INFORMATION TO USERS

University of Alberta

Search for a Standard Model Higgs Boson in 161 GeV Electron Positron Collisions using Neural Networks

by

Robert Allan Davis

©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

Department of Physics

Edmonton, Alberta

Spring 1997

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced with the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-21162-2

Canada

University of Alberta

Library Release Form

Name of Author: Robert Allan Davis

Title of Thesis: Search for a Standard Model Higgs Boson in 161 GeV Electron Positron Collisions using Neural Networks

Degree: Master of Science

Year this Degree Granted: 1997

Permission is hereby granted to the University of Alberta library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.
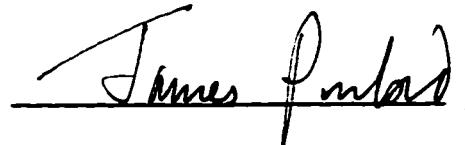
Robert Allan Davis
9422 - 95 Street
Edmonton, Alberta
T6C 3X3

Date: April 16, 1997

University of Alberta

Faculty of Graduate Studies and Research
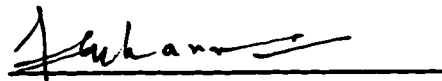
The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *Search for a Standard Model Higgs Boson in 161 GeV Electron Positron Collisions using Neural Networks* submitted by Robert Allan Davis in partial fulfillment of the requirements for the degree of Master of Science

Dr. James L. Pinfold, Supervisor

Dr. Douglas M. Gingrich

Dr. Faquir C. Khanna

Dr. W. Armstrong

Date: April 4, 1997

This thesis is dedicated to the memory of my grandparents,

Harry Allan and Marjorie Davis.

Preface

This thesis represents work performed within the OPAL collaboration while studying towards a Master of Science degree at the University of Alberta from May 1994 to April 1997. The thesis begins with an overview of the Standard Model of Particles and Interactions and the theoretical motivation for the existence of a Higgs Boson within this context. An introduction is then given to the European Laboratory for Particle Physics (CERN) and the LEP collider and OPAL detector are described in detail.

Next, an introduction to Artificial Neural Networks and their applications in data classification is given followed by a description of the network used for the Higgs search in this thesis. The training procedure and generalization performance of this network on simulated data is presented and optimized. Systematic effects are considered and the analysis is then applied to 161 GeV data from OPAL taken during the summer of 1996. The thesis concludes with a comparison of the neural network approach to other analysis methods and a discussion of the steps necessary to improve the analysis for the higher energy LEP runs of the near future.

Abstract

A search for a Standard Model Higgs Boson in the four jet final state using data from the OPAL detector at the Large Electron Positron (LEP) collider is presented. An Artificial Neural Network technique is used to isolate candidate Higgs events from the background created in 161 GeV centre of mass energy electron positron collisions.

The Artificial Neural Network method shows a greater sensitivity to the Higgs signal in the four jet channel than other analysis methods currently used within the OPAL collaboration. No candidate Higgs events are observed in the 161 GeV data, although a low integrated luminosity for the 161 GeV run precludes setting a meaningful mass limit on the Higgs.

Acknowledgments

I am indebted to many people, probably many more than I have listed below, for their help and support during the writing of this thesis. In particular, I would like to thank:

- Dr. J.L. Pinfold, for his interest in and support of the research presented in this thesis.

- Anthony Faust and Wayne Springer for their help and encouragement within the University of Alberta OPAL group.

- Satoru Yamashita and Ulrich Jost for helping with several technical matters regarding the OPAL software.

- Pierre Savard, co-founder of the *Canadian Task Force*, for his close collaboration on the ATLAS work completed during the course of my MSc.

- Mette Stuwe and Lisa Greenough of the OPAL Secretariat at CERN and Audrey Schaapman and Lynn Chandler of the Department of Physics for making navigation of the murky bureaucratic waters of trans-Atlantic science seem easy.

- The boys of P146: Alick, Anthony, Bryan, Dugan, Norm, Philip, and Trevor for helping with my teaching load during my trips to Geneva and for broadening my horizons through the daily "lunchtime discussion".

- My family for their continuing love and support.

- And finally, my love and thanks to Julie who gave the most sage advice of all: "Just get it done!"

Table of Contents

List of Tables

List of Figures

1. The Standard Model of Particles and Interactions

The Standard Model of Particles and their Interactions (the *Standard Model*) currently describes all observed physical interactions to within the limits of experimental accuracy. Despite its widespread success, however, there are some fundamental problems with the Standard Model which indicate that it is perhaps forms only a subset of a more fundamental theory. The shortcomings of the Standard Model will be discussed in section 1.3 below. The following sections describe the mathematical foundation of the Standard Model as well as the particles and forces that comprise the model.

1.1 Ingredients of the Standard Model

The Standard Model consists of two types of elementary particles: the spin ½ *fermions* and the integral spin *vector bosons* which mediate the interactions amongst the fermions. The fermions themselves are further subdivided into the quarks and the leptons. The vector bosons are associated with the local symmetry group $SU(3)_c \times SU(2)_L \times U(1)_Y$ in a manner which will be described below.

There are four fundamental forces in nature: the electromagnetic force, the weak force, the strong force, and the force of gravity. Table 1 shows the forces with their mediating gauge bosons and relative strengths. The electromagnetic force acts on charged particles and is mediated by the photon. The weak nuclear force, which is responsible for β decay, is mediated by the exchange of W and Z bosons which are massive. The strong force acts on quarks and is mediated by the eight massless gluons. Gravity has not yet been incorporated into the structure of the Standard Model although it is separately described by the theory of general relativity.

Force	Acts On	Gauge Boson(s)	Relative Strength	Range (m)
Gravity	All Particles	Graviton (massless, spin 2)	0.53×10^{-38}	∞
Electromagnetism	All Electrically Charged Particles	Photon (massless, spin 1)	7.3×10^{-3}	∞
Weak Interaction	Quarks, Leptons, Electroweak Gauge	$W^{\pm}, Z^0$ (heavy, spin 1)	1.02×10^{-5}	10^{-18}
Strong Interaction	All Coloured Particles	Eight Gluons (massless, spin 1)	~1	$< 10^{-15}$

Table 1: The forces of the standard model with their mediating gauge bosons.

The interactions within the Standard Model are described by quantum field theories. The electromagnetic force is described by Quantum Electrodynamics (QED). The weak interaction has been unified with the electromagnetic interaction into a full electro-weak theory [1]. The strong nuclear force is itself described by Quantum Chromodynamics, or QCD [2].

The quarks and the leptons can be divided into three families which exhibit identical properties with respect to the symmetries of the Standard Model. The masses of the particles in each of three families are different, however.

Quarks carry a colour charge from QCD while leptons carry no such charge. There are six different quarks and six different leptons, each of which is referred to as a "flavour". The six quarks are referred to as (in increasing order of mass) up (u), down (d), strange (s), charmed (c), bottom (b), and top (t). The six flavours of leptons are the electron (e), the electron neutrino (v_e), the muon (μ), the muon neutrino (v_μ), the tau (τ), and the tau neutrino (v_τ). All fermions except for the tau neutrino have been experimentally observed. While the masses of the neutrinos are currently consistent with zero, it is not generally expected that they will be equal to zero.

The three family structure of the fermions is shown below while the particle properties [3] are shown in Table 2.

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$$

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

Particle	Mass (GeV/c^2)	Charge (e)	Hypercharge
d	0.010±0.005	-1/3	1/3
u	0.005±0.003	+2/3	1/3
s	0.2±0.1	-1/3	1/3
c	1.5±0.2	+2/3	1/3
b	5.0±0.3	-1/3	1/3
t	171±10	+2/3	1/3
e	0.51	-1	0
μ	105.6	-1	0
τ	1776.9±0.5	-1	0
ν_e	$< 8 \times 10^{-6}$	0	0
ν_μ	< 0.25	0	0
ν_τ	< 35	0	0

Table 2: Properties of the fermions of the Standard Model.

If the SU(3)$_c$ × SU(2)$_L$ × U(1)$_Y$ symmetry which underlies the Standard Model were exact, all gauge bosons would be massless and the forces which they mediate would therefore have infinite range. The observed short range of the weak force, however, implies that the W and Z bosons must be massive. In order for the symmetry to be broken to allow these masses while still allowing the theory to be free of infinities (renormalizable), the symmetry of the Standard Model must be broken spontaneously. As will be seen in the next chapter, this spontaneous symmetry breaking leads to a real scalar particle known as the Higgs Boson. The theory predicts the coupling strengths of the Higgs boson, but not its mass.

The Higgs boson is yet to be experimentally observed, but lower limits on its mass have been set by the LEP experiments. The latest OPAL lower limit [4] on the Standard Model Higgs boson mass is 59.5 GeV/c^2 at the 95% confidence level.

1.2 Underlying Symmetries of the Theory

In this section, some of the mathematics which underlies gauge theories such as the Standard Model is presented. A general gauge invariant Lagrangian may be written as

$$L(x) = \overline{\psi}(x)i\gamma_\mu D^\mu\psi(x) - \frac{1}{4}F^a_{\mu\nu}F^{\mu\nu,a}(x)$$

where $\psi(x)$ is a fermion field. If we define a gauge boson field $A_\mu(x)$ then the field strength tensor $F_{\mu\nu}(x)$ is given by

$$F^a_{\mu\nu}(x) = \partial_\mu A^a_\nu(x) - \partial_\nu A^a_\mu(x) + gf^{abc}A^b_\mu(x)A^c_\nu(x)$$

and the covariant derivative D_μ is written

$$D_\mu = \partial_\mu - igT^a A^a_\mu(x).$$

In the above equations, g is the coupling constant of the theory and T^a are the generators of the Lie symmetry underlying the theory with structure constants f^{abc}. These generators satisfy the commutator relation $\left[T^a, T^b\right] = if^{abc}T^c$.

The particles of the Standard Model exist as either singlets or as examples of the fundamental representation of the symmetry group: doublets for SU(2) or triplets for SU(3). The gauge bosons exist in the adjoint representation of the symmetry group and hence their number is determined by the number of generators for the group: one for U(1), three for SU(2), and eight for SU(3). These bosons are the photon for U(1), the three massive vector gauge bosons W^+, W^-, and Z^0 for SU(2), and the eight gluons for SU(3).

In an Abelian theory such as QED, the generators commute and hence vector bosons have no self interactions. QED is based on the U(1) symmetry and its gauge boson (the photon) does not carry any charge. Within QED, the fermion fields transform as $\psi(x) \rightarrow e^{ix}\psi(x)$ and the coupling constant g is simply the electron charge e. In order to allow the theory to be renormalizable, that is to avoid infinities due to perturbative calculations, the coupling constant must change (or *run*) with the energy scale. In QED, the associated coupling constant $\alpha(Q^2=0)\sim 1/137$ increases slowly with the energy scale.

An example of a non-abelian theory is QCD. Non-abelian theories are much more complicated than abelian ones as the kinetic term in the Lagrangian $\left(-\frac{1}{4}F_{\mu\nu}^{a}(x)F^{\mu\nu,a}(x)\right)$ introduces gauge boson self interactions into the theory. QCD has SU(3) as an underlying symmetry group and hence has eight gauge bosons which are associated with the gluons. The fermion fields under QCD transform as $\psi(x) \rightarrow e^{ix^{a}T^{a}}\psi(x)$ with the T^{a} being the Gell-Mann Matrices. A similar transformation exists for SU(2) with the T^{a} being the Pauli spin matrices. The coupling constant of QCD, α_{s}, decreases as Q^2 increases which results in the property of asymptotic freedom where quarks behave as free particles at short distances, but become increasingly bound at large distances.

As was mentioned previously, the basic theory does not allow for massive gauge bosons. Attempts to introduce mass terms for the bosons by hand fail as the Lagrangian then violates the gauge invariance and destroys the renormalization of the theory. The Higgs mechanism, described in the next chapter, provides a method for allowing mass in the theory without these problems.

1.3 Unanswered Questions in the Standard Model

Although the Standard Model has successfully explained the results of every experiment performed thus far, there are several fundamental problems with the theory that cause physicists to believe the it is embedded in a more fundamental theory through which further unification of the forces may be achieved. Some of these problems with the Standard Model are:

- There is no fundamental reason for the theory to be explained by the three independent symmetry groups U(1), SU(2), and SU(3). It is thought that it should be possible to unify the theory under one grand symmetry group.

- There are at least 23 free parameters in the theory which must be inserted by hand. It is thought that a fundamental theory would have far fewer free parameters.

- There is no fundamental reason for the existence of three families of quarks and leptons or for the symmetry between quark and lepton flavours. It is possible that the quarks and leptons are composite particles of more fundamental objects.

- There is no fundamental reason in the theory for the fact that the proton and electron have exactly opposite electric charges.

- The energy scale of the Standard Model varies over many orders of magnitude. It is considered a difficult problem that the weak unification scale is so much smaller than the projected grand unification scale.

- Correcting for radiative effects on the Higgs boson and gauge boson masses involves terms many orders of magnitude larger than the masses themselves. This "unnatural" fine tuning is not considered to be a property of a fundamental theory.

- Gravity has not been explained in a format that is consistent with the rest of the Standard Model.

- The Standard Model does not contain an ideal "dark matter" candidate to explain certain astrophysical observations regarding the total amount of mass in the universe.

- The Higgs boson has yet to be experimentally discovered.

It is widely thought that the central problem facing the Standard Model today is the experimental confirmation of the existence of a Higgs boson. The next chapter will explore the theory of the Higgs mechanism and the experimental signatures of such a particle.

2. The Higgs Boson in Theory and Experiment

The main theme of this thesis is the search for one of the last remaining unobserved particles in the Standard Model, the Higgs Boson. The Higgs boson is a theoretical tool invoked to solve the problem of the lack of mass terms in the Standard Model Lagrangian for fermions and massive gauge bosons. Attempts to explicitly add mass terms to the Lagrangian result in the breaking of the SU(2) or gauge invariances and thus cannot be the correct solution. The Higgs mechanism, however, manages to break the SU(2) and gauge symmetries in a subtle way that manages to allow mass in the theory while maintaining the gauge invariance. Although the Higgs mechanism is a technically adequate solution to the mass problem, the physical interpretation of what the Higgs mechanism implies is somewhat less well understood. It is believed [5] that achieving a better understanding of the Higgs sector should be considered to be the central problem in modern particle physics.

The following sections describe the mechanism by which the Higgs boson breaks the gauge symmetry of the Standard Model in order to allow massive fermions and bosons. Several examples of this *Spontaneous Symmetry Breaking* are presented in section 2.1 and then the entire Standard Model Higgs mechanism is described in section 2.2.

2.1 Spontaneous Symmetry Breaking

As was mentioned above, the Higgs mechanism breaks the gauge symmetry of the theory in a subtle way that allows for massive fermions and bosons while still maintaining the good effects of the symmetry. This method of breaking the symmetry is *called spontaneous symmetry breaking* and refers to the situation when the symmetry is still valid for the Lagrangian but not for the ground state solution of the system. In the following sections,

scenarios of increasing complexity are considered in order to develop the mechanism of spontaneous symmetry breaking for application to the full Standard Model Lagrangian of fermions and bosons.

2.1.1 A Simple Case

In order to illustrate the concept of spontaneous symmetry breaking, we first consider a very simple case. We consider a Lagrangian

$$L = T - V = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \left(\frac{1}{2} \mu^2 \phi^2 + \frac{1}{4} \lambda \phi^4 \right)$$

where we can require that $\lambda > 0$ in order to bound the potential as $\phi \to \infty$. In order to find the spectrum of this Lagrangian, we proceed in the usual quantum mechanical way to find the classical minimum of the potential in order to find the ground state of the system. The fields are then expanded around their ground state values in order to determine the spectrum of solutions. In quantum field theory, the ground state of the system is referred to as the *vacuum* and the excitations around the minimum as *particles*. The masses of the particles are then deduced from the self interaction terms for the field in the Lagrangian near the classical minimum of the potential.

If we suppose that $\mu^2 > 0$ for our Lagrangian, there is no symmetry breaking. If $\mu^2 < 0$, we can find the minimum of the potential in our Lagrangian through setting $\partial V/\partial \phi = 0$ which gives the result

$$\phi \left(\mu^2 + \lambda \phi^2 \right) = 0.$$

This result, shown in Figure 1, gives a minimum not at $\phi = 0$, but at

$$\phi = \pm \sqrt{\frac{-\mu^2}{\lambda}} = v$$

where υ is called the *vacuum expectation value* of ϕ. The field ϕ corresponds to the Higgs field for this simple case.
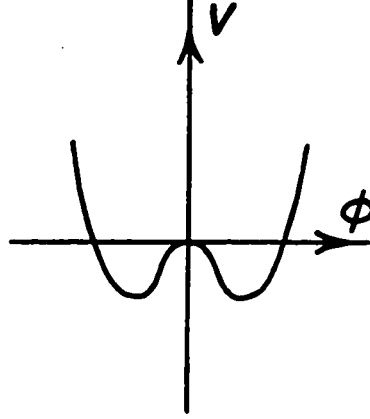
Figure 1: Potential energy versus ϕ for a simple example of spontaneous symmetry breaking.

In order to now determine the particle spectrum of our Lagrangian, we must expand the theory in the region of the minimum. To do this, we expand the Higgs field around $\eta=0$ to give

$$\phi(x) = \upsilon + \eta(x).$$

We could also have chosen $-\upsilon$ as our vacuum, but the reflection symmetry $\phi \rightarrow -\phi$ present in our original Lagrangian makes the physics independent of this choice. Substituting this expansion into our Lagrangian, we obtain the result

$$L = \frac{1}{2}\left(\partial_\mu\eta\partial^\mu\eta\right) - \left\{\frac{1}{2}\mu^2\left[\upsilon^2 + 2\eta\upsilon + \eta^2\right] + \frac{1}{4}\lambda\left[\upsilon^4 + 4\upsilon^3\eta + 6\upsilon^2\eta^2 + 4\upsilon\eta^3 + \eta^4\right]\right\}$$

$$= \frac{1}{2}\left(\partial_\mu\eta\partial^\mu\eta\right) - \left\{\frac{\upsilon^2}{2}\left(\mu^2 + \frac{1}{2}\lambda\upsilon^2\right) + \eta\upsilon\left(\mu^2 + \lambda\upsilon^2\right) + \frac{\eta^2}{2}\left(\mu^2 + 3\lambda\upsilon^2\right) + \lambda\upsilon\eta^3 + \frac{1}{4}\lambda\eta^4\right\}$$

which simplifies to

$$L = \frac{1}{2}\left(\partial_\mu \eta \partial^\mu \eta\right) - \left(\lambda v^2 \eta^2 + \lambda v \eta^3 + \frac{1}{4}\lambda \eta^4\right) + \text{constant.}$$

Thus, we see that the η field has acquired a mass[†] and therefore describes a particle with mass $m^2 = 2\lambda v^2 = -2\mu^2$. We note that the original reflection symmetry that was present in our initial Lagrangian is no longer present after we make the choice of vacuum to be $\phi = +v$ rather than $\phi = -v$. Thus, by making the choice of vacuum, we have broken the symmetry for the ground state while maintaining the reflection symmetry of the original Lagrangian. As a by product of this procedure, we gained a massive field η.

2.1.2 Complex Scalar Field

As the next step in our investigation of spontaneous symmetry breaking, we consider a complex scalar field $\phi = \left(\phi_1 + i\phi_2\right) / \sqrt{2}$ together with the Lagrangian

$$L = \left(\partial_\mu \phi\right)^* \left(\partial^\mu \phi\right) - \mu^2 \phi^* \phi - \lambda\left(\phi^* \phi\right)^2.$$

This Lagrangian is invariant under the global U(1) symmetry $\phi \rightarrow \phi' = e^{i\chi}\phi$ and may be written in terms of its real components as

$$L = \frac{1}{2}\left(\partial_\mu \phi_1\right)^2 + \frac{1}{2}\left(\partial_\mu \phi_2\right)^2 - \frac{1}{2}\mu^2\left(\phi_1^2 + \phi_2^2\right) - \frac{\lambda}{4}\left(\phi_1^2 + \phi_2^2\right)^2.$$

[†] Mass terms in the Lagrangian are identified as being quadratic in the field variable and opposite in sign to the kinetic term of the Lagrangian.
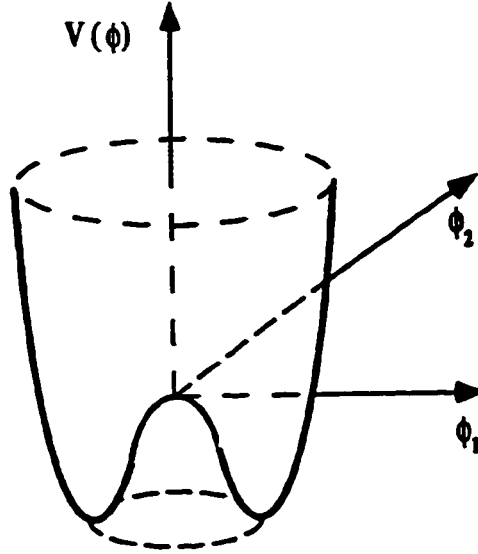
Figure 2: Potential energy as a function of the two components of a complex scalar field for $\mu^2 < 0$.

Proceeding as before and examining the potential in the Lagrangian for the case $\mu^2 < 0$, we see that the minimum potential energy is along a circle in the ϕ_1, ϕ_2 plane (see Figure 2) with radius

$$\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} = v^2$$

where we have again defined our vacuum expectation energy. We are free to select any point along the circle as our vacuum, but in doing so the symmetry of the ground state is broken. If we choose the point $\phi_1 = v$ and $\phi_2 = 0$, we may write (with $\eta(x)$ and $\rho(x)$ real fields)

$$\phi = \frac{\left(v + \eta(x) + i\rho(x)\right)}{\sqrt{2}}$$

as the expansion of the Higgs field about the vacuum. When this expansion is substituted into the Lagrangian, we obtain

$$L = \frac{1}{2}(\partial_\mu \rho)^2 + \frac{1}{2}(\partial_\mu \eta)^2 + \mu^2 \eta^2 - \lambda \upsilon(\eta \rho^2 + \eta^3) - \frac{\lambda}{2}\eta^2 \rho^2 - \frac{\lambda}{4}\eta^4 - \frac{\lambda}{4}\rho^4 + \text{constant.}$$

Analyzing this equation, we see that the η field has obtained a mass $m^2 = 2|\mu|^2$ but cancellation of the quadratic terms for the ρ field implies the existence of a gauge boson of zero mass. A massless boson produced in this way through spontaneous symmetry breaking is known as a Goldstone boson. Such a boson arises whenever a continuous global symmetry is broken while choosing a particular ground state. Intuitively, the massless Goldstone boson may be regarded as the excitation which corresponds to motion around the circular minimum in the potential energy. Since there is no resistance to motion around this circle, the resulting boson is massless.

2.1.3 The Higgs Mechanism for a Local Gauge Theory

As a final approximation to the real Higgs mechanism of the Standard Model, we will consider the Higgs mechanism where the global gauge invariance of the last section is changed to a local gauge invariance $\phi(x) \rightarrow \phi'(x) = e^{i\chi(x)}\phi(x)$. It is well known that in order to ensure such an invariance, we must introduce a massless vector field A_μ as well as write the Lagrangian in terms of the covariant derivative D.

$$\partial_\mu \rightarrow D_\mu = \partial_\mu - igA_\mu$$

The massless gauge field transforms under the local gauge symmetry as

$$A_\mu \rightarrow A_\mu' = A_\mu - \frac{1}{g}\partial_\mu \chi(x).$$

Our full Lagrangian for this case, then, is

$$L = (D_\mu \phi)^*(D^\mu \phi) - \mu^2 \phi^* \phi - \lambda(\phi^* \phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

with the last term being the kinetic energy term which does not enter into the analysis.

Due to the gauge invariance in this example, ϕ may be written in general as

$$\phi(x) = \eta(x)e^{-i\rho(x)}$$

where η and ρ are real. In addition, since we are dealing with a *local* gauge invariance, we may choose an appropriate transformation which will enable us to write ϕ as an expansion around the vacuum expectation value υ as

$$\phi(x) = \frac{(\upsilon + h(x))}{\sqrt{2}}.$$

It is important to note that such a choice of transformation can only be made in the case of a local gauge invariance and could not have been performed in the previous case of a global gauge invariance. Substituting this equation into the Lagrangian, we obtain

$$L = \frac{1}{2}\left[\left(\partial^\mu + igA^\mu\right)\!\left(\upsilon + h\right)\right]\!\left[\left(\partial_\mu - igA_\mu\right)\!\left(\upsilon + h\right)\right]$$

$$-\frac{\mu^2}{2}(\upsilon + h)^2 - \frac{\lambda}{4}(\upsilon + h)^4 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

$$= \frac{1}{2}\left(\partial_\mu h\right)\!\left(\partial^\mu h\right) + \frac{1}{2}g^2\upsilon^2 A_\mu A^\mu - \lambda\upsilon^2 h^2 - \lambda\upsilon h^3$$

$$-\frac{\lambda}{4}h^4 + g^2\upsilon h A^\mu A_\mu + \frac{1}{2}g^2 h^2 A_\mu A^\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

which gives a mass term of $g\upsilon$ for the gauge boson A_μ. So, the mass of the gauge boson is non-zero only when the symmetry is spontaneously broken by the choice of a specific vacuum. The massive gauge boson which comes out of the spontaneous symmetry breaking in this case is said to have "eaten" the Goldstone boson of the previous section.

The analysis of this section is essentially the entire Higgs mechanism. In the next section, the last additional complexities in order to incorporate the Higgs mechanism into the Standard Model are discussed.

2.2 The Higgs Mechanism and the Origin of Mass

When applied to the Standard Model Lagrangian, the Higgs field is assigned to an SU(2) doublet made up of two complex scalar fields:

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

with $\phi^+ = \dfrac{\phi_1 + i\phi_2}{\sqrt{2}}$ and $\phi^0 = \dfrac{\phi_3 + i\phi_4}{\sqrt{2}}$. Since the relationship between the electric charge Q, the weak isospin value T_3, and the U(1) hypercharge Y_H is given by $Q=T_3+Y_H/2$, the charge assignment given in the above doublet corresponds to $Y_H=1$. The Lagrangian has the form

$$L = \left(\partial_\mu \phi\right)^\dagger \left(\partial^\mu \phi\right) - \mu^2 \phi^\dagger \phi - \lambda\left(\phi^\dagger \phi\right)^2$$

with the local gauge transformation invariance

$$\phi(x) \rightarrow \phi'(x) = e^{i\vec{\alpha}(x)\cdot\vec{\tau}/2}\phi(x) \qquad\qquad\qquad \backslash$$

where τ_i are the Pauli matrices and α_i are a set of parameters.

Since the Higgs field is now an SU(2) doublet, we may write

$$\phi^\dagger \phi = \begin{pmatrix} \phi^{+*} & \phi^{0*} \end{pmatrix}\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \phi^{+*}\phi^+ + \phi^{0*}\phi^0$$

which may be re-written in terms of real components only as

$$\phi^{\dagger}\phi = \frac{\left(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2\right)}{2}.$$

We now consider the potential of our Lagrangian and observe that, for $\mu^2 < 0$, it has a minimum at

$$\phi^{\dagger}\phi = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2}.$$

In order to break the symmetry, we must again choose a specific vacuum. This time, we must select a specific direction in SU(2) space as our vacuum ϕ_0.

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}.$$

Since we again have a local gauge symmetry, we may choose our expansion around the vacuum to be of the form

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

We may now substitute this result into our Lagrangian to see the effects of spontaneous symmetry breaking in the Standard Model. In order for our Lagrangian to be invariant under our local gauge transformation, we must replace the derivatives ∂_μ with the covariant derivates D_μ. Including the gauge boson B_μ from the U(1) symmetry and the three gauge bosons W_μ from the SU(2) symmetry, the covariant derivative in this case is written as

$$D_\mu = \partial_\mu - ig_1\frac{Y}{2}B_\mu - ig_2\frac{\vec{\tau}}{2}\cdot\vec{W}^\mu.$$

When we substitute our ϕ with a chosen vacuum expectation value into the Lagrangian, we obtain the extra terms

$$\phi^\dagger \left(ig_1 \frac{Y}{2} B_\mu + ig_2 \frac{\vec{\tau}}{2} \cdot \vec{W}_\mu \right)^\dagger \left(ig_1 \frac{Y}{2} B_\mu + ig_2 \frac{\vec{\tau}}{2} \cdot \vec{W}^\mu \right) \phi \, .$$

If we now set Y=1 as we saw was the case above, write out the Pauli matrices explicity, and set $\phi = \phi_0$, the contribution to the Lagrangian becomes

$$\frac{1}{8} \left\| \begin{pmatrix} g_1 B_\mu + g_2 W_\mu^3 & g_2 \left(W_\mu^1 - iW_\mu^3 \right) \\ g_2 \left(W_\mu^1 + iW_\mu^2 \right) & g_1 B_\mu - g_2 W_\mu^3 \end{pmatrix} \begin{pmatrix} 0 \\ v \end{pmatrix} \right\|^2$$

$$= \frac{1}{8} v^2 g_2^2 \left(\left(W_\mu^1 \right)^2 + \left(W_\mu^2 \right)^2 \right) + \frac{1}{8} v^2 \left(g_1 B_\mu - g_2 W_\mu^3 \right)^2$$

2.2.1 The Boson Masses

The first term in the above equation may be written in terms of the real charge states of the W boson as

$$\left(\frac{1}{2} v g_2 \right)^2 W_\mu^+ W^{-\mu}$$

which gives the charged W bosons a mass $M_W = v g_2 / 2$. The second term in the contribution of the spontaneous symmetry breaking to the Lagrangian is a linear combination of B and W_3 forming the field $Z_\mu = \dfrac{g_1 B_\mu + g_2 W_\mu^0}{\sqrt{g_2^2 + g_1^2}}$. Thus, the Z boson

acquires a mass $M_Z = \dfrac{1}{2} v \sqrt{g_1^2 + g_2^2}$ after the symmetry breaking. As expected, no mass term for A_μ is found in the Lagrangian after spontaneous symmetry breaking and hence the photon remains massless.

2.2.2 The Fermion Masses

Since the Higgs field is now written as an SU(2) doublet, we may consider the interaction of fermions with the Higgs field. We can add a lepton interaction term to the Lagrangian

$$g_e\left(\overline{L}\phi e_R^- + \phi^\dagger \overline{e_R^-} L\right)$$

with $L = \begin{pmatrix} v_e \\ e^- \end{pmatrix}_L$. The coupling g_e is arbitrary and not determined by the invariance principle. If we replace the Higgs field by its expansion around the vacuum

$$\phi \rightarrow \begin{pmatrix} 0 \\ \dfrac{v+H}{\sqrt{2}} \end{pmatrix},$$

and then substitute this into our lepton interaction Lagrangian, we obtain the Yukawa terms

$$\frac{g_e v}{\sqrt{2}}\left(\overline{e_L^-}e_R^- + \overline{e_R^-}e_L^-\right) + \frac{g_e}{\sqrt{2}}\left(\overline{e_L^-}e_R^- + \overline{e_R^-}e_L^-\right)H.$$

The first term is of the form expected for a fermion mass, so the electron mass is given by $m_e = g_e v / \sqrt{2}$. Since g_e was arbitrary, however, the electron mass has not been calculated and must still be measured. The theory can now accommodate non-zero fermion masses. Once the electron mass is measured, the equation for the electron mass may be inverted to calculate the coupling constant g_e. The second term in the lepton interaction Lagrangian after spontaneous symmetry breaking indicates that there is an interaction between the Higgs and the electron of strength $g_e / \sqrt{2} = m_e / v$.

A similar calculation to the one performed above for leptons may be made with the quark interaction Lagrangian. Again, the masses of the quarks are accommodated, but not predicted, by the theory.

2.2.3 Higgs Boson Decays

As we saw in the above section, the strength of the coupling between the Higgs boson and fermion is given by $g_f = \sqrt{2}m_f / v$ which may be re-written using our knowledge of the vacuum expectation value in terms of the W mass from section 2.2.1 as $g_f = \dfrac{m_f g_2}{2M_w}$. We see that the strength of the coupling is proportional to the mass of the fermion and hence the Higgs boson will tend to decay into the heaviest fermions available. A similar interaction strength between the Higgs boson and other gauge bosons may be written as $g_w = g_2 M_w$. Higgs couplings to gluons or photons do not exist at the tree level but may be achieved through loops of fermions or W's.
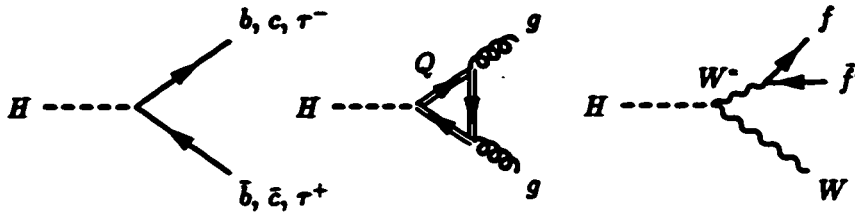
Figure 3: The Feynman diagrams for the main Higgs boson decay modes in the LEP2 energy range.

The Feynman diagrams for the main decay modes of the Higgs boson at LEP2 energies [6] [7] are shown in Figure 3 and consist of:

- quark decays (into b quark pairs and c quark pairs)

- lepton decays (primarily into pairs of τ leptons)

- gluon decay into pairs of gluons

- W boson and Z boson decay into pairs of these gauge bosons

The decay into b quark pairs is by far the most dominant process followed by decays into τ leptons, charmed quarks, and gluon decays. The last three decay channels contribute at a level less than 10% to the total number of Higgs boson decays. The branching ratios for the various decays of the Higgs are shown versus the Higgs mass in Figure 4. The branching ratios for Higgs decays into bosons is negligible for the range of Higgs masses of interest in this thesis. It can be clearly seen that decays into b quark pairs are the dominant process and hence any search for the Higgs boson must employ effective b quark tagging techniques.
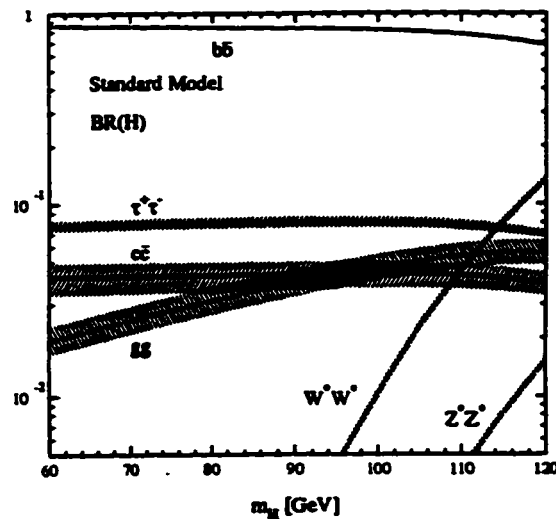
Figure 4: Branching ratios for the decay modes of the Standard Model Higgs Boson versus the Higgs mass. The shaded bands are the uncertainties in the branching ratios due to errors in the quark masses and the QCD coupling.

2.3 Higgs Boson Production at the Large Electron Positron Collider

Standard Model Higgs Boson production at LEP2 takes place through the two major mechanisms shown in Figure 5. The first, Higgs-strahlung, is the dominant process and results from Higgs bosons being radiated from a virtual Z boson. The second process, WW fusion, is suppressed by an extra electroweak vertex and therefore is only important in the region of phase space where the Z boson turns virtual in the Higgs-strahlung process.

In this region, however, the cross section of both processes is small and the WW fusion process will not add significantly to the available number of Higgs bosons.

The cross section for the Higgs-strahlung process for three different LEP centre of mass energies is shown in Figure 6. The cross section drops off rapidly at large Higgs masses.
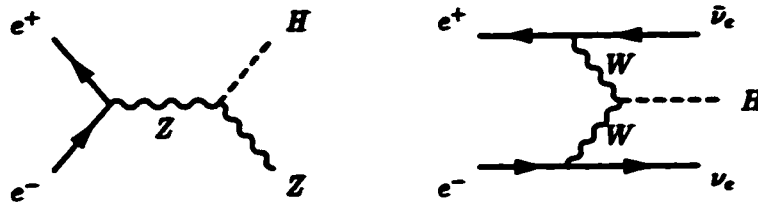
Figure 5: The two main Higgs production mechanisms at LEP. Higgs-strahlung is shown on the left and WW fusion on the right.
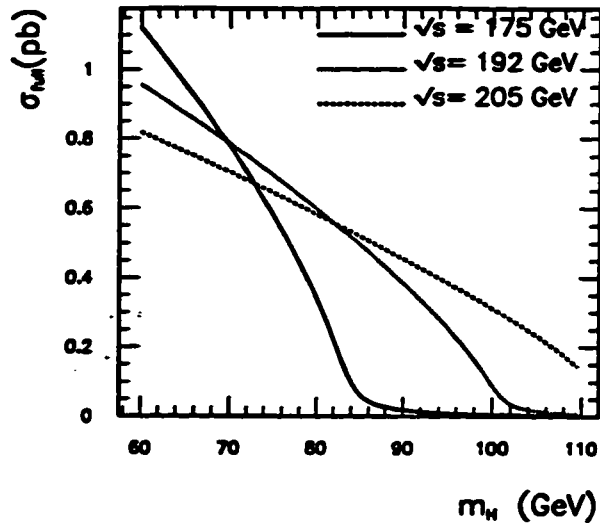
Figure 6: Cross sections for Higgs boson production through Higgs-strahlung for three representative LEP energies versus the Higgs mass.

2.3.1 The Four Jet Final State

The four jet final state of Higgs boson decay arises when the Higgs decays to a pair of quarks (in about 90% of all cases) while the accompanying Z boson also decays into quarks (about 70% of the time). These branching ratios make the four jet channel the most abundant final state of Higgs boson decay at LEP2 as it occurs in about 65% of the cases.

The procedure in searching for such a final state is to tag the heavy b quarks coming from the Higgs boson while fitting the event to a energy structure which includes a real Z boson. A more detailed explanation of the methods used to search for the Higgs in this channel may be found in chapter 5.

The cross section for Higgs production times the branching ratio for the four jet process at $\sqrt{s} = 161\,\text{GeV}$ are shown in Table 3 for various Higgs masses. As expected, the cross sections fall rapidly as the Higgs mass increases.

Higgs Mass (GeV/c²)	Cross Section (nb)
55	0.9530
60	0.7280
62	0.6270
64	0.5170
66	0.3980
68	0.2590
70	0.1170

Table 3: Cross sections for Higgs Boson production at a 161 GeV centre of mass energy at LEP.

2.3.2 Background Processes to the Four Jet Signal

The four jet Higgs decay channel is plagued by many background sources which mimic the four jet structure of the signal events. A large background comes from multi-jet QCD events from $\left(Z^0 / \gamma^* \rightarrow q\bar{q}(\gamma)\right)$ where the final state quarks radiate gluons to form four jets. Another large background comes from the production of W^+W^- pairs each of which subsequently decay into pairs of jets. The cross sections for the QCD and W pair backgrounds are shown in Figure 7. In addition, any four fermion final state event which

can mimic a four jet topology will form a background channel. Some of the Feynman diagrams which can produce these "four fermion" events are shown in Figure 8 and the resulting cross-sections are shown in Figure 9 for a centre of mass energy of 175 GeV.
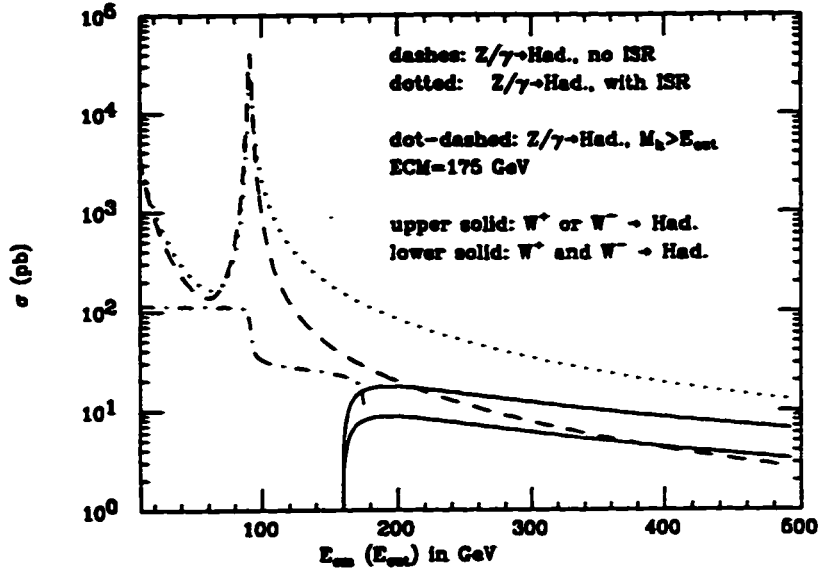
Figure 7: Cross sections for the hadronic Z decay background and W pair backgrounds versus centre of mass energy.
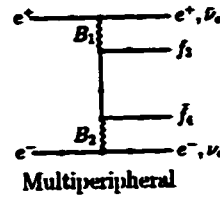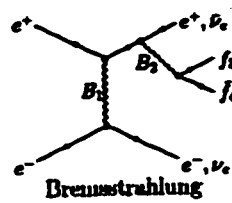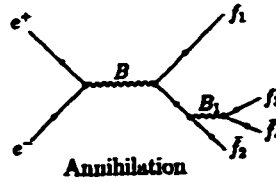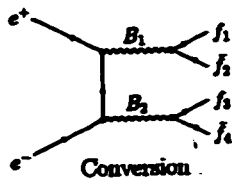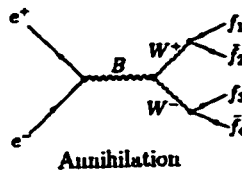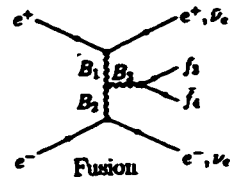
Figure 8: Some of the Feynman diagrams contributing to the four fermion events.

$$\sigma \left(e^+ e^- \rightarrow e^+ e^- b\, b \right)$$
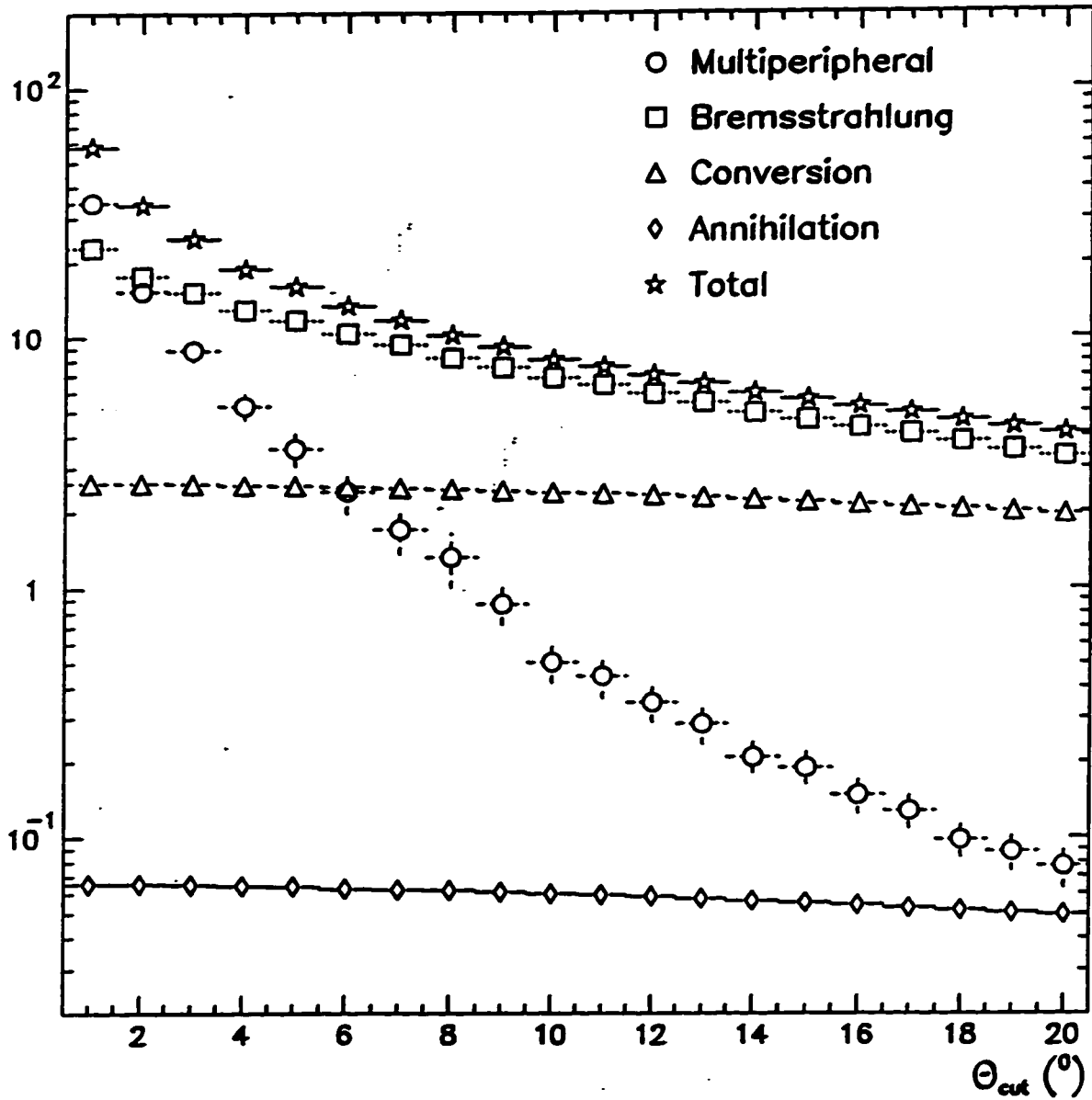
Figure 9: Cross sections for various four fermion processes (in fb) versus a cut on the polar angle of the fermions. The cut is designed to reject events with fermions close to the beam axis. The cross sections in a figure are for a centre of mass energy of 175 GeV.

The cross sections for the major background channels at $\sqrt{s} = 161$ are shown in Table 4. The methods used to reduce these background levels in order to allow detection of the Higgs boson signal are discussed in chapter 5.

Process	Cross Section (pb)
$\left(Z^0 / \gamma^* \to q\bar{q}(\gamma)\right)$	147.4
W^+W^-	3.457
ZZ	0.458
Ze^+e^-	2.53
$We\nu$	0.68

Table 4: Cross sections for background processes to the four jet Higgs channel at a centre of mass energy of 161 GeV.

3. The LEP Collider and the OPAL Experiment

In order to probe the microcosm at the energies necessary to detect ever more fundamental information about our universe, the complexity and size of the world's particle accelerators have increased greatly over the past forty years. This increase in the amount of energy available at particle colliders, and the attendant increase in the scope of accessible physics, has caused a dramatic increase in the expense and complexity of the detectors designed to observe and record the collisions. In fact, the expense of modern particle accelerators and detectors has become so large that only multi-national collaborations of scientists can pool sufficient resources to build them. The Large Electron Positron (LEP) collider and the Omni Purpose Apparatus for LEP (OPAL) detector are perfect examples of this multi-national cooperation in the pursuit of a better understanding of the fundamental particles and forces of our universe.

3.1 CERN - The European Laboratory for Particle Physics

The European Laboratory for Particle Physics (CERN[†]), was established in 1954 as a joint effort of twelve European nations. Situated astride the Franco-Swiss border just west of the city of Geneva, the laboratory was one of the first examples of international collaboration in postwar Europe. The laboratory provides research facilities to over 6500 scientists, some 1500 of which come from non-European states including Canada, the United States, and China. Half of the world's practicing particle physicists use CERN's facilities.

[†] The acronym CERN comes from the original French name for the laboratory, "Conseil Européen pour la Recherche Nucléaire".

CERN's main accelerator complex, shown in Figure 10, is centered on the venerable Proton Synchrotron (PS). Although originally designed as a proton accelerator for fixed target experiments, the PS has been continuously upgraded since its inception and is now capable of accelerating electrons, protons, and heavy ions simultaneously in order to feed the other accelerators in the complex. Several fixed target experiments and test beam apparatus still benefit from particles accelerated directly by the PS. The Super Proton Synchrotron (SPS), famous for producing the first W and Z bosons in the UA1 experiment, now accelerates protons and anti-protons for collider and fixed target experiments as well as accelerating electrons and positrons for injection into LEP. The complex is crowned by the 27 km in circumference Large Electron Positron (LEP) collider which now accelerates electrons and positrons to 172 GeV centre of mass collisions making LEP the most powerful e^+e^- collider in the world. Further energy upgrades are planned for LEP in the near future which will result in an accelerator capable of producing e^+e^- collisions at centre of mass energies equal to 192 GeV by 1998.

3.2 The Large Electron Positron (LEP) Collider

The Large Electron Positron (LEP) collider created its first e^+e^- collisions in 1989 and immediately began to fulfill its design goal as a factory for Z^0 bosons. LEP operated at its original design energy near the Z^0 peak (~45 GeV per beam) for seven years and at or near its design luminosity of 16×10^{31} cm^{-2}s^{-1} for much of this time. Over the last seven years, some 160 pb^{-1} of Z^0 decays were collected by the OPAL detector alone.

In November of 1995, the first superconducting magnets were added to the LEP ring in order to boost the accelerator's energy above the Z^0 peak. These sixty superconducting cavities allowed the beam energy to be increased from ~45 GeV to 70 GeV resulting in e^+e^- collisions with a centre of mass energy of 140 GeV. A further 84 superconducting cavities were installed in early 1996 allowing a beam energy of ~81 GeV and consistent collisions at $\sqrt{s}$ = 161 GeV to be achieved. This centre of mass energy allowed the production of

W^+W^- particle pairs, and thus opened up an entirely new area of research for LEP physicists. The Higgs boson search presented in this thesis makes use of the data taken by OPAL at $\sqrt{s} = 161$ GeV in the summer of 1996.

Figure 10: The CERN accelerator complex.

CERN makes use of all of its older accelerators in order to create highly energetic e^+e^- collisions since each machine can only operate over a fixed range of energies. The system of accelerators which feed the LEP ring are shown schematically in Figure 10. The electrons are first accelerated to 200 MeV in the electron linac (LIL) and directed onto a target in order to produce positrons. The positrons are then accelerated to an energy of 600 MeV in the second stage of LIL before being injected into the Electron-Positron Accumulator (EPA) where they are stored and compressed by synchrotron damping. This process is repeated until about 2×10^{11} positrons in eight bunches have been collected in the EPA, a goal which can take up to 1000 bursts from LIL to achieve. This beam is then injected into the Proton Synchrotron where it is ramped up to an energy of 3.5 GeV before being injected into the Super Proton Synchrotron (SPS) and further accelerated to 20 GeV. At this point, the positron beam is ready to be injected into LEP in a clockwise direction and ramped up to its final energy of ~81 GeV. The electrons follow the same acceleration program as the positrons except that the LIL target is removed and the SPS is set to inject them into LEP in a counter-clockwise direction.

LEP itself is divided into eight straight sections connected by curved sections which complete the circular shape. The straight sections contain both the experimental halls and the RF accelerating cavities while the curved sections hold the dipole magnets which direct the beam around the ring. Quadrupole magnets are situated at various positions throughout the LEP ring in order to maintain beam focus.

The electron and positron beams collide at four interaction points (IP's) on the LEP ring each of which contains a detector. The four detectors are shown at their respective LEP interaction points in Figure 11. Although the number of bunchlets in the beam allows collisions at more than the four instrumented interaction points, electrostatic separators at these potential collision sites keep the beams separated in order to extend the beam life.
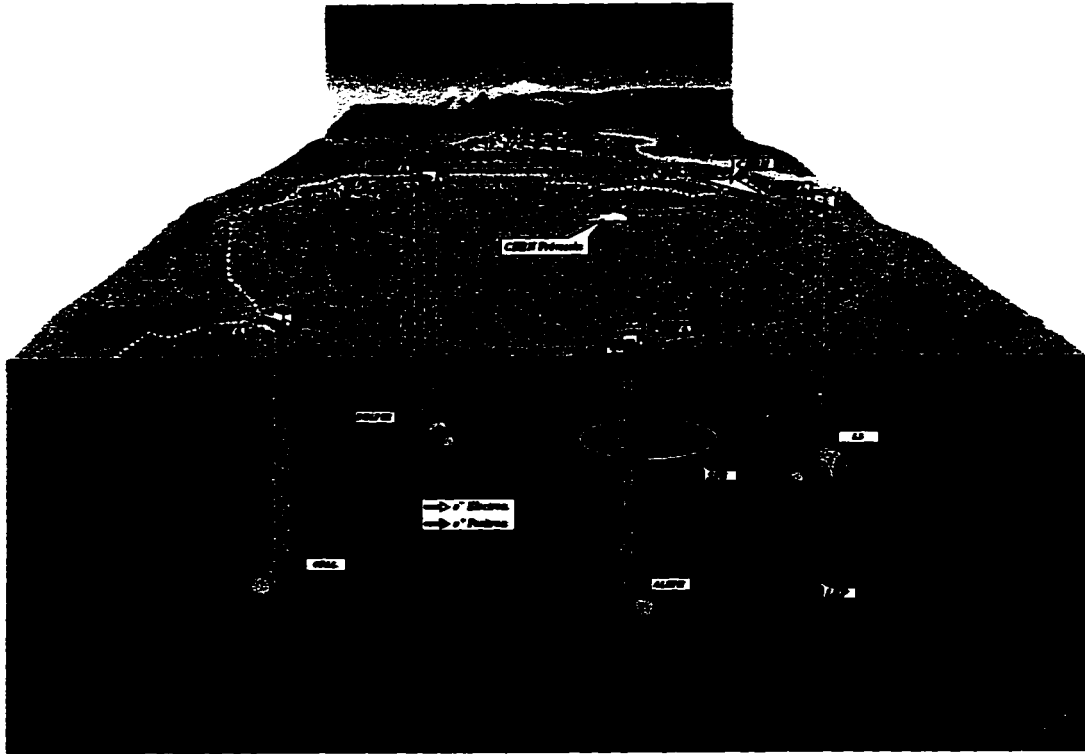
Figure 11: The location of the four LEP experiments at their e^+e^- interaction points. OPAL is located at Interaction Point 6.

The parameter of most importance to experimental physicists is the integrated luminosity ($\mathcal{L}$) which is delivered by the accelerator since the reaction rate, R, for any given physical process is given by:

$$R = \mathcal{L}\sigma$$

where σ is the cross-section (expressed in units of cm^2) of the process of interest. The luminosity itself is expressed in units of $cm^{-2}s^{-1}$ and is dependent on the beam characteristics through the relation (assuming the two beams overlap entirely):

$$\mathcal{L} = fn\frac{N_1 N_2}{A}$$

where f is the frequency of revolution of the beams, n is the number of bunchlets per beam, N_1 and N_2 are the number of particles per bunchlet for the first and second beams respectively, and A is the effective cross-sectional area of each beam.

From 1990 to 1995, the period during which LEP was running at the Z^0 peak, an ever improving understanding of the characteristics of the accelerator allowed an increase in the integrated luminosity† delivered by the machine as shown in Figure 12. With the addition of the superconducting magnets to the LEP ring and changes in the beam bunch structure, however, the properties of the accelerator had to be re-learned by the operators and initial integrated luminosities for the higher energy runs were somewhat low. A total of 10 pb⁻¹ of data were collected at a centre of mass energy equal to 161 GeV during the runs considered in this thesis.
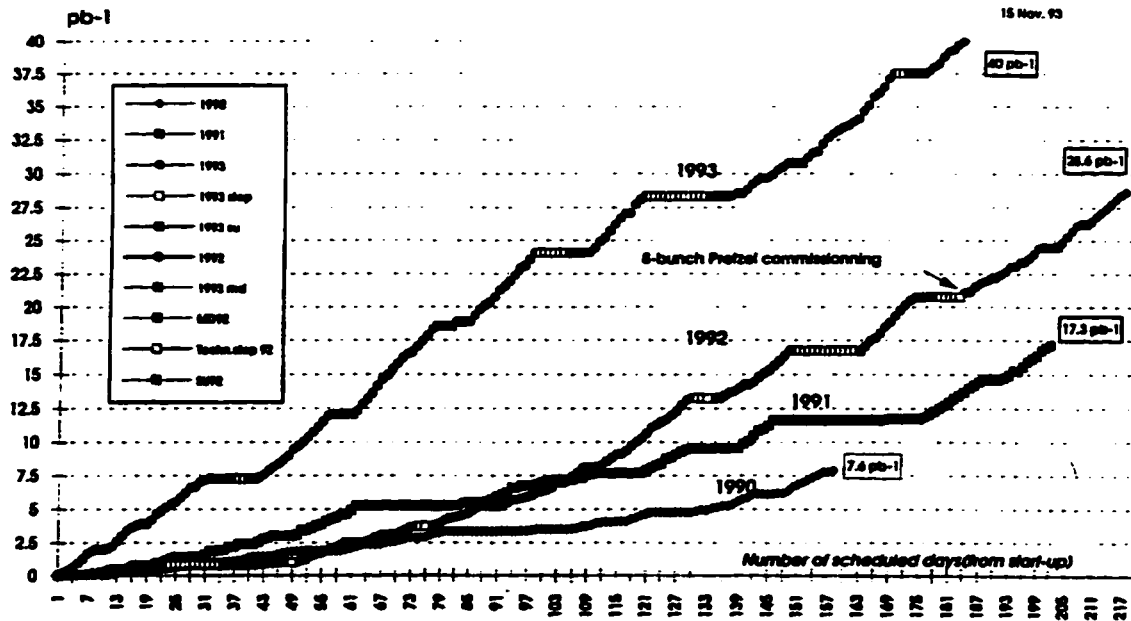
Figure 12: The integrated luminosity delivered by LEP from 1990 to 1993. An improved luminosity was achieved for each year of running at the Z^0 peak.

---

† Integrated luminosity is simply the time integral of the luminosity delivered by the accelerator. It is usually expressed in units of inverse picobarns (pb⁻¹) where 1 pb⁻¹ = 10^{-36} cm².

The luminosity observed by each of the experiments is usually determined through measuring the rate of small-angle Bhabba scattering events with dedicated subdetectors. This process is chosen for luminosity measurements as it is entirely electroweak in nature and hence its cross-section can be calculated very precisely. OPAL uses two dedicated subdetectors to measure the LEP luminosity both of which are described with the rest of the OPAL detector in the next section.

3.3 The Omni Purpose Apparatus for LEP (OPAL)

The Omni Purpose Apparatus for LEP (OPAL) is one of the four large detectors which are located at the interaction points of LEP. Each of the four detectors (ALEPH, DELPHI, L3, and OPAL) was built with different relative strengths in order to maximize the scope of the physics goals attainable at LEP. The OPAL collaboration (see Appendix A) has over 150 members from ten countries.

OPAL [8], shown in Figure 13, exhibits the traditional cylindrical symmetry of an e^+e^- collision detector. The apparatus is centered axially on the LEP beam pipe and is inclined at an angle of 13.9 mrad with respect to the horizontal plane following the slight slope of LEP. The OPAL coordinate system, also shown in Figure 13 is defined so that the z axis lies along the beam pipe in the direction of the e^- beam, the x axis points inward roughly towards the centre of the LEP ring, and the y axis points nearly vertically upward. The polar angle θ is defined as the angle from the $+z$ axis and the azimuthal angle ϕ as the angle from the $+x$ axis. In OPAL parlance, the x-y plane of the detector is referred to as the r-ϕ plane, and the x-z plane is known as the r-z plane.
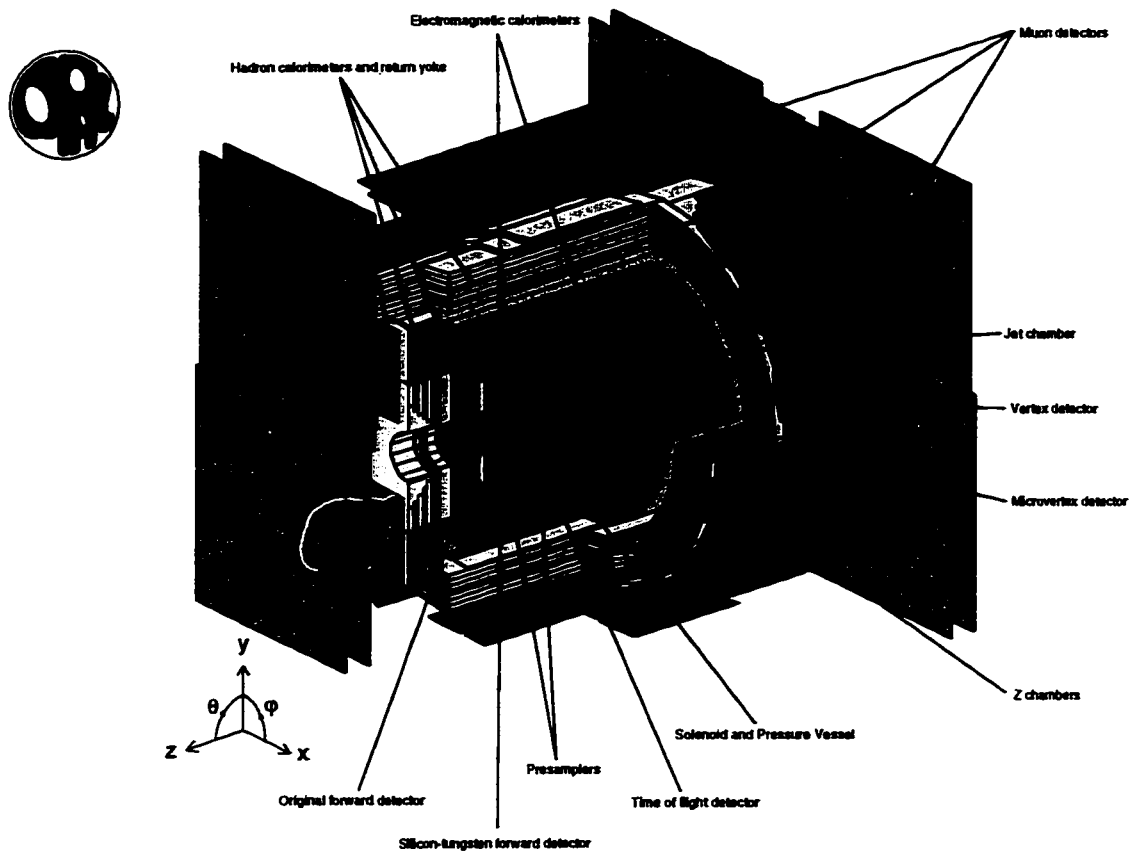
Figure 13: A cutaway view of the OPAL detector.

OPAL was designed to incorporate relatively conservative technology in order to provide a reliable multipurpose detector which would be able to detect, reconstruct, and unambiguously identify all types of interactions occurring in e^+e^- collisions. Specifically, the original detector was required to perform the following tasks over a solid angle of nearly 4π radians.

- Tracking of charged particles within the region of a central solenoidal magnetic coil thereby measuring their direction and momentum, particle identification through energy loss, and accurate reconstruction of primary and secondary vertices near the e^+e^- interaction point.

- Accurate identification of electrons and photons and measurement of their energies in the electromagnetic calorimetry.

- Measurement of the total hadronic energy of each event using the magnet yoke instrumented as a calorimeter.

- Identification of muons using position and direction measurements both within and outside of the hadron absorber.

- Measurement of the absolute LEP luminosity by detecting Bhabba[†] scattering events close to the beam line.

OPAL consists of several detector systems each of which fulfills some of the experiment's design goals. In order of increasing radial distance from the beam pipe, these systems consist of the central tracking system used to determine the direction and momentum of charged particles, the time of flight system which identifies some charged particles and rejects cosmic rays, the electromagnetic calorimeter system which measures the energy of photons and electrons, the hadronic calorimeter system used to measure the energy of hadronic showers, and the muon detector system which identifies muons at the outer edge of the detector. In addition to these main detector systems, OPAL also includes detectors close to the beam pipe in order to measure the LEP absolute luminosity.

Since the commissioning of the detector in 1989, several upgrades have enhanced OPAL's ability to satisfy its original specification. These improvements include:

- Installation of the Silicon Microvertex Detector close to the beam pipe. This detector allows even better resolution of the primary and secondary vertices near the e+e- interaction point. The original single dimensional readout Silicon Microvertex

---

[†] A Bhabba event is the electroweak scattering of the electron and the positron in the beam.

Detector was installed in 1991 and was replaced by an improved two dimensional readout model in 1993.

- Installation of the Silicon Tungsten Luminometer in order to provide a more accurate measurement of the absolute LEP luminosity at the OPAL interaction point.

- Installation of the Time of Flight Endcap detector which enlarged the time of flight system to cover more of the solid angle of the detector. This increase in coverage was designed to provide accurate prompt times for triggering in the LEP2 environment.

The hierarchy of the OPAL detector subsystems is summarized in Table 5. Each of the detector subsystems as well as the triggering and online data acquisition systems are described in detail in the following sections.

System	Detector
Central Detector	Silicon Microvertex Detector
	Vertex Chamber
	Jet Chamber
	Z Chambers
Time of Flight System	Time of Flight Barrel Detector
	Time of Flight Endcap Detector
Electromagnetic Calorimeter System	Electromagnetic Barrel Calorimeter
	Electromagnetc Endcap Calorimeter
Hadronic Calorimeter System	Hadronic Barrel Calorimeter
	Hadronic Endcap Calorimeter
	Hadronic Poletip Calorimeter
Muon Detection System	Muon Barrel Detector
	Muon Endcap Detector
Luminosity System	Forward Detectors
	Silicon Tungsten Luminometer

Table 5: Hierarchy of the OPAL subdetectors.

3.3.1 The Beam Pipe and Magnet

Prior to the 1991 run, the OPAL beam pipe consisted of a series of overlapping layers of carbon fibre epoxied to the outside of a 0.1 mm thick aluminum tube. This original beam pipe had a total thickness of 1.4 mm and was placed at an inner radius of 78 mm from the beam. In 1991, however, installation of the Silicon Microvertex Detector (described in

3.3.2) at an inner radius of 61 mm from the beam necessitated a new beam pipe geometry. This new pipe consisted of 0.11 mm thickness beryllium and was placed at a radius of 53.5 mm from the beam. The old beam pipe was replaced by a 2.0 mm thick carbon fibre tube at a radius of 80 mm from the beam. The Silicon Microvertex Detector was then placed in between the two structures. Beryllium was chosen as the material because of its long radiation length[†] as well as the ease with which it can be machined into a thin layer. The OPAL beam pipe causes only 0.31% of a radiation length of dead material when measured in a direction perpendicular to the beam direction.

The OPAL magnet consists of two parts: the solenoidal coil and the return yoke. The solenoidal coil lies at the interface of the central detector and the calorimetry at a mean diameter of 4.36 m from the beam. It is constructed of a hollow aluminum conductor fused together with glass-epoxy. The return yoke is made of iron and forms the absorber material for the hadron calorimeter. The apparatus operates at a current of about 7000 A and is water cooled. Within the solenoidal coil, the central detector magnetic field is maintained at 0.435 T with a uniformity of ±0.5%. The magnetic field in the calorimetry region between the solenoid and the return yoke is only a few tens of gauss which allows the photomultiplier tubes of the electromagnetic calorimetry (see section 3.3.5) to be operated without difficulty.

3.3.2 The Silicon Microvertex Detector

The OPAL silicon microvertex detector was a relative latecomer to the experiment being installed some two years after the initial commissioning of OPAL. As will be seen in the sections that follow, the silicon microvertex detector is of particular importance to searching for the Standard Model Higgs Boson in the four jet final state. For this reason, it

---

[†] One radiation length, X_0, of a given material is defined as the distance over which a high energy electron travelling in the material loses all but $1/e$ of its energy due to bremsstrahlung.

is described in a section of its own instead of being grouped with the rest of the central detector in section 3.3.3.

The power of the silicon microvertex detector lies in its ability to track particles at a point so close to the e^+e^- interaction point that these tracks may be traced back to their origin with very little error and with an extremely precise position resolution. The process of extrapolating particle tracks back to a common origin is known as vertex finding and has important applications in particle lifetime identification.

When a particle has a "long" lifetime (on the order of $c\tau=10$ nm), it can travel a considerable distance (up to the order of mm) from its point of production before decaying. The point of decay of this long lived particle is therefore well separated from the e^+e^- interaction point and is referred to as a secondary vertex. The particles resulting from this secondary decay can be traced back to the secondary vertex and the long lived particle's lifetime deduced from the distance of the secondary vertex from the primary e^+e^- interaction point.

Long lived particles of interest include the τ lepton and b flavoured hadrons. The lifetimes of these particles have been measured with great precision by OPAL since the installation of the microvertex detector. In addition, the secondary vertex finding by the silicon microvertex detector allows jets originating from b quarks to be efficiently tagged which, as was seen in section 2.3, is of great importance in separating signal from background in the four jet final state of HZ production at LEP.

Semiconductor detectors operate in a manner analogous to drift chamber devices in that ionization caused by passing charged particles is used for detection purposes. In a semiconductor device, the medium is a solid semiconductor in which passing ionizing particles produce electron-hole pairs which are then collected by an electric field. A minimum ionizing particle in a typical silicon detector creates on the order of 10,000 electron-hole pairs. Semiconductor detectors need 10 times less energy to cause an

electron-hole pair than is required in a typical drift chamber to produce an electron-ion pair. Thus, they are very sensitive devices for recording the passage of charged particles.

In order to accurately place secondary vertices, a detector must be very close to the beam interaction point and have excellent spatial resolution qualities. As will be seen below, the OPAL silicon microvertex detector satisfies both of these requirements well.

History of the OPAL Silicon Microvertex Detector

OPAL did not contain a microvertex detector when it was first used to collect data in 1989. After the first LEP run, however, it was found that background radiation close to the beam pipe was at a lower level than first expected. This allowed the beam pipe radius to be reduced and hence left room for a new detector in OPAL between the vertex chamber and the beam pipe.

The first OPAL silicon microvertex detector [9] was installed and began data collection in June 1991. This version of the detector was a single sided silicon device and which only allowed readout in one dimension (ϕ) at a known radius. Nonetheless, the new detector allowed a single hit resolution of 8 μm and gave OPAL the ability to resolve impact parameters[†] to within 15 μm [9]. Prior to 1991, OPAL was only capable of an impact parameter resolution of 42 μm (for a 45 GeV muon) using the jet chamber and the vertex chamber in concert. This improvement in the resolution of impact parameters, and hence in the accuracy of secondary vertex finding, directly resulted in the successful precision measurements of the τ lepton and b hadron lifetimes published by OPAL [10].

In 1993, the silicon microvertex team implemented their design for a two coordinate readout silicon microvertex detector [11]. This device used double sided silicon detectors

---

[†] The *impact parameter* (d_0) of a track is defined as the distance from the OPAL coordinate origin to the point of closest approach of the track. The impact parameter may be signed according to the side of the origin on which the track passes and is used to calculate the position of the track's origin. A more detailed description of the impact parameter can be found in section 3.5.1.

and was capable of readout in both ϕ and z coordinates. In 1995, the detector was enlarged in order to cover more of the interaction region and to close some gaps in ϕ. Finally, in 1996, the detector was again lengthened along the z axis in order to improve its coverage of the e^+e^- collision region. The sections below describe this latest version of the silicon microvertex detector which was present in OPAL for the 161 GeV LEP run during the summer of 1996.

Construction and Layout of the Detector

The OPAL silicon microvertex detector is shown in partial cutaway view in Figure 14. The basic modular unit is a 33 mm x 60 mm wafer which consists of back to back ϕ and z readout silicon detectors. These wafers are connected end to end to form structures called *ladders*. The detector as a whole is made up of two radial layers of these ladders. The inner detector layer consists of a cylinder formed from 12 two wafer ladders at the $-z$ end of the detector and 12 three wafer ladders at the $+z$ end of the detector. The inner layer is placed at a radius of 61 mm from the beam. Similarly, the outer detector layer is a cylinder formed from 15 two wafer ladders at the $-z$ end and 15 three wafer ladders at the $+z$ end. This outer cylinder is at a radius of 75 mm from the beam. Thus, each detector layer consists of a total of five detector wafers longitudinally with the e^+e^- interaction point located at the center of the five wafers. As can be seen in Figure 15, the ladders of the two layers are staggered so as to avoid lining up the small gaps in ϕ which allows a single hit efficiency of nearly 100%.
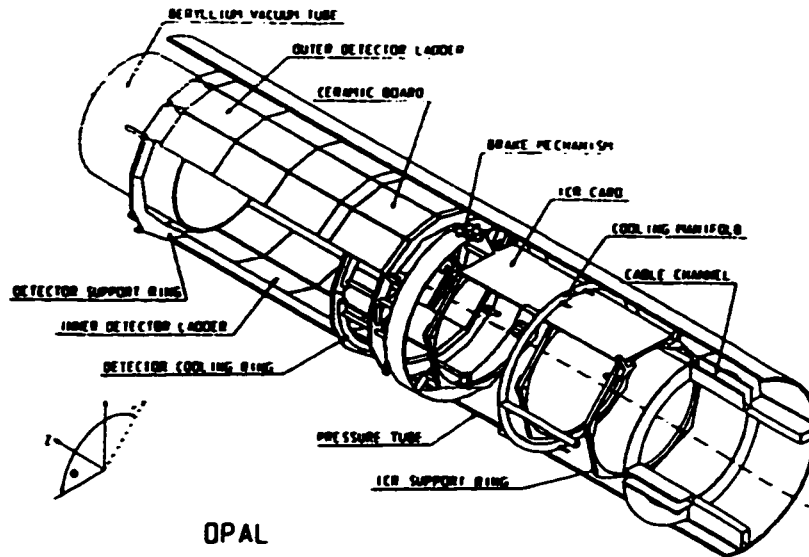
OPAL

Figure 14: A cutaway view of the OPAL silicon microvertex detector.
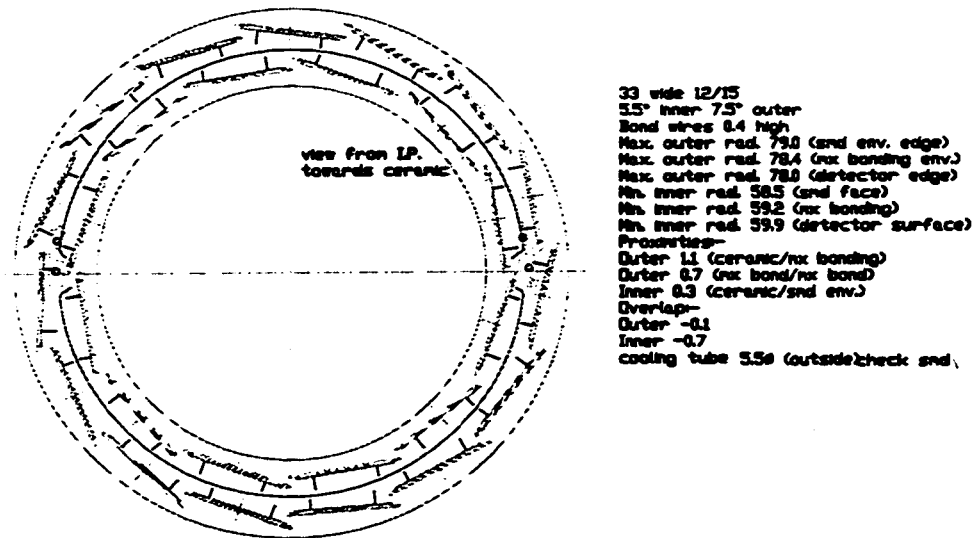
Figure 15: Cross-sectional view of the silicon microvertex detector in the r-φ plane. The radial overlap of the detector ladders is shown in both detector layers.

The three wafer ladder shown in Figure 16 gives an overview of the active detector layout. The active detector wafers are single sided AC coupled FoxFET (Field Oxide Field Effect Transistor) biased silicon strip detectors which were designed by OPAL and manufactured by Micron Semiconductor Ltd. The wafers are made of high resistivity (>5 kΩ cm) n-type bulk silicon with n$^+$ type doping and metalization on the backplane and p type implant strips on the front (readout) side. The inner side of the wafer forms the ϕ detectors with implant strips pointing in the z direction while the outer side makes up the z detectors with implant strips pointing in the direction of ϕ. Each of the active detectors (ϕ and z) has a thickness of 250 μm. Above every other implant strip on the ϕ detectors and every fourth strip on the z detectors, metalized aluminum readout strips are placed with a silicon oxide layer in between to form an AC coupling capacitor to allow multiplexing[†] of the detector readout. A p type implant drain runs around the active strip detection area which therefore forms a p-n-p junction for the entire detector. The implant strip pitch is 25 μm for both the ϕ and z detectors, but the effective readout pitch is 50 μm for the ϕ detectors and 100 μm for the z detectors due to the readout multiplexing.
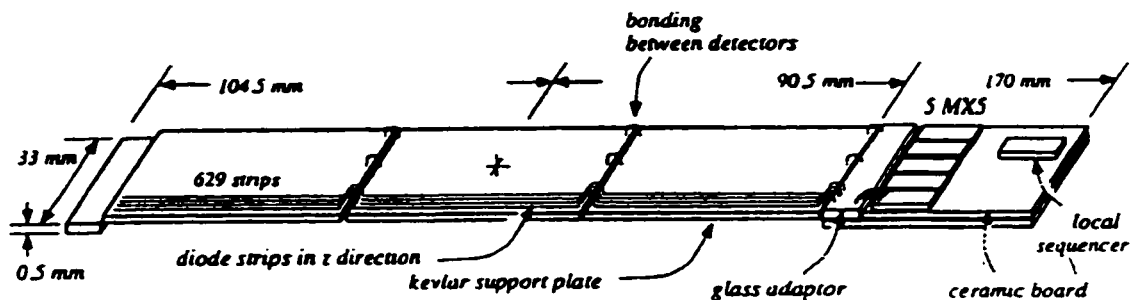
Figure 16: Schematic diagram of a silicon microvertex detector ladder.

The depletion voltage for the detectors lies in the range 15-40 V. The bias voltage applied between the drain line and the backplane, however, is in the range of 35-60 V. The extra

---

[†] Multiplexing a detector refers to reading more than one channel out through a single line thus saving both money and space.

voltage above the depletion voltage compensates for the voltage drop over the strip to drain channel and guarantees full depletion of the entire detection region for all detectors on a ladder.

The detector is supported on three support rings one of which carries the water cooling for the detector. Structural rigidity is supplied to the detector by two semi-cylindrical beryllium shells which were chosen for their low contribution to the effective radiation length of the detector. The entire detector fits between the 53.5 mm radius of the beam pipe and the 80 mm inner radius of the vertex chamber.

An important design consideration for the silicon microvertex detector was that it present as little material as possible to particles traversing the detector. This goal has been achieved with the detector only presenting 1.5% of a radiation length at $90^{\bullet}$ to the beam direction. The bulk of this amount (1.1% of a radiation length) comes directly from the silicon wafers.

The Readout Electronics

The ϕ readout strips on the active detector are connected to the front end electronics located on 300 µm thick ceramic circuit boards at the end of each ladder. The z detector strips are also read out into the front end electronics at the end of each ladder through the use of a gold printed circuit on a borosilicate glass support which daisy chains the signals of the z strips to the end of the ladder without placing an unnecessarily large amount of material in the way of the particles. The front end electronics consist of 5 MX7 Microplex VLSI readout chips which receive the detector signals and a local sequencer chip to control them. The local sequencers are controlled through Interconnecting Ring (ICR) card logic by a Fastbus Master Sequencer module located in the electronics hut. The local sequencers and MX7 chips also allow calibration pulses to be injected into the front end readout of each channel.

The analogue readout signals from the front end electronics exit the detector through the ICR cards and are processed by Fastbus SIROCCO IV modules in the electronics hut which digitize, store, and process the data.

Online Data Processing

The online processing is performed by 28 Digital Signal Processors (DSP's) in the SIROCCO modules. The DSP's perform a channel specific pedestal subtraction and then apply a cluster algorithm to find charge clusters in the SI channels. By only keeping good clusters, the data volume from the microvertex detector is reduced by a factor of 100. The clustering algorithm operates by adding the signals of two (or three) neighboring strips and comparing that value with the quadratically summed noise of the same strips. If the signal exceeds the noise by a factor ~3, both strips are marked as hits.

The entire triggering and online procedure results in a dead time of 1.6 ms for a nominal trigger rate of 4-10 Hz. This is very small compared to the overall OPAL deadtime and is dominated by the analogue readout and digitization sequence in the SIROCCO modules.

Detector Operation Considerations

The environment of the microvertex detector is very important to its stable operation. Temperature fluctuations are monitored by 14 thermistors which are accurate to ±0.05 °C and send warnings to the operator in case of overheating. Radiation is monitored by detectors attached to four of the seven ICR cards. These detectors are also connected to alarms in the control room.

The microvertex is operated in an environment of dry nitrogen. This gas minimizes humidity effects which could lead to large leakage currents in the detector which would destroy its data taking ability. Under normal operating conditions, the leakage current in the microvertex is less than 1 μA.

Detector Performance within OPAL

The OPAL silicon microvertex detector has proven to be very reliable and has a 99% working channel rate. The signal to noise ratios are on the order of 20 for both the ϕ and the z detectors.

The standard measure of resolution performance for the silicon microvertex detector is the impact parameter (d_0) resolution for dilepton events. In these events, the process $Z^0 \rightarrow \mu^+\mu^-$ gives two muons which originate at the primary vertex. By plotting the difference of the impact parameters calculated by the microvertex for these tracks, the impact parameter resolution is obtained.

Early results with the two coordinate readout silicon microvertex detector reported a resolution of 21 μm for d_0 in r-ϕ and 100 μm in z. This has improved [11] to 18 μm in r-ϕ for all 1995 data while the z resolution has remained the same.

Finally, Figure 17 and Figure 18 compare the impact parameter resolution for dimuons with and without the silicon microvertex detector respectively. The improvement in resolution is very clear.
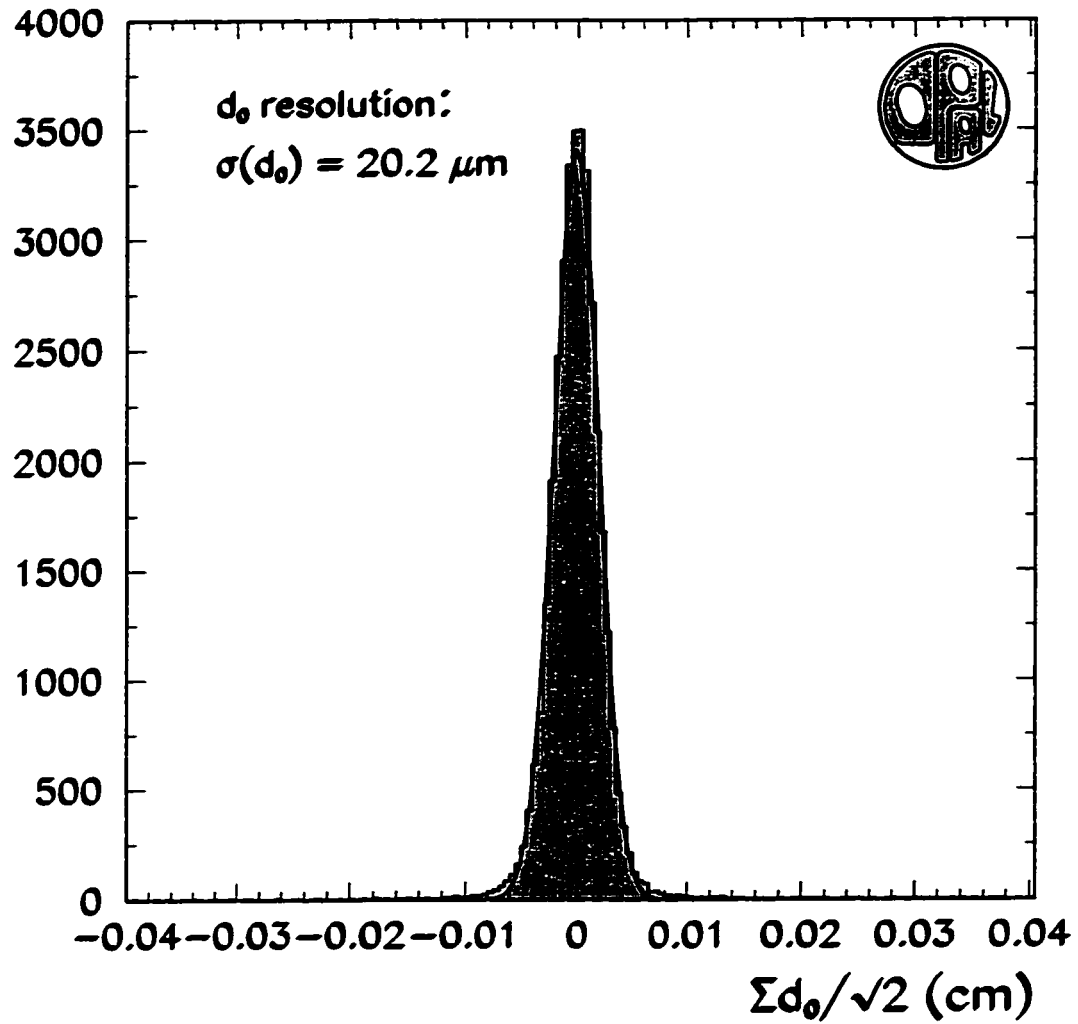
Figure 17: The impact parameter resolution for dimuons using silicon microvertex detector information in conjunction with the central tracking detectors. The data is taken from the 1993 LEP run.
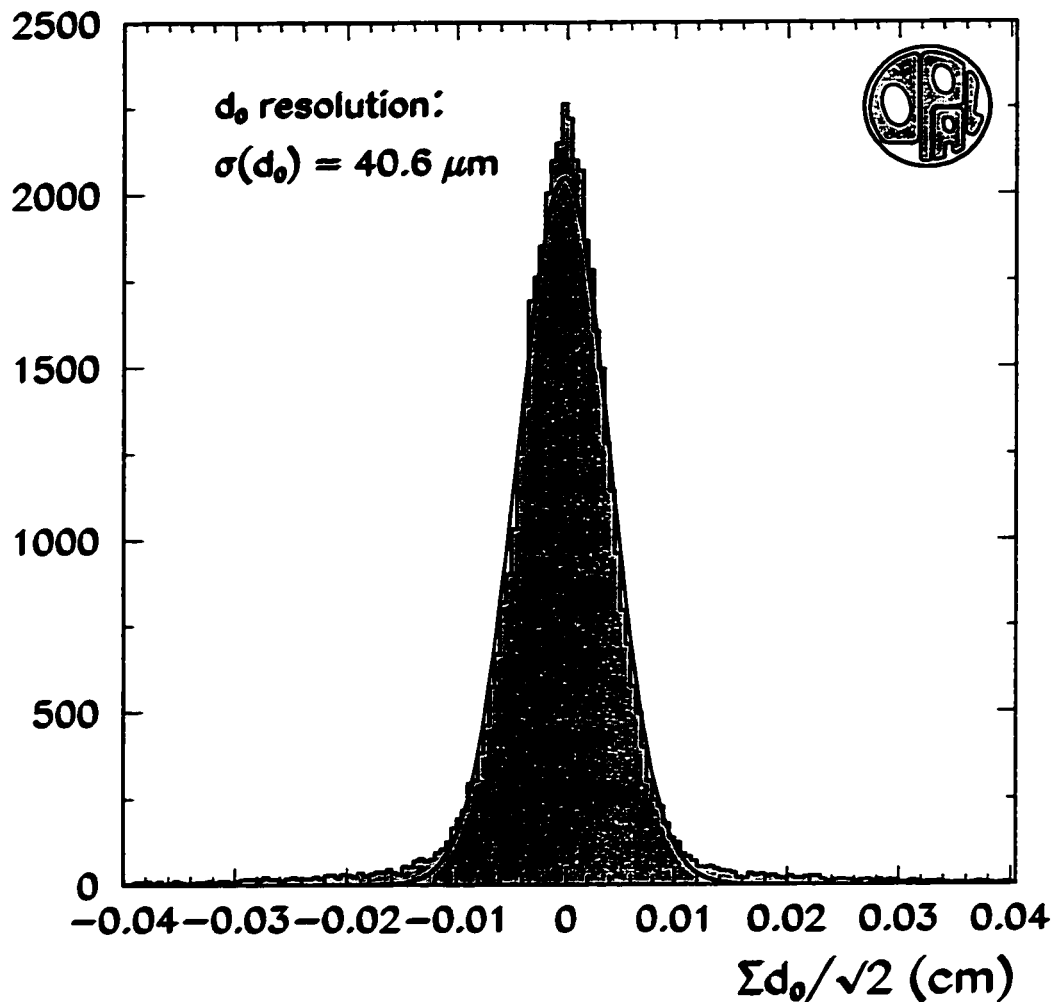
Figure 18: The impact parameter resolution for dimuons using only central tracking detector information. The data is taken from the 1993 LEP run.

3.3.3 The Central Detector

The components of the OPAL central detector, which provides all charged particle tracking information in the detector, may be seen in the cutaway detector view in Figure 13. In order of increasing radial distance from the beam pipe, the central detector consists of the silicon microvertex detector, the vertex chamber, the jet chamber, and the z

chambers. Acting in concert, these three detectors track charged particles from the interaction point until their entry into the calorimetry.

The vertex chamber, jet chamber, and z chambers are all *drift chambers* and therefore operate under the principle that a charged particle traversing an active gas volume will ionize the gas and liberate electrons. These electrons then drift in an applied electric field to anode wires and produce an avalanche signal which is detected some time after the passage of the charged particle. With a knowledge of the characteristic drift time for electrons in the detector gas, the position at which the particle traversed the chamber can be deduced from the time taken for the electrons to reach the anode wire. The drift chambers of the central detector use a mixture of gases consisting of 88.2% argon, 9.8% methane, and 2.0% isobutane in their active volumes.

The central detector drift chambers are located within the pressure bell and are kept at a pressure of 4 bar. The central detector is also within the 0.435 T field of the magnet solenoid. The curvature in charged particle trajectories due to the magnetic field is used to determine the momentum of the passing particle. Charged particle trajectories from all three drift chambers and the silicon microvertex detector are combined to calculate the final track parameters for each particle. These track parameters and their meanings will be described in section 3.5.1.

The individual components of the central detector are described below with the exception of the silicon microvertex detector which was described in section 3.3.2.

The Vertex Chamber

The OPAL vertex chamber [8] [12], pictured in Figure 19, was the closest tracking detector to the beam pipe prior to the installation of the silicon microvertex detector in 1991. The chamber itself is a 1 m long cylindrical drift chamber with inner and outer radii of 88 and 235 mm respectively. The main tasks of the vertex chamber are to measure the

position of secondary vertices from long-lived particles and to improve the overall momentum resolution of the central detector system.
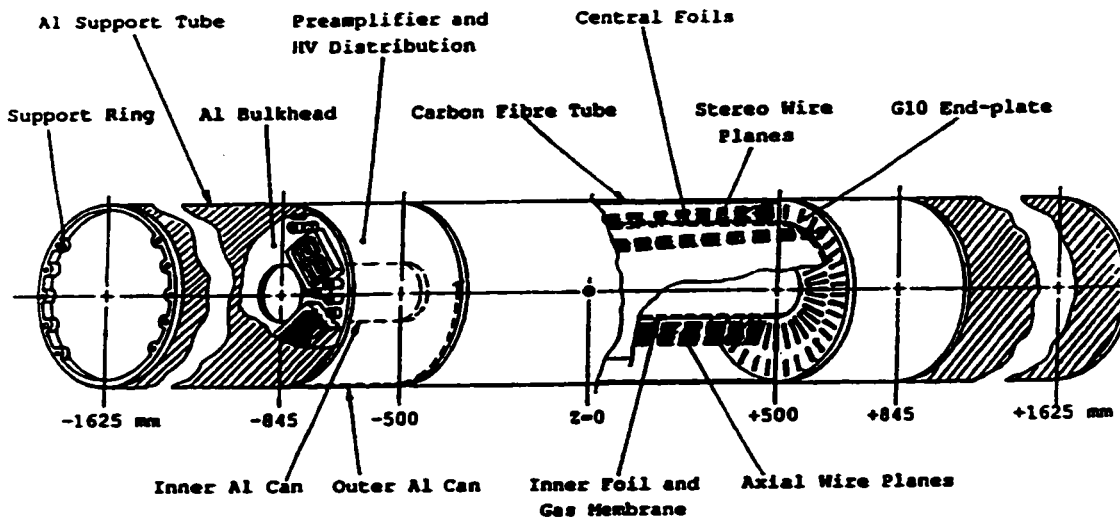
Figure 19: The OPAL central vertex chamber.

The wiring layout of the vertex chamber is shown in Figure 20. The volume of the chamber between radii of 90 and 170 mm is divided into 36 axial sectors with 12 anode wires each stretched parallel to the beam. The spacing of the anode wires is 5.3 mm. The remaining volume is divided into 36 stereo sectors with 6 anode wires each at a stereo angle of 4°. The spacing of the anode wires in these sectors is 5.0 mm. All anode wires are made of 20 μm diameter tungsten-rhodium alloy.

Each anode wire defines the centre of a readout cell which is bounded by the cathode wires and potential wires used to maintain the electric field in the chamber. The vertex detector contains 648 such cells. The potential wires are made of 200 μm diameter gold plated copper-beryllium while the cathode wires are 125 μm diameter unplated copper-beryllium. The cathode wires are spaced 1 mm apart in the r-φ plane. Since the cells are bidirectional, that is the electrons drift to the anode wires from either side, a given drift time does not uniquely specify on which side of the anode wire the track passed. In order to resolve this

left-right ambiguity, the anode wires are offset from the centre line of the cell by a ±41 μm alternating stagger. Thus, a given set of drift times will only reconstruct a good track on the correct side of the anode wire plane.

Cathode planes

Carbon fibre tube

r=235 mm

Foil

r=213 mm

Stereo cells

r=188 mm

r=175 mm

Foils (2)

r=162 mm

Axial cells

Field shaping rings

r=103 mm

Foil

r=88 mm

Gas membrane

Anode (×) - potential (•) wire planes
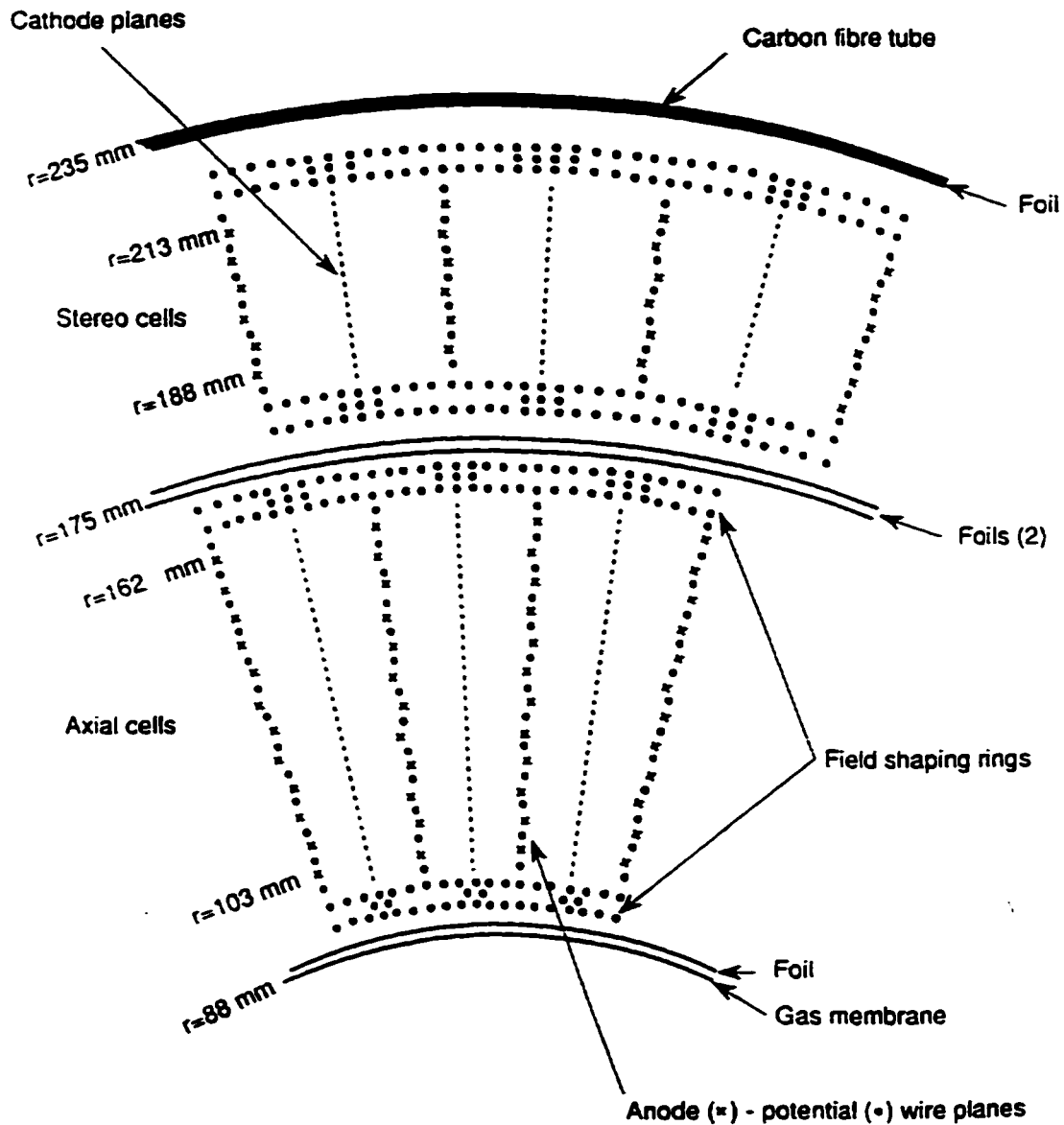
Figure 20: Wire layout of the vertex detector in the r-φ plane.

Each anode wire is connected to a preamplifer at both ends and the resulting signals from electron avalanches are transferred to constant fraction discriminators. The two outputs from each wire are then averaged in a mean timer and the average time digitized by a time to digital converter (TDC) with a bin size of 0.67 ns. This bin size corresponds to a drift distance of 26 μm and therefore is not the limiting performance factor for the detector. In addition to the mean time used in the drift time measurement, a measurement of the difference in arrival time of the signal at each end of the anode wire is performed in order to obtain a crude z position measurement. This measurement is made to an accuracy of 0.1 ns which corresponds to a rough z resolution of 4 cm.

The performance of the vertex chamber is determined by two electrostatic fields: the anode surface field and the drift field. The anode wires are held at ground and the fields are established by manipulation of the voltages on the cathode and potential wires. The anode surface field determines the gas gain and therefore the detection efficiency of the chamber. When operating with the standard anode surface field of 360 kV/cm, the single hit efficiency of the vertex chamber is close to 97%. The drift field within the chamber is 2.5 kV/cm which gives rise to a drift velocity of 39.5 μm/ns in the standard central detector gas. Adjustment of the drift field directly affects the spatial resolution of the detector which is measured to be around 50 μm in r-ϕ over most of the chamber volume under standard operating conditions. The z coordinate information contained in the drift times from the stereo wires is reconstructed in the offline analysis and hence only the end to end timing z position is available for online triggering. The maximum drift distance in the vertex chamber is about 1.5 cm which limits the effects of electron diffusion on the spatial resolution. A two particle seperation resolution of 2 mm is achieved through reducing the fall time of the anode wire pulses in the readout electronics to the point where detector dead time is only 40-50 ns.

The Jet Chamber

The jet chamber [8] [13] is a 4 m long cylindrical drift chamber with inner and outer radii of 0.25 m and 1.85 m respectively. The detector is made up of 24 azimuthal sectors each containing one anode wire plane consisting of 159 wires strung parallel to the beam direction. The anode wires themselves are grounded and alternate with potential wires between radii of 25.5 cm and 183.5 cm. The radial spacing of the anode wires is 1.0 cm. Like the vertex chamber, the anode wires in the jet chamber are staggered at ± 100 µm from the potential wire plane in order to resolve the left right ambiguity. The maximum drift distance in the jet chamber varies from 3 cm for the innermost anode wires to 25 cm for cells at the outer radius of the chamber.

The jet chamber geometry is arranged such that up to 159 individual points can be measured along each track in the polar angle range $43° < \theta < 137°$. At least eight points can be measured along any track for a solid angle of over 98% of the full 4π radians.

The signal from both ends of a given anode wire are read out into preamplifiers mounted on the chamber end plates and are then digitized by 100 MHz flash analog to digital converters (FADC) which measure the charge collected on the anode wire in $1/10^8$ s intervals. These high frequency FADC's were selected as their sampling speed greatly improves both the drift time measurement and the two particle separation distance of the detector. After the pulses have been digitized by the FADC's, the data is passed to twenty dedicated microprocessors where fast online track finding is performed for use in the trigger.

The r-ϕ position of the track is obtained directly from the drift time. This drift time is calculated in the dedicated microprocessors by analyzing the shape of the charge distribution in the FADC's. A crude z measurement with a resolution of a few centimeters is made possible by a charge division technique. This technique uses the ratio of the integrated charge collected on one end of an anode wire to that collected from the other

end to determine the longitudinal position of the hit along the wire. The accuracy of this technique increases with the magnitude of the drift field in the chamber as a higher gas gain increases the total amount of charge collected and hence the significance of the charge ratio between the two ends of the anode wires.

The integrated charge collected in the FADC's is also used in conjunction with the track momentum to perform particle identification. The chamber is therefore operated with a drift field chosen to optimize both the z coordinate resolution and the dE/dx particle identification potential. As described above, a high gas gain is necessary in order to optimize the z position resolution. For the dE/dx, however, high gas gain leads to saturation effects where the full drift field is screened by ions formed in the avalanche process thus blocking some charge from reaching the anode wires. The amount of this lost charge is difficult to correct for and hence increases the uncertainty in the dE/dx measurement. A compromise of 890 V/cm is chosen for the drift field which gives rise to a spatial resolution of 130 μm in r-ϕ and 10 cm in z.

The position resolution of the chamber allows the momentum of particles to be measured within a resolution determined by the relation:

$$\left(\sigma_p / p\right)^2_{xy} = 0.02^2 + 0.0015 p^2_{xy}$$

in the region $|\cos\theta| < 0.7$ where p_{xy} is the measured momentum in the x-y plane in GeV/c. The constant term in this equation does not come from the resolution limitations of the chamber as one might expect, but rather from the systematic effect of multiple Coulomb scattering in the jet chamber gas.

One of the strongest attributes of the OPAL detector is its ability to identify particles through dE/dx in the jet chamber. This procedure [14] is based on the Bethe-Bloch equation [15] for the energy loss of charged particles in matter. The energy loss of a charged particle decreases as $1/\beta^2$ at low values of $\beta\gamma$ until it reaches a minimum at around

βγ=4. The parameter b is simply the velocity of the particle in units of the speed of light

and $\gamma = 1 / \sqrt{1 - \beta^2}$. After this minimum, the energy loss rises logarithmically in what is

called the *relativistic rise* followed by a saturation which is called the *Fermi plateau*. Since

the energy loss is a function of βγ for all particles, the measurement of both the momentum

and the dE/dx (energy loss per unit distance) of the particle determines the particle's mass.

The curves in Figure 21 show the expected energy loss (dE/dx) for various particle types

versus momentum in the OPAL jet chamber.
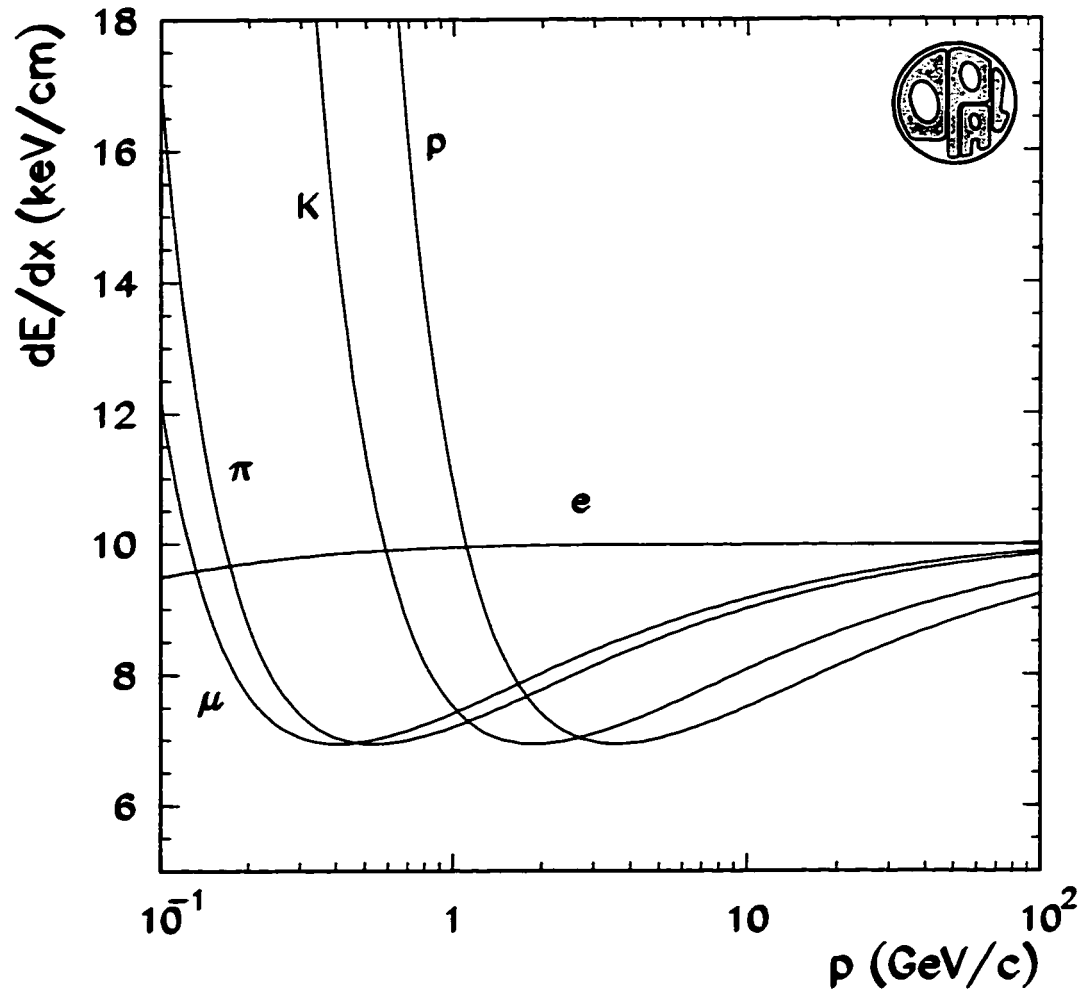
Figure 21: Energy loss per unit distance (dE/dx) versus momentum predicted for pions, muons, electrons, kaons, and protons in the OPAL jet chamber.

The particle separation power† of the jet chamber is maximized at a gas pressure between 3 and 4 bar and it is the optimization of the separation power that is one of the main reasons for operating the OPAL jet chamber at 4 bar. Many individual readings of the dE/dx along a particle's track are necessary in order to minimize the uncertainty of the mean dE/dx for that track when performing particle identification.

The dE/dx resolution for minimum ionizing pions in the OPAL jet chamber has been found to be 3.8% for tracks with at least 130 individual dE/dx measurements. This value is sufficient for OPAL to have the distinction of containing one of the most accurate jet chambers for particle identification in existence today.

The Z Chambers

The z chambers [8] [16] consist of 24 drift chambers each of which is 4 m long, 50 cm wide, and 59 mm thick. These chambers are mounted on the support structure of the jet chamber and collectively form a 4 m long barrel with a diameter of 3.85 m. A single z chamber is shown in Figure 22.

---

† The *particle separation power* of the jet chamber is defined as the dE/dx resolution divided by the difference in dE/dx between the two particles types that are to be separated.
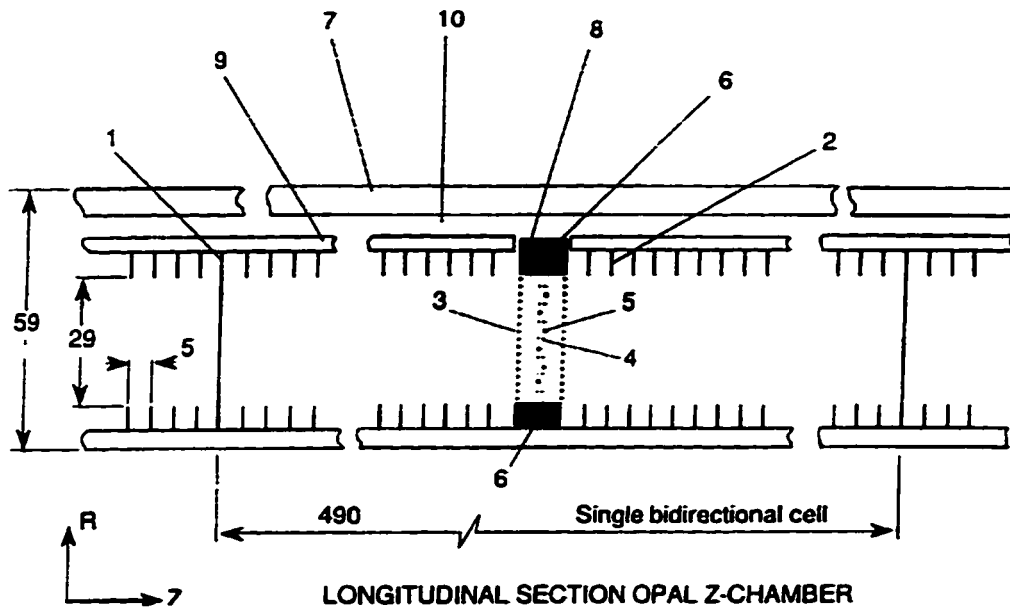
Figure 22: Cross-section of a single z chamber.

Each cell contains six anode wires spaced 4 mm apart at increasing radii. The anode wires are staggered at ±250 μm in order to resolve left-right tracking ambiguities. The chambers are operated with a drift field of 800 V/cm which allows a resolution of around 150 μm in z. A charge division technique gives an r-ϕ spatial resolution of 1.5 cm from the z chambers.

3.3.4 The Time of Flight System

The Time of Flight (TOF) systems was designed to allow charged particle identification through measurement of the time taken for the particle to reach the detector from the interaction point. Charged particles with energies in the range of 0.6 - 2.5 GeV may be identified in this way. The TOF counter also serves as an important veto for cosmic ray events as it protects against the recording of events which are out of time synchronization with the LEP beam crossings. The TOF detector is also essential for the trigger (described

in section 3.3.9) to be able to deduce the beam crossing in which an interesting event occurred.

The Time of Flight Barrel Detector

The barrel section of the TOF detector covers the polar angle range $|\cos\theta| < 0.82$. It consists of 160 scintillating counters which form a cylinder 6.84 m long with a radius of 2.36 m. Thus, the TOF barrel is situated just outside of the magnetic coil.

Each scintillating counter which makes up the TOF barrel is 6.84 m long and 4.5 cm thick. The cross-section of each counter is trapezoidal in shape and increases in width from 8.9 cm at the narrow end to 9.1 cm at the wide end. Light is collected at both ends of each counter by 30 cm long plexiglass light guides which are directly coupled to phototubes. The signal from the phototubes is used to both determine the time of arrival of the signal and the total charge deposited by the passing particle.

The timing resolution of the TOF barrel detector is determined by measuring the time of flight of muons from the decay $Z^0 \rightarrow \mu^+\mu^-$. This gives a time resolution of 300 - 400 ps depending on the position in z at which the measurement is taken.

The Scintillating Tile Endcap Detector

The Scintillating Tile Endcap (TE) design is based upon plastic scintillator "tiles" which are embedded with wave length shifting fibre in order to allow the photomultiplier tubes to reside outside of the detector. The detector was designed to complement the Time of Flight Barrel detector and to provide improved triggering information in the forward region of OPAL. The TE, in conjunction with the Time of Flight Barrel detector, also allows the accurate determination of the collision time of the LEP bunchlets at the centre of OPAL. This prompt time information is useful as it allows the correction of cluster energies in the endcap calorimeters which were recorded with a fixed gate time which is not tuned to the arrival of individual bunchlets.

The TE detector, shown in Figure 23, is placed in between the endcap presampler wire chambers (see section 3.3.5) and the electromagnetic endcap calorimeter in a space that was originally left as tolerance in the OPAL design. The TE is made up of three radial subsectors of tiles each of which is divided into 24 bins in ϕ. The tiles are constructed using 10 mm thick Bicron BC408 scintillator and are all embedded with 1 mm in diameter wavelength shifting (WLS) fibres with an emission peak at 500 nm. The geometry of the WLS fibres within the tiles was chosen to maximize the light collection efficiency while allowing for the maximum bending radius of 40 mm for the fibres. Each 1/24 sector of the TE consists of a single tile in the innermost radial layer and two tiles in each of the outer layers. This granularity of tiles was chosen due to the mechanical constraints of installing the TE next to the existing wire chambers in the endcap presampler.

The WLS fibre is attached to a 1 mm in diameter clear fibre using a custom connector immediately upon exiting the tile. The clear fibre then transmits the light an average of 15 m to the photomultiplier tubes.

Initial results from the TE detector indicate that its timing resolution goal of 5% will be met and perhaps exceeded. Full tests are being performed and the detector should be fully operational for the 184 GeV LEP run in the summer of 1997.

pressure
bell

presampler

TE scintillator

ECAL
lead-glass

1/24 sector

2 WLS fibres per groove
both ends of fibre exit tile.

PMT ≤32 fibres

phototubes outside
magnetic field

932mm

PMT ≤8 fibres

234mm
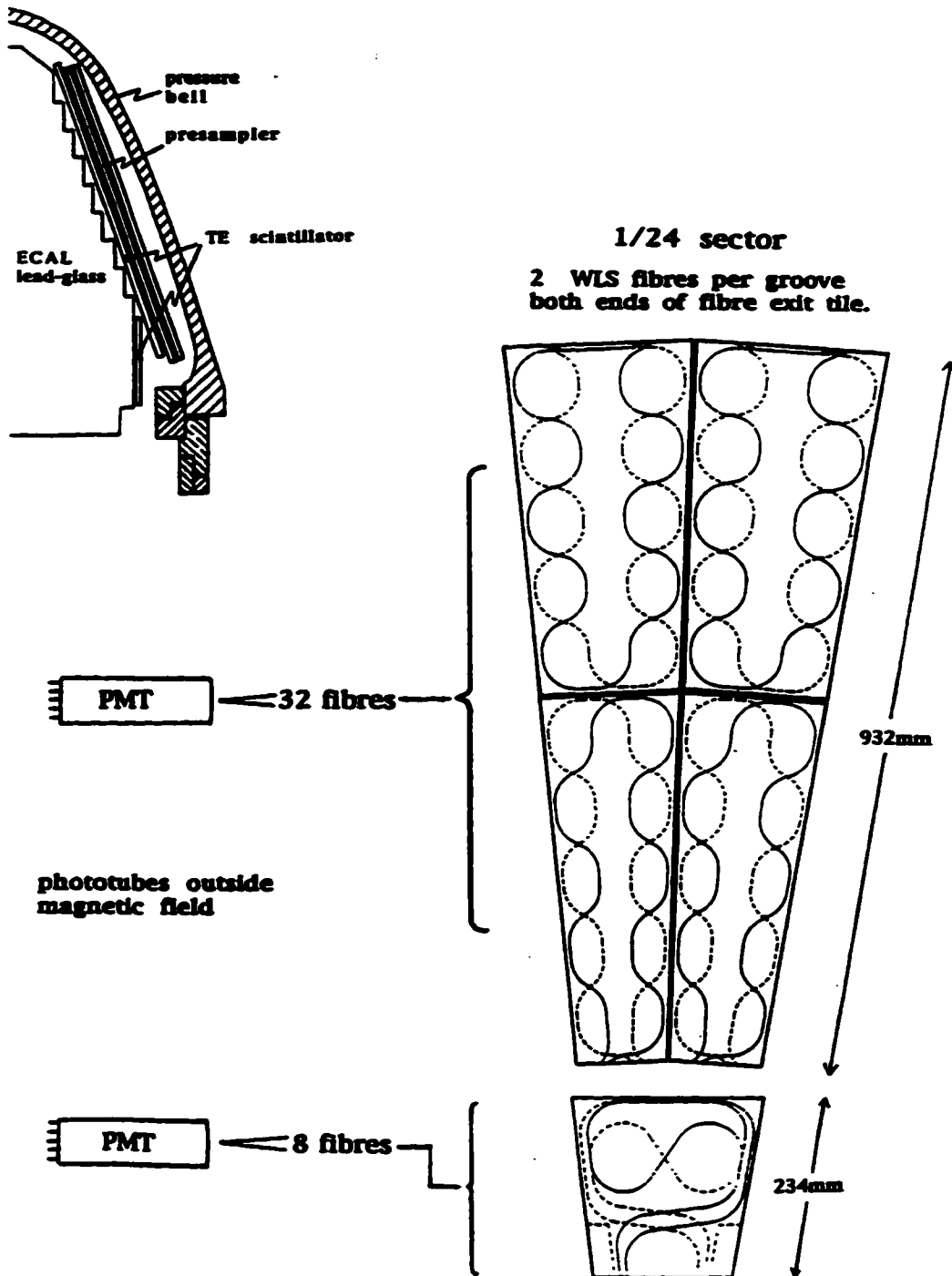
Figure 23: The Time of Flight Endcap (TE) detector. This figure shows both the location of the TE between the electromagnetic endcap calorimeter and the endcap presampler chambers as well as the detail of a 1/24 sector of the detector.

3.3.5 The Electromagnetic Calorimeter

The purpose of the electromagnetic calorimeter [8] is to measure the energy of electromagnetic showers originating from electrons and photons. This calorimeter is capable of measuring initial particle energies ranging from tens of MeV to over 100 GeV and also provides discrimination between hadrons and electrons when used with central tracking information. The electromagnetic calorimeter covers 98% of the solid angle and is divided into a barrel section with azimuthal coverage $|\cos\theta| < 0.82$ and two endcap sections with azimuthal coverage $0.81 < |\cos\theta| < 0.98$. Both sections measure the energy deposited in lead glass blocks by Cerenkov[†] photons which are generated by the particles entering the lead glass.

The barrel section of the electromagnetic calorimeter is placed just outside of the time of flight detector with an inner radius of 2.455 m from the beam pipe. It consists of 9440 lead glass blocks each of which having dimensions of 10 cm x 10 cm x 37 cm. These dimensions correspond to an amount of material equal to about 24.6 radiation lengths in each block. The blocks point towards the interaction region in an attempt to ensure that any entering particle does not share its energy between more than one calorimeter block.

The lead glass blocks are directly attached to a magnetic field tolerant phototube which is capable of operating in the magnetic fields present just outside of the solenoidal coil. In the absence of intervening material, the intrinsic energy resolution of the electromagnetic barrel calorimeter is given by the equation:

$$\frac{\sigma_E}{E} = 0.2\% + \frac{6.3\%}{\sqrt{E}}$$

[†] Cerenkov radiation of frequency ω is emitted when a particle traverses a medium with a velocity greater than the phase velocity of electromagnetic fields of frequency ω in that medium.

In practice, however, the amount of material in the detector preceding the electromagnetic calorimeter reduces this energy resolution by as much as a factor of two.

Each of the electromagnetic endcap calorimeters forms a dome shaped structure placed between the pressure bell and the pole tip hadronic calorimeters which are described in section 3.3.6. Each endcap consists of 1132 lead glass blocks which present a total depth of at least 20.5 radiation lengths to incoming particles. The blocks are placed coaxially with the beam axis and are read out through a single stage photomultiplier called *a vacuum photo-triode* which can operate in the full magnetic field of the solenoid. The energy resolution of the electromagnetic endcap calorimeter has been measured to be $5\% / \sqrt{E}$ for low energy particles.

The presence of material preceding the electromagnetic calorimeter causes many electromagnetic showers to be initiated prior to the particles entering the calorimeter. Thus, the installation of presamplers in front of the lead glass was necessary. Presamplers allow the position of the showers to be more accurately measured as well as the energy recorded in the calorimeter to be measured with a better resolution. This is possible due to the fact that the presampler signal is approximately proportional to the multiplicity of charged particles entering the device which is in turn proportional to the total energy deposited by the shower in the material preceding the calorimeter.

The barrel electromagnetic presampler [8] [17] is a 6.623 m long cylinder formed of tubular streamer chambers at a radius of 2.388 m from the beam pipe. There are two layers of tubes with each tube having a square cross section of area 9.6 mm^2. The ionization showers in the tubes are located by the induced charge on cathode strips placed along each side of each layer of tubes. The cathode strips of the inner layer are orthogonal to those of the outer layer. The position resolution of the barrel presampler is between 4 and 6 mm (depending on the particle energy) in a direction perpendicular to the shower direction. A position resolution of about 10 cm in z is also made possible through a charge division technique.

The endcap presampler is made up of thin multiwire proportional chambers placed between the pressure bell and the electromagnetic endcap calorimeter. The position of showers in these chambers is also determined from cathode strips and results in a resolution similar to that obtained for the barrel presampler.

3.3.6 The Hadron Calorimeter

The purpose of the hadronic calorimeter [8] is to determine the energy of showers of hadrons which exit the electromagnetic calorimeter. The hadronic calorimeter also acts as the return yoke for the magnet coil and consists of three distinct parts: the barrel, the endcaps, and the pole tips.

The barrel section of the calorimeter is 10 m long and surrounds the electromagnetic calorimeter. The inner radius of the barrel is 3.4 m, the outer radius is 4.4 m and it covers the polar angles $|\cos\theta| < 0.81$. The barrel calorimeter is composed of eight slabs of iron each of which is 10 cm thick. The central 4 m of the barrel is also covered by an additional two iron slabs in order to increase the number of interaction lengths[†] presented to particles passing through this region. The iron slabs are separated by 2.5 cm gaps which are filled with planes of streamer tubes. The anode wires in adjacent streamer tubes are 1 cm apart and run parallel to the beam axis. Large area cathode pads on one side of the streamer tubes read out the signal as do 4 mm wide aluminum strips running parallel to the anode wires on the opposite side from the cathode pads. Layers of cathode pads are grouped together to form hadronic calorimeter towers.

The endcap section of the calorimeter consists of seven 10 cm thick iron slabs and covers the angular region $0.81 < |\cos\theta| < 0.91$. The gaps between the iron slabs are 3.5 cm wide in the endcaps and are also filled with streamer tubes. The readout of the tubes in the

---

[†] An interaction length is defined similarly to the radiation length for electromagnetic showers and is the length scale which is appropriate to describe the evolution of hadronic showers.

endcap is performed similarly to the barrel readout scheme. When the cathode pad towers of the endcap are included with those of the barrel, the angular acceptance of the hadronic calorimeter is divided into 976 equal towers pointing towards the interaction region.

The pole tip hadronic calorimeters [18] cover the region in between the electromagnetic endcap calorimeter and the hadronic endcap calorimeter in the angular region $0.91 < |\cos\theta| < 0.99$. Each pole tip consists of nine 8 cm thick irons slabs with an inter-slab gap width of 1.0 cm. The gaps between the slabs are filled with thin multi-wire proportional chambers using a mixture of 55% carbon dioxide and 45% n-pentane gas. The chambers are read out on one side with 500 cm^2 cathode pads and by 2.5 cm wide strips on the other side. The pads from the ten layers of chambers are again grouped to form calorimeter towers pointing toward the interaction point.

The hadron calorimeter contains enough material to ensure that less than 0.1% of all pions will pass through the detector and hit the muon detectors. The calorimeter is also thick enough to reduce the amount of *punchthrough* (hadronic shower particles leaving the hadronic calorimeter) to less than 0.8% of all particles at 5 GeV/c. The energy resolution of the hadronic calorimeter varies from $100\% / \sqrt{E}$ for particles of energy less than 15 GeV to $140\% / \sqrt{E}$ for energies of 45 GeV.

3.3.7 The Muon Detector

The purpose of the muon detector [8] is to separate muons from the hadronic background in the OPAL detector. The detector is made up of a barrel section and two endcap sections and covers over 93% of the total solid angle.

The barrel section of the muon detector is made up of four layers of 10 m long by 1.2 m wide drift chambers which form a cylinder of radius 5 m. All four layers of the barrel muon detector cover the polar angle region $|\cos\theta| < 0.68$ while at least one layer covers the region $|\cos\theta| < 0.72$. Each chamber contains two drift cells with one anode wire each

running parallel to the beam axis. The r-ϕ coordinate of hits in the chambers is measured from the drift time while the z position is given by diamond shaped cathode pads [19] which run the length of each chamber along the anode wire.

The muon endcap detectors consist of 12 m × 12 m panels placed perpendicularly to the beam axis. Each covers the angular region $0.67 < |\cos\theta| < 0.98$ and is instrumented with four layers of streamer tubes operating with a gas mixture of 75% isobutane and 25% argon. Two layers of streamer tubes run vertically and two run horizontally in each endcap. The anode wires of the streamer tubes are read out using 8 mm wide aluminum cathode strips which run parallel to the anode wires on one side and perpendicular to the wires on the other side.

In order to identify muons, tracks seen in the central tracking detectors are extrapolated to past the hadronic calorimeter and matched with track segments in the muon chambers after correction for multiple scattering in the inner detector material. The positional accuracy required to match a muon chamber track segment with a track from the central detector is determined by the multiple scattering experienced by the highest energy muons of interest. This requirement gives a positional accuracy of about 2 mm which is made possible by the positional resolution of 1.5 mm in r-ϕ and 2 mm in z in the barrel muon chambers and 1 mm in x and y in the endcap muon chambers.

3.3.8 The Forward Detector and Luminometer

The measurement of the total integrated luminosity delivered to OPAL by LEP is made possible by the accurate measurement of the Bhabba scattering rate near the OPAL beam pipe. Two separate detector systems are used to measure this scattering rate: the forward detectors [8] [20] and the Silicon-Tungsten Calorimeter [21]. The forward detectors were part of the original OPAL design while the silicon-tungsten calorimeter was installed in 1994.

The two forward detectors surround the OPAL beam pipe just outside of the pressure bells and fill the holes at the center of the electromagnetic endcap calorimeters. Each forward detector is made of tube chambers and drift chambers in front of an independent electromagnetic calorimeter fitted with its own presampler. The tube and drift chambers measure the angles and positions of scattered electrons and positrons while the calorimeter measures their energies. The calorimeter consists of 35 layers of lead separated by layers of plastic scintillator. The forward detectors can accept incoming particles coming from the interaction region at angles ranging between 47 and 120 mrad from the beam axis.

The Silicon-Tungsten Calorimeter was designed to improve the luminosity measurement in OPAL. It consists of two detectors placed at ±238.94 cm from the interaction point in z. Each detector is made up of 19 layers of active silicon separated by 18 layers of tungsten. Each detector has a bare layer of silicon at the end facing the interaction point in order to detect preshowering.

Each active silicon layer consists of 16 wedge shaped silicon detectors each of which covers 22.5° in ϕ and radii between 6.2 cm and 14.2 cm. Each wedge is further subdivided into 64 pads which are read out individually. Adjacent wedges within a given layer are offset by 800 μm in z and consecutive layers are offset by half a wedge in ϕ in order to eliminate gaps in the detector.

3.3.9 Triggering and Online Data Acquisition System

In order to reduce the amount of data recorded by the detector, only physically interesting events as defined by trigger conditions are actually read out of OPAL. The OPAL trigger and the data flow after a trigger is registered are described in the following sections.

The Trigger

Trigger conditions [22] in OPAL may be divided into two general classes: stand-alone signals such as multiplicity counts and energy sums, and threshold signals from the trigger

matrix which is divided into 6 bins in θ and 24 bins in φ. Signals from the central tracks, the time of flight counters, the electromagnetic calorimeter, and the muon detector all serve as inputs to the trigger matrix. Certain geometrical correlations between detector inputs in the matrix are sufficient to trigger the detector. The list of correlations and stand-alone signals which are sufficient to trigger the detector are stored in the *trigger definition file* which may be updated for any given OPAL run.

OPAL also makes use of a pretrigger which takes into account the bunch crossing signal from LEP as well as many of the same inputs as the main trigger. The pretrigger matrix, however, has only 12 bins in φ and no segmentation in θ and thus the pretrigger serves only as a fast indication of the possibility of a good e^+e^- collision in OPAL. Only 1-2% of all bunch crossings generate a pretrigger signal thus reducing the dead time in the detector that would be induced by a full readout sequence.

When a trigger is generated, the General Trigger Unit (GTU) sends a trigger interrupt request to a Local Trigger Unit (LTU) located on each subdetector. The subdetector then reads the data out while the LTU holds a busy line. Only when the GTU detects that all subdetectors have read out their data does it resume triggering. The trigger reduces the data flow to the 10 Hz level necessary for the successful operation of the rest of the online readout chain.

Online Data Flow

After a trigger, each of the sixteen OPAL subdetectors is read out into its Local System Crate (LSC) and then assembled in the Event Builder (EVB). The EVB is connected to each of the eighteen LSC's (one for each subdetector and one each for the trigger and track trigger) by high speed memory map links.

After assembly, the event is passed to the filter which compresses the data and performs some data monitoring. Obviously badly measured events, some 15-20% of all triggers, are

rejected at this stage. All header information for events reaching the filter, whether they are rejected or not, are recorded on disk to help in later book-keeping.

Events which pass the filter are written to disk in 20 MB long partitions and then are copied to the reconstruction farm of Unix workstations where the event parameters (described in section 3.5.1) are determined. After reconstruction, the completed events are recorded on data cartridges as a permanent backup as well as being transferred to the OPAL offline analysis computers on the main CERN site. There, the events are stored on disk and again backed up to data cartridges. Data taken by OPAL is commonly available for offline analysis within several hours of its being recorded by the detector.

3.4 The OPAL Simulation Software

In order to examine the response of the OPAL detector to various interesting physics processes, a full computer simulation of the detector has been developed. This simulation package, known as GOPAL [23] is based on the CERN detector simulation package GEANT [24]. The simulation operates by defining the geometry of the OPAL detector and controlling the tracking of particles through this geometry. The energy loss of the particles in the various subdetectors is calculated and the output given in the same format as real OPAL data in order to facilitate analysis by the same offline analysis tools as used for real data.

Event Generators

Prior to using the OPAL simulation program, physics events must be simulated using an event generator program. Event generators calculate the cross sections for the requested physical process and generate the initial particle configuration and kinematics at the interaction point. The output of these generators can then be used as input to the OPAL simulation program.

Several different event generator programs exist for different physics interests. The generators used in this thesis are described in section 5.1. Much work has been invested by the OPAL collaboration in tuning the output of these event generators to match real OPAL data [25].

Detector Simulation

GOPAL uses GEANT defined geometrical shapes to specify the location and extent of each of the OPAL subdetectors. GEANT also allows the properties of the detector materials to be simulated. Particles arising from the event generators are then traced through the detector geometry in steps and the probability of an interaction during any step is calculated. If an interaction occurs, the energy loss of the particle in the subdetector is recorded as a "hit" and stored. Secondary physics effects such as photon pair production and particle decays are also accounted for in the GEANT framework.

After all particles in the event have been tracked through the detector, the hit information is collected for each subdetector and the detector response is determined. The resolution and inefficiencies of the OPAL subdetectors are simulated at this point. At the end of this stage, the simulated data is written out in the standard data summary tape (DST) format for later analysis by the OPAL offline code.

3.5 The OPAL Offline Analysis Software

The OPAL offline analysis software, known as ROPE, processes either raw OPAL data or simulated data summary tapes and calculates the physical variables which are later used in physics analysis. Information about the track parameters, the primary and secondary event vertices, the dE/dx of tracks in the jet chamber, calorimeter cluster energies, and muon candidates are all generated by ROPE. The ROPE process can be used to create DST from raw OPAL data or to analyze DST which has been produced previously.

3.5.1 Track Reconstruction

ROPE makes use of a central tracking (CT) subprocessor to match track segments in each of the central detectors in order to calculate parameters for each track. A track fit which includes the effects of multiple scattering in the detector is used to merge segments from each of the central detectors into a single track. The resulting parameters which are calculated for each track are:

- The curvature (κ) of the track as defined by $\kappa=1/2\rho$ where ρ is the radius of curvature of the track. Since the OPAL magnetic field is directed along the positive z axis, a positive value of κ corresponds to a particle with negative charge.

- ϕ_0, the azimuthal angle of the tangent to the track direction at the point of closest approach to the OPAL coordinate origin.

- d_0 (or b), the track impact parameter where $|d_0|$ is the distance from the origin to the point of closest approach of the track in the x-y plane. The impact parameter is given a sign as described in section 5.2.5.

- The dip angle of the track λ where $\tan(\lambda)=\cot(\theta)$.

- The value of the z coordinate at the point of closest approach to the origin, z_0.

With the above track parameters, and a knowledge of the OPAL magnetic field B in kilogauss, the total momentum of a track in the r-ϕ plane is given by $p_t = 1.5 \times 10^{-4} B / \kappa$ GeV/c when κ is measured in units of cm^{-1}. The three components of the momentum are then given by $p_x = p_t\cos\phi_0$, $p_y = p_t\sin\phi_0$, and $p_z = p_t\tan\lambda$.

3.5.2 Track Quality Selection

OPAL makes use of several quality selections in order to reduce the amount of data which need be considered for physics analysis. These selections aim to reject events with badly

reconstructed tracks as well as events which are not multihadronic in nature. Without such quality selections, the amount of computer time necessary to analyze the entire data set would be prohibitive. The following two sets of quality selection criteria are the most often used within OPAL as a starting point for physics analysis with further refinement in the selection being made by each researcher later in his or her analysis.

The Gold Plated Multihadron Selection

The Gold Plated Multihadron (GPMH) selection is performed online and hence is one of the first indicators of the quality of an event. This selection requires that:

- Electromagnetic calorimeter clusters must have at least a 100 MeV initiating block and at least one 50 MeV neighbor block in the barrel and at least a 200 MeV initiating block and at least one 100 MeV neighbor block in the endcap.

- The event must have an energy sum over all good electromagnetic clusters of at least 8 GeV.

- The event must have at least 6 good electromagnetic clusters.

- The sum of the good electromagnetic cluster energies in the hemisphere opposite the highest energy cluster must be at least 2 GeV.

- The event must pass the halo muon rejection cut.

- If the sum of the electromagnetic barrel cluster energies is greater than 2 GeV and the number of clusters if greater than or equal to 2, the number of time of flight counter hits is required to be at least 3.

The Tokyo Multihadron Selection

The Tokyo Multihadron Flag (TKMH) is set in the ROPE offline processor and is a much better event quality indicator than the GPMH criterion presented above. In the TKMH processor, good electromagnetic clusters are defined as having:

- a raw energy of at least 100 MeV in the barrel or at least 200 MeV in the endcap.

- at least one constituent block in the barrel and at least two in the endcap.

Good tracks are defined as having:

- at least 20 hits in the central detector.

- an impact parameter ($|d_0|$) no greater than 2.0 cm.

- a value of $|z_0|$ no greater than 40.0 cm.

- the radius of the first hit in the central detector no greater than 60.0 cm.

- a momentum in the x-y plane greater than or equal to 0.050 GeV/c.

- a value of $|\cos(\theta)|$ no greater than 0.995.

- a χ^2 on the r-ϕ track fit no greater than 999.

- a χ^2 on the z track fit no greater than 999.

Using good tracks and clusters selected by these cuts, two ratios are defined:

- R_{vis} which is equal to the sum of the raw electromagnetic cluster energies divided by two times the beam energy, and

- E_{bal} which is equal to the sum of the raw electromagnetic cluster energies times the cosine of the azimuthal angle of the cluster.

With these definitions, an event is selected by the TKMH processor if:

- R_{vis} is greater than or equal to 0.10.

- $|R_{bal}|$ is less than or equal to 0.65.

- There are at least 7 good electromagnetic calorimeter clusters.

- There are at least 5 good tracks.

The High Multiplicity Selection

The High Multiplicity Selection, HIMS, is the data selection used to search for the Standard Model Higgs Boson at OPAL. In order to be flagged by the HIMS selection, the event must have passed the GPMH selection described above or it must contain at least five tracks with the following properties:

- The track momentum in the x-y plane must be at least 100 MeV/c.

- The absolute value of the cosine of the azimuthal angle of the track must be no greater than 0.996.

- The magnitude of the impact parameter d_0 must be no greater than 2.5 cm.

- The magnitude of z_0 for the track must be no greater than 50.0 cm.

- The number of hits in the jet chamber must be at least 20.

- The number of hits in the jet chamber must comprise at least 50% of the total possible central detector hits for the track.

3.5.3 OPAL Offline Analysis Packages

In order to allow efficient physics analysis, several offline analysis packages have been developed to provide specialized tools for the various OPAL physics working groups. These offline analysis packages extract useful information from the DST for a particular purpose and standardize the definitions of various analysis methods within the collaboration.

Offline anlaysis packages exist for the identification of leptons in an event, for determining the primary vertex of an event, for tagging heavy quarks in jets, and for extracting variables relevant to searches for new particles to name just a few. The offline packages which have been used in the analysis presented in this thesis are described in section 5.2.

3.6 Personal Contributions to OPAL

A total of almost one year of my tenure as a MSc. student was spent at CERN where I was involved with several tasks related to the maintenance and operation of the OPAL detector. These tasks included shifts watching over the detector's safety equipment as well as helping to operate the online data analysis system. I was also involved with the early testing of scintillating tiles and the fabrication of the light guide fibres for the Scintillating Tile Endcap detector which was installed in OPAL in May 1996. Finally, I contributed to the OPAL Higgs search effort as a whole by helping to develop the offline analysis software (the DH [33] package) for the four jet decay channel.

4. Classification Using Artificial Neural Networks

An Artificial Neural Network (ANN) consists of a web of interconnected processing units known as *neurons*. Each neuron is capable of combining inputs from several other neurons in order to determine its output state. Data is presented to a set of input neurons which process and pass on various features of the data to other layers of neurons. Several such intermediate layers can process the data before a final output is presented back to the operator. This output is used to make a classification decision on a particular piece of data.

The ability of ANN's to recognize unique features in different classes of data makes them very well suited to high energy physics analysis which often strives to separate a signal from a background based on small differences in several variables. Through *associative recognition* of complex structures, multi-dimensional features in the signal and background signatures may be appreciated by the network that would be missed in a conventional multivariate analysis. ANN's are also *trainable,* that is they are designed to self-organize and to optimize their classification ability. This trait makes ANN's very efficient in terms of development time for high energy physics analysis as optimizing a cutting approach for separating signal from background in several variables can become a very lengthy procedure. The *fault tolerance* of ANN's is also beneficial as a well trained ANN can see past noisy or incomplete data to the underlying structure and thus complete its classification task unimpaired. Furthermore, the algorithms which govern the training and use of ANN's are inherently parallel and recursive which allows them to be implemented on widely available digital computing systems.

While Artificial Neural Networks have many beneficial properties when applied to particle physics analysis, there are some drawbacks to their use. The most important of these is the

closed nature of the ANN process where a classification decision is made directly from the input variables without any visible intermediate steps. The actual relationships between the input variables which are being exploited by the ANN are obscured in the inner workings of the network and thus the evaluation of systematic errors in the analysis is made difficult. In addition, ANN's can be *overtrained* to recognize features which are not indicative of the general class of data one wishes to isolate, but which belong to a particular subset of that data. Such a network will therefore be unable to recognize other sets of data from the same general class and will be of no use for data classification.

These problems have been well documented in the literature and the advent of customizable ANN software packages, such as JETNET [26], has made the diagnosis and cure of ANN development problems easier and faster. With careful design, implementation, and testing, problems can be avoided and the ANN can become a useful weapon in the analysis arsenal of the high energy physicist.

4.1 A Brief History of Artificial Neural Networks

This section is intended to give a flavour of the history of the development of Artificial Neural Networks. A comprehensive discussion of this history may be found in reference [27]. The emergence of the study of Artificial Neural Networks can be traced to the introduction of the dot product neuron (described in section 4.2 below) by McCulloch and Pitts in 1943. At the same time, Hebb postulated a learning mechanism for networks of such neurons which stated that the connection strength between any two neurons is reinforced when a high activity in one neuron is correlated to a large reaction in the other. This is the basis of the delta rule of network learning which will be discussed in more detail in section 4.5.2 below.

In 1957, an attempt to build the first artificial neuron was undertaken by Resenblatt at Cornell. Named the *Perceptron*, this neural network consisted of only two layers of neurons: a single input layer and a single output layer. As will be made clear below, this

limited the Perceptron to linear separation of data during classification. By 1959 this major limitation of the Perceptron was being widely criticized in the literature, and resulted in a loss of funding for many neural network research projects. As a result, work in the field slowed to a standstill for many years.

In 1982, however, Hopfield of the California Institute of Technology breathed new life into neural network research by proposing an associative memory model based on a network of fully connected neurons. This type of network would become the basis of the pattern recognition networks in use today. The upsurge of interest in neural networks caused by Hopfield's developments was strengthened by the development of a Multi-Layer Perceptron (MLP) by Rumelhart at the Massachusetts Institute of Technology. The MLP was capable of non-linear separation during classification and hence set the stage for the wide variety of classification problems accessible to neural network analysis today.

4.2 The Neuron

The basic building block of the Artificial Neural Network is the *neuron*. This unit emulates the functional description of the biological neuron found in the brain by replacing electro-chemical communication processes with the electrical impulses of a computer software or hardware implementation.

An Artificial Neural Network consists of a set of these single neurons connected together in such a way that they may share information. The connection between any two neurons i and j has an intrinsic strength characterized by a coupling constant w_{ij}. A neuron i is itself completely described by its activation function f and its current state s_i. There are two basic models of the neuron, both of which share the common features mentioned above. These two models, the *dot product* neuron and the *distance* neuron differ in how the weights of the connected neurons are combined in order to determine the state of the neuron.

4.2.1 The Dot Product Neuron

The most common neuron model, the dot product neuron, is drawn schematically in Figure 24. In this model, a given neuron i performs a dot product of the output states s_j of its input neurons with their associated weights w_{ij}:

$$P_i = \sum_j w_{ij} \cdot s_j$$

The resulting weighted sum, P_i, is then used as input to the activation function of neuron i with the output specifying the state of neuron i through the relation:

$$s_i = f(P_i)$$

The activation function f is chosen from several common functions and usually consists of applying a threshold and some amplification to P_i. The simplest example of an activation
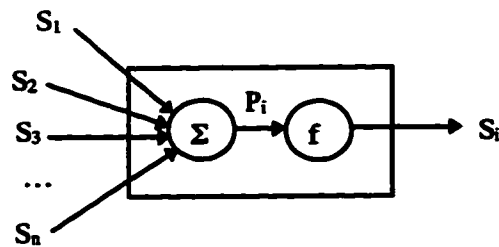
Figure 24: Schematic diagram of a dot product neuron. There are n input neurons each with state s_j and connection weight w_{ij}. The single output state of the neuron, s_i, is governed by the activation function f.

function for a dot product neuron is a step function. The most commonly used function, however, is a non-linear sigmoid characterized by the logistic function:

$$\sigma(z_i) = \frac{1}{1+e^{-z_i}}$$

Both types of activation function are shown in Figure 25.

threshold (step) function   **sigmoid function**
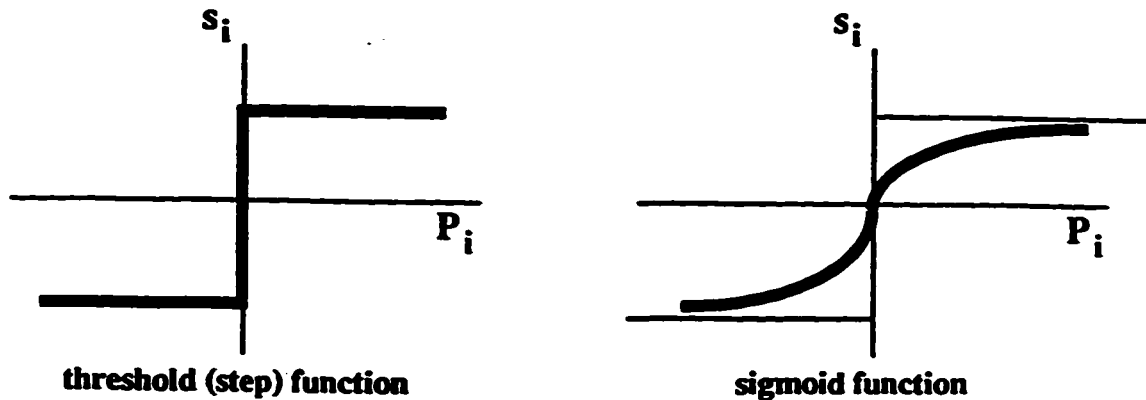
Figure 25: Two common activation functions used in the dot product neuron. The step function is shown on the left and the logistic function on the right.

The analogous nature of the dot product neuron to its biological cousin is quite clear. The weighting function between interconnected neurons in the ANN corresponds to the chemical strength in the synapses between the dendrites and neuron body in the biological model. Similarly, the output state of the ANN neuron corresponds to the response sent down the axon of a biological neuron. The magnitude of this output state is related to the frequency of firing of the biological neuron.

4.2.2 The Distance Neuron

A second type of neuron, known as the *distance* neuron, differs from the dot product neuron in the manner in which the incoming signals from connected neurons are processed and in the choice of activation function used in each neuron. A schematic drawing of the distance neuron is shown in Figure 26.

$$d_i = \| w_i \cdot s \|$$

$$s_i = f(d_i)$$

bell-shaped function
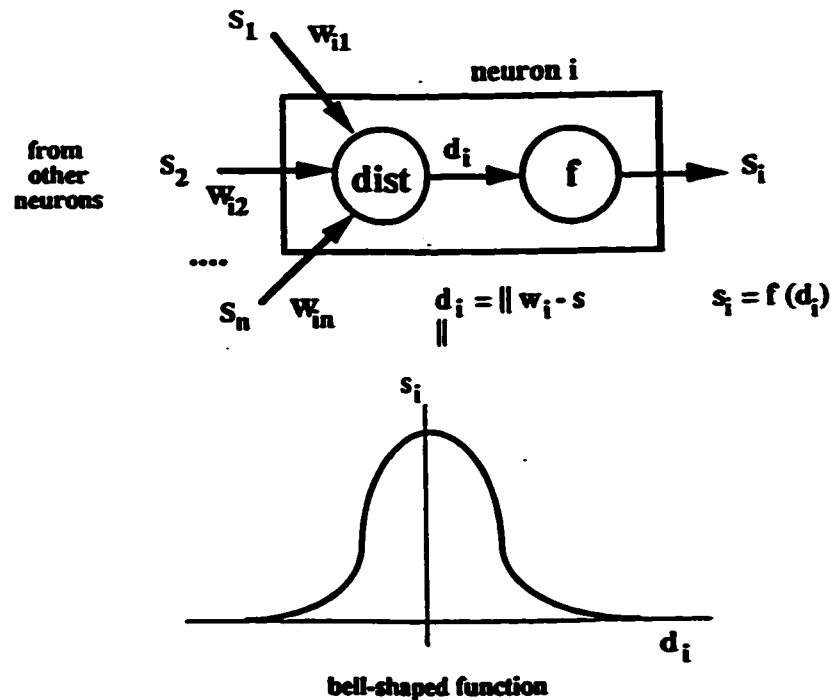
Figure 26: A schematic diagram of the distance neuron. The n input neurons each with state s_j are compared to their respective weights through a euclidean distance norm and this output is used in a bell shaped activation function.

Instead of the weighted sum of the input neuron states, the distance neuron computes the "distance" between the vector of input weights $\vec{W}$ and the actual input state vector $\vec{S}$. This distance can be specified in any metric, but is commonly given by the euclidean norm:

$$d_i = \left\| \vec{W} - \vec{S} \right\|$$

The activation function for such a neuron then takes on the form of a bell curve. This implies the notion that the output of a given neuron i is higher if its weight vector and input vectors are similar thus giving an activation function that peaks at an input of zero and decreases as $|d_i|$ increases.

The distance neuron is commonly used in competitive learning techniques in unsupervised self-organizing networks. As this type of network is not commonly used for classification

purposes in particle physics, it is not discussed further in this thesis. In depth discussions of self-organizing networks may be found elsewhere [27].

4.2.3 Comparison to the Biological Model

While the basic architecture of an Artificial Neural Network may closely emulate its biological counterpart, the vast difference between the capability of the human brain and even the best ANN may be explained as a matter of scale. While the speed of a "neural operation" (one neuron firing) in the biological neuron is only ~10 ms as compared the ~1 ns speed of a single hardware VLSI† chip network operation, the human brain has on average 10^{11} neurons each of which is connected to between 10^2 and 10^4 other neurons. In comparison, an average ANN contains only 10^2 to 10^3 neurons each of which is connected to less than 10 other neurons. This vast difference in the number of neurons and connectivity between the human brain and an ANN accounts for the superlative classification and pattern recognition capabilities of the brain. It is estimated that the immense number of neurons in the human brain [27] corresponds to a speed of 10^{11} operations per second and a total memory capacity of 10^8 MB!

With these descriptions of the basic function unit of Artificial Neural Networks, we are now ready to consider the history of the development of ANN's as well as their architecture and operation.

4.3 The Architecture of Neural Networks

A particular neural network is completely specified by the number of neurons it contains and the way in which they are interconnected. Two main architectures of ANN's may be

---

† Very Large Scale Integration. These types of microchips can integrate millions of individual transistors on a single small piece of silicon.

identified. These are the *feed forward* network and the *recurrent* network. These two different architectures are shown schematically in Figure 27.
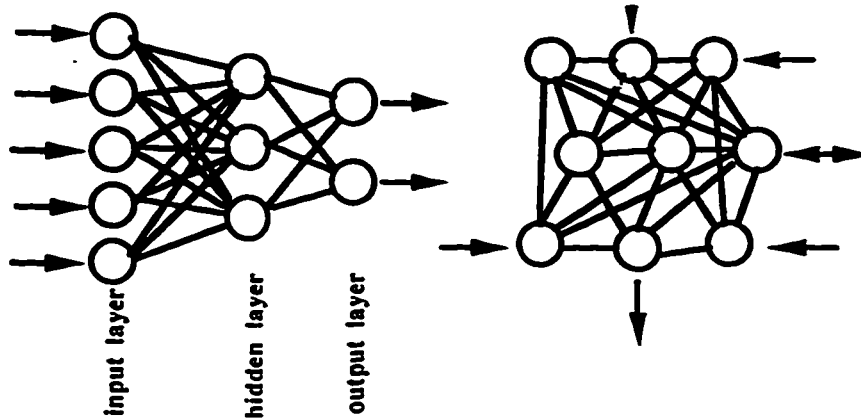
Figure 27: A feed forward network is shown on the left where information propagation occurs only in a specified direction and there are well defined input and output layers. On the right is a recurrent network where all neurons are bi-directionally connected and a neuron can simultaneously be an input and an output neuron.

In a feed forward network, each neuron belongs to a well defined layer within the network. Each layer can be classified as an *input layer* where raw data is accepted into the network, a *hidden layer* where non-linear separation of data classes takes place, or an *output layer* where the resulting output of the net is reported.

Neurons are almost always of the dot product type in a feed forward network. In such a network, no connection between neurons in the same layer is possible as the output of a particular neuron must feed uni-directionally to a neuron in the next layer. This type of network is representative of the general class known as Multi-Layer Perceptrons and is the one most commonly used networks for classification problems in high energy physics. The remainder of this chapter will concentrate mainly on this type of network.

In contrast, the neurons of a recurrent network may connect to any other neuron in the net and are capable of both receiving input from and sending output to the same connected neuron. Thus, any neuron may be both an input and output neuron. In principle, each neuron is connected to all of the others with the weight between any two neurons

becoming zero during the learning process if the two neurons in question do not interact with each other. Neural networks of this type find application in *associative memory* models. In such a configuration, the ANN produces an output which is related to the input configuration but which contains information that may have been lost or obscured by noise in the original input data. The obvious use of such networks in particle physics is pattern recognition in tracking detectors with the associative memory filling in the details of tracks from only a few hits. The possibilities of this type of network for pattern recognition will not be discussed in the present work. Excellent introductions to recurrent ANN's may be found in [27] and [28].

4.4 Significance of Hidden Neurons in Classification Tasks

As was mentioned in Section 4.1, a major stumbling block of the first Perceptron, which consisted of only an input and an output layer, was its inability to isolate sets of data which were not linearly separable in n-space. As can be seen in Figure 28, the separation of two classes of data, even in n dimensions, can take place even with the simple Perceptron as long as the two classes of data are linearly separable in n-space. In the example shown in Figure 28, the two classes of data, $\mathcal{A}$ and $\mathcal{B}$ are separable by the line $aX + bY + c = 0$. Thus, by using the X and Y values as inputs with weights of a and b respectively, a step activation function $\Phi(aX + bY + c)$ in the output neuron will successfully isolate the two data classes.
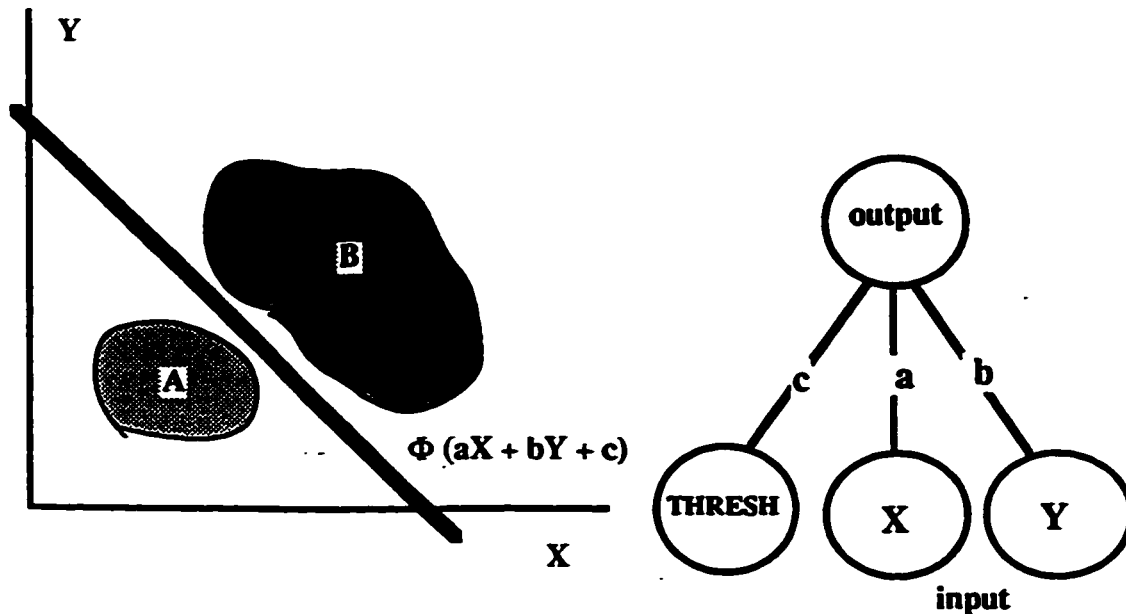
Figure 28: An example of linear separation of two classes of data by a two layer neural network. This is possible with a two layer network only because the two classes of data, *A* and *B*, are linearly separable by the line $aX + bY + c = 0$ in 2-space. The right hand section of the figure shows the resulting logic diagram for the network with the weights clearly marked on the connecting lines. The threshold input always reports a value of c to the output neuron.

For cases in which linear separation of the two classes is impossible, hidden layers become necessary in the network. The classic simple example of linearly non-separable classes is the logical *exclusive or* (XOR) operation. Figure 29 shows the network diagrams for the case of A.XOR.B. As can be seen in the diagram, the class of data for which A.XOR.B is true may not be separated from the class for which A.XOR.B is false by any single line in the two dimensional plane. This problem is overcome by the introduction of a hidden layer. Closer examination reveals that the hidden neuron detects whether or not A and B are both on, thus finding a more complex feature of the data. In more complex networks, the identification of these features is what gives neural networks their powerful classification abilities.
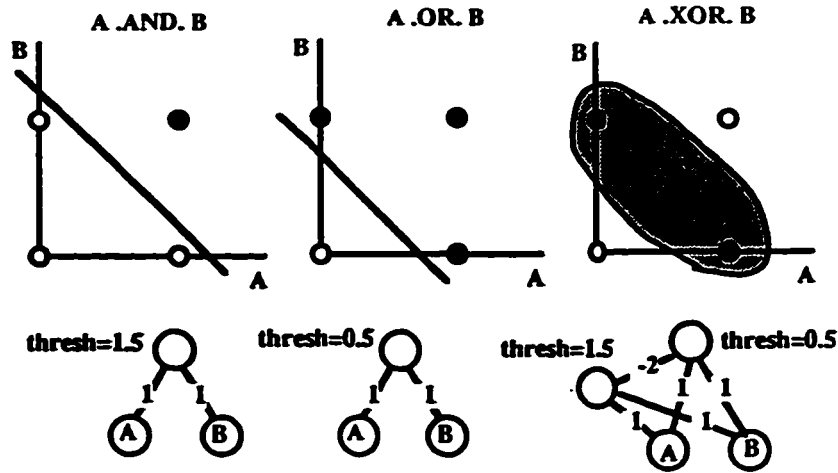
A	B	Hidden Neuron	Output
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

Figure 29: The top portion of the diagram shows the two dimensional data space for the XOR function. Clearly no single linear cut can separate the points for which A.XOR.B is true from the points for which A.XOR.B is false. With the addition of a hidden neuron, however, a non-linear separation can be achieved with the results shown in the table. In the schematic neural net diagram, the thresholds for the various neurons are marked as are the weights by which their outputs are scaled.

4.5 The Learning Process

In this thesis, we only deal with *supervised* learning by neural networks. In this type of learning, an external supervisor trains the network to recognize a certain class of data or pattern. The network is given a target output vector $\vec{T}$ and then is fed a set of training data as input. The output of the network from this training data is compared to $\vec{T}$ on an event by event basis and the weights between neurons updated in order to make the output more like the target vector. The weights in the network therefore change with time according to the following equation.

$$w_{ij}(t') = w_{ij}(t) + \Delta w_{ij}.$$

There are several rules which govern this learning process. These are detailed in the sections below.

4.5.1 Hebbian Learning

This postulate states that the weight between two neurons strengthens in proportion to the product of the activation status of the two neurons in question. This strengthening of weights is described by the equation:

$$\Delta w_{ij} = \eta s_{it} s_{jt}$$

When compared to the biological neural network model, Hebbian learning embodies the notion that a neural pathway strengthens with use. In practical terms within an ANN, Hebbian learning has the result that only neural connections which are successfully identifying features in the data will develop appreciable weights.

4.5.2 Delta Rule Learning

The *Delta Rule* is the mechanism by which the neural net reduces the differences between its output vector $\bar{O}$ and the target vector $\bar{T}$. At the outset of training, a network begins with randomly arranged weights between its constituent neurons. A given input event produces a certain output vector $\bar{O}$ when passed through the network. A quadratic form can be assigned to measure the degree of difference between the output and target vectors for a particular event p.

$$E_p = \frac{1}{2}\left|\bar{O}_p - \bar{T}_p\right|^2 = \frac{1}{2}\sum\left[O_p(i) - T_p(i)\right]^2$$

When summed over a set of m patterns, this function becomes a *global error function*

$E = \sum_{p=1}^{m} E_p$ and may be minimized to yield the optimal configuration of the weights for the

network.

The delta rule itself describes the manner in which the weights are varied in order to minimize E. Weights are changed in proportion to the effect they have on the minimization of the global error function. This weight change is called *gradient descent* and may be written:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \qquad (gradient\ descent)$$

The parameter η in the above equation is known as the *learning parameter* and is discussed in more detail below.

We now clarify the learning rules presented above by applying the delta rule to the simple two layer neural network shown in Figure 30. In this network, a set of p input vectors $\bar{A}_p$ is presented to the input layer of the ANN and associated output patterns $\bar{O}_p$ are generated at the output nodes. Note that the dimension of the output vector need not be the same as the dimension of the input vector; the input and output spaces can be of different dimensions. It is assumed that a desired target vector $\bar{T}_p$ is known for each input pattern. The weight between the input node j and each output node is given by w_{ij}.
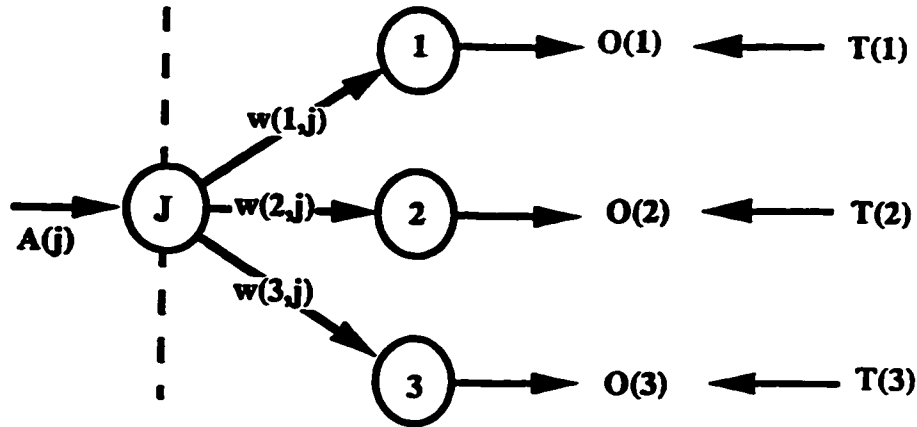
Figure 30: A simple two layer ANN. Inputs A(j) are presented at the input node and transferred through weights $w(i,j)=w_{ij}$ to the output nodes.

Recalling the form of the output of a dot-product neuron from section 4.2.1, we see that we can write the output vector of our sample ANN for a given input pattern as:

$$O_p(i) = f\left[\sum w_{ij}A_p(j)\right] = f\left[s_p(i)\right]$$

with the definition $s_p(i) = \sum w_{ij}A_p(j)$ and where f is the activation function. We can now calculate the gradient of the error function:

$$\frac{\partial E}{\partial w_{ij}} = \sum_p \frac{\partial E_p}{\partial w_{ij}}$$

$$= \sum_p \frac{\partial E}{\partial \bar{O}_p} \cdot \frac{\partial \bar{O}}{\partial s_p(i)} \cdot \frac{\partial s_p(i)}{\partial w_{ij}}$$

$$= \sum_p \left[\bar{O}_p(i) - \bar{T}_p(i)\right] \cdot f'\left[s_p(i)\right] \cdot A_p(j)$$

Thus, we may write the change in the weighting constants as:

$$\Delta w_{ij} = \sum_p \Delta_p w_{ij}$$

$$= -\eta \sum_p \left[\vec{O}_p(i) - \vec{T}_p(i) \right] f'\left[s_p(i) \right] A_p(j)$$

$$= -\eta \sum_p \delta_p(i) A_p(j) \quad (the\ Widrow\text{-}Hoff\ Formula)$$

where we have defined:

$$\delta_p(i) = \left[\vec{O}_p(i) - \vec{T}_p(i) \right] f'\left[s_p(i) \right]$$

We note that in the Widrow-Hoff formula, the change in the strength of the weights depends on both the strength of the input contained in $A_p(j)$ as well as the strength of the effect of the input on a given neuron as given by $\delta_p(i)$. This is the effect of Hebbian learning showing itself within the delta rule.

The fact that the first derivative of the activation function f appears in the expression for $\delta_p(i)$ and therefore in the expression for Δw_{ij} indicates that only continuous activation functions may be used in the network. This means that the threshold step function neurons may not be used with the delta learning rule. More commonly, a continuous sigmoid function such as $f(s) = \dfrac{1}{1 + e^{-s}}$ is used as the activation function.
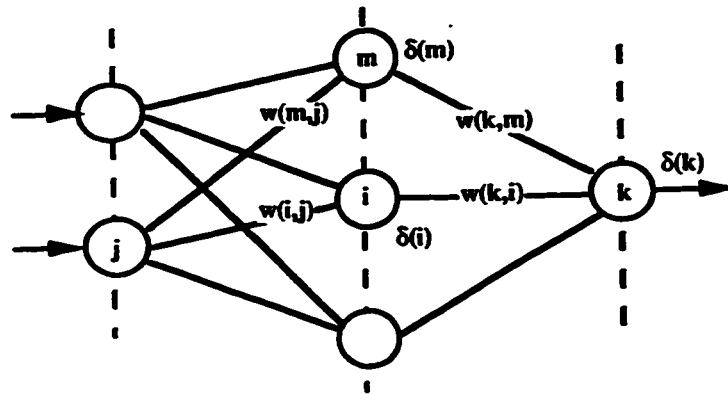
Figure 31: A simple three layer ANN with a hidden layer. Weights in the hidden layer are updated by backpropagation.

4.5.3 Backpropagation

The simple example of the two layer network in the section above may be expanded to the case of a multi-layer network with hidden layers through the introduction of *backpropagation*. This concept defines the manner in which the weights of hidden neurons are corrected in order to reduce the global error function. Applying the gradient descent equation of section 4.5.2 to the three layer network shown in Figure 31, we see that:

$$\Delta w_{ij} = -\eta \sum_{p} \frac{\partial E_{p}}{\partial w_{ij}}$$

with the gradient calculated as:

$$\frac{\partial E_{p}}{\partial w_{ij}} = \frac{\partial E}{\partial \vec{O}_{p}(i)} \cdot f'\left[s_{p}(i)\right] \cdot A_{p}(j)$$

This is the same equation that was obtained for the two layer ANN. However, in the three layer case, we cannot relate the $\dfrac{\partial E_{p}}{\partial \vec{O}(i)}$ term to the difference between the output and target vectors as the hidden neuron i cannot directly see the output layer and target. This term is instead calculated using an expansion of the errors computed for the output layer.

$$\frac{\partial E_{p}}{\partial \vec{O}_{p}(i)} = \sum_{k} \frac{\partial E_{p}}{\partial \vec{O}_{p}(k)} \cdot \frac{\partial \vec{O}_{p}(k)}{\partial s_{p}(k)} \cdot \frac{\partial s_{p}(k)}{\partial \vec{O}_{p}(i)}$$

$$= \sum_{k} \delta_{p}(k) w(k,i)$$

Recalling the definition of δ_{p} from section 4.5.2, we now have an expression for the evolution of the weights for the hidden neuron i.

$$\Delta w_{ij} = -\eta \sum_p \left(f'\left[s_p(i)\right] \cdot A_p(j) \cdot \left[\sum_k \delta_p(k)w(k,i)\right]\right)$$

Thus, the generalized error for a hidden neuron in a multi-layer network is calculated from a weighted sum of the errors in each subsequent layer times the derivative of the activation function. The fact that the error in downstream layers is used in the calculation of the error in an upstream layer gives rise to the term backpropogation.

4.5.4 Testing the Success of the Learning Process

The process of learning, or minimization of the global error function, can prove to be difficult in that many minima may exist for the error function. In order for the neural net to avoid local minima and settle upon the correct global minimum, the learning parameter η encountered earlier may be manipulated to control the "speed" of the gradient descent. The learning parameter is large at the outset of learning, when the steps between successive sets of neuron weights are large. When a minimum is neared, the value of η is decreased in order to ensure that the correct minimum is identified within appropriate accuracy. If the minimum reached is only a local minimum, there should still be a significant deviation in the error function which will cause the next update of the weights to jump out of the local minimum.

In addition to the learning parameter, the speed of the gradient descent can be modulated by a *momentum* term α. This term provides a "memory" of how quickly the descent was proceeding at the last iteration and conveys the information that if the last step was large, the current step cannot slow the descent too much. This allows the minimization to proceed at a faster rate without slowing for insignificant local effects. The cumulative effect of these two terms on the weight updating can be written as:

$$\Delta_n w(i, j) = -\eta \frac{\partial E_n}{\partial w(i, j)} + \alpha \Delta_{n-1} w(i, j)$$

In practice, when training an ANN with backpropogation, one divides the available set of input patterns into three sets, the *training sample*, the *testing sample*, and the *validation sample*. The training begins by presenting a number of the training samples to the network and then performing the gradient descent minimization to optimize the network weights. At the end of a certain number of epochs of training, the performance of the network is measured on the events of the testing sample which have not been used in the training process. This procedure is continued until the deviation of the network output from the target vector is smaller than some predefined value over the testing sample. The optimization of the weights is a very time consuming process and can involve iterating many thousands of times over the input patterns. Once the minimization is complete, the network is ready for validation.

Validation of the network performance takes place on a separate set of input data (the validation set) which does not contain any of the same input patterns used in the training or testing data sets. The success of training can be quantified by measuring the performance (ie the sum of the difference between output and target vectors) of the neural network on this validation data set. When the network consistently achieves the desired level of success on a series of validation data sets, it is ready for general use.

When training a network, it must be noted that the training data set should contain at least several times as many patterns as there are weights to be optimized in the network. This constraint has the obvious effect of making the smallest neural network which gives acceptable results the most attractive for a given application. There is no set rule which gives the minimum number of nodes (and therefore weights) which a neural network requires in order to give good performance. In fact, the number of hidden units in an ANN is usually found as the minimum number which gives a reasonable performance for a given problem on a trial and error basis.

The network must be guarded against overtraining as well. In this case, the number of hidden units in the ANN is too large for the problem at hand and the network actually

"memorizes" the input patterns of the training set. This, of course, results in excellent performance on the training data, but when tested on a separate set of input patterns, the network performance degrades drastically. When an ANN is overtrained, it is in essence looking at the input data so closely that it is no longer able to generalize and pick out identifying features for a given class of data.

The theoretical limit of the quality of classification by a neural network is given by the *Bayes classifier*. In a data set consisting of two classes of data, $\mathcal{A}$ and $\mathcal{B}$, the Bayes classifier gives the probability of a certain input pattern belonging to either class. To see this correspondence, we consider a neural network with an output neuron that tends towards the value 1 if the input pattern is of class $\mathcal{A}$ and tends towards a value of 0 if the input pattern is of class $\mathcal{B}$. We recall the error function for this type of network:

$$E = \frac{1}{2} \sum_i \left[O_i - T_i \right]^2$$

with the summation being over the i events in the sample used to update the weights. Now, the target vector T can only have a value of 0 or 1 as it corresponds to the desired network output for the $\mathcal{A}$ and $\mathcal{B}$ classes of data. If we define $\alpha_{\mathcal{A}}$ to be the fraction of events of type $\mathcal{A}$ in the sample and $\alpha_{\mathcal{B}}$ to be the fraction of events of type $\mathcal{B}$ in the sample along with $P_{\mathcal{A}}(i)$ and $P_{\mathcal{B}}(i)$ as the probabilities of an event i belonging to class $\mathcal{A}$ or $\mathcal{B}$ respectively, the error function above may be written as:

$$E = \left\{ \alpha_{\mathcal{A}} P_{\mathcal{A}}(i) \left[O(i) - 1 \right]^2 + \alpha_{\mathcal{B}} P_{\mathcal{B}}(i) O(i)^2 \right\}$$

In order to find the minimum of this error function, we differentiate with respect to the output $O(i)$ and equate the result to zero. This results in the expression:

$$\alpha_{\mathcal{A}} P_{\mathcal{A}}(i) \left[O(i) - 1 \right] + \alpha_{\mathcal{B}} P_{\mathcal{B}}(i) O(i) = 0$$

$$\Rightarrow O(i) = \frac{\alpha_{\mathcal{A}} P_{\mathcal{A}}(i)}{\alpha_{\mathcal{A}} P_{\mathcal{A}}(i) + \alpha_{\mathcal{B}} P_{\mathcal{B}}(i)}$$

$$\Rightarrow O(i) = \frac{n_{\mathcal{A}}}{n_{\mathcal{A}} + n_{\mathcal{B}}}$$

where $n_{\mathcal{A}}$ and $n_{\mathcal{B}}$ are simply the number of events of type $\mathcal{A}$ and $\mathcal{B}$ respectively in the training sample. So, we see that minimizing the error function in this simple example corresponds to the Bayesian probability of an event being in class $\mathcal{A}$. In multi-layer ANN's, the quality of classification approximates a Bayes classifier with the degree of accuracy depending on the number of hidden units.

5. The Higgs Boson Search Algorithm

In this chapter, the development of an Artificial Neural Network for use in the search for the Standard Model Higgs Boson is presented. Several preliminary tasks were necessary to prepare for the implementation of a neural network based approach to the analysis. First, sufficient simulated physics events were generated in order to train the network to separate the signal HZ events from the background hadronic Z decays and four fermion final state events. Then, an offline analysis of these events was constructed in order to extract useful experimental signatures which offered some means of differentiating the signal and background events. These experimental signatures were then examined in order to choose input variables for the network which would be of the most benefit to the analysis. Only then could a network be constructed and trained to search for the Standard Model Higgs Boson in OPAL data. Finally, an estimation of the effect of systematic uncertainties on the results of the network was performed in order to set the stage for the application of the network to real OPAL data. The application of the network to real data as well as a comparison of the ANN technique to existing search methods for the Higgs boson comprise the last two sections of this chapter.

5.1 Generation of Simulated Data

The generation of several sets of Monte Carlo data was necessary in order to both train and test the performance of the neural network analysis presented in this thesis. As was explained in chapter 4, Artificial Neural Networks require large numbers of template events during training in order to minimize the error function between the network output and the desired target output for a given class of data. In fact, ANN training is most effective when the number of template training events exceeds the number of interconnections in the network by several orders of magnitude. Thus, large amounts of computer time must be

invested in creating Monte Carlo events of both the signal and background channels in order to provide training patterns to the network.

The signal Monte Carlo events used in this thesis were generated using the HZHA [29] simulation program developed at CERN by P. Janot. This generator was programmed to specifically simulate both Standard Model and Supersymmetric Higgs Boson production and takes into account all initial state radiation effects. The background events were generated using several different Monte Carlo programs. The hadronic Z decay events $\left(Z^0/\gamma^* \to q\bar{q}(\gamma)\right)$ were generated using the PYTHIA [30][31] Monte Carlo program. PYTHIA was also used for the generation of the W^+W^- events used in this analysis. Finally, the EXCALIBUR [32] Monte Carlo program was used for the generation of the four fermion final states (including W^+W^- and ZZ) background events. After the generation of the original partons using these Monte Carlo programs, hadronization was performed using JETSET [30].

The resulting four vectors were passed through the OPAL detector simulation program GOPAL that was described previously in section 3.4. The version of the simulation reflected the geometry of the detector actually in place during the data taking at $\sqrt{s} = 161\,\mathrm{GeV}$.

As was described in chapter 4, several sets of data must be generated in order to effectively train and test an Artificial Neural Network. First, a large set of *training* data for each of the classes which the network is expected to identify must be generated. Secondly, some amount of data from each class must be reserved in order to monitor the success of the network during the training process. This data is referred to as the *test* set. Finally, several *validation* sets must be generated which will be presented to the finalized network in order to measure the actual performance of the ANN after training. With these three types of data in mind, Monte Carlo data was generated for each of the signal and background channels.

5.1.1 The Training and Testing Sets

The first set of Monte Carlo data to be produced was the training and testing set. This set also had to contain the largest number of events as a portion of it was used as templates for the ANN during training. The training and testing set consisted of a 4,300 event sample of HZ signal events, a 150,000 event sample of hadronic Z decays, and a 10,000 event sample of W^+W^- events which were used to approximate the four fermion final state background. The accuracy of using W^+W^- events to approximate the four fermion final state background is justified since after the precut (see section 5.3), the four fermion final state events consist mostly of W^+W^- events. When training an Artificial Neural Network, an attempt should be made to present the network with signal and background training samples in a similar ratio to that observed in the real data. With this in mind, far more hadronic Z decay background events were generated than either the signal or the W^+W^- background channels. Table 6 summarizes the Monte Carlo data generated for the training and testing sets.

Process	OPAL Run Number	Events	Generator
$e^+e^- \rightarrow H^0Z^0 \rightarrow b\bar{b}q\bar{q}$ $m_H = 65 \ GeV/c^2$	3432	1 - 4000	HZHA
$e^+e^- \rightarrow Z^0 / \gamma^* \rightarrow q\bar{q}(\gamma)$	4419	1 - 150000	PYTHIA
$e^+e^- \rightarrow W^+W^- \rightarrow$ anything	4420	40001 - 50000	PYTHIA

Table 6: Monte Carlo data produced for the ANN training and testing sets.

After the data in these sets was passed through the precut described in section 5.3 below, it was split up into events which were used to train the network and events which were used to monitor the performance of the network after each training epoch. This splitting of the training and testing sets is described below in section 5.5.

5.1.2 The Validation Sets

In order to test the final performance of the ANN, several validation Monte Carlo data sets were produced. HZHA was again used to generate several Higgs samples at different Higgs boson masses. PYTHIA was used to generate hadronic Z decays to measure the final response of the network to that background, and the EXCALIBUR generator was used to create a large bank of true four fermion background events. None of the validation events were used in either the training or testing sets. The details of the validation set Monte Carlo data are given in Table 7.

Process	OPAL Run Number	Events	Generator
$e^+e^- \rightarrow H^0Z^0 \rightarrow b\bar{b}q\bar{q}$ $m_H = 55 \text{ GeV}/c^2$	3416	2001 - 2500	HZHA
$e^+e^- \rightarrow H^0Z^0 \rightarrow b\bar{b}q\bar{q}$ $m_H = 60 \text{ GeV}/c^2$	3416	2501 - 3000	HZHA
$e^+e^- \rightarrow H^0Z^0 \rightarrow b\bar{b}q\bar{q}$ $m_H = 62 \text{ GeV}/c^2$	3416	3001 - 3500	HZHA
$e^+e^- \rightarrow H^0Z^0 \rightarrow b\bar{b}q\bar{q}$ $m_H = 64 \text{ GeV}/c^2$	3416	3501 - 4000	HZHA
$e^+e^- \rightarrow H^0Z^0 \rightarrow b\bar{b}q\bar{q}$ $m_H = 66 \text{ GeV}/c^2$	3416	4001 - 4500	HZHA
$e^+e^- \rightarrow H^0Z^0 \rightarrow b\bar{b}q\bar{q}$ $m_H = 68 \text{ GeV}/c^2$	3416	4501 - 5000	HZHA
$e^+e^- \rightarrow H^0Z^0 \rightarrow b\bar{b}q\bar{q}$ $m_H = 70 \text{ GeV}/c^2$	3416	5001 - 5500	HZHA
$e^+e^- \rightarrow Z^0 / \gamma^* \rightarrow q\bar{q}(\gamma)$	4419	150001 - 200000	PYTHIA
$e^+e^- \rightarrow$ 4fermion final state	4436	1 - 20000	EXCALIBUR

Table 7: Monte Carlo data used as validation sets to measure the final performance of the ANN analysis.

5.2 Offline Analysis

An offline analysis using the CERN *Discover Higgs* (DH) software package [33] was used to extract useful experimental signatures from the Monte Carlo data. The goal of any offline analysis is to reduce the data to a set of interesting variables which hold some discriminatory power for the analysis problem at hand. The main features of the offline analysis used in this thesis are described in the following sections.

5.2.1 Quality Selection of Tracks and Clusters

The initial stage of the offline analysis selected tracks and calorimeter clusters which were measured with sufficient accuracy by the detector to be considered reliable. Only the selected tracks and clusters were retained for use in the rest of the offline analysis. The quality requirements on the tracks and clusters consisted of the following requirements[†].

- The total number of hits of a given track in the vertex chamber, the jet chamber, and the z chambers was required to be greater than 20 for the track to be accepted.

- The minimum transverse momentum of an acceptable track was set at 0.10 GeV/c.

- The maximum value for $|\tan \lambda|$ for a good track was 100.

- The maximum $|z_0|$ for a good track was set at 50 cm.

- The maximum momentum of an acceptable track was set at 100 GeV/c.

- The maximum impact parameter ($|d_0|$) of an accepted track was set to 2.5 cm.

- The maximum χ^2 of the r-ϕ fit for a track to be good was 100.

[†] The definitions of the track parameters referred to in these quality cuts may be found in section 3.5.1.

- The maximum χ^2 of the z fit to the track was 100.

- The minimum fraction of hits along a track found in the jet chamber was required to be 50%.

- The maximum absolute value of $\cos \theta$ of a good track was set at 0.98 in order to keep tracks well measured within the active volume of the detector.

- A good electromagnetic barrel calorimeter cluster was required to have a minimum energy of 0.05 GeV after corrections and a minimum raw energy of 0.1 GeV.

- A good electromagnetic endcap calorimeter cluster was required to have a minimum corrected energy of 0.05 GeV and a minimum raw energy of 0.25 GeV.

- A minimum of 1 hit block was required to form a cluster in the electromagnetic barrel and a minimum of 2 hit blocks was required in the electromagnetic endcap.

- The total corrected cluster energy was required to be between 0.25 GeV and 100 GeV.

- The energy of a good hadronic calorimeter tower was required to be greater than 0.5 GeV.

- A minimum of 1 tower hit was required to form a cluster.

- The minimum uncorrected total energy for clusters in the barrel or endcap hadronic calorimeters was required to be 0.6 GeV.

- The minimum uncorrected total energy for clusters in the pole tip was required to be 2.0 GeV.

- The energy rescaling factor for the barrel and endcap hadronic calorimeter clusters was 1.05.

• The energy rescaling factor for the pole tip calorimeter clusters was 1.00.

5.2.2 Jet Finding

The next stage in the analysis identified hadronic jets in each event. This was achieved using the Durham jet finder [34]. The jet finder used all hits in the central tracking, electromagnetic calorimeter, and hadronic calorimeter which passed the quality selection presented in section 5.2.1. The Durham jet finder was constrained to find four jets in each event.

The Durham jet finder is typical of cluster algorithms used in e^+e^- environments based on *binary joining*. Initially, each final state particle is considered to be an individual cluster. Using a defined distance measure, the closest two clusters are found and the distance between them determined. If the distance is less than some cut off value, the two clusters are joined into one. This process is repeated until all clusters are separated by a distance larger than the specified cut off. The clusters which remain at the end of the process are then taken to be the final jets in the event. The Durham jet finder has the result common to all binary joining methods that each initial cluster or particle belongs, at the end of the procedure, to only one jet. In the Durham algorithm, the distance measure is given by

$$\tilde{y}_{ij} = \frac{2\min\left(E_i^2, E_j^2\right)\left(1 - \cos\theta_{ij}\right)}{E_{cm}^2}$$

where E_i and E_j are the energies of the two clusters being considered, θ_{ij} is the angle between them, and E_{cm} is the centre of mass energy of the combination. When an event is constrained to a certain number of jets using the Durham jet finder, the cut off value of the distance measure at which the number of jets in the events changes is recorded. For example, for events constrained to four jets, the reported value of y_{34} would indicate the distance measure cut off at which the event changes from containing three to containing

four jets. This variable will be used later in this analysis to help quantify the "four jet nature" of an event.

5.2.3 Energy Flow Correction

Since the energy of charged particles in the detector is measured both in the central tracking subdetectors and in the calorimetry, the possibility exists for a given charged particle to be assigned too high an energy due to double counting of energy loss in the detector. Several algorithms have been developed to eliminate this possible double counting. In this thesis, the OPAL Matching (MT) software package [35] was used to correct the double counting in the energy flow. This package matches tracks in the central tracking detector by extrapolating the track to the electromagnetic calorimeter and the hadronic calorimeter and comparing the location of the track intersection at each of the calorimeters to the cluster centre and boundary. Once an association between a track and a cluster is made, the expected energy response of that particle in the calorimeters is calculated from the track momentum.

Since the central tracking has a better momentum and energy resolution than the calorimetry (except for high energy electrons), the MT package then attempts to reduce the energy measured in a given cluster which has already been measured in the track associated with that cluster. Three distinct alternatives are possible at this point.

1. If a cluster has no associated track, both the energy of the cluster and the momentum measured along the track are kept in the corrected energy and momentum of the event.

2. If a cluster is matched to a track and the energy of the cluster is less than the expected energy response for the given track within a certain tolerance, the cluster energy is ignored and only the track momentum information is retained.

3. If a cluster is matched to a track and the energy of the cluster exceeds the expected energy response of the associated track and is outside of the tolerance, the energy of

the cluster is reduced by an amount equal to the expected energy response of the associated track. The track momentum and the reduced cluster energy are then both retained in the corrected energy and momentum of the event.

The MT package has been shown to improve the momentum-energy flow of OPAL events quite markedly and to bring the final event shape variables closer to those of an ideal detector [35].

5.2.4 Calculation of Event Shape Variables

After the correction of the momentum-energy flow for double counting, the event shape variables can be calculated. Some shape variables used in the analysis presented in this thesis are directly measured such as the total visible mass and energy of an event and the minimum number of electromagnetic calorimeter clusters associated with a jet. Others, however, have more involved definitions and are described in the sections below.

The Effective Centre of Mass Energy

The importance of the hadronic Z decay background is reduced by recognizing the fact that initial state radiation returns the hadronic energy of the event to the Z^0 pole at 90 GeV through the decay $e^+e^- \rightarrow Z^0\gamma \rightarrow q\bar{q}\gamma$ in many cases. Thus, by eliminating the energy of the radiated photon, these events will display an effective centre of mass energy, or $\sqrt{s'}$, around the mass of the Z^0 rather than close to the true centre of mass energy of 161 GeV.

The effective centre of mass energy is calculated by searching for an isolated photon cluster of energy greater than 3 GeV in the electromagnetic calorimeter. The event is then subjected to a kinematic fit which assumes that a photon has been missed in the OPAL beam pipe. The nominal energy of the event is then reduced by the energy of the isolated photon or the postulated beam pipe photon (whichever is larger) to give the value of $\sqrt{s'}$ for the event.

The Thrust

The thrust of an event is used to get an idea of the relative isotropy of the particles in that event. The magnitude of the thrust [36] is given by the equation

$$T = \max \frac{\sum_i \vec{n} \cdot \vec{p}_i}{\sum_i |\vec{p}_i|}$$

where n is a unit vector. The thrust axis is given by the direction of n for which T is maximum. The allowed range for the thrust is $0.5 < T < 1$. A two jet event corresponds to a thrust value of about 1 while a perfectly isotropic event would give a thrust of 0.5.

The Sphericity

The sphericity tensor is also used to measure the relative isotropy of an event. The tensor is defined [37] by the equation

$$S^{\alpha\beta} = \frac{\sum_i p_i^{\alpha} p_i^{\beta}}{\sum_i |p_i|^2}$$

where both α and β range over the values 1, 2, and 3 which correspond to the x, y, and z components of the momentum vector. The tensor may be diagonalized to yield three eigenvalues λ_1, λ_2, and λ_3 which are then used to calculate the sphericity S through the equation

$$S = \frac{3}{2}(\lambda_2 + \lambda_3).$$

The value of the sphericity ranges between 0 and 1 for a given event. A perfect two jet event corresponds to a sphericity of 0 while a perfectly isotropic event corresponds to a sphericity of 1.

5.2.5 Methods of b Quark Tagging

As was mentioned in section 3.3.2, the OPAL silicon microvertex detector makes several sophisticated methods of tagging the signatures of b quarks possible. The fact that the Higgs Boson decays to pairs of b quarks 90% of the time makes efficient b quark tagging extremely desireable for the analysis presented in this thesis. Several algorithms are provided by the OPAL b quark tagging (BT) software package [38] as tools for the lifetime tagging of heavy quarks. Some of these tools are described in the following sections.

Forward and Backward Tracks

The significance of the impact parameter of a given track in the OPAL detector is given by b/σ_b. In this equation, b is equal to the *impact parameter*, or the minimum distance in the r-ϕ plane between the track and the primary event vertex and σ_b is the error in this value. The value of b is signed as follows. If $\hat{j}$ is the vector passing through the primary vertex parallel to the momentum vector of the jet containing the track in question, and $\hat{y}$ is the vector from the primary vertex to the point of closest approach of the track to the primary vertex, the sign of b is equal to the sign of $\hat{j} \cdot \hat{y}$. Note that the determination of the primary vertex with respect to which the value of b for a given track is calculated never includes the track itself. This avoids an intrinsic difference in the calculation of b for tracks which actually form part of the primary vertex when compared to the calculation for tracks which are not part of the primary vertex.

The denominator in the impact parameter significance equation, σ_b, is the error in b. This error includes a contribution from the error in the track's trajectory near the primary vertex as well as from the error in the position of the primary vertex itself.

The number of forward tracks in a jet is defined as the number of tracks for which b/σ_b is greater than some S_{min}. S_{min} for the purposes of the present analysis was set at a typical value of 2.4. Backward tracks are similarly defined as tracks for which b/σ_b is less than -

S_{min}. Forward tracks are associated with long lived hadrons and thus with b quark jets while backward tracks hold no real lifetime information. Often, the light quark background in lifetime tagging of b quarks may be reduced by considering the number of forward tracks minus the number of backward tracks in an event.

Tear Down Vertex Finder

Another b quark tagging tool available in the BT software package is the tear down vertex finder. This algorithm attempts to fit all tracks in a given jet to a single common vertex position. The program then evaluates the contribution of each track to the χ^2 of the vertex fit. Any tracks that contribute more than the cut off value of W_{los} to the fit are discarded and the fit is repeated. The procedure is repeated until none of the tracks used to form the vertex contribute more than W_{los} to the χ^2 of the vertex fit or there are less than three tracks remaining, in which case the vertex finder fails. The value of W_{los} is typically 4.

The significance of the decay length of the newly found secondary vertex from the primary vertex, l/σ_l, is then used to tag heavy quarks. The vertices found in jets arising from b quarks will generally have a larger decay length significance than those arising from lighter quarks due to the longer relative lifetime of b hadrons.

The tear down vertex finder has been found [38] to be most effective in finding displaced secondary vertices in jets arising from b quarks when:

- The b hadron itself travels a large distance before decaying into secondary particles.

- The total charged track multiplicity in the jet arising from the b hadron itself is greater than the total charged track multiplicity arising from the primary vertex.

Build Up Vertex Finder

The build up vertex finder takes the opposite approach to the tear down vertex finder in an attempt to find displaced secondary vertices in a jet. The algorithm begins with a small

number of "seed" tracks consisting of all forward tracks (as defined above) in the jet. The initial proposed secondary vertices are fitted from these seed tracks and thus is displaced from the primary vertex even in light quark events.

Initially, the intersection of each pair of seed tracks is taken as a potential secondary vertex. Then, other tracks in the jet which have a greater separation significance from the primary vertex than from the secondary vertex candidate being considered are fitted one at a time to the seed vertex. The combination of three tracks which gives the best overall probability of the fit to the seed vertex is then chosen as the revised seed vertex as long as the probability of the fit is greater than 1%. This process is repeated until either all tracks in the jet are fitted to a seed vertex, or no more tracks can be fit to the vertex without causing a fit probability below 1%. This combination of N tracks is then taken as the final vertex for the given initial pair of seed tracks. This entire procedure is repeated starting with all combinations of initial seed tracks. All candidate vertices are then considered in order to determine the best choice of secondary vertex for the jet.

The best vertex among the final vertices is then taken as:

- The candidate with the largest number of seed tracks included in its fit.

If this choice is not unique, then the best vertex is determined by:

- The candidate with the largest decay length significance from the primary vertex, if the number of seed tracks in the vertex is greater than two, or

- The candidate with the highest multiplicity of seed tracks fit to it, if the multiplicity is equal to two.

If there are still two or more candidate vertices, then the final best vertex is taken as:

- The candidate with the largest decay length significance from the primary vertex.

The build up vertex finder has been shown [38] to give results which are not completely correlated with the results of the tear down vertex finder. Hence, the inclusion of both the tear down and build up vertex finders in the analysis is beneficial and does not simply result in two highly correlated variables.

5.3 Precut

A loose precut was applied to the Monte Carlo data in order to rid the data set of events which obviously did not fit the four jet profile required for the HZ search. This precut rids the analysis of two jet events from hadronic Z decays, four fermion final states and hadronic Z decays involving leptons, and a large number of radiative return Z decays.

The precut consisted of three subcuts:

1. The Tokyo Multihadron Flag (TKMH) as described in section 3.5.2 was required to be set.

2. The effective centre of mass energy ($\sqrt{s'}$) of the event was required to be greater than or equal to 100 GeV.

3. The value of the Durham jet finder parameter y_{34} where the event changes from a three jet to a four jet topology was required to be greater than or equal to 0.002.

The number of events remaining after the precut and the resulting reduction in the cross section for each process is shown in Table 8. We see that difference in the cross sections between the Excalibur four fermion set (which contains all four fermion final state diagrams including W^+W^-) and the W^+W^- set used in training the network is much reduced by the precut. This indicates that the four fermion final state background after the precut consists largely of W^+W^- events and thus justifies the use of W^+W^- Monte Carlo as training input to approximate the four fermion final state background. The precut was intentionally kept as

loose as possible in order to allow the ANN to do most of the work of separating the signal from the background processes.

Process	Initial Events	Events after Precut	Efficiency	Initial Cross Section (pb)	Cross Section after Precut (pb)
$e^+e^- \rightarrow H^0 Z^0 \rightarrow b\bar{b}q\bar{q}$ $m_H = 64$ GeV/c^2	500	488	97.6%	0.517	0.505
$e^+e^- \rightarrow Z^0 / \gamma^* \rightarrow q\bar{q}(\gamma)$	50000	9150	18.3%	147.4	26.974
$e^+e^- \rightarrow$ 4fermion final state	20000	4055	20.3%	17.13	3.47
$e^+e^- \rightarrow W^+W^- \rightarrow$ anything	38000	27189	71.6%	3.46	2.47

Table 8: The efficiencies and reduction in cross section for the precut when applied to Monte Carlo of the signal and background processes.

5.4 Selection of Network Input Variables

The selection of input variables for the Artificial Neural Network proceeded by considering the experimental signatures of the signal and background processes. As was described in section 2.3.2, hadronic decays of the Z^0 boson and four fermion events (in particular W^+W^- pairs) are capable of imitating the four energetic hadronic jets which form the main signature of HZ signal events. The types of variables which may be used to discriminate between the signal and each of these background types are described in the paragraphs which follow.

As was described in section 5.2.4, the effective centre of mass energy $\sqrt{s'}$ is useful in tagging hadronic Z decays in which the emission of an initial state photon has returned the Z boson to its pole. Thus, the inclusion of this variable in the network will allow these radiative return events to be rejected by the analysis.

After the exclusion of the radiative return events, the remaining hadronic Z decay background consists mainly of four jet events formed from two quarks and two gluons which have been radiated from the quarks. Since the two gluon jets are radiated from the original two quarks coming from the Z^0 decay, they display properties typical of bremsstrahlung processes, namely that they are emitted at a small angle with respect to their parent quark jets, and that they have a relatively soft energy spectrum when compared to the quark jets. In contrast, the four jets coming from a true signal HZ decay arise essentially from the independent decays of two heavy particles and as such, a signal event will be less collinear than a background hadronic Z decay event as well as having a harder energy spectrum for all four jets. Thus, variables which measure the angles between jets as well as the energies of jets will be useful in reducing the hadronic Z decay background.

The tagging of displaced secondary vertices arising from b quark jets forms the most powerful tool in reducing the hadronic Z decay background. Since only a small fraction (~15%) of Z^0 particles decay into $b\bar{b}$ pairs while $H^0 \rightarrow b\bar{b}$ in about 85% of all cases, tagging the signature of b quark production provides a very useful method for separating signal HZ decays from background. Thus, the use of variables derived from the b quark tagging methods described in section 5.2 is necessary to combat the hadronic Z decay background.

The four fermion final state background channels, characterized by W^+W^- decays, ZZ decays, and a multitude of other processes as described in section 2.3.2, are much harder to eliminate from the analysis using only topological variables. While b quark tagging is still effective in reducing the importance of these channels, the fact that W^+W^- and ZZ decays are truly the independent decay of two heavy particles makes the topology of these events

very similar to the signal HZ events in terms of event shape and jet energy profiles. Thus, with the exception of b tagging, the only variables which can effectively discriminate between these events and the desired signal make use of the fact that both heavy bosons in these background events have the same mass. A constrained fit which applies the condition that each of the two jet pairs in the event comes from a mother particle of equal mass is used to achieve this goal with the discriminating variable being the probability of such a fit being true for a given event. Another method of tagging W^+W^- events may be applied which makes use of the fact that the centre of mass energy of 161 GeV is just barely sufficient to produce such boson pairs. Since the resulting W^+ and W^- will be produced almost at rest, the resulting jets from the decay of either boson will be almost back to back. Thus, by reconstructing the mass of the jet pair with the largest angle between them and the other pair, events where the masses are almost equal can be tagged as W^+W^- candidates and thus rejected from the analysis.

A final consideration is the testing of each event with the hypothesis that it actually is a HZ decay. In order to do this, another constrained fit which imposed the nominal Z boson mass on one of the jet pairs was used to calculate the probability of an event containing a Z boson. The mass of the other jet pair was then taken as the candidate Higgs Boson mass. This fit was applied to every possible pairing of jets in a given event and the pairing which resulted in the highest probability for the constrained fit was taken to be the candidate pairing for the Higgs hypothesis.

With the above motivations for the choice of variables in mind, several variables were chosen as initial input variables for the Artificial Neural Network. In addition to variables which allowed good separation between the signal and background events, several variables which did not display a good separation were included in the analysis in order to observe how they would be incorporated into the network's learning procedure. The chosen variables are listed below in three groups: b quark tagging variables, event shape variables, and kinematic fit variables.

5.4.1 The b Quark Tagging Variables

The following variables were chosen as inputs involving the b quark tagging methods used in the offline analysis. Each of the variables is given a short name in bold face by which it is referred to hereafter in this document. The degree of separation between the spectra of these variables for 64 GeV/c^2 Higgs Monte Carlo and hadronic Z decay Monte Carlo is shown in Figure 32 while the separation between the Higgs Monte Carlo and Excalibur four fermion Monte Carlo is shown in Figure 33.

1. **(tdsum)** The sum of the b tagging significance given for each of the "Higgs" jets by the tear down b tagging method. The "Higgs" jets are determined by the most probable Z mass constrained fit as was described above.

2. **(fbmul)** The sum of the *forward* tracks in each of the "Higgs" jets minus the sum of the *backward* tracks in each of the "Higgs" jets. The forward and backward tracks are defined in terms of their impact parameters as was described in section 5.2 above. The "Higgs" jets are again determined by the most probable Z mass constrained fit to the event.

3. **(busig)** For each of the "Higgs" jets, the significance determined by the build up b tagging method is multiplied by the square root of the charged track multiplicity for that jet. The sum of this quantity over the two "Higgs" jets is then taken as the input variable. The scaling of the vertex significance by the square root of the multiplicity of the jet reflects the fact that the build up b tagging method provides a stronger result when more tracks are used in the vertex determination. The "Higgs" jets are again identified by the most probable Z mass constrained fit to the event.

5.4.2 The Event Shape Variables

The following paragraphs describe the event shape variables which were used as initial inputs to the Artificial Neural Network. Each variable is again referred to by an

abbreviated name shown in boldface in each entry. The relative separation between HZ signal Monte Carlo and hadronic Z decay Monte Carlo for these variables is shown in Figure 34 while the separation between the HZ Monte Carlo and four fermion Monte Carlo is shown in Figure 35.

Figure 32: The spectra of the three b tagging variables used as input to the Artificial Neural Network for the 64 GeV/c^2 signal (solid histograms) and for the hadronic Z decay background (dashed histograms). The variables are described in section 5.4.1. The distributions are all taken after the precut described in section 5.3 and equal numbers of signal and background events are plotted.

Figure 33: The spectra of the three b tagging variables used as input to the Artificial Neural Network for the 64 GeV/c^2 signal (solid histograms) and for the four fermion (Excalibur) background (dashed histograms). The variables are described in section 5.4.1. The distributions are all taken after the precut described in section 5.3 and equal numbers of signal and background events are plotted.

1. **(minct)** The minimum number of charged particles per jet in the event.

2. **(minem)** The minimum number of electromagnetic calorimeter clusters per jet in the event.

3. **(spher)** The *sphericity* of the event. Sphericity is defined in section 5.2 above.

4. **(cost)** The cosine of the polar angle (θ) of the thrust vector of the event. The thrust of an event is also defined in section 5.2 above. This variable shows little separation between the signal and background data classes and was included in order to test the response of the network to a non-discriminating variable.

5. **(evis)** The visible energy of the event. The visible energy is simply all the energy recorded by the detector in a given event after correction for tracker/calorimeter double counting as was described in section 5.2 above.

6. **(angl)** The minimum angle between any pair of jets in the event. This variable is useful as it discriminates between the relatively small emission angle of the gluon jets in the hadronic Z decay background when compared to the large angle production of the jets in the signal events.

7. **(mvis)** The visible mass of the event. This variable simply sums all of the mass in a given event which was detected by OPAL after energy corrections for double counting.

8. **(sprime)** The effective centre of mass energy $\sqrt{s'}$ as was calculated in the offline analysis. This variable is useful in rejection radiative return events in the hadronic Z decay background channel.

9. **(ntrack)** The total number of charged tracks in the event. Although the number of charged tracks in an event indicates the hadronic content, this variable was included as a low discimination power variable since the hadronic purity of the data sample is quite high following the selections of the precut described in section 5.3.

10. **(y34)** The Durham jet finder, when constrained to find a certain number of jets reports this parameter at which the event would change from containing three to containing four jets if the jet finder was left to its own devices. The higher the value of this

parameter, the more likely that the event is of four jet nature. This variable was also expected to be of limited benefit in the neural network since the precut selects four jet events very efficiently.

11. (y45) This variable is defined in an analogous way to y34 above except that it measures the threshold between the four jet and five jet nature of an event.

12. (y23) This parameter of the Durham jet finder indicates the threshold between the two jet and three jet nature of an event and was postulated to be of some use in rejecting two jet events which had a wide jet profile vulnerable to being interpreted as four individual jets.

5.4.3 Constrained Fitting Variables

In the paragraphs below, the input variables resulting from constrained fitting on the kinematics of the events are described. Again, each variable is referred to by an abbreviation indicated in boldface. The separation between the 64 GeV/c^2 Higgs Monte Carlo and the hadronic Z decay Monte Carlo for these variables in shown in Figure 36 while the separation between the HZ and four fermion Monte Carlo is shown in Figure 37.

1. (mdiff) The difference between the reconstructed masses of the jet pair with the largest angle between them and the other jet pair using the energy and momentum conservation constrained fit to calculate the masses. This variable is useful as a veto for W$^+$W$^-$ events which are produced at threshold as their lack of boost will cause each of them to decay into back-to-back jet pairs.

2. (mlow) The mass of the lowest mass jet pair. This variable is calculated using the energy and momentum conservation constrained fit. It is useful to reject hadronic Z decay events because of their softer gluon jets.

3. **(zprob)** The probability of the Z mass constrained fit for the jet pairings giving the highest probability. This tests the quality of the HZ hypothesis.

4. **(eqprob)** The probability of the equal mass constrained fit for the jet pairings which give the highest probability. This serves as a veto for W^+W^- and ZZ background events.
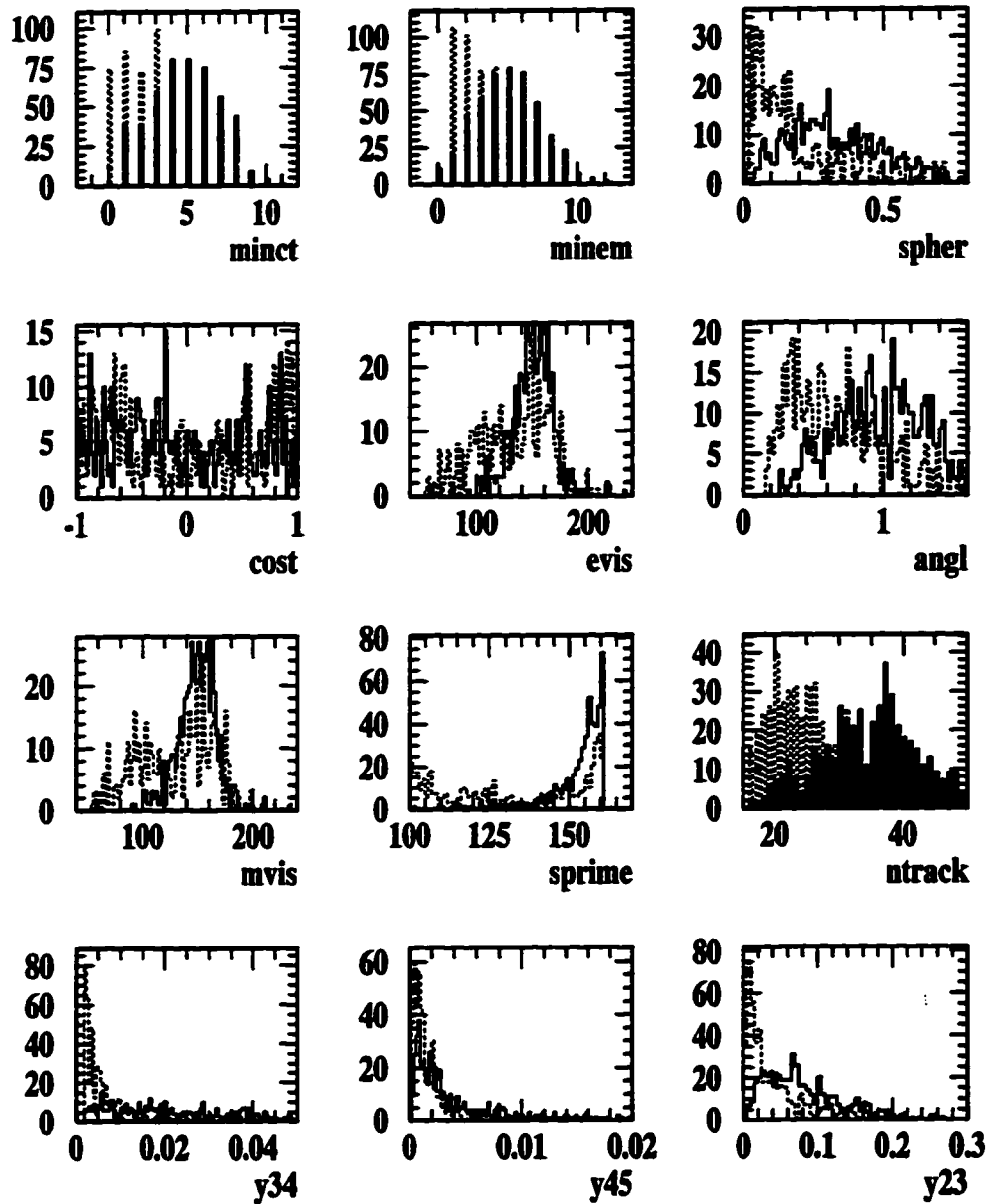
Figure 34: The spectra of the twelve event shape variables used as input to the Artificial Neural Network for the 64 GeV/c^2 signal (solid histograms) and for the hadronic Z decay background (dashed histograms). The variables are described in section 5.4.2. The distributions are all taken after the precut described in section 5.3 and equal numbers of signal and background events are plotted.
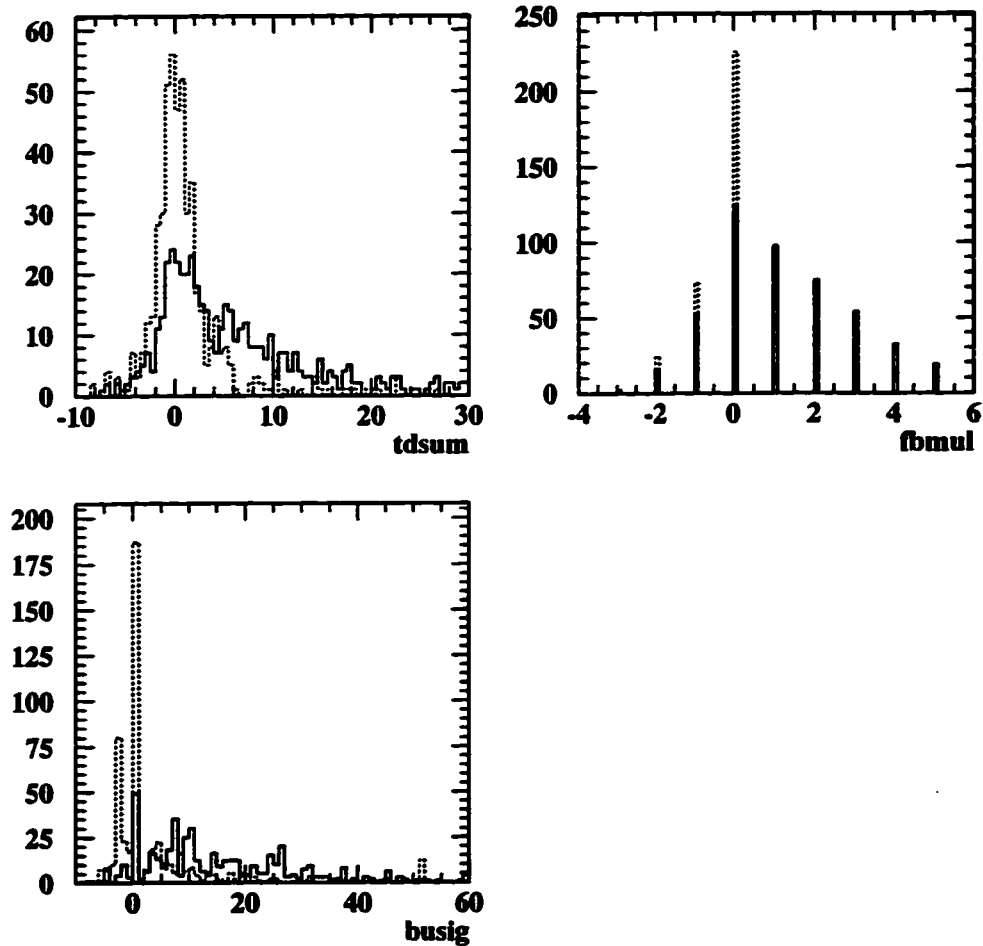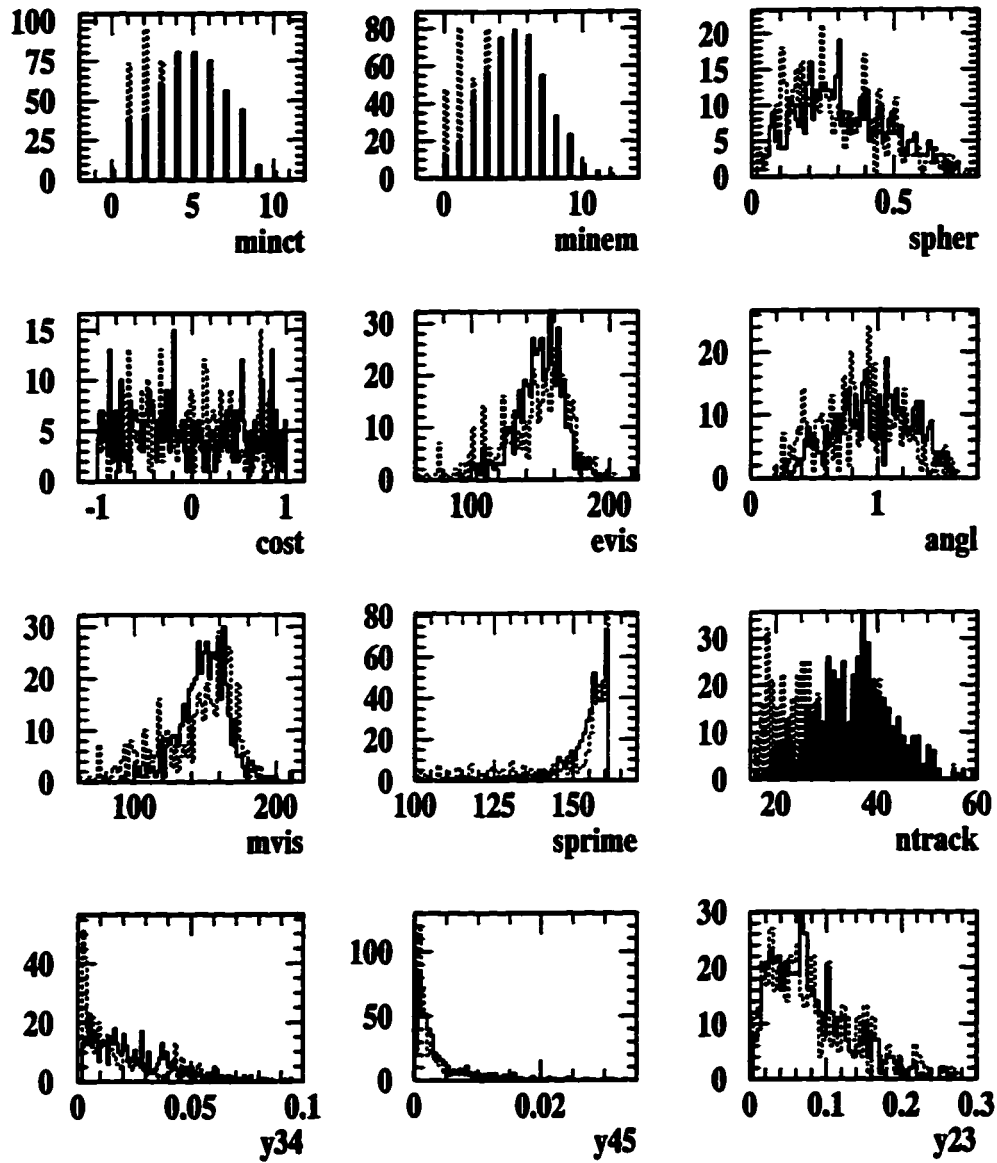
Figure 35: The spectra of the twelve event shape variables used as input to the Artificial Neural Network for the 64 GeV/c² signal (solid histograms) and for the four fermion (Excalibur) background (dashed histograms). The variables are described in section 5.4.2. The distributions are all taken after the precut described in section 5.3 and equal numbers of signal and background events are plotted.
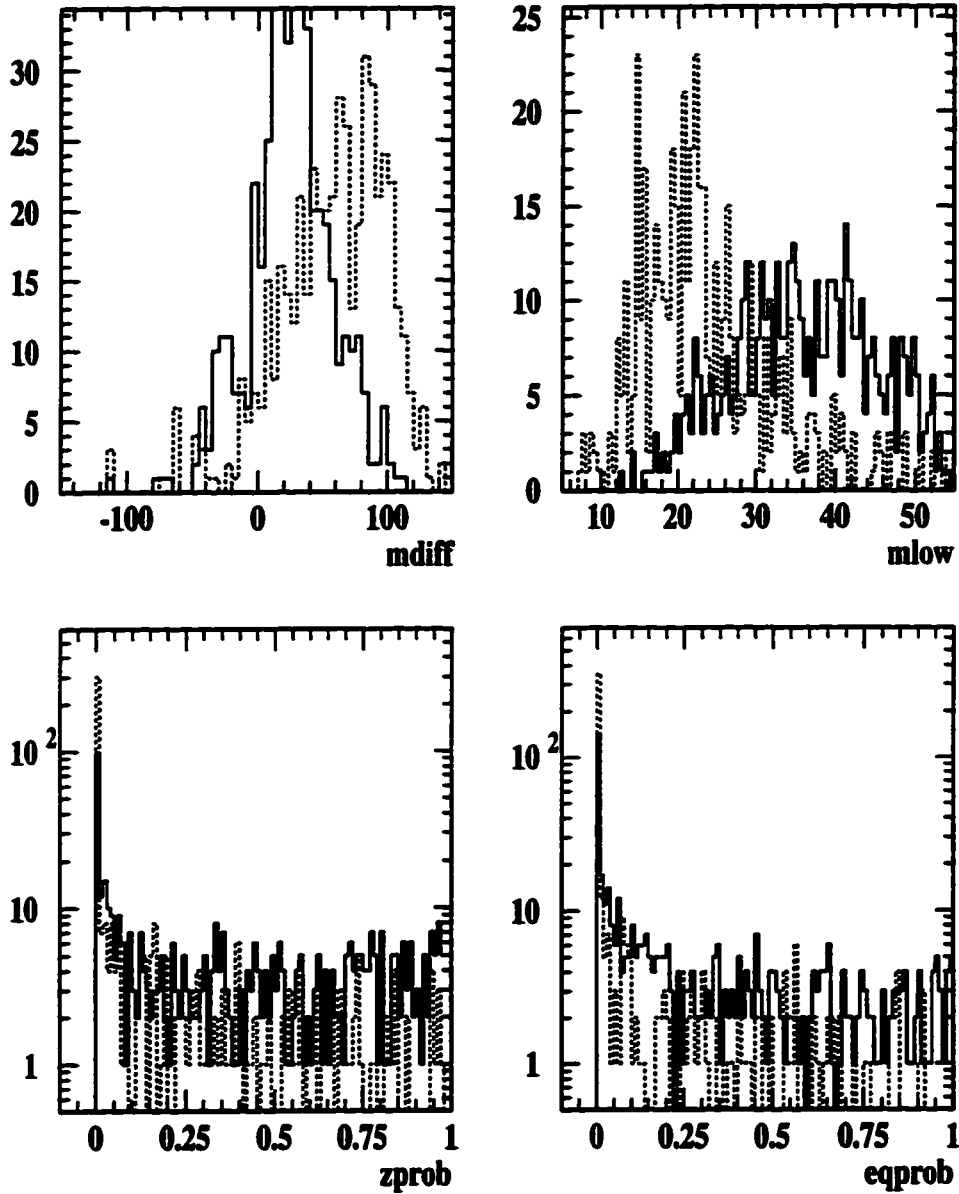
Figure 36: The spectra of the four constrained fitting variables used as input to the Artificial Neural Network for the 64 GeV/c^2 signal (solid histograms) and for the hadronic Z decay background (dashed histograms). The variables are described in section 5.4.3. The distributions are all taken after the precut described in section 5.3 and equal numbers of signal and background events are plotted.
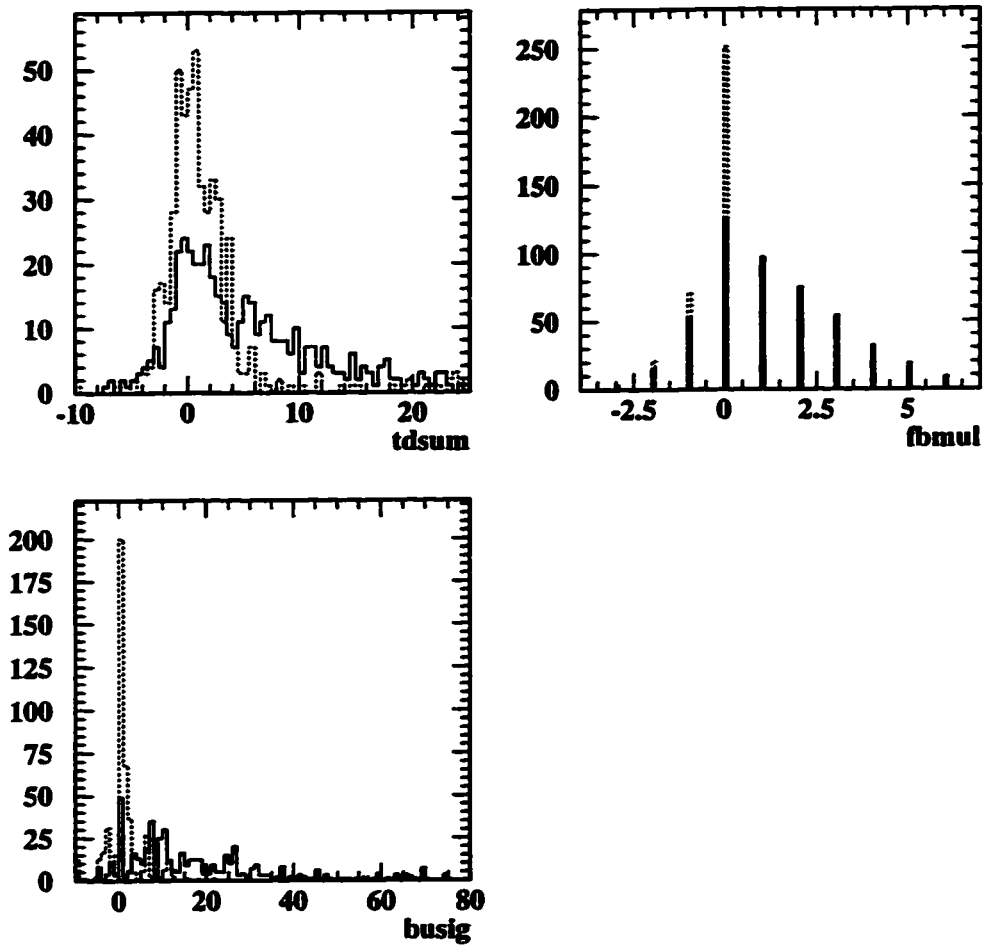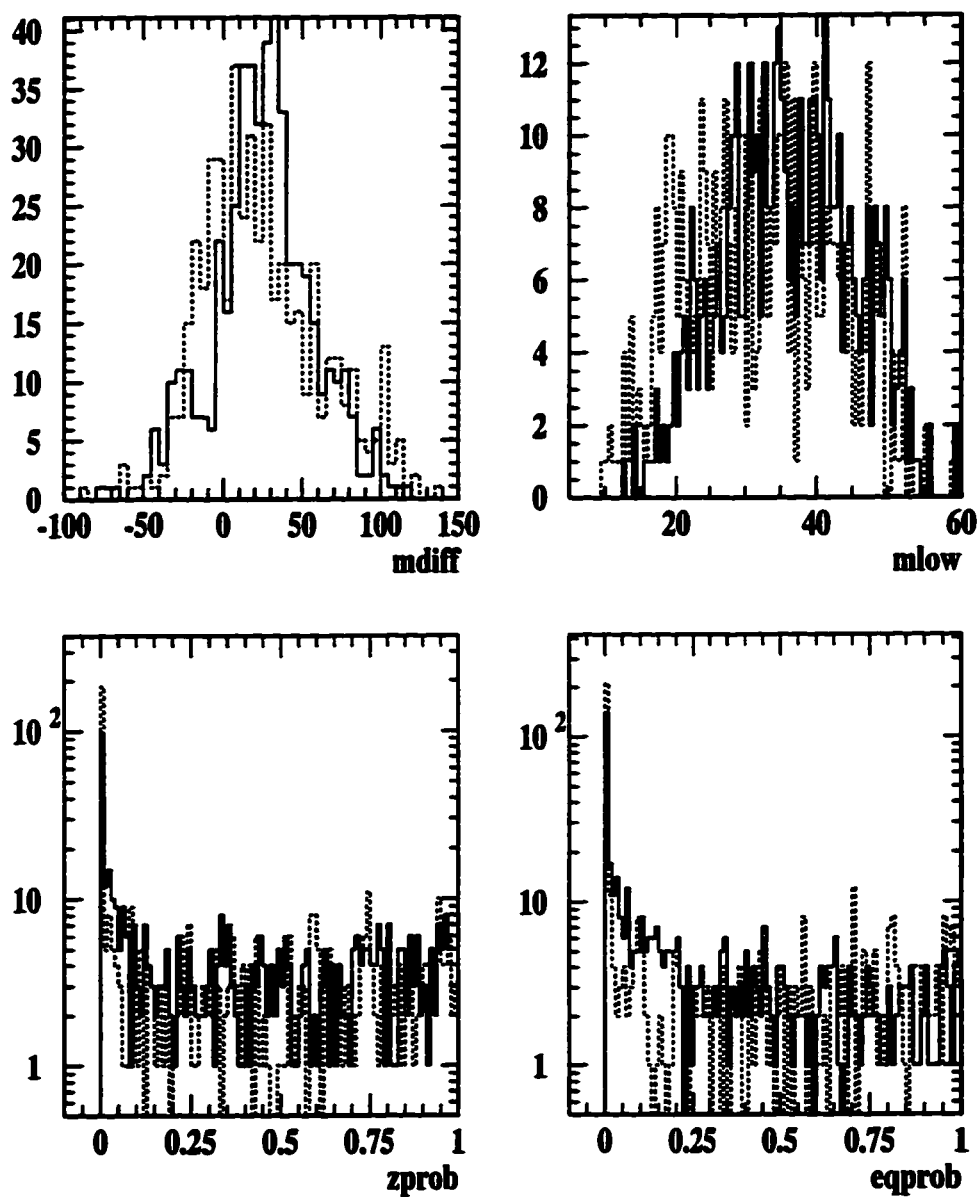
Figure 37: The spectra of the four constrained fitting variables used as input to the Artificial Neural Network for the 64 GeV/c² signal (solid histograms) and for the four fermion (Excalibur) background (dashed histograms). The variables are described in section 5.4.3. The distributions are all taken after the precut described in section 5.3 and equal numbers of signal and background events are plotted.

5.4.4 Discriminatory Power of Input Variables

In order to estimate the discriminatory power of the input variables prior to the process of training the Artificial Neural Network, the degree of separation between the 64 GeV/c^2 HZ signal sample and each of the hadronic Z decay and four fermion background Monte Carlo samples was quantified for each of the nineteen input variables. The separation between the signal and background samples was defined by λ_f where:

$$\lambda_f = \frac{(\mu_1 - \mu_2)^2}{\sigma_{12}^2}$$

In the preceeding equation, μ_1 and μ_2 are the mean values of the variable for the signal and background respectively and σ_{12} is the rms value for the distribution of both signal and background together. The larger the value of λ_f, the greater the separation between the signal and the background in that particular variable.

Table 9 shows the values of λ_f for each of the nineteen input variables for both the separation between the HZ signal and hadronic Z decay background and between the HZ signal and four fermion background. The table lists the variables in decreasing order of separation significance for each of the two background types. As expected, the topological variables top the list for successful discrimination between the signal and the hadronic Z decay background while the b tagging variables are important for reducing both the hadronic Z decay and four fermion backgrounds. Surprisingly, the probability of the equal mass constrained fit does not produce a large separation in the four fermion case, but this is most likely due to the large spike of zero probability events which skew the results of the λ_f calculation. It can also be seen in Table 9 that the degree of separation for variables such as the cosine of the polar angle of the thrust axis is low as expected. The relative utility of the variables as given by their values of λ_f as compared to their relative values when they are used to train the Artificial Neural Network will be examined in section 5.6 below.

Hadronic Z Decays qq(γ)			4 Fermion Background	
Variable	**λ**		**Variable**	**λ**
mlow	1.27		busig	0.818
y34	1.045		tdsum	0.513
ntrack	0.99		fbmul	0.479
sprime	0.89		ntrack	0.43
angl	0.72		minem	0.333
minct	0.698		minct	0.323
minem	0.593		sprime	0.192
busig	0.57		mlow	0.15
mdiff	0.502		zprob	0.0822
spher	0.489		mvis	0.078
zprob	0.475		y45	0.071
tdsum	0.416		angl	0.067
y45	0.394		evis	0.056
evis	0.377		y34	0.028
fbmul	0.371		mdiff	0.013
mvis	0.335		y23	0.01
eqprob	0.317		spher	0.005
y23	0.246		cost	0.004
cost	0.0032		eqprob	0.0002

Table 9: The degree of separation between the 64 GeV/c² HZ signal Monte Carlo and the two major background Monte Carlo distributions for the nineteen input variables. The variables are listed in decreasing order of separation.

5.5 Initial Network Training

Although some of its input variables were of questionable utility in separating the Higgs signal from background processes, the initial nineteen input network was trained in order to evaluate the degree of success that could be expected using an Artificial Neural Network search technique. The ANN presented in this section was programmed using the JETNET [26] Artificial Neural Network package that was described in chapter 4. The network consisted of nineteen input neurons, a hidden layer with twenty neurons, and a single output neuron. Each neuron was of the dot product type with a sigmoidal activation

function. During training, signal (HZ) events were assigned a target output of 1 while background events (qq(γ) and WW) were assigned a target of 0.

The mean squared error between the network output and the target vectors was used as the quadratic form for minimization. Standard backpropogation was used as the learning method and the following parameters were used during the training:

- Learning Parameter η=0.001

- Momentum α=0.5

- Temperature[t] T=1.0

- Weights updated once per every 10 training patterns processed.

The set of Monte Carlo data described in section 5.1.1 was divided into separate training and testing sets as follows. A training set of 4,000 HZ events, 21,000 hadronic Z decay (qq(γ)) events, and 4,000 WW events was used as input to the network during the minimization process. At the end of each training epoch, the network performance was evaluated using a test set of 1,000 HZ, 5,000 qq(γ), and 1,000 WW events which were completely independent of the training events. All events in both the training and testing sets had already passed the precut described in section 5.3. As was stated in chapter 4, all input variables to an Artificial Neural Network should range over a similar order of magnitude in order to avoid a "stiffening" of the weights which results in some neurons not learning properly. Thus, before being input to the ANN, all variables were shifted to a mean of zero and scaled to unit variance across the combination of the training and testing sets in order to provide a range of inputs over similar orders of magnitude to the network.

---

[t] The temperature used during training simply gives the slope of the central portion of the sigmoidal activation function used in the neurons.

The network was initially trained over 10,000 epochs in order to ascertain the correct number of epochs to use to achieve a minimal test set error. The mean squared error between the network output and the target value for the test set is shown versus the training epoch in Figure 38. The sharp increase in error after about 1,000 training epochs indicates that the network is becoming overtrained past this point. In order to optimally train the network, the training process was then repeated and terminated near the error function minimum at 1,000 epochs.
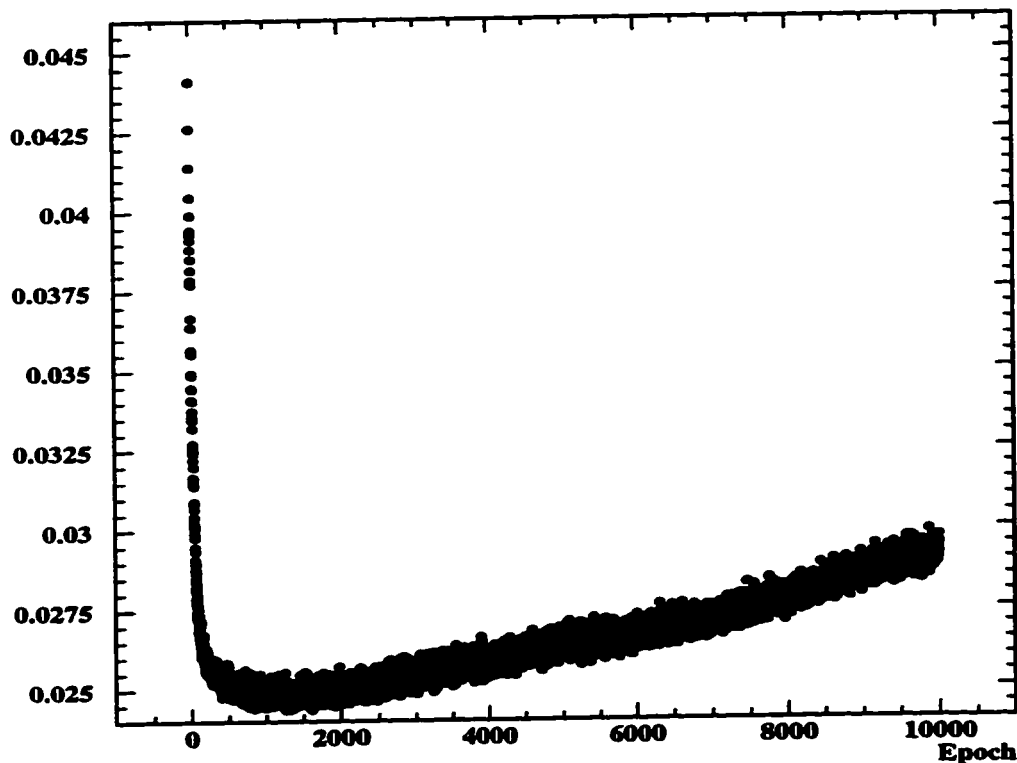
Figure 38: The mean squared error between the network output on the test set and the training target versus the training epoch. The gradual increase in the error after about 1,000 epochs indicates the onset of overtraining.

In order to see the if network was "learning" the signature of the HZ events, the number of known HZ events in the Monte Carlo test set which resulted in a network output greater than 0.7 was recorded at the end of each training epoch. This efficiency for the signal events is plotted in Figure 39. The fact that the network is successfully learning is quite evident as the HZ signal efficiency starts close to 0% in the first training epoch and reaches a nearly stable value close to 50% after 1000 epochs of training. The corresponding proportions of qq(γ) and WW test set events which resulted in network outputs of greater than 0.7 are shown in Figure 40. It is quite clear from these plots that the network is successfully learning to isolate the HZ signal events while not accepting many of the background events.
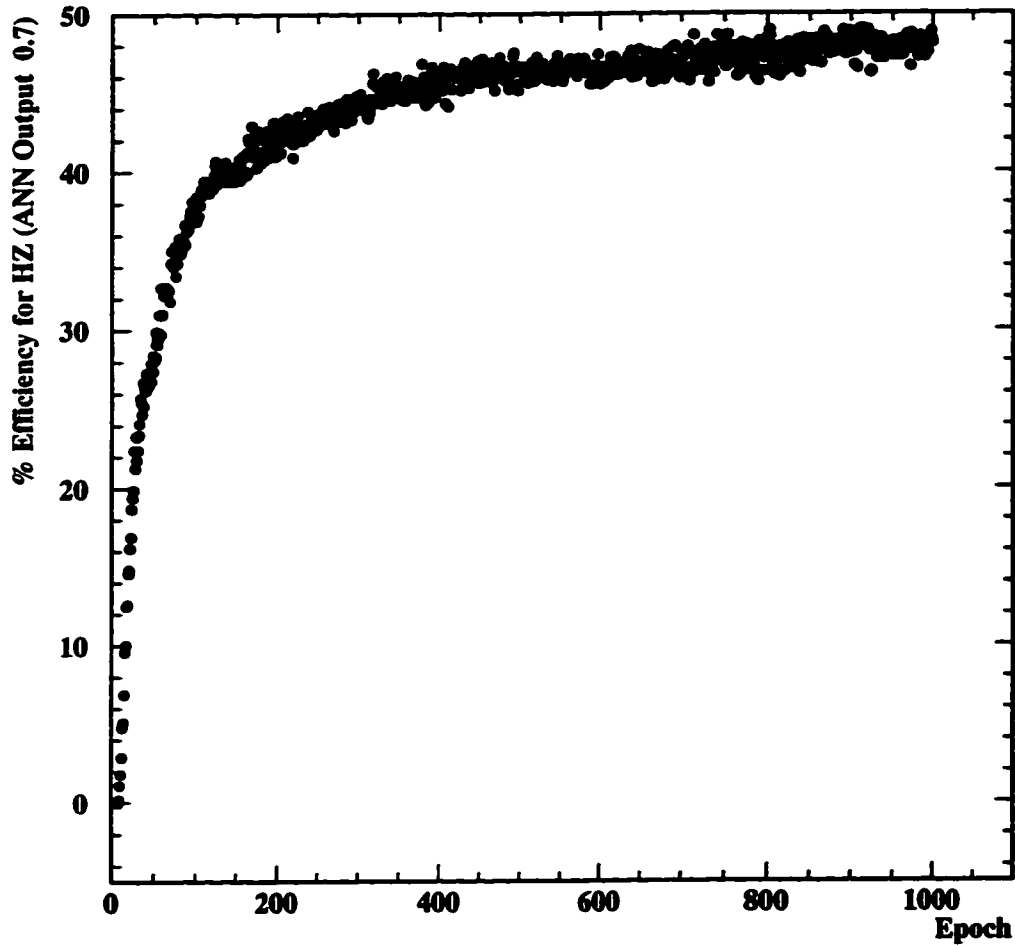
Figure 39: Efficiency for selection of HZ signal events when the output of the network is required to be greater than 0.7 versus the training epoch.

The successful identification of the HZ events by the network may also be seen in Figure 41. In this figure, the output of the network after 1,000 epochs of training is plotted for the signal events and for each of the background types. As was expected, the signal test set events result in an output distribution peaked near a value of 1 while both the $q\bar{q}(\gamma)$ and WW background test sets result in an output distribution peaked near 0.
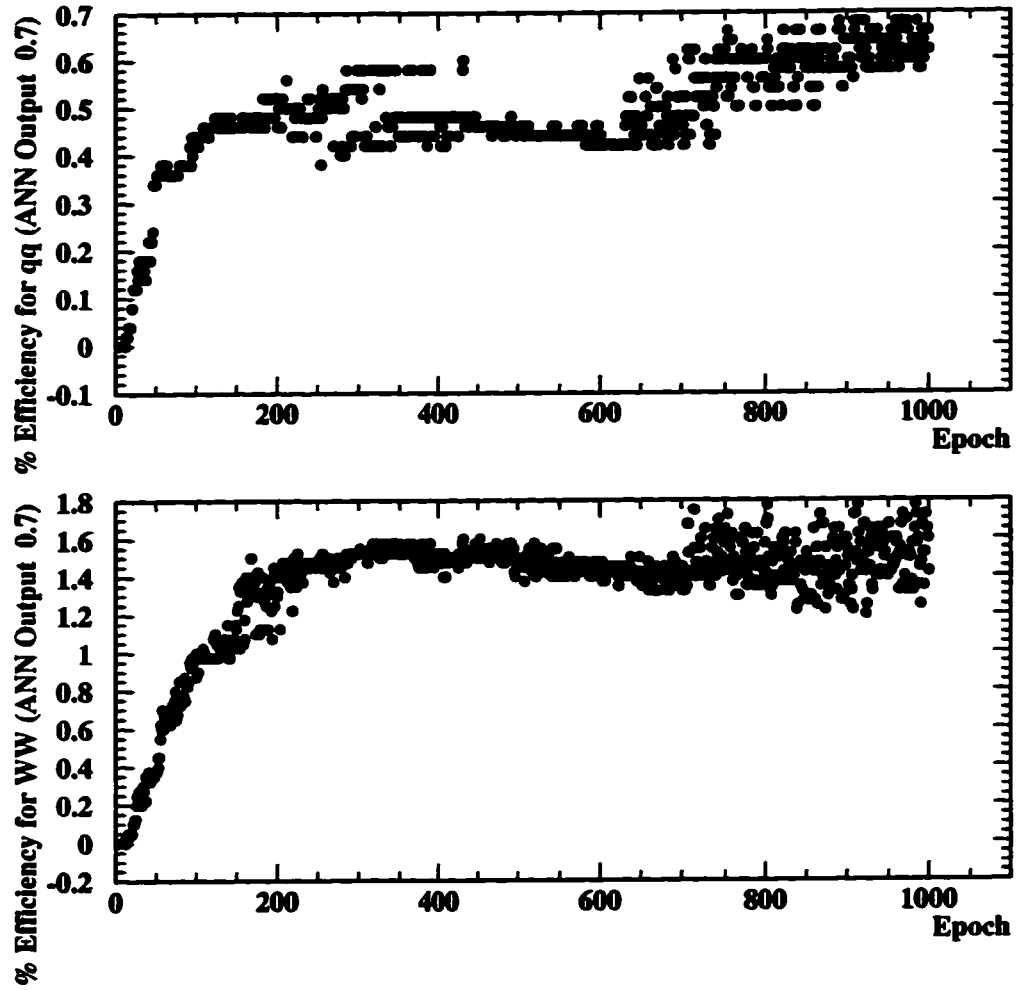
Figure 40: Percentage of qq(γ) events (top) and WW events (bottom) from the test set which are misidentified as HZ events by causing a network output greater than 0.7. Note that the network selects much less of both backgrounds than of the signal.
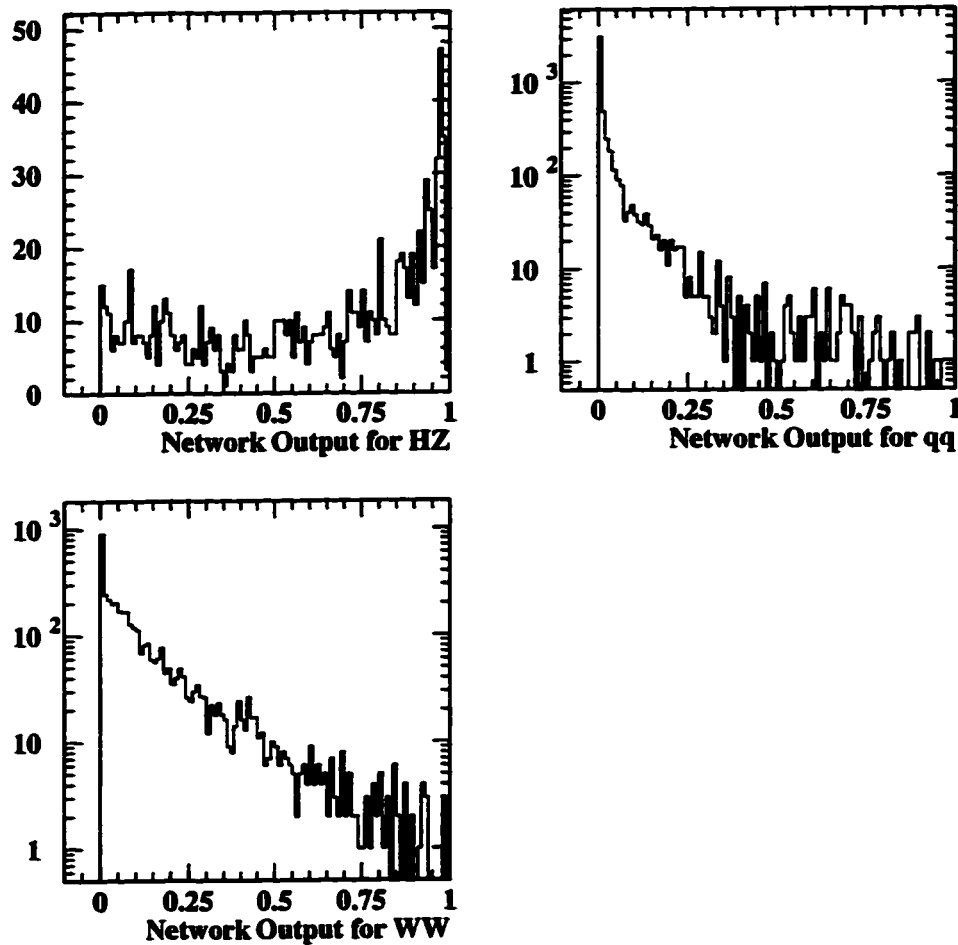
Figure 41: Network output after 1000 epochs of training for events in the test set. The output for HZ signal events is shown in the upper left, for qq̄(γ) background events in the upper right, and for WW background events in the lower left.

In order to finalize the performance of the nineteen input network, the weights were frozen after 1,000 epochs of training and the network was then applied to the validation sets. The HZ validation set with a Higgs mass of 64 GeV/c^2 was used to evaluate the signal efficiency while the 50,000 event qq̄(γ) sample and the 20,000 event four fermion sample from the Excalibur generator were used to determine the effect of the network on the

background. None of these validation events had been used previously in the training or testing of the network. The following cuts were performed on each of the validation sets in order to measure their response to the analysis.

1. The precut (described in section 5.3).

2. A cut on the value of the network output.

3. A *postcut* consisting of the requirement that the probability of the most probable five parameter Z mass constrained fit to the event be greater than 1%. This cut was applied to reject events that did not have a jet pair that was likely to have come from a Z decay. It was not applied prior to the neural network cut as it allowed an independent monitor of the quality of the events passing through the ANN. The quality of the events passing the ANN cut is shown by the small effect the precut has on the number of survivng events.

Table 10 shows the number of remaining events in each of the validation samples after the sequential application of the precut, several values of network output cut, and the postcut. It also shows the corresponding efficiencies and the remaining cross section of each data sample after the cuts. The remaining cross section of the signal set, each of the background sets, and the sum of the backgrounds calculated using these efficiencies and rejections is shown versus the value of the network output cut applied in Figure 42. The efficiency for the 64 GeV/c^2 Higgs sample and the signal to background (signal to noise) ratio is shown versus the network output cut in Figure 43. It can be seen that, for network output cuts greater than 0.7, the remaining amount of signal surpasses the impurity represented by the total background thus giving signal to noise values greater than 100%.

Signal HZ m_H=64 GeV/c²									
No Cut	500								
Precut	488								
ANN Cut	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Remaining	419	381	341	311	274	252	225	178	112
Postcut	348	316	288	267	241	220	197	161	102
Efficiency	69.6%	63.2%	57.6%	53.4%	48.2%	44.0%	39.4%	32.2%	20.4%

Hadronic Z Decay qq(γ) Background									
No Cut	50000								
Precut	9150								
ANN Cut	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Remaining	772	415	247	176	106	61	35	12	5
Postcut	518	298	176	124	74	39	27	10	5
Rejection	98.964%	99.404%	99.648%	99.752%	99.852%	99.922%	99.946%	99.980%	99.990%

4 Fermion Final State Background									
No Cut	20000								
Precut	4055								
ANN Cut	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Remaining	1268	753	498	332	227	147	79	49	16
Postcut	1079	662	443	302	210	135	72	47	15
Rejection	94.605%	96.690%	97.785%	98.490%	98.950%	99.325%	99.640%	99.765%	99.925%

Table 10: The number of remaining events in each of the validation sets is shown after sequential application of the precut, various values of network output cut for the nineteen input ANN, and the postcut. The efficiency for the signal and rejection for the backgrounds is also given for each cut.
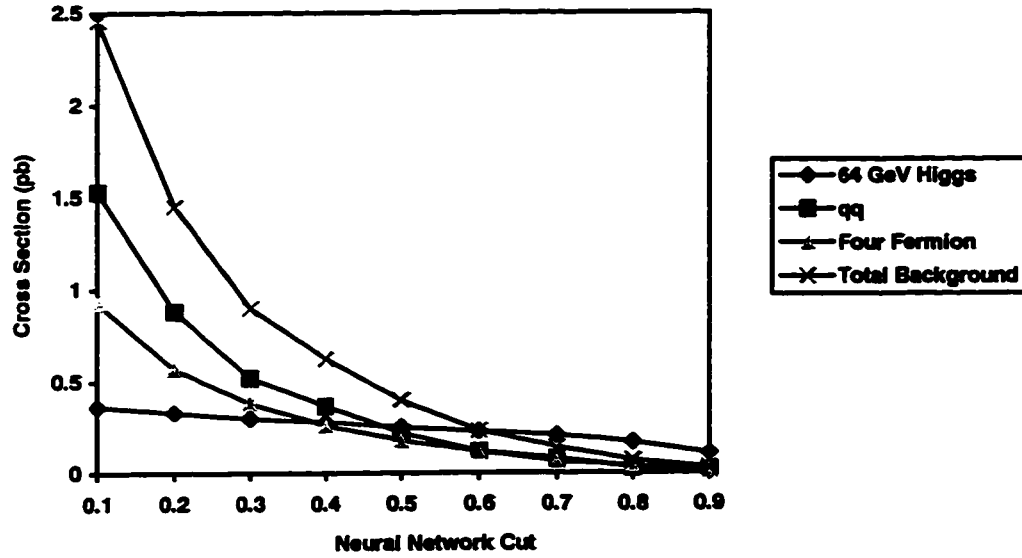
Figure 42: The remaining cross sections of HZ signal for a 64 GeV/c^2 Higgs, qq(γ) background, four fermion final state background, and total background after the sequential application of the precut, various values of network output cut, and the postcut.
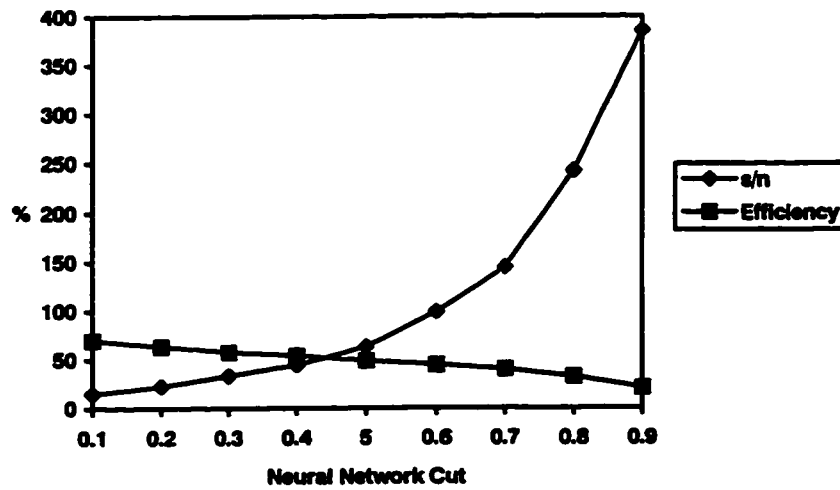
Figure 43: Selection efficiency for 64 GeV/c^2 Higgs events and signal to noise values after the sequential application of the precut, various values of the network output cut, and the postcut.

The success of the nineteen input network indicated that the Artificial Neural Network technique was capable of separating the HZ signal from the major backgrounds, and as such, further investigation of the possibilities of this method was warranted. The next section describes the pruning of the original network in order to achieve a minimal network which was still capable of the same quality of classification.

5.6 Network Pruning

The nineteen input ANN presented in the previous section had a total of 421 interconnections between its constituent neurons. As was alluded to in chapter 4, training of an ANN is made more efficient if the number of training template events exceeds the number of interconnections in the network by several orders of magnitude. Thus, it is in the best interests of an efficient analysis to remove extraneous inputs from the ANN in order to create the minimal network which will still perform an adequate classification of the signal and background events.

In order to successfully prune the network inputs without adversely affecting the classification quality of the network, the following algorithm was devised. The sum of the weights between a given input node i and each node in the hidden layer was reported at the end of the training run of 1,000 epochs. This sum S_i may be written as:

$$S_i = \sum_{j=1}^{n_h} w_{ij}$$

Where n_h is the number of neurons in the hidden layer of the network. Initially, the two input variables with the lowest two weight sums S were removed from the network and the ANN was then retrained with the reduced number of inputs and a number of hidden neurons equal to the number of inputs plus one. The final mean squared test set error between the network output and the target values of this reduced network was then compared to the mean squared error for the initial nineteen variable network. If the error

had increased by an amount less than 5% of the error of the original network, the process was repeated and the two variables with the lowest two weight sums in the reduced network were removed. The number of hidden neurons was also reduced by two and the training process was then repeated again with this further reduced network. Each reduced network was trained over 1,000 epochs. When the final mean squared test set error of the reduced network became more than 1.05 times the error of the original network, the pruning process was stopped and the iteration was taken as the final reduced network.

The final mean squared test set errors of each of the iteratively reduced networks as well as the input variables that were dropped and the fractional increase in the final mean squared error are shown in Table 11.

# Remaining Variables	Final Error	Increase in Error	Dropped Variables
19	2.47×10^{-2}	-	cost, fbmul
17	2.52×10^{-2}	2.02%	y23, evis
15	2.52×10^{-2}	2.02%	y45, minct
13	2.52×10^{-2}	2.02%	ntrack, minem
11	2.69×10^{-2}	8.91%	-

Table 11: The evolution of the network pruning process. The number of remaining variables, mean squared test set error after 1,000 epochs of training, increase in the test set error, and variables dropped at the end of each iteration are shown. The pruning was terminated when the increase in the test set error exceeded 5%. The thirteen input network was selected as the final pruned network.

The four variables with the highest weight sums remained the same in each iteration of the pruning process. These variables were (in decreasing order of weight sum), the build up b quark tagging variable (busig), the tear down b quark tagging variable (tdsum), the effective centre of mass energy of the event (sprime), and the mass of the lowest mass jet pair in the event (mlow). We see that the b tagging variables are most important as they help to tag both the hadronic Z decay and four fermion background and the two shape variables which identify the radiative Z return events and and the softer gluon jets coming

from the hadronic Z decays are second most important. The variables which are identified as being most important in the ANN, however, do not exactly agree with the importance of the variables as measured by the separation significance λ_f which was determined in section 5.4.4. For example, while the value of y34 was found to be a good variable for separating hadronic Z decay events from the signal by the separation significance method, it does not enter into the top four variables identified by the weight sums in the ANN. This indicates that relationships between the input variables are important in the ANN and that connections between variables are being used which are not immediately apparent when each input variable is considered separately.

The pruning process resulted in the thirteen input network being selected as the final pruned ANN. The remaining input variables consisted of:

1. The minimum number of electromagnetic calorimeter clusters per jet.

2. The sphericity of the event.

3. The minimum jet-jet angle.

4. The mass difference between the jet pair with the largest angle in between them and the other jet pair.

5. The sum of the tear down significances of the vertices associated with the "Higgs" jets.

6. The visible mass of the event.

7. The effective centre of mass energy ($\sqrt{s'}$) of the event.

8. The number of charged tracks in the event.

9. The Durham jet finder parameter y_{34}.

10. The probability of the most probable Z mass constrained fit to the event.

11. The probability of the most probable equal mass constrained fit to the event.

12. The build up b quark tagging variable as described in section 5.4.1.

13. The mass of the lowest mass jet pair.

This network had only 111 interconnections, a reduction of 74% from the original nineteen input network. The thirteen variable network was adopted as the ANN for the rest of the analysis presented in this thesis.

5.7 Final Network Training

The thirteen variable network was subjected to a training procedure similar to the initial training of the nineteen variable network. This network was again defined using the JETNET package with the thirteen input variables listed in the previous section, fourteen hidden neurons in a single layer, and a single output node. Each neuron was of the dot product type with a sigmoidal activation function and full connectivity to all neurons in directly adjoining layers. Standard backpropagation was used to update the weights during the training and the mean squared error was again used as the global error function.

The training and testing set of Monte Carlo data described in section 5.1.1 was again divided into a training set consisting of 4,000 HZ signal events, 21,000 $qq(\gamma)$ background events, and 4,000 W$^+$W$^-$ background events and a testing set consisting of 1,000 HZ signal events, 5,000 $qq(\gamma)$ events, and 1,000 W$^+$W$^-$ events. All of the training and testing events had passed the precut described in section 5.3.

During the training of the thirteen variable network, several training parameters in the ANN package were varied to asses the effect of non-standard choices on the learning process. The network was first trained with the following standard choices for the training parameters.

- Learning Parameter η=0.001

- Momentum α=0.5

- Temperature T=1.0

- Weights updated once per every 10 training patterns processed.

- Performance on the test set measured at the end of each training epoch.

The network was trained with these parameters over 10,000 epochs to see where the test set error function reached a minimum. The evolution of the error function on the test set error with the training epoch is shown in Figure 44. Again it can be seen that the error function reaches a minimum at about 1,000 epochs. This plot also shows a less dramatic increase in the test set error function at the onset of overtraining when compared to the same plot for the nineteen input network in Figure 38. This reduction in the severity of the overtraining for the thirteen input network can be explained by the fact that the reduced number of degrees of freedom in the weights of the smaller network provide less opportunity for the network to "overlearn" the signatures of the training set.

With the optimum training time set at 1,000 epochs, the learning parameter, momentum, number of events per weight update, and temperature were all varied in order to observe their effects on the test set error function as the network was trained.

Varying the temperature in the range of 0.5 to 1.5 had no visible effect on the test set error function except at the extreme high end of the range where the resulting large activation function caused the sigmoidal function to resemble a step function. At this point, the network did not minimize the error function at all. Varying the learning parameter, however, created some large differences in the evolution of the error function. Figure 45 shows the test set error function plotted against the training epoch for two different values of the learning parameter η. The crosses show the evolution of the error for the default

learning parameter of 0.001 while the circles show the same quantity for a learning parameter of 0.1. We see that a larger value of the learning parameter causes the test set error evolution to become flat with a wide variance when compared to the result for the smaller learning parameter. This reflects the fact that a large learning parameter allows the ANN to take large steps in the weight space when attempting to minimize the error function. While this means that the minimum is reached in fewer epochs, the resulting variance in the error function found in subsequent epochs makes the results of the network less accurate. The learning parameter was also changed to a value of 0.0001, but at this low step size the test set error did not progress towards a minimum at all in the 1,000 training epochs.
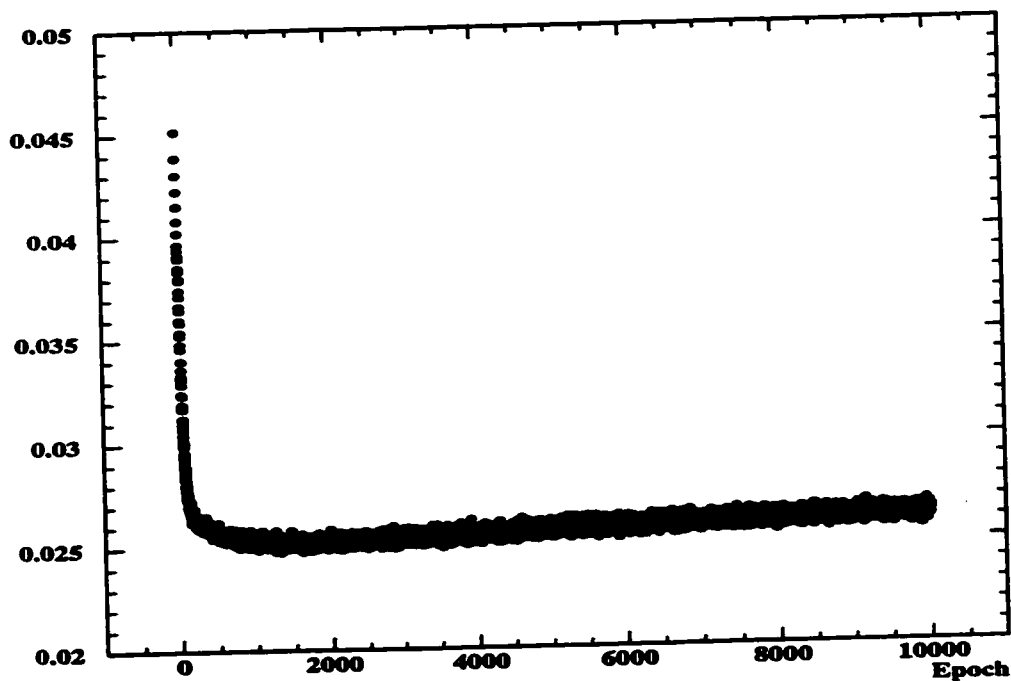
Figure 44: The evolution of the test set error function with the training epoch for the thirteen input network with the default training parameters. Note that the minimum is again reached around 1,000 epochs.
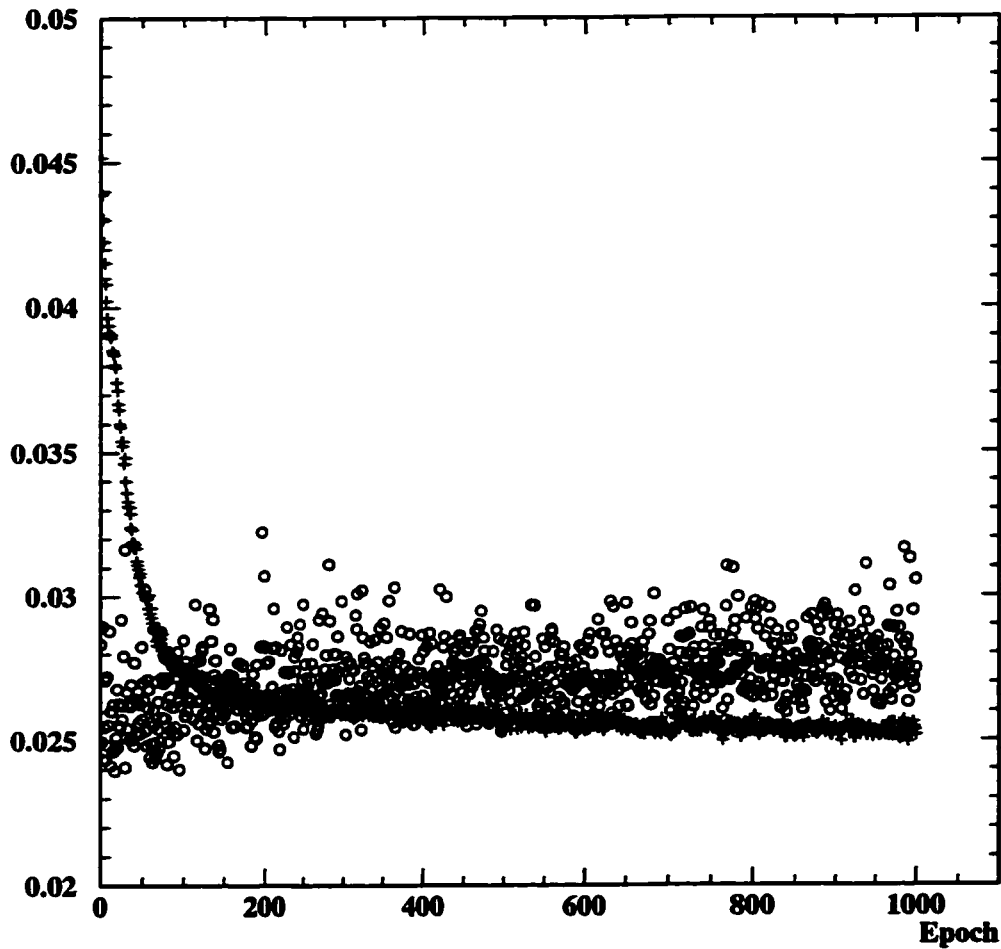
Figure 45: The evolution of the test set error function with the training epoch for two different values of the learning parameter. The crosses correspond to a learning parameter of 0.001 while the circles correspond to a learning parameter of 0.1.

Next, the momentum parameter α was varied in the range between 0 and 1. Figure 46 shows the evolution of the test set error function for two values of the momentum. The crosses show the evolution for a momentum of 0.5 while the circles correspond to a momentum of 0. Momenta closer to 1 resulted in the network not learning at all and the test set error function therefore did not proceed towards a minimum. These results indicate

that providing the network with a "memory" of the size of its last minimization step in weight space allows faster learning up to a certain value of α. Beyond this value, however, the network continues to take large steps even as it nears the minimum and hence the network does not effectively learn.
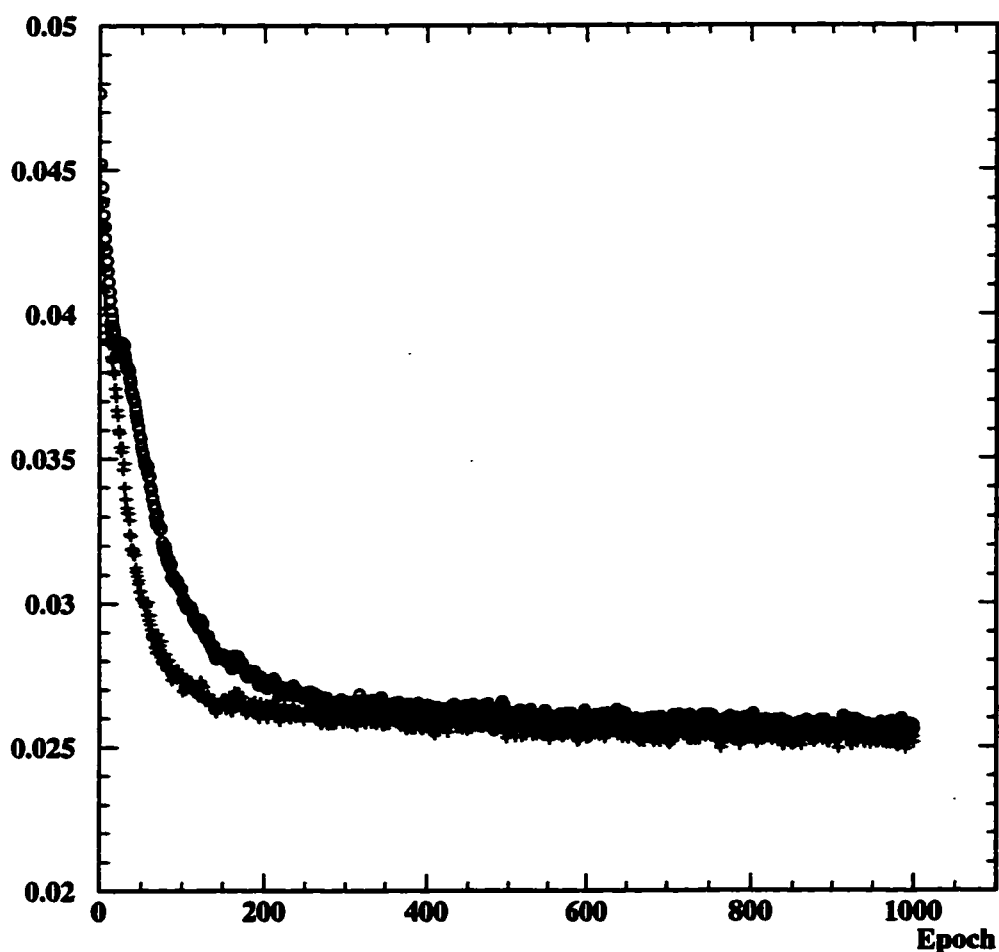
Figure 46: The evolution of the test set error function with the training epoch for two different values of the ANN momentum term. The crosses correspond to a momentum of 0.5 while the circles correspond to a momentum of 0.

The number of event processed before each update of the weights during the training process was also varied. Figure 47 shows the evolution of the test set error function for 1, 10, and 100 events per weight update. While setting this parameter at 1 event per update allows the network to reach an error minimum more quickly, the resulting higher variation in the error function between epochs makes the 10 event per update value more attractive. When the parameter is set to 100 events per update, the network is much slower to reach a minimum.
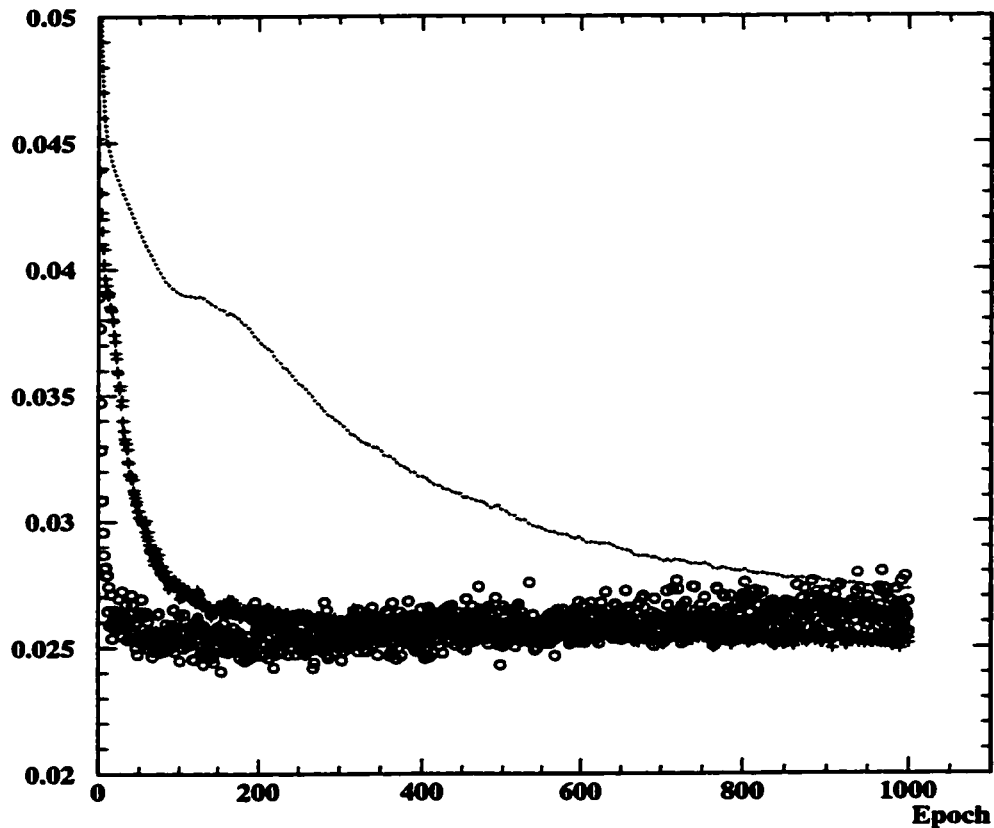
Figure 47: The evolution of the test set error function versus the training epoch for three different values of the number of events per weight update. The circles correspond to one event per update, the crosses to ten events per update, and the dotted line to 100 events per update.

As a final check on the success of the network training, the *saturation* of the neurons in the hidden layer and the output neuron was monitored over the 1000 training epochs. The saturation of a neuron is a measure of its current location on its activation function sigmoid. Nodes which have reached activations on the flat tail of the sigmoid are no longer actively learning and are a warning flag that the network is becoming saturated. A saturation measure of 1 indicates that this state of affairs exists in a given neuron. Figure 48 shows both the average saturation of the nodes in the hidden layer and the saturation of the output node over the 1000 training epochs. It can be clearly seen that at no point during the training is the network becoming saturated and thus all neurons remain active learners throughout the training process.
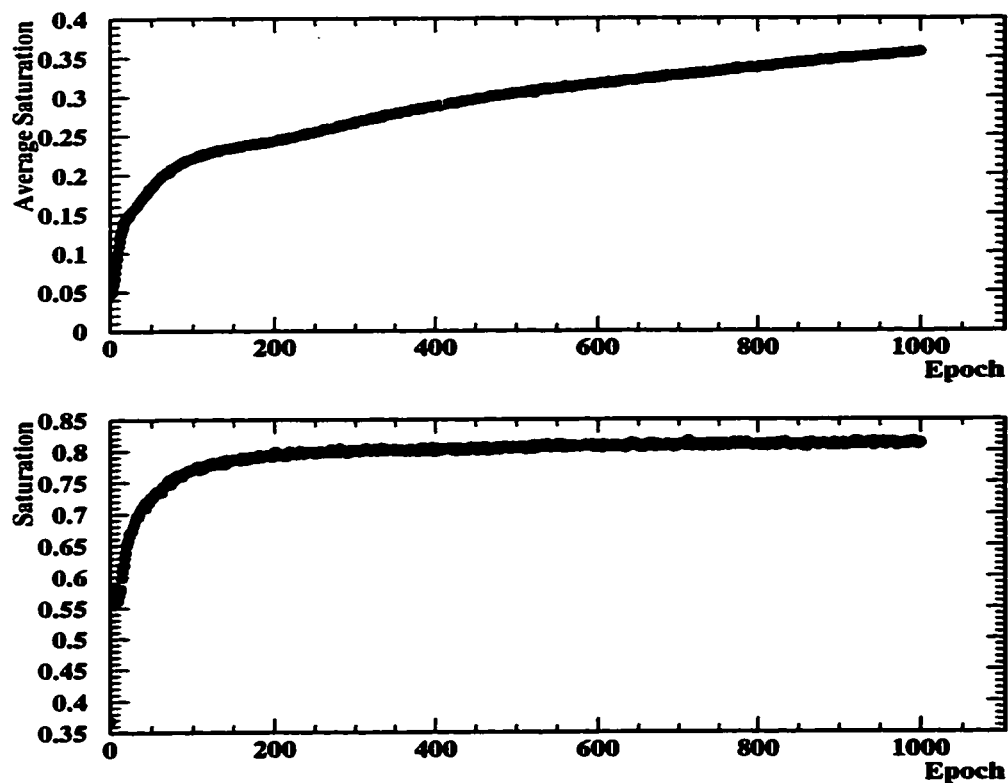
Figure 48: The average saturation of the neurons in the hidden layer (top) and of the output neuron (bottom). All neurons remain active learners throughout the training process.

The above considerations indicate that the original choices of the learning parameter, temperature, momentum, and number of events per weight update were good choices for the training of the Higgs search ANN. Parameters in the range of those selected for this analysis have been used in other search analyses [39] [40] [41] with success levels greater than standard multivariate cutting approaches.

In summary, the thirteen input network was finally trained over 1000 epochs using a learning parameter of 0.001, a momentum of 0.5, a temperature of 1.0, and updating the weights once per 10 events. The efficiency of the network on the HZ test set when an event was required to give an ANN output greater than 0.7 to be tagged as a Higgs event is shown in Figure 49. As was the case for the nineteen input network, the thirteen input ANN is again seen to learn with the initial efficiency of near 0% climbing to a pseudo-stable value near 50% after 1000 epochs of training. Figure 50 shows the corresponding acceptance for both the $q\bar{q}(\gamma)$ test set and the $W^{+}W^{-}$ test set. Again we see that the efficiency for the signal events is far greater than the contamination of background events which are accepted. Finally, Figure 51 shows the network output for the HZ signal of the test set as well as the output for the $q\bar{q}(\gamma)$ and $W^{+}W^{-}$ test sets. Again it can be seen that the network gives outputs peaked at one for signal events and at zero for background.
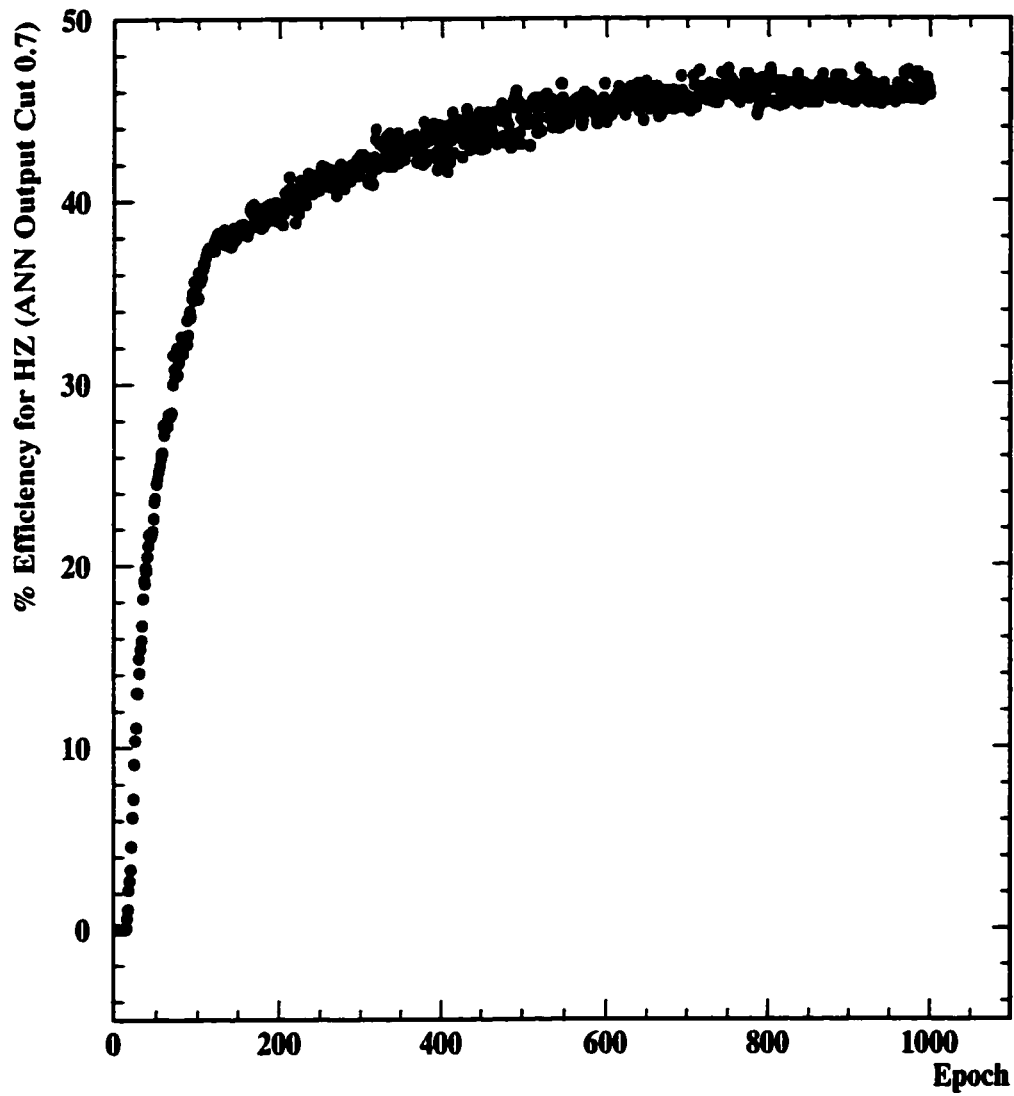
Figure 49: Efficiency for the selection of HZ signal events when the output of the thirteen input network is required to be greater than 0.7 versus the training epoch.
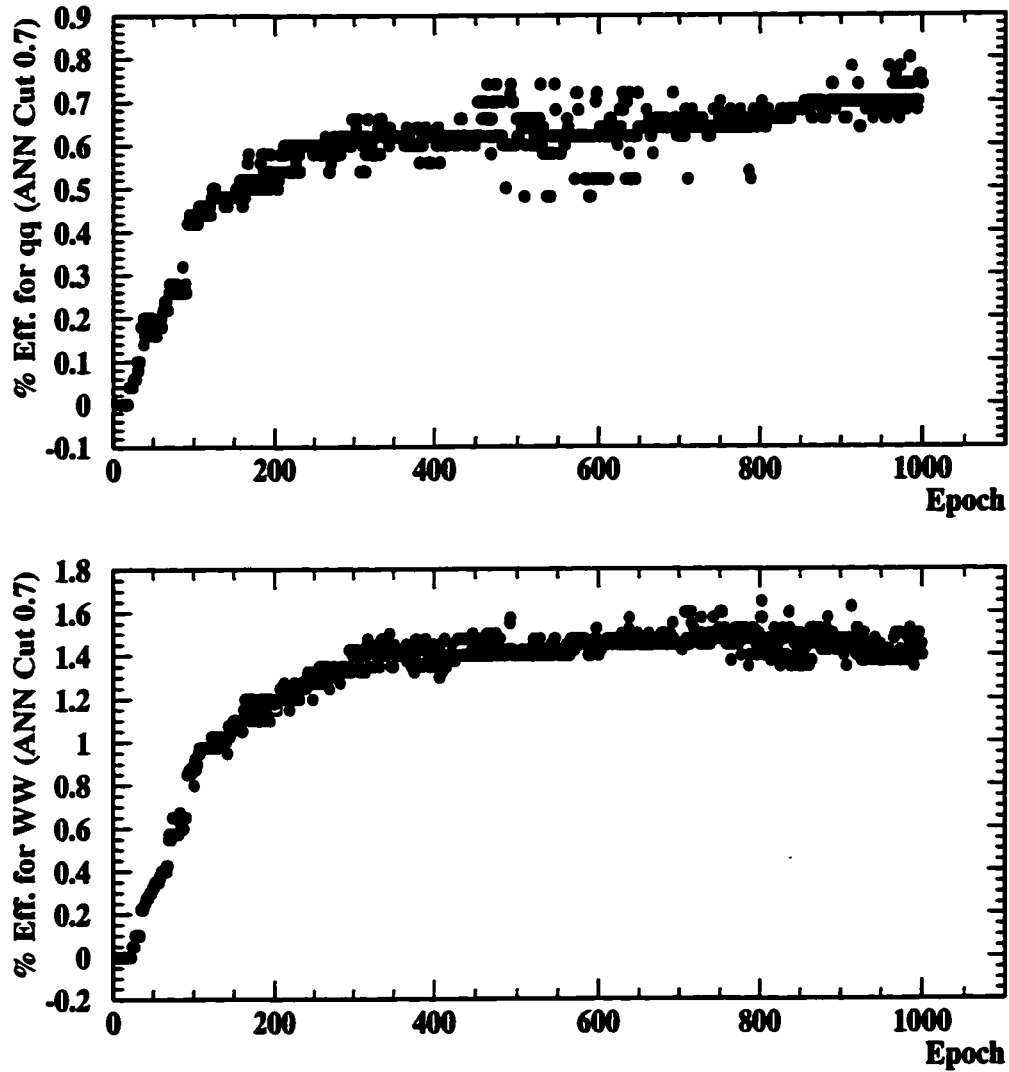
Figure 50: Percentage of qq(γ) events (top) and WW events (bottom) from the test set which are misidentified as HZ events by causing a thirteen input ANN output greater than 0.7. The network again selects much less of both backgrounds than of the signal.
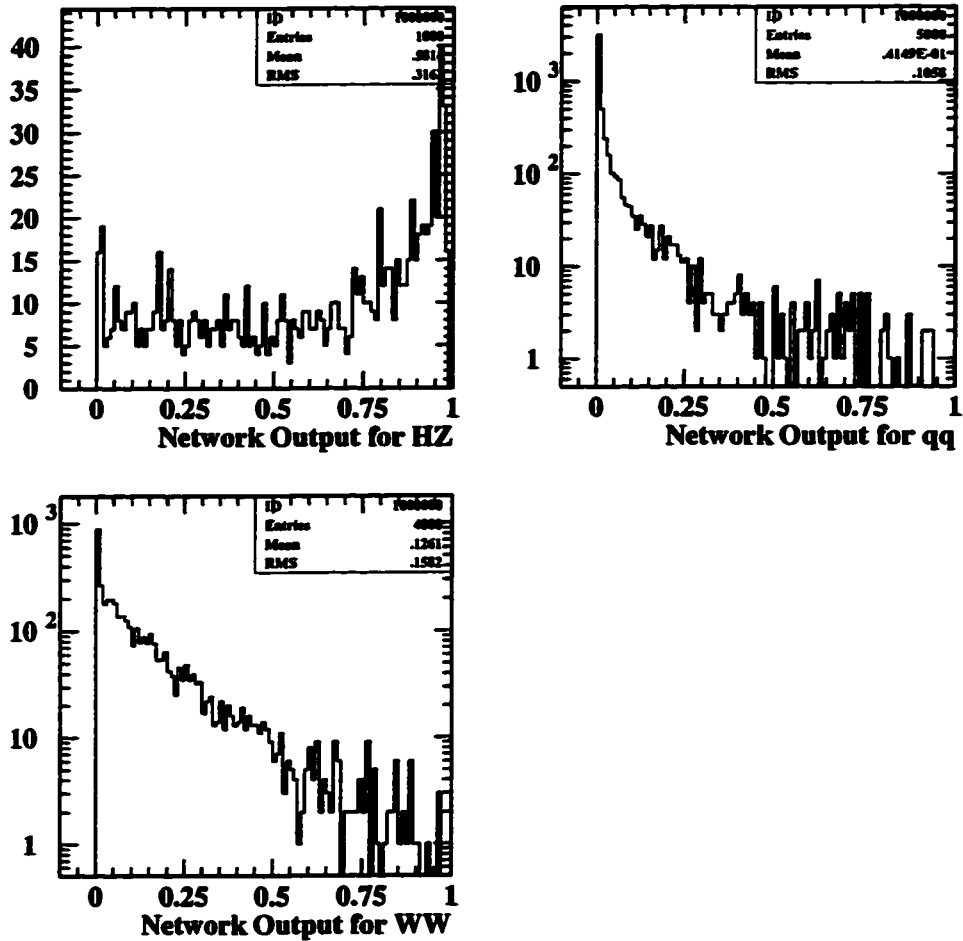
Figure 51: Network output after 1000 epochs of training for events in the test set using the thirteen input network. The output for HZ signal events is shown in the upper left, for qq(γ) events in the upper right, and for WW background events in the lower left.

5.8 Network Performance

The final performance of the thirteen input network was again evaluated using the validation sets of Monte Carlo data which were described in section 5.1.2. The final cut sequence was the same as that applied in the case of the nineteen input network and

consisted of sequential application of the precut (as described in section 5.3), a cut on the output of the thirteen input ANN, and the postcut which required that the probability of the most probable Z mass constrained fit to the event be greater than 1%. This postcut was again applied in order to provide a final quality check on the event and only rejects a small number of the events after the ANN cut. The number of events remaining in the 64 GeV/c^2 HZ signal validation set and the qq(γ) and four fermion final state validation sets are shown in Table 12. The efficiency for the signal and rejection for the background processes is shown for several cuts on the output of the network. Figure 52 shows the remaining cross sections of the 64 GeV/c^2 signal validation set, the qq(γ) validation set, the four fermion final state validation set, and the total background (the sum of the qq(γ) and four fermion final state sets) versus the magnitude of the network output cut. Figure 53 shows the signal efficiency and signal to noise ratio achieved for various ANN output cuts. It can be seen from these two figures that the amount of remaining signal again exceeds the background for ANN cuts greater than 0.7.

Signal HZ m_H=64 GeV/c²									
No Cut	500								
Precut	488								
ANN Cut	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Remaining	419	370	331	301	272	242	222	174	99
Postcut	349	310	280	258	236	213	196	160	92
Efficiency	69.8%	62.0%	56.0%	51.6%	47.2%	42.6%	39.2%	32.0%	18.4%

Hadronic Z Decay qq(γ) Background									
No Cut	50000								
Precut	9150								
ANN Cut	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Remaining	784	444	284	174	112	87	43	15	1
Postcut	537	306	194	122	72	54	36	13	1
Rejection	98.926%	99.388%	99.612%	99.756%	99.856%	99.892%	99.928%	99.974%	99.998%

4 Fermion Final State Background									
No Cut	20000								
Precut	4055								
ANN Cut	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Remaining	1381	841	536	343	205	146	92	41	10
Postcut	1172	743	480	310	185	134	85	40	10
Rejection	94.140%	96.285%	97.600%	98.450%	99.075%	99.330%	99.575%	99.800%	99.950%

Table 12: The number of remaining events in each of the validation sets is shown after sequential application of the precut, various values of network output cut for the thirteen input ANN, and the postcut. The efficiency for the signal and rejection for the backgrounds is also given for each cut.
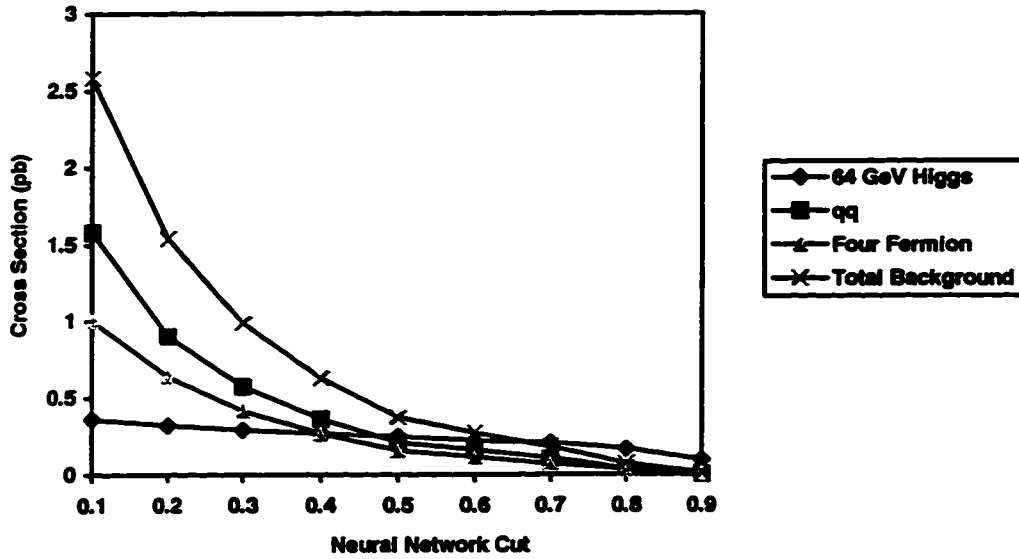
Figure 52: The remaining cross sections of HZ signal for a 64 GeV/c^2 Higgs, qq(γ) background, four fermion final state background, and total background after the sequential application of the precut, various values of thirteen input ANN output cut, and the postcut.
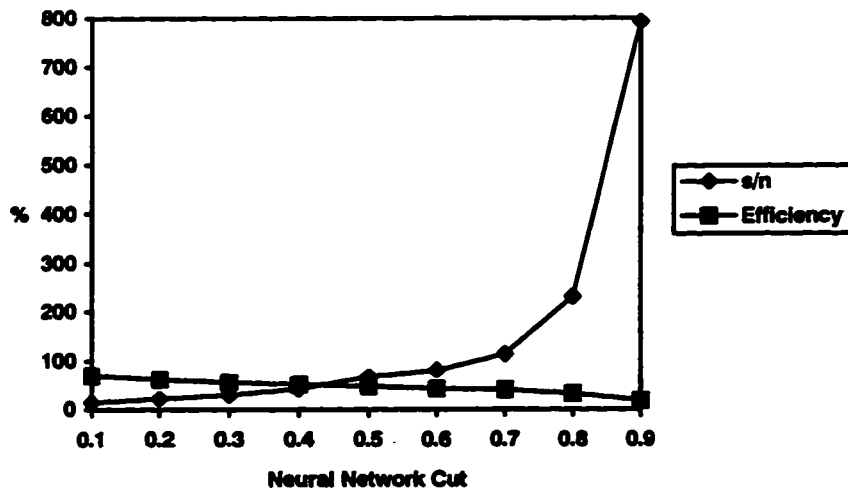
Figure 53: Selection efficiency for 64 GeV/c^2 Higgs events and signal to noise values after the sequential application of the precut, various values of the thirteen input ANN cut, and the postcut.

In order to keep the efficiency of the analysis at about the 25% level, a final cut of 0.85 on the output of the thirteen input ANN was decided upon. The rest of the validation sets were then processed using these final cuts:

1. The precut described in section 5.3.

2. A cut on the output of the thirteen input ANN at 0.85.

3. A postcut requirement that the probability of the most probable Z mass constrained fit to the event be greater than or equal to 1%.

Table 13 shows the number of remaining events, efficiencies, and remaining cross sections for the seven different Higgs masses when these cuts were applied. The efficiency (for all but the 55 GeV/c^2 point) remain around the 25% level as was expected. The efficiency remains fairly constant over the range of Higgs masses greater than 60 GeV/c^2.

Events:							
Higgs Mass	55	60	62	64	66	68	70
No Cut	500	500	500	500	500	500	500
Precut	479	478	489	488	491	492	488
nn > 0.85	92	129	123	135	135	133	129
Postcut	85	123	118	126	130	125	126

Efficiencies:							
Higgs Mass	55	60	62	64	66	68	70
No Cut	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Precut	95.8%	95.6%	97.8%	97.6%	98.2%	98.4%	97.6%
nn > 0.85	18.4%	25.8%	24.6%	27.0%	27.0%	26.6%	25.8%
Postcut	17.0%	24.6%	23.6%	25.2%	26.0%	25.0%	25.2%

Cross Sections in pb:							
Higgs Mass	55	60	62	64	66	68	70
No Cut	0.95	0.73	0.63	0.52	0.40	0.26	0.12
Precut	0.91	0.70	0.61	0.50	0.39	0.25	0.11
nn > 0.85	0.18	0.19	0.15	0.14	0.11	0.07	0.03
Postcut	0.16	0.18	0.15	0.13	0.10	0.06	0.03

Table 13: The number of events remaining, selection efficiency, and cross section remaining for the signal validation sets after sequential application of the final cuts. All Higgs masses are in GeV/c^2.

The remaining number of background events for the $qq(\gamma)$ and four fermion final state validation sets after application of the cuts as well as the selection efficiencies and remaining cross sections for these events is shown in Table 14. Finally, the remaining number of signal and background events for 10 pb^{-1} of integrated luminosity (roughly the amount of data that was recorded by OPAL during the 161 GeV LEP run) and the resulting signal to noise ratio for each Higgs mass is shown in Table 15. The decrease in efficiency at lower Higgs masses may be explained by the fact that the ANN was trained to recognize the jet profiles of m_H=65 GeV/c^2 template signal events. Thus, the reduced magnitude of the mass of the lowest mass jet pair and the reduced energies of the Higgs jets in lower mass Higgs events would appear to the ANN as not meeting the requirements for selection.

With the calculation of these final efficiencies and background rates, the analysis was ready for application to real OPAL data. Prior to considering the OPAL data, however, an estimation of the effects of systematic errors on the results of the neural network analysis is made in the next section.

Events:		
Channel	**qq(γ)**	**Excalibur**
No Cut	50000	20000
Precut	9150	4055
nn > 0.85	5	28
Postcut	5	27

Efficiencies:		
Channel	**qq(γ)**	**Excalibur**
No Cut	100.0%	100.0%
Precut	18.30%	20.28%
nn > 0.85	0.0100%	0.1400%
Postcut	0.0100%	0.1350%

Cross Sections in pb:		
Channel	**qq(γ)**	**Excalibur**
No Cut	147.4	17.132
Precut	26.974	3.474
nn > 0.85	0.0147	0.0240
Postcut	0.0147	0.0231

Table 14: The remaining number of events, selection efficiencies, and remaining cross sections for the qq(γ) and four fermion final state (Excalibur) background validation sets after the application of the final cuts.

Higgs Mass (GeV/c²)	Efficiency	Signal	qq(γ)	Excalibur	Total BG	s/n
55	17.0%	1.62	0.15	0.23	0.38	4.28
60	24.6%	1.79	0.15	0.23	0.38	4.73
62	23.6%	1.48	0.15	0.23	0.38	3.91
64	25.2%	1.30	0.15	0.23	0.38	3.44
66	26.0%	1.03	0.15	0.23	0.38	2.73
68	25.0%	0.65	0.15	0.23	0.38	1.71
70	25.2%	0.29	0.15	0.23	0.38	0.78

Table 15: The final selection efficiencies, number of remaining signal and background events, and signal to noise ratios for each of the Higgs masses for 10 pb^{-1} of integrated luminosity.

5.9 Estimation of Systematic Errors

An estimation of the systematic errors inherent in the ANN analysis is presented in this section. Evaluating the systematic behavior of neural network systems is a difficult task as the non-linear relationships between input variables in the network precludes the possibility of varying inputs in an independent way.

With these limitations in mind, the estimation of systematic errors presented in this chapter serves only as a first look at the effects of imperfect modelling of the input variables upon the output of the network. A list of items to be considered in order to give a more complete estimation of the systematic errors involved in this analysis is discussed in chapter 6.

The thirteen input variables to the ANN considered in the last section were compared in the background Monte Carlo and in the data. The background Monte Carlo was weighted between the qq(γ) events and the four fermion final state events in the expected ratios after the application of the precut. The mean of each input variable in the data was then compared to the mean in the background Monte Carlo and the deviation necessary for the Monte Carlo mean to match the data mean was calculated.

The input variables were divided into three groups: the b quark tagging variables, the event shape variables, and the constrained fitting variables. Table 16 shows the mean and rms values of the b quark tagging values in both the background Monte Carlo and in the OPAL data. The deviation between the Monte Carlo and data means is also shown. Similar results for the shape and constrained fitting variables are shown in Table 17 and Table 18 respectively. The definitions of the short variable names may be found in section 5.4.

Variable	Mean MC	Mean Data	Deviation	rms MC	rms Data
tdsum	1.118	1.287	+0.169	4.47	6.03
busig	5.541	5.000	-0.541	11.74	18.58

Table 16: Deviation between the mean values of the b quark tagging input variables in the data and the background Monte Carlo (MC). The rms values in the data and MC are also shown.

Variable	Mean MC	Mean Data	Deviation	rms MC	rms Data
sprime	135.6	135.0	-0.6	21.73	21.12
minem	3.22	3.42	+0.20	1.60	1.95
mvis	132.0	132.7	+0.7	31.23	31.34
y34	0.00993	0.01157	+0.00164	0.0113	0.0136
angl	0.692	0.735	+0.043	0.318	0.323
spher	0.213	0.228	+0.015	0.165	0.164
ntrack	26.71	27.72	+1.01	7.10	7.25

Table 17: Deviation between the mean values of the event shape input variables in the data and the background Monte Carlo (MC). The rms values in the data and MC are also shown.

Variable	Mean MC	Mean Data	Deviation	rms MC	rms Data
mlow	24.44	26.91	+2.47	8.44	9.31
mdiff	51.61	55.77	+4.16	42.04	40.77
zprob	0.292	0.170	-0.122	0.318	0.266
eqprob	0.256	0.109	-0.147	0.310	0.230

Table 18: Deviation between the mean values of the constrained fitting input variables in the data and the background Monte Carlo (MC). The rms values in the data and MC are also shown.

The mean values for each input variable in a given group were then shifted to match the mean in the data in each of the validation sets for the 64 GeV/c^2 HZ signal, the qq(γ) background and the four fermion final state (Excalibur) background. These shifted validation sets were then passed through the thirteen variable network and the performance appraised after application of the same final cuts as in section 5.8. This process was repeated for each of the three groups of input variables.

Since the change in the number of remaining signal and background events is positively correlated when the input variables are shifted, the change in the remaining number of signal events for a background level equal to that of the nominal analysis and the change in the remaining number of background events for a signal efficiency equal to that of the nominal analysis was recorded for each shift of the input variables. Thus, the signal efficiency after input variable shifting for a background level of 0.38 pb and the remaining background after input shifting for a signal efficiency of 25.2% were recorded for each of the three shifts of input variables.

Table 19 shows the relative percentage error in the signal for a background level of 0.38 pb and the percentage change in the number of surviving background events for a signal efficiency of 25.2% when each of the three types of input variables are shifted to their respective data means individually. The final estimate of systematic error in the signal and efficiency is taken to be the sum in quadrature of these values, thus giving a relative error of 7.7% on the signal efficiency and 32.3% on the background rate. The final efficiencies and background rates for several Higgs masses with statistical (calculated using Poisson statistics) and systematic errors are shown in Table 20 and Table 21 respectively. It can be seen from these two tables that the magnitudes of the statistical and systematic errors in this analysis are very similar.

Variable Type	Δ Signal Efficiency for Background = 0.38 pb	Δ Background for Signal Efficiency = 25.2%
b Quark Tagging	7.1%	2.9%
Event Shape	2.4%	31.0%
Constrained Fitting	1.6%	7.7%

Table 19: The percentage change in the signal efficiency for 64 GeV/c^2 HZ events at a background level of 0.38 pb and the percentage change in background for a signal efficiency of 25.2% when the input variables for each of three classes are shifted to their data means.

Higgs Mass (GeV/c^2)	Efficiency	Signal	qq(γ)	Excalibur	Total BG	s/n
55	17.0±1.3%	1.62±0.12	0.15	0.23	0.38±0.12	4.28±1.67
60	24.6±1.9%	1.79±0.14	0.15	0.23	0.38±0.12	4.73±1.90
62	23.6±1.8%	1.48±0.11	0.15	0.23	0.38±0.12	3.91±1.54
64	25.2±1.9%	1.30±0.10	0.15	0.23	0.38±0.12	3.44±1.35
66	26.0±2.0%	1.03±0.08	0.15	0.23	0.38±0.12	2.73±1.07
68	25.0±1.9%	0.65±0.05	0.15	0.23	0.38±0.12	1.71±0.67
70	25.2±1.9%	0.29±0.02	0.15	0.23	0.38±0.12	0.78±0.31

Table 20: Systematic errors on the final number of signal and background events as well as on the signal to noise ratio for various Higgs masses.

Higgs Mass (GeV/c²)	Efficiency	Signal	qq(γ)	Excalibur	Total BG	s/n
55	17.0±1.8%	1.62±0.17	0.15±0.07	0.23±0.04	0.38±0.11	4.28±1.69
60	24.6±2.2%	1.79±0.16	0.15±0.07	0.23±0.04	0.38±0.11	4.73±1.79
62	23.6±2.2%	1.48±0.14	0.15±0.07	0.23±0.04	0.38±0.11	3.91±1.50
64	25.2±2.2%	1.30±0.11	0.15±0.07	0.23±0.04	0.38±0.11	3.44±1.29
66	26.0±2.3%	1.03±0.09	0.15±0.07	0.23±0.04	0.38±0.11	2.73±1.03
68	25.0±2.2%	0.65±0.06	0.15±0.07	0.23±0.04	0.38±0.11	1.71±0.65
70	25.2±2.2%	0.29±0.03	0.15±0.07	0.23±0.04	0.38±0.11	0.78±0.31

Table 21: Statistical errors calculated using Poisson statistics for the signal efficiencies and background rates for several Higgs masses. The statistical error in the signal to noise ratios is also shown.

As a cross check of the systematic errors involved in this analysis, the 64 GeV/c² HZ signal validation set as well as the qq(γ) and four fermion final state validation sets were subjected to *track smearing* before being passed through the offline analysis. Track smearing involves altering the track parameters in the Monte Carlo by some resolution factor in order to simulate inaccuracies in the reconstruction from the detector. The track parameters for each of the three validation Monte Carlo data sets mentioned above were smeared by the maximum amount suggested for systematic studies in OPAL [42] and the resulting data was passed through the offline analysis and the trained thirteen input ANN. This resulted in a relative change of 3.2% in the number of HZ signal events accepted and a change of 15.5% in the number of accepted background events. Since the track smearing method reconstructs shape variables well when compared to the data but does not accurately account for errors in the b quark tagging variables, the systematic errors obtained from the method of varying the input variables to the network by their difference between data and Monte Carlo was retained as the most conservative estimate of systematic errors for this analysis.

5.10 Application to OPAL Data

The ANN analysis method presented in the previous sections was applied to the 10 pb^{-1} of real data recorded at $\sqrt{s}$ = 161 GeV in the summer of 1996. During this period, 1500 events were selected by the Tokyo Multihadron flag described in section 3.5.2. Table 22 shows the comparison of the number of data events passing each of the three final cuts compared to the number of expected events from the qq(γ) and four fermion final state backgrounds. The final cut sequence consisted of the following:

1. The precut described in section 5.3.

2. A cut on the thirteen input ANN output at 0.85.

3. A postcut requiring that the probability of the most probable Z mass constrained fit to the event be greater than 1%.

The slight excess of data events after the precut is due to the presence of two photon events in the real data which are not accounted for in the simulation. No candidate events were observed in the data. The ANN output for the 366 data events which passed the precut may be seen in Figure 54. Although no events are observed, a limit cannot be calculated for the low amount of integrated luminosity collected during this running period. In order to quote a limit at the 95% confidence level, Poisson statistics require that an experiment be sensitive to 3 events while observing none [43]. With the limited data available (see Table 15), such a sensitivity is not attained in this case.

Cut	Data	Expected Background
TKMH	1500	-
Precut	366	317
ANN > 0.85	0	0.387
Postcut	0	0.378

Table 22: Comparison of the number of observed and expected events in 10 pb^{-1} of OPAL data after the cuts used in this analysis.
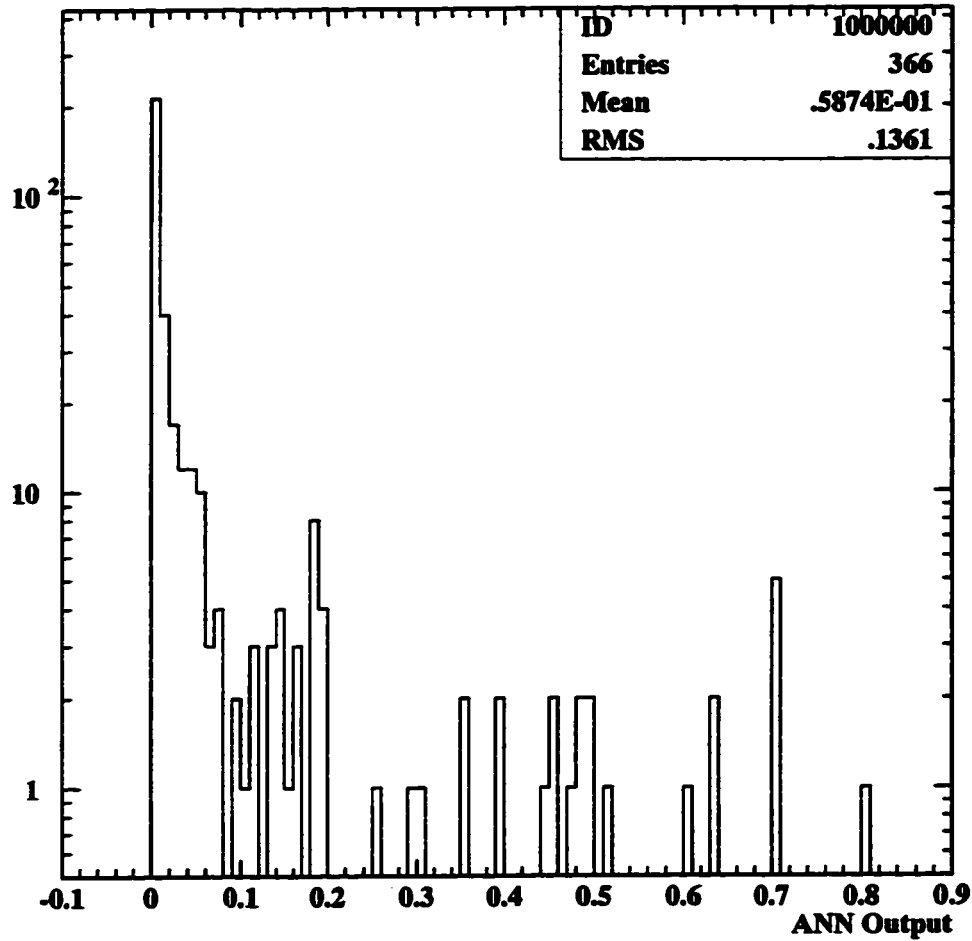
Figure 54: The thirteen input ANN output for the 366 events of real OPAL data which pass the precut. No event exceeds the ANN output cut at 0.85.

5.11 Comparison to Other Analysis Methods

In this section, a comparison of the ANN method to two other methods used to search for the Standard Model Higgs in the four jet channel at OPAL is performed. The first alternative method is the standard multivariate cutting approach used by the OPAL Higgs group [44] while the second is a maximum likelihood approach [45].

The standard cutting approach makes use of a series of kinematic cuts followed by strong cuts on the decay length significance of the vertices associated with the Higgs jets as identified by the most probable Z mass constrained fit. The efficiencies and background rates expected for 10 pb^{-1} of data as well as the resulting signal to noise ratios are shown in Table 23 and may be compared to the corresponding results for the ANN in Table 24. Both tables show only statistical errors. The systematic errors on the cutting approach are estimated at a relative change in signal efficiency of 5.5% and a relative background systematic error of 24%. This values compare favourably with the systematic errors calculated for the ANN approach in section 5.9. The cutting analysis was optimized for a 55 GeV/c^2 Higgs mass and this explains its better performance than the ANN for lower Higgs masses. At higher masses, however, the ANN approach has both a higher efficiency and a lower background rate.

Higgs Mass (GeV/c^2)	Efficiency	Signal	qq(γ)	Excalibur	Total BG	s/n
55	21.3±0.8%	2.03±0.08	0.57	0.23	0.8±0.08	2.54±0.35
60	22.4±0.8%	1.63±0.06	0.57	0.23	0.8±0.08	2.04±0.28
62	22.8±1.9%	1.43±0.12	0.57	0.23	0.8±0.08	1.79±0.33
64	23.6±1.9%	1.22±0.10	0.57	0.23	0.8±0.08	1.53±0.28
66	20.6±1.8%	0.82±0.07	0.57	0.23	0.8±0.08	1.03±0.20
68	16.2±1.6%	0.42±0.04	0.57	0.23	0.8±0.08	0.53±0.10
70	15.4±1.6%	0.18±0.02	0.57	0.23	0.8±0.08	0.23±0.05

Table 23: Signal selection efficiencies, background rates, and signal to noise ratios for various Higgs masses using a standard cutting approach. All numbers of events are quoted for 10pb^{-1} and all quoted errors are from Monte Carlo statistics only.

Higgs Mass (GeV/c^2)	Efficiency	Signal	qq(γ)	Excalibur	Total BG	s/n
55	17.0±1.8%	1.62±0.17	0.15±0.07	0.23±0.04	0.38±0.11	4.28±1.69
60	24.6±2.2%	1.79±0.16	0.15±0.07	0.23±0.04	0.38±0.11	4.73±1.79
62	23.6±2.2%	1.48±0.14	0.15±0.07	0.23±0.04	0.38±0.11	3.91±1.50
64	25.2±2.2%	1.30±0.11	0.15±0.07	0.23±0.04	0.38±0.11	3.44±1.29
66	26.0±2.3%	1.03±0.09	0.15±0.07	0.23±0.04	0.38±0.11	2.73±1.03
68	25.0±2.2%	0.65±0.06	0.15±0.07	0.23±0.04	0.38±0.11	1.71±0.65
70	25.2±2.2%	0.29±0.03	0.15±0.07	0.23±0.04	0.38±0.11	0.78±0.31

Table 24: Signal selection efficiencies, background rates, and signal to noise ratios for various Higgs masses using the ANN approach developed in this thesis. All numbers of events are quoted for 10 pb^{-1} and quoted errors are from Monte Carlo statistics only.

The maximum likelihood approach uses a much harder precut than the ANN analysis. The efficiency of the maximum likelihood precut for a 64 GeV/c^2 Higgs is 71% while the ANN precut gives an efficiency of 98%. The likelihood approach is also optimized for a 55 GeV/c2 Higgs mass and hence it obtains better results than the ANN at low Higgs masses. The signal efficiencies, background levels, and signal to noise ratios for the likelihood method are shown in Table 25 and indicate that the ANN has a higher sensitivity at higher Higgs masses.

Higgs Mass (GeV/c^2)	Signal	qq(γ)	Excalibur	Total BG	s/n
55	3.15±0.11	0.82±0.09	0.37±0.06	1.19±0.15	2.65±0.43
64	1.55±0.13	0.82±0.09	0.37±0.06	1.19±0.15	1.30±0.27
66	1.03±0.09	0.82±0.09	0.37±0.06	1.19±0.15	0.87±0.19

Table 25: Signal selection rates, background rates, and signal to noise ratios for various Higgs masses using the maximum likelihood approach. All numbers of events are quoted for 10 pb^{-1} and quoted errors are from Monte Carlo statistics only.

6. Conclusions

A search for the Standard Model Higgs Boson in $\sqrt{s} = 161$ GeV electron positron collisions was performed using data from the OPAL detector at LEP. The Artificial Neural Network technique employed resulted in better sensitivity and signal to noise ratios for the Higgs boson signal as compared to both a standard cutting approach and a maximum likelihood method.

In preparation for application of the ANN analysis to higher energy LEP runs, several possible improvements to the training process have been identified. Instead of using a W^+W^- Monte Carlo data set to approximate the four fermion final state background during the ANN training, sufficient Excalibur generator events will be produced to allow both training and validation of the network to take place on Excalibur data. In addition, several different Higgs masses will be used in the ANN training data in order to lessen the mass dependence of the selection efficiency for the network.

The systematic errors inherent in the ANN method were estimated and shown to be on a level comparable to conventional analyses. In future analyses, systematic errors stemming from uncertainties in the Monte Carlo model such as the hadronic fragmentation functions and the value of α_s will be evaluated for the network. The correlations between the network input variables will also be investigated in more detail in terms of their effects on the systematic error of the network.

LEP plans to deliver about 100 pb^{-1} of data to the OPAL detector in the $\sqrt{s} = 184$ GeV run which will take place in the summer of 1997. The ANN analysis of the four jet Standard Model Higgs signature presented in this thesis will be re-optimized for this run with the improvements suggested above. The sensitivity of the ANN analysis when

combined with the much improved integrated luminosity of $100\,\text{pb}^{-1}$ will allow a meaningful limit to be set on the Higgs boson mass during this run.

Bibliography

[1] S.L. Glashow, Nucl. Phys. **22** (1961) 579.
   S. Weinberg, Phys. Rev. Lett. **19** (1967) 1264.
   A. Salam, *Elementary Particle Theory*, N. Svartholm ed., Almquist and Wiksell, Stockholm (1968).

[2] H. Fritsch, M. Gell-Mann, and H. Leutwyler, Phys. Lett. **B47** (1973) 365.

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

[4] The OPAL Collaboration, OPAL Physics Note PN234, July 1996.

[5] Gordon Kane, Modern Elementary Particle Physics, Addison-Wesley (1993).

[6] A. Djouadi, M. Spira, and P.M. Zerwas, DESY 95-210, Z. Phys. C in print.

[7] E. Gross, B.A. Kniehl, and G. Wolf., Z. Phys. **C63** (1994) 417.

[8] OPAL Collaboration, K. Ahmet et. al., Nucl. Instr. and Meth., **A305** (1991) 275.

[9] P.P. Allport et. al., Nucl. Instr. and Meth., **A324** (1993) 34.

[10] OPAL Collaboration, P.D. Acton et. al., Z. Phys. **C59** (1993) 183;
    OPAL Collaboration, P.D. Acton et. al., Phys. Lett. **B307** (1993) 247;
    OPAL Collaboration, P.D. Acton et. al., Phys. Lett. **B312** (1993) 501.

[11] P.P. Allport et. al., Nucl. Instr. and Meth., **A346** (1994) 476.

[12] J.R. Carter et. al., Nucl. Instr. and Meth., **A286** (1990) 99.

[13] O. Biebel et. al., Nucl. Instr. and Meth., **A323** (1992) 169.

[14] M. Hauschild et. al., Nucl Instr. and Meth., **A314** (1992) 74.

[15] H.A. Bethe, *Handbuch der Physik 24/1*, J. Springer Verlag Berlin (1933) 491.

[16] H. Mes et. al., Nucl. Instr. and Meth., **A265** (1988) 445.

[17] J.D. Hobbs et. al., Nucl. Instr. and Meth., **A325** (1993) 494.

[18] S. Dado et. al., Nucl. Instr. and Meth., **A252** (1986) 511.

[19] J. Allison et. al., Nucl. Instr. and Meth. **A236** (1985) 284.

[20] OPAL Collaboration, G. Alexander et. al., Z. Phys. **C52** (1991) 175.

[21] B.E. Anderson et. al., IEEE Trans. On Nuclear Science, **41** (1994) 845.

[22] M. Arignon et. al., Nucl. Instr. and Meth., **A313** (1992) 103.

[23] J. Allison et. al., Nucl. Instr. and Meth., A317 (1992) 47.

[24] GEANT Detector Description and Simulation Tool, CERN Program Library Long Writeup W5013 (1994).

[25] OPAL Technical Note TN322 (1995).

[26] Leif Lönnblad, Carsten Peterson, and Thorsteinn Rögnvaldsson, *JETNET 3.0 - A Versatile Neural Network Package*, CERN-TH.7135/94, December 1993.

[27] S.R. Amendolia, *Neural Networks*, CERN 94-06 (1994).

[28] H. Kolanoski, *Application of Artificial Neural Networks in Particle Physics*, DESY 95-061 (1995).

[29] HZHA Generator: P. Janot, in *Physics at LEP2*, edited by G. Altarelli, T. Sjöstrand and F. Zwirner, CERN 96-01, vol. 2 (1996), p. 309.

[30] PYTHIA 5.721 and JETSET 7.408 Generators: T. Sjöstrand, Comp. Phys. Comm. **82** (1994) 74.

[31] T. Sjöstrand, LU TP 95-20 (1995).

[32] F.A. Berends, R. Kleiss, and R. Pittau, *EXCALIBUR - a Monte Carlo Program to Evaluate all Four Fermion Processes at LEP 200 and Beyond*, INLO-PUB-12/94, July 1994.

[33] S. Yamashita, DH Manual, OPAL Higgs Group Internal Note.

[34] S. Catani, Yu. L. Dokshitzer, M. Olsson, G. Turnock, and B.R. Webber, Phys. Lett. **B269** (1991) 432.

[35] Tsunehiko Omori, Shoji Asai, Isamu Nakamura, and Satoru Yamashita, *A Matching Algorithm : MT Package*, OPAL Technical Note TN-381, July 1996.

[36] S. Brandt, Ch. Peyrou, R. Sosnowski, and A. Wroblewski, Phys. Lett. **12** (1964) 57.

[37] J.D. Bjorken and S.J. Brodsky, Phys. Rev. **D1** (1970) 1416.

[38] R. Kowalewski, *Study of Lifetime Tagging of b Hadrons in OPAL*, OPAL Technical Note TN181, August 1993.
     R. Hawkings, *Vertex Taggingof b-Hadrons using 3D Tracking and Neural Networks*, OPAL Technical Note TN366, April 1996.

[39] K. Hultqvist, R. Jacobsson, and K.E. Johansson, Nucl. Inst. and Meth. A364 (1995) 193.

[40] Th, Lindblad, C.S. Linsey, F. Block, and A. Jayakumar, Nucl. Instr. and Meth. A356 (1995) 498.

[41] R.D. Field et. al., Phys. Rev. **D53** (1996) 2296.

[42] B. King, *A New Routine for Global Track Smearing of OPAL Monte Carlo Events*, OPAL Technical Note TN295, April 1995.

[43] Particle Data Group, Phys. Rev. **D50** (1994) 1281.

[44] S. Yamashita, *Study on HZ Production with 4 Jet Final State at LEP 161 GeV*, OPAL Higgs Group Internal Note, September 1996.

[45] K. Desch, M. Schumacher, and E.V. Toerne, *A Likelihood Selection for SM Higgs Search in the 4 Jet Channel*, OPAL Technical Note TN440, November 1996.

\

A. The OPAL Collaboration

University of Birmingham

Ashby S, Bell A.N, Bloodworth I., Bloomer J., Charlton D., Hillier S., Homer R.J., Jimack M.P., Jovanovic P., McMahon T., O'Neale S., Rigby D., Stokes W.N., Talbot S.D., Watkins P., Watson A.T., Watson N., Wilson J.

University of Bologna

Arcelli S., Bonacorsi D., Brasolin F., Brigliadori L., Calligola P., Capiluppi P., Ciocca C., Cuffiani M., Dallavalle M., Fabbri F., Fanfani A., Fanti M., Ferrari M., Fierro M., Giacomelli G., Giacomelli R., Grandi C., Marcellini S., Michelini A., Montanari A., Odorici F., Poli B., Rossi A.M., Siroli G.P.

Universitat Bonn

Bartoldus R., Baumann S., Desch K., Dieckmann A., Eyring A., Feld L., Fischer H.M., Geich-Gimbel Ch., Gross S., Hartmann Ch., Kocian M., Kokott T., Kunst B., Mader W., Maettig P., Markus C., Menke S., Meuser S., Mueller U., Nellen B., Posthaus A., Scharf F., Schmitz R.E., Schumacher M., Sieberg R.P.B., Stockhausen B., von Toerne E., Voss H., Wermes N.

Cambridge University

Altekamp N.D.R., Batley J.R., Biguzzi A., Carter J.R., Gibson V., Goodrick M.J., Hawkes C., Hill J.C., Hoare J., Hutchcroft D.E., Jones C.R., Rison M., Shaw R., Ward C.P., Ward D.R., Wotton S.

Carleton University

Beaulne P., Carnegie R.K., Donkers M., Estabrooks P.G., Hemingway R.J., Jones M., Karlen D., Koetke D., Krieger P., Towers S.

CERN

Ackerstaff K., Barberio E., Barillari T., Baudot J., Behnke T., Bentvelsen S., Burckhart H., Burgard C., Fleck I., Furtjes A., Gallno P., Gasparini O., Glessing W., Gruwe M.,

Hammarstrom R., Hansroul M., Hauschild M., Heuer R.D., Hildreth M., Hill J., Jimenez M., Junk T., Kluth S., Lorenzi R., Maettig P., Mannelli M., McPherson R.A., Meijers F., Neal H.A., Nisius R., Perez-Ochoa R., Plane D.E., Renoux A., Ros E., Runolfsson O., Scharff-Hansen P., Schleper P., Schmitt B., Schoning A., Schroeder M., Schwick C., Shepherd-Themistocleous C., Smith A., Teuscher R., Thomson M.A., Watson M., Wells P., Zankel K., Zer-Zion D., Znoy M.

University of Chicgao

Anderson K.J., Armstrong R.L., Coxe R., Evans H., Glenzinski D., Hocker A., Merritt F.S., Northrop R., Oreglia M.J., Pilcher J.E., Pod E., Sanders H., Turcot A.S.

Hamburg/DESY

Blobel V., Fiedler F., Hawkings R., Huentemeyer P., Meyer J., Oh A., Petzold S., Sittler A., Steuerer J., Wagner A., Wilson G.W.

Universitat Freiburg

Berlich P, Bobinski M, Buergin R., Degenhart C., Geppert R., Herten G., Hilse T., Joos D., Kobel M., Kolpin R., Kolrep M., Lauxtermann S., Lippert A., Ludwig J., Messmer I., Mohr W., Ohse B., Patt J.,Pfister J., Roeser H., Rogalla M., Rombach H., Roy B., Runge K., Sachs K., Schmid T., Schmidt W.D., Schultz-Coulon H., Schweizer B., Soeldner-Rembold S., Stoll K., Thiergen M., Van Der Bij J., Verzocchi M., Webel M., Weber H.Ch., Weigel S., Wilkens B.

Universitat Heidelberg

Bock P., Christen K., Fath P., Feuerstack M., Igo-Kemenes P., Jost U., Kim E., Kress T., Liebisch R., Matsumura H., Schieck J., Schleper P., Schmitt S., Soergel V., Steiert M., Utzat P., von Dobschuetz S., von Krogh J., von Walter P.

IndianaUniversity

Anderson S.H., Braibant S., Hanson G.G., Letts J., De Jong S.J., Ogren H.O., Rust D.R., Sulanke T., Van Kooten R.   .

Manchester University

Allison J., Barlow R.J., Beeston C., Bird S.D., Cooke O.C., Duerdoth I.P., Eatough D., Edwards J.E., Futyan D.I., Hughes-Jones R.E., Ingram M.R., Jones G.J., Lafferty G.D., Loebinger F.K., Marshall R., Needham M., Nijjhar B., Oldershaw N.J., Pater J., Phillips P.D., Shears T., Stephens K., Vokurka E., Wyatt T.

University of Maryland

Ball A., Bard R., Chang C.Y., Dallapiccola C., Fong D., Foucher M., Gascon-Shotkin S., Herndon M., Jawahery H., Kellogg R., Martinez G. R., Skuja A., Snow G.A.

University of Montreal

Azuelos G., Beaudoin G., Beaulieu M., Berichon J., Blais-Ouellette S., Capdevielle N., Courville J.F., Doucet M., Gascon J., Jeremie H., Joly A., Lefebvre E., Le Page F., Lewis M., Martin J.P., Tafirout R., Taras P., Theoret C., Trigger I., Vikas P., Zacek V.

Universitat Muenchen

Dubbert J., Duckeck G., Boutemeur M., Schaile D., Schaile O., Trefzger T., Vollmann U.

CRPP CANADA

Dixit M.S., Gagnon P., Gravelle P., Hargrove C.K., Losty M., Mes H., Miels K., Oakham F.G., O'Neill M., Watson N.K.

Queen Mary and Westfield College

Beck G.A., Carter A.A., Evans M.D.T., Gibson W.R., Hapke M., Kyberd P., Lloyd S.L., Markopoulos K.D., Martin A., McNab A.I., Morris J., Newman-Coburn D., Pritchard T.W., Robins S.A.

University of California Riverside

Chrisman D., del Pozo L.A., Gary J.W., Giacomelli P., Gorn W., Harin-Dirac M. Layter J.G., Minor C.P.D., Shen B.C.

Rutherford Appleton Laboratory

Bell K.W., Brown R.M., Brumfitt R., Connolly J.F., Denton L.G., Geddes N.I., Geralis T., Jacob F.R., Jaroslawski S., Jeffreys P.W., Jeffs M., Kennedy B.W., Patrick G.N., Rippington D., Saunders B.J., Scott W.G., Sharp P.H., Sproston M., Stephenson R., Szymanski P., White D.J.

Saclay

Gaidot A., Gentit F., Lafoux H., Le Du P., Malik A., Pansart J-P., Simon P, Vasseur G.

Israel Institute of Technology (Technion)

Dado S., Goldberg J., Grenitzki Y., Lazic D., Lupu N., Rozen Y., Tarem S.

Tel Aviv University

Alexander G., Bella G., Cohen I., Grunhaus J., Sarkisyan-Grinbaum E., Tsur E.

University of Tokyo

Asai S., Fukui H., Ishii K., Kawagoe K., Kawamoto T., Kobayashi T., Komamiya S., Mashimo T., Matsumoto S., Mori T., Morii M., Nakamura I., Nozaki M., Omori T., Orito S., Tanaka S., Tsukamoto T., Yamashita S.

University College London

Anderson B., Attree D.J., Barraclough G., Betts S. T., Charalambous A., Clarke P., Clay E., Conboy J., Cranfield R., Cresswell M., Crone G., Dumper J., Fraser T.J., Hayes D., Hoare T., Lauber J., Mckigney E., Miller D.J., Rooke A., Sherwood P., Skillman A.J.

Brunel University

Hobson P., Imrie D.C., Rylko R., Sang M.W., Selby C.

University of British Columbia

Axen D., Howard R., McKenna J., Sbarra C.

University of Victoria

Astbury A., Balasubramanian S., Beer G., Deatrich D.C., Dodd H., Graham K., Honma A., Keeler R., Lefebvre M., Lawson I.T., Long G., Mason G., Pitman D., Poffenberger P., Robertson L., Robertson S., Roney J.M., Rosvick M., Sobie R., White J.S.

Weizmann Institute of Science

Duchovni E., Folman R., Gross E., Lellouch D., Levinson L., Mikenberg G., Mir R., Nagai K., Shoa M., Wolf G., Yaari R., Yekutieli G.

University of Alberta

Caron B.L., Davis R.A., Gingrich D. , Faust A.A., Hewlett J., Macpherson A., McDonald J., Mullin S., Pinfold J., Rodning N.L., Routenburg P., Schaapman J., Springer W., Wampler L.

University of Oregon

Lai W-P., Strom D.M.

Duke University

Lautenschlager S.R., Lee A.

Aachen University

Bethke S., Bechtluft J., Biebel O., Hilgers R.M., Lanske D., Pfeifenschneider P., Randerath G., Ruppel U., Toennesmann M.

Budapest

Dienes B., Hajdu C., Horvath D., Pasztor G., Toth K.

Debrecen Atomki, Hungary

Palinkas J., Trocsanyi Z.