

Experimentelle Physik

Anwendung eines neuronalen Netzes zur Teilchenklassifizierung in einem Bleiglaskalorimeter

Diplomarbeit

von
Stephan Joerrens

Westfälische Wilhelms-Universität Münster
Institut für Kernphysik

März 1998

Inhaltsverzeichnis

1 Einleitung	1
1.1 Motivation	2
2 Physikalische Grundlagen	3
2.1 Aufbau der Materie	3
2.1.1 Hadronen	3
2.1.2 Leptonen	4
2.2 Quarks und Gluonen	5
2.3 Quark-Gluon-Plasma	6
2.3.1 Ultrarelativistische Schwerionenreaktionen	8
2.3.2 Nachweismöglichkeiten eines Quark-Gluon-Plasmas . .	10
3 WA98-Experiment	13
3.1 Aufbau des WA98-Experimentes	13
3.2 Bleiglas-Detektor LEDA	15
3.2.1 Physikalische Prozesse	16
3.2.2 Bearbeitung der Daten	18
4 Neuronale Netze	23
4.1 Biologisches Analogon	23
4.2 Architektur künstlicher neuronaler Netze	24
4.2.1 Einschichtperzeptron	26
4.2.2 Multilayer-Netzwerke	27

5 Lernverfahren	31
5.1 Lern-Algorithmen	33
5.1.1 Backpropagation-Algorithmus	33
5.1.2 <i>Rprop</i> -Algorithmus	36
5.2 Lernabbruch	39
6 Netzwerk Training	43
6.1 Begriffe zur Wahrscheinlichkeitsinterpretation	43
6.1.1 A-priori-Wahrscheinlichkeit	44
6.1.2 Likelihood	44
6.1.3 A-posteriori-Wahrscheinlichkeit	44
6.2 Netzwerktypen	45
6.3 Training der beiden Netzwerke	47
6.3.1 Netzwerk mit 28 Eingabeneuronen	48
6.3.2 Netzwerk mit 3 Eingabeneuronen	55
7 Teilchenklassifikation	57
7.1 Dispersionsmethode	57
7.2 Vergleich beider Methoden an simulierten Daten	61
7.3 Vergleich beider Methoden mit WA98 Daten	66
8 Zusammenfassung	71
Anhang	73
A.1 Rapidität	73
A.2 Gewichte des neuronalen Netzes	74

1 Einleitung

Bei den Erfindungen des Menschen diente die Natur schon immer als Vorbild. Die Funktionen des menschlichen Körpers werden schon lange mit Hilfe künstlicher Geräte nachgebaut. Angefangen bei einfachen Gegenständen, wie einer Holzprothese, bis zu komplexen Bewegungsabläufen, wie dem Gang auf zwei Beinen. In der Barockzeit wurden Puppen gebaut, die sehr menschlich aussahen und in einem relativ begrenzten Rahmen schreiben konnten. Das Ziel dieser Nachbauten ist, die Natur besser zu verstehen und kopieren zu können. Heutzutage werden in Fabriken Industrieroboter eingesetzt, um schwere, gefährliche oder monotone Arbeiten auszuführen. Ein *Gerät* aus der Natur wurde bis jetzt noch nicht kopiert, das Gehirn. Hier setzt die Forschung der KI (künstliche Intelligenz, im Englischen: AI, **A**rtificial **I**ntelligence) an. Bei der KI-Forschung werden verschiedene Wege untersucht, die Eigenschaften des Gehirns für Maschinen nutzbar zu machen. Einige dieser Fähigkeiten sind:

- Große Ausfalltoleranz, d.h. es können Zellen absterben oder ausfallen, ohne daß die Funktionsweise gestört oder beeinträchtigt wird.
- Verarbeitung fehlerhafter Daten.
- Schnelligkeit.
- Geringes Volumen.
- Lernfähigkeit, d.h. bei neuen Problemen werden selbständig Lösungen gefunden

Ein Gebiet in der KI sind die ANN (engl. **A**rtificial **N**eural **N**etwork, künstliches neuronales Netz, kurz NN). Besonders in dieser Richtung wurde die Forschung vorangetrieben. Mittlerweile existieren einige kommerzielle Anwendungen z.B. OCR (*optical character recognition*) oder für Prognosen an der Börse oder beim Wetter. Mit Hilfe neuronaler Netze kann man mit relativ geringem Aufwand sehr komplexe Aufgaben lösen, ohne sich allzuviiele Gedanken über den Lösungsweg zu machen, denn neuronale Netze können selbständig eine Problemlösung ausarbeiten.

1.1 Motivation

Bei dem **WA98**-Experiment am CERN¹ wurden am Super-Proton-Synchroton (SPS) ultrarelativistische Schwerionenreaktionen bei Energien von ca. 158 GeV pro Nukleon untersucht. Neben verschiedenen Tracking-Detektoren und Veto-Detektoren für geladene Teilchen wurde ein Bleiglaskalorimeter (LEDA²) aus 10080 einzelnen Bleiglasmodulen verwendet. Mit Hilfe des Detektors wurden Photonen in einem Energiebereich von 100 MeV bis 40 GeV gemessen. Diese Diplomarbeit befaßt sich mit einer alternativen Methode zur Klassifizierung der Signale in elektromagnetische oder hadronische Teilchen durch ein neuronales Netz. Eine gute Unterscheidung zwischen diesen beiden Klassen ist notwendig, um das γ/π^0 -Verhältnis möglichst genau zu bestimmen. In dem Bleiglas-Detektor erzeugen die eintreffenden Teilchen ein typisches Muster der Energieverteilung. Bei der Entwicklung dieser Arbeit wurde an die Fähigkeit der Mustererkennung von neuronalen Netzen gedacht.

Die schon bestehende Methode zur Unterscheidung dieser Signale soll zusammen mit der hier dargestellten Methode Aufschlüsse über mögliche systematische Fehler liefern. Außerdem erhofft man die Klassifikation des unteren Energiebereiches zu verbessern.

¹Conseil Européen de la Recherche Nucléaire

²LEadglass Detektor Array

2 Physikalische Grundlagen

2.1 Aufbau der Materie

Im Laufe unseres Jahrhunderts war die Entwicklung in der Kernphysik rasant. Am Anfang dieses Jahrhunderts waren nur die Bestandteile der Atome bekannt, das positiv geladene Proton, das negativ geladene Elektron und einige Jahre später das elektrisch neutrale Neutron (siehe Abbildung 2.1). Durch die Weiterentwicklung der Nachweismethoden ist die Zahl der bekannten Teilchen schnell auf mehr als 300 verschiedene Teilchen gestiegen. Um diese Vielfalt der Teilchen (*Teilchen-Zoo*) zu erklären, nahm man an, daß diese Teilchen aus noch kleineren Bausteinen aufgebaut sind. Man unterteilt die Materie heute in Hadronen (griech. hadros, stark) und Leptonen (griech. leptos, leicht).

2.1.1 Hadronen

Die Baryonen (griech. barys, schwer) und Mesonen (griech. meson, mittel) bilden zusammen die Gruppe der Hadronen. Die Hadronen zeichnen sich dadurch aus, daß sie neben der *schwachen* und elektromagnetischen Wechselwirkung, alleine an der *starken Wechselwirkung* teilnehmen. Murray Gell-Mann gelang es, durch Eichsymmetrien und der Quarktheorie, die Hadronen in Symmetriegruppen einzuteilen. Durch spätere Entdeckungen wurden die Vorhersagen aus dieser Theorie, speziell die Existenz des Ω -Teilchens, welches aus drei Strange-Quarks besteht, hervorragend bestätigt.

Baryonen

Zu den Baryonen gehören u.a. das Proton (Zeichen: p) und das Neutron (Zeichen: n). Die Vielfalt der Baryonen, führte unter anderem zur Quarktheorie. Mit dieser Theorie lassen sich viele Eigenschaften der Teilchen erklären. Ein Baryon besteht aus drei Quarks, die alle einen halbzahligen Spin¹ besitzen.

¹Spin: Der Spin ist der Drehimpuls eines Systems. Dieser liegt in gequantelten Einheiten von $\hbar = 6,5821220 \cdot 10^{-22}$ [MeVs] vor. Der Spin ist entweder halbzahlig (. . . , -3/2, -1/2, +1/2, +3/2, . . .) oder ganzzahlig (. . . , -2, -1, 0, +1, +2, . . .).

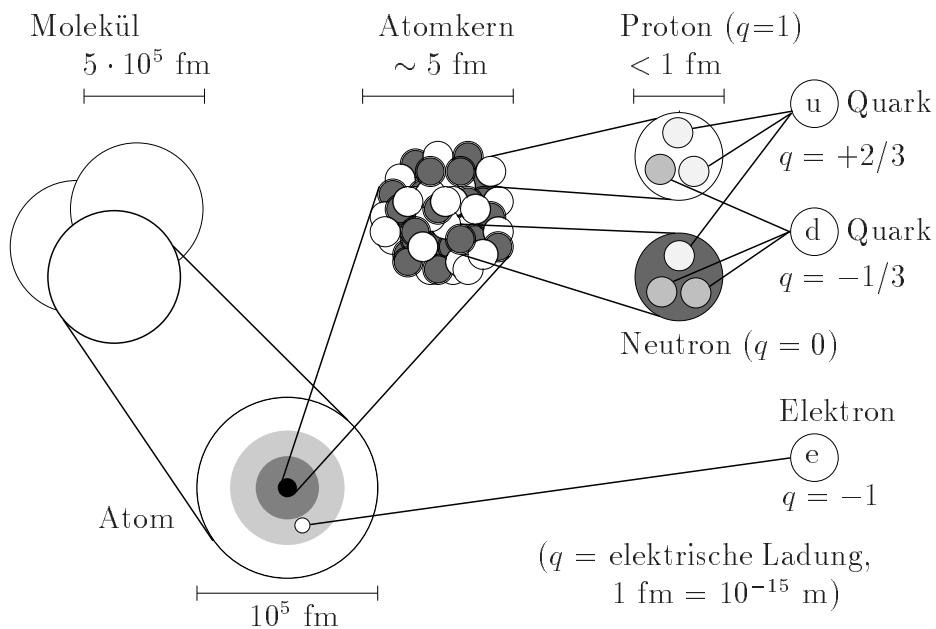


Abbildung 2.1: Schematischer Aufbau der Materie.

Daraus ergibt sich ebenfalls ein halbzahligiger Spin für die Baryonen. Die Quarks werden durch die Gluonen, die Trägerteilchen der *starken Kraft*, im Baryon gehalten.

Mesonen

Mesonen unterliegen ebenfalls der *starken Wechselwirkung*, haben aber einen ganzzahligen Spin. Sie bestehen aus einem Quark und einem Antiquark. Wie bei den Baryonen nimmt man an, daß beliebig viele Typen existieren, die sich durch Spin und Masse unterscheiden. Das leichteste Meson ist das Pion (π^\pm, π^0).

2.1.2 Leptonen

Die Leptonen unterliegen nicht der *starken Wechselwirkung*, weil die Leptonen nicht aus Quarks aufgebaut sind. Von den Leptonen und Quarks gibt es jeweils sechs verschiedene Elementarteilchen, die in drei Familien eingeteilt werden. Man nimmt an, daß die Leptonen wirkliche Elementarteilchen sind und nicht aus noch kleineren Teilchen aufgebaut sind.

$$\text{Quarks: } \begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}$$

$$\text{Leptonen: } \begin{pmatrix} e^- \\ \nu_e \end{pmatrix} \quad \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix} \quad \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}$$

2.2 Quarks und Gluonen

Die gesamte Materie ist aus Leptonen und Quarks aufgebaut. Die Quarks besitzen Spin 1/2 und gehören zu der Gruppe der Fermionen¹. Auf diese Fermionen wirken die verschiedenen Wechselwirkungen, die über Bosonen² (genauer Vektorbosonen) mit Spin 1 übermittelt werden.

Das Photon ist das Eichboson der Quantenelektrodynamik (QED), und es vermittelt die Coulomb-Wechselwirkung zwischen geladenen Teilchen. Die Eichbosonen der *schwachen Wechselwirkung* sind das W^\pm - und das Z^0 -Boson. Mit ihnen werden zum Beispiel die Reaktionen zwischen Elektronen und Neutrinos mit und ohne Ladungsaustausch (geladene und ungeladene Ströme) beschrieben. Das Gluon (engl. = *glue*, Leim) ist das Eichboson der *starken Wechselwirkung*.

In Analogie zur QED wurde die Quantenchromodynamik (QCD) entwickelt. Sind es bei der QED die elektrischen Ladungen die eine Wechselwirkung hervorrufen, so sind es in der QCD die *Farbladungen* der Quarks. Dabei ist der Begriff der *Farbe* nur als anschauliche Beschreibung zu verstehen. Im Gegensatz zur QED, tragen die Austauschteilchen der *starken Wechselwirkung* selbst eine *Farbladung* und können deshalb untereinander wechselwirken. In der QED können die elektrisch neutralen Photonen nicht untereinander wechselwirken. Alle natürlichen Teilchen zeigen nach außen keine *starke Wechselwirkung*. Diese Beobachtung wird als *Farbneutralität* bezeichnet. Nimmt man die drei Grundfarben Rot, Grün und Blau, so lässt sich die *Farbneutralität* durch Kombination von allen drei Farben erzielen, denn alle drei Grundfarben zusammen

¹Fermionen werden alle Teilchen mit ungeraden Spin bezeichnet. Sie unterliegen der Fermi-Dirac-Statistik, d.h., es kann immer nur ein Teilchen in einem bestimmten durch die Quantenzahlen beschriebenen Zustand existieren.

²Als Bosonen werden Teilchen mit ganzzahligen Spin bezeichnet. Sie unterliegen der Bose-Einstein-Statistik, d.h., es können sich beliebig viele Teilchen im selben Zustand befinden.

ergeben *weiß*. Bei den Mesonen wird die *Farbneutralität* erreicht, indem die *Farbe* des einen Quarks mit der passenden *Antifarbe* des Antiquarks kombiniert wird.

Die Potentiale, mit denen die QED und QCD beschrieben werden, sehen folgendermaßen aus:

$$\begin{aligned} V_{QED}(r) &= -\frac{\alpha}{r}, \\ V_{QCD}(r) &= -\frac{4}{3} \cdot \frac{\alpha_s}{r} + kr. \end{aligned}$$

Die Kopplungskonstante α ist in der QED die Feinstrukturkonstante $\alpha = 1/137$. Das Potential nimmt mit dem Abstand der wechselwirkenden Teilchen ab. In der QCD ist die Kopplungskonstante α_s nicht konstant, denn sie hängt vom Impulsübertrag q^2 ab:

$$\alpha_s(q^2) = \frac{12\pi}{(33 - 2N_f) \ln(\frac{q^2}{\lambda^2})},$$

(N_f : Anzahl der Quarksorten, $\lambda \approx 100$ MeV: Skalenfaktor der QCD).

Je größer der Impulsübertrag q^2 , d.h. je kleiner der Abstand wird, desto kleiner wird α_s (Asymptotische Freiheit). Bei kleinem q^2 , also großem Abstand, nimmt die Kraft zu ($V \sim kr$), die Teilchen werden stärker gebunden (engl. *confinement*, Einschluß). Wenn der Abstand groß genug wird, werden neue Quarks und Antiquarks aus der Potentielennergie erzeugt. Die erzeugten Quark-Anti-Quarkpaare verbinden sich sofort zu Mesonen. Es wurde noch kein einzelnes *freies* Quark beobachtet. In Tabelle 2.1 findet man eine Übersicht über die bekannten Elementarteilchen und Eichbosonen.

2.3 Quark-Gluon-Plasma

Mit Hilfe der Gittertheorie, bei der die quantenmechanischen Zustände in einem diskreten Raum-Zeit-Gitter berechnet werden, lassen sich Aussagen über Prozesse mit großen Kopplungskonstanten machen, die störungstheoretisch nicht mehr berechenbar sind.

Bei sehr hoher Temperatur oder hoher Kompression wird, nach Gitterrechnungen, ein Phasenübergang der hadronischen Materie in einen neuen

Quarks (Spin $\frac{1}{2}$)			
Name	Symbol	Masse	Ladung Q [e]
Up	u	5 MeV/c ²	+ $\frac{2}{3}$
Down	d	10 MeV/c ²	- $\frac{1}{3}$
Charme	c	1,5 GeV/c ²	+ $\frac{2}{3}$
Strange	s	100 MeV/c ²	- $\frac{1}{3}$
Top/Truth	t	180 GeV/c ²	+ $\frac{2}{3}$
Bottom/Beauty	b	4,7 GeV/c ²	- $\frac{1}{3}$

Leptonen (Spin $\frac{1}{2}$)			
Name	Symbol	Masse	Ladung Q [e]
Elektron	e ⁻	0,511 MeV/c ²	-1
Elektron-Neutrino	ν_e	< 7,3 MeV/c ²	0
Myon	μ^-	105,7 MeV/c ²	-1
Myon-Neutrino	ν_μ	< 0,27 MeV/c ²	0
Tau	τ^-	1,784 GeV/c ²	-1
Tau-Neutrino	ν_τ	< 35 MeV/c ²	0

Eichbosonen (Spin 1)			
Name	Masse	Ladung Q [e]	Naturkraft
Photon (γ)	0	0	el.-magnetische Kraft
W ⁻	83 GeV/c ²	-1	schwache Kraft
W ⁺	83 GeV/c ²	+1	schwache Kraft
Z ⁰	93 GeV/c ²	0	schwache Kraft
Gluon (g)	0	0 Farbe	starke Kraft
Graviton	0 (?)	0	Gravitation

Tabelle 2.1: Übersicht über Quarks, Leptonen und Eichbosonen.

Aggregat-Zustand der als Quark-Gluon-Plasma (QGP) bezeichnet wird, erwartet. Das Quark-Gluon-Plasma zeichnet sich dadurch aus, daß sich Quarks und Gluonen innerhalb des Plasmas frei bewegen können. Die Bedingungen für einen solchen Zustand könnten beispielsweise kurz nach dem Urknall existiert haben, oder im inneren von Neutronensternen existieren.

Das Phasendiagramm in Abbildung 2.2 zeigt verschiedene Szenarien bei denen es zur Ausbildung eines Quark-Gluon-Plasmas kommen könnte. Bei geringer Baryonendichte, wie sie beim Urknall vorherrschte, wird die kritische Tem-

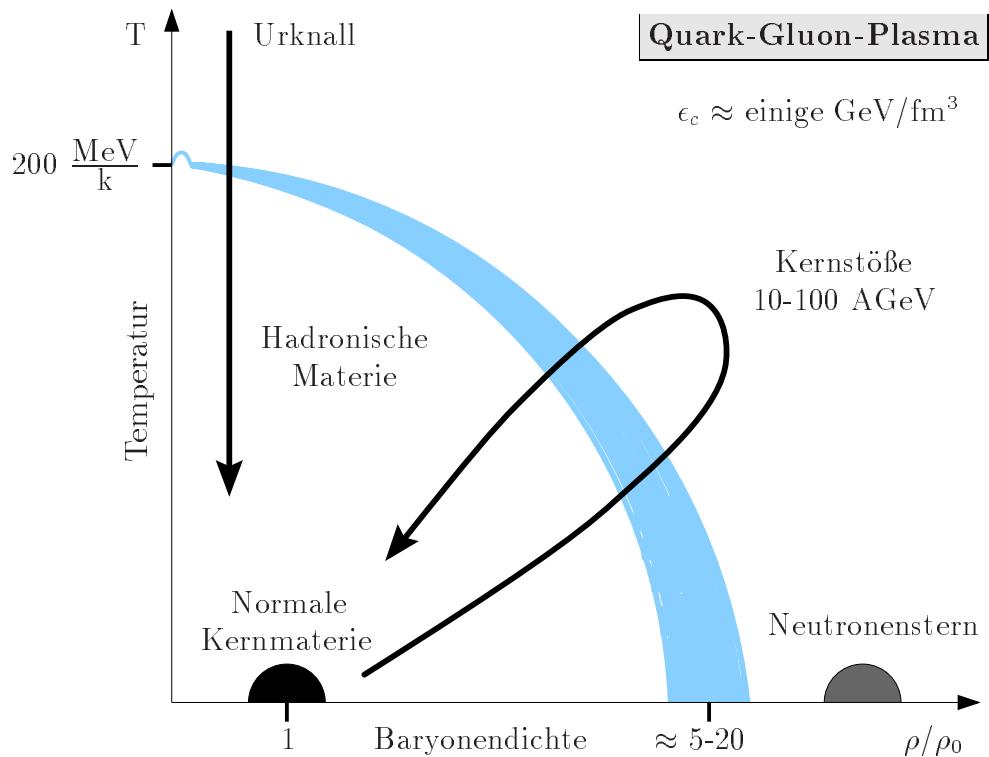


Abbildung 2.2: Phasendiagramm für den Übergang in ein Quark-Gluon-Plasma. Die Baryonendichte ist auf die normale Dichte der Kernmaterie, $\rho_0 = 0,17 \text{ fm}^{-3}$, normiert.

peratur bei $T_c = 150 - 200 \text{ MeV}/\text{k}$ erwartet. Umgekehrt wird bei geringer Temperatur eine kritische Dichte von $\rho_c = 2 - 20\rho_0$ vermutet. Solche Bedingungen können in einem Neutronenstern vorliegen. Aus diesen Werten lässt sich eine kritische Energiedichte von $\epsilon_c \approx 1 - 5 \text{ GeV}/\text{fm}^{-3}$ ableiten.

2.3.1 Ultrarelativistische Schwerionenreaktionen

Experimentell versucht man den Phasenübergang von normaler Materie zu einem Quark-Gluon-Plasma bei ultrarelativistischen Schwerionenreaktionen zu beobachten. Die Schwerpunktsenergien bei solchen Reaktionen liegen bei mehr als 10 GeV pro Nukleon. Bei solchen Energien können die Stoßpartner als unabhängige Teilchen betrachtet werden, da ihre Compton-Wellenlänge klein ist im Vergleich zu den Kernabmessungen. Um die Wahrscheinlichkeit der Bildung eines Quark-Gluon-Plasmas zu erhöhen, muß die Materie komprimiert

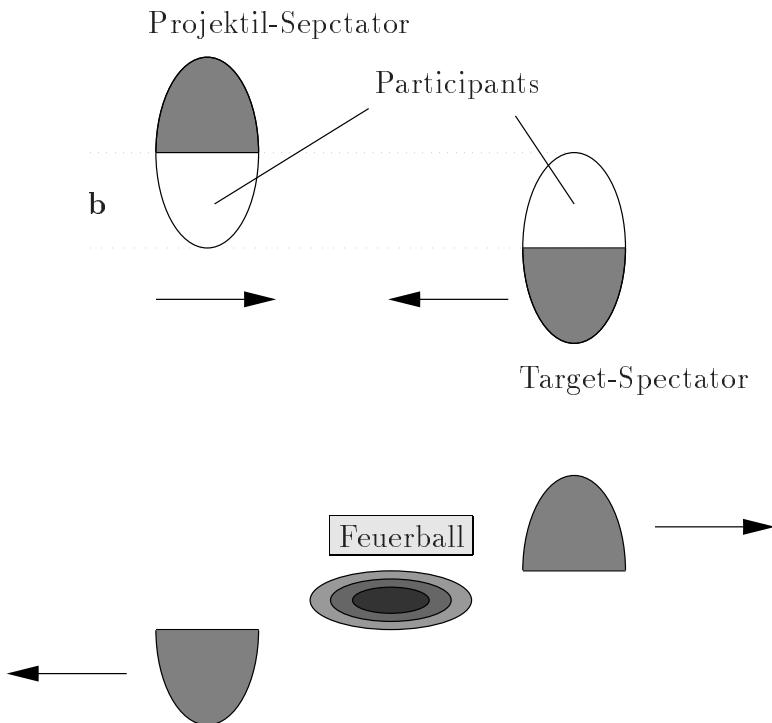


Abbildung 2.3: Veranschaulichung des *Participant-Spectator*-Modells.

und erhitzt werden. Wird dabei die Energie auf möglichst viele Freiheitsgrade verteilt, kann sich ein thermodynamisches Gleichgewicht bilden. Die Tatsache, daß es sich um unabhängige Teilchen handelt, läßt einfache geometrische Modelle zu. Das einfachste ist das *Participant-Spectator*-Modell bei dem sich ein *Feuerball* bildet. Hierbei werden Target- und Projektilkern aufgeteilt in Nukleonen, die an der Reaktion teilnehmen (engl. *participants*, Teilnehmer), und welche, die an der Reaktion nicht teilnehmen (engl. *spectators*, Zuschauer). Projektil und Target laufen unter dem Stoßparameter \mathbf{b} aufeinander zu. Im Bereich der Kollision entwickelt sich ein Gebiet, in dem die kinematische Energie in thermische Energie umgewandelt wird, es entsteht ein *Feuerball*.

In Abbildung 2.3 wird gezeigt, wie man sich so eine Reaktion vereinfacht vorstellt. Im Gebiet der Überdeckung entsteht eine Zone mit hoher Energiedichte, der *Feuerball*.

Ein möglicher Raum-Zeit-Verlauf einer Schwerionenreaktion mit hoher Energie und der Ausbildung eines Quark-Gluon-Plasmas wird in Abbildung 2.4 gezeigt. Die Reaktion durchläuft verschiedene Phasen. Nach der Kollision der Kerne und einer kurzen Vorgleichgewichtsphase bildet sich eventuell ein Quark-Gluon-Plasma aus. Bei der folgenden Expansion kühlt sich das Plasma ab. Es

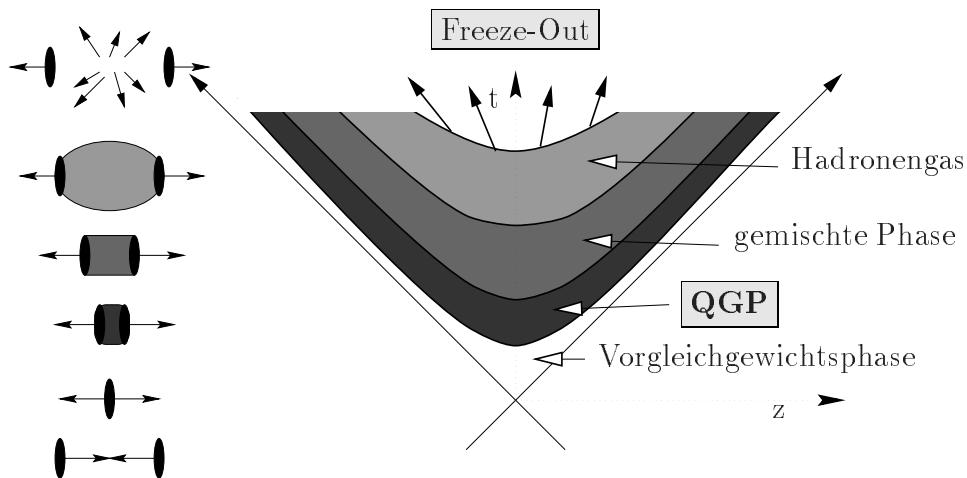


Abbildung 2.4: Raum-Zeit-Diagramm für eine Schwerionenreaktion mit Ausbildung eines Quark-Gluon-Plasmas. Die Kerne bewegen sich entlang des Lichtkegels.

geht in eine gemischte Phase über und kühlt weiter ab, um schließlich komplett als Hadronengas vorzuliegen. Das Ende der Reaktion, wenn die entstandenen Hadronen als freie Teilchen vorliegen und nicht mehr mit anderen Hadronen wechselwirken, wird mit *Ausfrieren* (engl. *freeze-Out*) bezeichnet.

2.3.2 Nachweismöglichkeiten eines QGP

Um bei der Schwerionenreaktion einen Phasenübergang der hadronischen Materie in ein Quark-Gluon-Plasma nachzuweisen, sind im Experiment eindeutige Signale zu messen. Viele der Signale sind aber durch Reaktionsmodelle ohne Ausbildung eines Quark-Gluon-Plasmas zu erklären. Deswegen setzt man nicht auf ein einzelnes Signal, sondern auf eine Kombination von verschiedenen Signalen.

Als mögliche Signaturen eines Quark-Gluon-Plasmas werden folgende Signale gesucht:

- Verringerte Produktion von J/Ψ -Teilchen:

Ähnlich der Debye-Abschirmung der QED werden die Farbladungen der einzelnen c -Quarks durch freie Quarks im Plasma abgeschirmt. Bei höher werdender Plasmatemperatur kommt es deshalb immer seltener zur Bildung eines gebundenen $c\bar{c}$ -Zustandes.

- Verstärkte Produktion von Strangeness:
Die Schwellenenergie zur Produktion eines $s\bar{s}$ -Paars liegt im Hadronengas bei ca. 700 MeV. In einem QGP liegt die Schwellenenergie bei etwa 300 MeV, ungefähr der doppelten Ruhemasse des s-Quarks.
- Produktion von Dileptonen:
In einem Quark-Gluon-Plasma kann bei einer Quark-Antiquark-Vernichtung ein virtuelles Photon produziert werden. Dieses zerfällt dann in ein Dilepton:

$$q + \bar{q} \rightarrow \gamma^* \rightarrow l^+ + l^- .$$

- Produktion von *direkten* Photonen:
Ein weiteres Signal direkt aus dem QGP sind Photonen. Die hauptsächlichen Prozesse sind:
Quark-Antiquark-Vernichtung

$$q + \bar{q} \rightarrow \gamma + g$$

und die Quark-Gluon-Comptonstreuung

$$q + g \rightarrow \gamma + q .$$

Besonders die beiden letzten Punkte sind interessant, da es sich um elektromagnetische Signale handelt. Diese Signale können den Reaktionsbereich relativ ungehindert verlassen, weil sie nicht *stark wechselwirken*. Außerdem können Aussagen über die Temperatur in der Frühphase der Reaktion gemacht werden. Ihre Messung ist aber wegen des starken Untergrundes nicht einfach. Bei den *direkten* Photonen wird der Untergrund durch Photonen aus den verschiedenen Entwicklungsphasen der Reaktion und den Zerfällen von Mesonen ($\pi^0 \rightarrow \gamma\gamma$) erzeugt. Die Dileptonen-Messung wird durch andere Prozesse, die auch Leptonen-Paare oder virtuelle Photonen erzeugen, gestört (Drell-Yan-Leptonen-Paare, Hadron-Antihadron-Reaktionen oder hadronische Zerfälle von Resonanzen).

3 WA98-Experiment

Beim WA98-Experiment am CERN wurde eine Reihe von Detektoren eingesetzt. Es wurden bei ultrarelativistischen Reaktionen von Bleiionen bei einer Energie von 158 GeV pro Nukleon nach Signaturen eines Quark-Gluon-Plasmas gesucht. Das Experiment wurde im Herbst 1994 am Super-Proton-Synchrotron (SPS) begonnen und lief bis Herbst 1996. Aufgrund der großen Anzahl von Daten sind effektive Untersuchungsmethoden erforderlich. Im Experiment wurde mit Hilfe von fast vollständig ionisierten Bleikernen durch Stöße mit einem festen Blei-Target, hochangeregte, komprimierte Kernmaterie, erzeugt. In diesem Gebiet könnte sich ein Quark-Gluon-Plasma ausgebildet haben. In den Vorgängerexperimenten, WA80 und WA93, wurden ^{16}O - bzw. ^{32}S -Ionen auf ein Gold-Target geschossen. Durch die höhere Ordnungszahl von Blei kann sich ein viel größeres Reaktionsvolumen ausbilden. Dadurch wird die Wahrscheinlichkeit, daß ein Quark-Gluon-Plasma entsteht, gesteigert. Durch die spezielle Kombination der Detektoren im WA98-Experiment erhofft man sich umfassende Informationen über den Verlauf der Reaktion. Durch Kombination der einzelnen Detektorsignale sollen verschiedene Signaturen für den Nachweis des Quark-Gluon-Plasmas gefunden werden [WA80, WA93, WA98].

3.1 Aufbau des WA98-Experimentes

In Abbildung 3.1 werden die verschiedenen Detektoren, die am WA98-Experiment beteiligt waren, dargestellt. Die Bleiionen treffen vor der Reaktion auf einen Gas-Čerenkov-Detektor mit sehr guter Zeitauflösung, der als Startdetektor verwendet wird. Das Blei-Target befindet sich im Inneren des Plastic-Ball-Spektrometers, mit dem die Teilchenproduktion um das Target untersucht wird. Der aus 655 kugelförmig angeordneten $\Delta E - E$ -Teleskopen bestehende Plastic-Ball erlaubt, durch die Energie- und Energieverlustmessung, die Identifikation von Hadronen bis zu einer Energie von 300 MeV. Außerdem kann der Energiefluß in Rückwärtsrichtung gemessen werden.

Mit dem Goliath-Magneten und den beiden *Tracking*-Armen wird versucht, anhand der rekonstruierten Bahnen, geladene Teilchen zu identifizieren. Dazu dienen zum einem sechs hintereinander stehende **Multi-Step-Avalance**-

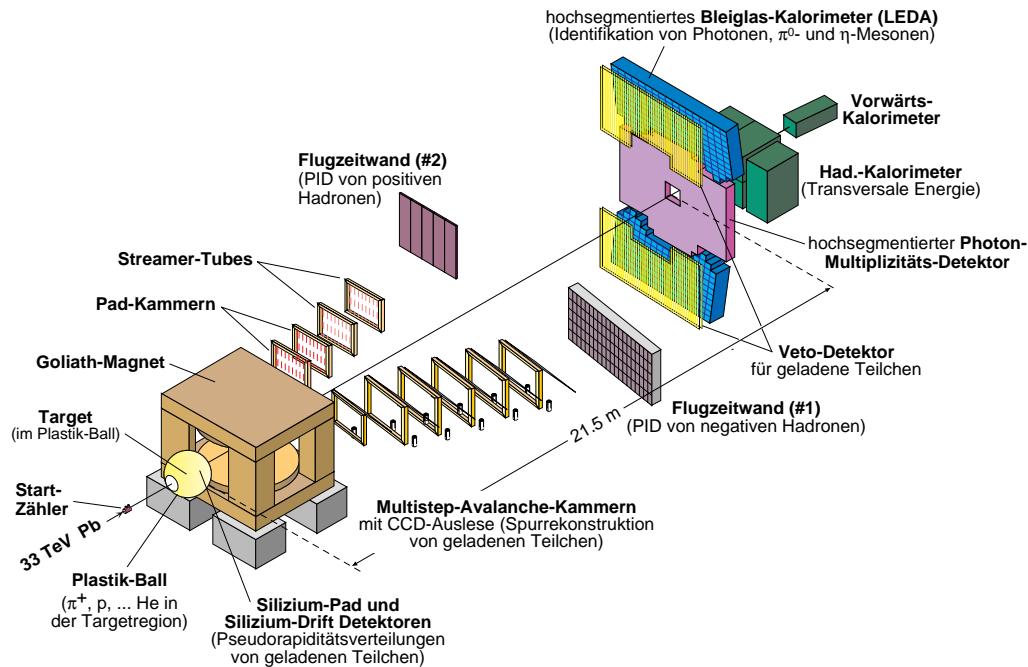


Abbildung 3.1: Der Aufbau des WA98-Experimentes.

Chambers (MSAC) im rechten Arm. Im linken *Tracking*-Arm stehen zwei Streamer-Tube-Detektoren und zwei Pad-Chambers. Am Ende jedes Armes befindet sich jeweils eine Flugzeitwand.

Am hinteren Ende des gesamten Aufbaues stehen verschiedene Kalorimeter, das **Zero-Degree-Calorimeter** (ZDC) und das **Midrapidity-Calorimeter** (MIRAC). Im MIRAC wird die transversale Energie gemessen. Das ZDC misst die Energie der Teilchen, die an der Reaktion kaum oder gar nicht teilgenommen haben. Mit diesen beiden Werten kann eine Aussage über die Zentralität der Reaktion gemacht werden.

Die Teilchenmultiplizitäten von geladenen Teilchen die bei der Reaktion entstehen, werden mit dem **Silizium-Drift-Detektor** (SDD) und einem **Silizium-Pad-Multiplizitäts-Detektor** (SPMD) direkt im Plastic-Ball gemessen. Die Multiplizität der Photonen wird mit dem **Photonen-Multiplizitäts-Detektor** (PMD) ermittelt.

Mit dem **Charged-Particle-Veto-Detektor** (CPV) sollen alle geladenen Teilchen, die im LEDA-Detektor gemessen wurden, verworfen werden, denn der LEDA-Detektor soll ausschließlich die bei der Reaktion entstandenen Photonen messen. Mit dem CPV kann auch die Multiplizität geladener Teilchen

Name	Pseudorapidität η	Zweck
Plastic-Ball	-1,7 bis 1,3	Spektrometrie geladener Teilchen
Silizium-Pad	2,4 bis 3,75	Multiplizität geladener Teilchen
PMD	2,9 bis 5,4	Photonen- Multiplizität
ZDC	$> 5,9$ $(\theta < 0,3^\circ)$	Bestimmung der Zentralität
MIRAC	3,7 bis 5,3 $(0,5^\circ < \theta < 3,0^\circ)$	Bestimmung der Zentralität
LEDA	2,28 bis 2,96 $(6,0^\circ < \theta < 11,6^\circ)$	Messung von Photonen

Tabelle 3.1: Daten der im WA98-Experiment verwendeten Detektoren. Zur Definition der Pseudorapidität siehe Anhang A.1

bestimmt werden. In Tabelle 3.1 sind einige der Detektoren mit ihrem Verwendungszweck und Pseudorapiditätsbereich aufgeführt.

3.2 Bleiglas-Detektor LEDA

Der LEDA-Detektor besteht aus 10080 Bleiglasmodulen. Jedes Modul ist ein $4 \times 4 \times 40$ cm³ großer TF1-Bleiglasblock¹. 6 × 4 dieser Einzelmodule sind zu einem **Supermodul** zusammengefasst. Jedes dieser insgesamt 420 Supermodule hat ein eigenes Referenzsystem, so daß bei einem späteren neuen Aufbau (z.B. beim PHENIX²-Experiment am RHIC³) eine neue Anordnung der Supermodule einfach wird. In [SCH94] werden Konstruktion und Eigenschaften der Supermodule beschrieben.

¹Zusammensetzung von TF1-Bleiglas: 51% PbO und 49% SiO₂.

²Pioneering High Energy Ion Experiment.

³Relativistic Heavy Ion Collider des National Laboratory in Brookhaven, New York.

3.2.1 Physikalische Prozesse

Verschiedene Teilchenarten wechselwirken unterschiedlich mit dem Detektormaterial. In der Regel entwickelt sich eine Kaskade von Sekundärteilchen, auch Schauer genannt. Die Schauerteilchen erzeugen Čerenkov-Photonen, welche mit Hilfe von Photovervielfachern ausgewertet werden. Der Schauer von hadronischen Teilchen (Pionen, Kaonen, Protonen, ...) unterscheidet sich von elektromagnetisch wechselwirkenden Teilchen (Photonen, Elektronen).

Elektromagnetische Schauer

Bei hohen Energien verlieren Photonen ihre Energie durch Paarbildung, Elektron und Positron durch Bremsstrahlung. Comptonstreuung und Ionisation können vernachlässigt werden.

Die Schichtdicke eines Materials in der ein Teilchen 63% seiner Energie depo-niert, wird Strahlungslänge X_0 genannt. Mit einer Wahrscheinlichkeit von 54% erzeugt ein Photon in einer Strahlungslänge X_0 ein Elektron- Positronpaar. Elektronen und Positronen erzeugen Bremsstrahlungsphotonen, welche weitere Elektron- Positronpaare bilden können. So bildet sich ein elektromagnetischer Schauer aus, der erst abbricht, wenn die Energie der entstehenden Photonen die kritische Energie E_c unterschreitet. E_c ist die Energie, bei der der Energieverlust durch Bremsstrahlung gleich dem durch Ionisation ist¹:

$$\left(\frac{dE}{dx} \right)_{ion} = \left(\frac{dE}{dx} \right)_{brems} = -\frac{E}{X_0} .$$

Von Photonen und Elektronen induzierte Schauer lassen sich nicht unterscheiden, da beide einen zu ähnlichen Verlauf haben.

E_c bestimmt auch die Schauertiefe X_{max} mit der maximalen Teilchenzahl:

$$\frac{X_{max}}{X_0} \sim \ln \left(\frac{E}{E_c} \right) - t$$

(mit $t = 1,1$ für Elektronen und $t = 0,3$ für Photonen).

Die radiale Energiedeposition fällt exponentiell ab und wird durch den Molière-Radius R_M beschrieben:

¹Formeln und Daten stammen im folgendem aus [KLE92].

$$R_M \approx \frac{21\text{MeV} \cdot X_0}{E_c} .$$

Legt man um die Schauerachse einen Zylinder mit $2R_M$, so werden etwa 95% der Energie innerhalb des Zylinders deponiert. Das Bleiglas TF1 im LEDA-Detektor besitzt eine Strahlungslänge von $X_0 = 2,78$ cm und einen Mollière-Radius von $R_M = 3,9$ cm.

Hadronische Schauer

Hadronische Teilchen, wie etwa Pionen oder Protonen, erzeugen durch inelastische Stöße mit dem Detektormaterial sekundäre Hadronen, meistens Pionen. Die Sekundärteilchen interagieren wiederum mit dem Detektormaterial, so daß ein hadronischer Schauer entsteht. Die weitere Schauerbildung bricht ab, wenn die Energie der erzeugten Teilchen so klein wird, daß sie vollständig abgebremst werden. Bei diesen Prozessen werden durch Bremsstrahlung oder π^0 -Zerfällen auch elektromagnetisch wechselwirkende Teilchen erzeugt. Mit λ_{int} bezeichnet man die Strecke, in der mit einer Wahrscheinlichkeit von 63% eine hadronische Wechselwirkung stattfindet:

$$\lambda_{int} = \frac{A}{\sigma \cdot N_A \cdot \rho}$$

(A : Molemasse, ρ : Dichte des Detektormaterials, N_A : Avogadrozahl, σ : inelastischer Wirkungsquerschnitt).

Die hadronische Wechselwirkungslänge λ_{int} ist in dem verwendetem Bleiglas mit $\lambda_{int} = 38$ cm wesentlich größer als die Strahlungslänge X_0 . Da die Bleiglasmodule mit 40 cm nur etwas mehr als eine hadronische Wechselwirkungslänge haben, wechselwirken nur etwa 64% aller Hadronen innerhalb des Detektors. Dabei deponieren sie meistens nur einen Bruchteil ihrer Energie im Detektor. Der Rest durchdringt den Detektor als minimalionisierendes Teilchen. Die mittlere Schauertiefe ist größer als bei elektromagnetischen Schauern, genauso wie die radiale Ausdehnung. Diese Merkmale, besonders die radiale Ausdehnung, werden zur Unterscheidung von elektromagnetischen und hadronischen Schauern benutzt.

Čerenkov-Photonen

Hat ein geladenes Teilchen im Detektormaterial eine höhere Geschwindigkeit als das Licht in dem Medium mit der Dispersionszahl n , so erzeugt es entlang seiner Bahn Čerenkov-Photonen:

$$v > \frac{c}{n} = c_{Phas.}$$

(c : Lichtgeschwindigkeit, $c_{Phas.}$: Lichtgeschwindigkeit im Medium).

Die Čerenkov-Strahlung wird, ähnlich dem Schall beim Durchbruch der Schallmauer eines Flugzeugs, unter einem von der Geschwindigkeit abhängigem Winkel θ abgestrahlt:

$$\cos \theta = \frac{1}{\beta \cdot n} \leq \frac{1}{n}$$

mit $\beta = \frac{v}{c}$.

Der Winkel beträgt für das TF1-Bleiglas $\theta \simeq 53^\circ$ bei einem Brechungsindex von $n = 1,647$. Durch Totalreflexion werden die Čerenkov-Photonen zum Photovervielfacher am Ende des Bleiglasmoduls geführt. Die Anzahl der Čerenkov-Photonen ist proportional zur Primärenergie der Teilchen. Die Photovervielfacher bilden aus den einfallenden Photonen ein analoges Spannungssignal, welches anschließend digitalisiert (Analog-Digital-Converter, ADC) wird. Diese ADC-Werte können im folgendem weiter verarbeitet werden.

3.2.2 Bearbeitung der Daten

Abbildung 3.2 zeigt einen Ausschnitt des Detektors mit einer simulierten Verteilung der ADC-Werte eines Photos (γ) und eines Pions (π^+). Das Photon wurde mit einer Energie von 10 GeV auf den Detektor geschossen und deponierte dabei insgesamt 9,921 GeV in den Modulen des Detektors. Das Pion hingegen wurde mit einer Energie von 20 GeV simuliert, deponierte dabei aber nur 2,057 GeV im Detektor. Gut ist die unterschiedliche Energieverteilung in den Modulen zu erkennen.

Nun müssen die ADC-Werte einzelnen Teilchen zugeordnet werden. Dazu wird im Detektor nach Clustern gesucht. Mit Cluster wird eine Gruppe von Modulen

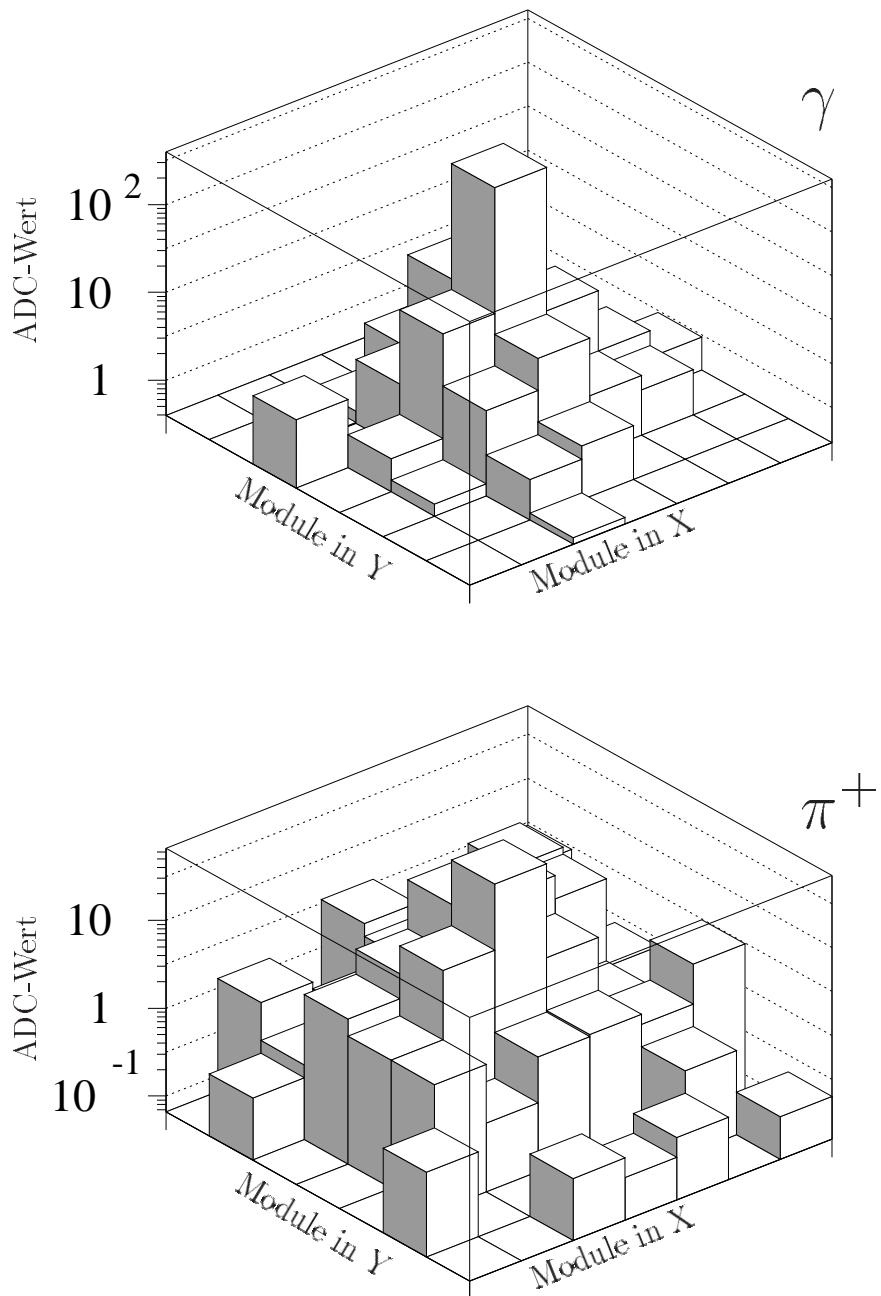


Abbildung 3.2: Exemplarische Verteilung der ADC-Werte eines simulierten Photons (oberes Bild) und eines geladenen Pions (unteres Bild).

genannt deren ADC-Werte größer als 0,0 und zusammenhängend sind. Um diese Cluster zu finden wird zunächst im Detektor nach einem Modul gesucht dessen ADC-Wert über 6,0 liegt. Dieses Modul wird als Maximum betrachtet und als Anfang eines Clusters angenommen. Als nächstes wird überprüft, ob

die direkt angrenzenden Module einen höheren ADC-Wert als 0,0 haben. Falls ein Modul mit einem höheren ADC-Wert als das Maximum gefunden wurde, wird dieses nun als neues Maximum des Clusters angesehen. Im der Umgebung des neuen Maximums wird ebenfalls nach angesprochenen Modulen gesucht. Im Laufe dieses Prozesses werden alle Module, die einen von Null verschiedenen ADC-Wert haben, in den Cluster aufgenommen und das Modul mit dem maximalen ADC-Wert bestimmt. Treten zwei Maxima in einem Cluster auf, d.h. beide Maxima sind durch eine Reihe von Modulen mit geringeren ADC-Werten verbunden, so werden den beiden Maxima die einzelnen Module zugeordnet, die Maxima werden entfaltet. Die dann einzeln vorliegenden Cluster werden jeweils als ADC-Verteilung eines Teilchens angesehen und weiter analysiert.

Flugzeit für Pionen und Photonen

Die Flugzeit ist für Pionen und Photonen unterschiedlich, das Signal der Pionen wird im Mittel etwas eher gemessen. Der Grund liegt in der unterschiedlichen Eindringtiefe. Pionen dringen im Durchschnitt viel tiefer in den Detektor ein bevor sie einen Schauer bilden. Die dabei entstehenden Čerenkov-Photonen haben nur einen kurzen Weg bis zum Photovervielfacher. Bei den Photonen fängt die Schauerbildung sehr früh an. Dadurch legen die Čerenkov-Photonen einen längeren Weg zurück, bis sie am Photovervielfacher ein Signal liefern. Durch die hohe Energie der Pionen bewegen sie sich fast mit Lichtgeschwindigkeit (Siehe dazu [BUS97]). Zum Zeitpunkt dieser Arbeit wurde an der Anwendbarkeit dieser Information noch gearbeitet.

π^0 -Meson Rekonstruktion

Die π^0 -Mesonen zerfallen direkt nach der Reaktion überwiegend in zwei Photonen. Anhand des Spektrums der invarianten Masse, von allen paarweise kombinierten als Photonen klassifizierten Teilchen, werden die π^0 -Mesonen bestimmt.

$$m_{\gamma\gamma} = \sqrt{2 \cdot E_1 E_2 (1 - \cos \Phi)} ,$$

(E_1, E_2 : Energien der beiden Photonen,
 Φ : Winkel zwischen den Photonen).

Im Spektrum tritt bei der π^0 -Masse ein Peak auf. Der Flächeninhalt dieses Peaks entspricht, nach Abzug des kombinatorischen Untergrundes, der Zahl der π^0 -Mesonen.

4 Neuronale Netze

Die häufigste Anwendung für neuronale Netze liegt in der Muster- und Bilderkennung. Weitere Anwendungen sind Spracherkennung, Optimierung und Funktionsapproximation sowie Vorhersagen. Gerade in der Bilderkennung ist das menschliche Gehirn allen Rechnern weit überlegen. Besondere Vorteile des Gehirns sind Robustheit, Kompaktheit, Flexibilität, Ausfalltoleranz und die Verarbeitung von fehlerhaften Daten. Hinzu kommt eine hohe Bearbeitungsgeschwindigkeit bei intensiven Parallelbetrieb (z.B. Koordination von Hand und Auge beim Lesen). Mit Hilfe von künstlichen neuronalen Netzen versucht man, einige dieser Eigenschaften bei Problemlösungen nutzbar zu machen. Mit sogenannten Basiselementen werden die Funktionsweisen und häufig auch die Topologie biologischer Netzwerke übernommen.

Ein simuliertes neuronales Netz ist dem Gehirn zwar weiterhin unterlegen, wenn es um die einzelne Bilderkennung geht, aber bei Klassifikationsproblemen mit hoher Statistik ist es eindeutig überlegen.

4.1 Biologisches Analogon

Schaut man sich das natürliche Vorbild an, merkt man rasch, daß eine komplette Nachbildung quasi unmöglich ist. Das menschliche Gehirn besteht aus ca. 10^{11} Nervenzellen, auch Neuronen genannt. Ein solches Neuron ist in Abbildung 4.1 schematisch dargestellt.

Über die Dendriten sammelt das Neuron die Signale anderer Neuronen in seiner Umgebung. Diese Signale bestehen aus komplexen chemischen Reaktionen bei denen Ionen (Ca^+ , Na^+) freigesetzt werden. Im Zellkörper werden die Ionen gesammelt und wenn das elektrische Potential einen Schwellwert überschreitet wird ein Signal variabler Länge, Frequenz und Intensität (Pulszug) über das Axon der Zelle abgegeben. Dieser Pulszug wird wieder von anderen Neuronen über deren Dendriten aufgenommen und z.B. zur Steuerung von Muskeln verwandt. Das Axon verstärkt oder schwächt den Pulszug bei der Übertragung. Diese Übertragung der Informationen erfolgt mittels Ionen. Die Stelle, wo ein Axon auf einen Dendriten einer anderen Zelle trifft, wird Synapse genannt.

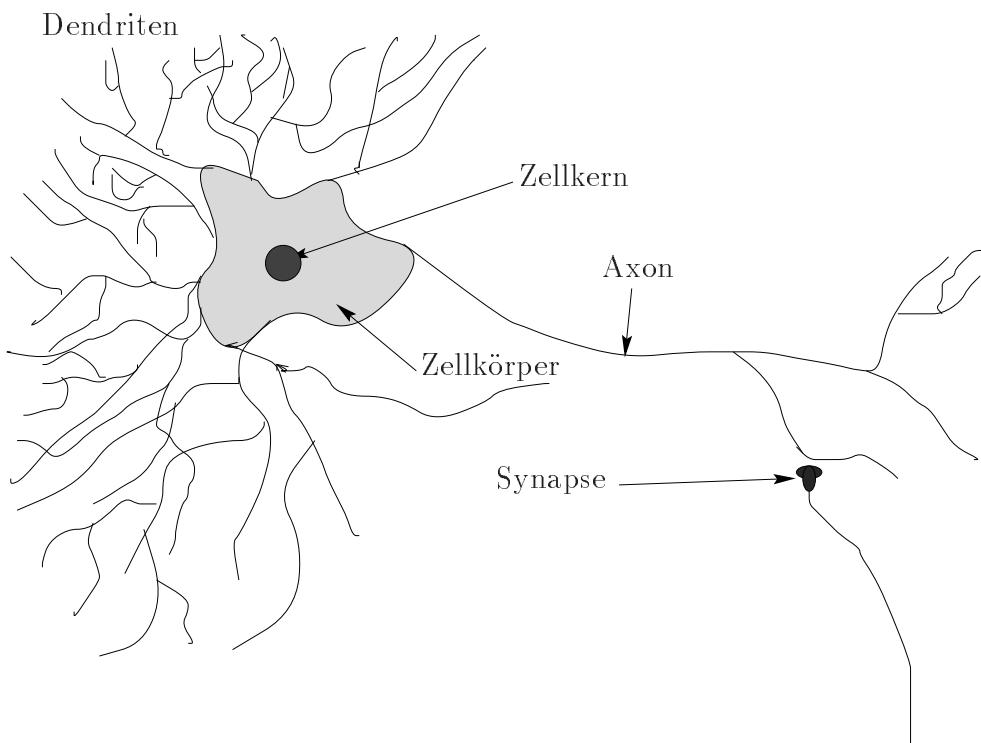


Abbildung 4.1: Schematische Darstellung des biologischen Neurons.

An dieser Stelle treten die Ionen aus dem Axon aus und diffundieren zum Dendriten. Mittels Enzymen kann hier die Stärke des Pulszuges geregelt werden [HER91].

4.2 Architektur künstlicher neuronaler Netze

Ein erstes, mathematisch abstraktes Modell eines Neurons wurde 1943 von McCulloch und Pitts entworfen. In Abbildung 4.2 ist dieses einfache Modell dargestellt. Die n Eingangssignale (I_k), die mit dem Neuron j verbunden sind, werden summiert, nachdem sie mit den *Gewichten* w_{jk} multipliziert wurden (siehe Gleichung 4.1). Die Eingangssignale stehen für die verschiedenen Dendriten des Neurons und die Gewichte für eine Abschwächung oder Verstärkung des Signals bei der Übertragung. Das Ausgangssignal O_j wird mit Hilfe der *Transferfunktion* g und den summierten gewichteten Eingangssignalen, abzüglich des Schwellwerts Θ_j , gebildet:

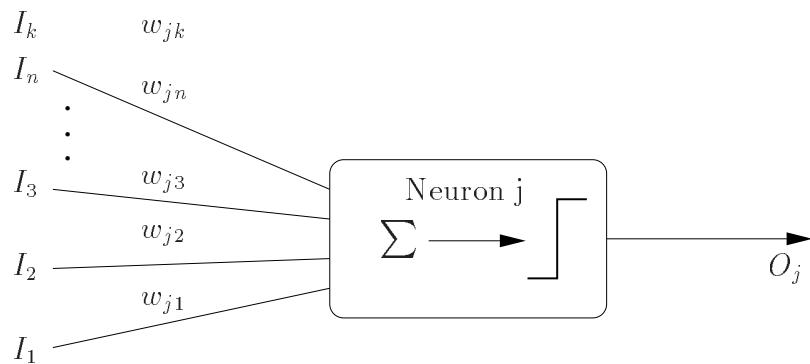


Abbildung 4.2: Darstellung des mathematischen Modells von McCulloch und Pitts.

$$O_j = g \left(\sum_{k=1,n} w_{jk} \cdot I_k - \Theta_j \right) . \quad (4.1)$$

Die einfachste Transferfunktion $g(x)$ ist die Heaviside-Funktion:

$$g(x) = H(x) = \begin{cases} 1 & \text{falls } x \geq 0, \\ 0 & \text{falls } x < 0. \end{cases}$$

Die Transferfunktion wird um den Schwellwert Θ_j auf der x-Achse verschoben. Der Schwellwert wird auch *Bias* genannt und in der Regel mit den Gewichten zusammengefaßt. Anstelle des Bias wird dann ein zusätzliches Eingangssignal erzeugt, daß immer auf -1,0 gesetzt wird ($I_0 = -1,0$). Das so hinzukommende Gewicht ist dem Bias gleichzusetzen ($w_{j0} = \Theta_j$). Die Summation läuft nun von $k = 0$ bis n .

Die Heaviside-Funktion hat den Nachteil, daß sie nicht stetig differenzierbar ist. Deshalb wird als Transferfunktion eine weichere, S-förmige (sigmoide) Funktion gewählt. Die folgende sigmoide Funktion wurde in dieser Arbeit verwendet:

$$g(x) = \frac{1}{1 + e^{-\frac{2x}{T}}} , \quad (4.2)$$

$$g'(x) = \frac{2}{T} g(x) \cdot (1 - g(x)) . \quad (4.3)$$

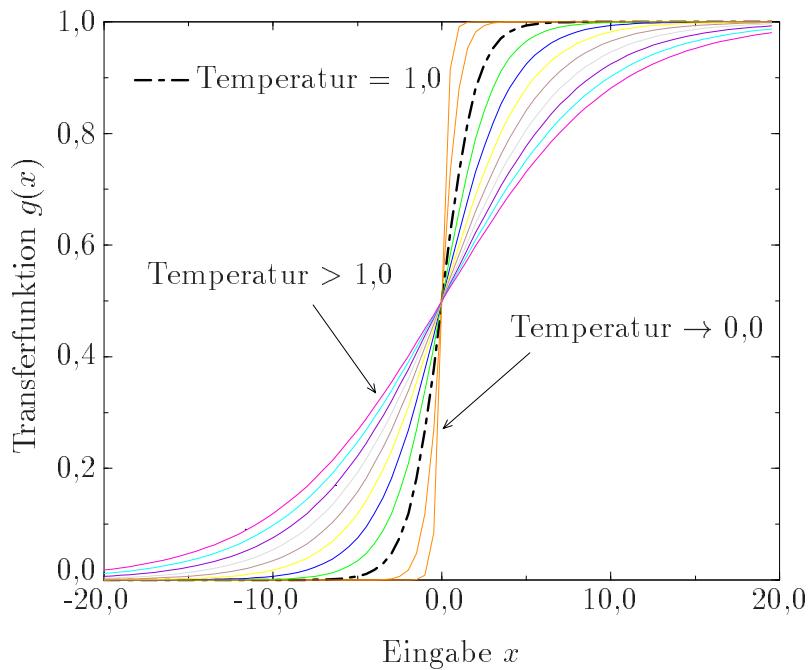


Abbildung 4.3: Veränderung der sigmoiden Funktion 4.2 für verschiedene Werte von T .

Der Parameter T sorgt für eine Abflachung ($T > 1$) der Funktion oder bei $T \rightarrow 0$ den Übergang in die Heaviside-Funktion. T wird auch als Temperatur bezeichnet. In Abbildung 4.3 ist die Transferfunktion für verschiedene Werte von T dargestellt. In der Regel wird T auf 1,0 oder 2,0 gesetzt (im folgendem mit $T = 1,0$).

4.2.1 Einschichtperzeptron

Das Einschichtperzeptron oder auch Perzeptron (SLP, *Single Layer Perzeptron*) genannt, ist das einfachste neuronale Netzwerk. Als Beispiel sei hier ein Perzeptron mit zwei Eingängen und einem Ausgang angeführt. Die Ausgabe des Netzes lässt sich dann mit folgender Formel beschreiben:

$$O = g\left(\sum_{n=1,2} w_n \cdot I_n - \Theta\right).$$

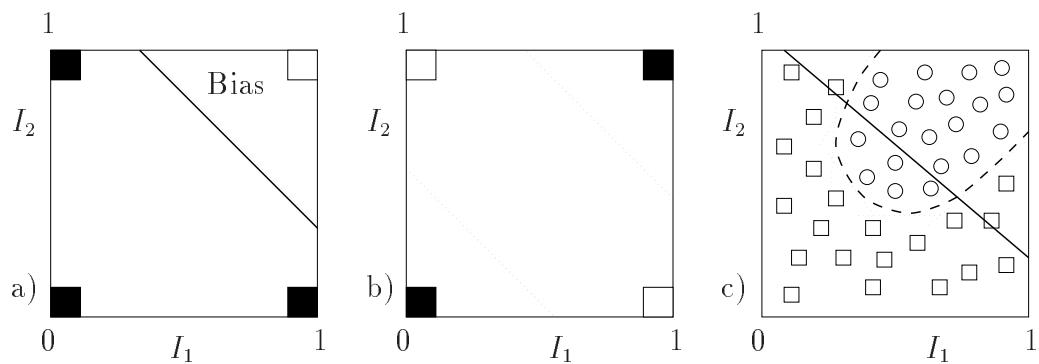


Abbildung 4.4: a) AND-Verknüpfung: Die Linie stellt den Bias des Neurons dar. b) XOR-Verknüpfung: Eine mögliche Trennung ist gestrichelt eingezeichnet. c) Nur durch einen nichtlinearen Schnitt (gestrichelt) ist eine Trennung Mengen (Kreise und Quadrate) möglich.

Der Index n steht für die beiden Eingänge I_n und Θ ist der Bias des Neurons. Für die Transferfunktion $g(x)$ nehmen wir die Heaviside-Funktion. In Abbildung 4.4 sind drei Beispiele dargestellt. Bild 4.4 a zeigt das einfache AND-Problem. An der Ausgabe soll nur eine 1 anliegen, wenn an beiden Eingängen eine 1 anliegt. Mit dem Perzeptron ist dieses Problem lösbar. In Abbildung 4.4 b wird gezeigt, daß das XOR-Problem, d.h. nur an einem Eingang darf eine 1 vorliegen, schon nicht mehr vom einfachen Perzeptron zu lösen ist. Mit einem Perzeptron lassen sich zwei Mengen nur linear trennen. Wird die Anzahl der Neuronen erhöht, können auch nichtlineare Schnitte, approximiert durch viele lineare Schnitte, durchgeführt werden (Siehe Abbildung 4.4 c).

Werden dem Perzeptron durch weitere Eingänge zusätzliche Informationen gegeben, könnte eine Trennung der Mengen erfolgen. Im Grunde gilt: Die Größe der Dimension des Eingaberaums (Anzahl der Eingänge) ist entscheidend für die Fähigkeit zur Klassifikation. So lassen sich auch Mengen trennen, die sich mit weniger Neuronen nicht trennen lassen. Da das Perzeptron in seinen Fähigkeiten recht beschränkt ist, dient es oft nur als Referenznetzwerk.

4.2.2 Multilayer-Netzwerke

Ein Multilayer Netzwerk besteht aus mehreren Schichten (engl. *layer*, Schicht) von Neuronen. Die erste Schicht sind die Eingänge (engl. *input layer*, Eingabeschicht), auch Eingangsneuronen genannt. Beim Zählen der Schichten

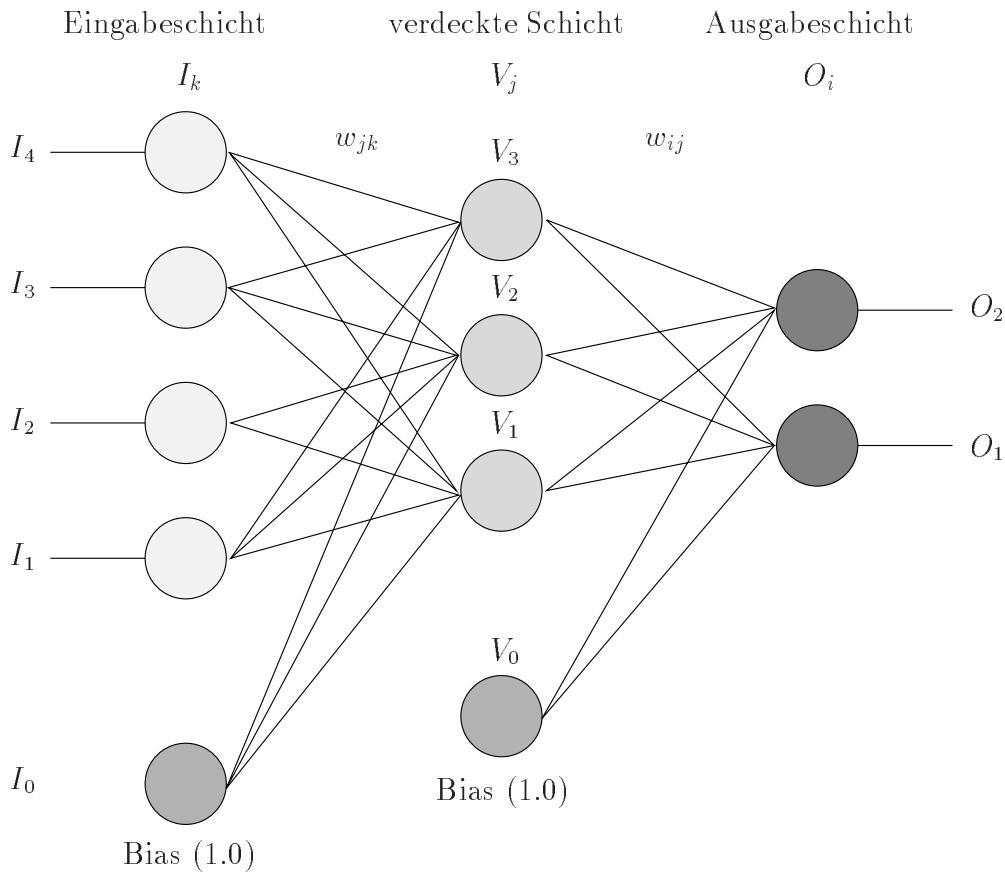


Abbildung 4.5: Einfaches Beispiel für ein 2-schichtiges (2 layer) Feed-Forward-Netzwerk. Der Bias der nächsten Schicht wird als zusätzliches Neuron der vorherigen Schicht eingeführt.

wird die Eingabeschicht nicht mitgezählt. Die folgende Lage von Neuronen wird verdeckte Schicht (engl. *hidden layer*) genannt, weil sie keine direkte Verbindung zur Außenwelt hat. Die verdeckte Schicht kann aus mehr als einer Schicht von Neuronen bestehen. Zum Schluß kommt die Ausgabeschicht (*output layer*) die die Ausgabeneuronen enthält. Jedes Neuron einer Lage hat zu jedem Neuron der nächsten Lage eine Verbindung. Die Informationen werden von der Eingabeschicht zur Ausgabeschicht geleitet, daher der Begriff Feed-Forward-Netze. In Abbildung 4.5 ist ein 2-Schicht Feed-Forward-Netzwerk dargestellt.

Von den Eingabeneuronen wird der Input (I_1 bis I_4 auch als Eingabevektor \vec{I} bezeichnet) mit den Gewichten w_{jk} multipliziert und von den Neuronen in der verdeckten Schicht summiert. Die Ausgabe der verdeckten Neuronen wird

mit den Gewichten w_{ij} multipliziert und von den Ausgabeneuronen erneut summiert. Die Ausgabeneuronen bilden dann die Ausgabe des Netzes:

$$\begin{aligned} h_j &= \sum_{k=0,4} w_{jk} \cdot I_k & V_j &= g(h_j) && (\text{mit } j = 1, 3), \\ l_i &= \sum_{j=0,3} w_{ij} \cdot V_j & O_i &= g(l_i) && (\text{mit } i = 1, 2). \end{aligned}$$

Die Netzwerkausgabe O_i lässt sich folgendermaßen zusammenfassen:

$$O_i(\vec{I}) = g\left(\sum_{j=1,3} w_{ij} \cdot g\left(\sum_{h=1,4} w_{hk} \cdot I_h\right)\right).$$

Die Gewichte w_{ij} und w_{jk} zwischen den einzelnen Neuronenschichten sind der Schlüssel zur Mustererkennung. Mit Hilfe dieser frei wählbaren Parameter lassen sich unterschiedliche Eingabevektoren auf bestimmte Ausgaben abbilden. Hat man einen vorgegebenen Datensatz \vec{I}^p von N verschiedenen ($p = 1 \dots N$) Mustern (engl. *pattern*), so gibt es zu jedem Eingabevektor ein Zielvektor \vec{T}^p (engl. *target*). Das Ziel ist es, daß Zielvektoren \vec{T}^p und Netzwerkausgabe \vec{O}^p möglichst gut übereinstimmen. Dazu muß das neuronale Netz *trainiert* werden, d.h. die Gewichte müssen entsprechend geändert werden. Dieses Training wird im folgenden Kapitel 5 erklärt.

5 Lernverfahren

Grundsätzlich gibt es zwei Arten, wie ein neuronales Netz trainiert werden kann, mit Lehrer und ohne Lehrer. Beim Training mit Lehrer (engl. *supervised learning*) sind die Zielvektoren \vec{T}^p bekannt. Die Gewichte werden so angepaßt, daß der Fehler zwischen Ausgabe und gewünschtem Ziel minimiert wird. Beim Training ohne Lehrer (engl. *unsupervised learning*) ordnet sich das Netz selbstständig an, um sich ähnelnde Eingaben zu einer Ausgabe zusammenzufassen (z.B. das selbstorganisierende Kohonenmodell [BIS95]).

Ein weiterer Unterschied wird bei der Präsentation der Daten gemacht. Es gibt die Möglichkeit, daß nach jedem präsentiertem Muster das Netz angepasst wird (Online-Verfahren), oder nach Präsentation eines kompletten Datensatzes (Batch-Verfahren).

Für das Training werden drei Datensätze benötigt:

- Trainingsdaten,
- Testdaten,
- Validierungsdaten.

Je größer die Anzahl der Muster im Trainingsdatensatz ist, desto besser ist es für das Trainingsverfahren. Da aber nur eine begrenzte Anzahl von Mustern zur Verfügung steht, braucht man eine Möglichkeit, um die Generalisierungsfähigkeit zu testen. Das ist die Fähigkeit des Netzes, Muster richtig zuzuordnen, die nicht am Training beteiligt waren.

Die Trainingsdaten werden benutzt, um die Gewichte im Netz anzupassen. Mit den Testdaten wird während des Trainings der Erfolg des Lernens beobachtet und die Generalisierungsfähigkeit getestet. Diese Testdaten werden nicht zum Ändern der Gewichte benutzt. Der dritte Datensatz dient zum Testen des Lernerfolges. Das Trainings- und Auswahlverfahren umfaßt folgende Punkte:

1. Training:

- (a) Initialisierung der Gewichte:

Für alle Gewichte wird ein Zufallswert berechnet.

- (b) Feed-Forward Berechnung für die Testdaten:

Die Ausgabe der Testdaten und deren Testfehler

$$\epsilon_{test} = \sum_p (\vec{T}^p - \vec{O}^p)^2$$

werden berechnet.

- (c) Feed-Forward Berechnung für die Trainingsdaten:

Die Ausgabe der Trainingsdaten und deren Trainingsfehler

$$\epsilon_{train} = \sum_p (\vec{T}^p - \vec{O}^p)^2$$

werden berechnet.

- (d) Anpassung der Gewichte.

- (e) Nächster Iterationsschritt:

Zurück zu Punkt 1-b, bis eine Konvergenz des Testfehlers auftritt oder die Anzahl der Iterationen ein Limit überschreitet.

2. Test:

- (a) Test der Klassifikationsfähigkeit mit simulierten Daten.

- (b) Test der Generalisierungsfähigkeit mit simulierten Daten.

3. Modellwahl an Hand von folgenden Eigenschaften:

- (a) Klassifikationsfähigkeit,

- (b) Generalisierungsfähigkeit,

- (c) verbrauchte Rechenzeit,

- (d) zukünftige Erweiterbarkeit.

4. Anwendung auf reale Meßdaten und Vergleich mit anderen Methoden.

Die Lernphase wird abgebrochen, wenn die zeitliche Entwicklung des Testfehlers eine Konvergenz zeigt oder die Anzahl der Iterationen, auch als Epoche bezeichnet, ein vorher festgesetztes Limit erreicht. Damit soll eine Überanpassung (engl. *Overfitting*) des Netzes an die Trainingsdaten verhindert werden. Die Abbruchbedingungen werden ausführlicher in Kapitel 5.2 beschrieben.

Die nun trainierten Netze werden auf Klassifikations- und Generalisierungsfähigkeit getestet. Anschließend wird das neuronale Netz mit den gewünschten Eigenschaften ausgesucht. Mit diesem Netz können nun reale Daten vom WA98-Experiment analysiert werden.

5.1 Lern-Algorithmen

Erst viele Jahre nach der Einführung von Multilayer-Netzwerken wurde ein effektiver Lern-Algorithmus entwickelt. Der Algorithmus wird *Backpropagation* genannt, da der Fehler von der Ausgabeschicht zurück zur Eingabeschicht durch das Netz geführt wird. Dieser Algorithmus kann bei jeder Art von *Feed-Forward* Netzwerken angewendet werden.

Im folgenden wird der Backpropagation-Algorithmus erklärt, da sich diese Routine als quasi-Standard etabliert hat. Im Anschluß folgt eine Erläuterung des Lern-Algorithmuses der in dieser Diplomarbeit angewendet wurde und eine Weiterentwicklung des Backpropagation-Algorithmus darstellt.

5.1.1 Backpropagation-Algorithmus

Der Backpropagation-Algorithmus (engl. *back propagation*, zurück Übertragung) wurde zuerst von Bryson und Ho im Jahre 1969 entdeckt. Doch erst Rumelhart, Hinton und Williams [RUM86] brachten mit seiner Wiederentdeckung und Weiterentwicklung im Jahre 1986 einen effektiven Algorithmus für das Training der Gewichte für alle Arten von Feed-Forward Netzwerken. Der Algorithmus beruht auf einem Gradientenabstiegsverfahren, um den euklidischen Abstand $\|\vec{T}^p - \vec{O}^p\|$ vom Ausgabe- und Zielvektor zu minimieren. Hierzu wird die quadratische Fehlerfunktion ϵ definiert¹:

$$\epsilon(w) = \sum_{p=1}^N (\vec{T}^p - \vec{O}^p)^2 . \quad (5.1)$$

Die Fehlerfunktion wird über alle Muster $p = 1, \dots, N$ gebildet, wobei \vec{T}^p die Zielvektoren und \vec{O}^p die Ausgabevektoren des neuronalen Netzes sind. Die Änderung der Gewichte von der Zeit t nach $t + 1$ erfolgt nach den Regeln:

$$\begin{aligned} w_{jk}^{t+1} &= w_{jk}^t + \Delta w_{jk} , \\ w_{ij}^{t+1} &= w_{ij}^t + \Delta w_{ij} . \end{aligned}$$

¹Die Formeln dieses Abschnittes können in [BIS95, HER91, ROJ93] nachgeschlagen werden.

Die Größe der Änderungen wird direkt über die Ableitung der Fehlerfunktion bestimmt:

$$\begin{aligned}\Delta w_{jk} &= -\eta \frac{\partial \epsilon}{\partial w_{jk}}, \\ \Delta w_{ij} &= -\eta \frac{\partial \epsilon}{\partial w_{ij}}.\end{aligned}$$

Der Parameter η wird als Lernrate bezeichnet und wird zur Regulierung der Lerngeschwindigkeit benutzt. Eine hohe Lernrate beschleunigt das Lernverfahren. Ein zu großes η kann aber dazu führen, daß das Minimum der Fehlerfunktion übersprungen wird. Eine niedrige Lernrate verlangsamt das Lernen, und man kann in einem lokalem Minimum der Fehlerfunktion stecken bleiben. Typische Werte für η sind 0,1 oder 0,2.

Für diesen Algorithmus ist eine stetig differenzierbare Transferfunktion, wie zum Beispiel die sigmoide Funktion in Gleichung 4.2 nötig. Nun lassen sich die Änderungen der einzelnen Gewichte zwischen den Neuronen berechnen. Für die Gewichte zwischen Ausgabeschicht (i) und verdeckter Schicht (j) gilt:

$$\Delta w_{ij} = -\eta \frac{\partial \epsilon}{\partial w_{ij}} = \eta \sum_{p=1}^N (T_i^p - O_i^p) \cdot g'(l_i^p) \cdot V_j^p.$$

Mit folgender Vereinfachung kommt man auf die in der Literatur häufig zu findende Darstellung:

$$\delta_i^p = (T_i^p - O_i^p) \cdot g'(l_i^p), \quad (5.2)$$

$$\Delta w_{ij} = \eta \sum_{p=1}^N \delta_i^p \cdot V_j^p.$$

Mit der Kettenregel werden nun die Änderung der Gewichte zwischen der verdeckten Schicht (j) und der Eingabeschicht (k) berechnet:

$$\begin{aligned}\Delta w_{jk} &= -\eta \frac{\partial \epsilon}{\partial w_{jk}} = -\eta \sum_{p=1}^N \frac{\partial \epsilon}{\partial V_j^p} \cdot \frac{\partial V_j^p}{\partial w_{jk}} \\ &= \eta \sum_{p=1}^N \sum_i (T_i^p - O_i^p) \cdot g'(l_i^p) \cdot w_{ij} \cdot g'(h_j^p) \cdot I_k^p.\end{aligned}$$

Mit der Vereinfachung aus Gleichung 5.2 erhält man

$$\Delta w_{jk} = \eta \sum_{p=1}^N \sum_i \delta_i^p \cdot w_{ij} \cdot g'(h_j^p) \cdot I_k^p ,$$

und mit folgender Zusammenfassung gelangt man schließlich zu einer sehr kompakten Schreibweise:

$$\begin{aligned}\delta_j^p &= g'(h_j^p) \cdot \sum_i w_{ij} \cdot \delta_i^p , \\ \Delta w_{jk} &= \eta \sum_{p=1}^N \delta_j^p \cdot I_k^p .\end{aligned}$$

Mit diesem Verfahren lassen sich die Gewichte so ändern, daß die Fehlerfunktion 5.1 minimiert wird. Dabei werden die Neuronen des Netzes nicht verändert oder neu angeordnet, wie es in anderen Netzwerktypen der Fall ist. Um den Lernvorgang zu beschleunigen und der Gefahr, in einem lokalen Minimum der Fehlerfunktion stecken zu bleiben, zu entgehen, wurden viele Erweiterungen und Variationen des normalen Backpropagation-Algorithmus entwickelt. Die drei häufigsten werden nun kurz erläutert:

- Hinzufügen eines *Impulsterms*:

Bei der Berechnung der Gewichtsänderung zum Zeitpunkt $t + 1$ wird ein Teil der vorherigen Änderung addiert:

$$\Delta w_{uv}(t + 1) = -\eta \frac{\partial \epsilon}{\partial w_{uv}} + \alpha \cdot \Delta w_{uv}(t) \quad (\text{mit } \alpha = 0,01).$$

Damit lassen sich Oszillationen in Gebieten mit einem steilen Anstieg oder Abfall in der Fehlerfunktion verhindern, da die Gewichtsänderung verkleinert wird.

- Änderung der Lernrate η :

Die Lernrate η wird am Anfang des Lernprozesses recht groß gewählt und im Laufe des Trainings verkleinert. Damit möchte man erreichen, daß am Anfang die Änderung der Gewichte stark ist und im Laufe des Trainings, wenn das Minimum nicht mehr so weit entfernt ist, die Gewichte nicht mehr so stark geändert werden.

- Änderung der Temperatur T :

Eine Änderung der Temperatur T bewirkt den gleichen Effekt. Ist die Temperatur hoch, so ist die Transferfunktion (Gleichung 4.2) relativ flach. Mit sinkender Temperatur wird die Funktion immer steiler (Siehe auch Abbildung 4.3).

Der nun folgende *Rprop*-Algorithmus ist eine Variante des Bakpropagation-Algorithmus. Da der Algorithmus in dieser Arbeit angewandt wurde, wird er nun etwas ausführlicher beschrieben.

5.1.2 *Rprop*-Algorithmus

Rprop steht für **R**esilient **b**ack**p**ropagation (engl. *resilient*, Rücksprung) und beschreibt einen Lernalgorithmus, der lokal die Gewichte ändert. Ein Problem beim Backpropagation-Algorithmus ist die Abhängigkeit von der Größe der Ableitung, da der Wert der Ableitung direkt zur Änderung der Gewichte benutzt wird. Beim *Rprop*-Algorithmus wird deswegen nur das Vorzeichen der Ableitung der Fehlerfunktion benutzt. Anhand dessen wird die Richtung der Gewichtsänderung bestimmt. Die Größe der Änderung wird für jedes Gewicht einzeln bestimmt, dem *Update*-Wert Δ_{ij} . Zuerst wird anhand des Vorzeichens der Ableitung das Vorzeichen für den *Update*-Wert bestimmt (für die Gewichte einer anderen Schicht kann einfach ij durch jk ausgetauscht werden)¹:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t), \quad (5.3)$$

$$\Delta w_{ij}(t) = \begin{cases} -\Delta_{ij}(t) & , \text{ wenn } \frac{\partial \epsilon}{\partial w_{ij}} > 0, \\ +\Delta_{ij}(t) & , \text{ wenn } \frac{\partial \epsilon}{\partial w_{ij}} < 0, \\ 0 & , \text{ sonst.} \end{cases} \quad (5.4)$$

Bei dem *Manhattan*-Algorithmus wird in Gleichung 5.4 Δ_{ij} durch einen konstanten Wert Δ ersetzt. Im Unterschied dazu werden nun die neuen *Update*-Werte Δ_{ij} ermittelt:

¹Die Formeln zum *Rprop*-Algorithmus stammen, soweit nicht anderweitig angegeben, aus [RIE93].

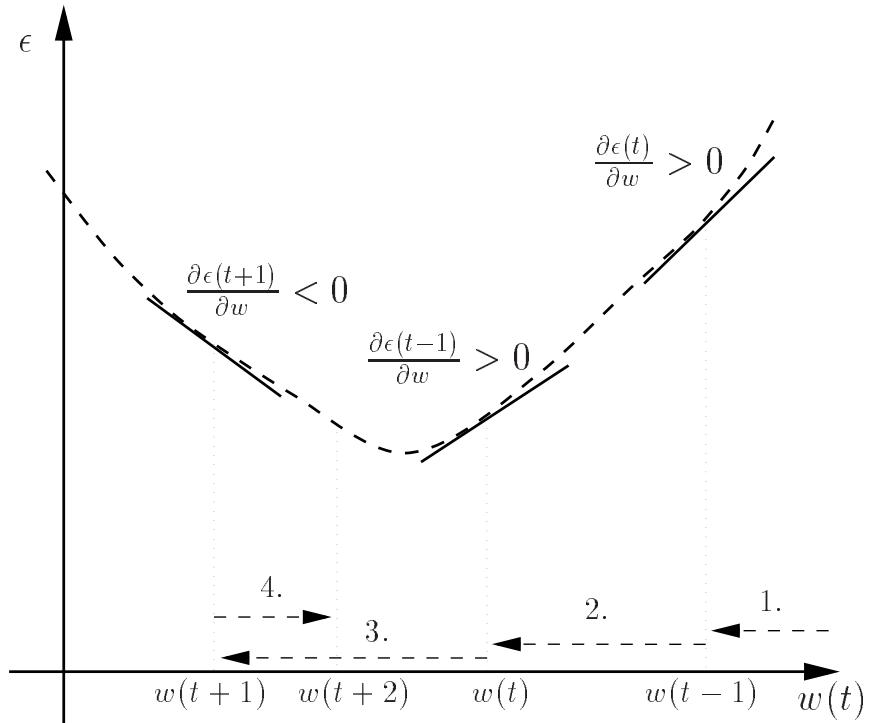


Abbildung 5.1: Ein einfaches Beispiel des *Rprop*-Algorithmus.

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^+ \cdot \Delta_{ij}^{(t-1)} & , \text{ wenn } \frac{\partial \epsilon}{\partial w_{ij}} \cdot \frac{\partial \epsilon^{(t)}}{\partial w_{ij}} > 0, \\ \eta^- \cdot \Delta_{ij}^{(t-1)} & , \text{ wenn } \frac{\partial \epsilon}{\partial w_{ij}} \cdot \frac{\partial \epsilon^{(t)}}{\partial w_{ij}} < 0, \\ \Delta_{ij}^{(t-1)} & , \text{ sonst,} \end{cases} \quad (5.5)$$

(mit $0 < \eta^- < \eta^+$) .

Als einfache Anschauung wird nun ein eindimensionales Problem betrachtet. In der Abbildung 5.1 ist die Fehlerfunktion ϵ gegen die Gewichte $w(t)$ zu verschiedenen Zeitpunkten aufgetragen. Im ersten Schritt wird das Gewicht $w(t-2)$ folgendermaßen zu $w(t-1)$ geändert, da die Ableitung der Fehlerfunktion größer als Null ist (auf dem Bild nicht zu sehen):

$$w(t-1) = w(t-2) + \eta^+ \cdot \Delta w(t-2) .$$

Nun wird an der Stelle $w(t-1)$ erneut die Ableitung der Fehlerfunktion berechnet und falls kein Vorzeichenwechsel eintritt, wird die Schrittweite vergrößert, und man gelangt zu $w(t)$. $w(t+1)$ erhält man auf dem gleichen Weg:

$$\begin{aligned} w(t) &= w(t-1) + \eta^+ \cdot \Delta w(t-1) , \\ w(t+1) &= w(t) + \eta^+ \cdot \Delta w(t) . \end{aligned}$$

Bei $w(t+1)$ ist man nun über das Minimum der Fehlerfunktion hinaus gekommen und ein Vorzeichenwechsel der Ableitung ist aufgetreten. Nun geht die Änderung in die andere Richtung ($\Delta w(t+1)$ ist negativ) und die Schrittweite wird verringert:

$$w(t+2) = w(t+1) + \eta^- \cdot \Delta w(t+1) .$$

Dieses Verfahren wird nun fortgesetzt bis man zum Minimum gelangt.

Anfangs sind alle *Update*-Werte auf den Anfangswert $\Delta_0 = 0,1$ gesetzt. Der Wert Δ_0 wird direkt zur ersten Änderung genommen. Die Wahl des Anfangswertes beeinflußt den Lernprozeß kaum, da Δ_{ij} ständig angepaßt wird. Außerdem ist der *Update*-Wert noch durch ein $\Delta_{max} = 20,0$ und $\Delta_{min} = 10^{-5}$ begrenzt. Damit soll verhindert werden, daß die Gewichte zu groß werden oder daß der Algorithmus zu schnell stecken bleibt. Auch eine Änderung dieser Parameter ist für den Lernprozeß unkritisch. Die beiden Faktoren zum Vergrößern und Verkleinern von Δ_{ij} werden mit $\eta^+ = 1,2$ und $\eta^- = 0,5$ festgelegt. Weitere Informationen zu diesem speziellem Lernalgorithmus lassen sich in [RIE93] finden.

Vergleicht man nun die beiden Lern-Algorithmen anhand eines einfachen 3-Bit-XOR Problems sieht man den Unterschied zwischen den beiden Algorithmen. Daß der *Rprop*-Algorithmus dieses Problem schneller löst kann in Abbildung 5.2 deutlich gezeigt werden.

Als Erweiterung des *Rprop*-Algorithmus wird an Gleichung 5.1 ein sogenannter *weight-decay* Term angefügt. Dieser wirkt einer Überanpassung entgegen (Siehe Kapitel 5.2) und soll den Wert der Gewichte klein halten [SNNS4]:

$$\epsilon = \sum_p (\vec{T}^p - \vec{O}^p)^2 + 10^{-\alpha} \cdot \sum w^2 ,$$

(mit $\alpha = 5$).

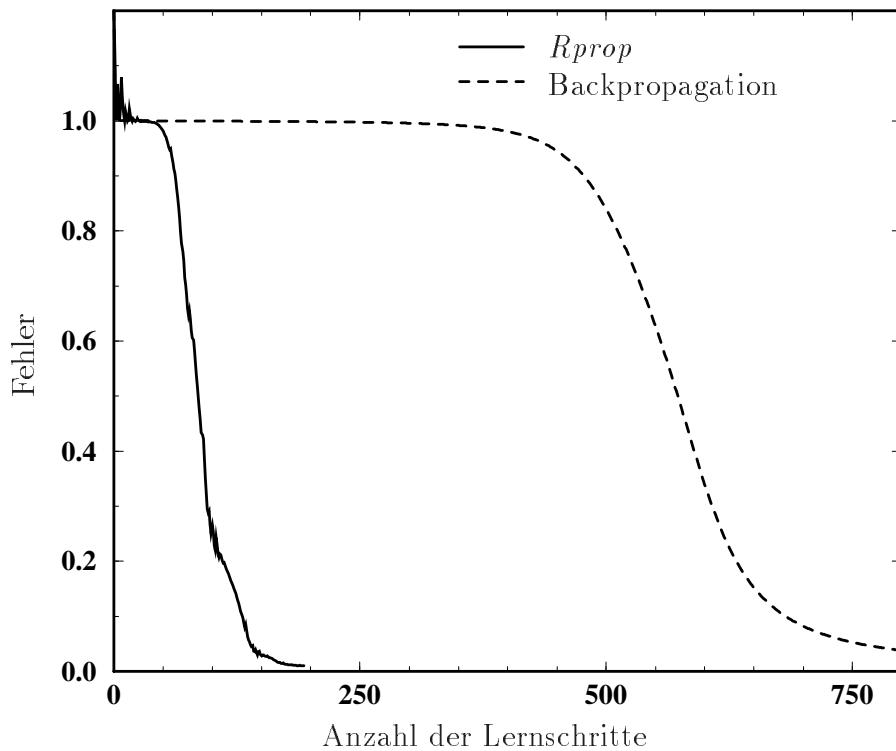


Abbildung 5.2: Vergleich der Lerngeschwindigkeit zwischen Bakpropagation- und *Rprop*-Algorithmus.

Außerdem wurde bei der Entwicklung des Programms festgestellt, daß der *Rprop*-Algorithmus auf Unterschiede in der Anzahl der Muster für die verschiedenen Klassen (hier: Photonen und Pionen) in den Trainingsdaten reagiert. Um das auszugleichen, wurde bei der Bildung der Ableitung der Fehlerfunktion ein Normierungsfaktor integriert. So wird beim Training keine der Klassen aufgrund einer höheren Anzahl von Mustern bevorzugt gelernt.

5.2 Lernabbruch

Der Zeitpunkt für den Abbruch des Lernens, beim Training eines neuronalen Netzes, ist sehr wichtig. Wird das Training zu früh eingestellt, ist das Netz oft nicht in der Lage, akzeptable Ergebnisse bei der späteren Anwendung zu erreichen. Wird das Training zu lange fortgesetzt, tritt das Problem der *Überanpassung* (engl. *Overfitting*) auf. Überanpassung bedeutet, daß die Ge-

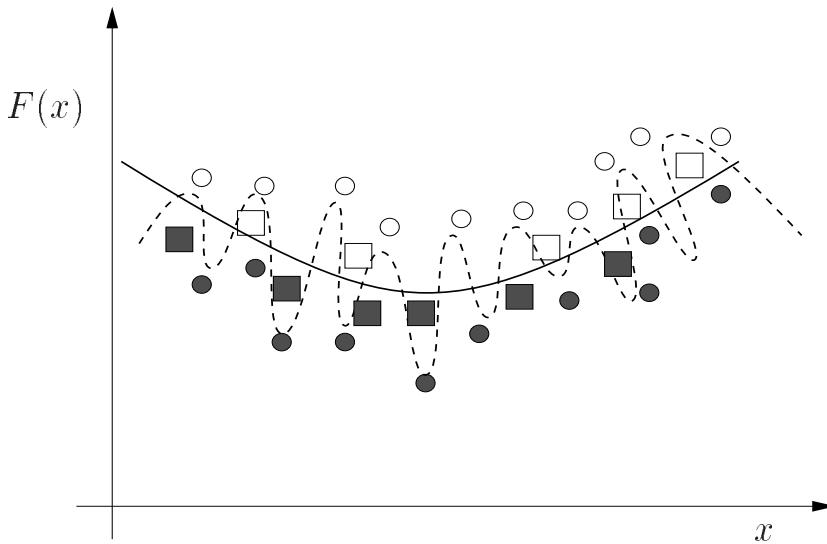


Abbildung 5.3: Beispiel zur Überanpassung. Die dunklen und hellen Objekte stehen für zwei zu trennende Klassen. Die Kreise waren am Training beteiligt, die Quadrate nicht.

neralisierungsfähigkeit stark eingeschränkt ist. Am Training beteiligte Muster werden sehr gut klassifiziert. Muster die am Training nicht beteiligt waren, werden kaum richtig klassifiziert.

Als Beispiel ist in Abbildung 5.3 gezeigt wie man sich eine Überanpassung bei einem eindimensionalem Problem vorzustellen hat. Die dunklen und hellen Kreise stellen Muster aus den Trainingsdaten und die dunklen und hellen Quadrate Muster, die nicht trainiert wurden. Die durchgezogene Linie stellt die ideale Trennlinie zwischen den beiden Klassen (dunkel, hell) dar. Die Überanpassung wird durch die gestrichelte Linie verdeutlicht. Einige Muster der unabhängigen Daten werden falsch klassifiziert, weil die Trennungsgeraden zu stark an die speziellen Eigenschaften der Trainingsdaten angeglichen wurde. Die Generalisierung, eine Anwendung auf Muster die nicht in den Trainingsdaten enthalten waren, ist nicht mehr möglich.

Den Effekt der Überanpassung kann man während der Lernphase überprüfen. Wenn man nach jeder Anpassung der Gewichte dem neuronalen Netz einen zweiten, vom ersten unabhängigen Datensatz, präsentiert geben die Fehler (Trainings- und Testfehler) Auskunft über eine mögliche Überanpassung. Der Trainingsfehler wird durch die Muster erzeugt, die am Training beteiligt sind, der Testfehler von den am Training unbeteiligten Mustern. Zunächst sinken

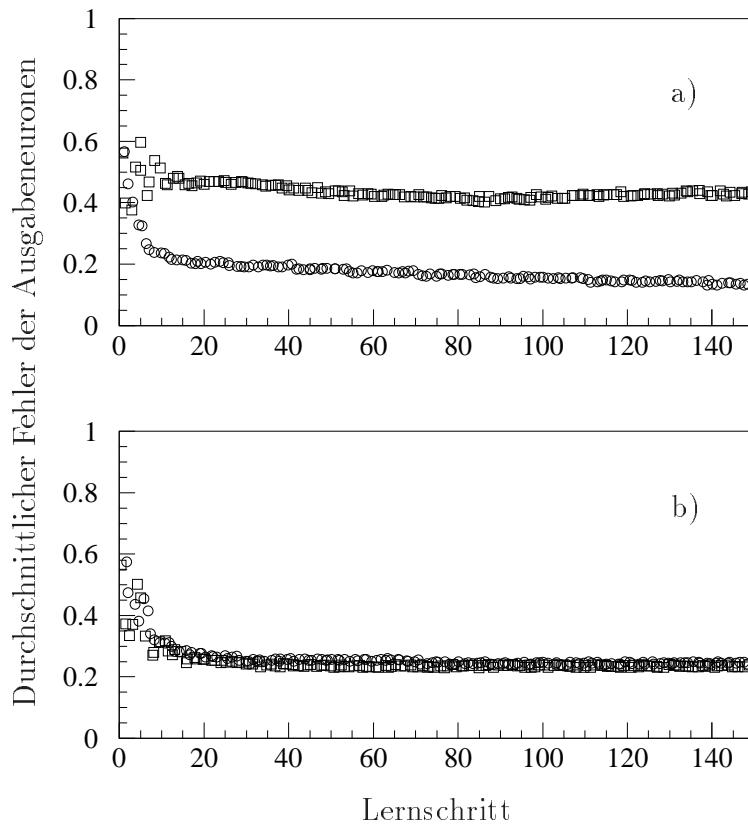


Abbildung 5.4: In Bild a) ist der *Rprop*-Algorithmus ohne den *weight-decay*-Term beim Lernprozeß eingesetzt worden und in Bild b) mit. Die Kreise gehören zu den am Training beteiligten Daten, die Quadrate zu den Testdaten.

beide Fehler bis zu einem bestimmten Grenzwert, wobei der Testfehler meistens leicht höher liegt. Wird das Training nun nicht abgebrochen, kann der Testfehler wieder ansteigen, während der Trainingsfehler weiter sinkt, die Überanpassung an die Trainingsmuster fängt nun an. Der Anstieg des Testfehlers vom zweiten Datensatz wird als Abbruchbedingung für die Lernphase benutzt (engl. *early stopping*).

In Abbildung 5.4 ist der durchschnittliche Fehler pro Ausgabeneuron gegenüber dem Lernschritt aufgetragen. Das obere Bild zeigt den *Rprop*-Algorithmus ohne *weight-decay*-Term, das untere mit. Die Überanpassung ist in Bild 5.4 a) deutlich zu sehen. Der Fehler der Trainingsdaten (Kreise) sinkt kontinuierlich, während der Fehler der Testdaten (Quadrate) relativ konstant, aber deutlich höher als der Fehler der Trainingsdaten liegt. Im unteren Bild mit *weight-*

decay-Term sind beide Fehler fast gleich groß, und es tritt keine signifikante Überanpassung auf. Bei dem in dieser Arbeit angewendeten *Rprop*-Algorithmus mit *weight-decay*-Term findet keine Überanpassung statt.

Bei anderen Lernalgorithmen kann die Überanpassung nur durch vorzeitigen Lernabbruch gelöst werden. Die theoretische Lösung durch beliebig viele Trainingsmuster ist nicht anwendbar, da man immer nur einen begrenzten Datensatz zur Verfügung stehen hat. Bei einer beliebig großen Anzahl von Mustern zum Trainieren, würde eine Überanpassung nicht auftreten, da immer wieder verschiedene Muster als Eingabe dienen würden.

6 Netzwerk Training

Mit dem *Rprop*-Algorithmus aus dem letzten Kapitel wurden verschiedene neuronale Netze untersucht. Die neuronalen Netze unterscheiden sich in der Anzahl der Eingabeneuronen und der Neuronen in der verdeckten Schicht. Zunächst werden einige Begriffe erklärt, die bei der folgenden Untersuchung benötigt werden. Nach der Beschreibung der verschiedenen neuronalen Netzwerke, wird deren Training beschrieben.

6.1 Begriffe zur Wahrscheinlichkeitsinterpretation

Mit der Ausgabe des neuronalen Netzes wird die a-posteriori-Wahrscheinlichkeit (lat. posterus, nachfolgend), auch Bayes-Wahrscheinlichkeit genannt, *geschätzt*. Im folgenden werden die Begriffe zur probabilistischen Klassifikation erklärt.

In der formalen Beschreibung sind die Begriffe Muster und Klasse genau definiert:

- Merkmalsraum:

Der Merkmalsraum enthält die Menge aller Merkmalsvektoren. Die Muster $\vec{I} = (I_0, \dots, I_{28})$ sind Elemente eines 29-dimensionalen Merkmalsraumes.

- Zustandsraum:

In dem Zustandsraum sind die beiden Klassen Γ und Π enthalten. Die Elemente der beiden Klassen sind die Muster \vec{I} . Hierbei steht Γ für die Klasse der Photonen und Π für die Klasse der Pionen.

Es wird bei der Berechnung der Wahrscheinlichkeiten ein Vorwissen in Form von a-priori-Wahrscheinlichkeiten (lat. prior, voriger) benutzt.

6.1.1 A-priori-Wahrscheinlichkeit

Die a-priori-Wahrscheinlichkeit wird aus der Verteilung der Muster des Trainingsdatensatzes ermittelt. $P(\Gamma)$ ist die a-priori-Wahrscheinlichkeit dafür, daß ein Teilchen der Klasse Γ angehört. $P(\Pi)$ ist die a-priori-Wahrscheinlichkeit für die Zugehörigkeit eines Teilchens zur Klasse Π . Bei einem Trainingsdatensatz mit 9991 Photonen und 6793 Pionen ergibt sich eine a-priori Wahrscheinlichkeit von $P(\Gamma) = 60\%$ und $P(\Pi) = 40\%$. Diese Verteilung entspricht nicht der real erwarteten Verteilung, ist aber bei dem hier angewandtem Lernalgorithmus durchaus vernünftig.

6.1.2 Likelihood

Der Likelihood ist die bedingte Wahrscheinlichkeit $P(\vec{I}|\Gamma)$ und $P(\vec{I}|\Pi)$, mit der die Klassen Γ bzw. Π ein bestimmtes Muster \vec{I} erzeugen. Die bedingten Wahrscheinlichkeiten $P(I_{13}|\Gamma)$ und $P(I_{13}|\Pi)$ sind in Abbildung 6.1 zu sehen. Diese bedingten Wahrscheinlichkeiten wurden vom Netz gelernt. Die Eingabevektoren \vec{I} wurden auf Werte zwischen 0,0 und 4,0 normiert. Weiterhin wurden nur Muster die ihre Energie in mehr als 3 Modulen deponieren verwendet.

6.1.3 A-posteriori-Wahrscheinlichkeit

Mit den a-priori-Wahrscheinlichkeiten $P(\Gamma)$ und $P(\Pi)$ und den bedingten Wahrscheinlichkeiten $P(\vec{I}|\Gamma)$ und $P(\vec{I}|\Pi)$ können die a-posteriori-Wahrscheinlichkeiten, $P(\Gamma|\vec{I})$ und $P(\Pi|\vec{I})$, berechnet werden. Bei einem gegebenen \vec{I} gibt

$$P(\Gamma|\vec{I}) = \frac{P(\vec{I}|\Gamma) \cdot P(\Gamma)}{P(\vec{I}|\Gamma) \cdot P(\Gamma) + P(\vec{I}|\Pi) \cdot P(\Pi)},$$

die a-posteriori-Wahrscheinlichkeit an, daß \vec{I} zur Klasse Γ gehört und

$$P(\Pi|\vec{I}) = \frac{P(\vec{I}|\Pi) \cdot P(\Pi)}{P(\vec{I}|\Gamma) \cdot P(\Gamma) + P(\vec{I}|\Pi) \cdot P(\Pi)},$$

die a-posteriori-Wahrscheinlichkeit an, daß \vec{I} zur Klasse Π gehört.

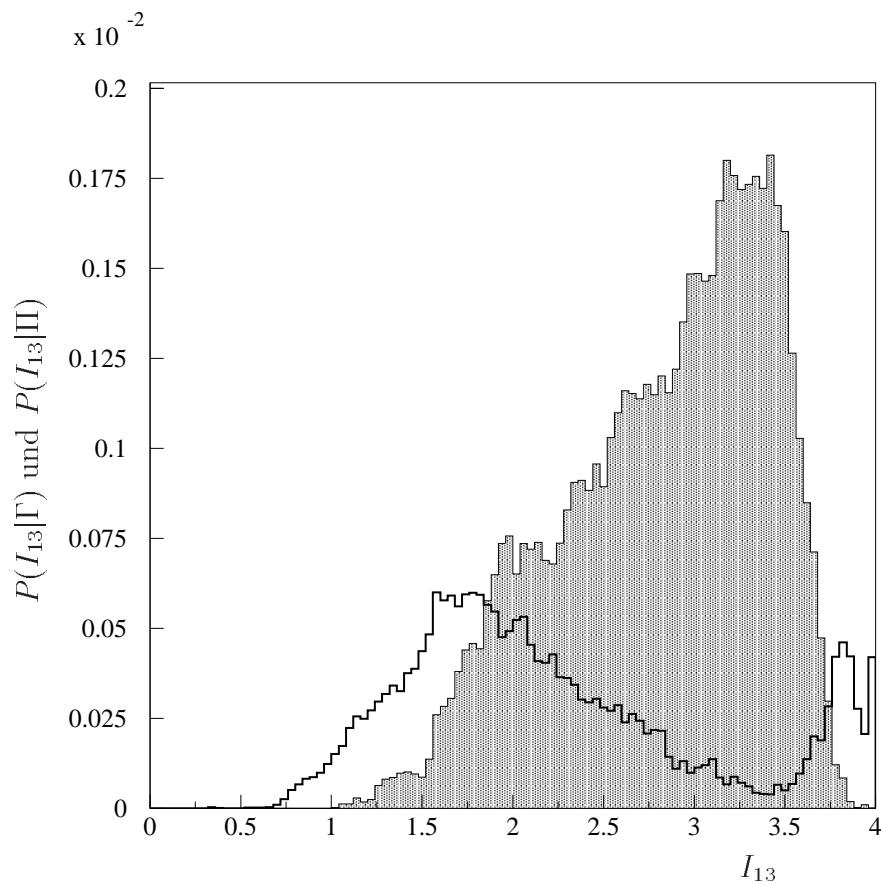


Abbildung 6.1: Bedingte Wahrscheinlichkeit (likelihood) der 13. Komponente des Inputvektors für Photonen (dunkel) und Pionen (hell).

6.2 Netzwerktypen

Im weiteren werden zwei unterschiedliche Netzwerke mit jeweils einer unterschiedlichen Anzahl von Neuronen untersucht. Das erste Netzwerk (Netz 1) besteht aus 28 Eingabeneuronen. Es wurden Varianten mit einer verdeckten Schicht aus 15 Neuronen oder keiner verdeckten Schicht untersucht. Das zweite Netzwerk (Netz 2) besteht aus 3 Eingabeneuronen, und die verdeckte Schicht bestand aus 5 Neuronen, wenn sie vorhanden war. Die Ausgabeschicht bestand aus zwei Ausgabeneuronen, eines für Photonen, eines für Nicht-Photonen. Der Vergleich zwischen Netz 1 und Netz 2 soll zeigen, ob die Informationen, die Netz 2 als Eingabe dienen, schon genügen um eine Klassifikation zu vollziehen. Die Netze ohne verdeckter Schicht dienen als Vergleichsnetze. In [OTT96] wurden verschiedene Netzwerk-Topologien bei einem ähnlichen Problem un-

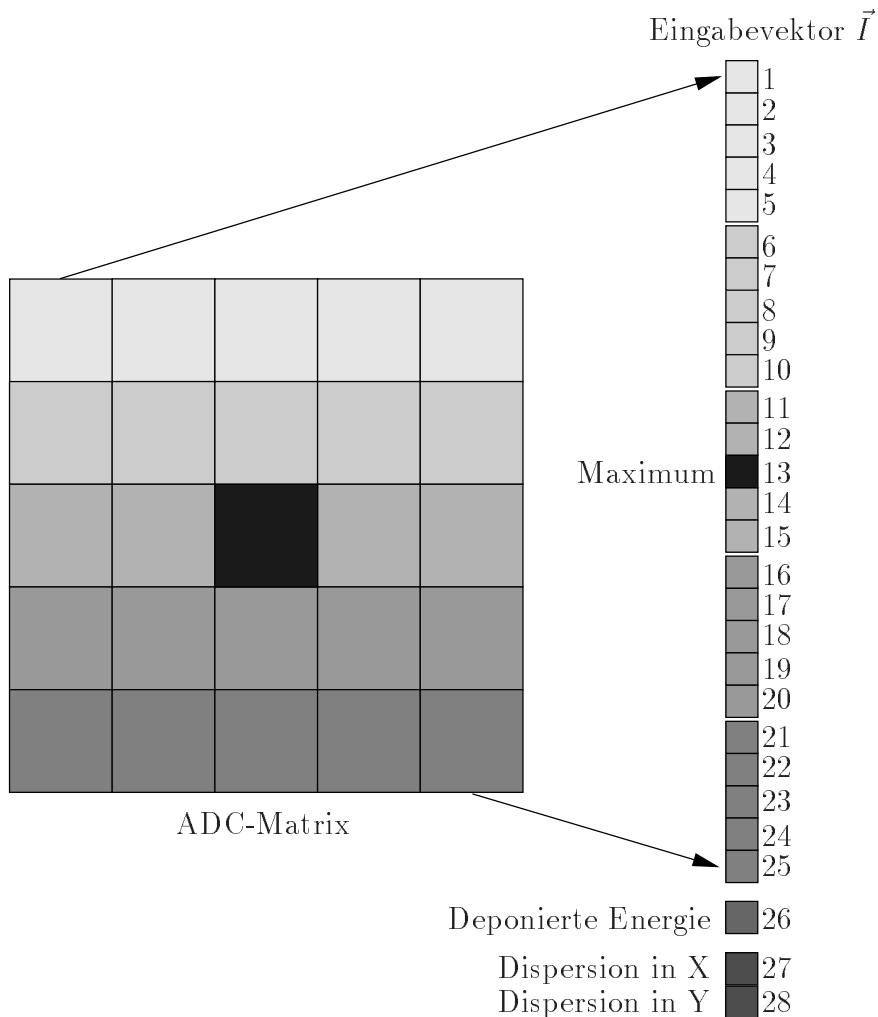


Abbildung 6.2: Zur Veranschaulichung der Verbung zwischen ADC-Matrix und Eingabevektors.

tersucht. Die dort vorgeschlagene Topologie von 15 verdeckten Neuronen hat sich auch hier als sehr vorteilhaft herausgestellt.

Bei Netz 1 werden die ADC-Werte der einzelnen Cluster, in einer 5×5 -Matrix um das Maximum, zur Bildung der ersten 25 Eingabewerte benutzt. Als weitere Eingabe kommt die deponierte Energie und die Dispersionswerte (laterale Ausdehnung) in X- und Y-Richtung hinzu (Siehe dazu Kapitel 7.1). Der Eingabevektor \vec{I} und die ADC-Matrix sind eindeutig über das Maximum des Clusters, das in der Mitte der ADC-Matrix liegt, verbunden (Siehe dazu Abbildung 6.2). Die ADC-Werte im Eingabevektor wurden durch einviertel

der Cluster-Summe¹ geteilt. Damit sollen die Eingaben auf Werte zwischen 0,0 und 40, normiert werden. Eine Normierung auf Werte zwischen 0,0 und 1,0 hat sich als unpraktikabel erwiesen, da die Eingabewerte dann oft zu kleine Werte annehmen und infolgedessen nicht mehr verarbeitet werden können.

Der Eingabevektor für das Netz 2 wird nur aus der deponierten Energie und den beiden Werten für die Dispersion gebildet.

Zur Klassifikation von zwei Klassen, elektromagnetisch (Photonen) und hadronisch (Pionen), wird im Prinzip nur ein Ausgabeneuron benötigt. Mit einem weiteren Neuron in der Ausgabeschicht ist dann auch eine Wahrscheinlichkeitsaussage über die Zugehörigkeit zu einer der beiden Klassen möglich (*1 aus M Kodierung*) [BIS95] [LEC91] [LIP91]. Eine *1 aus M Kodierung* bedeutet, daß für jede Klasse ein Ausgabeneuron zur Verfügung steht. Die Wahrscheinlichkeit, daß der Eingabevektor \vec{I} von einem Photon stammt, wird mit der Ausgabe $O_1(\vec{I})$ angegeben. $O_2(\vec{I})$ steht für die Wahrscheinlichkeit, daß der Eingabevektor \vec{I} von einem Pion kam. Somit liegen die Zielvektoren für die beiden Klassen fest, $\vec{T} = (1,0)$ für Photonen und $\vec{T} = (0,1)$ für Pionen. Mit der *1 aus M Kodierung* läßt sich die Wahrscheinlichkeitsaussage der Netze überprüfen, da sich beide Ausgabeneuronen zu 1 addieren sollten.

6.3 Training der beiden Netzwerke

Das Training der Netzwerke erfolgte mit einem Trainingsdatensatz und einem Testdatensatz. In Tabelle 6.1 ist die Zusammensetzung der einzelnen Datensätze aufgeführt. Die unterschiedlichen Datensätze wurden mit dem Monte-Carlo-Programm [GEANT] erstellt. Mit GEANT lassen sich ganze Detektoraufbauten beschreiben und deren physikalischen Prozesse simulieren. Den gesamten LEDA-Detektor mit seinen 10080 Bleiglasmodulen zu berechnen ist zu aufwendig und für diese Aufgabenstellung auch nicht nötig. Ein Block aus 9 Supermodulen, was 216 Bleiglasmodulen entspricht, genügt. Die simulierten Teilchen werden in die Mitte des Blokes mit Energien zwischen 0,2 und 10,0 GeV eingeschossen. Der Einschußort wird mit einer Gaußverteilung verschmiert. Die simulierten ADC-Werte werden anschließend ebenfalls verschmiert, was dem Grundrauschen des Analog-Digital-Converters um 1 bis 2 Kanäle entspricht. Eine ausführliche Beschreibung des Simulationsablaufes und dessen Anpassung an das reale Experiment findet sich in [BUS97].

¹Cluster-Summe: Summe der ADC-Werte aller zum Cluster gehörenden Module.

Zweck	Anzahl an γ	Anzahl an π^+	Winkel θ
Training	9991	6793	0°
Test	2307	896	0°
Validation	12298	7689	0°
Training	5176	3478	4°
Test	2036	791	4°
Validation	7212	4269	4°
Training	4890	3356	8°
Test	1735	673	8°
Validation	6625	3073	8°

Tabelle 6.1: Beim Training benutzte Datensätze. Die Einschußenergie der Teilchen liegt zwischen 0,2 und 10,0 GeV in einer dem realen Experiment angepaßten Verteilung.

In diesem Kapitel werden die unterschiedlichen Einschußwinkel noch einzeln betrachtet. In Kapitel 7.3 werden die beiden neuronalen Netze, um mit den realen Daten besser arbeiten zu können, mit allen Datensätzen trainiert. Mit den so trainierten Netzen sollen später Daten vom WA98-Experiment untersucht werden.

6.3.1 Netzwerk mit 28 Eingabeneuronen

Bei diesem Netzwerk werden die ADC-Werte als Eingabe benutzt. Weil die meisten physikalischen Prozesse im Mittel rotationssymmetrisch sind, kann durch Symmetrietransformation die Anzahl der Muster erhöht werden. Dadurch kann das Training verbessert werden. Hier wurden sieben Symmetrietransformationen verwendet (Siehe dazu Abbildung 6.3):

- Spiegelung an der
 - Abzisse (b)
 - Ordinate (c)
 - Diagonalen im 4. Quadranten (d)
 - Diagonalen im 1. Quadranten (e)

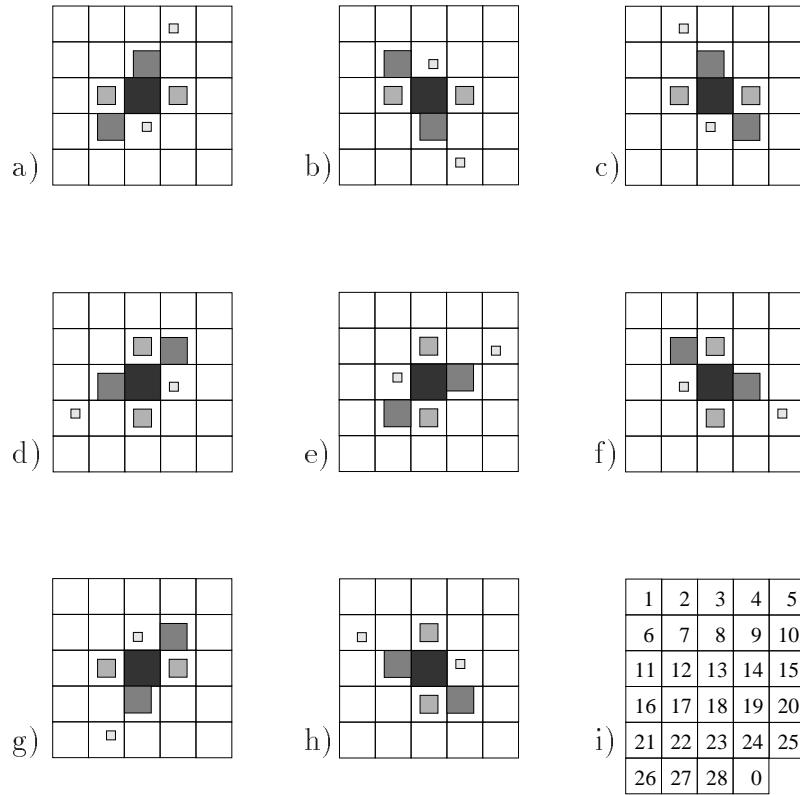


Abbildung 6.3: Beispiel für die Symmetrietransformationen. In a) ist die Originalmatrix zu sehen und in b) bis h) die verschiedenen Transformationen. Den kompletten Eingabevektor gibt Matrix i) noch einmal wieder. Die Vektorelemente I_{26} bis I_{28} enthalten die Informationen über die deponierte Energie und die Dispersion in x- und y-Richtung. I_0 wird konstant auf 1,0 gesetzt und dient der verdeckten Schicht als Bias (Vergleiche dazu Abbildung 6.2).

- Drehung um 90° (f)
- Drehung um 180° (g)
- Drehung um 270° (h)

Das einfache Muster \vec{I} und seine sieben Symmetrietransformationen $\mathcal{T}(\vec{I}_i)$ ($i = 1, 7$) werden dem neuronalem Netz beim Training präsentiert. Durch Hinweise auf bestimmte Eigenschaften der Eingabemuster, wie der Rotationsinvarianz, kann das Trainingsergebnis verbessert werden.

In Abbildung 6.4 und 6.5 sind zu verschiedenen Zeitpunkten die beiden Ausgabeneuronen des Netzes 1 gegeneinander aufgetragen, im ersten Bild vor dem

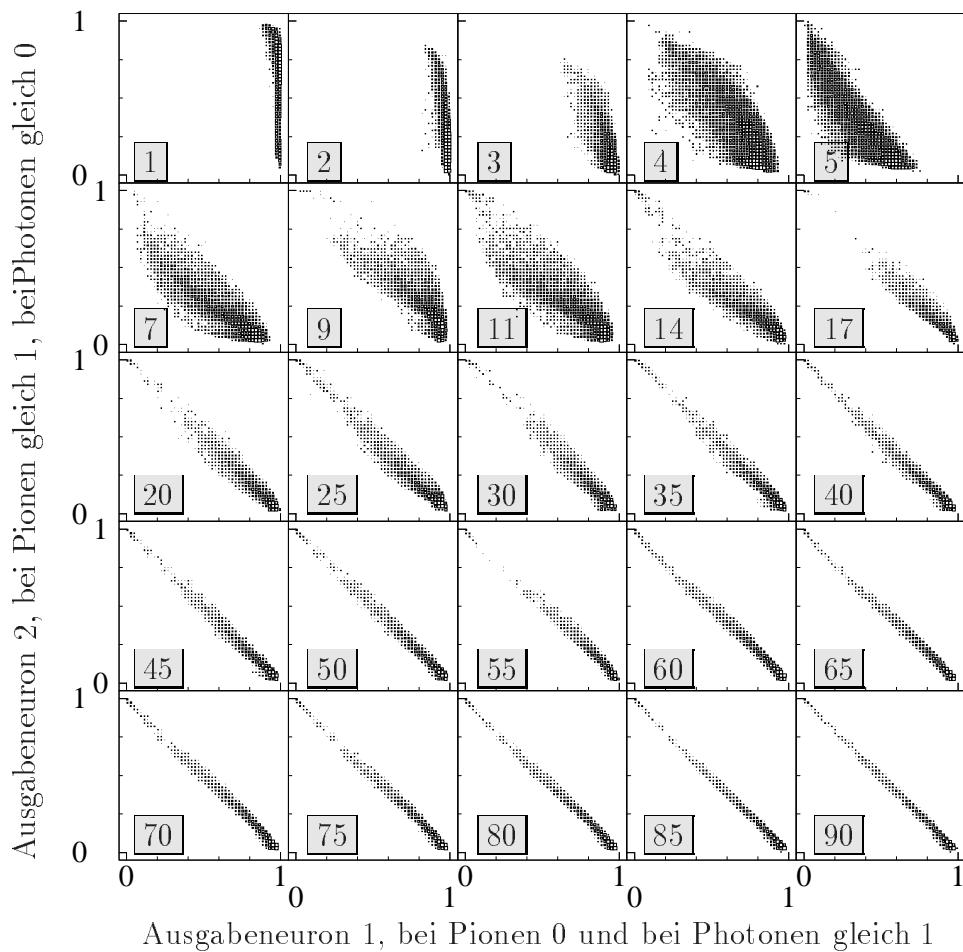


Abbildung 6.4: Ausgabe der beiden Ausgabeneuronen während des Trainings mit Photonen von Netz 1 mit verdeckter Schicht. In der linken unteren Ecke ist jeweils die Nummer des momentanen Lernschrittes gezeigt.

ersten Lernschritt und im letzten vor dem 90-ten Lernschritt. Deutlich ist der Lerneffekt zu erkennen. Die Ausgaben werden mit jedem Lernschritt mehr zu den gewünschten Sollwerten hingezogen, der bei Photonen in der rechten unteren Ecke liegt ($x = 1,0, y = 0,0$) und für Pionen in der oberen linken Ecke liegt ($x = 0,0, y = 1,0$).

Die Anzahl der verdeckten Neuronen hängt auch mit dem Lernerfolg des Netzes zusammen. In Abbildung 6.6 ist der durchschnittliche Fehler pro Ausgabeneuron während des Trainings für das Netz 1 mit und ohne verdeckter Schicht dargestellt. Die Kreise stammen von den Daten die am Training beteiligt waren und die Quadrate von den Testdaten.

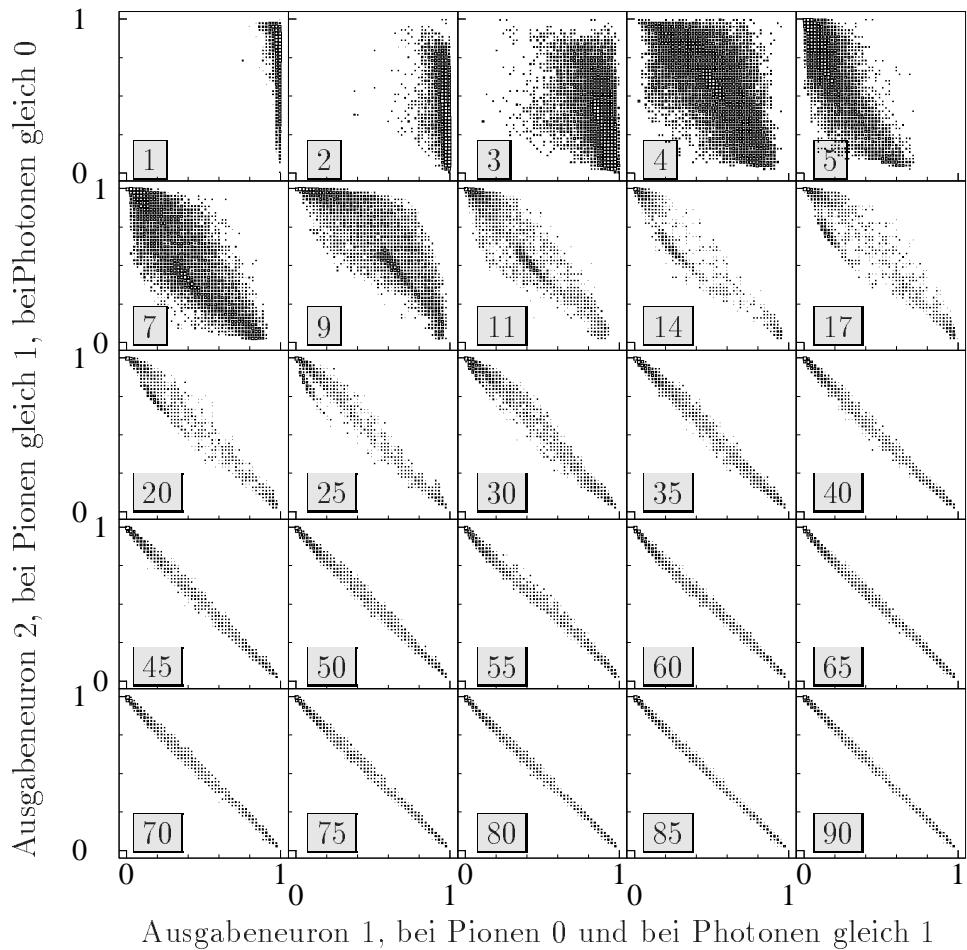


Abbildung 6.5: Ausgabe der beiden Ausgabeneuronen während des Trainings mit geladenen Pionen von Netz 1 mit verdeckter Schicht. In der linken unteren Ecke ist jeweils die Nummer des momentanen Lernschrittes gezeigt.

Bei dem Netz ohne verdeckter Schicht ist für einen Einschußwinkel von $\theta = 0^\circ$ und 4° ein leicht höherer Trainingsfehler zu beobachten, als bei dem Netzwerk mit einer verdeckten Schicht von 15 Neuronen. Bei einem Winkel von $\theta = 8^\circ$ ist der Unterschied nicht mehr so groß und auch der Fehler ist insgesamt deutlich geringer. Das hängt mit dem Einschußwinkel zusammen. Die Pionen fangen viel später an im Bleiglas aufzuschauern und dadurch dringen sie viel tiefer in den Detektor ein. Treffen sie nun in einem schrägen Winkel auf den Detektor, so bildet sich auch der Schauer in eine bestimmte Richtung aus. Bei Photonen, die schon sehr früh nach dem Eindringen in den Detektor anfangen einen Schauer zu bilden, ist der Effekt nicht so deutlich. In Abbildung 7.1 ist deutlich die

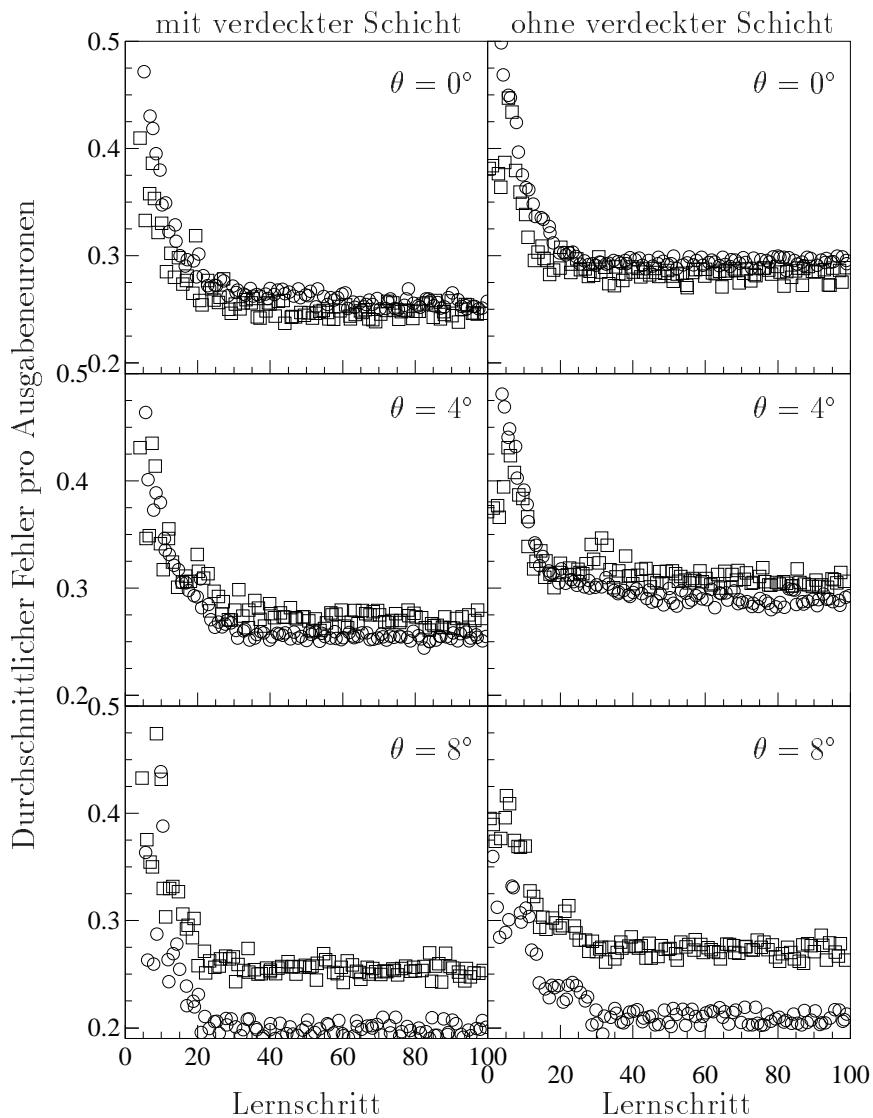


Abbildung 6.6: Entwicklung des Ausgabefehlers bei verschiedenen Einschußwinkeln für Netz 1, links mit verdeckter und rechts ohne verdeckter Schicht.

vergrößerte Schauerbreite der Pionen im Verhältnis zu den Photonen, die sich in einer größeren Dispersion äußert, zu erkennen.

Weiterhin wurde überprüft, ob die Gesamtwahrscheinlichkeit der beiden Ausgabeneuronen sich zu 1,0 addiert. Tabelle 6.2 listet die Mittelwerte der summierten Ausgaben (Σ) mit deren Standardabweichung (σ) auf. Der Mittelwert wird wie folgt ermittelt:

		neuronales Netzwerk			
		Netz 1 mit verdeckter Schicht	Netz 1 ohne verdeckter Schicht	Netz 2 mit verdeckter Schicht	Netz 2 ohne verdeckter Schicht
Einfallswinkel $\theta = 0^\circ$					
Σ	0,9976		1,002	1,006	1,002
σ	$0,167 \cdot 10^{-1}$		$0,1388 \cdot 10^{-1}$	$0,125 \cdot 10^{-1}$	$0,9806 \cdot 10^{-2}$
Einfallswinkel $\theta = 4^\circ$					
Σ	0,9995		0,9986	1,000	0,9971
σ	$0,168 \cdot 10^{-1}$		$0,1338 \cdot 10^{-1}$	$0,139 \cdot 10^{-1}$	$0,1152 \cdot 10^{-1}$
Einfallswinkel $\theta = 8^\circ$					
Σ	1,002		0,9986	1,008	1,001
σ	$0,2138 \cdot 10^{-1}$		$0,1863 \cdot 10^{-1}$	$0,2755 \cdot 10^{-1}$	$0,1001 \cdot 10^{-1}$

Tabelle 6.2: Gesamtwahrscheinlichkeit für die verschiedenen Einschußwinkel und verschiedenen Netzwerktypen mit 28 und 3 Eingabeneuronen, mit und ohne verdeckter Schicht.

$$\Sigma = \frac{1}{n} \cdot \sum_{p=1,n} O_1^p + O_2^p ,$$

(n : Anzahl der Muster, $O_i(i = 1, 2)$: Wert der Ausgabeneuronen).

Die Summation zu 1,0 ist von allen Netzen gut erfüllt. Diese Eigenschaft ist typisch für mehrschichtige Feed-Forward Netze [LIP91]. Als Ursache ist meistens eine Antisymmetrie in den Gewichten zu finden. Die Gewichte von der verdeckten zur Ausgabeschicht sind folgendermaßen verknüpft:

$$w_{1k} = -w_{2k} ,$$

der Index k steht für die einzelnen Neuronen in der verdeckten Schicht und den Bias der Ausgabeneuronen. Für das Einschicht-Perzeptron gilt:

$$w_{1j} = -w_{2j} ,$$

		Klassenzugehörigkeit			
		Π	Γ		
geschätzte Klassenzugehörigkeit $\hat{\Gamma}$	$\hat{\Pi}$	N_{Π}^{Π}	N_{Γ}^{Π}	\hat{N}_{Π}	geschätzte Teilchenzahlen
	$\hat{\Gamma}$	N_{Π}^{Γ}	N_{Γ}^{Γ}	\hat{N}_{Γ}	Teilchenzahlen
		N_{Π}	N_{Γ}		
		Teilchenzahlen			

Tabelle 6.3: Zusammenhang der wahren und geschätzten Teilchenzahlen, bei den vier möglichen Fällen der Klassifikation.

wobei j für die Eingabeneuronen und den Bias steht.

Für das Netz ohne verdeckter Schicht ist ein im Schnitt größerer Trainingsfehler zu beobachten. Dadurch wird eine etwas schlechtere Klassifikationsfähigkeit erwartet.

Ein weiteres Entscheidungskriterium für die Güte eines neuronalen Netzes ist die Fähigkeit, die Teilchenzahlen korrekt zu *schätzen*. Dazu werden die vier möglichen Fälle für die Klassifizierung betrachtet, die in Tabelle 6.3 dargestellt sind.

Der obere Index gibt die geschätzte Klassenzugehörigkeit an, der untere die wahre Klasse. N_{Π}^{Π} ist also die Anzahl der echten Pionen, die auch als Pionen klassifiziert werden, N_{Γ}^{Π} ist die Anzahl der echten Photonen, die als Pionen falsch klassifiziert werden. Ebenso sind N_{Γ}^{Γ} , die als Photonen richtig klassifizierten Photonen und N_{Π}^{Γ} die als Photonen falsch klassifizierten Pionen. Wird über einen der Indizes summiert, um z.B. die wahre Teilchenzahl zu erhalten, fehlt der Index. N_{Γ} gibt die Gesamtzahl der Photonen an, die richtig und falsch klassifiziert wurden, $N_{\Gamma} = N_{\Gamma}^{\Gamma} + N_{\Gamma}^{\Pi}$. Die Gesamtzahl der Teilchen, die als Photon klassifiziert wurden, ist $N^{\Gamma} = N_{\Gamma}^{\Gamma} + N_{\Pi}^{\Gamma}$. Die Gesamtteilchenzahl wird mit $N = N_{\Gamma} + N_{\Pi}$ bestimmt. Die geschätzten Werte werden mit einem $\hat{}$ markiert. Je besser die Ausgabe des Netzes mit den wahren Teilchenzahlen übereinstimmt, desto besser approximiert das neuronale Netz die wahren a-posteriori-Wahrscheinlichkeiten. Die Ergebnisse sind in Tabelle 6.4 aufgeführt. Bei allen Netzen ist die Schätzung der Teilchenzahl meistens gut erfüllt, nur bei dem Netz 2 gibt es einige Werte, die stark von der wahren Teilchenzahl abweichen.

	Netz 1 mit verdeckter Schicht	Netz 1 ohne verdeckter Schicht	Netz 2 mit verdeckter Schicht	Netz 2 ohne verdeckter Schicht	Untersuchte Teilchen
Winkel $\theta = 0^\circ$					
\hat{N}_Γ	6175,0	6185,1	6216,9	6181,6	$N_\Gamma = 6181$
\hat{N}_Π	2164,7	2169,9	2174,2	2166,9	$N_\Pi = 2168$
Winkel $\theta = 4^\circ$					
\hat{N}_Γ	4485,6	4467,0	4483,4	4443,4	$N_\Gamma = 4476$
\hat{N}_Π	1419,9	1425,8	1422,4	1423,2	$N_\Pi = 1423$
Winkel $\theta = 8^\circ$					
\hat{N}_Γ	4396,2	4375,9	4411,6	4381,2	$N_\Gamma = 4381$
\hat{N}_Π	1201,7	1205,7	1213,2	1201,4	$N_\Pi = 1204$

Tabelle 6.4: Auflistung der geschätzten Teilchenzahlen für die verschiedenen Netzwerke.

6.3.2 Netzwerk mit 3 Eingabeneuronen

Bei diesem Netzwerk sind keine Symmetrietransformationen gemacht worden, da die ADC-Werte nicht als Eingabe dienen. In Abbildung 6.7 ist wieder die Entwicklung des Trainingsfehlers für verschiedene verdeckte Schichten dargestellt. Die Summation der beiden Ausgabeneuronen ist ebenfalls in Tabelle 6.2 zu finden.

Wiederum ist zu erkennen, daß die Netze ohne verdeckter Schicht einen höheren Fehler aufweisen. Auch ist wieder der Zusammenhang mit dem Einfallswinkel θ zu erkennen.

Die Schätzungen der Teilchenzahlen und die Entwicklungen der Trainingsfehler lassen den Schluß zu, daß die Netze ohne verdeckter Schicht sich nicht für eine Anwendung eignen. Die Untersuchung der Güte der verschiedenen neuronalen Netze wird in Kapitel 7.2 weiter fortgeführt.

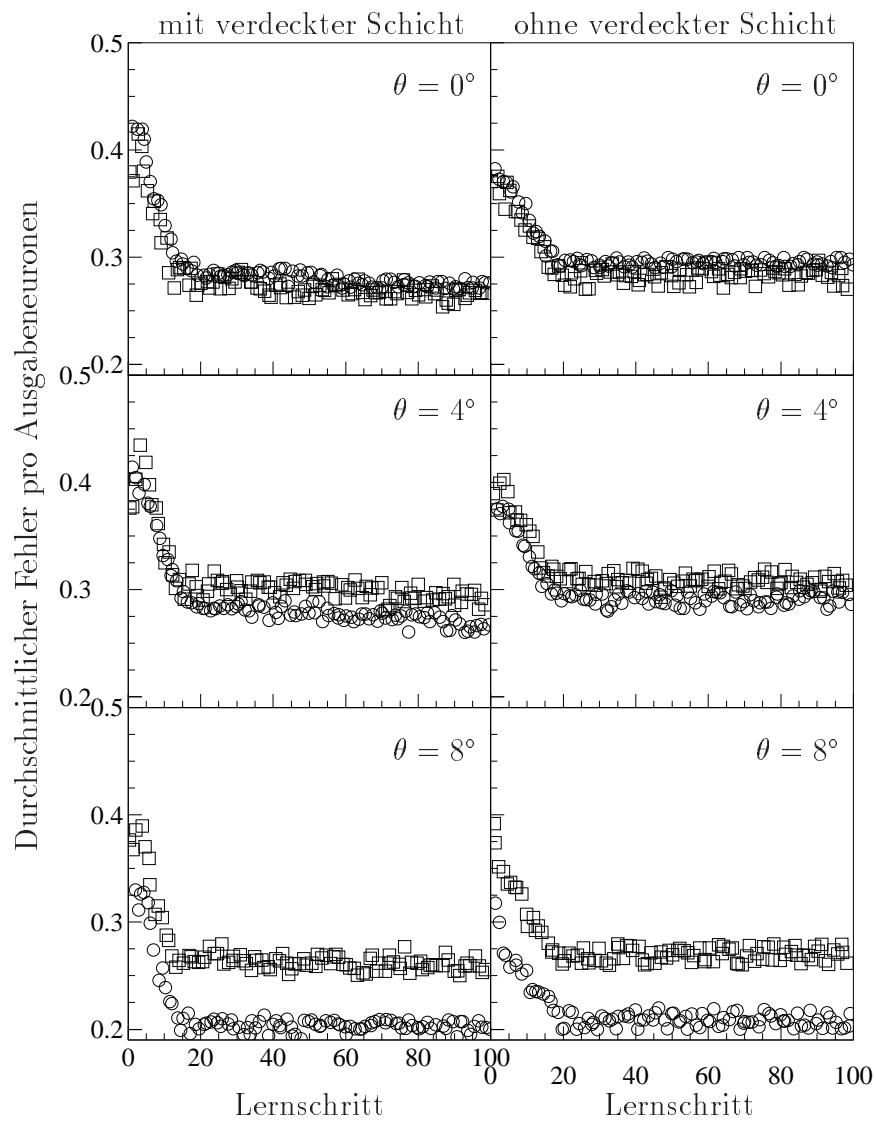


Abbildung 6.7: Entwicklung des Ausgabefehlers bei verschiedenen Einschußwinkeln für Netz 2, links mit verdeckter und rechts ohne verdeckter Schicht.

7 Teilchenklassifikation

In diesem Kapitel geht es um den Vergleich der Klassifikationseigenschaften zwischen den neuronalen Netzen und der Dispersionsmethode. Zunächst wird die Dispersionsmethode erläutert und anschließend werden die beiden Methoden anhand von simulierten Datensätzen verglichen. Darauf folgt ein Test mit einem Teil der Daten des WA98-Experiments, um einen ersten Eindruck über die Unterschiede beider Methoden zu erhalten.

7.1 Dispersionsmethode

Bisher wurde ein Verfahren benutzt, daß eine möglichst gute Trennung zwischen elektromagnetischen und hadronischen Schauern bei geringem Aufwand bietet. Der Unterschied zwischen den beiden Schauerarten liegt größerenteils in deren lateralen Ausdehnung (Siehe dazu Kapitel 3.2.1). Um ein vom Koordinatensystem unabhängiges Klassifikationsmerkmal zu erlangen, wird die Dispersion benutzt. Die Dispersion ist ein Maß für die laterale Ausdehnung des Schauers, das zentrale zweite Moment [BER92]:

$$D_x = \frac{\sum_i A_i \cdot x_i^2}{\sum_i A_i} - \left(\frac{\sum_i A_i \cdot x_i}{\sum_i A_i} \right)^2$$

(x_i : Relativkoordinate des i -ten Moduls bezüglich des Einschußmoduls ($x = 0$) in Moduleinheiten, A_i : ADC-Wert des i -ten Moduls).

Die Summation erfolgt über alle Module i eines Clusters. Für die y-Koordinate gilt der gleiche Zusammenhang (Austausch von $x \leftrightarrow y$). In Abbildung 7.1 ist die Dispersion in Abhängigkeit des ersten Momentes anhand von Photonen und Pionen bei verschiedenen Einschußwinkeln dargestellt. Aufgrund der vielen Einträge bei $D_x = 0,0$ wurde die Fläche proportional zum Logarithmus der Teilchenzahl gewählt. Deutlich ist die wesentlich größere Dispersion der Pionen zu erkennen. Bei dieser Darstellung kann keine einfache lineare Schwelle zur Klassifikation gesetzt werden. Man müßte eine parabolische Schwelle ansetzen. Durch eine Korrektur des Dispersionswertes D_x wird die Nutzung einer linearen Schwelle zur Unterscheidung möglich:

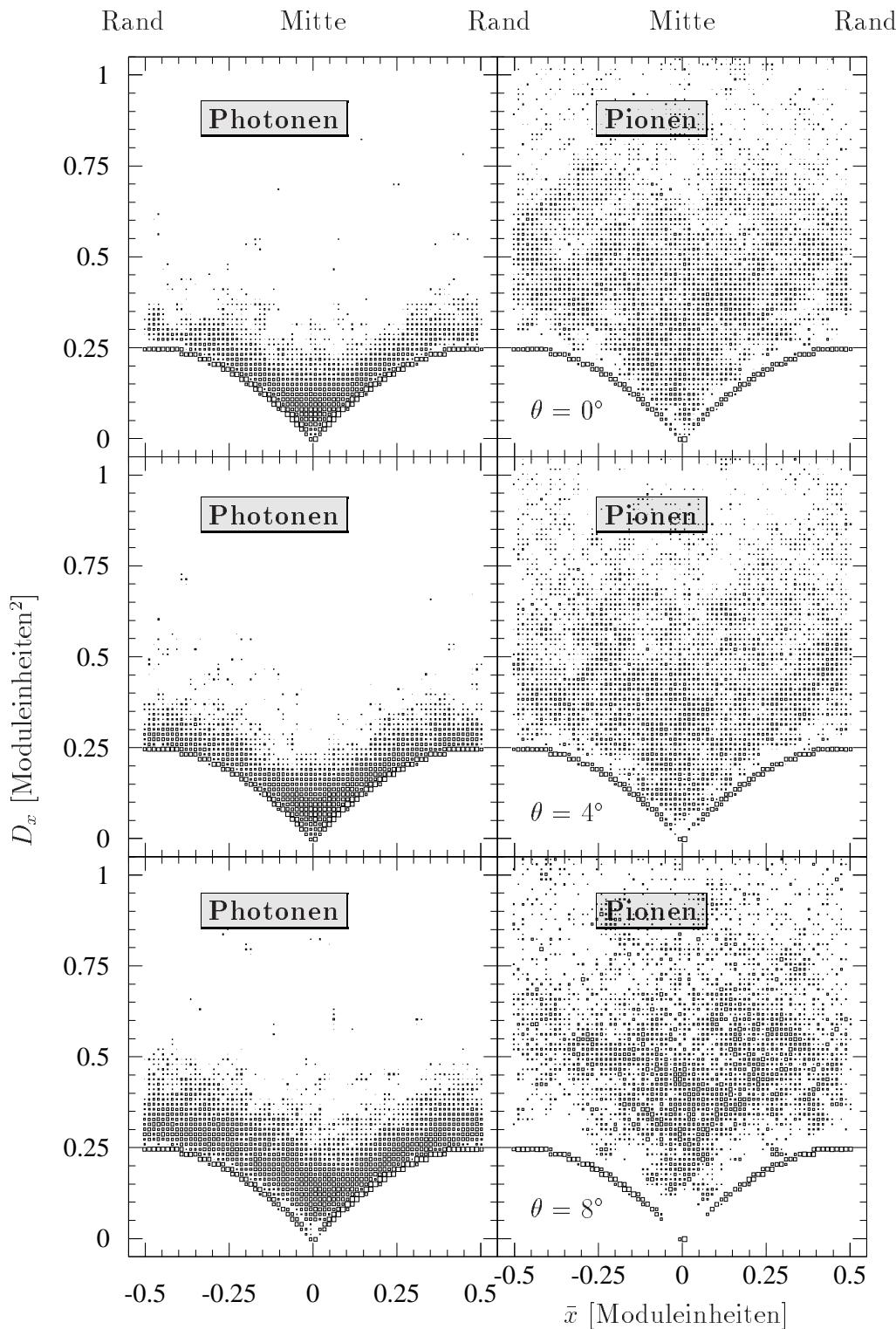


Abbildung 7.1: Abhängigkeit der Dispersion D_x vom Schwerpunkt \bar{x} für simulierte Photonen (links) und Pionen (rechts) der Trainingsmenge bei verschiedenen Einschußwinkeln.

$$D_x^{corr} = D_x - D_x^{min},$$

$$D_x^{min} = 0,25 - \left(\frac{\sum_i A_i \cdot x_i}{\sum_i A_i} \right)^2 - \left| \left(\frac{\sum_i A_i \cdot x_i}{\sum_i A_i} \right) \right|.$$

Für Teilchen, die ihre Energie in nur zwei oder weniger Modulen deponieren gilt:

$$D_x = D_x^{min}.$$

Hier wird deutlich, daß die Dispersionsmethode bei Ereignissen, die nur in einem oder zwei Modulen ADC-Werte liefern, keine Aussagekraft besitzt.

Zur Klassifikation wird mit:

$$D_{max}^{corr} = Max(D_x^{corr}, D_y^{corr})$$

eine Variable definiert, die durch eine Schwelle (*cut*) zur Unterscheidung benutzt wird. In Abbildung 7.2 sind die korrigierten maximalen Dispersionswerte für Photonen (dunkel) und Pionen (hell) aufgetragen. Außerdem ist die zur Analyse benutzte Dispersionsschwelle von $D_{cut} = 0,267$ eingetragen. Alle Teilchen mit einer höheren Dispersion als D_{cut} werden verworfen.

Da hadronische Teilchen im wesentlichen minimal Ionisierend den Detektor durchdringen (MIP-Peak bei ca. 500MeV), wird eine Energieschwelle eingesetzt. Alle Teilchen, die eine Energie von unter $E_{cut} = 750$ MeV deponieren werden verworfen. Dabei verwirft man aber auch alle Photonen unterhalb der Energieschwelle. Abbildung 7.3 zeigt die Energieverteilung anhand von simulierten Photonen und Pionen.

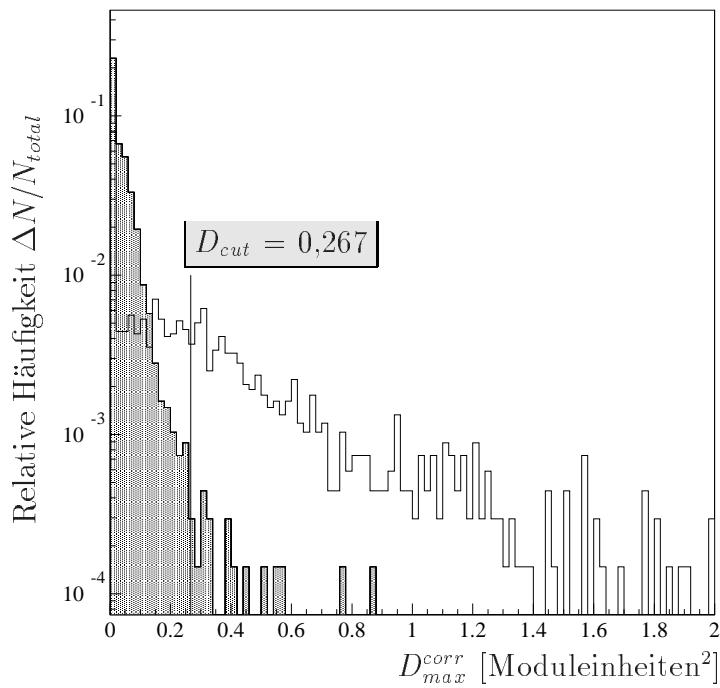


Abbildung 7.2: Maximale Korrigierte Dispersion für simulierte Photonen (dunkel) und Pionen (hell).

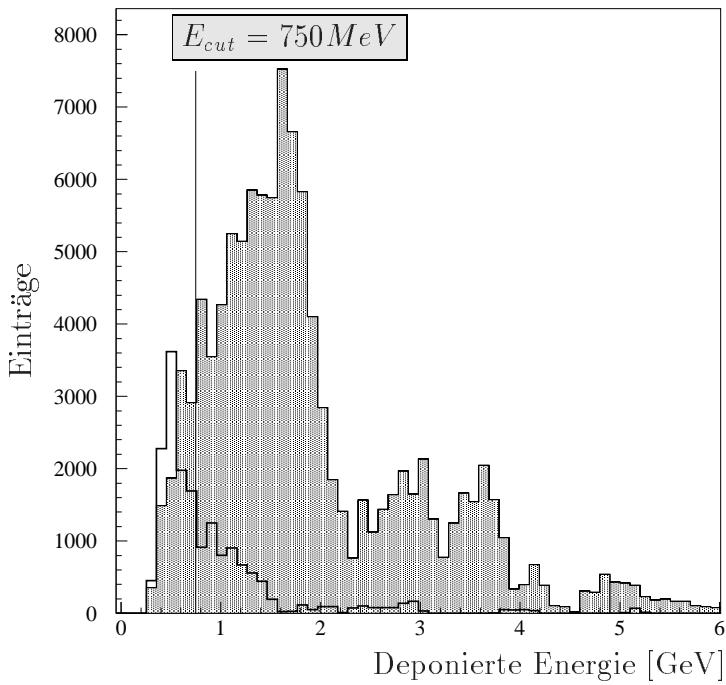


Abbildung 7.3: Deponierte Energie für Photonen (dunkel) und Pionen (hell).

7.2 Vergleich beider Methoden an simulierten Daten

Der Vergleich der Dispersionsmethode mit einem neuronalen Netz wird über die Veränderung der Klassifikationsparameter gemacht. Bei der Dispersionsmethode wird der Schwellwert D_{cut} , der zur Unterscheidung der Teilchen dient, verändert. Bei den neuronalen Netzen wird die Schwelle, die am ersten Ausgabeneuron zwischen den Klassen differenziert, verändert. Der Anteil an richtig klassifizierten Pionen, bei einem festgelegten Verlust an Photonen, gibt nun Auskunft über die Qualität einer Methode. Dieser Vergleich wird in den folgenden Diagrammen 7.4 und 7.5 dargestellt. In den Bildern sind die beiden Methoden bei verschiedenen Einfallswinkeln der simulierten Photonen und Pionen dargestellt. Links sind jeweils die Ergebnisse des neuronalen Netzes mit verdeckter Schicht aufgetragen, in der Mitte die des Netzes ohne verdeckter Schicht. Rechts ist die entsprechende Verteilung der Dispersionsmethode gezeigt. Die Untersuchung wurde in sechs verschiedene Energiebereiche unterteilt.

Betrachtet man nun den Anteil verworfener Pionen bei 1% Verlust an Photonen, so ist, besonders in dem Energiebereich zwischen 1,0 und 4,0 GeV, ein Vorteil der neuronalen Netze zu beobachten. Oberhalb von 4,0 GeV und unterhalb von 1,0 GeV sind die Ergebnisse beider Methoden gleich gut, wenn man die Netze mit verdeckter Schicht betrachtet. Die Netze ohne verdeckter Schicht haben besonders bei den höheren Energien große Probleme. Für die Netze ohne verdeckter Schicht ist das Problem zu komplex, sie können es mit der geringen Anzahl von Neuronen nicht lösen, das Netz lernt nur eine Teillösung des Problems. Das Ergebnis des unteren Energiebereiches, lässt sich durch die höhere Anzahl von Mustern, gegenüber des Bereiches mit höherer Energie, erklären.

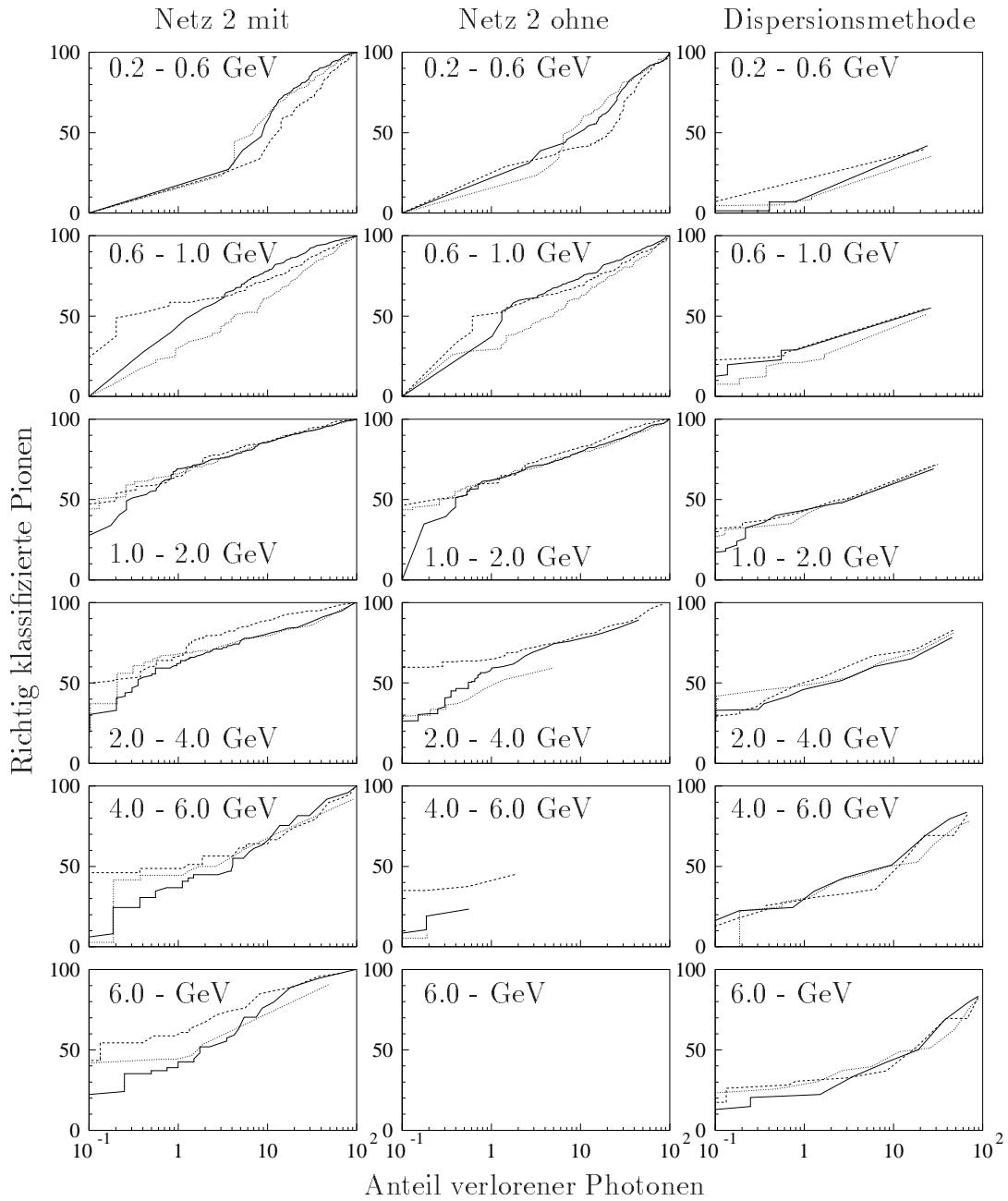


Abbildung 7.4: Vergleich der beiden Methoden, rechts Netz 1 mit verdeckter Schicht, in der mitte Netz 1 ohne verdeckter Schicht und links die Dispersionsmethode. Die unterschiedlichen Einfallswinkel werden durch die verschiedenen Linien dargestellt, die durchgezogene Linie steht für $\theta = 0^\circ$, gestrichelte Linie für $\theta = 4^\circ$ und die gepunktete Linie für $\theta = 8^\circ$.

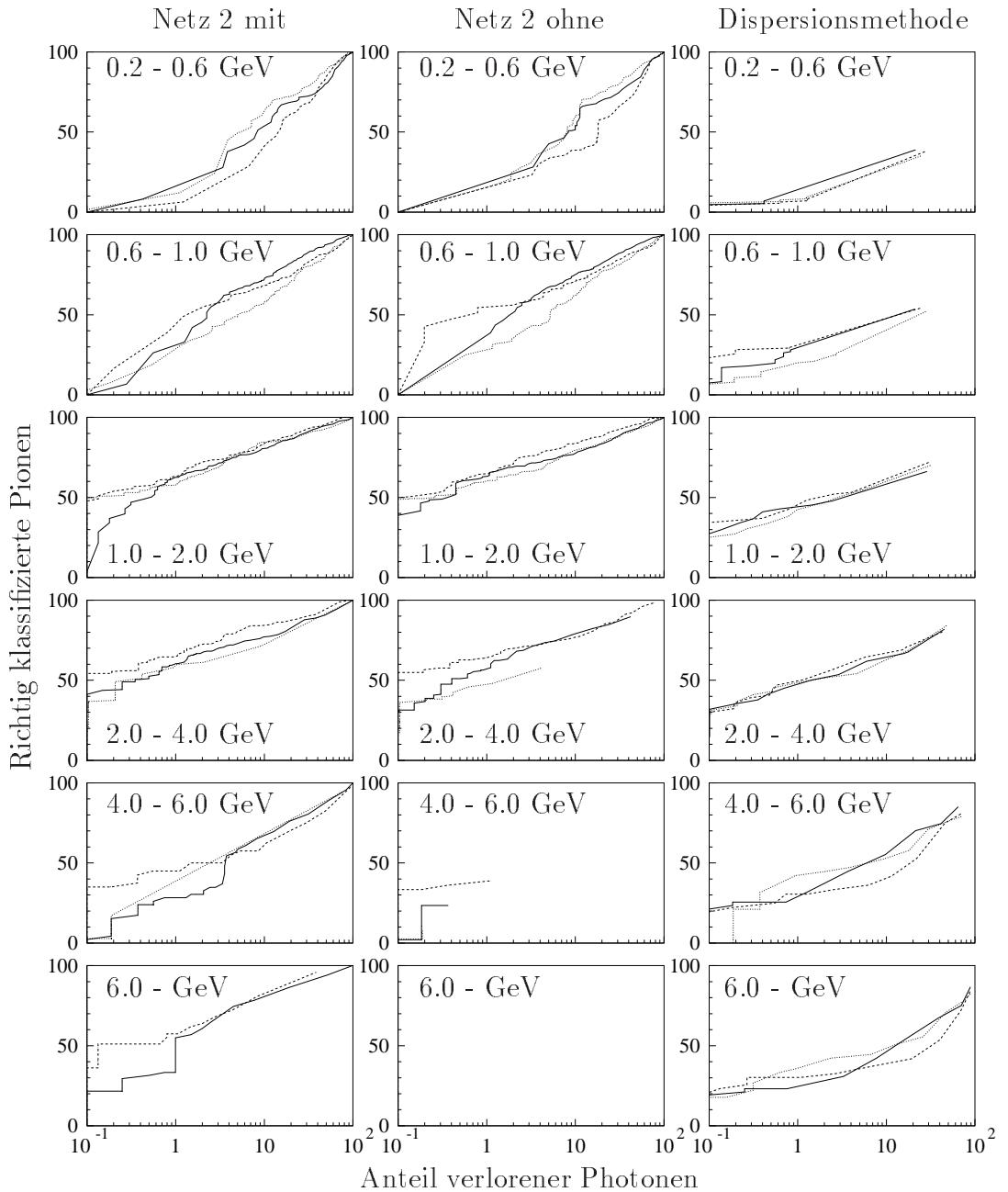


Abbildung 7.5: Vergleich der beiden Methoden, rechts Netz 2 mit verdeckter Schicht, in der Mitte Netz 2 ohne verdeckter Schicht und links die Dispersionsmethode. Die unterschiedlichen Einfallswinkel werden durch die verschiedenen Linien dargestellt, die durchgezogene Linie steht für $\theta = 0^\circ$, gestrichelte Linie für $\theta = 4^\circ$ und die gepunktete Linie für $\theta = 8^\circ$.

Energie- bereich in [GeV]	neurolale Netze Verworfene Pionen in [%]				Dispersions- methode Verworfene Pionen in [%]
	Netz 1		Netz 2		
	mit verdeckter Schicht	ohne verdeckter Schicht	mit verdeckter Schicht	ohne verdeckter Schicht	
Winkel $\theta = 0^\circ$					
0.2 - 0.6	13% ± 2	17% ± 2	17% ± 2	14% ± 5	9% ± 1
0.6 - 1.0	45% ± 2	35% ± 2	30% ± 2	36% ± 5	29% ± 1
1.0 - 2.0	67% ± 3	61% ± 3	62% ± 3	61% ± 5	47% ± 2
2.0 - 4.0	62% ± 3	57% ± 3	57% ± 3	54% ± 5	44% ± 3
4.0 - 6.0	45% ± 7	10% ± 7	27% ± 7	24% ± 5	30% ± 5
6.0 -	39% ± 6	0% ± 6	42% ± 6	0% ± 5	21% ± 7
Winkel $\theta = 4^\circ$					
0.2 - 0.6	13% ± 2	21% ± 2	5% ± 2	12% ± 5	22% ± 1
0.6 - 1.0	55% ± 2	50% ± 2	41% ± 2	54% ± 5	29% ± 1
1.0 - 2.0	60% ± 3	60% ± 3	62% ± 3	63% ± 5	46% ± 2
2.0 - 4.0	67% ± 3	63% ± 3	64% ± 3	64% ± 5	49% ± 3
4.0 - 6.0	50% ± 6	0% ± 7	45% ± 7	24% ± 5	30% ± 5
6.0 -	60% ± 7	0% ± 6	57% ± 6	0% ± 5	31% ± 7
Winkel $\theta = 8^\circ$					
0.2 - 0.6	13% ± 2	11% ± 2	10% ± 5	14% ± 5	7% ± 1
0.6 - 1.0	30% ± 2	27% ± 2	27% ± 5	28% ± 5	21% ± 1
1.0 - 2.0	65% ± 3	61% ± 3	55% ± 5	57% ± 5	38% ± 2
2.0 - 4.0	69% ± 3	43% ± 3	56% ± 5	45% ± 5	47% ± 3
4.0 - 6.0	44% ± 6	5% ± 6	37% ± 5	0% ± 5	31% ± 5
6.0 -	52% ± 7	0% ± 7	0% ± 5	0% ± 5	29% ± 7

Tabelle 7.1: Vergleich der verschiedenen Untersuchungen bei einem Verlust von 1% der Photonen.

In Tabelle 7.1 sind die Ergebnisse der einzelnen Untersuchungen aufgeführt, und in Abbildung 7.6 sind die Werte graphisch wiedergegeben. Zur besseren Unterscheidung sind die Symbole versetzt dargestellt.

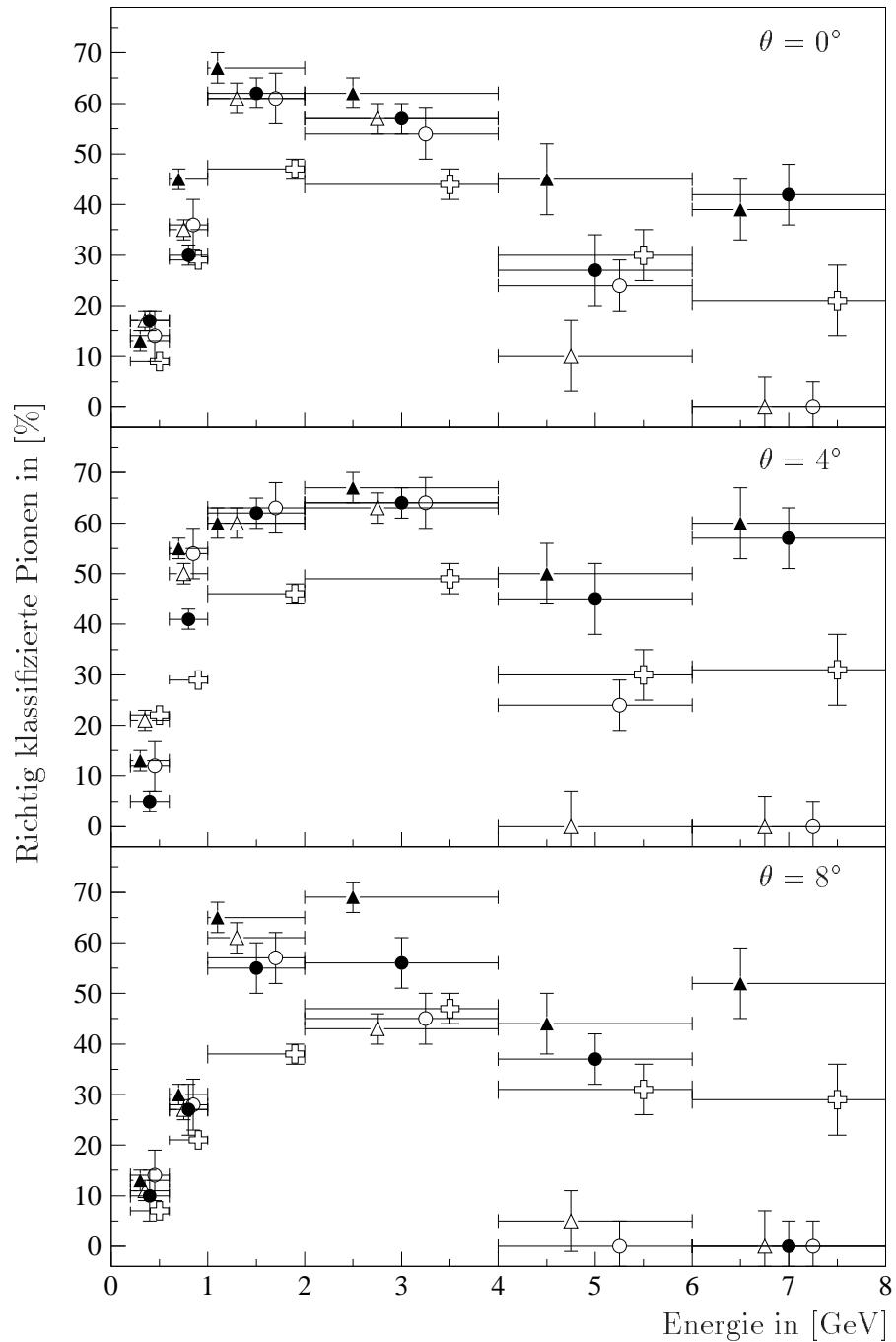


Abbildung 7.6: Abhangigkeit der richtig klassifizierten Pionen in [%] von der deponierten Energie eines Teilchens in [GeV]. Netz 1 stellt die Dreiecke und Netz 2 die Kreise dar. Die Netze mit verdeckter Schicht stellen die ausgefullten Symbole, die ohne verdeckter Schicht die leeren Symbole dar. Die Kreuze stehen fur die Dispersionsmethode.

7.3 Vergleich beider Methoden mit WA98 Daten

Um mit den Daten des WA98-Experiments arbeiten zu können wurden nur die beiden Netze mit verdeckter Schicht benutzt. Beim Training wurden alle Trainingsdatensätze mit den verschiedenen Einschusswinkeln genommen. Zunächst wurde mit den Datensätzen das Netz trainiert, um mit den Validationsdaten den Schwellparameter zu bestimmen, der zwischen Photonen und Pionen entscheidet. Dazu wurde in den verschiedenen Energiebereichen nach dem Schwellparameter gesucht, der zu einem Verlust von 1% der Photonen führt. Der unterste Energiebereich von 0,2 bis 0,6 GeV wurde nicht betrachtet, da in diesem Energiebereich kaum vernünftige Aussagen machbar sind. Es wurde ein durchschnittlicher Schwellparameter bestimmt, unter besonderer Berücksichtigung des Energiebereiches von 1,0 bis 4,0 GeV, da hier die besten Ergebnisse erzielt wurden.

Dies führt zu den Schwellparametern von $C_{nn1} = 0,43$ für Netz 1 für Netz 2 $C_{nn2} = 0,42$.

Der Vergleich zwischen den neuronalen Netzen und der Dispersionsmethode ist in Abbildung 7.7 dargestellt. Die Ergebnisse von Netz 1 sind in der linken Spalte der Bildes zu sehen. In der rechten Spalte sind die Ergebnisse von Netz 2 dargestellt. Die Werte der neuronalen Netze werden jeweils durch die durchgezogene Linie repräsentiert. Die gestrichelte Linie stammt jeweils von der Dispersionsmethode.

Mit diesen Schwellparametern wurde die nun folgende Untersuchung an realen WA98 Daten vorgenommen. Eine Aussage, ob die eine oder andere Methode besser funktioniert, lässt sich nun nicht machen. Man kann aber überprüfen, ob die beiden Methoden nicht zu völlig verschiedenen Ergebnissen führen.

In Abbildung 7.8 und 7.9 sind die Ausgaben der beiden Netzwerke gegen die Ausgabe der Dispersionsmethode aufgetragen. Die verschiedenen Energiebereiche werden durch die Schattierung verdeutlicht. Betrachtet man die komplette Verteilung, so klassifiziert die Dispersionsmethode die Daten in 75,4% Photonen und 24,6% Nicht-Photonen. Netz 1 kommt auf eine Verteilung von 68,8% Photonen und 31,2% Nicht-Photonen, Netz 2 auf 23,6% Photonen und 76,4% Nicht-Photonen.

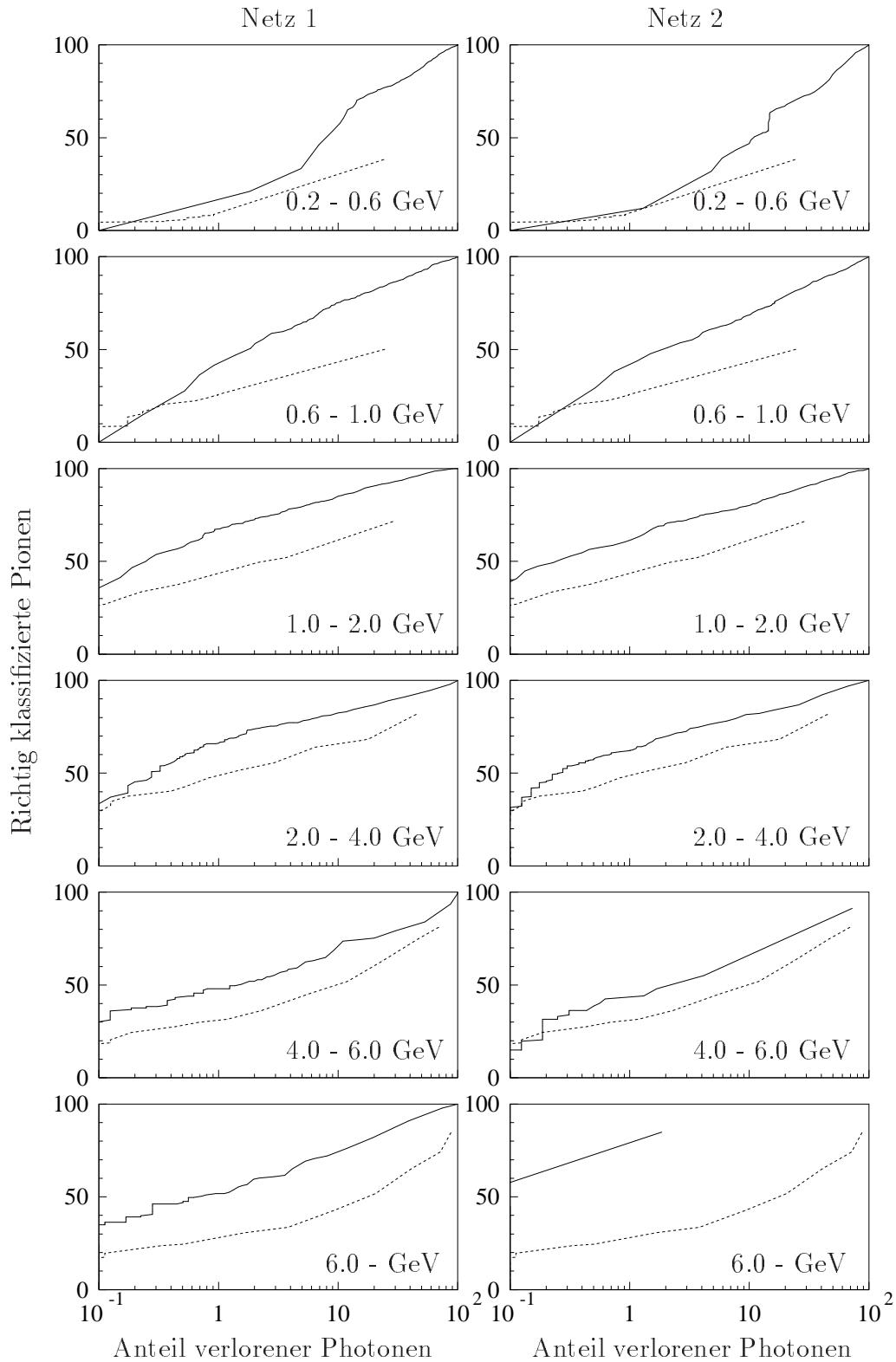


Abbildung 7.7: Anteil von richtig klassifizierten Pionen zu verlorenen Photonen bei verschiedenen Schwellwerten, für Netz 1 (rechts) und Netz 2 (links). Die neuronalen Netze werden durch die durchgezogenen Linien dargestellt, die Dispersionsmethode durch die gestrichelte Linie.

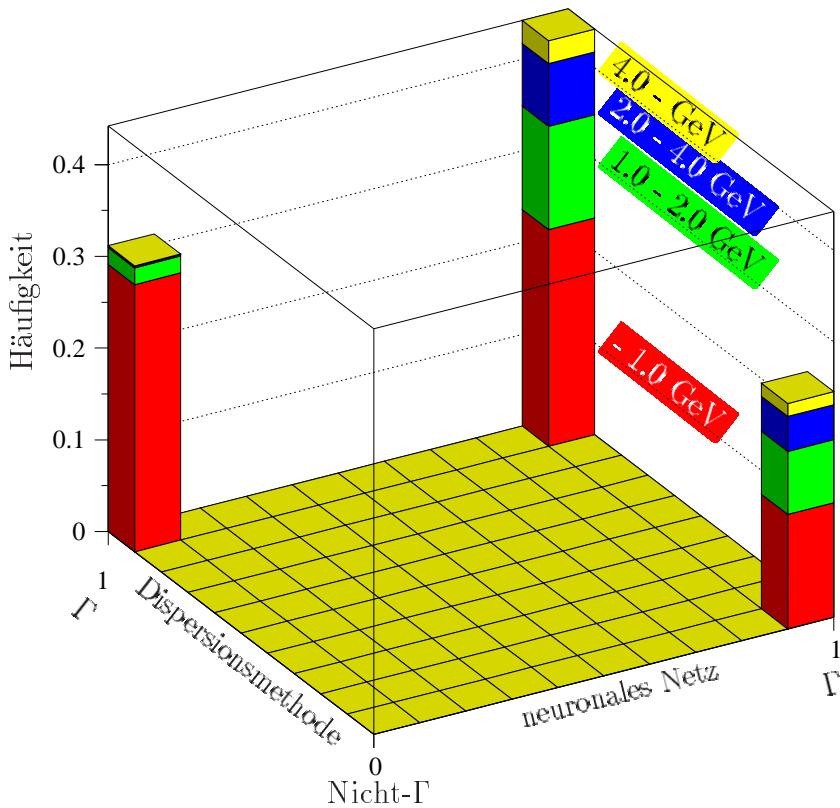


Abbildung 7.8: Direkter Vergleich beider Methoden mit Daten vom WA98-Experiment. Auf der X-Achse ist die Klassifizierung von Netz 1, auf der Y-Achse die der Dispersionsmethode aufgetragen und in Z-Richtung die Anzahl der Einträge. Die farbigen Schattierungen stellen die verschiedenen Energiebereiche dar.

Betrachtet man Bild 7.8 so sieht man, daß es keine Teilchen gibt, die die Dispersionsmethode und Netz 1 gleichzeitig als Nicht-Photon klassifizieren. Es gibt aber einen großen Anteil an Teilchen, die nach beiden Methoden als Photonen klassifiziert werden. Dieser Effekt muß noch weiter untersucht werden, eine mögliche Erklärung wäre folgende: In den ADC-Werten, mit denen das Netz trainiert wurde, stecken viel mehr Informationen als in den Werten der Dispersion alleine. Diese Informationen führen nun zu völlig unterschiedlichen Ergebnissen. Das wird besonders im unteren Energiebereich sichtbar.

In Bild 7.9 ist der Vergleich zwischen Netz 2 und der Dispersionsmethode zu sehen. Hier sieht man eine gute Übereinstimmung zwischen den beiden Methoden. Besonders auffällig ist in den unteren Energiebereichen bis 2,0 GeV,

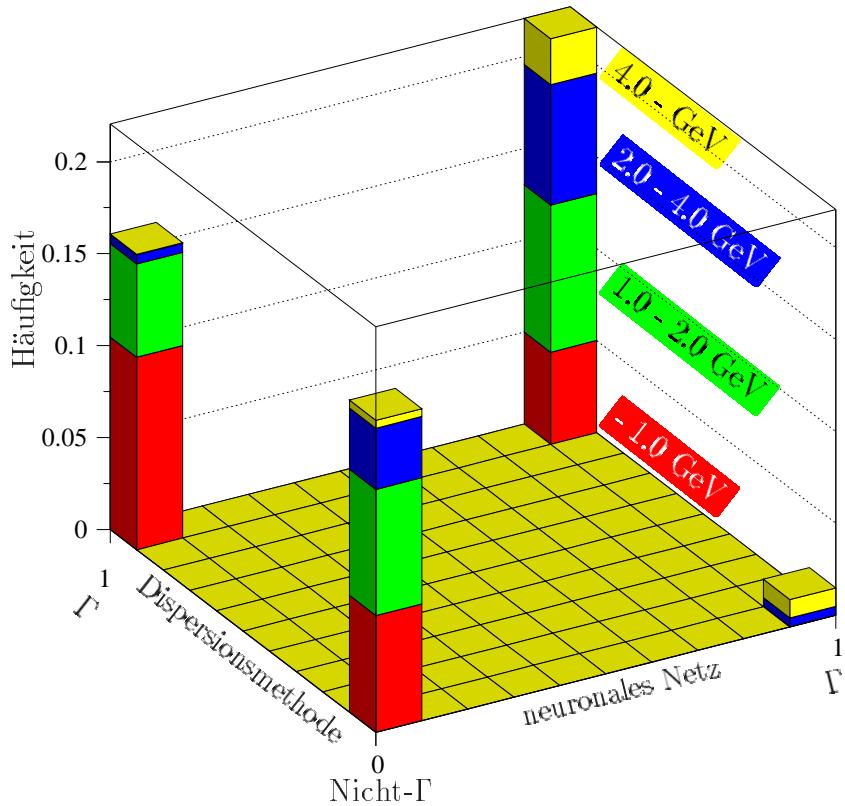


Abbildung 7.9: Direkter Vergleich beider Methoden mit Daten vom WA98-Experiment. Auf der X-Achse ist die Klassifizierung von Netz 2, auf der Y-Achse die der Dispersionsmethode aufgetragen und in Z-Richtung die Anzahl der Einträge. Die farbigen Schattierungen stellen die verschiedenen Energiebereiche dar.

daß das neuronale Netz viele Teilchen als Nicht-Photon klassifiziert, die die Dispersionsmethode als Photon erkennt. Bei diesem Netz sind, im Gegensatz zum vorherigen, die Unterschiede bei der Klassifikation sehr gravierend. Das kann daran liegen, daß die Informationen, die diesem Netz vorliegen, nicht ausreichen um eine eindeutige Klassifikation zu erzielen.

Die Parameter von Netz 1 sind im Anhang A.2 aufgeführt.

8 Zusammenfassung

Die bisher benutzte Dispersionsmethode zur Klassifizierung von elektromagnetischen und hadronischen Schauern im Bleiglaskalorimeter LEDA des WA98-Experimentes weist, besonders im unteren Energiebereich, einige Mängel auf. Ein neuronales Netz kann hier bessere Ergebnisse erzielen. Während einer Trainingsphase werden dem neuronalen Netz Muster von bekannten Teilchen vorgeführt, und anhand dessen werden die veränderbaren Parameter (Gewichte) im Netz so angepasst, daß ein späteres Präsentieren von ähnlichen Mustern eine Unterscheidung der Muster möglich macht. Diese Trainingsmuster werden vorzugsweise mit einem Monte-Carlo-Programm (GEANT) erstellt, dadurch sind alle Teilchenparameter bekannt. Die in dieser Arbeit verwandten Datensätze bestanden aus Photonen, für die elektromagnetisch wechselwirkenden Teilchen, und Pionen, für die hadronisch wechselwirkenden Teilchen. Die Energieverteilung der einzelnen Datensätze wurde den real zu erwarteten angepasst.

Zusätzlich wurden Datensätze mit verschiedenen Einschußwinkeln erstellt, um den Einfluß des Einfallswinkels auf die Klassifikationsmethoden zu testen. Ein Testdatensatz und ein Datensatz zur Überprüfung der Generalisierungsfähigkeit wurden außerdem noch erstellt. Mit diesen Datensätzen wurden die verschiedenen neuronalen Netzwerke und die Dispersionmethode untersucht und verglichen. Als Lernalgorithmus wurde der *Rprop*-Algorithmus in einem Batch-Verfahren angewendet. Die Ausgabe des Netzes kann als Wahrscheinlichkeitsaussage bezüglich der Zugehörigkeit zu einer der beiden Klassen angesehen werden. Um die Anzahl der Muster während des Trainings zu erhöhen und um den neuronalen Netzen einen Hinweis auf die Symmetrie der Reaktion zu geben, wurde während der Lernphase den Netzen 7 Symmetrietransformationen, zusätzlich zum Original, als Eingabe präsentiert.

Ab einer Energie von 0,6 GeV können die neuronalen Netze eingesetzt werden. Weiterhin scheint sowohl das Netzwerk mit 28 Eingabeneuronen und das mit 3 Eingabeneuronen zur Anwendung geeignet zu sein. Aber in den Bereichen höherer Energie hat das Netz mit 28 Eingabeneuronen bessere Ergebnisse erzielt, so daß die Anwendung dieses Netzwerktypes anzuraten ist. Vergleicht man die Dispersionsmethode mit den neuronalen Netzen direkt, so erhält man,

bei einem Verlust von 1% an Photonen, im Bereich von 0,6 bis 1,0 GeV für Dispersionsmethode 29%, für Netz 1 45% und Netz 2 30% an richtig klassifizierten Pionen. Im Bereich von 1,0 bis 2,0 GeV für die Dispersionsmethode 47%, 67% für Netz 1 und 63% für Netz 2. Und im Bereich von 2,0 bis 4,0 GeV, 44% für die Dispersionsmethode, 62% für Netz 1 und 57% für Netz 2 an richtig klassifizierten Pionen. Der Anteil der richtig erkannten Pionen liegt für die neuronalen Netze immer leicht höher, als der der Dispersionsmethode.

Die folgende Anwendung auf Daten aus dem WA98-Experiment zeigt, daß dieses neue Verfahren zur Teilchenklassifizierung verwendet werden kann. Die Klassifizierung von Netz 1 wich in der Gesamtheit nicht sonderlich von der der Dispersionsmethode ab. Netz 2 hingegen kommt auf eine völlig entgegengesetzte Verteilung. Betrachtet man jedoch die beiden Methoden genauer, so ist bei Netz 1 keine Übereinstimmung bei denen als Nicht-Photonen klassifizierten Teilchen zu sehen. Besonders deutlich ist, daß im unterem Energiebereich viele von der Dispersionsmethode als Photon erkannten Teilchen vom Netz 1 verworfen werden. Der Vergleich von Netz 2 mit der Dispersionsmethode sieht hingegen völlig anders aus. Hier ist sowohl bei den verworfenen Teilchen, als auch bei den als Photonen klassifizierten Teilchen, eine starke Übereinstimmung beider Methoden. Diese Unterschiede können auch auf einen möglichen systematischen Fehler der Dispersionsmethode hinweisen, der noch weiter untersucht werden muß. Eine weitere Erklärung wäre, daß es einen Unterschied für die beiden Teilchensorten in der Simulation gibt, den es in realen Experiment nicht gibt. Dieser Unterschied wurde vom Netz gelernt und führt nun zu falschen Ergebnissen.

Abschließend sei noch erwähnt, daß ein Hinzufügen von weiteren Informationen, insbesondere der Flugzeit, das Ergebnis noch verbessern kann. Mit dieser zusätzlichen Information wäre sogar eine Klassifizierung von Ereignissen möglich, die nur in einem Modul ihre Energie deponieren.

Anhang

A.1 Rapidity

Um bei ultrarelativistische Schwerionenreaktionen die dabei entstehenden Teilchen mit ihren extrem hohen Geschwindigkeiten beschreiben zu können ist die spezielle Relativitätstheorie notwendig. Dabei wird die Energie E eines Teilchens folgendermaßen beschrieben¹:

$$E^2 = m_0^2 + \vec{p}^2 , \\ (m_0^2: \text{Ruhemasse des Teilchens}, \vec{p}: \text{Dreierimpuls}).$$

Der Dreierimpuls \vec{p} wird nun in eine Komponente p_{\perp} senkrecht und p_{\parallel} zur Strahlrichtung aufgeteilt:

$$p_{\perp} = p \sin \theta , \\ p_{\parallel} = p \cos \theta ,$$

(p : Betrag des Dreierimpulses \vec{p} , θ : Polarwinkel zur Strahlachse).

Der Transversalimpuls p_{\perp} ist in Richtung der Strahlachse invariant gegenüber Lorenztransformationen, p_{\parallel} ist dies nicht. Eine neue dimensionslose Größe, die Rapidity y , wird eingeführt:

$$y = \frac{1}{2} \ln \left(\frac{E + p_{\parallel}}{E - p_{\parallel}} \right) . \quad (\text{A.1})$$

Die Rapidity verhält sich unter einer Lorentztransformation additiv. Bei nicht-relativistischen Geschwindigkeiten geht die Rapidity in die longitudinale Geschwindigkeit v_{\parallel} , in Einheiten von c , über. Energie E und Longitudinalimpuls p_{\parallel} lassen sich durch Umformen von Gleichung A.1 beschreiben als:

¹In allen folgenden Gleichungen wird die Lichtgeschwindigkeit $c = 1$ gesetzt.

$$\begin{aligned} E &= m_{\perp} \cosh y , \\ p_{\parallel} &= m_{\perp} \sinh y , \\ (\text{mit } m_{\perp} &= \sqrt{m_0^2 + p_{\perp}^2}). \end{aligned} \tag{A.2}$$

Da die Rapidity im Experiment nur schwer zu bestimmen ist, hierzu würde die Masse der Teilchen benötigt, wird die Pseudorapidity ermittelt. Wenn die Energie $E \gg m_0$ wird, geht die Rapidity y in die Pseudorapidity η über, bei dessen Bestimmung nur der Winkel θ zur Strahlachse bekannt sein muß:

$$\eta = \frac{1}{2} \ln \left(\frac{p + p_{\parallel}}{p - p_{\parallel}} \right) = -\ln \left(\tan \frac{\theta}{2} \right) .$$

In Analogie zu Gleichung A.2 gilt:

$$\begin{aligned} p &= p_{\perp} \cosh \eta , \\ p_{\parallel} &= p_{\perp} \sinh \eta . \end{aligned}$$

Für Photonen ($m_0 = 0$) sind Rapidity und Pseudorapidity identisch.

A.2 Gewichte des neuronalen Netzes

In der folgenden Tabelle A.1 sind die Gewichte des neuronalen Netzes mit 28 Eingabeneuronen und 15 Neuronen in der verdeckten Schicht aufgeführt. Dieses Netz wurde bei der Analyse der WA98 Daten verwendet.

A.2 Gewichte des neuronalen Netzes

Tabelle A.1: Gewichte des verwendeten neuronalen Netzes.

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}	V_0
I_1	-0.043	0.188	-0.020	-0.052	-0.322	-0.669	-0.140	-0.180	-0.047	-0.192	-0.005	0.156	-0.045	0.274	0.040	
I_2	-0.025	-0.403	0.039	0.099	-0.462	-0.945	-0.035	-0.156	-0.062	-0.700	0.156	0.268	-0.134	0.098	0.078	
I_3	0.020	0.575	0.002	-0.051	-0.368	-1.218	-0.522	-0.482	-0.069	-0.214	-0.078	0.273	-0.113	0.010	0.195	
I_4	-0.007	0.013	-0.059	-0.012	-0.355	-1.062	-0.117	-0.249	-0.043	-0.563	0.008	0.259	-0.127	0.124	0.054	
I_5	-0.026	-0.127	-0.035	-0.072	-0.316	-0.652	-0.112	-0.225	-0.049	-0.047	0.037	0.074	-0.046	-0.073	0.066	
I_6	-0.012	0.060	-0.047	0.164	-0.525	-1.136	-0.044	-0.240	-0.065	-0.592	0.017	0.295	-0.112	0.097	0.061	
I_7	-0.151	0.335	-0.243	0.870	-0.399	-0.648	-0.087	-1.726	0.001	-1.969	1.151	0.410	-0.035	0.232	0.403	
I_8	-0.170	-0.038	1.601	1.151	-1.466	-1.273	1.775	0.758	-0.682	0.669	-2.507	0.758	-1.400	-0.414	0.360	
I_9	-0.283	0.496	-0.385	0.361	0.858	-0.720	-0.387	-1.562	-0.050	-3.350	1.071	0.019	0.028	1.086	-1.016	
I_{10}	-0.027	-0.265	-0.072	0.031	-0.206	-0.781	-0.041	-0.199	-0.077	-0.279	0.225	0.140	-0.134	0.124	0.013	
I_{11}	-0.026	0.638	-0.019	0.128	-0.676	-1.087	-0.084	-0.098	-0.063	-0.597	0.019	0.261	-0.214	0.026	0.104	
I_{12}	-0.312	-0.468	0.003	0.925	0.200	-1.483	-0.822	1.379	-0.071	-0.550	0.264	1.086	0.953	0.993	0.777	
I_{13}	1.399	-0.683	-0.568	-1.215	0.902	1.681	-0.008	1.346	1.038	0.211	-1.377	0.176	-0.251	-0.026	1.510	
I_{14}	-1.029	1.009	0.617	-0.141	-0.754	-1.626	0.016	1.101	-1.074	0.101	1.064	1.729	1.107	0.061	-1.144	
I_{15}	-0.034	-0.451	-0.025	-0.098	-0.274	-1.101	-0.210	-0.282	-0.079	-0.455	-0.089	0.210	-0.181	-0.172	0.132	
I_{16}	-0.013	0.376	-0.067	0.004	-0.481	-1.250	-0.156	-0.263	-0.055	-0.798	-0.022	0.212	-0.117	0.006	0.063	
I_{17}	-0.209	0.608	-0.574	0.059	0.511	-0.165	-0.517	-1.449	-0.364	-2.122	0.122	0.049	0.782	-0.491	-0.022	
I_{18}	-0.257	0.010	1.550	-0.694	-1.306	-0.833	-0.209	1.121	-0.971	0.413	-1.611	1.892	-0.615	1.321	-0.687	
I_{19}	-0.738	0.800	-1.090	-0.265	0.711	-0.873	-0.199	-1.537	-1.000	-1.201	-0.668	-0.429	0.623	0.324	-0.746	
I_{20}	0.028	-0.273	-0.022	-0.085	-0.225	-0.533	-0.340	-0.160	0.063	-0.325	0.212	0.183	-0.134	0.259	0.013	
I_{21}	-0.025	0.121	0.010	-0.046	-0.195	-0.679	-0.046	-0.148	-0.047	-0.113	0.055	0.108	-0.041	0.073	0.004	
I_{22}	0.030	0.216	-0.042	-0.045	-0.310	-1.158	-0.144	-0.283	-0.043	-0.469	-0.004	0.266	-0.087	0.118	0.036	
I_{23}	-0.003	-0.194	0.213	-0.115	-0.437	-1.203	-0.558	-0.342	-0.074	-0.233	0.238	0.160	-0.155	-0.062	0.136	
I_{24}	-0.024	0.173	-0.047	-0.059	-0.174	-1.131	-0.011	-0.279	-0.070	-0.250	0.086	0.212	-0.066	0.133	0.022	
I_{25}	-0.009	0.402	-0.009	0.023	-0.211	-0.845	-0.054	-0.231	-0.056	-0.187	-0.025	0.127	-0.051	0.041	0.047	
I_{26}	0.389	3.904	0.927	-0.623	-3.562	0.083	0.650	-0.306	-1.271	-2.451	-1.027	1.799	-0.120	2.302	-0.609	
I_{27}	0.027	0.837	0.313	-0.051	-0.382	-3.173	-2.190	-1.239	-0.243	-0.680	0.051	0.059	-0.527	0.029	0.926	
I_{28}	0.135	0.473	0.337	-0.080	-0.163	-2.130	-1.718	-1.163	-0.158	0.006	0.180	0.070	-0.669	0.315	0.652	
I_0	0.578	0.147	-1.630	0.169	-0.808	-0.732	-1.211	-1.441	1.331	0.241	1.128	-1.544	0.569	-1.319	0.339	
O_1	0.006	0.059	0.053	0.035	0.037	0.043	-0.109	-1.008	0.320	0.041	0.013	0.722	0.400	0.677	0.097	0.017
O_2	0.114	-1.053	-0.144	0.072	0.032	0.020	0.096	0.022	0.055	0.477	0.112	0.035	-0.334	0.004	-0.059	0.001

Literaturverzeichnis

- [BUS97] Diplomarbeit, H. Busching, Untersuchung nichtlinearer Effekte im Bleiglaskalorimeter LEDA - Eine Computersimulation -, Universität Münster (1997)
- [BIS95] C. Bishop, Neural Networks for Pattern Recognition, Oxford University Press (1995)
- [BER92] F. Berger et. al., *Particle identification in modular electromagnetic calorimeters*, Nucl. Instr. and Meth. **A321** (1992) 152-164
- [CLA96] Dissertation, A. CLausen, Einsatz des Bleiglaskalorimeters LEDA zur Untersuchung ultrarelativistischer bleiinduzierter Schwerionenreaktionen in Hinblick auf die Produktion inklusiver Photonen, Universität Münster (1996)
- [CLE93] Dissertation, G. Clewing, Untersuchung zur Produktion direkter Photonen in ultrarelativistischen Schwerionenreaktionen bei 200 AGeV (1993)
- [FUK90] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press (1990)
- [GEANT] R. Brun et. al., GEANT 3.21, CERN Data Handling Division
- [HER91] J. Hertz, A. Krogh, R. G. Palmer, *Introduction to the Theory of Neural Computation* , Addison-Wesley (1991)
- [KLE92] K. Kleinknecht, Detektoren für Teilchenstrahlung, Teubner Verlag Stuttgart (1992)
- [KUN] Habilitationsschrift, M. Kunze, Über die Anwendung Neuronaler Systeme in der Teilchenphysik, Ruhr-Universität Bochum

- [LEC91] J. S. Denker, Y. leCun, Transforming Neural-Net Output Levels to Probability Distributions, Neural Information Processing System3, 853-859 (1991)
- [LIP91] M. D. Richard, R. P. Lippmann, Neural Network Classifiers Estimate Bayesian a posteriori Probabilities, Neural Computation3 (4) (1991)
- [LON75] E. Longo, I. Sesteli, Nuclear Instruments and Methods **128** (1975) 283-307
- [MUE95] B. Müller, J. Reinhardt, M.T. Strickland, Neural Networks, Springer Verlag (1995)
- [OTT96] Diplomarbeit, R. Ottenhues, Teilchenklassifizierung im Bleiglaskalorimeter mit künstlich neuronalen Netzen, Universität Münster (1996)
- [PER87] D. H. Perkins, Introduction to High Energie Physics, Addison Wesley, Menlo Park (1987)
- [RIE93] M. Riedmiller und H. Braun, A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *Proceedings of the IEEE International Conference on Neural Networks (ICNN)*, Seiten 586 - 591, San Francisco, 1993
- [RUM86] D.E. Rumelhart, G. Hinton und R. Williams. Learning internal representations by error propagation. *Parallel Distributed Processing, Vol. I Foundations*, Seiten 318 - 362. MIT Press, Cambridge, MA, 1986
- [ROJ93] R. Rojas, Theorie der neuronalen Netze, Springer-Verlag, Berlin (1993)
- [ROT91] Diplomarbeit, B. Roters, Untersuchung von Ortsrekonstruktionsmethoden und Teilchenidentifizierung bei einem segmentierten Kalorimeter, Universität Münster (1991)
- [SCH94] Diplomarbeit, H. Schlagheck, Konstruktion und Kalibration eines Bleiglaskalorimeters für ultrarelativistische Schwerionenreaktionen, Universität Münster (1994)

- [SIH90] B. Sinha, S. Pal und S. Raha, Quark-Gluon-Plasma, Springer Verlag (1990)
- [SIN96] R. Sinkus und T. Voss, *Particle identification with neural networks using a rotational invariant moment representation*, Deutsches Elektron-Synchrotron in Hamburg, DESY 96-264, December 1996
- [SNNS4] SNNSv4.1, Neural Network Simulator, User Manual, Version 4.1, Universität Stuttgart (1995)
- [WA80] WA80-Kollaboration, *Nucl. Instr. and Meth. /bf A* 292 (1990), 81
- [WA93] WA93-Kollaboration, GSI-Scientific Report 1992, I-93, S. 386, (1993)
- [WA98] WA98-Kollaboration, *Proposal for a large acceptance hadron and photon spectrometer*, CERN/SPSLC 91-17, SPSLC/P260, CERN, Genf, 1991

Danksagung

Zum Schluß möchte ich mich bei allen bedanken, die bei der Erstellung der Arbeit geholfen haben.

Herrn Prof. Dr. R. Santo für die interessante Aufgabenstellung, die mir sehr entgegenkam.

Bei Dr. habil. T. Peitzmann, für die Unterstützung und die vielen Tips.

Für die Mühe bei der Korrektur meiner Orthographie und Interpunktions sowie für zahlreiche Hinweise bedanke ich mich besonders bei Dipl.-Phys. S. Bathe, Dipl.-Phys. H. Büsching, Z. Böröcz, Dipl.-Math. T. Dierkes, S. Gorus, H. Köhncke, E. Szentmártoni, C. Thomas, sowie U. Weißenfels.

Desweiteren möchte ich der gesamten Arbeitsgruppe danken, für lustige und hilfreiche Diskussionen:

Dipl.-Phys. C. Barlag, Dipl.-Phys. C. Blume, Dipl.-Phys. D. Bucher, Dr. R. Glasow, N. Heine, Dr. A. Khoukaz, T. Krümpel, Dipl.-Phys. J. Langheinrich, Dipl.-Phys. T. Lister, I. Pellmann, Dipl.-Phys. C. Quentmeier, Dipl.-Phys. K. Reygers, Dipl.-Phys. G. Schepers, Dipl.-Phys. H. Schlagheck, Dipl.-Phys. D. Stüken, W. Verhoeven.

Ein ganz besonderer Dank geht an meine Familie für deren Unterstützung, besonders an meine Mutter G. Joerrens und E. Szentmártoni, ohne sie wäre diese Arbeit nicht zustandegekommen.

*Hiermit bestätige ich, daß ich diese Arbeit selbstständig
verfaßt und keine anderen als die angegebenen
Quellen und Hilfsmittel benutzt habe.*

Münster, 31. März 1998

STEPHAN JOERRENS

