluon	g	0	1	0
Electromagnetic	photon	γ	0	1	0
Weak	W boson	$W^{\pm}$	± 1	1	80
	Z boson	Z^0	0	1	91

Table 2.1: Fundamental forces and gauge bosons.

The weak and electromagnetic interactions of both quarks and leptons are described by the *Electroweak* theory, which is a generalization of *Quantum Electrodynamics (QED)*. The strong interaction of quarks is described by *Quantum Chromodynamics (QCD)*.

The forces between matter units (leptons and quarks) are transmitted by specific fields or particles (equivalent concepts in relativistic quantum theory). The electromagnetic force is transmitted by photons, the strong force by massless spin-1 gluons and the weak force by massive W and Z bosons. These forces are all described by *gauge theories* and the transmitting particles are called *gauge bosons*. The Quantum treatment of gravity is still an open question and even though the General Theory of Relativity is a gauge theory, gravitation is not included in the Standard Model. Table 2.1 summarizes the characteristics of the fundamental forces and their corresponding gauge bosons [1].

The muon and the tau are leptons which, apart from mass, are identical to the electron. All three leptons have an associated light neutrino. There are six flavors

	name	symbol	charge	mass(GeV)	type
first generation	up	u	$\frac{2}{3}$	$2\text{-}8 \times 10^{-3}$	quark
	down	d	$-\frac{1}{3}$	$5 - 15 \times 10^{-3}$	quark
	e-neutrino	ν_e	0	$< 5.1 \times 10^{-9}$	lepton
	electron	e	-1	5.1×10^{-4}	lepton
second generation	charm	c	$\frac{2}{3}$	1.0-1.6	quark
	strange	s	$-\frac{1}{3}$	0.1-0.3	quark
	μ-neutrino	ν_μ	0	$< 2.7 \times 10^{-4}$	lepton
	muon	μ	-1	0.106	lepton
third generation	top(truth)	t	$\frac{2}{3}$	$175 - 200$	quark
	bottom(beauty)	b	$-\frac{1}{3}$	4.1-4.5	quark
	τ-neutrino	ν_τ	0	$< 3.1 \times 10^{-2}$	lepton
	tau	τ	-1	1.78	lepton

Table 2.2: Three generations of fundamental spin-$\frac{1}{2}$ particles.

of quarks. The up, charm and top flavors of quarks have electric charge $\frac{2}{3}$, while the down, strange and bottom quarks have charge $-\frac{1}{3}$. These six quarks and six leptons can be grouped into three families (generations). As shown in Table 2.2 apart from mass all generations are identical [1].

2.2 Color and Color Confinement

The quark model is very successful, but without assigning a new degree of freedom to quarks this model would contradict the Pauli principle. *The Pauli principle states that the wavefunction of any quantum state must be antisymmetric under the interchange of any two identical spin $\frac{1}{2}$ fermions.* The spectrum of light baryons

requires that the combined space and spin wavefunction be symmetric under the interchange of any two quarks with the same flavor, which is in conflict with the Pauli principle. This contradiction between quark model and Pauli principle can be resolved by assuming an additional degree of freedom for quarks. The wavefunction for this new attribute, which is called color, is antisymmetric. So the combined space, spin and color wavefunction is antisymmetric, which is consistent with the Pauli principle and the space and spin wavefunction remains symmetric, which explains the spectrum of light baryons.

According to the color idea any quark can exist in three different colors "red", "green" and "blue". The cross section for producing a quark pair of any flavor is proportional to the number of colors, so the number of colors can be tested against experiment. The experimental data are in excellent agreement with the existence of three colors. Color is an additive quantum number like electric charge, whose values are opposite in sign for particles and antiparticles.

The quark model alone does not rule out the existence of hadrons with fractional electric charges. Hadrons can be composed of any combination of quarks and antiquarks such as qq, $qq\bar{q}$, $qqqq$. Hadrons with fractional electric charges, however, can be excluded by imposing color confinement on hadron substructure. *The color confinement hypothesis states that isolated hadrons as free particles must have no color charge.* The direct consequence of this hypothesis is that the only allowed combinations of quarks and antiquarks are :

$$(3q)^i(q\bar{q})^j \qquad i,j \geq 0 \tag{2.1}$$

such as qqq, $q\bar{q}$, $qqq\bar{q}q$, ... Since the electric charge of the quarks is a multiple of $\frac{1}{3}$, free hadrons cannot have fractional electric charge. Color confinement is analogous to allowing only neutral atoms to exist yet forbidding ionization. The interaction between quarks is weaker at short distances *(asymptotic freedom)* so quarks can be treated as free particles and perturbative methods apply. On the other hand, at large distances the interaction strength grows, making the ionization energy infinite *(infrared slavery)*. Since isolated quarks and gluons cannot exist, injecting energy into a hadron creates quark-antiquark pairs, which materialize as new hadrons. "Breaking" hadrons to produce new hadrons (not free quarks) is very similar to breaking a bar magnet to give smaller bar magnets (not single S or N poles).

2.3 The KLN Theorem and Jets

At high energy (short distances) perturbative QCD is applicable. The presence of massless particles (gluons) can cause "mass singularities" in the series expansion of perturbative QCD. This is closely related to the high degree of degeneracy in the system i.e. the same total momentum implies the same total energy for parallel moving massless particles. Such infinities in the QCD series can be removed if we average over an appropriate ensemble of degenerate states. This is described in the Kinoshita, Lee, Nauenberg theorem [2]. We call such an ensemble of parallel moving particles a jet. *Each jet is characterized by a total momentum and a narrow cone of*

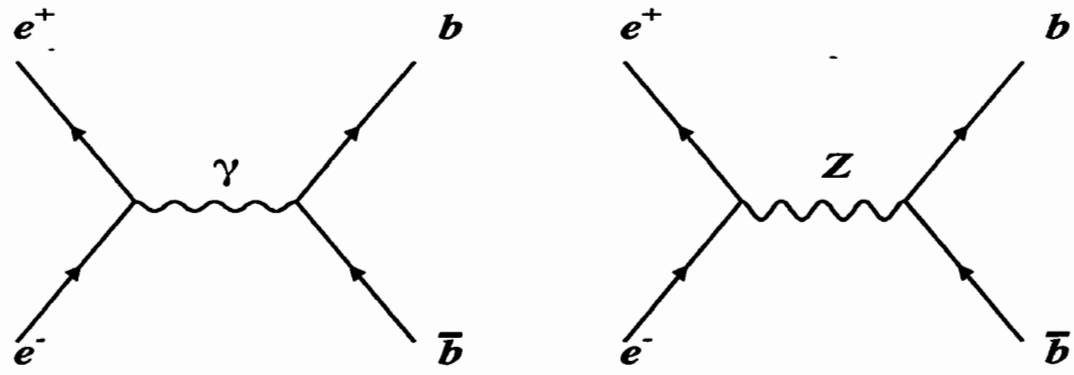


Figure 2.1: Feynman diagrams for $e^+e^- \rightarrow b\bar{b}$

fixed size which can be arbitrary small. Since gluons are massless, and light quarks approximately so, the total energy of a jet is the magnitude of its total momentum within a sufficiently small cone.

2.4 Why The Top Quark Should Exist

Besides the direct observation of the top quark by the CDF and DØ collaborations at FNAL [43, 44], there is strong indirect evidence to believe in the existence of the top quark. The Standard Model requires the b quark to be a member of an $SU(2)_L$ doublet, i.e. it should have a partner. There are experiments which support this prediction of the Standard Model.

Consider the reaction $e^+e^- \rightarrow b\bar{b}$. This process can be mediated by either a photon or a Z boson (fig 2.1). The contribution of photon exchange to the angular distribution of the b about the plane perpendicular to the beam in the center of mass frame is symmetric; the contribution of the Z should be asymmetric, if the b is in

an $SU(2)_L$ doublet. It can be shown that the forward-backward asymmetry for this process is :

$$A_{FB} \sim (T_{3L} - T_{3R})(T_{3L} + T_{3R} + \frac{2}{3} cos^2\theta_W) \tag{2.2}$$

where $A_{FB} \equiv \frac{\sigma_F - \sigma_B}{\sigma_F + \sigma_B}$ and **T** is the generator of SU(2) symmetry [3].

Now, consider the Z decay to $b\bar{b}$ (fig 2.2). The Feynman rule for Z decay vertex gives $\frac{-ie}{sin\theta_W cos\theta_W}\gamma_\mu(c_L\frac{1-\gamma_5}{2} + c_R\frac{1+\gamma_5}{2})$, where $c_{L,R} = T_{3L,R} + \frac{1}{3}sin^2\theta_W$

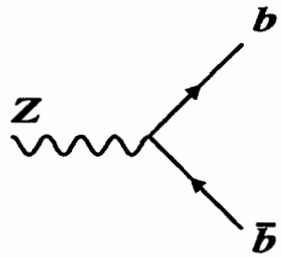


Figure 2.2: Feynman diagram for $Z \rightarrow b\bar{b}$

This results in a width with left and right-handed contributions [3]:

$$\Gamma(Z \rightarrow b\bar{b}) \sim (T_{3L} + \frac{1}{3}sin^2\theta_W)^2 + (T_{3R} + \frac{1}{3}sin^2\theta_W)^2 \tag{2.3}$$

The combined measured values for A_{FB} and $\Gamma(Z \rightarrow b\bar{b})$ yields [4]:

$T_{3L} = -0.504 \pm^{0.018}_{0.011}$ $\quad T_{3R} = -0.008 \pm^{0.056}_{0.056}$

Theory predicts $\quad T_{3L} = -0.5,\ T_{3R} = 0$

The results strongly suggest that the b quark is a member of an $SU(2)_L$ doublet ($2T_{3L} + 1 = 2 \times 0.5 + 1 = 2$). A theory with no top would predict $T_{3L} = 0$. The other member of the doublet is the top quark, by definition.

There are also some theoretical arguments in favor of the top quark, such as SM anomalies which can be eliminated by assuming the existence of the t quark and $\bar{B}^0$-B^0 mixing which needs a t loop diagram to generate the observed values. Even though the existence of the t quark is sufficient to explain these phenomena, it is not a necessary condition. However, top does provide the simplest way to explain the observed $\bar{B}^0$-B^0 mixing and rectify SM triangular anomalies.

2.5 Top Production and Decay

The lowest order Feynman diagrams for top production due to proton-antiproton collision are shown in fig 2.3 and fig 2.4. Figures 2.3 correspond to single top production[1] [5, 6], whereas Figure 2.4 corresponds to pair production. As shown in Figure 2.5, pair production is the dominant process [5]. Among the lowest order processes for QCD production of a $t\bar{t}$ pair, quark-antiquark annihilation dominates for heavy top (fig 2.6). According to the Standard Model, the possible t decay modes with $m_t > m_W + m_b$ are :

$$t \rightarrow W + b$$

$$t \rightarrow W + s$$

$$t \rightarrow W + d$$

[1]The contribution of processes such as $gb \rightarrow tW$ are negligible at $\sqrt{s} = 1.8\ TeV$.

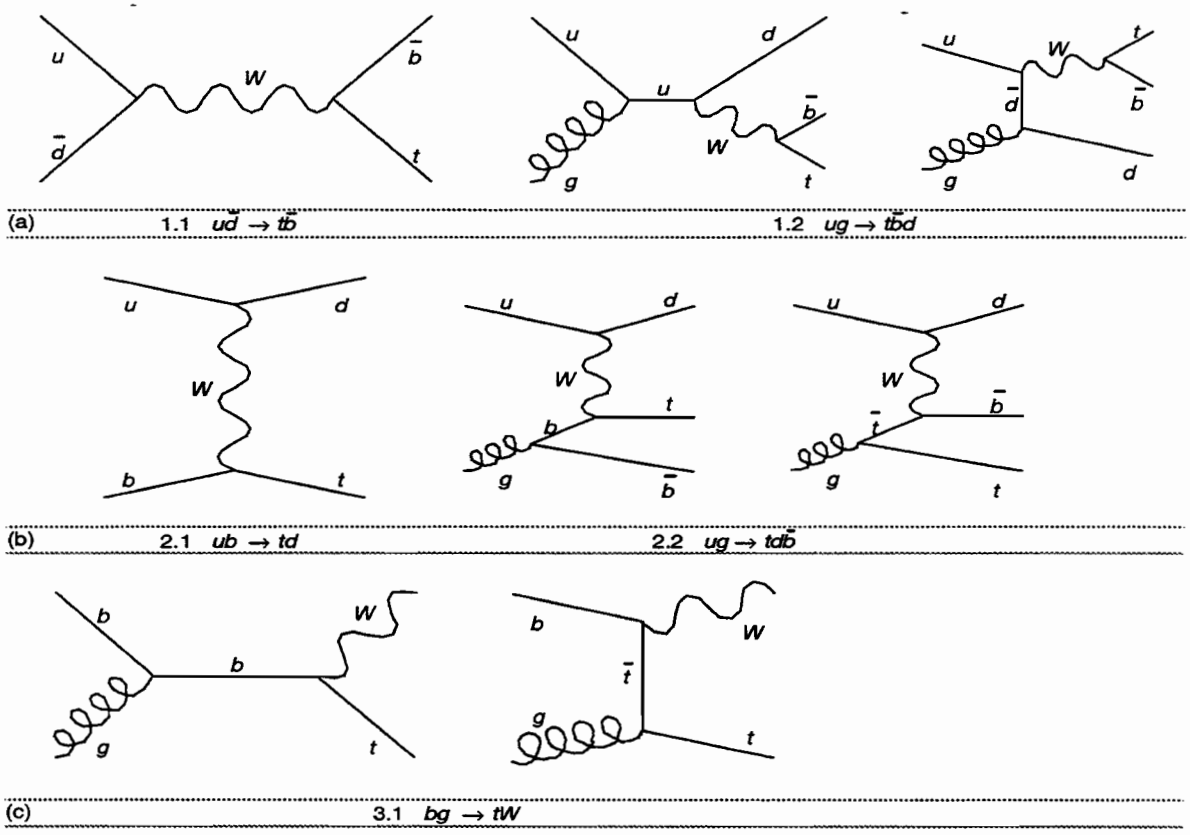


Figure 2.3: Feynman diagrams for single top production processes.

The branching ratio for $t \to W + q$ is proportional to $|V_{tq}|^2$, where V is the Cabibbo-Kobayashi-Maskawa mixing matrix. Since with 90% $C.L.$

$$V_{td} \sim 0.004 - 0.015$$

$$V_{ts} \sim 0.030 - 0.048$$

$$V_{tb} \sim 0.9988 - 0.9995$$

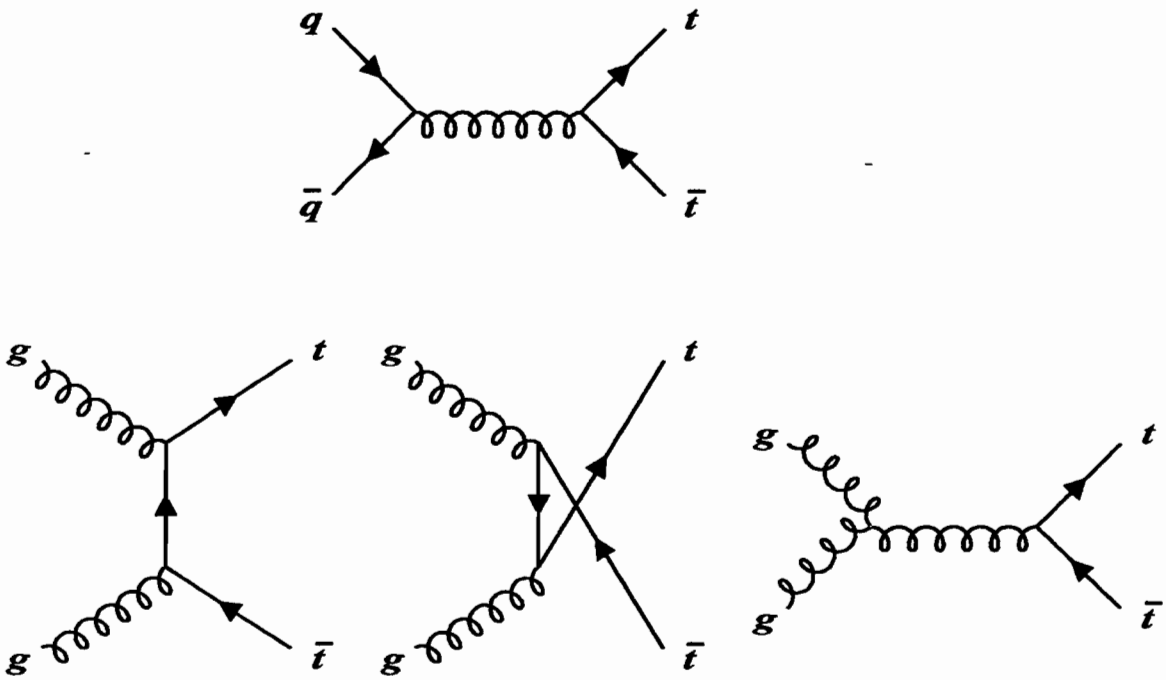


Figure 2.4: Feynman diagrams for top pair production processes

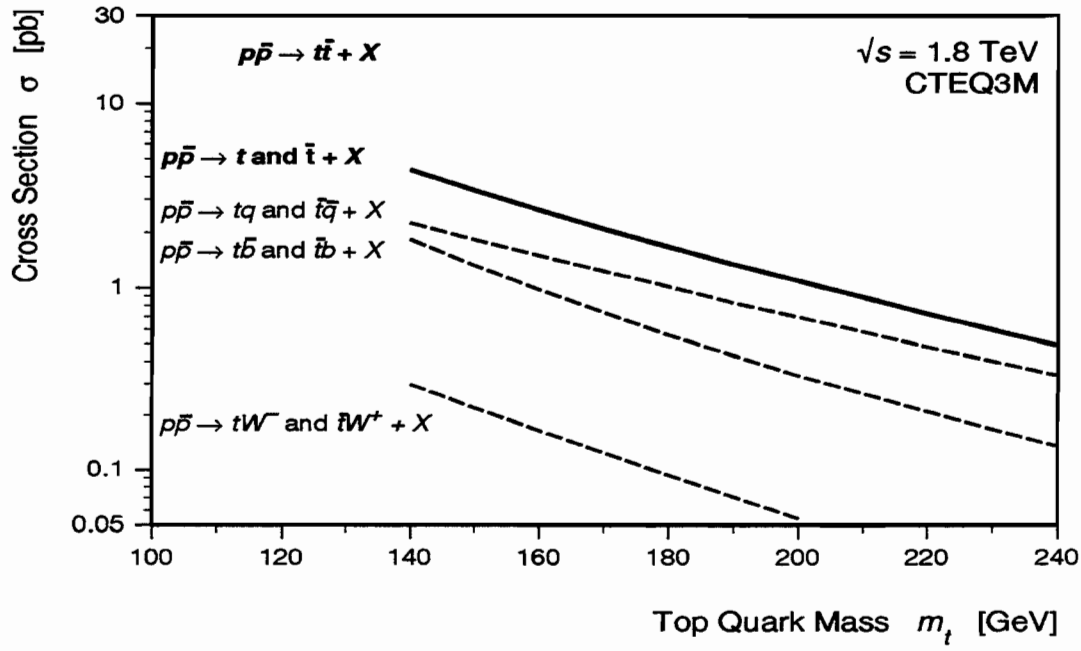


Figure 2.5: Top cross section for pair and single top production processes.

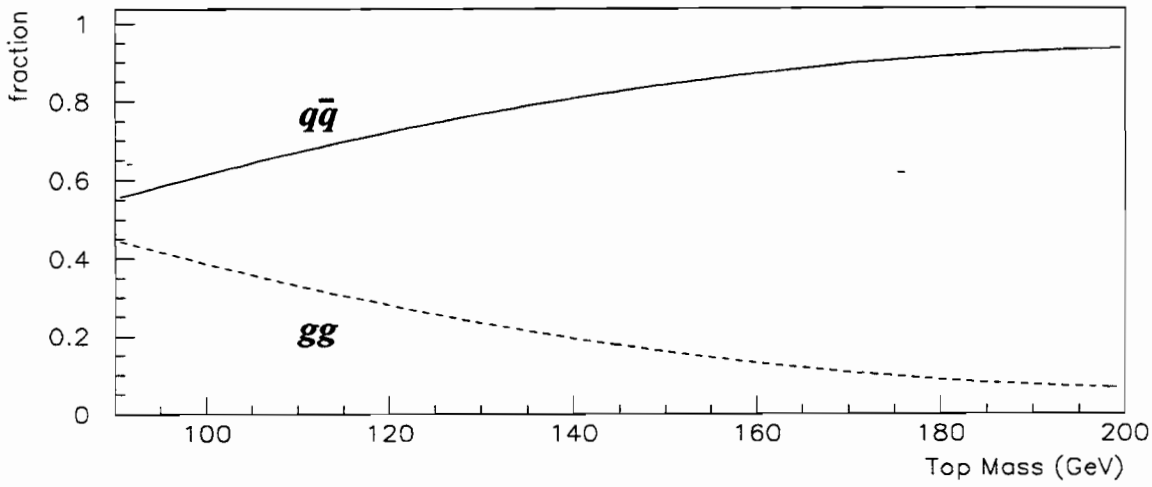


Figure 2.6: Contributions from gluon pair fusion and quark-antiquark annihilation in next to leading order cross section calculations for $t\bar{t}$ production [7].

$t \rightarrow W + b$ is highly dominant [1]. The b quark forms a jet and the W decays to either a quark pair(two jets) or a lepton-neutrino pair. Thus the total number of channels is :

$$\begin{aligned} W^+ \rightarrow \quad & e^+\nu_e \quad 1 \\ \rightarrow \quad & u\bar{d} \quad 3 \\ \rightarrow \quad & \mu^+\nu_\mu \quad 1 \\ \rightarrow \quad & c\bar{s} \quad 3 \\ \rightarrow \quad & \tau^+\nu_\tau \quad 1 \\ & Total: \quad 9 \end{aligned}$$

$t\bar{b}$ is kinematically excluded and the quark channels contribute three times as much as lepton-neutrino because of the three color degrees of freedom. Therefore, the branching ratio for decay into jets is $\frac{6}{9}$ and into each lepton-neutrino it is $\frac{1}{9}$. Hence, the branching ratio for each channel is :

$$t\bar{t} \rightarrow \quad \text{two identical leptons} \; + \; \text{two neutrinos} \; + \text{jets} \quad 1/81 = 1.23\%$$

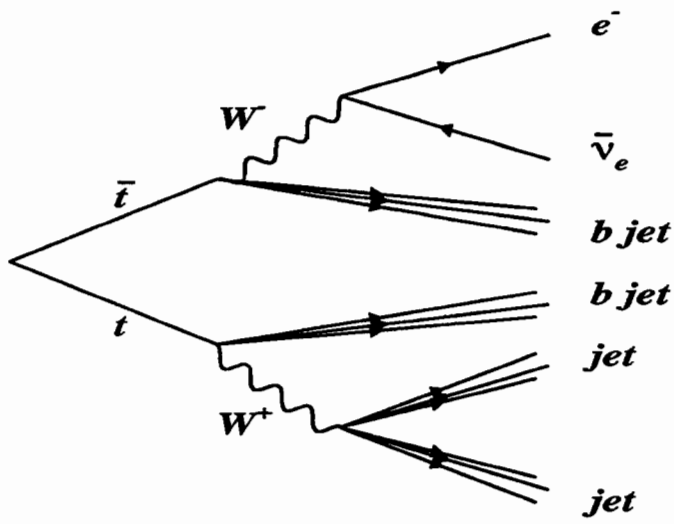


Figure 2.7: $t\bar{t} \rightarrow e + \bar{\nu_e} + jets$

$t\bar{t} \rightarrow$	two different leptons + two neutrinos + jets	$2/81 = 2.47\%$
$t\bar{t} \rightarrow$	lepton + neutrino + jets	$12/81 = 14.8\%$
$t\bar{t} \rightarrow$	jets	$36/81 = 44.4\%$

Since there are 3 leptons,

$$3 \times 1.23\% + 3 \times 2.47\% + 3 \times 14.8\% + 44.4\% = 100\% \ .$$

Even though $t\bar{t} \rightarrow jets$ has the highest branching ratio, it suffers from a huge QCD multijet background. $t\bar{t} \rightarrow lepton + jets + neutrino$ has the second highest branching ratio. In the rest of this thesis, I will be discussing characteristics of the $electron + jets + neutrino$ channel (fig 2.7) and the methods to isolate the signal from background for this channel. Backgrounds to this process are a single W along with several jets and QCD multijet events where a jet is misidentified as an electron.

Chapter 3

Experimental Apparatus

The data analyzed in this thesis was collected using the DØ detector, at the Tevatron $p\bar{p}$ collider located at Fermi National Accelerator Laboratory.

3.1 The Tevatron

The Tevatron can reach a center of mass energy of 1.8 TeV, currently the highest energy attainable by particle colliders in the world. The long Tevatron circumference of 3.7 miles reduces the energy loss due to radiation. As shown in Figure 3.1, the Tevatron consists of the following parts:

- A Cockroft-Walton accelerator, which accelerates H^- ions obtained by ionizing hydrogen gas.

- The Linac, which generates an oscillating electric field to accelerate the ions. The Linac is 150 meters long and employs an oscillating electric field synchronized to the passage of the ions through electrodes. A carbon foil is then utilized to strip the electrons from the ions to obtain proton.

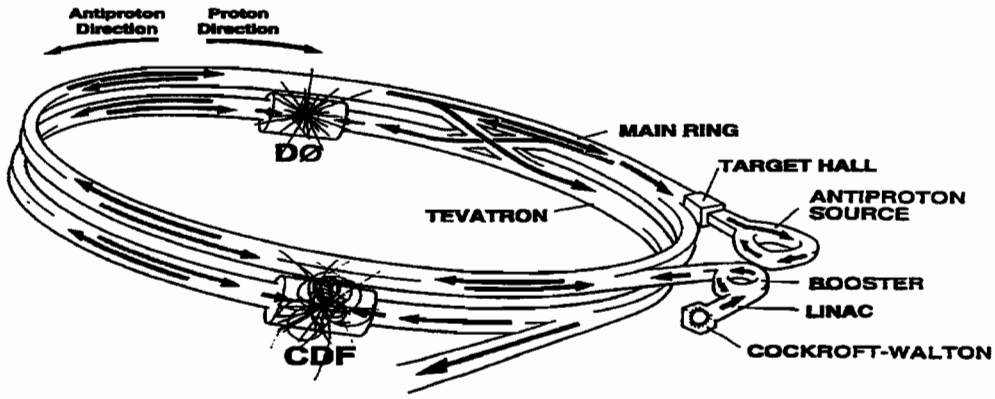


Figure 3.1: Tevatron collider

- The Booster synchrotron ring, which confines the protons to a closed orbit using bending magnets. The energy of the particles is increased by acceleration in a synchronized RF cavity. The energy of the protons is increased to 8 GeV in this stage. As the energy of the protons increases in the synchrotron ring the magnetic field is increased accordingly to keep the protons in the ring. The circumference of the Booster synchrotron ring is 500 meters.

- The Main Ring which is employed to further accelerate protons up to 120 GeV. The protons circulate around the 3.7 mile long Main Ring in bunches containing 2×10^{12} protons each. Six bunches circulate around the Main Ring simultaneously. The bunch crossing time, τ, for any point on the ring is

$$\tau \sim \frac{\text{Circumference}}{cN_{bunch}} = 3.5\ \mu\text{s} \tag{3.1}$$

- The Target Hall in which 2×10^7 antiprotons are produced by extracting proton bunches onto a nickle/copper target. Then a magnetic lens is used to focus and inject the antiprotons into the debuncher in which a coherent beam is formed. This process is known as the "cooling" process.

This process continues until about 4×10^{11} antiprotons are stored. The antiprotons are then injected into the Main Ring where they are accelerated to 150 GeV and then transferred to the Tevatron ring. In the Tevatron, antiprotons circulate in the opposite direction to the protons. Both protons and antiprotons are accelerated to acquire an energy of 900 GeV before they collide at DØ. The instantaneous luminosity is given by

$$L_{ins} = \frac{N_p N_{\bar{p}}}{\tau S} \rightarrow L_{ins} \sim 5 \times 10^{30} \text{ sec}^{-1}\text{cm}^{-2} = 0.5 \text{ nb}^{-1}/\text{sec}, \tag{3.2}$$

where N_p and $N_{\bar{p}}$ are number of the protons and antiprotons per bunch, respectively and S is the geometrical area of the interaction. The integrated luminosity $\mathcal{L}$ is defined as

$$\mathcal{L} = \int L_{ins} dt. \tag{3.3}$$

The cross section, σ, for a process is given by

$$\sigma = \frac{N}{\mathcal{L}}, \tag{3.4}$$

where N is the number of events produced by the process.

3.2 The DØ Detector

The DØ detector is a multipurpose detector specially designed to provide good electron and muon resolution, superior electromagnetic and hadronic energy resolution through highly segmented calorimetry , and full solid angle coverage. The major physics goals of the DØ collaboration include the study of electroweak physics and the perturbative QCD and the search for high mass states, high P_T phenomena and

new physics beyond the Standard Model.

The basic components of the DØ detector are :

- The Central Detector;
- The Calorimeter;
- The Muon Detector.

A cutaway isometric view of the detector is shown in fig 3.2.

The central detector is designed to trace the trajectory of charged particles (tracking). This tracking system is designed to be the closest to the point of interaction to thereby minimize multiple scattering and identify secondary vertices. The central detector is surrounded by the calorimeter. The calorimeter is designed to measure the energy of particles and should be thick enough to stop particles[1] and measure the deposited energy. A tracking system, on the contrary, should contain as little material as possible to minimize the probability of inelastic interactions before particles reach the calorimeter. Since the energy deposited by a lepton in matter is inversely proportional to the mass of the particle, muons escape the calorimeter with little energy deposition. Muons are detected by a 3-layer proportional drift chamber surrounding the calorimeter. Neutrinos interact only weakly and cannot be stopped by the calorimeter. Neutrinos are identified by balancing the energy flow transverse to the beam.

[1] Except ν's and μ's which escape the calorimeter.

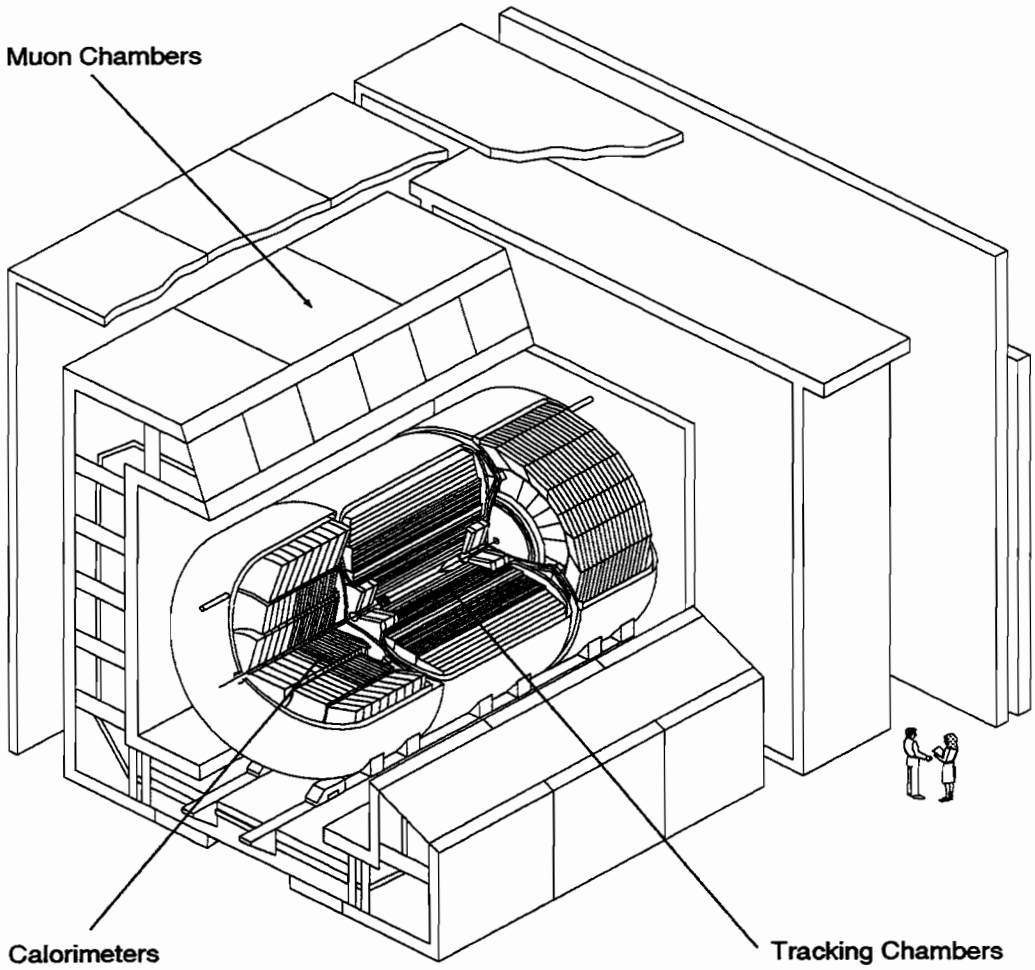


Figure 3.2: DØ detector

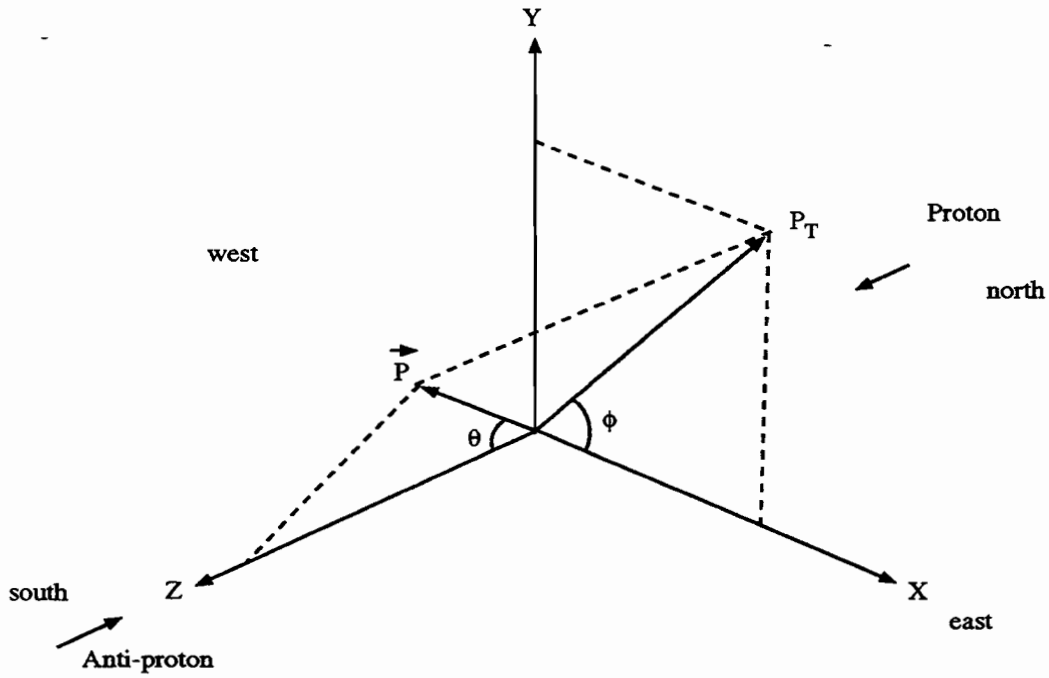


Figure 3.3: DØ detector coordinate system

The DØ detector weighs 5500 tons and the absence of a central magnetic field makes it possible for the detector to fit in a compact volume of 13m high $\times$11m wide $\times$17m long. The DØ detector is not designed for tracking and identification of individual particles within jets. A better jet energy measurement is achieved by the calorimeter because of the absence of a central magnetic field. A magnetic field would deflect the charged particle out of the jets. In order to be consistent throughout this thesis we will adopt a right-handed coordinate system with the positive z-axis along the beam and in the direction of the protons and y-axis pointing up (fig 3.3). In $p\bar{p}$ collisions, the total secondary momentum along the beam direction is not known since

secondary particles may escape down the beam pipe. The transverse momentum[2], $\vec{p_T}$, for these particles is small, so momentum conservation can be applied in the transverse plane. p_T is defined to be:

$$p_T \equiv |\vec{p_T}| = psin\theta \tag{3.5}$$

The *transverse energy* is defined as a vector whose direction is the direction of $\vec{p_T}$ in the transverse plane and it's magnitude is:

$$E_T = Esin\theta \tag{3.6}$$

Pseudorapidity η is frequently used instead of θ.The pseudorapidity is defined as:

$$\eta = -ln(tan\frac{\theta}{2}), \tag{3.7}$$

which is an approximation of the rapidity defined as:

$$y = \frac{1}{2}ln(\frac{E+p_z}{E-p_z}), \tag{3.8}$$

in the limit that $\frac{m}{E} \ll 1$, where $m = \sqrt{E^2 - p^2}$. Figure 3.4 shows the η values corresponding to different θ's in one quarter of the calorimeter and the tracking system.

3.3 The Central Detector (CD)

The major role of the central detector is to reconstruct the 3-dimensional trajectory of each charged particle passing through. Tracking in the central detector is important because :

[2]The momentum vector projected on the x-y plane

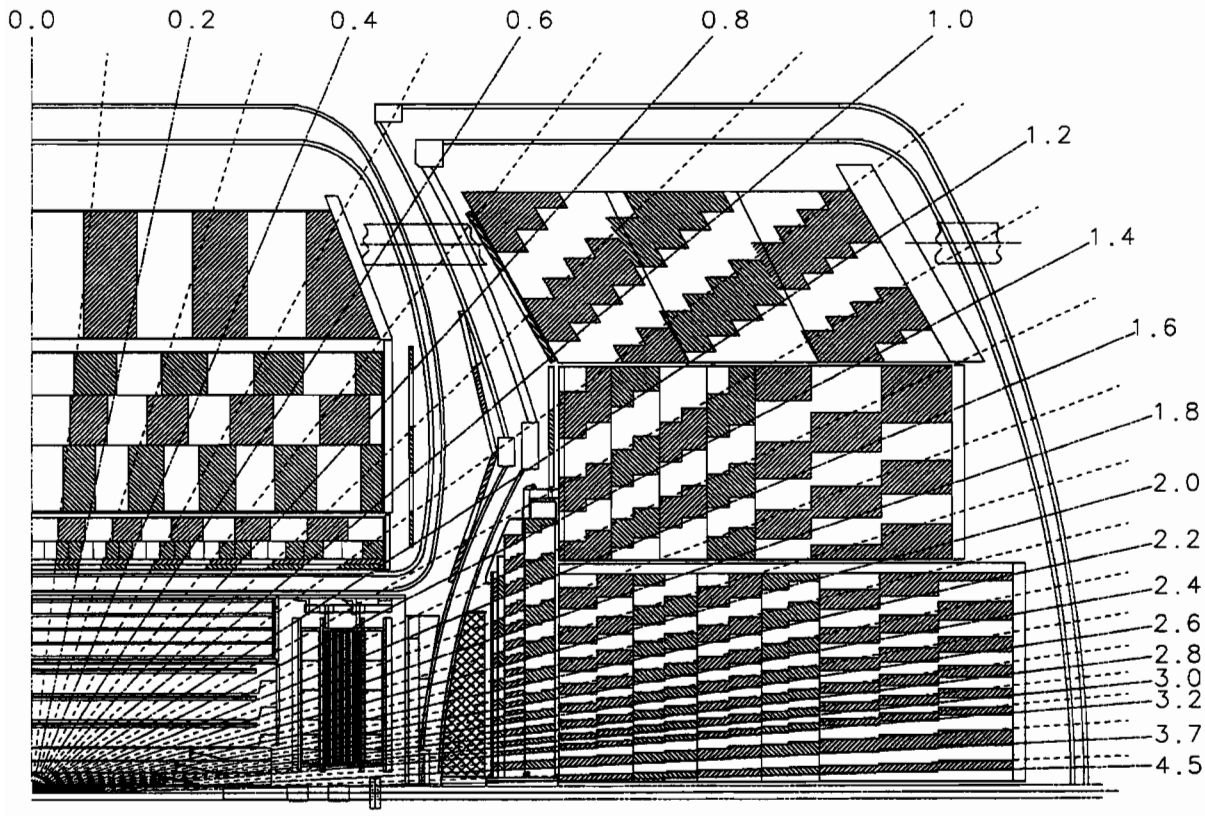


Figure 3.4: Side view of one quarter of the calorimeter and the central detector. The numbers show pseudorapidity (η) values.

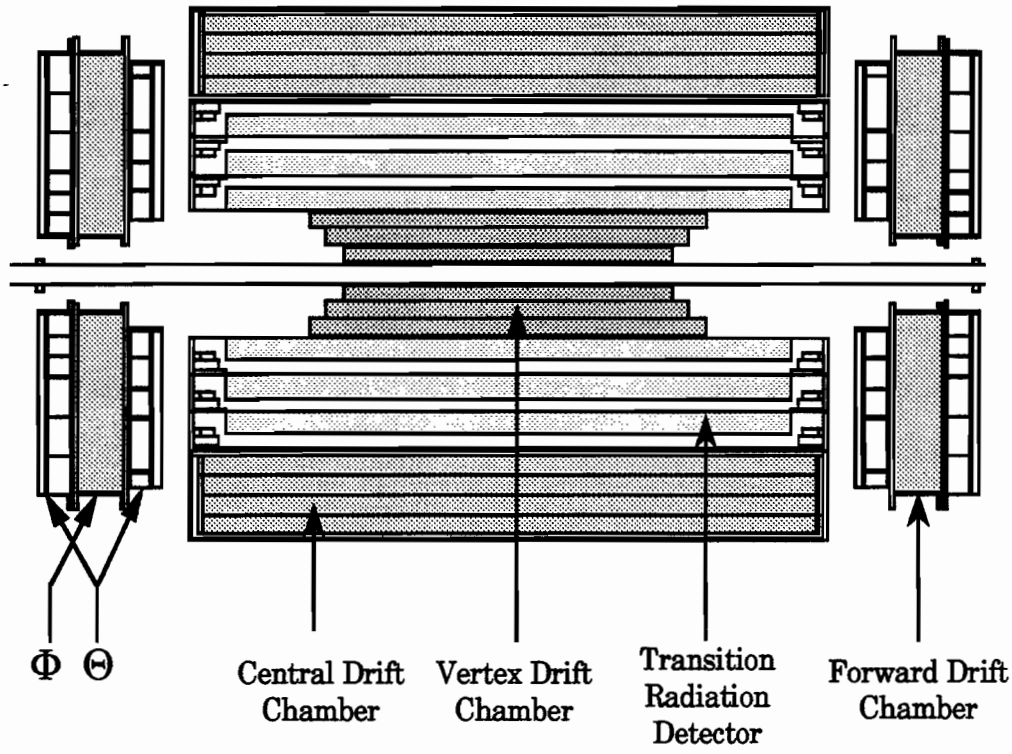


Figure 3.5: Side view of the central detector

- Using the tracking information, one can determine whether an electromagnetic shower in the calorimeter is produced by an electron, a photon or π^0;

- The precise measurement of the location of the interaction vertex is done using the CD tracking information. The precise vertex measurements can be used for the calorimeter position measurements;

- By measuring dE/dx for a track, one can decide if a track is caused by photon conversion, $\gamma \rightarrow e^+e^-$.

The central detector consists of four subdetectors (fig 3.5). These four subdetectors, ordered from inside to outside, are:

- The vertex drift chamber (VTX) surrounding the beryllium beam pipe;
- The transition radiation detector (TRD) surrounding the VTX;
- The central drift chamber (CDC) surrounding the TRD;
- The forward drift chambers (FDC) at each end of the central detector.

These four subdetectors are contained in cylindrical volume of radius $r = 75$ cm and length $l = 270$ cm.

3.3.1 Operation Principles of The Drift Chambers

The fact that three out of the four subdetectors in CD are drift chamber motivates us to have a brief overview of these devices. For a detailed discussion of drift chambers, references [8, 10] are recommended. The working principle of the drift chambers is based on the fact that energetic charged particles cause ionization along their path as they pass through a gas. When a charged particle passes through a gas, it will interact with nearby atomic electrons, creating electron/ion pairs along its path. The number of electron/ion pairs created depends on the energy of the particle and the type of gas. An electric field is used to collect the liberated electrons and cause them to drift through the gas towards the positive electrode (sense wire). The drifting electron causes further ionization along the way to the positive electrode. As the accelerated electron gets closer to the anode it experiences a stronger electric field causing electron to accelerate faster and gain enough energy to cause further ionization. This phenomenon, in which the number of the electrons increases exponentially, is called the *avalanche* effect. This effect gives rise to a measurable current which is proportional to the original number of ions created. The ratio of the

final number of electrons collected by the anode to the initial number deposited is called the *gas gain*. The gas gain is of the order of $10^4 - 10^6$ for a typical drift chamber.

The velocity of a drift electron is a known quantity determined by the strength of the field and the density, pressure and temperature of the gas. The fact that the velocity of the electron is known along its path to the anode enables us to measure the position of the source particle knowing the drift time. In order to obtain a linear relationship between the electric field and the velocity of the electron, it is necessary to have an electric field which is constant over a large volume. The large electric field needed to drift electrons far away from the anode is generated by a very thin wire (20-100 μm in diameter). Additional electrodes, known as *field-shaping* electrodes are used to make the electric field more uniform.

3.3.2 The Vertex Drift Chamber (VTX)

The vertex chamber is the innermost drift chamber used for vertex position measurement. It consists of four carbon fiber cylinders surrounding 3 concentric layers. The VTX extends from $r = 3.7$cm to $r = 16.2$cm radially. The length of the innermost layer is 97cm and each successive layer is about 10cm longer. Figure 3.6 shows an end view of the VTX chamber. The VTX chamber is a *jet* chamber. In a jet chamber sense wires are strung in planes parallel to the path of the particles from the interaction vertex. The inner layer is divided into 16, while the two outer layers into 32 cells each. Each layer is rotated in ϕ with respect to the adjacent layer to

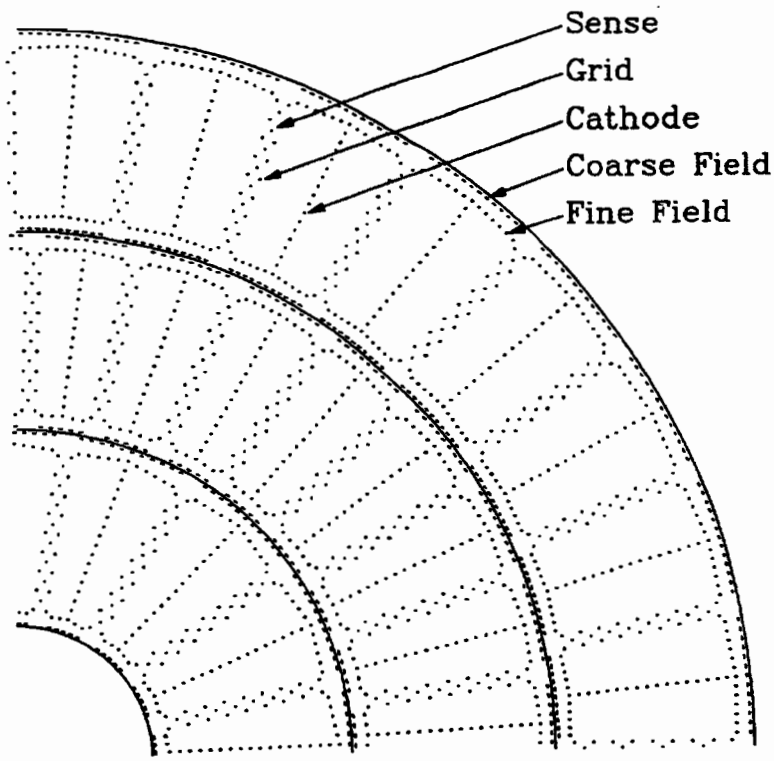


Figure 3.6: End view of the VTX chamber

eliminate dead regions and left-right ambiguities[3]. Each cell contains 8 sense wires, which are staggered out of the r-ϕ plane by 100 μm to lessen left-right ambiguities. The (r, ϕ) position of the track is determined from the drift time. The z position is determined using charge division, in which the sense wire is read out at both ends. The parameters of the VTX chamber are summarized in Table 3.1.

3.3.3 The Transition Radiation Detector (TRD)

The working principle of the transition radiation detector is based on the fact that charged particles radiate photons in the forward direction as they traverse the boundary between two media with different dielectric constants. The radiation inten-

[3]The drift time measurements only yield the distance electrons have drifted, since the drift can be from either left or right, the position of any single hit is ambiguous.

Length of Active Volume: Layer 1	96.6 cm
Layer 2	106.6 cm
Layer 3	116.8 cm
Phi sectors/layer	16, 32, 32
Radial Interval (active)	3.7-16.2 cm
Radial Wire Interval	4.57 mm
Number of Sense Wires/Cell	8
Number of Sense Wires	640
Stagger of Sense Wires	$\pm$ 100 μm
Gas Mixture	CO_2(95%)-Ethane(5%)
Gas Pressure	1 atm
Drift Field	1.0-1.6 kV/cm
Average Drift Velocity	7.3-12.8 μm/ns
Gas Gain at Sense Wires	4×10^4
Sense Wire Potential	+2.5 kV
Sense Wire Diameter	25 μm NiCoTin
Guard Wire Diameter	152 μm Au-plated Al

Table 3.1: VTX chamber parameters

sity is proportional to $\gamma \equiv \frac{1}{\sqrt{1-v^2/c^2}} = E/mc^2$ and concentrated in a cone with a half angle proportional to $1/\gamma$. For highly relativistic particles, the radiation is in the X-ray frequency range. Using these characteristics, a transition radiation detector discriminates particles with different masses which have similar energies. In order to obtain a reasonable signal, the charged particle has to traverse a large number of boundaries. The DØ transition radiation detector is utilized to discriminate electrons from heavier particles. Electrons are the only particles at the Tevatron likely to cause

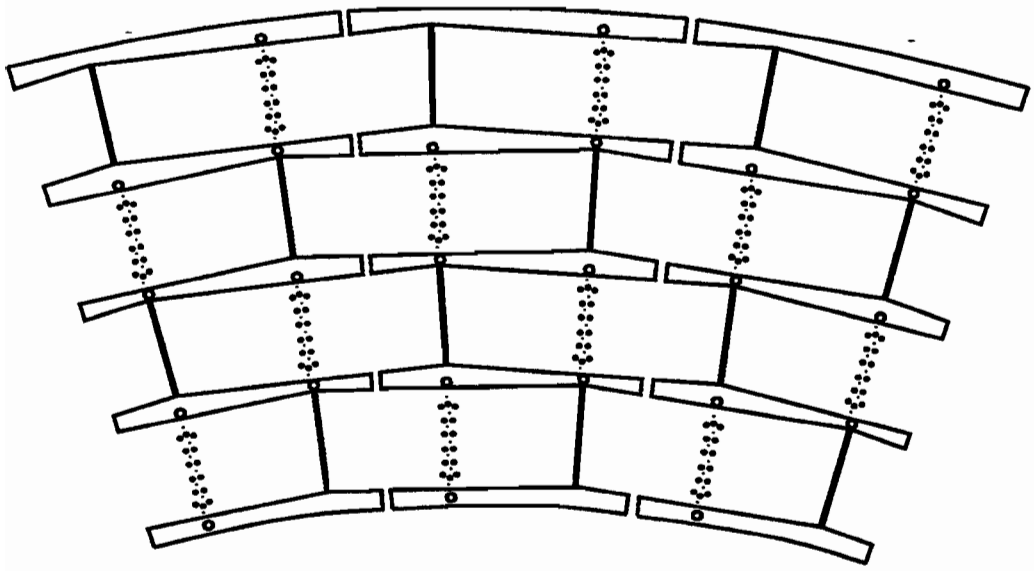


Figure 3.7: An end view of 3 of the 32 CDC modules.

detectable transition radiation[4] . The TRD has 3 layers, each layer containing 393 sublayers of 18 μm polyethylene foil with 150 μm separations. The gaps between the foils are filled with dry nitrogen. Each radiator is surrounded by a Xenon-filled drift chamber to detect the transition X-ray radiation. The TRD provides a factor of 10 rejection against pions with a high efficiency of 90% for isolated electrons.

3.3.4 The Central Drift Chamber (CDC)

The central drift chamber is the outermost subdetector of the CD. It covers the pseudorapidity range of $|\eta| \leq 1.2$. The CDC consists of 4 layers which extend radially from $r = 49.5$ cm to $r = 74.5$ cm and are 184 cm long. Figure 3.7 shows an end view of a portion of the CDC. Each layer of the CDC is divided into 32 identical sectors which are arranged in a cylindrical ring. Within each module, there are 7 sense wires,

[4]The transition radiation occurs when $\gamma > 10^3$.

Length of Active Volume	179.4 cm
Radial Interval (active)	51.8-71.9 cm
Number of Layers	4
Radial Wire Interval	6 mm
Number of Sense Wires/Cell	7
Number of Sense Wires	896
Stagger of Sense Wires	$\pm$ 200 μm
Number of Delay Lines	256
Gas Mixture	Ar(93%)-CH_4(4%)-CO_2(3%)
Gas Pressure	1 atm
Drift Field	620 V/cm
Average Drift Velocity	34 μm/ns
Gas Gain at Sense Wires	$2,6 \times 10^4$
Sense Wire Potential	+1.5 kV
Sense Wire Diameter	30 μm Au-plated W
Guard Wire Diameter	125 μm Au-plated CuBe

Table 3.2: CDC detector parameters

staggered 200 μm relative to each other to resolve left-right ambiguities. The CDC has a jet geometry similar to the vertex chamber. The (r, ϕ) position of a hit is determined using the drift time and the z position is measured by comparing the arrival times of the avalanche induced pulse at both ends of the inductive delay lines placed in the module walls in the sense wire plane. The relevant parameters of the CDC are listed in the Table 3.2.

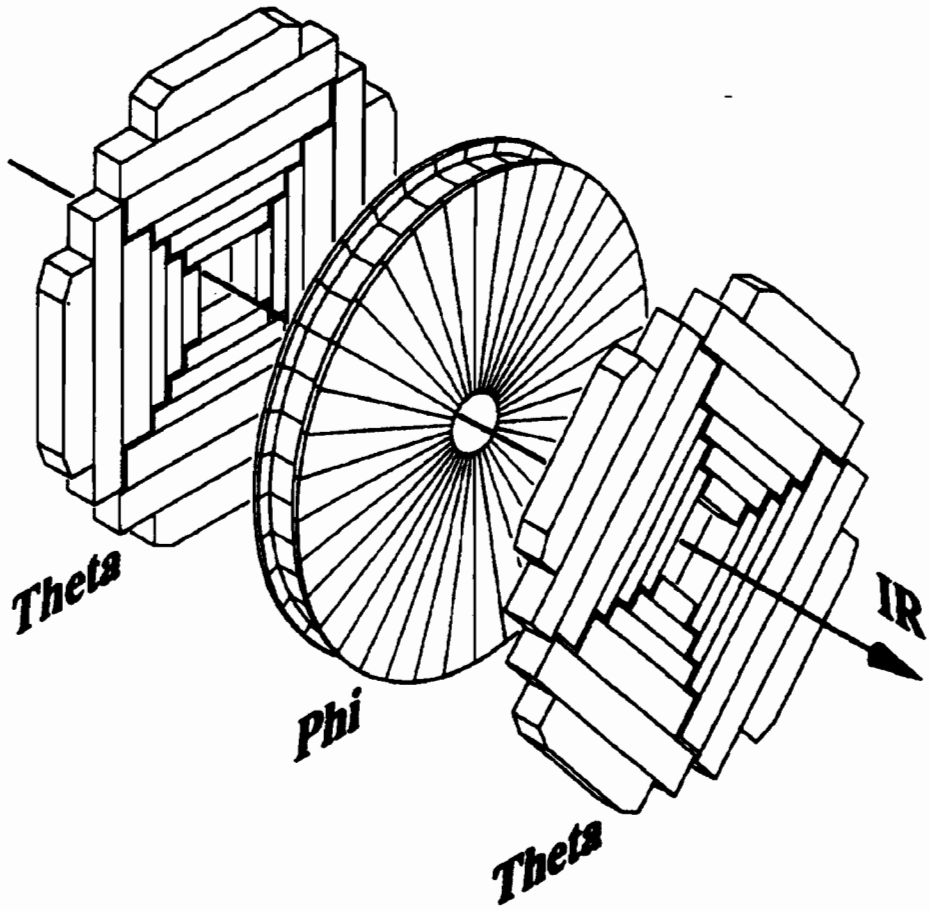


Figure 3.8: An exploded end view of an FDC

3.3.5 The Forward Drift Chambers (FDC)

The forward drift chambers are located at both ends of the CDC, covering $1.2 \leq |\eta| \leq 3.1$. This translates to a θ range of $\sim$ 5°-34°. Each FDC consists of 3 layers of chambers, two Θ layers sandwiching a Φ layer. Figure 3.8 shows an exploded view of one of the forward drift chambers. The Φ layer is divided into 36 azimuthal drift cells, each containing 16 sense wires strung radially. The two Θ layers consist of 4 separate quadrants, each containing 6 rectangular drift cells. The sense wires in each cell are oriented parallel to the z-axis. In each rectangular drift cell there is a

delay line similar to that of CDC to measure the position along the length of the cell. However, there is no delay lines in the Φ layer. The outer Θ chambers are rotated by 45° with respect to the inner ones.

The parameters of the FDC are given in Table 3.3.

3.4 The Calorimeter

The absence of a central magnetic field makes the calorimeter the only available device for measuring the energy of most types of particles. Due to the importance of the calorimeter in the DØ detector, we will have a brief overview of calorimetry. A more detailed discussion can be found elsewhere [8, 9].

3.4.1 Operation Principles of A Calorimeter

In calorimetry, one measures the energy deposited by a particle, by stopping it in an absorber. As a high energy electron ($E^e \gg 10$ MeV) passes through a dense material, it interacts electromagnetically with atomic nuclei in the material and emits energetic photons (Bremsstrahlung). A high energy photon, in turn, produces electron-positron pairs. The photons and electron-positron pairs created through Bremsstrahlung or pair production produce more electrons, positrons and photons undergoing the same processes. Therefore, an energetic photon or electron passing through dense media can produce a shower of electrons, positrons and photons known as an *electromagnetic shower*. As the shower develops, it loses energy mostly due to ionization until it does not have enough energy to go through showering processes. The rate of energy loss in a material is constant and depends only on the type of the

material. The rate is expressed as:

$$\frac{(\frac{dE}{dx})}{E} = -\frac{1}{X_0} \quad or \quad E = E_0 e^{-\frac{x}{X_0}}, \tag{3.9}$$

where X_0 is called the *radiation* length.

Hadrons, on the other hand, lose energy by colliding inelastically with atomic nuclei. The hadrons produced by these collisions can cause more inelastic collisions to produce a *hadronic shower*. The hadronic shower continues to develop until it loses its energy due to ionization and inelastic collisions. The rate of energy loss for hadronic showers is the same as in equation 3.9 with X_0 replaced by the *nuclear absorbtion length* λ. For uranium, $\lambda \sim 10.5$ cm, whereas $X_0 \sim 3.2$ mm. So, hadronic showers are generally both longitudinally and transversly larger than electromagnetic showers.

In order to measure the energy of the low energy particles produced through showering, layers of an ionization-sensitive material can be inserted in the dense particle absorber. Since this active medium only sees a fraction of the energy lost by the incident particle, this type of calorimeter is called a *sampling calorimeter*. The fraction of the energy detected is known as the *sampling fraction*. The statistical nature of this type of calorimetry degrades the energy resolution.

The response of the calorimeter to electromagnetic and hadronic showers is different. Since neutrinos and muons produced by K and π decays escape the calorimeter and the nuclei break-up energy is not measured, the calorimeter response to hadronic showers is smaller. This difference in calorimeter response to electromagnetic and

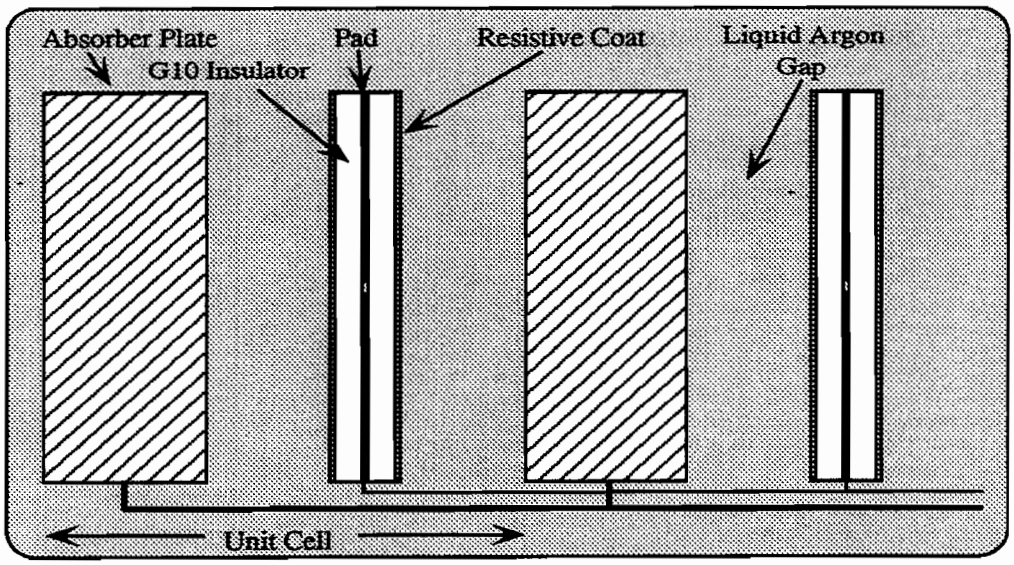


Figure 3.9: A unit cell of the liquid argon calorimeter

hadronic showers is quantified by measuring the ratio of the responses to electrons and pions, known as the e/π ratio. Hadronic showers can have electromagnetic components through η and π decays to photons. The fraction of hadrons which cause electromagnetic showers may change from shower to shower. In order to have an energy resolution independent of this, it is desirable to have $e/\pi = 1$. A calorimeter with $e/\pi \sim 1$ is called a *compensating* calorimeter.

3.4.2 Calorimeter Configuration

The DØ calorimeter is a sampling calorimeter. The primary absorber material used in the calorimeter is depleted uranium; copper and stainless steel are also used. The sampling medium is liquid argon (LAr). The calorimeter is divided into a large number of modules, each containing layers of absorber plates and signal boards. Figure 3.9 shows a part of a module. The 2.3 mm gaps between absorber plates and signal boards are filled with LAr as the ionization material. The signal boards are copper readout pads placed between two layers of G10 covered by a resistive epoxy.

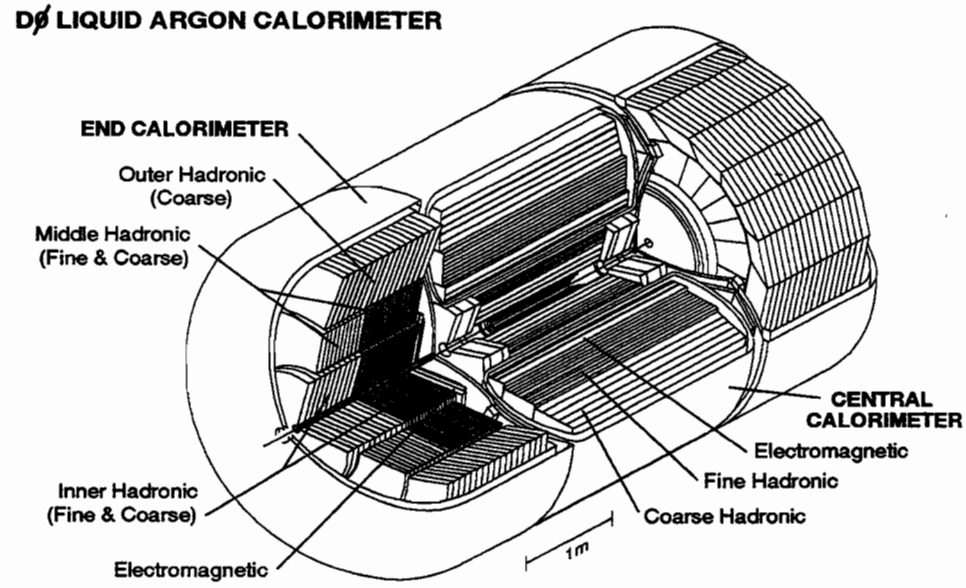


Figure 3.10: DØ calorimeter

The electric potential difference between the resistive coating and the absorber plates is about 2.0-2.5kV with the absorber plates grounded. During operation, as a shower develops, charged particles in the shower cause ionization in the LAr and the collected electrons on the signal boards induce a signal on the copper pads. Each readout pad is further divided into smaller cells to measure the transverse position of the shower.

The calorimeter is divided into a central calorimeter (CC), covering $|\eta| < 1.2$ and two end calorimeters (EC) with a coverage of $1.3 < |\eta| < 4$. Since the sampling material used in the calorimeter is LAr, the calorimeter has to be kept cold. Therefore, both CC and EC are placed inside cryostats. Figure 3.10 shows an isometric view of the DØ calorimeter.

The CC consists of 3 layers of modules. The inner layer contains 32 electromagnetic (EM) modules for electromagnetic shower measurement. The middle layer contains 16 fine hadronic (FH) modules, measuring hadronic showers. The outer lay-

ers consists of 16 coarse hadronic (CH) modules, to reduce the leakage(punchthrough) out of the calorimeter to the muon system. The EC contains 3 layers of electromagnetic, fine hadronic and coarse hadronic modules. The relevant parameters of the central and end calorimeters are summarized in Tables 3.4 and 3.5.

	Θ *modules*	Φ *modules*
z interval	104.8-111.2 cm	113.0-127.0 cm
	128.8-135.2 cm	
Radial Interval	11-62 cm	11-61.3 cm
Number of Cells in Radius	6	
Maximum Drift Distance	5.3 cm	5.3 cm
Stagger of Sense Wires	0.2 mm	0.2 mm
Sense Wire Separation	8 mm	8 mm
Angular Interval/cell		10°
Number of Sense Wires/Cell	8	16
Number of Delay Lines/Cell	1	0
Number of Sense Wires/End	384	576
Number of Delay Lines Readout/End	96	
Gas Mixture	Ar(93%)-CH_4(4%)-CO_2(3%)	
Gas Pressure	1 atm	1 atm
Drift Field	1.0 kV/cm	1.0 kV/cm
Average Drift Velocity	37 μm/ns	40 μm/ns
Gas Gain at Sense Wire	$2.3, 5.3 \times 10^4$	3.6×10^4
Sense Wire Potential	+1.5 kV	+1.5 kV
Sense Wire Diameter	30 μm NiCoTin	
Guard Wire Diameter	163 μm Au-plated Al	

Table 3.3: FDC detector parameters

	EM	FH	CH
Rapidity Coverage	$\pm$ 1.2	$\pm$ 1.0	$\pm$ 0.6
Number of Modules	32	16	16
Absorber[a]	Uranium	Uranium	Copper
Absorber Thickness (cm)	0.3	0.6	4.65
Argon Gap (cm)	0.23	0.23	0.23
Number of Cells/Module	21	50	9
Longitudinal Depth	20.5 X_o	3.24 λ_o	2.93 λ_o
Number of Readout Layers	4	3	1
Cells/Readout Layer	2, 2, 7, 10	21, 16, 13	9
Total Radiation Lengths	20.5	96.0	32.9
Radiation Length/cell	0.975	1.92	3.29
Total Absorption Lengths (Λ)	0.76	3.2	3.2
Absorption Length/Cell	0.036	0.0645	0.317
Sampling Fraction (%)	11.79	6.79	1.45
Segmentation ($\phi \times \eta$)[b]	0.1 $\times$ 0.1	0.1 $\times$ 0.1	0.1 $\times$ 0.1
Total Number of Readout Cells	10,368	3456	768

Table 3.4: CC detector parameters

[a] Uranium is depleted and FH absorbers contain 1.7% Niobium alloy

[b] Layer 3 of the EM has 0.05 $\times$ 0.05

	EM	IFH	ICH	MFH	MCH	OH
Rapidity Coverage	1.3-4.1	1.6-4.5	2.0-4.5	1.0-1.7	1.3-2.0	0.7-1.4
Number of Modules/End Calor.	1	1	1	16	16	16
Absorber[a]	U	U	SS[b]	U	SS	SS
Absorb Thickness (cm)	0.4	0.6	4.6	0.6	4.6	4.6
Argon Gap (cm)	0.23	0.21	0.21	0.22	0.22	0.22
Number of Cells/Module	18	64	12	60	12	24
Longitudinal Depth	$20.5X_0$	$4.4\lambda_0$	$4.1\lambda_0$	$3.6\lambda_0$	$4.4\lambda_0$	$4.4\lambda_0$
Number of Readout Layers	4	4	1	4	1	3
Cells/Readout layer	2, 2, 6, 8	16	12	15	12	8
Total Radiation Lengths	20.5	121.8	32.8	115.5	37.9	65.1
Total Absorption Length (Λ)	0.95	4.9	3.6	4.0	4.1	7.0
Sampling Fraction (%)	11.9	5.7	1.5	6.7	1.6	1.6
$\Delta\phi$ Segmentation[c]	0.1	0.1	0.1	0.1	0.1	0.1
$\Delta\eta$ Segmentation[d]	0.1	0.1	0.1	0.1	0.1	0.1
Total Number of Readout Ch.[e]	14976	8576	1856	2944	768	1784

Table 3.5: EC detector parameters

[a] Uranium is depleted and FH (IFH and MFH) absorbers contain 1.7% Niobium alloy

[b] Stainless Steel

[c] Layer 3 of the EM has $\Delta\phi \times \Delta\eta = 0.05 \times 0.05$ for $|\eta| < 2.6$

[d] For $|\eta| > 3.2$, $\Delta\phi = 0.2$ and $\Delta\eta \approx 0.2$

[e] MCH and OH are summed together at $|\eta| = 1.4$

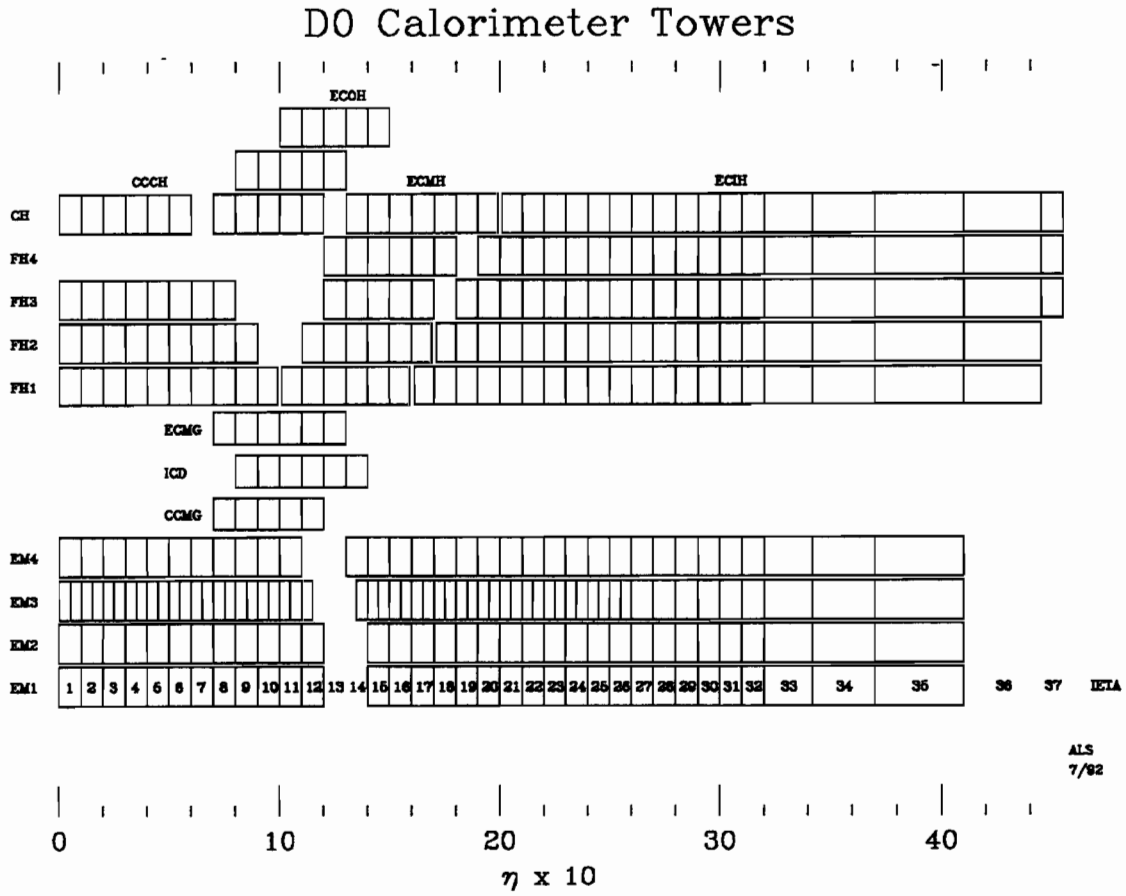


Figure 3.11: Calorimeter modules in r-η.

The typical coverage of a readout cell is 0.1×0.1 in η-ϕ space. The cells in the third-layer of the electromagnetic modules are smaller (0.05×0.05), and the cells in pseudorapidity beyond 3.2 are larger. Figure 3.11 shows the calorimeter segmentation for full η coverage of 0-4 in all layers.

The region $0.8 < |\eta| < 1.4$ between the CC and the EC in the calorimeter is uninstrumented due to the cryostat walls and support hardware (see fig 3.4). In order to sample the shower development in this region, two subsystems are employed: the massless gaps(MG) and the intercryostat detector(ICD). The ICD is an array of 384 scintillation counter tiles of size 0.1×0.1 in η-ϕ space. The ICD modules are mounted on the front surface of each EC cryostat. Phototubes are utilized for ICD readout. The modules of the massless gaps contain two signal boards with LAr-filled gaps. The size of the readout cells of the massless gaps is 0.1×0.1 in η-ϕ space. The massless gap modules are mounted on the end plates of the CCFH, ECMH, and ECOH modules [5] (see fig 3.11).

3.4.3 Calorimeter Performance

The calorimeter energy resolution is given by

$$(\frac{\sigma}{E})^2 = C^2 + \frac{S^2}{E} + \frac{N^2}{E^2}, \tag{3.10}$$

where C, S, and N are constants reflecting the error due to calibration, statistical fluctuations, and noise, respectively. The measured values for these constants are:

$$C = 0.003 \pm 0.002,\ S = 0.157 \pm 0.005\ (\text{GeV})^{\frac{1}{2}},\ N = 0.140\ \text{GeV}$$

[5] The CCFH, ECMH, and ECOH stand for CC fine hadronic, EC middle hadronic, and EC outer hadronic, respectively.

for electrons and:

$$C = 0.032 \pm 0.004,\ S = 0.41 \pm 0.04\ (\mathrm{GeV})^{\frac{1}{2}},\ N = 1.28\ \mathrm{GeV}$$

for pions. The e/π ratio ranges from 1.04(150 GeV) to 1.11(10 GeV) and the resolution for position measurements is about 0.8-1.2 mm.

3.5 The Muon System

Although muons are not studied in the analysis presented in this thesis, a brief overview of the muon system is included in this Section. A detailed discussion of the muon system can be found in [11, 12, 13].

The muon system consists of 5 solid iron toroidal magnets sandwiched between 3 layers of proportional drift tubes (PDT).

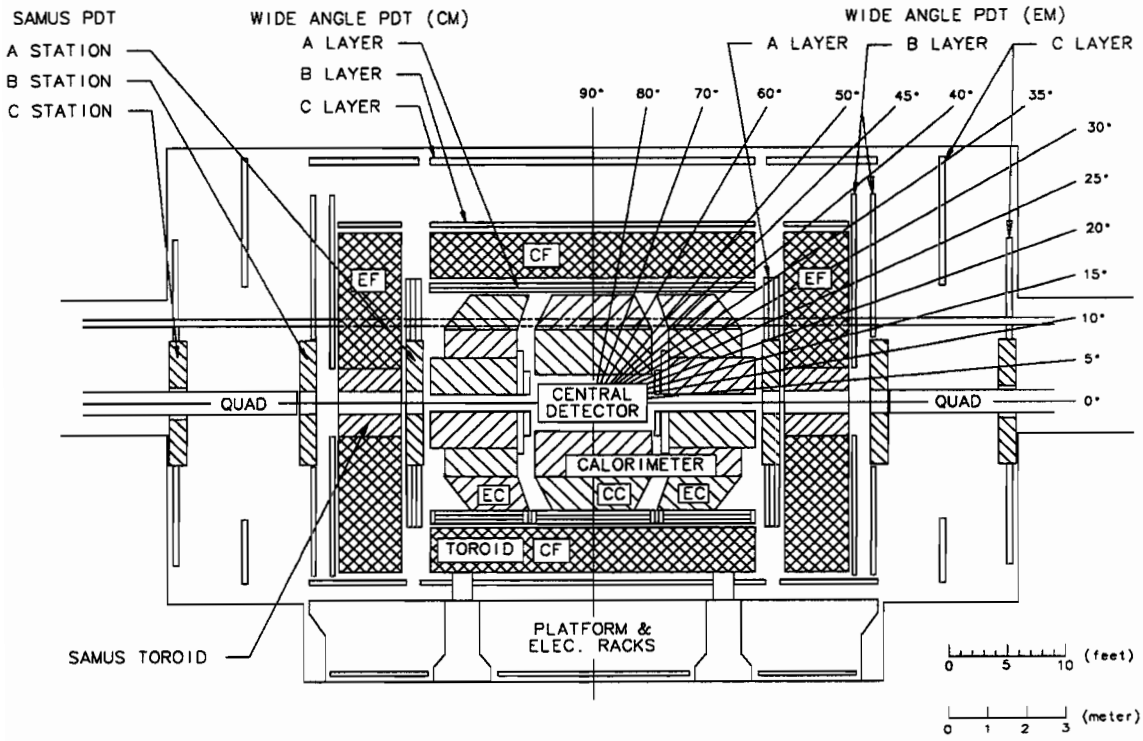


Figure 3.12: Muon system.

The initial and final directions of the muon are measured and the muon momentum is determined by measuring the deflection of the muons in the magnets. The muons are deflected in the r-z plane as they experience a 1.9 T magnetic field. The three major parts of the muon system are (fig 3.12):

- A magnet in the central region called CF (Central Fe) covering $|\eta| \leq 1.0$.
- Two magnets in the end regions called EF (End Fe) covering $1.0 < |\eta| \leq 2.5$.
- Two magnets in the Small Angle Muon System (SAMUS) covering $2.5 < |\eta| \leq 3.6$.

The CF and the two EF's together are known as the Wide Angle Muon System (WAMUS). One layer of PDT is just inside the magnet and two layers outside the magnet with a 1-3m air-gap between them. The inner and the two outer layers are referred to as A, B, and C layers, respectively. The muons should have at least a minimum energy of 3.5 GeV to reach the muon system in the central region. This minimum energy becomes about 5 GeV for higher η as the muon has to go through more material in the calorimeter. The momentum resolution for the muon is given by

$$(\frac{\Delta p}{p})^2 = (0.18)^2 + (0.01p)^2, \tag{3.11}$$

where p is the momentum of the muon in GeV. Table 3.6 summarizes the relevant parameters of the muon system.

3.6 The Trigger System

As derived in the beginning of this Chapter, the instantaneous luminosity is about $0.5 \times \text{nb}^{-1}/\text{sec}$. The total cross section for $p\bar{p} \rightarrow X$ is 70 mb. Therefore, the

	WAMUS	SAMUS
Rapidity Coverage	$\|\eta\| \leq 2.3$	$2.3 \leq \|\eta\| \leq 3.6$
Magnetic Field	2 T	2 T
Number of Chambers	164	448
Interaction Lengths	13.4	18.7
Bend View Resolution[a]	$\pm$ 0.53 mm	$\pm$ 0.35 mm
Non-Bend Resolution	$\pm$ 3 mm	3.5 mm
$\delta P/P$[b]	18%	18%
Gas	Ar(90%)-CF_4(5%)-CO_2(5%)	CF_4(90%)-CH_4(10%)
Average Drift Velocity	6.5 cm/μs	9.7 cm/μs
Anode Wire Voltage	+4.56 kV	+4.0 kV
Cathode Pad Voltage	+2.3 kV	-
Number of Cells	11,386	5308
Sense Wire Diameter	50 μm	50 μm

Table 3.6: Muon system parameters

[a]The diffusion limit is 0.2-0.3 mm.

[b]Multiple Coulomb Scattering limit - assumes 100% chamber efficiency.

rate of $p\bar{p}$ interactions is about 350 kHz. Since it is not feasible to record all of the interactions, a filtering system, known as the *trigger* system has been implemented to select events relevant to the physics analyses at DØ. Triggering is carried out in 3 levels; level 0, level 1 and level 2.

Level 0

Level 0 is a scintillator-based trigger designed for fast vertex position measurement and to indicate the occurrence of inelastic collisions. The Level 0 trigger consists

of two hodoscopes mounted on the front surface of the EC cryostats. The hodoscopes contain two sets of scintillator counters arranged in planes perpendicular to the beam and rotated 90°with respect to each other. Since inelastic collisions cause a significant activity in the forward regions, the coincidence of the signals from two scintillator arrays is attributed to inelastic collisions. Moreover, the interaction vertex is determined with a resolution of ± 15 cm, by comparing the arrival times of the signals from two scintillator arrays. The hodoscpes give a partial coverage for $1.9 < |\eta| < 4.3$ and an almost complete coverage for $2.2 < |\eta| < 3.9$.

The Level 0 trigger is also used for instantaneous luminosity measurements [15]. The rate of the events at the Level 0, N_{L0}, is given by:

$$N_{L0} = \epsilon_{L0} L_{ins} \sigma_{tot}, \tag{3.12}$$

where ϵ_{L0} is the acceptance of the Level 0 trigger. Using a Monte Carlo simulation ϵ_{L0} is determined to be 69%. The cross-section σ_{tot} is measured by experiments E710 [16, 17] and CDF [18, 19] at the Tevatron, the average of which used by DØ is 70 mb. The relative error on the instantaneous luminosity is 5.4% [20].

Level 1

The level 1 trigger decides within 3.5 μs (bunch crossing time) whether an event should be kept or not, based on the information collected from Level 0, the calorimeter, and the muon system. These pieces of information are stored in a 32-bit logical trigger word. Events with at least one nonzero trigger bit are accepted. A programmable prescale is associated with each trigger bit. When a prescale N is assigned to a trigger, the trigger bit is fired every once in N times that events satisfy the trigger condition.

Level 2

The events which pass the level 1 trigger requirements are transferred to level 2. Level 2 is a software trigger system which employs 48 nodes of VAXstations (4000/M60 and 4000/M90) for fast reconstruction of the digitized raw data collected from the different parts of the detector which passed level 1. At this level a set of algorithms is implemented to look for electromagnetic jets, hadronic jets, missing transverse energy and muons, *etc.* As a result, 128 software filters are built out of these algorithms for the specific physics analyses. Any event passing level 2 must satisfy the requirements of at least one of these filters. The output event rate for the full trigger system is about 1-2 Hz. The trigger rate at each level is shown in Figure 3.13.

Events passing Level 2 are sent to the host system where they are recorded on magnetic tapes. In the next Chapter we will see how this digitized data is used to derive the various parameters of the physical objects. For further study of the experimental apparatus at DØ, references [12, 13] are recommended.

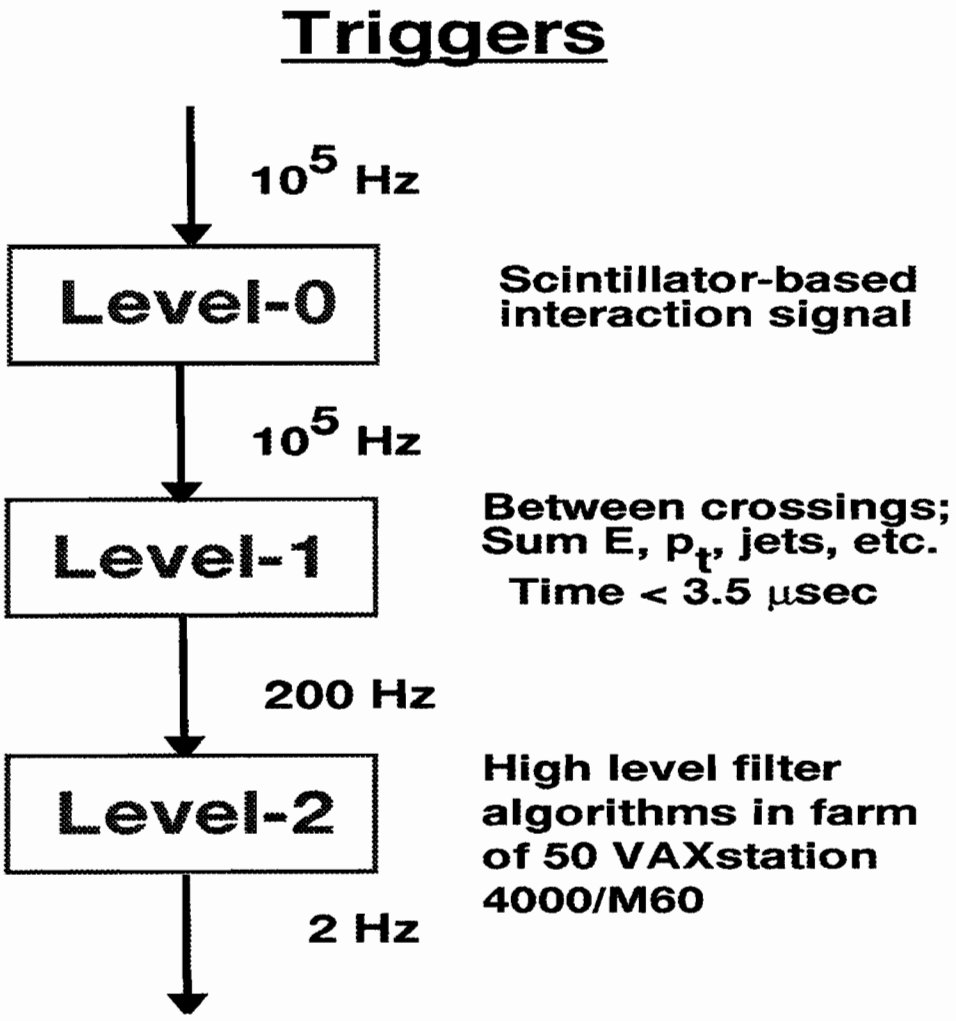


Figure 3.13: Trigger rates.

Chapter 4

Event Reconstruction

The signals collected from different parts of the detector are due to the particles produced by the $p\bar{p}$ collisions. However, one cannot describe the kinematic parameters of these particles, using the raw information from the detector directly. The process of turning the raw data into the kinematic parameters of the particles is known as event *reconstruction*. The reconstruction process in DØ is carried out by a software package called DØRECO.

4.1 Event Vertex

As explained in the previous Chapter, the transverse components of the momenta and energies of the particles are the quantities of interest in physics analyses. In order to calculate the transverse energies, one needs to know the energy and direction of the particles. The calorimeter determines the energy and the position of the particles where they hit. To find the direction of the particles, the position of the interaction vertex from which the particles originate is needed.

The vertex position is reconstructed with the following procedure:

- The drift chamber hit are fitted to reconstruct a track in the r-ϕ plane[1];
- A reconstructed r-z track is associated with the reconstructed r-ϕ track;
- The intersections of these reconstructed space tracks with the z-axis form a distribution in z. The estimated z position of the vertex is the mean of a gaussian fitted to this z intercept distribution. In the case of a z distribution with more than one peak and therefore multiple vertices, the vertex with the maximum number of tracks is considered as the primary vertex. The resolution of the vertex z-component is about 1-2 cm.

4.2 Electrons

An electron is identified as localized energy in the electromagnetic calorimeter with a track in the central detector pointing to a vertex. Electrons are reconstructed by the following procedure:

- A cluster is constructed by starting with the highest-E_T EM tower[2] and adding nearby towers to the cluster with E_T above a certain threshold;
- At least 90% of the energy of the cluster should be from the electromagnetic calorimeter and 40% of its energy should be contributed by a single tower;

[1]The stagger in the sense wire helps finding the correct solution by choosing the best fit.

[2]An EM tower is defined as 4 layers of the EM calorimeter plus the first layer of the fine hadronic calorimeter.

- The centroid of the cluster is calculated as the ln(E)-weighted mean of the cell positions in the third EM layer(the third layer is the most finely segmented EM layer);
- If there is a track in the central detector within a solid angle [3] of $\Delta\eta = \pm 0.1$ and $\Delta\phi = \pm 0.1$, pointing from a vertex to the cluster, the cluster will be identified as an electron; otherwise, it is considered as a photon.

The electron energy resolution is

$$(\frac{\sigma}{E})^2 = C^2 + \frac{S^2}{E} + \frac{N^2}{E^2}, \tag{4.1}$$

where E is the mean energy, C reflects the calibration errors due to the thickness variation in the LAr gaps and the momentum variation of the calibration test beam, S is the error due to statistical fluctuations, and N reflects the noise due to the electronics and the radioactivity of the absorber.

4.2.1 Electron Identification

The identification procedure explained above favors efficiency rather than rejection. The following quantities can further discriminate electrons from other objects.

- Isolation fraction:

  The electrons coming from W's should not be too close in space to other objects. Therefore the fraction of the electron energy outside a certain well-difined cone should be small. In order to quantify this idea the *isolation fraction* is defined as:

$$f_{isol} = \frac{E_{TOT}(0.4) - E_{EM}(0.2)}{E_{EM}(0.2)}, \tag{4.2}$$

[3] Frequently referred to as the "road".

where $E_{TOT}(0.4)$ is the energy deposited in all calorimeter cells within $\Delta R = 0.4$ around the electron direction and $E_{EM}(0.2)$ is the energy deposited in the electromagnetic calorimeter within $\Delta R = 0.2$. Requiring $f_{isol} < 0.1$ enhances electron identification.

- Track ionization:

  Photon conversion, $\gamma \rightarrow e^+e^-$, can produce close electron-positron pairs. Since there is no magnetic field to move the electron and positron further apart, they may be reconstructed as a single track. However, the energy deposition per distance, $\frac{dE}{dx}$, would be twice the $\frac{dE}{dx}$ for a single electron. The unit usually used for $\frac{dE}{dx}$ is that of the "minimum ionization particle" (mip). One can have a better electron identification rejection by excluding electron candidates with $\frac{dE}{dx}$ around 2 mips. The actual cuts used for $\frac{dE}{dx}$ are:

  $1.5 \leq \frac{dE}{dx} \leq 3.0$ for the CDC;

  $1.3 \leq \frac{dE}{dx} \leq 2.7$ for the FDC

- Track match significance:

  Since photons do not leave a track in the central detector, one can enhance the electron identification by reducing the chance of reconstructing a track for photons due to nearby charged particles. In order to quantify the accuracy with which a track points to a calorimeter cluster, the *track match significance* is defined as:

$$S \equiv \sqrt{(\frac{\Delta\phi}{\sigma_{\Delta\phi}})^2 + (\frac{\Delta z}{\sigma_{\Delta z}})^2} \quad \text{for the CC;} \tag{4.3}$$

$$S \equiv \sqrt{(\frac{\Delta\phi}{\sigma_{\Delta\phi}})^2 + (\frac{\Delta r}{\sigma_{\Delta r}})^2} \quad \text{for the EC,} \tag{4.4}$$

  where $\Delta\phi$, Δz, and Δr denote the coordinate differences between the track hit and the cluster centroid, and $\sigma_{\Delta\phi}$, $\sigma_{\Delta z}$, and $\sigma_{\Delta r}$ are the corresponding resolu-

tions. A cut of $S < 5$ will further discriminate electrons from the background.

- Covariance Matrix χ^2:

  Suppose $|x>$ is a vector whose components are a set of M linearly independent variables. The covariance matrix[4] , V, for a sample of N events is defined as

$$V = \frac{1}{N}\sum_{i=1}^{N}(|x_i> - |\bar{x}>)(<x_i| - <\bar{x}|), \tag{4.5}$$

  where

$$|\bar{x}> = \frac{1}{N}\sum_{i}^{N}|x_i> . \tag{4.6}$$

  Then the χ^2 for an event $|x_k>$ not in the sample is defined as:

$$\chi^2 = (<x_k| - <\bar{x}|)V^{-1}(|x_k> - |\bar{x}>). \tag{4.7}$$

  The components of $|x>$ are linearly independent, so V is invertible and the above definition is well-defined. Since V is symmetric, it is diagonalizable. Suppose the $|y_i>$, $i = 1, ..., M$ are eigenvectors of V and the σ_i^2 $i = 1, ..., M$ are its eigenvalues[5]. Then

$$\chi^2 = \sum_{i=1}^{M}\frac{(y_k^i - \bar{y}^i)^2}{\sigma_i^2}. \tag{4.8}$$

  A small χ^2 requires that the event k be close to the means $\bar{y}^i$ if the $|y_i>$ distributions of the sample events are approximated by gaussians with variances σ_i.

$$e^{-\frac{1}{2}\chi^2} = \prod_{i=1}^{M} e^{-\frac{(y_k^i - \bar{y}^i)^2}{2\sigma_i^2}} \tag{4.9}$$

---

[4]The multi-dimensional analogue of the variance.

[5]Suppose λ_i is an eigenvalue of V corresponding to $|y_i>$. Hence,

$$\lambda_i = <y_i|V|y_i> = \frac{1}{N}\sum_{i=1}^{N}|<y_i|x_i> - <y_i|\bar{x}>|^2 \geq 0 \rightarrow \lambda_i = \sigma_i^2.$$

To have a better electron identification, the χ^2 has been calculated for a set of 41 variables. These variables are:

- The fractional energies in layers 1, 2 and 4 of the EM calorimeter;
- The fractional energy in each cell of the 6×6 array in the third layer of the EM calorimeter;
- The vertex z-component;
- The logarithm of the total energy of the cluster.

The covariance matrix V is obtained using Monte Carlo electrons. A cut of $\chi^2 < 100$ is usually applied for better background rejection.

- The electron likelihood:

  In order to obtain a better background rejection one can optimize a multivariate cut in the 4-dimensional space of χ^2, S, $\frac{dE}{dx}$, and f_{EM} where $f_{EM} = \frac{EM\ energy}{Total\ energy}$. The likelihood method is applied to obtain this cut. The optimized likelihood cuts are [14],

  $$L^e < 2.5 \text{ for the CC and } L^e < 3.0 \text{ for the EC}$$

  where $L^e \equiv L^e(\chi^2, S, \frac{dE}{dx}, f_{EM})$.

Table 4.1 summarizes the efficiencies for different electron-id cuts relevant to this thesis.

4.3 Jets

Jet reconstruction is carried out using the "cone algorithm", in which a jet is considered as the energy inside of a cone with a fixed size in η-ϕ. In this algorithm the

	Efficiency in CC	Efficiency in EC
electron likelihood	0.8135 ± 0.0127	0.4913 ± 0.0200
tracking efficiency(data)	0.8267 ± 0.0107	0.8519 ± 0.0102
tracking efficiency(Monte Carlo)	0.9475 ± 0.0064	0.9069 ± 0.0157

Table 4.1: Electron tracking and likelihood efficiencies.

calorimeter towers are sorted in E_T to form a set of "seed" clusters. A precluster is constructed using all of the towers above an E_T-threshold of 1 GeV within $|\Delta\eta| < 0.5$ and $|\Delta\phi| < 0.5$ around the highest-E_T tower which has not yet been assigned to a precluster. The axis of the jet is defined as the E_T-weighted centroid. Some cells may contribute to more than one jet. If the fraction of the energy shared by the jets is more than 50%, the two jets will be merged and a new jet axis is defined using all the cells in the merged jet. If the fraction of the energy shared is less than 50%, the two jets are considered as separate jets and each shared cell is assigned to the closest jet. A threshold of 8 GeV is required for the E_T of the jets. However, the response of the calorimeter is nonlinear for low-energy (< 2 GeV) particles and the sum of the calorimeter responses does not give the correct total energy. Moreover, a hadronic shower may develop beyond the jet cone. Some corrections are applied to take these effects into account [21].

The jet energy resolution is expressed by:

$$(\frac{\sigma}{E_T^j})^2 = C^2 + \frac{S^2}{E_T^j} + \frac{N^2}{{E_T^j}^2}, \tag{4.10}$$

where C, S, and N are constants depending on the calibration error, the shower

fluctuations in the sampling gap and the detector noise, respectively. The value of these constants obtained from the data are:

$C = 0.01 \pm 0.005,\ S = 0.74 \pm 0.07\ (\text{GeV})^{\frac{1}{2}},\ N = 2.16 \pm 0.22$ GeV.

4.4 Missing Transverse Energy

According to conservation of transverse momentum, the sum of the transverse momenta of the particles produced by $p\bar{p}$ collisions should be zero. If the total transverse momentum is significantly different from zero, the difference is attributed to neutrinos. In order to calculate the transverse energy of the neutrinos, a vector $\vec{E}_T$ is assigned to each calorimeter cell, including the ICD, whose magnitude is the measured energy in the cell and it points from the interaction vertex to the center of the cell. Then the calorimeter missing energy is defined as:

$$\vec{\not\!\!E}_T^{\,cal} \equiv - \sum_{all\ cells} \vec{E}_{cell}. \tag{4.11}$$

The muons deposit small amounts of energy in the calorimeter, so the transverse momenta of all good muon tracks should be subtracted from $\vec{\not\!\!E}_T^{\,cal}$ to get the total transverse missing energy, $\vec{\not\!\!E}_T$. Since all the objects in the calorimeter contribute to $\not\!\!E_T^{\,cal}$ any mismeasurement in the energy of these objects would cause a mismeasurement in $\not\!\!E_T^{\,cal}$. Therefore when a correction is applied to the calorimeter objects, a corresponding correction should be applied to $\not\!\!E_T^{\,cal}$.

The $\not\!\!E_T$ resolution has been studied using QCD dijet data samples. The $\not\!\!E_T$ resolution is parameterized as follows [22]:

$$\sigma = a + bS_T + cS_T^2, \tag{4.12}$$

where S_T is the summed transverse energy in the calorimeter and

$a = 1.89 \pm 0.05$ GeV, $b = (6.7 \pm 0.7) \times 10^{-3}$, $c = (9.9 \pm 2.1) \times 10^{-6}$ $(\text{GeV})^{-1}$

Chapter 5

Classification and Neural Networks

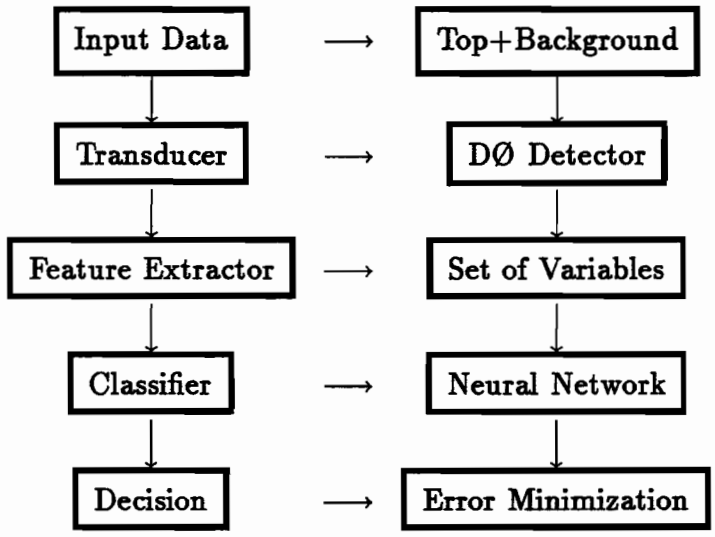
This Chapter provides a brief overview of classification methods, and neural networks formalism, theorems and rules of thumb which are relevant to this thesis. The first Section contains a general discussion about the classification.

5.1 Classification

Any classification model consists of three parts: 1) a transducer; 2) a feature extractor; 3) and a classifier. The transducer senses the input data and converts these data to a form suitable for processing. The feature extractor[1] extracts relevant information from the data. The classifier uses this information to assign input data to one or another of the categories of interest. The distinction between feature extractor and classifier is based on practical, not theoretical considerations. Theoretically, an ideal feature extractor would make the job of the classifier trivial and an extremely powerful classifier would not need a feature extractor.

[1] also known as the receptor, property filter, attribute detector or preprocessor.

The corresponding elements of the classification model used in this thesis (shown below) are the DØ detector as the transducer, a set of variables as the feature extractor and the neural network as the classifier. As explained above, a good choice of variables would simplify the classifier job and a strong classifier would be able to partition the input data using any combination of input variables, so long as they carry all the necessary information and then there would be no need to look for the best choice of variables.



Ideally, one would like to classify the data events such that no decision would be ever wrong. When that is not possible, one would like to minimize the probability of error.

Any classification consists of two steps :

- Estimation of the probability to observe an instance of each class in the data.
- Making a decision based on the estimated probabilities.

First we discuss the decision making and we see the relevance of the probability density function estimation.

5.1.1 Bayes Decision Theory

Bayes Decision theory is based on the assumption that a decision problem is described in probabilistic terms and all the relevant probabilities are known. Let C_i denote the possible *classes* of the data or *states of nature*. Since the state of nature is unpredictable, the C_i's are considered as random variables. For example you can think of the C_i's as the possible classes of the data in the analysis in this thesis, *i.e.* QCD multijet, W+jets and top. We assume there is some *a priori* probability $P(C_i)$ that the state of the data is C_i. If we were to make a decision about the state of the data at this level, with so little information, it would be reasonable to adopt a decision rule as follows :

$$\textit{Decision Rule} \ : \ \textit{decide} \ C_i \quad \textit{if} \ P(C_i) > P(C_j) \quad \forall \, j \neq i \tag{5.1}$$

This may look strange because we assume that the state of the data is always of one type even though we know it contains other types. However, this decision rule guarantees the smallest probability of error. Fortunately, we are not forced to make a decision with so little information about the data. Usually, we have some Monte Carlo samples that represent the patterns we intend to classify and a set of features to distinguish states of the data from each other. Let $\vec{x}$ denote this set of features. For the classification problem in this thesis, the components of $\vec{x}$ can be, for example, the energy of the jets, variables describing the shape of the event, *etc.* We consider $\vec{x}$ as a set of continuous random variables whose distributions depend on the state

of the data. Let $p(\vec{x}|C_i)$ be the *class-conditional* or the *state-conditional* probability density function for $\vec{x}$, given that the state of the data is C_i.

As an example, consider the classification problem we are trying to solve in this thesis, *i.e.* classifying the events in our data-sample into top, QCD multijet and W+jets events. The possible classes of the data are $C = \{top, qcd, wj\}$. We apply certain procedures to estimate the amount of each signal and its background in the data. This would provide *a priori* probabilities to observe each class in the data : $P(top)$, $P(qcd)$ and $P(wj)$. As explained above these *a priori* probabilities do not carry enough information for making a useful decision. So we study certain features of the Monte Carlo samples (simulation), such as $\vec{x} = (E_T^j, \not\!\!E_T, E_T^e, ...)$ for each class. The distributions for the features of the Monte Carlo samples provide more information about each class through $p(\vec{x}|C_i)$.

How can we update our *a priori* probabilities $P(C_i)$ by using the information we obtain through the class-conditional probabilities $p(\vec{x}|C_i)$ derived from the Monte Carlo samples ? The Bayes rule provides the answer. According to the Bayes rule:

$$p(C_i|\vec{x}) = \frac{p(\vec{x}|C_i)P(C_i)}{p(\vec{x})}, \tag{5.2}$$

where

$$p(\vec{x}) = \sum_i p(\vec{x}|C_i)P(C_i). \tag{5.3}$$

Therefore the updated *a priori* probability is $p(C_i|\vec{x})$. This is known as the *a posteriori* probability; $p(\vec{x}|C_i)P(C_i)$ is normalized by $p(\vec{x})$ to make sure that $\sum_i p(C_i|\vec{x}) = 1$.

Since we have more information about each class of the data we can make a more

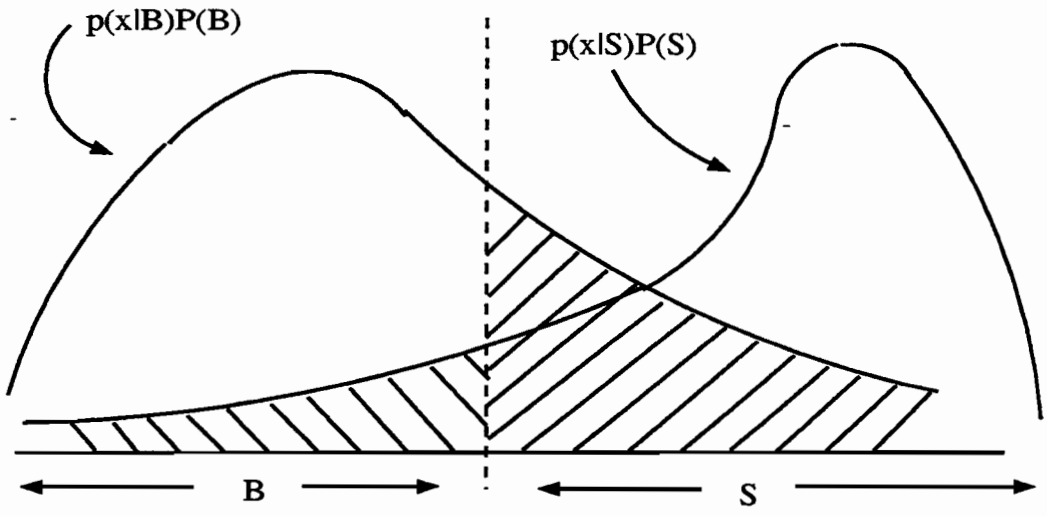


Figure 5.1: Risk for a given one dimensional cut.

useful decision. The Bayes decision rule suggests :

$$decide\ C_i\ if\ p(C_i|\vec{x}) > p(C_j|\vec{x}) \quad for\ \forall\ j \neq i \tag{5.4}$$

This decision rule assures us that the misclassification probability or the probability of error is minimal. Suppose the decision α_i is made where the true state of the data is C_j then this decision would be wrong if $i \neq j$. So

$$p(error) \equiv R(\alpha_i|\vec{x}) = \sum_{j \neq i} p(C_j|\vec{x}), \tag{5.5}$$

where $R(\alpha_i|\vec{x})$ is the risk taken by the decision α_i. Hence :

$$R(\alpha_i|\vec{x}) = \sum_{j \neq i} p(C_j|\vec{x}) = 1 - p(C_i|\vec{x}) \tag{5.6}$$

Since the Bayes decision rule maximizes the *a posteriori* probabilities it minimizes the risk or the misclassification probability. One may think of a classifier as machinery which divides the feature space into decision regions. Suppose a classifier divides the feature space into regions S and B for the signal and the background, then the risk

function is :

$$\begin{aligned} Risk(\vec{x}) = & \quad p(\vec{x}\epsilon S|B)P(B) + p(\vec{x}\epsilon B|S)P(\bar{S}) \\ = & \quad \int_S p(\vec{x}|B)P(B)d\vec{x} + \int_B p(\vec{x}|S)P(S)d\vec{x}. \end{aligned} \tag{5.7}$$

When $p(\vec{x}|B)P(B) > p(\vec{x}|S)P(S)$, the risk would be less if we moved the boundary between the signal and the background regions toward the signal region and vice versa. This is exactly what Bayes decision rule suggests. Figure (5.1) shows the risk where x is one dimensional. However, the decision might be made according to the goal of the classification. The goal might be a certain measurement or the observation of a rare signal. In the latter case one might want to maximize the significance[2] , whereas the former would be in favor of a decision which results in a smaller relative error on the measurement.

As discussed in this Section we can have an optimal classifier if we know the *a priori* probabilities $P(C_i)$ and the class-conditional probabilities $p(\vec{x}|C_i)$. In the next Section we discuss methods which attempt to provide us these probabilities.

5.1.2 Classification Methods

In this Section, we present a brief overview on methods which approximate unknown probabilities. The estimation of $P(C_i)$ is not difficult as compared with the estimation of $p(\vec{x}|C_i)$, where we have serious problems due to limited statistics of the

[2] *Significance* is the number of standard deviations away from the mean of the unit normal distribution in the positive direction above which the integrated probability equals the probability that a normally distributed background with mean B and error ΔB would fluctuate above the signal according to a poisson distribution. This concept will be revisited in the next Chapters.

Monte Carlo samples and large dimensionality of the feature space. So the major focus of the classification methods is on the class-conditional probability density approximation.

The general knowledge about the data will help to guess the functional form of the function $p(\vec{x}|C_i)$ so one could reduce the probability function estimation to a parameter estimation, where $p(\vec{x}|C_i)$ is described as a particular functional form depending on several parameters. For example, based on knowledge about the problem one may consider $p(\vec{x}|C_i)$ as a normal distribution with mean μ and standard deviation σ. As a result, the problem would be reduced to the problem of estimating μ and σ using distributions of Monte Carlo samples in the feature space. This class of techniques is known as *parametric* techniques as opposed to *non-parametric* techniques which estimate probability functions directly. The PDE (probability density estimation) method is an example of a non-parametric technique and the maximum likelihood method and the neural networks are parametric techniques.

Even though non-parametric techniques are more general than parametric techniques, they require much larger statistics to approximate the probability function as compared to the parametric techniques. The only drawback of the parametric technique is the loss of generality due to the particular functional form assumed for the probability function. In the next Section we show how neural networks method minimizes this loss of generality by using a functional form for the probability function which is close to the most general form.

It is worth mentioning that even though the maximum likelihood method and

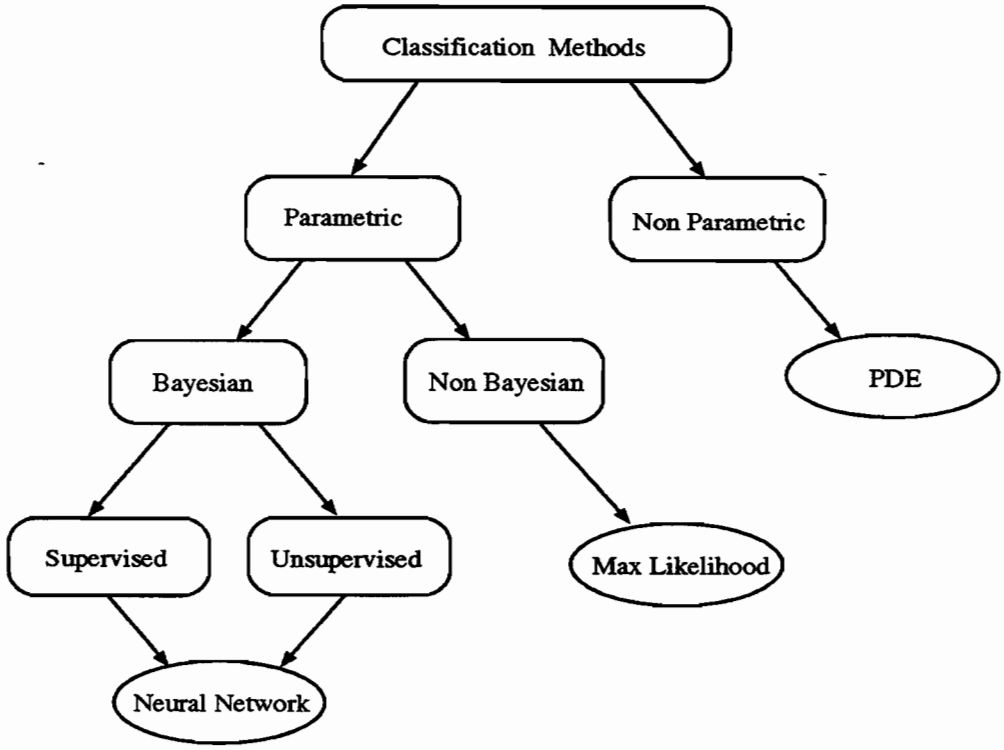


Figure 5.2: Classification methods

the neural networks are both parametric techniques, there is a conceptual difference between them. In the Maximum Likelihood method parameters are considered fixed but unknown and one tries to estimate the best values of the parameters by maximizing the probability (likelihood) of obtaining the observed samples. In the neural network or *Bayesian* estimation, the parameters are random variables with some *a priori* distribution. We then fit these parameters by sharpening the *a posteriori* density function. This process is called *Bayesian learning*. If the samples we select for the training are labelled, then the learning is called *supervised* learning as opposed to *unsupervised* learning where the training samples are not labelled. In the rest of this Chapter we focus on supervised neural networks as a Bayesian estimator.

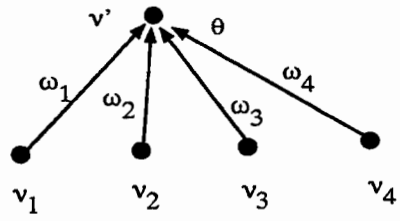


Figure 5.3: Nodes v_i feeding the node v'

5.2 Artificial Neural Networks

Artificial Neural Networks (ANN) is a mathematical model which exhibits features and functionality of the Vertebrate Central Nervous system (CNS), whose details were uncovered in the 1940's with the invention of the electron microscope [23, 24].

Artificial neural networks consist of building blocks called *nodes*[3]. A node is an abstraction of the biological neuron. Mathematically, each node can be considered as a variable whose value is a non-linear function of the superposition of the values of the other nodes. This can be shown analytically as follows :

$$v' = g(\sum_i \omega_i v_i + \theta) \tag{5.8}$$

where v' and the v_i's are nodes, the ω_i's are called *weights* and $-\theta$ is called the *threshold*[4]. The threshold and the weights may be collectively symbolized as weights $\omega = \{\omega_i, \theta\}$. Figure 5.3 represents equation (5.8) schematically. In the neural networks terminology, the nodes v_i are feeding v'. This term would make more sense if you thought of $g(x)$ as a step function or a sigmoid function $g(x) = (1 + e^{-2x})^{-1}$ (fig 5.4). As shown in Figure 5.4, if linear combinations of v_i's exceed the threshold $(-\theta)$, the value of v' will be significant, otherwise v' is negligible and is said to be not *activated*. As we will discuss later, this feature of non-linear functions widely used in

[3]Also called *unit*.

[4]θ is referred to as *bias*.

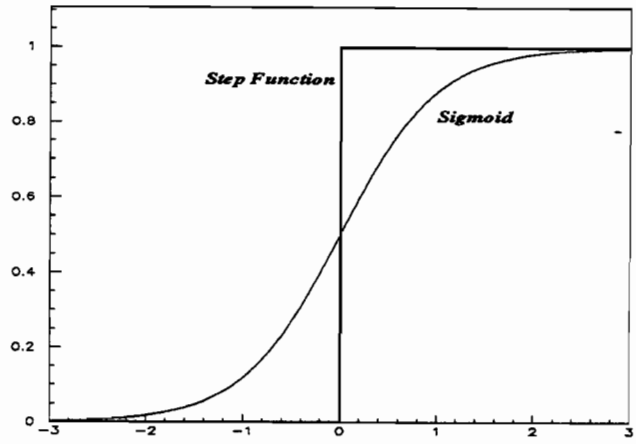


Figure 5.4: Examples of activation functions

the neural networks method plays a major role in classification problems.

By connecting nodes one can design any arbitrary architecture for a neural network. There are two distinct architectures for the neural networks, *feed-forward* and *feed-back*. In the feed-forward networks a signal is processed from a set of input nodes through *layers* of nodes to the output nodes. In the feed-back networks, on the contrary, activation is bidirectional. In this thesis we restrict consideration to feed-forward networks with 3 layers of nodes; an input layer, an output layer, and an intermediate layer known as the *hidden* layer. Figure 5.5 is an example of a feed-forward neural network with 3 input nodes in the input layer, 5 hidden nodes in the hidden layer and one output node. The mathematical expression for a general feed-forward network with one hidden layer and one output is :

$$y = g(\sum_{i=1}^{N_h} \omega_i g(\sum_{j=1}^{N_i} \omega_{ij} x_j + \theta_i) + \theta) \tag{5.9}$$

where the x_j's are input nodes and y is the output node. N_i is the number of input nodes, N_h is the number of hidden nodes and the function $g(x)$ is known as the *activation* function. Therefore, the feed-forward neural network can be considered as

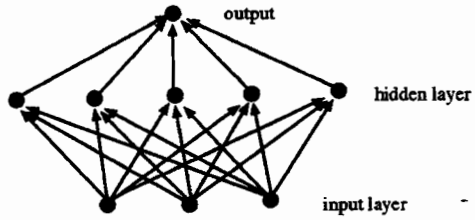


Figure 5.5: A feed-forward network

a non-linear function of $\vec{x}$ with several parameters ω,

$$y = F(\vec{x}, \omega). \tag{5.10}$$

The function $F(\vec{x}, \omega)$ is a candidate for the probability function we try to approximate in the classification problems. This non-linear function has the specific functional form (5.9). How general is the functional form proposed by the feed-forward neural networks?

5.3 Kolmogorov's Theorem and the Existence of Neural Networks

In the 1950's , solving the 13th problem of Hilbert, two prominent Soviet mathematicians, Kolmogorov and V.I.Arnold, were drawn into a friendly competition. Kolmogorov's powerful 1957 theorem [25] ended this competition. Although this theorem did not help mathematicians in proving other important theorems in mathematics, the implications of this theorem concerning the representation of arbitrary functions from the n-dimensional cube to the real numbers in terms of one dimensional functions can be interpreted as an existence theorem for feed-forward neural networks [26]. An improved version of the Kolmogorov's theorem due to Sprecher [27]

states that :

Given any continuous function $f : I^n \rightarrow R^m$ $y = f(x)$ *where* I *is the closed unit interval* $[0, 1]$, f *can be implemented exactly by a 3-layer neural network having* n *input nodes,* $(2n + 1)$ *hidden nodes and* m *output nodes.*

Although this theorem is a powerful theorem, it does not provide a procedure to implement the function f or the activation function $g(x)$ of the network. One can restrict himself/herself to the class of continuous *squashing* functions.

A function $g{:}R \rightarrow [0, 1]$ *is a squashing function if it is non-decreasing ,* $\lim_{x \rightarrow \infty} g(x) = 1$, *and* $\lim_{x \rightarrow -\infty} g(x) = 0$.

It has been proven mathematically [28, 29] that:

The multilayer feed-forward networks with as few as one hidden layer using arbitrary continuous squashing functions are capable of approximating any continuous function from one finite dimensional space to another to any desired degree of accuracy, provided sufficient number of hidden nodes are available.

One of the most frequently used squashing activation functions is the sigmoid function [30]:

$$g(x) = \frac{1}{1 + e^{-2\beta x}}, \quad \beta = \frac{1}{T}. \tag{5.11}$$

T is called the *temperature* and β is referred to as the *inverse temperature*. According to the above theorem if we use a feed-forward neural network with one hidden

layer and a sufficient number of hidden nodes, using the sigmoid activation function, we will have the most general parameterized form for the probability function we intend to approximate in the classification problems. How do we fit the weights ω appearing in the feed-forward neural network general functional form ?

5.4 Training

In the *training* processes the ω's are fitted using the *gradient descent* method on a suitable error function.[5] In this process the *training patterns* (events in the Monte Carlo sample) are presented over and over again with successive updating of the weights. As soon as this iterative process reaches an acceptable level of low error the training process is stopped and the weights are frozen. We then feed the network by patterns that the network has never seen before to classify the events in our data-set, *e.g.* as signal and background. One of the most widely used error functions is the *summed square* error function:

$$E(\omega) = \frac{1}{2N_p} \sum_{i}^{N_p} (y_i(\omega) - d_i)^2 \tag{5.12}$$

where N_p is the number of patterns used for updating the weights, y is the output of the network and d is the desired output. In supervised training, training samples are labelled, so the desired outputs are :

$$d_i \equiv d(C_i) = \delta_{ij} \tag{5.13}$$

[5]Also called the *cost function*.

For example, in the case that we have two classes, *e.g.* the signal and background :

$$d = \begin{cases} 1 & \text{for signal} \\ 0 & \text{for background} \end{cases}$$

We force the network output to be close to the desired output by minimizing the error function $E(\omega)$. In the gradient descent algorithm the ω's are randomly chosen initially. We pick N_p patterns from the training patterns and calculate $E(\omega)$ and $\frac{\partial E}{\partial \omega}$. We then change the ω's in the direction along which E decreases fastest, *i.e.* $\Delta\omega \propto -\frac{\partial E}{\partial \omega}$, so:

$$\Delta\omega(t) = -\eta \frac{\partial E}{\partial \omega}, \tag{5.14}$$

where t is the iteration index and η is a multiplicative factor known as the *learning rate*. This procedure is convergent mathematically [31]. This means that after a sufficient number of iterations E reaches a minimum and the ω's are fixed. In order to avoid oscillation around the minimum and reach the minimum error faster, an extra term known as the *momentum* term is usually added to the equation (5.14) :

$$\Delta\omega(t+1) = -\eta \frac{\partial E}{\partial \omega} + \alpha \Delta\omega(t), \tag{5.15}$$

where t refers to the previous iteration. Now, one may ask how many cycles of training do we have to go through, practically ?

5.5 Cross Validation

Theoretically, for ideal training samples with an infinite number of patterns we can make $|E - E_{min}|$ as small as we wish by going through more cycles. However, in practical cases the number of patterns in the training sample is finite. Finite

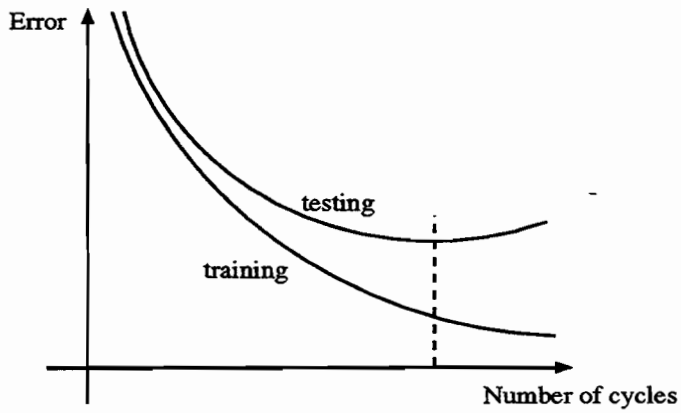


Figure 5.6: Overtraining

samples may contain misleading regularities due to sampling. After a certain number of cycles the network starts to take advantage of these idiosyncrasies to decrease the error. This causes the error on a separate test sample to increase since the network is trying to classify the patterns based on artifacts of the training sample rather than on the general features (fig 5.6). To avoid this overfitting known as *overtraining*, one may use *cross-validation* methods to measure the generalization ability of the network during the training. The two most popular cross-validation methods are the *leave-one-out* cross-validation and the *two-fold* cross-validation. In the leave-one-out cross-validation method one updates the weights in the network using $N-1$ patterns out of N patterns leaving one event out for the testing. This method is difficult to implement in the neural networks technique as compared to the two-fold cross-validation method where the patterns are divided into two disjoint subsets of equal size; training and the testing samples. The error on the testing sample is monitored while the training is in progress on the training sample. As soon as the error on the testing sample starts rising the training is stopped. This then gives us the number of cycles the network has to go through in order to acquire a good generalization ability

by avoiding overtraining. However, overtraining is less relevant when the size of the training sample is much bigger than the number of weights in the network.

5.6 The Optimal Network and The Pruning Algorithm

As discussed earlier in this Chapter, feed-forward networks with one hidden layer and the sigmoid activation function are capable of approximating any continuous function provided a sufficient number of hidden nodes are available. To obtain a good generalization one has to use a small network with a small number of hidden nodes ($N_{weights} \ll N_{patterns}$). On the other hand, a small network may be sensitive to initial conditions and learning parameters or it may get stuck in some local minimum of the error function. One can start with a large network and then reduce the size of the network by trimming the less active hidden nodes during training. This process is called *pruning*. One of the most popular pruning algorithms is to add an extra term to the error function known as the *penalty* term. The penalty term used in this thesis has the following form :

$$E \rightarrow E + \lambda \sum_{ij} \frac{\omega_{ij}^2/\omega_0^2}{1+\omega_{ij}^2/\omega_0^2} \tag{5.16}$$

where the summation is over all weights. For large ω_{ij}'s, *i.e.* $|\omega_{ij}| \gg \omega_0$, the penalty is λ, whereas for $|\omega_{ij}| \ll \omega_0$ the penalty is negligible. So the penalty term is a measure of the number of active weights. Since the training algorithm forces the error function to decrease, those ω's which are not active get pruned by this algorithm during the training until an optimal network architecture is obtained.

Another way of looking at pruning is that the penalty term adds the following extra term to the equation (5.15):

$$-\frac{2\eta\lambda\omega_{ij}/\omega_0^2}{(1+\omega_{ij}^2/\omega_0^2)^2}. \tag{5.17}$$

This term causes the weights to decay with a weight-dependent decay rate:

$$\text{decay rate} = \frac{2\eta\lambda/\omega_0^2}{(1+\omega_{ij}^2/\omega_0^2)^2}. \tag{5.18}$$

As the decay rate implies, smaller weights decay faster while larger weights survive.

5.7 ANN as Bayesian *a posteriori* probability estimator

In the first Section of this Chapter it was explained that the goal of the classification is to estimate *a posteriori* probabilities $p(C_i|\vec{x})$ using $P(C_i)$, the estimation of $p(\vec{x}|C_i)$ from simulation, and applying the Bayes rule $p(C_i|\vec{x}) = \frac{p(\vec{x}|C_i)P(C_i)}{p(x)}$.

One of the most powerful characteristics of neural networks is that under very general conditions a network is capable of estimating $p(C_i|\vec{x})$ directly. Simple proofs [32] show that :

The outputs of the multilayer networks estimate Bayesian a posteriori probabilities when the desired network outputs are 1 *for a given class and* 0 *for all others*

($d(C_i) = \delta_{ij}$) and the summed square or cross-entropy[6] error function is used.

The proof of this theorem assumes the availability of infinite training data. Therefore the accuracy of this estimation depends on the size of the training sample [32]. However, one may choose to consider the output of the neural network as a random variable to discriminate the signal from the background rather than a Bayesian a posteriori probability.

5.8 Rules of Thumb

The last Section of this Chapter is devoted to practical issues concerning training parameters. We discuss some rules of thumb which enhance the quality of the training and the convergence speed.

- Input scale : The change of weights in each updating is proportional to the learning rate η. If we are to use a global learning rate for the network, the weights should be of the same order of magnitude. The outputs of the hidden nodes are sigmoid functions so $< h_i > \sim O(1) \; \forall \; i$. It is natural to rescale the input such that $< x_i > \sim O(1) \; \forall \; i$ so the weights connecting input nodes to hidden nodes would be of the same order of magnitude as the weights connecting hidden nodes to the output node. Therefore a global η can be used across all

[6] Cross-entropy error function:

$$E(\omega) = -\sum_{i}^{N_p}[d_i log y_i + (1 - d_i) log(1 - y_i)] \tag{5.19}$$

layers.

- Learning rate : As a rule of thumb the optimal learning rate scales like $\eta \sim \frac{1}{\text{fan-in}}$ or stronger [33]. The fan-in to a node is defined as the number of nodes connected to that node in the forward direction.

- Momentum α : It is helpful to pick an α close to unity $(0.5 < \alpha < 1)$.

- Initial weights : In order to have a large initial change in weights, $g'(x)$ should be large $(\Delta\omega \propto E' \propto g')$. $g(x)$ is steeper for small x so $\beta \sum \omega_{ij} h_i$ should be small. As a rule of thumb, weights are initially chosen such that $\beta \sum \omega_{ij} h_i \sim 0.1$, hence :

$$\text{initial } \omega \sim \frac{0.1}{\beta(\text{max fan} - \text{in})} \tag{5.20}$$

- Temperature T : A low T corresponds to a steep $g(x)$ and frustrates the weights. A good choice is $T = 1$ which corresponds to a moderate slope of 1 for $g(x)$ at $x = 0$.

- Number of training patterns, N_p : The generalization error is of order $(\frac{N_\omega}{N_p})$. For a good generalization $\frac{N_p}{N_\omega} \geq 10$.

Although the above rules recommend certain values for learning parameters, the best values are determined by the analyst according to the level of generalization ability and convergence speed needed.

Chapter 6

Background and Signal Estimations

This Chapter discusses background and signal samples, selection cuts applied on each sample and the expected contribution of signal and each background to the data sample. The data-set used in this analysis was collected during run 1b, from December 1993 to July 1995, with an integrated luminosity of $74.86 \pm 4.04\ pb^{-1}$.

6.1 Background and Signal Samples

As discussed in Chapter 1, $W + jets$ and QCD multijet events are background to $t\bar{t} \rightarrow e + \bar{\nu}_e + jets$. In this Chapter we simulate each of these using both Monte Carlo techniques and the data.

6.1.1 W+jets

The W+jets sample is generated using the VECBOS Monte Carlo program. The VECBOS generator contains leading order parton-level calculations using exact matrix elements of $W, Z + njet$ processes for $n \leq 4$ [34]. VECBOS calculations are based

on *perturbative QCD* (*PQCD*) and require:

- $E_T > 10$ GeV for all final partons;
- $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} > 0.5$ for every pair of jets;
- CTEQ1M structure functions;
- $< Q^2 >= {m_W}^2$ for dynamical scale.

To estimate the higher order processes ($n > 4$) and fragmentation effects, events generated by VECBOS are passed through ISAJET, the DØGEANT detector simulator (shower library) and the event reconstruction program. VECBOS ignores diagrams beyond leading order. This causes a significant uncertainty in cross section. Therefore, we estimate the $W + jets$ cross section by normalizing the VECBOS generated events with respect to the data. So then we can rely only on the distributions of events generated by VECBOS. The observed distributions agree with the background calculation, which includes contributions from both $W + jets$ as calculated by VECBOS and QCD multijet events [44].

6.1.2 QCD Multijet

The QCD multijet events sample is derived from data. QCD multijet events are background to the $electron + jets$ events where one of the jets is misidentified as an electron and a significant amount of $\not\!\!E_T$ is observed due to mismeasurement. QCD jets fragmented in a highly electromagnetic mode may deceive the electron-id algorithm into misidentifying $\pi^{\pm}$ or π^0's as electrons. Moreover, the photons from isolated

π^0's may convert in the central tracking system and produce electrons. Even though the probability for electron misidentification due to statistical fluctuations is small, the large cross section for multijet events makes this background significant. Since events with high EM fraction jets and $\not\!\!E_T$ are considered further to fake an electron, the QCD multijet background sample is obtained by requiring a "bad" electron in the data sample (an electromagnetic cluster satisfying electron triggers but failing electron-id) convoluted with the probability of faking an electron [37, 38].

6.1.3 Top Monte Carlo Sample

Our top Monte Carlo sample is generated by ISAJET. The output of ISAJET is fed into the DØGEANT detector simulation, trigger simulation and reconstruction routines. To study the systematic error due to event generation, another popular event generator known as HERWIG is used for comparison.

6.2 Offline Cuts on The Samples

The first set of offline cuts that we make on our samples before feeding them into the neural network classifier is listed below. These cuts reject unwanted events and enhance the characteristics of each sample whether it is signal or background.

- Main Ring loss and Microblank [35]

  One of the unusual features of the DØ detector is that two accelerators pass through it! This causes about 25% deadtime. The Main Ring is constantly employed in $\bar{p}$ production to supply the anti-protons which have been ex-

tracted from the stack for $\bar{p}p$ collisions. The major Main Ring loss is due to injection/transition[1]. To blank out the injection/transition period plus calorimeter and muon high voltage recovery time we use two trigger terms :

MRBS_LOSS A gate from 0.1-0.5 sec during the Main Ring full cycle ($\sim$ 2.4 sec) to blank the injection/transition period and allow a recovery time for muon high voltage and calorimeter preamps. This brings about 17% (0.4/2.4) deadtime.

MICROBLANK vetoes events when the Main Ring beam is present at DØ and the muon system is functioning. The muon lifetime for Main Ring particles is $\pm$800 nsec around the $\bar{p}p$ crossing time. Since the calorimeter lifetime is about 2 μsec, MICROBLANK will not be fully efficient for vetoing events with the Main Ring energy in calorimeter. MICROBLANK causes a deadtime of about 7-9%.

- Trigger requirements

  We require **ELE_JET_HIGH** and **EM1_EISTRKCC_MS** trigger terms for our data-set. These triggers enhance $t\bar{t} \rightarrow e + jets$ and $W \rightarrow e + \bar{\nu}_e$ characteristics in the data by applying the following cuts on the data samples :

  ELE_JET_HIGH selects:

[1] *Transition* is the period of time when the particle beam enters the highly relativistic regime where the head of the bunch should be accelerated to keep the bunch together.

- One or more electromagnetic towers with $E_T > 15$ GeV, $|\eta| < 2.5$, tight longitudinal and transverse shape cuts and isolation in a cone with $R = 0.4$;
- Two or more hadronic plus electromagnetic trigger towers with $E_T > 10$ GeV, $|\eta| < 2.5$ and $R = 0.3$;
- $\not\!\!E_T > 14$.

EM1_EISTRKCC_MS selects:

- One or more electromagnetic towers with $E_T > 20$ GeV, isolation in a cone with $R = 0.4$ and a track cut in the CC region;
- $\not\!\!E_T > 15$ GeV .

- Cut on $\not\!\!E_T$

  We require a cut of 25 GeV on $\not\!\!E_T$ for all samples. This cut reduces the QCD multijet background in the data-sets.

- Cut on electrons

  We require a "good" electron in the data sample. Our criteria for a good electron are:

$$|\eta^e| < 2.0$$
$$E_T^e > 20 \text{ GeV}$$
$$f_{isol} \leq 0.1$$
$$L^e < 2.5 \text{ for the CC and } L^e < 3.0 \text{ for the EC}$$

  Since the electron likelihood is not well-modeled by Monte Carlo in the CC region, the electron likelihood cut is not applied to the Monte Carlo for CC electrons. We require a bad electron for the QCD multijet sample.

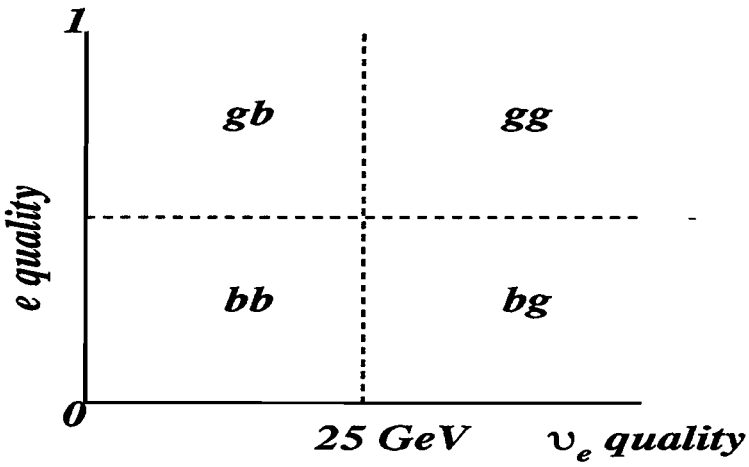


Figure 6.1: Electron and $\not\!\!E_T$ quality plane

- Cut on jets

  For a sample with jet multiplicity N, we require N jets with $E_T > 15$ GeV, $|\eta| < 2.0$ and cone size $R = 0.5$.

6.3 Background and Signal Estimation

To study the discrimination power of the variables and train the neural network, we need to have a realistic background sample with the proportion of $W + jets$ and QCD multijet events we expect to see in the data. We use the following procedure to estimate the proportion of each background and signal. To separate QCD multijet events (non-W background) from electrons and neutrinos coming from real W's, we study the quality of electron and neutrino for each event and divide the events into four different categories (fig 6.1) :

$gg \equiv$ good electron and good neutrino $\equiv$ passed electron id and $\not\!\!E_T > 25$ GeV
$gb \equiv$ good electron and bad neutrino $\equiv$ passed electron id and $\not\!\!E_T \leq 25$ GeV
$bb \equiv$ bad electron and bad neutrino $\equiv$ failed electron id and $\not\!\!E_T \leq 25$ GeV
$bg \equiv$ bad electron and good neutrino $\equiv$ failed electron id and $\not\!\!E_T > 25$ GeV

The total number of events is :

$$N_i = N_i^W + N_i^Q \quad i = gg, gb, bb, bg \tag{6.1}$$

where :

$$\begin{aligned} N_i &\equiv \text{number of data events in quadrant } i \\ N_i^Q &\equiv \text{number of fake W events in quadrant } i \\ N_i^W &\equiv \text{number of real W events in quadrant } i \end{aligned}$$

equation 6.1 can be rewritten as

$$N_i = a_i N_{gg}^W + N_i^Q \text{ where } a_i = \frac{N_i^W}{N_{gg}^W} \tag{6.2}$$

The a_i's are determined from the Monte Carlo and take into account the electron-id efficiency.

To solve for N_{gg}^W and N_{gg}^Q, we need one more constraint on N_i^Q and N_i^W. Since the electron and neutrino in QCD multijet events are not decay products of a real W, we assume that the electron and $\not\!\!E_T$ quality are independent for QCD multijet background. Therefore :

$$\frac{N_{gg}^Q}{N_{bg}^Q} = \frac{N_{gb}^Q}{N_{bb}^Q} \tag{6.3}$$

Using equations 6.2 and 6.3 we can solve for N_{gg}^W and N_{gg}^Q. N_{gg}^W and N_{gg}^Q for each jet multiplicity and their corresponding statistical errors are shown in Table 6.1.

The contribution to real W's comes from either the $W + jets$ background or the top quark. Berend scaling [36] is the basis of our estimation for top events in the data sample. According to Berend scaling:

$$\frac{\text{number } of\ W + \ n\ jets}{\text{number of } W + \ (n-1)\ jets} = \alpha = \text{constant} \quad \text{for } n \leq 5 \tag{6.4}$$

Jet Multiplicity	N_{gg}^{Q}	N_{gg}^{W}
1	236.75 ± 22.44	7185.25 ± 90.00
2	202.71 ± 14.78	1152.29 ± 40.07
3	51.86 ± 7.21	199.14 ± 17.61
4	10.93 ± 3.32	37.07 ± 7.89

Table 6.1: Number of fake and real W's.

In other words, an extra jet in $W+jets$ sample adds an extra strong coupling constant factor to the cross section for samples with low jet multiplicity. So:

$$N_{gg,i}^{W} = N_{i}^{W} + N_{i}^{t} \quad i = 1, ..., 4 \tag{6.5}$$

where:

$$\begin{aligned} N_{gg,i}^{W} &\equiv \text{number of real } W \text{ events with jet multiplicity } i \\ N_{i}^{W} &\equiv \text{number of } W + i \text{ jets events} \\ N_{i}^{t} &\equiv \text{number of top events with jet multiplicity } i \end{aligned}$$

Using (6.4)

$$N_{gg,i}^{W} = \alpha^{i-1} N_{1}^{W} + f_{i} N^{t} \quad i = 1, ..., 4 \tag{6.6}$$

where f_i is the fraction of top events with jet multiplicity i and is derived from the Monte Carlo. We find α, N_1^W and N^t by fitting equation (6.6) to $N_{gg,i}^{W}$ in Table 6.1. The proportion of QCD multijet, W+jets and expected top events for each jet multiplicity is shown in Table 6.2. The number obtained for the amount of top in the data is not used in any calculation in this analysis, however, the proportion of QCD

Jet Multiplicity	N^Q	N^W	N^t
1	236.75 ± 22.44	7168.48 ± 89.66	16.77 ± 17.50
2	202.71 ± 14.78	1136.26 ± 49.44	16.26 ± 16.98
3	51.86 ± 7.21	185.39 ± 15.62	13.75 ± 14.35
4	10.93 ± 3.32	28.02 ± 3.52	9.05 ± 9.45

Table 6.2: Contributions of QCD multijet, W+jets and top.

and W background is used for normalizing the corresponding Monte Carlo samples to the data.

In the next Chapter we study the discrimination power of variables and try to select a set of variables for inclusive 3-jet analysis using the neural network. Higher statistics is the only incentive for requiring 3 jets or more in the data sample. This not only promises lower statistical errors but also improves the performance of the optimization process and training algorithm of the neural network. The selection cuts applied to the samples and their contribution to the 3-jet data-set are summarized in Table 6.3. References [39, 40, 41, 42] were major source of inspiration for methods used for background and signal estimation in this Chapter.

Selection Cut	QCD Multijet	W+jets	Top	Data
Main Ring	√	×	×	√
trigger	√	×	×	√
good e	×	√	√	√
bad e	√	×	×	×
$\not\!\!E_T > 25$ GeV	√	√	√	√
$E_T^j > 15$ GeV, $\|\eta^j\| < 2.0$	√	√	√	√
Proportion(3j)	51.86	185.39	13.75	251

Table 6.3: Selection cuts on the Monte Carlo samples and the data.

Chapter 7

Feature Extraction

In this Chapter, we study the distribution of signal and background events for different variables and choose a set of variables to be fed into the neural network to separate the signal from background. The background sample is a mixture of QCD multijet and W+3jets combined with the proportions estimated for inclusive 3-jet events in the previous Chapter. The signal sample is the top Monte Carlo sample with a top mass of 180 GeV. Theoretically, for an infinite training sample one can use all the variables describing the events without losing generalization ability, but for finite samples the generalization error $\sim \frac{N_\omega}{N_p}$. As discussed in Chapter 3, in order to have a good generalization one should have $\frac{N_\omega}{N_p} \leq 0.1$. The number of patterns in our signal and background is about 2000 events each. Since we apply the 2-fold cross-validation method, half of this sample will be used for training and the other half for testing. Therefore the training sample contains about 1000 events for each of signal and background. To have a good generalization, the number of weights, N_ω, should be about 100. If we use a network with n input nodes and $2n + 1$ hidden

nodes[1], then

$$N_\omega = (n+1)(2n+3), \tag{7.1}$$

hence

$$(n+1)(2n+3) \sim 100 \;\rightarrow\; n \sim 6. \tag{7.2}$$

Therefore, we select up to 6 variables to train the network, without loss of generalization ability. In the next two Sections we focus on two important features of the top events; high mass and event shape.

7.1 H_T

One of the most pronounced characteristics of the top quark is its high mass. Recent top mass estimations by both CDF and DØ at FNAL confirm this assumption. In March 1995, both experiments claimed observation of the top quark, CDF with a mass of 176 ± 13 GeV [43] and DØ with a mass of $199^{+19}_{-21} \pm 22$ [44]. The final hadronic decay products of a heavy top tend to have higher E_T compared to the background. Therefore, H_T, defined below is a good discriminator.

$$H_T = \sum_{|\eta|<2, E_T^j>15} E_T^j \tag{7.3}$$

Figure 7.1 shows the H_T distribution for the top and the background. All distributions are normalized to 1. The y axis represents the relative frequency for an event with a given H_T. Figures 7.2, 7.3, and 7.4 show the E_T distribution for the first three leading jets.

[1]To comply with Kolmogorov's theorem.

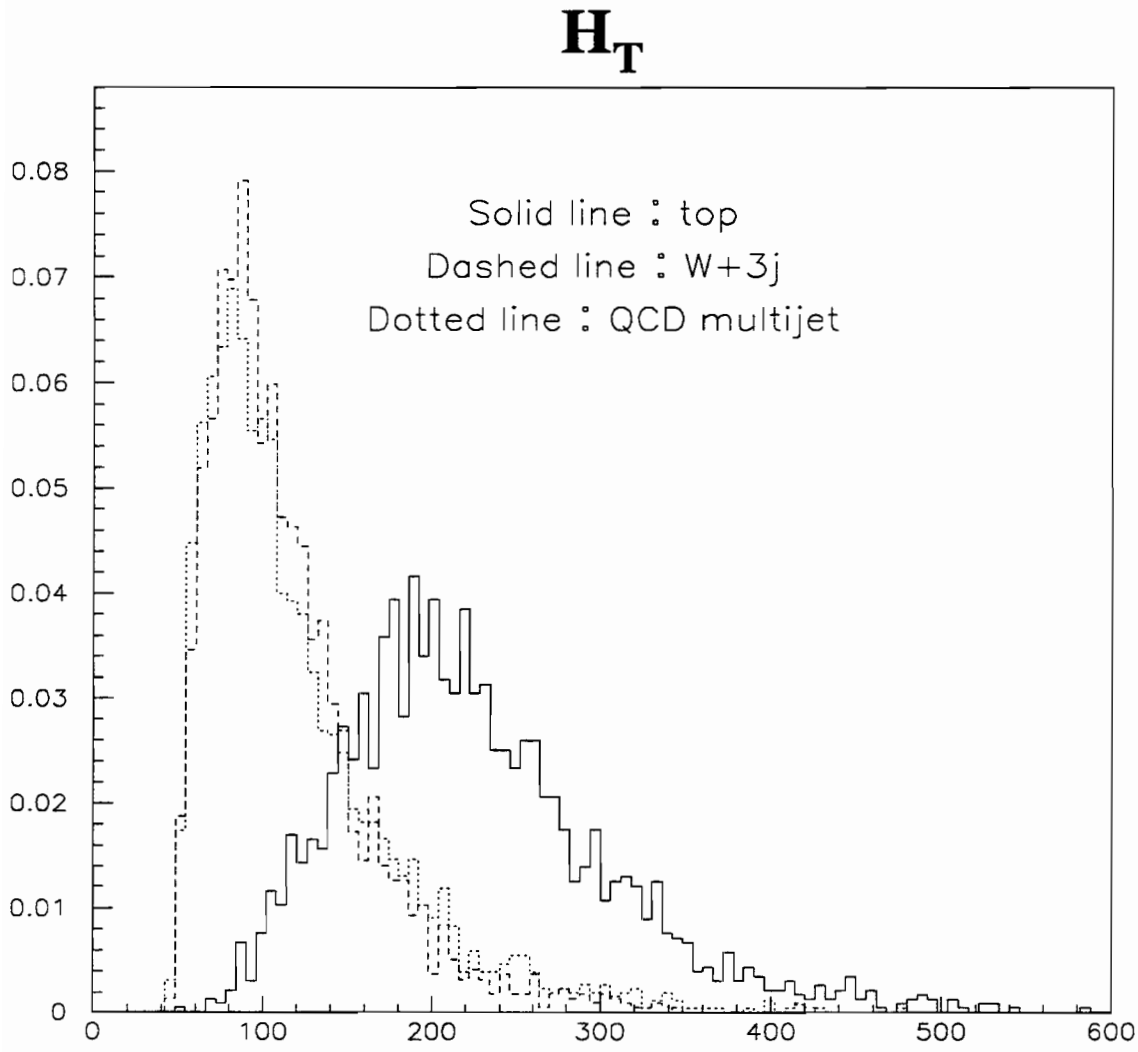


Figure 7.1: H_T distribution for signal and backgrounds

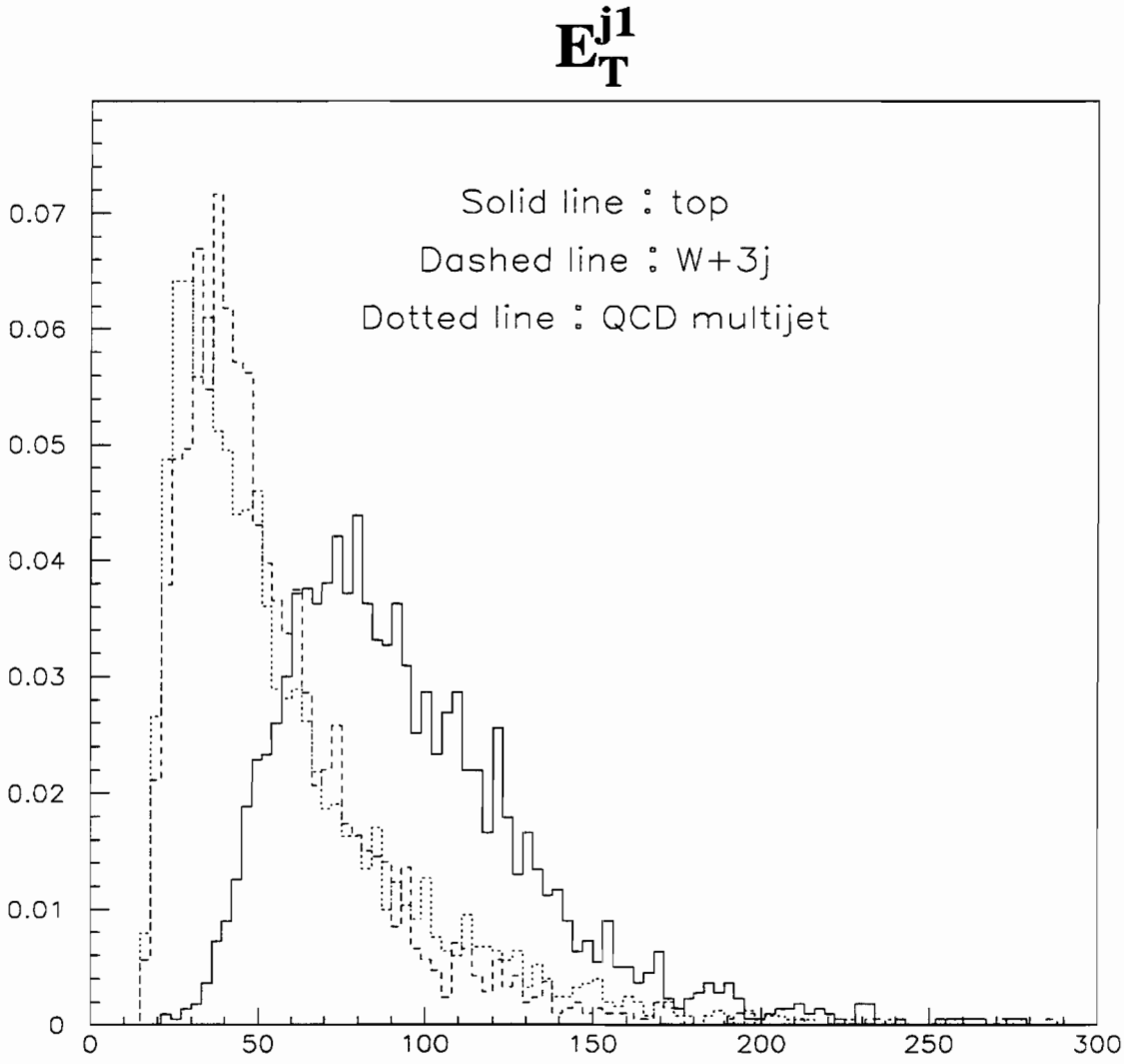


Figure 7.2: E_T^{j1} distribution for signal and backgrounds

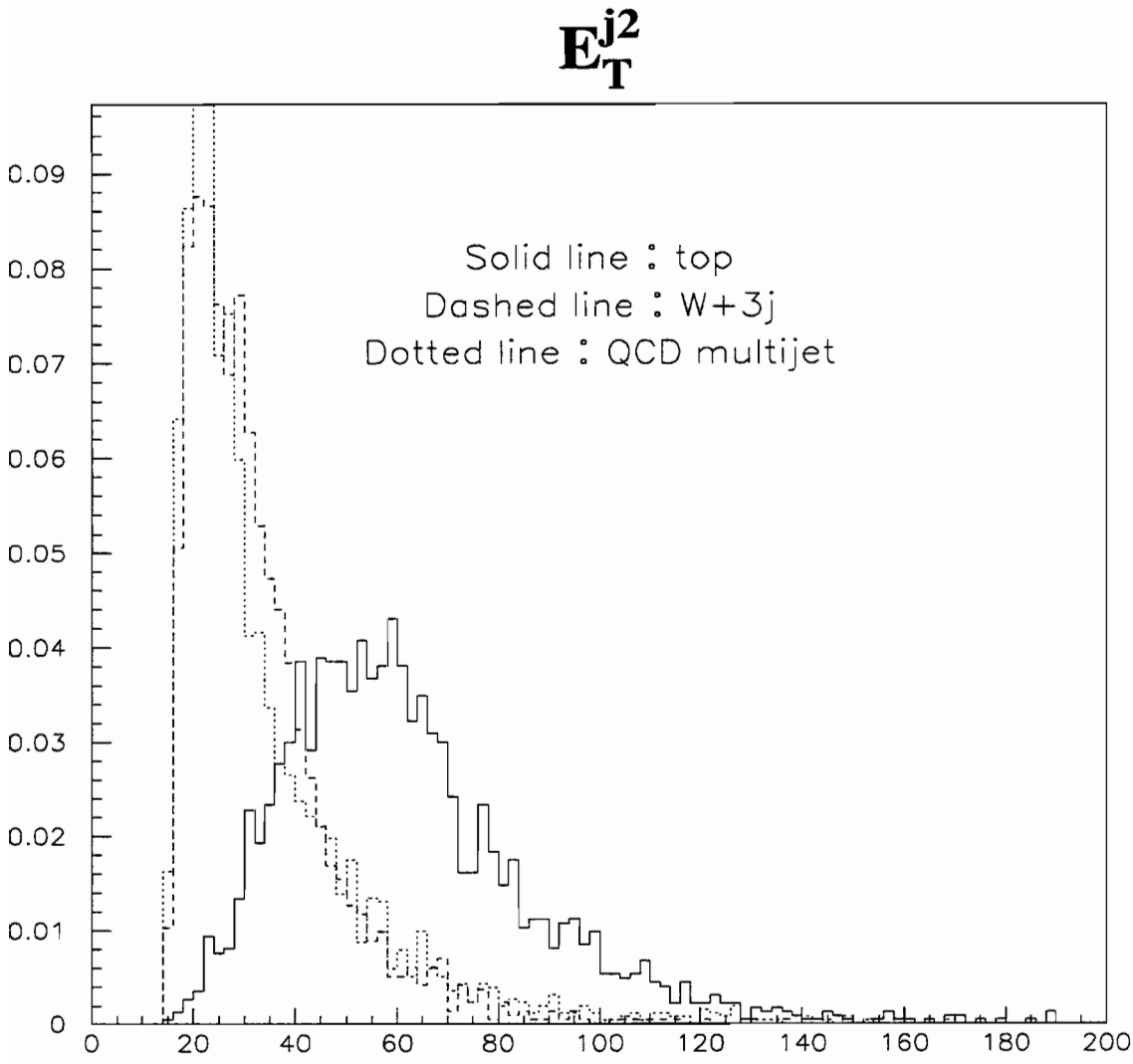


Figure 7.3: $E_T^{j_2}$ distribution for signal and backgrounds

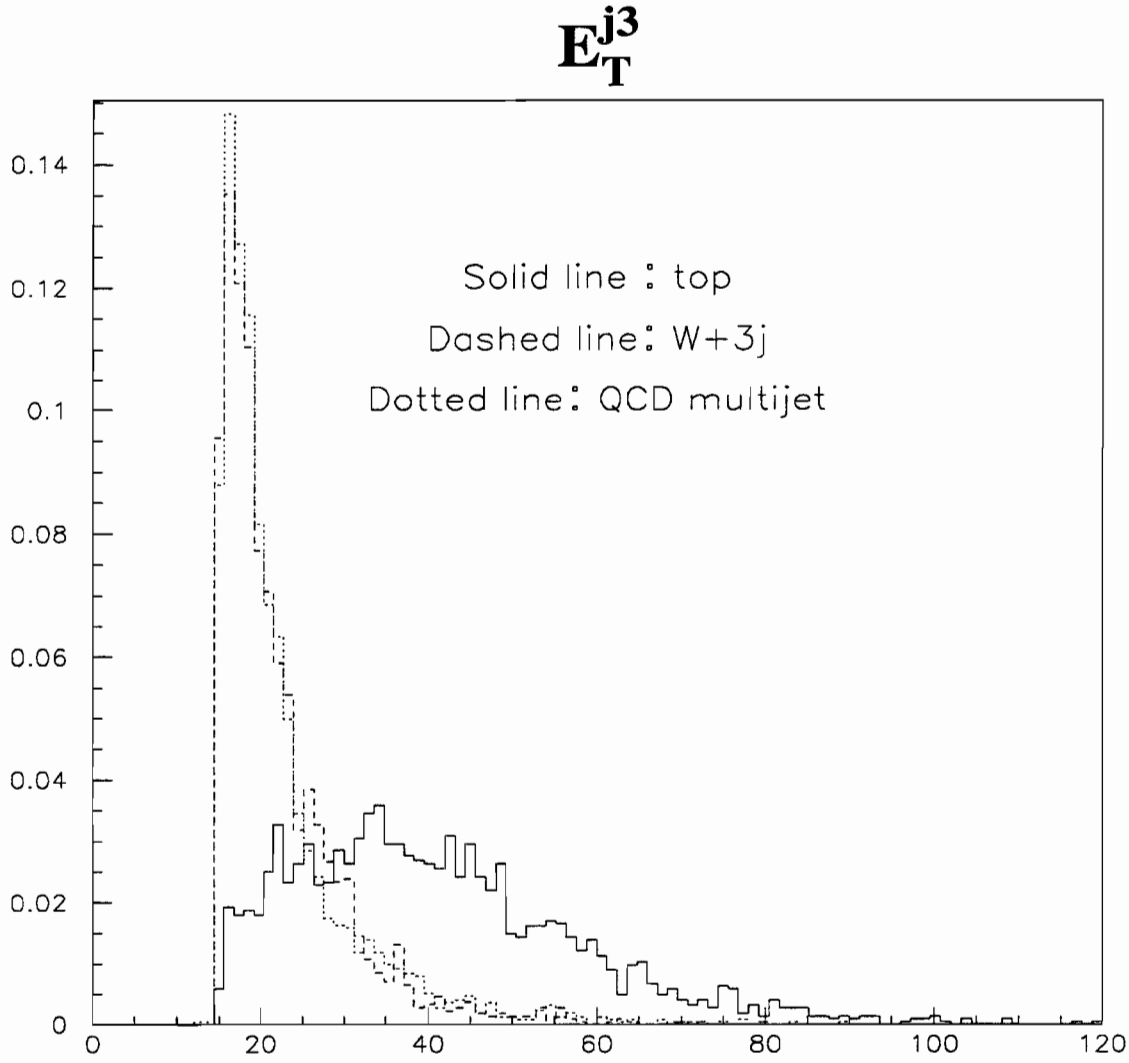


Figure 7.4: E_T^{j3} distribution for signal and backgrounds

7.2 Aplanarity and Sphericity

Another distinctive characteristic of the top quark is the event shape. Monte Carlo studies show that top events tend to be more spherical compared to background events. In order to quantify this observation, the momentum tensor M is defined as :

$$M = \sum_{i=jets} |\vec{p_i}><\vec{p_i}| \tag{7.4}$$

To rely only on the shape of the event rather than its energy we normalize M by its trace:

$$tr(M) = \sum_{j=1}^{3} < j|(\sum_{i=jets} |\vec{p_i}><\vec{p_i}|)|j> = \sum_{i=jets} |\vec{p_i}|^2 \tag{7.5}$$

so we define the normalized momentum tensor $\hat{M}$ as:

$$\hat{M} = \frac{M}{tr(M)} \tag{7.6}$$

The way we constructed $\hat{M}$ can assure us that $\hat{M}$ depends on the shape of the event only and has a small correlation with H_T and E_T of the jets. $\hat{M}$ is symmetric and positive so it has to have 3 real eigenvalues $\lambda_1, \lambda_2, \lambda_3$. Suppose $|\vec{\lambda_i}>$ is the normalized eigenvector corresponding to an eigenvalue λ_i. Therefore :

$$\hat{M}|\vec{\lambda_i}> = \lambda_i|\vec{\lambda_i}> . \tag{7.7}$$

Hence,

$$\lambda_i = <\vec{\lambda_i}|\hat{M}|\vec{\lambda_i}> = \frac{\sum_{i=jets} |<\vec{p_i}|\vec{\lambda_i}>|^2}{\sum_{i=jets} |\vec{p_i}|^2}. \tag{7.8}$$

As a result

$$\lambda_i \geq 0 \quad i = 1, 2, 3. \tag{7.9}$$

Considering the fact that $tr(\hat{M}) = 1$, for an ordered set of λ_i's we have:

$$\lambda_3 \geq \lambda_2 \geq \lambda_1 \geq 0 \ , \quad \lambda_1 + \lambda_2 + \lambda_3 = 1 \tag{7.10}$$

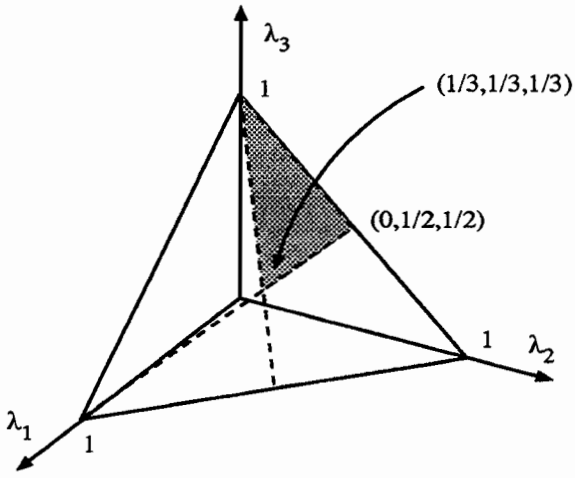


Figure 7.5: Allowed values for λ_1, λ_2, and λ_3.

The shaded area in Figure 7.5 shows the allowed values of the λ_i. Since λ_3 is not independent of λ_1 and λ_2, we can find the range of the λ_i by looking at the projection of the shaded triangle in Figure 7.5 on the λ_1-λ_2 plane (fig 7.6) :

- $0 \leq \lambda_1 \leq \frac{1}{3}$
- $0 \leq \lambda_2 \leq \frac{1}{2}$
- $\frac{1}{3} \leq \lambda_3 \leq 1$

Each corner of the shaded triangle corresponds to a degeneracy in $\hat{M}$, which in turn corresponds to a distinct event shape.

- Linear : $\lambda_1 = \lambda_2 = 0$ and $\lambda_3 = 1$
- Planar : $\lambda_1 = 0$ and $\lambda_2 = \lambda_3 = \frac{1}{2}$
- Spherical : $\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{3}$

Aplanarity and *Sphericity* are defined as :

$$A = \frac{3}{2}\lambda_1, \quad 0 \leq A \leq \frac{1}{2} \tag{7.11}$$

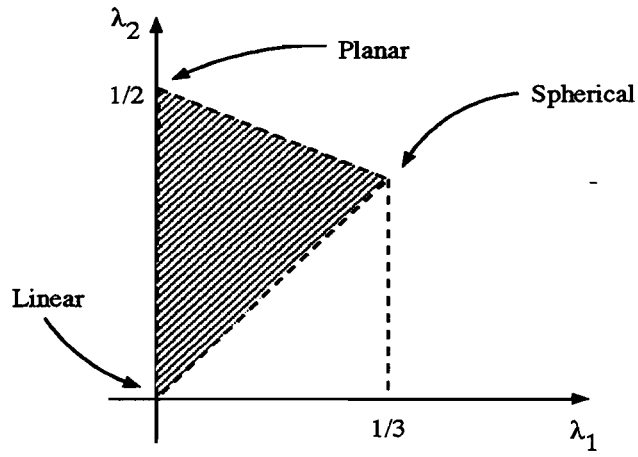


Figure 7.6: Allowed values for λ_1 and λ_2.

$$S = \frac{3}{2}(\lambda_1 + \lambda_2), \quad 0 \leq S \leq 1 \tag{7.12}$$

As shown in Figures 7.7 and 7.8, top events tend to have higher aplanarity and sphericity as compared to background events. This can also be seen in the λ_1-λ_2 plane where background events tend to be far from the spherical corner of the triangle and closer to the linear corner see(fig 7.6). The discrimination provided by aplanarity and sphericity is not as good as that of H_T and E_T of the jets. However, the event shape variables can improve the discrimination when they are used along with the energy of the jets because of their independence from the energy of the jets.

7.3 E_T of electrons and neutrinos

Electrons and neutrinos coming from real W's have higher E_T as compared to electrons and neutrinos from W's faked by a QCD multijet. So, any cut on $\not\!\!E_T$ or E_T^e will further suppress the QCD multijet background. Figures 7.10 and 7.11 illustrate this fact.

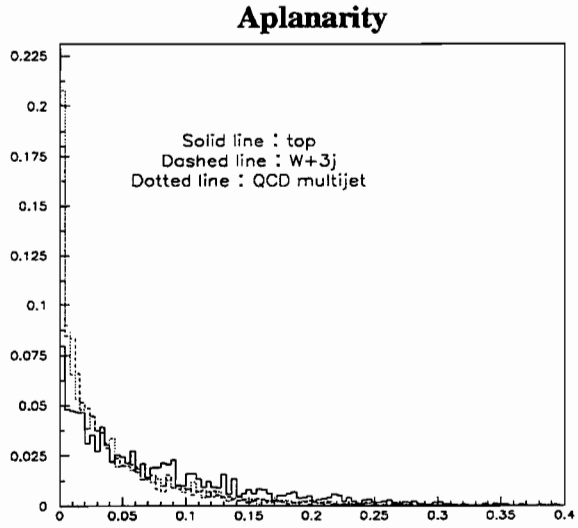


Figure 7.7: Aplanarity distribution for signal and backgrounds

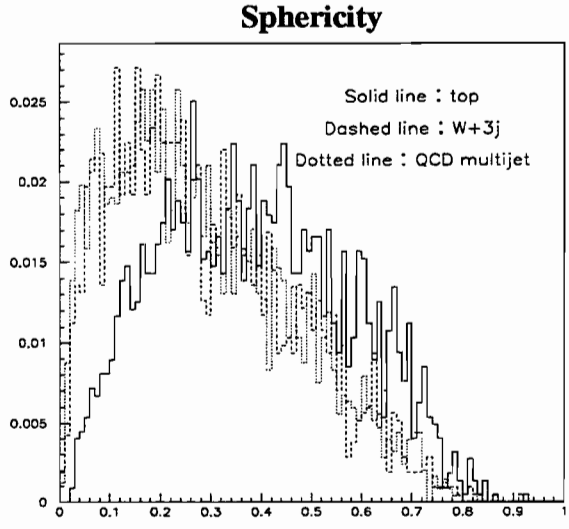


Figure 7.8: Sphericity distribution for signal and backgrounds

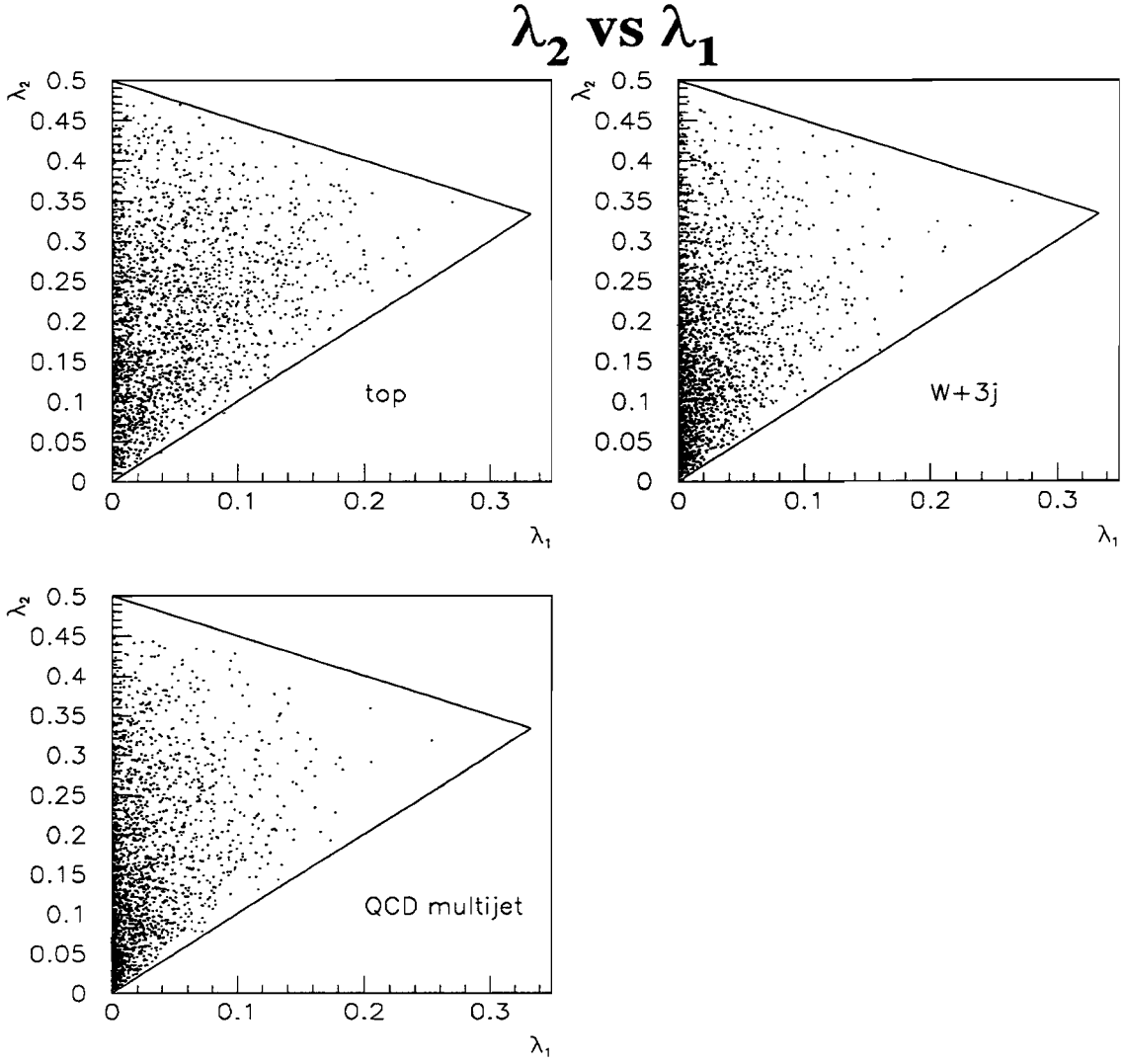


Figure 7.9: Distribution of events in λ_1-λ_2 plane for signal and backgrounds. Plots are normalized to the same number of events.

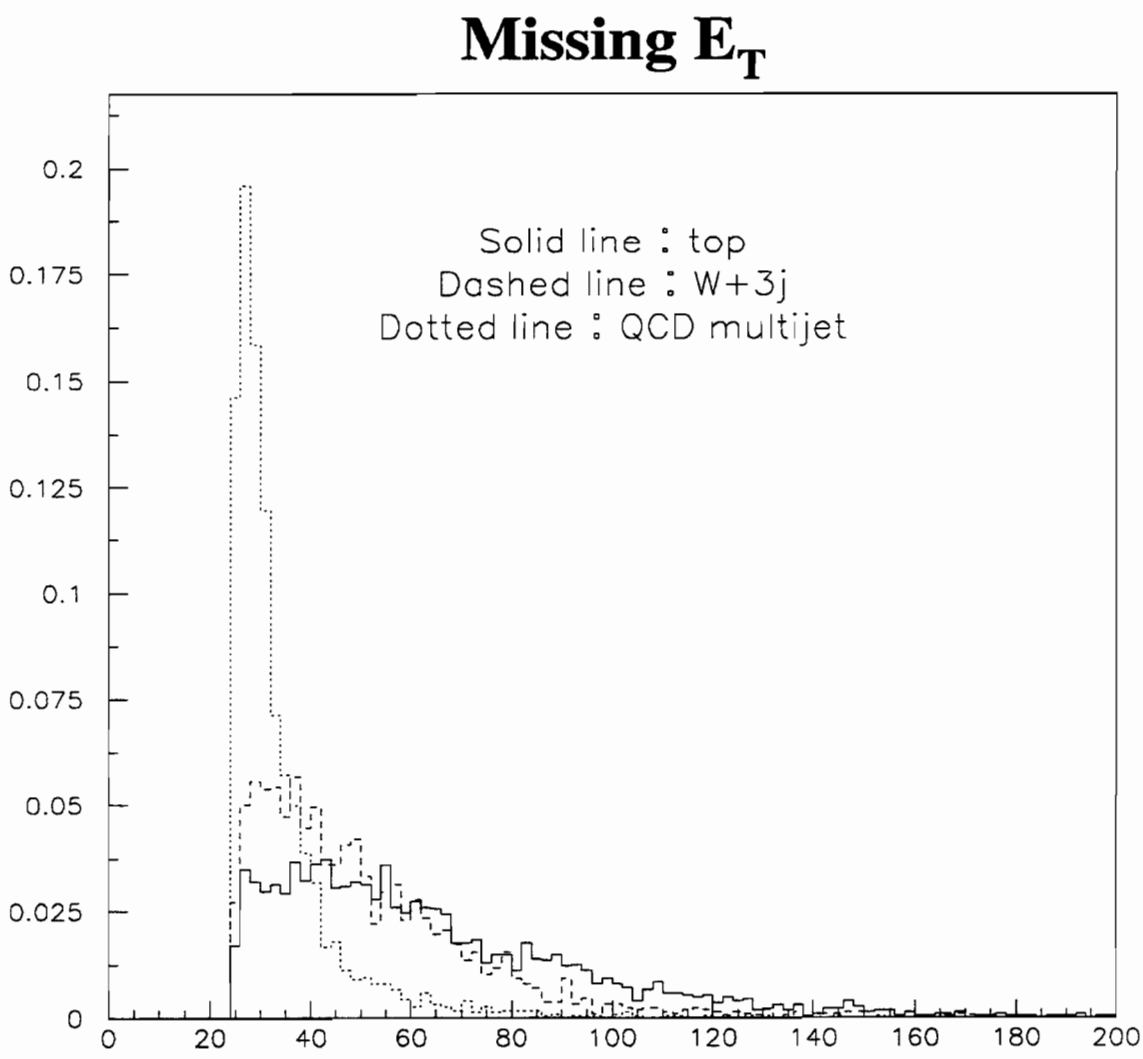


Figure 7.10: $\not\!\!E_T$ distribution for the signal and backgrounds.

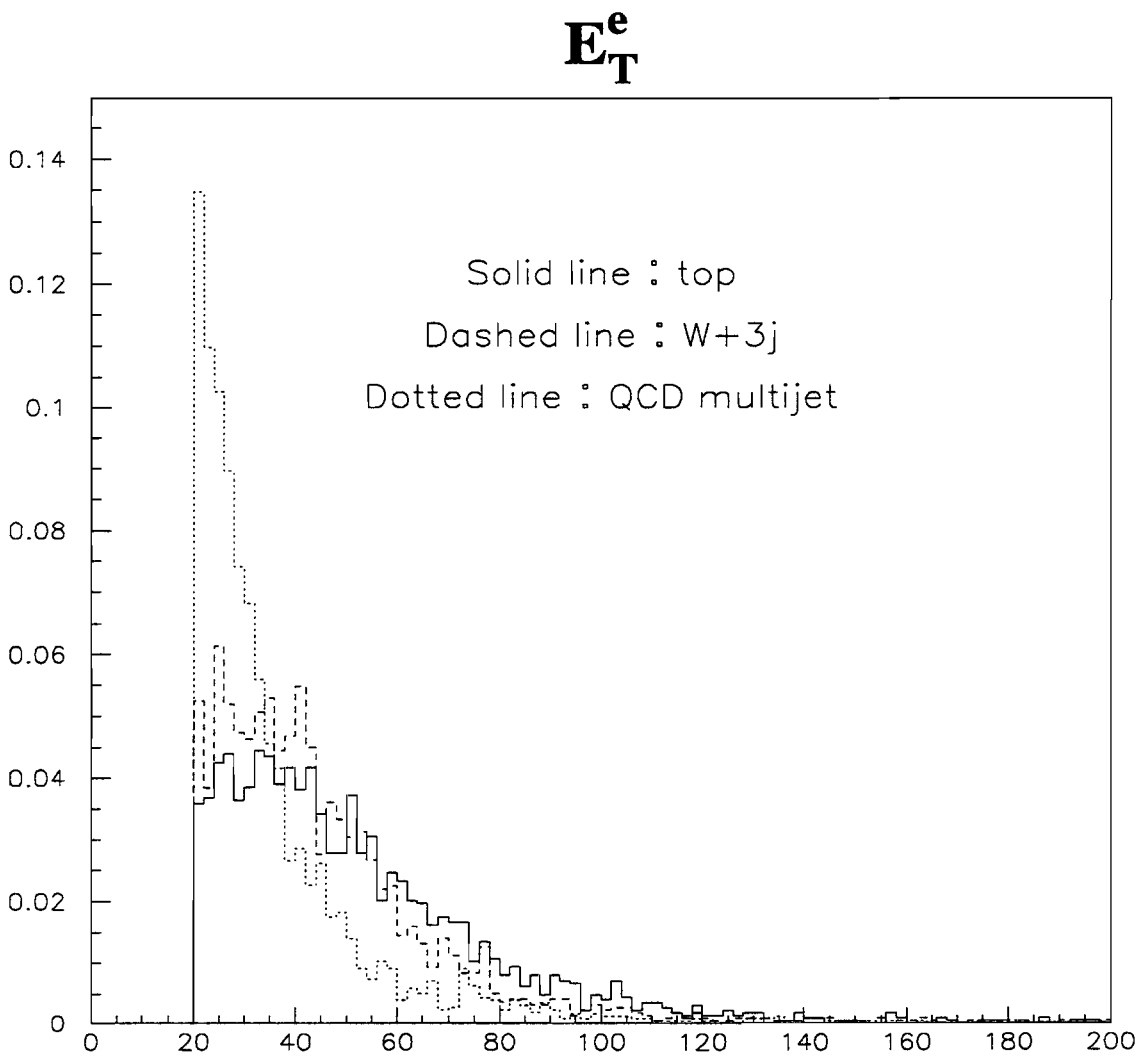


Figure 7.11: E_T^e distribution for the signal and backgrounds.

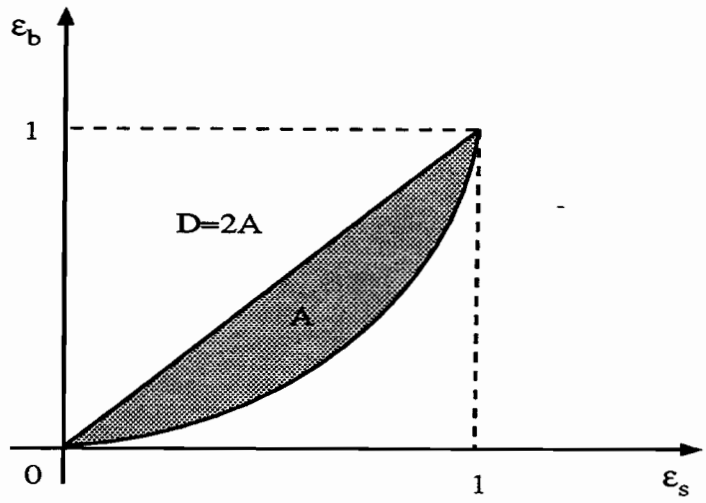


Figure 7.12: Discrimination

7.4 Discrimination Power

In order to rank the variables according to their discrimination power between signal and background one can study the efficiency for the signal and the background for various cuts on the variables. If ϵ_s and ϵ_b represent the efficiency for signal and background, respectively, then variables with a higher positive curvature for ϵ_b as a function of ϵ_s give a better discrimination. Figure 7.13 shows ϵ_b versus ϵ_s for some of the variables discussed in this Chapter. The background used for calculating ϵ_b is a combination of $W + 3jets$ and QCD multijet with the proportions obtained in the previous Chapter.

One can quantify the discrimination power using ϵ_s-ϵ_b plots. A good candidate is,

$$D(x) = 2 \times (\text{area between } \epsilon_b(\epsilon_s) \text{ and the line } \epsilon_b = \epsilon_s) \tag{7.13}$$

Figure 7.12 shows the area mentioned in the above definition. The range of values for D is $[0, 1]$. $D = 0$ occurs when the distribution of the background and the signal for the variable x coincide, and $D = 1$ when signal and background are completely

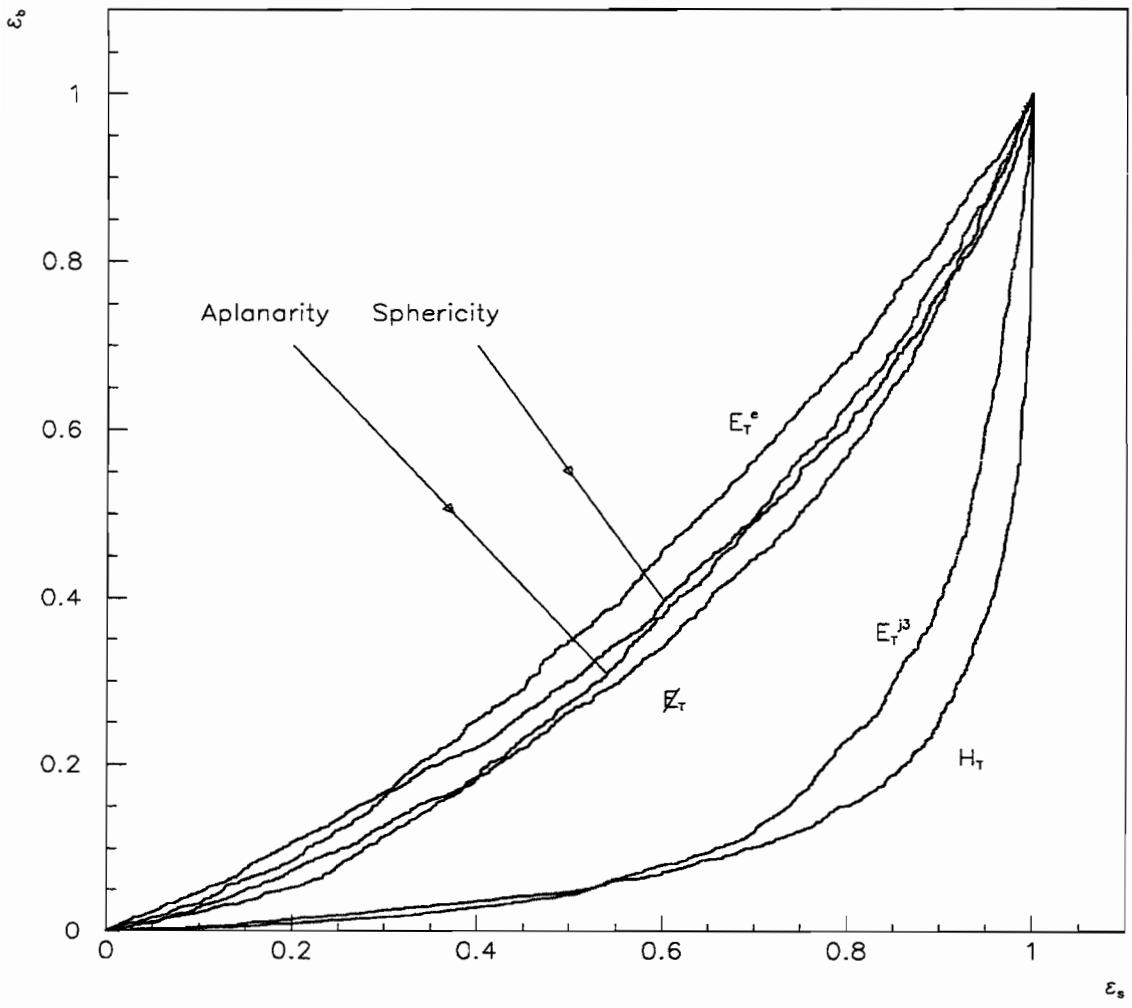


Figure 7.13: Contamination as a function of efficiency.

separated[2] in x. An analytical expression for $D(x)$ is

$$D(x) = 2\int_0^1 (\epsilon_s - \epsilon_b) d\epsilon_s \tag{7.14}$$

As shown in the Figure 7.13 the variables ordered in descending discrimination power are H_T, $E_T^{j_3}$, $\not{E}_T$, S, A, and E_T^e. This ranking is according to the discrimination power of individual variables. The discrimination power of the whole set, considering the correlations among variables, will be discussed in the next Chapter where these six variables are fed into a neural network to separate the signal from background.

[2]It is assumed that ϵ_b as a function of ϵ_s is concave ($\frac{d^2\epsilon_b}{d\epsilon_s^2} \geq 0$) *i.e.* the cut on the variable x favors signal.

Chapter 8

Event Classification and Conclusion

In this Chapter we use the six variables selected in the previous Chapter, H_T, E_T^{j3}, $\not\!\!E_T$, A, S, and E_T^e, to train a neural network. Then the trained network is used to classify the events in the data as signal and background. The top cross section is calculated based on this classification and various sources of systematic errors are studied.

8.1 Training

The network architecture chosen consists of 6 input nodes, 13 hidden nodes, and 1 output node. This architecture contains 105($= 7 \times 15$) weights to be fitted. The training and the testing samples contain 1116 events each for signal and background. The network is trained on one of the samples while the error on the other sample is monitored. The learning rate is chosen to be $\eta = 0.02$ and the momentum $\alpha = 0.5$ with pruning on. The 2-fold cross-validation method is used to avoid overtraining. Figures 8.1 and 8.2 show the error on the testing and training samples as the number of training cycles increases. As shown in Figure 8.1, after 700 cycles the network

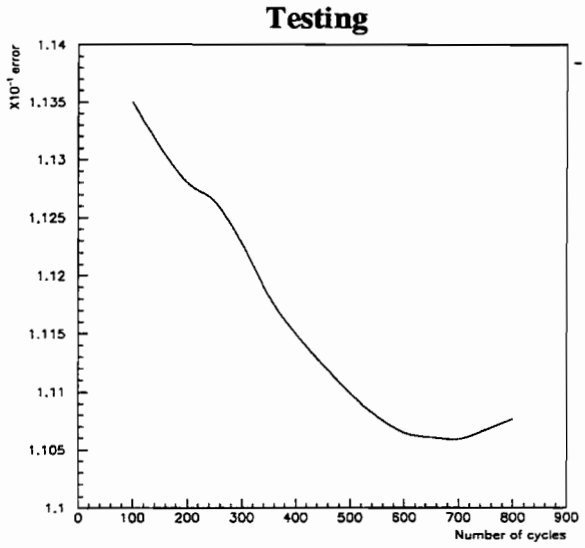


Figure 8.1: Error for testing sample

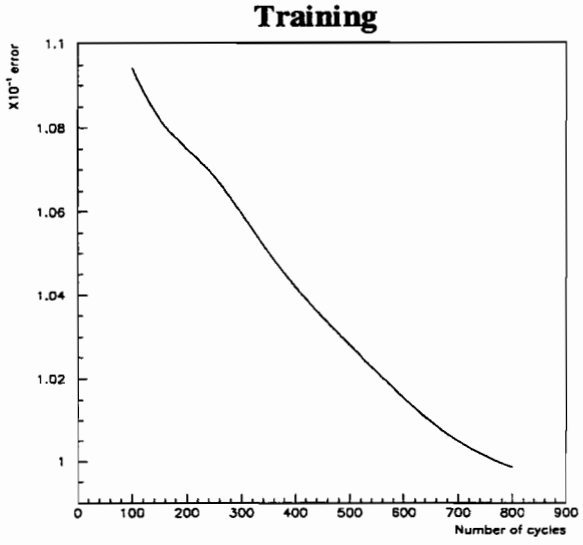


Figure 8.2: Error for training sample

starts taking advantage of the artifacts of the training sample to decrease the error. Therefore, the training is stopped at 700 cycles and the weights are frozen.

Next we treat the network output, P_{out}, as a random variable (like H_T, A, *etc*). P_{out} can be considered as a new variable which is a nonlinear function of H_T, E_T^{j3}, $\not\!\!E_T$, A, S, and E_T^e:

$$P_{out} = P_{out}(H_T, E_T^{j3}, \not\!\!E_T, A, S, E_T^e) \tag{8.1}$$

The analytic functional form of P_{out} in terms of H_T, E_T^{j3}, $\not\!\!E_T$, A, S, and E_T^e is given in Appendix A. One may also think of P_{out} as a probability function $P_{out} : R^6 \rightarrow [0,1]$ for being the top quark. Any cut on P_{out} corresponds to a 5-dimensional surface which divides the 6-dimensional space of (H_T,E_T^{j3},$\not\!\!E_T$,A,S,E_T^e) into signal and background regions. Treating P_{out} as a new variable, we can study its discrimination power, ϵ_s, ϵ_b, signal-to-background ratio S/B, *etc.* for various cuts between 0 and 1. Figure 8.3 shows the distribution of signal and backgrounds for P_{out}. All distributions are normalized to the same number of events. In Figure 8.4 the ϵ_b-ϵ_s curve for P_{out} is compared with that of H_T (as discussed in the previous Chapter H_T is the best single discriminator among the 6 variables used). Figure 8.5 shows that for $\epsilon_s < 0.7$ (where we will make the final cut), P_{out} achieves to suppress the background significantly by combining other variables with H_T. The S/B as a function of (total efficiency) $\times$ (branching ratio), $\epsilon_{tot}Br$, is shown in Figure 8.6. As calculated in Chapter 1, $Br = 14.8\%$ for $e + jets + \not\!\!E_T$ channel.

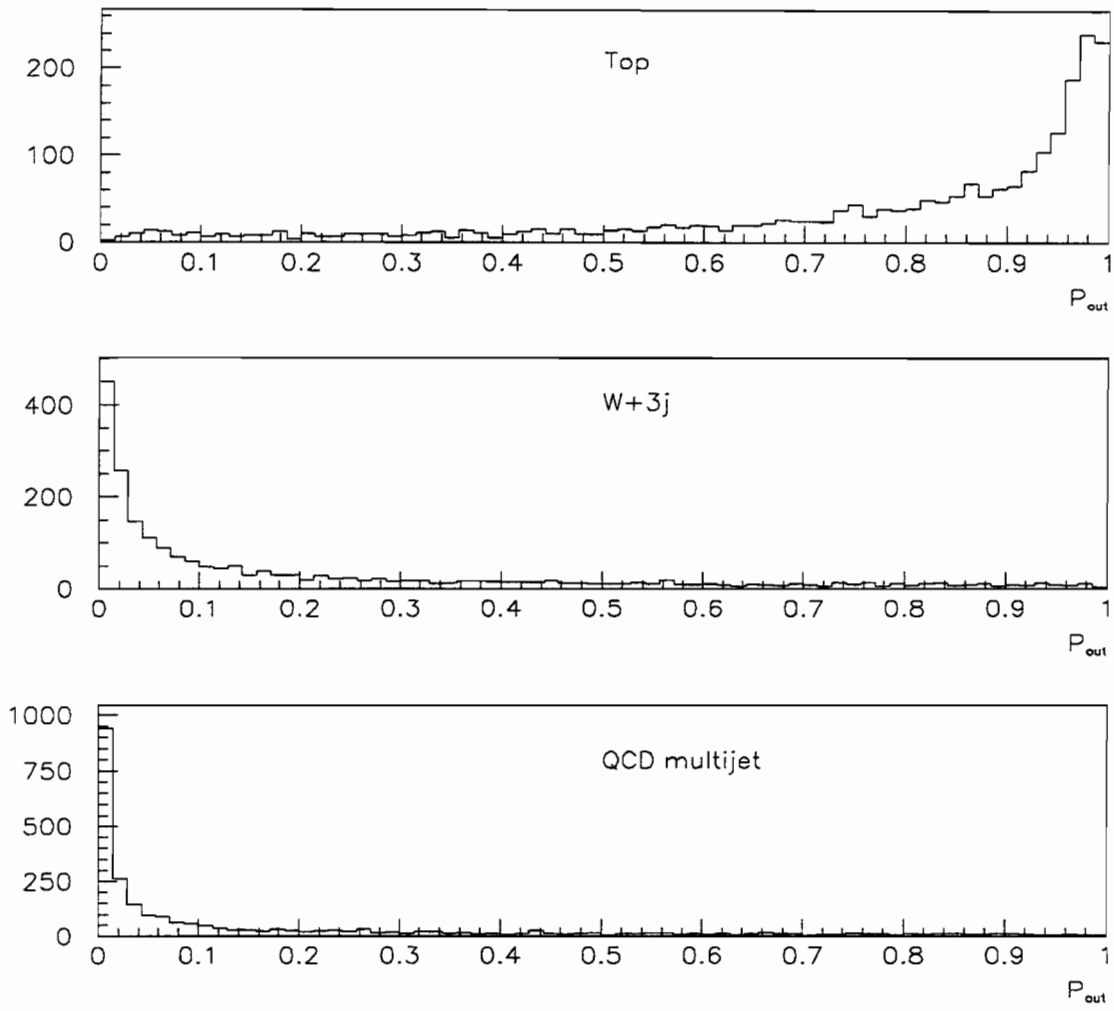


Figure 8.3: Neural network output for the signal and backgrounds

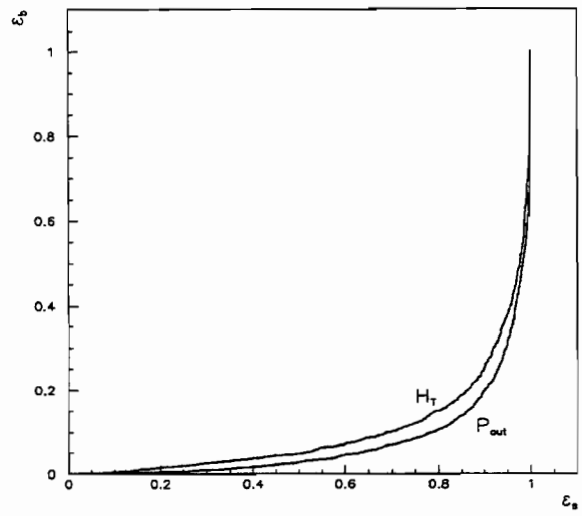


Figure 8.4: ϵ_s-ϵ_b curve for H_T and the network output.

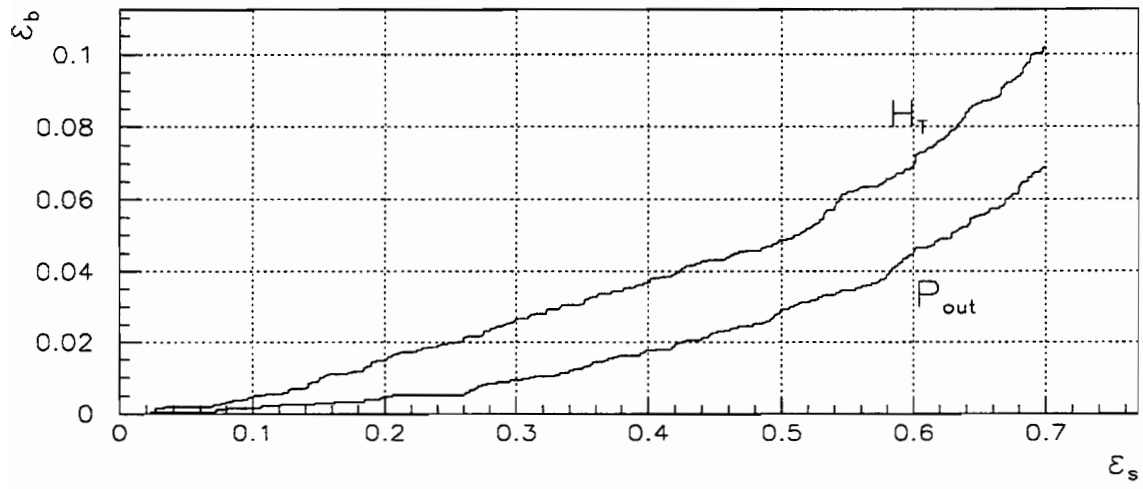


Figure 8.5: ϵ_s-ϵ_b curve for H_T and the network output. The plot is magnified for the low efficiency region of Figure 8.4.

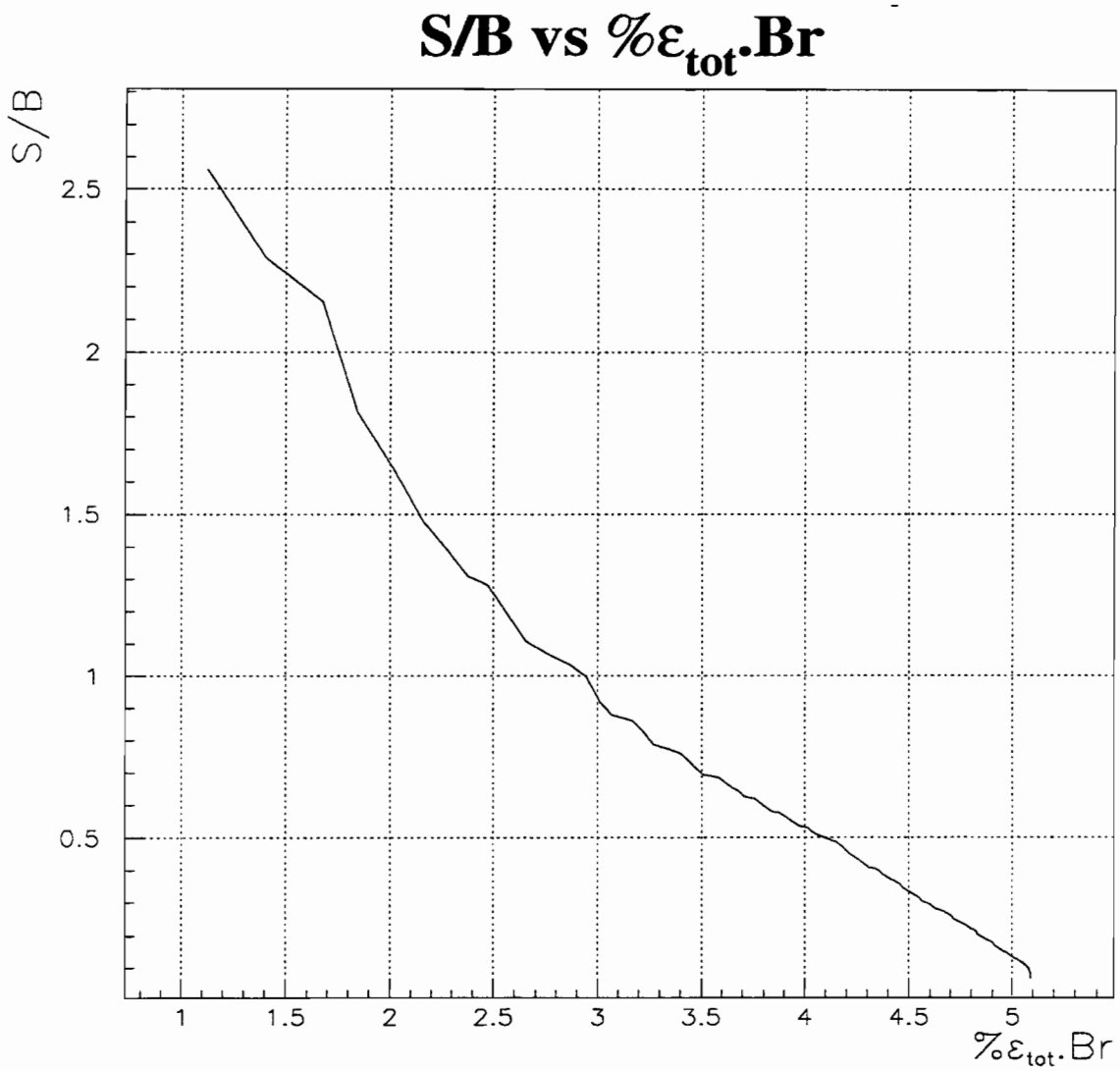


Figure 8.6: Ratio of the expected signal to expected background as a function of total efficiency multiplied by $\bar{t}t \rightarrow e + jets + \not\!\!E_T$ branching ratio.

8.2 Decision

The network trained in the previous Section provides an infinite set of optimized 6-dimensional cuts, each corresponding to a cut on P_{out}. In order to best classify the events observed in the data into top and background events, the cut which minimizes the misclassification probability or the risk function is selected. Since we use P_{out} as a random variable, the risk function discussed in Chapter 4 is ($x \rightarrow P_{out}$):

$$Risk(P_{out}) = \int_S [p(P_{out}|wj)P(wj)+p(P_{out}|qcd)P(qcd)]dP_{out}+\int_B p(P_{out}|top)P(top)dP_{out}. \tag{8.2}$$

Replacing

$$\begin{aligned} \int_B p(P_{out}|top) &\rightarrow (1-\epsilon_{top}(P_{out})) \\ \int_S p(P_{out}|wj) &\rightarrow \epsilon_{wj}(P_{out}) \\ \int_S p(P_{out}|qcd) &\rightarrow \epsilon_{qcd}(P_{out}) \end{aligned}$$

we have:

$$Risk(P_{out}) = \epsilon_{wj}(P_{out})P(wj) + \epsilon_{qcd}(P_{out})P(qcd) + (1-\epsilon_{top}(P_{out}))P(top), \tag{8.3}$$

where $P(wj)$ and $P(qcd)$ are the expected proportions of $W+3jets$ and QCD multijet backgrounds in the data. $P(top)$ is derived from Monte Carlo. Figure 8.7 shows the risk as a function of P_{out}. Figure 8.8 focuses on a part of Figure 8.7 where the risk function approaches the minimum. The risk function is minimal at $P_{out} = 0.95$ and the minimum risk is 0.052. Therefore, the cut $P_{out} \geq 0.95$ assures us that our classification of events in the data as top quark and background is 94.8% correct. In other words, on average only 5.2% of the time a background event may be misclassified as

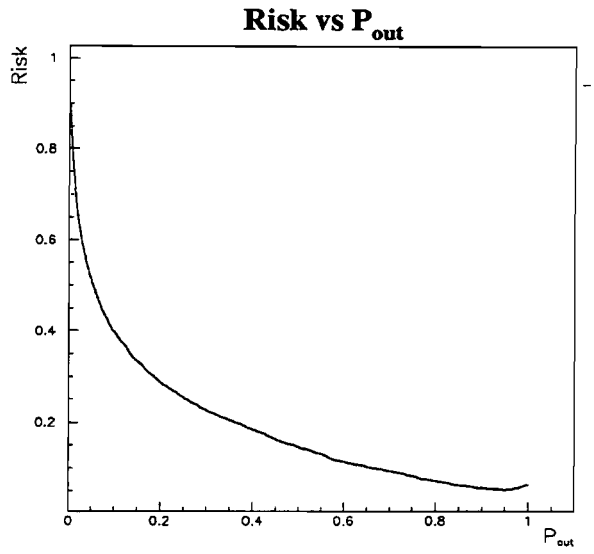


Figure 8.7: Risk or misclassification probability as a function of the network output.

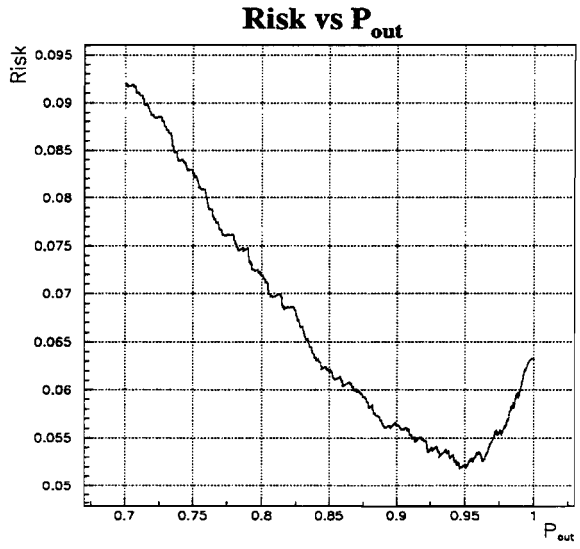


Figure 8.8: Risk or misclassification probability as a function of the network output. The plot is magnified for the low efficiency region of Figure 8.7.

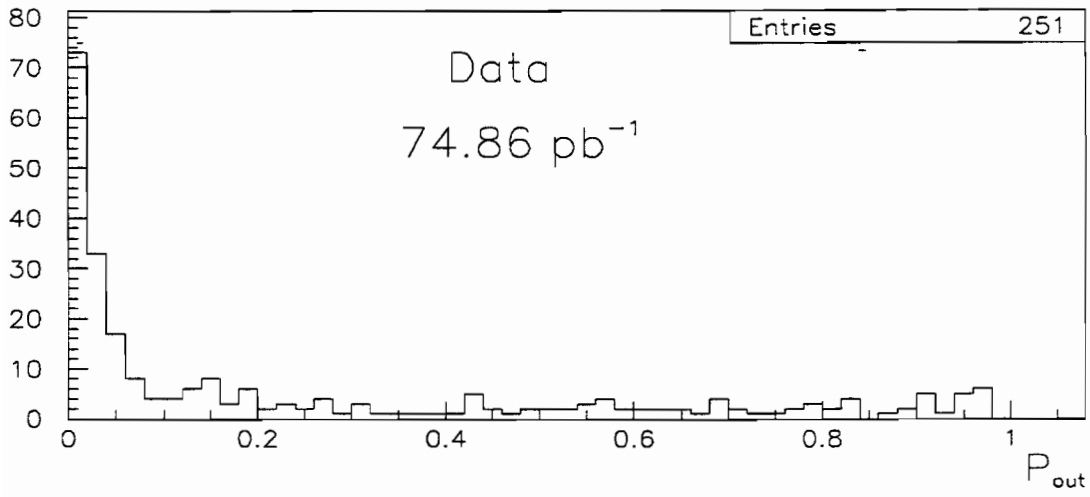


Figure 8.9: Network output for the data events.

top event or a top event mistaken as background.

Now, we apply the cut $P_{out} \geq 0.95$ to the data. The P_{out} distribution for the data is shown in Figure 8.9. The number of observed events in the data, N_{obs}, the expected number of signal and background events, S_{exp} and B_{exp}, $\epsilon_{tot}Br$ and the cross section are listed below. The error on the quantities are statistical.

$P_{out} \geq 0.95$

$p(error) = Risk = 5.2\%$

$N_{obs} = 10$

$S_{exp} = 5.44 \pm 0.20$

$B_{exp} = 2.51 \pm 0.43$ ($B_{exp}^{wj} = 2.08 \pm 0.42, B_{exp}^{qcd} = 0.43 \pm 0.09$)

$\%\epsilon_{tot}.Br = 1.73 \pm 0.06$

$\sigma = 5.78 \pm 2.47\ pb$

8.3 Systematic Errors

The following systematic errors on the efficiency and cross section have been studied[1]:

- The systematic error on the integrated luminosity is 5.4% [20];
- The systematic error on efficiencies due to the jet energy corrections is 4.4% for the top Monte Carlo and 13.8% for VECBOS;
- The systematic error on the top efficiency due to the Monte Carlo generator is estimated to be 5.8% (comparing ISAJET and HERWIG);
- A systematic error of 2.9% has been assigned to the application of berend scaling by looking at the deviation of the fitted values from the berend scaling values.

Considering the above systematic effects, the cross section and the efficiency for top production with a mass of 180 GeV are:

$$\%\epsilon_{tot}.Br = 1.73 \pm 0.06(stat) \pm 0.16(sys)$$

$$\sigma = 5.78 \pm 2.47(stat) \pm 0.59(sys)\ pb$$

[1]There is no systematics associated with the neural networks method because the network output has been treated as a random variable which is a function of H_T, E_T^{j3}, $\not\!\!E_T$, A, S, and E_T^e similar to H_T being a function of E_T of the jets. Any change to this functional form will result in a new variable.

N_{obs}		10
B_{exp}		$2.51 \pm 0.43(stat) \pm 0.35(sys)$
Significance		3.3
S_{exp}	top 160 GeV	$8.14 \pm 0.34(stat) \pm 1.77(sys)$
	top 180 GeV	$5.44 \pm 0.20(stat) \pm 0.58(sys)$
	top 200 GeV	$4.11 \pm 0.13(stat) \pm 0.89(sys)$
$\%\epsilon Br$	top 160 GeV	$1.33 \pm 0.05(stat) \pm 0.28(sys)$
	top 180 GeV	$1.73 \pm 0.06(stat) \pm 0.16(sys)$
	top 200 GeV	$2.43 \pm 0.07(stat) \pm 0.52(sys)$
σ (pb)	top 160 GeV	$7.52 \pm 3.22(stat) \pm 1.69(sys)$
	top 180 GeV	$5.78 \pm 2.47(stat) \pm 0.59(sys)$
	top 200 GeV	$4.12 \pm 1.76(stat) \pm 0.91(sys)$

Table 8.1: Summary of the event classification using neural networks.

8.4 Conclusion

The result of applying the neural network cut on the top Monte Carlo with masses 200 GeV and 160 GeV is summarized in Table 8.1 along with that of 180 GeV. The observation of 10 events shows a significant excess of events over the estimated background of $B_{exp} = 2.51$. The probability that a normally distributed background, with a mean $B_{exp} = 2.51$ and variance $\Delta B_{exp} = 0.55$ $(0.43 \oplus 0.35)$, could have fluctuated above 10 according to a Poisson distribution to bring about the observation $N_{obs} = 10$ is 7×10^{-4} which yields a significance of 3.3. The theoretical cross section for top quark production in $p\bar{p}$ collision has been calculated up to $O(\alpha_s^3)$ by Laenen

m_t (GeV)	Laenen *et al.* (pb)	Berger *et al.* (pb)	experiment (pb)
160	$8.16\pm^{1.37}_{0.75}$	$9.01\pm^{0.13}_{0.68}$	7.52 ± 3.64
180	$4.21\pm^{0.57}_{0.35}$	$4.71\pm^{0.07}_{0.35}$	5.78 ± 2.54
200	$2.26\pm^{0.26}_{0.17}$	$2.57\pm^{0.03}_{0.20}$	4.12 ± 1.98

Table 8.2: The calculated and measured top cross section.

et al. and Berger *et al.* [45, 46]. The difference between these two calculations is due to different approaches in setting infra-red cutoffs[2]. The calculated cross section for m_t =160, 180, 200 GeV and corresponding measured cross sections are shown in Table 8.2. The uncertainty in theoretical calculations is due to the change in renormalization scale[3], μ. The lower limits correspond to $\mu = \frac{m_t}{2}$, central value to $\mu = m_t$, and upper limit to $\mu = 2m_t$. Since $\frac{\mu}{m_t}$ is a dimensionless scale in the renormalization group equation, the range $\frac{\mu}{m_t} \in \{\frac{1}{2}, 2\}$ is considered to be a reasonable measure of the theoretical perturbative uncertainty. The variation of the strong coupling strength , α_s, is about $\pm 10\%$ in this range. In Figure 8.10 the measured values for the cross section are compared with the central values of theoretical perturbative calculations.

[2]In both the $q\bar{q}$ and the $g\bar{g}$ channels, the contribution of the $O(\alpha_s^3)$ terms to partonic cross section is much larger than that of $O(\alpha_s^2)$ counterparts in kinematic region of small η where $\eta = \hat{s}/4m_t^2 - 1$ and $\hat{s}$ is the square of the energy of parton-parton subprocess. Therefore the notion, that higher order terms in the perturbation series should be smaller, is not valid at small η. The significant contribution of this near-threshold region makes this region of phase space important for top production at the Tevatron.

[3]The physical quantities are independent of renormalization scale μ. Unfortunately, however, S-matrix can be calculated only through a perturbation series which is truncated after first few orders. This makes the physical predictions μ-dependent.

One surmises that, within errors, the calculated cross section tend to be more consistent with the measured values in the mass range 160-180 GeV than the mass range 180-200 GeV.

Comparison of the DØ published results in the e+jets channel [44] with the neural network results indicates that for all three masses, 160, 180, and 200 GeV the signal efficiency provided by the neural network is higher than that of the conventional method for the same signal-to-background ratio, and, for the same signal efficiency the neural network achieves a higher signal-to-background ratio as compared to the conventional method. For example, for top with a mass of 180 GeV, the signal efficiency provided by the neural network is 46% higher than that of the conventional cut for the same signal-to-background ratio, and, for the same efficiency, the signal-to-background ratio for the neural network cut is 34% higher than this ratio for the conventional cut. It is conclusive that the neural network method is far superior to the conventional method of event classification.

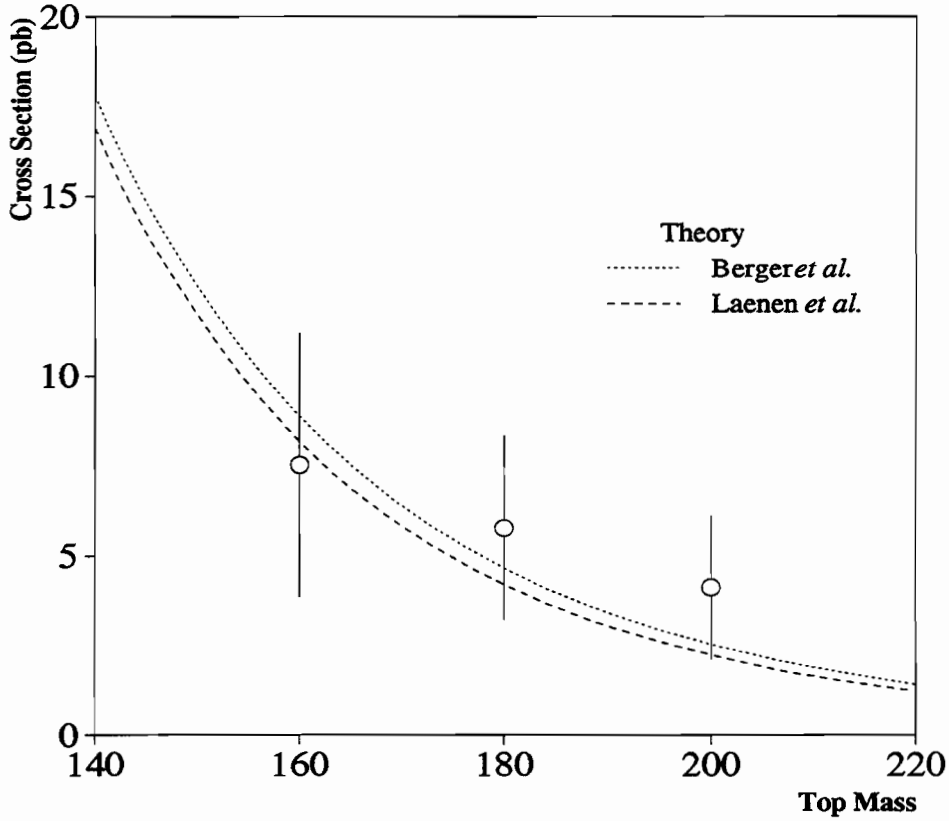


Figure 8.10: The measured cross sections as compared to theoretical perturbative calculations.

Bibliography

[1] Particle Data Group. *Phys. Rev.*, D50(3), August 1994.

[2] T.Kinoshita, *J.Math.Phys.* 3, 650, 1962. T. D. Lee and M. Nauenberg, *Phys. Rev.* 133, B1549, 1964.

[3] Gordon L. Kane. Top quark topics. *Gauge Bosons and Heavy Quarks: Proceedings of the Eighteenth SLAC Summer Institute on Particle Physics,* 125. SLAC-REPORT-378, July 1990.

[4] Gordon L. Kane. *Modern Elementary Particle Physics.* Addison-Wesley, 1993.

[5] G. Bordes and B. van Eijk, *Nucl. Phys.* B435, 23, 1995.

[6] P. Baringer and A. P. Heinsen . *Single Top Physics with a High Luminosity Tevatron,* DØnote 2600.

[7] E. Laenen, J. Smith, and W. L. van Neerven, *Top Quark Production Cross Section,* FERMILAB-PUB-93/270-T, 1993.

[8] R. C. Fernow. *Introduction to Experimental Particle Physics.* Cambridge University Press, 1986.

[9] C. Fabjan. Calorimetry in high-energy physics. Thomas Ferbel, *Experimental Techniques in High Energy Physics.*

[10] F. Sauli Principles of operation of multiwire proportional and drift chambers. Thomas Ferbel, *Experimental Techniques in High Energy Physics.* Addison-Wesley, 1987.

[11] C. Brown *et al., Nucl. Inst. Meth.,* A279:331-338, 1989.

[12] S. Abachi *et al., Nucl. Inst. Meth.,* A324:53, 1993.

[13] S. Snyder. *Measurement of the Top Quark Mass at DØ,* PhD thesis, State University of New York at stony Brook, May 1995.

[14] M. Narain, U. Heintz, *A Likelihood Test for Electron ID,* DØ Note 2386, 1994.

[15] N. Amos *et al., Luminosity Calculations for DØ,* DØ Note 2031, 1994.

[16] N. Amos *et al.*, Physics Letters B 243 158, 1990.

[17] N. Amos *et al.*, Phys. Rev. Letters B 63 2784, 1989.

[18] The CDF Collaboration, G. Chiarelli *et al.*, Fermilab preprint, FERMILAB-Conf-93/360-E, 1993.

[19] The CDF Collaboration, S. Belforte *et al.*, Fermilab preprint, FERMILAB-Conf-93/358-E, 1993.

[20] J. Bantly *et al., Improvement to the DØ Luminosity Monitor Constant,* DØ Note 2544, FERMILAB-TM-1930, 1995.

[21] R. Kehoe. *Resolution bias in jet response measurement* , DØ Note 2025, 1994.

[22] M. Paterno. *A study of the DØ calorimeter $\not\!\!E_T$ resolution using low E_T jet triggers.* DØ Note 1782, Fermi National Laboratory, 1993, (unpublished).

[23] E.R Kandal and J.H. Schwartz, *Principles of Neural Science,* 3rd ed, Elsevier Pub, New York (1991).

[24] *Scientific American* 3, 241, Sept 1979.

[25] Kolmogorov, Andrei Nikolaevic, *On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of One Variable and Addition*, Dokl. Akad. Nauk USSR, 114, 953-956, 1957.

[26] R. Hecht-Nielson, *IEEE First International Conference on Neural Networks San Diego, CA*, June 21-24, 1987

[27] Sprecher, David A.,*On the Structure of Continuous Functions of Several Variables* , Trans Amer. Math. Soc., 115, 340-355, March 1965.

[28] K. Hornik, M. Stinchcombe, and H. White, *Multilayer Feedforward Networks are Universal Approximators, Neural Networks,* Vol 2, pp. 359-366, 1989.

[29] K. Funashi, *On the Approximate Realization of Continuous Mappings by Neural Networks, Neural Networks,* Vol 2, pp. 183-192 (1989).

[30] G. Cybenko, *Approximation By Superposition of a Sigmoidal Function, Math. Control Signals Systems* 2, 303 (1989).

[31] J. Hertz, A. Krogh and R. G. Palmer,*Introduction to the Theory of Neural Computation,* Addison-Wesley, Redwood City, CA. (1991).

[32] M. D. Richard, R. P. Lippmann,*Neural Network Classifiers Estimate Bayesian a posteriori Probabilities, Neural Comput.* 3, 461 (1991).

[33] G. Tesauro, B. Janssens,*Scaling Relationships in Back-propagation Learning, Complex Systems*, 2, 39 (1988).

[34] W. Giele, E. Glover, and D. Kosower. *Nucl. Phys.,* B403: 633-667,1993.

[35] John M. Butler. *Main Ring Deadtime,* DØnote 1682.

[36] F. A. Berends, H. Kuijf, B. Tausk, and W. T. Giele. *Nucl. Phys.*, B357:32-64,1991

[37] S. Chopra, R. Raja. *Estimation of the QCD background to* $W \rightarrow e\nu + jets$, DØnote 2098.

[38] S. Chopra, M. Narain. *Determination of QCD backgrounds in* $e + jets$ *events*, DØnote 2105.

[39] Dhiman Chakraborty. *A search for* $t\bar{t} \rightarrow electron + \not{E}_T + jets$ *signature in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.8$ *TeV with the DØ detector.* PhD thesis, State University of New York at Stony Brook, September 1994.

[40] Meenakshi Narain, private communication.

[41] Sailesh Chopra, private communication.

[42] William G. Cobau, private communication.

[43] F. Abe *et al.*, (CDF). *Phys. Rev. Lett.*, 74:2626, April 1995.

[44] S. Abachi *et al.*, (DØ). *Phys. Rev. Lett.*, 74:2632, April 1995.

[45] E. Laenen, J. Smith, and W. L. van Neerven. *Top Quark Production Cross Section*, Fermilab-Pub-93/270-T, August 1993.

[46] E. Berger, H. Contopanagos. *Top Quark Production Dynamics in QCD.* ANL-HEP-CP-95-85, November, 1995.

Appendix A

The Functional Form of the Network Output

The analytic expression of P_{out} for the network trained in chapter 7 is,

$$P_{out}(\not{E}_T, E_T^e, E_T^{j_3}, H_T, A, S) = g(\vec{\omega}.g(\Omega\vec{x} + \vec{\Theta}) + \theta),$$

where

$$g(x) = \frac{1}{1 + e^{-2x}}$$

and

$$\vec{x} = \begin{pmatrix} \not{E}_T \\ E_T^e \\ E_T^{j_3} \\ H_T - E_T^{j_3} \\ A \\ S \end{pmatrix},$$

$$\Omega = \begin{pmatrix}
+0.2922E-01 & -0.1504E-01 & +0.3685E-01 & +0.2741E-02 & -0.4952E+01 & +0.2498E+00 \\
+0.8418E-03 & -0.1673E-01 & +0.2884E-02 & -0.4526E-02 & -0.1316E+01 & -0.2175E+00 \\
+0.4056E-02 & +0.8896E-03 & +0.2543E-01 & -0.3715E-02 & +0.3014E+01 & +0.4433E+00 \\
-0.4507E-02 & -0.1091E-01 & +0.1306E-03 & -0.6537E-02 & +0.1096E+01 & -0.1744E+00 \\
-0.3831E-01 & -0.4027E-01 & +0.2750E-01 & +0.1987E-03 & -0.3424E+00 & -0.3512E+00 \\
-0.7949E-02 & -0.8831E-02 & -0.8364E-02 & -0.6048E-02 & +0.2180E+01 & -0.1322E+00 \\
-0.4081E-02 & -0.1090E-01 & +0.3954E-02 & -0.8154E-02 & +0.1378E+01 & -0.2222E+00 \\
-0.4664E-01 & +0.3638E-01 & -0.4880E-02 & +0.1986E-02 & -0.7754E+01 & -0.9766E+00 \\
-0.3261E-01 & +0.2391E-01 & +0.3379E-02 & -0.1712E-03 & +0.8160E+01 & +0.2987E+00 \\
+0.2656E-01 & -0.7661E-02 & +0.1789E-01 & -0.1203E-02 & -0.2026E+01 & +0.3770E+00 \\
+0.4419E-01 & +0.8900E-02 & -0.5998E-01 & -0.1381E-01 & +0.8518E+00 & -0.9107E-01 \\
-0.9001E-02 & -0.6398E-02 & -0.1073E-01 & -0.7483E-02 & +0.3382E+01 & -0.5247E-01 \\
-0.2976E-02 & -0.1684E-01 & +0.2328E-02 & -0.3261E-02 & -0.1122E+01 & -0.2379E+00
\end{pmatrix},$$

$$\vec{\Theta} = \begin{pmatrix}
-0.5366 \\ +0.6075 \\ -0.4679 \\ +0.7340 \\ +0.7250 \\ +0.8228 \\ +0.9278 \\ +0.7508 \\ -0.4960 \\ -0.5256 \\ +0.3938 \\ +0.8909 \\ +0.6163
\end{pmatrix}, \quad \vec{\omega} = \begin{pmatrix}
+0.1046E+01 \\ -0.1029E+01 \\ +0.9053E+00 \\ -0.1166E+01 \\ -0.1303E+01 \\ -0.1266E+01 \\ -0.1334E+01 \\ -0.1292E+01 \\ +0.8451E+00 \\ +0.9228E+00 \\ -0.1338E+01 \\ -0.1381E+01 \\ -0.1053E+01
\end{pmatrix}, \quad \textit{and } \theta = -0.2293.$$