

Fermilab Library



0 1160 0063239 2

KA-IEKP-92-9

IEKP-KA-92-9

B-Quark-Tagging
mit
neuronalen Netzwerken

Martin Büchert



FERMILAB

SEP 23 1998

LIBRARY

INSTITUT FÜR EXPERIMENTELLE KERNPHYSIK

UNIVERSITÄT KARLSRUHE

FERMILAB

SEP 23 1998

LIBRARY

IEKP-KA/92-9

**B-Quark-Tagging
mit
neuronalen Netzwerken**

Martin Büchert

Diplomarbeit

Institut für experimentelle Kernphysik
Fakultät für Physik
Universität Karlsruhe (TH)

Karlsruhe, März 1992

Zusammenfassung

In dieser Diplomarbeit wurde gezeigt, daß die Separation von multihardonischen Z^0 -Zerfällen in die zwei Klassen von schweren b -Quark und *nicht* - b -Quark Ereignissen durch ein dreischichtiges vorwärtsvermittelndes neuronales Netzwerk möglich ist. Die Studie wurde mit rein topologischen Observablen anhand vollrekonstruierter Monte-Carlo Daten durchgeführt. Es lassen sich Reinheiten von 50% bei 50% Effizienz erreichen. Reinheiten deutlich über 60% bei noch akzeptablen Effizienzen sind nur durch zusätzliche Eingangsinformation erzielbar.

Die Handhabung von neuronalen Netzwerken wurde unter Berücksichtigung der speziellen Erfordernisse des b -Tagging studiert. Es wurde eine Methode zur Normierung der Eingangsdaten entwickelt und ein Algorithmus zur Reduktion der Knoten in der Zwischenschicht eines dreischichtigen Netzwerkes unter Beibehaltung der Leistungsfähigkeit implementiert.

Inhaltsverzeichnis

1	Einführung	1
2	Theoretische Beschreibung	6
2.1	Das Standardmodell	6
2.2	Grundlagen und die daraus entwickelten Tagging-Methoden	9
2.2.1	Die Fragmentation	11
2.2.2	Die Zerfallslänge des <i>B</i> -Mesons	14
2.2.3	Der semileptonische Zerfall des <i>B</i> -Mesons	16
3	Neuronale Netzwerke	18
3.1	Biologische Vorbilder	18
3.2	Künstliche Modelle (artificial neural networks)	22
3.2.1	Das Modell eines mehrschichtigen Vorwärtsvermittlungsnetzes	27
3.2.2	Die Mathematik eines mehrschichtigen Vorwärtsvermittlungsnetzes	35
3.2.3	Algorithmus zur Variation der Anzahl der versteckten Knoten	41
3.2.4	Das Testen eines Netzwerkes	43
4	Das DELPHI-Experiment	45
4.1	Die supraleitende Spule	46
4.2	Die Spurkammern	50
4.3	Die elektromagnetischen Kalorimeter	52
4.4	Die Flugzeitzähler (Time Of Flight counters)	54
4.5	Die Müonkammern	54
4.6	Die Kontroll-, Auslese- und Triggersysteme	55
4.7	Die Monte Carlo Simulation	57
4.8	Die Rekonstruktion der Ereignisse	59

INHALTSVERZEICHNIS

5	Studie zur Auswahl von b-Quark-sensitiven Variablen	63
5.1	Die Monte-Carlo-Daten	64
5.2	Die untersuchten Variablen	66
5.2.1	Verschiedene Variablensätze	74
6	Ergebnisse und Interpretation	75
6.1	Zusammenfassung	87
A	Literaturhinweise	89
	Abbildungsverzeichnis	90
	Tabellenverzeichnis	90
	Literaturverzeichnis	92

Kapitel 1

Einführung

Wohl seit seiner Bewußtwerdung ist der Mensch auf der Suche nach Antworten über das Wesen seiner Umwelt. Die Fragen, die sich ihm immer wieder stellen sind :

- Wie ist die Welt entstanden und woraus ist sie aufgebaut ?
- Was sind die kleinsten Bausteine, so es sie überhaupt gibt ?
- Was hält die Welt im Innersten zusammen ?

Mit der systematischen Suche nach Antworten auf diese grundlegenden Fragen befaßt sich die moderne Elementarteilchenphysik. Der synonym verwendete Begriff Hochenergiephysik hat seinen Ursprung darin, daß zum Vorstoß zu immer kleineren Strukturen immer höhere Energien notwendig sind. Das Bild, welches sich der Menschen von der ihn umgebenden Welt macht, war und ist einem ständigen Wandel unterworfen. Betrachten wir die Grundfesten der Physik, wie sie sich uns heute darbieten, so ist die Rede von elementaren Teilchen und den zwischen ihnen wirkenden Kräften. Die elementaren Teilchen lassen sich zu drei Teilchenfamilien gruppieren, siehe Tabelle 1.2, welche jeweils wieder in Leptonen und Quarks aufgeteilt sind. Es gibt sechs verschiedene Quarksorten (Flavours), wobei der direkte experimentelle Nachweis des top-Quarks noch aussteht. Korrespondierend zu den Quarks gibt es die drei Leptonen Elektron, Müon und Tau, denen jeweils ein Neutrino zugeordnet ist.

Kapitel 1. Einführung

Diese Grundbausteine, sowie die aus ihnen zusammengesetzten größeren Strukturen, können untereinander über Mittelteilchen (Bosonen) wechselwirken. Von den vier Grundkräften, oder besser Wechselwirkungen, Tabelle 1.1, ist die uns im Alltag vertrauteste die Gravitation. Obwohl sie als die relativ schwächste der vier Wechselwirkungen im makroskopischen Bereich die dominierende Rolle spielt, konnte sie bisher noch nicht in einer umfassenderen Theorie mit den anderen Grundkräften verknüpft werden. Sie wird durch die Gesetze von Newton und durch die allgemeine Relativitätstheorie Albert Einsteins beschrieben. Die Quantenelektrodynamik (QED), welche die elektromagnetische Wechselwirkung beschreibt, konnte in der Synthese mit der schwachen Wechselwirkung zum Glashow-Weinberg-Salam (GWS) Modell der elektroschwachen Wechselwirkung zusammengefaßt werden. Die Theorie der starken Wechselwirkung ist die Quantenchromodynamik, die zusammen mit der GWS-Theorie das sogenannte Standardmodell bildet. Am Ringbeschleuniger LEP, an dem Elektronen und Positronen zur Kol-

Wechselwirkung	physikalisches Phänomen	Reichweite	relative Stärke ($\hbar = c = M_P = 1$)	Mittelteilchen (Bosonen)
starke	Kopplung zwischen Quarks und Gluonen	effektiv kurz (10^{-13} cm)	1	acht Gluonen (g)
elektromagnetische	Atomkräfte Optik Elektrizität	lang ∞	1/137	Photon (γ)
schwache	β -Zerfall	extrem kurz (10^{-16} cm)	10^{-5}	W^{\pm} - und Z^0 -Boson
Gravitation	Planetenbewegung	lang ∞	$6 \cdot 10^{-39}$	Graviton (?)

Tabelle 1.1: Die vier fundamentalen Wechselwirkungen

lision gebracht werden, versucht man auf einige der vielen noch offenen Fragen über die Natur der elementaren Teilchen und ihren Wechselwirkungen Antworten zu finden. Eine wichtige Rolle spielt dabei die Physik des 'bottom-'(Grund,Boden) oder auch 'beauty-'(Schönheit) Quarks. Dieses kurz b -Physik genannte Gebiet beschäftigt sich bei LEP, um einige Beispiele zu nennen, mit

- der Messung der partiellen Breite des Z^0 in das b ,

-
- der Bestimmung der Lebensdauer der B -Mesonen,
 - der Untersuchung der Vorwärts-Rückwärtsasymmetrien für die b -Quark-erzeugung,
 - der Untersuchung des Übergangs von B^0 -Mesonen in ihre Antiteilchen \bar{B}^0 ($B^0\bar{B}^0$ -Mixing), durch Austausch von mindestens zwei schweren W -Bosonen.
 - Fragmentationsstudien, das heißt dem Studium des Übergangs vom Quark zum Hadron,
 - dem Studium der B -Zerfälle,
 -

Um b -Physik betreiben zu können, ist es notwendig, Ereignisse an denen ein b -Quark beteiligt ist, zu erkennen und zu markieren¹. Thema dieser Arbeit ist es, das b -Quark-Tagging mittels neuronaler Netzwerke unter Verwendung rein topologischer Observablen durchzuführen. Der Begriff topologische oder globale Ereignis-Observable bedeutet in diesem Zusammenhang, daß keine explizite Teilchenidentifikation (d.h. e, μ, π ...-Erkennung) benutzt wird um z.B. anhand typischer Zerfallskanäle eines Teilchens eine Ereignisklassifikation vorzunehmen. Vielmehr wird das Gesamtereignis betrachtet und anhand von Charakteristika auf Ereignis- und Jetlevel eine Entscheidung getroffen. Die Schwierigkeit der expliziten Identifikation eines Ereignisses liegt unter anderem darin begründet, daß nur die Endzustände eines Ereignisses beobachtet werden können und auch diese nur unvollständig, was in Detektorakzeptanzen, Effizienzen, Meßfehlern etc. begründet liegt. Der perfekte Detektor, der 'alles' sehen kann, ist nicht existent. Da von 10^{27} Neutrinos die ein Nukleon durchqueren nur eines wechselwirkt, sind z.B. ν 's prinzipiell 'nicht' zu sehen. Die in dieser Arbeit verwendeten Ereignismengen enthalten keine neutralen Spuren ($\gamma, K^0, n...$), da das e.m. Kalorimeter (HPC) noch nicht die erhoffte Energieauflösung erbringt und so die Messung von neutralen Teilchen zu unzuverlässig ist. Wie jeder andere Detektor auch verliert DELPHI konstruktionsbedingt z.B. durch sogenannte Löcher an den Nahtstellen zweier Detektorkomponenten und durch fehlerhafte Rekonstruktion von einzelnen Spuren einige Teilchen. Zwischen den im beobachteten Endzustand vorhandenen Teilchen und dem Primärprozeß,

¹engl. to tag = markieren, etikettieren ; daher b -Quark Tagging

auf den man dann zurückschließen muß, können lange Zerfallsketten liegen, die eine Informationsreduktion bewirken.

Die 'kleinen Informationsfragmente', die im Endzustand noch vorhanden sind, müssen durch ein geeignetes Verfahren 'herausgepickt' und bewertet werden, damit eine Klassifikation doch noch möglich ist. An diesem Punkt kommen die neuronalen Netzwerke ins Spiel.

Künstliche neuronale Netzwerke (N.N.) finden in zunehmendem Maße Anwendung in der Informationsverarbeitung, speziell zur Mustererkennung in den verschiedensten Bereichen (z.B. Sprach- und Handschriftenerkennung). Ein Typ von N.N., die vorwärtsvermittelnden sind besonders geeignet, wenn es gilt eine Grundgesamtheit nach speziellen Kriterien in Unterklassen aufzuspalten. Das b -Quark Tagging stellt ein solches Problem dar. Die Grundgesamtheit der hadronischen Z^0 -Zerfälle soll in die Unterklassen der b -Quark Ereignisse und die der leichten Quarks aufgeteilt werden. Da die Anwendung von neuronalen Netzwerken im allgemeinen und insbesondere in der Hochenergiephysik noch sehr jung ist, existieren nur sehr wenige Erfahrungen über ihre Handhabung.

Die Anwendung einer solchen Tagging-Methode ist in drei Phasen unterteilt.

- Phase 1 : Das Konditionieren eines Netzwerkes auf das Erkennen von b -Quark Ereignissen anhand von topologischen Größen aus Monte-Carlo-Daten.
- Phase 2 : Die Leistungsverifikation des N.N. mit statistisch unabhängigen Monte-Carlo-Daten.
- Phase 3 : Die Anwendung auf reale Daten mit unbekannter Klassenaufteilung aus dem Experiment.

Im Rahmen dieser Arbeit wurden hauptsächlich die beiden ersten Phasen durchgeführt, während die ausführliche Behandlung der dritten Phase einer Nachfolgearbeit überlassen bleibt.

In den nachfolgenden Kapiteln wird die zugrundeliegende Physik, das Experiment, sowie die Grundlagen von Netzwerken dargestellt. Es schließt sich eine Beschreibung der benutzten Größen an. Im abschließenden Kapitel werden die mit ihnen erzielten Ergebnisse dargestellt, interpretiert und zusammengefaßt.

Tabelle 1.2: Das Standardmodell : Es werden die Quarks, Leptonen und Bosonen, ihre Ladungszustände, sowie ihre Wechselwirkungen mit den drei Grundkräften des Standardmodells und mit der noch nicht einbezogenen Gravitation dargestellt.

	Quarks		Leptonen							
erste Generation	u (up)	d (down)	e	ν_e						
zweite Generation	c (charm)	s (strange)	μ	ν_μ						
dritte Generation	t (top)	b (bottom)	τ	ν_τ						
elektrische Ladung	+2/3	-1/3	-1	0						
Farbladung	'rot'	'rot'	farblos	farblos						
	'gelb'	'gelb'								
	'blau'	'blau'								
					Bosonen	Masse (GeV)	Farb- ldg.	e.m. Ldg.	schw. Ldg.	
(starke) Farbkraft	⊗	⊗			Gluonen	0	⊗			
elektromagnetische Kraft	⊗	⊗	⊗		Photon	0				
schwache Kraft	⊗	⊗	⊗	⊗	W^\pm , Z^0	80, 91		± 1 , 0	⊗	
Gravitation	⊗	⊗	⊗	⊗	'Graviton'	0				

Kapitel 2

Theoretische Beschreibung

Wie schon in der Einführung erwähnt, beschreibt das sogenannte Standardmodell die elektromagnetische, die schwache und die starke Wechselwirkung. Ich möchte im folgende eine kurze Übersicht über dieses Modell geben.

2.1 Das Standardmodell

Der mathematische Formalismus, durch den im Standardmodell die drei auf Quarks und Leptonen wirkenden Kräfte beschrieben werden, ist prinzipiell derselbe. Es handelt sich um eichinvariante Quantenfeldtheorien oder kurz Eichtheorien. Die Hoffnung, die dahinter steht, ist, daß alle Kräfte nur verschiedene Aspekte einer einzigen 'vereinheitlichten' Wechselwirkung sind. Eine renormierbare Eichtheorie wurde erstmals in der Quantenelektrodynamik (QED), der Theorie der elektromagnetischen Kraft, erfolgreich eingeführt. Die dabei benutzte unitäre Eichgruppe $U(1)$ führt zu einem masselosen Eichboson. Das Photon, welches als dieses Eichboson identifiziert wurde, koppelt mit der Stärke $e = \sqrt{4\pi\alpha}$ an die elektrisch geladenen Konstituenten der Gruppe.

Fast analog dazu läßt sich die schwache W.W. durch die Gruppe $SU(2)$ beschreiben. Dabei spielt der schwache Isospin die Rolle, die die elektrische Ladung in der QED spielt. Schließlich gelang die Vereinigung der beiden Wechselwirkung zu einer einheitlichen Beschreibung. Die nach ihren Begründern benannte Glashow-Weinberg-Salam Theorie (GWS), ist die

Theorie der elektroschwachen Wechselwirkung. Die ihr zugrundeliegenden Gruppe $SU(2) \times U(1)$ erfordert vier Eichbosonen, welche sich in Form der realen Mittelteilchen W^+, W^-, Z^0 der schwachen und dem γ der elektromagnetischen Wechselwirkung im Experiment nachweisen lassen. Die W^+, W^-, Z^0 -Massen werden in der Theorie durch Anwendung des Prinzips der spontanen Symmetriebrechung erzeugt. Dieser sogenannte Higgsmechanismus erzwingt die Existenz mindestens eines zusätzlichen Bosons, das Higgs-Boson, dessen Nachweis allerdings noch aussteht.

Die Quantenchromodynamik (QCD) bildet eine weitere Komponente des Standardmodells. Die QCD wurde, nachdem Gell-Mann das Quarkmodell und u.a. Feynmann das Partonmodell entwickelt hatte, wie die QED als Eichtheorie aufgebaut. Die verwendete Eichgruppe ist die im Gegensatz zur $U(1)$ der QED nicht abelsche $SU(3)$ -Gruppe. Als Konsequenz davon sind die Bosonen nicht nur Überträger sondern auch Teilnehmer der Wechselwirkung, das heißt, daß sie auch untereinander koppeln können. Die Bosonen bilden ein Oktett, die acht Gluonen, welche die starke oder Farbwechselwirkung zwischen den Quarks vermitteln. Die Quarks tragen einen zusätzlichen Farbfreiheitsgrad, das heißt jedes Quark kommt in drei Farbladungszuständen vor, die in Analogie zur Farbenlehre als rot, grün und blau bezeichnet werden. Nach dem Quarkmodell lassen sich die Mesonen und Baryonen, welche aus zwei bzw. drei Quarks bestehen, in Form von Multipletts systematisieren. Da die Mesonen nach Außen immer farbneutral erscheinen, sind nur solche Quarkkombinationen zulässig, bei denen sich die Farbladungen gegenseitig aufheben. Das ist Farbe und Antifarbe ($q_{Farbe} \bar{q}_{\bar{Farbe}}$) für Mesonen oder drei verschiedenen Farben bzw. Antifarben ($q_r q_g q_b$ bzw. $\bar{q}_r \bar{q}_g \bar{q}_b$) in Baryonen, was analog zur Farbtheorie jeweils einen weißen oder farblosen Zustand ergibt. Eine weitere Besonderheit ist, daß die Quarks nicht als freie Teilchen sondern nur innerhalb von Hadronen auftreten, da sie dem sogenannten Quarkeinschluß (Confinement) unterliegen. Die Confinement-Hypothese trägt der Tatsache Rechnung, daß in der Natur offenbar nur farbneutrale Teilchen auftreten.

Zusammengefaßt ergibt sich folgendes Bild. Die elementaren Teilchen lassen sich in zwei Gruppen aufteilen :

- Fermionen mit halbzahligem Spin, die die Grundbausteine der Materie darstellen und
- Bosonen mit ganzzahligem Spin die die Wechselwirkung vermitteln.

Kapitel 2. Theoretische Beschreibung

Zu den elementaren Fermionen gehören die 6 verschiedenen Quarkflavour (udcsbt) in je drei verschiedenen Farbladungszuständen und die sechs Leptonen (e^- , ν_e , μ , ν_μ , τ , ν_τ), sowie die entsprechenden Antiteilchen. Diese Teilchen lassen sich in drei Generationen oder Familien aufteilen. Unsere Umwelt scheint nur aus den Teilchen der ersten Generation aufgebaut zu sein. Die beiden anderen Generationen sind offenbar nur schwerere Kopien der ersten und konnten bisher nur im Laborexperiment nachgewiesen werden. Durch die Ergebnisse von LEP konnte eine mögliche vierte Generation noch schwererer Teilchen ausgeschlossen werden, falls sie ein leichtes Neutrino haben soll. Zu den Feldquanten zählen

- das Photon der elektromagnetischen Wechselwirkung,
- die Bosonen der schwachen Kraft $W^+W^-Z^0$ (schwere intermediäre Vektorbosonen),
- die acht Gluonen der starken Wechselwirkung,
- mindestens ein hypothetisches skalares Higgsboson, welches von der Theorie zum Zwecke der Massenerzeugung vorhergesagt wird, jedoch kein Eichboson ist, und außerhalb des Standardmodells das ebenfalls noch nicht nachgewiesene
- Graviton, welches die Gravitation überträgt.

Tabelle 1.2 zeigt im Überblick die elementaren Fermionen und Bosonen.

2.2 Grundlagen und die daraus entwickelten Tagging-Methoden

Die aus Experimenten bei niederen Energien bekannten grundlegenden Kriterien zur Quarkselektion seien hier stichwortartig erwähnt. Es erfolgte die Separation von :

- b -Quarks über B -Mesonen und über Flugstrecken,
- c -Quarks über den Nachweis durch D -Mesonen,
- s -Quarks durch führende Kaonen während
- u - und d -Quarks nicht weiter getrennt werden konnten.

Bei LEP-Energien können die Quarks u, d, s, c und b erzeugt werden. In der vorliegenden Arbeit sollen, wie bereits oben erwähnt, Ereignisse an denen b -Quarks beteiligt sind durch eine geeignete Taggingmethode selektiert werden.

Bei harten e^+e^- -Stößen im Bereich um 91 GeV (LEP- Energien), ist der in Abbildung 2.1 gezeigte Prozeß vorherrschend. Das Elektron-Positron-Paar geht über ein Z^0 in ein Fermion-Antifermion-Paar über. Da wir uns auf der Z^0 -Resonanz befinden, ist im Gegensatz zu Experimenten im niederen und mittleren Energiebereich (z.B. DORIS), der ebenfalls mögliche Übergang über ein Photon stark unterdrückt. Die bestimmenden Z^0 -Zerfallskanäle sind [PD 90] :

70.9 %	Z^0	$\rightarrow q\bar{q}$
3.2 %	Z^0	$\rightarrow e^+e^-$
3.3 %	Z^0	$\rightarrow \mu^+\mu^-$
3.3 %	Z^0	$\rightarrow \tau^+\tau^-$
19.2 %	Z^0	\rightarrow 'unsichtbare' Zerfälle $\nu\bar{\nu}$

Der hier wichtige Z^0 -Zerfallskanal ist der in ein Quark-Antiquarkpaar, welches schließlich in einen multihadronischen Endzustand übergeht. Von diesen etwa 71 % entfallen circa 21 % auf $b\bar{b}$ -Ereignisse, das sind 14.6 % bezogen auf die Gesamtheit der Z^0 -Zerfälle. Diese $b\bar{b}$ -Ereignisse zu erkennen, ist die

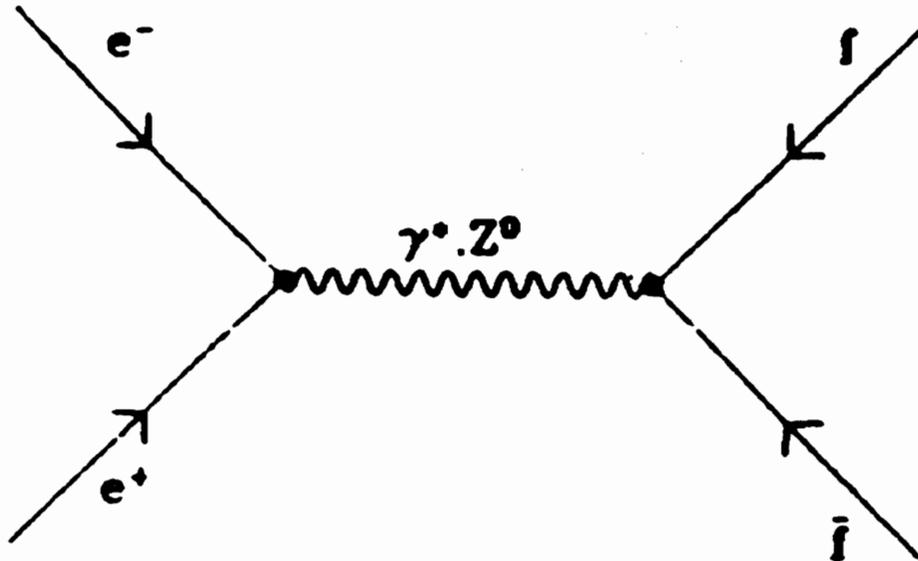


Abbildung 2.1: Z^0 -Graph

gestellte Aufgabe, wobei vorselektierte multihadronische Ereignisse, wie in Kapitel 5 erläutert, den Ausgangspunkt darstellen. Dies bedeutet aus den multihadronischen Ereignissen sollen die 21 % $b\bar{b}$ möglichst vollständig und sauber herausgefiltert werden.

Die im Zerfall primär erzeugten Quarks bewegen sich als quasi freie Teilchen in entgegengesetzte Richtungen. Dabei steigt die Energiedichte im sie verbindenden Farbfeld so weit an, bis weitere $q\bar{q}$ -Paare aus dem Vakuum erzeugt werden können. Es setzt die Bildung von Hadronen ein. Auf diesen Fragmentationsprozeß wird im nächsten Unterkapitel eingegangen. Die aus der Fragmentation und dem anschließenden Zerfall hervorgehenden Teilchen bilden sogenannte Jetstrukturen. Unter einem Teilchenjet, kurz Jet genannt, versteht man eine Gruppe von Teilchen mit einer gemeinsamen Hauptflugrichtung (Jetachse). Die Teilchen eines Jets befinden sich innerhalb eines relativ kleinen Raumkegels von etwa 30° und besitzen bezüglich der Jetachse kleine Transversalimpulse (< 350 MeV). Das Gegenteil eines durch Jets strukturierten Ereignisses wäre ein solches mit sphärischer Ge-

stalt. In der vergleichsweise hohen Masse des b -Quarks ist die Mehrheit der Charakteristika der b -Ereignisse begründet. Sie zeichnen sich durch eine ausgeprägte Jetstruktur, durch hohe Multiplizitäten, höhere Spurenergien und hohe transversale Impulskomponenten der Zerfallsteilchen aus. Diese allgemeinen topologischen Eigenschaften sind ein Ansatzpunkt zur Entwicklung einer Taggingmethode. Einen zweiten Ansatzpunkt bildet der schwache Zerfall des B -Mesons. Es wird dabei das Auftreten eines signifikanten Sekundärvertex bzw. der Nachweis von Leptonen aus dem semileptonischen Zerfall ausgenutzt. Im Folgenden soll der Weg von den physikalischen Grundlagen zu den einzelnen erwähnten Taggingmethoden nachvollzogen werden.

2.2.1 Die Fragmentation

Die primär beim Zerfall des Z^0 erzeugten Quarkpaare fragmentieren in Hadronen. Dieser Vorgang der starken Wechselwirkung ist mit dem Ansatz der Störungsrechnung nicht mehr berechenbar. Man ist daher auf phänomenologische Modelle beschränkt. Alle existierenden Modelle beruhen auf Wahrscheinlichkeitsaussagen und sind in ihrem Ablauf von iterativem Charakter. Das heißt sie gehen nur von einem (oder mehreren) zugrundeliegenden einfachen Prozeß aus, die wiederholt ausgeführt werden. In jeder Iteration wird das Erzeugen neuer Flavours, die Aufteilung von Energie und Impuls nach Wahrscheinlichkeitsfunktionen durchgeführt.

Es sollen hier als die beiden meistverwendeten Modelle das der unabhängigen Fragmentation (Independent Fragmentation) und die Stringfragmentation angesprochen werden. Im Modell der unabhängigen Fragmentation, welches der älteste Ansatz zur Beschreibung des Fragmentationsprozesse ist, wird die Annahme gemacht, daß die Fragmentation eines Systems von Partonen als inkohärente Summe der voneinander unabhängigen Fragmentationsprozesse der einzelnen Partonen beschrieben werden kann. Der Ausgangspunkt ist ein Quark mit wohl definierter Energie und Impuls. Der sich wiederholende Prozeß ist $Quark \rightarrow Hadron + verbleibendes Quark$. Jedes Parton wird separat behandelt. Es fragmentiert unabhängig von den anderen, indem aus dem Vakuum ein Quark-Antiquark Paar erzeugt wird und es mit dem Antiquark ein Meson bildet. Dieser Vorgang wiederholt sich mit dem jeweils verbleibenden Quark, bis die restliche Energie nicht mehr zur erneuten Paarbildung ausreicht.

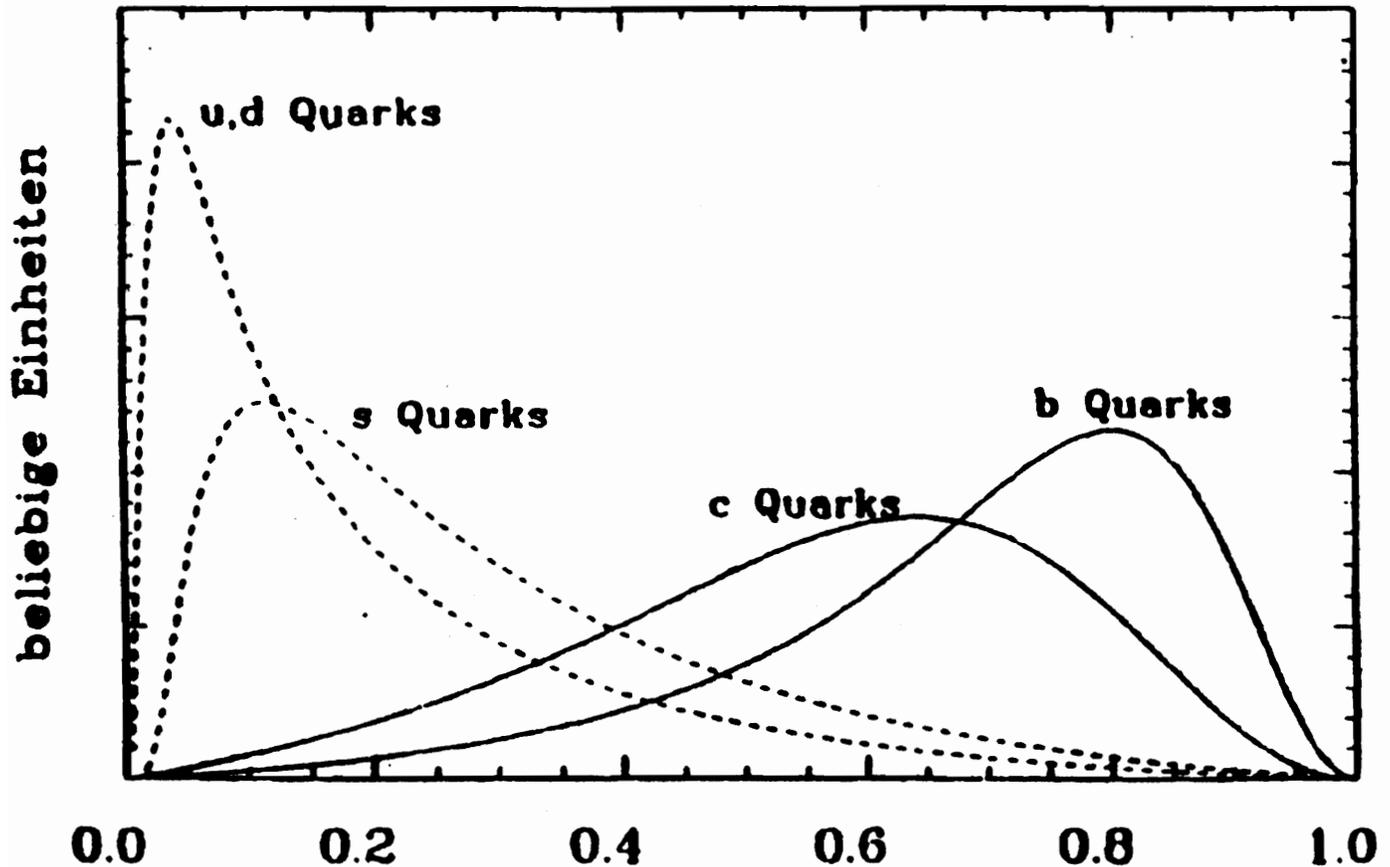


Abbildung 2.2: Die Fragmentationsfunktionen für die verschiedenen Quarkflavour

Das in der QCD angenommene linear Confinement der Quarks bei größeren Abständen bildete den Ausgangspunkt des Stringfragmentationsmodell. Die Vorstellung ist im einfachsten Falle, daß sich zwischen den Partonen ein lineares Farbfeld ('colour flux tube' oder String) ausbildet, das durch das gegenseitige sich Entfernen der Partonen an potentieller Energie gewinnt. Ist die Energie groß genug, bricht der String an mehreren Stellen und es bilden sich Quark-Antiquark Paare. Dieser Prozeß : *String* \rightarrow *Hadron* + *verbleibender String* wiederholt sich wie in der unabhängigen

2.2. Grundlagen und die daraus entwickelten Tagging-Methoden

Fragmentation, bis die verbleibende Energie nicht mehr dazu ausreicht ein weiteres $q\bar{q}$ - Paar zu erzeugen.

In beiden Modellen beschreiben sogenannte 'Fragmentationsfunktionen' den Anteil des Longitudinalimpulses des primären Partons, der auf das führende Meson übergeht und damit auch den noch verbleibenden Anteil. In Abbildung 2.2 sind typische Fragmentationsfunktionen der verschiedenen Quarkflavours zu sehen. Die Beobachtung zeigt, daß die Quarks mit zunehmender Masse immer härter fragmentieren. Eine harte Fragmentation heißt, daß der Großteil des longitudinalen Impulses des Partons an das erste¹ entstehende Hadron weitergegeben wird. Dies spiegelt sich in den Funktionen aus Abbildung 2.2 wieder. Die Massenabhängigkeit der Fragmentation kann man sich anschaulich als kinematischen Effekt erklären. Ein schweres Quark hat am Anfang die Strahlenergie und wird beim "Aufpicken" eines leichten Quarks aus dem virtuellen Quarksee nur wenig aus der Bahn gebracht, vergleichbar dem Stoß einer Kugel mit großer Masse mit einer leichten Kugel. Eine deutliche Auswirkung auf die Flugbahn ist jedoch zu sehen, falls zwei annähernd gleiche Massen (Quarks) aufeinander treffen. Als Folge der harten Fragmentation des b -Quarks sollten sich b -Quark-Jets in ihren Eigenschaften von den Jets leichter Quarks deutlich unterscheiden. Die hohe b -Quarkmasse und damit die harte Fragmentation hat also direkten Einfluß auf die Impulsspektren. Der anschließende Zerfall des B -Mesons verursacht höhere Transversalimpulse in den entstehenden Zerfallsprodukten, was gleichfalls eine direkte Folge der hohen Masse ist.

Die Methode, die sich aus dem oben angeführten Grundlagen entwickeln läßt, besteht darin die so erzeugten Unterschiede in Observablen nachzuweisen und sie zur Separation der b -Ereignisse auszunutzen. Dies kann durch die direkte Messung von Impulskomponenten und/oder auf einem höheren Level durch Jet- und Ereignisvariablen geschehen. Der limitierende Faktor bei diesem Vorgehen ist jedoch, daß nur der hadronische Endzustand beobachtet werden kann, wie er sich nach Ablauf von unter Umständen langen Zerfallsketten präsentiert; dies gilt insbesondere für das schwere B -Meson. Der Zerfall der instabilen Teilchen hat den Effekt einer Informationsreduktion, das heißt die ursprünglichen Unterschiede werden stark verwischt. Bei diesem Prozeß werden nicht nur die Unterschiede zwischen den Quarkflavours sondern auch zwischen Quark und Antiquark verwischt. Bildlich ausgedrückt verschwinden die Unterschiede "im hadronischen Sumpf". Daß

¹Auch leading particle (= führendes Teilchen) genannt.

die Fragmentationseigenschaften nicht völlig verwischt werden, sondern in einigen physikalischen Größen doch noch Unterschiede meßbar sind, ist die Voraussetzung, daß dies überhaupt zu einer Methode der Trennung der Ereignisse in verschiedene Klassen werden kann. In Kapitel 6 jedoch wird anhand eines Beispiels gezeigt, daß es nicht möglich ist, durch einen einfachen Schnitt auf eine einzelne Größe eine gute Trennung zu erreichen. Es ist vielmehr ein Verfahren wie das der neuronalen Netzwerke nötig, das möglichst alle zur Verfügung stehende Information ausnutzt.

2.2.2 Die Zerfallslänge des B -Mesons

Der schwache Zerfall des B -Mesons ist ein universeller W -Zerfall. Im Standardmodell, wie es im Augenblick diskutiert wird, sind nur folgende Zerfälle möglich :

$$b \longrightarrow cW^-$$

$$b \longrightarrow uW^-$$

Das W^- koppelt anschließend universell an ein Fermionpaar. In Abbildung 2.3 ist Vorgang im Spectator Modell dargestellt. Das quasi freie b -Quark zerfällt in ein W^- und ein c oder u -Quark. Das W^- zerfällt weiter in ein Lepton-Neutrino Paar oder in leichtere Quarks. Das zweite Quark des B -Mesons ist in diesem Modell nur als Zuschauer² beteiligt und bleibt unverändert. Abbildung 2.3 zeigt, das auch ein D -Meson, das heißt das in ihm enthaltene c -Quark entsprechend dem b -Quark eines B -Mesons schwach zerfallen kann. Betrachtet man die Werte der Quark-Mischungsmatrix, die Cabbibo-Kobayashi-Maskawa (CKM) Matrix, wie wir sie heute kennen, so ist praktisch eine Entkopplung³ des dritten Quark-Dubletts von den anderen beiden festzustellen. Dies manifestiert sich in einer relativ langen Lebenszeit des B -Mesons, die im Pikosekundenbereich ($11.8 \cdot 10^{-13} s$ [PD 90]) liegt. Da das Meson sich hier mit annähernd Lichtgeschwindigkeit bewegt, legt es bevor es zerfällt eine Wegstrecke, die Zerfallslänge genannt wird, von einigen mm zurück. Die Lebenszeit des D -Mesons ist im Vergleich dazu um etwa $1/3$ kürzer und entsprechend auch die dabei zurückgelegte Wegstrecke. Könnte man diese Strecke, welche wie gezeigt für das B -Mesonen

²engl. spectator

³Das Verhältnis der Matrixelemente V_{cb}/V_{ub} ist klein ($\sim 1/4$) [TH 85].

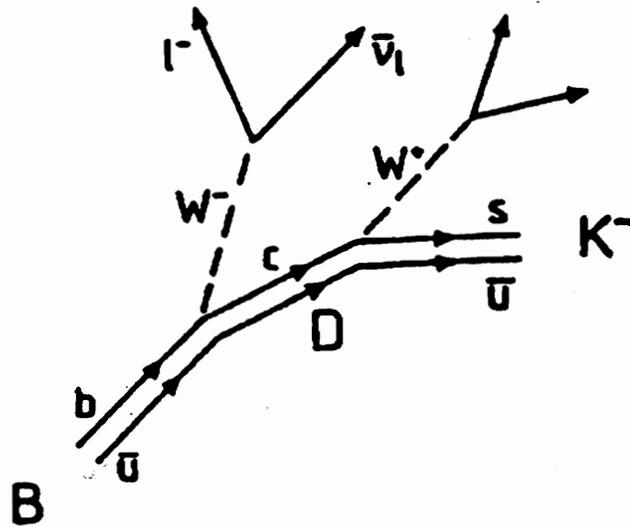


Abbildung 2.3: Der universelle W-Zerfall im Spectator-Modell am Beispiel eines B^- -Mesons ($b\bar{u}$). Das quasi freie b -Quark des Mesons zerfällt in ein c -Quark und ein W^- -Boson, welches weiter in ein Lepton-Neutrino-Paar zerfällt. Das c bildet mit dem in diesem Modell unbeteiligten \bar{u} ein D^0 -Meson. Das c -Quark kann anschließend über einen entsprechenden W -Zerfall weiter zerfallen, wobei ein weiteres Lepton-Neutrino-Paar entsteht.

signifikant größer ist, messen hätte man eine ausgezeichnete Methode zum b -Quark-Tagging.

In der Realität versucht man dieses über die Rekonstruktion eines Sekundärvertex, dessen Abstand zum Wechselwirkungspunkt die Zerfallslänge ist oder über die Messung des Impaktparameters. Die Geometrie, auf der die Rekonstruktion des Sekundärvertex beruht, zeigt Abbildung 2.4. Der Impaktparameter ist als kleinster Abstand einer Spur zum Hauptvertex definiert. Spuren die ihren Ursprung in einem Sekundärvertex haben, sollten sich durch einen im Mittel höheren Wert des Impaktparameters auszeichnen. Wie man sich geometrisch veranschaulichen kann, besteht ein direkter Zusammenhang zwischen dem Abstand des Sekundärvertex zum Hauptvertex und dem Impaktparameter, wodurch prinzipiell eine Trennung von Prozessen mit verschiedenen Zerfallslängen möglich ist. Eine solche Methode kann

jedoch nur erfolgreich sein, wenn die Zerfallslänge groß genug ist um überhaupt gemessen zu werden und ein entsprechend hochauflösender Detektor zur Verfügung steht. Mit dem im DELPHI-Experiment verwendeten Vertexdetektor, dessen geplante Auflösung in der Größenordnung von μm liegt, sollte dies prinzipiell gelingen.

Der Vertexdetektor wurde im Sommer 1990 erstmals in Betrieb genommen und erreichte erst 1991 seine volle Ausbaustufe. Deshalb und wegen der anfangs noch unzureichenden Detektorbeschreibung in den Monte-Carlo Programmen, sowie der noch nicht ausgereiften Auswertesoftware kam diese Methode für die vorliegende Arbeit leider nicht mehr zum tragen. Sie ist

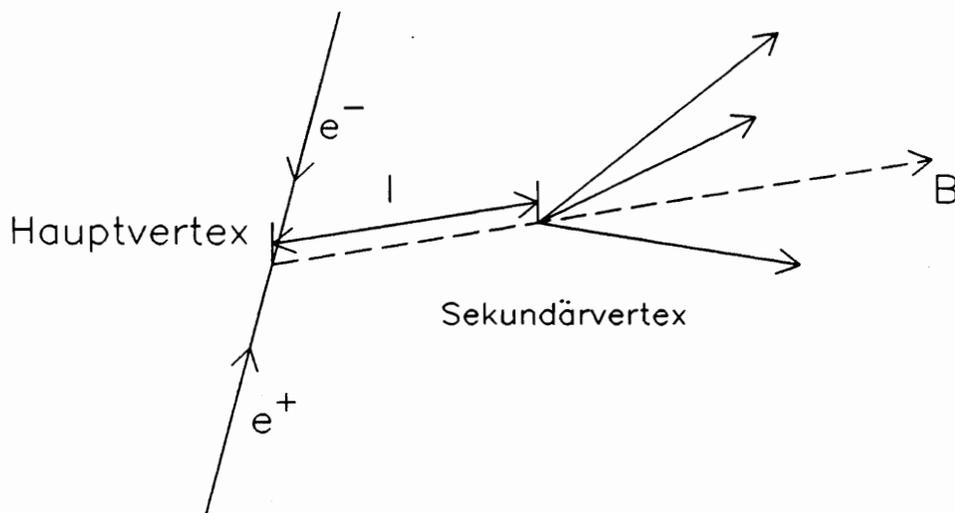


Abbildung 2.4: Geometrie des B -Zerfalls. Die Zerfallslänge l wird durch die dazu proportionale Lebenszeit τ des B -Mesons bestimmt, und beträgt einige mm .

jedoch, wie aufgezeigt wurde, sehr vielversprechend für die Zukunft.

2.2.3 Der semileptonische Zerfall des B -Mesons

Die zweite Möglichkeit den schwachen Zerfall des B -Mesons auszunutzen, besteht darin nicht auf die Geometrie des Zerfalls zu achten sondern auf

2.2. Grundlagen und die daraus entwickelten Tagging-Methoden

die Zerfallsprodukte. Wie schon bei kleineren Energien angewandt, wird der semileptonische Zerfall des B -Mesons benutzt um das b -Quark zu taggen, indem das dabei entstehende Lepton nachgewiesen wird. Etwa 20 % der hadronischen Ereignisse auf der Z^0 -Resonanz sind $b\bar{b}$ - und rund 17 % $c\bar{c}$ -Ereignisse. In beiden Fällen ist ein semileptonischer Zerfall in ein leichteres Quark, ein Lepton und ein Neutrino möglich. Das semileptonische Verzweigungsverhältnis von $b \rightarrow cl$ ist etwa 11 %, das von $c \rightarrow sl$ etwa 9 % [KU 89]. Aus diesen Werten ist zu ersehen das der semileptonische Zerfall des B -Mesons das häufigere Ereignis ist. Eine gute Trennung zwischen $b \rightarrow c$ einerseits und $c \rightarrow s$ bzw. der Zerfallskaskade $b \rightarrow c \rightarrow s$ andererseits ist aber erst durch die unterschiedlichen Transversalimpulsspektren der resultierenden Leptonen möglich. Die Leptonen aus dem b -Quark Zerfall zeigen ein wesentlich härteres transversales Impulsspektrum bezüglich der jeweiligen Jetachse als die Leptonen aus dem c -Zerfall, was in der hohen Masse des b -Quarks seine Ursache hat.

Die daraus entwickelte Tagging-Methode besteht aus der Identifikation von Leptonen, wobei ein Schnitt auf p_t die Absonderung der Leptonen aus c -Zerfällen, sowie derer aus weiteren Untergrundereignissen (z.B. Kaskadenzerfälle) ermöglicht. Man erhält so sehr reine b -Ereignissamples. Diese Methode kann jedoch nach den oben genannten Zahlen prinzipiell keine sehr große Effizienz liefern. Beschränkt man sich nur auf Elektronen oder nur auf Müonen so halbiert sich die Zahl der Ereignisse nochmals. Mit dem Nachweis von Leptonen mit einem hohen p_t hat man jedoch eine Methode die nach theoretischen Berechnungen [KU 89] sehr hohe Reinheiten verspricht, wie sie mit keiner der anderen Methoden im Augenblick erreichbar sind.

In dieser Arbeit wird das b -Tagging nur mit topologischen Größen durchgeführt. Die Unterscheidungskriterien haben ihre Ursache hauptsächlich in den verschiedenen Quarkmassen und den beschriebenen Fragmentationseigenschaften. Die verwendeten physikalischen Größen werden in Kapitel 5 beschrieben.

Kapitel 3

Neuronale Netzwerke

” Ein Wort vorweg: Im Wissensgebiet der *Neuronalen Netze* ist noch kein 'Königspfad', noch keine Methodologie bekanntgeworden, der/die objektiv zur erfolgreichen Realisierung führt. Aus diesem Grund feiert die empirische Vorgehensweise hier fröhliche Urstände’.”
(aus [KR 90] Seite 144).

Mit **neuronalen Netzwerken** hat in die Lehre der Informationsverarbeitung ein Begriff Einzug gehalten, der ursprünglich aus dem medizinisch biologischen bzw. neurophysiologischen Bereich kommt.

3.1 Biologische Vorbilder

Ein **Neuron** bezeichnet eine Nervenzelle. Sie ist die kleinste Einheit, aus der das Nervensystem einschließlich der Gehirne der Lebewesen auf unserem Planeten aufgebaut ist. Neuronen stehen über Fortsätze untereinander in informationellem Kontakt. In den so aufgebauten **Netzwerken** findet die Informationsverarbeitung und -speicherung statt. **Eingangsinformationen** erhält das biologische Netzwerk über die **Rezeptoren** des Organismus. Für äußerliche Reize sind das die Sinnesorgane und für Reize aus dem Körperinneren spezielle Zellen des Nervensystems. Das Ergebnis der Informations-

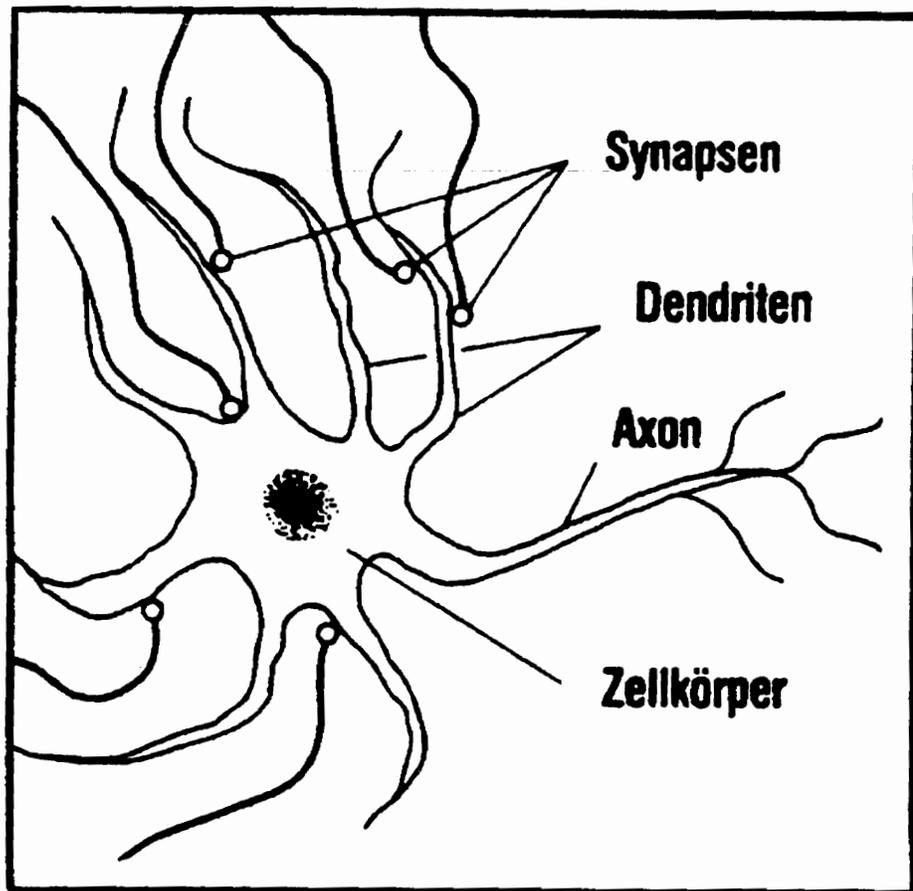


Abbildung 3.1: Das vereinfachte Schema einer Nervenzelle. Die von anderen Zellen kommenden Signale gelangen über die Synapsen in die Nervenzelle. Die Dendriten leiten die Eingangssignale über den Zellkörper, wo die Signalverarbeitung geschieht, zum Axon, welches das Ausgangssignal an nachfolgende Nervenzellen weiterleitet.

Kapitel 3. Neuronale Netzwerke

verarbeitung, das heißt der Ausgang des neuronalen Netzwerks, wird den Effektoren (z.B. Muskeln, Hormondrüsen ...) zugeleitet, welche dann entsprechend reagieren.

Ein biologisches Neuron läßt sich, wie Abbildung 3.1 zeigt, ganz grob in drei funktionale Bestandteile unterteilen :

- zentraler Zelleib (Soma)
- Dendriten und
- Axon

Vom zentralen Zelleib gehen Fortsätze, sogenannte Axone und Dendriten aus, die in mehr oder weniger großer Entfernung verdickte Enden bilden. Ein Neuron nimmt über zu den Dendriten gelangte Neurotransmitter Informationen von anderen Nervenzellen auf. Die Informationen werden nach bestimmten Regeln in der Zelle verarbeitet. Das Ergebnis wird über den Zellkörper und ein Axon an andere Neuronen weitergeleitet. Die Kontaktstelle zweier Neuronen, d.h. des Axons der presynaptischen Zelle und des Dendriten der postsynaptischen Zelle, wird Synapse genannt. Im Detail beruht die Reizweiterleitung auf elektrochemischen Reaktionen. Die Zeit für die Übertragung von einer Zelle zur nächsten ist kleiner als 1/1000 Sekunde.

Die Großhirnrinde des Menschen ist aus etwa 10^{11} - 10^{12} Neuronen mit durchschnittlich 10^3 Dendriten aufgebaut. Das gesamte Netzwerk enthält etwa 10^{15} synaptische Verbindungen ! Ein Vergleich der Funktionsweise von Gehirn und Computer zeigt die grundlegenden Unterschiede auf. Informationsverarbeitung mit neuronalen Netzwerken ist mit den Attributen **parallel, asynchron** und **stochastisch** verknüpft. Im Gegensatz dazu können herkömmliche Rechner mit **sequentiell, getaktet** und **deterministisch** umschrieben werden. Die Information wird nicht in genau spezifizierten Speicherbereichen abgelegt, wie dies in Rechenmaschinen geschieht, sondern ist inhärent über das ganze Netzwerk verteilt.

Der konzeptionelle Unterschied von Neuronalen Netzwerken und einer konventionellen von Neumann Rechnerarchitektur wirkt sich auf die differenzierte Leistungsfähigkeit in den unterschiedlichen Problemfeldern aus :

- Computer : präzise mathematische Berechnungen, sich wiederholende Algorithmen.

- Gehirn : symbolische Informationsverarbeitung (Mustererkennung), Assoziativspeicher, Informationsverarbeitung mit unvollständigen Eingangsdaten.

Trotz des verhältnismäßig einfachen Aufbaus eines einzelnen Neurons ist ihre Wechselwirkung untereinander im Gesamtsystem Gehirn, wegen der enormen Komplexität des Netzwerkes (noch?) nicht durchschaut. Über die prinzipielle Frage, ob es dem Menschen jemals vergönnt sein wird, sein eigenes Gehirn zu verstehen, sind sich Fachleute wie Philosophen auch heute noch uneinig.

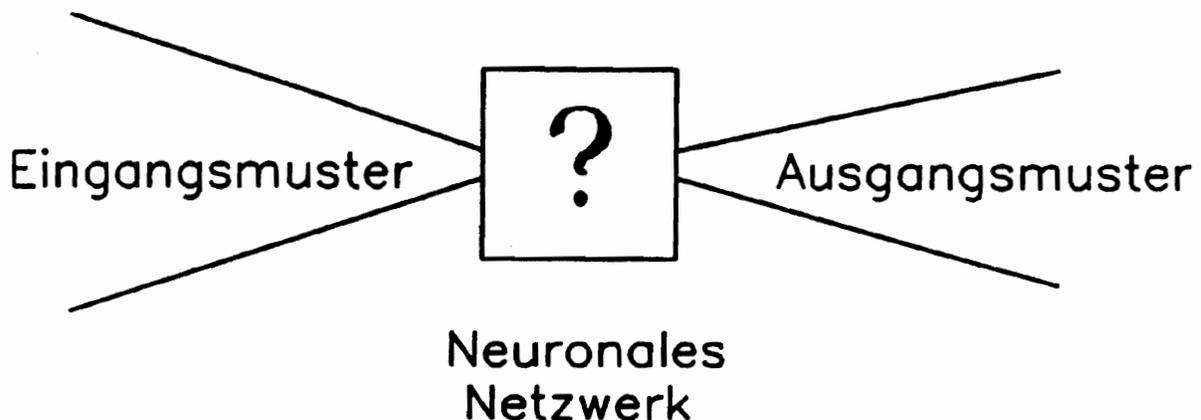


Abbildung 3.2: Das neuronale Netzwerk als 'black box'- Schema.

Die entwickelten künstlichen Modelle sind was Größe (Anzahl der Neuronen), Aufbau (Art der Vernetzung) und differenzierte Funktionsweise der Neuronen angeht, ganz grobe Annäherungen an das in der Natur angewandte Konzept der Informationsverarbeitung. Sie sind jedoch sehr hilfreich, wenn es um die Anwendung von neuronalen Netzwerken in der Forschung und Technik geht.

3.2 Künstliche Modelle (artificial neural networks)

Abbildung 3.2 zeigt die allgemeinste Darstellung des Systems als "black-box", die ein Eingangsmuster auf ein Ausgangsmuster abbildet. Um die Funktionsweise dieser "black-box" zu beschreiben, gibt es verschiedene mathematische Ansätze. Die verwendeten Komponenten sind jedoch im wesentlichen immer dieselben und lassen sich in die zwei Gruppen Hardware

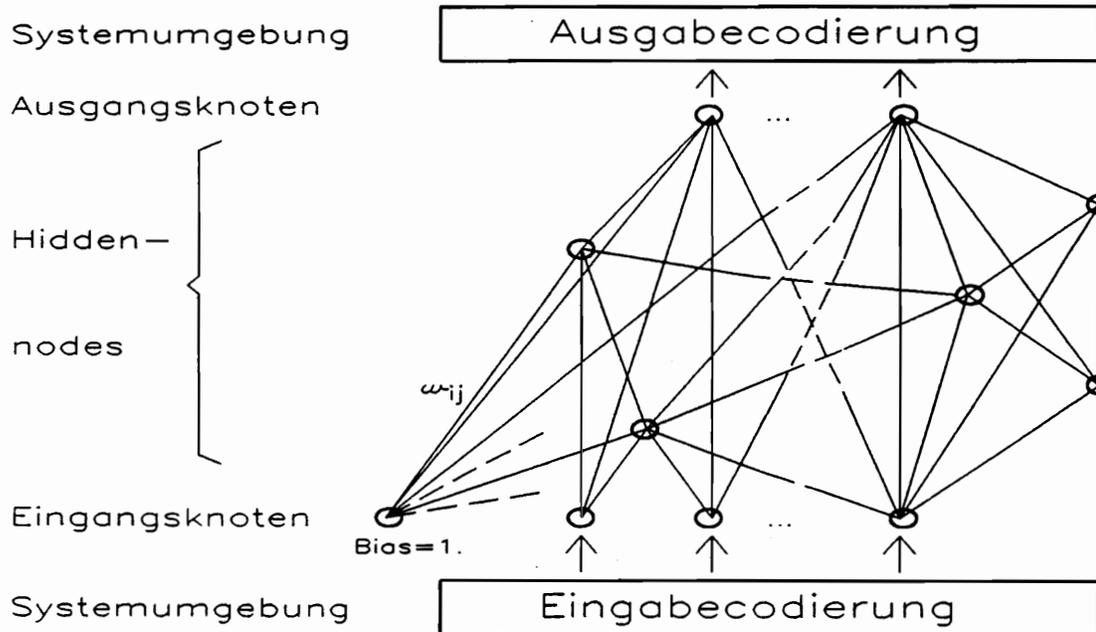


Abbildung 3.3: Hardwarekomponenten eines künstlichen neuronalen Netzwerkes. Die Kreise stellen die Knoten und die Linien die mit ω gewichteten Verbindungen zwischen ihnen dar.

und Software unterteilen :

- Hardwarekomponenten (siehe Abbildung 3.3)
 - Eingangsknoten¹ (input nodes), entsprechend den Rezeptoren

¹Bei künstlichen N.N. spricht man von Knoten statt von Neuronen, als den kleinsten strukturellen und funktionalen Elementen.

3.2. Künstliche Modelle (artificial neural networks)

- Knoten im Innern des Netzwerkes (hidden nodes)
 - Ausgangsknoten (output nodes)
 - Gewichtungsfaktoren (weights); sie bestimmen ob und wenn ja, wie stark zwei Knoten miteinander in Verbindung stehen. Positive Gewichte haben erregenden (excitatory) und negative Gewichte hemmenden (inhibitory) Einfluß auf den nachfolgenden Knoten.
 - Schwellwerte (bias), die ein Maß für die Tendenz eines Knotens i zur Aktivierung oder Deaktivierung darstellen, entsprechend einem Membranpotential im biologischen Vorbild. Ein Bias kann als eine Verschiebung der Aktivierungsfunktion auf der x -Achse angesehen werden, wodurch die 'interessanten' Stellen der Eingangsverteilung eines Knotens mit der Stelle, an der die Sigmoidfunktion ihre größte Steigung hat, in Übereinstimmung gebracht werden kann. Die Realisierung sieht meist so aus, daß die Schwellwerte θ_i ; das Gewicht der Verbindung zwischen einer immer voll aktivierten virtuellen Zelle und einer Zelle i sind.
 - Systemumgebung; sie beinhaltet hauptsächlich die Ein- und Ausgabecodierung, das heißt die Schnittstellenverwaltung zu Rahmenprogrammen in die das Netz eingebunden ist. Auch eventuelle notwendige Vorab-Transformationen (Normalisierung), sowie Voranalysen fallen unter diesen Punkt.
- Software
- Verknüpfungsmuster; es regelt wie die Knoten untereinander verknüpft sind.
 - Aktivierungsregel (activation rule); sie bestimmt aus den Eingangsgrößen das Aktivitätspotential (Aktivitätszustand) des Knotens. Gebräuchliche Aktivierungsregeln sind :
 - *Identität* oder *lineare Abbildung*
 - *Schwellwert-* oder *Treppenfunktion*, dazu gehört aus der Klasse der *semilinearen Aktivierungsfunktionen* auch die sogenannte *Sigmoidfunktion*, wie sie in dieser Arbeit verwendet wurde. Die Sigmoidfunktion bringt als willkommene Eigenschaft eine Limitierung des Aktivitätszustandes mit (hier zwischen 0 und 1), welches in ihrem asymptotischen

Verhalten im Unendlichen begründet ist (siehe Abbildung 3.4).

- *Stochastische Aktivierung.*

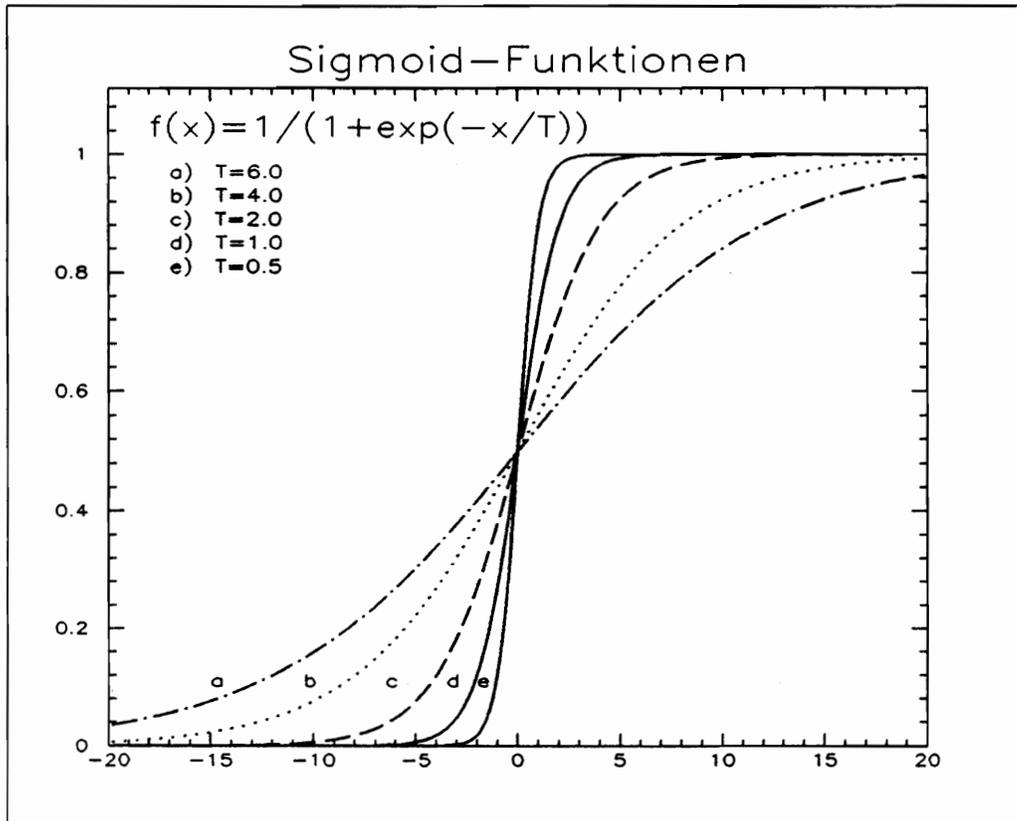


Abbildung 3.4: Sigmoid-Funktionen für verschiedene T-Werte. Je kleiner T ist, desto ausgeprägter ist das Schwellwertverhalten der Sigmoidfunktion.

- Ausgangsfunktion (output function); sie berechnet aus dem Aktivitätspotential des Knotens den Output.
- Gewichtsregel (Vermittlungsregel, propagation rule); sie bestimmt wie der Ausgang (Output) eines Knotens mit dem Gewichtungsfaktor der Verbindung zum Eingang (Input) des nächsten Knotens verknüpft wird. Bekannte Beispiele sind die gewichtete Summe und die multiplikative Vermittlungsregeln wie die *Sigma-Pi-Einheiten*.
- Lernregel (Adaptionsregel, learning rule); sie regelt die Änderung des in den Gewichten repräsentierten Wissens in der Lernphase.

Die Adaption umfaßt im einzelnen :

- Änderung der Gewichte bestehender Verbindungen.
- Auf- und Abbau von Verbindungen.
- Änderung der Netztopologie; z.B. Anzahl der Zwischenzellen.

Es werden grob zwei Gruppen von Lernregeln unterschieden:

- überwachtes Lernen (supervised learning); mit speziellen Trainingsmustern wird das Netzwerk auf die Anwendung trainiert.
- unüberwachtes Lernen (unsupervised learning); das Netzwerk versucht selbständig die präsentierten Eingangsvektoren oder Muster in Untergruppen zu klassifizieren.

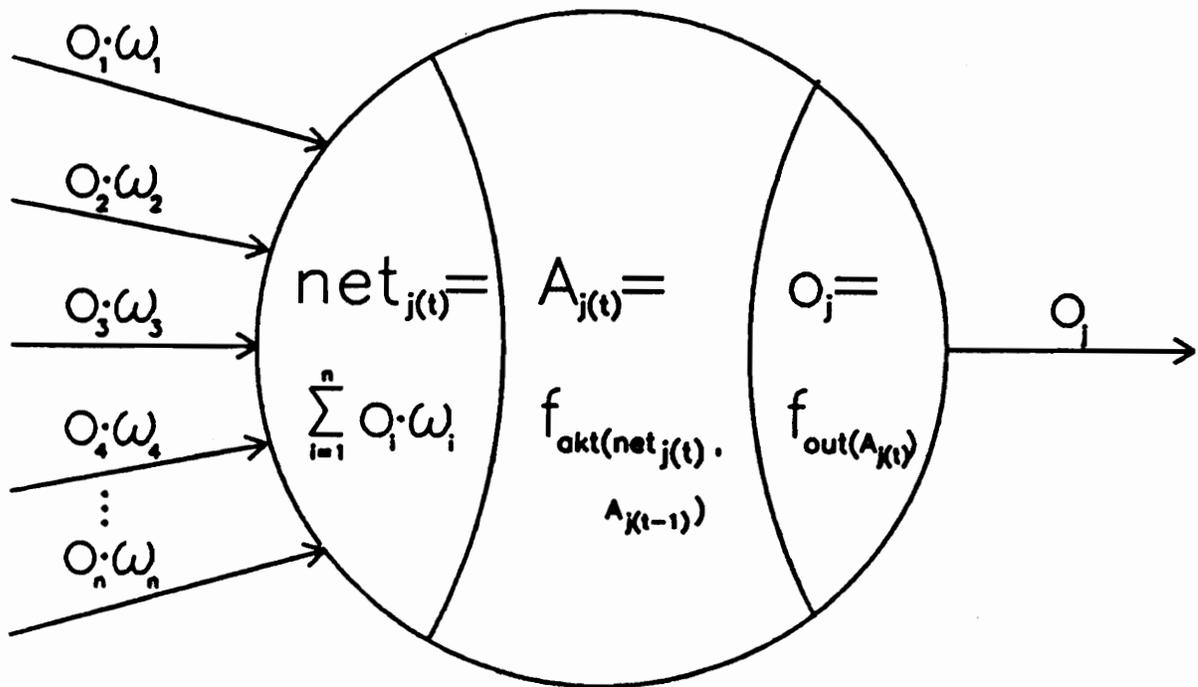
Die prinzipielle Funktionsweise eines Knotens, der einem Neuron im Gehirn entspricht, ist in Abbildung 3.5 zu sehen. Das Aktivitätspotential $A_{j(t)}$ eines Knotens zum Zeitpunkt t wird nach der Aktivierungsregel $f_{akt}(net_i(t), A_i(t-1))$ mit dem Aktivitätszustand und den Eingängen der Zelle als Argumente berechnet. Das Aktivitätspotential ist seinerseits wieder Argument der Ausgangsregel (output function) $f_{out}(A_i(t))$, die den Ausgangswert des Knotens bestimmt. Die Ausgangsfunktion besteht oft nur aus der Identität. Der Ausgang $o_{i(t)}$ des Knotens wird dann nach der Vermittlungsregel (propagation rule) $f_{i(o_{i(t)}, \omega_{ij})}$ mit dem jeweiligen Gewichtungsfaktor einer Verbindung verknüpft und als Eingang dem nächsten Knoten zugeleitet. Die Vermittlungsregel besteht in vielen Fällen aus dem inneren Vektorprodukt des Aktivierungsvektors $o_{i(t)}$ mit dem Gewichtsvektor ω_{ij} ; was allgemein unter dem Begriff gewichtete Summe bekannt ist. In den nachfolgenden Knoten wiederholt sich der ganze Vorgang von vorne.

Wie teilweise schon angedeutet, liegen die Unterschiede der verschiedenen Modelle in der Wahl der oben unter dem Punkt Software aufgelisteten Komponenten :

- Inhalt der Lernregel.
- Verbindungsstruktur: gerichtet Verbindungen, rekursive Verbindungen.
- Topologie; schichtorientierter Aufbau (layers), vollvernetzte Strukturen.

Kapitel 3. Neuronale Netzwerke

- Kontrollstrategie; synchroner/asynchroner, getakteter(gleichmäßig/ungleichmäßig) / kontinuierliche Berechnung (Update) der Gewichtungen.
- lineare/nicht lineare Funktionen als Aktivierungs- und Ausgangsfunktion.
- ...



Eingänge

Knoten

Ausgang

Abbildung 3.5: Funktionsweise eines Netzwerkknotens

Der Einsatz eines N.N. gliedert sich, wie in Abbildung 3.6 gezeigt, in die drei Phasen des Lernens, Testens und Anwendens.

3.2. Künstliche Modelle (artificial neural networks)

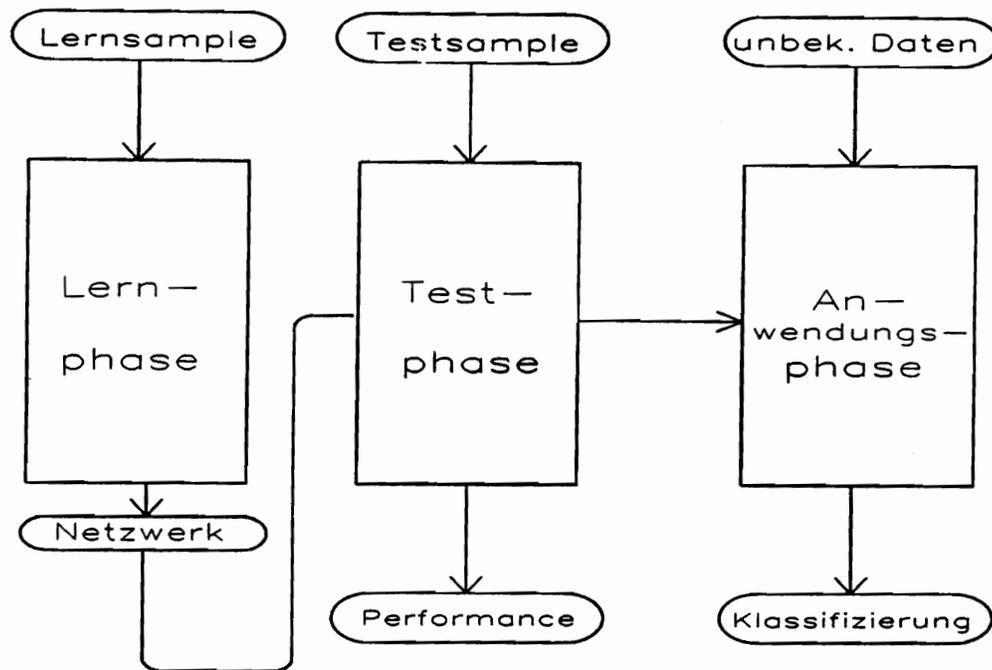


Abbildung 3.6: Die drei Phasen im Umgang mit neuronalen Netzwerken.

Im Folgenden wird nur noch auf die von David E. Rumelhardt zurückgehende Beschreibung eines Vorwärtsvermittlungsnetzes (**feed-forward-neuralnetworks**) eingegangen, da sie die Grundlage der in dieser Arbeit verwendeten Methode darstellt.

3.2.1 Das Modell eines mehrschichtigen Vorwärtsvermittlungsnetzes

Nach der Anfangseuphorie in den Vierzigern und Fünfzigern und der Stagnation in den Siebzigern erlebte das Gebiet der neuronalen Netzwerke in den Achtziger Jahren eine regelrechte Renaissance. Dies ist zu einem Großteil der Entwicklung und Einführung der generalisierten Delta-Regel durch David Rummelhart und Geoffrey Hinton im Jahre 1985 zuzuschreiben. Darauf aufbauend wurde das universelle Vorwärtsvermittlungsnetz (**feed-forward-network**) entwickelt.

Das Netz, wie es in Abbildung 3.7 dargestellt ist, hat einen schichtartigen Aufbau, bestehend aus den Eingangsknoten in der ersten Schicht, dem oder

den Ausgangsknoten in der abschließenden Schicht und den Hiddenknoten

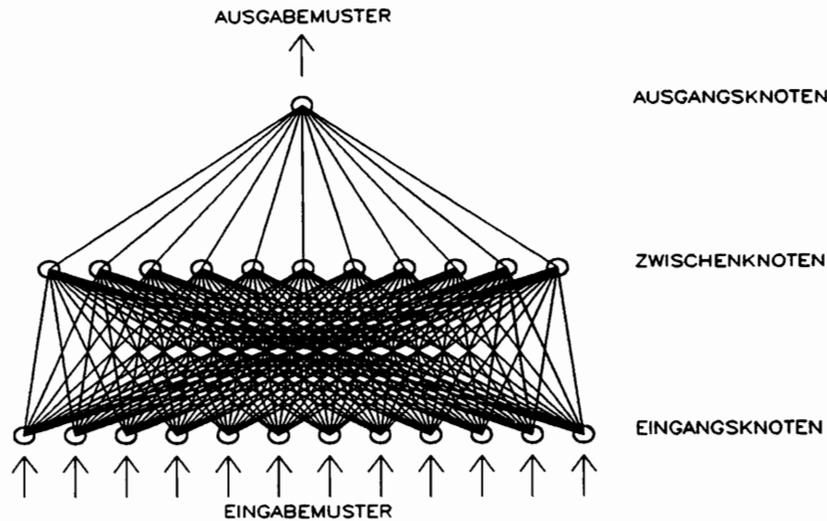


Abbildung 3.7: Dreischichtiges Vorwärtsvermittlungsnetz; die Kreise stellen die Knoten und die Linien die gewichteten Verbindungen dazwischen dar.

in der oder den Schichten dazwischen. In [KR 90] wird gezeigt, daß jede beliebige Separation durch ein vierschichtiges Netzwerk geleistet werden kann. Dies schließt nicht-konvexe und unzusammenhängende Bereiche, sowie Einschlüsse im Eigenschaftsraum mit ein. Sind die zu trennenden Klassen nicht verschränkt und in ihrer Gesamtheit jeweils zusammenhängend, so ist eine Zwischenschicht ausreichend. Daß dies, für das b -Tagging, der Fall zu sein scheint, zeigen die Ergebnisse mit vierschichtigen Netzwerken, die nicht besser sind, als die mit dreischichtigen. Ein weiteres Indiz für diese Annahme sind die mit dieser Arbeit übereinstimmenden Ergebnisse des Fisher-Diskriminanten-Verfahrens von Steffi Scheu [SCH 91], welches nur eine lineare Separation zuläßt. Auch die notwendige und ausreichende Zahl der versteckten Knoten zu ermitteln, ist kein triviales Problem bei der Anwendung von N.N., weshalb ich es in einem eigenen Unterkapitel (3.2.3) behandeln möchte. Die Zahl der Knoten in der Eingangs- und Ausgangsebene ist meist durch die Problemstellung vorgegeben. In unserem Fall werden so viele Eingangsknoten benötigen, wie Variablen in das Netzwerk eingegeben werden sollen. Es genügt ein Ausgangsknoten, da nur eine Klassifizierung in zwei Klassen beabsichtigt wird. Die Linien stellen die mit einem Gewicht ω versehenen Verbindungen zwischen den Knoten dar. In diesem Modell gibt es nur Verbindungen von Knoten eines Layers zu allen Knoten der nachfol-

3.2. Künstliche Modelle (artificial neural networks)

genden Schicht. Es existieren weder rekursive Verbindungen, noch solche innerhalb einer Ebene oder über mehrere Ebenen hinweg.

Die weiteren Charakteristika sind :

- Gewichtsregel: multiplikativ, d.h. der Ausgangswert o_i der Zelle i wird mit der Gewichtung ω_{ij} der Verbindung zum Eingang der nachfolgenden Zell j multipliziert. Die Netzaktivität am Eingang der empfangenden Zell j berechnet sich also wie : $net_j = \sum_i o_i \omega_{ij}$.
- Aktivierungsfunktion: die Sigmoidfunktion mit $T=1$.

$$f(x) = \frac{1}{1 + e^{-x/T}}$$

- Ausgangsfunktion: die Identität.

$$o_j = f_{out}(A_{j(t)}) = A_{j(t)}$$

- Bias: mit und ohne.
- Systemumgebung:
 - Eingangsnormierung,
 - Monitoring.

Die Eingangsnormierung ist ein weiteres Problem in der Anwendung der Netzwerke. Durch die Faltung mit der Sigmoidfunktion wird der gesamte Wertebereich einer Eingangsgröße auf das Gebiet zwischen Null und Eins abgebildet. Dies ist jedoch keine lineare Abbildung. Sehr niedrigen und sehr hohen Werten wird automatisch immer ein Wert nahe Null bzw. Eins zugewiesen. Nur die Umgebung um die Null erfährt eine differenziertere Behandlung. Dies erschwert den Lernprozeß erheblich oder macht ihn gar unmöglich. Die oft angewendete lineare Normierung auf Werte zwischen Null und Eins ist sehr problematisch und unbefriedigend, da die Wertebereiche bei vielen Größen nicht exakt definierbar sind und auch einzelne Ausreißer vernünftig behandelt werden müssen. Dazu bietet sich eine modifizierte Sigmoidfunktion an. Die von mir entwickelte Normierung, das Schema ist in Abbildung 3.8 dargestellt, verändert die Standard-Sigmoidfunktion mit $T=1$ so, daß die zu verarbeitenden Verteilungen gleichmäßig auf den Bereich

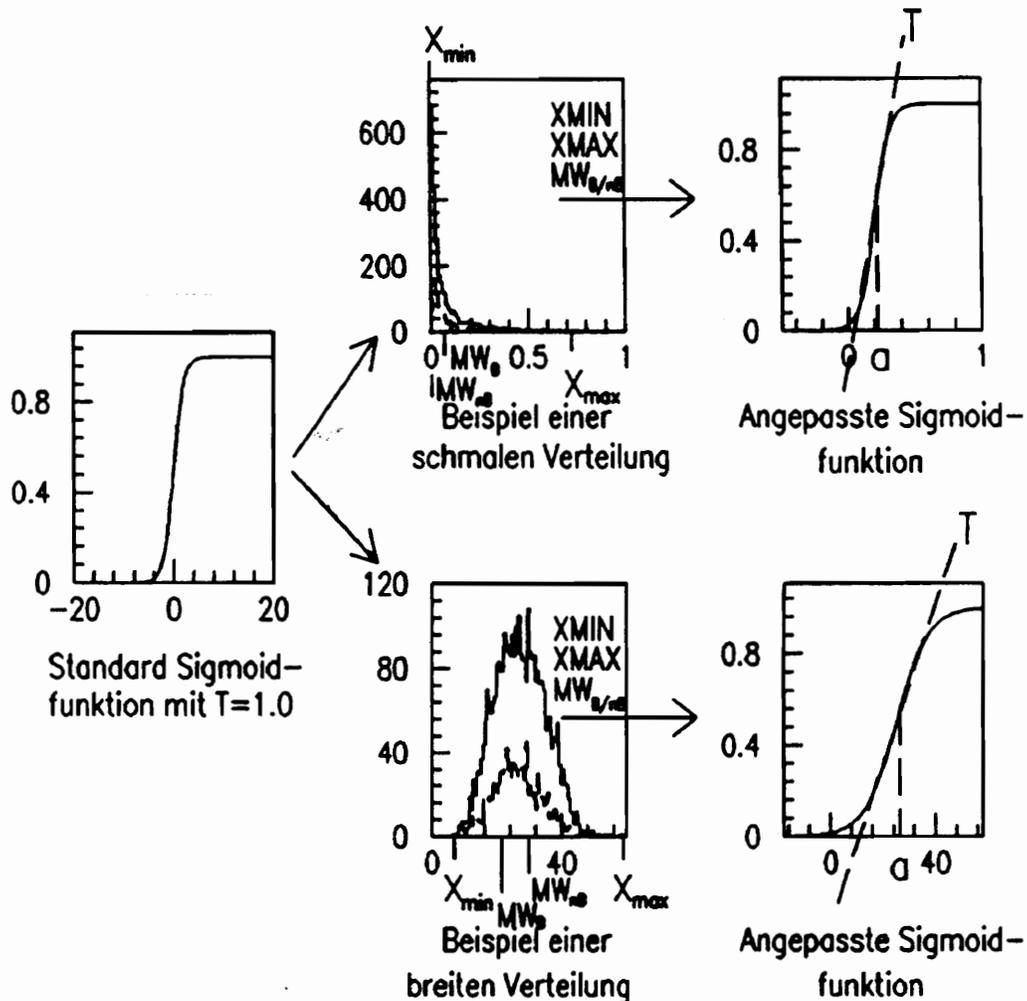


Abbildung 3.8: Normierungsschema für die Eingabevariablen. Die Standard-Sigmoidfunktion mit $T=1$ soll so verändert werden, daß die zu verarbeitenden Verteilungen gleichmäßig auf den Bereich zwischen 0 und 1 abgebildet werden. Das arithmetische Mittel α der Mittelwerte der beiden Klassen MW_B und MW_{nB} bestimmt die Verschiebung der Sigmoidfunktion in x -Richtung. Aus den Extremwerten X_{min} und X_{max} , d.h. aus dem x -Wertebereich der Verteilungen, wird der 'Temperaturwert' T der Sigmoidfunktion berechnet, der das Schwellwertverhalten der Funktion bestimmt. So wird für jede Eingangsgröße eine 'maßgeschneiderte' Sigmoidfunktion errechnet.

3.2. Künstliche Modelle (artificial neural networks)

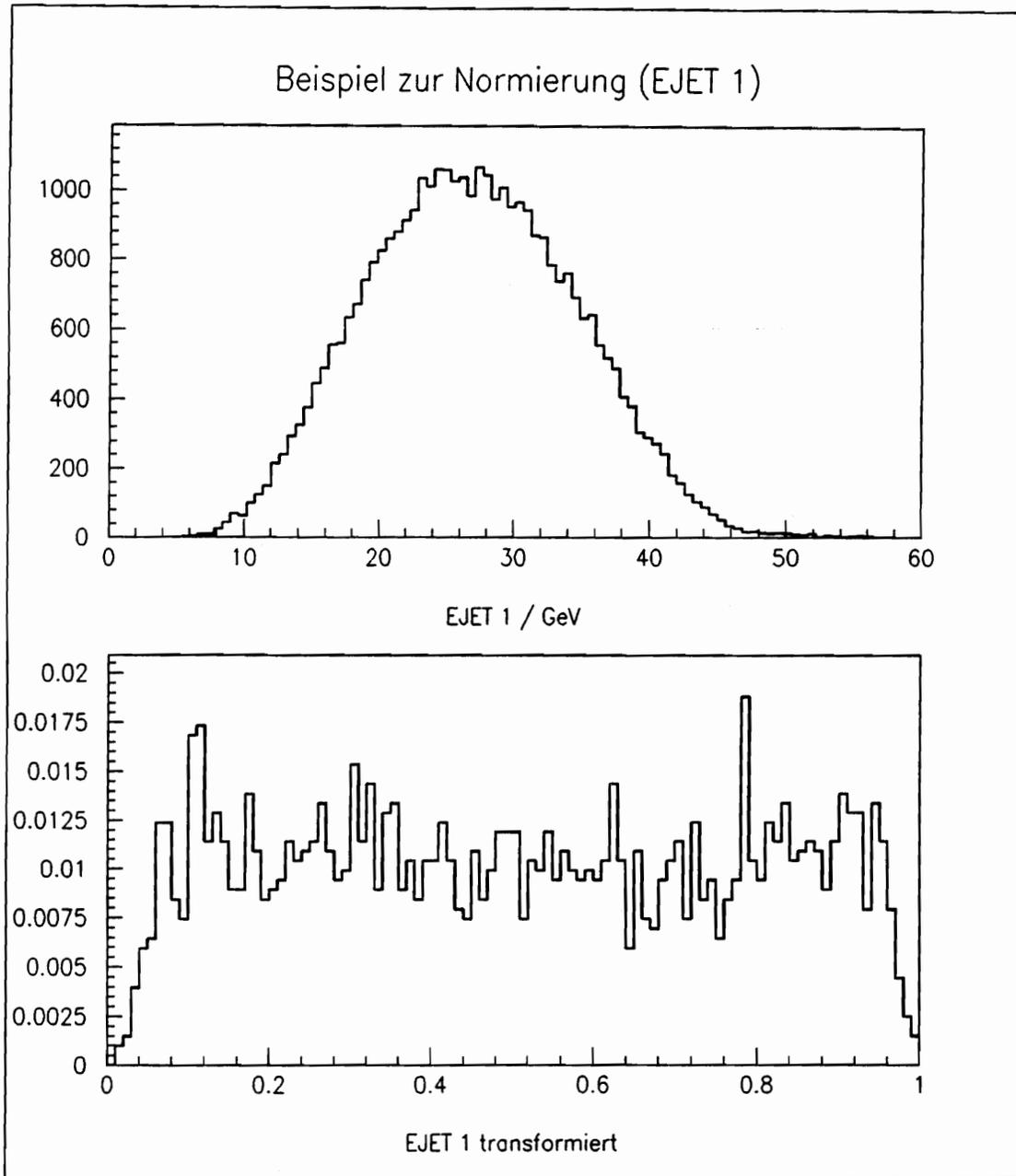


Abbildung 3.9: Oben ist das Spektrum der Jetenergie des energiereichsten Teilchenjets aufgetragen. Der untere Teil zeigt die Verteilung nachdem das im Text erläuterte Normierungsverfahren angewendet wurde. Es ist im Vergleich zu ursprünglichen Verteilung gut die gewünschte gleichmäßige Ausnutzung des gesamten Wertebereichs zwischen 0 und 1 zu sehen.

zwischen 0 und 1 abgebildet werden, indem die Breite und die Mittelwerte der zu normierenden Verteilungen berücksichtigt werden. Das arithmetische Mittel $a = \frac{1}{2}(MW_B + MW_{nB})$ der Mittelwerte der beiden Klassen MW_B und MW_{nB} bestimmt die Verschiebung der Sigmoidfunktion in x-Richtung. Aus dem arithmetischen Mittel der Breiten der Verteilungen für die beiden Klassen, welche wie die Mittelwerte aus Monte-Carlo Verteilungen bestimmt werden, errechnet sich der 'Temperaturfaktor' T . Der funktionale Zusammenhang ist

$$T = \frac{1}{12} \frac{1}{2} (|X_{Bmax} - X_{Bmin}| + |X_{nBmax} - X_{nBmin}|)$$

Mit der Verschiebung a und dem Temperaturfaktor T erhält man für jede Eingangsgröße eine 'maßgeschneiderte' Sigmoidfunktion

$$f(x) = \frac{1}{1 + e^{-(x-a)/T}}$$

In Abbildung 3.9 sind als Beispiel die Spektren der Jetenergie des energiereichsten Jet vor und nach der Normierung aufgetragen. Das Monitorieren wird durch das Abspeichern von Fehler- und Ergebniswerten zur späteren Auswertung bewerkstelligt. Leider ist es durch den Batchjob Betrieb an der IBM 3090 nicht möglich, sich ständig wichtige Parameter während des Lernens bzw. Testens direkt auf einem Monitor anzeigen zu lassen, um eventuelle Änderungen direkt vornehmen zu können. Will man den systematischen Einfluß von einzelnen Parametern studieren, ist dies ein gravierender Nachteil, der nur teilweise durch erhöhten Zeitaufwand kompensiert werden kann. Eine Workstation von entsprechender Leistungsfähigkeit, auf der Programme interaktiv betrieben werden können, ist, falls verfügbar, die effektivere Lösung.

Jeder Satz von Eingangsgrößen (Eingabevektor) wird nach der Normierungsprozedur über die Eingangsknoten in das Netzwerk eingegeben und danach von unten nach oben durch das gesamte Netzwerk hindurchprozessiert. Dabei läuft in jedem Knoten der schon beschriebene Prozeß ab. Am Ausgangsknoten ergibt sich ein Wert zwischen Null und Eins, der zur Klassifikation dient. Soweit die Beschreibung der Funktionsweise eines neuronalen Netzwerkes. Zwei sehr wichtige Punkte wurde noch nicht angesprochen :

- Wo ist das 'Wissen' des Netzwerkes lokalisiert ?
- Wie kommt das N.N. zu diesem 'Wissen', das heißt, wie lernt es ?

3.2. Künstliche Modelle (artificial neural networks)

Das 'Wissen', d.h. die Fähigkeit zur Unterscheidung, sitzt in diesem Modell nicht wie man vielleicht vermuten könnte in den Knoten, sondern in den Verbindungen zwischen ihnen ! Die Knoten dienen nur als Verarbeitungseinheiten (Prozessoren). Die Gewichtungen ω der Verbindungen repräsentieren die Information. Durch die Komplexität eines Netzwerkes ist es praktisch nicht möglich die Aufgabe oder die Trennfähigkeit einer einzelnen Verbindung zu bestimmen. Dies macht auch gar keinen Sinn, da die Information inhärent im Geflecht des gesamten Netzwerkes gespeichert ist². Im Gegensatz dazu sitzt in herkömmlichen Rechenanlagen jede Informationseinheit in einem genau lokalisierbaren Speicherbereich. Die Funktion eines mehrschichtigen Netzwerkes ist im einzelnen also nicht verifizierbar, sondern nur validierbar.

Wie die Werte der Gewichtungsfaktoren ω zu wählen sind, läßt man das Netzwerk selbst herausfinden. Neuronale Netzwerke sind selbstorganisierend. In einem Trainingsmodus, der noch genau beschrieben wird, wird das Netzwerk aus einem Anfangszustand mit zufällig gewählten Gewichtungsfaktoren in einen Zustand überführt, der es im Idealfall in die Lage versetzt, die gestellte Aufgabe perfekt zu lösen. Wird ein N.N. mit einer symmetrischen Belegung der Gewichte gestartet und der zu erreichende stabile Endzustand besteht aus einer unsymmetrischen Belegung, so kann das Netzwerk diesen Zustand aufgrund des Lernalgorithmus nie erreichen [RU 86]. Die symmetriebrechende Wirkung einer zufälligen Anfangsbelegung der Gewichtungen beugt dieser Gefahr vor. Das Schema, das den vorwärtsvermittelnden Netzwerken zugrundeliegenden Lern-Algorithmuses, sieht im einfachsten Falle folgendermaßen aus :

Belege die Gewichte ω_{ij} mit zufälligen Werten.

Wiederhole bis der Fehler des N.N. konvergiert ...

Wähle einen Eingabe-Aktivierungsvektor aus, zu dem eine Soll-Ausgabe existiert.

Phase 1: Vorwärtsvermittlung

Aktiviere die Eingabeschicht mit dem gewählten Eingangsvektor.

Ermittle schrittweise die Aktivierung der nachfolgenden Zwischenschichten und der abschließenden Ausgabeschicht.

Phase 2: Fehlerrückvermittlung (back propagation)

Ermittle die δ -Werte³ für die Ausgabeschicht

²Es wird auch von einer holistischen Verteilung des Wissens innerhalb des neuronalen Netzwerkes gesprochen.

Ermittle schrittweise die δ -Werte für die vorgelagerten Zwischenschichten bis hinab zur ersten Zwischenschicht.

Führe die Änderung (update) der Gewichte durch.

Der δ -Wert oder Fehler wird in einem rekursiven Prozeß von oben nach unten durch das Netzwerk zurück prozessiert. Dabei wird für jeden Knoten ein Fehler und schließlich die Änderung der Gewichte berechnet. Dieser Vorgang wird mit einem Lernsample, wobei jeder Eingabevektor dem Netzwerk mehrmals präsentiert wird, so lange wiederholt bis der Fehler innerhalb einer eingestellten Grenze gegen einen konstanten Wert konvergiert und sich die Gewichtungen damit in einem stabilen Zustand befinden. Die Werte der Gewichte werden abgespeichert und damit ist die Lernphase beendet.

Von diesem Algorithmus ausgehend ergeben sich aus praktischen Gründen einige Verfeinerungen. Es wird nicht für jeden Eingabevektor eine Gewichtsänderung durchgeführt, sondern nur wenn der δ -Wert des Ausgangsknotens größer als eine einstellbare Grenze ist. Das heißt nichts anderes, als daß der Output des Netzwerk nicht exakt Null oder Eins sein muß, was außer bei logischen Problemstellungen auch nicht der Fall sein wird, sondern innerhalb gewisser Schranken variieren darf. Ein typischer Wert hierfür ist 0.2, damit wird alles im Bereich 0.0- 0.2 als *nicht - b* und zwischen 0.8 und 1.0 als *b* angesehen. Eine weitere Änderung liegt darin, das die Trainingsergebnisse nicht zufallsverteilt in das N.N. eingegeben werden, sondern immer abwechselnd ein Teilchen der Klasse A und ein Teilchen der Klasse B. Nun werden nach jedem Teilchen die δ -Werte berechnet, ohne daß die Gewichte verändert werden. Dies geschieht nur dann, wenn mindestens für eines der beiden Ereignisse der Ausgangswert des Netzwerkes außerhalb der angegebenen Schranke liegt. Die Änderung der Gewichte erfolgt dann mit dem Mittelwert der Änderungen, die für die beiden Eingangsvektoren berechnet wurden. Das paarweise geordnete Lernsample verhindert, daß ein Netzwerk nacheinander sehr viele Eingabevektoren nur einer Klasse präsentiert bekommt und damit automatische alle Ereignisse als dieser Klasse zugehörig ansieht. Dieser Fehler wurde zu Beginn dieser Arbeit gemacht und mit den so trainierten Netzwerken konnte keine Separation erreicht werden, da entweder alle Ereignisse als *b* oder als *nicht - b* getaggt wurden. Daß der Update der Gewichte nach zwei Eingabevektoren und nicht nach 4, 8, 10 oder

³Der δ -Wert der Ausgangsschicht ist die Differenz zwischen dem IST-Wert und dem SOLL-Wert des Ausgangsknotens.

mehr geschieht, liegt daran, daß eine möglichst schnell Konvergenz erreicht werden soll. In der ursprüngliche Version des Algorithmus ist ein Update immer erst nach der Präsentation des gesamten Trainingssamples vorgesehen. Es wird der Fehler über alle Eingabemuster berechnet und daraus die Gewichtungsänderung. Dies hat aber einen sehr hohen Rechenaufwand zur Folge und ist daher nicht praktikabel, zumal mit dem abgeänderten Verfahren gleichfalls stabile Endzustände erreicht werden.

3.2.2 Die Mathematik eines mehrschichtigen Vorwärtsvermittlungsnetzes

Mathematisch exakt, dafür etwas unanschaulicher wird der Algorithmus wie folgt dargestellt. Die dabei verwendeten Variablen sind

- p Index, der den Aktivierungsvektor angibt (p =pattern)
- j Index, der die Komponente des jeweiligen Vektors angibt
- E_p Fehler des Netzwerks
- i_{pj} Netzwerkinput
- o_{pj} berechneter Netzwerkoutput (Ist)
- t_{pj} Zielwert des Netzwerkoutputs (Soll) (t =target)
- net_{pj} Netzaktivität am Eingang eines Knotens
- ω_{ij} Gewichtungsfaktor zwischen den Knoten i und j
- $\Delta\omega_{ij}$ Änderung des Gewichtungsfaktors zwischen den Knoten i und j
- η Lernrate
- α Konstante des Momentum-Term
- f Aktivierungsfunktion
- f' Ableitung der Aktivierungsfunktion

Ziel des iterativen Verfahrens ist die Minimierung des quadratischen Abstandes (Fehler) zwischen Soll und Ist bezogen auf alle im Trainingsmodus präsentierten Ein- und Ausgabemuster. Diese Minimierung vollzieht sich im virtuellen vieldimensionalen Raum, der durch die Gewichtungen ω aufgespannt wird (Gewichtungsraum). Um dies zu erreichen, steht einem nur das Mittel der Gewichtsänderung zur Verfügung. Das Problem liegt in der Bestimmung der Größe dieser Gewichtsänderung. Beginnen wir mit der Definition des Gesamtfehlers

$$E = \sum_p E_p \quad ,$$

wobei der Fehler E_p für ein Muster (pattern p) $E_p = \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2$ ist.

Leitet man E_p nach ω_{ij} ab, so ergibt sich

$$\frac{\partial E_p}{\partial \omega_{ij}} = \frac{\partial E_p}{\partial o_{pj}} \frac{\partial o_{pj}}{\partial \omega_{ij}} \quad \text{mit} \quad \frac{\partial E_p}{\partial o_{pj}} = -(t_{pj} - o_{pj}) := -\delta_{pj} \quad . \quad (3.1)$$

Für Knoten mit einer linearen Aktivierungsfunktion gilt :

$$o_{pj} = \sum_i \omega_{ji} i_{pi} \quad \Rightarrow \quad \frac{\partial o_{pj}}{\partial \omega_{ij}} = i_{pi} \quad (3.2)$$

$$\Rightarrow \quad E'_p = \frac{\partial E_p}{\partial \omega_{ij}} = -\delta_{pj} i_{pi} \quad \sim \quad \Delta \omega_{ij} \quad \Rightarrow \quad \Delta_p \omega_{ij} = \eta (t_{pj} - o_{pj}) i_{pi}$$

Beobachtung [RU 86]:

$$\frac{\partial E}{\partial \omega_{ij}} = \sum_p \frac{\partial E_p}{\partial \omega_{ij}} = \sum_p -\delta_{pj} i_{pi}$$

Dieses Grundlagenwissen kann man nun für nichtlineare Aktivierungsfunktionen verallgemeinern. Es sind folgende Forderungen an eine solche Funktion f zu stellen :

- Differenzierbarkeit an jeder Stelle und damit auch Stetigkeit.
- f sollte streng monoton steigend sein, um die Gefahr des 'Festfressens' des Netzwerkes an einem Punkt mit $f'_{(x)} = 0$, d.h. einem Sattelpunkte zu vermeiden.

3.2. Künstliche Modelle (artificial neural networks)

- f sollte nach oben und unten (z.B. durch 0 und 1) beschränkt sein.

Die Klasse von Funktionen, die diesen Forderungen genügen, bezeichnet man als semi-lineare Funktionen. Die im Bereich der Fehlerrückvermittlung populärste Aktivierungsfunktion ist die auch hier verwendete Sigmoidfunktion

$$f(x) = \frac{1}{1 + e^{-x/T}}$$

In Anlehnung an die in der Thermodynamik gebräuchliche Boltzmann-Verteilung wird T als Temperatur bezeichnet. Abbildung 3.4 zeigt eine Kurvenschar von Sigmoidfunktionen für verschiedene Temperaturwerte.

Verallgemeinern wir (3.2), so ersetzen wir o_{pj} durch

$$o_{pj} = f_j(\text{net}_{pj}) = f_j(\sum_i \omega_{ij} o_{pi}) \quad , \text{ d.h. es ist } \quad \text{net}_{pj} = \sum_i \omega_{ij} o_{pi}$$

Weiter setzen wir, wie im linearen Fall für die Änderung der Gewichte

$$\Delta_p \omega_{ij} \propto - \frac{\partial E_p}{\partial \omega_{ij}}$$

mit

$$\frac{\partial E_p}{\partial \omega_{ij}} = \frac{\partial E_p}{\partial \text{net}_{pj}} \frac{\partial \text{net}_{pj}}{\partial \omega_{ij}}$$

mit

$$\frac{\partial \text{net}_{pj}}{\partial \omega_{ji}} = \frac{\partial}{\partial \omega_{ji}} \sum_i \omega_{ij} o_{pi} = o_{pi}$$

Def.:

$$\delta_{pj} = \frac{\partial E_p}{\partial \text{net}_{pj}} \quad \text{analog zu (3.1)} \quad (3.3)$$

$$\Rightarrow \frac{\partial E_p}{\partial \omega_{ij}} = -\delta_{pj} o_{pi} \quad \text{mit (3.1)} \Rightarrow \quad \Delta_p \omega_{ij} = \eta (t_{pj} - o_{pj}) o_{pi} \quad (3.4)$$

Die hier eingeführte frei wählbare Konstante η wird Lernrate genannt, da mit ihr die Lerngeschwindigkeit, das heißt der Grad der Proportionalität der Gewichtsänderung zur Steigung der Fehlerebene gesteuert werden kann.

Gesucht wird nun

$$\delta_{pj} = \frac{\partial E_p}{\partial net_{pj}} = \frac{\partial E_p}{\partial o_{pj}} \underbrace{\frac{\partial o_{pj}}{\partial net_{pj}}}_{f'_{j(net_{oj})}} \quad (3.5)$$

$$\Rightarrow \text{ges. } \frac{\partial E_p}{\partial o_{pj}}$$

An diesem Punkt wird eine Fallunterscheidung notwendig :

(1) für Verbindungen zwischen Hidden- und Output-Nodes :

$$\frac{\partial E_p}{\partial o_{pj}} = -(t_{pj} - o_{pj})$$

eingesetzt in (3.5) $\Rightarrow \delta_{pj} = (t_{pj} - o_{pj}) f'_{j(net_{oj})}$

(2) für Verbindungen zwischen Input-(bzw. Hidden-) und Hidden-Nodes :

$$\frac{\partial E_p}{\partial o_{pj}} = \sum_k \underbrace{\frac{\partial E_p}{\partial net_{pk}}}_{-\delta_{pk} \text{ vgl. (3.5)}} \frac{\partial net_{pk}}{\partial o_{pj}} = \sum_k -\delta_{pk} \frac{\partial \sum_i \omega_{kj} o_{pi}}{\partial o_{pj}} = -\sum_k \delta_{pk} \omega_{kj}$$

eingesetzt in (3.5) $\Rightarrow \delta_{pj} = f'_{j(net_{oj})} \sum_k \delta_{pk} \omega_{kj}$

Eingesetzt in (3.4) und zusammengefaßt erhalten wir als Ergebnis folgenden Ausdruck für die Größe der Gewichtsänderung :

$$\Delta_p \omega_{ij} = \eta \underbrace{(t_{pj} - o_{pj})}_{\delta_{pj} \text{ vgl. (3.1)}} o_{pi} = \begin{cases} \eta f'_{j(net_{pj})} (t_{pj} - o_{pj}) & \text{für (1)} \\ \eta f'_{j(net_{pj})} \sum_k \delta_{pk} \omega_{kj} & \text{für (2)} \end{cases}$$

Um mögliche Oszillationen bei zum Beispiel zu groß gewählter Lernrate η zu dämpfen addieren wir noch einen sogenannten Momentum-Term. Dieser zusätzliche Term besteht aus der zuletzt durchgeführten Gewichtsänderung und einem frei wählbaren Momentenfaktor α über den der Einfluß der

3.2. Künstliche Modelle (artificial neural networks)

Dämpfung bestimmt werden kann. Die Gewichtsänderung zu einem Zeitpunkt $n + 1$ berechnet sich schließlich so :

$$\Rightarrow \boxed{\Delta_p \omega_{ij}(n+1) = \eta \delta_{pj} o_{pi} + \underbrace{\alpha \Delta_p \omega_{ij}(n)}_{\text{Momentum-Term}}}$$

Diese Formel ohne den Momentum-Term ist unter dem Namen **generalisierte Delta-Regel (generalized delta-rule)** [RU 86] bekannt und wurde von D.E.Rumelhardt und G.E.Hinton 1985 erstmals vorgestellt. Sie bestimmt die Änderung $\Delta \omega_{ij}$ der Verbindung zwischen dem Knoten i und j , die sich als Produkt einer konstanten Lernrate η , des Fehlers δ_j des empfangenden und des Ausgangwertes o_i des sendenden Knotens berechnet, für den Aktivierungsvektor p . Das Training eines neuronalen Netzwerkes mit dieser Regel, kann man geometrisch als die Suche nach einem Minimum der Fehlerfläche im Gewichtsraum interpretieren. Im Idealfall ist es das absolute Minimum. $\delta_{pj} o_{pi}$ stellt einen lokal 'abwärts' weisenden Richtungsvektor auf der Fehlerebene in diesem Raum dar, und mit η wird die absolute Schrittweite festgelegt. Werden die Gewichte mit den berechneten $\Delta \omega$ -Werten verändert, so heißt das in unserem Bild, daß sich der Zustandsvektor des N.N. in diesem imaginären Raum auf der Fehlerebene um den berechneten ω -Vektor weiterbewegt. Das Konvergenzverhalten dieses Verfahrens hängt ganz entscheidend von einer geschickten Wahl der Faktoren η und α ab. Ist η zu klein gewählt, ergibt sich nur eine sehr langsame Konvergenz. Dabei besteht auch noch die Gefahr des 'Steckenbleibens' in einem lokalen Extremum. Ist η allerdings zu groß, ist es dem Netzwerk unter Umständen unmöglich in ein Minimum zu gelangen, da dieses immer mit zu großen Schritten übersprungen wird. Wie das Netzwerk in einem solchen Fall über einem Minimum oszilliert, ist in Abbildung 3.10 zu sehen. Zur Dämpfung solcher und hochfrequenter Oszillationen, welche durch spezielle Topologien des ω -Raumes begünstigt werden können, wird der Momentum-Term hinzugefügt. Hierdurch wird die Verwendung höherer Lernraten möglich.

Man spricht von einem statischen oder 'fixen' Lernmodus, wenn während der Trainingsphase die Faktoren η und α konstant bleiben. Dem gegenüber beginnt man beim dynamischen Lernen mit einem großen η und vermindert dieses im Laufe des Lernprozesses. So nimmt man jeweils die Vorteile eines großen bzw. eines kleinen η in Anspruch. Am Anfang bewegt sich das Netzwerk mit großen 'Schritten' auf der Fehlerebene und kann so kleinere lokale Minima überspringen. In der Nähe des absoluten Minimums ange-

langt, ist durch das Schrumpfen der Lernrate eine gute Konvergenz möglich. In der Praxis hat sich während dieser Arbeit das dynamische Lernen durch eine sehr viel schnellere Konvergenz gegenüber dem statischen Trainieren ausgezeichnet.

Hat ein Netzwerk eine stabile Lage erreicht, kann man nie sicher sein, daß dies nicht ein lokales Minimum ist. Um dem N.N. die Chance zu geben sich aus einem vermeintlichen relativen Minimum zu befreien, kann man

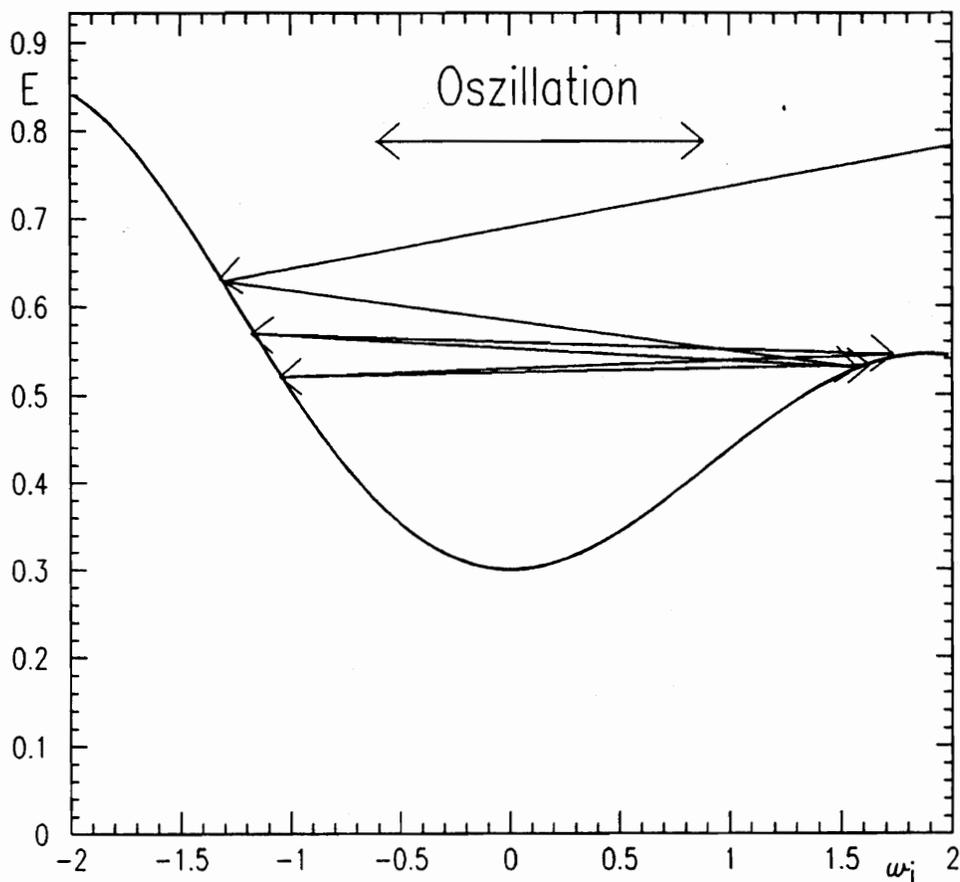


Abbildung 3.10: Das Oszillieren eines Netzwerkes über einem Minimum. Da die Lernrate η zu groß gewählt wurde, ergibt sich eine zu große Schrittweite und das Netzwerk ist nicht in der Lage den Zustand am Boden des Minimums einzunehmen. Dargestellt ist der zu minimierende Netzwerkfehler E über (aus Darstellungsgründen nur) einem Gewichtungsfaktor ω_i . E ist natürlich von allen ω_i abhängig.

zum Abschluß der Lernphase das Netz noch einmal mit einem großen η

trainieren. Dies entspricht in einem mechanischen Analogon dem kräftigen Anstoßen einer Kugel, die in einer Mulde liegt. Ist diese Position nur ein relatives Minimum, so besteht bei einem kräftigen Stoß die Hoffnung, daß sich die Kugel über die begrenzenden Wälle hinwegsetzt und in eine tiefere Position rollt. Andernfalls wird sie die Ausgangslage wieder einnehmen.

Der Ausgangspunkt der Arbeit war ein Satz von neun Unterprogrammen, welcher von O. Barring an der Universität LUND in Schweden programmiert wurde. Er enthält im wesentlichen den Backpropagation Algorithmus. Während der Durchführung dieser Arbeit entwickelte sich durch den Einbau verschiedener Funktionen (Systemumgebung, Normierung, Monitorierung ...) ein immer umfangreicheres Simulationsprogramm für neuronale Netzwerke.

3.2.3 Algorithmus zur Variation der Anzahl der versteckten Knoten

Eines der vielen ungelösten Probleme im Bereich der neuronalen Netzwerke ist die Ermittlung der hinsichtlich Laufzeit und/oder Trenngüte optimalen Anzahl von Knoten in der/den Zwischenschicht(en). Es gibt bis heute keine allgemeingültige Regel für dieses Problem; vielmehr kursieren mehr oder weniger fundierte 'Faustregeln'. Es gibt keine Theorie, die eine optimale Anzahl festlegt. Im Allgemeinen gilt je höher die Anzahl der versteckten Knoten, desto mehr Wege zum absoluten Minimum gibt es, da mit jedem zusätzlichen Knoten die Dimension des Gewichtungsraumes anwächst. Viele Knoten brauchen jedoch viel Rechenzeit. Das Ergebnis sollte daher mit möglichst wenig Knoten erreicht werden. Es ist also sinnvoll mit mehreren Knoten zu lernen und nach erfolgter Konvergenz die Anzahl so lange zu reduzieren, wie noch eine ausreichende Leistung erreicht wird. Von den Fujitsu Lab.Ltd. wurde ein weniger mathematisch schlüssiger Algorithmus als vielmehr ein pragmatischer Ansatz, der genau dieses leisten soll, entwickelt [HI 91]. Die Grundidee ist, während des Lernens die Anzahl der Knoten in der Zwischenschicht zu variieren und die Auswirkung auf die Performance (Fehler) des Netzwerks zu registrieren. Aufbauend auf dieser Idee wurde im Rahmen dieser Diplomarbeit in das entwickelte Netzwerkprogramm ein Algorithmus implementiert, der zur Aufgabe hat, die Anzahl der Knoten in der Zwischenschicht eines existierenden Netzwerkes bei einem vorgegebenen Maximalwert des Fehlers zu optimieren, das heißt zu minimieren. In Ab-

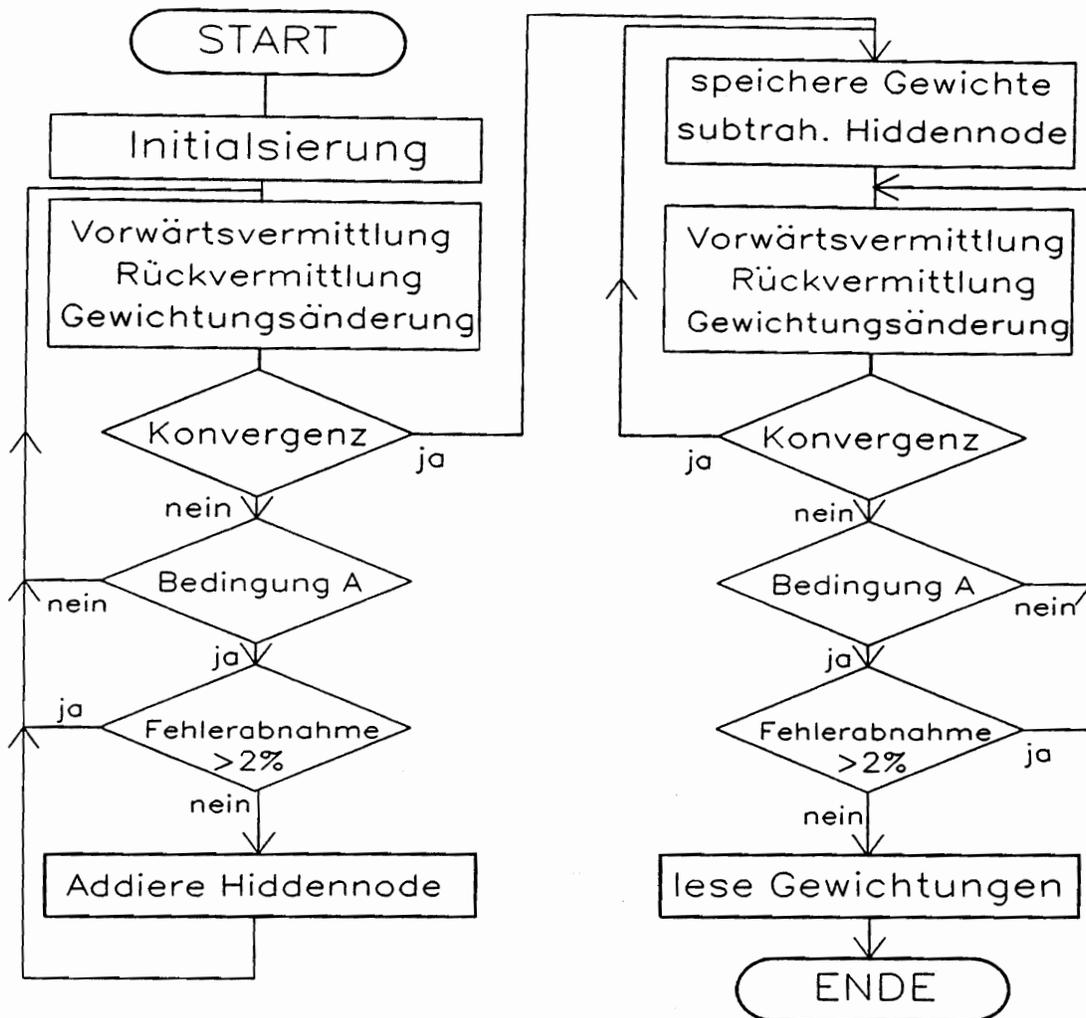


Abbildung 3.11: Flußdiagramm des zweigeteilten Algorithmus zur Variation der Hiddennodeanzahl. In der linken Hälfte, wird die Anzahl der Hiddenknoten solange erhöht bis das Netzwerk konvergiert. Die Bedingung A : 'Ist die Anzahl der Gewichtsänderungen ein vielfaches von 100 ?' regelt dabei, daß in sinnvollen Abständen die Fehlerabnahme überprüft wird. Zeigt das Netzwerk konvergentes Verhalten, wird nach dem Abspeichern der aktuellen Gewichte und nach Entfernen eines Hiddenknoten, in der rechten Hälfte des Diagramms weitergelernt. Konvergiert das Netz immer noch wiederholt sich dieser Vorgang. Ist die Fehlerabnahme über einen in Bedingung A festgelegten 'Zeitraum' (hier 100 Gewichtsänderungen) kleiner als eine bestimmte Grenze (hier 2%), so wird das zuletzt abgespeicherte Netzwerk wieder geladen und der Algorithmus ist beendet.

bildung 3.11 ist der Algorithmus in Form eines Flußdiagramms dargestellt. Der Algorithmus bringt im Allgemeinen⁴ keine Steigerung der Leistungsfähigkeit des Netzwerkes bezüglich der Trenngüte, ermöglicht jedoch die Ermittlung der minimalen Anzahl von Knoten in der Zwischenschicht, ohne daß die Leistungsfähigkeit des Netzwerkes deutlich abnimmt. Eine solche Optimierung verhindert eine Überdimensionierung, welche unnötig CPU-Zeit kostet. Dies spielt bei dem hier behandelten Problem des *b*-taggings mit dem verwendeten schnellen Großrechner vom Typ IBM 3090 eine untergeordnete Rolle, gewinnt aber sicher dann an Bedeutung, wenn Analysen auf weniger leistungsfähigen Rechnern gemacht werden sollen. Ein sehr viel wichtigeres Anwendungsgebiet wird sich wohl ergeben, wenn neuronale Netzwerke (z.B. als Hardwareimplementation) zur online Analyse bzw. zum triggern von bestimmten Ereignissen verwendet werden sollen, wie dies am SSC [LA 90] geplant ist, wo man nach jedem Ereignis einer immensen Datenflut gegenüberstehen wird, die es zu verarbeiten gilt.

3.2.4 Das Testen eines Netzwerkes

Ein N.N. wird die Daten mit denen es trainiert wurde richtig klassifizieren als unbekannt. Da in unserem Fall das Netzwerk aber in der Anwendung unbekannt Ereignisse richtig zuordnen soll, ist es notwendig die Leistungsfähigkeit ebenfalls mit Daten, die es noch nicht kennt, zu ermitteln. Beim *b*-Tagging geht es weniger um das eindeutige Wiedererkennen einzelner Ereignisse, wie z.B. beim Sortieren spezieller Werkstücke, sondern um die Generalisierung von Charakteristika. Das Testen geschieht mit einem vom Lernsample statistisch unabhängigen Testsample. Die Daten werden wie im Trainingsmodus durch das Netzwerk hindurch prozessiert. Es wird nun aber nur protokolliert, ob die Klassifikation des Netzwerkes richtig oder falsch ist. Die Gewichtungen bleiben unverändert. Die Ergebnisse werden durch Reinheits- und Effizienzwerte angegeben, die sich mit dem jeweiligen Netzwerk erreichen lassen. Die Definition dieser Werte sind :

Effizienz (efficiency) = Der Prozentsatz aller tatsächlichen *b*-Ereignisse, die als solche auch identifiziert wurden.

⁴Falls sich das Netzwerk in einem lokalen Minimum befindet, kann es durch die Anwendung dieses Algorithmus aus diesem in das absolute Minimum gelangen, und sich so die Leistungsfähigkeit doch erhöhen.

Kapitel 3. Neuronale Netzwerke

Reinheit (purity) = Der Prozentsatz der als b identifizierten Ereignisse, welche tatsächlich b -Ereignisse sind.

Der noch verbleibende Freiheitsgrad ist den Schnitt auf die Ausgangsverteilung des N.N. zu variieren. So ergibt sich, je nach spezifischer Zielsetzung, eine höhere Reinheit oder Effizienz. Abbildung 3.12 zeigt die Ausgangsverteilungen der beiden Klassen im Idealfall. Hier erbringt ein Schnitt in der Mitte bei 0.5 eine praktisch perfekte Trennung. Dies sieht in der Realität des b -Taggings, wie später in Kapitel 6 gezeigt wird, ganz anders aus.

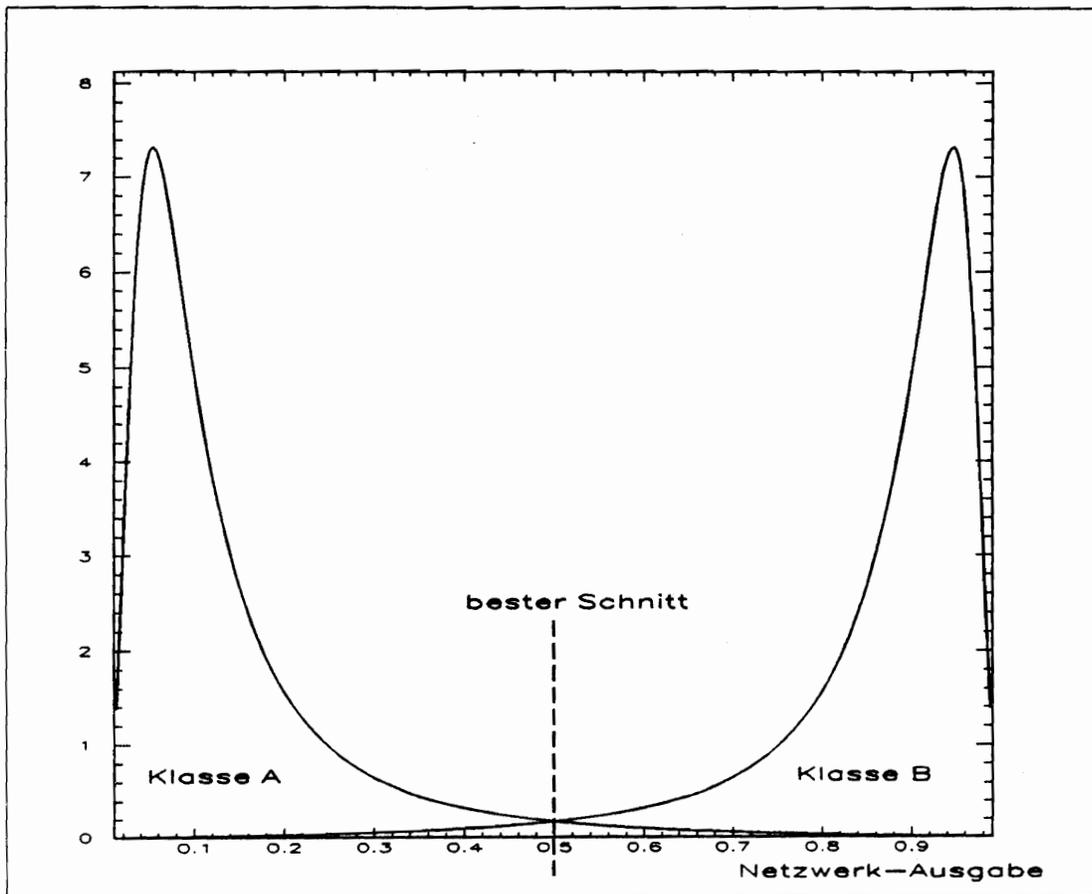


Abbildung 3.12: Der ideale Output eines neuronalen Netzwerkes. Es ist deutlich zu sehen, wie mit einem Schnitt bei 0.5 auf die Ausgangsverteilung des Netzwerkes, die beiden Klassen A und B annähernd perfekt separiert werden können.

Kapitel 4

Das DELPHI-Experiment

Diese Arbeit wurde im Rahmen der DELPHI-Kollaboration angefertigt. **DELPHI** (**DE**tector with **L**epton, **P**hoton and **H**adron **I**dentification) ist eines der vier unterirdisch gelegenen Experimente (ALEPH, DELPHI, L3, OPAL) am Elektron-Positron-Speicherring LEP. LEP ist Teil des CERN (Europäisches Laboratorium für Teilchenphysik), welches sich in der Nähe von Genf befindet. Mit 27km Umfang ist LEP (Large Electron Positron collider) der zur Zeit größte Speicherring. Er ging am 14.07.1989 in Betrieb. In seiner jetzigen Ausbaustufe kann er Schwerpunktsenergien bis 100 GeV erreichen; dies genügt zur Erzeugung reeller Z^0 -Bosonen (Schwerpunktsenergie = 91.6 GeV). In einer geplanten zweiten Ausbaustufe (LEP200) sollen die normalleitenden Kupfer-Resonatoren gegen supraleitende Resonatoren ausgetauscht werden, um eine Schwerpunktsenergie von 200 GeV zu erreichen. Dann ist die Produktion von reellen W^+W^- -Paaren möglich. Die bis Ende 1989 erreichte mittlere Luminosität beträgt etwa $\mathcal{L} = 10^{30} \text{cm}^{-2} \text{s}^{-1}$. Die Größenordnung, sowie den schalenartigen Aufbau des zylindrischen Mittelteils (Barrel) und einer der zwei Endkappen (end caps) aus denen DELPHI besteht, zeigt Abbildung 4.1. DELPHI ist ein 4π Detektor, dessen Aufgabenschwerpunkt auf Teilchenidentifikation, präziser Vertexbestimmung und hochauflösender dreidimensionaler Information über ein Ereignis liegt [DE 90/1].

Die verschiedenen Aufgaben der einzelnen Komponenten sind in Tabelle 4.1 aufgelistet. Am Ende des Kapitels sind in Tabelle 4.2 einige charakteristische Daten zusammengefaßt.

Aufgabe	Detektortyp	Detektorkomponente bei DELPHI
Impulsmessung	Si-Dioden Spurkammern	Vertexdetektor VD Innerdetektor ID Outerdetektor OD Zeitprojektionskammer TPC Vorwärtskammer A FCA Vorwärtskammer B FCB
Geschwindigkeitsbestimmung	Čerenkovzähler	Barrel-RICH Forward-RICH
Teilchenidentifikation	Driftkammern Szintillationszähler	Barrel-Müonkammern B-MU Vorwärts-Müonkammern F-MU Flugzeitzähler TOF
Energiemessung	e.m. Kalorimeter Hadronkalorimeter	HPC Vorwärts e.m. Kalorimeter FEMC Hadronkalorimeter HCAL

Tabelle 4.1: Die Aufgabenverteilung der verschiedenen Detektorkomponenten bei DELPHI. Die Detektorenkomponenten wurden in der Rubrik Aufgabe dem Punkt zugeordnet, für den sie von besonderer Bedeutung sind. Dies bedeutet jedoch nicht, daß sie für die anderen Aufgaben nicht auch wichtige Informationen liefern.

Die nachfolgenden Abschnitte gehen detaillierter auf einige Detektorkomponenten ein.

4.1 Die supraleitende Spule

Die supraleitende Spule, welche mit flüssigem Helium auf 4.5 K gekühlt wird, hat eine axiale Länge von 7.4 m und einen inneren Durchmesser von 5.2 m. Die Hauptspule, welche bei einem Strom von bis zu 5000 A ein Magnetfeld von ca. 1.2 T erzeugen kann, wird an den beiden Enden von zwei 35 cm lange Kompensationsspulen ergänzt, um ein möglichst homogenes Feld zu erreichen. Messungen innerhalb der **Time Projection Chamber** (TPC) ergaben ein longitudinales Magnetfeld von $1.2334^{+0.0001}_{-0.0010}$ T [DE 90/1]. Die kleinen Abweichungen nach oben und unten lassen auf eine schwache Axialasymmetrie schließen. Während die radiale Änderung des Magnetfeldes weniger als 5 G beträgt, ist der azimutale Fehler vernachlässigbar.

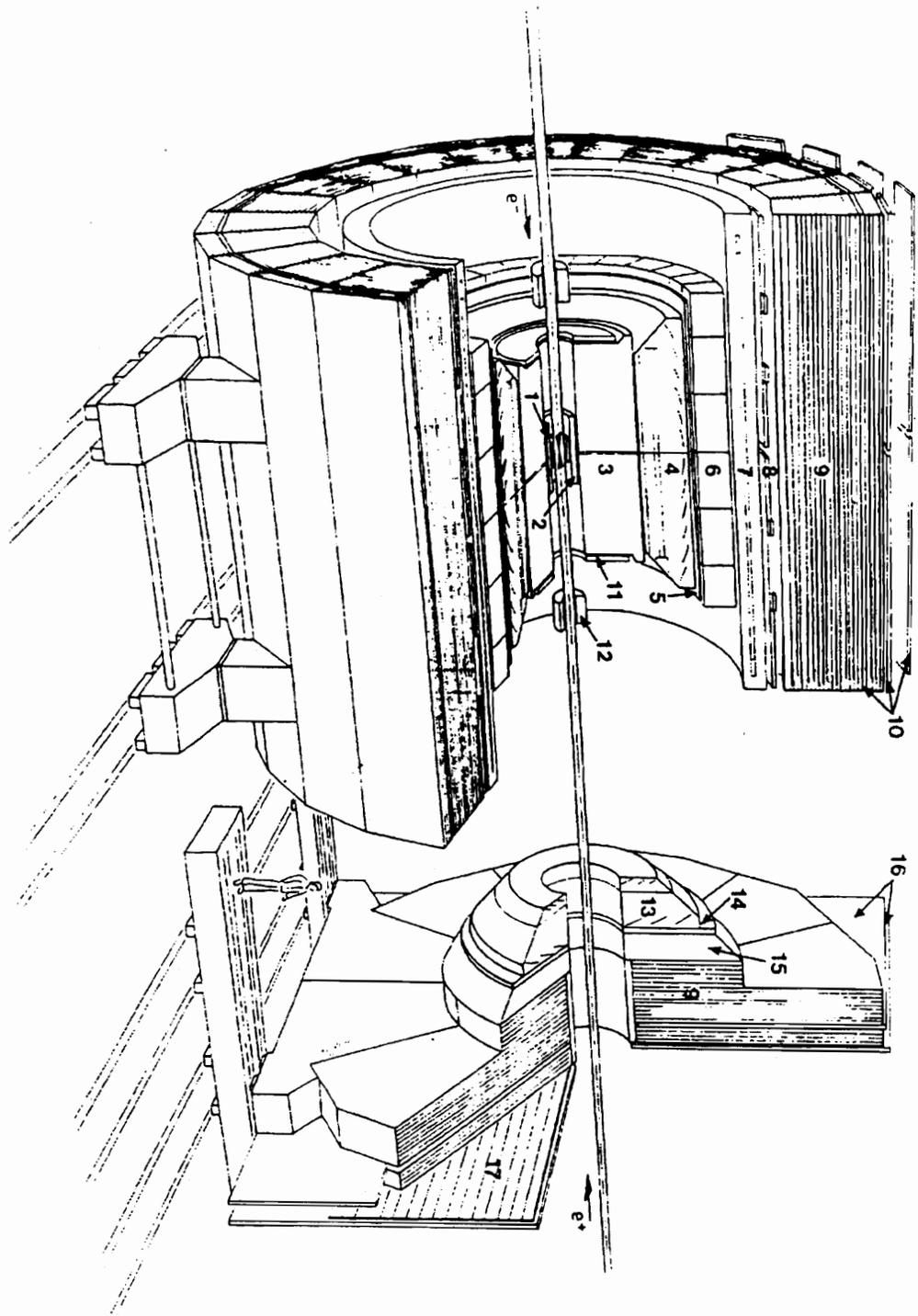


Abbildung 4.1: Perspektivische Aufrißzeichnung des **DELPHI** Detektors. Die einzelnen Detektorkomponenten sind: 1=Vertexchamber, 2=Inner Detector, 3=TPC, 4=Barrel Rich, 5=Outer Detector, 6=HPC, 7=Solenoid, 8=TOF, 9=Hadron Calorimeter, 10=Myon Chambers, 11=Forward Chamber A, 12=SAT, 13=Forward Rich, 14=Forward Chamber B, 15=Forward Electromagnetic Calorimeter, 16=Forward Myon Chambers, 17=Forward Scintillators.

Kapitel 4. Das DELPHI-Experiment

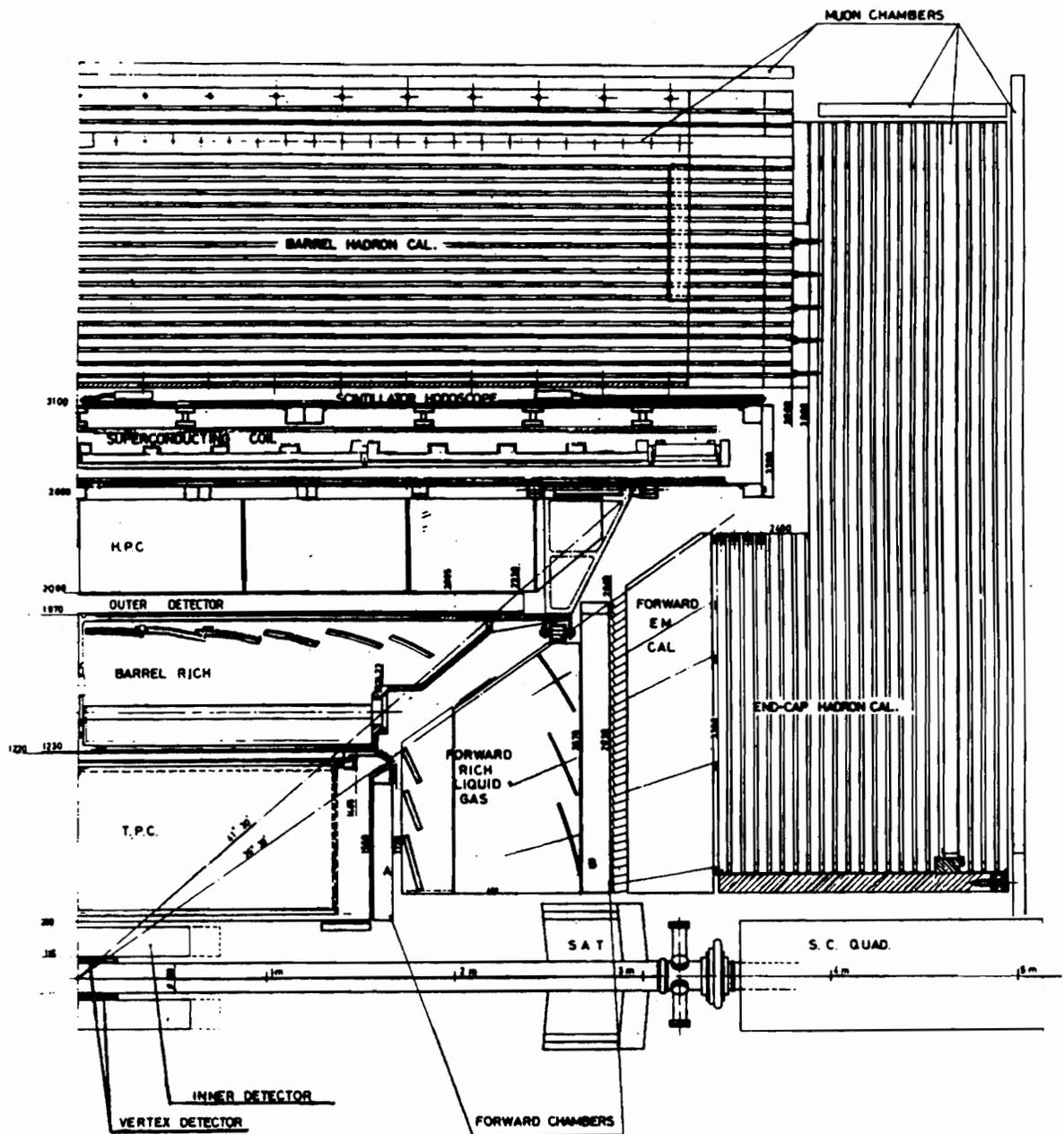


Abbildung 4.2: Längsschnitt durch den DELPHI Detektor

4.1. Die supraleitende Spule

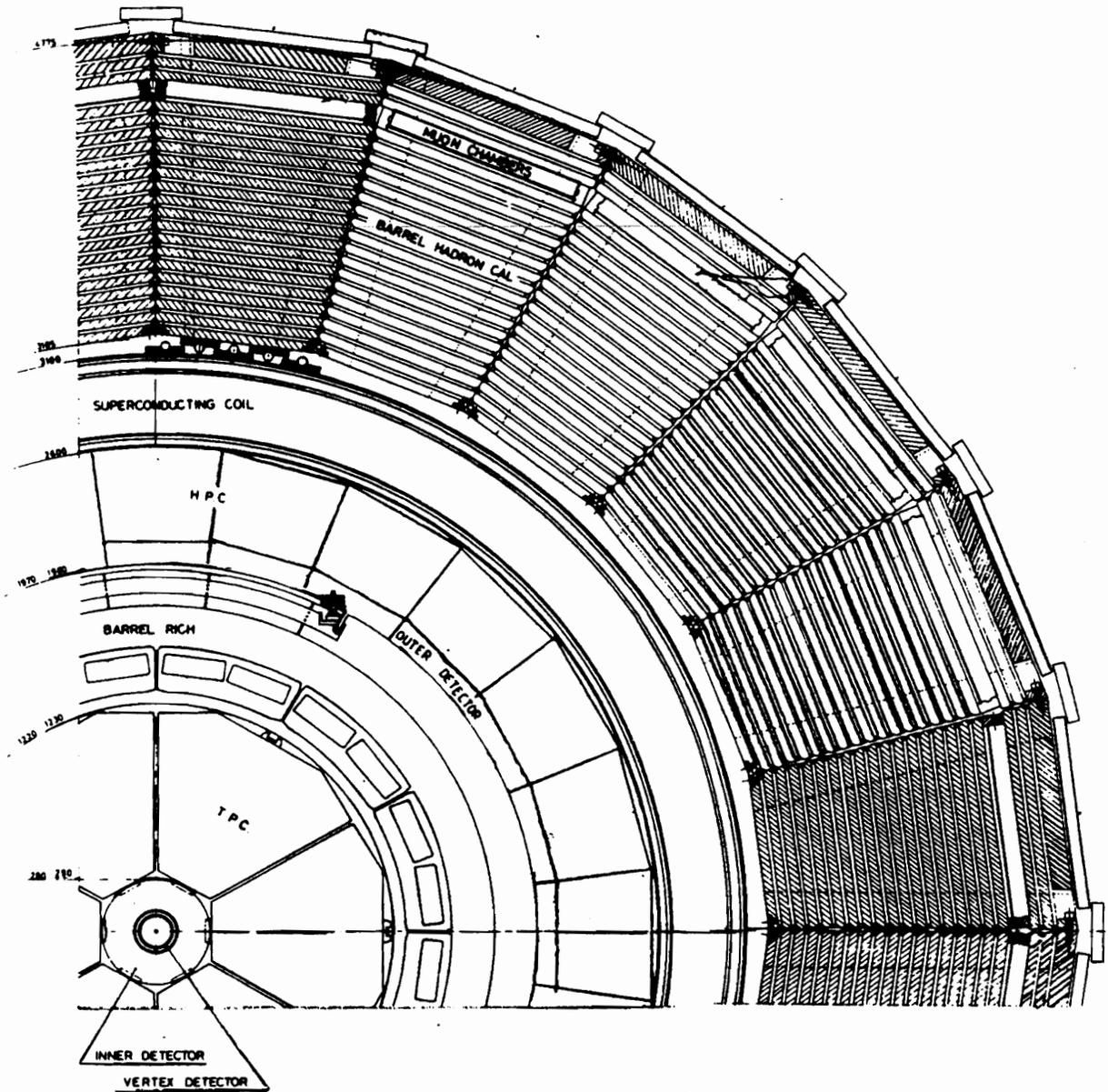


Abbildung 4.3: Schnittbild eines Quadranten des DELPHI Detektor

4.2 Die Spurkammern

Um das Strahlrohr herum ist als erste Detektorkomponente auf 24 cm Länge der **Vertex-Detektor (VD)** aufgebaut. Seine beiden konzentrischen Schalen, die sich im Abstand von 9 und 11 cm um das Strahlrohr befinden, bestehen aus jeweils 24 Modulen, die in ϕ mit einem Überlapp von ca. 10% angeordnet sind, um Akzeptanzlöcher im Detektor zu vermeiden. Jedes Modul beinhaltet vier in z -Richtung angeordnete Silizium-Micro-Strip-Detektoren mit einer Dicke von jeweils $300 \mu\text{m}$. Die einzelnen Detektoren sind auf einer Länge von 59 mm und einer Breite von 25.6 bzw. 32 mm empfindlich. Dies entspricht jeweils 512 bzw. 640 Auslesekanälen. Die Gesamtzahl der Ausleseleitungen beträgt allein für den Vertex-Detektor 54254. Mit der vom Vertexdetektor gelieferten Information ist es möglich die $r\phi$ -Koordinate geladener Teilchen sehr genau zu bestimmen, und so Sekundärvertices zu rekonstruieren. Die Ortsauflösung beträgt ca. $5 \mu\text{m}$ für einzelne Spuren, wobei zwei Spuren noch im Abstand von $\leq 100 \mu\text{m}$ getrennt werden.

Im Anschluß an den Vertexdetektor befindet sich der Innerdetektor (**ID**). Er besteht aus zwei konzentrischen Teilen :

- einer inneren Jetkammer und
- fünf zylindrische Lagen Vieldrahtproportionalkammern.

Die Jetkammer, die den Bereich von $r = 11.8 \text{ cm}$ bis $r = 22.3 \text{ cm}$ einnimmt, ist aus 24 identischen Sektoren mit je 24 Drähten zusammengesetzt. Die Bestandteile des Füllgases, sowie das elektrische Feld, welches zwischen 1 und 2 kV/cm variieren kann, sind so gewählt, daß sich eine zu r proportionale Driftgeschwindigkeit ergibt. Die sich so ergebenden annähernd gleichen Driftzeiten ermöglichen einen schnellen Trigger auf geladene Spuren. Pro Draht ergibt sich eine Ortsauflösung von ca. $110 \mu\text{m}$. Die fünf Lagen Proportionalkammern, die sich von $r = 23 \text{ cm}$ bis $r = 28 \text{ cm}$ erstrecken, bilden den zweiten Teil des Innerdetektors. Sie sind ein wesentlicher Teil im DELPHI Spur-Trigger, dienen aber auch zur eindeutigen Ortsbestimmung von Spuren. Neben der $r\phi$ -Information wird mittels Kathodenauslese auch die z -Koordinate der Teilchen mit einer Auflösung $< 1\text{mm}$ gemessen.

Direkt auf dem Barrel-RICH sitzen die 24 Module, aus denen der **Outerdetektor** aufgebaut ist. Die aus je 145 Driftkammern in fünf Lagen kon-

struierten Module, haben eine Länge von 4.7 m. Der überlappende Aufbau der verschiedenen Lagen ist in Abbildung 4.4 zu sehen. Die Driftkammern werden im limited streamer mode betrieben.

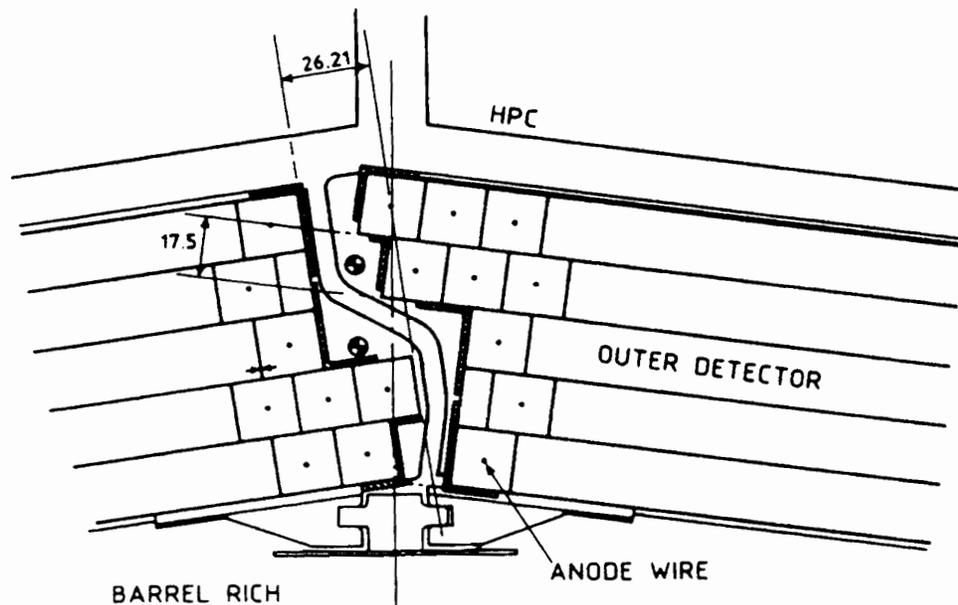


Abbildung 4.4: Anordnung der Module des Outer Detector

Der Outerdetektor versorgt das Triggersystem von DELPHI mit schneller $R\phi$ - und z -Information. Des weiteren dienen seine Meßwerte dazu die Impulsaufösung um einen großen Faktor, welcher für schnelle Teilchen > 5 sein kann, zu verbessern. Um dies erreichen zu können, stellen alle Lagen $R\phi$ -Information bereit, und aus den Werten von drei Lagen läßt sich durch Zeitvergleich der Signale eine schnelle z Information gewinnen.

Ausgelegt für eine Auflösung von $\delta R\phi = 300\mu m$ und $\delta z = 7cm$, ergaben Kalibrationsmessungen $\delta R\phi = 80\mu m$ und $\delta z = 2cm$. Aus dem bisherigen Betrieb während der Laufzeit von LEP ergeben sich folgende vorläufige Ergebnisse $\delta R\phi = 110\mu m$ und $\delta z = 4.4cm$. Die Leistungsfähigkeit dieser Detektorkomponente zeigt auch die Tatsache, daß es möglich ist selbst die Ladung von Spuren mit höchsten Impulsen nur aus dem Spurelement des Outerdetektors zu bestimmen.

Die Zeit-Projektionskammer (TPC) ist der zentrale Spur- Detektor des DELPHI-Experiments. Die von ihr gelieferten Informationen bilden den Ausgangspunkt der Mustererkennung zum Zwecke der Spurrekonstruktion. Ihr räumliches Ausmaß beträgt in longitudinaler Richtung $2 \times 150\text{cm}$ und reicht radial von $r = 28\text{cm}$ bis $r = 122\text{cm}$. Der Aufbau gliedert sich in zwei separate Hälften, die aus jeweils sechs Sektoren bestehen. Die Sektoren sind aus 16 Reihen Kathodenpads und 183 Anodendrähten, die alle mit FADCs (sehr schnelle Analog-Digital Wandler) ausgelesen werden, aufgebaut. Die vorläufigen Werte der Ortsauflösung für einen Raumpunkt innerhalb der TPC betragen $180\text{-}280 \mu\text{m}$ (abhängig von ϕ und z) in der $r\phi$ - Koordinate und unter $900 \mu\text{m}$ in der z -Koordinate. Dabei wird die z -Komponente einer Teilchenspur mit Hilfe der Driftzeit bestimmt und die $r\phi$ -Projektion über den Ladungsschwerpunkt auf den Pads gemessen. Aus den Pulshöhen der Anodendrähte wird die spezifische Ionisation $\frac{dE}{dx}$ bestimmt, die für kleine Impulse zur Teilchenidentifikation beiträgt.

4.3 Die elektromagnetischen Kalorimeter

Die High-Density Projection Chamber (HPC) ist als Teil des Barrelbereichs von DELPHI die in dieser Größenordnung erste Anwendung des Zeit-Projektions-Prinzips in der Kalorimetrie. Innerhalb des Magnetfeldes der supraleitenden Spule gelegen umfaßt die HPC das Barrel-Innere. Die 144 Module aus denen sie aufgebaut ist sind in z -Richtung in 6 hintereinander liegenden Ringen angeordnet, welche aus je 24 Segmenten bestehen. Den schematischen Aufbau eines einzelnen Moduls zeigt Abbildung 4.5 . Der auf diese Weise erfaßte Winkelbereich ist $43^\circ < \theta < 137^\circ$. Jedes Modul enthält Blei als Konvertermaterial, das in 41 Lagen, getrennt durch einen jeweils 8 mm dicken gasgefüllten Zwischenraum ($\text{Ar} - \text{CH}_4 - \text{Gemisch}$), angeordnet ist. Ein Spannungsgefälle von 100V/cm zwischen zwei benachbarten Drähten erzeugt ein longitudinales Driftfeld. Dadurch werden Ionisationsladungen von Schauern und Spuren auf eine Proportionaldrahtebene am Ende des Moduls geleitet. Um die zum schnellen Triggern der Detektorauslese notwendige Information zu gewinnen, ist nach 4.5 Strahlungslängen eine Ebene mit Szintillationszähler in die HPC integriert. Im Vergleich zur Auslesezeit der HPC von $18 \mu\text{sec}$ liefern sie ein sehr schnelles Signal ($< 1\mu\text{sec}$) für elektromagnetische Schauer. Das mit der HPC verfolgte Ziel, ist die genaue Messung der räumlichen Ladungsverteilung, welche von elek-

4.3. Die elektromagnetischen Kalorimeter

tromagnetischen Schauern und geladenen Hadronen stammen. Dies ist auch bei komplexen Ereignistopologien wie Mehr-Jet-Ereignissen möglich. Teststrahlraten zur Schauerachsenrekonstruktion ergaben eine Winkelaufösung von $(36/\sqrt{E} + 2.5)$ mrad in θ und $(97/\sqrt{E} + 10)$ mrad in ϕ für festes θ und ϕ .

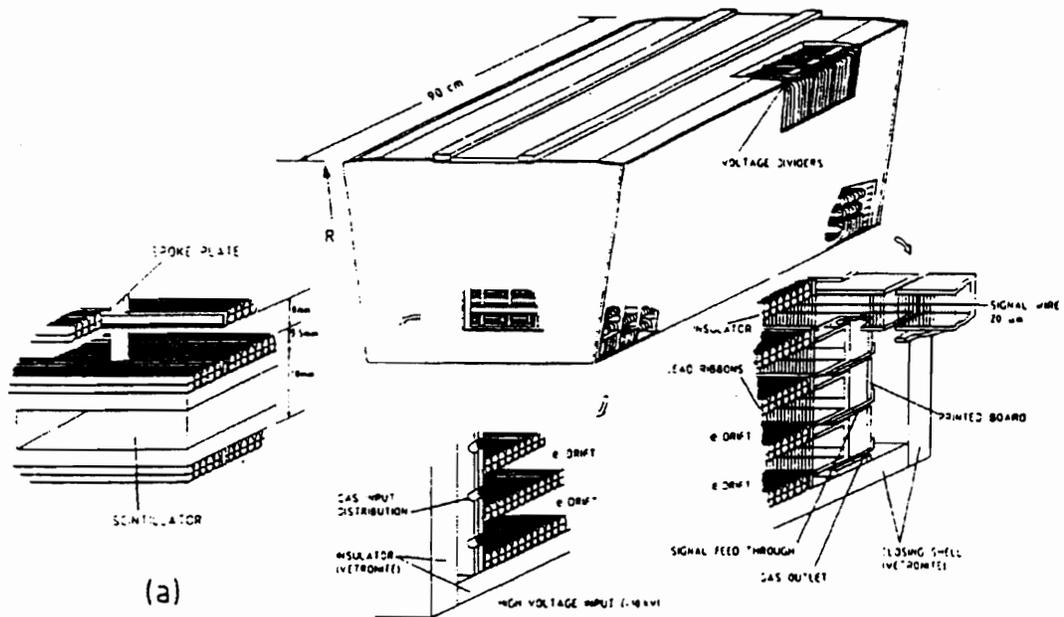


Abbildung 4.5: Der Aufbau eines HPC-Modules

Um im Vorwärtsbereich eine gute Energieauflösung, sowie eine gute räumliche Auflösung zu erreichen, wurde ein Bleiglaszähler (FEMC) als elektromagnetisches Kalorimeter entwickelt. Es besteht aus zwei ringförmigen Scheiben mit 5 m Durchmesser, die aus zusammen 9064 Bleiglasblöcken in Form von Pyramidenstümpfen aufgebaut sind. Die Bleiglasblöcke, welche eine Dicke von 20 Strahlungslängen und eine Grundfläche von $5 \times 5 \text{ cm}^2 \sim 1^\circ \times 1^\circ$ besitzen, werden mit neuentwickelten Vakuum-Phototrioden ausgelesen. Aus Teststrahlraten wurde die Energieauflösung $\sigma_E/E = [(0.35 + 5/\sqrt{E})^2 + (6/E)^2]^{1/2} \%$ mit E in GeV ermittelt.

4.4 Die Flugzeitähler (Time Of Flight counters)

Von den in DELPHI vorhandenen Szintillationszählern sollen hier nur noch die Flugzeitähler kurz erwähnt werden.

Gerade noch außerhalb der Spule ist eine Lage Flugzeitähler angebracht. Die einzelnen Zähler (siehe Abbildung 4.6) überdecken zusammen den Bereich von $\theta = 41^\circ$ bis $\theta = 139^\circ$ mit einer 6 cm breiten Aussparung um $\theta = 90^\circ$. Ihrem Verwendungszweck als schneller Trigger für kosmische Ereignisse, sowie Ereignisse aus der Strahlwechselwirkung werden Sie mit einer Zeitauflösung von $\delta_t = 1.2ns$ gerecht. Die Flugzeitähler können so als Veto auf den Untergrund aus der Höhenstrahlung (kosmische Müonen) verwendet werden.

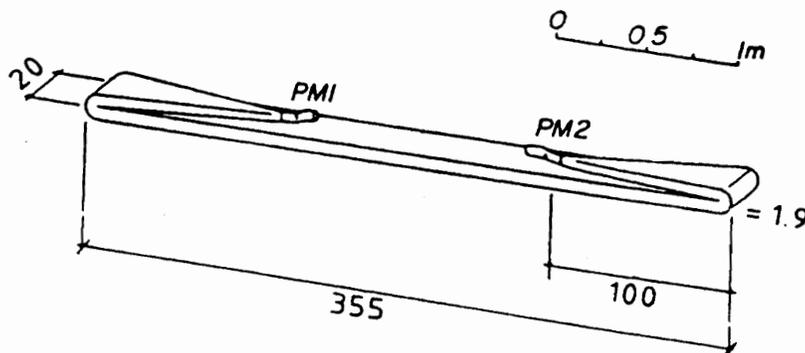


Abbildung 4.6: Ein einzelner TOF-Zähler

4.5 Die Müonkammern

Der Barrel- (B-MU) und der Vorwärts Müon-Detektor (F-MU) bestehen aus je zwei Ebenen von Driftkammern; eine erste befindet sich nach 90 cm Eisen innerhalb des Eisenjochs und eine zweite außerhalb des Jochs hinter weiteren 20 cm Eisen. Das Eisen hat die Funktion eines Hadronabsorbers.

Da die Müonen im Vergleich zu den Hadronen eine geringe Absorptionswahrscheinlichkeit in Materie besitzen, durchqueren sie das Eisen — sieht man von Vielfachstreuungen ab — praktisch unbeeinflusst.

Die daraus zur Teilchenidentifikation gewonnene Information genügt einer Auflösung der extrapolierten Spuren von

$$\begin{aligned}\delta_{R\phi} &\sim 4mm \text{ und } \delta_z \sim 25mm \text{ im B-MU und} \\ \delta_{xy} &\sim 1mm \text{ im F-MU.}\end{aligned}$$

4.6 Die Kontroll-, Auslese- und Triggersysteme

Basierend auf einem lokalen Computer-Netzwerk (DELPHI local area network) mit μ VAX Kontrollrechnern gibt es bei DELPHI ein automatisches System zum Anzeigen (Monitoring) und Kontrollieren der technischen Aspekte (Temperaturen, Spannungen, Gasdrücke etc.) des Experiments.

Die Ausleseprozedur ist in drei Phasen unterteilt und mit einem vierstufigen Triggersystem verknüpft. Der zugrundeliegende Systemgedanke ist, die Ereignisse nach passieren einer Verarbeitungsstufe in lokalen Zwischenspeichern abzulegen, damit schon mit dem nächsten Event gearbeitet werden kann, während auf höherer Stufe noch die vorhergehenden Annihilationsergebnisse prozessiert werden. Auf diese Weise soll die Totzeit des Detektors auch bei hohen Luminositäten und großem Untergrund möglichst klein gehalten werden.

Während die ersten beiden Triggerstufen, die aus festverdrahteten Hardwarekomponenten bestehen, synchron zur alle $22\mu s$ stattfindenden Strahlwechselwirkung arbeiten, laufen die Programme des 3. und 4. Triggerlevels dazu asynchron. Die Triggerelektronik ist im FASTBUS-Standard ausgeführt. Da im ersten Betriebsjahr noch nicht alle FASTBUS-Komponenten installiert waren, wurden ihre Aufgaben von einer NIM-Elektronik übernommen.

In Phase 1 des Auslesens werden die Front-End-Buffer (FEB) der Detek-

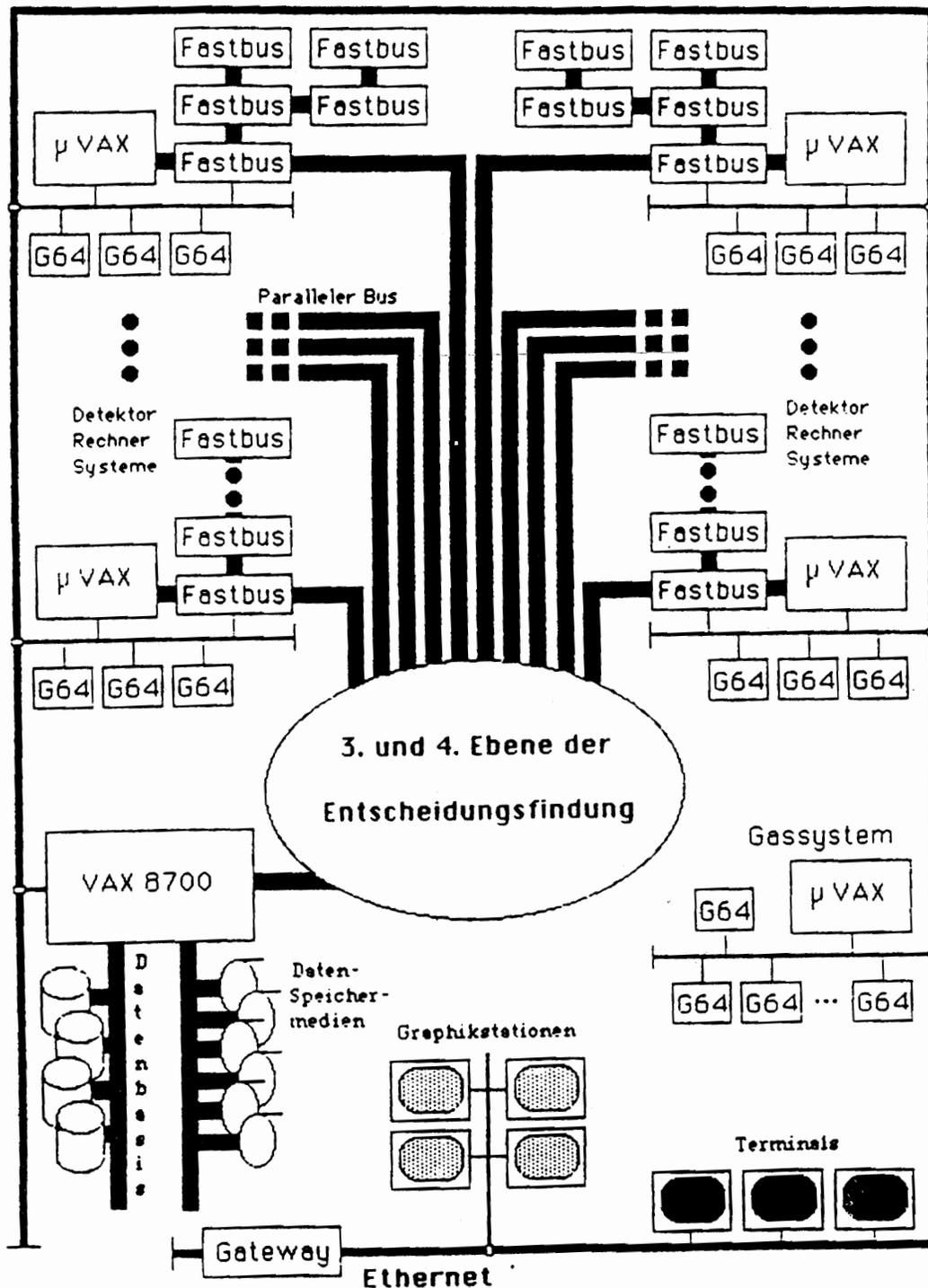


Abbildung 4.7: Die Architektur des Auslesesystems

torkomponenten gefüllt, was je nach Detektor zwischen einigen hundert Nanosekunden und bis zu 23 μs dauert. Die anschließende Entscheidung ist in der ersten und zweiten Triggerstufe nach 3 μs bzw. 40 μs gefallen. Bei einer positiven Entscheidung werden die Daten in der 2.Phase, der Hauptauslesephase, je nach Detektor in einen der 15 Multi-Event-Buffer (MEB) übertragen.

Typische Verarbeitungszeiten für die nun folgenden 3. und 4. Triggerstufen sind 30ms bzw. 300-500ms. In der 3. Phase des Auslesens werden die Meßwerte aus den verschiedenen MEB's zum Global-Event-Buffer (GEB) transferiert. Für eine vorläufige schnelle Physikanalyse und zur Untergrundbeseitigung wird auf der 4. Triggerstufe eine erste Identifikation der unterschiedlichen Ereignistypen (multihadronisch, $e^+e^- \rightarrow e^+e^-$, ...) durchgeführt. Die Ergebnisse dieser Analyse werden schließlich zusammen mit dem GEB von einer VAX8700, welche über der Erdoberfläche, das heißt 100m vom Experiment entfernt steht, auf Magnetplatten und -bänder geschrieben. Die am Auslesesystem (Abbildung 4.7) beteiligten Rechner laufen unter dem Betriebssystem OS/9.

Zur weiteren Analyse und Rekonstruktion der Daten werden die Bänder zu einer 'VAX-Farm' transferiert.

4.7 Die Monte Carlo Simulation

Anhand der von Simulationsprogrammen generierten Ereignisse, die im Idealfall der Realität entsprechen sollen, können Analyseprogramme entwickelt und getestet, sowie Effizienz und Akzeptanz des Experiments bestimmt werden. Aufbauend auf die vorhandenen Programme aus anderen Experimenten (ISR, UA1 etc.) hat DELPHI das Programmpaket DELSIM (DELphi eventgeneration and detector SIMulation) [DE 89/1] entwickelt, während die drei anderen Experimente am LEP die Software GEANT benutzen.

Die Generierung des Prozesses

$$e^+e^- \longrightarrow \gamma, Z^0 \longrightarrow q\bar{q} \longrightarrow \text{Hadronen}$$

und der dabei möglichen Abstrahlung eines oder mehrerer Photonen im Anfangszustand und/oder eines Photons im Endzustand geschieht mit dem Generator DYMU3. Den Endzustand dieses primären Prozesses bilden Photonen und entsprechend ihrem Anteil am hadronischen Wirkungsquer-

schnitt die verschiedenen Quarks. Die so erzeugten primären Quarkpaare werden mit Routinen des Programmpaketes Jetset 7.2[SJ 89] in Gluonen und sekundäre Quarks überführt. Zur Berechnung der Physik (QCD) stehen dazu die zwei Modelle Matrixelement (ME) und Parton Shower (PS) zur Auswahl. Im ME-Modell werden die Feynmangraphen bis einschließlich der Ordnung α_s^2 exakt berechnet. Damit wird die Kinematik, sowie die Interferenz- und Helizitätsstruktur der Partonen genau beschrieben. Das Parton-Shower-Modell dagegen beruht auf der iterativen Abfolge der einfachen Zerfälle :

$$q \longrightarrow q + g,$$

$$g \longrightarrow g + g \text{ und}$$

$$g \longrightarrow q + \bar{q}.$$

Diese grundlegenden Quark- und Gluonzerfälle bilden eine Kaskade, welche abgesehen von der ersten Ordnung als leading-log-Näherung berechnet wird, bis die verbleibende Energie eine einstellbare Grenze (Standardwert 1 GeV) unterschreitet. Einer der Vorteile ist, daß die Anzahl der erzeugten Partonen nicht wie im ME-Modell auf ≤ 4 beschränkt ist. Jedoch werden Interferenzen und virtuelle Effekte nicht korrekt berücksichtigt. Beide Modelle haben, wie bereits angedeutet im Detail ihre spezifischen Vor- und Nachteile, beschreiben die Daten jedoch hinreichend gut. Die nächste Phase bei der Generierung des hadronischen Endzustandes bildet die Fragmentation der Partonen in farblose Hadronen nach dem Modell der Stringfragmentation. Dabei wird der Zerfall von instabilen kurzlebigen Hadronen noch innerhalb des Programs Jetset ausgeführt. Die innerhalb von DELSIM nun folgende Simulation des Detektors beinhaltet die Sekundärwechselwirkungen der Teilchen mit dem Detektormaterial und dem Magnetfeld. Auf ihrem Weg durch die verschiedenen Detektorbestandteile werden dabei neben dem Zerfall der Teilchen folgende Prozesse berücksichtigt :

- Vielfachstreuung in Materie
- Bremsstrahlung
- δ -Strahlen
- hadronische Wechselwirkung

- e^+e^- -Paarproduktion
- Positronvernichtung
- Comptonstreuung
- Photoeffekt.

Die dazu notwendigen detaillierten Informationen über den Detektor werden aus einer umfangreichen Datenbank geholt. Das Ergebnis dieses gesamten Prozesses ist ein multihadronischer Endzustand entsprechend den im Experiment beobachtbaren Teilchen. Die Schnittstelle zur Rekonstruktion und Analyse ist eine Datenstruktur, die der der echten Detektorauslese gleicht. Somit ist die Weiterverarbeitung mit denselben Programmen, denen auch die realen Daten unterworfen werden, gewährleistet.

4.8 Die Rekonstruktion der Ereignisse

Der ersten schnellen Analyse nach dem Auslesen folgt mit dem Programmpaket DELANA (DELphi data ANALysis program) [DE 89/2] eine in mehreren Stufen ablaufende, genaue Rekonstruktion der Ereignisse. Die Ergebnisse jeder Stufe werden mit dem Programm TANAGRA (Track ANALysis and GRAPhics package) [DE 89/3], [BR 87] in einem genormten Format, sogenannten T(ANAGRA)-Bänken, abgespeichert. Die sechs Level dieser Datenstruktur stellen eine immer weiterführende Reduktion der ursprünglichen Rohdaten und Konzentration auf abstrakterem Niveau dar. Abbildung 4.8 zeigt das Schema der Rekonstruktion und den Informationsfluß über die TANAGRA-Bänke.

Mit dem Programm DELGRA (DELphi GRAPhics package) können diese TANAGRA-Bänke eingelesen und die Ereignisse mitsamt den Detektorkomponenten auf einem Grafikbildschirm dreidimensional dargestellt werden.

Nachdem die Daten mit den Informationen aus einer Detektordatenbank kalibriert worden sind (TD), findet die erste Stufe der Spurerkennung statt. Diese lokale Spurfindung beschränkt sich zunächst auf Informationen aus jeweils einem Detektor. Nur in einigen Ausnahmefällen (z.B. Kalorimeter) werden schon Gruppen von Detektoren zusammengefaßt. Die gefundenen

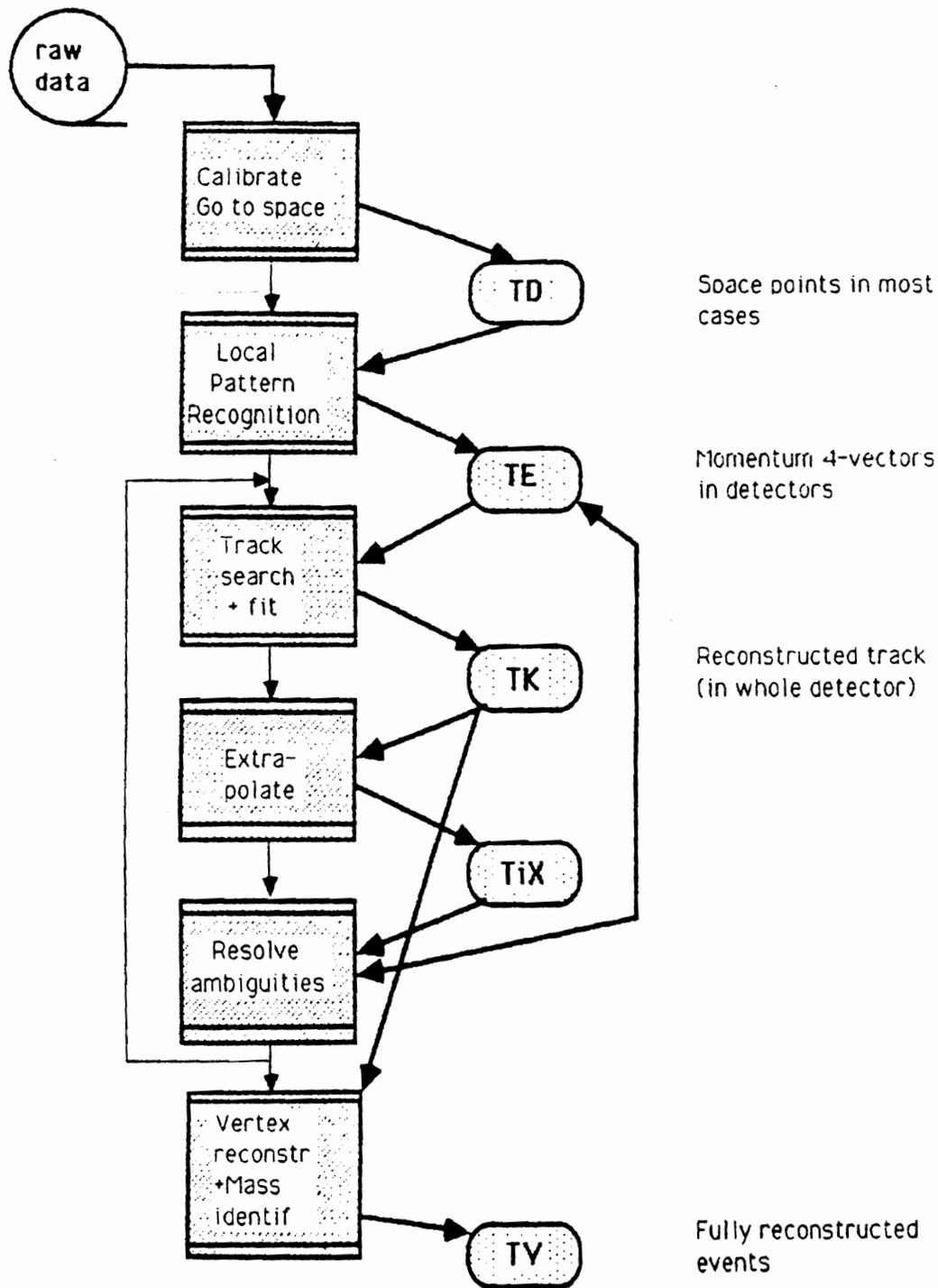


Abbildung 4.8: Die Datenrekonstruktion und der Informationsfluß über die TANAGRA-Bänke

Spurelemente werden in TE-Bänken abgelegt. Im nächsten Schritt wird versucht aus den Spurelemente der einzelnen Komponenten sogenannte Strings zu konstruieren, welche dann in TS-Bänken stehen. Die nachfolgende

Spuranpassung,
Extrapolation der Spuren zu anderen Detektorkomponenten und
nochmalige Spursuche und -anpassung

beseitigt eventuell auftretenden Mehrdeutigkeiten, falls Spurelemente mehreren Strings zugeordnet wurden. In dieser zweiten Stufe der Spurerkennung werden auch bisher nicht assoziierte Elemente verarbeitet und Schauer aus den Kalorimetern mit den rekonstruierten Spuren in Verbindung gebracht. Ist ein Schauer keiner Spur zuzuordnen, so wird angenommen, er stammt von einem neutralen Teilchen. Die Resultate der Spuranpassung werden in den TK-Bänken abgespeichert, wo jede Spur eine eigene Bank belegt. Als nächster logischer Schritt folgt das Zusammenfassen der Spuren, die vermutlich einen gemeinsamen Schnittpunkt haben, zu Bündeln. Darauf aufbauend wird mit geometrischen und kinematischen Mitteln eine Suche nach primären und sekundären Vertices durchgeführt. Zur Massenbestimmung werden hier auch Informationen der Detektoren zur Teilchenidentifizierung verwendet. Nachdem die Werte der Spur-Bündel in den TB-Bänken gesichert wurden, werden die Vertexdaten schließlich in den TV-Bänken abgespeichert.

Zusammengefaßt und auf Band geschrieben ergeben die ursprünglichen Rohdaten und die aus der Rekonstruktion resultierende Bankstruktur ein sogenanntes Master-DST¹, welches den Mitgliedern der Kollaboration für ihre Analysen zur Verfügung steht. Da dieses Master DST jedoch sehr umfangreich ist, wird daraus ein sogenanntes Mini-DST herauskopiert. Dabei werden die Rohdaten weggelassen und die Informationen der verschiedenen T-Bänke in den neu gebildeten PV- und PA-Bänken für Vertices und Teilchen konzentriert. In einer weiteren Reduktionsstufe werden mit dem Programm VECSUB die physikalischen Größen wie Impulse, Massen und Energien der Spuren in ein handlicheres Tabellenformat extrahiert. Diese Tabelle enthält die für eine Reihe von Analysen, wie auch die für diese Arbeit ausreichende Information.

¹DST = Data Summary Tape

Kapitel 4. Das DELPHI-Experiment

Spur- kammer	Akzeptanz			Auflösung $\sigma(mm)$	Energieauf- lösung $\frac{\sigma_E}{E}(\%)$
	$R(cm)$	$ z (cm)$	$\theta(^{\circ})$		
VD	9/11	12	37-143	$r\phi : 0.007$	—
ID	11.8-22.3 23.0-28.0	40 ≤ 50	17-163 30-150	$r\phi : 0.11$ $z : < 1$	—
OD	198-206	≤ 232	43-137	$r\phi : 0.11$ $z : 44$	—
B-MU	~ 445	≤ 185	52-138	$r\phi : 1.5$ $z : 10$	—
F-MU	70-460	463	9-43	$x, y : 1.0$	—
FCA	30-103	155-165	11-33	$x, u, v : 0.3$	—
FCB	53-195	267-283	11-35	$x, u, v : 0.25$	—
TPC	35-111	≤ 134	20- 160	$r\phi : 0.23$ $z : 0.9$	—
Kalori- meter					
HPC	208-260	≤ 254	43-137	$z : 4$ $\phi : 1^{\circ}$	$\frac{23}{\sqrt{E}} + 1.1$
FEMC	46-240	284-340	10-36.5	$\phi : 1^{\circ}$ $\theta : 1^{\circ}$	$\left[\left(0.35 + \frac{5}{\sqrt{E}} \right)^2 + \left(\frac{6}{E} \right)^2 \right]^{1/2}$
HCAL	B: 320-479 FW: 65-460	< 380 340-489	10-170	$\phi : 3.75^{\circ}$ $\theta : 3.0^{\circ}$ $\theta : 2.6^{\circ}$	$\frac{120}{\sqrt{E}}$

Tabelle 4.2: Akzeptanz, Spur- und Energieauflösung der einzelnen Detektorkomponenten.

Kapitel 5

Studie zur Auswahl von b-Quark-sensitiven Variablen

Unabhängig davon welche Methode man benutzt, um eine größere Menge von Ereignissen in kleinere Teilklassen einzuteilen, ist man darauf angewiesen, zuerst die Unterschiede der zu bildenden Klassen herauszufinden. Um b -Ereignisse aus der Gesamtheit der Z^0 -Zerfälle herauszufiltern, ist es notwendig, Variablen zu finden, welche sich für B-Ereignisse im Vergleich zum Untergrund signifikant unterscheiden. Nach der theoretischen Betrachtung des Prozesses bedient man sich, um solche charakteristischen Variablen zu finden, der Daten aus Monte-Carlo Programmen, bei denen man genau weiß zu welcher Klasse ein einzelnes Ereignis gehört. Trägt man die Verteilungen der einzelnen Größen getrennt nach Klassen auf, so ist eine erste grobe Beurteilung meist schon mit dem Auge möglich. Mit Beurteilung ist die Entscheidung gemeint, ob eine Größe zur Diskriminierung geeignet ist oder nicht. Einen genaueren Anhaltspunkt ergeben die Mittelwerte und die Standardabweichungen der Verteilungen. Ob die Information zweier Variablen jedoch redundant ist, läßt sich auf diese Weise genausowenig feststellen, wie die genau Trenngüte, die sich erreichen läßt.

5.1 Die Monte-Carlo-Daten

Diese Studie wurde mit verschiedene Sätzen von vollrekonstruierten Monte-Carlo Ereignissen (siehe Tabelle 5.1) angefertigt, welche von den Gruppen innerhalb der Kollaboration mit dem Programm JETSET 7.2 generiert wurden. Der Grund für diese Aufteilung der Monte-Carlo Produktion ist der hohe Bedarf an Rechenzeit solcher Simulationen, den eine einzelne Gruppe nicht aufbringen kann. Vollrekonstruiert bedeutet, daß die Ereignisse nach der Generierung die Detektorsimulation mit all ihren in Kapitel 4 beschriebenen Sekundärwechselwirkungen durchlaufen haben. Es wurden Datensätze, die nach dem optimierten Matrixelement Modell und solche die nach dem Partonschauer-Modell generiert wurden, verwendet.

Um mit den realen Daten vergleichbare Samples multihadronischer Ereignisse zu erhalten, wurden mit Rücksicht auf die Leistungsfähigkeit des DELPHI-Detektors verschiedene Schnitte auf Spur- bzw. Ereignisebene durchgeführt.

Monte Carlo	Herkunft		nach multihad. Ereignisselekt.			
					+ Ejet > 15 GeV	
			b	nicht b	b	nicht b
LUND opt. ME	Karlsruhe	Lernsample	8438	27561	5360	17375
		Testsample	1010	2249	628	2129
LUND PS	Milano	Lernsample	2937	9526	1899	5826
		Testsample	656	2223	407	1376
LUND PS	Saclay	Lernsample	2337	7655		
		Testsample	467	1592		

Tabelle 5.1: Die benutzten Monte Carlo Samples; ME = Matrix Element Modell, PS = Parton Shower Modell.

Die Definition 'guter' Spuren basiert auf folgenden Anforderungen (track cuts) :

- keine neutralen Spuren
- minimaler Impulsbetrag : 0.1 GeV
- maximaler Impulsbetrag : 50 GeV

- minimaler Polarwinkel θ : 25°
- maximaler Polarwinkel θ : 155°
- kleinste Spurlänge : 50 cm

Die minimale Spurlänge soll gewährleisten, daß genügend Raumpunkte zur genauen Rekonstruktion der Spur vorhanden sind. Der Bereich des polaren Winkels der Spur ist durch den Akzeptanzbereich der Zeitprojektionskammer (TPC), welche der zentrale Spur-Detektor ist, vorgegeben. Die Impulsschnitte dienen dazu unphysikalische Spuren mit zu hohem oder zu kleinem Impuls auszuschließen. Solche Spuren können durch fehlerhafte Rekonstruktion bzw. Fehlinterpretation von Untergrundrauschen als Signal auftreten.

Für die Schnitte auf Ereignisebene (event cuts), die multihadronische Ereignisse separieren sollen, werden nur noch 'gute' Spuren verwendet. Die Kriterien sind im Einzelnen :

- minimale Anzahl guter geladener Spuren: 5 .
- minimale Summe der Impulse d. geladenen Spuren: $\sum |p_i| \geq 15 \text{ GeV}$.
- minimale Summe der Impulse der geladenen Spuren je z-Hemisphäre: $\sum_{z\text{-Hemi.}} |p_i| \geq 3 \text{ GeV}$.
- maximal fehlender Impuls der geladene Teilchen: $|\sum \vec{p}_i| < 20 \text{ GeV}$.
- minimaler Polarwinkel der Sphärisitätsachsen zum Strahl: 40° .

Die minimale Anzahl und der minimale Impuls guter geladener Spuren trägt der Tatsache Rechnung, daß multihadronische Ereignisse durch hohe Multiplizitäten und hohe sichtbare Energie gekennzeichnet sind. Die mittlere geladene Multiplizität liegt, wie die Monte-Carlo Simulation zeigt, bei 21. Obwohl der fehlende Impuls aufgrund des Fehlens der neutralen Teilchen (siehe Kapitel 1) groß ist, wird ein Schnitt auf die obere Grenze durchgeführt. Mit diesem Schnitt sollen hinreichend 'gute' Ereignisse vorselektiert werden. Ereignisse bei denen z.B. Spuren aus irgendwelchen Gründen verloren gegangen sind und die daher einen vergleichsweise großen fehlenden Impuls aufweisen, werden aussortiert. Dasselbe gilt für Ereignisse, deren Energie

sich durch Abstrahlung von γ 's reduziert hat und für solche mit einer $\gamma\gamma$ -Wechselwirkung, die ebenfalls einen multihadronischen Endzustand bilden. Aus den so gewonnenen Samples wurden durch einen 15 GeV Schnitt auf die Jetenergien weitere Teilmengen erzeugt. Diese Teilsamples dienen der Untersuchung, welchen Einfluß das Aussortieren niederenergetischer Jets auf das Ergebnis hat.

Die in Tabelle 5.1 aufgeführten Monte Carlo Ereignisse standen in Form von in Kapitel 4 beschriebenen DST-Bändern zur Verfügung. Aus den physikalischen Größen, welche aus diesen Bändern mit Hilfe des Programms VEC-SUB gewonnen wurden, sind mit einem weiteren im Rahmen dieser Arbeit geschriebenen Analyseprogramm NTP (wie N-TuPels) die im nachfolgenden Abschnitt aufgeführten Variablen berechnet worden. Die Ergebnisse dieser Berechnungen wurden zu verschiedenen Variablensätzen zusammengefaßt und in N-Tupeln abgespeichert. Die N-Tupel dienen danach als Eingabe für das Netzwerkprogramm. Das in anfänglichen Tests praktizierte Verfahren, daß das Netzwerkprogramm direkt an VEC-SUB und NTP angehängt wurde, hat den gravierenden Nachteil, daß die rechenzeitintensive Analyse jedesmal durchgeführt werden muß. Die abgespeicherten N-Tupel beinhalten dagegen alle Informationen, um immer wieder zum Lernen und Testen der Netzwerke verwendet werden zu können.

Zur Definition der Jets wurde der LUND-Cluster-Algorithmus PUCLUS verwendet.

5.2 Die untersuchten Variablen

Nach den in Kapitel 2 angestellten Überlegungen muß man sich nun fragen in welchen meß- oder berechenbaren Größen sich die angeführten Eigenschaften äußern. Die Fragmentationseigenschaften haben Auswirkung auf die longitudinalen Impulsanteile p_L der Teilchen. Es ist jedoch daran zu erinnern, daß nicht die p_L direkt nach der Fragmentation sondern nur die p_L der Spuren des hadronischen Endzustandes gemessen werden können. Wegen der dadurch verursachten starken Verwischung der Unterschiede ist es noch am ehesten Erfolg versprechend die energiereichsten Teilchen zu untersuchen. Dies sind entweder direkt die Impulsanteile p_x, p_y, p_z oder die transversalen und longitudinalen Impulsanteile p_L, p_T bezüglich der Jetachse. Die Verwendung von p_L und p_T hat den Vorteil, daß bei praktisch

gleichem Informationsgehalt, aufgrund der Drehinvarianz der Ereignisse, ein Eingangsknoten des N.N. pro Teilchen eingespart wird. Vielversprechend sind daher auch die komplexeren topologischen Größen, in deren Berechnung die Impulskomponenten eingehen. Betrachtet wurden die Spherizität, die Aplanarität, Oblateness, Thrust, und das Boosted Sphericity Product.

Die geladenen Multiplizitäten lassen sich direkt messen, entweder als Gesamtmultiplizitäten oder Multiplizitäten nach bestimmten Auswahlkriterien.

Die p_T aus der Fragmentation sind quarkmassenabhängig. Dies läßt sich direkt durch die Berechnung der Impulskomponenten der einzelnen Spuren bezüglich der jeweiligen Jetachse untersuchen. Die folgende Aufzählung umfasst eine Beschreibung aller untersuchten Variablen :

- Jet-Multiplizität ($NJET$)
- BSP (boosted sphericity product); das Produkt der Spherizitäten der beiden energiereichsten Jets nach einer Lorentztransformation entlang ihrer Achsen in das Ruhesystem des vermeintlichen B -Mesons.
- Jetenergien ($EJET1, EJET2, EJET3$)

Die Jets sind nach der Jetenergie geordnet, wobei JET 1 der energiereichste Jet ist.

- Spherizität ($SPHER$) ; $0. \leq SPHER \leq 1.$

$SPHER = 0.$ entspricht einem idealen 2-Jetereignis.

$SPHER = 1.$ entspricht einem ideal kugelförmigen Ereignis.

Die Spherizitätsachse ist als Achse mit dem kleinsten 'Trägheitsmoment' definiert:

$$SPHER = \min_{\vec{n}} S(\vec{n}) = S(\vec{n}_3) \text{ mit } S(\vec{n}) = 3 \sum_{i=1}^N p_{i\perp}^2 / 2 \sum_{i=1}^N p_i^2.$$

- Aplanarität ($APLAN$)

Für planare Ereignisse gilt, daß alle Impulsvektoren in einer Ebene senkrecht zur Richtung des größten Trägheitsmomentes liegen.

Die Aplanarität ist ein Maß für die Abweichung eines Ereignisses von einem solchen ideal ebenen (planaren) Ereignis.

$$APLAN = 3 \sum_{i=1}^N (\vec{p}_i * \vec{n}_1)^2 / 2 \sum_{i=1}^N p_i^2.$$

- transversale/longitudinale Impulskomponente bezüglich der Jetachse der drei schnellsten Teilchen der beiden energiereichsten Jets ($PTjet/teilchen, PLjet/teilchen$)

- Oblateness (OBL)

Die Oblateness ist ein Maß für die "Ebenheit" eines Ereignisses und ist mit der Spherizität hoch korreliert.

$OBL = F_{major} - F_{minor}$ mit

$$F_{major} = \max_{\vec{e}_2} \sum_i |\vec{E}_i * \vec{e}_2| / \sum_i |\vec{E}_i|, \vec{e}_2 \perp \vec{e}_1$$

$$F_{minor} = \sum_i |\vec{E}_i * \vec{e}_3| / \sum_i |\vec{E}_i| \approx \min_{\vec{e}} \sum_i |\vec{E}_i * \vec{e}| / \sum_i |\vec{E}_i|.$$

- geladene Multiplizität (NCHG)
- Summe der Energie der geladenen Spuren (ECHAG)
- NTC

Die geladene Multiplizität der beiden energiereichsten Jets nach den zusätzlichen Winkel- und Impulsschnitten :

$$|P| > 0.5 \text{ GeV und}$$

$$\alpha < 15^\circ \text{ mit } \alpha \text{ dem Winkel der Spuren zur Jetachse.}$$

Mit diesen beiden zusätzlichen Schnitten wird versucht, die Zerfallsprodukte des Mesons von den begleitenden Teilchen aus der Fragmentation zu separieren, welche unter Umständen eine größere α und/oder einen kleineren Impuls haben können. Für Spuren die ihren Ursprung im Zerfall des B -Mesons haben ist dies weniger der Fall.

- PTNTC

Die Summe der Impulsquadrate der Teilchen, die zur Multiplizität NTC beitragen.

- Missing Impuls (MISP)

Die vektorielle Impulssumme aller Spuren eines Ereignisses ist aufgrund des Impulserhaltungssatzes im Idealfall gleich Null. Der Betrag der Größe MISP ist die Abweichung von diesem Idealfall.

Nicht zuletzt wegen der hervorzuhebenden Leistungsfähigkeit bedarf das BSP einer ausführlicheren Erläuterung. Das BSP ist das Produkt der Sphärizitäten in den beiden Hemisphären eines Ereignisses nach einer Lorentztransformation (boost) mit einem β , welches der Geschwindigkeit entspricht, die für ein b -Quark am LEP erwartet wird. Diese topologische Ereignisvariable wurde erstmals beim TASSO-Experiment benutzt [BN 90]. Wie oben erklärt ist die normale Sphärizität ein Maß für die Kugelförmigkeit eines Ereignisses. Die Überlegung ist nun die folgende : Das B -Meson zerfällt, betrachtet man es in seinem Ruhesystem wie jedes Teilchen mit Spin 0 isotrop. Nun sind die resultierenden Teilchen aus der e^+e^- -Kollision nicht in Ruhe. Sie besitzen abhängig davon in welcher Hemisphäre sie sich befinden eine zusätzliche positive bzw. negative Geschwindigkeitskomponente bezüglich der Sphärizitätsachse. Diese Komponenten müssen mittels einer Lorentztransformation mit einer entsprechenden Geschwindigkeit v und damit mit einem bestimmten β kompensiert werden, will man die Teilchen im Ruhesystem des B -Mesons betrachten. Dieses geeignete β gilt es experimentell zu ermitteln. Da die Fragmentationsfunktion der b -Quarks sehr scharf 'gepeakt' ist, ist es zulässig und vor allem praktikabel für alle b -Ereignisse ein gemeinsames β anzunehmen. Im Ruhesystem betrachtet wird aufgrund der hohen Masse des b -Quarks und damit auch des B -Mesons erwartet, daß b -Quark-Jets sphärischer sind als die von leichten Quarks. Die Sphärizität sollte daher im Mittel mehr gegen eins verschoben sein, und damit ein höheres BSP ergeben. Da alle Ereignisse mit dem demselben für b -Ereignisse typischen β transformiert werden, sollte sich dieser Effekt noch verstärken, da *nicht- b* -Ereignisse weiter 'verformt' werden, ihre Sphärizität dadurch kleiner wird. Ein Schnitt auf das BSP sollte theoretisch eine gute Trennung liefern. Das Problem ist jedoch, daß man die Zerfallsprodukte des B -Mesons nicht getrennt betrachten und transformieren kann. Sie bilden mit den begleitenden Teilchen aus der Fragmentation jeweils einen Jet. Da man von einem $b\bar{b}$ -Paar ausgeht, transformiert man die beiden energiereichsten Jets getrennt, berechnet jeweils die Sphärizitäten und multipliziert diese dann miteinander zum BSP. Ein nicht- b Ereignis sollte sich durch ein BSP nahe Null auszeichnen, während für ein b -Ereignis das BSP näher zur Eins verschoben sein sollte. Die Tatsache, daß die Separationsfähigkeit des BSP, welche sich hauptsächlich auf die p_t der Teilchen und damit auf die hohe Masse des b -Quarks stützt, nicht so gut wie vielleicht erwartet ist, liegt u.a. an den begleitenden Teilchen aus der Fragmentation, die sich dem isotropen B -Mesonzerfall überlagern. Ein weiterer Punkt ist, daß jedes nicht nachgewiesene Teilchen aus dem Zerfall natürlich direkten Einfluß auf die

Sphärizität und damit das BSP hat. Trotzdem zeigt sich für ein geeignet gewähltes β ($\beta = 0.92$) [SCH 91] eine kleine, im Vergleich zu den anderen Größen aber gute Unterscheidung der beiden zu trennenden Klassen. In der Abbildung 5.1 sind die Verteilungen für b und nicht- b Ereignisse dargestellt. Anhand des BSP läßt sich noch einmal die Misere darstellen, falls man sich nur auf topologische Größen zum b -Quark Tagging beschränkt. Die Unterschiede in den Variablen der beiden zu trennenden Klassen sind sehr klein. Wir brauchen ein Verfahren, welches bei seiner Entscheidung alle kleinen Differenzen zusammen berücksichtigt.

In den Abbildungen 5.1 bis 5.3 sind die Verteilungen für die oben besprochenen Observablen nach Klassen getrennt dargestellt. Es zeigte sich, daß in den Monte-Carlo Daten die nach dem Parton Shower Modell generiert wurden, die Unterschiede der beiden Klassen in einzelnen Größen schwächer ausgeprägt sind, als in den Daten die nach dem Matrixelement Modell berechnet wurden. Die trifft besonders auf folgende Variablen zu :

- PT und PL, SPHER, OBL, NCHG, ECHAG, NTC ,PTNTC

Die Ursache für diese Unterschiede sind wohl die verschieden gut oder schlecht an die realen Daten angepaßten Parametersätze der Modelle, was die Vergleichbarkeit der damit erreichbaren Ergebnisse einschränkt. Diese Unterschiede haben natürlich auch direkte Auswirkungen auf das erreichbare Ergebnis, wie in Kapitel 6 bestätigt wird.

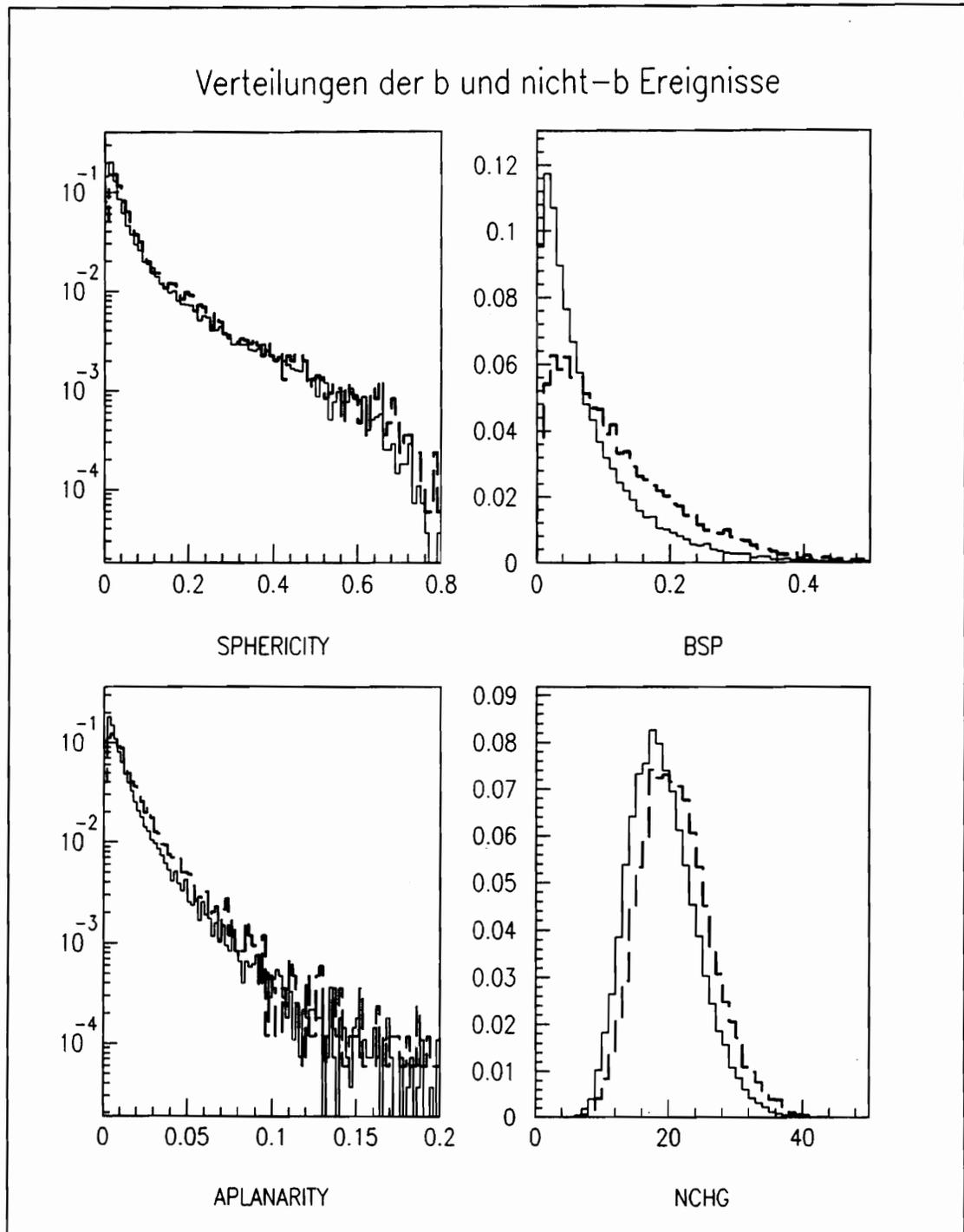


Abbildung 5.1: Verteilungen für $---$ b und $—$ nicht- b Ereignisse.

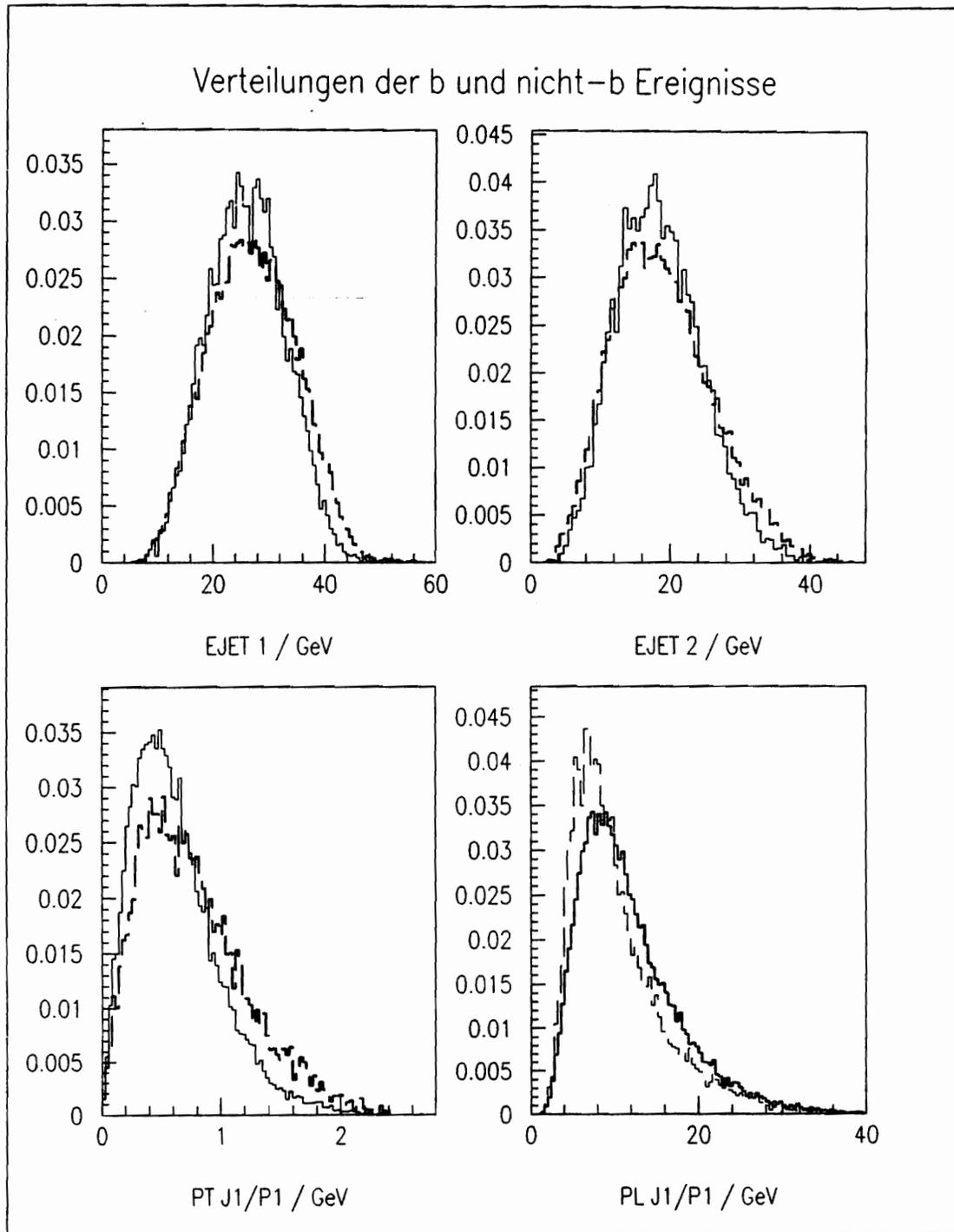


Abbildung 5.2: Verteilungen für $---$ b und $---$ nicht- b Ereignisse.

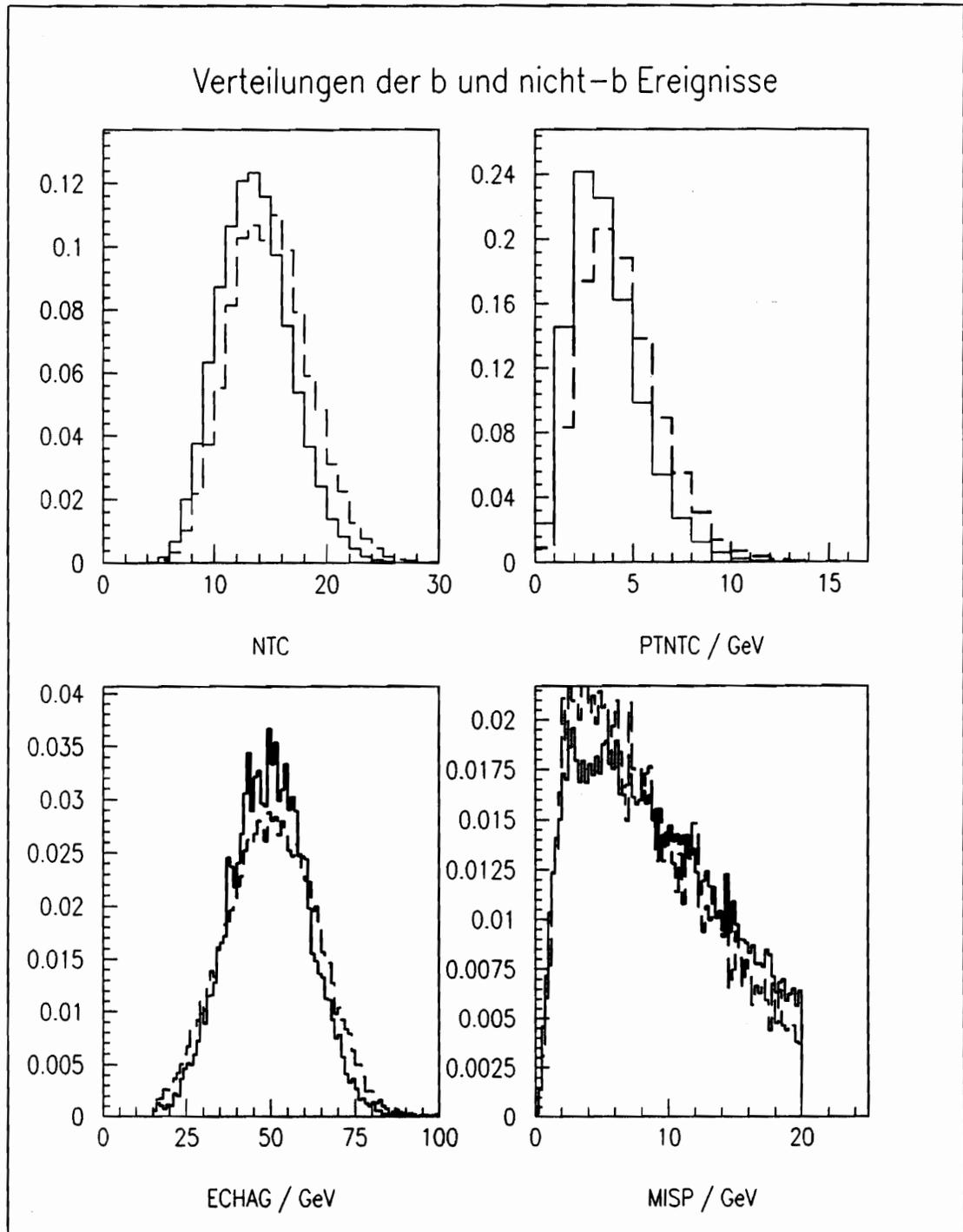


Abbildung 5.3: Verteilungen für $---$ b und $---$ nicht- b Ereignisse.

5.2.1 Verschiedene Variablensätze

Die Variablen wurden nach einer Vorselektion in verschiedene Variablensätze eingeteilt. Die Kriterien waren :

- die Differenz der Verteilungen, wie sie mit dem Auge beurteilt werden können und die Differenz ihrer Mittelwerte;
- die Rangfolge, die das Fisher-Diskriminaten-Verfahren aufstellte, welches in einer parallel durchgeführten Arbeit von Steffi Scheu [SCH 91] benutzt wurde. Dieses Verfahren berücksichtigt Korrelationen zwischen den Größen. Es liefert deshalb Hinweise auf redundante Informationen.
- Veröffentlichungen ([WU 91], [BE 90], [BO 90], [CS 90], [DE 90/2]), deren Ergebnisse überprüft werden sollten.

Die benutzten Variablensätze sind :

Satz *A* ; 23 Variablen

BSP, EJET1, EJET2, SPHER, APLAN,
PT J1/P1, PL J1/P1, PT J1/P2, PL J1/P2, PT J1/P3, PL J1/P3,
PT J2/P1, PL J2/P1, PT J2/P2, PL J2/P2, PT J2/P3, PL J2/P3,
OBL, NCHG, ECHAG, MISP, NTC, PTNTC.

Satz *B* ; 9 Variablen

BSP, SPHER, NTC, PTNTC
PT J1/P1, PT J1/P2, PL J1/2,
PT J2/P1, PT J2/P2.

Satz *C1/C2/C3* ; 12 bzw. 22 Variablen

PX J1/P1, PX J1/P2, PX J1/P3,
PY J1/P1, PY J1/P2, PY J1/P3,
PZ J1/P1, PZ J1/P2, PZ J1/P3.
EJET1, NCHG, SPHER.

Für den zweiten Jet ist Satz *C2* entsprechend definiert.

Satz *C3* ist als die Vereinigung der beiden Variablensätze *C1* und *C2* definiert.

Während in den Variablensätzen *A* und *B* ein Eingabevektor Informationen über das gesamte Ereignis enthält, sind in den Variablensätzen *Cx* jeweils nur Informationen aus dem ersten und/oder zweiten Jet enthalten.

Kapitel 6

Ergebnisse und Interpretation

Die durch einen einfachen Schnitt auf die am besten separierende Einzelvariable *BSP* erreichbaren Ergebnisse sind :

Schnitt	Effizienz	Reinheit
0.06	64 %	31 %
0.19	17 %	40 %

Reinheiten deutlich über 40 % sind mit diesem simplen Verfahren nicht zu erreichen. Es folgen Resultate, die mit einem dreischichtigen vorwärtsvermittelnden Netzwerk ermittelt wurden. Eine typische Verteilung der Output-Werte eines Netzwerkes für die beiden Klassen *b* und *nicht - b* Ereignisse ist in Abbildung 6.1 zu sehen. Im Gegensatz zur idealen Trennung, wie sie in Kapitel 3 dargestellt wurde, besteht hier eine immer noch große Überlappung der beiden Klassen. Die Ergebnistabellen enthalten folgende Angaben :

Netzwerk	:	Anzahl der Knoten, verwendete Monte-Carlo Samples und verwendeter Variablensatz
Parameter	:	Lernrate η und Dämpfungsfaktor α
Besonderheiten	:	Anmerkungen über Bias, Energieschnitte etc.
Schnitt	:	Schnitt auf die Ausgangsverteilung des Netzwerkes
Effizienz	:	Effizienzwerte bei entsprechendem Schnitt
Reinheit	:	Reinheitwerte bei entsprechendem Schnitt

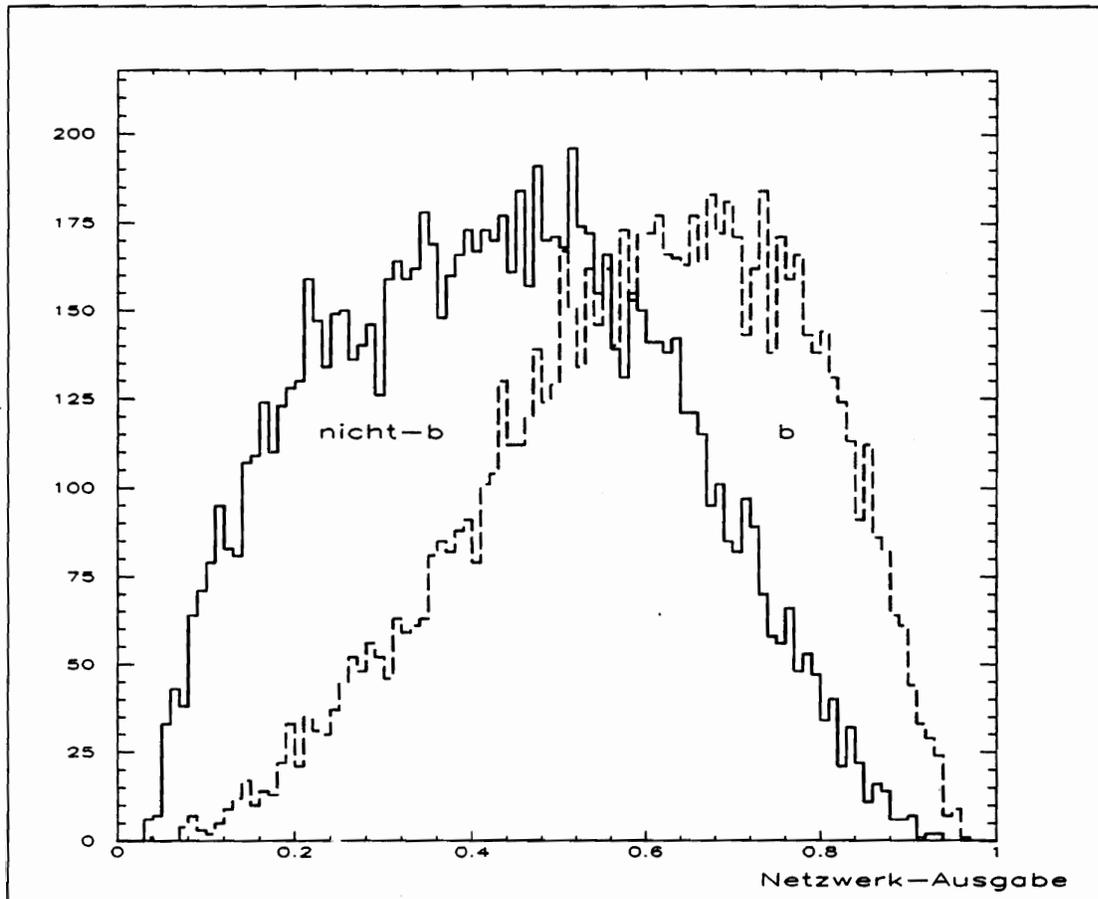


Abbildung 6.1: Eine typische Verteilung der Output-Werte eines neuronalen Netzwerkes (23/15/1) für die beiden Klassen b und $nicht - b$ Ereignisse. Im Vergleich zum 'idealen' Output (siehe Kapitel 3) ist ein großer Überlapp vorhanden, der die erreichbare Leistungsfähigkeit begrenzt.

Ist für α, η ein Bereich angegeben (z.B. $\eta = 0.01 - 0.1$), so wurde die Größe während des Trainings vom Anfangswert 0.01 bis zum Endwert 0.1 schrittweise verändert (dynamisches Lernen).

Tabelle 6.3 führt die Ergebnisse für die Variablensätze A und B auf. Die Reinheitswerte sind bei gleicher Effizienz für ME-Daten um bis zu 8% besser als für mit PS-Daten trainierte Netze. In den Ergebnissen, die mit Daten, welche dem 15 GeV Schnitt auf die Jetenergie unterliegen, ermittelt wurden, betragen die Unterschiede teilweise über 10%. Da die Unterschiede der beiden Klassen in den einzelnen Größen, wie in Kapitel 5 beschrieben,

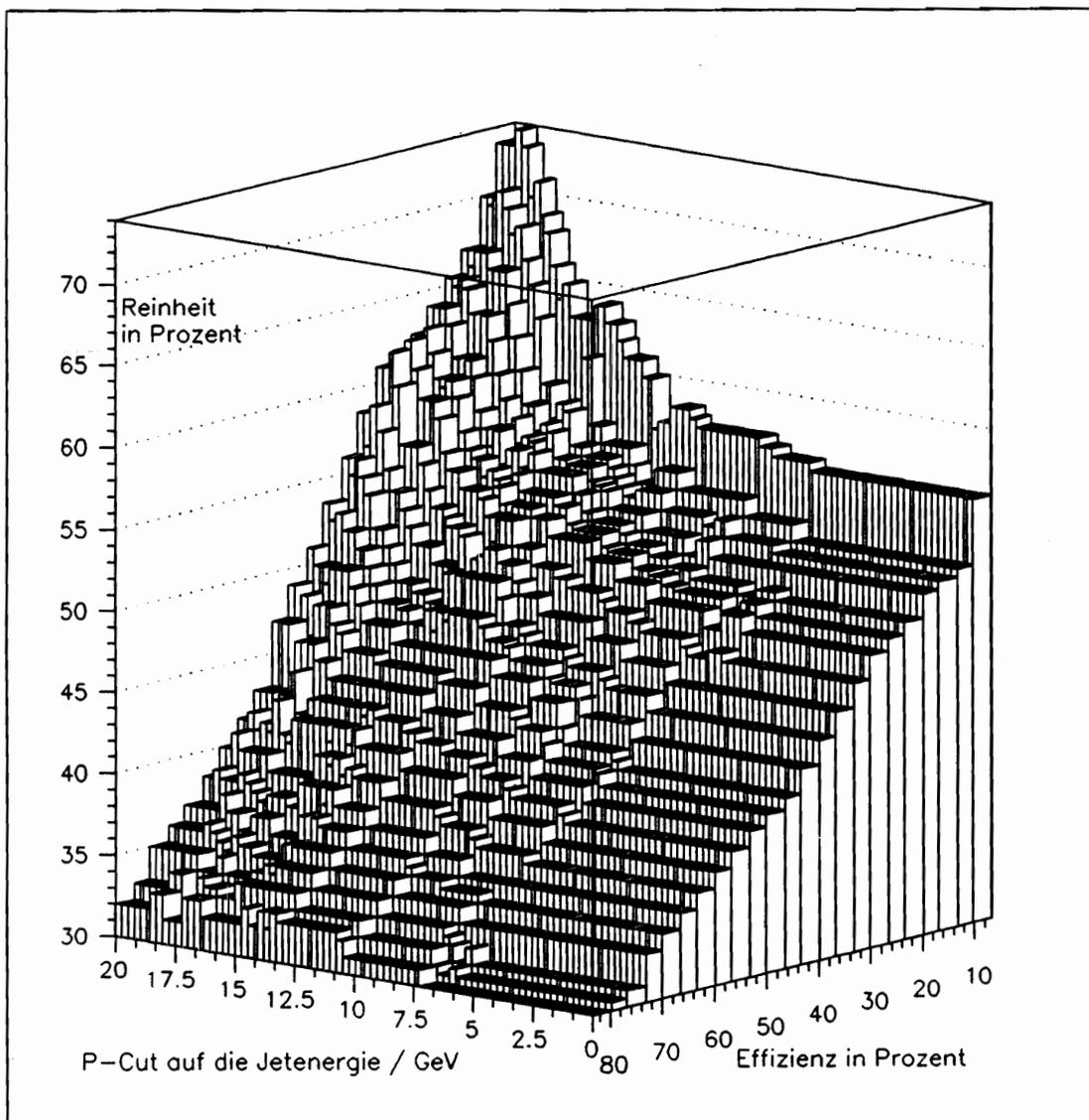


Abbildung 6.2: Die Abhängigkeit der Ergebnisse von einem Schnitt auf die Jetenergie. Es ist die Reinheit über der Effizienz und der Größe des Schnittes auf die Jetenergie aufgetragen.

im PS-Sample geringer sind, ist eine geringere Trenngüte die logische Folge. Die Unterschiede bei Verwendung von 9 oder 23 Eingangsvariablen liegen im Allgemeinen unter 5%. Dies bedeutet, daß beim Sprung von 9 auf 23 Variablen noch ein Informationsgewinn zu verzeichnen ist, wodurch eine bessere Trennung möglich wird. Ergebnisse mit 12 Eingabevariablen, die

hier nicht explizit aufgeführt sind, zeigen in ihrer Separationsfähigkeit kaum Unterschiede zu Netzwerken mit 23 Eingängen. Das heißt die Variablen, die im Schritt von 12 auf 23 Größen dazukommen, enthalten größtenteils redundante bzw. nicht zur Trennung beitragende Information.

Die Ergebnisse, welche mit dem Variablensatz *C1* bzw. *C2* erzielt wurden, sind in Tabelle 6.5 aufgelistet. Mit dem Variablensatz *C1* ergaben sich die gleichen Werte für Effizienz und Reinheit, wie sie Nicolò de Groot und Martin Los in [GR 91] präsentierten. Für den Satz *C2* liegen keine Vergleichszahlen vor. Die Ergebnisse sind jedoch deutlich schlechter im Vergleich zu den Daten aus Jet 1. Das bedeutet, daß die Unterschiede zwischen den beiden Klassen im zweiten Jet der nach der Jetenergie sortierten Jets wie erwartet geringer sind. Verarbeitet man die Information aus Jet 1 und Jet 2 gleichzeitig, so sind die Ergebnisse praktisch die gleichen, wie mit Satz *C1*. Das bedeutet, daß die Information des Variablensatzes *C2* bereits in *C1* enthalten ist. Dies verwundert nicht, da die Trennfähigkeit der Information im Variablensatz *C2* doch äußerst klein ist.

Die Abhängigkeit der Ergebnisse, insbesondere der Reinheitswerte von einem Schnitt auf die Jetenergie ist in Abbildung 6.2 dargestellt. Der Schnitt wurden nicht beim Lernen durchgeführt, sondern beim Testen des Netzwerkes. Für Schnitte bis etwa 7.5 GeV ändert sich am Ergebnis nichts. Für größere Schnitte erhält man bei gleicher Effizienz eine höhere Reinheit. Die Erklärung für dieses Verhalten ist, daß durch den Schnitt auf die Jetenergie die 'flavourblinden' niederenergetischeren Jets im voraus aussortiert werden. Daraus resultiert der Schluß, daß die Unterschiede zwischen *b*- und *nicht-b* Ereignissen in den niederenergetischen Jets nicht so stark ausgeprägt sind. Die Studie der systematischen bzw. statistischen Fehler der Ergebnisse ist Aufgabe einer Nachfolgearbeit.

In Abbildung 6.3 sind zwei Verläufe des Netzwerkfehlers während eines Trainings zu sehen. Es handelt sich um das Netzwerk, dessen Ergebnisse in Tabelle 6.3 ganz oben stehen. Der Netzwerkfehler ist nach Kapitel 3 als Mittelwert, der quadratischen Differenz zwischen dem Soll- und dem Istausgangswert des N.N. über eine Anzahl von Eingabevektoren definiert. In der Theorie sollte diese Mittelung über das gesamte Lernsample gemacht werden. Der so gewonnene Fehler ist jedoch zum Monitorieren des Netzwerkes unbrauchbar, da sich schon während eines Trainingszyklus eine Veränderung der Gewichtungen und damit des Netzwerkes ergibt. In der Praxis wird über eine kleinere Anzahl gemittelt, deren Größe je nach Zweck so

gewählt wird, daß sich feinere bzw. gröbere Effekte während des Lernens be-

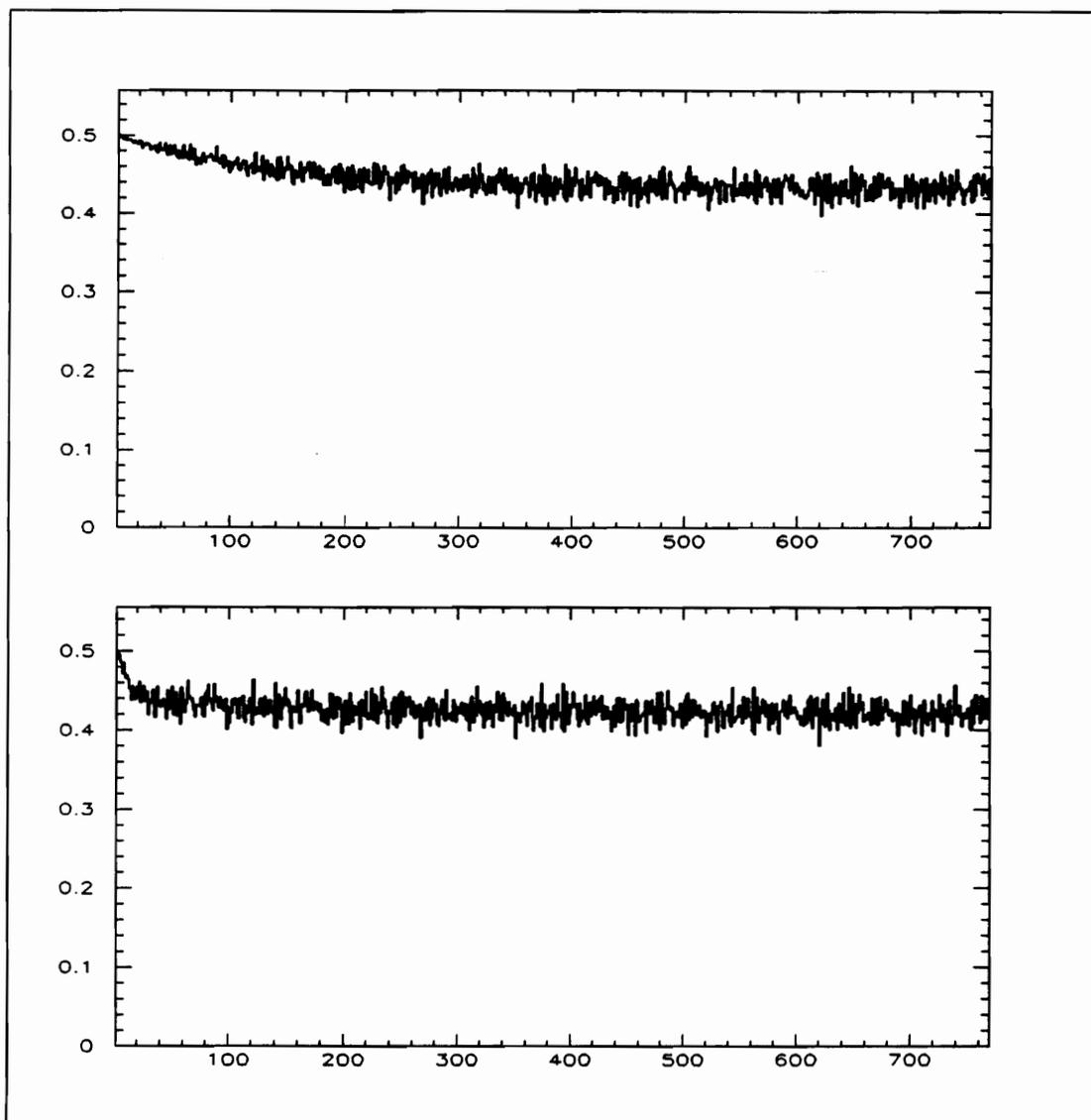


Abbildung 6.3: Verlauf des Fehlers im 'fixed' bzw. im dynamischen Lernmodus, bei ansonsten gleichen Parametern. Erklärung der Kurven im Text!

obachten lassen. In den Kurven der Abbildung 6.3 wurde für jeden Eintrag über 200 Ereignisse gemittelt, das heißt ein Skalenteil der x-Achse entspricht 200 Ereignissen. Beide Trainingsdurchgänge führen zum selben Ergebnis, wobei beim dynamischen Lernen (untere Kurve) eine wesentliche schnell-

lere Konvergenz zu beobachten ist. Bereits nach etwa 60000 verarbeiteten Ereignissen — das entspricht etwa 4 Trainingszyklen — ist beim dynamischen Lernen ein statischer Zustand erreicht, der noch von Fluktuationen einzelner Ereignisse überlagert ist. Beim festen Lernen ist auch nach 140000 Ereignissen noch ein leichtes Absinken der Kurve zu beobachten.

Der CPU-Zeitverbrauch hängt von vielen Faktoren ab (verwendete Hardware, Samplegröße, Variablenanzahl, Knotenanzahl, Netzwerkparameter ...). Deshalb gestaltet sich eine Aussage darüber schwierig. Die Größenordnung der verarbeiteten Ereignisse pro Sekunde ist für folgende Netzwerke :

9/9/1	fixer Lernmodus	~630	Ereignisse/s
23/15/1	fixer Lernmodus	~170	Ereignisse/s
9/9/1	dym. Lernmodus	>1200	Ereignisse/s
23/15/1	dym. Lernmodus	>350	Ereignisse/s .

Die Werte beim dynamischen Lernen liegen im Mittel doppelt so hoch. Dieser Effekt ist darin begründet, daß das Netzwerk beim dynamischen Lernen schneller konvergiert und so weniger Back-Propagation Prozesse durchgeführt werden müssen, womit die Zahl der in derselben Zeit verarbeiteten Ereignisse ansteigt. Die entsprechenden Zahlen im Testmodus sind :

9/9/1	Testmodus	~1600	Ereignisse/s
23/15/1	Testmodus	~570	Ereignisse/s .

Sie sind höher, da hier jedes Ereignis nur durch das Netzwerk hindurchprozesiert und kein Lernprozeß durchgeführt wird. Da diese Werte — wie bereits erwähnt — von vielen Faktoren abhängen, sollte man weniger den absoluten Zahlenwerten, als vielmehr den Zahlenverhältnissen Beachtung schenken. Sie sind an dieser Stelle lediglich angegeben worden, um eine ungefähre Vorstellung des Rechenzeitaufwands zu geben und den großen Einfluß der Netzgröße, sowie der Lernmethode auf die Rechenzeit aufzuzeigen.

Netzwerke, die mit einem zusätzlichen Biasknoten in der Eingangsebene arbeiten, sonst aber identisch mit den bisher verwendeten N.N. sind, zeigten nur mit dem Variablensatz C3 eine deutlich bessere Leistung. Das Netzwerk mit dem zusätzlichen Bias-Knoten zeigt bei gleicher Reinheit eine um 5 bis 15 % höhere Effizienz.

Netzwerk	22/20/1					PS mit Satz C3		
Parameter	$\eta = 0.01 - 0.1$					$\alpha = 0.5$		
Besonderheit	Teilchen d. beiden energiereichsten Jets & 15 GeV Cut & Bias							

Schnitt	.30	.36	.41	.47	.56	.59	.64	.66	.67	.69	.70	.77	.78
Effizienz in %	90	83	72	69	60	58	48	37	28	25	24	19	10
Reinheit in %	26	28	31	32	33	34	35	36	37	39	40	44	48
Mittelwerte	$O_b = 0.56$			$O_{nb} = 0.46$			\Rightarrow		$\Delta(O_b - O_{nb}) = 0.10$				

Tabelle 6.1: Reinheiten und Effizienzen mit dem Variablensatz C3 mit einem Netzwerk mit zusätzlichem Biasknoten.

Wie in Kapitel 3 beschrieben, wurde ein Algorithmus zur Reduktion der Hiddennodeanzahl implementiert. Den Einfluß der Hiddennodeanzahl auf das Ergebnis soll in Tabelle 6.2 anhand von drei Netzwerken aufgezeigt werden. Als Kriterium der Netzwerkeistungsfähigkeit dient dabei die Differenz der Ausgangsverteilungen der beiden Klassen. Zusätzlich ist ein Zeitfaktor angegeben, welcher der zur Verarbeitung des Testsamples benötigten Zeit entspricht. Im ersten Fall wird deutlich, daß die Reduktion der Hiddenkno-

Netzwerk	$\Delta(MW_{b,nb})$	t	Netzwerk	$\Delta(MW_{b,nb})$	t	Netzwerk	$\Delta(MW_{b,nb})$	t
23/15/1	0.18	1	9/9/1	0.15	1	8/8/1	0.17	1
23/10/1	0.18	0.75	9/7/1	0.14	0.83	8/6/1	0.17	0.78
23/4/1	0.11	0.44	9/5/1	0.11	0.66	8/5/1	0.12	0.78
						8/3/1	0.11	0.78

Tabelle 6.2: Ergebnisse in Abhängigkeit von der Hiddennodeanzahl.

ten von 15 auf 10 das Ergebnis nicht beeinträchtigt, jedoch eine Zeitersparnis von 25% erbringt. Eine weitere Verminderung verschlechtert das Ergebnis. Im zweiten und dritten Fall sieht man, daß bei 7 bzw. 6 Hiddenknoten das Ergebnis einer geringen bzw. keiner Verschlechterung unterliegt, wobei der Zeitgewinn $\sim 20\%$ beträgt. Eine weitere Reduktion hat drastische Einbußen in der Leistungsfähigkeit zur Folge und bringt kaum noch einen Zeitgewinn, da das Entfernen eines weiteren Knotens nur relativ wenige Verbindungen und damit Rechenschritte einsparen würde. Es stellt sich heraus, daß die ursprünglich durch 'Erfahrung' gewählten Anzahlen der Hiddenknoten im Allgemeinen zu hoch sind. Durch Anwendung des beschriebenen Algorithmus ist in den meisten Fällen eine Reduktion um 'einige' Knoten bei nahezu gleicher Leistungsfähigkeit möglich.

Kapitel 6. Ergebnisse und Interpretation

Netzwerk	23/15/1 ¹										ME mit Satz A		
Parameter	$\eta = 0.01 - 0.1$					$\alpha = 0.5$							

Schnitt	.24	.31	.38	.45	.48	.55	.59	.62	.66	.69	.72	.76	.79
Effizienz in %	93	88	83	75	71	60	55	49	42	35	28	20	15
Reinheit in %	27	29	31	34	36	40	43	45	47	50	52	55	61
Mittelwerte	$O_b = 0.59$					$O_{nb} = 0.41$					$\Rightarrow \Delta(O_b - O_{nb}) = 0.18$		

Netzwerk	9/9/1										ME mit Satz B		
Parameter	$\eta = 0.01 - 0.1$					$\alpha = 0.5$							

Schnitt	.21	.28	.34	.38	.41	.45	.52	.55	.59	.62	.66	.69	.72
Effizienz in %	96	90	85	80	75	69	55	49	43	35	28	20	14
Reinheit in %	26	27	30	31	33	34	38	41	45	47	50	54	57
Mittelwerte	$O_b = 0.53$					$O_{nb} = 0.39$					$\Rightarrow \Delta(O_b - O_{nb}) = 0.14$		

Netzwerk	23/15/1										PS mit Satz A		
Parameter	$\eta = 0.01 - 0.1$					$\alpha = 0.5$							

Schnitt	.28	.34	.41	.45	.48	.52	.55	.59	.62	.66	.69	.72	.76
Effizienz in %	93	89	82	77	71	67	62	53	44	35	27	21	16
Reinheit in %	26	28	29	30	31	33	35	38	40	43	45	50	53
Mittelwerte	$O_b = 0.58$					$O_{nb} = 0.44$					$\Rightarrow \Delta(O_b - O_{nb}) = 0.14$		

Netzwerk	9/9/1										PS mit Satz B		
Parameter	$\eta = 0.01 - 0.1$					$\alpha = 0.5$							

Schnitt	.31	.34	.38	.41	.45	.48	.52	.55	.59	.62	.66	.69	.72
Effizienz in %	93	91	87	84	78	73	66	58	49	41	31	21	15
Reinheit in %	26	27	28	29	30	32	33	36	38	41	44	48	55
Mittelwerte	$O_b = 0.57$					$O_{nb} = 0.45$					$\Rightarrow \Delta(O_b - O_{nb}) = 0.12$		

Tabelle 6.3: Netzwerk-Ergebnisse. ¹ Gebräuchliche Schreibweise zur Beschreibung des Aufbaus eines Netzwerkes :

Eingangsknoten / Hiddenknoten / Ausgangsknoten

Netzwerk	23/15/1						ME		mit Satz A			
Parameter	$\eta = 0.01 - 0.1$						$\alpha = 0.5$					
Besonderheit	15 GeV Schnitt auf die Jetenergie											

Schnitt	.17	.24	.28	.34	.41	.48	.52	.55	.62	.69	.72	.76	.79
Effizienz in %	96	92	90	85	79	70	65	61	49	36	31	24	17
Reinheit in %	26	28	30	33	36	40	43	46	49	54	58	63	68
Mittelwerte	$O_b = 0.59$			$O_{nb} = 0.38$				\Rightarrow		$\Delta(O_b - O_{nb}) = 0.21$			

Netzwerk	9/9/1						ME		mit Satz B			
Parameter	$\eta = 0.01 - 0.1$						$\alpha = 0.5$					
Besonderheit	15 GeV Schnitt auf die Jetenergie											

Schnitt	.24	.31	.38	.45	.48	.52	.55	.59	.62	.66	.69	.72	.76
Effizienz in %	97	93	87	77	71	65	59	53	44	37	30	21	13
Reinheit in %	26	28	31	34	36	39	42	44	46	50	53	55	57
Mittelwerte	$O_b = 0.58$			$O_{nb} = 0.43$				\Rightarrow		$\Delta(O_b - O_{nb}) = 0.15$			

Netzwerk	23/15/1						PS		mit Satz A			
Parameter	$\eta = 0.01 - 0.1$						$\alpha = 0.5$					
Besonderheit	15 GeV Schnitt auf die Jetenergie											

Schnitt	.24	.31	.34	.41	.48	.55	.59	.62	.66	.69	.72	.76	.79
Effizienz in %	91	88	86	80	73	63	56	48	40	32	25	19	13
Reinheit in %	26	28	29	30	33	36	38	39	41	43	48	51	56
Mittelwerte	$O_b = 0.58$			$O_{nb} = 0.42$				\Rightarrow		$\Delta(O_b - O_{nb}) = 0.16$			

Netzwerk	9/9/1						PS		mit Satz B			
Parameter	$\eta = 0.01 - 0.1$						$\alpha = 0.5$					
Besonderheit	15 GeV Schnitt auf die Jetenergie											

Schnitt	.28	.31	.34	.38	.41	.45	.48	.52	.55	.59	.62	.66	.69
Effizienz in %	88	84	82	79	75	71	65	56	47	39	29	20	11
Reinheit in %	26	26	28	29	31	33	36	37	38	40	44	50	53
Mittelwerte	$O_b = 0.51$			$O_{nb} = 0.39$				\Rightarrow		$\Delta(O_b - O_{nb}) = 0.12$			

Tabelle 6.4: Ergebnisse mit 15 GeV Schnitt auf die Jetenergie.

Kapitel 6. Ergebnisse und Interpretation

Netzwerk	12/6/1						PS mit Satz C1					
Parameter	$\eta = 0.01 - 0.1$						$\alpha = 0.5$					
Besonderheit	Teilchen des energiereichsten Jet & 15 GeV Cut											

Schnitt	.32	.36	.40	.43	.44	.47	.49	.53	.55	.56	.57	.60	.70
Effizienz in %	91	84	78	74	71	65	61	53	48	44	43	36	11
Reinheit in %	26	27	28	30	31	32	33	34	35	36	38	40	40
Mittelwerte	$O_b = 0.52$			$O_{nb} = 0.44$			\Rightarrow			$\Delta(O_b - O_{nb}) = 0.08$			

Netzwerk	12/6/1						PS mit Satz C2					
Parameter	$\eta = 0.01 - 0.1$						$\alpha = 0.5$					
Besonderheit	Teilchen des zweitenergiereichsten Jet & 15 GeV Cut											

Schnitt	.34	.37	.40	.44	.47	.51	.54	.55	.57	.60			
Effizienz in %	92	89	84	73	64	54	45	43	35	25			
Reinheit in %	24	25	26	27	28	29	30	31	32	33			
Mittelwerte	$O_b = 0.51$			$O_{nb} = 0.47$			\Rightarrow			$\Delta(O_b - O_{nb}) = 0.04$			

Netzwerk	22/20/1						PS mit Satz C3					
Parameter	$\eta = 0.01 - 0.1$						$\alpha = 0.5$					
Besonderheit	Teilchen aus den beiden energiereichsten Jets & 15 GeV Cut											

Schnitt	.30	.36	.41	.47	.56	.59	.64	.66	.67	.69	.70	.77	.78
Effizienz in %	88	79	72	61	54	37	27	24	24	21	19	9	8
Reinheit in %	26	27	28	31	32	33	34	35	36	37	38	39	40
Mittelwerte	$O_b = 0.52$			$O_{nb} = 0.44$			\Rightarrow			$\Delta(O_b - O_{nb}) = 0.08$			

Tabelle 6.5: Reinheiten und Effizienzen mit den Variablensätzen C1, C2 und C3.

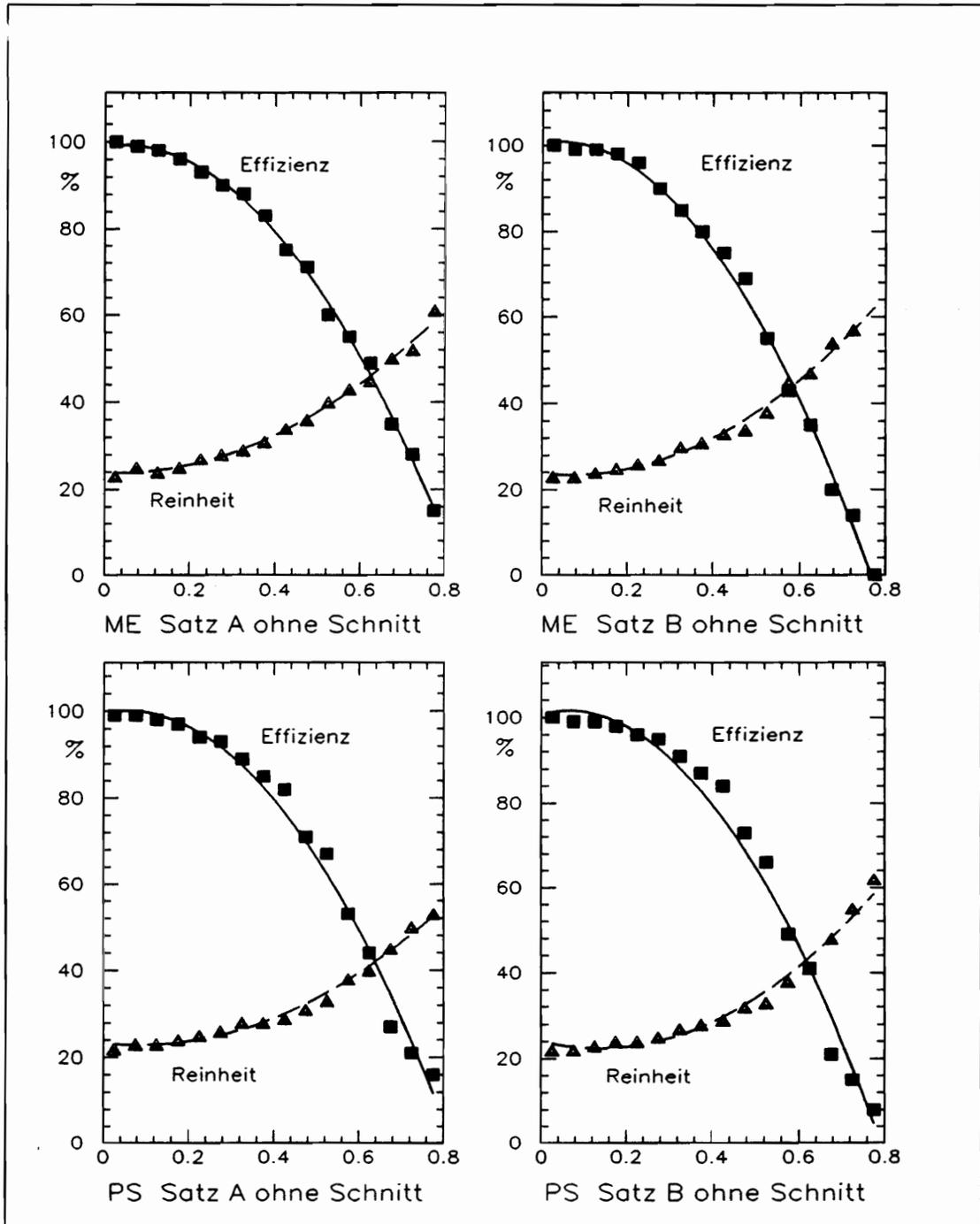


Abbildung 6.4: Plots der Ergebnisse aus Tabelle 6.3. Es ist die Effizienz und die Reinheit über dem Schnitt aufgetragen.

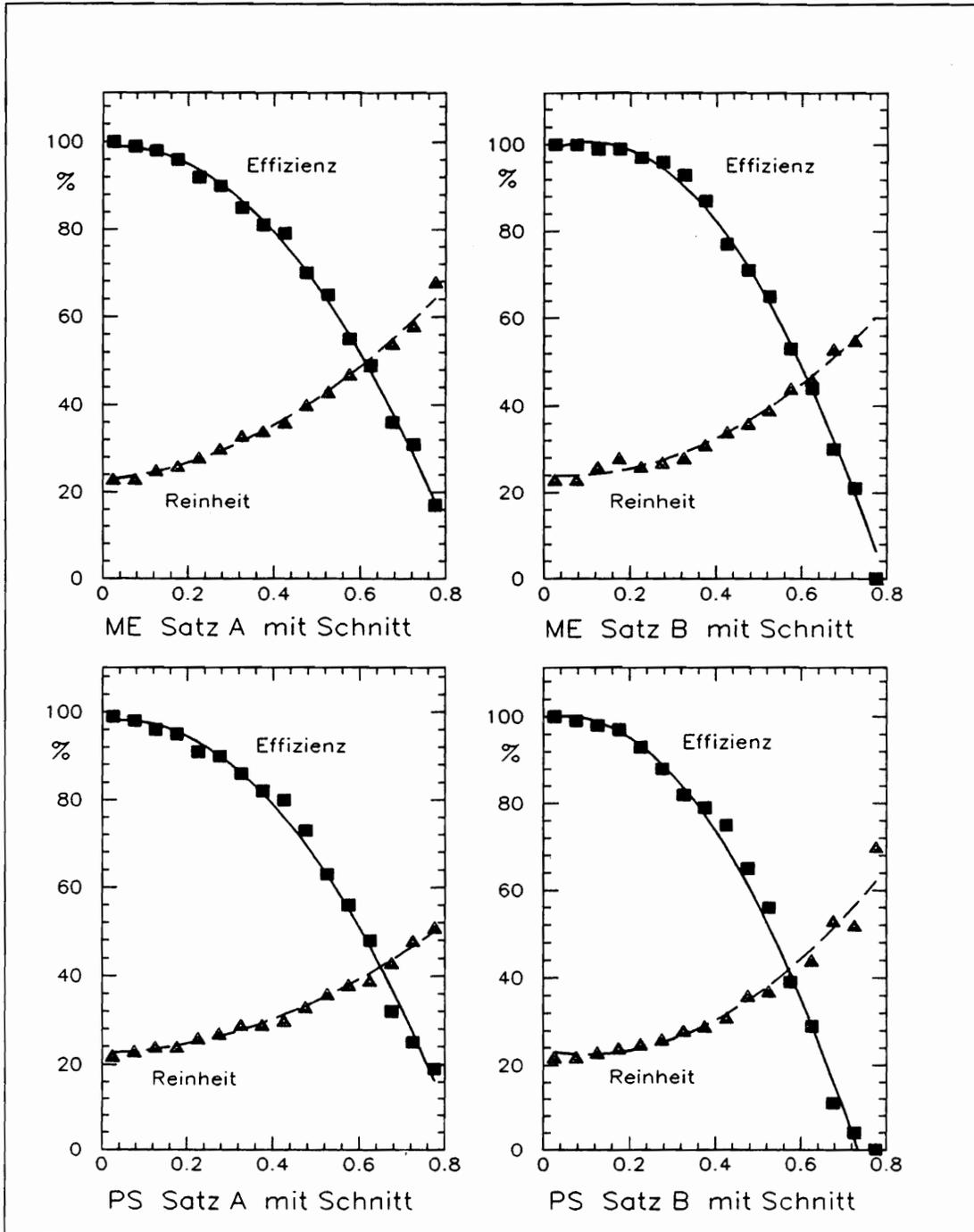


Abbildung 6.5: Plots der Ergebnisse aus Tabelle 6.4. Es ist die Effizienz und die Reinheit über dem Schnitt aufgetragen.

6.1 Zusammenfassung

Es wurde in einer Monte-Carlo-Studie gezeigt, daß es prinzipiell möglich ist, mit neuronalen Netzwerken eine Separation von multihadronischen Z^0 -Zerfällen in b und *nicht- b* Ereignisse vorzunehmen. Die Qualität der Trennung ist neben Netzwerkeigenschaften und Parametern hauptsächlich von der Wahl der Eingangsinformation abhängig.

Wie jedes Trennverfahren können auch die Netzwerke eine Trennung nur dann durchführen, wenn die zu separierenden Klassen sich in den zur Beurteilung zur Verfügung stehenden n Eigenschaften auch 'genügend' unterscheiden. Das bedeutet, daß eine hundertprozentige Trennung grundsätzlich nur dann durchführbar ist, wenn sich die Bereiche der beiden Klassen im n -dimensionalen Eigenschaftsraum nicht überlagern. Wie gut eine tatsächliche Trennung dann wirklich ausfällt, hängt von der Wahl des Trennverfahrens und dessen korrekter Anwendung ab. Überlagern sich die beiden Klassen teilweise, so ist maximal eine Trennung der sich nicht überschneidenden Bereiche möglich. Die von beiden Klassen besetzten Bereiche sind bei der zur Verfügung stehenden Information indifferent. Eine Verbesserung des Ergebnisses ist nur durch zusätzliche, nicht redundante Information möglich. Das heißt es werden weitere Eigenschaften zur Beschreibung der beiden Klassen benötigt, womit der Eigenschaftsraum um zusätzliche Dimensionen erweitert wird.

Die Tatsache, daß das nur zu einer linearen Trennung fähige Fisher-Diskriminanten Verfahren und die neuronalen Netzwerke, die aufgrund ihrer Struktur in der Lage sind beliebig komplexe Separationsfunktionen zu modellieren, die gleichen, weit von einer perfekten Trennung entfernten Ergebnisse liefern, zwingt zu folgendem Schluß. Die beiden Klassen sind durch die bearbeitete Eingangsinformation nicht exakt zu trennen, das heißt sie zeigen im Eigenschaftsraum eine große Überlagerung. Die maximal mögliche Trennung ist durch einen linearen Schnitt möglich und damit durch Netzwerke mit nur einer Hiddennodeebene durchführbar.

Erreichbare Werte sind 85% Effizienz bei einer Reinheit von 30% und 50% Reinheit bei 35% Effizienz. Dies entspricht einer Anreicherung um mehr als 100 % in der ursprünglich zu 21 % b -Ereignisse enthaltenden Menge.

Durch die Aussonderung von niederenergetischen Jets unter 15 GeV ist eine höhere Reinheit erreichbar, da die Eigenschaften dieser Jets weit weniger

Kapitel 6. Ergebnisse und Interpretation

eindeutig sind als die energiereicherer Jets. Die Ergebnisse liegen im Bereich von 65%(50%) Reinheit bei 20%(50%) Effizienz.

Dies ist die bestmögliche Trennung, die mit den hier verwendeten topologischen Größen erreichbar ist. Eine neue 'Qualität', das heißt zusätzliche Information über die beiden zu trennenden Klassen, ist erforderlich um eine wesentliche Steigerung dieser Werte zu erreichen. Vielversprechend sind in diesem Zusammenhang vor allem die Vertexinformation und der semileptonische Zerfall. Beide Methoden sind eigenständig oder in Kombination mit Netzwerken vorstellbar. Besonders die Vertexinformation kann ohne großen Aufwand als zusätzliche Eingangsinformation für ein neuronales Netzwerk verwendet werden, wovon eine deutliche Ergebnisverbesserung zu erwarten ist.

Anhang A

Literaturhinweise

Einführende, vertiefende Literatur sowie Fachaufsätze zu den Bereichen :

- Elementarteilchenphysik :
 - [EB 89], [NA 86], [PE 82].
- Neurobiologie:
 - [EC 87], [VE 91], [WO 76].
- Neuronale Netzwerke:
 - Grundlagen: [DE 86], [DO 90], [JO 87], [KR 90], [RU 86].
 - N.N. in der Hochenergiephysik : [DE 90/3], [DE 90/4], [DE 90/5], [DE 90/6], [GO 90], [PE 89].

Abbildungsverzeichnis

2.1	Z^0 -Graph	10
2.2	Die Fragmentationsfunktionen für die verschiedenen Quark-flavour	12
2.3	Der universelle W-Zerfall	15
2.4	Geometrie des B-Zerfalls	16
3.1	Das vereinfachte Schema einer Nervenzelle.	19
3.2	Das neuronale Netzwerk als 'black box'- Schema.	21
3.3	Hardwarekomponenten eines künstlichen neuronalen Netzwerkes	22
3.4	Sigmoid-Funktionen für verschiedene T-Werte.	24
3.5	Funktionsweise eines Netzwerkknotens	26
3.6	Die drei Phasen im Umgang mit neuronalen Netzwerken. . .	27
3.7	Vorwärtsvermittlungsnetz	28
3.8	Normierungsschema für die Eingabevariablen	30
3.9	Beispiel zur Normierung der Eingabevariablen	31
3.10	Das Oszillieren eines Netzwerkes über einem Minimum. . . .	40
3.11	Flußdiagramm des zweigeteilten Algorithmus zur Variation der Hiddenodeanzahl	42

3.12	Der ideale Output eines Neuronalen Netzwerkes.	44
4.1	Perspektivische Aufrißzeichnung des DELPHI Detektors . .	47
4.2	Längsschnitt durch den DELPHI Detektor	48
4.3	Schnittbild eines Quadranten des DELPHI Detektor	49
4.4	Anordnung der Module des Outer Detector	51
4.5	Der Aufbau eines HPC-Modules	53
4.6	Ein einzelner TOF-Zähler	54
4.7	Die Architektur des Auslesesystems	56
4.8	Die DELPHI-Datenrekonstruktion	60
5.1	Sphericity, BSP, Aplanarity, NCHG	71
5.2	EJET 1, EJET 2, PT J1/P1, PL J1/P1	72
5.3	NTC, PTNTC, ECHAG, MISP	73
6.1	Der Output eines neuronalen Netzwerkes in der Realität. . .	76
6.2	Abhängigkeit der Ergebnisse von einem Schnitt auf die Jet- energie.	77
6.3	Fehlerplot für dynamisches und 'fixed' Lernen.	79
6.4	Plots der Ergebnisse aus Tabelle 6.3.	85
6.5	Plots der Ergebnisse aus Tabelle 6.4.	86

Tabellenverzeichnis

1.1	Die vier fundamentalen Wechselwirkungen	2
1.2	Das Standardmodell der Elementarteilchenphysik	5
4.1	Die Aufgabenverteilung der Detektorkomponenten bei DEL- PHI.	46
4.2	Akzeptanz, Spur- und Energieauflösung der einzelnen Detek- torkomponenten.	62
5.1	Die benutzten Monte Carlo Datenmengen	64
6.1	Reinheiten und Effizienzen mit dem Variablensatz <i>C3</i> mit einem Netzwerk mit zusätzlichem Biasknoten	81
6.2	Ergebnisse in Abhängigkeit von der Hiddennodeanzahl. . .	81
6.3	Netzwerk-Ergebnisse für verschiedene MC- Daten	82
6.4	Reinheiten und Effizienzen mit 15 GeV Schnitt auf die Jet- energie	83
6.5	Reinheiten und Effizienzen mit den Variablensätzen <i>C1</i> , <i>C2</i> und <i>C3</i>	84

-
- [GO 90] Gottschalk, T.D. und Nolty, R., *Identification Of Physics Processes Using Neural Network Classifiers*
California Institute of Technology, (1990)
- [GR 89] Grotz, Klapdor, Volker, *Die schwache Wechselwirkung in Kern-, Teilchen- und Astrophysik*
B.G. Teubner , Stuttgart (1989)
- [GR 91] Groot de Nicolo, Los Martin, *b-Tagging with a Neural Network*
Vortrag am 25.4.91 im Neural-Network-meeting, CERN (1991)
- [HI 91] Hirose, Y. *Back-Propagation Algorithm Wich Varies the Number of Hidden Units*
Neural Networks, Vol.4, pp. 61-66, 1991
- [JO 87] Jones, P. *Back-Propagation*
Byte, October 1987
- [KR 90] Kratzer, Klaus Peter., *NEURONALE NETZE*
Carl Hanser Verlag , München (1990)
- [KU 89] Kühn, J.H., Zerwas, P.M., *Heavy Flavours at LEP*
MPI-PAE/PTh 49/89, München (August 1989)
- [LA 90] Lackner, Klaus S. u.a., *Data Processing at the SSC with Structured Neural Nets*
Los Alamos N. Lab., LA-UR -90-3774, Los Alamos (1990)
- [LO 83] Lohrmann, E., *Einführung in die Elementarteilchenphysik*
Teubner Verlag, Stuttgart (1983)
- [LO 90] Lönnblad, Leif, *Finding Gluon Jets with a Neural Trigger*
LU TP 90-3, Lund (1990)
- [NA 86] Nachtmann, O., *Elementarteilchenphysik, Phänomene und Konzepte*
Vieweg Verlag, Braunschweig (1986)
- [PD 90] Particle Data Group, *Review of Particle Properties*
Physics Letters **B239** (April 1990)
- [PE 82] Perkins, D. H., *Introduction to High Energy Physics*
2nd Edition, Addison-Wesley Publishing Company, (1982)
-

- [PE 89] Peterson, Carsten, *Track Finding with Neural Networks*
Nuclear Instruments and Methods in Physics Research A279
(1989) 537-545, North Holland, Amsterdam
- [RU 86] Rumelhardt, David E., *Parallel Distributed Processing*
Eighth printing, The Massachusetts Institute of Technology,
(1986)
- [SCH 91] Scheu, S., *Diplomarbeit in Vorbereitung.*
Inst. für Experimentelle Kernphysik, Karlsruhe (1991)
- [SJ 89] Sjöstrand, T., *THE LUND MONTE CARLO PROGRAMS*
CERN, (1.11.1989)
- [TH 85] Thorndike, H., *Production and decay of the b quark*
Nucl. Phys. Sci. 1985.35:195-243, (1985)
- [VE 91] Vester Frederic, *Denken, Lernen, Vergessen*
dtv 1327, (1991)
- [WO 76] Wolf, Gerald, *Neurobiologie*
Gusatv Fischer Verlag, Stuttgart (1976)
- [WU 91] Wu, Sau Lan; Bellantoni, L., *Using Neural Networks with Jet
Shapes to Identify b Jets in e^+e^- Interactions*
CERN-PPE/91-80 (24.5.1991)

Danksagung

Ich möchte der Karlsruher DELPHI-Gruppe für die freundliche Aufnahme, die ständige Hilfsbereitschaft, sowie für die sehr angenehme Arbeitsatmosphäre auch über das Institut hinaus danken.

Mein spezieller Dank gilt :

- Herrn Prof. Dr. Wim de Boer für die interessante Aufgabenstellung.
- Herrn Dr. W.-D. Apel für die Betreuung. Er hatte beständig ein offenes Ohr bei Fragen zur Physik und zu Computerproblemen und hat durch seine Anregungen und kritischen Fragen sehr viel zum Gelingen dieser Arbeit beigetragen.
- Steffi für die angenehme Zusammenarbeit.
- Jobst, der durch viele interessante Gespräche den Fortgang der Arbeit begleitete.
- Herrn Prof. Dr. D.C. Fries, für das Korreferat.
- Herrn Prof. Dr. K.R. Schubert, dem Leiter des Instituts für experimentelle Kernphysik, für die Aufnahme in seinem Institut.

Schließen möchte ich mit einem besonders herzlichen Dank an meine liebe Mutter, die mir das Studium ermöglichte und mich beständig unterstützt hat.

Versicherung

**Ich versichere, diese Arbeit selbstständig,
unter Angabe aller wesentlichen Hilfsmittel und
Referenzen, angefertigt zu haben.**

Martin Büchert