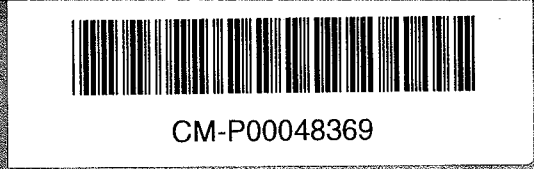


PITHA

94/24

CERN LIBRARIES, GENEVA



Determination of the Cross Section Ratio
 $\sigma(Z^0 \rightarrow b\bar{b})/\sigma(Z^0 \rightarrow \text{hadrons})$
using a Neural Network Classifier
for Event Selection

Martin Röhrner

Thesis-1994-Röhrner

PHYSIKALISCHE INSTITUTE
RWTH AACHEN
Sommerfeldstr.
52056 AACHEN, GERMANY

Determination of the Cross Section Ratio
 $\sigma(Z^0 \rightarrow b\bar{b})/\sigma(Z^0 \rightarrow \text{hadrons})$
using a Neural Network Classifier
for Event Selection

Von der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Physiker Martin Röhner
aus Düsseldorf

Referent : Professor W. Wallraff
Korreferent: Universitätsprofessor K. Lübelmeyer

Tag der mündlichen Prüfung: 1.7.1994

222/184



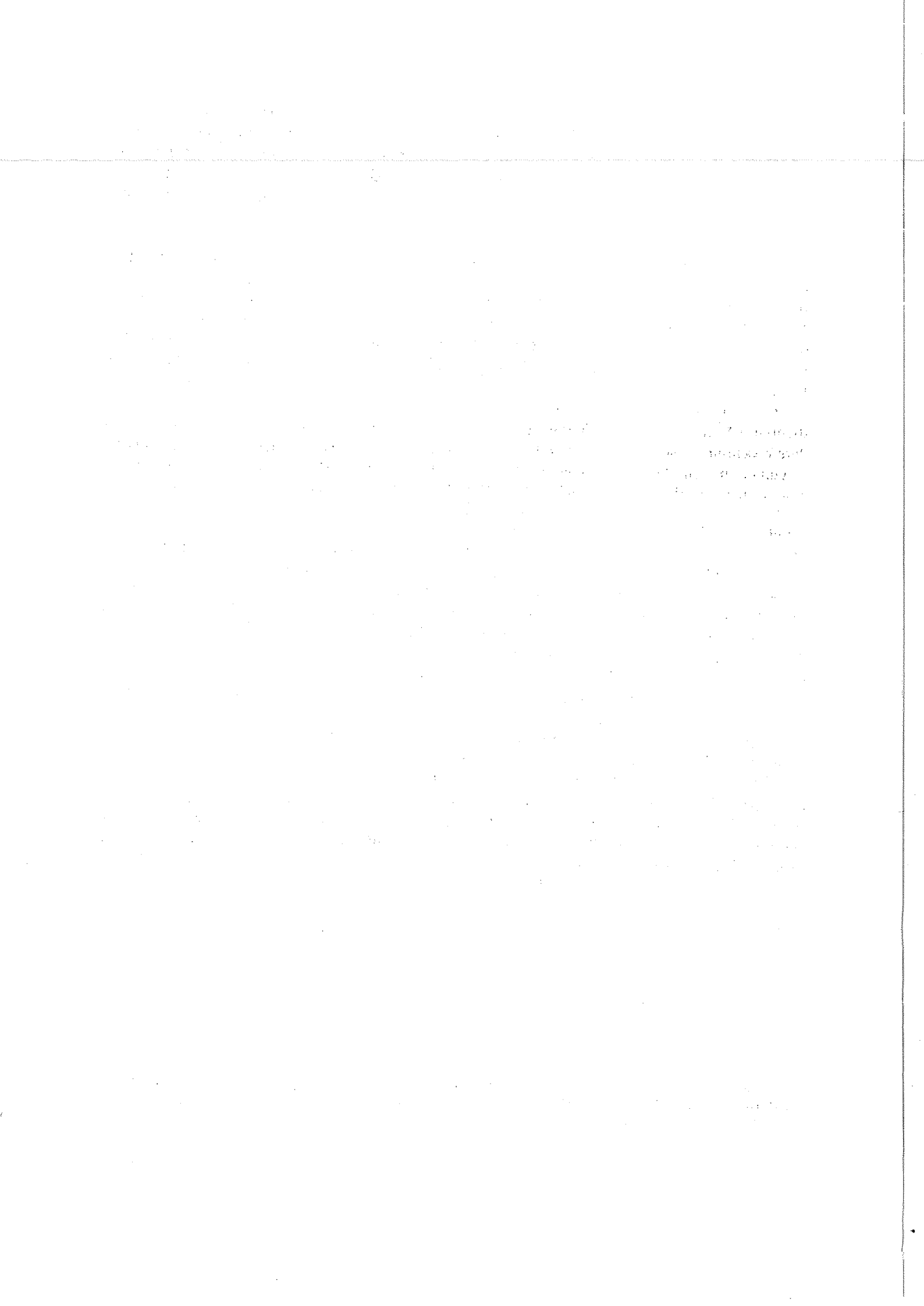
Zusammenfassung: In dieser Arbeit wird das Verhältnis R_{bb} des Wirkungsquerschnitts für den Prozess $e^+e^- \rightarrow Z^0 \rightarrow b\bar{b}$ zum Wirkungsquerschnitt $e^+e^- \rightarrow Z^0 \rightarrow \text{Hadronen}$ für 9 Energien in der Umgebung des Z^0 -Pols bestimmt. Dieses Verhältnis kann zu einer Abschätzung der Masse des noch nicht entdeckten¹ top-Quarks benutzt werden. Die Analyse basiert auf 900000 hadronischen Z^0 -Ereignissen, die in den Jahren 1991 und 1992 mit dem L3 Detektor bei LEP aufgezeichnet wurden. Für die Ereignisse auf dem Z^0 -Pol wurde ein Mittelwert von $R_{bb} = 0.2200 \pm 0.002$ (stat. Fehler) ± 0.009 (sys. Fehler) bestimmt. Dies ergibt eine obere Grenze für die Masse des top-Quarks von $m_{top} < 375 \text{ GeV}$ (95 % CL). Ein modellunabhängiger Fit, der das Verhältnis der Wirkungsquerschnitte für alle 9 Energien berücksichtigte, ergab für die partielle Breite $\Gamma_{bb} = (381 \pm 16) \text{ MeV}$. Unter Einbeziehung der von L3 gemessenen Werte für die Vorwärts-Rückwärts Asymmetrie des b -Quarks konnten die schwachen Kopplungskonstanten für das b in einem modellunabhängigen 2-Parameter Fit bestimmt werden zu $g_{ab} = -0.518 \pm 0.054$ und $g_{vb} = -0.323 \pm 0.083$. Diese Werte bestätigen die Annahme eines schwachen Isospinpartners für das b -Quark.

Zur Selektion der b -Ereignisse wurde ein Neuronales Netz verwendet. Als Eingabegrößen dienten 5 Variablen, die jeweils die Eigenschaften der beiden höchstenergetischen hadronischen Jets beschreiben, sowie eine globale Ereignisvariable. Damit war es möglich einen Datensatz zu gewinnen, in dem der b Anteil gegenüber dem im ursprünglichen Hadrondatensatz um ca. 140 % angereichert wurde bei einer b -Effizienz von 50 %.

Abstract: The topic of this thesis has been the determination of the cross section ratio R_{bb} of the process $e^+e^- \rightarrow Z^0 \rightarrow b\bar{b}$ to $e^+e^- \rightarrow Z^0 \rightarrow \text{hadrons}$ at 9 energies in the vicinity of the Z^0 pole. This ratio can be used in an estimate for the mass of the top quark, which is still undetected¹. The analysis is based on 900000 hadronic events, taken during 1991 and 1992 with the L3 detector at LEP. For the events taken on the Z^0 pole the mean value for the fraction has been measured to $R_{bb} = 0.2200 \pm 0.002$ (stat. error) ± 0.009 (sys. error). This leads in turn to an estimate for the upper limit of the top mass $m_{top} < 375 \text{ GeV}$ (95 % CL). All 9 cross section ratios have been used as input to a model-independent fit yielding the partial width $\Gamma_{bb} = (381 \pm 16) \text{ MeV}$. Using in addition the values for the forward-backward asymmetry of the b quark, which have been published by L3, the weak coupling constants of the b could be determined to $g_{ab} = -0.518 \pm 0.054$ and $g_{vb} = -0.323 \pm 0.083$. These numbers confirm the necessity of a weak isospin partner for the b quark.

A Neural Network has been applied for the selection of the b events. 5 input variables have been used, describing respectively the properties of the two most energetic hadronic jets, and furthermore one global eventshape variable. Using the net as discriminant, it was possible to achieve a datasample with the b contents enriched by 140 % compared to the original hadron sample with a b efficiency of still 50 %.

¹Meanwhile the CDF collaboration has reported strong evidence for the top quark at $174 \pm 10 \pm 13 \text{ GeV}$. (FERMILAB-PUB-94-116-E, May 1994)



Contents

1	Introduction and Overview	1
2	Fermion Pair Production in Electron-Positron Collisions	3
2.1	The process $e^+e^- \rightarrow f\bar{f}$ in lowest order	3
2.2	Higher order corrections	5
3	Neural Networks	10
3.1	Bayes classification	11
3.2	Linear discriminants	12
3.3	Training of a linear discriminant	13
3.4	Non-linear discriminants	15
4	The L3 Detector	19
5	Selection of Hadronic Events	24
6	The Classifier used for the b Selection	28
6.1	Variables used as input to the classifier	28
6.2	Training and performance of the classifier	33
7	Determination of R_{bb}	42
8	Systematic Effects	46
8.1	Effects from the classifier and the fit to the output	46
8.2	Uncertainties of the Monte-Carlo predictions and detector effects	49
8.3	Different fragmentation models	51
8.4	Summary	51
9	Standard Model Parameters	53
9.1	Properties of the Z^0 and the partial width of $Z^0 \rightarrow b\bar{b}$	53
9.2	Weak coupling constants of the b quark	54
9.3	Top mass	57
10	Summary and Conclusion	58
A	The Rummelhart Algorithm for Multi-layered Neural Networks	61
B	Performance of a Multilayered Neural Network	65
C	b Selection with low-level Variables	73

D Further Methods for Quark Flavour Separation	76
E The Program Packages ZFITTER and FUNPLO	79
E.1 ZFITTER	79
E.2 FUNPLO	80

Chapter 1

Introduction and Overview

The *Large Electron-Positron Collider* (LEP) at CERN, Geneva started its operations in August 1989. The main task during its first phase (LEP1) with a center-of-mass energy around 91 GeV has been a precise measurement of the properties of the Z^0 gauge boson. Another topic has been the search for new particles, above all the top quark and the Higgs boson. None of them was found and in the meantime the lower mass limit for the top has been set to 108 GeV [1]. Therefore it has become a matter of interest to give at least an estimate to the upper limit of the top mass.

One of the physical processes, which are well suited for this task, is the decay of the Z^0 boson into a pair of b quarks. Because the top is the weak isospin partner of the b , it couples relatively strong virtually to the $Zb\bar{b}$ vertex, introducing large corrections to the tree-level diagram [2] (Figure 2.4). A particular clean observable in this context is R_{bb} , the fraction of b events in the entire hadron sample. This quantity is sensitive only to vertex effects¹, because the oblique contributions in the Z^0 propagator [3] are the same for all final states and cancel out in the ratio. The Z^0 is a copious source of b , thus the underlying statistics will be high.

The subject of this thesis is to describe the measurement of R_{bb} for the data taken with the L3 detector during the years 1991-1992. The result obtained has been used in order to determine several parameters of the Standard Model.

Another emphasis of this work is a closer analysis of the applied selection method. In order to distinguish the b events from the ones coming from lighter quarks a *Functional-Link Net* has been used. This algorithm constitutes a very fast implementation of a Bayes Classifier. Such classifiers divide a dataset into disjunct classes and try to reduce the number of misclassifications to its optimal value. Another classification algorithm widely used, is the *Feedforward-Backpropagation Net*. The behaviour of such a net has been examined but because of the high computational effort required by this algorithm, it had not been applied in the physics analysis. The results of these studies can be found in appendix B.

The thesis is structured as follows. The next two chapters will give a short review of the theoretical basis needed for the later analysis. The first one will deal with the process of fermion pair production in e^+e^- collisions. The second one will cover Bayes classification and Neural Networks. The experimental apparatus, i.e. the L3 detector, is briefly presented in chapter 4. After these introductory remarks the description of the actual analysis starts with the extraction of a hadronic sample out of the raw data. Out of this sample the b events have been selected with the help of the Functional-Link Net. Several variables, describing the characteristic energy flux pattern of a b event in contrast to events from hadrons made up of lighter quarks,

¹This is true only for one-loop corrections.

were defined and used as input to the net. The definitions of these variables and the performance achieved by the classifier are discussed in chapter 6. The value for R_{bb} is determined in chapter 7 and systematic uncertainties are discussed in the following section. In the end the measured value for R_{bb} is used in order to determine the partial width of the decay $Z^0 \rightarrow b\bar{b}$, the weak coupling constants g_{ab} and g_{vb} , and an upper limit for the top mass. All results are then summarized in the last chapter.

To provide a basis for the understanding of the tests performed in appendix B, a derivation of the Backpropagation algorithm by Rummelhardt can be found in the first appendix. The results of studies, dealing with the training performance of a Feedforward-Backpropagation Net, are presented in the following appendix. Appendix C reports of a further approach to b selection using only the energy and the angular distribution of the clusters deposited in the calorimeter as input to a Neural Net. In order to judge the performance of the Functional-Link Net used in this thesis a review of conventional methods for the b selection is given in appendix D. Appendix E describes briefly the program packages ZFITTER and FUNPLO, which were used to determine the Standard Model parameters.

Throughout the thesis holds $\hbar = c = 1$.

Chapter 2

Fermion Pair Production in Electron-Positron Collisions

2.1 The process $e^+e^- \rightarrow f\bar{f}$ in lowest order

The process $e^+e^- \rightarrow f\bar{f}$ is described by the *Standard Model* (SM) in the frame of a relativistic gauge field theory with a spontaneous broken $SU(2)_L \otimes U(1)_Y$ symmetry [4]. Herein the electroweak interactions between the fermions are mediated by vector bosons corresponding to the generators of the gauge groups, namely the photon, the Z^0 , and the W^\pm .

In general the s-channel exchange of Z^0 and γ is the essential process for electron-positron scattering at center-of-mass energies around the Z^0 mass ($\approx 91 \text{ GeV}$). For Bhabha scattering though the pure QED t-channel exchange of a γ becomes the dominating fraction of the cross section for small scattering angles. The exchange of a Higgs is small and therefore will be neglected in the following.

For unpolarized initial states, summing over all helicities of the final states, and neglecting the electron mass, the differential cross section for $e^+e^- \rightarrow f\bar{f}$ ($f \neq e, \nu_e$) in the lowest order can be written¹ as [5]

$$\frac{d\sigma}{d\cos\vartheta} = \frac{3}{8} N_C^f \sigma_{QED} \beta \left\{ G_1(s)(1 + \cos^2\vartheta) + G_2(s)(1 - \beta^2) \sin^2\vartheta + G_3(s) \beta \cos\vartheta \right\} \quad (2.1)$$

with the QCD colour factor N_C^f , the QED cross section $\sigma_{QED} = 4\pi\alpha^2/3s$, the velocity $\beta = \sqrt{1 - 4m_f^2/s}$, and the scattering angle ϑ as given in figure 2.1. For LEP1-energies ($\approx 91 \text{ GeV}$) the fermion masses can be neglected ($\beta = 1$) except for the b -quark, for which $\beta \approx 0.993$. The expression for the total cross section follows immediately

$$\sigma_{tot} = \int_{-1}^1 \frac{d\sigma}{d\cos\vartheta} d\cos\vartheta = N_C^f \frac{\sigma_{QED}}{2} \beta (2G_1(s) + (1 - \beta^2)G_2(s)) \quad (2.2)$$

The cofactors G_1 , G_2 , and G_3 contain the various couplings via the electrical charges q as well as via the vector and axial-vector coupling constants g_v and g_a . They consist of three terms, the first one describing the photon exchange ($\sim q_e^2 q_f^2$), the second one the Z^0 exchange

¹In order to avoid divergencies during the calculation of the higher-order terms the various elements of the Lagrangian — such as the coupling constants, spinors and propagators — have to be redefined properly, a process called *renormalisation*. The renormalization scheme used for the calculation of (2.1) is the *on-shell* scheme [5]. Herein α_S , α , G_μ , m_Z , m_f , and m_H serve as input parameter to the theory.

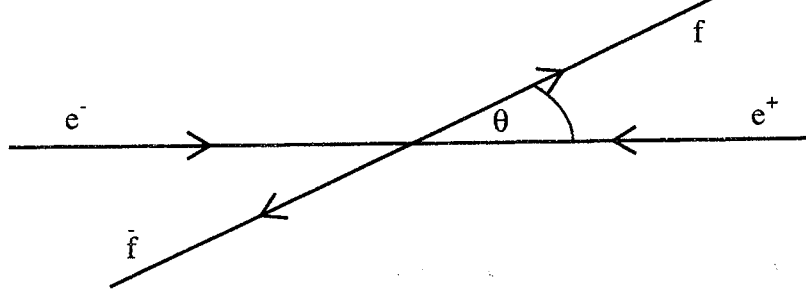


Figure 2.1: The scattering angle ϑ is defined as the angle between the momentum of the incoming electron and the outgoing fermion.

($\sim [g_{ve}^2 + g_{ae}^2][g_{vf}^2 + g_{af}^2]$), and finally an γ - Z interference term ($\sim q_e q_f g_{ve} g_{vf}$).

$$\begin{aligned}
 G_1(s) &= q_e^2 q_f^2 + 2g_{ve} g_{vf} q_e q_f \Re(\chi(s)) + (g_{ve}^2 + g_{ae}^2)(g_{vf}^2 + g_{af}^2 - (1 - \beta^2)g_{af}^2) |\chi(s)|^2 \\
 G_2(s) &= -2g_{ae} g_{af} q_e q_f \Re(\chi(s)) + (g_{ve}^2 + g_{ae}^2) g_{vf}^2 |\chi(s)|^2 \\
 G_3(s) &= -2g_{ae} g_{af} q_e q_f \Re(\chi(s)) + g_{ve} g_{ae} g_{vf} g_{af} |\chi(s)|^2
 \end{aligned} \tag{2.3}$$

Herein $\chi(s)$ is the ratio of the photon and the Z propagator

$$\chi(s) = \frac{1}{4 \sin^2 \Theta_W \cos^2 \Theta_W} \cdot \frac{s}{s - m_Z^2 + i m_Z \Gamma_Z}$$

and the weak coupling constants g_{af} and g_{vf} have been defined as

$$\begin{aligned}
 g_{af} &= I_3^f \\
 g_{vf} &= I_3^f - 2q_f \sin^2 \Theta_W
 \end{aligned} \tag{2.4}$$

where I_3^f is the third component of the weak isospin. In the on-shell scheme the *Weinberg angle* $\sin^2 \Theta_W$ is determined by the boson masses

$$\sin^2 \Theta_W = 1 - \frac{m_W^2}{\rho_0 m_Z^2} \tag{2.5}$$

(2.5) holds for all orders of perturbation theory [6]. $\rho_0 = 1$ in the lowest order of the SM.

For cms-energies \sqrt{s} close to the Z^0 mass, the terms for the photon exchange as well as the interference term can be neglected in (2.3) and (2.2) reduces to

$$\sigma = N_C^f \frac{G_\mu^2}{32\pi} (g_{ve}^2 + g_{ae}^2) \cdot \left(\beta \frac{3 - \beta^2}{2} g_{vf}^2 + \beta^3 g_{af}^2 \right) \cdot \frac{m_Z^4 s}{(s - m_Z^2)^2 + m_Z^2 \Gamma_Z^2}$$

where G_μ is the coupling constant in the Fermi model of four-fermion interactions. G_μ is measured with high precision in μ decays. The last equation defines the partial decay widths Γ_f of the Z^0 boson into a fermion pair $f\bar{f}$

$$\Gamma_f = \frac{G_\mu m_Z^3}{\sqrt{8}\pi} \cdot \left(\beta \frac{3 - \beta^2}{2} g_{vf} + \beta^3 g_{af} \right) \tag{2.6}$$

After radiative corrections are taken into account, the mass m_Z , the total width Γ_Z , and the partial decay widths Γ_f into fermions can be derived from the position, width, and height of the resonance curve, describing the cross section as a function of \sqrt{s} as shown in figure 2.2.

Because of an interference of the axial-vector current ($J^{PC} = 1^{++}$) with the vector current ($J^{PC} = 1^{--}$) the differential cross section varies asymmetrically with ϑ . This behaviour is reflected by the linear cosine term in (2.1) and has been specified as *Forward-Backward Asymmetry*

$$A_{fb} = \frac{\sigma_F - \sigma_B}{\sigma_F + \sigma_B}$$

where σ_F and σ_B are the cross sections in forward and backward hemisphere, respectively. Neglecting again the photon exchange and the interference terms in G_1 , G_2 , and G_3 for the situation on the Z pole, the forward-backward asymmetry can be written as [5]

$$A_{fb}(m_Z) = \frac{3}{4} \frac{2g_{ve}g_{ae}}{g_{ve}^2 + g_{ae}^2} \cdot \frac{2\beta g_{vf}g_{af}}{\frac{(3-\beta^2)}{2}g_{vf}^2 + \beta^2g_{af}^2} \approx \frac{3}{4} \frac{2g_{ve}g_{ae}}{g_{ve}^2 + g_{ae}^2} \cdot \frac{2\beta g_{vf}g_{af}}{g_{vf}^2 + g_{af}^2} (1 + \delta_f)$$

with the mass correction term

$$\delta_f = \frac{4m_f^2}{s} \cdot \frac{-g_{vf}^2 + g_{af}^2/2}{g_{vf}^2 + g_{af}^2}$$

The correction for the b quark turns out to $|\delta_b| \leq 10^{-4}$ as a result of an accidental cancellation of the coupling constants ($g_{vb}^2 \approx g_{ab}^2/2$). Therefore the mass correction for the asymmetry can be neglected even for the b quark. A measurement of the partial width and of the asymmetry can be used for an independent determination of the absolute value of the weak coupling constants.

2.2 Higher order corrections

The formulas given in the last section are not sufficient to describe the experimental results correctly at higher energies and for precise measurements. Diagrams of higher order of the perturbation expansion have to be taken into account [5]. They consist of photon bremsstrahlung in the initial and final states (QED corrections), gluon radiation in the final states (QCD corrections), and electroweak corrections, which themselves again fall into three classes, namely propagator corrections, vertex corrections, and box diagrams.

QED corrections

For the treatment of the photonic corrections the bremsstrahlung is divided into two parts, a soft contribution with the momentum $k < \Delta E$ and a hard one, for which holds $\Delta E < k < \frac{1}{2}\sqrt{s}$. The hard initial bremsstrahlung is described by a radiatorfunction $R(z, s)$ with $z = (p_f + p_{\bar{f}})^2/s$, which gives the probability of the radiation of a photon with the energy $k = E_{beam}(1-z)$. It is one of the advantages of the on-shell renormalization scheme, that the QED correction can be handled simply by convoluting the improved Born cross section σ_0 (see below) with the radiator function. $R(z, s)$ has been calculated up to $\mathcal{O}(\alpha^2)$ [7].

$$\sigma(s) = \int_{4m_f^2/s}^1 \sigma_0(zs)R(z, s)dz \quad (2.7)$$

The impact of initial photon radiation is large on the Z^0 pole, because the initial radiation leads to a shift in the effective cms-energy. Due to this reason, the peak cross section is reduced by

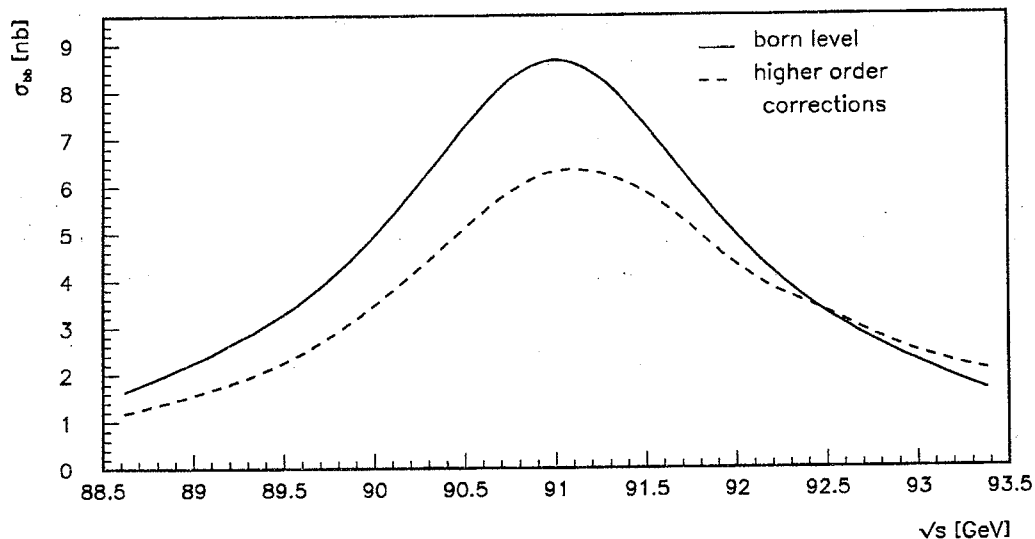


Figure 2.2: The cross section for the process $Z^0 \rightarrow b\bar{b}$ calculated at tree level and including radiative corrections to $\mathcal{O}(\alpha)$ and $\mathcal{O}(\alpha_s^2)$. The corrections have been calculated with the program package ZFITTER 4.5 [75].

about 25% and the maximum is moved by about 122 MeV toward higher energies. Moreover the resonance curve becomes asymmetric (Figure 2.2). The final state radiation exhibits a much smaller effect, which corresponds to an overall correction factor of

$$(1 + \delta_{QED}) = 1 + \frac{3}{4} \frac{\alpha}{\pi} q_f^2 + \mathcal{O}(\alpha^2)$$

It amounts a 0.019% effect for the b -quark. The contribution of the bremsstrahlung interference is also very small and can be neglected.

QCD corrections

Beside the photon bremsstrahlung also gluon radiation is possible in the final state of a hadronic decay of the Z^0 . Because of the large coupling constant of the strong interaction α_s , the QCD contributions are rather large compared to the final state photon bremsstrahlung. Until now the higher-order terms for $m_q = 0$ are calculated up to $\mathcal{O}(\alpha_s^3)$ and up to $\mathcal{O}(\alpha_s^2)$ for $m_q \neq 0$ [8]. As for the final photon bremsstrahlung, the correction can be collected into one multiplicative factor

$$(1 + \delta_{QCD}^{a,v}) = 1 + C_1^{a,v} \left(\frac{\alpha_s}{\pi} \right) + C_2^{a,v} \left(\frac{\alpha_s}{\pi} \right)^2 + C_3^{a,v} \left(\frac{\alpha_s}{\pi} \right)^3 + \dots$$

	C_1	C_2	C_3
Vector	$1 + 3\mu^2$	1.41	-12.8
Axialvector	$1 + 3\mu^2 \ln(4/\mu^2)$	$1.41 \pm f(m_t)$	-12.8

Table 2.1: Due to the breaking of the chiral symmetry, the correction coefficients for the axial-vector and the vector coupling are different. $f(m_t)$ can be found in [9]. $\mu^2 = 4m_f^2/s$.

Weak corrections [10]

Weak corrections arise from non-photonic, i.e., genuine weak higher-order diagrams. There are three groups of weak corrections

1. Propagator (also called self-energy) corrections (Figure 2.3)

Propagator corrections amount for the biggest part of the electroweak corrections. They are the same for all final states. Because a fermion loop decouples from a photon for higher fermion masses, whereas the coupling of a fermion to the Z^0 is proportional to m_f , the largest contribution for cms-energies $\approx m_Z$ results from the top quark.

2. Vertex corrections (Figure 2.4)

The vertex corrections do depend on the fermion species in the final state. But for light fermions up to the mass of the c quark the vertex corrections can be neglected. This is not true for the b , where large corrections arise from top quark effects due to the high top quark mass. The reason for these large corrections in case of the $b\bar{b}$ production lies in the fact, that the top is the weak isospin partner of the b . Therefore, the respective diagrams are not suppressed by off-diagonal elements of the CKM-matrix.

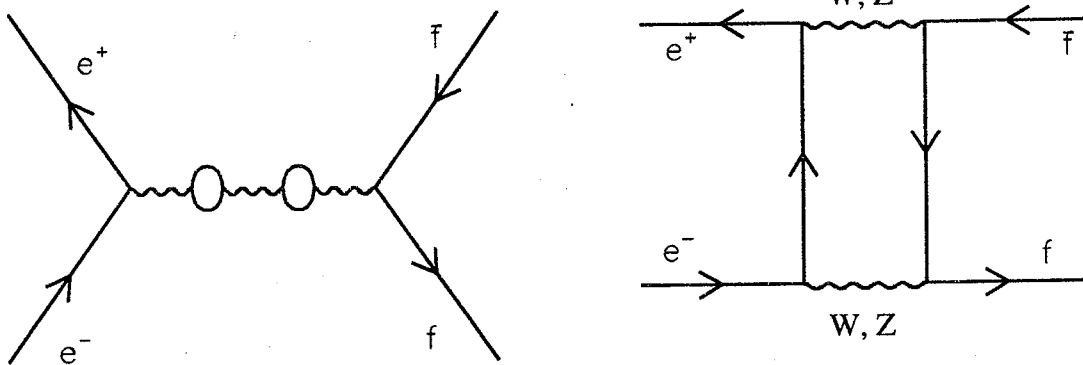


Figure 2.3: Propagator and box corrections

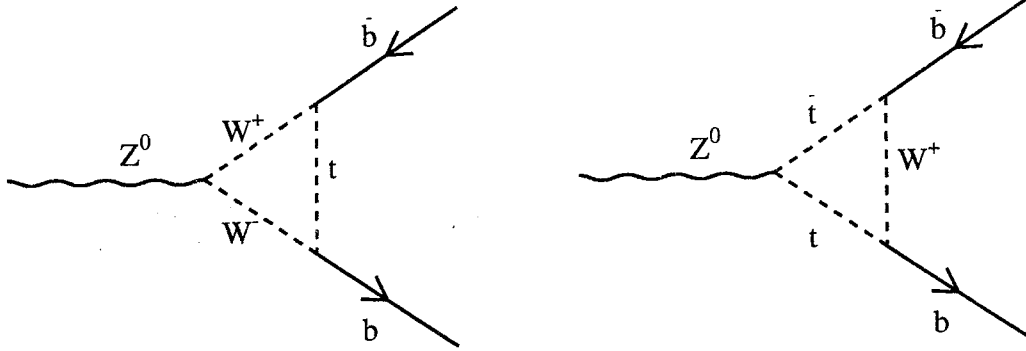


Figure 2.4: Vertex corrections for the process $Z^0 \rightarrow b\bar{b}$ in the Standard Model.

3. Box diagrams (Figure 2.3)

The genuine weak box diagrams are those containing WW and ZZ exchange. Their contributions are non-resonant, hence their effect is small ($\leq 0.02\%$) for energies close to the Z^0 mass.

The weak corrections are treated by an adjustment of the various constants in the lowest-order Born formulae, yielding the so-called *improved Born approximation* [11]. In particular, the corrections to the γ, Z^0 -propagators can be absorbed in an adjustment of the coupling constant α and the propagator χ . The effects on the γ - Z^0 -mixing are handled by an adjustment of the weak mixing angle or of the weak coupling constants g_{af} and g_{vf} respectively. This leads in turn to the so-called *effective mixing angle* or the *effective coupling constants*. Higher order corrections to the self energies are treated by an expansion of the ϱ_0 in equation (2.5) with the $\Delta\varrho$ parameter [3]

$$\varrho_0 \equiv 1 \rightarrow \bar{\varrho} = 1 + \Delta\varrho \quad \text{with} \quad \Delta\varrho = \frac{3\sqrt{2}G_\mu m_t^2}{16\pi^2} + \dots$$

Vertex corrections are incorporated again by corrections to the effective coupling constants. In this framework the coupling constants and the Z^0 width become s dependent [11].

$$\alpha \rightarrow \alpha(s) = \frac{\alpha}{1 - \Delta\alpha(s)} \quad \left(\alpha = \frac{1}{137} \rightarrow \alpha(m_Z) = \frac{1}{129} \right)$$

$$\Gamma_Z \rightarrow \Gamma_Z(s) = \frac{s}{m_Z^2} \Gamma_Z$$

$$\sin^2 \Theta_W \rightarrow \sin^2 \bar{\Theta}_W = \sin^2 \Theta_W + \cos^2 \Theta_W \Delta\varrho$$

$$g_{af} \rightarrow \bar{g}_{af}(s) = \sqrt{\varrho_f(s)} \cdot T_3^f \quad (2.8)$$

$$g_{vf} \rightarrow \bar{g}_{vf}(s) = \sqrt{\varrho_f(s)} \cdot (T_3^f - 2\kappa_f(s)q_f \sin^2 \Theta_W) \quad (2.9)$$

Both, ϱ and κ , although in principle dependent on the final fermion species f , are to a good approximation equal for all light fermions.

The $Zb\bar{b}$ vertex

Vertex corrections due to the top quark cannot be neglected any longer in the decay $Z^0 \rightarrow b\bar{b}$ and a further adjustment of the effective quantities is required

$$\begin{aligned}\bar{\rho}_b &= \bar{\rho}\left(1 - \frac{4}{3}\Delta\rho\right) \\ \sin^2 \overline{\Theta}_W^b &= \left(1 + \frac{2}{3}\Delta\rho\right) \sin^2 \overline{\Theta}_W\end{aligned}\quad (2.10)$$

The impact of the vertex corrections to the $Zb\bar{b}$ vertex can be observed best in the partial width Γ_{bb} . Whereas the partial width for lighter fermions becomes broader with rising top mass, propagator and vertex effects just cancel out for Γ_{bb} [2], leaving Γ_{bb} almost unchanged with m_t as plotted in figure 2.5. Thus a measurement of the difference $\Gamma_{q\bar{q}} - \Gamma_{bb}$ or the cross section ratio $R_{bb} = \frac{\sigma_{bb}}{\sigma_{had}}$ can be used for a determination of the top mass. Due to the small variation of R_{bb} within the expected range of the top mass a precision of 5 MeV on Γ_{bb} or 1% on R_{bb} is required for a meaningful estimate [12].

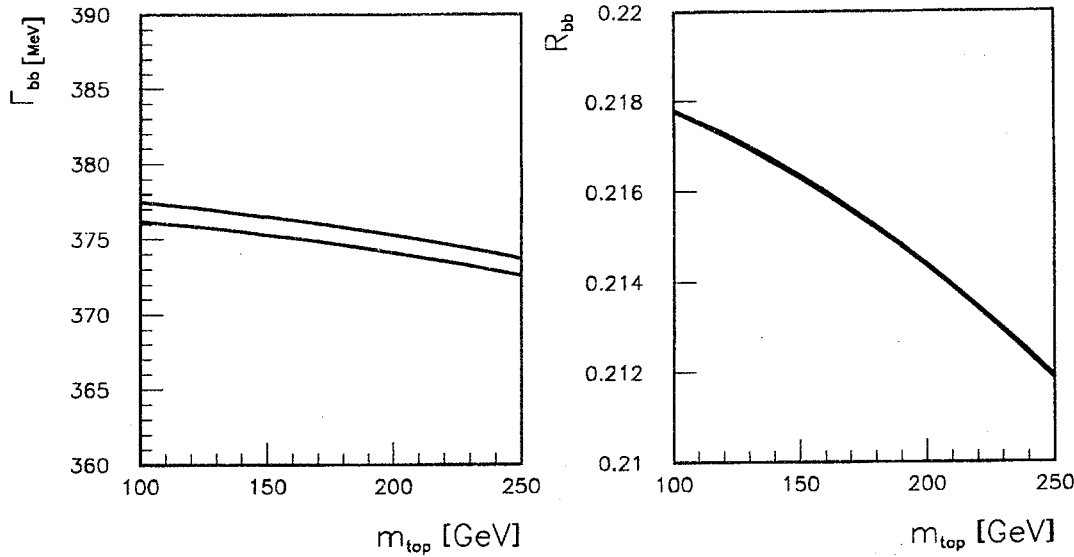


Figure 2.5: Γ_{bb} and R_{bb} as a function of the top mass. The band indicates the theoretical error due to a variation of the Higgs mass from 100 GeV to 1000 GeV.

Chapter 3

Neural Networks

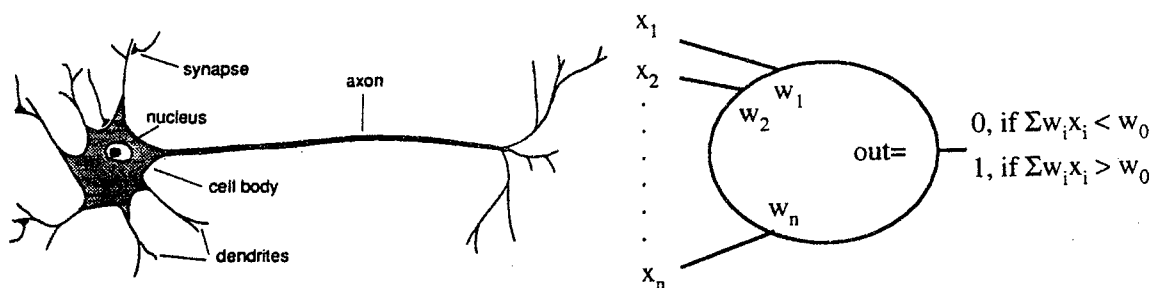


Figure 3.1: Real neuron and the model of McCulloch and Pitt

“Neural Network” is the generic term for a whole set of mathematical algorithms applied in multi-variate analysis. They are used for such different tasks like classification [13], optimization problems [14], clustering [15], track fitting [16], image processing [17], associative memories [18], or expert systems. All these algorithms have in common that they implement a heavily parallel and linked processing of the input data. This approach is contrary to the purely sequential processing of a von-Neumann machine and thought to be similar to the working of the brain. This equivalence is not limited to the global view. The basic quantity, which is calculated in many (but not in all) Neural Network algorithms is equivalent to a very simple model of a neuron suggested by McCulloch and Pitts [19]. Herein a neuron possesses several inputs x_i , which couple with the strength w_i , respectively. The neuron outputs a signal (it “fires”), when the weighted sum of all inputs becomes higher than some given threshold value w_0 .

$$\text{output} = \begin{cases} 1, & \text{if } \sum_i w_i \cdot x_i > w_0 \\ 0, & \text{if } \sum_i w_i \cdot x_i < w_0 \end{cases} \quad (3.1)$$

Nevertheless the equivalence for the majority of Neural Network algorithms and neuroscience is weak. This is particular true for the algorithm applied in this work. Because of that the

biological aspect won't be treated here in more detail.

For the analysis performed in this thesis, a Neural Network has been used for the selection of b events out of a data sample containing hadrons of all flavours. This constitutes a problem of optimal classification and therefore a short summary of decision theory is given in this chapter. First the Bayesian approach is reviewed. A special kind of Neural Network, the *Perceptron*, is applied in order to implement a discriminant function. Afterwards so-called *training algorithms* are outlined, which are used to adapt the Perceptron to a particular problem. The two most popular training algorithms, the pseudoinverse and error-minimization procedures, will be discussed. The classification performance of a linear discriminant decreases on non-gaussian probability distributions. Thus the concept of the linear Perceptron has to be extended, leading either to a multi-layered network or to the Functional-Link Net.

3.1 Bayes classification

The basic entities in decision theory are *events* and *classes*. An event k is described by several parameters x_{ki} ($i = 1 \dots n$), which make up the components of the *event vector* $\vec{x}_k \in \mathcal{F} \subseteq \mathbb{R}^n$. $\Omega = \{\omega_1, \omega_2 \dots \omega_c\}$ is a set of labels, i.e. isomorph to a finite subset of \mathbb{N} , and a unique¹ mapping $\mathcal{F} \rightarrow \Omega : \vec{x}_k \rightarrow \omega_l$ might exist. \mathcal{F} is called *feature space*. The elements ω_l of Ω are referred to as classes, Ω itself as *class-member space*.

If the class membership of an event is unknown, a minimum-risk analysis founded on classical probability theory [21] suggests to decide always in favour for the one assignment, which yields the highest class conditionally probability² $p(\omega_l|\vec{x}_k)$ in order to achieve a minimal overall number of misclassifications.

Bayes Decision Rule Assign \vec{x}_k to class $\omega = \omega_m$, if and only if

$$p(\omega_m|\vec{x}_k) > p(\omega_l|\vec{x}_k) \quad (3.2)$$

for all $l = 1 \dots c$, $m \neq l$, where $p(\omega_l|\vec{x}_k)$ are the class conditionally probabilities for class l .

The class conditionally probability $p(\omega_l|\vec{x}_k)$ can be derived from

$$\begin{aligned} P(\omega_l) &\equiv \text{the } a \text{ priori probability that an event belongs to } \omega_l, \text{ regardless of the identity} \\ &\quad \text{of the event. } (\sum_l P(\omega_l) = 1) \\ p(\vec{x}_k) &\equiv \text{the probability that an event is } \vec{x}_k, \text{ regardless of its class membership.} \\ &\quad (\sum_k p(\vec{x}_k) = \sum_k \sum_l p(\vec{x}_k|\omega_l) \cdot P(\omega_l) = 1) \\ p(\vec{x}_k|\omega_l) &\equiv \text{the conditional probability that an event is } \vec{x}_k, \text{ given that it belongs to class} \\ &\quad \omega_l \end{aligned}$$

The observation of the event \vec{x}_k changes the a priori probability $P(\omega_l)$ into the *a posteriori conditional probability* $P(\omega_l|\vec{x}_k)$, which can be computed with the *Bayes Relation*

$$p(\omega_l|\vec{x}_k) = \frac{p(\vec{x}_k|\omega_l) \cdot P(\omega_l)}{p(\vec{x}_k)} \quad (3.3)$$

Because $p(\vec{x}_k)$ is class-independent, the *joint probability* $p(\omega_l, \vec{x}_k) = P(\omega_l) \cdot p(\vec{x}_k|\omega_l)$ might be used instead of $p(\omega_l|\vec{x}_k)$. The application of (3.2) leads in fact to a maximum overall probability

¹The more general case of overlapping classes is a subject of *Fuzzy Logic*. See e.g. [20]

²Discrete probability distributions are labeled with capital P , continuous ones with small p .

for a correct classification. If the feature space is partitioned into r ($r \geq c$) arbitrary regions \mathcal{R}_i , than holds

$$\begin{aligned}
\langle P(\text{correct}) \rangle_{\mathcal{F}} &= \sum_{i=1 \dots r} \int_{\mathcal{R}_i} p(\omega_l | \vec{x}_k) \cdot p(\vec{x}_k) dx \\
&= \sum_{i=1 \dots r} \int_{\mathcal{R}_i} \frac{p(\vec{x}_k | \omega_l) \cdot P(\omega_l)}{p(\vec{x}_k)} \cdot p(\vec{x}_k) dx \\
&= \sum_{i=1 \dots r} \int_{\mathcal{R}_i} p(\vec{x}_k, \omega_l) dx \\
&= \sum_{i=1 \dots r} P(\vec{x} \in \mathcal{R}_i, \omega_i)
\end{aligned}$$

The expression will become maximal, when the regions \mathcal{R}_i are chosen in such a manner that the integrands become maximum, i.e. when (3.2) holds.

3.2 Linear discriminants

Bayes classification can be implemented in several ways. In case a closed representation of the a posteriori probability $p(\omega_l | \vec{x}_k)$ can be given, the most direct one would be an explicit computation [22]. Another approach uses the boundaries of the decision regions \mathcal{R}_i as discriminant. Herein a discriminant is a function or an operator, that, when applied to an event, yields an output that is an estimate of the class membership. In other words, the action of a discriminant is to produce a mapping from feature space to class-member space. An event \vec{x} is classified according to its position relative to the boundaries.

The boundary between two classes can be described analytically by the condition

$$p(\omega_i, \vec{x}_k) = p(\omega_j, \vec{x}_k) \quad \text{for all } j \neq i \text{ and any event } k$$

In the case of only two classes, where both probability distributions are n -dimensional Gaussians

$$p(\omega_l, \vec{x}) = P(\omega_l) \cdot \frac{\exp \left[-\frac{1}{2} (\vec{x} - \vec{\mu}_l)^t \Sigma_l^{-1} (\vec{x} - \vec{\mu}_l) \right]}{(2\pi)^{N/2} \sqrt{|\Sigma_l|}} \quad l = 1, 2$$

with the covariance matrices $\Sigma_1 = \Sigma_2 = \Sigma$, the boundary is given by

$$\vec{x}^t \cdot \left[\Sigma^{-1} (\vec{\mu}_2 - \vec{\mu}_1) \right] = \ln \left(\frac{P(\omega_1)}{P(\omega_2)} \right) \quad (3.4)$$

This is the equation of a hyperplane with the distance $\ln \left(\frac{P(\omega_1)}{P(\omega_2)} \right)$ from zero. (3.4) is called a *linear discriminant*. The position of \vec{x} relative to the hyperplane, i.e. its class membership, is given by the sign of the dot product.

A *classifier*, i.e. a classification algorithm, which exploits the concept of a linear discriminant, is the *Perceptron* by Rosenblatt [23]. In principle, it constitutes a slight reformulation of the model of McCulloch and Pitts. The weight vector \vec{w} contains the properties of the distributions, i.e. $\Sigma^{-1} \cdot (\vec{\mu}_2 - \vec{\mu}_1)$, and the threshold w_0 is given by the ratio of the a priori probabilities $\ln(P(\omega_1)/P(\omega_2))$. The position of \vec{x} is indicated by a *threshold function*³.

$$\text{signum}(\vec{w} \cdot \vec{x} - w_0) = \begin{cases} 0 \Leftrightarrow & \vec{x} \text{ belongs to class 1} \\ 1 \Leftrightarrow & \vec{x} \text{ belongs to class 2} \end{cases} \quad (3.5)$$

³The terms *transfer function* or *activation function* are often used synonymously.

Usually the weight vector \vec{w} and the threshold w_0 are not known and have to be determined with so-called *training procedures*.

3.3 Training of a linear discriminant

A linear discriminant is determined entirely by its normal vector \vec{w} and its distance to zero w_0 , which themselves are given essentially by the joint probabilities $p(\vec{x}_k, \omega_l)$. In order to adapt the classifier to a given problem, one uses a set (*trainingsample*) of events, which are distributed in the same manner as the events in the real classification problem. Two approaches exist, *supervised learning* and *unsupervised learning*. The distinction is that with supervised learning, one knows the class assignment for each event, whereas with unsupervised learning one does not. Only two procedures for supervised learning a two-class problem, i.e. $\omega \in \{1, 2\}$, will be presented in the following

The pseudoinverse

Given a trainingsample $\mathcal{T} = \{(\vec{x}, \omega)_k | k = 1 \dots t\}$ with the event vector $\vec{x} = (x_1, x_2 \dots x_n)$ and let $\omega \in \{1, 2\}$ according to the class membership of \vec{x} . Then the pairs $(\vec{x}, \omega)_k$ have to fulfill the following system of linear equations (The learning of the threshold w_0 has been included by extending \vec{x} with a dummy component $x_0 \equiv 1$. To simplify writing, this notification will be kept in future and no distinction will be made between weights and threshold any longer.)

$$\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{t1} & x_{t2} & \cdots & x_{tn} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_t \end{bmatrix} \quad (3.6)$$

(3.6) can be solved by

$$X^T X \vec{w} = X^T \vec{\omega} \quad \rightarrow \quad \vec{w} = (X^T X)^{-1} X^T \vec{\omega} = X^\dagger \vec{\omega}$$

where X^\dagger is called the *pseudoinverse*. Though this procedure yields an exact result for \vec{w} (when $X^T X$ is nonsingular), it is not yet widely applied. The reason is that one would like to have a high statistics⁴ in order to reproduce the probability distributions $p(\vec{x}_k, \omega_l)$ to a sufficient degree, causing the problem to invert a very large matrix $X^T X$. Nowadays programs like *Mathematica* [23] are able to handle this task rather easily [24]. Because for the analysis presented in this thesis, such a powerful package was not available, the gradient-descent algorithm presented in the next section has been applied.

The Widrow-Hoff Learning Algorithm

The original Widrow-Hoff Learning [25] was developed for a Perceptron (3.5), which uses a *signum* function as transfer function. This version is of no big interest for practical applications, thus only the version for a generalized Perceptron shall be outlined in the following.

The approach of the Widrow-Hoff Learning is founded on a risk analysis where every misclassification yields some loss. The overall loss for a given configuration of the Perceptron, i.e. a

⁴In this analysis $n = \mathcal{O}(10)$ and $t = \mathcal{O}(100000)$.

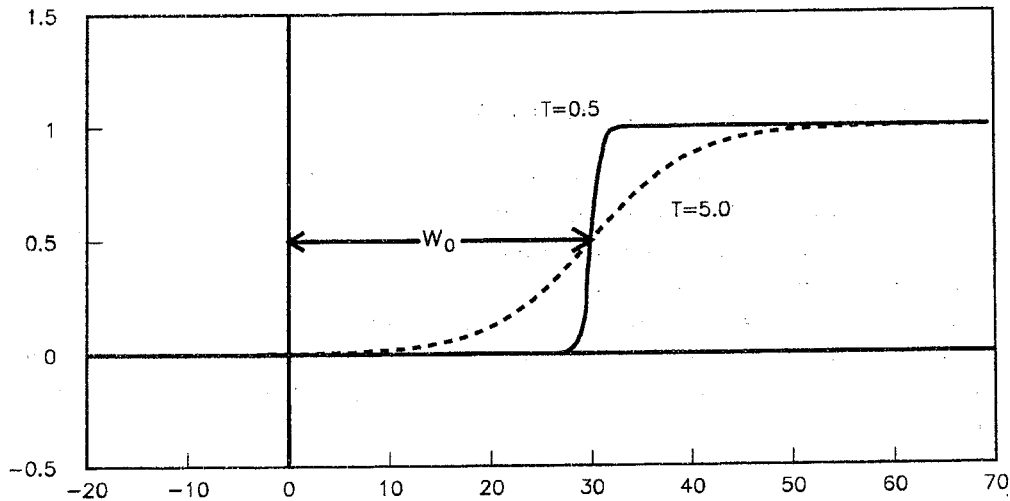


Figure 3.2: The Fermi function

given set of weights \vec{w} , is determined and then \vec{w} is modified in order to minimize this loss. The following quantities can be used to give an estimate for the loss

Mean-Squared Error:⁵

$$err = \frac{1}{2} \sum_{k=1}^t (\omega_k - out(\vec{x}_k))^2 \quad (3.7)$$

Entropy Error [26]:

$$err = \sum_{k=1}^t \{ \omega_k \cdot \log(out(\vec{x}_k)) + (1 - \omega_k) \cdot \log(1 - out(\vec{x}_k)) \} \quad (3.8)$$

with the (\vec{x}_k, ω_k) drawn from a training sample \mathcal{T} as defined in the previous section and $out(\vec{x}_k)$ is the output of the Perceptron as computed with the actual set of weights \vec{w} .

Usually the mean-square error is applied as a measure for the obtained loss. Some tests to study the application of the entropy error have been undertaken yielding no significant improvement. Results are presented briefly in appendix B.

In order to minimize (3.7) or (3.8), \vec{w} is modified proportional to $-\nabla_{\vec{w}} err$. Because the **signum** function is not differentiable, it is approximated by a *sigmoidal* function, which is continuous and differentiable. Usually a Fermi Function (Figure 3.2)

$$\mathcal{F}(\vec{x}) = \frac{1}{\exp\left(-\frac{\sum_i w_i \cdot x_i - w_0}{T}\right) + 1} \quad (3.9)$$

or a **tangens** are chosen for this purpose. For the Fermi function holds

$$\lim_{T \rightarrow 0} \mathcal{F}(\vec{x}) = \text{signum}(\vec{w} \cdot \vec{x} - w_0)$$

⁵This minimization of a mean-square error is a common source of the misunderstanding a Neural Network being "something like an n-dimensional fit". Neural Networks presented in this chapter are realizations of a decision algorithm, and one out of several methods to obtain its parameters is a fit.

The training starts with a random set of weights. The elements of the trainingsample \mathcal{T} are presented as input to the Perceptron one at a time and the deviation between the desired output ω_k and the actual output value is calculated. Afterwards the weights are changed by

$$\Delta_k w_i = -\eta \frac{\partial err}{\partial w_i} = -\eta (\omega_k - out(\vec{x}_k)) \cdot f'(\vec{w} \cdot \vec{x}_k) \cdot x_i \quad (3.10)$$

for the k -th event of the trainingsample and with any (differentiable) Perceptron transfer function f . When a Fermi Function (3.9) is used as transfer function, (3.10) simplifies to

$$\Delta_k w_i = -\eta (\omega_k - out(\vec{x}_k)) out(\vec{x}_k) (1 - out(\vec{x}_k)) \cdot x_i \quad (3.11)$$

The change of the weights may take place immediately after each presentation or is summed up until all elements of \mathcal{T} have been processed. Experience shows that the global minimum is not found in one presentation of \mathcal{T} . Thus the same sample may be presented several times in order to improve the accuracy (Figures 6.6, 6.7, 6.9, 6.12).

The tricks to accelerate the training and to avoid local minima are countless. Most of them yield a notifiable impact only for very special situations. One of the more useful ones is the addition of a damping term to (3.10) or (3.11). In order to smooth statistical fluctuations during the descent, the term takes into account the direction of the most recent modification $\Delta_{k-1} w_i$.

$$\Delta_k w_i = -\eta (\omega_k - out(\vec{x}_k)) f'(\vec{w} \cdot \vec{x}_k) x_i + \mu \cdot \Delta_{k-1} w_i \quad 0 < \mu < 1 \quad (3.12)$$

η is a real positive number. Its size is dependent on the problem. (See table 6.1 and in contrary table B.3.)

3.4 Non-linear discriminants

To design a linear discriminant very stringent restrictions have been made with respect to the properties of the probability distributions. In order to generalize the concept of discriminants to any distribution the event vector \vec{x} is mapped into an intermediate space, where the probability distributions are linear separable [27]. It was shown by Kolmogorov [28] that such a mapping always exists.

Theorem 1 (Kolmogorov) *Given any continuous function $f : [0, 1]^n \rightarrow \mathfrak{R}^m$, $f(\vec{x}) = \vec{y}$, f can be made up of two functions $g(\vec{x}) = \vec{z}$ and $h(\vec{z}) = \vec{y}$*

$$f : [0, 1]^n \xrightarrow{g} \mathfrak{R}^{(2n+1)} \xrightarrow{h} \mathfrak{R}^m$$

where holds

$$g_k(\vec{x}) = \sum_{j=1}^n \lambda^k \Psi(x_j + k\epsilon) + k \quad k = 1 \dots 2n + 1$$

with the real constant λ and the continuous real monotonically increasing function Ψ independent of f (though dependent on n) and ϵ a rational constant with $0 < \epsilon \leq \delta$ for an arbitrarily chosen positive constant δ . And

$$h_i(\vec{z}) = \sum_{k=1}^{2n+1} \Phi_i(z_k) \quad i = 1 \dots m$$

with the Φ_i are real and continuous and depend on f and ϵ .

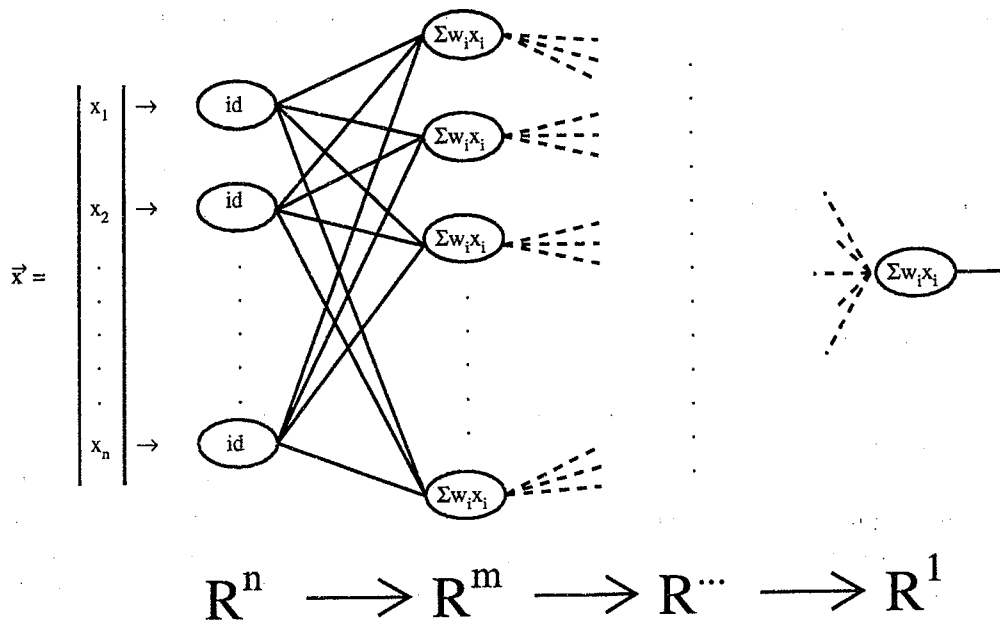


Figure 3.3: A Feedforward-Backpropagation Network consists of at least three layers. The hidden layers perform the mapping of the input vector into an intermediate space, where the classification problem is linearly separable. The output of each neuron in the layers $R^m \dots R^1$ is given by $f(\sum w_i \cdot x_i - w_0)$ with the transfer function f and the threshold w_0 . The action of f is not shown in the picture due to the limited space. Usually f is the same for every node.

Unfortunately the proof of the theorem is merely an existence proof and nothing is known about the very nature of the functions Ψ and Φ or about the ϵ . Nevertheless it is always possible to work with an expansion into a series of orthonormal functions. Two networks of this type will be presented in the following. The *Functional-Link Net* constitutes the most general realization of this idea. The *Feedforward-Backpropagation Net* exploits only the first order of the expansion. Because this net is still the most widely applied classifier in HEP analysis, it will be briefly discussed though.

The Feedforward-Backpropagation Net

It is a very neat approach to imagine the Feedforward-Backpropagation Net as being built of connected processing units instead of working with mapping functions. In this picture, the Feedforward-Backpropagation Net is made up of several layers of processing units (Figure 3.3). All units in one layer are processed in parallel and their outputs serve as input to the nodes in the next layer. The nodes in the first layer are merely fan-out units distributing the components of the event vector \vec{x} to the nodes of the second layer. The units in the upper layers consist of generalized Perceptrons. In order to deal with multi-class problems, the output layer may possess several units, each one representing one class, respectively. The layers between the input

layer and the output layer don't have any connection to the outer world and thus are referred to as *hidden layers*. They perform an approximation to the desired mapping of the input vector \vec{x} into a space, where the classification problem is linearly separable. One hidden layer is sufficient for most classification purposes [29].

The training is a generalization of the Widrow-Hoff Learning (3.10) and takes place in two steps. First one vector of the the trainingsample is fed into the net and the output is computed for the actual setting of the weights (*feedforward*). The weights of the output nodes are changed according to (3.10). The weights in the hidden layers are then modified according the changes in the output layer. Hence the error is *backpropagated* through the network. Because a Feedforward-Backpropagation Net has not been used for the physics analysis in this thesis, a detailed outline of the algorithm was banished into appendix A.

The Functional-Link Net

A processing unit in a Functional-Link Net [30] works similar to a generalized Perceptron, i.e. it calculates the value of a transfer function $f(\sum_i w_i \cdot x_i - w_0)$. But contrary to a Perceptron, it provides an entire output vector. The components of this output vector are the terms of an expansion of $f(\sum_i w_i \cdot x_i - w_0)$ into a series of orthonormal functions (Figure 3.4). All terms up to a certain order serve as input to succeeding nodes. Thus it has become possible to approximate the (unknown) functions Ψ and g , which have been demanded in theorem (1).

If already the components of the event vector \vec{x} are expanded in the described manner and this expanded vector \vec{x}' is used as input to a generalized Perceptron (Figure 3.5), a non-linear discriminant can already be implemented with just one single node. This expansion can be regarded as an *explicit transformation* of \vec{x} into a space, which is spanned by a basis of orthonormal functions and where the classification problem is linearly separable. This is in contrast to the *implicit transformation* of \vec{x} in the hidden layer of a Feedforward-Backpropagation Net. It was proven in [31] that \vec{x} can always be expanded in such a way, so that the matrix X in equation (3.6) becomes rectangular with $\det X \neq 0$. For real-life applications, the expansion

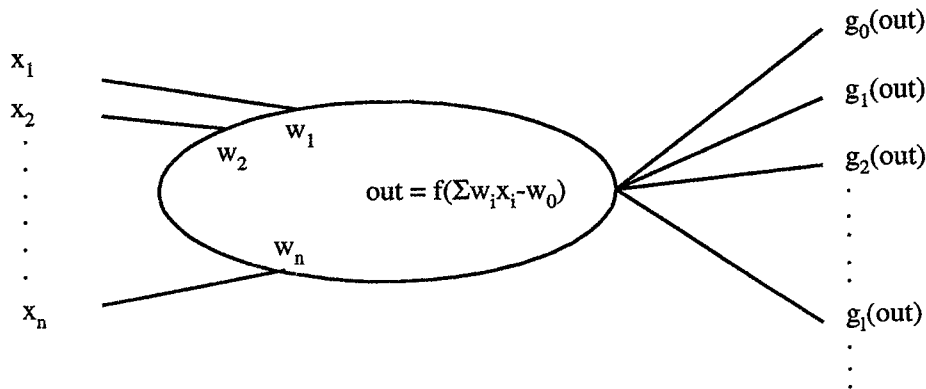


Figure 3.4: Concept of a *functional link*. The g_i are a set of orthonormal functions. (f is the Perceptron transfer function.)

will be terminated, when the desired discrimination precision has been reached. The training of a single-node Functional-Link Net is identical to the training of a Perceptron, hence either one solves the system of linear equations (3.6) or applies the Widrow-Hoff Learning.

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \rightarrow \vec{x}' = \begin{pmatrix} x_1 \\ g_0(x_1) \\ g_1(x_1) \\ \vdots \\ g_l(x_1) \\ x_2 \\ g_0(x_2) \\ g_1(x_2) \\ \vdots \\ g_l(x_2) \\ \vdots \\ x_n \\ g_0(x_n) \\ g_1(x_n) \\ \vdots \\ g_l(x_n) \end{pmatrix} \rightarrow out(\vec{x}') = f\left(\sum_{i=1}^n w'_i \cdot x'_i - w'_0\right)$$

Figure 3.5: In a single-node Functional-Link Net already the input vector is represented explicitly in a space, which is spanned by orthonormal functions $g_{i=0\dots l}$.

Chapter 4

The L3 Detector

The L3 detector shown in figure 4.1 is one of the four main experiments at the 'Large Electron Positron Collider' (LEP) [32] at CERN, Geneva. It is designed to study e^+e^- collision up to 200 GeV with emphasis on high resolution energy measurements of electrons, photons, muons, and jets. Viewed from the interaction point, L3 consists of the following operating subdetector components (at the end of 1992):

- a time expansion chamber for tracking charged particles
- an electromagnetic calorimeter composed of bismuth germanium oxide (BGO) crystals
- a luminosity monitor composed of BGO crystals
- an array of 30 plastic scintillation counters for trigger purposes
- a hadron calorimeter with depleted uranium as absorber and proportional wire chamber readout
- a muon spectrometer made of multi-wire drift chambers
- a large magnet providing a 0.5 T field

For the analysis performed in this thesis only the two calorimeter systems are of interest. The other detector components are described in [33].

The electromagnetic calorimeter

The electromagnetic calorimeter uses bismuth germanium oxide (BGO) as both the showering and detecting medium. The remarkable features of BGO in contrast to other scintillator materials are its high density (7.13 g/cm^3), its narrow Molière radius of 2.3 cm, and its excellent energy resolution of $\approx 5\%$ at 100 MeV and about 1.4% at high energies [33]. Due to the high density, the radiation length X_0 is only 1.12 cm. 24 cm ($\approx 22 X_0$) of BGO crystals in the electromagnetic calorimeter are sufficient to absorb almost completely a 50 GeV electromagnetic shower. The energy deposit of hadrons in the BGO is about 45% of their total energy.

The narrow lateral shower profile leads to the good spatial resolution of the energy clusters required for the presented analysis. It has been determined to [34]

$$\sigma_\phi = \frac{(7.1 \pm 0.4) \text{ mrad}}{\sqrt{E}} \quad \text{and} \quad \sigma_\vartheta = \sigma_\phi \sin \vartheta$$

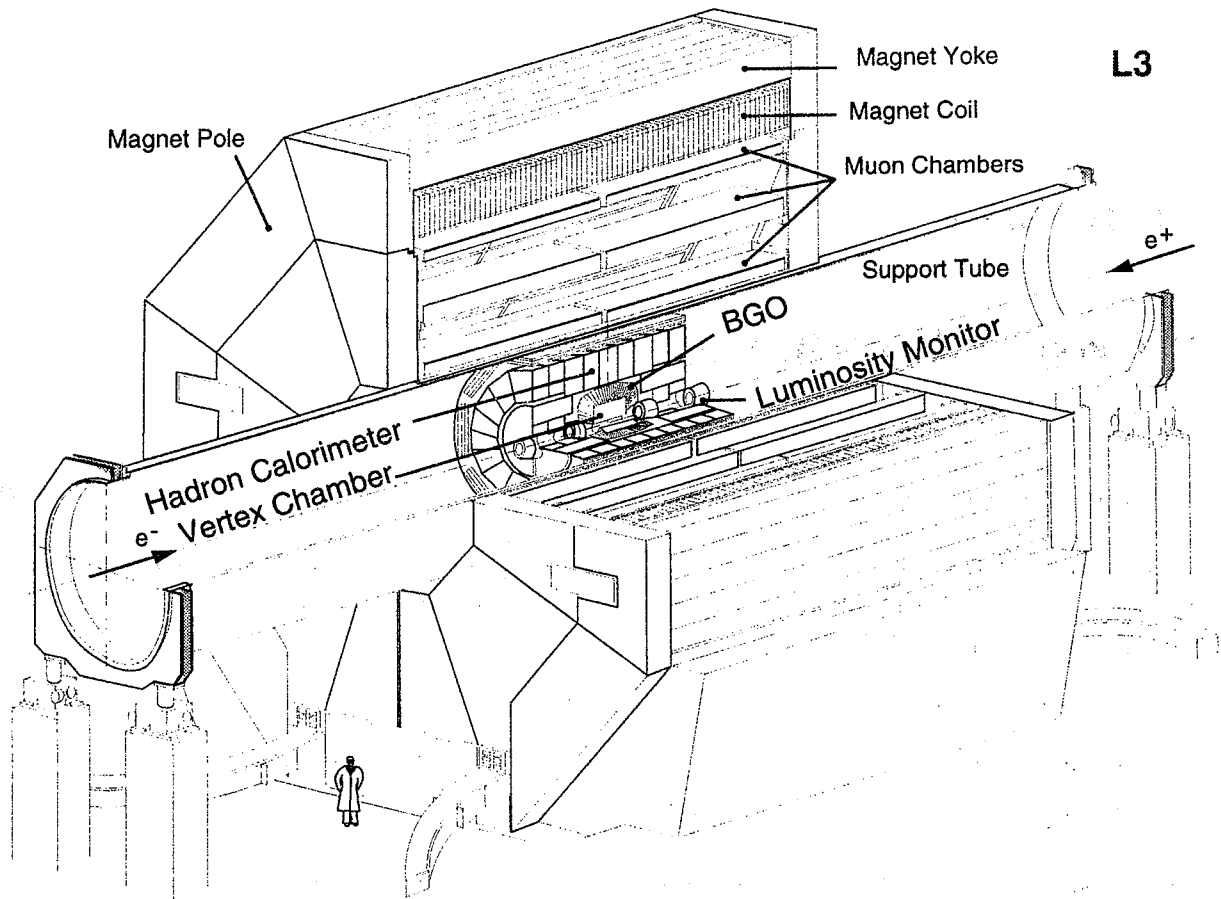


Figure 4.1: Perspective view of the L3 detector

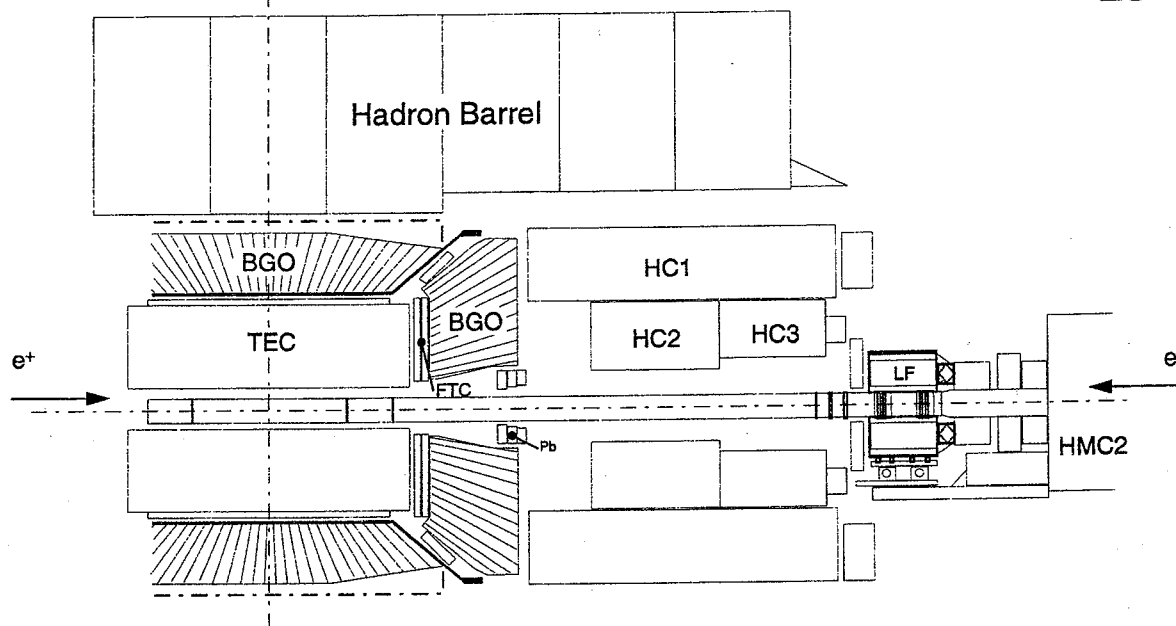
Further advantages of BGO are its chemical stability and the linear energy response.

The calorimeter itself consists of about 11000 BGO crystals. The inner radius of the barrel part is 50 cm. It is made up of 7680 crystals and covers a polar angle of $42^\circ < \vartheta < 138^\circ$. The two endcaps contain 1527 crystals each, and cover a polar angle of $11.6^\circ < \vartheta < 38^\circ$. Each BGO crystal has the shape of a truncated pyramid of about $2 \cdot 2$ cm at the inner end and $3 \cdot 3$ cm at the outer end. They point to the interaction point with a small angular offset to suppress photon leakage. The readout of the scintillation light is done by two 1.5 cm² photodiodes glued to its rear face. The diodes are insensitive to the magnetic field and have a quantum efficiency of about 70 %.

The hadron calorimeter

The hadron calorimeter system is built around the electromagnetic calorimeter and uses depleted uranium as absorber and proportional wire chambers as readout. Uranium was chosen as sampling material, because of its high density and in an effort to optimize the energy resolution due to compensation in the response for electromagnetic and hadronic showers [35].

The entire system is made up of three components, the barrel part, the forward-backward system, and the muon filter. The barrel covers the central region $35^\circ < \vartheta < 145^\circ$. It is a fine sampling calorimeter made of depleted uranium absorber plates interspersed with 7968



(RB26 Side)

LEP212 XP 1 0081 4

Figure 4.2: The central part of the L3 detector with the electromagnetic and the hadron calorimeters. The energy deposited in the luminosity monitor (LF) has not been used in the analysis.

proportional wire chambers and a total of 370000 wires. It has a modular structure consisting of 9 rings of 16 modules each. The wires are grouped to form readout towers. In the ϕ projection the towers point to the beam axis with a constant angular interval. Typically, a tower covers $\Delta\theta = 2^\circ$, $\Delta\Phi = 2^\circ$. The segmentation is 9 layers in ϕ and z and 10 (8) in the radial direction for the long (short) modules. The thickness including electromagnetic calorimeter and support structure is at least six nuclear absorption lengths in the barrel part.

The endcaps of the hadron calorimeter, which have been built at the I. Physikalisches Institut of the RWTH Aachen in collaboration with the Tata Institute in Bombay, cover the polar angle regions $5.5^\circ \leq \vartheta \leq 35^\circ$ and $145^\circ \leq \vartheta \leq 174.5^\circ$ over the full azimuthal range, and thus extend the coverage of the hadronic calorimetry to 99.5% of 4π . Each endcap consists of three separate rings, an outer ring and two inner rings. The rings are stainless steel containers filled with alternating layers of brass tube proportional chambers and 5 mm thick absorber plates of depleted uranium. The amount of material traversed by a particle originating at the interaction point varies between 6 and 7 nuclear absorption lengths. The wire signals are grouped to form 3960 towers for each endcap, with $\Delta\Phi = 22.5^\circ$, $\Delta\theta = 1.5^\circ - 1.8^\circ$.

A muon filter is mounted between the hadron calorimeter barrel and the support tube¹ and adds 1.03 absorption lengths. It consists of 8 identical octants, each made of six 1 cm thick brass

¹The support tube houses the inner detector components.

absorber plates, interleaved with five layers of proportional tubes and followed by 1.5 cm thick absorber plates matching the circular shape of the supporting tube.

The proportional wire chambers are operated with a gas mixture of 80 % Argon and 20 % CO₂. The energy resolution of the calorimeter has been determined to [33]

$$\frac{\sigma}{E} = \frac{55\%}{\sqrt{E}} + 5\%$$

The fine segmentation allows the measurement of the axis of jets with an angular resolution of approximately 2.5°, and the total energy of hadronic events from Z⁰ decay with a resolution of better than 10 %.

Trigger and readout

After each LEP beam crossing ($\approx 22 \mu s$ in 4-bunch mode), the data collected by the various subdetectors are read as analog signals within a few μs . When the level-1 trigger decision (see below) is positive, the analog signals are digitized for each detector component, collected by a FASTBUS system and, after passing the level-2 and level-3 trigger, written to tape. The entire data acquisition is steered by a VAX cluster, consisting of a central VAX 8800, several smaller VAX computers for each sub detector, and multiple workstations for online control.

It is the purpose of the trigger system to decide whether an event should be recorded. The trigger itself consists of three levels. The level-1 trigger is built with fast logic devices and uses the combined analog² signals from the various subdetectors. Based on its decision, either the digitization of the analog data starts or the readout electronic is reset. The decision and the reset has to be done before the next beam crossing; this is within 22 μs . There are essentially four subtrigger in level-1, the *energy trigger*, the *scintillator trigger*, the *charged track trigger*, and the *muon trigger*. Hadronic events are triggered by the first three of these subsystems

1. **Energy trigger** The energy trigger uses the informations of 256 trigger cells within the electromagnetic calorimeter and of 384 trigger cells within the hadron calorimeter. It is set, when 10 GeV have been deposited in the BGO barrel, 15 GeV in BGO barrel and endcaps, or 20 GeV in both calorimeter systems. The rate is about 2 Hz.
2. **Scintillator trigger** At least 6 out of 30 scintillator are required to have pulsed. The rate is about 0.2 Hz.
3. **Charged track trigger** At least two tracks with a transverse momentum (relative to the beam axis) higher than 0.15 GeV are required within the TEC. The angle between these tracks has to be larger than 120°. The rate is typically 1.5 Hz.

The rate of the level-1 trigger is about 5 Hz and largely dependent on the beam conditions. The dead time of the detector is smaller than 3%. Typically two of the four subtrigger are set during one event. This redundance guarantees a high trigger efficiency.

The level-2 trigger decision is working on the same data as the first level, but is implemented with fast processors allowing for a more complicated analysis. After a positive decision of the level-2 trigger the data of the various detector components are collected (event building) via a FASTBUS system and then fed into the level-3 trigger. The level-3 trigger consists of three processing units, which each emulate an IBM/3081 E computer. The analysis of this trigger level is able to consider all the data of the whole event. Only after a positive decision of level-3 the event is written to tape.

²The signals are digitized by very fast Analog-Digital Converters with less resolution (9-11 bit) than in the actual readout (12-20 bit).

Data processing

After the tapes have been written by the VAX cluster, they are processed on an IBM 3090/180J computer. In a first step (validate step), the integrity of the data is checked. In the next phase (Pass 1), the events on tape are reconstructed. This reconstruction takes place in two steps [38]. First the data of all subdetectors is treated independently. Herein the basic entities are collected into larger objects. Time dependent informations like calibration constants or different detector setups had been tracked in a database and are used in the reconstruction. In a second step, the objects reconstructed for each detector component are put together to form even larger objects spreading across the entire detector. An example for this procedure is the building of *ASRCs* and *ASJT's*, which is described in detail in chapter 5. This modular approach bears the advantage, that e.g. new calibration constants for one detector component don't require the reprocessing of the entire tape, but only for this particular subdetector.

Detector simulation

Simulated hadron events are necessary in order to train the classifier applied for the *b* selection and to determine the detector acceptances for *b* and for light quark events. For this purpose the JETSET 7.3 program [39] has been used. This program generates the energies and momenta of the fragmentation products coming from the process $e^+e^- \rightarrow \text{hadrons}$. A more detailed discussion can be found in chapter 8.

Then the generated particles are tracked through a detector simulation using the GEANT program [36]. This program takes into account effects like energy loss, multiple scattering, interactions and decay inside the detector material or the beam pipe. The processes inside the hadron calorimeter are simulated in particular by the program library GHEISHA [37]. The simulation programs generate an output format of hits and energy depositions, which is identical to the format written by the real data acquisition. Thus these simulated data can later be processed in the same manner like the real data.

Chapter 5

Selection of Hadronic Events

The selection of the b events is done in two steps. First a preselection takes place to obtain a datasample, which contains almost only hadronic events, i.e. processes of type $Z^0 \rightarrow q\bar{q}$. Then the actual b selection is carried out with the help of a Functional-Link Net. In this chapter the cuts applied in the preselection are presented. Because the hadron cuts, as well as the variables used as input to the classifier, are built out of energy clusters and jets, the process of finding these entities shall be discussed to some more detail now.

In a first step, the hits in the BGO and the HCAL, which are higher than a certain threshold energy (2 MeV for BGO, 9 MeV for HCAL), are grouped together into clusters. This is done separately for the two calorimeters. A BGO cluster (bump) consists of a most energetic crystal, the central crystal, and the neighbouring ones. In the hadron calorimeter hits are grouped geometrically (HGCL) [38]. Afterwards, the bumps are combined with the HGCLs which are located in a 3° cone behind them, viewed from the interaction point, in order to build an *ASRC*. (The *A* indicates, that it is about an object, which extends "Across L3" and the *SRC* stands for

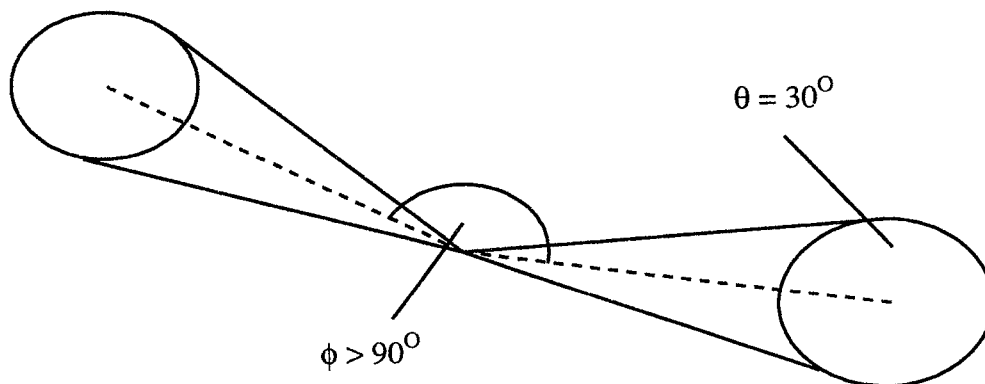


Figure 5.1: A jet is defined as the 30° cone around the radius vector of the most energetic *ASRC* not yet assigned to another jet. In order to avoid double assignment of an *ASRC* it has been required, that the angle between the two most energetic jets is larger than 90° .

“Smallest Resolvable Cluster”.) The direction of an *ASRC* is the vector sum of the direction of its constituting hits. Herein the “length” of a hit is defined by its energy. Its direction is given by its polar coordinates. The nature of the shower, either electromagnetic or hadronic, can be inferred from the longitudinal and transversal shape. An isolated bump without any HGCLs behind is taken as a electromagnetic *ASRC*. In general, one *ASRC* is supposed to correspond to one physical particle. The angular position of a hadronic *ASRC* is known with a precision of 10 *mrad* [40].

Jets are built from *ASRCs* and muons and are referred to as *ASJT*. The algorithm starts with the most energetic *ASRC* in the event. Its direction is taken as the jet axis. All *ASRCs* and muons in a 30° cone around this direction are gathered and the new direction is calculated from the momenta of all the included particles. This procedure is repeated as long until no new *ASRC* can be found in the 30° cone. After one jet is found, the procedure is applied again on the remaining *ASRCs* until all are used up. (A minimal energy of 1 *GeV* is required for an *ASJT*.) The angular resolution for jets was determined to be 36 *mrad* [41].

As described in chapter 4, all events taken with the L3 detector are written to one single data stream without any distinction. Thus the raw data tape contains besides the desired hadron events also leptonic processes of the type $Z \rightarrow ee$, $\mu\mu$, and $\tau\tau$. Moreover, the non-resonant background consists of two photon events, beamgas and beampipe interactions, and cosmics. The following cuts have been used in order to discriminate against the background

- (i) $0.6 < E_{tot}/\sqrt{s} < 1.4$
- (ii) $|E_{||}|/\sqrt{s} < 0.4$ with $E_{||}$ the energy imbalance parallel to the beampipe
- (iii) $E_{\perp}/\sqrt{s} < 0.4$ with E_{\perp} the energy imbalance perpendicular to the beampipe
- (iv) $N_{ASRC} > 18$ with N_{ASRC} the number of *ASRCs* in the event

where E_{tot} is defined as the sum over all *ASRCs*, i.e. the whole energy deposited in the two calorimeters. The energy of an *ASRC* was required to be larger then 100 *MeV* to suppress noise.

running period	b	light quarks
91a	0.9651 ± 0.0007	0.9486 ± 0.0005
91bc	0.9591 ± 0.0007	0.9460 ± 0.0004
92	0.9586 ± 0.0006	0.9447 ± 0.0003

Table 5.1: The acceptances for b and light quark events in the three running periods after the cuts have been applied. Due to modifications of the L3 detector the numbers are slightly different for the running periods, respectively.

Due to the special definitions of the input variables of the Functional-Link Net (see next chapter) two further cuts were required

(v) at least 2 jets with $E_{jet} > 10 GeV$ and $N_{ASRC}/jet > 4$

(vi) $\angle(Jet_1, Jet_2) \geq 90^\circ$

The distributions of the variables are shown in figure 5.2. The bullets are the distributions for the data taken in the second half of 1991 (91bc), the histograms show the Monte-Carlo prediction for the variables. The MC events were generated with the JETSET 7.3 program [39] with string fragmentation and full detector simulation. The agreement between data and MC is very good in the signal region, except for the number of *ASRCs*. Here the multiplicity is higher than expected due to low energetic bumps, which are not contained in the Monte-Carlo¹.

Leptonic events have a low cluster multiplicity and are discriminated by the N_{ASRC} -cut. A further reduction of the process $Z \rightarrow \mu\mu, \tau\tau$ is achieved by a cut in E_{tot}/\sqrt{s} . Cosmic muons also have a low multiplicity and a low energy deposit in the calorimeters. Beamgas and beam-wall interactions as well as two-photon-processes feature a low total energy and show also an imbalance in the energy deposit. After the cuts (i)-(iv), the fraction of remaining background processes is smaller than 0.5 % [41]. Since the calorimeters cover 99.5 % of the full solid angle, the trigger efficiency is very high, $(99.15 \pm 0.03) \%$ [41].

running period	$\sqrt{s} [GeV]$	N_{event}
91a	91.254	151871
	88.480	3885
91bc	89.470	8338
	90.228	12157
	91.222	87936
	91.967	16805
	92.966	10399
	93.716	8067
92	91.294	658619

Table 5.2: The number of hadron events collected during 1991 and 1992 at the various energy points. The 1991 data are split into two parts reflecting the improved LEP energy calibration after August 1991.

¹In the meantime the problem has been overcome using an improved version of the GEANT program.

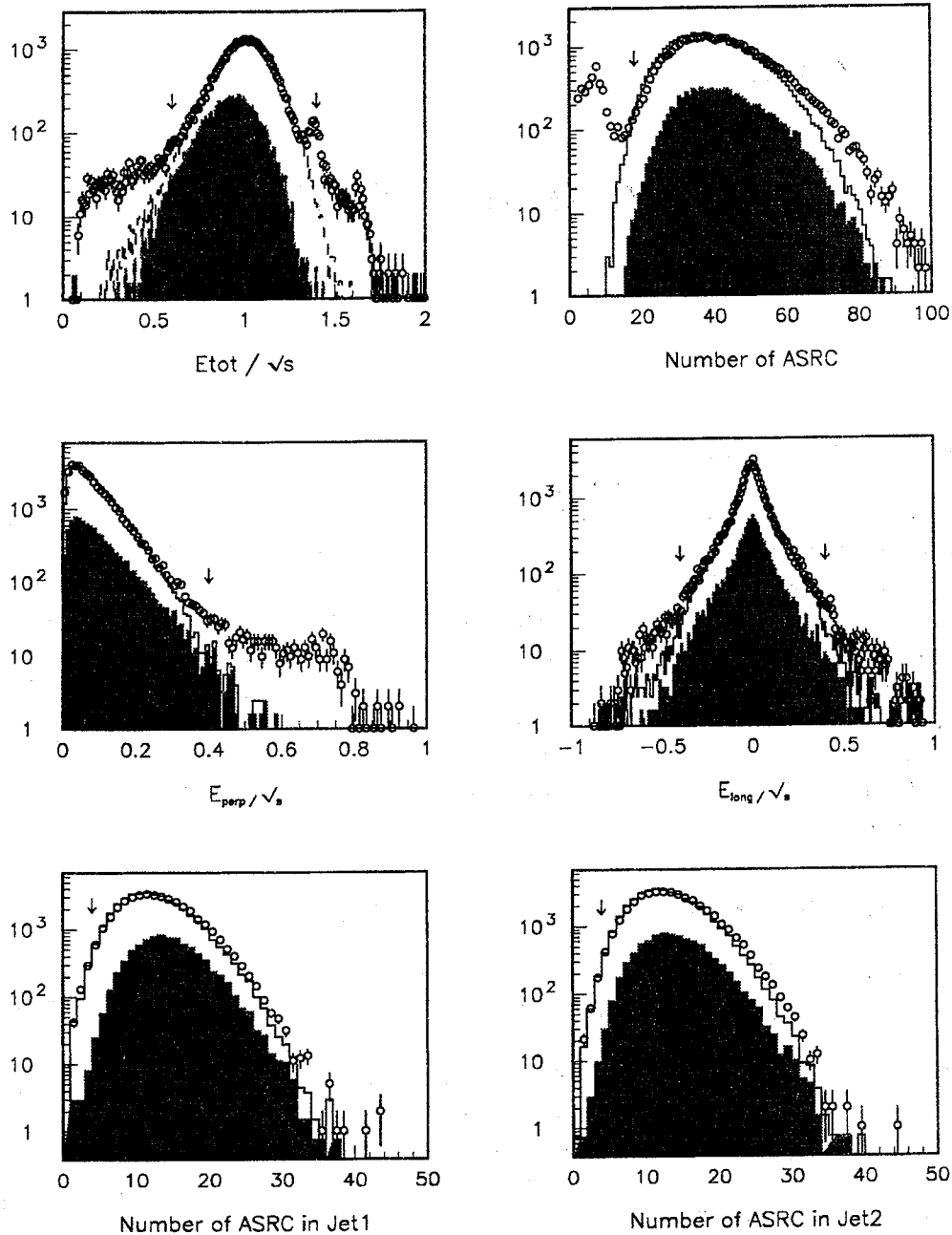


Figure 5.2: The plots show the distributions of the cut variables. The bullets are for the data taken in the second half of 1991, the histograms (white for all events, shaded *b*) give the Monte-Carlo prediction. Due to the missing energy carried away by the neutrino, the mean of the E_{tot}/\sqrt{s} distribution for *b* events is less than 1. The peak in the data distribution at $E_{tot}/\sqrt{s} \approx 1.5$ is due to Bhabha's and of no further interest.

Chapter 6

The Classifier used for the b Selection

Inclusive leptons have already been used in the L3 analysis to tag b events [42]. This method is reviewed shortly in appendix D. The analysis presented in this thesis however is based on event and jet shape variables. Shape variables are built from the energy flux pattern measured in the calorimeters. The relatively high mass of the b quark results into an event shape, which is different from the event topologies made from lighter quarks. Six variables have been defined in order to describe the differences. These variables are calculated respectively for each single event and then serve as input to a Neural Network. According to the output of the net, the event is classified as coming from a b or a light quark. The advantages of this approach are its independence of the semileptonic branching ratio and the high underlying statistics.

In the first part of the chapter the differences in the shape of b events in contrast to events from lighter quarks are discussed as well as the variables, which have been designed in order to quantify these differences. The second part will report of the training and the achieved performance of the classifier.

6.1 Variables used as input to the classifier

In the hadronization of the b quark, very little energy is radiated in the form of gluons due to its high mass [44]. Thus, the fragmentation of the b is hard and the resulting b hadron carries off a large fraction of the beam energy. At LEP1 energies this is typically 70 % for a b hadron but only 50 % for a prompt c hadron [43]. The b hadron then decays into several particles in a weak process with the energy evenly distributed at the resulting decay products. In contrast, most of the energy of a light quark is radiated out in the form of gluons leading to a broad energy spectrum of particles. Thus, the energy gaps between the leading particles are larger for light quark jets. Also the light leading hadron being stable can still be fairly energetic. Figure 6.1 shows the most probable energy of the four leading clusters in the most energetic jet.

In addition, the mass of the b quark affects the jet characteristics. The jet sphericity is high as a consequence of the large rest mass of the b hadrons [45]. This effect can be enhanced further if the jet sphericities are measured in a Lorentz frame, which is a result of a boost toward the b hadron restframe. The jet boost ($\gamma\beta = p/\mu$ with μ the invariant mass) is also smaller for b quark jets. For LEP1 energies $\gamma\beta \approx 8$ for a b jet.

Based on these considerations five variables have been refined. They are calculated for each of the two most energetic jets of the event. The information used is based on calorimetric clusters

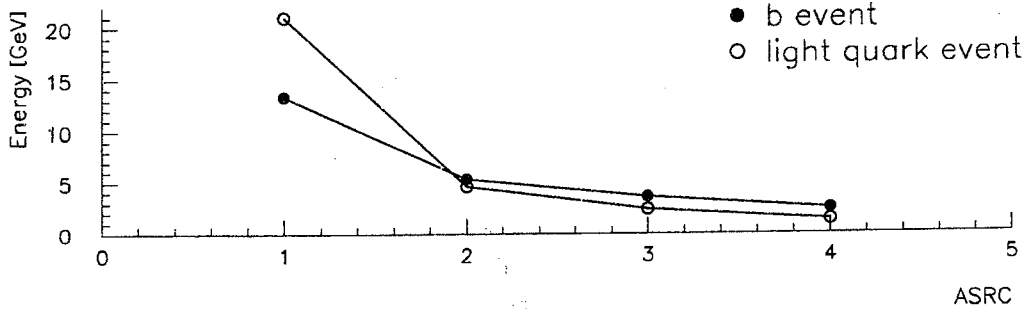


Figure 6.1: Most probable energy of the four most energetic clusters in a jet. The descend in energy for successive clusters is much steeper for lighter quarks.

only. For the calculation of the invariant mass, the clusters, i.e. the final particles, are assumed to be massless, hence $E = p$. In particular, the variables are defined as follows

1. (**ESRC**) The energy of the most energetic cluster in the jet.
2. (**EGAP**) The difference between the first and the fourth most energetic cluster in the jet.

$$\mathbf{EGAP} = \frac{E_1 - E_4}{E_{Jet}}$$

The expression has been normalized to cancel out uncertainties in the energy calibration constants (g-factors). Several differences were tested and the one between the first and the fourth cluster was found to yield the best selection performance, when used as input to the classifier.

3. (**SPHE**) The *boosted jet sphericity* is defined as

$$\mathbf{SPHE} = \frac{\sum p_i^2}{\sum p^2}$$

where the momenta are in the jet rest frame and the transverse momentum is with respect to the jet direction. (For the definition of a jet see chapter 5.)

4. (**GBJET**) The jet boost $\gamma\beta$ has been calculated for the entire jet
5. (**GBSRC**) The jet boost $\gamma\beta$ of the set of the four leading clusters in the jet.

In addition, it was found that the fraction of visible energy outside the two most energetic jets improved the selection performance quite a lot though there is no difference between the distributions for b and light quarks (Figure 6.13).

6. (**DJET**)

$$\mathbf{DJET} = \frac{E_{tot} - (E_{Jet1} + E_{Jet2})}{E_{tot}}$$

Again the expression has been normalized in order to cancel out uncertainties in the g -factors.

The distributions of the variables are plotted in figure 6.3 and 6.4. In general, data and MC agree rather well, except for the **SPHE** and the **EGAP** variable. These disagreements result from the occurrence of low energetic clusters in the data, which is higher than predicted in the MC. (See footnote on page 26.) They yield a higher energy difference between the leading clusters in the jet. Also the sphericity is lower due to faked particles with lower momentum.

It is claimed one of the features of a Neural Network algorithm, that it is still capable to classify even noisy data in a correct manner. In fact this statement holds also true in the actual case. The bad agreement for the sphericity and the energy gap is not propagated to the output distribution of the net (Figure 7.1). Only this distribution is of interest for the b selection and here data and MC agree very well. Because the net is obviously able to handle sufficiently the noisy data and on the other hand the application of these quantities improves the selection performance of the classifier quite a bit (Figure 6.2), it seems acceptable to keep them as input variables.

The MC predictions for the b have been checked using data events with a high p_t muon ($p_t > 1 \text{ GeV}$, $p_\mu > 4 \text{ GeV}$), where p_t is the momentum of the muon transverse to the jet axis. The b purity of a sample tagged with inclusive muons is almost 80 % [46].

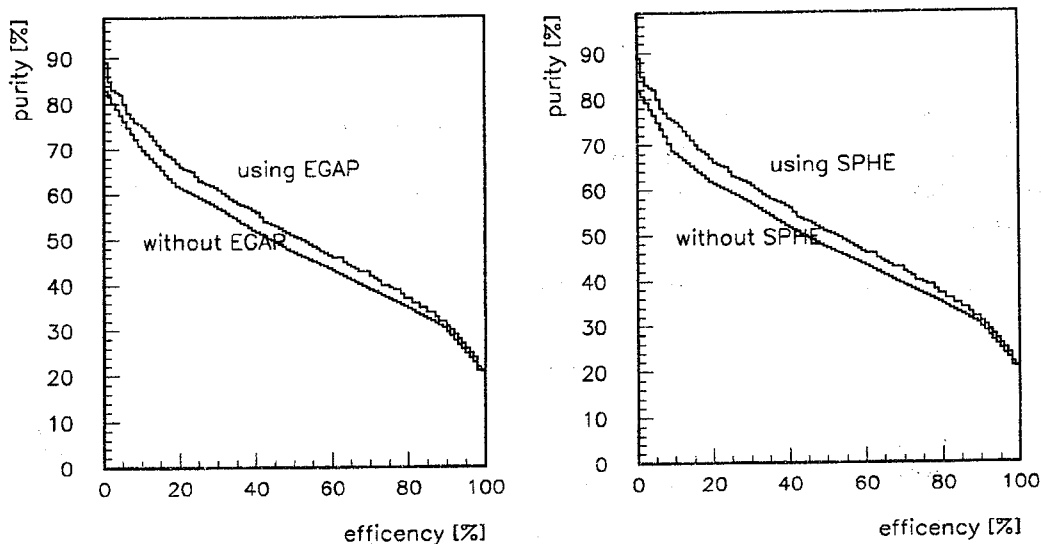


Figure 6.2: The selection performance is notably improved by the variables **EGAP** (left) and **SPHE** (right).

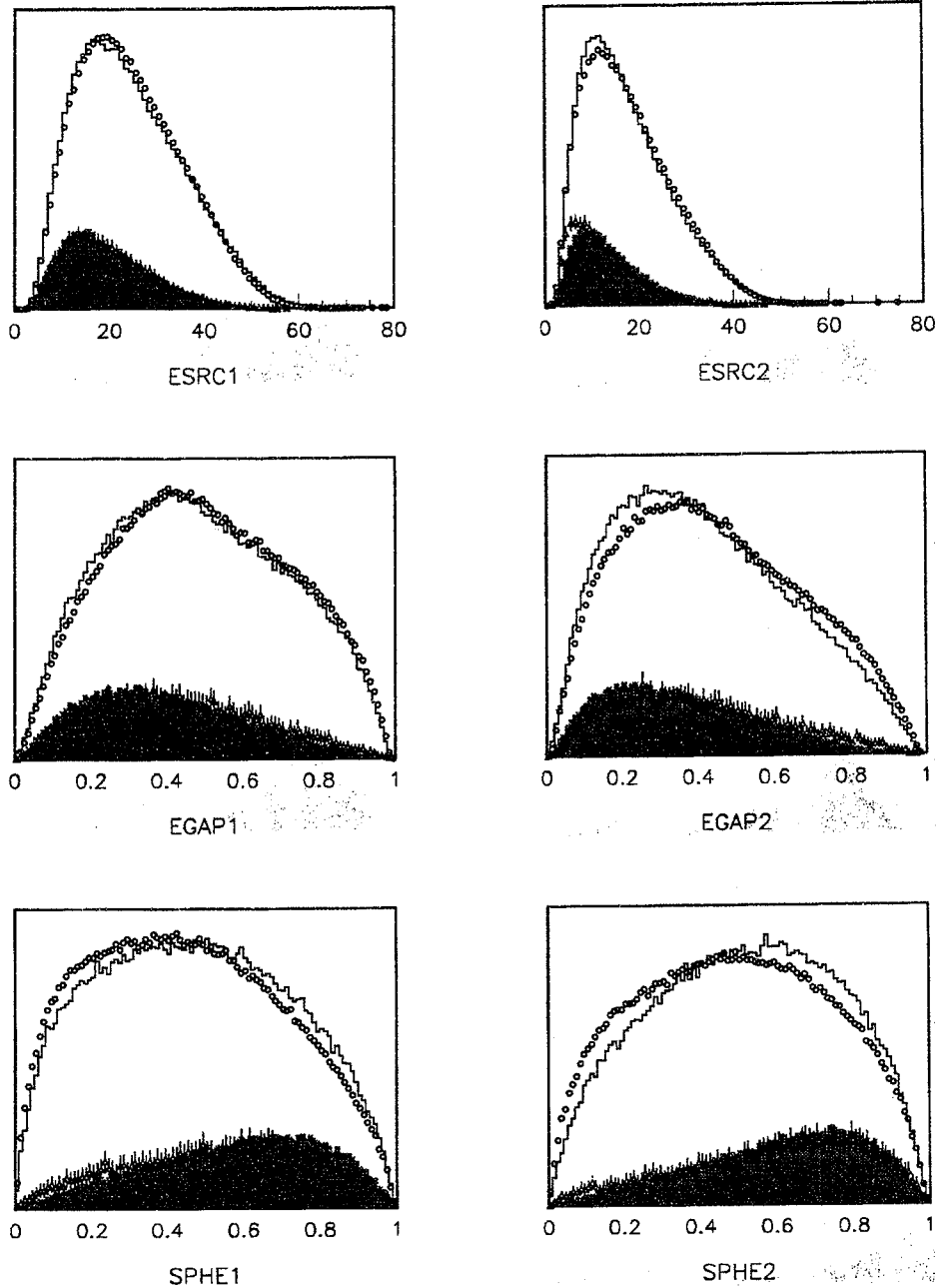
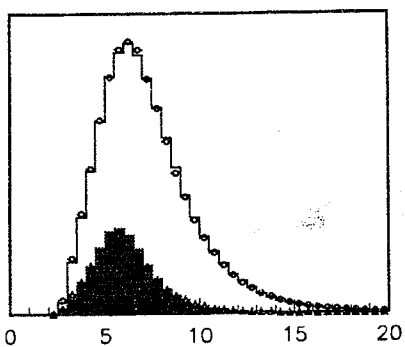
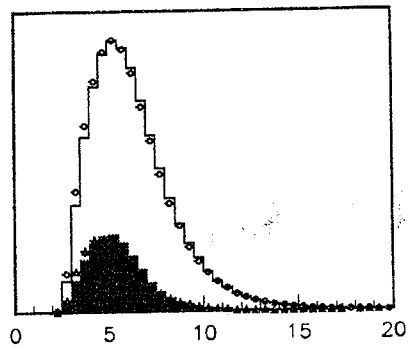


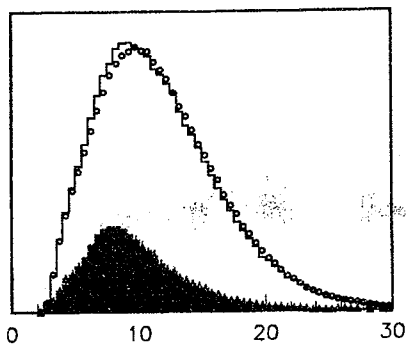
Figure 6.3: The histograms show the distributions of the input variables as predicted by MC. The white bullets are the hadronic data taken during 1992. The disagreement between data and MC for **SPHE** and **EGAP** are due to low energetic clusters found in the data. The triangles are those events, which contain an inclusive muon ($\approx 80\%$ b purity). The numbers at the end of the variable name indicate the jet.



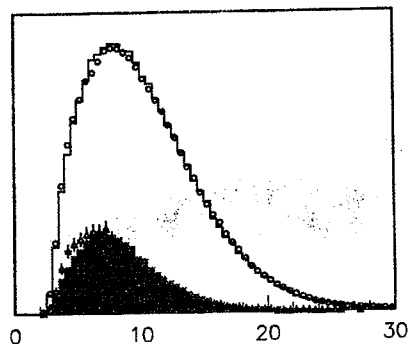
GBJET1



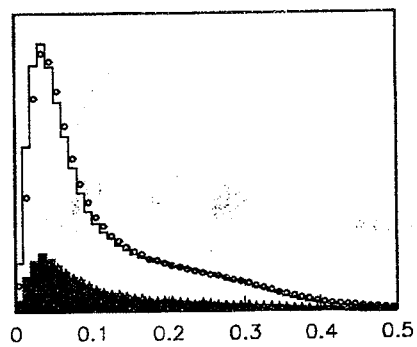
GBJET2



GBSRC1



GBSRC2



DJET/ETOT

Figure 6.4: The distributions of the input variables 4-6.

$$\vec{x} = \begin{pmatrix} \text{ESRC}_{Jet1} \\ \text{ESRC}_{Jet2} \\ \text{EGAP}_{Jet1} \\ \text{EGAP}_{Jet2} \\ \text{SPHE}_{Jet1} \\ \text{SPHE}_{Jet2} \\ \text{GBJET}_{Jet1} \\ \text{GBJET}_{Jet2} \\ \text{GBSRC}_{Jet1} \\ \text{GBSRC}_{Jet2} \\ \text{DJET} \end{pmatrix} \rightarrow \vec{x}' = \begin{pmatrix} \text{ESRC}_{Jet1} \\ \sin(\pi \cdot \text{ESRC}_{Jet1}) \\ \cos(\pi \cdot \text{ESRC}_{Jet1}) \\ \sin(2\pi \cdot \text{ESRC}_{Jet1}) \\ \cos(2\pi \cdot \text{ESRC}_{Jet1}) \\ \sin(3\pi \cdot \text{ESRC}_{Jet1}) \\ \cos(3\pi \cdot \text{ESRC}_{Jet1}) \\ \sin(4\pi \cdot \text{ESRC}_{Jet1}) \\ \cos(4\pi \cdot \text{ESRC}_{Jet1}) \\ \text{ESRC}_{Jet2} \\ \vdots \\ \sin(4\pi \cdot \text{DJET}) \\ \cos(4\pi \cdot \text{DJET}) \end{pmatrix} \rightarrow \text{out}(\vec{x}') = \frac{1}{\exp(-\frac{\vec{w}' \cdot \vec{x}' - w_0}{T}) + 1}$$

Figure 6.5: The way the classifier works

6.2 Training and performance of the classifier

A single-node *Functional-Link Net*, as described in chapter 3, has been used as classifier for the b selection. The 11 variables presented above are calculated for each event, respectively, and then normalized to their mean. The purpose of this normalization is twofold. First the values of the variables have to stay within some limits to take into account the periodic nature of the Fourier terms. Second, uncertainties in g -factors cancel out. Afterwards the first four terms of the Fourier series are calculated for each variable and this, so called, *expanded event vector* \vec{x}' serves eventually as input to the genuine classifier, which consists of a single Perceptron node. The Perceptron evaluates the weighted sum of the components of \vec{x}' together with a threshold w_0 and issues a value $\text{out}(\vec{x}') \in [0, 1]$.

In order to obtain different distributions of $\text{out}(\vec{x}')$ for b and for light quarks, the classifier has to be trained, this is, the weights of the Perceptron have to be determined properly. This was done with the *Widrow-Hoff Learning Algorithm* (3.11) using in addition a damping term (3.12). A *trainingsample* consisting of 40000 JETSET 7.3 hadron events, fully reconstructed for the detector setup during the running period 1991bc and a cms-energy of the Z^0 mass, was presented to the classifier about 300 times¹. To obtain a balanced training, the sample contained equal numbers of b and light quark events. The weights were updated according equation (3.12) after every single sweep. The best training performance was achieved with the parameter setting listed in table 6.1. For the events of the training set the squared-error measure (3.7) reaches a stable value after at least 100 cycles (Figure 6.6). For an independent *testsample* (see below) the error continues to decrease slightly even at 300 cycles. But because no further significant improvement was expected, the training was terminated at this point.

After the training, the classifier remains unchanged for the rest of the analysis. This means in particular that the same weights were used the whole time. (Actually, the set of weights obtained after 284 cycles has been used in the further analysis, because the network output at this time yielded the best χ^2 in the fit described in the next chapter.) It has not been

¹In the slang of Neural Network people a single presentation of an event vector is called a *sweep*. A presentation of the whole training set is referred to as *cycle*. Hence the training consisted of 300 cycles with 40000 sweeps, respectively.

Orthonormal functions	$\sin(i \cdot \pi \cdot x) + \cos(i \cdot \pi \cdot x)$
Order of expansion	4
Learning parameter η	0.001
Damping factor μ	0.5
Temperature T	5.5
Size of training sample	40000 events (20000 b , 20000 light quarks)
Number of training cycles	284

Table 6.1: The parameters of the Functional-Link Net applied for the analysis.

retrained for input data taken with different detector setups, other cms-energies, or the checks for systematic errors described in chapter 8.

The selection performance of the classifier was tested with an independent sample of 300000 fully reconstructed JETSET 7.3 events. This time, the fraction of the quark flavours was as predicted by the Standard Model. ($R_{bb} \approx 0.22$). Figure 6.7 shows the evolution of the classifier decisions during the training period. Due to the use of a Fermi Function as an approximation of the **signum** function, the crisp $\{0,1\}$ -output of an ideal classifier has been smeared into the whole intervall $[0,1]$. In order to regain the optimal behaviour of a Bayes Decision Machine, as defined in chapter 3, it has become necessary to apply a cut, where the Fermi Function intersects the leap of the signum function. This is the case just at the turning point of the Fermi Function, for which holds: $\vec{w} \cdot \vec{x}' = w_0$, so that the denominator in the exponential becomes zero and $\mathcal{F}(\vec{x}) = 0.5$. Thus a cut on the output distributions at 0.5 yields an optimal classification by the net. Figure 6.8 shows in fact, that the number of misclassifications becomes minimal at 0.5, so that the used Functional-Link Net really implements a Bayes Classifier. Figure 6.9 shows, that though the maximum of *purity · efficiency* still rises slightly after 300 cycles, the location of the maximum becomes stable almost immediatley after a few trainings cycles. It should be stressed, that the choice of the “temperature” T of the Fermi function does not affect the location of the cut, so that this choice has no impact on the optimal error rate of the classifier (Figure 6.10). Its choice is of great importance for the training performance though (Appendix B).

Figure 6.11 shows the possible combinations of b efficiency and the sample purity², when the output is used as a continuous discriminant.

Contrary to a selection founded on cuts in 1-dimensional distributions, Neural Networks are also able to take into account relations between the input variables. This capability shows up for the application of variable DJET. Although there is no difference between the distributions of signal and background for this variable, used together with the other variables the discrimination power of the classifier is notably improved (Figure 6.13).

The use of orthogonal functions other than the Fourier terms, eg. polynomials (10th order) or Bessel functions (4th order) in order to expand the event vector, resulted in a worse classification performance of the net. Variations of the training procedure, as for instance accumulating the

²Let b the number of b events and x the number of light quark events in the sample obtained with a cut at the classifier output, and let B the number of all b events in the total hadron sample. Then *efficiency* and *purity* are defined: $efficiency = \frac{b}{B}$ and $purity = \frac{b}{b+x}$.

updates and changing the weights only after each cycle or decreasing the learning parameter η from 0.01 to 0.001 during the training yielded no improvement. The same holds true for the sequence of the presentation. It did not matter, if the events were presented sequentially or selected randomly during one cycle. Further tests for systematic effects are presented in chapter 8.

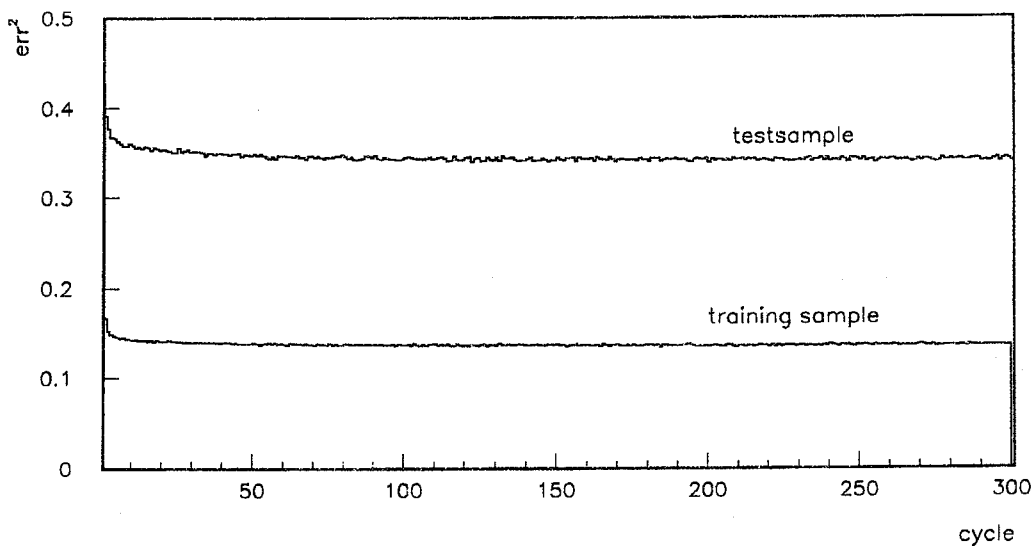


Figure 6.6: The squared-error measure (3.7) normalized to the number of events has been calculated after each cycle, once for the training sample itself and twice for an independent test sample. Although the error decreases still very slightly, the training was terminated after 300 cycles.

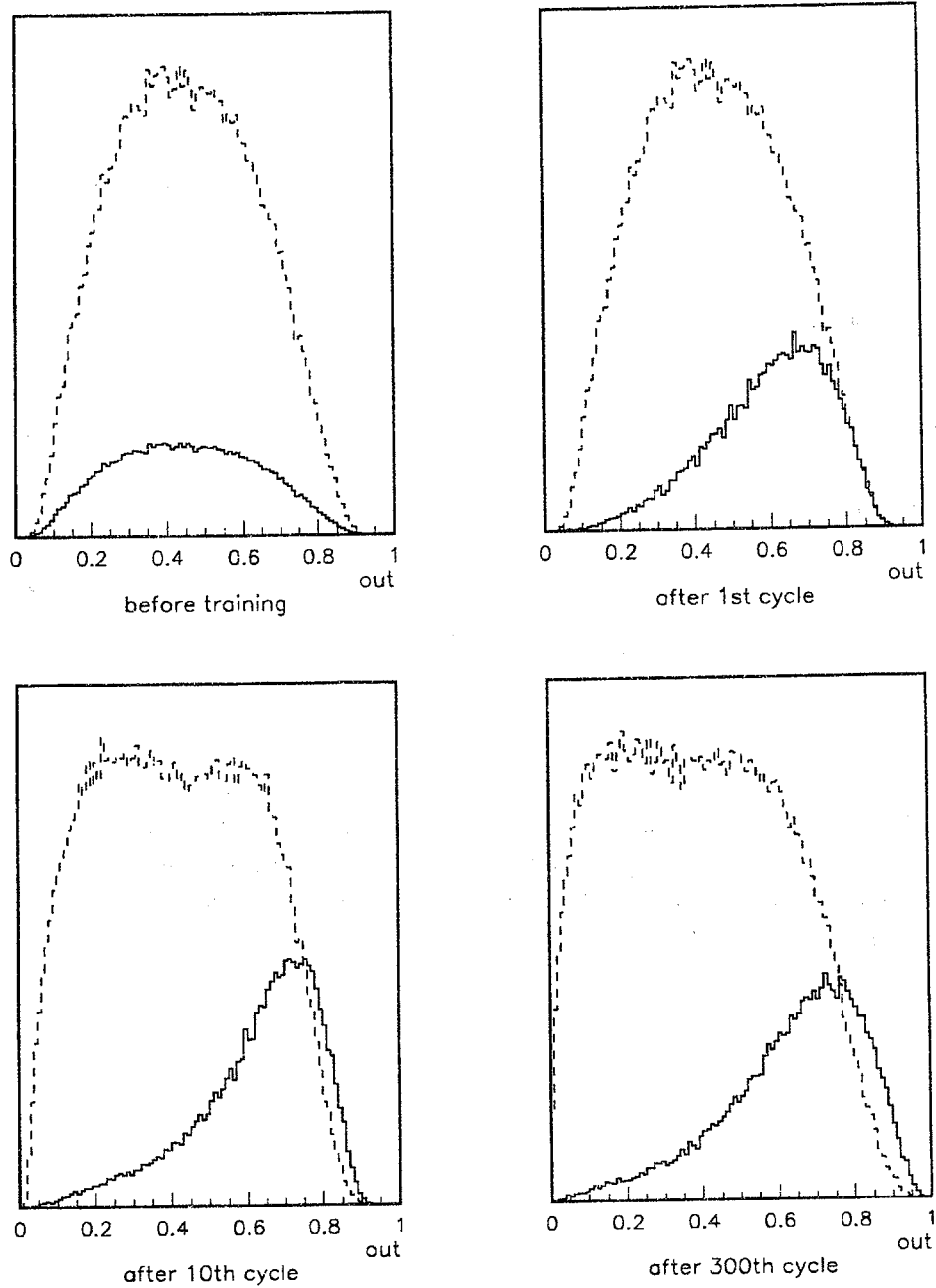


Figure 6.7: The evolution of the classifier output for signal (straight line) and background (dashed line) during training. The output distributions are for an independent testsample.

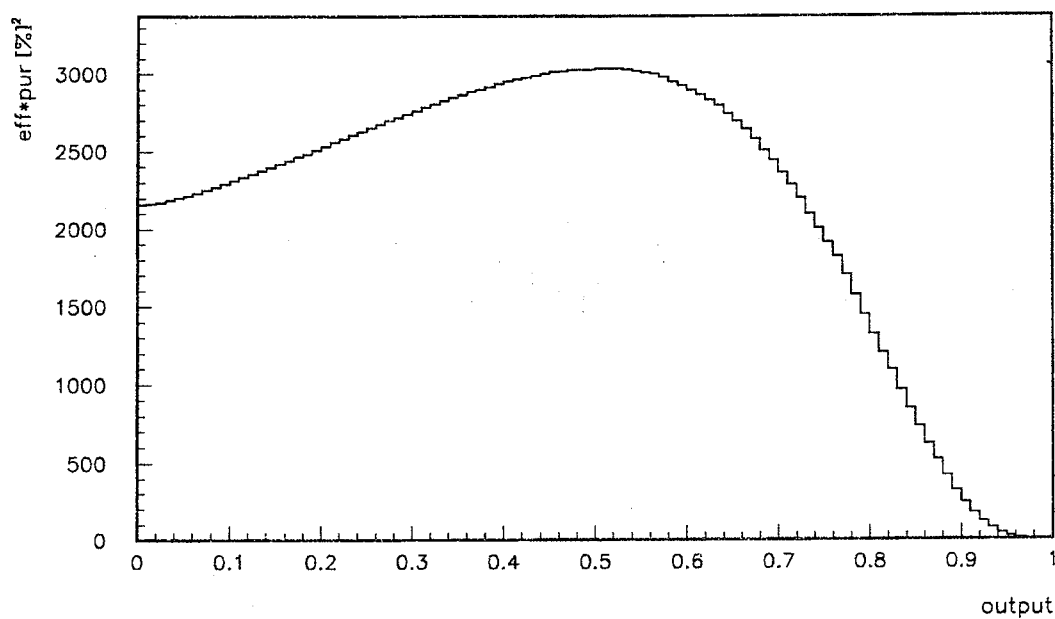
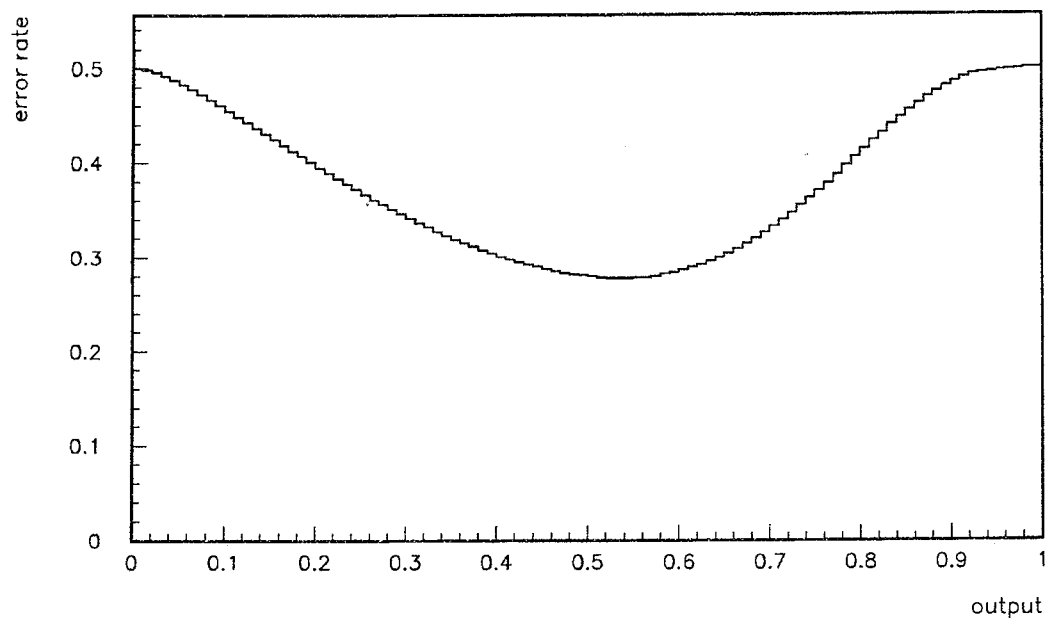


Figure 6.8: For a cut at 0.5 on the output the error rate becomes minimal. The product of efficiency and purity is an equivalent measure but can be calculated more easily. Therefore this quantity will be used in the further analysis.

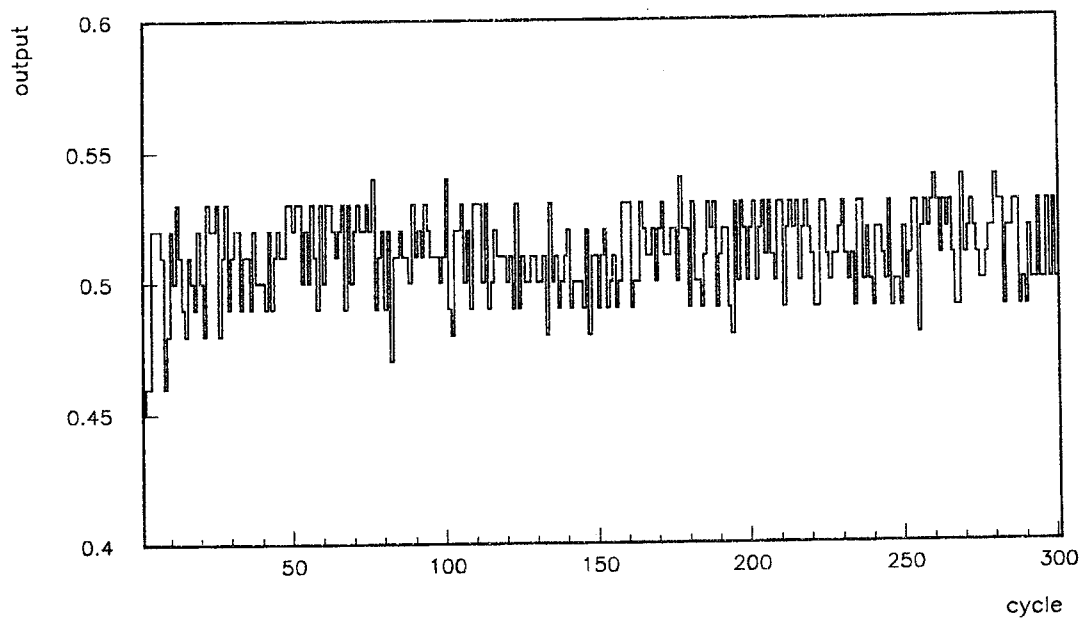
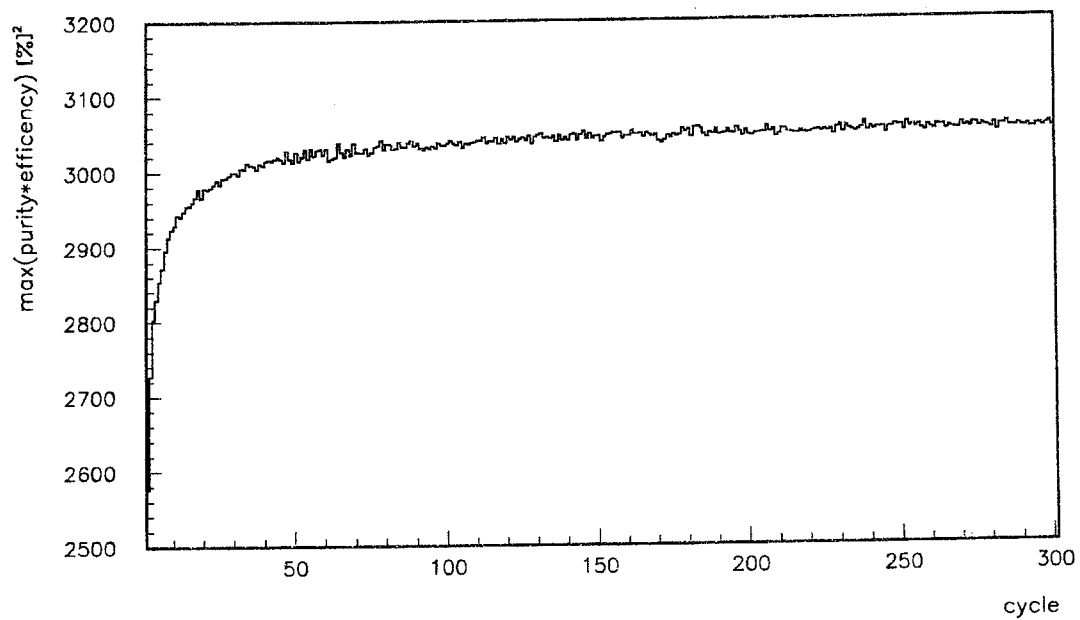


Figure 6.9: Though the product of *efficiency* · *purity* as a measure of the error rate is still improving after 300 cycles, the location of its maximal value is almost immediately stable at 0.5.

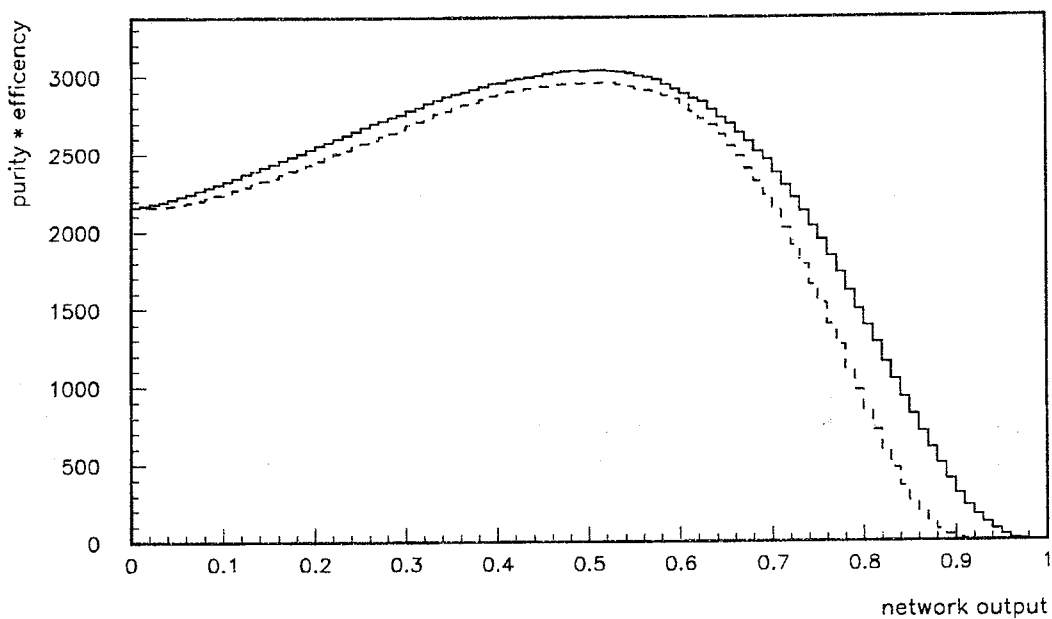
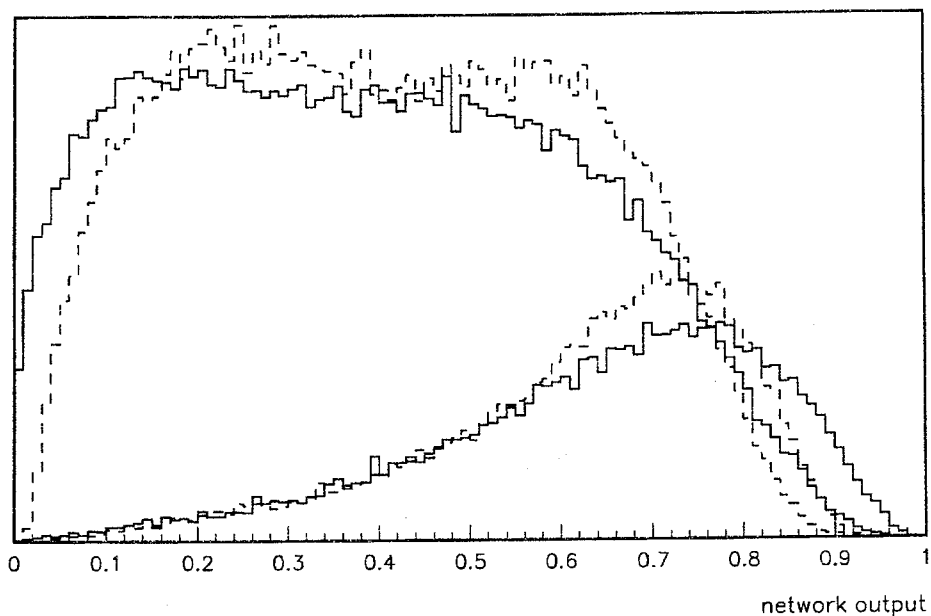


Figure 6.10: Though the output distributions for higher “temperatures” (dashed line, $T = 1000$) are less separated than for lower ones ($T = 1$), the error rate is always minimal at 0.5. Nevertheless the smaller separation yields a lower overall classification performance. Therefore the curve for the classifier trained at higher “temperature” is always below the one trained at lower “temperature”.

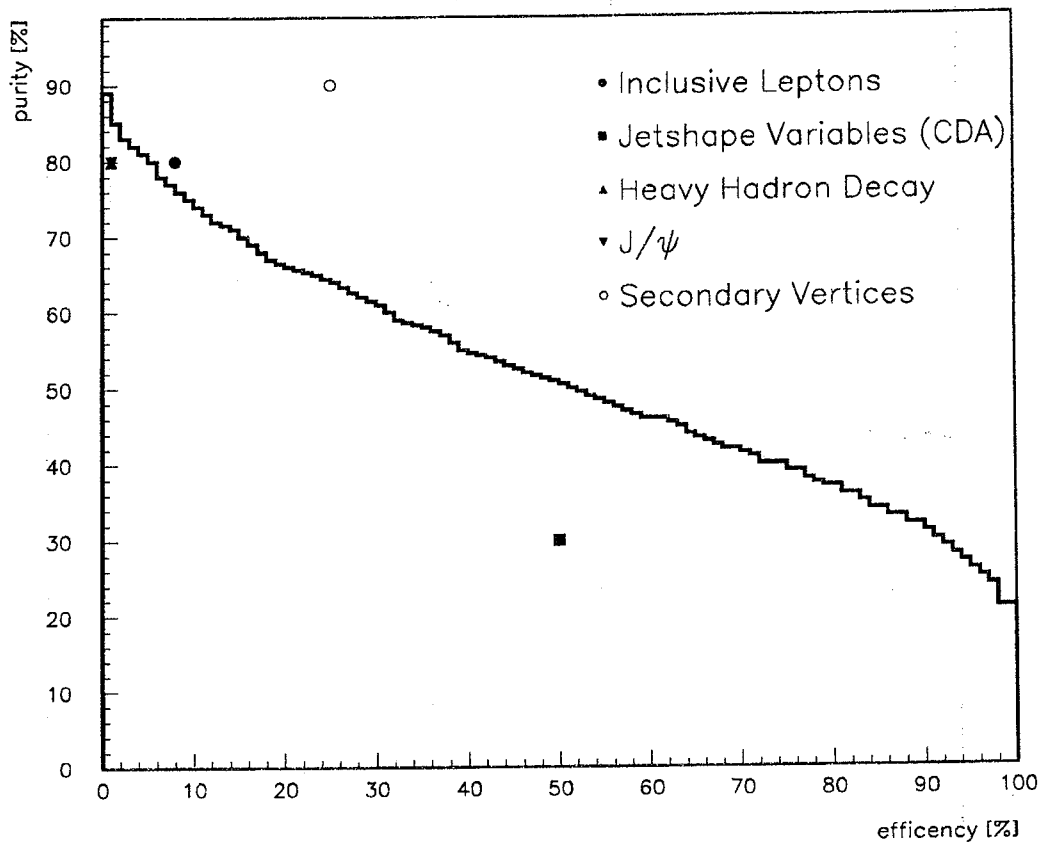


Figure 6.11: The possible combinations of sample purity and b efficiency, which can be achieved with the classifier used. The original hadron sample contains $\approx 20\%$ b events. The markers show the performances of conventional selection methods (Appendix D).

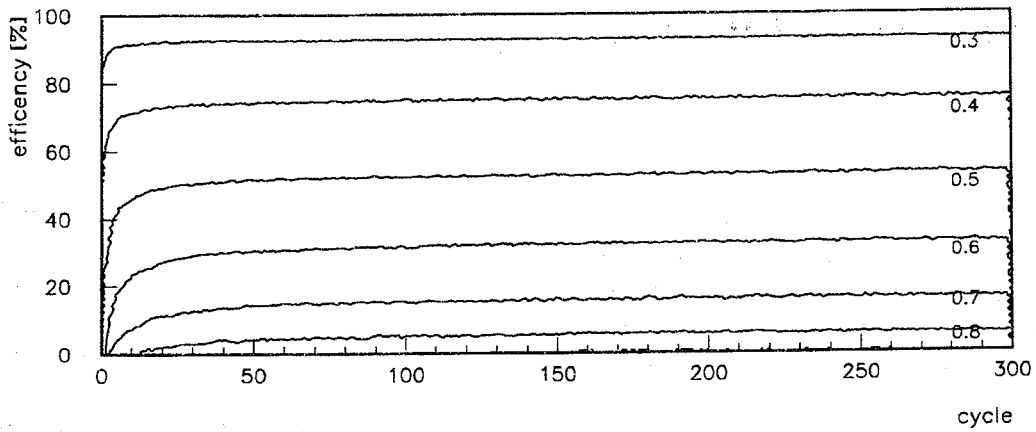


Figure 6.12: Classification performance of the classifier as function of training cycle. The numbers at the contour lines indicate the purity achieved at this efficiency.

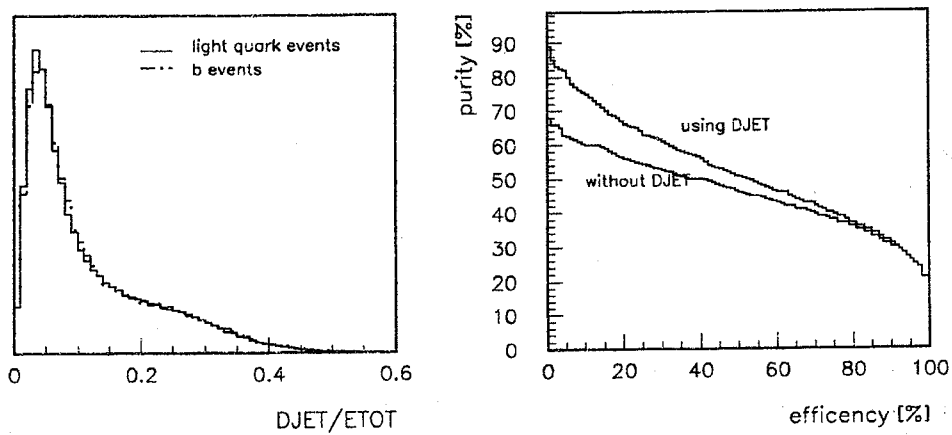


Figure 6.13: Though there is virtually no difference in the normalized **DJET** distributions of *b* and light quarks, the application of **DJET** yields a decisive improvement in the selection performance of the classifier.

Chapter 7

Determination of R_{bb}

In principle, the number of b events can now be determined by counting all hadron events, which yield a classifier output of at least 0.5, correcting this number for the classifier efficiencies for b and light quarks as well as for the detector acceptances. The cut at 0.5 would ensure the minimal possible number of misidentifications. The statistical error is influenced quite a bit by the uncertainty in the efficiencies though. Because the classifier suppresses more than 60 % of all light quark events and still 20 % of the b events, these uncertainties are intrinsically rather high.

In another approach, which exploits the whole underlying statistics, the trained classifier is used merely to map the multi-dimensional probability distribution of the event vectors \vec{x} onto the interval $[0, 1] \in \mathbb{R}$. Then a fit is performed to make up the output distribution out_{Data} , as obtained from the hadron data, of the two distributions for b and light quarks, out_b and out_x , as predicted from Monte-Carlo. out_b and out_x contribute just with the actual fraction of b and light quarks in the hadron sample.

$$out_{Data} = R_{bb} \cdot out_b + R_{xx} \cdot out_x = R_{bb} \cdot out_b + (1 - R_{bb}) \cdot out_x \quad (7.1)$$

Herein out_b and out_x are normalized to the data and still have to be corrected for the detector acceptances acc_b and acc_x , respectively. R_{bb} can be determined by minimizing the following χ^2 quantity using the CERN program MINUIT [47]. The output was divided into 100 bins to take into account properly the steep rise of the distributions at the edges of the interval.

$$\chi^2 = \sum_{i=1}^{100} \frac{(out_{Data}(i) - out_{MC}(i))^2}{\sigma_{Data}^2 + \sigma_{MC}^2} \quad (7.2)$$

assuming the error to be Gaussian distributed, i.e. $\sigma = \sqrt{n_i}$, and the normalized linear combination of the output distributions

$$out_{MC}(i) = \frac{acc_b \cdot R_{bb} \cdot out_b(i) + acc_x \cdot (1 - R_{bb}) \cdot out_x(i)}{acc_b \cdot R_{bb} + acc_x \cdot (1 - R_{bb})} \quad (7.3)$$

The normalization of out_{MC} takes into account, that the overall acceptance is dependent from R_{bb} and therefore cannot be determined from MC but has to be part of the fit itself. A loss in the acceptances cancels out exactly, if it is the same for b and non- b .

A useful side effect of the fit method consists in the resulting χ^2 , which constitutes a measure of the agreement between data and Monte-Carlo predictions.

Because of these advantages, the fit method has been given preference to the usage of the Functional-Link Net as a Bayes Classifier. A whole sample of approximately 900000 hadron

events collected on the Z^0 peak during 1991 and 1992 has been used in this analysis (Table 5.2). The output distributions were corrected with the detector acceptances listed in table 5.1. For each running period, the appropriate Monte-Carlo, describing the particular detector setup, was used to determine the estimated output for signal and background. The classifier was trained only once with the MC describing the L3 detector in the period 1991bc running at a cms-energy of 91.222 GeV, but is used for all three periods as well as for data taken at offpeak energies. This procedure is reasonable, because later the relevant MC was used to determine out_b and out_x for the deviant detector setups or energies, respectively. In order to simulate events for offpeak energies with sufficient statistics, the standard MC was used but the cluster energies were scaled with respect to the beam energy. Afterwards the various input variables for the classifier were calculated from the scaled clusters as described in chapter 6. The output distributions for MC and data agree very well for all periods and energies and are plotted in figures 7.1 and 7.2.

The results for all three running periods as well as for the offpeak energies are listed in table 7.1. The cross sections σ_{bb} have been calculated using the L3 results for the hadronic cross sections [41], [48]. Because the errors on the hadronic cross section are rather small compared to the ones of the b cross section, their effect can be neglected.

running period	\sqrt{s} [GeV]	R_{bb}	χ^2/dof	σ_{bb} [nb]
91a	91.254	0.224 ± 0.005	2.00	6.82 ± 0.14
	88.480	0.221 ± 0.021	1.25	1.14 ± 0.11
	89.470	0.205 ± 0.015	1.16	2.07 ± 0.15
	90.228	0.201 ± 0.012	1.24	3.64 ± 0.23
91bc	91.222	0.214 ± 0.005	1.64	6.48 ± 0.14
	91.967	0.216 ± 0.011	1.21	5.29 ± 0.26
	92.966	0.207 ± 0.014	1.32	2.97 ± 0.20
	93.716	0.206 ± 0.015	1.51	2.06 ± 0.15
92	91.294	0.220 ± 0.003	1.53	6.69 ± 0.09

$$\bar{R}_{bb} = 0.220 \pm 0.002$$

Table 7.1: The values of R_{bb} for the various energy points together with the χ^2 of the fit (7.2). \bar{R}_{bb} is calculated from the three values measured at the Z^0 peak, each weighted with its statistical error. The too small MINUIT estimate for the statistical error has already been corrected for (Chapter 8).

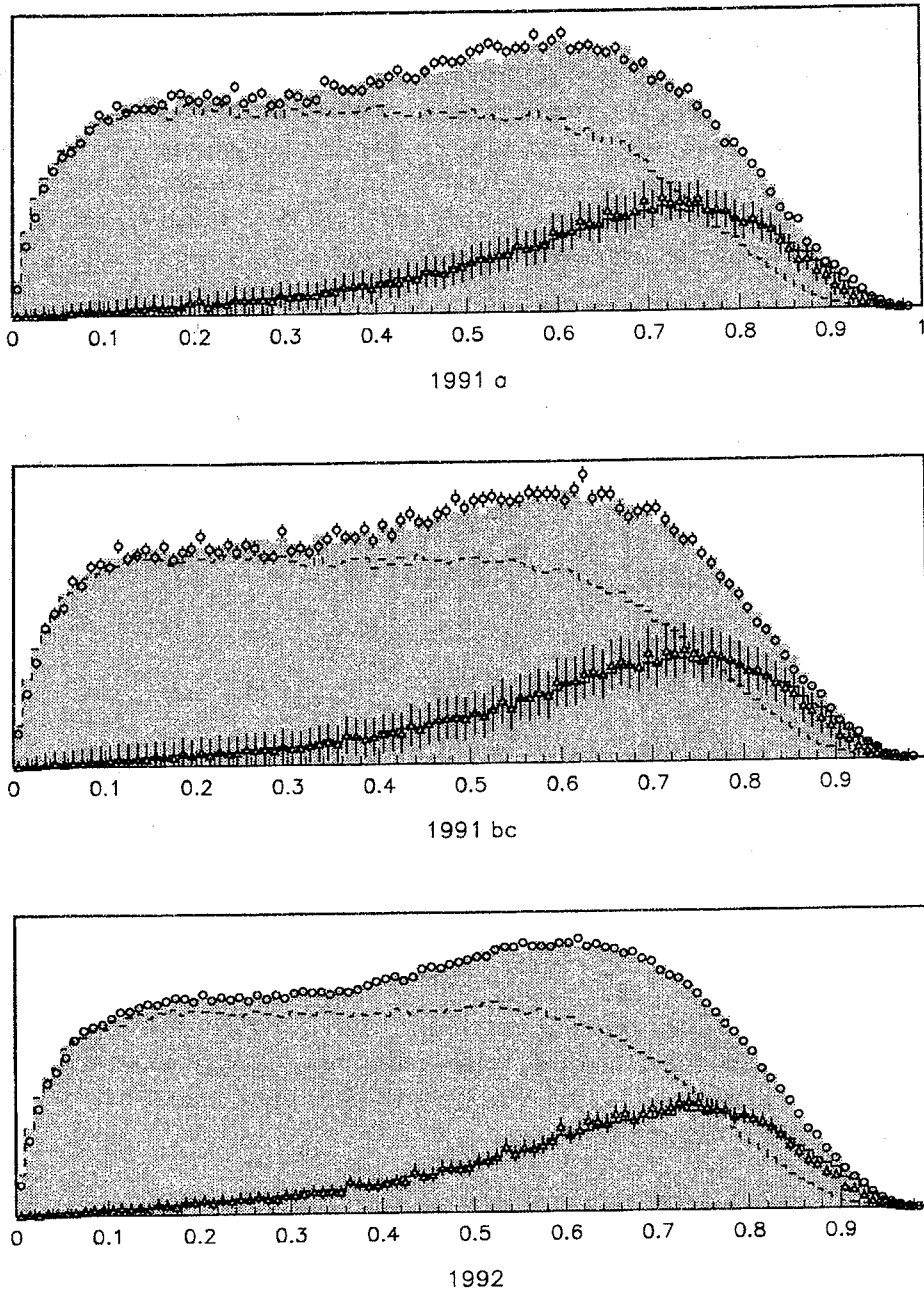


Figure 7.1: The bullets give the classifier output for the data taken on the Z^0 pole in 1991a, 1991bc, and 1992. The dashed line is the distribution for light quarks, the straight one for the b . The light shaded area shows the sum of the two MC distributions. The MC distribution for the b has been checked with inclusive μ events as was done in chapter 6.

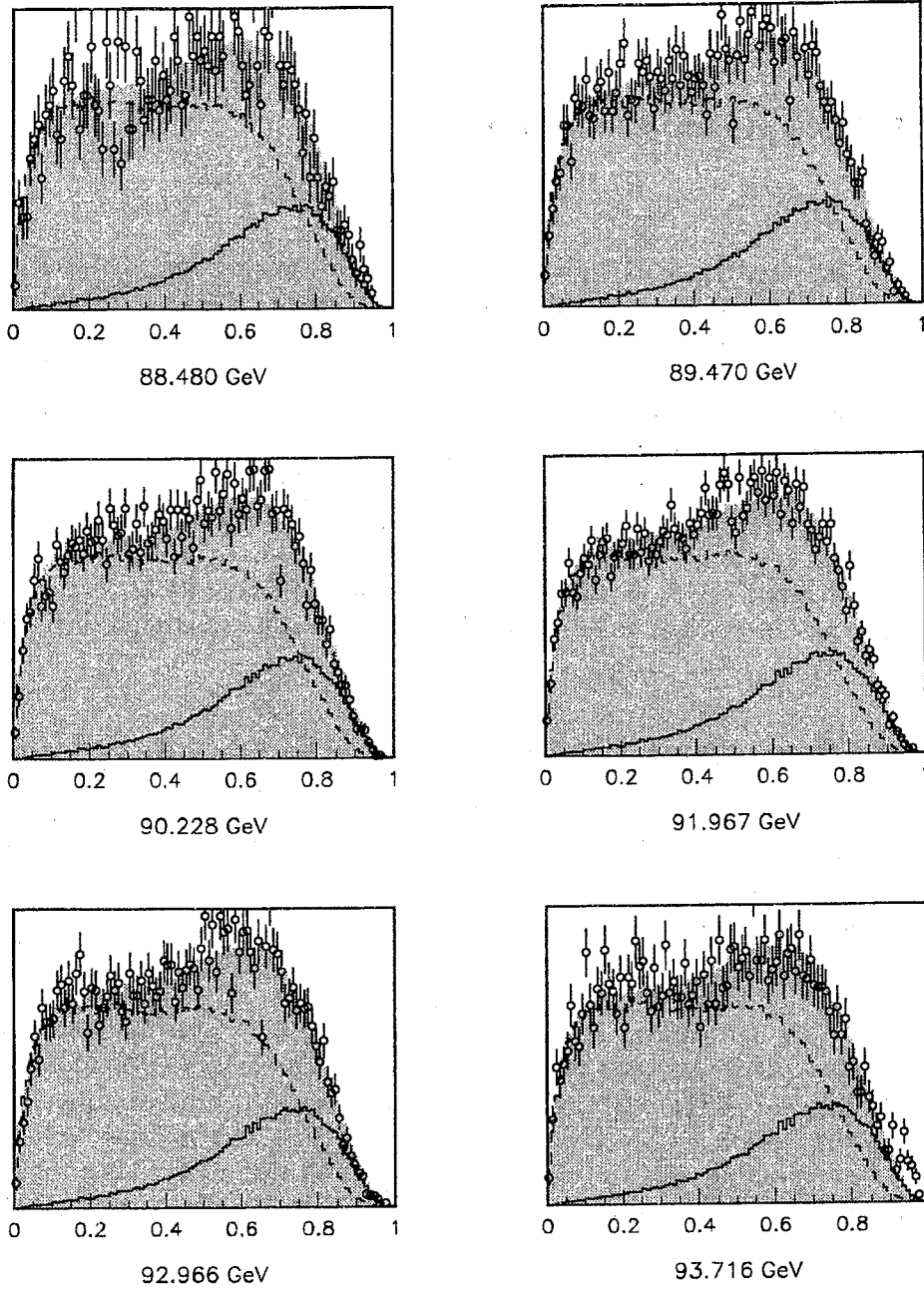


Figure 7.2: The classifier output for the data taken offpeak during 1991.

Chapter 8

Systematic Effects

Several sources might possibly contribute to the systematic error on R_{bb} . They can be divided into two categories. First, the applied analysis method, this is the Functional-Link Net and the fit to the output distributions, could already influence the result. Second, the errors in the theoretical predictions due to uncertainties in the Monte-Carlo parameters as well as the errors in the data due to limited detector resolution, will propagate into the final value of R_{bb} .

Except for the check concerning the systematic effect of the “temperature” to the fit result, all studies have been performed with the same set of weights, which has been used in the actual determination of R_{bb} in chapter 7. The impact of uncertainties in the input variables has rather been studied in spoiling $out_{MC}(\vec{x})$. Therefore MC events have been generated, where the input parameter of interest had been varied by $\pm 1 \sigma$. The half difference between the fit results obtained with each pair of extreme values allowed for the parameters is then taken as systematic error.

8.1 Effects from the classifier and the fit to the output

Several studies concerning the stability of the classifier with respect to the training period have already been presented in chapter 6 (Figures 6.6, 6.7, 6.9, 6.12). All in all, the system seems to have reached a more or less stable state after about 200 training cycle.

Two further checks have been performed to examine the dependency of the fit to the training period. As for the other stability studies, the weights obtained after every training cycle have been used in order to generate $out_{MC}(\vec{x})$ and $out_{Data}(\vec{x})$. The upper plot in figure 8.2 shows the result of the fit for data taken in the 91bc running period. The band indicates the 1σ error of the fit for ≈ 300000 MC and 90000 data events. The fit becomes stable after ≈ 250 cycles. Probably because of the slight disagreement in the output distributions for data and MC (Figure 7.1), the system settles down at a later time than for the other tests, which were done with MC only. To obtain the lower plot, half of the MC events have been taken as pseudo data, i.e. $out_{Data}(\vec{x})$ has been generated with MC instead of real data. Because the R_{bb} of a MC sample is of course well known, the reliability of the fit can be tested. Obviously the true value, indicated by the thick line, lies off more than 1σ of the fit result. For a more detailed study of this finding, the fit has been performed for 10 pairs of statistically independent out_{MC} and out_{Data} distributions. (out_{Data} generated by pseudo data.) The deviation of the true value and the fit result $(R_{bbtrue} - R_{bbFit})/\sigma_{Fit}$ was found to be Gaussian distributed around zero with $\sigma \approx 1.5$ (Figure 8.1). This number has been confirmed by an F -test [49] yielding a 98 % CL. Hence the assumption, that the errors in (7.2) would be Gaussian distributed, is not correct and

T	R_{bb}	χ^2
1	0.221 ± 0.003	89/99
100	0.221 ± 0.003	109/99
1000	0.221 ± 0.003	101/99

Table 8.1: The fit result is not dependent on the “temperature” T .

the number given by MINUIT was multiplied with a correction factor of 1.5.

As a last check concerning the fit and the classifier, the impact of the “temperature” T in the Fermi transfer function (3.9) has been examined. It was shown already (Figure 6.10) that the choice of T has no impact on the location of the optimal cut at the output, though a major one on the training speed (Figures B.1-B.6). In general, the shape of the output distributions is determined by T , which might yield some effects on the fit. The net has been trained for three “temperatures” yielding three different sets of weight. Then the fit was done with the output generated by these three sets. No effect has been found. The results of the fit are listed in table 8.1.

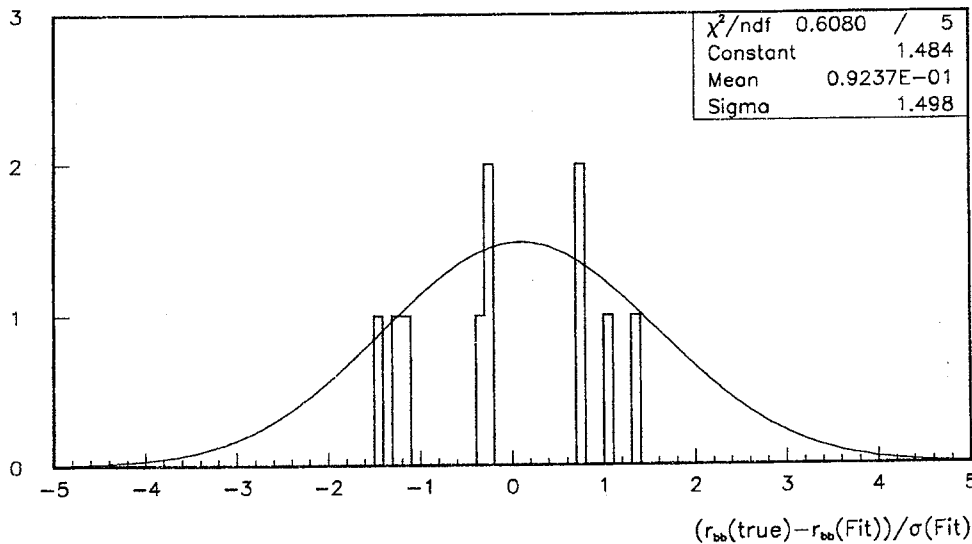


Figure 8.1: The deviation of the fit result for R_{bb} and the true value.

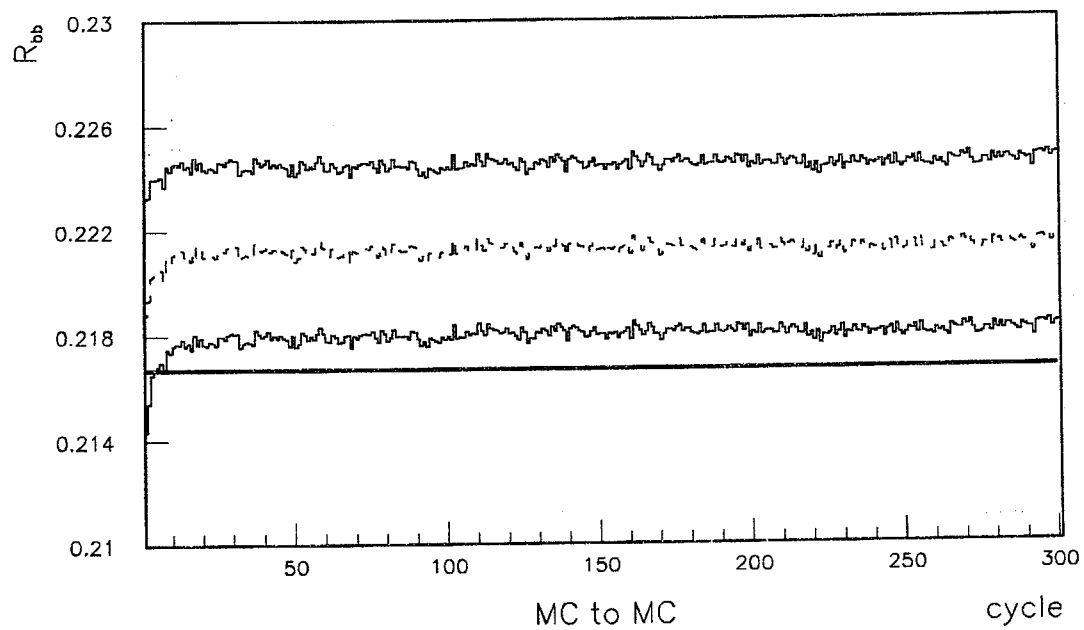
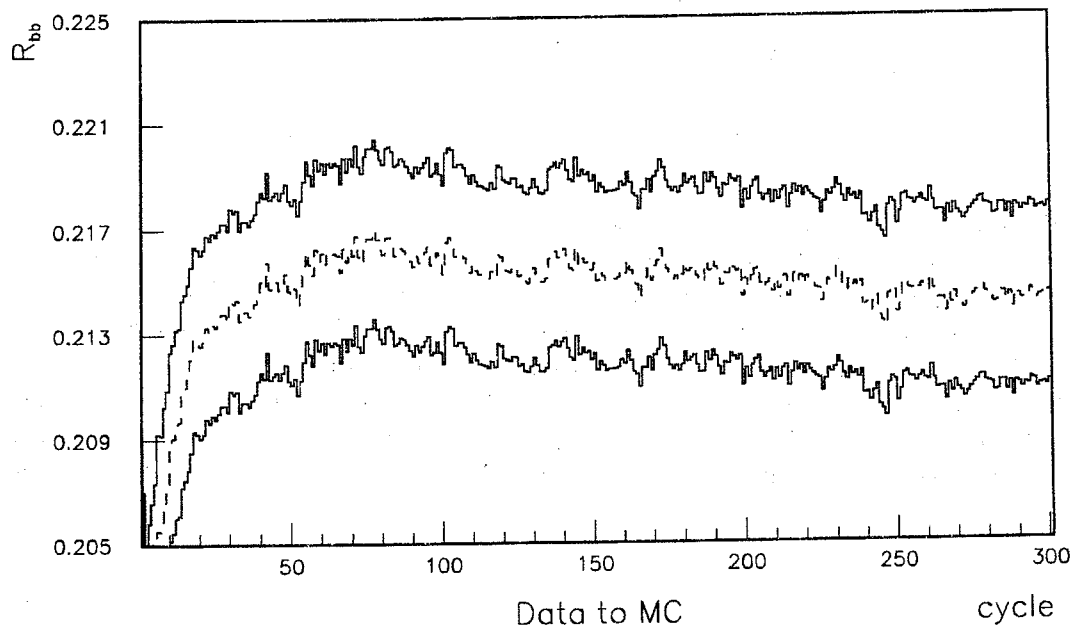


Figure 8.2: R_{bb} has been fitted from the output distributions for Monte-Carlo and data obtained after every training cycle, respectively. The upper plot shows the result for 91bc data. For the plot below MC was used as pseudo data. The band indicates the 1σ error of the fit, the thick line the true R_{bb} value of the pseudo data.

8.2 Uncertainties of the Monte-Carlo predictions and detector effects

The values of the shape variables serving as input to the classifier are determined by the energies and the angular distributions of the final particles in the process $Z^0 \rightarrow q\bar{q} \rightarrow \text{hadrons}$. This process is simulated by Monte Carlo programs according to four different steps

1. production of $q\bar{q}(\gamma)$ (electroweak)
2. gluon radiation (perturbative QCD)
3. hadronization of quarks and gluons (phenomenological models)
4. decay of unstable particles

In order to model the perturbative QCD the JETSET Monte-Carlo [39] program applies the *Parton Shower* method. This approach is based on an approximation of the full matrix element expressions. The version of the program used works within the framework of the Leading Log Approximation (LLA), i.e. only the leading terms in the perturbative expansion are kept.

Because QCD is not well understood at low energy scales, the fragmentation of colored quarks and gluons cannot be calculated by perturbative QCD. One needs to rely on phenomenological models. The separation between perturbative and fragmentation phase is generally characterized by an energy scale (Q_0) with a typical value of a few GeV . Several different fragmentation models have been developed. The one, which was found to describe the data in the best manner [50], is the *String Fragmentation* model [51]. The string model is derived from the QCD inspired idea that a color flux tube (string) is stretched between quark and anti-quark pairs, with gluons corresponding to kinks in the string. Particles are generated in the formalism of string breaking.

The following sources of systematic errors have been examined for the different stages of the MC simulation listed above

1. Because the mass of the c quark ($1.3 GeV$ [53]) is the closest to the mass of the b , a change in the fraction of c events with respect to all light quarks might also have some influence on the measured R_{bb} . The c contents of the MC sample was enriched or reduced at the expense of the lighter flavours respectively by the error given in [53].
2. The energy dependence of the strong coupling constant α_S is given bei Λ_{QCD} .

$$\alpha_S = \frac{12\pi}{(33 - 2N_f) \ln Q^2 / \Lambda_{QCD}^2}$$

with N_f the number of flavours and Q^2 the momentum transfer. Due to several ambiguities in the LLA description, especially in the renormalization, the parton shower scale parameter Λ_{LL} extracted from LLA models through comparison with data does not correspond exactly to the QCD scale parameter Λ_{QCD} . The value and the errors for Λ_{LL} used for the systematic check were taken from [50].

3. The process of hadronization is determined by several parameters.

- The fragmentation of light quarks is given by the LUND *symmetric function* [52]

$$f(z) \propto \frac{1}{z} (1-z)^a \exp\left(-\frac{b m_t^2}{z}\right)$$

where $z = (E + p_{||})_{had}/(E + p_{||})_{quark}$ is the fraction of energy taken away by the resulting hadron, m_t is the transverse mass of the system, and a and b are the fragmentation parameters.

- The fragmentation of the c and the b quark can be described by the *Peterson Fragmentation* function [54]

$$f(z) = \frac{N}{z} \left(1 - \frac{1}{z} - \frac{\epsilon_q^z}{1-z} \right)^{-2}$$

where N is a normalization constant and ϵ_q ($q = c, b$) are the fragmentation parameters for the heavy quarks.

- The transverse momentum spectrum p_t is described by the Gaussian function

$$f(p_t) \propto \exp\left(-\frac{p_t^2}{2 \cdot \sigma_q^2}\right)$$

where σ_q is a free parameter.

a , b , and σ_q were set to the values published in [50]. Due to the strong correlations between a and b , only b has been varied with $a \equiv 0.5$. The JETSET default values for ϵ_c and ϵ_b have been used and changed by the errors derived from [42].

4. The error in the branching ratio for the decay $b \rightarrow \mu \bar{\nu} X$ yields an uncertainty in the amount of the missing energy carried away by the neutrino. Similar to the check of Γ_{cc} the b fraction of the MC sample was enriched or reduced with inclusive muon events by the value given in [42] as error on the branching ratio.

The value of the b mass together with the error were taken from [53]. The parton shower termination parameter Q_0 was set to $Q_0 = 1 \text{ GeV}$.

The errors on R_{bb} , which are inflicted by the uncertainties in the listed parameters, have been determined as described in the prolog of this chapter. In order to distinguish a deviation due to systematic effects from statistical fluctuations a very high number of MC events is needed. Because a full blown detector simulation would require an immense amount of CPU time, the systematic effects caused by uncertainties in Λ_{LL} , σ_q , m_b , and all fragmentation parameters have been tested with a fast detector simulation. Herein events had been generated with the JETSET program and the energies and angular positions of the stable particles have been smeared according the detector resolutions. The checks concerning the branching ratios Γ_{cc} and $\Gamma_{b \rightarrow l}$ have been performed with full reconstructed MC and real data.

The impact of the finite detector resolution has been examined in the same way but (of course) with full detector simulation. The energy resolution of 13.2% [56] of the calorimeters has been worsened by using a different set of g-factors. The angular coordinates of the clusters have been smeared according a Gaussian with a σ of 10 *mrad*, the ones of the jet axis with a σ of $\approx 36.7 \text{ mrad}$. Because dead cells in the hadron calorimeter haven't taken into account during the reconstruction of the data at that time, their impact is estimated by setting to zero the hadronic part of 0.1% of all clusters.

For the tests concerning the QCD parameters, 200000 MC events have been generated with respectively one parameter varied by 1 σ in one direction. The checks for the systematic effects by Γ_{cc} , $\Gamma_{b \rightarrow l}$, and the lower energy resolution were based on 300000 MC events. 150000 MC events and 150000 pseudo data have been used to test the impact of the uncertainties in the angular position of clusters and jet axis and of the dead cells in the hadron calorimeter.

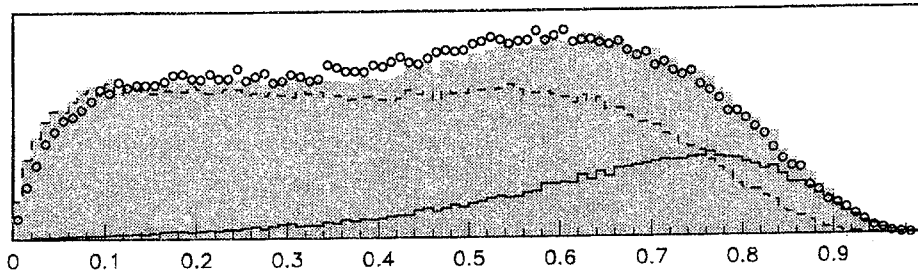


Figure 8.3: The agreement between data and the HERWIG Monte-Carlo is bad. The fit yields a $\chi^2 = 557$ for 99 degrees of freedom.

8.3 Different fragmentation models

Contrary to JETSET the HERWIG Monte-Carlo program [55] works with the *Cluster Fragmentation Model*. To estimate the impact of a different fragmentation model out_{MC} was generated using ≈ 170000 fully reconstructed HERWIG events. The agreement between the output for data and HERWIG Monte-Carlo is bad (Figure 8.3), the fit yields a $\chi^2 = 557$ for 99 degrees of freedom. Thus it had been abstained from including the deviation caused by the application of HERWIG model into the systematic error.

8.4 Summary

All in all, the biggest contributions to the systematic error are due to the uncertainties in Λ_{LL} and in the energy resolution. The first result has been found also by [57] and [58].

The final number for the error has been obtained by adding the deviations given in table 8.2 in quadrature. The error due to the application of a different fragmentation model (HERWIG) has been ignored. The number for the systematic error becomes eventually

$$\Delta_{R_{bb}} = \pm 0.009$$

Source of systematic error	Value	Variation	ΔR_{bb}
branching ratio $Z^0 \rightarrow c\bar{c}$	0.126	± 0.003	0.002
Λ_{LL} [GeV]	0.3	± 0.03	0.004
σ_q [GeV]	0.39	± 0.03	0.002
b [GeV ⁻²] (a=0.5)	0.76	± 0.08	0.003
ϵ_b	0.008	± 0.0024	0.003
ϵ_c	0.07	± 0.002	0.002
m_b [GeV]	5.0	± 0.5	0.000
branching ratio $b \rightarrow \mu\nu X$	0.113	± 0.012	0.001
energy resolution of the calorimeters	$\Delta E = 13.6\%$	$-10\% \cdot \Delta E$	0.005
angular position of jet axis		37 mrad	0.001
angular position of ASRC		10 mrad	0.001
dead cells in HCAL		0.1%	0.001
fragmentation model (HERWIG 5.1)			0.021 ($\chi^2 = 557/99$)
MC error			0.001

$$\Delta R_{bb} = \pm 0.009$$

Table 8.2: Systematic errors on R_{bb} . All parameters have been respectively modified by $\pm 1\sigma$ according to the errors given in the cited literature.

Chapter 9

Standard Model Parameters

The values of R_{bb} measured at the various energy points (Table 7.1) have been used to calculate several parameters of the Standard Model. Herein the FUNPLO program has been applied as an interface to ZFITTER 4.5 in order to perform the necessary fits. Both programs are described in appendix E. The fits take into account only the QED radiative corrections. This method is called *model-independent*, since no relations predicted by the SM are imposed between the measured quantities. This approach is possible because the radiative corrections can be separated into QED corrections and weak corrections (Equation (2.7)). The QED corrections, which depend on the acceptance on the detector and on cuts applied in the analysis, are always taken into account for calculating the theoretical predictions. Since the weak corrections cannot be calculated exactly outside the SM, these corrections are absorbed into the fitted parameters. These parameters are the *effective parameters* introduced in section 2.2.

9.1 Properties of the Z^0 and the partial width of $Z^0 \rightarrow b\bar{b}$

Already the cross sections for the b pair production alone can be used in order to constrain the basic properties of the Z^0 , including mass, total width and the partial decay width (2.6) into b . A model-independent fit with the cross sections measured at all energy points scanned during 1991 and 1992 as input, has been performed. The 1-parameter fit uses the values for m_Z and Γ_Z published in [41]. The results are collected in table 9.1. The one of major interest is the value of Γ_{bb} :

$$\Gamma_{bb} = 381 \pm 16 \text{ MeV}$$

	1 parameter	3 parameter
m_Z	(91.195 GeV)	$91.203 \pm 0.043 \text{ GeV}$
Γ_Z	(2.490 GeV)	$2.38 \pm 0.07 \text{ GeV}$
Γ_{bb}	$381 \pm 16 \text{ MeV}$	$351 \pm 24 \text{ MeV}$
	$\chi^2 = 6.9/8 \text{ dof}$	$\chi^2 = 4.4/6 \text{ dof}$

Table 9.1: Properties of the Z^0 obtained from fits to the cross section data of the b . The values in brackets have been fixed to the values published in [41].

Energy [GeV]	$\sigma_{bb}[nb]$	$A_{fb}^b \pm (\text{stat.}) \pm (\text{syst.})$
89.67	2.07 ± 0.15	$0.051 \pm 0.051 \pm 0.007$
91.24	$\left\{ \begin{array}{l} 6.82 \pm 0.14 \\ 6.48 \pm 0.14 \\ 6.69 \pm 0.09 \end{array} \right\}$	$0.097 \pm 0.017 \pm 0.007$
92.81	2.97 ± 0.20	$0.062 \pm 0.042 \pm 0.007$

Table 9.2: The numbers for the cross section and the forward-backward asymmetry applied in the fit to $g_{ab}g_{vb}$. The errors for the cross section are statistical only.

$G_\mu [GeV^{-2}]$	α	$\alpha_S [60]$	$g_{ae}[61]$	$g_{ve}[61]$	$\sin^2 \overline{\Theta}_W [62]$
$1.166388 \cdot 10^{-5}$	137.0359895^{-1}	0.117	-0.4996	-0.0360	0.2336

Table 9.3: The numbers for the SM parameters used in the fit. (The ZFITTER default values have been applied for G_μ and α .)

The error combines the statistical and the systematic error, as well as the energy error of LEP. The Higgs mass was fixed at $300 GeV$ and the top mass at $150 GeV$. The result improves the number given by the prior L3 measurement [42], which has used inclusive leptons to tag B events, by more than 30%. The prediction of the SM is $378 \pm 3 MeV$ [5], so that the result of the fit agrees within the measurement errors.

9.2 Weak coupling constants of the b quark

A model-independent 2-parameter fit has been performed using the b forward-backward asymmetry measured by L3 at three energies [41] and the corresponding b cross sections determined in this thesis. The values are given in table 9.2. The fit program assumes negative signs for both coupling constants by default, taking into account recent results from PETRA, PEP, and TRISTAN [59]. For the effective mixing angle the number published in [62] has been applied. The value for $\sin^2 \overline{\Theta}_W^b$ is about 0.4% higher for a top mass of $150 GeV$ due to weak corrections in the $Zb\bar{b}$ vertex and can be assumed equal within the errors. The results of the fit is given in table 9.4. The good agreement with the SM predictions confirms the existence of a weak isospin partner for the b quark, in other words the existence of the top.

	Fit	SM
g_{ab}	-0.518 ± 0.054	-0.50
g_{vb}	-0.323 ± 0.083	-0.34
$\chi^2 = 5.9/6 \text{ dof}$		

Table 9.4: The results for the coupling constants g_{ab} and g_{vb} . The correlation between the two variables is $-0.4273 \cdot 10^{-2}$.

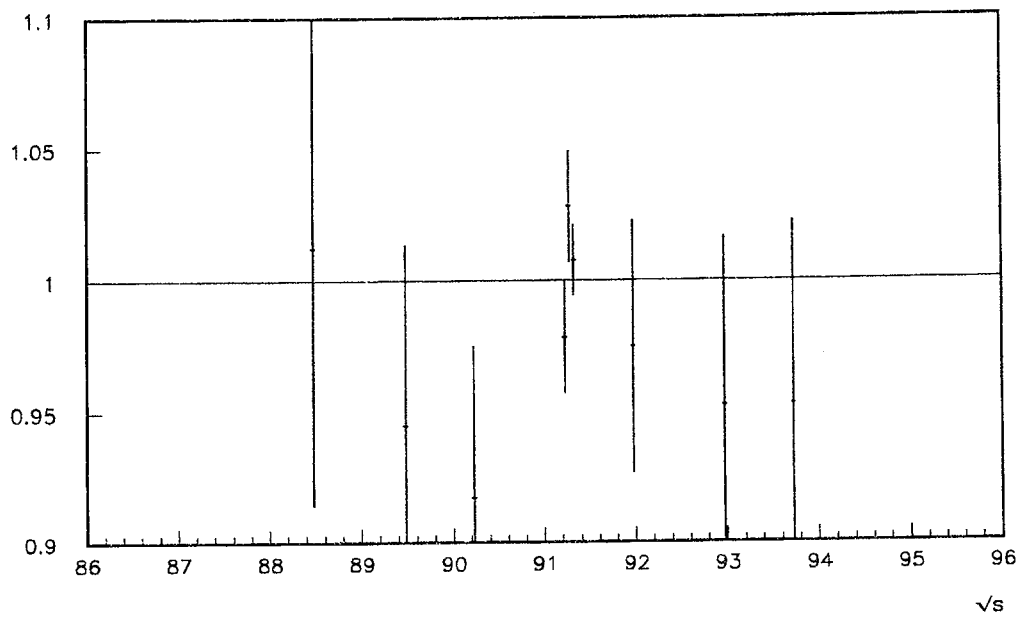
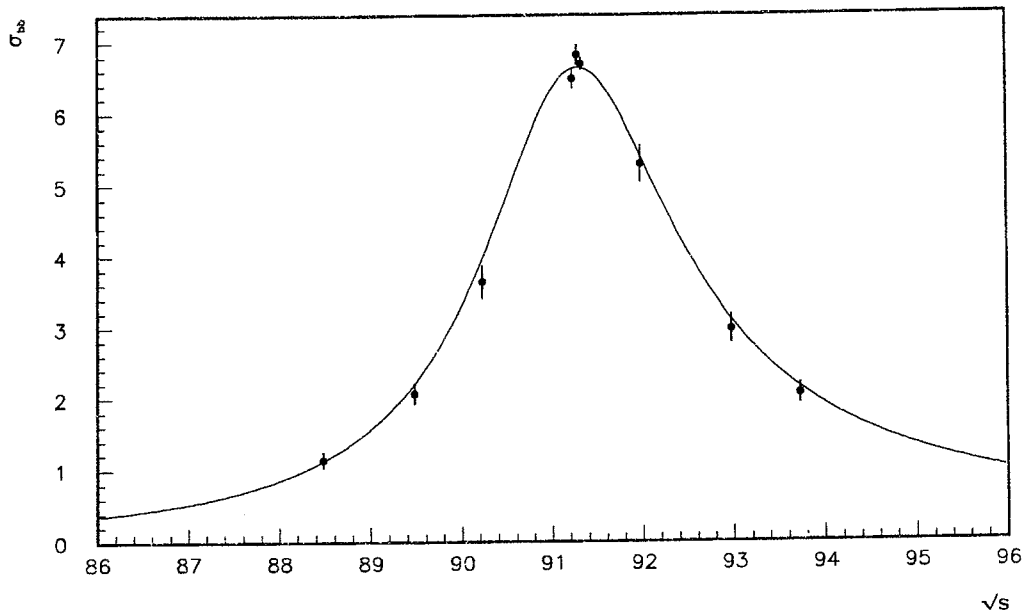


Figure 9.1: The b lineshape from the energies scanned in 1991 and 1992. The plot below shows the deviation from the theoretical forecasts. The errors include the statistical and the systematic error as well as the energy error of LEP [76].

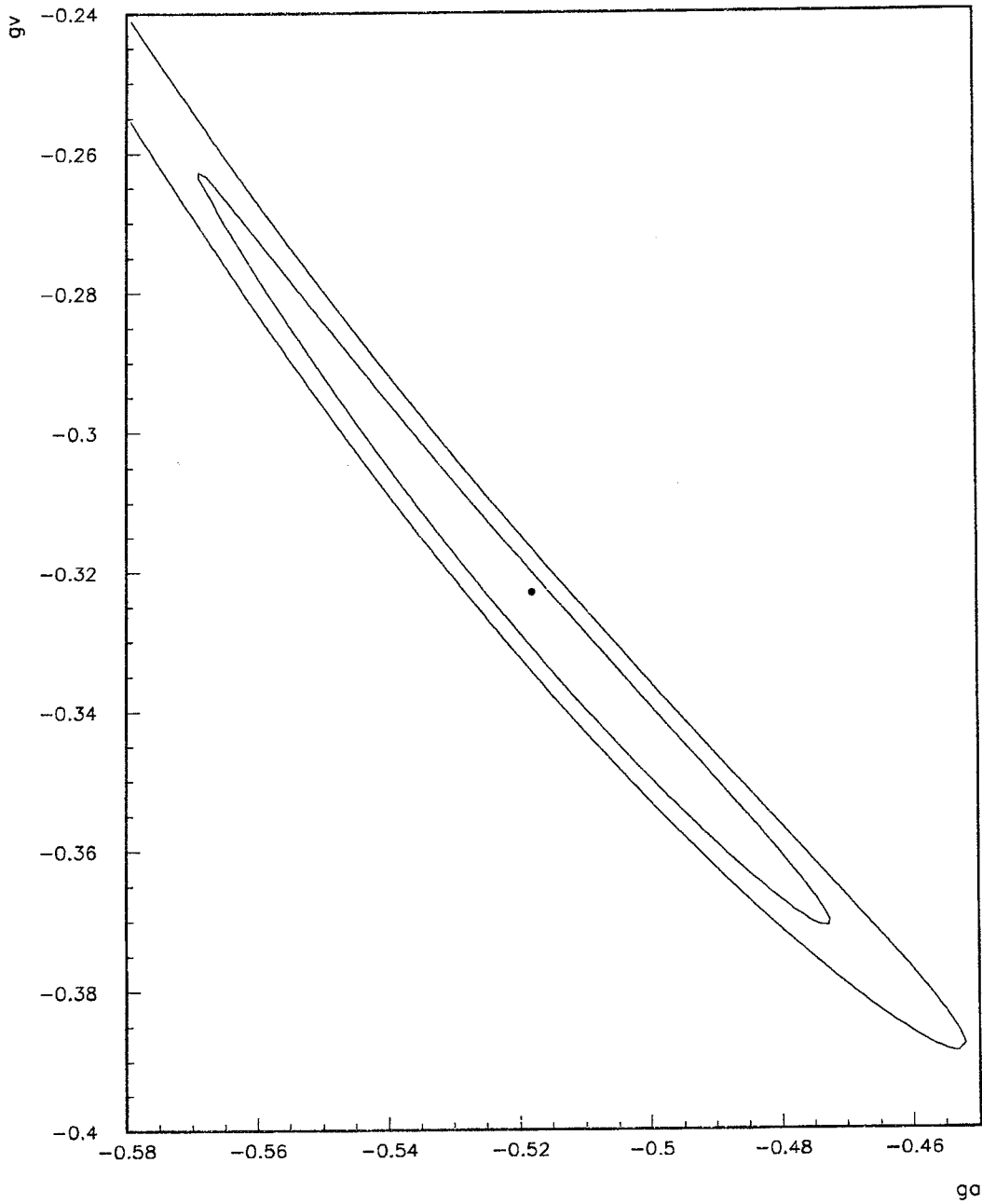


Figure 9.2: The weak coupling constants of the b quark. The contour shows the 68% and 95% confidence level of the fit result.

9.3 Top mass

The mean value for $R_{bb} = 0.22$ as determined in this thesis is correlated with a top mass already been excluded by the Tevatron experiments [1]. But at least an estimate for an upper limit can be given

$$m_{top} < 375 \text{ GeV} \quad (95\% \text{ CL})$$

Herein the Tevatron exclusion has been considered by taking as upper limit the value, where the integral of the residual distribution above 108 GeV becomes 95 % [53].

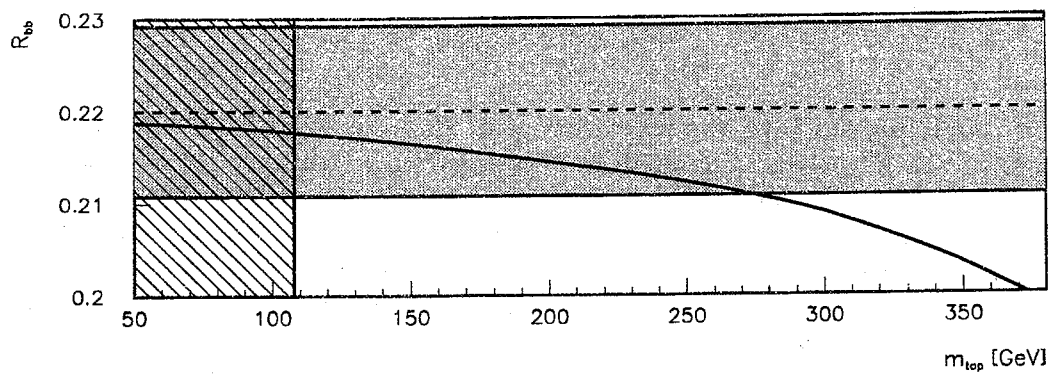


Figure 9.3: R_{bb} as a function of the top mass. The curve shows the prediction by the SM. The number for R_{bb} determined in this work is indicated by the dashed line. The grey band is the 1σ error. The region below 108 GeV has already been excluded by the Tevatron experiments.

Chapter 10

Summary and Conclusion

The subject of this thesis has been the measurement of R_{bb} , the fraction of b quarks in all hadronic decays of the Z^0 boson. The analysis is founded on a sample of 900000 hadrons, collected with the L3 detector during the years 1991 and 1992. One event shape and 5 jet shape variables, calculated respectively for the first two jets, have been used for the selection of b events. These quantities served as input to a Functional-Link Net. In general, the output distributions of a Neural Network allow for the best possible separation of signal and background events for a given set of discriminating variables. This behaviour has been exploited to fit the net output for the hadron data to the Monte-Carlo distributions for b and light quarks in order to determine the b fraction.

R_{bb} has been measured at 9 energy points, in particular three on the Z^0 pole. The result obtained, averaging only the values on the Z^0 peak, is

$$R_{bb} = 0.220 \pm 0.002 \text{ (stat.)} \pm 0.009 \text{ (syst.)}$$

This result is in good agreement with the average of all four LEP experiments (Figure 10.2), but too high for a reasonable top mass (Figure 9.3). [63] have suggested additional $Zb\bar{b}$ vertex corrections in the frame of the *Minimal Supersymmetric Model* by light charginos and an s-top (Figure 10.1) with their masses close to the experimental bounds [64]. These contributions would increase the value of R_{bb} . If this is the case then the chargino and the s-top should be visible at LEP200.



Figure 10.1: Vertex corrections due to 'New Physics' [63]. The left process yields a negative contribution to R_{bb} , the right one a positive.

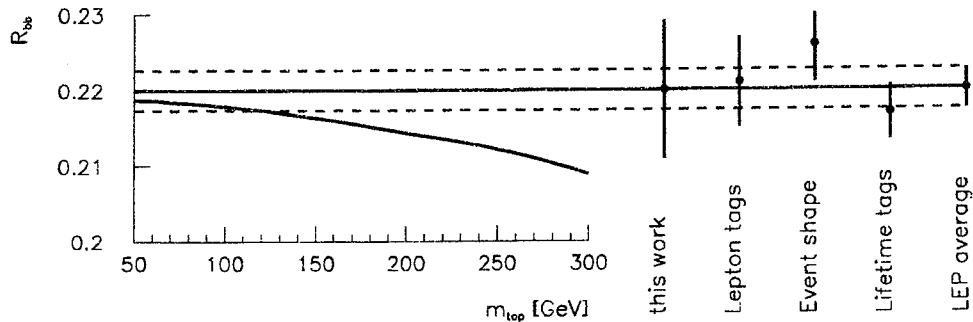


Figure 10.2: LEP averages for R_{bb}

When the output of the net is taken as discriminant to select b events, the b fraction of the hadron sample can be enhanced from originally 20 % up to 52 % with still a 50 % b efficiency. This is a 75% improvement compared to algorithms commonly applied in multi-variant analysis, eg. the ‘Canonical Discriminant Analysis’ [68]. Purities up to 80 % can be achieved with about 8 % efficiency, a number comparable to lepton tagging.

The b cross sections for the 9 points were used as input to a model-independent fit to the partial width of the decay $Z^0 \rightarrow b\bar{b}$

$$\Gamma_{bb} = 381 \pm 16 \text{ MeV}$$

The error improves the number given by the prior L3 measurement [41], which has used inclusive leptons to tag b events, by 30 %.

Together with the L3 results for the forward-backward asymmetry of the b quark, the weak coupling constants g_{ab} and g_{vb} have been fitted.

$$g_{ab} = -0.518 \pm 0.054$$

$$g_{vb} = -0.323 \pm 0.083$$

This result proves the existence of an isospin partner for the b quark, with other words the existence of the top.

The error on R_{bb} is still rather high, so that the upper limit for the top mass [65] cannot be improved. With the result obtained for R_{bb} one gets

$$m_{top} < 375 \text{ GeV} \quad (95\% \text{ CL})$$

The application of a Neural Network for b selection yielded a notable improvement to previous measurements based on lepton tagging [42]. Even when the input data were rather noisy due to too many low energetic clusters in the calorimeters, the events were still classified in a correct manner. Furthermore the net was able to exploit correlations between the event variables, allowing for use of variables with no difference between signal and background in their 1-dimensional distributions.

But further progress in the determination of R_{bb} is obviously limited by the large uncertainties in the QCD parameters (Chapter 8). Therefore future work in this field should concentrate first on a more precise determination of these numbers. As another approach, the combination of lepton tagging with event shape analysis [57] provides a possibility of a less model dependent analysis.

Appendix A

The Rummelhart Algorithm for Multi-layered Neural Networks

The concept of a multi-layered net of Perceptrons as processing units was already known in 1965 [27], but it was only in 1986 that a learning algorithm was found [66]. Because the knowledge of the concept is needed for the understanding of the tests discussed in the next appendices, the Rummelhart Algorithm will be outlined in the following.

The algorithm of Rumelhart et al. treats multi-layered networks with any number of hidden layers and any number of output nodes. It is a generalisation of the Widrow-Hoff Learning, i.e. it is founded on the minimization of a mean-square error measure. For m output nodes the

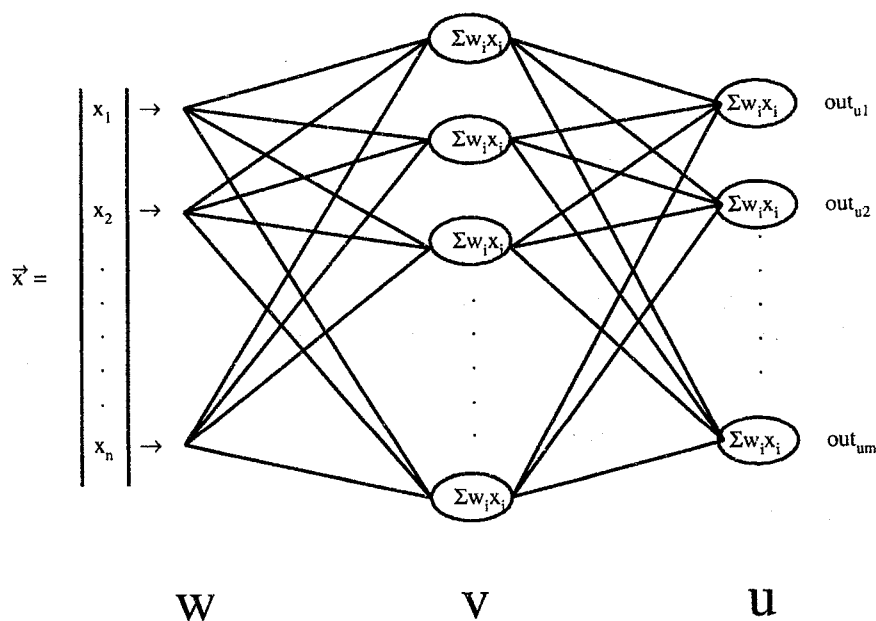


Figure A.1: The architecture of the network used in the derivation.

definition (3.7) can be generalized

$$err = \frac{1}{2} \sum_{k=1}^t \sum_{i=1}^m (\omega_{ki} - out(\vec{x}_k)_i)^2 \quad (A.1)$$

with the $(\vec{x}_k, \vec{\omega}_k)$ drawn from a training sample $\mathcal{T} = \{(\vec{x}, \vec{\omega})_k | k = 1 \dots t\}$ and $out(\vec{x}_k)_i$ the output of the i -th node in the output layer as computed with the actual set of weights $\vec{\omega}$. The regarded net (Figure A.1) shall consist of an input layer w , one hidden layer v , and an output layer u . (The final result can be generalized immediately to any number of hidden layers.) Then the input to and output of a node in layer v is

$$net_v = \sum w_{vw} out_w \quad out_v = f(net_v)$$

for any transfer function f . The same holds for the nodes in layer u

$$net_u = \sum w_{uv} out_v \quad out_u = f(net_u) \quad (A.2)$$

For each single presentation the weights of a node in layer u shall be changed proportional to the gradient of (A.1)

$$\Delta w_{uv} = -\eta \frac{\partial err}{\partial w_{uv}}$$

The error is expressed in terms of the outputs out_u , which itself is a nonlinear function of the input net_u . Thus the partial derivate $\frac{\partial err}{\partial w_{uv}}$ can be evaluated using the chain rule

$$\frac{\partial err}{\partial w_{uv}} = \frac{\partial err}{\partial net_u} \frac{\partial net_u}{\partial w_{uv}}$$

Using (A.2), we obtain

$$\frac{\partial net_u}{\partial w_{uv}} = \frac{\partial}{\partial w_{uv}} \sum w_{uv} out_v = out_v$$

We now define

$$\delta_u = -\frac{\partial err}{\partial net_u}$$

and write

$$\Delta w_{uv} = \eta \delta_u out_v \quad (A.3)$$

similar to (3.10). To compute δ_u , we use the chain rule to express the partial derivate in terms of two factors, one expressing the rate of change of error with respect to the output out_u , and the other expressing the rate of change of out_u with respect to the input to that same node. That is, we have

$$\delta_u = -\frac{\partial err}{\partial net_u} = -\frac{\partial err}{\partial out_u} \frac{\partial out_u}{\partial net_u}$$

The two factors are obtained as follows

$$\frac{\partial err}{\partial out_u} = -(\omega_u - out_u)$$

and

$$\frac{\partial out_u}{\partial net_u} = f'(net_u)$$

From which we obtain for δ_u

$$\delta_u = -(\omega_u - out_u) f'(net_u)$$

for any output layer node u , and we have for (A.3)

$$\Delta w_{uv} = \eta (w_u - out_u) f'(net_u) out_v$$

Circumstances are different if the weights do not affect output nodes directly. We still make the ansatz:

$$\begin{aligned} \Delta w_{vw} &= -\eta \frac{\partial err}{\partial w_{vw}} \\ &= -\eta \frac{\partial err}{\partial net_w} \frac{\partial net_w}{\partial w_{vw}} \\ &= -\eta \frac{\partial err}{\partial net_w} out_w \\ &= \eta \left(-\frac{\partial err}{\partial out_v} \frac{\partial out_v}{\partial net_v} \right) out_w \\ &= \eta \left(-\frac{\partial err}{\partial out_v} \right) f'(net_v) out_w = \eta \delta_v out_w \end{aligned}$$

However, the factor $\partial err / \partial out_v$ cannot be evaluated directly. Instead, we write in terms of quantities that are known and other quantities that can be evaluated. In particular, we write for m nodes in the output layer

$$\begin{aligned} -\frac{\partial err}{\partial out_v} &= -\sum_u \frac{\partial err}{\partial net_u} \frac{\partial net_u}{\partial out_v} \\ &= \sum_u \left(-\frac{\partial err}{\partial net_u} \right) \frac{\partial}{\partial out_v} \sum_m w_{um} out_v \\ &= \sum_u \left(-\frac{\partial err}{\partial net_u} \right) w_{uv} = \sum_u \delta_u w_{uv} \end{aligned}$$

We see that, in this case

$$\delta_v = f'_v(net_v) \sum_u \delta_u w_{uv}$$

That is, the deltas at an internal nodes can be evaluated of the deltas at an upper layer, they are *backpropagated*.

Summarizing, and using the index t to denote the event number from the training sample, we have for the α -th weight of node β

$$\Delta_t w_{\beta\alpha} = \eta \delta_{t\beta} out_{t\alpha} \quad (\text{A.4})$$

If the nodes β are output layer nodes, we have

$$\delta_{t\beta} = (w_{t\beta} - out(\vec{x}_t)_\beta) f'_\beta(net_{t\beta}) \quad (\text{A.5})$$

However, if the nodes β are internal nodes, we need to evaluate $\delta_{t\beta}$ in terms of the deltas at a upper layer γ ; that is

$$\delta_{t\beta} = f'_\beta(net_{t\beta}) \sum_m \delta_{t\gamma} w_{\gamma\beta} \quad (\text{A.6})$$

For a Fermi function (3.9) as transfer function, it holds

$$\frac{\partial out_\beta}{\partial net_\beta} = out_\beta (1 - out_\beta)$$

Now for the nodes in the output layer (A.5) simplifies to

$$(A.5) \rightarrow \delta_t \beta = (\omega_{t\beta} - out(\vec{x}_t)_\beta) out_{t\beta}(1 - out_{t\beta}) \quad (A.7)$$

and for the inner nodes (A.6) becomes

$$(A.6) \rightarrow \delta_t \beta = out_{t\beta}(1 - out_{t\beta}) \sum_m \delta_{t\gamma} w_{\gamma\beta} \quad (A.8)$$

Appendix B

Performance of a Multilayered Neural Network

The behaviour of a Feedforward-Backpropagation Net is determined by a rather large number of parameters, in particular, the number of hidden layers, the number of nodes in each hidden layer, the transfer functions of the units, and the “temperature” T . The training might be influenced by the initialization of the weights, the applied error measure, and the parameters arising in the formulas given by Rumelhard (Appendix A), hence the η , the “temperature” T , and, if a damping term (3.12) has been used, the μ .

To gain a first impression of the behaviour of the Feedforward-Backpropagation Net, several tests were done for a toy problem. In particular, the effect of several parameters on the training performance has been studied. The toy problem consisted in the learning of the XOR function (Table B.1). The XOR problem has become famous, after Minsky and Papert could show [67], that it is not solvable by a single Perceptron but that one hidden layer with at least two nodes is needed [27]. Such a net is able to solve the task exactly. Therefore the XOR problem is well suited to study the performance of a Neural Net, because the end of the training is well defined, i.e. after all four states have been learned.

The Feedforward-Backpropagation Net, which has been studied, possessed one hidden layer and used a Fermi Function (3.9) as transfer function. If not explicitly notified, the error measure has been the mean-square (3.7). The parameters η , μ , the “temperature” T , and the number of nodes n in the hidden layer have been varied within the range tabulated in B.2. The net was trained on a Cray X-MP/48 for every combination of η , μ , T , and n until either the problem had been learned completely or 250 sweeps¹ had been reached. After the optimal combination

¹A *sweep* is a presentation of one single training pattern.

x_1	x_2	$out(x_1, x_2)$
0	0	1
0	1	0
1	0	0
1	1	1

Table B.1: Definition of the XOR function

Parameter	Range	Stepsize
η	0.1...5	0.1
μ	0.1...1	0.1
T	0.1...5	0.1
n	2...24	1

Table B.2: The parameters of the studied Feedforward-Backpropagation Net.

Configuration	n	T	η	μ	Numbers of training sweeps
—	18	0.3	2.6	0.5	70
different initialization	20	0.4	3.5	0.5	82
entropy error	20	0.3	0.9	0.8	28

Table B.3: The optimal parameters for the three configurations studied.

of parameters had been found, the training was repeated, now varying only one parameter and leaving the other three fixed at their optimal value, in order to study the vicinity of the optimal parameter setting. The parameters, which yielded the optimal performance are listed in table B.3. Figures B.1-B.4 show the performances for two different initializations of the weights, figures B.5 and B.6 the one for a training, which applied the entropy error (3.8). The entropy error yielded the best performance (28 sweeps), though only for a very small parameter region. The result that a rather large number of internal nodes yields a faster training than a small number with less weights to be determined, comes a bit as a surprise. Another curious point is the rather large value of the optimal η . The reason therefore can probably be found in the small number of states, which had to be learned.

These studies show that looking for the one parameter set, which yields the optimal training performance, requires a big effort and reminds on alchemy. The only promising way to obtain a fast training seems to be to keep the number of weights small. This is the case for the application of a Functional-Link Net. In spite of the expansion of the input vector (Figure 6.5), the complexity of the system still remains small compared to the one arising in a typical Feedforward-Backpropagation Net. No studies on the XOR problem have been done with a Functional-Link Net, but appendix C shows that for a rather complicated problem the training performance of a Functional-Link Net is superior to the one of a Feedforward-Backpropagation Net.

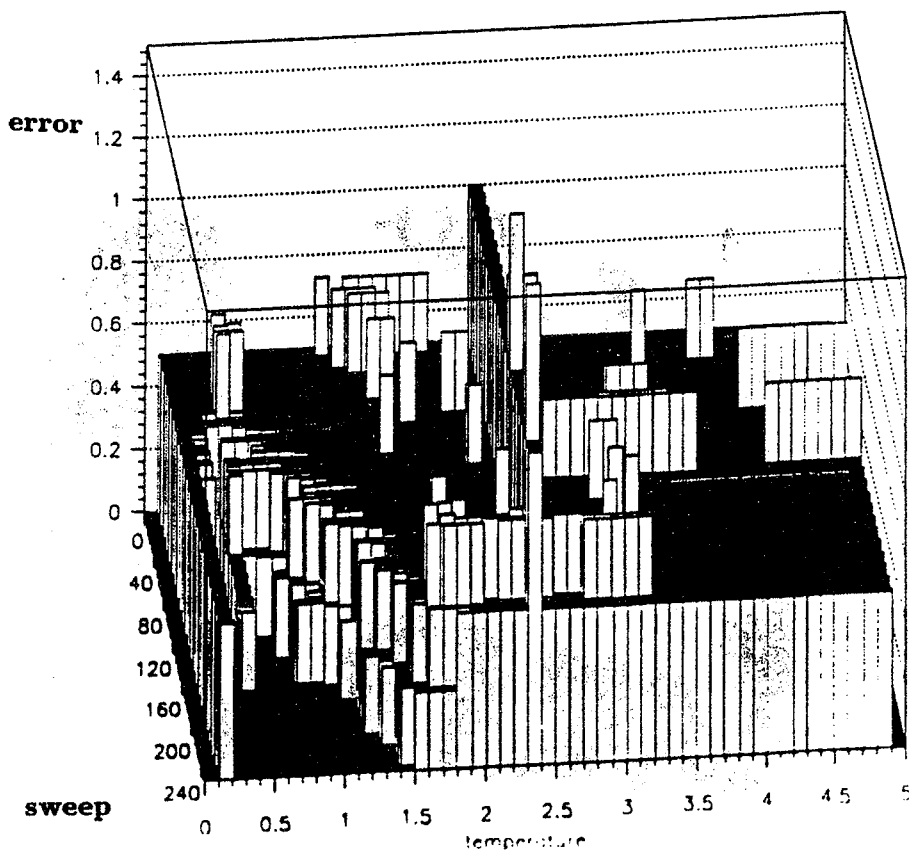
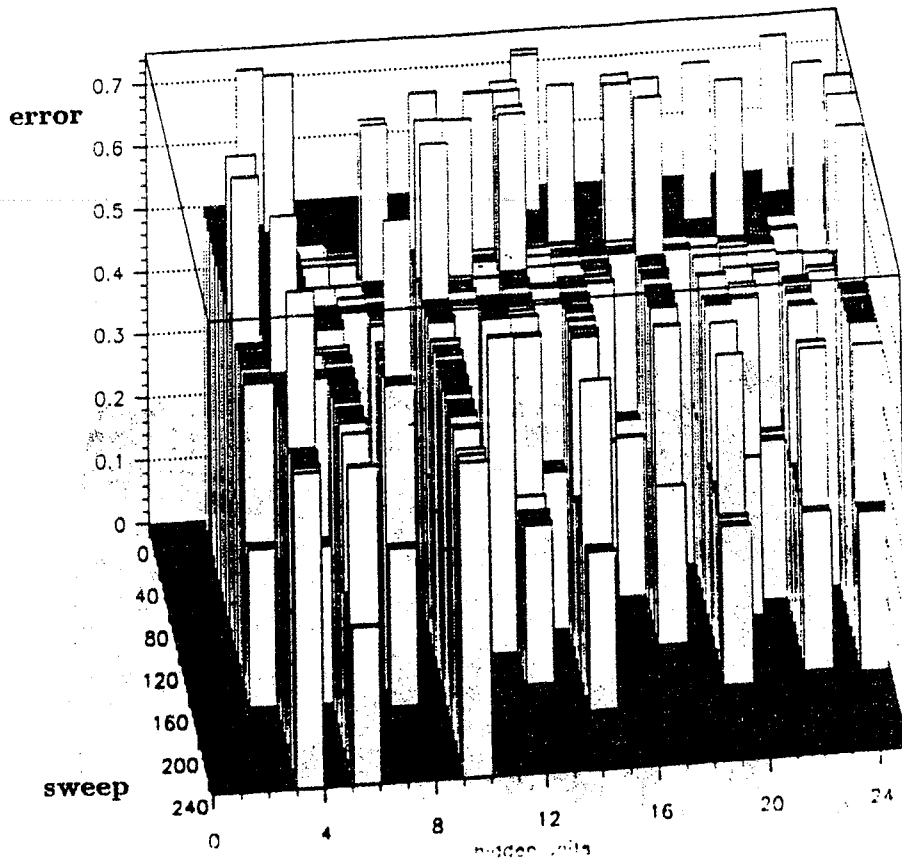


Figure B.1: The plots show the training performance of the net. The number of the nodes in the hidden layer n (upper plot) and the "temperature" T (below) have been varied, respectively, with the other three parameters kept at their optimal setting.

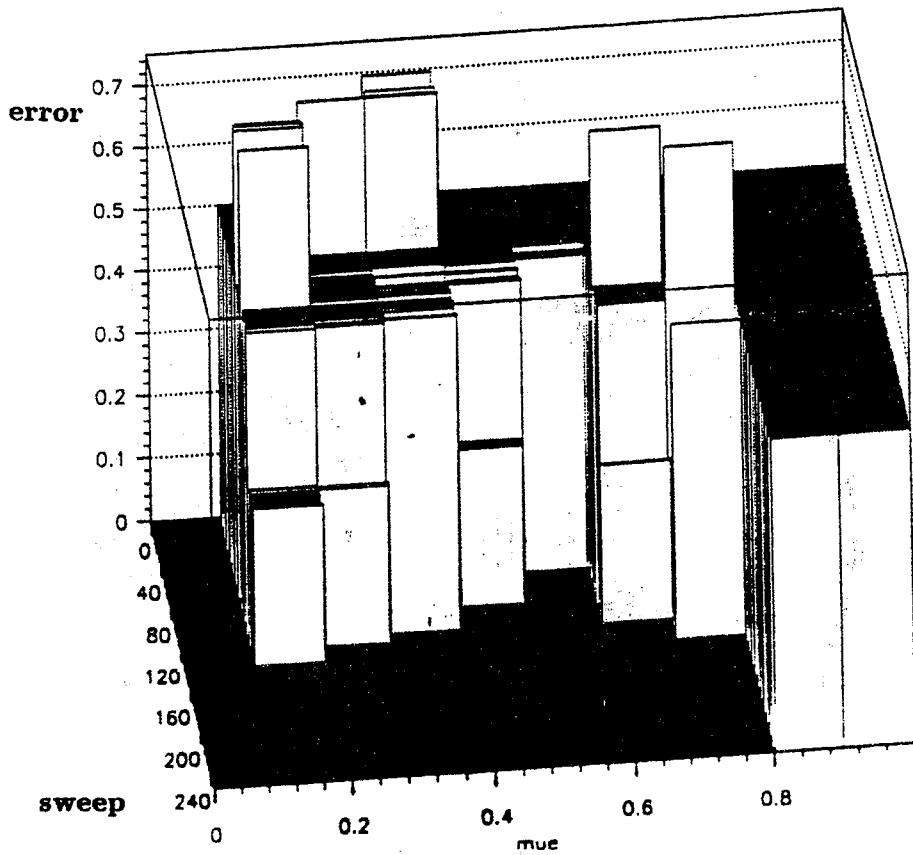
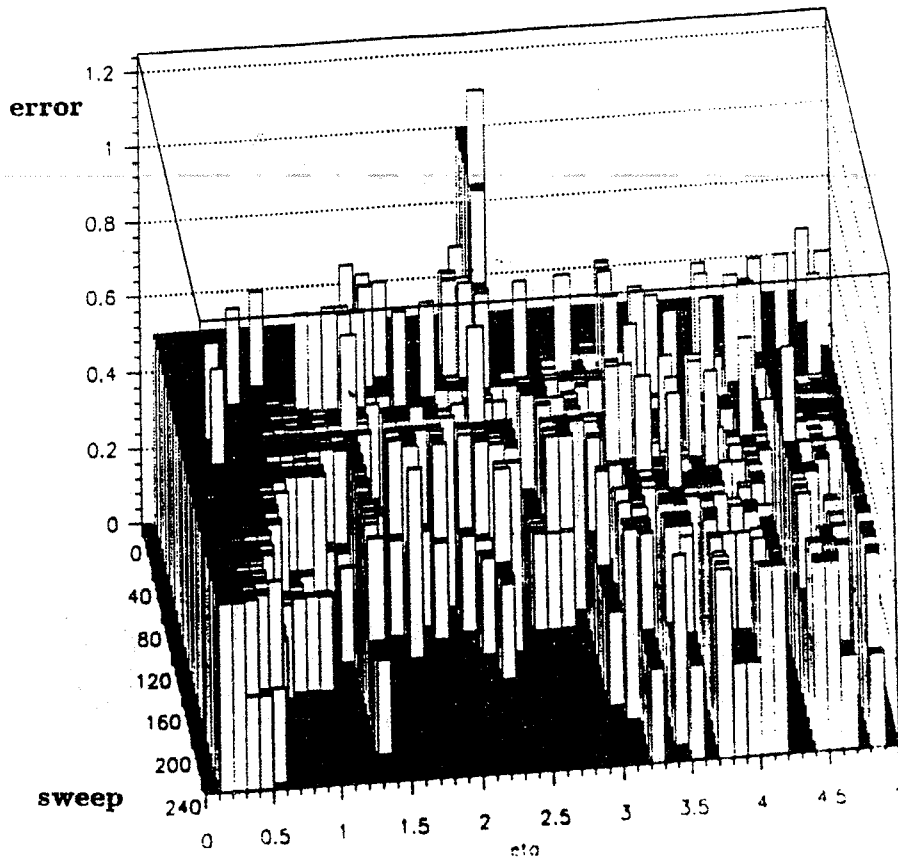


Figure B.2: The same situation as in figure B.1, but now η (upper plot) and μ (below) have been varied.

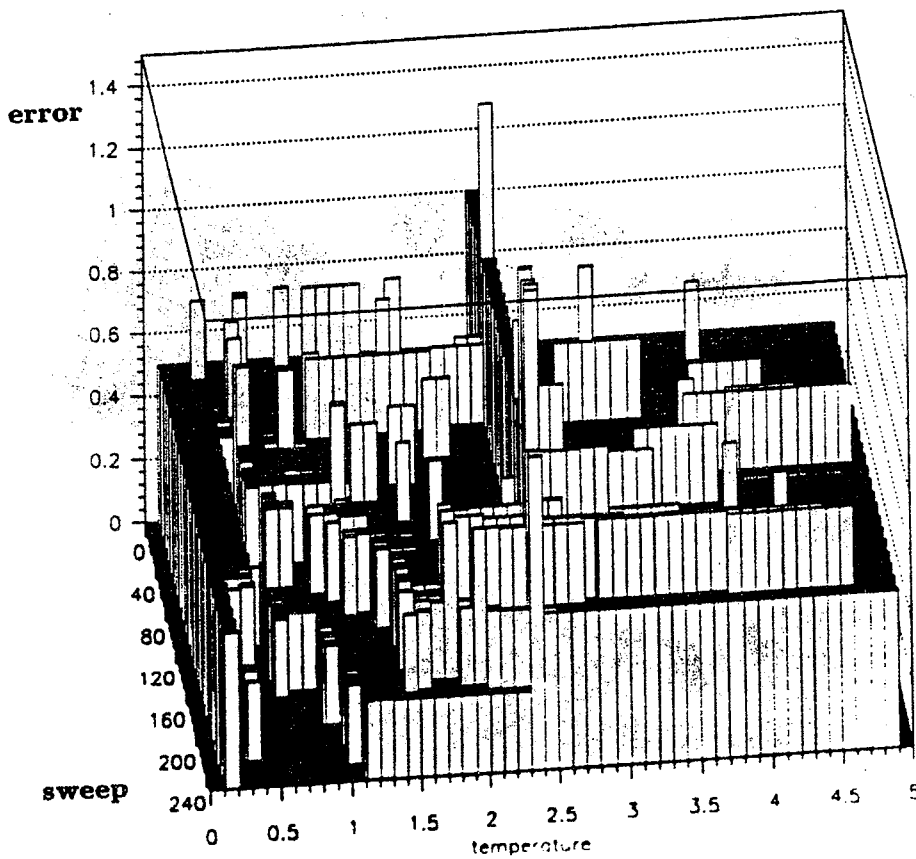
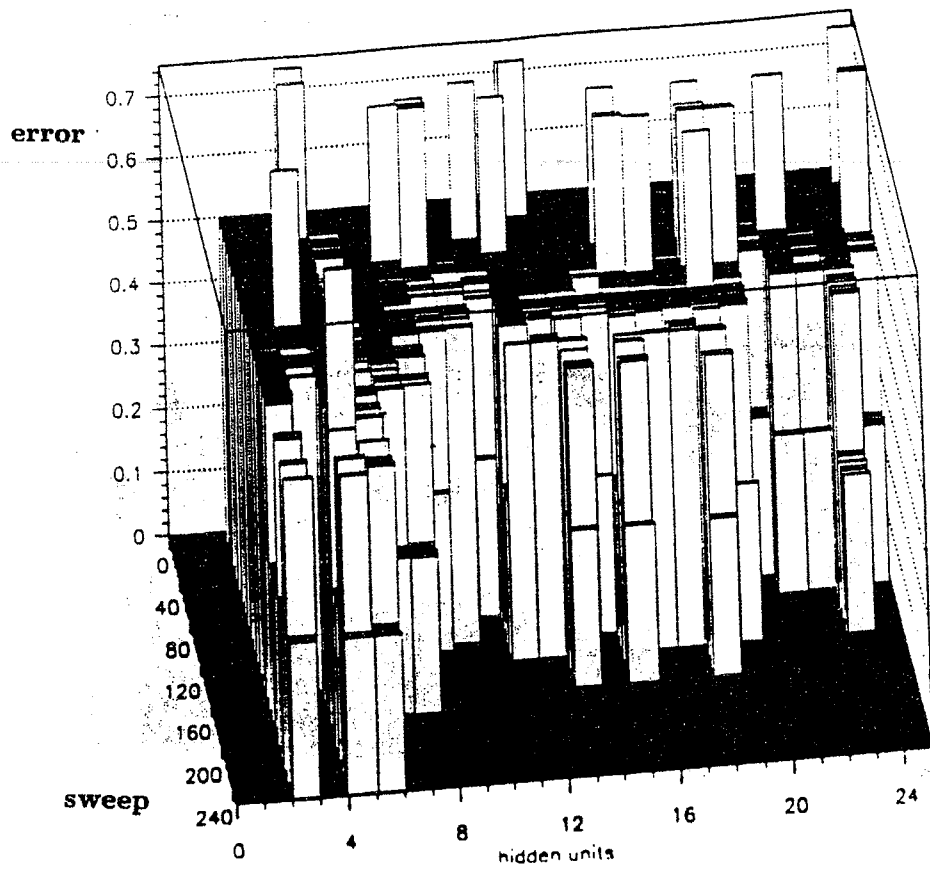


Figure B.3: The training performance as function of n (up) and T (below), but using a different initial setting of the weights.

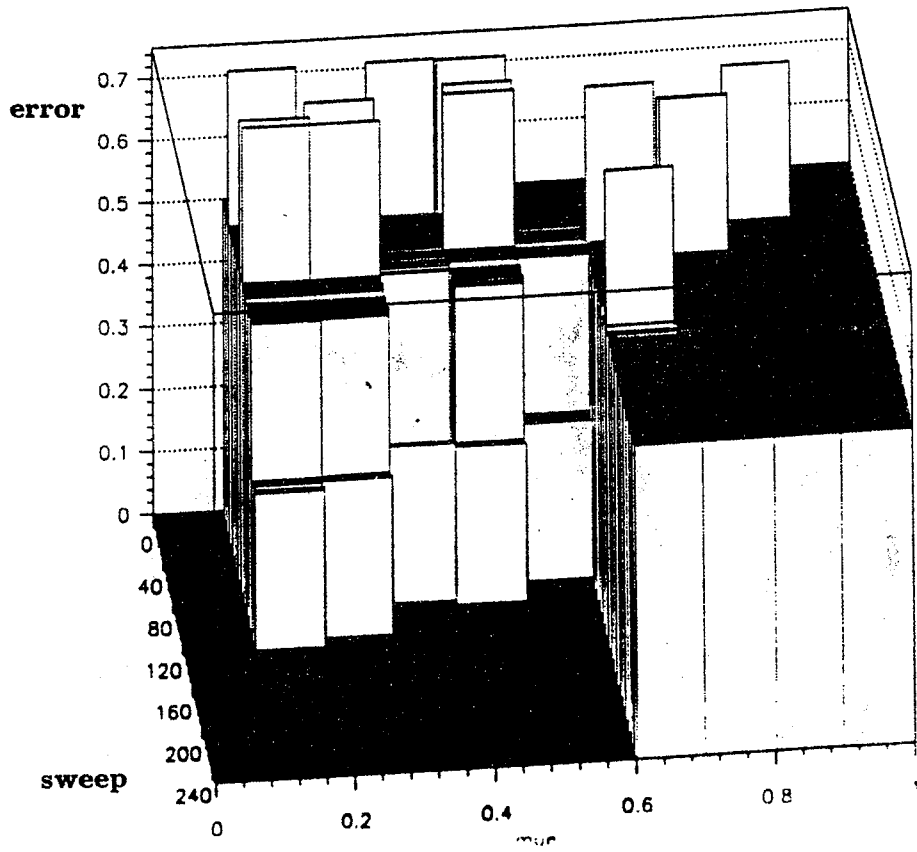
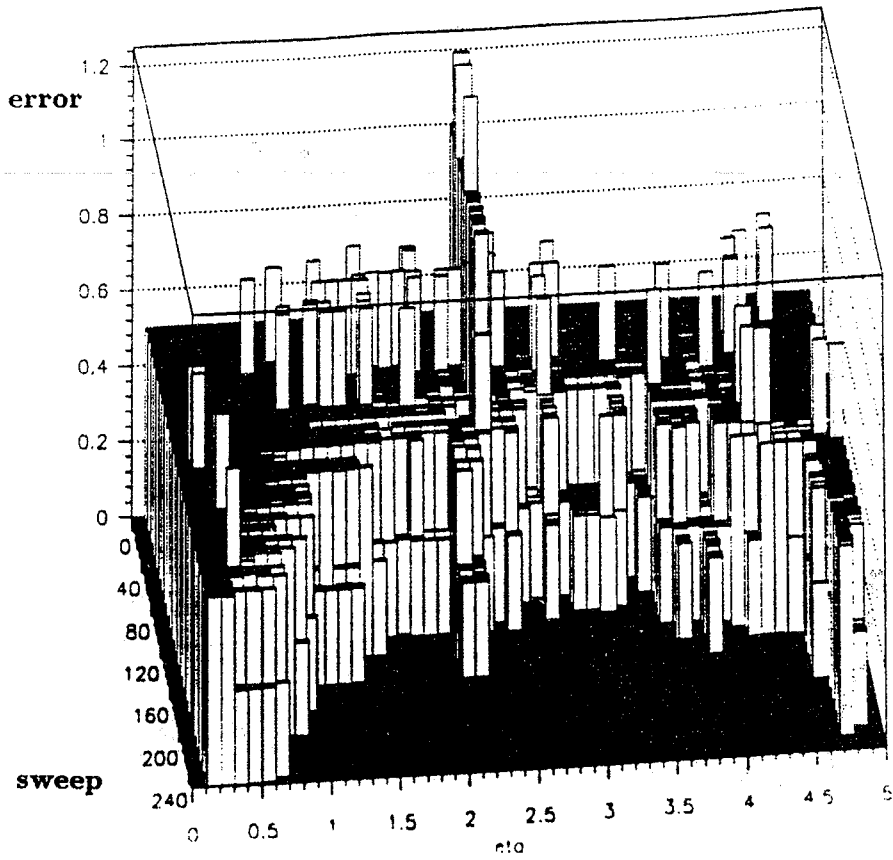


Figure B.4: The training performance as function of η (up) and μ (below), but using a different initial setting of the weights.

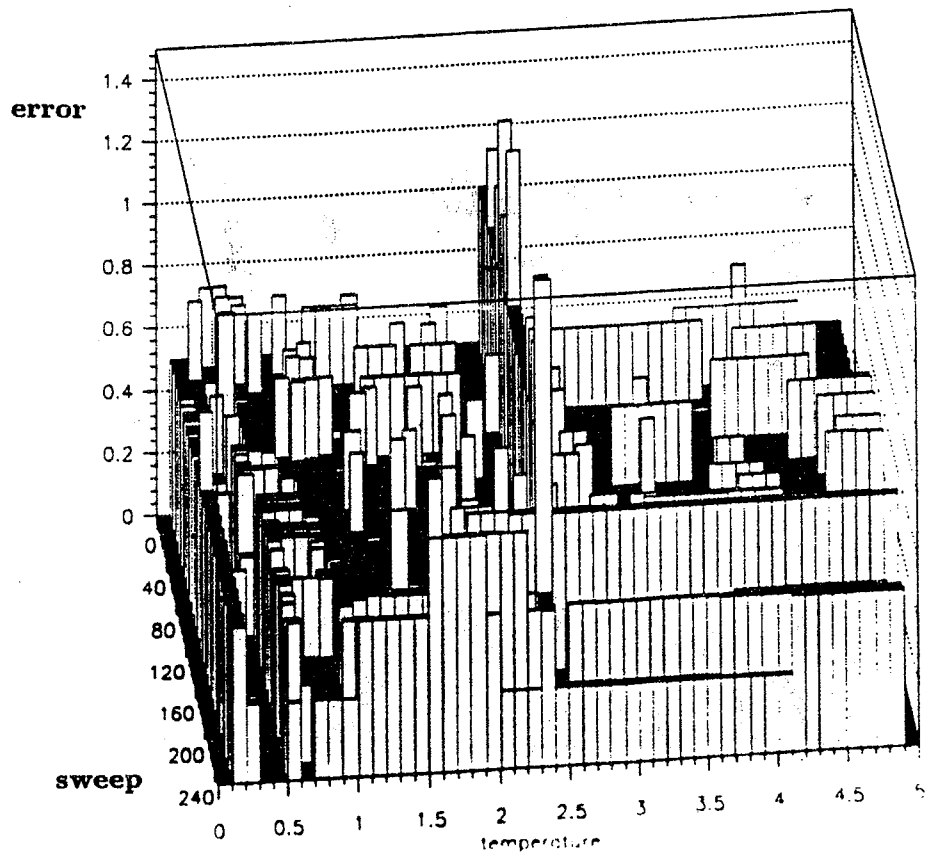
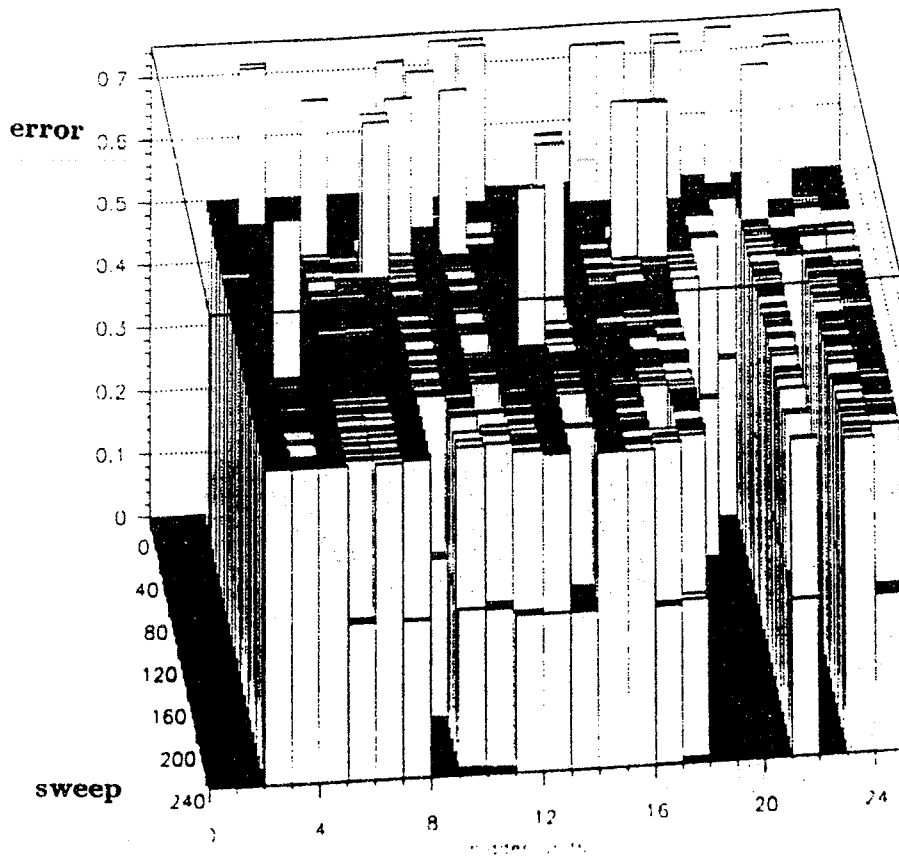


Figure B.5: The same situation as in figure B.1, but the weights have been modified by minimizing an entropy error (3.8).

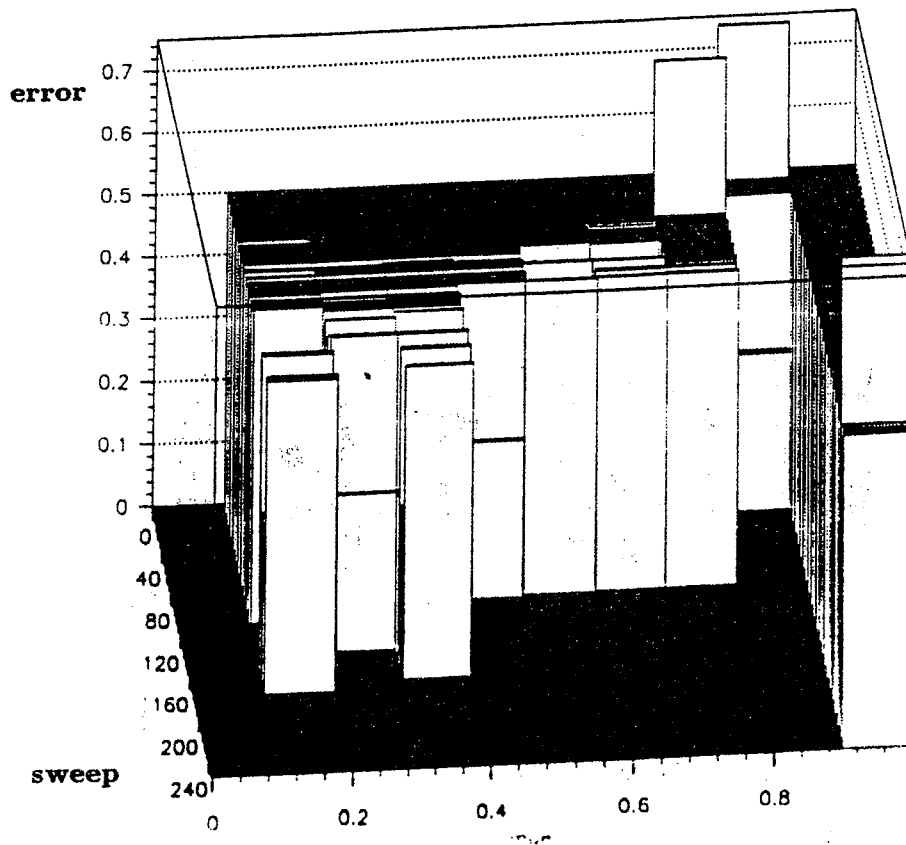
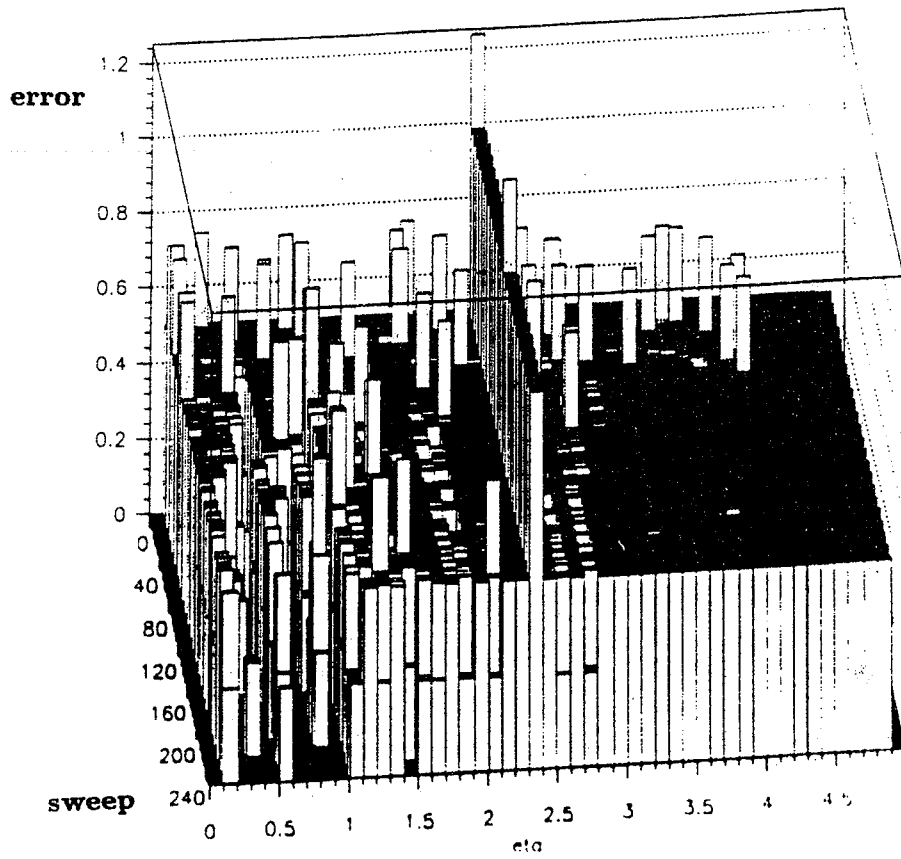


Figure B.6: The same situation as in figure B.2, but the weights have been modified by minimizing an entropy error (3.8).

Appendix C

b Selection with low-level Variables

The input variables of the classifier, which have been used in the main part of the thesis, have been rather complicated quantities built from the energy clusters in the calorimeters. On the other hand, biological systems receive only very primitive informations, for instance the location and the brightness of an object. Thus the grouping of these low-level input in some useful manner is apparently done by the system itself. (If at all!)

With this in mind, it had been studied, if a Neural Net is able to find meaningful relations itself, when presented merely very simple variables. Inspired from the example above, the input vector was made up by the energy E_i and the polar angles (ϑ_i, ϕ_i) of the 10 most energetic clusters of an event. To ease the task of the net, the positions of the clusters have been transformed into a coordinate system, which uses the thrust axis of the event as z -axis and its major axis as x -axis. Herein the thrust is defined as [50]

$$T = \max\left(\frac{\sum_i |\vec{p}_i \cdot \vec{n}|}{\sum_i |\vec{p}_i|}\right)$$

where \vec{p}_i is the momentum of the i -th particle. The direction \vec{n} , which is called the *thrust axis*, is chosen to maximize the expression. The major is defined the same way but is maximized in a plane perpendicular to the thrust axis.

The studies have been performed with 20000 fully reconstructed JETSET 7.3 events for training and 20000 events for testing. The same cuts have been applied as were in chapter 5. The distributions of the clusters in the (ϑ, ϕ) -plane is shown in figure C.1.

The classification performance of a single-node Functional-Link Net, a Feedforward-Backpropagation Net, and a single-node Perceptron have been studied. The Functional-Link Net used the 20 first terms of the Fourier series in order to expand the input vector, hence ≈ 1200 weights had to be determined. The achieved performances are shown in figure C.2. Though the performance of the input variables used in the thesis has not been reached, the one of the Functional-Link Net is still amazingly good. Probably a Feedforward-Backpropagation Net would have been able to obtain a similar performance like the Functional-Link Net, but due to its more complicated structure and the resulting numerical effort (matrix multiplication) a long (real) time of training is required. The Feedforward-Backpropagation Net was trained about a day on an Apollo 10000, compared with 1-2 hours of the training required for the Functional-Link Net. Nevertheless its performance remains poor. The poor ranking of the Perceptron is certainly founded on the fact, that the problem is not linear separable.

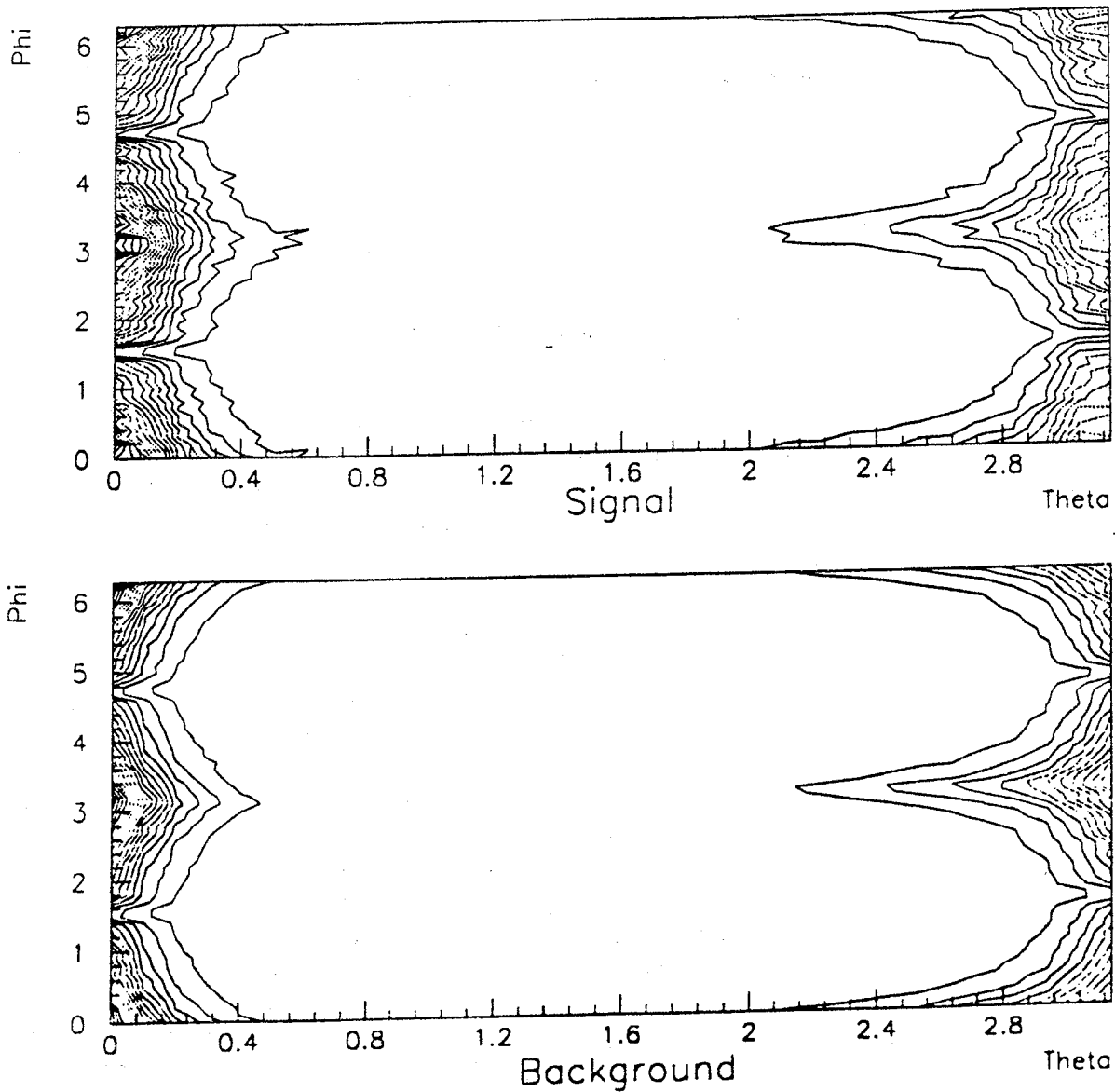


Figure C.1: Distribution of the cluster energies in the (ϑ, ϕ) -plane for b and light quarks. There is virtually no difference.

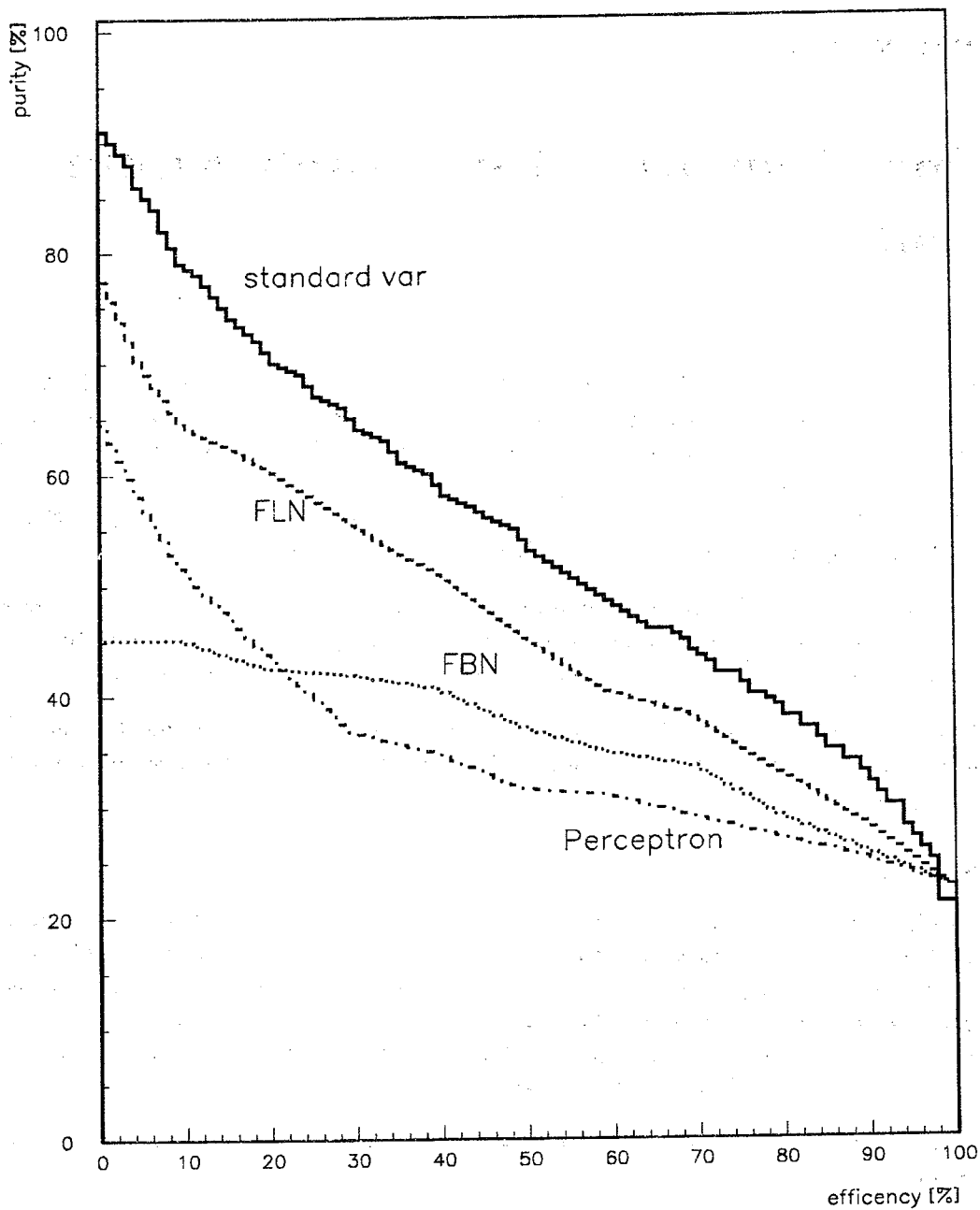


Figure C.2: Classification performance of a Functional-Link Net, a Feedforward-Backpropagation Net, and a Perceptron using low-level variables as input, compared with the performance of the variables applied in the thesis.

Appendix D

Further Methods for Quark Flavour Separation

About 70% of the Z^0 produced at LEP1 decay into hadrons. Whereas the selection of hadronic events is a rather easy task, the further separation of the hadrons into the respective quark flavours constitutes a far more complicated problem. This appendix shall give a short review of the methods commonly applied for flavour separation.

The properties of a quark, which can be exploited for separation, are its mass, its lifetime, its charge, and its decay. In general, the methods fall into two categories

1. Identification of the quark flavour by classifying the hadron, which contains the respective quark. For LEP1-energies this is only meaningful for the heavier quarks, hence the b and the c , because the lighter hadrons are also generated during fragmentation.
2. Identifying the type of quark, i.e. up-type or down-type, by measuring the charge. This assumes the universality of the coupling of the quarks in the up-sector and down-sector ($g_{a,v}^u = g_{a,v}^c$ and $g_{a,v}^d = g_{a,v}^s = g_{a,v}^b$).

Inclusive leptons

Because the semileptonic branching ratio $BR(b \rightarrow l)$ is about 11 %, about 22% of the b hadrons decay into $e\nu X$ or $\mu\nu X$, i.e. at least one electron or muon can be found in about 44% of the $Z^0 \rightarrow b\bar{b}$ -events. Unfortunately, the cuts necessary to obtain a sample, which is enriched with inclusive lepton events, reduce the overall efficiency for this type of event to 4 %.

The use of inclusive leptons for b and c -tagging is based on the specific properties of the leptons resulting from the high mass of their mother quark.

- Due to the hard fragmentation of the heavy quarks, the prompt leptons possess a high momentum.
- The transverse momentum of the leptons relative to the jet axis is high. ($\langle \vec{p}_\perp \rangle \approx m_q/4$)

The sample can be enriched with $b\bar{b}$ -events by applying cuts in the p, \vec{p}_\perp -spectra. The typical purity achieved is $\approx 80\%$. A further advantage of this tagging procedure consists in the possibility of the quark charge determination by measuring the charge of the lepton. Beside the limited statistics the uncertainty of the semi-leptonic branching ratio of about 10 % adds a relative large systematic error to the result of R_{bb} . [46]

Reconstruction of heavy hadrons

The reconstruction of exclusive decays of hadrons is a direct method to enrich a hadron sample with a desired quark flavour and is used mainly on b and c hadrons. The obtained statistic is rather low but the advantages of this approach is the low background and the full information concerning the heavy hadron.

Jet- and eventshape variables

As already discussed to some detail in chapter 6 the relative high mass of the b -quark leads to characteristic shapes of the jets and events. This manifests into a broader b jets due to the higher \vec{p}_\perp of the decay particles of the b and into different distributions in the energies of the clusters due to the harder fragmentation of the heavy quarks. Variables frequently used are eg. thrust, sphericity, energy of the first clusters, invariant mass of the jets etc. These variables are used as input to a multivariant-analysis, for instance CDA (*Canonical Discriminant Analysis* [68]) or a Neural Network. In general, Neural Network have shown a much better performance than CDA. Typical values for CDA are 30 % purity for 50 % efficiency.

The use of shape variables for b tagging has the advantage, that the measurement is independent of the error of the leptonic branching ratio. Also, the statistic is very high, because virtually all hadronic events can be used. On the other hand, uncertainties in the QCD parameters highly dominate the systematic error.

Secondary vertices [69]

The typical lifetime of b and c hadrons is $\approx 10^{-12}$ s. For LEP1-energies this results in decay lengths of several mm . Contrary, the decay length for strange particles (K^0, Λ^0) is in the magnitude of several cm . Therefore the reconstruction of secondary vertices (i.e. the decay vertex of the original hadron) can be used to separate the quark flavors. Provided that the detector is equipped with a micro-vertex detector, which allows a precise measurement of secondary vertices and impact parameters, this method is probably the most promising.

J/Ψ -mesons [70]

On the Z^0 -resonance J/Ψ -mesons are created almost exclusively from decays of b hadrons according to $b \rightarrow c\bar{c}s \rightarrow \Psi + X$. The creation of J/Ψ out of gluons is small and can be neglected. The J/Ψ s can be identified rather easily by their decay into lepton pairs. At LEP the process $b \rightarrow J/\Psi$ has already been used for the determination of the b lifetime and of the fragmentation parameter ϵ_b .

Mesurement of the jet charge

The total charge of all the particles in a jet is correlated to the charge of the primary quark. This effect is washed out by low-momentum particles coming from the fragmentation of the opposite quark and have been scattered back into the observed jet. In order to reduce this effect for the calculation of the total jet charge, each particle is weighted by its momentum

$$Q_{Jet} = \frac{\sum_i |\vec{p}_i \cdot \vec{e}_T|^\kappa \cdot q_i}{\sum_i |\vec{p}_i \cdot \vec{e}_T|}$$

where $|\vec{p}_i \cdot \vec{e}_T|$ is the component of momentum of the particle i parallel to the thrustaxis, q_i its charge, and κ is a weighting parameter and dependent on the sensitivity of the measurement. The main source of uncertainty is the limited knowledge of the process of fragmentation.

A measurement of the forward-backward jet charge asymmetry allows a determination of $\sin^2 \Theta_W$ [71]. A further application is the determination of the mixing parameter χ_B . Herein the b events are identified with inclusive leptons. The charge of the first quark is determined by the charge of the lepton and the charge of the second by the jet charge. The results [72] are in good agreement with the ones obtained by using events with two inclusive leptons. The jet charge method is dominated by the systematic error and the dilepton measurement by the limited statistic.

Prompt photons [73]

Photon bremsstrahlung can occur after the decay of the Z into a quark pair and before their fragmentation. The probability for this process is proportional to the squared quark charge. Assuming universality for the coupling of the up-type and down-type quarks to the Z , the number of expected $q\bar{q}\gamma$ -events can be cast like

$$N_{q\bar{q}\gamma} \propto \sum_{i=1}^5 q_i^2 \Gamma_i = \frac{1}{9} (3\Gamma_{\frac{1}{3}} + 8\Gamma_{\frac{2}{3}}) \quad (\text{D.1})$$

where q_i is the quark charge, Γ_i the partial width of a single quark flavour, and $\Gamma_{\frac{2}{3}}, \Gamma_{\frac{1}{3}}$ are the partial widths of all quarks from the up- ($q_i = \frac{2}{3}$) and the down-sector ($q_i = \frac{1}{3}$), respectively. On the other hand the total hadronic width of the Z is given by

$$\Gamma_{had} = \sum_{i=1}^5 \Gamma_i = 3\Gamma_{\frac{1}{3}} + 2\Gamma_{\frac{2}{3}} \quad (\text{D.2})$$

(D.1) and (D.2) can be used in order to determine the partial widths. Also for this method, the systematic error is dominated by our limited knowledge of the fragmentation process. The main background for prompt photons consists of isolated, high energetic π^0 - and η -mesons, which can fake a photon. In order to estimate the error, simulation of the energy spectra and production rate of these particle are necessary. Therefore further understanding of these processes is needed.

Appendix E

The Program Packages ZFITTER and FUNPLO

Many analytical programs for specialized applications in e^+e^- annihilation are available. These programs are to fit for parameters of the Standard Model (SM). The input to the fits consists of cross sections and asymmetries measured at a series of \sqrt{s} values, already corrected for detector effects, together with the theoretical predictions of the analytical programs. One of the most popular of these programs for LEP1 physics is the ZFITTER package. The program FUNPLO, written by a member of the L3 collaboration [74], actually performs these fits, taking into account correlations in the input data and their errors.

E.1 ZFITTER

The analytical program ZFITTER 4.5 [75] performs calculations for fermion-pair production in e^+e^- annihilation. The following models are implemented in ZFITTER:

- analytical SM formulae with higher order corrections
- model independent ansatz using effective axial-vector and vector couplings of fermions to the Z^0 boson
- model independent ansatz using partial decay widths of the Z^0 to fermion species
- S-matrix ansatz using S-matrix theory for a global description of the hard scattering process

ZFITTER calculates the following observables:

- SM cross sections and forward-backward asymmetries as functions of m_Z , m_{top} , m_H , and α_S
- SM τ -polarization and forward-backward τ -polarization as functions of m_Z , m_{top} , m_H , and α_S
- model-independent cross sections and asymmetries as functions of the normalization form factors (ρ) and the effective vector and axial-vector coupling constants (\bar{g}_{af} , \bar{g}_{vf}), respectively

- model-independent final-state polarization in τ -pair production as a function of the normalization factors and the effective coupling constants
- model-independent cross sections as functions of the partial and the total Z width
- model-independent cross sections, based on an S-matrix inspired ansatz

ZFITTER 4.5 contains a complete $\mathcal{O}(\alpha)$ electroweak treatment using the DIZET 4 weak library, including:

- soft photon exponentiation
- higher order QED corrections for initial-state radiation
- complete m_t dependent $\mathcal{O}(\alpha)$ terms
- leading $\mathcal{O}(\alpha^2 m_t^4)$ terms
- complete $\mathcal{O}(\alpha\alpha_S)$ terms plus leading parts in $\mathcal{O}(\alpha\alpha_S m_t^2)$

ZFITTER is the standard analytical program used by three LEP collaborations, including L3, to make fits of parameters of interest to the data.

E.2 FUNPLO

The ZFITTER program calculates observables measured at LEP as functions of interesting parameters of the underlying model. A fit has to be performed in order to determine the best values of these parameters, i.e. those, which best reproduce the measured data. For this purpose, fit programs like FUNPLO have been written, which calculate the appropriate χ^2 or likelihood function for the data set used and minimize these quantities by calling the CEN program MINUIT [47] in order to extract the best parameter values. FUNPLO takes into account common and correlated systematic errors, effects of the beam-energy calibration of LEP and effects of the beam-energy spread.

Fitting Function

The FUNPLO fit is performed by minimizing a χ^2 quantity. The χ^2 contains measurements obtained from data in channels f at cms-energy point i of year n , usually measured values for cross sections, $\sigma(f, i, n)$, and asymmetries (forward-backward and polarization, though the latter is not of interest in the frame of this thesis), $A(f, i, n)$. In the case of quarks, only the b quark can be treated separately. Otherwise the hadronic cross section is summed over all quark flavours.

The fit variables form a vector \vec{x} with components x_j , which contain two sets: First, there are the variables, which have a physical meaning within the theoretical model, for instance the mass and width of the Z^0 in the SM. Second, there are the variables, which are of a more technical nature. They are also called *correction terms*, because they implement common correlated errors between several measurements used in the fit. Variables of both sets can be constrained to a mean \bar{x} within a range $\delta\bar{x}$.

Furthermore, “external” measurements, giving additional constraints $y_k \pm \delta y_k$, for instance measurements of the $\frac{W}{Z^0}$ mass ratio from collider experiments, can be taken into account in the fit.

After all the χ^2 function has the following form:

$$\begin{aligned}\chi^2 = & \sum_{f,i,n} \alpha(f,i,n) \left[\frac{\sigma'(f,i,n) - \sigma_{th}(f,i,n,\vec{x})}{\delta\bar{\sigma}(f,i,n,\vec{x})} \right]^2 \\ & + \sum_{f,i,n} \beta(f,i,n) \left[\frac{A'(f,i,n) - A_{th}(f,i,n,\vec{x})}{\delta\bar{A}(f,i,n,\vec{x})} \right]^2 \\ & + \sum_j a_j \left[\frac{x_j - \bar{x}_j}{\delta\bar{x}_j} \right]^2 + \sum_k b_k \left[\frac{y_k - \bar{y}_k}{\delta\bar{y}_k(\vec{x})} \right]^2\end{aligned}$$

The factors α , β , a , and b are set either to 0 or 1 to determine if the term is included in the χ^2 function or not, respectively.

The cross sections $\sigma'(f,i,n)$ and asymmetries $A'(f,i,n)$ entering the χ^2 contain the correction terms x mentioned above, taking the correlated part of the total measurement errors, i.e., the common systematic errors, into account:

$$\begin{aligned}\sigma'(f,i,n) &= x_{Lum}(n) \cdot x_{sys}^\sigma(f,n) \cdot \sigma(f,i,n) \\ A'(f,i,n) &= A(f,i,n) + x_{sys}^A(f,n)\end{aligned}$$

The correction terms x are treated as Gaussian distributed random variables with mean \bar{x} either 0 or 1 for additive or multiplicative corrections, respectively, and a $\delta\bar{x}$ corresponding to the value of the correlated error they describe. Usually, for a given channel f , the correction terms expressing the systematic errors are identical for all \sqrt{s} points i of a given year n .

The correction terms x_{sys} refer to the systematic errors of event selection. In the case of fitting a cross section, an additional common systematic error arises from the luminosity measurement, corresponding to x_{Lum} . This error consists of two parts, one due to the common theoretical error in the Bhabha cross section determination, and another from the systematic error in luminosity event selection.

The theoretical predictions $\sigma_{th}(f,i,n,\vec{x})$ and $A_{th}(f,i,n,\vec{x})$, calculated by ZFITTER, depend of course only indirectly on i and n via the corresponding cms-energy $\sqrt{s}(i,n)$. The values of $\sqrt{s}(i,n)$ are given by the LEP energy calibration. They in turn depend on several variables X describing the fluctuations around the luminosity weighted mean energy of the fills belonging to energy point i . These variables are therefore included also in the vector \vec{x} . The finite spread of the LEP beam energy of about 51 MeV [76] is included by convoluting the theoretical cross section as a function of \sqrt{s} with a normalized Gaussian of 51 MeV width around the mean cms-energy.

The errors $\delta\bar{\sigma}$ and $\delta\bar{A}$ contain the uncorrelated part of the total error of each measurement. The experimental errors are corrected by a factor to obtain the error as expected by the theoretical prediction for the measurement:

$$\begin{aligned}\delta\bar{\sigma}(f,i,n) &= \delta\sigma(f,i,n) \cdot \sqrt{\frac{\sigma_{th}(f,i,n,\vec{x})}{\sigma(f,i,n)}} \\ \delta\bar{A}(f,i,n) &= \delta A(f,i,n) \cdot \sqrt{\frac{1 - A_{th}^2(f,i,n,\vec{x})}{1 - A^2(f,i,n,\vec{x})}} \cdot \sqrt{\frac{A_{th}(f,i,n,\vec{x})}{A(f,i,n)}}\end{aligned}$$

Standard Fits

Usually, the following three types of fits are performed

- a model-independent fit to the cross section data alone, to determine the mass m_Z as well as the total width Γ_Z and the various partial widths Γ_f of the Z^0
- a model independent fit to the combined cross section and asymmetry data, to determine the effective coupling constants g_{af} and g_{vf}
- a SM fit to the combined data set to determine the as yet unknown parameters of the SM: m_t and m_H , along with α_S and m_Z .

These fits may or may not impose lepton universality or use "external" constraints.

Bibliography

- [1] A. Barbaro-Galtieri, *Search for the Top at the TEVATRON*, LBL EPS, Marseille, 26 July 1993
N.N. Shaw, *Recent Results on Top, Bottom and Exotic Physics at the TEVATRON*, FERMILAB-Conf-93/263-E
- [2] W.Beenakker, W.Hollik, *The width of the Z boson*, Zeitschrift für Physik C40, 141-148 (1988)
J.Bernabéu, A.Pich, S.Santamaria, *Top quark mass from radiative corrections to the $Z^0 \rightarrow b\bar{b}$ decay*, Nuclear Physics B363 (1991) 326-344
- [3] M.Veltman, Nucl.Phys. B123 (1977) 89
- [4] S.Glashow, Nucl.Phys. 22 (1961) 579
A.Salam, Phy.Review 127 (1962) 331
S.Weinberg, Phy.Rev.Letter 19 (1967) 1264
- [5] M.Consoli, W.Hollik, *Electroweak radiative corrections for Z physics*, ZPhysics at LEP1, Vol 1, CERN 89-08
- [6] G.Altarelli, *The standard electroweak theory and its experimental tests*, Lectures given at the 32. Internationale Universitätswochen für Kern- und Teilchenphysik, Schladming, Austria (24. February - 5. March 1993), CERN-TH6867/93
- [7] F.Berends, *Z line shape*, ZPhysics at LEP1, Vol 1, CERN 89-08
- [8] J.H.Kühn, P.M.Zerwas, *Heavy Flavours*, ZPhysics at LEP1, Vol 1, CERN 89-08
- [9] *Heavy Flavours at LEP*, J.H.Kühn, P.M.Zerwas (Convenors), MPI-PAE/PTh 49/89 (1989)
- [10] A.A.Akhundov, D.Y.Bardin, T.Riemann, *Electroweak one-loop corrections to the decay of the neutral vector boson*, Nuc.Phys. B 276 (1986) 1-13
D.Y.Bardin, P.Ch.Christova, O.M.Fedorenko, *On the lowest order electroweak corrections to spin- $\frac{1}{2}$ fermion scattering*, Nuc.Phys. B 197 (1982) 1-44
D.C.Kennedy, B.W.Lynn, *Electroweak radiative corrections with an effective Lagrangian: four-fermion processes*, Nuc.Physics B 322 (1989) 1-54
D.C.Kennedy, B.W.Lynn, C.J.C. Im, R.G.Stuart, *Electroweak cross sections and asymmetries at the Z^0* , Nuc.Physics B 321 (1989) 83-107
- [11] *Z physics at LEP*, G.Altarelli, R.Kleiss, C.Verzegnassi (Eds.), CERN 89-08
- [12] R.Kleiss, *Electro-weak interactions*, Lecture Series for postgraduate students, CERN, November 1990

- [13] Leif Lönnblad, Carsten Peterson, Thorsteinn Rögnvaldsson, *Using Neural Networks to Identify Jets*, LU TP 90-8 (May 1990)
V.Innocente, Y.F.Wang, Z.P.Zhang, *Identification of Tau Decays Using a Neural Network*, CERN-PPE/92-98 (June 1992)
- [14] C.Peterson, B.Söderberg, *A new method for mapping optimization problems onto Neural Networks*, Int.Journal of Neural Systems, Vol.1, No.1 (1989) 3-22
- [15] T.Kohonen, *Self-Organized Formation of Topologically Correct Feature Maps*, Biological Cybernetics 43, 59-69
G.A.Carpenter, S.Grossberg, *ART2: Self-organization of stable category recognition codes for analog input patterns*, Applied Optics, Vol.26, 4919-4930
- [16] C.Peterson, *Track finding with Neural Networks*, Nuc.Inst.Meth. A 279 (1989) 537-545
- [17] M.A.Mahowald, C.Mead, *Die Silicium Netzhaut*, Spektrum der Wissenschaft Juli 1991
- [18] E.Gardner, B.Derrida, *Optimal storage properties of neural network models*, J.Phys.A — Math.Gen. 21 (1988) 271-284
- [19] W.S.McCulloch, W.Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bulletin of Math.Bio. 5 (1943) 225-133
- [20] Kandel, *Fuzzy Techniques in Pattern Recognition*, Wiley-Interscience, 1982
- [21] J.Neyman, E.S.Pearson, *On the use and interpretation of certain test criteria for purposes of statistical inference*, Biometrika 20 A (1928), 175-240
J.Neyman, E.S.Pearson, *Of the problem of the most efficient tests of statistical hypotheses*, Phil.Trans.Royal Soc. London 231 (1933) 289-337 A.Wald, *Contributions to the theory of statistical estimation and testing of hypotheses*, Ann.Math.Stat. 10 (1939) 299-366
- [22] R.O.Duda, P.E.Hart, *Pattern Classification and Scene Analysis*, Wiley (1973)
- [23] F.Rosenblatt, *Principles of Neurodynamics: Perceptrons and The Theory of Brain Mechanisms*, Spartan, New York
- [23] *Mathematica* is a registered trademark from Wolfram Research Inc., 199 Trade Center Drive, Champaign, Illinois 61820-7237
- [24] Frank Raupach, *private communication*
- [25] B.Widrow, M.E.Hoff, IRE Wescon Convention Record – Part 4 – pp. 96-104 (1960), Institute of Radio Engineers, New York
B.Widrow, *Generalization and information storage in networks of Adaline 'Neurons'*, M.C.Yovits, G.T.Jacobi, and G..D.Goldstein (Eds.) *Self-Organizing Systems* – pp. 435-461 (1962), Spartan Books, Ney York
- [26] S.A.Solla, E.Levin, M.Fleisher, *Accelerated Learning in Layered Neural Networks*, Complex Systems 2 (1988) 625-640
- [27] N.J.Nilsson, *Learning Machines* (1965), McGraw-Hill, New York
- [28] A.N.Kolmogorov, *On the Representation of Continuous Functions Of Many Variables by Superposition of Continuous Functions of One Variable and Addition* [in Russian], Dokl.Akad.Nauk USSR, 114 (1957), pp. 953-956

- [29] Leif Lönnblad, Carsten Peterson, Thorsteinn Rögnvaldsson, *Pattern Recognition in High Energy Physics with Artificial Neural Networks — JETNET 2.0 Manual*, LU TP 91-18 (August 1991)
- [30] Yoh-Han Pao, *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley 1989
- [31] D.Sobajic, *Neural Nets for Control of Power Systems*, Ph.D.Thesis (1988), Computer Science Dept., Case Western Reserve University, Cleveland, OH.
- [32] C.Camilleri et.al., *Physics with Very High Energy e^+e^- Colliding Beams*, CERN 76-18 (1976)
- [33] *The Construction of the L3 Experiment*, L3 collaboration, Nucl.Inst. A 289 (1990) 35-102
- [34] H.T.Meinholz, *Erkennung von Elektronen und Photonen im elektromagnetischem BGO-Kalorimeter des L3-Detektors*, Diploma thesis, PITHA 87/01, RWTH Aachen, May 1987
- [35] J.F.Zhou, *Untersuchung an einem Uran Hadron Kalorimeter mit Proportionalkammerauslese für den L3 Detektor am e^+e^- Spricherring LEP*, phd-thesis, PITHA 90/06, RWTH Aachen, Mai 1989
- [36] R.Brun et.al., *GEANT 3*, CERN-DD/EE/84-1 (Revised September 1987)
- [37] H.Fesefeldt, *The Simulation of Hadronic Showers: Physics and Applications*, RWTH-Aachen PITHA 85/02 (1985)
- [38] O.Adriani et.al., Nucl.Inst.Meth. A 302 (1991) 53
- [39] T.Sjöstrand, *Pythia 5.6 and JETSET 7.3 — Physics and Manual*, CERN-TH6488/92 (September 1992)
- [40] D.Stickland, *private communication*
- [41] The L3 Collaboration, *Results from the L3 Experiment at LEP*, 22.February 1993, submitted to *Physics Report*
- [42] *Measurements of the $Z^0 \rightarrow b\bar{b}$ Decays and the Semileptonic Branching Ratio $BR(b \rightarrow l + X)$* , L3 collaboration, Phys. Lett. B 261 (1991) 177
- [43] *Measurement of $Z^0 \rightarrow b\bar{b}$ Decay Properties*, L3 collaboration, Phys. Lett. B 241 (1990) 416
- [44] G.Altarelli, G.Parisi, Nucl.Phys. B126 (1977) 298
- [45] W.Braunschweig et al., TASSO Collab., Z.Phys. C43 (1989)
- [46] Yi-Jin Pei, *Schwache Wechselwirkungen von b- und c-Quarks untersucht in Z^0 -Zerfällen beobachtet mit dem L3-Detektor*, Ph.D.Thesis (1992), Phy.Inst., RWTH Aachen
- [47] F.James, M.Roos, *MINUIT manual*, CERN-D506 (1989)
- [48] J.Mnich, Ch.Paus *private communications*
- [49] Bronstein, Semendjajew, *Taschenbuch der Mathematik*, Vol 1, Verlag Harri Deutsch, Thun 1980, pg.726

- [50] *Studies of Hadronic Event Structure and Comparison with QCD Models at the Z^0 Resonance*, L3 Collab., CERN-PPE/92-50, 26.March 1992
- [51] G.Andersson, G.Gustafson, G.Ingelmann, T.Sjöstrand, Phys.Rep. 97 (1983) 33
- [52] T.Sjöstrand et al., Physics at LEP1, CERN 89-08, Vol 3, 143
- [53] *Review of Particle Properties*, Physical Review D45 (June 1992)
- [54] P.M.Stevenson, Phys.Rev. D23 (1981) 2916
- [55] G.Marchesini et.al., *HERWIG 5.1 — a Monte Carlo Event Generator for Simulating Hadron Emission Reactions with Interfering Gluons*, Comp.Phys.Comm. 67 (1992) 465
- [56] A.Ricker, *G-Factors*, L3 Internal Note 1394 (8.April 1993)
- [57] *Mesurement of the Ratio Γ_{bb}/Γ_{had} using Event Shape Variables*, ALEPH collaboration, CERN-PPE/93-113 (5.July 1993)
- [58] *Classification of the Hadronic Decays of the Z^0 into b and c Quark Pairs using a Neural Network*, DELPHI collaboration, CERN-PPE/92-151 (15.September 1992)
- [59] W.Braunschweig et al, TASSO collaboration, DESY preprint, DESY90-047 (1990)
 B.Adeva et al., MARK-J collaboration, Phys.Rep. 109 C (1984) 132
 E.Elsen et al., JADE collaboration, Z.Phys. 46 C (1990) 349
 H.J.Behrend et al., CELLO collaboration, DESY89-125 (1989)
 C.R.Ng et.al., HRS collaboration, ANL-HEP-PR-88-11 (1988)
 H.R.Band et.al., MAC collaboration, Phys.Lett.218 B (1989) 369
- [60] S.Bethke, St.Catani, *A Summary of α_S Measurements*, CERN-Th.6484/92
- [61] J.Mnich, *private communications*
- [62] O.Adriani et.al., *Determination of the effective electroweak mixing angle from Z decays*, Phy.Lett. B 309 (1993) 451-462
- [63] G.Girardi, W.Hollik, C.Verzegnassi, *A "no-lose" measurement of the hadronic and of the leptonic Z widths*, Phy.Letters B 240, No.3,4 (1990) 492-496
 G.Altarelli, R.Barbieri, F.Caravaglios, *The Minimal Supersymmetric Standard Model and Electroweak Precision Data*, CERN-TH.6902-93 (June 1993)
- [64] A.Weber, *Suche nach radiativen Neutralino Zerfällen mit dem L3 Detektor*, Ph.D. thesis, to be published
- [65] The LEP Collaborations ALEPH, DELPHI, L3, OPAL and The LEP Electroweak Working Group, *Updated Parameters of the Z^0 Resonance from Combined Preliminary Data of the LEP Experiments*, CERN/PPE/93-157 (26 August 1993)
- [66] D.E.Rumelhart, G.E.Hinton, R.J.Williams, *Learning internal representations by error propagations*, in D.E.Rumelhart and J.L.McClelland (Eds.) *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* Vol. 1: *Foundations* (1986) pp. 318-362, MIT Press, Cambridge, D.C.
- [67] M.Minsky, S.Papert, *Perceptron: An Introduction to Computational Systems* (1969), MIT Press, Cambridge, MA.

- [68] T.W.Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York (1958)
 Y.Fujikoshi, P.R.Krishnaia (eds.), *Discriminant Analysis and Canonical Correlation Analysis*
- [69] *Measurement of the average lifetime of B hadrons*, TASSO collaboration, Z.Phys. C – Particles and Fields 44, 1-13 (1989)
Measurement of the Average Lifetime of b Hadrons, L3 collaboration, CERN-PPE/93-158; August 30, 1993
- [70] *Inclusive J production in Z⁰ Decays*, L3 collaboration, Phys. Lett. B 288 (1992) 412
- [71] ALEPH collaboration, Phys.Lett. 259 B (1991) 377
 DELPHI collaboration, Phys.Lett. 277 B (1992) 371
- [72] *A measurement of the charge asymmetry of hadronic events in electron positron annihilation*, JADE collaboration, Z.Phys. C — Particles and Fields 42, 1-6 (1998)
- [73] *Determination of Quark Electroweak Couplings from Direct Photon Production in Hadronic Z Decays*, L3 collaboration, CERN-PPE/92-209 (4.December 1992)
- [74] J.Rose, *Bestimmung von Parametern der elektroschwachen Theorie mit dem L3-Detektor bei LEP*, Ph.D.Thesis (1992), Phy.Inst., RWTH Aachen
- [75] D.Bardin et al., *ZFITTER — An Analytical Program for Fermion Pair Production in e⁺e⁻ Annihilation*, CERN-TH 6443/92, May 1992
- [76] The Working Group on EP Energy, *The Energy Calibration of LEP in 1991*, CERN-PPE/92-125, CERN-SL/92-37(DI)



Acknowledgements

I would like to thank my supervisor, Prof. Wolfgang Wallraff, who made many helpful suggestions during the analysis and the writing of the thesis itself. He has been a valuable source of informations, not only concerning physics in general and L3 in particular, but he also offered his help everytime, when I had to come in closer contact with the French speaking world. (Though I always preferred to bother Gabriele with these problems.)

Furthermore I want to express my gratitude to the director of my institute, Prof. Klaus Lübelmeyer, for his support and mutual patience. He made possible to me a 3-year stay at CERN and, from the very beginning, encouraged my attempts to apply Neural Network algorithms in data analysis.

I owe special thanks to Tariq Aziz from the Tata Institute in Bombay. We worked together on the determination of R_{bb} and he taught me so much about physics analysis and its nasty little tricks.

Martin Grünwald gave countless hints and suggestions, that one is tempted to say, that this thesis is partially his work. I hope that I could return his help to some minor extend in initiating him into the art of cooking eggs.

I would like to thank Ian Brock, Joachim Mnich, Yi-Jin Pei, and Andreas Ricker for long and mostly fruitful discussions.

Jian-Feng Zhou explained physics to me in a very clear and neat manner. He should give up research soon and become a teacher.

Joachim Rose introduced me into his FUNPLO program and always supported me during modification of its code.

David Sticklands invention of the DVN data format accelerated the slow L3 analysis software to performances never been dreamed of. His efforts and engagement to get and keep things running have been really a very big help.

Markus Möller and Alfons Weber have proven valuable allies in the daily fight against computers and analysis software. They read the preprint of my thesis, drawing my attention to obscurities, mistakes, and typos.

Stephan Müller read the preprint and was impressed.

Georg Schwering closed convertibles during thunderstorms and provided me with lots of gossips. Ulrich Uwer showed me that there is more to cook than eggs. Thanks to my brother Stephan for logistics, tea, and cookies.

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is essential for ensuring transparency and accountability in the organization's operations.

2. The second part of the document outlines the various methods and tools used to collect and analyze data. It highlights the need for consistent and reliable data collection processes to support informed decision-making.

3. The third part of the document focuses on the role of technology in modern data management. It discusses how advanced software solutions can streamline data collection, storage, and analysis, thereby improving efficiency and accuracy.

4. The fourth part of the document addresses the challenges associated with data security and privacy. It stresses the importance of implementing robust security measures to protect sensitive information from unauthorized access and breaches.

5. The fifth part of the document explores the ethical implications of data collection and analysis. It discusses the need for transparency in data practices and the importance of obtaining informed consent from individuals whose data is being collected.

6. The sixth part of the document provides a summary of the key findings and recommendations. It reiterates the importance of a data-driven approach and the need for continuous improvement in data management practices.

7. The final part of the document includes a list of references and a glossary of key terms. This section is intended to provide additional context and resources for readers interested in the topics discussed in the document.

Lebenslauf

24. Mai 1961 geboren in Düsseldorf als erster Sohn des Dr.med. Gottfried Röhner und Dr.med. Hanne-Lore Röhner, geb. Andree
- 1967-1971 Besuch der Grundschule
- 1971-1980 Besuch des Gymnasiums
23. Mai 1980 Abitur
- 1980-1983 Grundstudium Diplomphysik an der Heinrich-Heine-Universität in Düsseldorf
18. Mai 1982 Vordiplom in Physik
- 1983-1986 Hauptstudium an der RWTH Aachen
- 1986-1987 Diplomarbeit
21. April 1988 Diplom in Physik
- 1988-1994 Anstellung als wissenschaftlicher Mitarbeiter an der RWTH Aachen mit dem Ziel der Promotion
- 1989-1992 3 jähriger Forschungsaufenthalt im CERN

