

Musterverarbeitung kernphysikalischer Meßdaten
mit
künstlichen neuronalen Netzen

Diplomarbeit

von

Matthias Kaspar

Institut für Kernphysik
der
Johann Wolfgang Goethe-Universität
Frankfurt am Main

November 1994

"Neural networks are the second best way of doing just about anything."

John Denker

Inhaltsverzeichnis

1	Einleitung	1
2	Künstliche neuronale Netze	3
2.1	Biologische Motivation	4
2.2	Abstrakte Struktur	6
2.3	Mehrschichtige Perzeptrons	7
2.4	Error Backpropagation-Algorithmus	12
2.5	Anmerkungen zur Anwendung	14
3	Approximation am Beispiel eines ortsempfindlichen Teilchenzählers	18
3.1	Aufbau und Funktionsweise des Detektors	18
3.2	Rahmenbedingungen des Trainings	19
3.3	Netzstrukturen und Ergebnisse	21
4	Musterklassifikation am Beispiel segmentierter γ-Detektoren	31
4.1	Bestimmung des ersten Segments	32
4.1.1	Physikalische Anmerkungen	32
4.1.2	Zweifach segmentierter Detektor	35
4.1.3	Vierfach segmentierter Detektor	44
4.1.4	16fach segmentierter Detektor	56
4.2	Doppeltreffererkennung	62
5	Zusammenfassung und Ausblick	70
A	Software	72
B	Bond-Diagramme	74
	Literatur	

1 Einleitung

Im Rahmen heutiger Experimente der Kern-, Schwerionen- oder Hochenergiephysik kommen in zunehmendem Maße komplexere Detektorsysteme zu Nachweis und Identifikation unterschiedlicher Reaktionsprodukte zum Einsatz. Dabei ist etwa die Energie emittierter Teilchen und γ -Quanten zu messen oder die Flugbahn von Teilchen - seien es Reaktionsprodukte oder Projektile - sowie ggfs. deren Ladung und Masse zu bestimmen. Um darüberhinaus Aufschluß über die räumliche Verteilung der Reaktionsprodukte zu erhalten und einen möglichst vollständigen Nachweis zu gewährleisten, sollten die verwendeten Detektorsysteme einen möglichst großen Bereich des interessierenden Raumwinkels abdecken und zum Zwecke der exakten Lokalisation über eine hohe Granularität verfügen. Dadurch fällt pro nachzuweisendem Ereignis eine Vielzahl zu verarbeitender Signale an, so daß derart große Datenmengen entstehen, daß nicht mehr alle Ereignisse durch ein Datenaufnahmesystem aufgezeichnet werden können. Daher müssen einzelne Ereignisse in einem Experiment anhand der Gesamtheit der sie charakterisierenden Signale möglichst schnell beurteilt und klassifiziert werden, um wesentliches und interessierendes von unwesentlichem und störendem, wie etwa von Untergrundprozessen, zu trennen. Auch bei der Analyse der Daten nach Abschluß des Experiments bestehen vielfältige Probleme der Klassifikation der Daten oder etwa des Auffindens geeigneter, die Daten beschreibender funktionaler Zusammenhänge. Oftmals stehen zur Lösung derartiger Probleme keine dezidierten Algorithmen zur Verfügung. Dies mag darin begründet sein, daß diese prinzipiell nicht aufgestellt werden können, beispielsweise, wenn keine geschlossene Lösung einer zugrundeliegenden Gleichung existiert, andererseits kann eine exakte Lösung möglicherweise nur mit unverhältnismäßig hohem Aufwand gefunden werden.

In den letzten Jahren fanden neue Methoden zur Musterverarbeitung durch unterschiedliche Modelle künstlicher neuronaler Netze in verschiedensten Bereichen Anwendungen. Der Einsatz eines künstlichen neuronalen Netzes erscheint lohnend, sofern zur Verarbeitung eines Musters keine detaillierten Regeln zur Verfügung stehen. Mustervervollständigung, Musterassoziation, die Klassifizierung von Mustern oder das Approximieren von Abbildungen, deren Funktionswerte nur auf einer begrenzten Zahl von Stützstellen bekannt sind, gehören dabei zu den Aufgaben künstlicher neuronaler Netze. Der Kerngedanke besteht darin, eine derartige Aufgabe nicht durch Erstellen eines geeigneten Algorithmus zu lösen, sondern daß vielmehr das neuronale Netz durch eine Analyse vorliegender Daten diesen Daten zugrundeliegenden Zusammenhang erkennt und diesen approximiert, so daß neue Daten der gleichen Problemstellung später sinnvoll bewertet werden können.

Ein Modell eines künstlichen neuronalen Netzes kann sowohl auf mathematischem Wege betrachtet, durch Simulationen untersucht als auch auf beispielsweise elektronischem Wege realisiert werden. Ist ein optimales Netz gefunden und realisiert, so kann mit seiner Hilfe eine sehr schnelle Bewertung der zu verarbeitenden Daten erfolgen. Im Rahmen von Hochenergiephysik-Experimenten - wie beispielsweise am CERN oder bei DESY - sind daher seit einiger Zeit Untersuchungen wie beispielsweise [Mö94] im Gange, künstliche neuronale Netze zur Beurteilung von Detektorsignalen einzusetzen.

Motiviert durch derartige Überlegungen wird in dieser Arbeit untersucht, inwieweit auch im Bereich der Kernphysik künstliche neuronale Netze zur Verarbeitung von Detektorsignalen angewendet werden können. Dabei wird keine Realisation eines Netzes auf elektronischem Wege zur Anwendung während eines Experiments angestrebt, vielmehr wird eine Simulation der Netze durch einen Rechner unternommen. Somit ist es möglich, im Rahmen einer Analyse nach Abschluß des eigentlichen Experiments die aufgezeichneten Daten zu bewerten und weiterzuverarbeiten. Dazu werden exemplarisch zwei unterschiedliche Problemstellungen betrachtet:

Im Rahmen der Eichung eines in Zusammenarbeit zwischen dem Institut für Kernphysik und der EUROBALL-Gruppe der GSI entwickelten Parallelplattenzählers [Vet94] ist eine unbekannte Abbildung zur Umrechnung von vier gemessenen Ladungen in zweidimensionale Ortskoordinaten zu finden. Da die in diesem Fall benötigte Lösung der Poissongleichung prinzipiell nicht geschlossen dargestellt werden kann, jedoch das Verhalten des Systems an einigen Stützstellen durch eine Eichung bekannt ist, ist zu untersuchen, inwiefern ein künstliches neuronales Netz zur Approximation des gesuchten Zusammenhangs eingesetzt werden kann.

Im Rahmen des Nachweises von γ -Quanten mit segmentierten Germaniumdetektoren - beispielsweise des EUROBALL-III Projektes - treten unerwünschte Comptoneffekte auf, welche zu Streuungen von γ -Quanten zwischen benachbarten Segmenten eines Germaniumkristalls oder gar zwischen benachbarten Kristallen führen. Dadurch wird die durch die Segmentierung verbesserte räumliche Auflösung teilweise reduziert, da im Falle des Ansprechens mehrerer Segmente nicht ersichtlich ist, in welchem Segment die erste Wechselwirkung des γ -Quants mit dem Germaniumkristall stattfand, zumal die Reihenfolge des Ansprechens der einzelnen Segmente meßtechnisch nicht mehr aufgelöst werden kann. Daher wird untersucht, inwieweit sich alleine aus der Kenntnis der in den einzelnen Segmenten deponierten Energien Rückschlüsse auf das Segment ziehen lassen, in dem die erste Wechselwirkung stattfand. Desweiteren treten bei vielen Kernreaktionen hohe γ -Multiplizitäten auf, so daß zwei Quanten simultan in unterschiedliche Segmente eines Detektors gelangen können. Es ist zu untersuchen, inwieweit ein derartiger Doppeltreffer von einem Eintreffter mit anschließender Comptonstreuung unterschieden werden kann.

Sowohl die Approximation einer unbekanntem Funktion als auch die Klassifikation von Mustern kann durch ein einziges Modell eines künstlichen neuronalen Netzes, das Modell des mehrschichtigen Perzeptrons [Her91], erfolgen. Da dieses Modell leicht zu simulieren ist und eine Vielzahl von Vorschlägen zur Modifikation des zum Training dieser Struktur eingesetzten Error Backpropagation-Algorithmus [Her91, Rum86] existiert, wurde dieses Modell im Rahmen der vorliegenden Arbeit angewendet. Dazu wurde ein Programm entwickelt, mit dessen Hilfe mehrschichtige Perzeptrons beliebiger Struktur durch mehrere Verfahren auf beliebige Datensätze trainiert werden können. Die Trainingsdaten standen einerseits durch die Eichung des Parallelplattenzählers mit einer Spaltquelle, andererseits durch Simulation der Germaniumdetektoren durch das Monte Carlo-Programm GEANT [GEA93] zur Verfügung.

2 Künstliche neuronale Netze

Ein künstliches neuronales Netz (KNN) ist eine mathematische Struktur, die durch Aufbau und Funktionsweise von Nervenzellen und deren Kombination zu komplexen Netzwerken in lebenden Organismen motiviert ist. Momentan werden einige Dutzend¹ Modelle künstlicher neuronaler Netze diskutiert, welche diese Struktur auf sehr unterschiedliche Weise realisieren. Obgleich die übergeordnete Struktur am biologischen Vorbild orientiert ist, müssen die konstruierten Modelle nicht notwendigerweise in Strenge biologisch bzw. biochemisch plausibel sein. Es ist reizvoll, durch ein KNN ein reales Netz aus Neuronen so detailgetreu als irgend möglich zu modellieren, jedoch ist dies - nicht zuletzt aufgrund der äußerst hohen Komplexität der biochemischen Abläufe in einer Nervenzelle - nicht Gegenstand der meisten heute angewandten Modelle.

Die Lösung eines Problems durch Netzwerke aus Nervenzellen in lebenden Organismen - von einfachen Verbänden aus nur wenigen Neuronen bis hin zum komplexen menschlichen Gehirn - vollzieht sich auf andere Art und Weise als seine Lösung bei Einsatz eines Digitalrechners.

Um ein Problem durch einen Computer zu lösen, ist durch seine iterative Zerlegung in Teilprobleme ein geeigneter Algorithmus zu finden, welcher von einem Prozessor sequentiell - oder im Rahmen parallel arbeitender Systeme von einem Verband mehrerer Prozessoren - abgearbeitet wird. Der Ausfall des Prozessors muß - der Ausfall eines Prozessors in einem parallelen System kann - zum Totalausfall des Systems führen. Heutige Prozessoren sind in ihrem Aufbau komplex und arbeiten unter Umsetzung einer Leistung von mehreren Watt mit Taktraten bis zu etwa 300 MHz.

Ein lebender Organismus dagegen wird die Lösung eines Problems zunächst erlernen, wobei nach heutiger Ansicht Parameter variiert werden, welche - wie im folgenden Abschnitt skizziert - die Kopplung jeweils zweier Nervenzellen charakterisieren. Während des Lernvorgangs wird der Organismus mit Stimuli konfrontiert und wird auf diese Eingabegrößen reagieren. Diese Reaktion kann extern bewertet und diese Bewertung dem Organismus wiederum mitgeteilt werden, so daß dieser seine Reaktion auf die ursprünglichen Stimuli ggfs. korrigieren kann. Dabei sind unterschiedliche Abstufungen der Bewertung denkbar; so kann dem Organismus beispielsweise neben dem bloßen Anzeigen einer falschen Reaktion etwa auch eine Musterlösung präsentiert werden, oder die Bewertung kann lediglich aus einer binären Aussage bestehen. Sofern keinerlei Bewertung der Reaktion erfolgt, müssen mögliche Strukturen der Stimuli selbständig erkannt werden. Das Netzwerk aus Nervenzellen selbst ist im allgemeinen massiv parallel; es besteht aus einer sehr großen Zahl von Neuronen. Die in der Natur realisierten Neuronen sind im Vergleich zu einem heutigen Prozessor sehr viel einfacher in ihrer Struktur und arbeiten unter Umsetzung einer sehr geringen Leistung. Die Verarbeitungs- und Übertragungsgeschwindigkeit von Signalen in bzw. zwischen Neuronen ist deutlich geringer als die innerhalb heutiger Prozessoren. Dies wird durch die massive Parallelität der Nervenzellen ausgeglichen, welche darüberhinaus eine Redundanz birgt; der Ausfall einzelner Neuronen führt nicht schlagartig zu einem Totalausfall des Systems; das System ist robust.

Um die Struktur eines KNNs besser verstehen zu können, werden im folgenden Abschnitt auszugsweise biochemische Grundlagen einer typischen Nervenzelle und eines gängigen Kopplungsmechanismus zweier Nervenzellen über eine chemische Synapse skizziert. Auf eine ausführliche Darstellung der sehr interessanten und zahlreichen molekularbiologischen bzw. elektrochemischen Vorgänge in der

¹Seit Mitte der achtziger Jahre finden KNNs in zunehmendem Maße in unterschiedlichsten Gebieten Anwendungen. Nicht zuletzt durch die dadurch entstandene Aktivität steigt die Zahl vorgeschlagener Modelle bzw. vorgeschlagener Modifikationen bestehender Modelle momentan an.

Vielzahl der unterschiedlichen Typen von Nervenzellen muß im Rahmen dieser Arbeit verzichtet werden, Interessenten seien dazu auf einführende Literatur der molekularen Neurobiologie wie beispielsweise [Hal92, Nic92] verwiesen. Kürzere Darstellungen können auch den einführenden Kapiteln einiger Lehrbücher über künstliche neuronale Netze [Mü91, Per92, Ami92] entnommen werden.

2.1 Biologische Motivation

In diesem Abschnitt werden zunächst der Aufbau eines Neurons und die Vernetzung mehrerer Neuronen, im Anschluß daran die Entstehung eines elektrischen Signals und dessen Übermittlung zwischen zwei Neuronen skizziert.

Eine typische Nervenzelle besteht aus einem einige zehn Mikrometer großen Zellkörper, der in ein bis zu einem Meter langgestrecktes Axon mündet, siehe Abbildung 1. Von der restlichen Oberfläche des Zellkörpers geht eine Vielzahl feinverästelter Dendriten aus. Das Axon teilt sich in eine baumartige Struktur auf, jedes der entstehenden Enden führt zu einem synaptischen Endknopf. Die Kopplung zweier Nervenzellen über eine chemische Synapse kommt dadurch zustande, daß ein derartiger synaptischer Endknopf sehr dicht - etwa 20 bis 200 Nanometer - an einem Dendrit oder direkt am

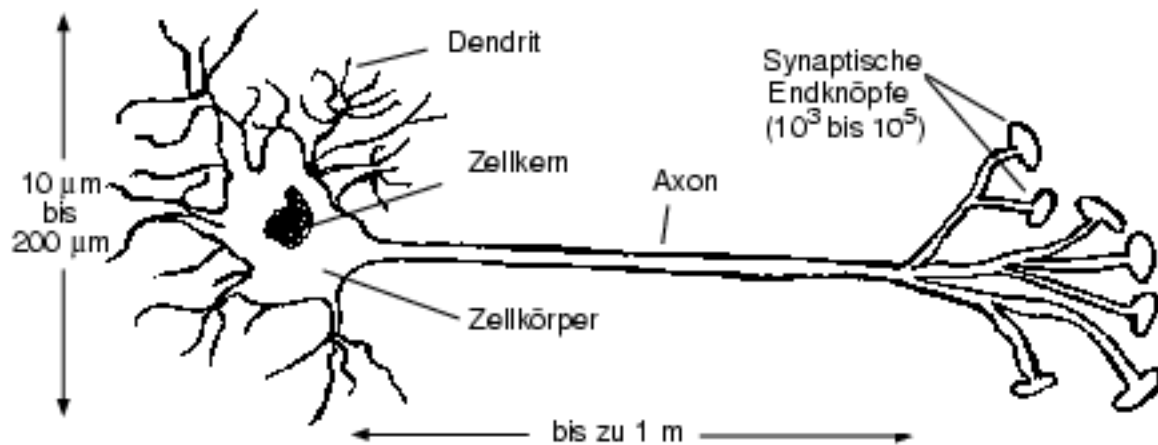


Abbildung 1: Schematische Darstellung eines Neurons.

Zellkörper eines anderen Neurons anliegt. Im cerebralen Cortex liegen schätzungsweise 10^{10} bis 10^{11} Neuronen innerhalb einer näherungsweise zweidimensionalen Struktur von etwa einem Quadratmeter vor, von denen jedes mit 10^3 bis 10^4 weiteren Neuronen verbunden ist; dadurch ergeben sich 10^{13} bis 10^{15} synaptische Verbindungen.

Informationsverarbeitung und -weitergabe vollzieht sich in Neuronen auf elektro- bzw. biochemischem Wege. Das Innere einer Nervenzelle ist von dem sie umgebenden Milieu durch Lipiddoppelschichten

elektrisch isoliert, so daß durch räumliche Trennung von - im wesentlichen - Natrium-, Calcium- und Chloridionen Potentialdifferenzen gebildet, aufrecht erhalten und wieder abgebaut werden können. Dazu sind ionenspezifische Kanäle in den Membranen vorhanden, welche durch Öffnung oder Schließung eine passive Ionenausbreitung entlang eines Konzentrationsgradienten ermöglichen bzw. verhindern. Zusätzlich stehen aktive Pumpen zur Verfügung, welche unter Umsetzung von Energie Ionen aktiv entgegen einen Konzentrationsgradienten transportieren. Ohne das Wechselspiel dieser Mechanismen detailliert zu diskutieren sei nun der Kreislauf der Entstehung eines lokalen elektrischen Signals am Anfang des Axons einer Nervenzelle, seine Ausbreitung entlang des Axons, seine Übertragung auf benachbarte Neuronen und die dortige ggfs. erneute Entstehung eines lokalen elektrischen Signals skizziert. Im Ruhezustand mißt man zwischen Innen- und Außenraum des Axons eine Potentialdifferenz von -70 mV, das Ruhepotential. Wird dieses am Anfang des Axon künstlich oder durch später erläuterte Umstände um einen gewissen Betrag erhöht, so entsteht eine lokale Depolarisation auf $+40$ mV und etwa eine halbe Millisekunde später eine Hyperpolarisation auf -80 mV. Der somit entstehende Puls einer Dauer von etwa einer Millisekunde wird Aktionspotential genannt; für die Dauer von drei bis vier Millisekunden ist nun das Auslösen weiterer Aktionspotentiale blockiert. Das Aktionspotential läuft unter Beibehaltung seiner Form mit einer Geschwindigkeit von etwa 5 m/sec das Axon herunter² und wird im synaptischen Baum vervielfältigt. Somit gelangt es auf eine Vielzahl synaptischer Endknöpfe, wo jeweils durch Öffnung von Calciumkanälen die Verschmelzung kleiner Vesikel mit der Membranaußenseite der Endknöpfe ausgelöst wird, so daß sehr kleine Mengen von etwa 10^{-17} mol der in den Vesikeln enthaltenen Neurotransmitter³ freigesetzt werden⁴. Diese diffundieren innerhalb einer halben Millisekunde von der präsynaptischen Seite aus durch den synaptischen Spalt und werden auf der postsynaptischen Seite von Neurorezeptoren gebunden. Dadurch ist das Signal von einem Neuron bis zur Außenseite der Membran eines benachbarten Neurons weitergegeben worden. Durch das Binden der Neurotransmitter verändert sich kurzzeitig die - wiederum ionenspezifische - Permeabilität der Membran des postsynaptischen Neurons, so daß dort Ionen ein- bzw. ausströmen können und somit ein Puls, das postsynaptische Potential, entsteht. Der Grad dieser Veränderung ist dabei ein veränderbarer Parameter; man spricht von synaptischer Plastizität. Ihn auf einer größeren als der momentan betrachteten Zeitskala für jede Synapse so einzustellen, daß das Gesamtsystem auf Stimuli sinnvoll zu reagieren vermag, ist nach heutiger Ansicht das Ziel eines jeden Lernvorgangs. Der Betrag des postsynaptischen Potentials richtet sich nach dem Grad der Veränderung der Permeabilität, sein Vorzeichen nach der Art des Neurotransmitters. Sofern das postsynaptische Potential an einem Dendriten und nicht direkt am Zellkörper des benachbarten Neurons entstand, bewegt es sich zum Zellkörper; dabei geht die ursprüngliche Form des Pulses verloren. Im Zellkörper treffen je nach Vorzeichen erregend oder hemmend wirkende postsynaptische Potentiale in großer Zahl - bis zu einigen 10^4 - ein und werden während einiger Millisekunden aufintegriert. Dieses Integral dient nun erneut als lokaler elektrischer Reiz am Anfang des Axons dieses Neurons; sofern ein gewisser Schwellwert nicht überschritten wird, kann kein neues Aktionspotential entstehen. Erreicht oder übersteigt das Integral dagegen diesen Schwellwert, so entsteht ein neues Aktionspotential, welches nach dem skizzierten Mechanismus wiederum an eine Vielzahl benachbarter Neuronen weitergegeben wird.

²Der Ausbreitungsprozeß wird detailliert durch die Goldman-Hodgkin-Huxley Gleichungen beschrieben, welche durch eine Kombination des das Axon beschreibender Kabelgleichungen mit die Dynamik der beteiligten Ionenkanäle beschreibenden Gleichungen entstehen. 1963 erhielten Eccles, Hodgkin und Huxley den Nobelpreis für Medizin für ihre Arbeiten auf diesem Gebiet.

³Beispielsweise Gammaaminobuttersäure oder Acetylcholin.

⁴Dieser Freisetzungsprozess ist stochastischer Natur.

2.2 Abstrakte Struktur

Motiviert durch die skizzierten biologischen Vorgänge folgt nun eine formale Definition eines künstlichen neuronalen Netzes in Form einer abstrakten Struktur. Dabei werden nur generelle Prinzipien des vergangenen Abschnitts, nicht aber biochemische Details verwendet. Die hier vorgestellte Struktur ist mit den von Rummelhart und McClelland [Rum86] vorgestellten Prinzipien des Modells einer parallel verteilten Prozessführung identisch.

Ein künstliches neuronales Netz (KNN) besteht aus

- einer Menge formaler Knoten,
- einer aus den Aktivitäten der einzelnen Knoten bestehenden Aktivität des Gesamtsystems,
- einer Menge und Anordnung von Verbindungen der Knoten untereinander,
- einer Aktivitätsfunktion, anhand derer jeder Knoten seine Aktivität aus den Ausgabesignalen anderer Knoten und den zwischen den Knoten bestehenden Verbindungen berechnet,
- einer Ausgabefunktion, anhand derer jeder Knoten aus seiner aktuellen Aktivität ein Ausgabesignal berechnet,
- einer Vorschrift, die die Dynamik des Gesamtsystems beschreibt, so daß ein Schema, nach dem die einzelnen Knoten ihre Aktivitäten zeitlich aktualisieren, festgelegt ist,
- einer Umgebung, in der das System existiert und mit der es kommuniziert, dazu sind einige Knoten als Eingabeknoten, einige als Ausgabeknoten zu charakterisieren,
- einer Lernregel, anhand derer im Rahmen eines Lernvorgangs die Verbindungen der Knoten untereinander modifiziert werden.

Dabei können Aktivitäts- und Ausgabefunktion durch Verkettung zu einer Transferfunktion zusammengefaßt werden. Unterschiedliche Knoten können über unterschiedliche Transferfunktionen verfügen. Sofern Zahl und/oder Anordnung der formalen Knoten zeitlich veränderlich sind, sind ein Ausgangszustand sowie eine Vorschrift, die die zeitlichen Modifikationen definiert, anzugeben.

Einige Lehrbücher definieren ein KNN als gerichteten Graphen aus Knotenmenge aller Knoten und Kantenmenge aller Verbindungen [Bra91], andere als gewichteten Graphen [Mü91]. Diese Definitionen umfassen jedoch in der Regel weder die Definition einer geeigneten Netzumgebung, noch schließen sie die Lernregel mit ein; daher erscheint die von Rummelhart und McClelland vorgestellte Struktur angemessener. Eine interessante, alle dargelegten Aspekte umfassende formale Definition schlägt Fiesler in [Fie94] vor.

Ein Modell eines KNNs entsteht durch Realisation obiger Struktur. Dabei kann die Wahl von Topologie oder Transferfunktion biochemisch motiviert sein, sofern versucht wird, neurologische Prozesse detailgetreu zu modellieren; jedoch besteht dazu keinerlei Notwendigkeit. Gängige Modelle künstlicher neuronaler Netze verwenden sehr einfache Topologien und Transferfunktionen, welche biologisch nicht mehr als plausibel angesehen werden können. Nichtsdestoweniger sind diese Modelle mathematisch, physikalisch sowie unter dem Gesichtspunkt ihrer Anwendbarkeit außerordentlich interessant. Weit verbreitet sind - neben den im folgenden Abschnitt ausführlich vorgestellten mehrschichtigen Perzeptrons - auch das mit Methoden der statistischen Physik sehr detailliert untersuchte Hopfieldsche Modell

der Assoziativspeicher mit seinen Analogien zu Spin-Gläsern, sowie Boltzmann-Maschinen, Netzwerke aus radialen Basisfunktionen und zahlreiche selbstorganisierende Strukturen. Interessenten seien dazu auf einführende Literatur wie beispielsweise [Her91, Mü91, Hay94, Rum86, McC86, Per92, Ami92] verwiesen.

2.3 Mehrschichtige Perzeptrons

Da in dieser Arbeit das Modell der mehrschichtigen Perzeptrons verwendet wird, um eine Funktion zu approximieren und um Muster zu klassifizieren, wird dieses Modell eines künstlichen neuronalen Netzes nun diskutiert.

Mehrschichtige Perzeptrons (MSPs) sind vorwärtsgerichtete, deterministische Modelle künstlicher neuronaler Netze, bei denen formale Knoten in Schichten angeordnet sind. Die gewählte Topologie bleibt zeitlich erhalten. Synaptische Verbindungen bestehen lediglich zwischen unmittelbar benachbarten Schichten; auf dieser Ebene ist die Verknüpfung in der Regel jedoch vollständig, so daß in diesem Fall jeder Knoten einer Schicht mit jedem Knoten der beiden benachbarten Schichten verbunden ist, siehe Abbildung 2. Jeder Knoten verfügt über eine skalare Aktivität und ein skalares Ausgabesignal, jede formale Synapse über ein skalares Gewicht; beide müssen nicht notwendigerweise reell sein, werden in dieser Arbeit jedoch als reell angenommen. Für die folgende Diskussion verfüge das Perzeptron über L Schichten, in jeder Schicht seien $N_{l,l=1..L}$ Knoten vorhanden. Die zwischen zwei aufeinanderfolgenden Schichten vermittelnden Gewichte⁵ werden durch ω_{ij} charakterisiert; j kennzeichnet den prä-, i den postsynaptischen Knoten⁶.

Die erste Schicht des Perzeptrons dient lediglich der Eingabe der Information, sie hat sonst keinerlei Funktion. Die Ausgabesignale der Knoten dieser Schicht werden auf die gewünschten Eingabegrößen gesetzt und später von Knoten der folgenden Schicht gelesen, eine Verrechnung der Signale wie in allen anderen Schichten findet hier nicht statt. Die somit an das Netz angelegte Information fließt streng vorwärtsgerichtet von einer Schicht zur nächsten durch das Netz hindurch. Die Reihenfolge, in der die Aktivitäten der Knoten innerhalb einer Schicht aktualisiert werden, spielt dabei keine Rolle. Da das System rückkopplungsfrei ist, muß seine Dynamik nicht gesondert betrachtet werden. Die Ausgabesignale der Knoten der letzten Schicht - der Ausgabeschicht - stellen das Resultat der Rechnung dar. Alle Schichten zwischen Eingabe- und Ausgabeschicht werden als versteckte Schichten, alle Knoten in ihnen als versteckte Knoten bezeichnet.

Die Aktivität h eines Knotens wird über eine Aktivitätsfunktion g berechnet. Der einfachste und gebräuchlichste Fall ist der eines Sigma-Knotens. Dabei wird ein k -ter Knoten in der l -ten Schicht des Perzeptrons eine gewichtete Summe aus den Ausgabesignalen O_i der Knoten in der ihm vorausgehenden $(l-1)$ -ten Schicht nach Gleichung (1) berechnet. Gewöhnlich wird ein zusätzlicher Freiheitsgrad für jeden Knoten in der Form einer additiven Konstanten ϕ_k eingeführt.

$$h_k = g(\omega_{ki}, O_i) = \sum_{i=1}^{N_{l-1}} \omega_{ki} O_i - \phi_k \quad (1)$$

⁵Synapsen, synaptische Effizienzen und (formale) Gewichte werden synonym gebraucht.

⁶Um die Nomenklatur nicht durch Einführung drei- oder gar vierfach indizierter Größen zur Kernzeichnung von Ausgangs- und Zielschicht einer Synapse zu komplizieren, werden die zwischen unterschiedlichen Schichten vermittelnden Gewichte stets mit ω_{ij} bezeichnet. Sie dürfen keinesfalls miteinander verwechselt werden.

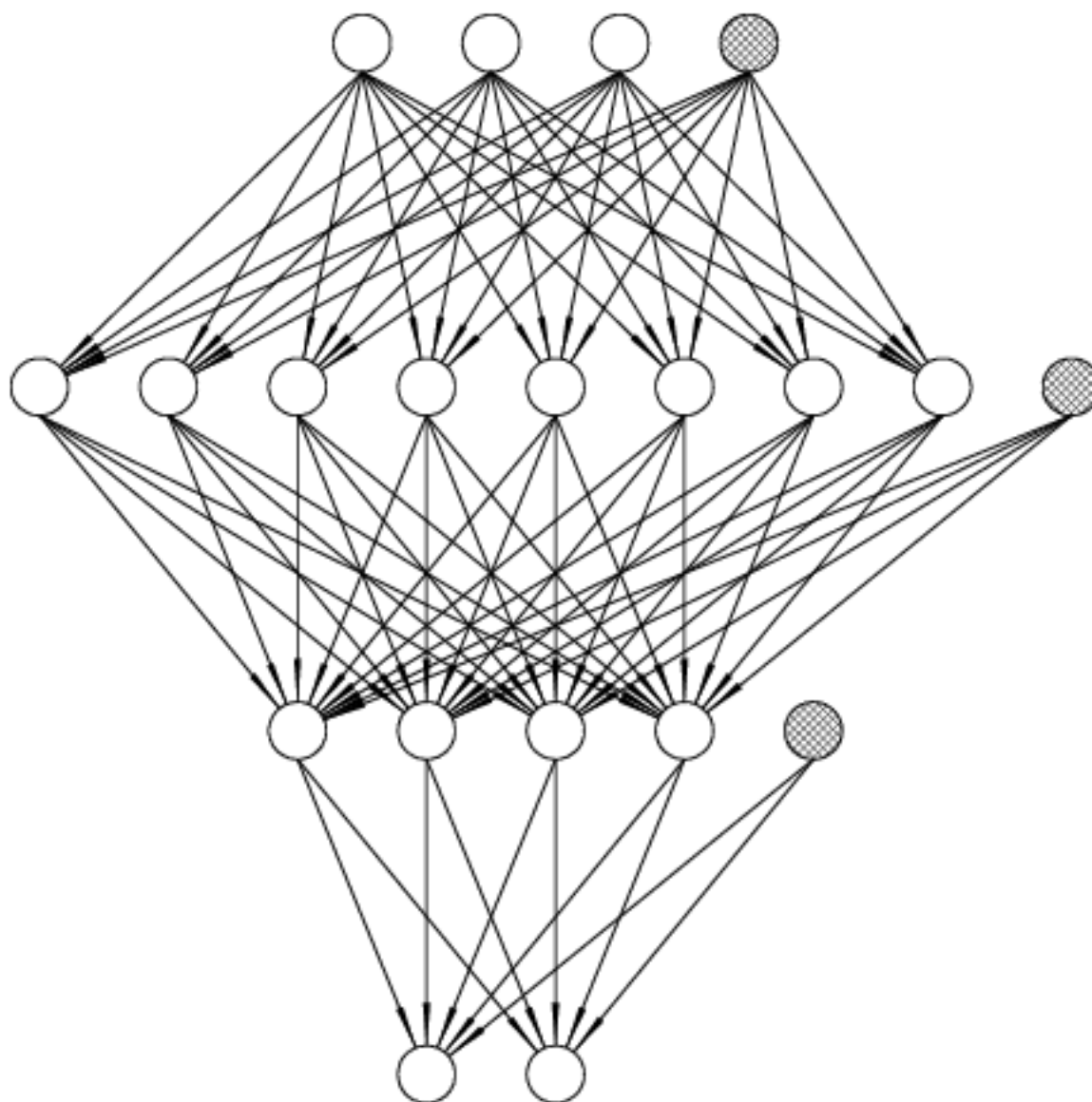


Abbildung 2: Schematische Darstellung eines mehrschichtigen Perzeptrons mit 3 Eingabeknoten, 2 versteckten Schichten zu 8 bzw. 4 Knoten und einer Ausgabeschicht zu zwei Knoten. Die Aktivität der künstlich eingeführten schraffierten Knoten ist identisch -1 .

ϕ_k kann jedoch elegant als Synapse zu einem künstlich in die vorausgehende $(l-1)$ -ste Schicht eingeführten Knoten formuliert werden, dessen Ausgabesignal konstant auf -1 liegt.

Neben derartigen Sigma-Knoten kommen gelegentlich auch Knoten höherer Ordnung, Sigma-Pi-Knoten zum Einsatz [Mü91]. Dabei werden neben den gewichteten Ausgabesignalen der Knoten der vorausgehenden Schicht auch gewichtete mehrfache Produkte dieser Signale aufaddiert. Die Ordnung eines Sigma-Pi-Knotens gibt an, bis zum wievielfachen Produkt der Signale dabei summiert wird. Dieses kann von Interesse sein, wenn MSPs zur Verarbeitung von Mustern eingesetzt werden sollen, und diese Muster unabhängig von ihrer Lage oder Größe erkannt werden müssen. Da durch Sigma-Pi-Knoten der Ordnung zwei eine Translationsinvarianz der zu verrechnenden Signale erreicht werden kann und sich auch Bedingungen für Rotations- oder Skaleninvarianz sowie Kombinationen dieser Invarianzen unter Verwendung von Sigma-Pi-Knoten höherer Ordnungen formulieren lassen, können mit Sigma-Pi Knoten auch verschobene, verdrehte oder skalierte Muster verarbeitet werden; Details sind in [Mü91] zu finden.

Das Ausgabesignal eines Knotens wird aus dessen Aktivität durch eine Ausgabefunktion berechnet und dient anderen Knoten als Eingabe bzw. wird in der Ausgabeschicht als Ergebnis abgegriffen und der Umgebung zur Verfügung gestellt. Die Wahl der Ausgabefunktion beeinflusst somit die mögliche Form der weitergegebenen und auszugehenden Signale erheblich. Eine Vielzahl unterschiedlicher Ausgabefunktionen sind denkbar, gewöhnlich kommen in MSPs differenzierbare⁷, sigmoide Funktionen zur Anwendung. Sie sind auf ganz R definiert, streng monoton und verfügen über einen beschränkten Wertebereich, wie beispielsweise der Tangens Hyperbolicus oder die Fermifunktion. Beide können durch Stauchung und Verschiebung nach Gleichung (2) bequem ineinander umgeformt werden, so daß die Knotenausgabe in einen gewünschten Wertebereich fällt. Der Parameter β in Gleichung (2) charakterisiert die Steilheit der Funktionen, große Werte führen zu steilen, stufenähnlichen, kleine Werte dagegen zu flachen Ausgabefunktionen. Gebräuchlich sind $\beta = 1$ oder $\beta = \frac{1}{2}$. In einigen Fällen werden auch die Identität⁸, binäre Stufenfunktionen wie die Vorzeichenfunktion, stückweise lineare Funktionen oder die Cosinus-Quetschfunktion [Bra91] verwendet. Auch sind stochastische Funktionen denkbar; sie kommen allerdings im Gegensatz zum Hopfield-Modell in MSPs kaum zur Anwendung.

$$O_k = O_k(h_k) = \frac{1}{2} \tanh(\beta h_k) + \frac{1}{2} = \frac{1}{1 + e^{-2\beta h_k}} \quad (2)$$

Eine Ausgabefunktion nach (2) erweist sich in der Praxis als besonders günstig, da ihre später benötigte Ableitung bequem auf die Funktion selbst zurückgeführt werden kann, was eine programmtechnische Umsetzung erleichtert. Zudem ist die Wirkungsweise eines Sigma-Knotens in Verbindung mit dieser Ausgabefunktion anschaulich darstellbar. Im Grenzfall $\beta \rightarrow \infty$ geht die Ausgabefunktion (2) in eine Sprungfunktion über und liefert nur noch zwei Funktionswerte, Null oder Eins. Betrachten wir einen Sigma-Knoten in der l -ten Schicht des Perzeptrons. Er berechnet eine gewichtete Summe der N_{l-1} Ausgabesignale der $(l-1)$ -ten Schicht; anhand dieser Summe wird einer der beiden möglichen Funktionswerte der Sprungfunktion ausgewählt. Die Gesamtheit der N_{l-1} Ausgabesignale wird also in zwei Klassen aufgeteilt, eine Klasse, auf der die Sprungfunktion den Wert Null, und eine Klasse, auf der sie den Funktionswert Eins annimmt. Durch Umformung der gewichteten Summe in ein Skalarprodukt aus einem Aktivitäts- und einem Gewichtsvektor läßt sich eine Ebenengleichung erkennen, wodurch die Art und Weise, auf der diese Klassifikation erfolgt, geometrisch gedeutet werden kann: Die Gesamtheit der N_{l-1} Ausgabesignale bildet einen Punkt in einem N_{l-1} -dimensionalen Raum, alle zu dem betrachteten Knoten hinführenden Synapsen

⁷Die Differenzierbarkeit wird später im Rahmen der Lernregel benötigt.

⁸Oftmals bei Knoten in einer Ausgabeschicht, deren Ausgabesignal nicht beschränkt sein soll.

kennzeichnen eine Ebene in diesem Raum. Die Klassifikation erfolgt nun einfach danach, ob der Punkt auf der einen oder der anderen Seite dieser Ebene liegt. Für endliche β ist diese Klassifikation nicht streng, es erfolgt ein stetiger Übergang von der einen zur anderen Klasse; liegt das Muster der Ausgabesignale nahe bei der trennenden Ebene, so wird die Knotenausgabe einen Wert zwischen den beiden Sättigungswerten der sigmoiden Ausgabefunktion annehmen.

Durch geschickte Anordnung von versteckten Schichten unterschiedlicher Breite kann somit eine beliebige Manipulation der Dimension eines Eingabesignals erreicht werden, was anhand von Abbildung 2 exemplarisch skizziert werden soll. Abbildung 2 zeigt ein willkürlich gewähltes MSP mit einer Eingabeschicht von drei, zwei versteckten Schichten von acht und vier und einer Ausgabeschicht von zwei Knoten. In diesem Fall werden also dreidimensionale Muster verarbeitet. Jeder der acht Knoten der ersten versteckten Schicht konstruiert mit den drei Synapsen, die zu ihm hinführen eine Ebene in diesem dreidimensionalen Raum der Eingabemuster. Die acht Ausgabesignale dieser Knoten informieren damit darüber, auf jeweils welcher Seite der acht Ebenen - bzw. bei endlichem β wie nahe bei diesen Ebenen - sich das zur Zeit an der Eingabeschicht anliegende Muster befindet. Diese acht Signale bilden einen Punkt in einem achtdimensionalen Raum. Die vier Knoten der zweiten versteckten Schicht bilden nun erneut über die zu ihnen hinführenden Synapsen vier Ebenen in diesem achtdimensionalen Raum. Ihre Ausgabesignale legen nun fest, wo sich der Punkt aus der Gesamtheit der Ausgabesignale der ersten versteckten Schicht bezüglich dieser vier Ebenen befindet. Die Ausgabesignale der zweiten versteckten Schicht bilden schließlich einen Punkt in einem vierdimensionalen Raum, der durch die beiden Ausgabeknoten jeweils erneut durch eine Ebene geteilt wird. Die Ausgabesignale des Netzes bestimmen, wo sich der aus den vier Ausgabesignalen der zweiten versteckten Schicht gebildete Punkt bezüglich der beiden in diesem vierdimensionalen Raum aufgespannten Ebenen befindet. Durch entsprechende Wahl von Zahl und Breite der versteckten Schichten kann die Dimension der zu verarbeitenden Muster somit beliebig erhöht oder verringert und auch schwierige Klassifikationsprobleme - zumindest prinzipiell - durch MSPs gelöst werden.

Um Muster durch MSPs zu klassifizieren wird in der Regel die Breite der Ausgabeschicht entsprechend der Zahl der gewünschten Klassen gewählt. Verwendet man als Ausgabefunktion eine Fermifunktion, so wird die Zugehörigkeit eines Musters zur i -ten Klasse oftmals durch das Tupel $(0, 0, \dots, 0, 1, 0, \dots, 0)$ gekennzeichnet, wobei die Eins die i -te Komponente des Tupels kennzeichnet. Die Signale der Ausgabeknoten dagegen können bei endlichen β jedoch auch alle Werte zwischen Null und Eins annehmen, so daß sich die Netzausgabe entsprechend nicht nur auf ausgewählten Ecken, sondern auch im Inneren eines Einheitswürfels bewegen kann. Sollten also beispielsweise Muster in zwei Klassen durch die Netzausgaben $(1,0)$ und $(0,1)$ sortiert werden, so kann die Netzausgabe durchaus auch Werte wie etwa $(0,3,0,7)$ annehmen. Eine derartige Ausgabe ist nach [Ric91] bei überlappenden Musterklassen ggfs. als Bayessche Wahrscheinlichkeit zu deuten.

Um den Begriff der Bayesschen Wahrscheinlichkeit zu erläutern, betrachten wir nach Abbildung 3 zwei überlappende, eindimensionale gaußförmige Wahrscheinlichkeitsverteilungen $P(x|C_1)$ in rot und $P(x|C_2)$ in blau. Diese geben eine bedingte Wahrscheinlichkeit für das Auftreten eines Musters x an, unter der Voraussetzung, daß dieses der Klasse C_1 oder C_2 angehört. Klassifizieren eines Musters x bedeutet nun umgekehrt, dieses entweder der Klasse C_1 oder C_2 zuzuordnen. Da diese überlappen, ist dies nicht eindeutig möglich. Sinnvoll ist die Bestimmung der Bayesschen Wahrscheinlichkeiten $P(C_1|x)$ und $P(C_2|x)$. Beide Bayesschen Wahrscheinlichkeiten lassen sich aus den Wahrscheinlichkeitsverteilungen leicht bestimmen; Ableitungen allgemeiner Formeln zur Berechnung Bayesscher Wahrscheinlichkeiten sind in Standardwerken wie [Bro89] zu finden und seien hier übergangen. Anschaulich stellen die Bayesschen Wahrscheinlichkeiten in Abbildung 3 an jeder Stelle x den prozentualen Anteil der

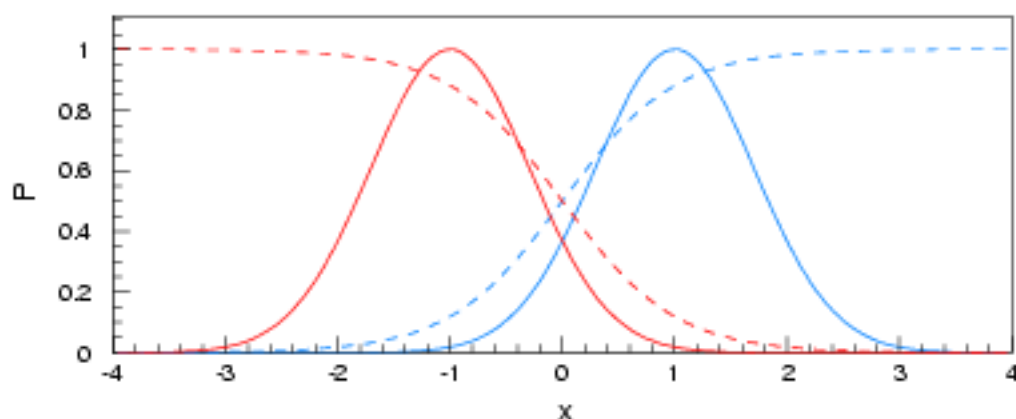


Abbildung 3: Zwei eindimensionale Wahrscheinlichkeitsverteilungen (durchgezogen) und die zugehörigen Bayesschen Wahrscheinlichkeiten (gebrochen).

jeweiligen Wahrscheinlichkeitsverteilung an der Summe aller Wahrscheinlichkeitsverteilungen an der Stelle x dar. Unter in [Ric91] diskutierten Voraussetzungen - wie u.a. einer ausreichenden Komplexität der verwendeten Netze oder adäquat gewählten Lerndatensätzen - kann die Ausgabe eines MSPs bei Klassifikation von überlappenden Musterverteilungen nun gerade die sich aus den Häufigkeitsverteilungen ergebenden Bayesschen Wahrscheinlichkeiten approximieren. Sofern dies der Fall ist, werden sich die Ausgabesignale der N_L Knoten der Ausgangschiicht zu Eins addieren. Weiterhin sollten die Erwartungswerte der Netzausgaben den N_L Klassenwahrscheinlichkeiten $P(C_i)$ entsprechen.

$$D = \sum_{i=1}^{N_L} P(C_i) \ln\left(\frac{P(C_i)}{O_i}\right) \quad (3)$$

Erstere können durch gemittelte Ausgabesignale \bar{O}_i der Ausgabeknoten genähert werden. Beispielsweise mit Gleichung (3) kann dann ein relativer entropischer Abstand zwischen gemittelten Netzausgaben und Klassenwahrscheinlichkeiten quantifiziert werden. Für Details sei auf [Ric91] verwiesen. Es bleibt anzumerken, daß diese Kriterien erfüllt sind, sofern die Netzausgabe einer Bayesschen Wahrscheinlichkeit entspricht, daß die Umkehrung jedoch nicht zulässig ist.

Neben ihrer Verwendung als Klassifizierer können MSPs auch zum Approximieren von Funktionen zwischen endlichdimensionalen Vektorräumen eingesetzt werden. Dazu existieren unterschiedliche Beweise, die einerseits eine, andererseits zwei versteckte Schichten verwenden. Referenzen hierzu sind beispielsweise in [Mü91, Bra91] oder in der Einleitung eines derartigen Beweises wie beispielsweise [Fun89] zu finden. Prinzipiell sind die mathematischen Voraussetzungen an die zu approximierenden Funktionen wie beispielsweise Stetigkeit gering. Im Wesentlichen basieren die Beweise auf der Zerlegbarkeit der zu approximierenden Funktion in einfachere Funktionen, welche durch versteckte Knoten eines MSPs erzeugt werden können. Das MSP stellt dann die zu approximierende Funktion umgekehrt als Überlagerung einer Vielzahl dieser einfachen Funktionen dar.

2.4 Error Backpropagation-Algorithmus

Nachdem der prinzipielle Aufbau eines mehrschichtigen Perzeptrons dargelegt wurde, ist schließlich eine Lernregel zu diskutieren. Unabhängig vom gewählten Modell eines künstlichen neuronalen Netzes bestimmt die Lernregel das Verfahren, mit dem die freien Parameter des Systems, die die Verbindungen der Knoten kennzeichnenden Gewichte, eingestellt werden. Im Rahmen eines überwachten Lernvorgangs wird dabei die - zunächst fehlerhafte - Netzausgabe extern bewertet und dem Netz diese Bewertung zurück gemeldet, so daß es seine Parameter korrigieren kann, um eine bessere Ausgabe zu erreichen. Im Rahmen unüberwachter Lernvorgänge muß ein Netz die Einstellung seiner Parameter selbständig anhand der Eingangsdaten ohne externe Beurteilung vornehmen.

Mehrschichtige Perzeptrons werden im allgemeinen durch einen überwachten Lernvorgang trainiert. Dieser setzt einen Trainingsdatensatz aus P Eingabegrößen und den dazugehörigen Musterlösungen voraus. Jede Eingabegröße sei ein - entsprechend der Breite der Eingabeschicht - N_1 -dimensionaler Vektor \underline{M}^μ mit Komponenten M_i^μ , jede dazu zu assoziierende Sollausgabegröße ein - entsprechend der Breite der Ausgabeschicht - N_L -dimensionaler Vektor \underline{S}^μ mit Komponenten S_i^μ . Der Lernvorgang ist vollständig überwacht; zu jedem Zeitpunkt wird die Leistung des Netzes auf dem Trainingsdatensatz ermittelt und Abweichungen der tatsächlichen Netzausgaben von den Sollausgaben zur Korrektur der Gewichte verwendet. Dazu wird eine Fehlerfunktion erstellt, welche während der Lernphase den Fehler des Netzes in Abhängigkeit der aktuellen Gewichte darstellt. Durch Gradientenabstieg oder eine andere Minimierungsmethode kann dieser Fehler durch Variation der Gewichte minimiert werden.

Prinzipiell sind alle den Fehler des Netzes messenden, differenzierbaren Funktionen denkbar, in der Praxis wird oftmals ein einfacher quadratischer Fehler nach Gleichung (4) verwendet. Der Faktor $\frac{1}{2}$ dient der Eliminierung des bei der folgenden Differentiation entstehenden Faktors Zwei. Die Normierung des Fehlers ist uneinheitlich, in dieser Arbeit wird er durch die Zahl der Trainingsmuster P und die Zahl der Knoten in der Ausgabeschicht N_L dividiert.

$$E = \frac{1}{2PN_L} \sum_{\mu=1}^P \sum_{i=1}^{N_L} (S_i^\mu - O_i^\mu)^2 \quad (4)$$

In Gleichung (4) hängt E über die Netzausgabe O_i^μ - das ist der Ausgabewert des i -ten Knotens der Ausgabeschicht bei Anlegen des μ -ten Trainingsmusters an die Eingabeschicht des Perzeptrons - von - nicht notwendigerweise allen - im Netz verwendeten Synapsen ab. Da die verwendeten Aktivitäts- und Ausgabefunktionen (1,2) stetig differenzierbar sind, kann (4) in beliebigen MSPs nach jedem gewünschten Gewicht differenziert werden. Mit zunehmender Zahl von Knoten in den versteckten Schichten und zunehmender Zahl der versteckten Schichten wird die analytische Differentiation jedoch schnell unübersichtlich.

Für den Spezialfall eines einfachen Perzeptrons, eines Perzeptrons ohne versteckte Schichten, ist die Gradientenbildung leicht durchzuführen und führt zu (5). Die Ableitung der Fehlerfunktion (4) nach einem synaptischen Gewicht ergibt sich - bis auf eine rechen technisch unkomplizierte innere Ableitung - als Summe über Produkte aus dem Signal, welches auf der präsynaptischen Seite der betreffenden Synapse anliegt und dem Fehler des Ausgabeknotens auf der postsynaptischen Seite.

$$\frac{\partial E}{\partial \omega_{ij}} = \frac{-1}{PN_L} \sum_{\mu=1}^P (S_i^\mu - g(h_i^\mu)) g'(h_i^\mu) M_j^\mu \quad (5)$$

Das Training einfacher Perzeptrons ist somit durch einen einfachen Gradientenabstieg leicht durchzuführen; theoretische Überlegungen zur Konvergenz des Verfahrens oder der maximalen Kapazität dieser Strukturen sind beispielsweise in [Her91] zu finden. Bedenkt man jedoch, daß jeder Knoten der Ausgangsschicht eines einfachen Perzeptrons nach obigen Überlegungen lediglich den Raum der Eingabemuster durch eine Ebene teilen kann, so wird klar, daß mit dieser Struktur nur linear separable Klassifikationsprobleme gelöst werden können. Dies wurde 1969 von Minsky und Papert diskutiert [Min69] und führte - da bereits so elementare Muster wie beispielsweise die Wahrheitstafel des exklusiven Oders nicht linear separabel sind - mangels eines geeigneten Algorithmus zur Gradientenberechnung in mehrschichtigen Perzeptrons - und vermutlich auch mangels geeigneter Rechner - dazu, daß das Interesse an einfachen Perzeptrons verloren ging.

Im Laufe der siebziger Jahre und zuletzt 1986 wurde mehrmals und unabhängig voneinander⁹ ein einfacher Algorithmus zur Gradientenbestimmung in beliebig komplexen mehrschichtigen Perzeptrons gefunden. Dieser Error Backpropagation-Algorithmus basiert auf einer geschickten Anordnung der durch iterative Anwendung der Kettenregel während der Differentiation entstehenden Terme. Seine Ableitung ist leicht und erfordert lediglich Fleißarbeit beim sortieren der einzelnen Terme. Sie sei hier übergangen; Interessenten können sie beispielsweise in [Her91, Rum86] nachvollziehen. Es zeigt sich, daß die Ableitung der Fehlerfunktion (4) nach einer beliebigen Synapse ähnlich wie in (5) wiederum - bis auf eine rechenstechnisch leicht handzuhabende innere Ableitung - als Summe über Produkte eines präsynaptischen Signals und eines postsynaptischen Fehlers geschrieben werden kann. Das präsynaptische Signal steht leicht nach Propagation eines Eingabemusters von der Eingabe- hin zur Ausgangsschicht zur Verfügung. Der postsynaptische Fehler eines Knotens in der Ausgangsschicht ist wiederum die Differenz aus Sollausgabe- und tatsächlichem Ausgabemuster des jeweiligen Ausgabeknotens. Der postsynaptische Fehler aller anderen Knoten innerhalb der versteckten Schichten im Perzeptron ergibt sich - und das ist der eigentliche Inhalt des Algorithmus - auf außerordentlich elegante Weise durch rückwärtige Propagation der Fehler der Knoten der Ausgangsschicht von der Ausgabe- hin zur Eingabeschicht des Netzes. Daher stammt auch der Name Error Backpropagation-Algorithmus. Somit liegt ein einfaches Schema vor, mit dessen Hilfe MSPs auf einen gegebenen Lerndatensatz trainiert werden können. Der Gradientenabstieg zur Minimierung von (4) vollzieht sich wie folgt: Innerhalb einer Schleife werden alle Muster des Trainingsdatensatzes nacheinander an die Eingabeschicht des Netzes angelegt und vorwärts durch das Netz propagiert. In der Ausgangsschicht wird für jeden Knoten die Differenz zwischen Soll- und tatsächlicher Knotenausgabe bestimmt und eine dieser Differenz proportionale Größe rückwärts durch das Netz bis zur ersten versteckten Schicht propagiert. Danach können alle Gradienten durch Produktbildung leicht ermittelt werden. Nach (4) erfolgt nach Summation dieser Gradienten über jeweils alle Muster des Trainingsdatensatzes ein Schritt des Gradientenabstiegs. Dazu werden alle Gewichte ω_{ij} um ein $\delta_{\omega_{ij}}$ nach Gleichung (6) variiert.

$$\delta_{\omega_{ij}} = -\eta \frac{\partial E}{\partial \omega_{ij}} \quad (6)$$

Die Proportionalitätskonstante η wird Lernrate genannt; sie ist ein Maß für die Größe der Schritte, die beim Abstieg durch das von (4) über dem Raum aller Gewichte aufgespannte Fehlergebirge vollzogen werden.

Obgleich dieser Algorithmus sehr leicht handzuhaben ist und daher gerne eingesetzt wird, existieren keinerlei allgemeingültige Beweise bezüglich seiner Konvergenzeigenschaften. Diese hängen von vielerlei Parametern und Details ab, auf die im nächsten Abschnitt hingewiesen werden soll.

⁹Referenzen hierzu sind z.B. in [Her91] zu finden.

2.5 Anmerkungen zur Anwendung

Will man in der Praxis mehrschichtige Perzeptrons auf einen gegebenen Datensatz trainieren, so sind eine Vielzahl weiterer Fragen zu klären.

- Welche Netztopologie soll verwendet werden?
- Welche Musterrepräsentation soll gewählt werden?
- Kann der Gradientenabstieg zwecks Konvergenzbeschleunigung modifiziert werden?
- Wie sollen die anfänglichen Synapsen gewählt werden?
- Wie viele Trainingsvorgänge sollen durchlaufen werden?
- Wann soll das Training abgebrochen werden?
- Wie kann die Generalisierungsfähigkeit eines Netzes, also seine Reaktion auf in der Lernphase nicht präsentierte Eingaben, bestimmt werden?

Die Topologie der verwendeten Netze entscheidet maßgeblich über Erfolg oder Mißerfolg eines Lernvorgangs. Ist das Netz zu klein, so verfügt es nicht über die zur Lösung der gestellten Klassifikations- oder Approximationsaufgabe notwendige Zahl freier Parameter bzw. kann diese nicht mit der gewünschten Genauigkeit erreichen. Ist das Netz dagegen zu groß gewählt und verfügt entsprechend über zu viele freie Parameter, so besteht die Gefahr, daß der Trainingsdatensatz, ein ihm überlagertes Rauschen oder andere Störungen auswendiggelernt werden. Wäre beispielsweise ein linearer Zusammenhang zwischen einer Ein- und einer Ausgabegröße zu approximieren, dem ein Rauschen überlagert ist, so kann ein überdimensioniertes Netz - ähnlich einem Polynom höherer Ordnung - dazu führen, daß die verrauschten Daten des Lerndatensatzes exakt reproduziert werden. Jedoch kann in diesem Fall der eigentliche lineare Zusammenhang möglicherweise nicht erkannt sein, was sich durch mangelnde Generalisierungsfähigkeit des Netzes - in diesem Fall etwa die Interpolationsfähigkeit zwischen gelernten Stützstellen - bemerkbar macht. Die Zahl der Knoten in Ein- und Ausgabeschicht werden durch Wahl einer Musterrepräsentation weitgehend festgelegt, eine optimale Zahl von versteckten Schichten sowie von versteckten Knoten ist im allgemeinen nicht bekannt. Zwar existieren - wie später angesprochen - Abschätzungen, die Obergrenzen für eine sinnvolle Zahl von Synapsen angeben, jedoch wird dadurch noch keine konkrete Netztopologie ausgezeichnet. In der Praxis haben sich zwei Verfahren herausgebildet, um iterativ eine problemangemessene Struktur zu finden. Einerseits kann von minimalen Strukturen ausgehend ein Netz schrittweise vergrößert werden, bis eine optimale Struktur erreicht ist, andererseits kann umgekehrt mit einem überdimensionierten Netz begonnen werden und dieses durch Beschneiden schrittweise verkleinert werden. Dabei ist zu beachten, daß eine Architektur unter verschiedenen Gesichtspunkten betrachtet werden muß und nicht bezüglich all dieser Gesichtspunkte gleichzeitig optimal wählbar sein wird; so können beispielsweise Rahmenbedingungen bestehen, die die Zahl der verwendeten Parameter als möglichst minimal, die Trainingszeiten als möglichst kurz oder die Generalisierungsfähigkeit als möglichst hoch festschreiben. Das Vergrößern bzw. Verkleinern der Netze selbst kann selbständig während des Lernvorgangs oder manuell nach abgeschlossenem Training erfolgen. So baut beispielsweise der Cascade Correlation-Algorithmus [Fah91] ein Netz während des Trainings auf oder es stehen verschiedene Methoden, Synapsen während des Trainings zerfallen zu lassen [Her91], zur Verfügung. Alternativ können auch unterschiedliche Netze nebeneinander trainiert und nach Abschluß der Lernphasen miteinander verglichen und dadurch schrittweise größere Architekturen erreicht werden. Zum Entfernen von Synapsen aus bereits trainierten Netzen stehen Algorithmen wie beispielsweise Optimal Brain Damage [LeC89]

oder der auch in dieser Arbeit eingesetzte Optimal Brain Surgeon-Algorithmus (OBS) [Has93] zur Verfügung¹⁰. OBS basiert auf folgender Überlegung: Eine Fehlerfunktion wie Gleichung (4) kann lokal in eine Taylorreihe entwickelt werden. Wurde unter Anwendung einer Lernregel ein Minimum gefunden, so verschwindet der Gradient in dieser Entwicklung. Alle Terme dritter und höherer Ordnung werden vernachlässigt, so daß ein Term zweiter Ordnung, eine mit der entsprechenden Hessematrix aller zweifachen partiellen Ableitungen gebildete quadratische Form, verbleibt. Ziel ist nun die Eliminierung eines Gewichtes, so daß der dadurch entstehende Anstieg der Fehlerfunktion minimal wird. Dieses Extremalproblem kann unter Einführung eines Lagrangemultiplikators gelöst werden; daraus ergibt sich ein Schema zur Bestimmung einer zur Eliminierung geeigneten Synapse. Dieses Schema setzt jedoch die Bestimmung der inversen Hessematrix voraus, was bei Netzen mit einigen hundert bis tausend oder mehr Synapsen zu rechentechnischen Problemen führt. OBS zeigt stattdessen ein iteratives Verfahren zur Approximation der benötigten Matrix durch Summation äußerer Produkte von Gradientenvektoren auf. Diese Gradientenvektoren stehen wiederum durch den Error Backpropagation-Algorithmus leicht zur Verfügung. Details sind in [Has93] zu finden und seien hier übergangen. Bestehend erscheint weiterhin die Möglichkeit, die Netztopologie durch einen genetischen Algorithmus optimieren zu lassen [Whi90, Ang94], jedoch erscheint der Rechenaufwand dazu - sofern komplexere als einfache Lehrbuch-Probleme betrachtet werden sollen - noch relativ hoch.

Liegen Daten vor, die mit einem Netz verarbeitet werden sollen, so ist zu klären, in welcher Repräsentation diese Daten dem Netz zur Verfügung gestellt werden sollen. In einfachen Fällen sind lediglich Normierungen oder Skalierungen erforderlich, sollen jedoch abstrakte Konzepte wie Wörter oder Sätze verarbeitet werden, muß erst eine geeignete Abbildung dieser Konzepte auf numerische Größen gefunden werden. Weiter ist zu klären, inwieweit eine Vorverarbeitung der Eingabedaten stattfinden soll. Diese ist entscheidend, kann sie doch die zur Lösung des Problems notwendige Information auslöschen, oder aber bereits selbst die gesamte Problemlösung beinhalten. Im ersten Fall kann das neuronale Netz das Problem nicht mehr lösen, im zweiten Fall muß es lediglich als Identität fungieren.

Um die Generalisierungsfähigkeit eines künstlichen neuronalen Netzes zu bestimmen, wird in der Praxis seine Leistungsfähigkeit auf einem vom verwendeten Lerndatensatz unabhängigen Kontrolldatensatz betrachtet. Da gerade zu groß gewählte Netze dazu neigen, mit fortschreitendem Training Trainingsbeispiele auswendig zu lernen, ist es erforderlich, die Entwicklung des Fehlers nach Gleichung (4) auf diesen beiden Datensätzen zu beobachten. Wenn der Fehler auf dem Kontrolldatensatz eine Sättigung erreicht hat, wird ein weiteres Training mit verbundener Abnahme des Fehlers auf dem Trainingsdatensatz nicht mehr zu einer Verbesserung der Generalisierungsfähigkeit beitragen. Es ist darauf zu achten, daß Lern- und Kontrolldatensatz voneinander unabhängig gewählt werden. Wird der Kontrolldatensatz zum Abbruch des Trainings oder zur Bewertung mehrerer und zur schlußendlichen Auswahl eines Netzes eingesetzt, so sollte eine abschließende Beurteilung der Generalisierungsfähigkeit auf einem weiteren, wiederum von Lern- und Kontrolldatensatz unabhängig gewählten dritten Datensatz erfolgen, da die Elemente des Kontrolldatensatzes in diesem Fall in Strenge nicht mehr als vom Lernprozeß unabhängig angesehen werden dürfen. Schließlich existieren Abschätzungen, die die Zahl der freien Parameter eines Netzes mit der Zahl der zu lernenden Muster des Trainingsdatensatzes in Beziehung setzen, um eine akzeptable Generalisierungsfähigkeit zu erreichen [Bau89, Wid63]. Eine resultierende Faustformel von Widrow besagt [Bau89], daß die Zahl der Muster im Trainingsdatensatz in etwa gleich der Zahl der Gewichte im Netz dividiert durch den gewünschten prozentualen Fehler in der späteren Generalisierungsphase sein sollte. So müßten bei einer erwünschten 90 prozentigen Quote richtiger Generalisierungen und entsprechend einer 10

¹⁰Eine Übersicht über weitere Algorithmen bietet [Res93].

prozentigen Fehlinterpretationsquote etwa zehn mal mehr Muster im Lerndatensatz als Gewichte im Netz vorhanden sein.

Die Dauer des Trainingsvorgangs sollte neben einer Obergrenze durch weitere Abbruchkriterien begrenzt werden. So erscheint es sinnvoll, das Training bei Eintritt sowohl des Fehlers auf dem Lern- als auch auf dem Kontrolldatensatz in eine Sättigung abubrechen. Im ersten Fall befindet sich der Algorithmus in einem flachen Bereich des Fehlergebirges, dessen Ausmaß in der Regel unklar ist, im zweiten Fall hat die Generalisierungsfähigkeit des Systems bereits ein Maximum erreicht und wird mit weiter fortschreitendem Training nicht mehr ansteigen, sondern kann im Gegenteil wieder abnehmen. Auch sind andere Abbruchkriterien, die etwa die zeitliche Variation der Synapsen messen, sowie Kombinationen mehrerer dieser Kriterien denkbar.

Generell besteht die Gefahr, daß der Lernvorgang in einem ungünstigen lokalen Minimum zum Stillstand kommt. Daher werden in der Praxis oftmals mehrere Startpunkte des Gradientenabstiegs in Form unterschiedlicher Startsynapsensätze zufällig gewählt und die Fehlerminimierung mehrmals durchgeführt. Die Häufigkeit lokaler Minima ist problemabhängig, tritt in der Praxis jedoch nicht in dem Maße auf, wie prinzipiell denkbar wäre [Her91]. Die Wahl geeigneter Startsynapsen ist nicht unkritisch [Lee93]. Zu große Startsynapsen führen dazu, daß sich die Knoten schon zu Beginn des Trainings im Sättigungsbereich der sigmoiden Ausgabefunktion (2) befinden. Dort liegen nur geringe Gradienten vor, so daß eine unnötig große Zahl von Schritten erforderlich ist, um aus diesem Sättigungsbereich heraus in den linearen Teil der Ausgabefunktion zu kommen. Gebräuchlich sind Werte im Intervall $[-1, +1]$ oder an der mittleren Zahl der im Netz zu einem Knoten führenden Synapsen orientierte Werte.

Schließlich wird in der Praxis der Gradientenabstieg zum Zwecke einer Konvergenzbeschleunigung modifiziert. Hierzu existiert eine Fülle von Vorschlägen, auf die an dieser Stelle nur hingewiesen werden kann. Führt man die Gradientenbestimmung nach Gleichung (5) in Strenge durch, so ist für jede Modifikation der Gewichte eine Summation über alle Muster des Trainingsdatensatzes durchzuführen. Diesen Modus bezeichnet man als Batch-Modus, die Präsentation aller Muster mit anschließender Synapsenmodifikation als eine Epoche. Um eine Beschleunigung zu erzielen, wird häufig auf die Summation verzichtet und stattdessen lediglich zufällig ein Muster des Trainingsdatensatzes ausgewählt und die Synapsen alleine aufgrund des mit diesem Muster bestimmten Gradienten modifiziert. Dieses Verfahren eines stochastischen Gradientenabstiegs bezeichnet man auch als Online-Modus, das einmalige Präsentieren eines Musters mit anschließender Synapsenmodifikation als Lernzyklus. Die zeitliche Entwicklung des Fehlers ist in diesem Fall nicht mehr monoton fallend, vielmehr ist der mittleren Fehlerabnahme eine unregelmäßige Oszillation überlagert.

Entscheidenden Einfluß auf die Konvergenzgeschwindigkeit des Algorithmus und damit die Dauer des Lernvorgangs hat die Wahl der Lernrate η . Ist sie zu klein, so dauert der Gradientenabstieg unverhältnismäßig lange, da nur kleine Schritte zum Durchwandern des Fehlergebirges verwendet werden. Ist sie dagegen zu groß gewählt, so besteht die Gefahr, daß Minima übersprungen werden. Daher ist es ratsam, eine variable Lernrate einzuführen. In der einfachsten Form wird diese dazu im Laufe des Trainingsvorgangs immer weiter reduziert, um gegen Ende der Lernphase ein gefundenes Minimum nicht durch einen grossen Schritt wieder zu verlassen. Effektiver erscheint jedoch, die Lernrate der jeweils lokalen Struktur des Fehlergebirges anzupassen. Dabei wird nach jedem Schritt die tatsächlich erfolgte Fehleränderung bestimmt. Nimmt der Fehler dabei sukzessive ab, so wird die Lernrate erhöht. Sobald eine Vergrößerung des Fehlers festgestellt wird, wird die Lernrate reduziert. Allerdings sollte erwähnt werden, daß die ständige Kontrolle des Fehlers zusätzliche Rechenzeit

erfordert. Die Änderung der Lernrate selbst wird dabei durch neue Größen parametrisiert [Her91]. Mit ähnlicher Absicht wird oftmals auch als weiterer Parameter ein additiver Impulsterm in Gleichung (6) eingeführt. Dabei wird zu jeder Synapsenmodifikation ein gewisser Bruchteil der zeitlich unmittelbar vorangegangenen Synapsenmodifikation zusätzlich addiert. Dies führt zu einer Verkleinerung der effektiven Schrittweite in zerklüfteten Gebieten, zu einer effektiven Vergrößerung in langgestreckten Tälern des Fehlergebirges. Eine andere Technik zur Minimierung einer vielparametrisierten Funktion wie (4) besteht darin, nicht permanent entlang des jeweils aktuellen Gradienten, sondern immer stückweise entlang einer geraden Linie im Fehlergebirge zu gleiten. Ist eine Minimierung entlang einer Geraden abgeschlossen, so wird eine neue Gerade gewählt und der Vorgang wiederholt. Gewöhnlich wird zur Bestimmung der Suchrichtung wiederum der Gradient verwendet, so daß jeweils aufeinander folgende Richtungen notwendigerweise senkrecht aufeinander stehen. Dadurch wird das Minimum über einen Zick-Zack förmigen Weg erreicht [Her91]; eine günstigere Wahl der Suchrichtungen stellt die Methode der konjugierten Gradienten dar [Bar92]. Neben der ersten Ableitung der Fehlerfunktion können auch höhere Ableitungen verwendet werden, beispielsweise nach dem Newtonschen Verfahren zur Extremwertbestimmung. Allerdings ist die Berechnung der in diesem Fall entscheidenden Hessematrix aller zweifachen partiellen Ableitungen von Gleichung (4) - insbesondere für größere Netze mit einigen tausend oder mehr Parametern - sehr rechenaufwendig. Interessanter erscheint der von Fahlman vorgeschlagene QuickProp-Algorithmus, der die Fehlerfläche lokal quadratisch nähert und sich zum Minimum einer dort konstruierten Parabel begibt [Fah88]. Auch sind Minimierungsmechanismen denkbar, die ohne jegliche Gradientenmethode auskommen. So kann beispielsweise wie im Rahmen des Threshold Accepting-Algorithmus [Duc90] oder von Random Optimization Methods [Sun90] das Fehlergebirge auf stochastischem Wege abgetastet werden oder aber die Wahl geeigneter Gewichte kann durch einen genetischen Algorithmus erfolgen [Whi90].

Eine spürbare Konvergenzbeschleunigung wird auch durch die Einführung einer für jedes Gewicht im Netz individuellen Lernrate erreicht. Diese - ebenfalls zeitlich variablen - Lernraten können bequem aufgrund des Vorzeichenverhaltens der jeweiligen Komponente des Gradienten modifiziert werden. So wird die Lernrate einer Synapse erhöht, wenn kein Vorzeichenwechsel deren partieller Ableitung erfolgt, umgekehrt erfolgt eine Lernratenverminderung bei Auftreten eines Vorzeichenwechsels während zweier sukzessiver Schritte. Die Anpassung der Lernraten wird dabei durch weitere Parameter bestimmt. Im Rahmen der Delta-Bar-Delta Lernregel erfolgt eine lineare Erhöhung, aber ein exponentieller Abfall der Lernraten [Jac88], im Rahmen des SuperSAB-Algorithmus erfolgen beide Änderungen exponentiell [Rie92]. Die tatsächliche Synapsenänderung ist jedoch nach wie vor der jeweiligen partiellen Ableitung proportional, so daß geringe Gradienten in flachen Gebieten des Fehlergebirges zu kleinen Synapsenänderungen und damit zu kleinen effektiven Schritten führen. Um derartige Probleme zu umgehen, kann eine vom Betrag der partiellen Ableitung unabhängige Synapsenänderung durchgeführt werden; je nach Vorzeichenverhalten der partiellen Ableitungen werden die Synapsen um eine konstante Größe modifiziert. Diese konstante Größe ist jedoch im Rahmen des ursprünglichen Manhattan-Trainings für alle Synapsen gleich und zeitlich konstant [Rie92]. Der Resilient Propagation-Algorithmus (RPROP) kombiniert die Vorteile beider Szenarien: Er verwendet zur Modifikation jeder Synapse eine eigene Größe, welche je nach Vorzeichenverhalten der betreffenden partiellen Ableitung exponentiell erhöht oder verringert wird. In dieser Arbeit wird hauptsächlich der RPROP-Algorithmus eingesetzt. Für weitere Details bezüglich RPROP sei auf [Rie94, Rie92], bezüglich der angedeuteten sowie weitergehender Modifikationen des Gradientenabstiegs auf [SNN94] oder Lehrbücher wie [Her91] verwiesen.

3 Approximation am Beispiel eines ortsempfindlichen Teilchenzählers

In diesem Kapitel wird untersucht, inwiefern sich die skizzierten mehrschichtigen Perzeptrons in der Praxis zur Approximation einer nur auf Stützstellen bekannten Funktion verwenden lassen. Dies erfolgt am Beispiel der Eichung eines von Kai Vetter et al. entwickelten Parallelplattenzählers PYRAMIDE [Vet94]. Dazu werden im folgenden zunächst Aufbau und Funktionsweise der PYRAMIDE skizziert und das Problem der Eichung des Zählers erläutert. Darauf werden die untersuchten mehrschichtigen Perzeptrons vorgestellt, ihre Lernphasen dokumentiert und exemplarisch das Beschneiden eines trainierten Netzes durch den Optimal Brain Surgeon-Algorithmus vorgestellt.

3.1 Aufbau und Funktionsweise des Detektors

In Schwerionestreuexperimenten ist zur Bestimmung der Kinematik eine Messung der Streuwinkel der Reaktionspartner erforderlich. Insbesondere im Rahmen von γ -Spektroskopie nach direkten Kernreaktionen in der Nähe der Coulombbarriere wie z.B. Coulombanregung oder Mehrkuleonentransfer werden positionsempfindliche Teilchenzähler zur Bestimmung des Stoßparameters und zur Dopplerkorrektur der γ -Spektren eingesetzt.

Der Parallelplattenzähler PYRAMIDE wurde für den Betrieb in Verbindung mit dem 4π EUROBALL-I Gammaskpektrometer GASP in Legnaro bei Padua konzipiert. Um den interessierenden Raumwinkel maximal zu überdecken, wurde die Geometrie einer dreiseitigen Pyramide gewählt, wobei die drei sensitiven Seitenflächen trapezförmig ausgebildet sind, siehe Abbildung 4. Ein aus der Reaktions-

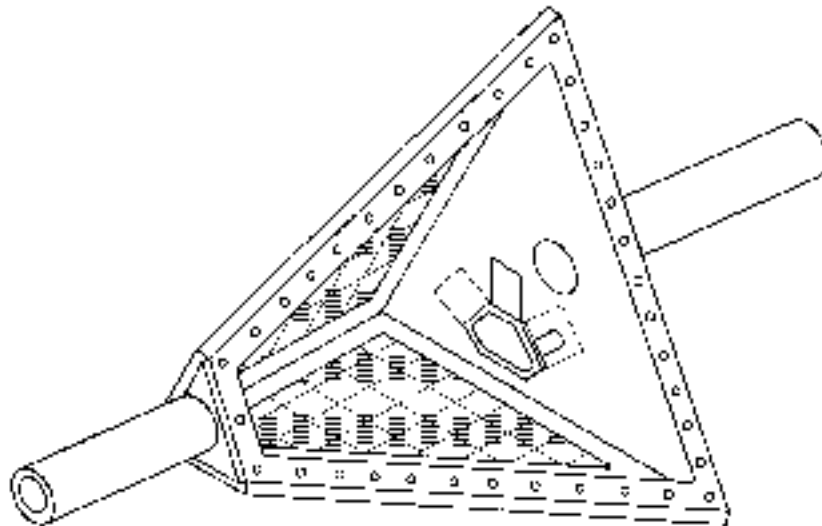


Abbildung 4: Blick in den geöffneten Parallelplattenzähler PYRAMIDE.

zone kommendes Teilchen durchquert zunächst eine dünne Eintrittsfolie, welche das den Zähler durchspülende Iso-Butan Gas vom Vakuum des Beschleunigers trennt. Zwischen einer aus einem dünnen Gitter gebildeten Wolframkathode und einer 4 mm entfernten als Anode dienenden Widerstandsschicht liegt eine Hochspannung. Einfallende Ionen werden auf dieser Strecke beschleunigt und ionisieren in einem Primärprozeß Gasmoleküle. Die freigesetzten Elektronen werden ebenfalls beschleunigt; in einem Sekundärprozeß entsteht eine Lawine. Die über die trapezförmige Widerstandsschicht fließende Ladung teilt sich je nach Ort auf vier Komponenten auf, welche schließlich über die vier Kanten des Trapezes abfließen und nachgewiesen werden. Aus diesen vier gemessenen Signalen soll nun auf den ursprünglichen Ort des nachgewiesenen Teilchens rückgeschlossen werden. Aufgrund der Asymmetrie des Trapezes ist ein analytischer Zusammenhang zwischen gemessenen Signalen und Ort durch Lösung der Poissongleichung nicht mehr darstellbar, wie beispielsweise im Falle von quadratischen Geometrien. Alleine aus diesem Grunde muß nach Methoden zur Konstruktion einer geeigneten Abbildung gesucht werden. Darüberhinaus zeigte sich, daß das Aufdampfen der Widerstandsschicht in der Praxis nicht mit der benötigten Homogenität erfolgt, daß die genaue Struktur der Inhomogenitäten unbekannt ist und daß diese darüberhinaus von Zähler zu Zähler variiert.

Somit ist eine Eichung jedes der drei Zähler erforderlich. Dazu kann die sensitive Fläche eines Zählers mit einer Lochmaske bedeckt und der Zähler gleichmäßig ausgeleuchtet werden, siehe Abbildung 5. Um aus den gemessenen Signalen näherungsweise eine Ortsinformation zu extrahieren, können die Gleichungen (7, 8) verwendet werden. Dabei wird der Ort in horizontaler bzw. vertikaler Richtung durch Differenzbildung der in der jeweiligen Richtung gemessenen Ladungen bestimmt. Auf diese Weise entsteht ein verzerrtes Bild der Lochmaske nach Abbildung 5.

$$x = \frac{q_2 - q_4}{q_2 + q_4} \quad (7) \quad y = \frac{q_1 - q_3}{q_1 + q_3} \quad (8)$$

In der Abbildung sind jeweils zwanzig Teilchen pro Loch der 91 Löcher der Lochmaske zu sehen. Man erkennt deutlich, daß besonders an den Rändern Nichtlinearitäten vorliegen und daß nach der Parametrisierung einer geeigneten entzerrenden Abbildung zu suchen wäre. Dieses erweist sich in der Praxis als schwierig, so daß stattdessen im folgenden Abschnitt mehrschichtige Perzeptrons darauf trainiert werden, die zur Erzeugung von Abbildung 5 verwendeten Ladungsquadrupel direkt auf die gewünschten Ortskoordinaten abzubilden.

3.2 Rahmenbedingungen des Trainings

Da die Bilder der einzelnen Löcher der Lochmaske in Abbildung 5 - wenn auch an den Rändern mit Mühe - zu trennen und die realen Positionen aller 91 Löcher bekannt sind, können den in der Eichmessung gemessenen Ladungsquadrupeln die jeweiligen bekannten Sollpositionen der Löcher manuell zugeordnet werden. Eine Vorverarbeitung der gemessenen Ladungen fand nur insofern statt, als daß die Summe aus den vier pro nachgewiesenem Teilchen gemessenen Ladungen auf eins normiert wurde. Eine weitergehende Verarbeitung, wie etwa eine Reduktion der Dimension des Ladungstupels, wurde nicht untersucht. Die gewünschten Sollpositionen $(x, y) \in [-5cm, +5cm] \times [-4cm, +4cm]$ wurden auf einen Teilbereich des Einheitsquadrates $[-0.8, +0.9] \times [-0.6, +0.6]$ skaliert, so daß auch in den Ausgabeschichten der Netze der Tangens Hyperbolicus als Ausgabefunktion eingesetzt werden konnte. Ebenso hätte hier auch die Identität verwendet werden können, wobei die Skalierung der Sollpositionen entfiel. Dieser Datensatz aus reellen vierdimensionalen Eingabe- und zweidimensio-

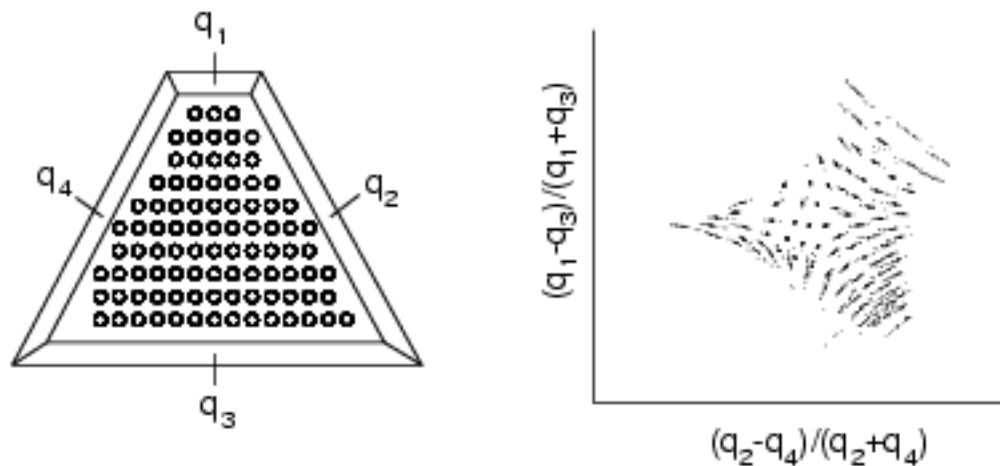


Abbildung 5: Schematische Darstellung der auf den Zähler aufgelegten Lochmaske und der vier Elektroden (links), sowie das mit Gleichungen (7, 8) gewonnene Bild des ausgeleuchteten Zählers (rechts).

nenal Ausgabedaten wurde in einen Lerndatensatz aus 50 Löchern zu je 20 Ereignissen und einen Kontrolldatensatz von 41 Löchern zu ebenfalls je 20 Ereignissen aufgeteilt. Der Lerndatensatz dient dem Training der MSPs, der Kontrolldatensatz zur Untersuchung der Generalisierungsfähigkeit der Netze während und nach der Trainingsphase¹¹. Generalisierungsfähigkeit bedeutet in diesem Fall, daß ein Netz zwischen im Lerndatensatz enthaltenen Punkten zu interpolieren oder sogar darüberhinaus zu extrapolieren vermag.

Somit war es naheliegend, der Struktur der Datensätze entsprechend MSPs mit einer vier Knoten breiten Eingabeschicht und einer zwei Knoten breiten Ausgabeschicht zu untersuchen. Eine geeignete Zahl und die ideale Breite der versteckten Schichten sind zunächst unbekannt, daher werden unterschiedliche Netze untersucht, wobei von einem minimalen Netz ausgehend schrittweise größere Strukturen betrachtet werden. Neben der minimalen 4-2 Struktur ohne versteckte Schicht, welche - wie nicht anders zu erwarten war - nicht in der Lage ist, die gewünschte Abbildung in der geforderten Genauigkeit darzustellen, wurden sieben weitere MSPs der Topologien 4-4-2, 4-7-2, 4-10-2, 4-15-2, 4-4-4-2, 4-7-7-2 und 4-10-10-2 untersucht. Mehr als zwei versteckte Schichten wären zwar denkbar und von dem aufgestellten Programm zur Simulation der Perzeptrons auch ohne Schwierigkeiten handhabbar gewesen. Da die zur Verfügung stehende Rechenzeit jedoch begrenzt ist und sich zwei versteckte Schichten als bereits ausreichend erwiesen haben, wurden keine komplexeren Strukturen untersucht. Daß die Zahl der Knoten in den beiden Schichten aller verwendeten Netze jeweils identisch ist, ist willkürlich; ebenso hätten weniger symmetrische Topologien gewählt werden können. Um den Effekt des Auswendiglernens zu demonstrieren, werden absichtlich Strukturen wie das 4-10-10-2 Netz betrachtet, welches mit $(4 + 1) \times 10 + (10 + 1) \times 10 + (10 + 1) \times 2 = 182$ freien Parametern im

¹¹In Strenge ist der Kontrolldatensatz dadurch nicht mehr unabhängig vom Lerndatensatz zu sehen, da er zur Bewertung der Netze - sei es zum Abbruch des Trainings, sei es zur Auswahl einer geeigneten Topologie - verwendet wurde.

Vergleich zu den 980 Mustern des Trainingsdatensatzes bereits als in seiner Größe kritisch erscheint. Gleichzeitig kann die Wirkung des Optimal Brain Surgeon Algorithmus zum Beschneiden des Netzes demonstriert werden.

Zum Training der Netze wurde der skizzierte RPROP-Algorithmus eingesetzt. Um darüberhinaus einen Vergleich mit einem stochastischen Gradientenabstieg und dem gewöhnlichen Batch-Modus zu erhalten, wurden für die 4-10-10-2 Struktur exemplarisch auch ein Online- und ein Batch-Training durchgeführt. In allen Fällen erfolgte die Berechnung der benötigten Gradienten mit dem Error Backpropagation-Algorithmus. Für jedes Netz wurden fünf Startpunkte durch zufällige Wahl der Startsynapsen aus dem Intervall $[-1, +1]$ willkürlich gewählt. Startsynapsen größerer Beträge wurden nicht verwendet, um frühzeitige Sättigungseffekte zu vermeiden; allerdings wurden keine systematischen Untersuchungen zur Konvergenz des Lernverfahrens in Abhängigkeit von der Startsynapsengröße durchgeführt.

Jede Lernphase wurde zunächst auf maximal 10^4 Epochen - im Rahmen des Online-Trainings Lernzyklen - begrenzt. Darüberhinaus wurde der Lernvorgang abgebrochen, wenn die Differenz des Fehlers auf dem Trainingsdatensatz zwischen jeweils 100 Epochen einen kritischen Wert von 10^{-7} nicht mehr unterschritt. In diesem Fall befindet sich der Algorithmus in einem sehr flachen Teil des Fehlergebirges. Es ist zwar nicht auszuschließen, daß ein günstiges Minimum nach Durchqueren dieses Bereiches noch erreicht werden könnte, aus Gründen der Rechenzeit erscheint es jedoch nicht als praktikabel, mit einem derartigen Szenario zu rechnen, sondern lieber einen neuen Startpunkt im Fehlergebirge zu bestimmen und einen neuen Gradientenabstieg vorzunehmen. Aus der Änderung des Fehlers auf dem Lerndatensatz kann nicht uneindeutig auf die Änderung des Fehlers auf dem Kontrolldatensatz geschlossen werden. Sofern der Fehler auf dem Lerndatensatz eine Sättigung erreicht hat, wird auch der Fehler auf dem Kontrolldatensatz nicht weiter abnehmen, jedoch trifft die Umkehrung nicht zu. Daher sollten generell die zeitlichen Entwicklungen beider Fehler betrachtet werden, um einem Auswendiglernen des Lerndatensatzes vorzubeugen.

Die Bewertung der Netze erfolgt neben einer Betrachtung der Fehler auf Trainings- und Kontrolldatensatz auch durch die der Verteilung der geometrischen Fehler der Netzausgabe, die nach Umrechnung der Signale der Ausgabeknoten in Zentimeter als Abstände zwischen dem tatsächlichen Ort und dem vom Perzeptron ausgegebenen Ort bestimmt wird. Darüberhinaus kann die Netzausgabe in einer Ebene dargestellt werden, so daß ein Vergleich der Güte der Netze mit der einfachen Faustformel (7) und (8) über Abbildung 5 möglich ist.

3.3 Netzstrukturen und Ergebnisse

Abbildung 6 zeigt die zeitliche Entwicklung des Fehlers auf dem Trainingsdatensatz mit durchgezogener, die des Fehlers auf dem Kontrolldatensatz mit gebrochener Linie für jeweils fünf unterschiedlich gewählte Startsynapsensätze. Dabei werden drei Netze gegenübergestellt; 4-4-2 in rot, 4-15-2 in grün und 4-10-10-2 in blau; zum Training wurde der RPROP-Algorithmus verwendet. Kleinere Netze mit wenigen Parametern zeigen generell schneller den typischen Sättigungseffekt als größere Strukturen mit mehr freien Parametern. Da größere Netze mit kürzeren Lernphasen zu vergleichbaren Fehlern kommen, erscheint es lohnend, auch diese Netze zu untersuchen, um Trainingszeiten zu reduzieren^{1,2}.

^{1,2}Dabei muß jedoch beachtet werden, daß mit zunehmender Zahl von versteckten Knoten auch die pro Epoche erforderliche Rechenzeit ansteigt und daß nach erfolgtem Training unnötige Parameter eliminiert werden sollten.

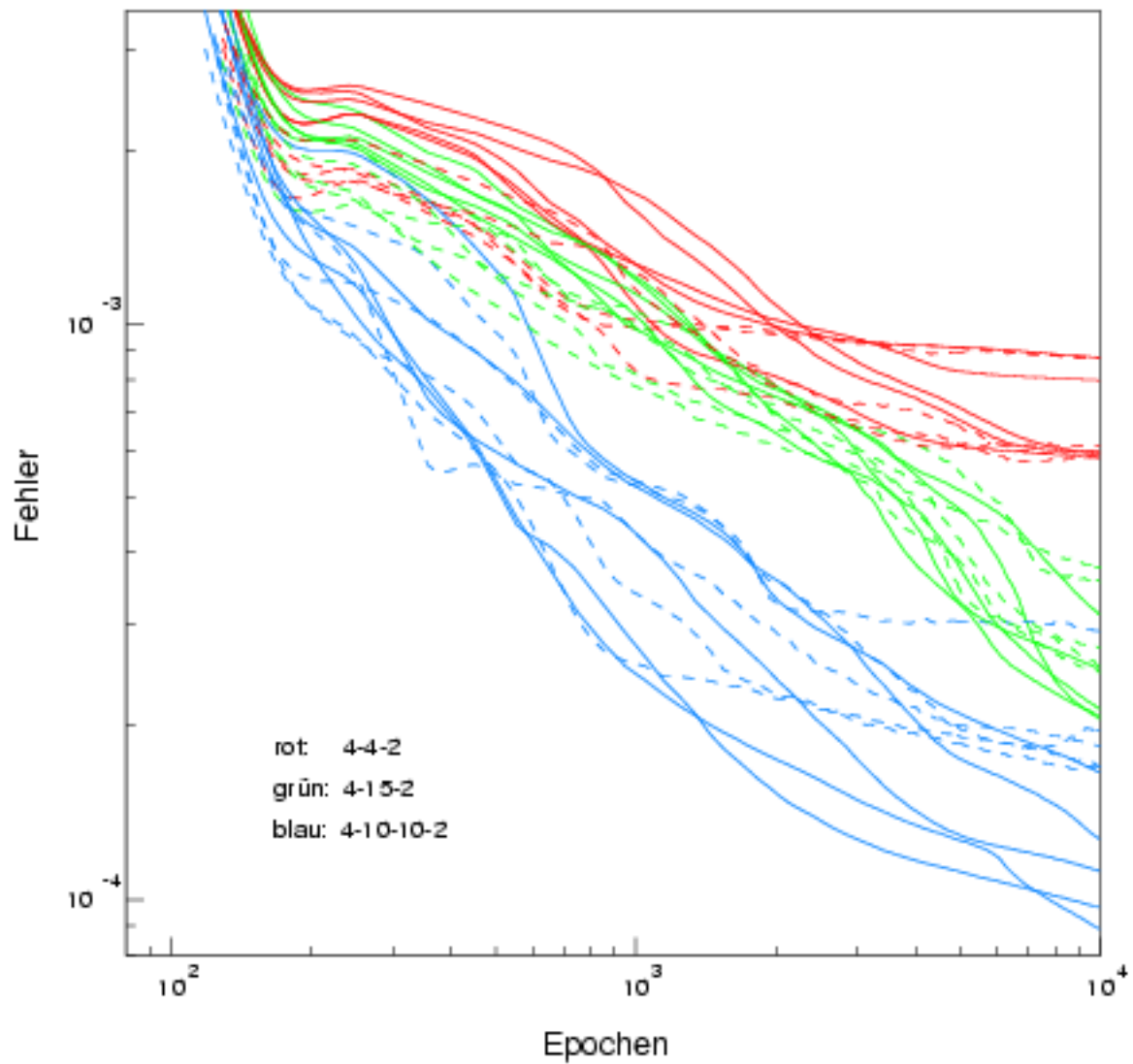


Abbildung 6: Zeitliche Abnahme des Fehlers auf dem Lerndatensatz (durchgezogen) und Kontrolldatensatz (gebrochen) dreier durch den RPROP-Algorithmus trainierter Netze für jeweils fünf unterschiedliche Startpositionen im Fehlergebirge.

Die Schwankungsbreite der fünf unterschiedlichen Läufe jeweils eines Netzes beträgt nach 10^4 Epochen etwa zwei Zehntel der Größenordnung des Fehlers. Damit sind die einzelnen Netze anhand ihrer Fehlerentwicklungen in Abbildung 6 noch zu trennen und es besteht noch nicht die Gefahr, durch eine ungünstige Startsynapsenwahl und der damit verbundenen Fehlerentwicklung die Netze falsch zu beurteilen. Unabhängig von der Wahl eines konkreten Trainingslaufes eines Netzes wird man zu der Aussage gelangen, daß der Fehler der 4-4-2 Struktur größer als der der 4-15-2 und dieser wiederum größer als der der 4-10-10-2 Struktur ist.

Aus den fünf pro Topologie trainierten Netzen wurden jeweils die mit dem geringsten Fehler auf dem Trainingsdatensatz weiter trainiert und wiederum die Entwicklung des Fehlers auf Lern- und Kontrolldatensatz betrachtet. Es zeigt sich, daß erwartungsgemäß die Diskrepanz zwischen Fehlerentwicklung auf Lern- und Kontrolldatensatz bei allen größeren Netzen mit vielen freien Parametern mit fortschreitender Dauer der Lernphase zunimmt. Dabei fällt der Fehler während zusätzlicher 250000 Trainingsepochen auf dem Trainingsdatensatz für die Netze 4-10-2, 4-7-7-2 und 4-10-10-2 auf $8,6 \times 10^{-5}$, $7,6 \times 10^{-5}$ bzw. $2,8 \times 10^{-5}$, wohingegen der Fehler auf dem Kontrolldatensatz während dieser Zeit nur unwesentlich von 4×10^{-4} auf knapp 2×10^{-4} absinkt. Dies ist ein Zeichen, daß der Trainingsvorgang während der zusätzlichen Epochen keine Verbesserung der Generalisierungsfähigkeit der Netze mehr erreicht. Allerdings nimmt diese während dieses Zeitraums nicht ab, wie in einem späteren Beispiel in Kapitel 4.2 zu beobachten ist. Exemplarisch werden die beiden Fehlerkurven für das 4-10-10-2 Netz in Abbildung 7 gezeigt, wobei darüberhinaus zwei weitere Variationen der Fehlerminimierung - der Batch-Modus mit einer adaptiven Lernrate und der Online-Modus - dem sonst in diesem Kapitel verwendeten RPROP-Algorithmus gegenübergestellt werden. Der Fehler auf dem Trainingsdatensatz ist in rot, der auf dem Kontrolldatensatz in blau eingezeichnet. Man erkennt deutlich, daß unter Verwendung des RPROP-Algorithmus ab etwa 3×10^4 Epochen keine weitere Verringerung des Fehlers auf dem Kontrolldatensatz mehr erreicht wird und das Training an dieser Stelle abgebrochen werden kann. Ebenfalls klar zu sehen ist der Vorteil des RPROP-Algorithmus gegenüber dem herkömmlichen Batch-Betrieb, obwohl dieser durch Einführung einer adaptiven Lernrate im Vergleich zu einer sonst konstanten Lernrate beschleunigt wurde. Die anfängliche Lernrate betrug $\frac{1}{2}$, sie wurde nach drei sukzessiven Schritten mit kontinuierlicher Fehlerverringerng auf 120 Prozent ihres Wertes vergrößert, sobald ein Schritt jedoch zu einer Fehlererhöhung führte, wurde sie auf die Hälfte ihres aktuellen Wertes reduziert. Das eingesetzte Diagramm zeigt die zeitliche Entwicklung der Lernrate, man erkennt eine auch nach 200000 Epochen noch nicht abgeschlossene Oszillation. Daraus folgt, daß noch kein endgültiges Minimum im Fehlergebirge erreicht ist¹³, was auch durch die noch deutlich fallende Fehlerkurve des Trainingsdatensatzes gestützt wird. Der stochastische Gradientenabstieg konvergiert nicht erst sehr spät, wie Abbildung 7 suggerieren könnte. Vielmehr muß in Rechnung gestellt werden, daß im Online-Modus bereits nach Präsentation nur eines Musters die Synapsen aktualisiert werden, so daß hier nur ein Bruchteil des Rechenaufwandes vorliegt. Allerdings muß in diesem Fall aufgrund der unregelmäßigen Fehlerentwicklung ständig eine Fehlerberechnung durchgeführt werden, um das jeweils aktuelle Minimum der Fehlerkurve zu bestimmen. Zur Berechnung des Fehlers müssen wiederum sämtliche Muster des Lern- bzw. Kontrolldatensatzes durch das Netz propagiert werden, was im Rahmen des monoton fallenden Fehlers im Batch-Modus nicht notwendigerweise nach jeder Epoche durchzuführen ist. Daher ergibt sich nicht die zunächst erhoffte deutliche Verringerung der aufgewendeten Rechenzeit¹⁴.

Neben den Fehlern auf Trainings- und Kontrolldatensatz können zur Kontrolle der Netze - da die

¹³Eine analoge Kurve in [Vet94] für das Training eines 4-7-7-2 Netzes zeigt das Erreichen eines Minimums durch einen Abfall der Lernrate.

¹⁴Allerdings besteht an dieser Stelle noch Potential für zukünftige Optimierungen.

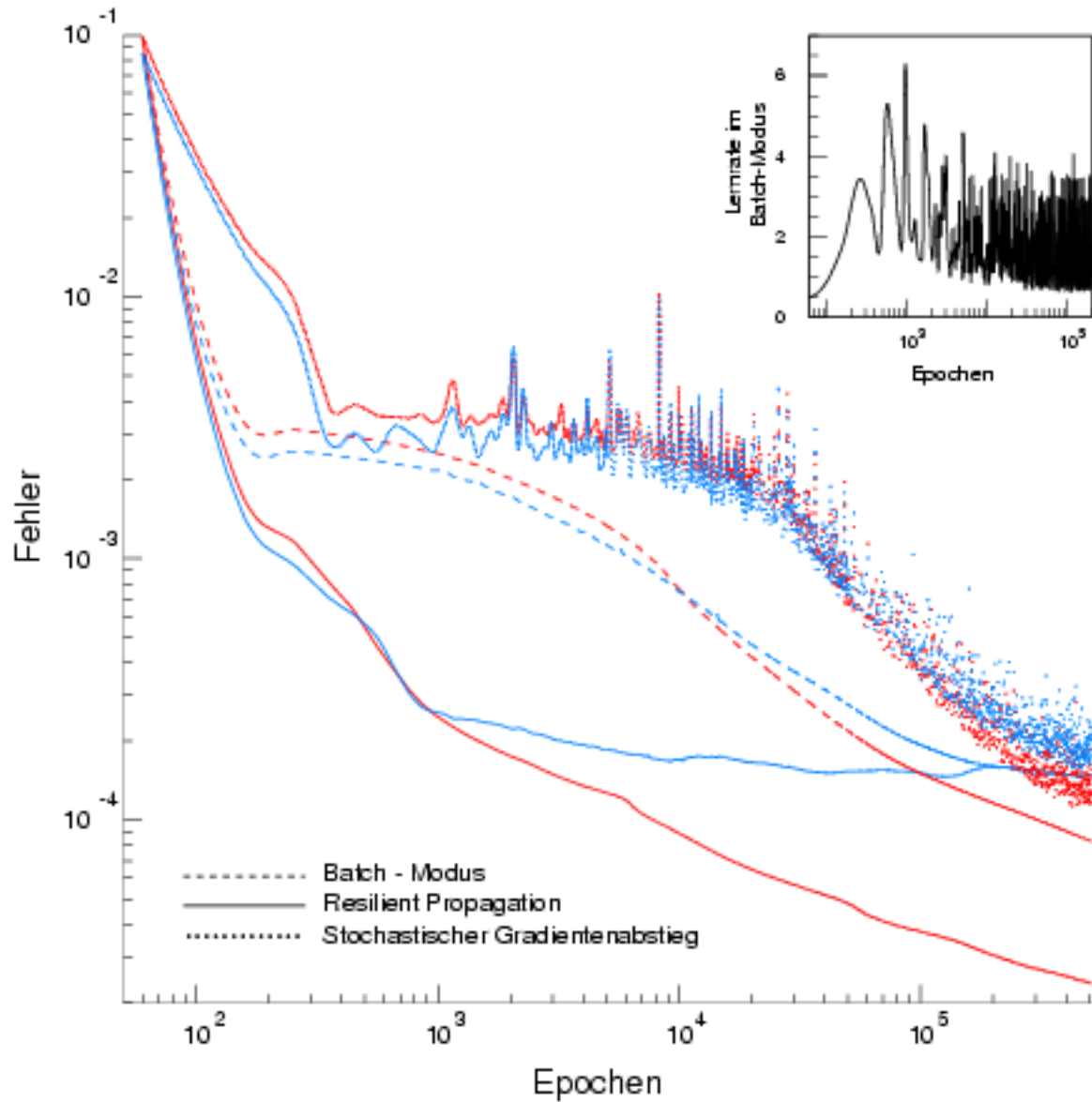


Abbildung 7: Zeitliche Entwicklung des Fehlers auf dem Trainingsdatensatz (rot) und Kontrolldatensatz (blau) für ein 4-10-10-2 Netz für drei Variationen des Gradientenabstiegs. Im Falle des stochastischen Gradientenabstiegs zeigt die Abszisse Lernzyklen. Oben rechts: Entwicklung der Lernrate η im Batch-Modus.

Netzausgabe ja einen zweidimensionalen Ort darstellt - auch direkt die Signale der Ausgabeknoten in einer Ebene dargestellt werden. Zusätzlich kann zu jedem so bestimmten Punkt auch der räumliche Abstand zur zugrundeliegenden Sollposition berechnet werden. Abbildung 8 zeigt wiederum für das 4-10-10-2 Netz neben der Netzausgabe auch die - bezüglich des Radius korrigierten - Häufigkeitsverteilungen dieser Abstände jeweils für Lern- und Kontrolldatensatz für fünf ausgewählte Zeitpunkte der Lernphase. Nicht abgebildet ist der Beginn der Lernphase, in dem die Bilder der einzelnen Löcher der Lochmaske nicht zu trennen sind. Bereits nach nur fünf Epochen sind die Strukturen der Lochmaske deutlich zu erkennen; nach etwa 10^3 Epochen sind diese soweit auseinandergezogen, daß sie die ganze Fläche des Trapezes - insbesondere die Ecken und Randbereiche - ausfüllen. Nach etwa 3×10^4 Epochen - zu dem Zeitpunkt, ab dem keine weitere Verringerung des Fehlers auf dem Kontrolldatensatz mehr eintritt - ist eine ausreichende Qualität erreicht. Auch bei fortschreitendem Lernvorgang ändert sich das Bild nicht mehr signifikant. Das durchgezogene Histogramm zeigt jeweils die Verteilung der Abstände auf dem Lern-, das durchbrochene Histogramm die Verteilung der Abstände auf dem Kontrolldatensatz. Während der ersten 3×10^4 Epochen rücken sowohl der Schwerpunkt der Verteilung auf dem Kontroll- als auch der der Verteilung auf dem Lerndatensatz zu kleineren Werten nach links. Ab diesem Zeitpunkt jedoch behält die Fehlerverteilung des Kontrolldatensatzes ihre Form bei, wohingegen die des Lerndatensatzes weiter nach links geschoben wird, was deutlich nach 5×10^5 Epochen zu sehen ist. Somit stützt auch diese Darstellung die Vermutung, daß nach etwa 3×10^4 Epochen des Trainings des 4-10-10-2 Netzes mit dem RPROP Algorithmus bei einem Fehler von knapp 2×10^{-4} auf dem Kontrolldatensatz das Training zu beenden ist und optimale Gewichte gefunden sind. Es ist nicht auszuschließen, daß der Lernvorgang an dieser Stelle in einem lokalen Minimum der Fehlerfunktion (4) zum Stillstand kam, jedoch entspricht das durch die dortige Synapsenkombination charakterisierte Netz allen Anforderungen bezüglich der darzustellenden Abbildung, so daß eine weitere Suche nach einem etwaigen globalen Minimum nicht erforderlich ist.

Mit den oben mit unterschiedlichen Netzen angegebenen minimal erreichbaren Fehlern von etwa 2×10^{-4} auf dem Kontrolldatensatz und der Tatsache, daß auch für unterschiedliche Minimierungsverfahren im 4-10-10-2 Netz dieser Fehler nicht wesentlich unter diesen Wert gedrückt werden konnte, ist zu vermuten, daß dieser Fehler einen optimalen Wert darstellt. Netze wie die 4-4-2 Struktur, die diesen Wert nicht erreichen, sondern bei etwa 6×10^{-4} in die Sättigung eintreten, verfügen über zu wenige freie Parameter. Bei Netzen mittlerer Größe wie der 4-4-4-2 oder der 4-7-2 Struktur wird dieser Wert erreicht, ohne daß ein Auseinanderlaufen der beiden Fehlerkurven von Trainings- und Kontrolldatensatz zu beobachten wäre; die Zahl ihrer freien Parameter erscheint dem Problem angepaßt. Größere Netze wie insbesondere die 4-7-7-2 und 4-10-10-2 Strukturen sind in der Lage, den Lerndatensatz auswendig zu lernen. Jedoch ist dabei keine Beeinträchtigung der Generalisierungsfähigkeit zu beobachten, so daß der Effekt in diesem Fall unkritisch erscheint. Weiter ist zu berücksichtigen, daß diese größeren Netze schneller konvergieren als die mittlerer Größe. Da Methoden zur Elimination überflüssiger Parameter aus fertig trainierten Netzen zur Verfügung stehen, kann es in diesem Fall lohnend sein, ein derartig überdimensioniertes Netz zu trainieren und anschließend zu beschneiden, was nun für den Fall des 4-10-10-2 Netzes demonstriert wird.

Dazu wurde das in Abbildung 7 trainierte 4-10-10-2 Netz durch den Optimal Brain Surgeon-Algorithmus beschnitten. Dabei wurde in jedem Schritt eine Synapse entfernt, indem ihr Wert auf Null gesetzt wurde. Die im Netz verbleibenden Synapsen wurden dabei durch OBS in jedem Schritt leicht korrigiert. Dieser Vorgang sollte nur solange iterativ wiederholt werden, bis ein nicht mehr tragbarer Anstieg des Fehlers festzustellen ist. Dessen ungeachtet zeigt Abbildung 9 den Anstieg des Fehlers auf dem Trainings- und Kontrolldatensatz in rot bzw. blau bis zum Entfernen der letzten noch im Netz verbleibenden Synapse. Insgesamt besteht das System aus 182 Parametern, es kann nicht

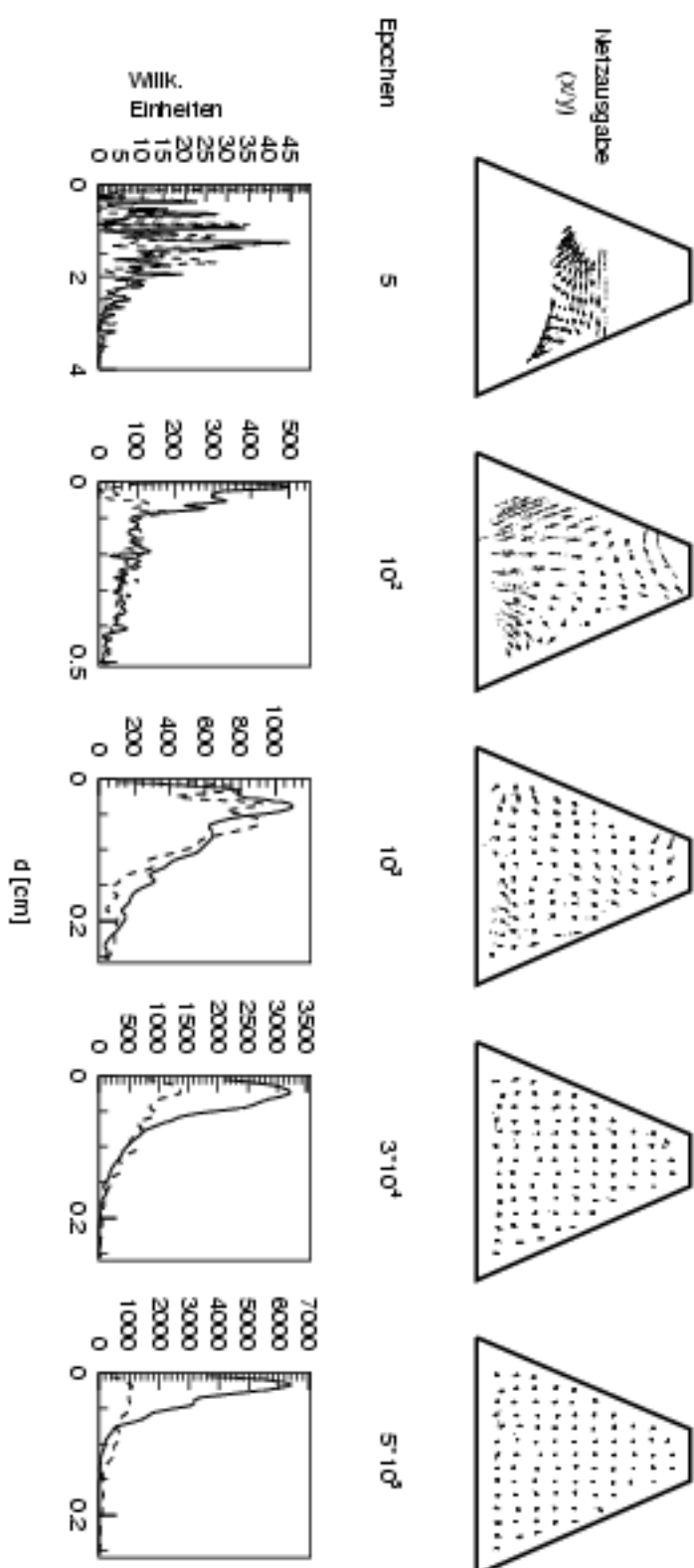


Abbildung 8: Zeitliche Entwicklung von Netzausgabe (oben) und korrigierter Verteilung der Abstände zwischen Soll- und Istposition (unten) auf Lerndatensatz (durchgezogen) und Kontrolldatensatz (gestrichelt).

erwartet werden, daß alle Gewichte eliminiert werden können - in diesem Falle wäre die Netzausgabe identisch Null. Das übermäßige Beschneiden wird nur durchgeführt, um den mit fortschreitender Elimination von Parametern verbundenen graduellen Verlust der in den Synapsen festgehaltenen Information über die gesuchte Abbildung zu demonstrieren.

Die Entfernung der ersten zehn Parameter vollzieht sich ohne Auswirkung auf den Fehler von Trainings- oder Kontrolldatensatz. Nach Entfernen von zwölf Gewichten steigt der Fehler des Trainingsdatensatzes sprunghaft auf den des Kontrolldatensatzes an, nach 18 entfernten Parametern ist ein Anstieg um eineinhalb Größenordnungen festzustellen. Nach 40 eliminierten Synapsen zeigt sich ein erneuter Anstieg um eine Größenordnung. Durch ein erneutes Training dieser beschnittenen Strukturen¹⁵ ist der Ausgangsfehler sowohl auf Trainings- als auch auf dem Kontrolldatensatz nach nur 10^2 bzw. 10^3 Epochen wieder hergestellt. Nach Elimination von 50, 80 oder gar 100 Synapsen steigt der Fehler auf Werte, die vor Beginn des Trainings bei zufälliger Wahl der Startsynapsen vorlagen. Ein erneutes Training dauert daher auch wieder etwa 10^4 Epochen, bis erneut ein akzeptabler Fehler erreicht wird¹⁶. Werden 100, 120 oder gar 150 Synapsen entfernt, so führt auch ein erneutes Training nicht mehr zu einem akzeptablen Fehler. Offenbar liegen durch die an diesen Stellen eliminierten Synapsen derartige Randbedingungen vor, daß die verbleibende Struktur die gewünschte Abbildung nicht mehr darstellen kann. Prinzipiell wären Netze mit der an diesen Stellen noch verbleibenden Zahl freier Parameter von 80 oder 60 durchaus dazu in der Lage, jedoch unterscheiden sich diese in ihrer Vernetzung von den hier beschnittenen Strukturen.

Darüberhinaus wurden vier weitere der untersuchten Netze durch OBS beschnitten. Als Ausgangspunkt dienten dazu die Synapsensätze der mehrerer durchgeführter Fehlerminimierungen nach 10^4 Epochen. Auf diese Weise konnten aus den 51 Parameter umfassenden 4-7-2 Netzen 9, 9, 12, 13 und 14, aus den 72 Parameter umfassenden 4-10-2 Netzen 17, 17, 18, 20 und 21 Synapsen entfernt werden. Aus den 50 Parameter umfassenden 4-4-4-2 Netzen wurden 6, 9, 12 und 15, aus den 4-7-7-2 Netzen mit 107 Parametern dagegen 3, 22, 23 und 30 Parameter entfernt. Schließlich konnten aus den 182 Parameter umfassenden 4-10-10-2 Netzen 30, 40 bzw. 60 Synapsen entfernt werden. Damit verbleiben im Falle der 4-7-2 und 4-4-4-2 Netze etwa 40, im Falle der 4-10-2 Netze etwa 55 Parameter. Bei den größeren 4-7-7-2 und 4-10-10-2 Netzen ist kein einheitliches Verhalten zu beobachten, die Zahl verbleibender Parameter schwankt zwischen etwa 80 und 150. Es ist der Trend erkennbar, daß in ursprünglich großen Netzen generell mehr Parameter verbleiben als in ursprünglich kleinen Netzen, was dadurch gedeutet werden könnte, daß in größeren Netzen die Approximation der gesuchten Funktion über mehr Parameter verteilt ist als in kleinen Netzen. Jedoch bleibt anzumerken, daß die Netze nicht iterativ erneut trainiert und erneut beschnitten wurden. Daher geben die genannten Daten lediglich einen groben Anhaltspunkt für die Zahl überflüssiger Parameter, eine kritische Mindestzahl von Synapsen darf aus ihnen nicht abgeleitet werden.

Schließlich wurde untersucht, inwieweit das 4-10-10-2 Netz auch über weitere Entfernungen als die Abstände der bisher zum Training verwendeten Punkte festlegten, zu interpolieren und darüberhinaus zu extrapolieren vermag. Dazu wurde der Datensatz aus gemessenen Ladungen und Sollausgabepositionen erneut aufgeteilt: Ein Datensatz umfaßt dabei Positionen aus der Mitte, der andere solche aus den Randbereichen des Detektors. Die Ausgabefunktion der Knoten in der Ausgabeschicht wurde durch die Identität ersetzt, so daß die Netzausgabe nicht mehr prinzipiell durch den Wertebereich

¹⁵ währenddessen die eliminierten Gewichte natürlich auf Null gehalten werden -

¹⁶ Es ist anzumerken, daß dies nicht im Sinne des OBS-Algorithmus ist, da nach [Has93] das Beschneiden bereits dann zu beenden ist, wenn ein deutlicher Anstieg des Fehlers zu beobachten ist, was im vorliegenden Fall nach 18, spätestens aber nach 30 eliminierten Synapsen gegeben wäre.

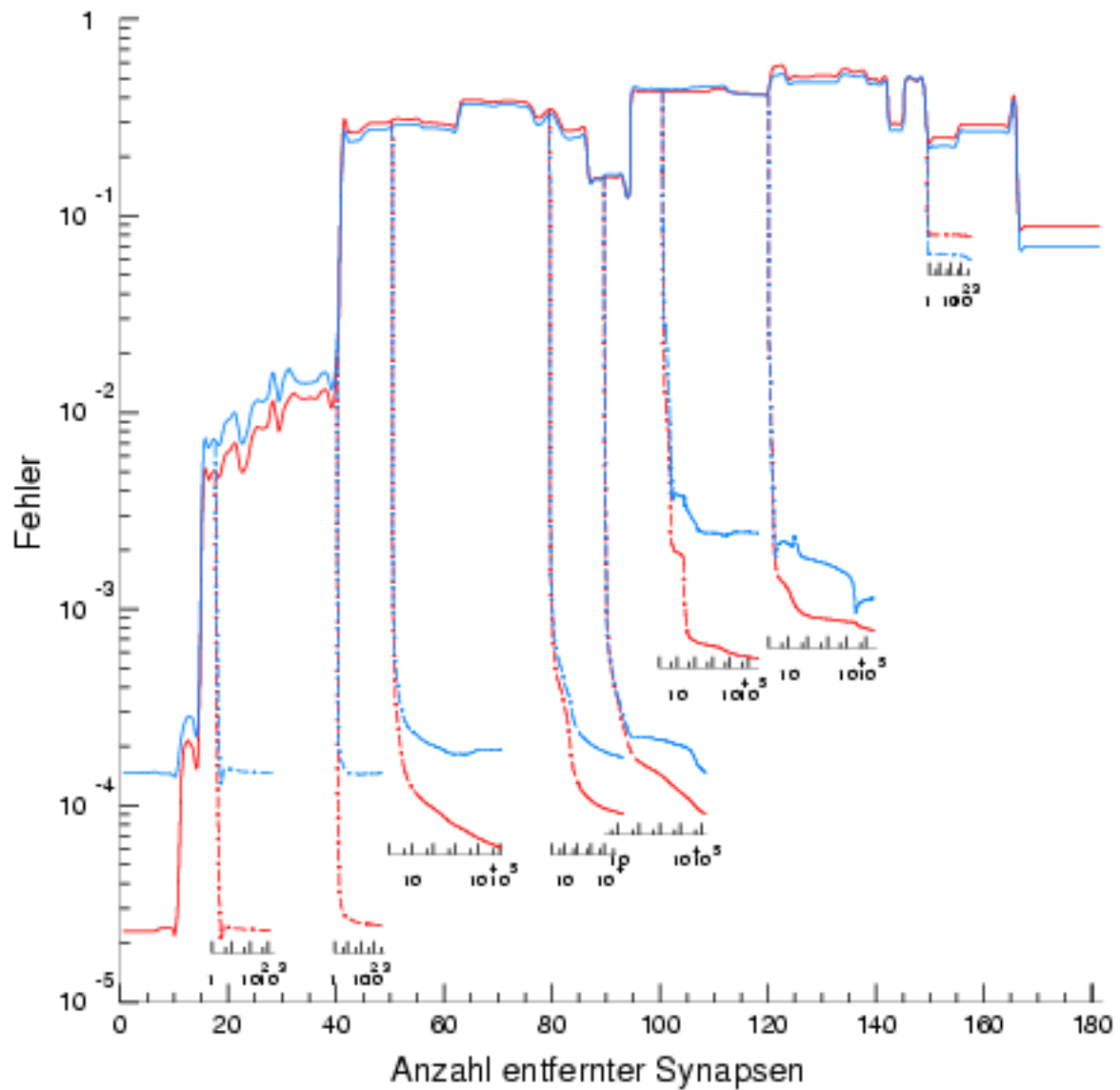


Abbildung 9: Anstieg des Fehlers bei Entfernen von Synapsen durch den OBS-Algorithmus und Abfall des Fehlers auf Lerndatensatz (rot) und Kontrolldatensatz (blau) bei erneutem Training. Trainingszeiten in Epochen.

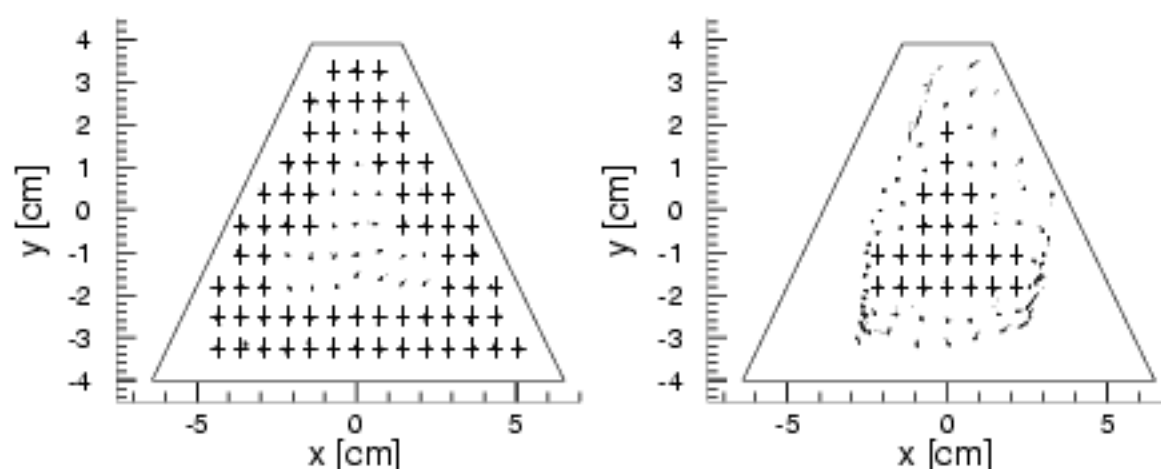


Abbildung 10: Netzausgabe bei Training auf den durch Kreuze markierten Positionen.

einer sigmoiden Ausgabefunktion begrenzt wird. Als Sollpositionen dienten die Koordinaten der Meßpunkte in Zentimetern. Abbildung 10 zeigt links die Netzausgabe für den Fall, daß zum Training des Netzes ausschließlich die durch die Kreuze markierten Punkte aus dem Randbereich, rechts dagegen die durch die Kreuze markierten Punkte aus dem inneren Bereich der aktiven Fläche des Zählers verwendet wurden. Es ist deutlich zu erkennen, daß das MSP zwischen den gelernten Punkten zu interpolieren vermag und daß die Extrapolationsfähigkeiten begrenzt sind. Eine Extrapolation über die gelernten Positionen hinaus ist im oberen und unteren Teil des Zählers gegeben, in den Ecken jedoch nicht mehr vorhanden. Da prinzipiell jedoch Meßpunkte der gesamten aktiven Zählerfläche zum Training zur Verfügung stehen, stellt dies keinerlei Einschränkung der Leistungsfähigkeit dar.

Abschließend ist festzuhalten, daß sich die untersuchten mehrschichtigen Perzeptrons zur Eichung des Parallelplattenzählers PYRAMIDE verwenden lassen. Netze mit etwa zehn bis fünfzehn versteckten Knoten reichen aus, um die gewünschte Abbildung von gemessenen Ladungen auf zweidimensionale Ortskoordinaten darzustellen. Die Wahl der anfänglichen Startsynapsen erweist sich nicht als kritisch. Die Dauer des Lernvorgangs bei 980 Mustern im Lerndatensatz hängt vom verwendeten Netz ab und beträgt für ein 4-10-10-2 Netz etwa 3×10^4 Epochen. Die benötigten Rechenzeiten sind unkritisch, vergleiche mit Anhang A. Abbildung 11 zeigt detailliert die Netzausgabe sowohl auf dem ursprünglichen Lern- als auch auf dem ursprünglichen Kontrolldatensatz. Die Kreuze kennzeichnen die zu lernenden Sollpositionen. Zur Bestimmung einer mittleren Abweichung der vom Netz ausgegebenen Position zur jeweiligen Sollposition wurde die Häufigkeitsverteilung der Abstände zwischen diesen Positionen verwendet. Die Radiusabhängigkeit dieser Verteilung wurde durch eine entsprechende Korrektur berücksichtigt. Der Schwerpunkt der sich daraus ergebenden Verteilung liegt für den Lerndatensatz bei 0.7 Millimetern, für den Kontrolldatensatz bei 1.1 Millimetern. Dies entspricht einem Fehler von etwa einem Prozent.

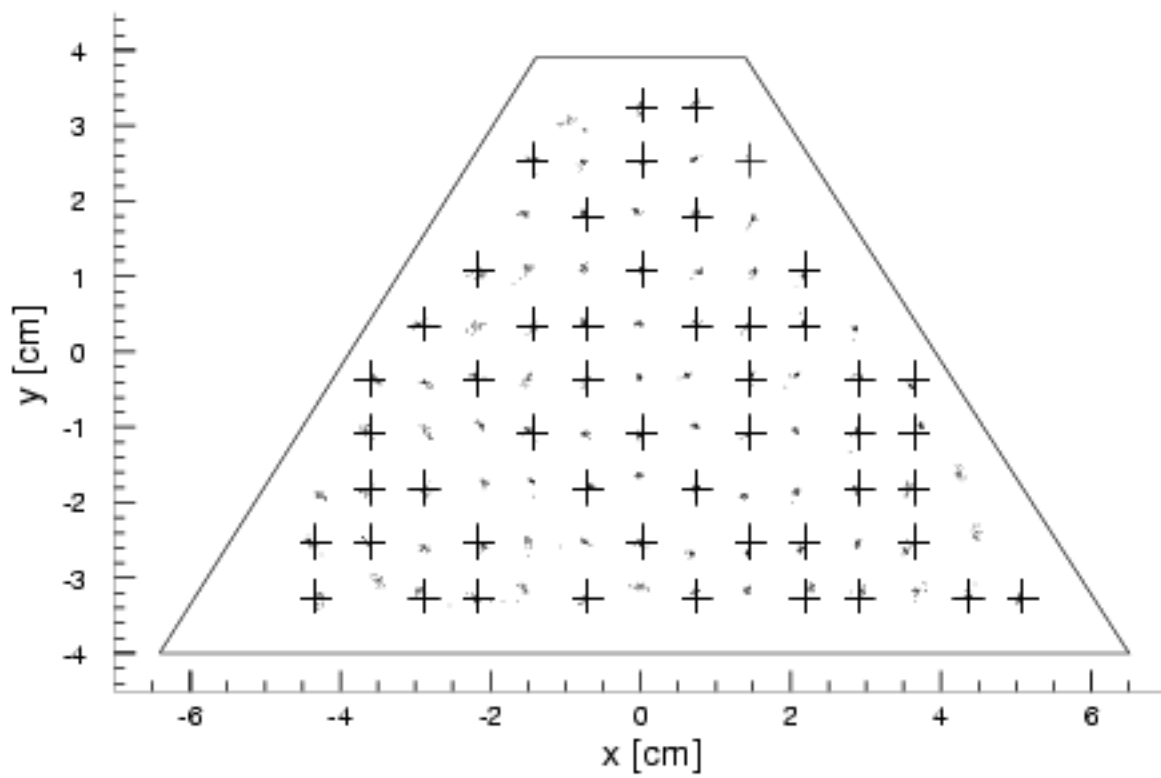


Abbildung 11: Netzausgabe auf durch Kreuze markiertem Lerndatensatz und Kontrolldatensatz.

4 Musterklassifikation am Beispiel segmentierter γ -Detektoren

Nachdem mehrschichtige Perzeptrons im vorherigen Kapitel zur Approximation einer unbekanntes Abbildung eingesetzt wurden, soll nun untersucht werden, inwieweit sie sich zur Musterklassifikation eignen. Dazu werden in diesem und den kommenden Abschnitten Fälle betrachtet, in denen Energiesignale von segmentierten Germaniumdetektoren zu verarbeiten sind.

Bei schwerioneninduzierten Kernreaktionen bewegen sich die zu spektroskopierenden Kerne mit Geschwindigkeiten von einigen Prozent der Lichtgeschwindigkeit. Dies führt durch den Doppler-Effekt zu einer Verschiebung der wahren Energie ausgesandter γ -Quanten und zu einer entsprechenden Verbreiterung der γ -Linien. Um diese Effekte korrigieren zu können, muß der Ort der ersten Wechselwirkung der γ -Quanten mit dem Detektor möglichst eng eingegrenzt werden. Dazu werden die eingesetzten Germaniumdetektoren - beispielsweise bei EUROBALL-III - segmentiert¹⁷. Aufgrund der Segmentierung machen sich allerdings unvermeidbare Comptoneffekte störend bemerkbar, welche zu Streuungen von γ -Quanten zwischen benachbarten Segmenten eines Kristalls, zwischen benachbarten Kristallen oder sogar zu einer Streuung aus dem Detektor heraus¹⁸ führen können. Verbleibt das Quant innerhalb des Detektors, so kann seine Energie über verschiedene Segmente eines Kristalls oder auch über mehrere Kristalle verteilt werden. Die zeitliche Reihenfolge des Ansprechens der einzelnen Segmente kann meßtechnisch nicht aufgelöst werden; es ist somit unklar, in welchem Segment die erste Wechselwirkung des γ -Quants mit dem Detektor stattgefunden hat und welche Koordinaten somit zur Dopplerkorrektur verwendet werden sollen. Somit ist es interessant zu untersuchen, inwieweit aus einem gemessenen Energiemuster ein Rückschluß auf das Segment gezogen werden kann, in dem die erste Wechselwirkung stattfand. Dazu sind die Muster in Klassen einzuteilen. Die Zahl der Klassen entspricht der Zahl der Segmente; wird ein Muster einer Klasse zugeordnet, so bedeutet dies, daß das entsprechende Segment zuerst anspricht. Die zur Dopplerkorrektur zu verwendenden Ortskoordinaten sind dann an diesem Segment zu orientieren. Es ist zu beachten, daß eine in allen Fällen eindeutige Klassifikation aufgrund durch den Comptoneffekt entstehender Mehrdeutigkeiten physikalisch nicht möglich ist; vielmehr ist die Zahl der richtigen Klassifikationen zu maximieren.

Darüberhinaus tritt im Rahmen von Hochspinexperimenten das Problem der Mehrfachtreffer auf. Bei typischen γ -Multiplizitäten von etwa 30 besteht bei EUROBALL-III eine Wahrscheinlichkeit von mehreren Prozent, daß zwei γ -Quanten verschiedene Segmente eines Germaniumdetektors treffen. Dieser Fall eines Mehrfachtreffers sollte vom Falle eines einzelnen, gestreuten γ -Quants unterschieden werden; erneut sind gemessene Energiemuster zu klassifizieren.

Die interessierenden Detektorgeometrien können mit Hilfe des Simulationsprogramms GEANT [GEA93] durch Monte Carlo-Simulationen untersucht werden. Dabei werden gewünschte Geometrien definiert, mit Materialeigenschaften versehen und schließlich γ -Quanten beliebiger Energien durch diesen Aufbau bewegt. Die erforderlichen Wechselwirkungen der Quanten mit Materie, wie Photo-, Compton- oder Paareffekt werden entsprechend ihrer Wirkungsquerschnitte berücksichtigt. Somit sind im Rahmen einer derartigen Simulation sowohl die Flugbahn der Quanten als auch die in den einzelnen Segmenten deponierten Energien bekannt und können zur Konstruktion von Trainings- und Kontrolldatensätzen für die untersuchten MSPs verwendet werden.

¹⁷Die Segmentierung der zylinderförmigen Kristalle erfolgt entlang der ϕ -Koordinate, darüberhinaus kann eine zusätzliche Information über die radiale Komponente des Ortes der ersten Wechselwirkung aus einer Analyse der Pulsform der Energiesignale gewonnen werden [Pet94].

¹⁸In diesem Fall können sie durch aktive Kollimatoren, Anti-Compton Abschirmungen, nachgewiesen werden [Fl⁹³].

4.1 Bestimmung des ersten Segments

In den folgenden Abschnitten werden die Signale segmentierter Germaniumdetektoren mit dem Ziel klassifiziert, das Segment zu bestimmen, in dem die erste Wechselwirkung des nachgewiesenen γ -Quants mit dem Detektormaterial stattfand. Dazu werden nach einleitenden Überlegungen zur Comptonstreuung ein zweifach, ein vierfach und schließlich ein 16fach segmentierter Germaniumdetektor, der SEGMENTED CLOVER-Detektor [Ger92], betrachtet.

4.1.1 Physikalische Anmerkungen

Wird die Energie eines γ -Quants über mehrere Segmente eines Detektors verteilt, so kann die erste Wechselwirkung kein Photoeffekt gewesen sein, da das einfallende Quant in diesem Fall vernichtet und seine Energie lokal in nur einem Segment deponiert worden wäre. Vielmehr muß ein Comptoneffekt aufgetreten sein. Um ein prinzipielles Verständnis der im Rahmen einer Comptonstreuung von γ -Quanten zwischen benachbarten Segmenten auftretenden Effekte zu bekommen, wird die zu untersuchende Geometrie des aus vier Kristallen zu je vier Segmenten bestehenden SEGMENTED CLOVER-Detektors zunächst auf den wesentlich einfacheren Fall eines zweifach segmentierten Zylinders aus Germanium reduziert. Die gewonnenen Resultate lassen sich dann auf einen zunächst vierfach segmentierten Kristall und schließlich auf den SEGMENTED CLOVER übertragen.

Wir betrachten einen zylinderförmigen Germaniumkristall, der gemäß Abbildung 12 in zwei Segmente unterteilt sei; γ -Quanten sollen auf der Stirnfläche einfallen und für die folgenden Überlegungen nicht aus dem Detektor herausgestreut werden. Erleidet ein γ -Quant der Energie E eine Comptonstreuung unter einem Winkel θ , so läßt sich die Energie E' des Quants nach der Streuung durch Gleichung (9) beschreiben. Dabei bedeutet m_0c^2 die Ruheenergie des Elektrons von 511 keV.

$$E' = \frac{E}{1 + \frac{E}{m_0c^2}(1 - \cos(\theta))} \quad (9)$$

Um eine Abschätzung über den Einfluß einer Comptonstreuung, die ein γ -Quant aus Segment A nach B streut zu bekommen, lassen wir in Gedanken die Länge l des Kristalls gegen Null gehen. Dadurch wird der mögliche Streuwinkel θ der γ -Quanten auf 90° beschränkt. Infolgedessen lassen sich leicht drei Fälle unterscheiden.

- $E < m_0c^2 \Rightarrow E' > \frac{E}{2}$
Im zeitlich ersten Segment wird weniger Energie als im zeitlich zweiten Segment deponiert.
- $E = m_0c^2 \Rightarrow E' = \frac{E}{2}$
In beiden Segmenten wird gleichviel Energie deponiert.
- $E > m_0c^2 \Rightarrow E' < \frac{E}{2}$
Im zeitlich ersten Segment wird mehr Energie als im zeitlich zweiten Segment deponiert.

Somit läßt sich in diesem Grenzfall verschwindenden Volumens das Segment in dem die erste Wechselwirkung stattfand eindeutig anhand der γ -Energie durch eine sehr einfache Fallunterscheidung bestimmen; bis zu einer γ -Energie von 511 keV sprach das Segment, in dem die kleinere der beiden, ab 511 keV dagegen das, in dem die größere der beiden Segmentenergien deponiert wurde, zuerst an. Bei genau 511 keV ist keine eindeutige Aussage möglich, die Wahrscheinlichkeit einer richtigen

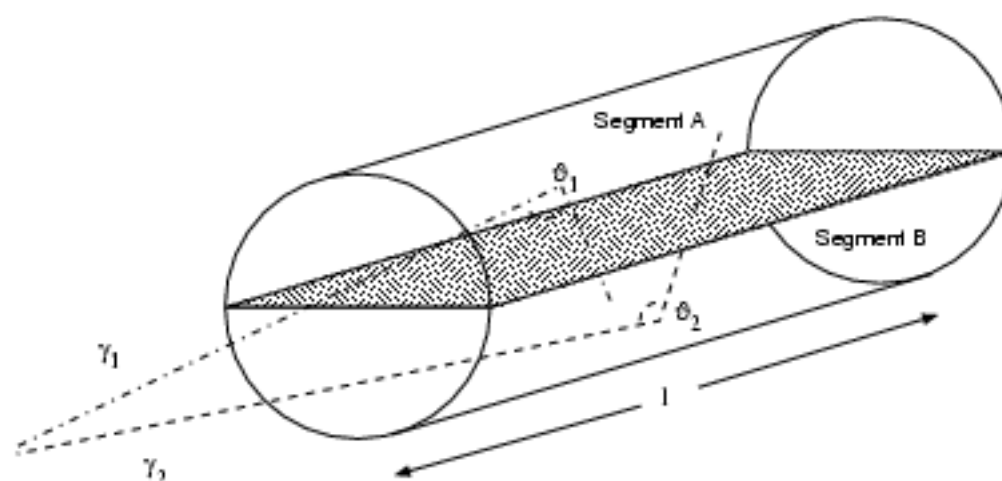


Abbildung 12: Schematische Darstellung eines zweifach segmentierten Germaniumkristalls.

Segmentvorhersage beträgt 50 Prozent. Lassen wir nun wieder schrittweise endliche Kristalllängen l und damit endliche Volumina zu, so geht diese Eindeutigkeit in zunehmendem Maße verloren. Es werden dabei für jede γ -Energie konkurrierende Prozesse denkbar, bei denen einmal ein Quant aus Segment A nach B und zum anderen umgekehrt aus Segment B nach A gestreut wurde, siehe Abbildung 12. Dabei wird sowohl in Segment A als auch in Segment B immer die gleiche Energie deponiert, obwohl einmal Segment A, einmal Segment B zuerst anspricht. So würden beispielsweise zwei γ -Quanten mit einer Energie von 1 MeV, von denen eines aus Segment A unter 45° nach Segment B, das andere aus Segment B unter 83° nach Segment A gestreut wird, beide in Segment A 365 keV, in Segment B 635 keV deponieren. Derartige Prozesse können zunächst durch Messung der Segmentenergien nicht mehr unterschieden werden; jedoch treten sie mit unterschiedlichen Häufigkeiten auf, so daß eine Unterscheidung im Rahmen einer Wahrscheinlichkeitsaussage möglich ist.

Darüberhinaus ist zu berücksichtigen, daß bei den in der Kernspektroskopie üblichen γ -Energien multiple Comptonstreuungen in Germanium vorherrschen. Abbildung 13 zeigt dazu die relativen Häufigkeiten, mit denen eine unterschiedliche Zahl von Comptoneffekten in Abhängigkeit von der γ -Energie auftreten, sofern beide Segmente ansprechen. Es ist zu erkennen, daß ab etwa 300 keV zwei-, drei- und vierfache Streuungen dominieren, fünffache Streuungen etwa gleichhäufig wie einfache und darüberhinaus auch sechsfache Prozesse auftreten. Eine Verkettung mehrerer Comptonstreuungen mit dem Ziel einer analytischen Vorhersage des zeitlich ersten Segments erscheint - insbesondere bei komplexeren als der momentan betrachteten Geometrie - nicht mehr praktikabel.

Daher wird in den folgenden Abschnitten der Transport von γ -Quanten durch unterschiedliche Detektorgeometrien mit GEANT simuliert. Dadurch sind die Verteilung der Energie über die Segmente und das zeitlich als erstes ansprechende Segment bekannt, so daß interessierende Häufigkeitsverteilungen aufgestellt und sich daraus ergebende Systematiken betrachtet werden können.

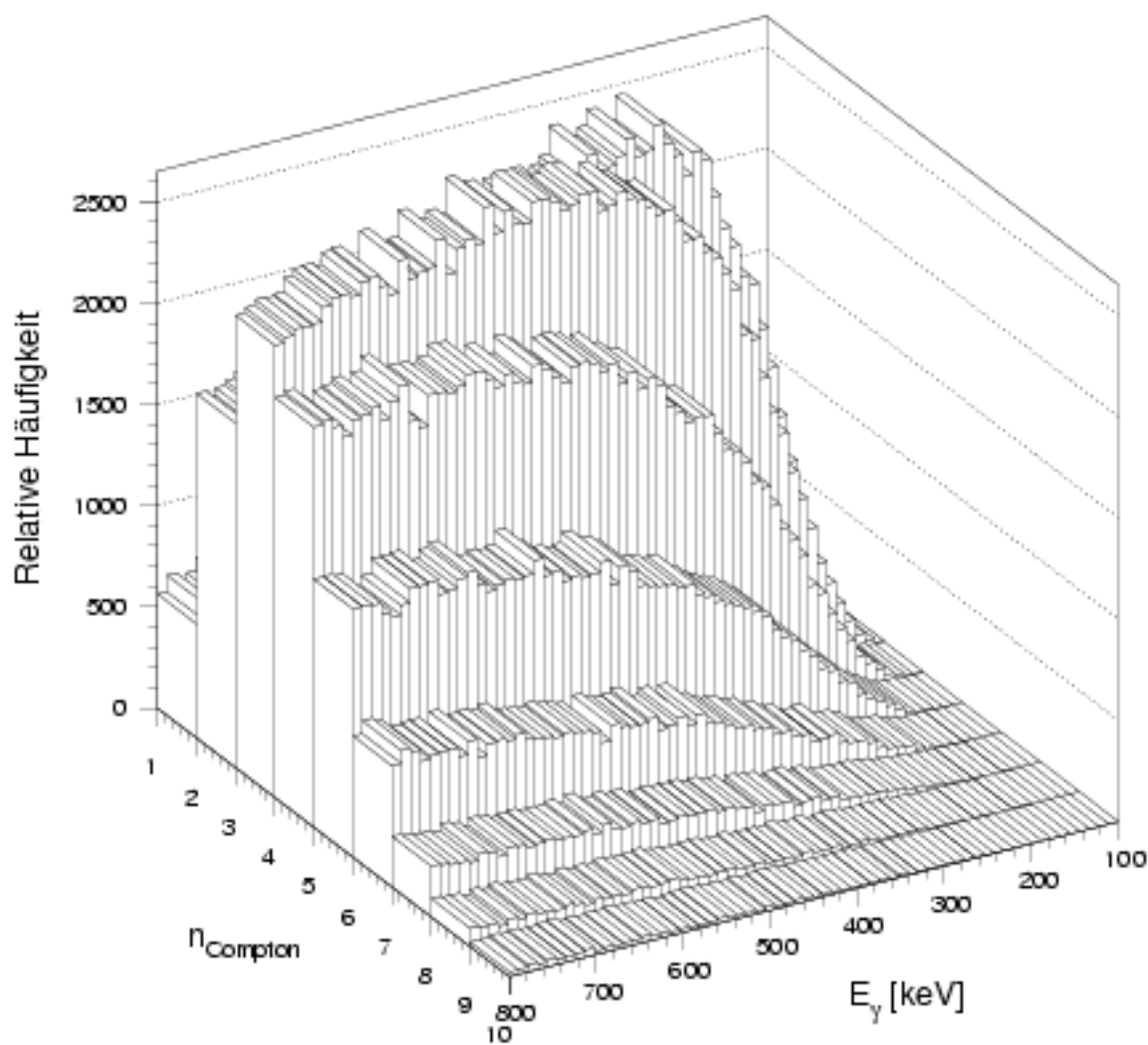


Abbildung 13: Verteilung der Zahl der Comptonstreuungen in einem zweifach segmentierten Germaniumzylinder von Länge und Durchmesser von 7 cm, sofern beide Segmente ansprechen.

4.1.2 Zweifach segmentierter Detektor

Mit Hilfe von GEANT wird in diesem Abschnitt der Transport von γ -Quanten durch einen zweifach segmentierten Germaniumkristall nach Abbildung 12 mit für heutige Germaniumdetektoren typischen Dimensionen von Länge und Durchmesser von jeweils sieben Zentimetern simuliert. Im Abstand von 25 cm von der Stirnfläche wurde dazu eine Quelle positioniert, welche γ -Quanten mit einer zwischen 100 und 800 keV gleichverteilten Energie isotrop emittierte. Abbildung 14 zeigt das energieabhängige Streuverhalten der Quanten. Man erkennt, daß ab etwa 300 keV in knapp zehn Prozent aller Fälle mit dem Ansprechen von zwei Segmenten zu rechnen ist, und daß die Wahrscheinlichkeit, daß ein Quant aus dem Detektor herausgestreut wird, mit zunehmender Energie größer wird.

Im folgenden ist lediglich der Fall einer Segmentmultiplizität von zwei von Interesse. Nach den Überlegungen des vorherigen Abschnitts sollte nun bei γ -Energien unterhalb einer gewissen Schwelle das Segment, in dem die geringere der beiden Segmentenergien deponiert wurde im Mittel häufiger als erstes ansprechen. Für γ -Energien oberhalb der Schwelle sollte das umgekehrte Verhalten auftreten. Dazu sind in Abbildung 15 die beiden interessierenden Wahrscheinlichkeiten in Abhängigkeit von der γ -Energie aufgetragen. Die beiden Kurven schneiden sich bei einer von der Geometrie abhängigen kritischen Energie, bei der die Wahrscheinlichkeit, daß das eine oder das andere der beiden Segmente zuerst anspricht, jeweils 50 Prozent beträgt. Somit werden die Überlegungen des vergangenen Abschnitts gestützt; gleichzeitig liegt eine einfache Abschätzung vor, mit der das zeitlich erste Segment mit einer gewissen Trefferquote vorhergesagt werden kann.

Um im betrachteten Fall einer Segmentmultiplizität von zwei ein Ereignis jedoch eindeutig beschreiben zu können, ist neben der Angabe der γ -Energie E_γ auch die Aufteilung dieser Energie auf die beiden Segmente des Detektors von Bedeutung. Da der Germaniumkristall symmetrisch geteilt wurde, können die beiden Segmente nun willkürlich mit A und B, die in diesen Segmenten deponierten Energien mit E_A und E_B bezeichnet werden. Jedes Ereignis kann somit in einer Ebene, in der die Abszisse die γ -Energie und die Ordinate das Verhältnis E_A/E_γ angibt, durch einen Punkt charakterisiert werden. Werte über 0.5 auf der Ordinate zeigen, daß mehr als die Hälfte der Energie in Segment A und somit entsprechend weniger Energie in Segment B deponiert wurde, Werte unter 0.5 bedeuten entsprechend, daß weniger als die Hälfte der γ -Energie in A und entsprechend mehr in B deponiert wurde. Abbildung 16 zeigt rechts, mit welchen absoluten Häufigkeiten die Ereignisse über diese Ebene verteilt sind. Links ist die Differenz der Häufigkeiten, mit denen Segment A bzw. Segment B als erstes anspricht zu sehen. Bereiche, in denen Segment A häufiger als Segment B als erstes anspricht, sind rot, die, in denen der umgekehrte Fall vorliegt, blau eingefärbt. Die Farbtiefe mißt die Häufigkeitsdifferenz; dunkle Farben kennzeichnen große Unterschiede in den Häufigkeiten, so daß das Segment der ersten Wechselwirkung mit hoher Wahrscheinlichkeit richtig vorhergesagt werden kann, weiße Bereiche dagegen kennzeichnen gleiche Häufigkeiten, in denen das erste Segment nur mit 50 prozentiger Wahrscheinlichkeit bestimmt werden kann.

Die Verteilung in Abbildung 16 bildet die physikalische Grenze der Trefferquote einer jeden Methode zur Vorhersage des ersten Segments. Es wird deutlich, daß die skizzierte Abschätzung, welche je nach γ -Energie das Segment der geringeren bzw. das der größeren Energie als erstes annimmt, nicht optimal gewählt ist. Sie teilt die Ebene mit Hilfe von zwei orthogonalen Geraden parallel zu Abszisse und Ordinate in vier Rechtecke und muß folglich in einem Energiebereich zwischen 300 bis 600 keV eine geringere als die theoretisch denkbare Vorhersagekraft aufweisen. Da die Verteilung jedoch von einfacher Struktur ist, kann sie beispielsweise durch Teilung der Ebene mittels dreier Strecken sehr gut approximiert werden, so daß an dieser Stelle ein MSP zur Vorhersage des ersten Segments nicht

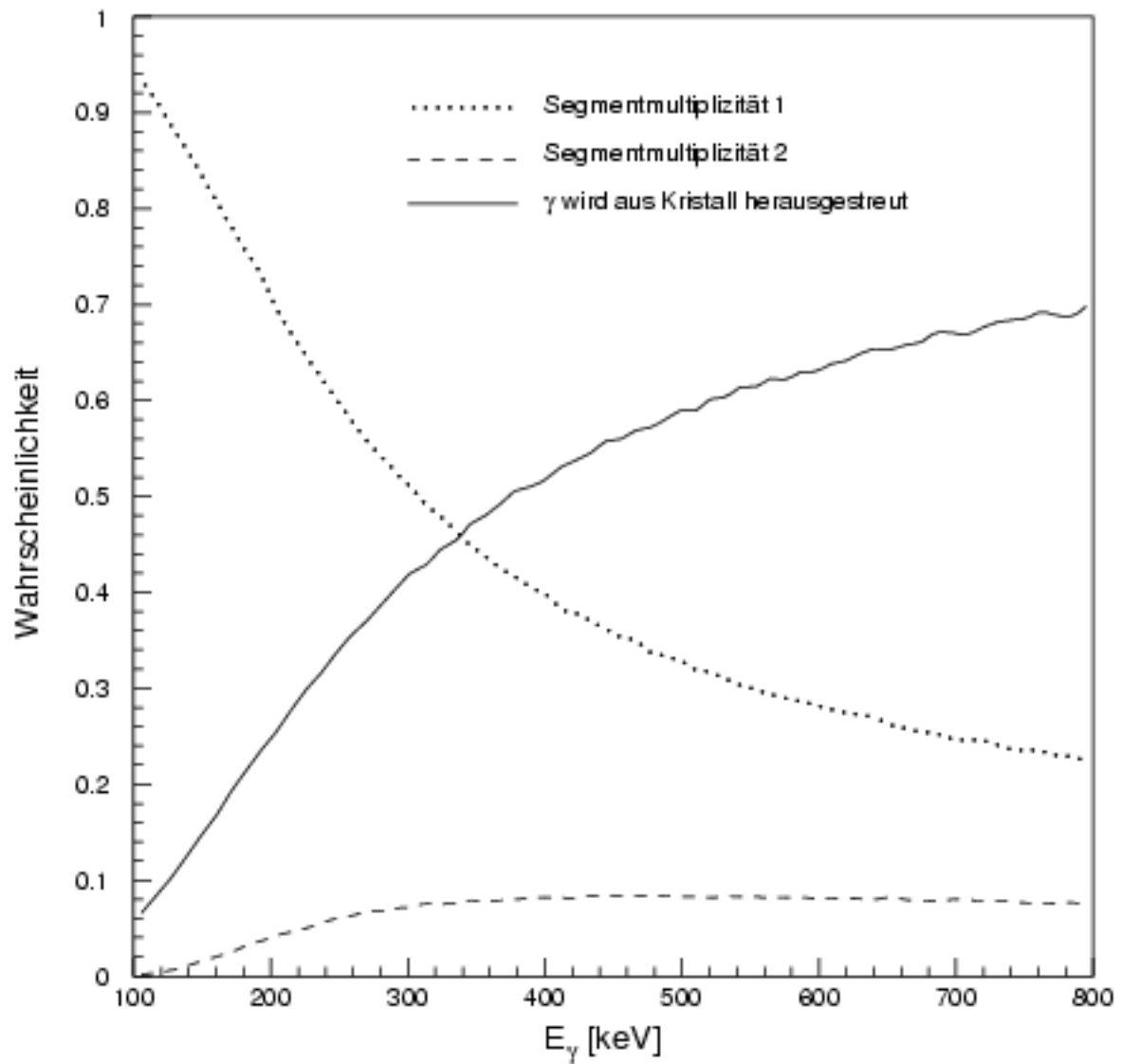


Abbildung 14: Verteilung der Segmentmultiplizitäten im zweifach segmentierten Detektor in Abhängigkeit von der γ -Energie.

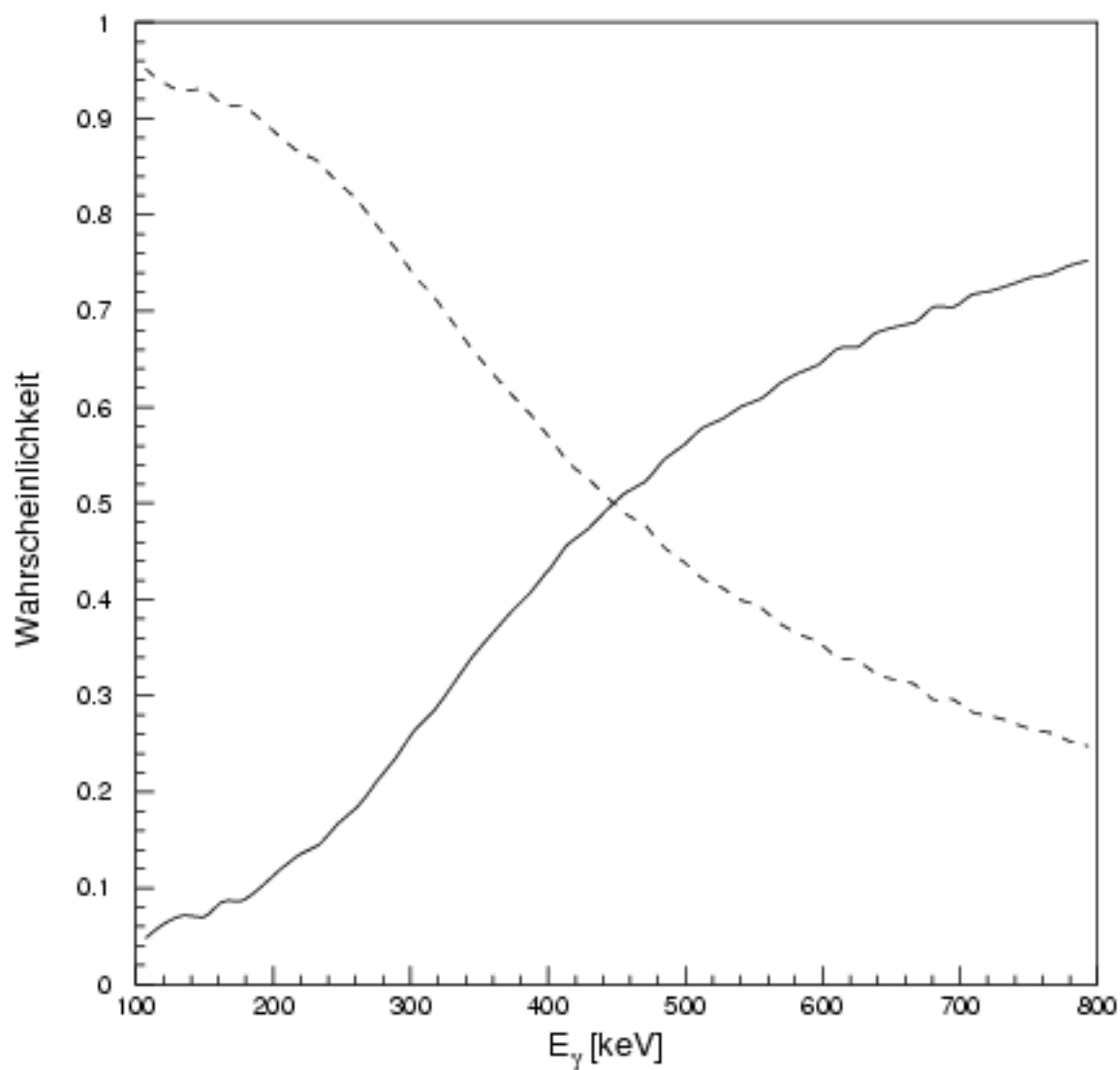


Abbildung 15: *Wahrscheinlichkeiten, daß das Segment mit der kleineren (gebrochen) bzw. der größeren (durchgezogen) der beiden Segmentenergien zuerst anspricht.*

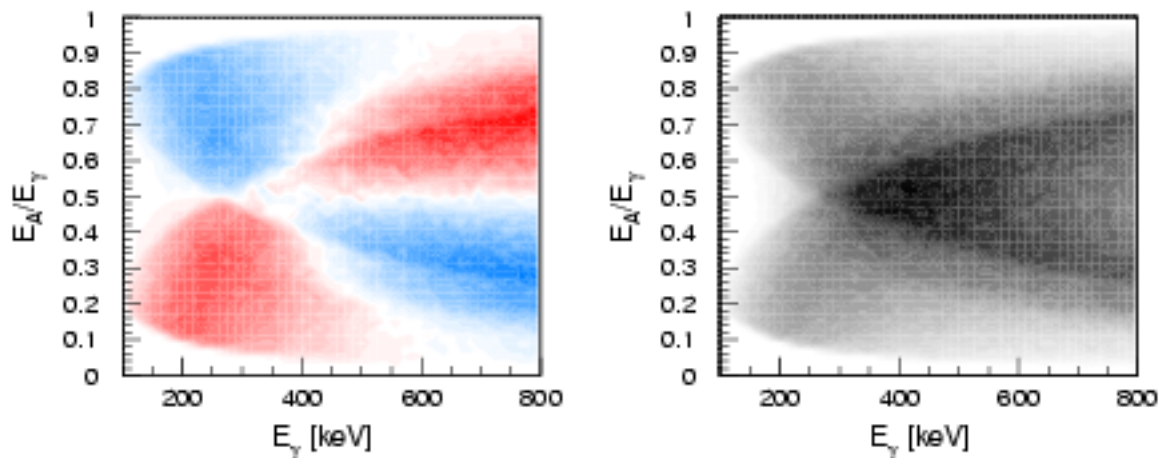


Abbildung 16: Bereiche, in denen Segment A (rot) bzw. B (blau) häufiger als erstes ansprach (links) und absolute Häufigkeiten der Ereignisse (rechts).

notwendigerweise zum Einsatz käme. Im Falle höherer Segmentmultiplizitäten sind jedoch komplexere Verteilungen zu erwarten, die nicht mehr durch derart einfache Verfahren angenähert werden können. Um beurteilen zu können, ob MSPs überhaupt in der Lage sind, derartige überlappende Verteilungen erfolgreich zu trennen, wird ihre Fähigkeit zur Klassifikation der Energiemuster zunächst am vorliegenden Fall untersucht. Dazu werden nun MSPs unterschiedlicher Struktur trainiert und verglichen.

Die in den beiden Segmenten deponierten Energien entsprechen Meßgrößen eines Experiments und können daher als Eingabegrößen für ein mehrschichtiges Perzeptron verwendet werden. Gewünscht wird eine Einteilung dieser zweidimensionalen Eingabemuster in zwei Klassen, nämlich einer Klasse, in der Segment A, und einer anderen Klasse, in der Segment B als erstes ansprach. Folglich ist es naheliegend, MSPs mit einer Eingabeschicht von zwei Knoten für die beiden Segmentenergien und einer Ausgabeschicht von zwei Knoten für die beiden Klassen zu verwenden. Durch GEANT wurden drei voneinander unabhängige, in ihrer Struktur jedoch identische Datensätze generiert. Jedes Element dieser Datensätze besteht aus zwei Eingabegrößen - den in den beiden Segmenten deponierten Energien - und zwei Ausgabegrößen. Die Segmentenergien wurden direkt in MeV eingegeben, eine Vorverarbeitung fand nicht statt. Als Sollausgabegrößen zur Kennzeichnung der beiden Klassen wurden die geordneten Paare (0,1) und (1,0) verwendet, als Ausgabefunktion aller Knoten im Netz die Fermifunktion. Somit kann später bequem überprüft werden, inwiefern die Netzausgabe einer Bayesschen Wahrscheinlichkeit entspricht. Sowohl der Lern- als auch der Kontrolldatensatz umfaßten jeweils 10000 Muster. Zur abschließenden Beurteilung des Netzes wurde zusätzlich ein dritter Datensatz mit einer Million Mustern generiert.

Da weder Zahl noch Größe der versteckten Schichten von vornherein ersichtlich sind, wurden verschiedene Strukturen mit einer und zwei versteckten Schichten untersucht. Im Falle einer versteckten Schicht wurden Breiten von zwei, vier, fünf, sieben, zehn und zwanzig Knoten, im Falle zweier versteckter Schichten Breiten von jeweils zwei und vier Knoten untersucht. Die Wahl dieser Werte

wurde willkürlich getroffen. Da sich die Zahl der in diesen Strukturen enthaltenen freien Parameter als ausreichend erwiesen hat, wurden keine größeren Netze betrachtet.

Für jedes Netz wurden jeweils fünf Startsynapsensätze mit zwischen -1 und +1 zufällig gewählten Synapsen bestimmt und ein Gradientenabstieg mit dem RPROP-Algorithmus durchgeführt. Das Training wurde abgebrochen, sobald die Differenz der Fehler auf dem Trainingsdatensatz zwischen 100 Epochen weniger als 10^{-7} betrug. Die Abnahme des Fehlers nach Gleichung (4) im Laufe der Lernphase ist für alle Netze und jeweils drei Startsynapsensätze in Abbildung 17 sowohl für den Lern- als auch für den Kontrolldatensatz zu sehen. Die Fehler des Netzes 2-2-2 laufen bereits nach 400 Epochen sowohl auf dem Lern- als auch auf dem Kontrolldatensatz unabhängig von den Startsynapsen in eine Sättigung bei 0,112. Das gleiche Verhalten zeigen zwei der fünf Fehlerminimierungen des 2-4-2 Netzes, das 2-2-2-2 Netz zeigt diesen Effekt bei einem Fehler von 0,107. Wie sich im Anschluß herausstellen wird, ist bei diesen Werten eine adäquate Klassifizierung der Eingabemuster nicht möglich. Das Langzeitverhalten der anderen drei Durchgänge des 2-4-2 Netzes ist mit dem aller anderen größeren Strukturen identisch, dabei fällt der Fehler auf dem Lerndatensatz auf Werte zwischen 0,101 und 0,102, der auf dem Kontrolldatensatz auf Werte zwischen 0,102 und 0,103. Innerhalb dieser Intervalle ist keine ausgeprägte Systematik in Abhängigkeit von der Netzgröße zu erkennen.

Somit erscheinen die 2-2-2 und 2-2-2-2 Netze als zu klein, das 2-4-2 Netz bezüglich der Startsynapsenwahl als kritisch. Alle anderen Netze sind alleine aufgrund des erreichbaren Fehlers nicht sinnvoll zu unterscheiden, allerdings treten geringfügige Unterschiede in der Geschwindigkeit der Konvergenz auf. So erreichen das 2-4-2 und das 2-7-2 Netz die Sättigung erst nach etwa 1100 Epochen, die Mehrheit der anderen Netze dagegen bereits nach 500 Epochen, was aufgrund der unerheblichen Rechenzeiten¹⁹ jedoch nicht ins Gewicht fällt. Da im Rahmen einer Anwendung nicht unnötig große Netze verwendet werden sollten, erscheinen die Netze, die einen möglichst kleinen Fehler auf dem Kontrolldatensatz erreichen und gleichzeitig über möglichst wenige Parameter verfügen, das sind im vorliegenden Fall die 2-5-2, 2-7-2 und 2-10-2 Netze, als korrekt dimensioniert.

Um dies zu überprüfen und die Leistungsfähigkeit der Netze zu beurteilen, werden nun exemplarisch für das 2-2-2, das 2-7-2 und das 2-20-2 Netz die energieabhängigen Wahrscheinlichkeiten der richtigen Klassifikation der Energiemuster in der unteren Zeile in Abbildung 18 dargestellt. Dabei sind die Trefferquoten der Netze in Grün und zum Vergleich die der skizzierten Abschätzung in Rot sowie die durch die Verteilung in Abbildung 16 bestimmte maximale mögliche Quote in Blau eingetragen. Man erkennt, daß das 2-7-2 und das 2-10-2 Netz im Rahmen der statistischen Fehler die maximal mögliche Vorhersagekraft aufweisen.

Da sich die Netzausgaben an jedem Punkt der eingeführten Ebene zu eins addieren, diese also komplementär zueinander sind, kann darüberhinaus das Ausgabesignal des Segment A zugeordneten Knotens dargestellt und mit der Verteilung nach Abbildung 16 verglichen werden. Dieses Signal liegt im Intervall $[0, +1]$. Werte über 0.5 bedeuten, daß Segment A, Werte unter 0.5, - da in diesem Fall das Ausgabesignal des Segment B zugeordneten Ausgabeknotens entsprechend über 0.5 liegt - daß Segment B als erstes angesprochen haben soll. In der oberen Zeile von Abbildung 18 sind die entsprechenden Bereiche für die drei diskutierten Netze wieder in rot bzw. blau eingefärbt. Die Farbtiefe gibt direkt die Größe des Ausgabesignals an, von Null über $\frac{1}{2}$ zu Eins ändert sich die Farbe von Blau über Weiß zu Rot. Man erkennt, daß das 2-2-2 Netz die Verteilung aus Abbildung 16 nicht in erforderlichem Maße approximiert. Das 2-7-2 Netz dagegen verfügt über eine ausreichende Zahl freier Parameter, um sogar feine Strukturen der Verteilung nachzubilden. Im Falle des 2-20-2

¹⁹. vergleiche mit Anhang A -

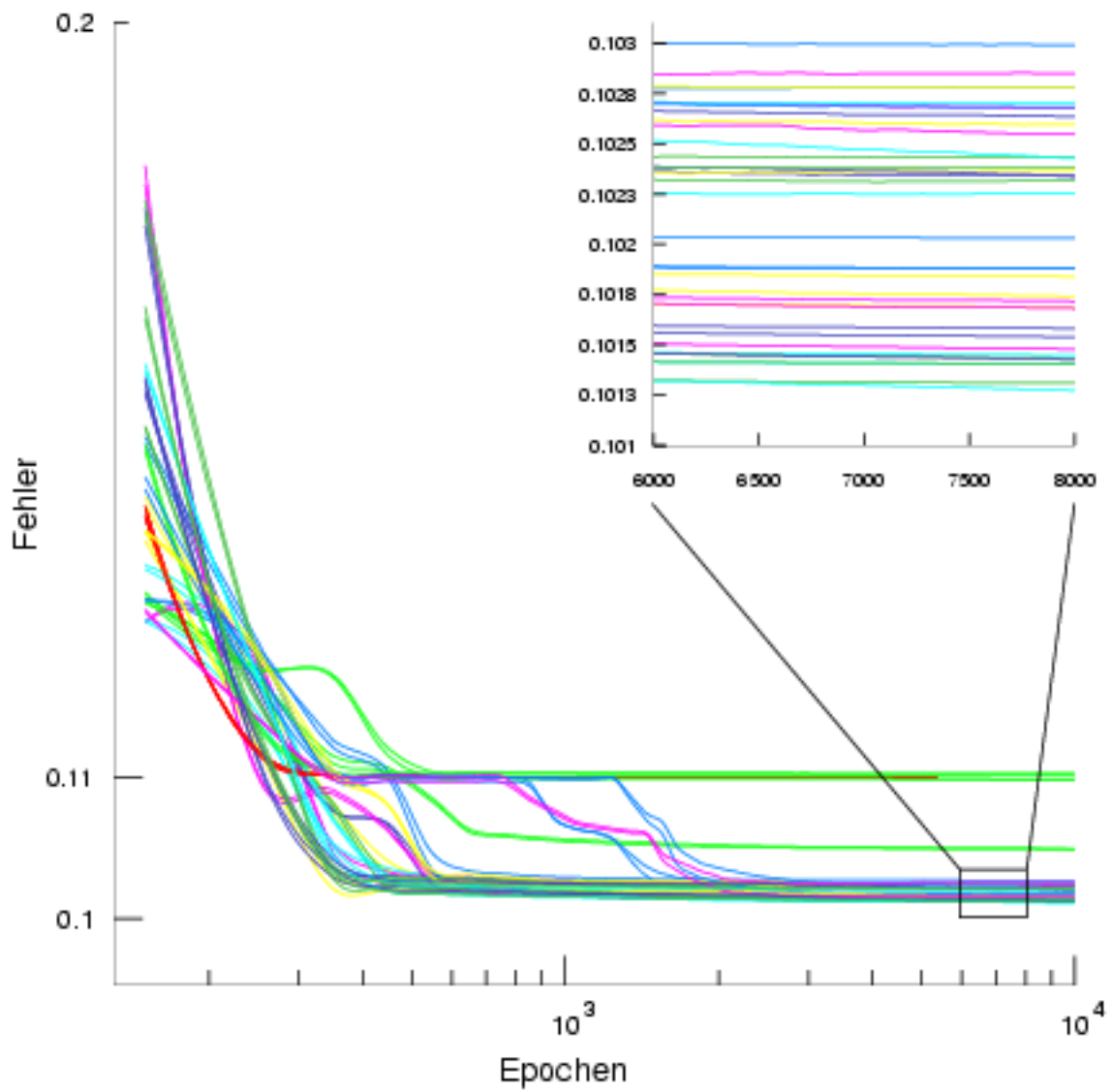


Abbildung 17: Abnahme des Fehlers auf Lern- und Kontrolldatensatz für jeweils drei Gradientenabstiege von acht verschiedenen MSPs.

Netzes zeigt sich die Wirkung der unnötig vielen freien Parameter, die vom 2-7-2 Netz adäquat reproduzierten feinen Strukturen werden übertrieben wiedergegeben, es erfolgt eine Überbewertung an und für sich nebensächlicher Elemente der zu lernenden Verteilung. Bei genauer Betrachtung der Randbereiche der Verteilungen fällt auf, daß die Netzausgabe der Häufigkeitsverteilung der zu trennenden Ereignisse zwar ähnelt, jedoch nicht exakt entspricht. Die Netze lernen also nicht, die Häufigkeitsverteilungen zu reproduzieren. Daß sich die Knotenausgaben darüberhinaus mit hoher Genauigkeit zu eins addieren und nicht - wie im Rahmen des Trainings verlangt - die diskrete Werte null und eins annehmen, deutet bereits auf eine mögliche Interpretation der Netzausgaben als Bayessche Wahrscheinlichkeiten hin. Dazu wird die Netzausgabe nun mit den aus den Häufigkeitsverteilungen bestimmten Bayesschen Wahrscheinlichkeiten verglichen. In Abbildung 19 sind in der ersten Zeile die Häufigkeitsverteilung der zu klassifizierenden Muster, in der zweiten Zeile die daraus berechneten Bayesschen Wahrscheinlichkeiten und in der dritten Zeile die Ausgabesignale der beiden Ausgabeknoten des 2-7-2 Netzes gegenübergestellt. Die linke Spalte kennzeichnet dabei die Klasse, in der Segment A, die rechte Spalte die Klasse, in der Segment B als erstes anspricht. Man erkennt deutlich, daß sich die Verteilungen der zu klassifizierenden Muster überlappen und daß die Netzausgaben die abgebildeten Bayesschen Wahrscheinlichkeiten approximieren. Die Summe der Ausgabesignale der beiden Knoten der Ausgabeschicht addieren sich an jedem Punkt der Ebene mit einer absoluten Genauigkeit von 4×10^{-4} zu Eins. Der nach Gleichung (3) bestimmte Abstand beträgt -3.2×10^{-5} . Auch diese Werte stützen eine Interpretation der Netzausgaben als Bayessche Wahrscheinlichkeiten.

Die Netze mehrerer Trainingsdurchgänge von jeweils 10^4 Epochen wurden schließlich wiederum mit Hilfe des Optimal Brain Surgeon-Algorithmus beschnitten. Aus den 22 Parameter umfassenden 2-4-2 Netzen konnten dabei 3, 4, 4, 5 und 6, aus den 27 Parameter umfassenden 2-5-2 Netzen 4, 4, 5 und 6 Parameter entfernt werden. Aus den 2-7-2 Netzen mit 37 Parametern konnten 5, 9, 10, 13 und 18 Synapsen, aus den 2-4-4-2 Netzen mit 42 Parametern 8, 10, 10, 14 und 25 Parameter eliminiert werden. Aus den 52 Synapsen umfassenden 2-10-2 Netzen entfernte OBS 9, 15, 16 und 17, aus den 102 Parameter umfassenden 2-20-2 Netzen 57 bzw. 60 Parameter. Es zeigt sich somit ein zu Kapitel 3.3 analoges Verhalten, daß in ursprünglich größeren Netzen generell mehr Parameter verbleiben, allerdings bleibt auch an dieser Stelle anzumerken, daß es nicht angestrebt wurde, iterativ möglichst minimale Strukturen zu erreichen, so daß diese Daten nur einen groben Anhaltspunkt der Zahl der überflüssigen Parameter darstellen.

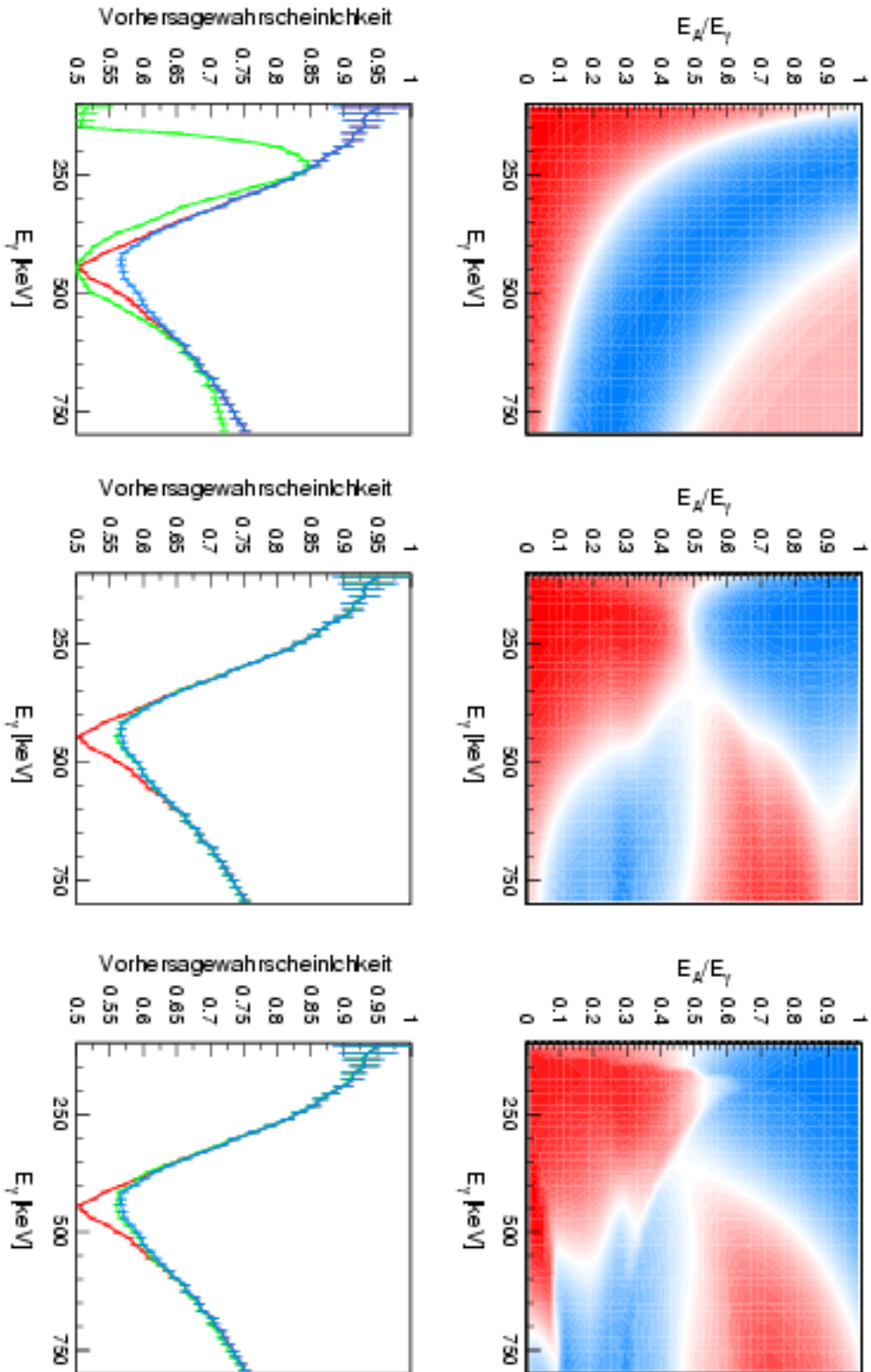


Abbildung 18: Netzausgaben der 2-2-2, 2-7-2 und 2-20-2 Netze (oben) und die daraus folgenden Vorhersagewahrscheinlichkeiten (unten) der Netze (grün); zum Vergleich sind die Vorhersagewahrscheinlichkeiten der Abschätzung (rot) und das theoretische Limit (blau) angegeben.

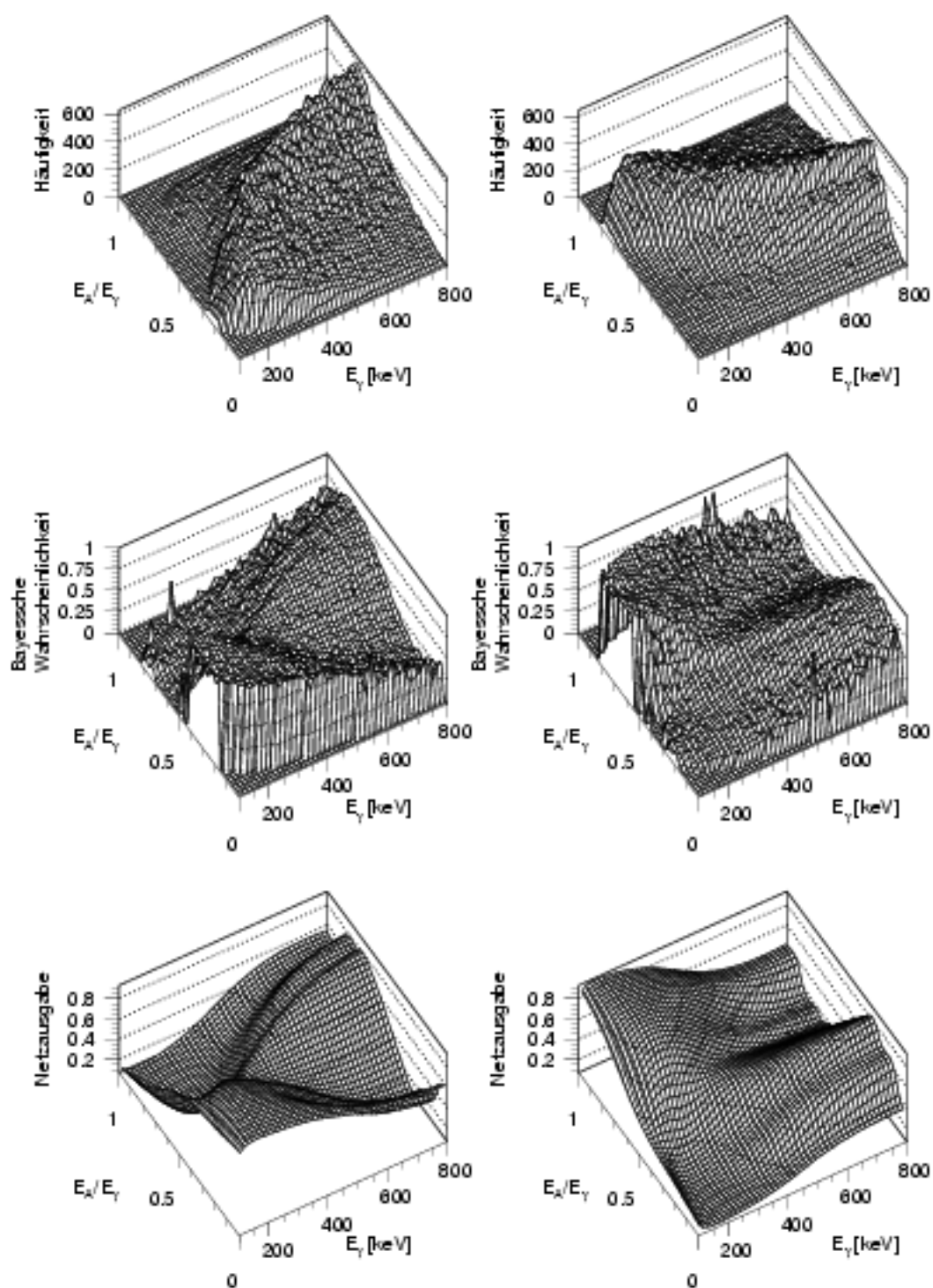


Abbildung 19: Häufigkeitsverteilungen der Ereignisse, in denen A bzw. B als erstes ansprach (oben), daraus berechnete Bayesche Wahrscheinlichkeiten (mitte) und Netzausgabe des linken bzw. rechten Ausgabeknotens des 2-7-2 Netzes (unten).

4.1.3 Vierfach segmentierter Detektor

In diesem Kapitel wird die Segmentierung des Germaniumkristalls weiter verfeinert. Zusätzlich zur in Abbildung 12 eingezeichneten Ebene wird senkrecht auf diese eine weitere Ebene eingeführt, wodurch der Kristall nunmehr symmetrisch in vier Segmente unterteilt wird. Somit können neben zwei auch drei oder gar alle vier Segmente ansprechen. Im Falle von zwei ansprechenden Segmenten kann darüberhinaus zwischen benachbarten und diagonal liegenden Segmenten weiter unterschieden werden. Abbildung 20 zeigt die Verteilung der Segmentmultiplizitäten für einen vierfachsegmentierten Kristall mit Länge und Durchmesser von sieben Zentimetern in Abhängigkeit von der γ -Energie anhand einer Simulation durch GEANT. Die Quelle wurde dazu wiederum 25 cm von der Stirnfläche entfernt angebracht, sie emittierte erneut isotrop γ -Quanten zwischen 100 und 800 keV gleichverteilter Energien. Es ist zu erkennen, daß wiederum ab etwa 300 keV zehn Prozent aller Ereignisse zwei Segmente ansprechen lassen. Dabei kommt der Fall, daß diese Segmente benachbart liegen etwa eine Größenordnung häufiger vor als der, in dem diese Segmente diagonal gegenüber liegen. Ab etwa 400 keV beträgt der Anteil der Ereignisse mit Segmentmultiplizität drei etwa ein Prozent. Ereignisse mit Segmentmultiplizität vier kommen nur mit Anteilen in der Größenordnung weniger Promille vor und werden daher nicht berücksichtigt.

Im Falle einer Segmentmultiplizität von zwei ist nicht zu erwarten, daß ein im Vergleich zum zweifach segmentierten Kristall qualitativ neues Verhalten auftritt, wenngleich die exakte Struktur der Häufigkeitsverteilungen aus Abbildung 16 für die beiden Fälle benachbarter und diagonalen Segmente nicht identisch und jeweils gegenüber Abbildung 16 verschoben oder gestaucht sein wird. Abbildung 21 zeigt dazu in der rechten Spalte die absoluten Häufigkeiten aller betrachteten Ereignisse, in der linken Spalte wiederum die Differenz der Häufigkeiten, mit denen das eine bzw. das andere der beiden Segmente als erstes anspricht. Die obere Zeile zeigt diese Verteilungen für den Fall zweier benachbarter Segmente, die untere für den Fall zweier diagonalen Segmente. Die Abszisse kennzeichnet analog zu Kapitel 4.1.2 die γ -Energie E_γ , die Ordinate die Verteilung dieser Energie über die beiden angesprochenen Segmente A und B durch den Quotient E_A/E_γ . Wie zu erwarten sind die Verteilungen sowohl untereinander ähnlich und zeigen die gleiche Struktur wie im Fall des zweifach segmentierten Detektors.

Bei einer Segmentmultiplizität von drei ist eine Systematik der Häufigkeiten, mit denen einzelne dieser Segmente zuerst ansprechen, aus den bisherigen Überlegungen und Abschätzungen nicht ersichtlich. Wiederum treten im Mittel drei bis vier Comptonstreuungen auf, so daß eine analytische Bestimmung der Häufigkeitsverteilungen nicht mehr praktikabel erscheint. Im Fall des zweifach segmentierten Detektors im vorangegangenen Abschnitt wurden im Rahmen einer ersten Abschätzung nach Abbildung 15 die Wahrscheinlichkeiten betrachtet, daß das Segment der geringeren oder der höheren Energie als erstes anspricht. Analog dazu zeigt Abbildung 22 die Wahrscheinlichkeiten, daß das Segment mit der geringsten, das der mittleren und das der höchsten Segmentenergie als erstes anspricht. Es ist ein Verhalten wie in Abbildung 15 zu erkennen; ab einer kritischen γ -Energie von etwa 420 keV spricht das Segment der größten Energie im Mittel häufiger als erstes an, unterhalb dieser Schwelle dagegen die übrigen beiden Segmente mit etwa gleichen Wahrscheinlichkeiten. Dadurch liegt auch für eine Segmentmultiplizität von drei eine einfache Abschätzung vor, mit deren Hilfe das zeitlich erste Segment grob bestimmt werden kann.

Prinzipiell wäre nun denkbar, für die beiden Fälle von zwei ansprechenden Segmenten und die Fälle unterschiedlicher Segmentmultiplizitäten einzelne Netze zu trainieren. Nach den Ergebnissen

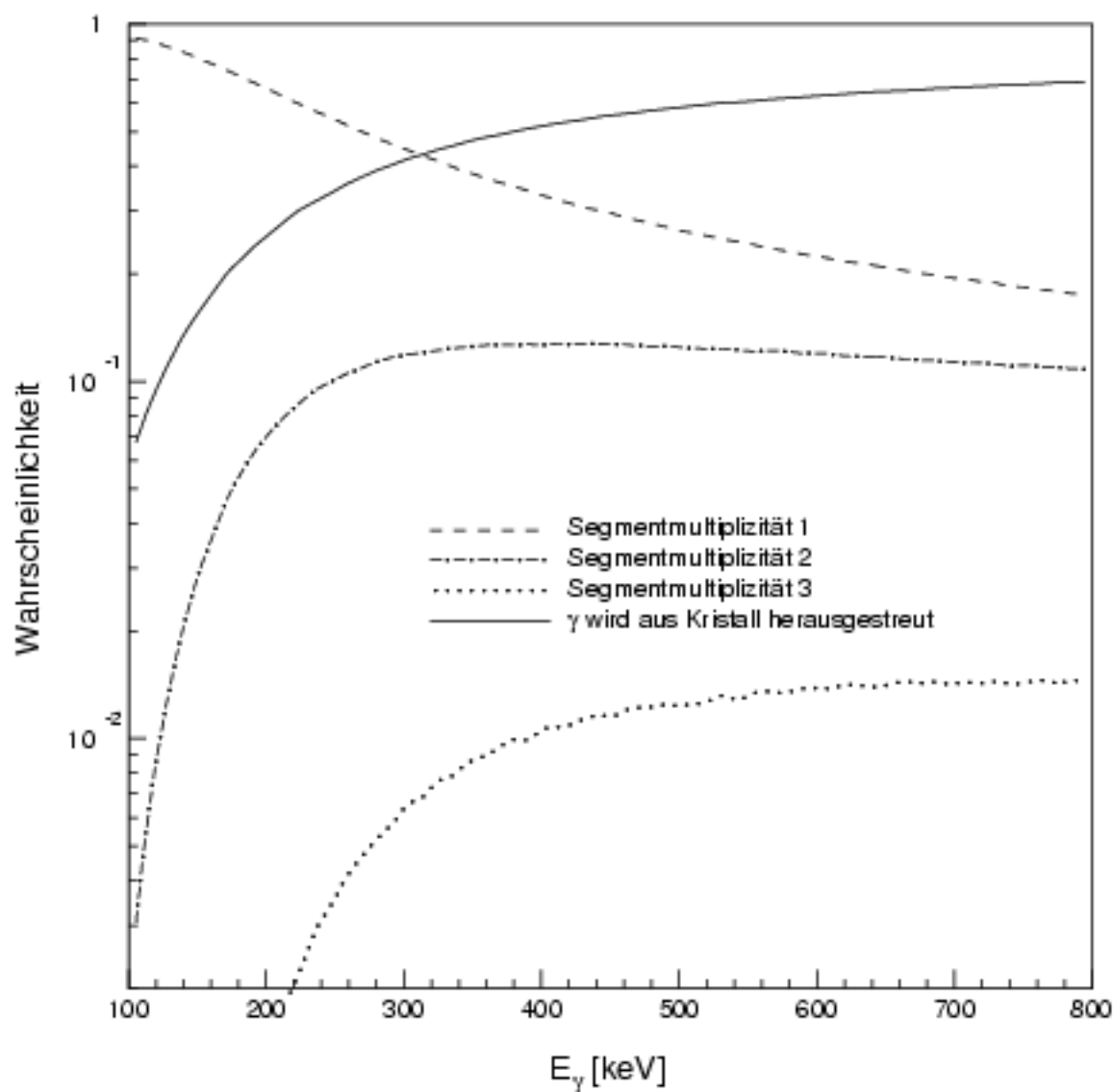


Abbildung 20: Verteilung der Segmentmultiplizitäten im vierfach segmentierten Detektor in Abhängigkeit von der γ -Energie.

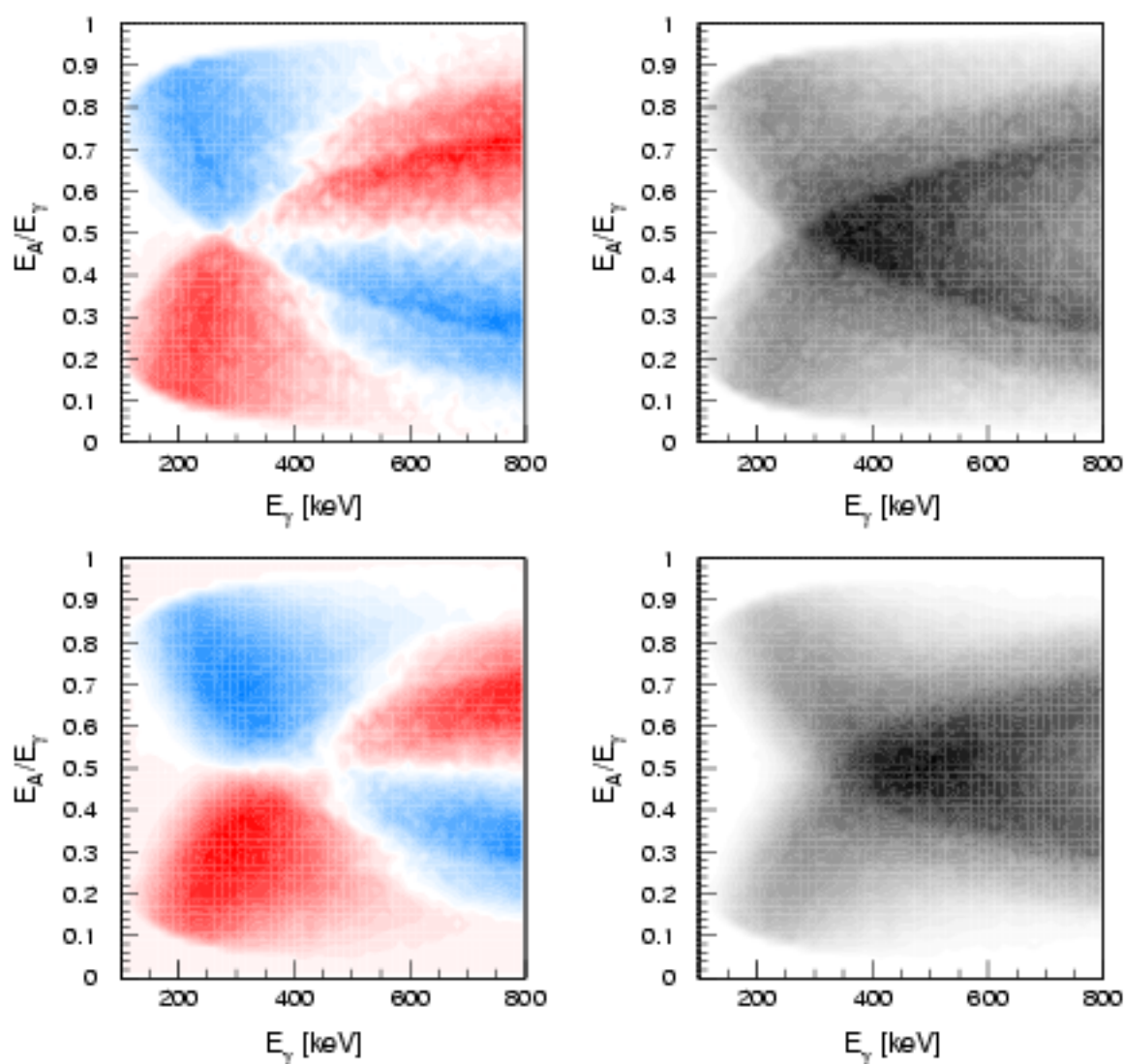


Abbildung 21: Verteilung der Bereiche, in denen Segment A (rot) bzw. B (blau) im Mittel häufiger als erstes ansprach (links) und absolute Häufigkeiten der Ereignisse (rechts) für benachbarte (oben) und diagonale (unten) Segmente.

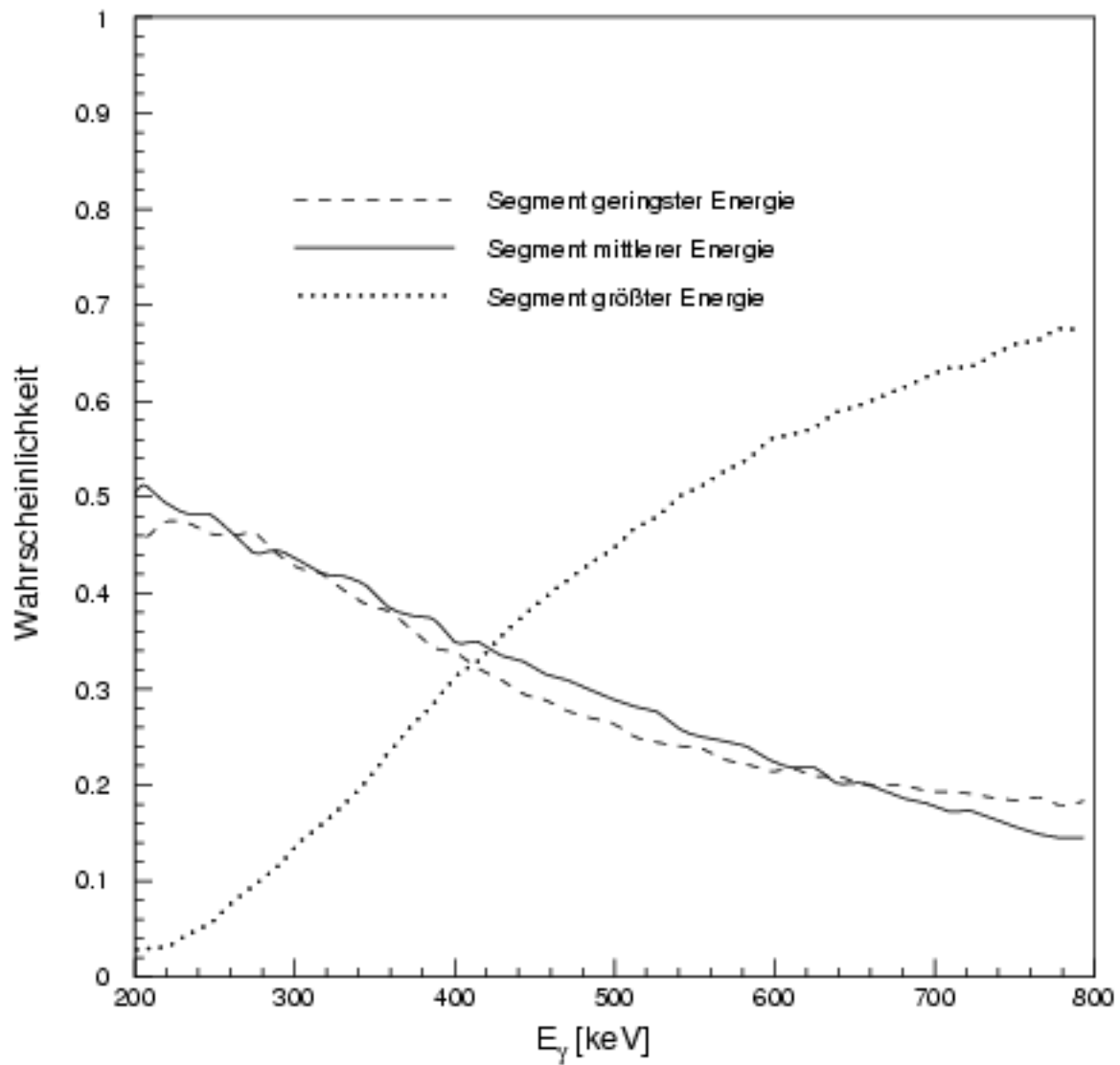


Abbildung 22: Wahrscheinlichkeiten, daß das Segment der geringsten, mittleren oder der größten Energie als erstes anspricht.

des vorhergehenden Kapitels ist zu erwarten²⁰, daß diese Fälle jeweils isoliert betrachtet keinerlei Schwierigkeiten darstellen werden. Stattdessen wird nun der Versuch unternommen, ein einziges Netz zu verwenden um alle diese Fälle gleichzeitig bearbeiten zu können.

Als Eingabegrößen der MSPs wurden erneut die Segmentenergien in MeV verwendet. Da vier Segmente vorliegen, besteht die Eingabeschicht aller untersuchten Netze naheliegenderweise aus vier Knoten. Zur Repräsentation der vier möglichen Klassen wurden die Vierertupel $(1,0,0,0)$, $(0,1,0,0)$, $(0,0,1,0)$ und $(0,0,0,1)$ verwendet. Als Ausgabefunktion aller Knoten wurde wiederum die Fermifunktion eingesetzt. Als Trainings- und Kontrolldatensätze wurden durch GEANT zwei Datensätze zu je 20000 Mustern generiert. Werden dabei die auftretenden Segmentmultiplizitäten entsprechend der Simulation durch GEANT gewichtet, so ist zu beobachten, daß die später diskutierte Leistung der Netze auf dem am häufigsten vorkommenden Fall zweier benachbarter Segmente zwar tadellos, in den anderen wesentlich seltener auftretenden Fällen jedoch unzureichend ist²¹. Dieser Effekt ist verständlich, da im Rahmen der Fehlerminimierung die häufiger vorkommenden Muster stärker ins Gewicht fallen. Daher wurde der Lerndatensatz aus 5000 Ereignissen einer Segmentmultiplizität von drei, 5000 Ereignissen von zwei diagonal liegenden und 10000 Ereignissen benachbart liegender Segmente zusammengesetzt. Zur abschließenden Kontrolle wurden Datensätze mit 500000 Ereignissen zweier benachbarter Segmente, 160000 Ereignissen zweier diagonaler Segmente und 500000 Ereignissen dreier angesprochener Segmente verwendet.

Es wurden vier verschiedene Netze untersucht, nämlich 4-4-4, 4-4-4-4, 4-10-4 und 4-20-4 Strukturen. Zum Training wurde wieder der RPROP-Algorithmus verwendet, das Training wurde abgebrochen, sobald die Differenz des Fehlers auf dem Lerndatensatz zwischen 100 Epochen geringer als 10^{-7} betrug. Für jedes Netz wurden fünf Startpunkte im Fehlergebirge zufällig gewählt und ein Gradientenabstieg durchgeführt. Abbildung 23 zeigt die Abnahme des Fehlers auf dem Lerndatensatz für die unterschiedlichen Netze in Farbe, die Abnahme des Fehlers auf dem Kontrolldatensatz zum Vergleich in schwarz. Man erkennt, daß die kleineren 4-4-4 und 4-4-4-4 Netze nach bereits 500 Epochen in eine Sättigung von 0,0608 bzw. 0,063 laufen, die beiden größeren 4-10-4 und 4-20-4 Strukturen erst ab etwa 2000 Epochen einen Fehler von etwa 0,057 bzw. 0,053 erreichen. Generell ist der Trend zu beobachten, daß der minimale Fehler mit zunehmender Netzgröße geringer wird. Die Verringerung des Fehlers auf dem Lerndatensatz geht wie im Falle des zweifach segmentierten Kristalls mit einer Abnahme des Fehlers auf dem Kontrolldatensatz einher. Ein Abfall des Fehlers auf dem Lerndatensatz mit gleichbleibendem oder gar ansteigendem Fehler auf dem Kontrolldatensatz wie in Kapitel 3.3 bzw. 4.2 ist nicht festzustellen. Im folgenden zeigt sich, daß das 4-20-4 Netz eine ausreichende Größe hat, um das gestellte Klassifikationsproblem zu lösen. Es kann zu diesem Zeitpunkt nicht ausgeschlossen werden, daß auch aus dieser Struktur - wie später demonstriert - noch überflüssige Parameter entfernt werden können oder daß auch kleinere Netze im Rahmen einer längeren Lernphase eine vergleichbare Leistung erreichen könnten.

Abbildung 24 zeigt die Netzausgabe des 4-20-4 Netzes für die beiden Fälle benachbarter und diagonalen Segmente. Dazu wurde zunächst überprüft, daß sich die Netzausgaben wiederum mit guter Genauigkeit zu eins addieren. In einem weiteren Schritt wurde sichergestellt, daß die Ausgabesignale der beiden Knoten, die nicht zu den beiden in Frage kommenden Klassen gehören, nur zu vernachlässigende Werte annehmen. Diese beiden Kriterien sind bis auf Randbereiche, in denen keine Ereignisse in den verwendeten Datensätzen enthalten sind, gut erfüllt. Daher können wiederum die beiden interessierenden Knotenausgaben zusammengefaßt und dargestellt werden. Bis auf die

²⁰Mit Ausnahme der Segmentmultiplizität von drei, die dort nicht untersucht werden konnte.

²¹Vergleiche mit dem folgenden Abschnitt.

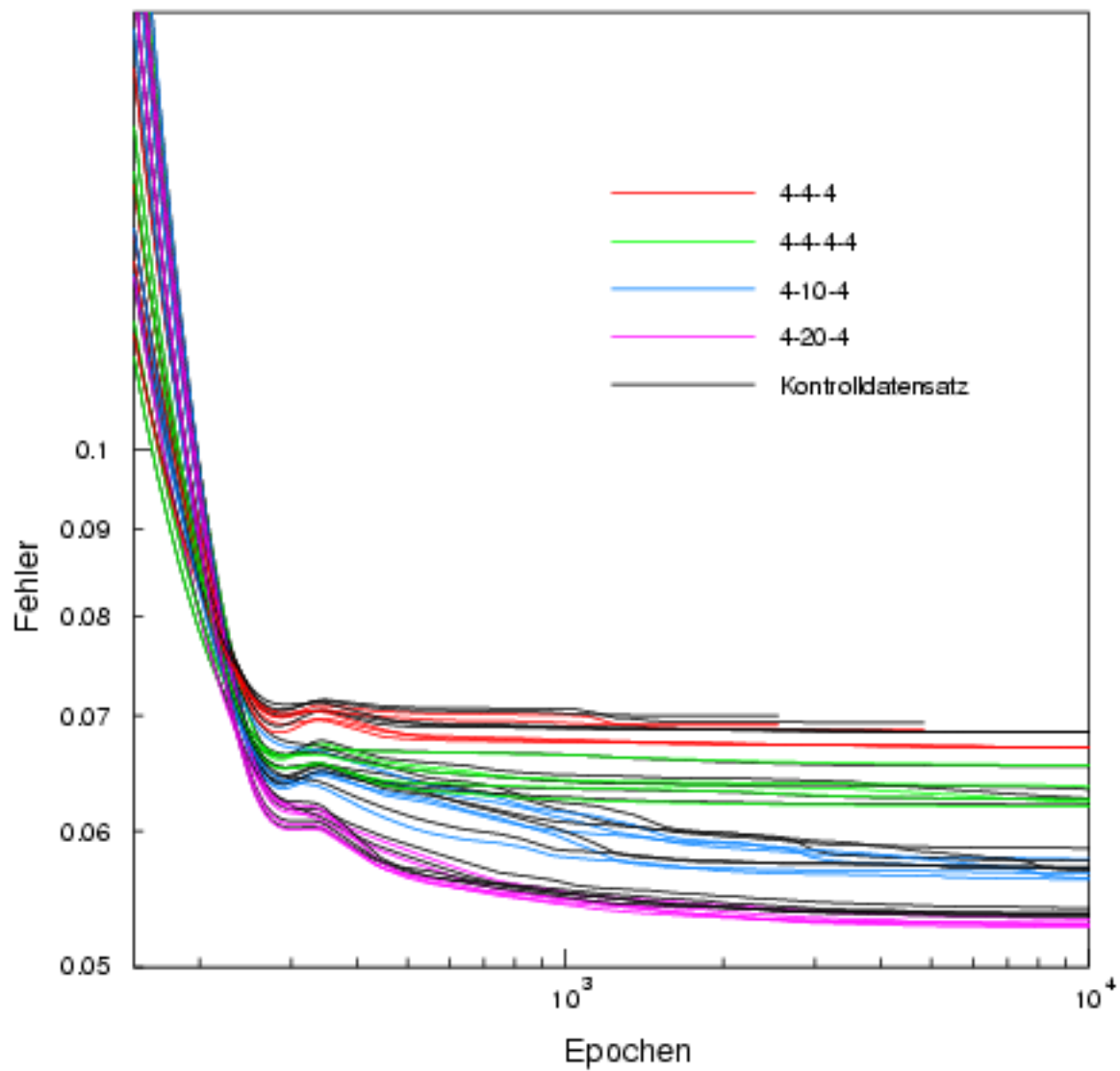


Abbildung 23: Abnahme des Fehlers auf Lern- (farbig) und Kontrolldatensatz (schwarz) für verschiedene MSPs

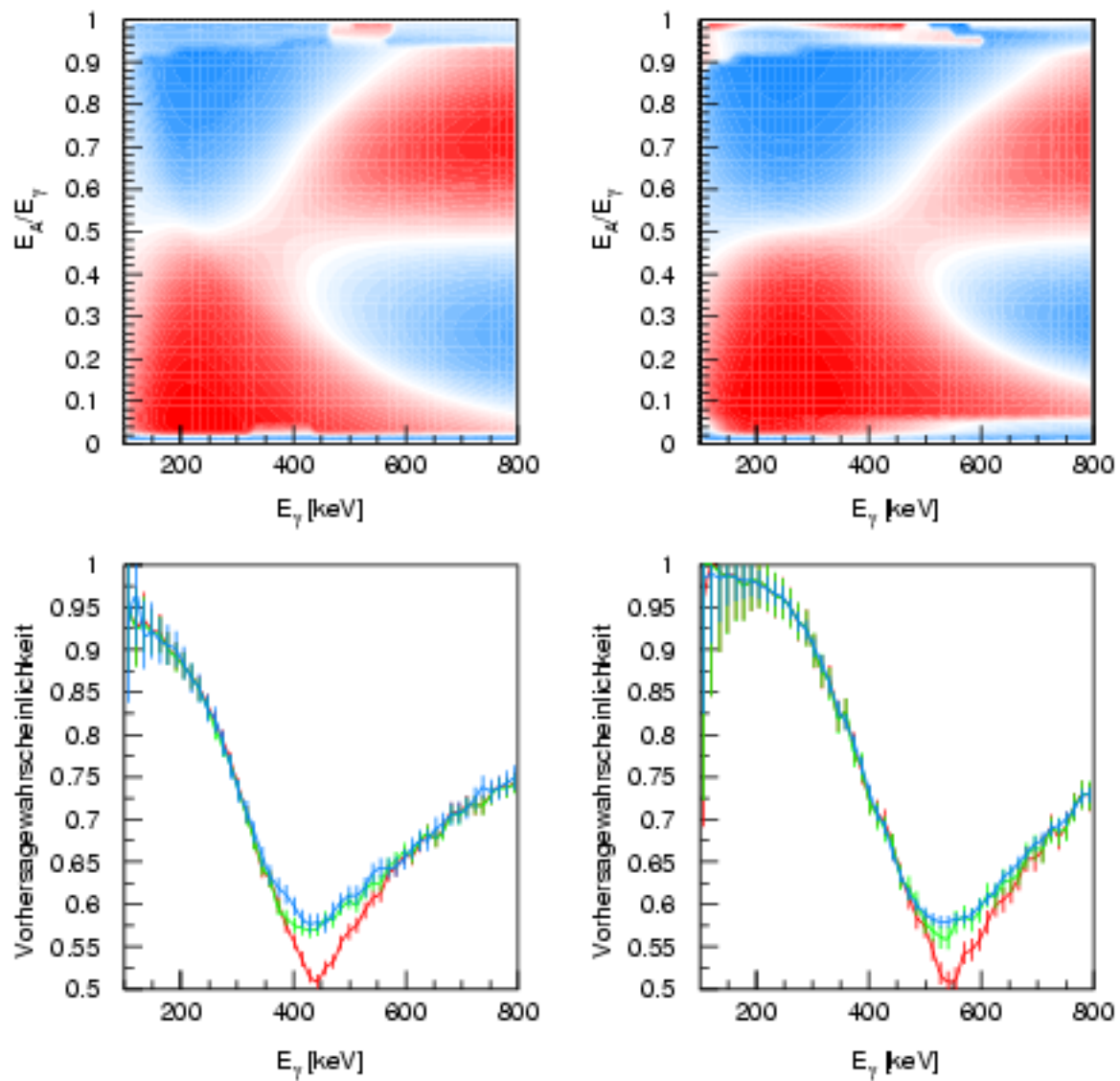


Abbildung 24: Ausgabe des 4-20-4 Netzes für benachbarte (links) und diagonale (rechts) Segmente (oben) und daraus folgende Vorhersagewahrscheinlichkeiten (unten), MSP (grün), Abschätzung (rot) und theoretisches Limit (blau).

Randbereiche ist eine ausreichende Ähnlichkeit der Netzausgaben mit den von GEANT generierten Verteilungen, genauer der daraus zu erwartenden Bayesschen Wahrscheinlichkeiten, festzustellen. Quantitativ ergeben sich die ebenfalls in Abbildung 24 dargestellten Trefferwahrscheinlichkeiten. Die Fehlerbalken stellen den statistischen Fehler dar. Die roten Kurven kennzeichnen jeweils die Wahrscheinlichkeiten einer richtigen Vorhersage durch eine Abschätzung nach Kapitel 4.1.1, die grüne Kurve die Wahrscheinlichkeit einer richtigen Klassifikation durch das MSP. Die blaue Kurve schließlich gibt das aus den Verteilungen durch Abbildung 21 ermittelte theoretische Maximum an. Wiederum erreicht das MSP im Rahmen der statistischen Fehler die maximal mögliche Vorhersagekraft.

Nachdem der Fall einer Segmentmultiplizität von zwei diskutiert wurde, ist nun die Netzausgabe für die Segmentmultiplizität drei zu untersuchen. Abbildung 25 zeigt dazu die Trefferwahrscheinlichkeit des Netzes in Abhängigkeit von der γ -Energie in grün, die einer Abschätzung anhand Abbildung 22 in rot. Die blauen Punkte geben für sieben diskrete Energien die jeweils maximal mögliche Trefferwahrscheinlichkeit an; ihre Ableitung wird später angedeutet. Über weite Energiebereiche erreicht das Netz die maximal mögliche Vorhersagewahrscheinlichkeit und liegt damit über der skizzierten Abschätzung.

Eine Abbildung analog zu Abbildung 16 bzw. Abbildung 21 ist nun nicht mehr möglich, da neben der γ -Energie zwei weitere Achsen zur Kennzeichnung der Verteilung der γ -Energie über die drei Segmente notwendig wären²². Stattdessen werden in Abbildung 26 für drei verschiedene γ -Energien einzelne Bilder gezeigt. Zur Vereinfachung werden die drei betrachteten Segmente mit A, B und C bezeichnet. B kennzeichnet dabei immer das mittlere der drei Segmente, A das ihm im Uhrzeigersinn folgende, C das verbleibende Segment. Die Abszisse gibt jeweils den Bruchteil der γ -Energie E_γ an, welche in Segment A, die Ordinate den Bruchteil der Energie, welche in Segment B deponiert wurde. In dieser Ebene liegen alle Ereignisse unterhalb einer Diagonalen durch die Punkte (0,1) und (1,0). Bei Ereignissen auf der Diagonalen addieren sich die beiden Energiebruchteile bereits zu eins, so daß im dritten Segment C keine Energie mehr deponiert sein kann und somit kein Ereignis mit Segmentmultiplizität drei vorliegt. Je weiter man sich von dieser Diagonalen zum Ursprung hin bewegt, desto weniger Energie ist in den Segmenten A und B, desto mehr in Segment C deponiert. Bei einer gegebenen γ -Energie ist also durch Wahl eines Punktes in der Ebene die Verteilung der Energie über die drei Segmente eindeutig festgelegt. In dieser Ebene können nun bequem drei Häufigkeitsverteilungen eingetragen werden, nämlich die Verteilungen der Ereignisse, bei denen jeweils Segment A, B oder C als erstes anspricht. Um nicht drei sich schneidende Häufigkeitsverteilungen betrachten zu müssen, wurde jeder Verteilung eine der Farben Magenta, Cyan und Gelb zugeordnet. Die Farbe eines Punktes in der skizzierten Ebene ergibt sich nun durch Mischung dieser drei Farben, wobei die Anteile der drei Farben in jedem Punkt den Anteilen der jeweiligen Häufigkeitsverteilung in diesem Punkt entsprechen.

Die obere Zeile von Abbildung 26 zeigt die Überlagerung dieser Verteilungen für γ -Energien von 300, 500 und 800 keV. Nach Abbildung 22 ist zu erwarten, daß bei hohen Energien mit großer Wahrscheinlichkeit das Segment als erstes anspricht, in dem die meiste Energie deponiert wurde. Dies wird durch das entsprechende Bild bei 800 keV in Abbildung 26, obere Zeile rechts, bestätigt. Die drei Verteilungen häufen sich in den Ecken des durch die beschriebene Diagonale gebildeten Dreiecks, so daß diese Ecken in den in den drei reinen, nicht gemischten Farben erscheinen. Die obere Ecke - sie kennzeichnet einen Bereich, in dem in Segment B die meiste Energie deponiert wurde - erscheint in Magenta, der Farbe, die den Bereich kennzeichnet, in dem Segment B zuerst anspricht. Entsprechendes gilt für linke untere Ecke, die Farbe Gelb und Segment C, sowie für die rechte untere Ecke, Cyan und Segment A. Im Falle einer kleinen γ -Energie wie etwa bei 300 keV

²²Da davon ausgegangen wird, daß das Quant vollständig im Detektor absorbiert wird, kann auf die Darstellung der Energie eines Segments verzichtet werden, da diese über die Summe der Einzelenergien festgelegt wird.

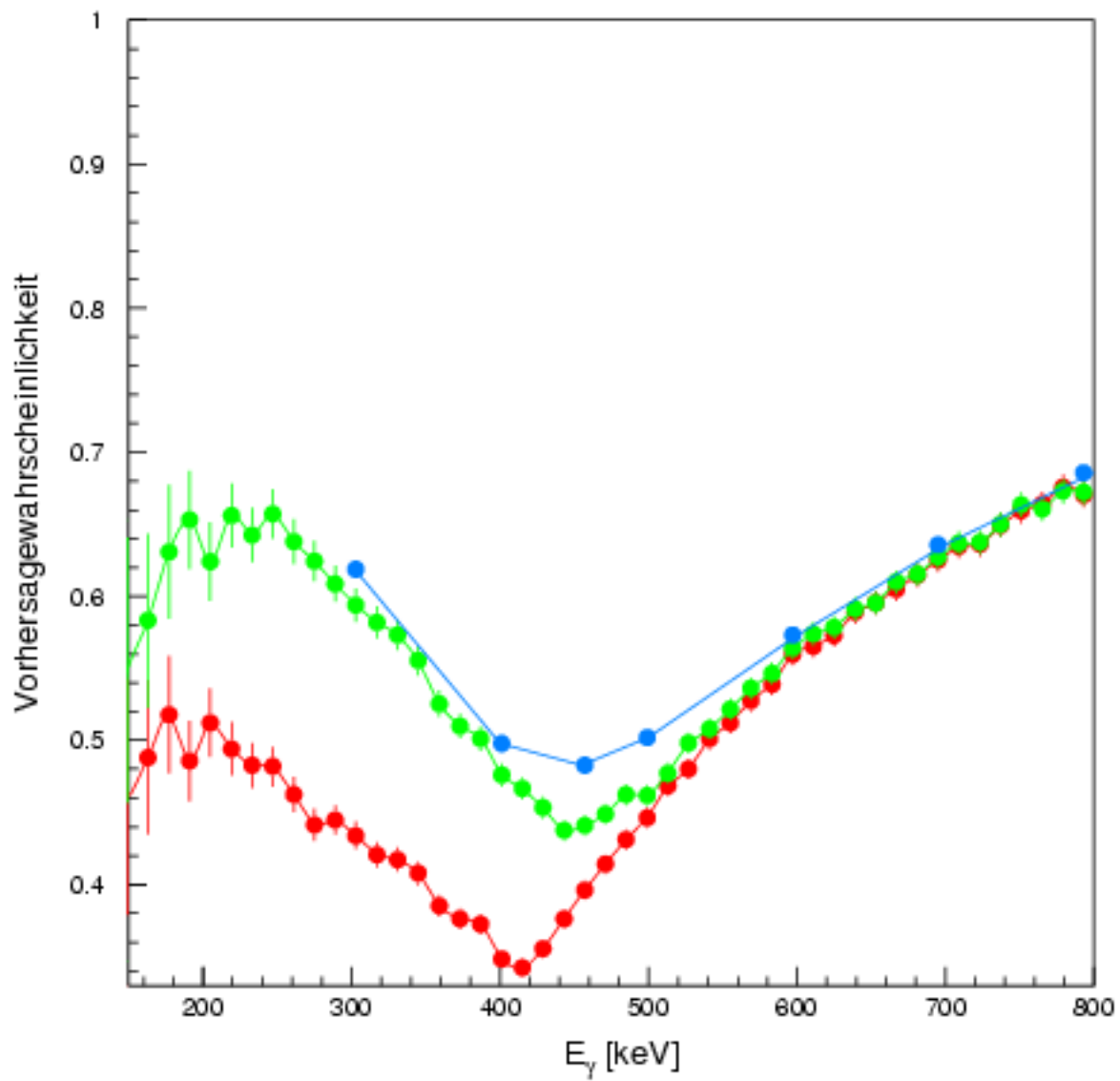


Abbildung 25: Vorhersagewahrscheinlichkeiten bei Segmentmultiplizität drei; MSP (grün), Abschätzung (rot) und theoretisches Limit (blau).

dagegen sprachen nach Abbildung 22 mit etwa gleicher Wahrscheinlichkeit die Segmente, in denen die mittlere bzw. geringste Energie deponiert wurde, zuerst an. Dabei blieb die Verteilung der Energie über die Segmente jedoch unberücksichtigt; es ergäbe sich eine Trefferquote richtiger Vorhersagen des ersten Segments von 50 Prozent. Die Verteilung in Abbildung 26, obere Zeile links zeigt, daß bei dieser Energie jedoch eine höhere Trefferquote möglich ist. Die obere Ecke des Dreiecks erscheint in Grün, das aus gleichen Teilen Gelb und Cyan zusammengesetzt ist. Sofern also in Segment B die größte Energie deponiert wurde, sprachen mit gleicher Wahrscheinlichkeit Segment A oder C zuerst an, es folgt eine lokale Trefferquote von 50 Prozent. Die linke untere Ecke des Dreiecks erscheint in Blaugrün, welches aus zwei Teilen Cyan und einem Teil Magenta zusammengesetzt ist. Daher sprachen, sofern in Segment C die größte Energie deponiert wurde, nun nicht mehr A und B mit gleicher Häufigkeit zuerst an, sondern im Verhältnis 2:1, so daß hier eine Trefferquote von 66 Prozent möglich ist. Die untere rechte Ecke schließlich erscheint in Orange, welches aus zwei Teilen Gelb und einem Teil Magenta entsteht. Daher sprachen Segmente B und C mit Wahrscheinlichkeiten im Verhältnis 1:2 an, sofern in A die größte Energie deponiert wurde; erneut beträgt die maximale Trefferquote 66 Prozent. In erster Näherung folgt somit bei 300 keV eine globale Trefferquote von etwa 60 Prozent, was durch Abbildung 22 gestützt wird. Die mittlere der beiden Verteilungen in der oberen Zeile zeigt die Situation bei einer Energie von 500 keV. Hier ergeben sich nach Abbildung 22 ohne Berücksichtigung der Verteilung der Energie über die drei Segmente in etwa gleiche Wahrscheinlichkeiten, daß die Segmente größter, mittlerer und geringster Energie als erstes ansprechen; es folgt somit eine Trefferquote von 33 Prozent. Die Verteilung zeigt jedoch deutliche Strukturen, die zur Verbesserung der Trefferquote ausgenutzt werden können.

Die untere Zeile in Abbildung 26 zeigt nun die Ausgabesignale des MSPs bei den drei entsprechenden Energien von 300, 500 und 800 keV. Es ist deutlich zu sehen, daß die Ausgaben des Netzes wie im Falle einer Segmentmultiplizität von zwei die durch GEANT bestimmten Strukturen approximieren und dadurch eine größere Vorhersagekraft als durch die Abschätzung zustande kommt. In den Randbereichen der skizzierten Dreiecke fallen die zugrundeliegenden Häufigkeitsverteilungen ab, die Netzausgaben jedoch nicht; sie kennzeichnen - im Bild durch eine tiefe Farbe - die dort wahrscheinlichste Klasse, so daß erneut eine Interpretation der Netzausgabe als Bayessche Wahrscheinlichkeit naheliegt. Der dazu nach Gleichung (3) bestimmte Abstand beträgt 8.16×10^{-4} , die Summe der Ausgabesignale der Knoten der Ausgabeschicht summieren sich mit einer absoluten Genauigkeit von fünf Prozent zu eins. Dadurch wird die Interpretation der Netzausgaben als Bayessche Wahrscheinlichkeiten gestützt. Die Netzausgabe bei 500 keV entspricht der vorgegebenen Verteilung noch nicht in allen Einzelheiten, was zur Abweichung der Trefferquote von dem theoretischen Maximum bei dieser Energie in Abbildung 25 führt.

Die theoretisch maximale Trefferquote in Abbildung 25 wurde durch Auswertung von sieben Diagrammen analog zu denen in der oberen Zeile von Abbildung 26 bestimmt. Dazu wurde für jeden Punkt der Ebene die maximale der drei durch GEANT bestimmten Verteilungen verwendet und die somit bestimmten lokalen Trefferwahrscheinlichkeiten mit der Verteilung aller Ereignisse über die Ebene gewichtet.

Vier der fünf 4-20-4 Netze mit jeweils 184 Parametern wurden weiterhin mittels OBS beschnitten, dabei wurden 18, 24, 25 und erneut 25 Synapsen entfernt, bevor ein Anstieg des Fehlers auf dem Lerndatensatz zu verzeichnen war. Erneut ist anzumerken, daß nicht angestrebt wurde, systematisch ein möglichst minimales Netz zu erreichen.

Abschließend bleibt festzuhalten, daß die Netzausgabe im Fall einer Segmentmultiplizität von drei

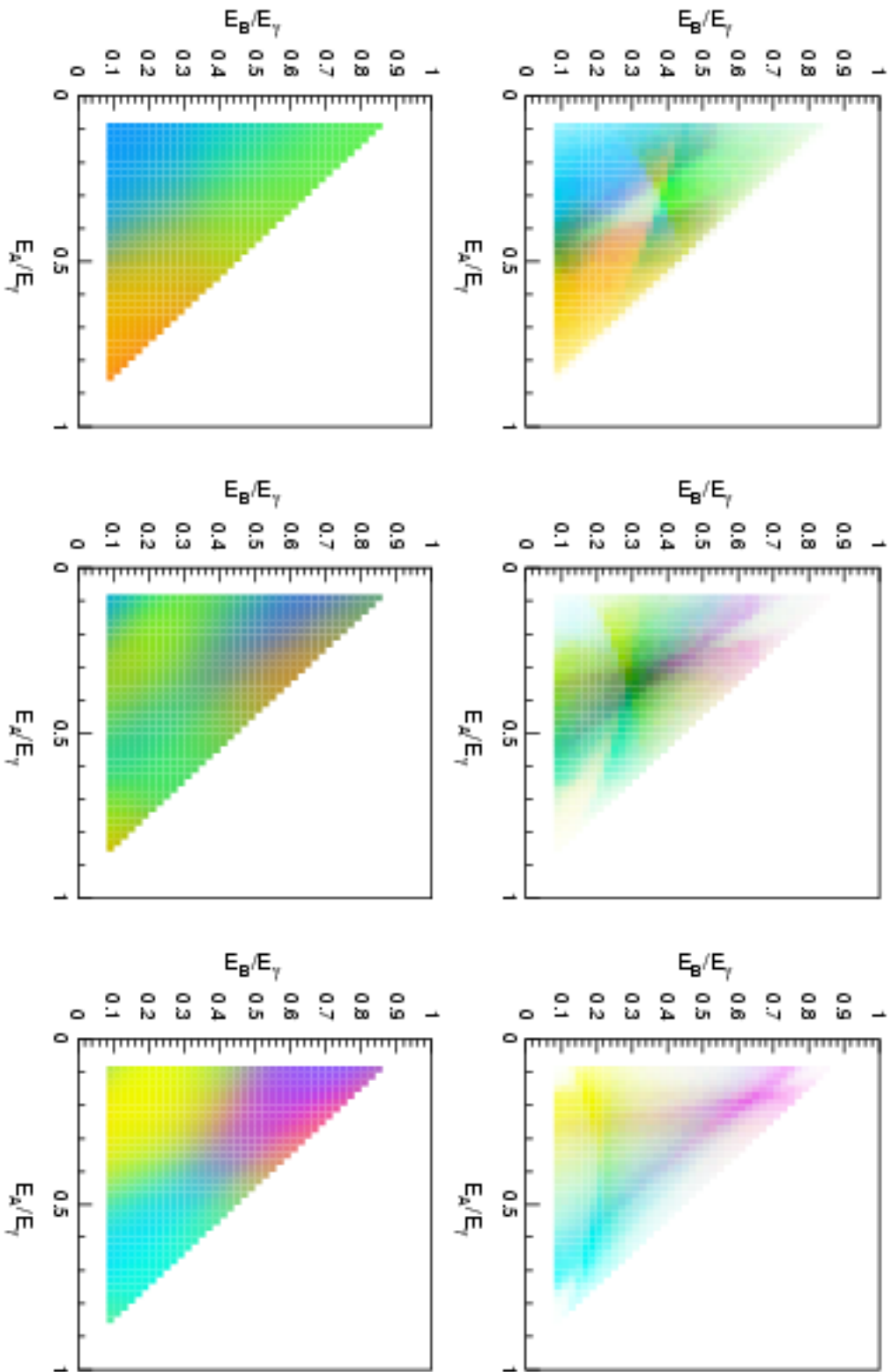


Abbildung 26: Überlagerungen der Verteilungen, mit denen Segment A, B und C als erstes anspricht (oben) und Ausgabe des 4-20-1 Netzes (unten) bei 300, 500 und 800 keV.

die durch GEANT bestimmten Verteilungen deutlich charakterisiert. Diese Verteilungen können nun nicht mehr auf einfache Weise und mit zugleich hoher Genauigkeit wie im Falle einer Segmentmultiplizität von zwei durch wenige Geraden approximiert werden. Mag dies bei einzelnen γ -Energien nach Abbildung 26 noch durchführbar erscheinen, so müßte zur Überdeckung weiter Energiebereiche eine Vielzahl derartiger Geraden für diskrete Energien konstruiert und ggfs. zwischen diesen interpoliert werden. Insbesondere im Fall komplexerer Geometrien wie der des SEGMENTED CLOVER-Detektors im folgenden Abschnitt ist ein derartiges Vorgehen nicht mehr praktikabel, so daß der Vorteil, den ein MSP zur Vorhersage des Segments der ersten Wechselwirkung bei unterschiedlichen Segmentmultiplizitäten und den dabei auftretenden unterschiedlichen Kombinationen ansprechender Segmente über weite Energiebereiche bietet, deutlich hervortritt.

4.1.4 16fach segmentierter Detektor

Aus vier der im letzten Kapitel vorgestellten vierfach segmentierten Germaniumkristalle wird nun ein insgesamt 16fach segmentierter Detektor, der SEGMENTED CLOVER, zusammengesetzt. Bei diesem Detektor sind die einzelnen zylinderförmigen Kristalle keilförmig geschnitten, so daß sie zur Struktur eines vierblättrigen Kleeblattes nach Abbildung 27 zusammengesetzt werden können. An dieser Stelle sind lediglich die in den vorherigen Kapiteln vorgestellten Effekte der Streuung einzelner Quanten zwischen den einzelnen Segmenten von Interesse. Für Details der Konstruktion des Detektors oder eine Diskussion seiner Eigenschaften sei auf [Ger92] verwiesen.

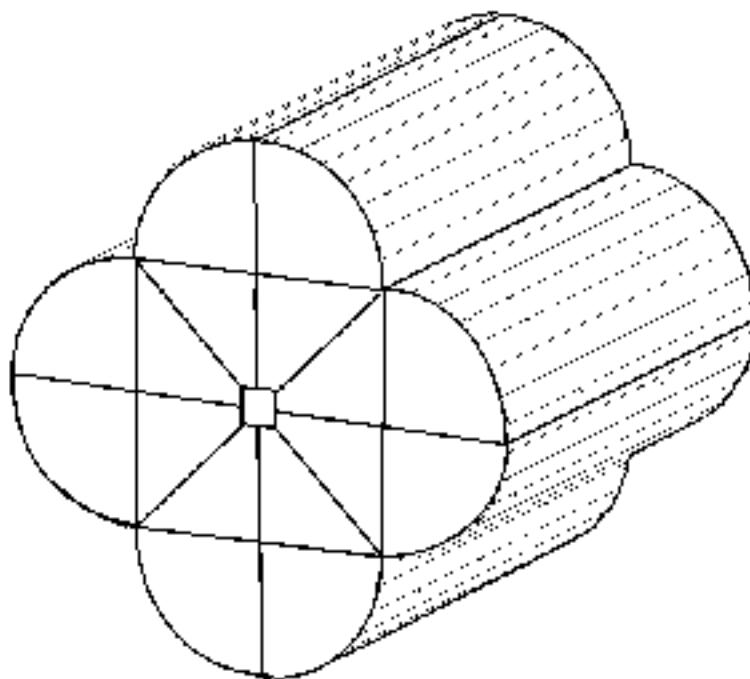


Abbildung 27: Schematische Darstellung des SEGMENTED CLOVER-Detektors aus vier Kristallen zu je vier Segmenten.

Mit Hilfe von GEANT wurde der Transport von γ -Quanten durch diese Geometrie untersucht, die Quelle befand sich wiederum 25 cm von der Stirnfläche des Detektors entfernt und emittierte isotrop Quanten einer zwischen 100 und 800 keV gleichverteilten Energie. Die einzelnen Kristalle hatten einen Durchmesser von sieben Zentimetern und eine Länge von 14 cm. Dabei ergaben sich Verteilungen der Segmentmultiplizitäten in Abhängigkeit von der Einschußenergie nach Abbildung 28. Die Wahrscheinlichkeit für zwei ansprechende Segmente beträgt bereits ab 250 keV zehn Prozent, die für drei ansprechende Segmente erreicht bei dieser Energie ein, ab etwa

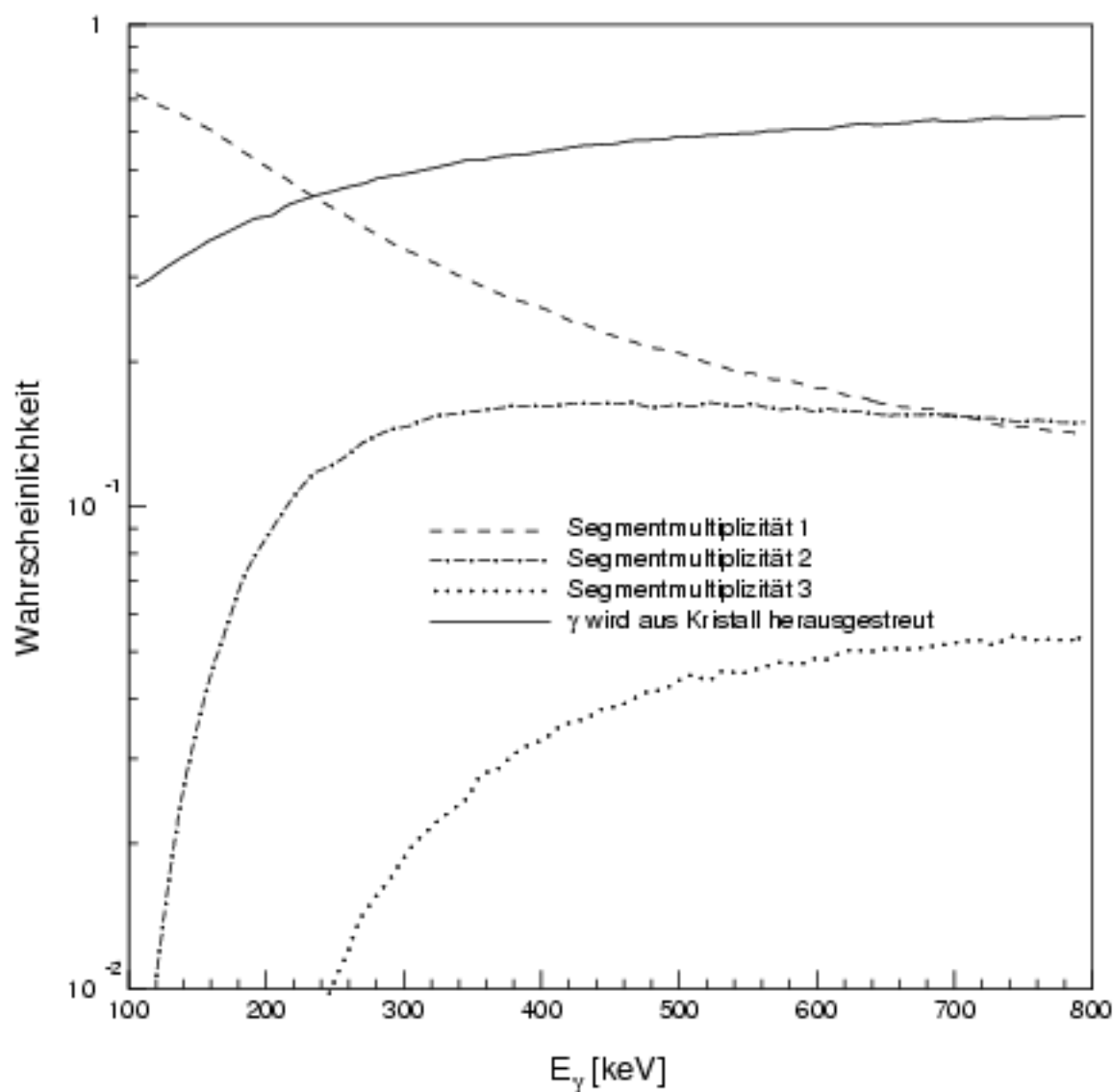


Abbildung 28: Multiplizitäten ansprechender Segmente im SEGMENTED CLOVER-Detektor.

600 keV fünf Prozent. Höhere Segmentmultiplizitäten liegen bei dieser Geometrie unter einem Prozent.

Zum Training der mehrschichtigen Perzeptrons wurden wiederum drei unabhängige Datensätze generiert, ein Lern- und ein Kontrolldatensatz von jeweils 30000 Mustern und ein weiterer Datensatz zur abschließenden Auswertung von 500000 Ereignissen einer Segmentmultiplizität von zwei und 500000 Ereignissen einer Segmentmultiplizität von drei. Die Verteilung der Segmentmultiplizitäten in Lern- und Kontrolldatensatz wurde nicht wie im Falle des vierfach segmentierten Detektors verzerrt; daher ist zu erwarten, daß die Vorhersagewahrscheinlichkeit der Netze für die seltener vorkommenden Ereignisse mit Segmentmultiplizität drei nicht optimal sein wird. Die Struktur der Datensätze bestand erneut aus den Segmentenergien in MeV und den jedem Segment zugeordneten Klassen. Somit ergeben sich jeweils 16 Ein- und Ausgabeknoten der untersuchten Netze. Die Zugehörigkeit bzw. Nichtzugehörigkeit eines Energiemusters zu einer Klasse wurde dabei nicht wie bislang durch die Ausgabesignale 1 und 0, sondern nunmehr durch die Werte 1 und -1 gekennzeichnet. Als Ausgabefunktion aller Knoten im Netz diente der Tangens Hyperbolicus.

Es wurden drei verschiedene Netze, ein 16-30-16, ein 16-50-16 und ein 16-80-16 Netz mit dem RPROP-Algorithmus trainiert und die Entwicklung des Fehlers auf Lern- und Kontrolldatensatz betrachtet. Wiederum wurden jeweils fünf Startpunkte im Fehlergebirge durch zufällige Wahl der Startsynapsen aus dem Intervall $[-1,+1]$ gewählt und eine Fehlerminimierung durchgeführt, wiederum wurde das Training abgebrochen, sobald sich der Fehler auf dem Lerndatensatz zwischen 100 Epochen um weniger als 10^{-7} änderte. Das anfängliche Training wurde auf 10^4 Epochen begrenzt.

Es zeigt sich, daß ähnlich zum Fall des vierfach segmentierten Detektors keine signifikanten Unterschiede zwischen den fünf Durchgängen pro Netz bestehen. Der Fehler auf dem Kontrolldatensatz liegt ebenfalls knapp über dem des Lerndatensatzes; auch ist kein Auseinanderlaufen der beiden Fehler zu beobachten. Im Falle des 16-30-16 Netzes fallen die Fehler auf dem Kontrolldatensatz auf 0,0466 bis 0,0469, auf dem Lerndatensatz auf 0,0464 bis 0,0468 im Fall des 16-50-16 Netzes auf 0,0459 bis 0,0462 bzw. 0,0456 bis 0,0462. Die geringsten Fehler zeigt das 16-80-16 Netz mit Werten von 0,0455 bis 0,0460 auf dem Lern- und 0,0459 bis 0,0463 auf dem Kontrolldatensatz. Nach Fortsetzen des Trainings um weitere 30000 Epochen war eine Synapsenkonfiguration gefunden, die einen Fehler von 0,0442 auf dem Lern- und 0,0451 auf dem Kontrolldatensatz zeigt. Mit diesem Netz wurde die Vorhersagewahrscheinlichkeit in Abhängigkeit von der γ -Energie für die Fälle einer Segmentmultiplizität von zwei und drei bestimmt und mit der skizzierten Abschätzungen verglichen, siehe Abbildung 29, untere Zeile. Es zeigt sich erneut ein Vorteil des Netzes gegenüber der Abschätzung. Im Falle einer Segmentmultiplizität von drei bleibt die Wahrscheinlichkeit des Netzes für hohe Energien knapp unter der der Abschätzung zurück, was darauf zurückzuführen ist, daß diese Muster im Lerndatensatz um etwa eine Größenordnung seltener vorkommen als die Muster einer Segmentmultiplizität von zwei. Die obere Zeile von Abbildung 29 zeigt die unzureichende Vorhersagekraft des 16-30-16 Netzes.

Eine Unterscheidung zwischen verschiedenen Kombinationen, in denen zwei oder mehr Segmente ansprechen können - wie im Falle benachbarter und diagonalen Segmente des vierfach segmentierten Detektors - wäre zwar prinzipiell denkbar, soll aber an dieser Stelle unterbleiben. Es hat sich gezeigt, daß die entsprechenden Verteilungen im Rahmen der vorliegenden Statistik keine signifikanten Unterschiede aufweisen, sofern die beiden beteiligten Segmente sich jeweils entlang einer Kante berühren. Dabei ist es ohne Bedeutung, ob die Segmente viertelkreis- oder trapezförmig sind. Daher ist lediglich mit dem Auftreten zweier Klassen wie im Fall des vierfach segmentierten Detektors zu rechnen; einer Klasse, in der sich die Segmente entlang einer Kante berühren und einer anderen, in

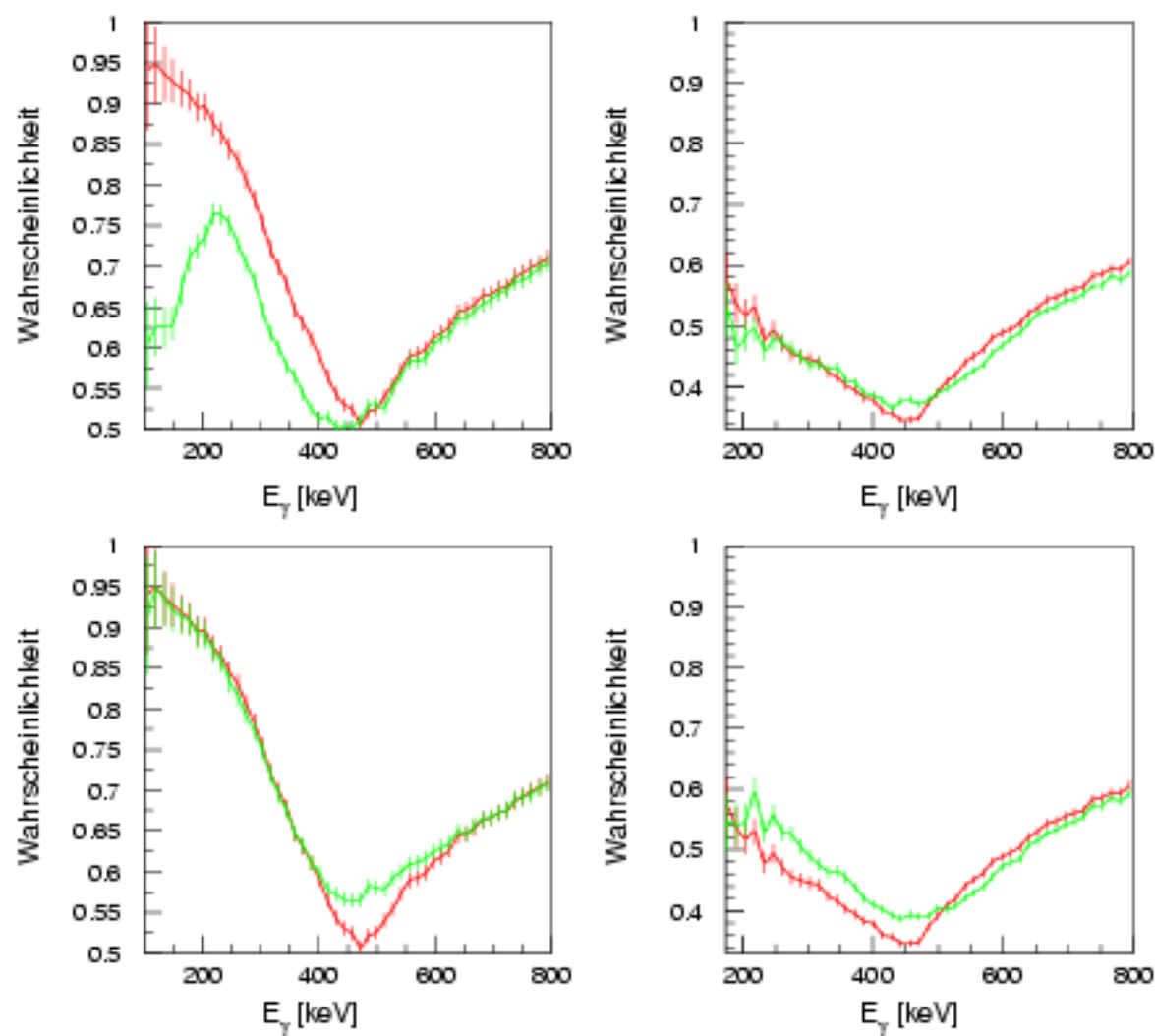


Abbildung 29: Vorhersagewahrscheinlichkeiten (grün) des 16-30-16 (oben) und 16-80-16 Netzes (unten) bei Segmentmultiplizität zwei (links) und drei (rechts) und den jeweiligen Abschätzungen (rot).

der sie sich in nur einer Ecke berühren. Da bereits im vorigen Kapitel gezeigt wurde, daß ein MSP diese Klassen zu unterscheiden vermag, wird an dieser Stelle diese Untersuchung nicht wiederholt. Der Fall, daß zwei Segmente weiter auseinanderliegen und sich in keinem Punkt mehr berühren, kommt zu selten vor, als daß er an dieser Stelle als eigene Klasse betrachtet und statistisch signifikante Aussagen gemacht werden könnten.

Stattdessen wird untersucht, wie genau sich der tatsächliche Ort der ersten Wechselwirkung eines γ -Quants mit dem Detektormaterial eingrenzen lassen kann. Dazu ist in Abbildung 30 die Verteilung der Koordinaten des durch GEANT bekannten, wirklichen Ortes der ersten Wechselwirkung für eine Segmentmultiplizität von zwei dargestellt. Die Kanten der einzelnen Segmente treten erwartungsgemäß hervor. Dies kommt dadurch zustande, daß die Auftreffpunkte aller Quanten zwar gleichmäßig über die Stimfläche des Detektors verteilt sind, daß genau zwei Segmente aber bevorzugt dann ansprechen, wenn dieser Ort in der Nähe einer Segmentgrenze liegt und das Quant durch eine Streuung aus einem Segment in ein anderes gelangt. Bewegt man sich entlang einer Senkrechten auf einer Segmentgrenze weg von dieser in ein Segment hinein, so nimmt die Häufigkeitsverteilung der Orte der ersten Wechselwirkung exponentiell ab. Die Unterschiede dieses exponentiellen Abfalls in Abhängigkeit vom Segmenttyp - Viertelkreis oder Trapez an unterschiedlichen Stellen des Detektors - sind zu gering, als daß sie an dieser Stelle diskutiert werden müssen. Hat man also - sei es durch eine Abschätzung, sei es durch ein MSP - im Falle einer Segmentmultiplizität von zwei das Segment, in dem die erste Wechselwirkung stattfand mit einer gewissen Wahrscheinlichkeit bestimmt, so kann man zur Dopplerkorrektur der γ -Spektren auf den Schwerpunkt der skizzierten Verteilung korrigieren; man verwendet als Ort der ersten Wechselwirkung also den Schwerpunkt der skizzierten Verteilung.

Darüberhinaus kann der Versuch unternommen werden, mit Hilfe eines MSPs nicht nur das Segment vorhersagen zu lassen, in dem die erste Wechselwirkung stattgefunden hat, sondern dieses vielmehr direkt dazu zu verwenden, die Koordinaten des Ortes der ersten Wechselwirkung zu bestimmen. Dazu wurden zwei Strukturen des Typs 16-16-2 und 16-50-2 untersucht. Als Eingabegrößen dienten wiederum die 16 Segmentenergien in MeV, als Sollaussabegrößen die durch GEANT bestimmten (x,y) Koordinaten des Ortes der ersten Wechselwirkung in Zentimetern. Als Ausgabefunktion wurde der Tangens Hyperbolicus verwendet, er wurde lediglich in der Ausgabeschicht durch die Identität ersetzt. Die sonstigen Rahmenbedingungen des Trainings wurden übernommen, der Lern- und Kontrolldatensatz bestanden aus jeweils 30000 Mustern, die abschließende Beurteilung nach Abbildung 30 erfolgte mit Datensätzen mit jeweils 500000 Mustern für Segmentmultiplizitäten von zwei und drei. Nach 10^4 Epochen fielen die Fehler von fünf Gradientenabstiegen auf Lern- und Kontrolldatensatz des 16-16-2 Netzes auf Werte zwischen 0,6494 und 0,6591 bzw. 0,6550 und 0,6666. Bei beiden Fehlern ist eine Sättigung festzustellen, ein Auseinanderlaufen ist nicht zu beobachten. Während 10^4 Epochen des Trainings des 16-50-2 Netzes lief der Fehler des Lerndatensatzes in eine Sättigung von 0,6310 bis 0,6405, der Fehler des Lerndatensatzes dagegen beginnt nach etwa 3000 Epochen deutlich unter den des Kontrolldatensatzes zu fallen und beträgt nach 10^4 Epochen 0,5973 bis 0,6018. Hier ist also wieder ein Auswendiglernen des Lerndatensatzes erkennbar.

Der - sei es durch Korrektur auf ein Segment, sei es direkt vom MSP ausgegebene - Ort wird nun nicht mit dem wirklichen Ort der ersten Wechselwirkung zusammenfallen, da einerseits nur mit einer gewissen Wahrscheinlichkeit das erste Segment korrekt vorausgesagt werden kann, andererseits auch die Koordinatenbestimmung durch das MSP - da ja lediglich Segmentenergien zur Verfügung stehen - fehlerbehaftet sein muß. Es entsteht folglich ein Fehler in Form eines Abstandes des vorhergesagten Punktes zum wirklichen Ort der ersten Wechselwirkung. Abbildung 31 zeigt die Häufigkeitsverteilung auftretender Abstände für verschiedene Korrekturmethode. Die Radiusabhängigkeit der

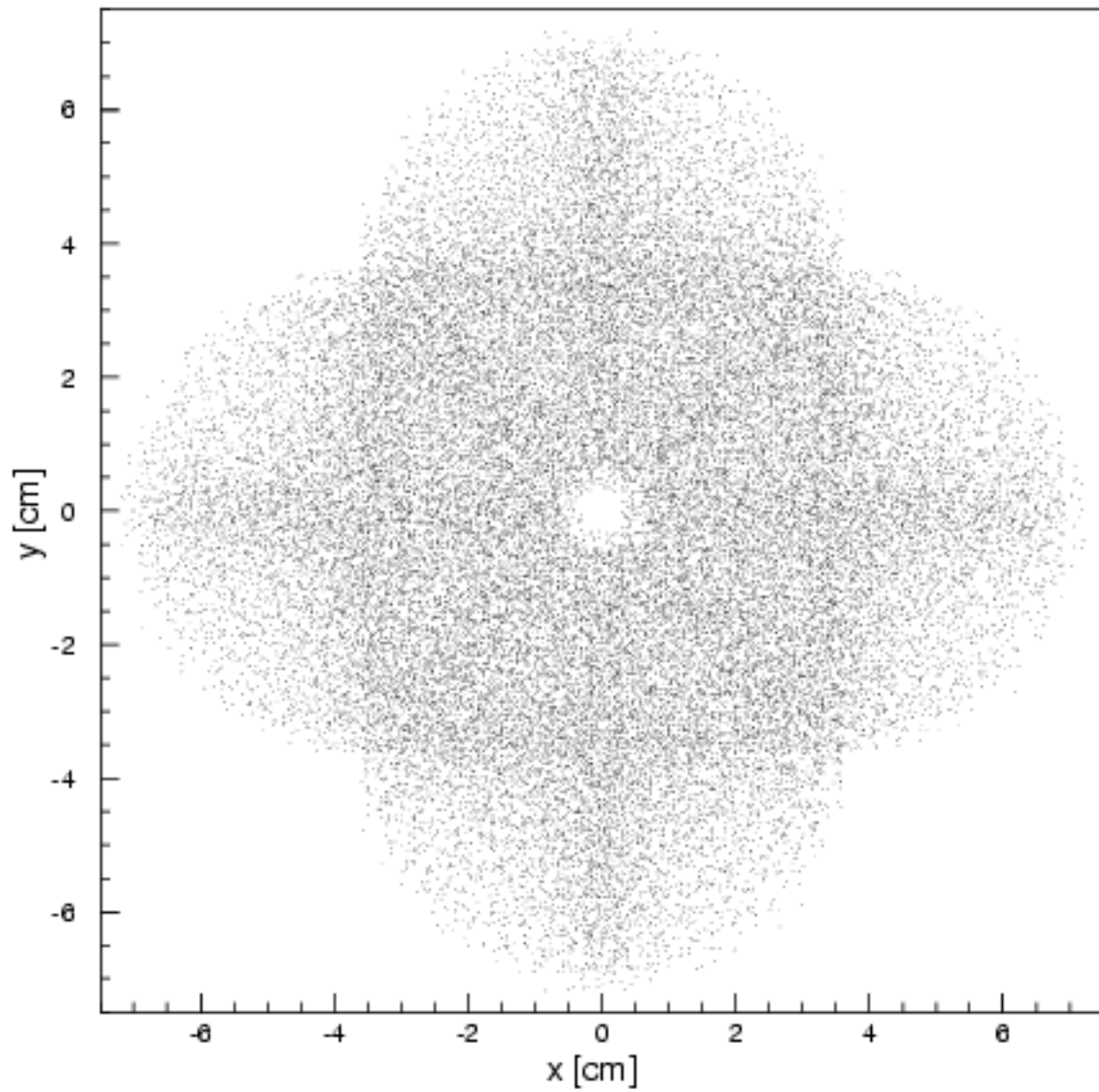


Abbildung 30: Verteilung der (x,y) -Koordinaten des Ortes der ersten Wechselwirkung im SEGMENTED CLOVER-Detektor bei einer Segmentmultiplizität von zwei.

Häufigkeitsverteilungen wurde korrigiert. Einerseits wird das durch das MSP vorhergesagte Segment verwendet, andererseits wird zum Vergleich das theoretische Limit im hypothetischen Fall einer Vorhersagewahrscheinlichkeit von eins gezeigt. Darüberhinaus ist eine Abstandsverteilung zu sehen, die entsteht, sofern beim Ansprechen zweier Segmente nach der Strassbourg-Methode generell auf die Mitte der Verbindungslinie der beiden geometrischen Segmentenschwerpunkte korrigiert wird. Schließlich sind zwei weitere Verteilungen für den Fall einer Segmentmultiplizität von zwei und drei zu sehen, welche durch Bestimmung der Koordinaten durch das skizzierte 16-50-2 Netz entstehen. Bei einer Segmentmultiplizität von zwei sind keine signifikanten Unterschiede zwischen den einzelnen Abstandsverteilungen - ihre Schwerpunkte liegen zwischen 0.8 und 0.9 cm - erkennbar, so daß die Qualität einer Dopplerkorrektur unter Einsatz eines MSPs der konventioneller Methoden vergleichbar ist. Allerdings sind alle bisherigen Korrekturmethode nicht in der Lage, eine Ortskorrektur für den Fall einer Segmentmultiplizität von drei durchzuführen. Mit dem vorgestellten MSP ist die Bestimmung des Ortes der ersten Wechselwirkung dagegen mit einem mittleren Fehler von 1.2 cm möglich.

4.2 Doppeltreffererkennung

Abschließend wird in diesem Abschnitt untersucht, inwieweit mit Hilfe eines mehrschichtigen Perzeptrons zwischen Einfach- und Doppeltreffern im Falle des SEGMENTED CLOVER-Detektors unterschieden werden kann. Dazu werden zunächst geometriebedingte Wahrscheinlichkeiten dafür bestimmt, daß ein derartiger Detektor simultan von mehreren γ -Quanten getroffen wird. Im Anschluß werden MSPs trainiert, die Muster der Segmentenergien in zwei Klassen, die eines Eintreffers und die eines Doppeltreffers, einzuteilen.

Der SEGMENTED CLOVER nach Abbildung 27 überdeckt etwa $1/72$ des vollen Raumwinkels von 4π . Daher besteht eine nichtverschwindende Wahrscheinlichkeit, daß bei hohen γ -Multiplizitäten gleichzeitig mehrere γ -Quanten den Detektor treffen. Unter der Annahme einer isotropen Emission der γ -Quanten und alleine durch geometrische Betrachtungen ohne Berücksichtigung der Detektoreffizienz zeigt Abbildung 32 die Wahrscheinlichkeiten, daß null bis vier Quanten simultan in den Detektor gelangen in Abhängigkeit von der Ereignismultiplizität. Diese entspricht der Zahl der Quanten, welche von einer Quelle zu einem Zeitpunkt isotrop emittiert werden sollen. Die Quelle befand sich 25 cm von der Stirnfläche des Detektors entfernt. Danach beträgt das Verhältnis von Einfach- zu Doppeltreffern bei einer Ereignismultiplizität von 10 etwa 16, bei 15 10, bei 20 7,5, bei 25 6 und bei 30 etwa 4,8.

Im Rahmen eines Doppeltreffers wird Energie über mehrere Segmente des Detektors verteilt, was jedoch auch durch einen Eintreffer mit anschließender Comptonstreuung erfolgen kann. Ziel einer Klassifikation ist die Unterscheidung dieser beiden Szenarien. Dazu wurden Datensätze generiert, welche wie bisher als Eingabegrößen die Energien der 16 Segmente enthalten. Als Eintreffer konnten die bisher verwendeten Ereignisse mit einer Segmentmultiplizität größer eins benutzt werden. Durch segmentweise Addition der jeweiligen Segmentenergien zufällig gewählter Eintreffer beliebiger Segmentmultiplizität wurden Doppeltreffer erzeugt. Die Ausgabe des untersuchten Netzes besaß zwei Knoten, um durch die Paare (1,-1) und (-1,1) bzw. (1,0) und (0,1) die beiden Klassen Einfach- und Doppeltreffer zu unterscheiden. Einfach- und Doppeltreffer wurden in den je 30000 Muster umfassenden Lern- und Kontrolldatensätzen und im 1000000 Muster umfassenden Datensatz zur abschließenden Beurteilung gleich gewichtet.

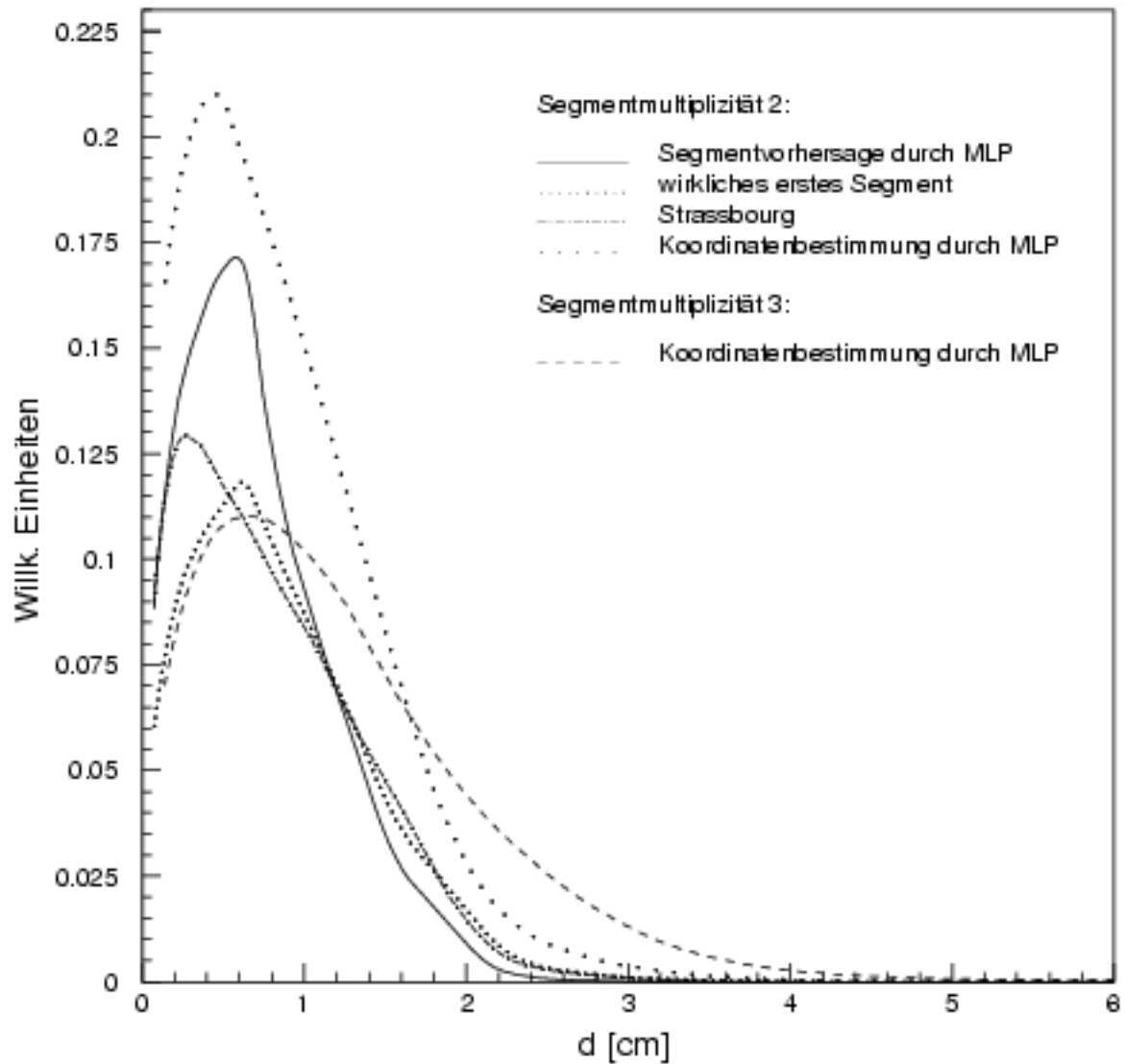


Abbildung 31: Korrigierte Verteilungen der Abstände zwischen bestimmtem und wirklichem Ort der ersten Wechselwirkung für verschiedene Vorhersagemethoden.

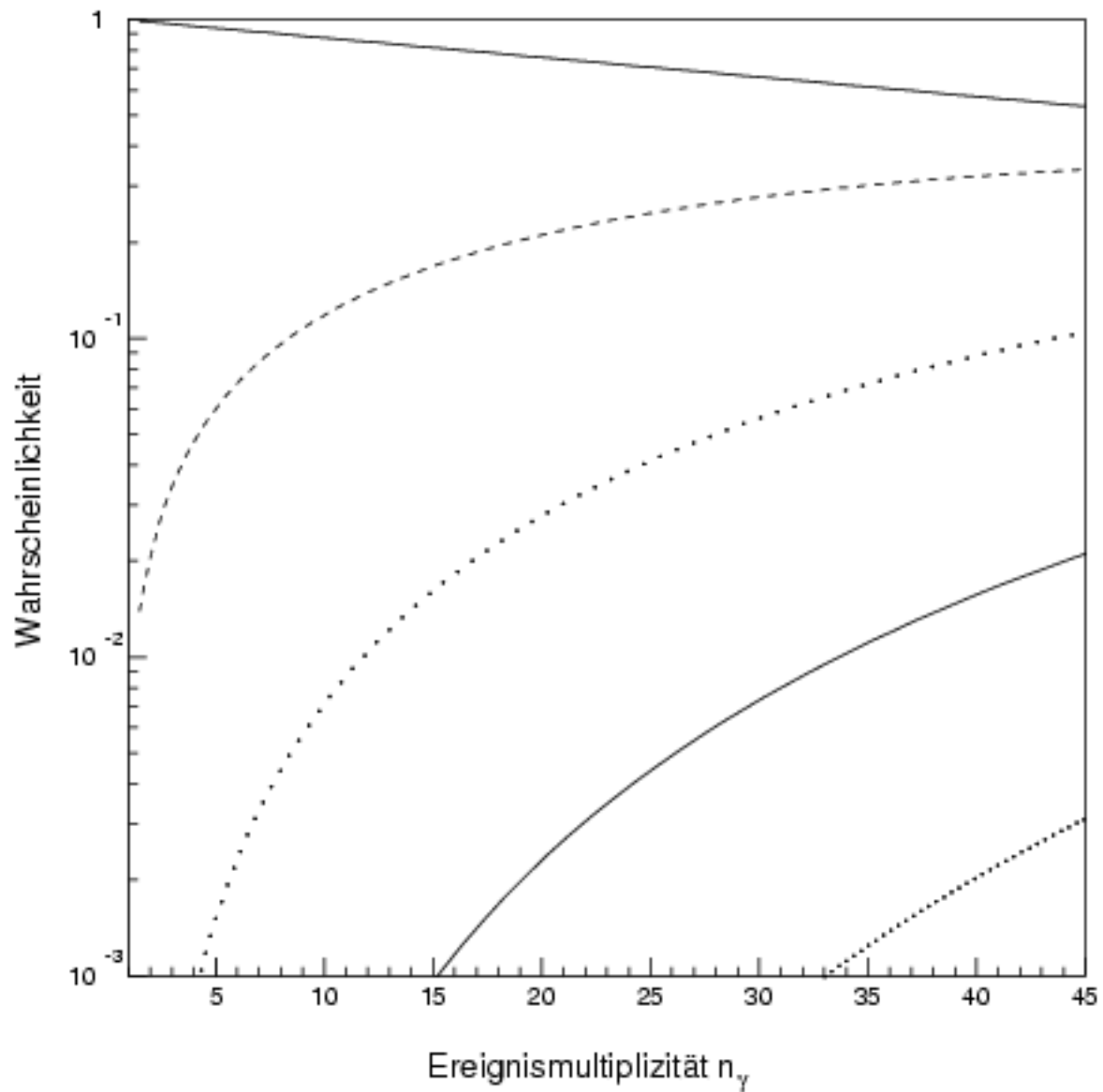


Abbildung 32: Geometriebedingte Wahrscheinlichkeiten, daß null bis vier γ -Quanten in einen SEGMENTED CLOVER-Detektor gelangen.

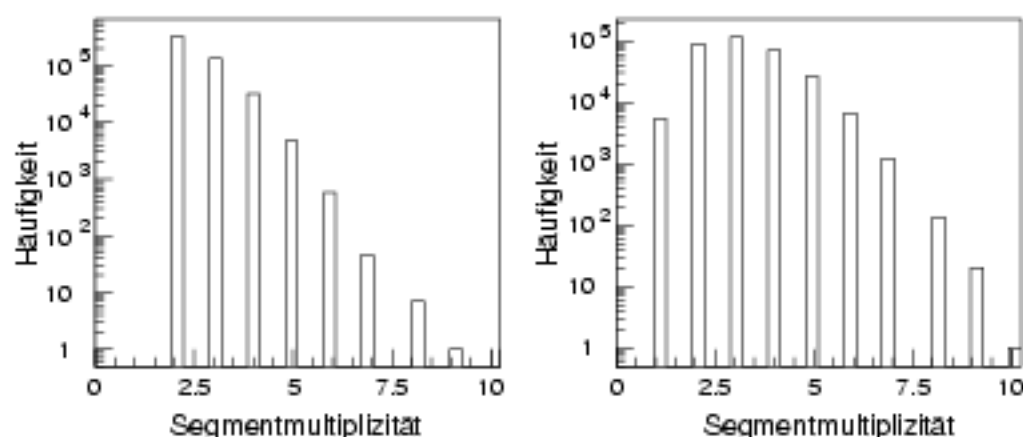


Abbildung 33: Verteilung der verwendeten Segmentmultiplizitäten für Einfach- (links) und Doppeltreffereignisse (rechts).

Abbildung 33 zeigt die Verteilung der Segmentmultiplizitäten der verwendeten Einfach- und Doppeltreffer. Um die Leistung der Netze später beurteilen zu können, wurde nach Abschätzungen gesucht, mit deren Hilfe ebenfalls eine Klassifizierung der Ereignisse nach Einfach- und Doppeltreffern möglich ist. Für den Fall einer totalen Segmentmultiplizität von zwei ist dies leicht möglich. Die im Falle eines Doppeltreffers nachgewiesenen γ -Quanten sind voneinander unabhängig und werden daher zwei zufällig aus den 16 möglichen Segmenten ausgewählte Segmente treffen. Trägt man den räumlichen Abstand der geometrischen Schwerpunkte der beiden angesprochenen Segmente auf, so ergibt sich in diesem Fall eine Gleichverteilung. Sprechen dagegen durch Comptonstreuung eines Quants zwei Segmente an, so werden diese nahe beieinander liegen, die Verteilung der Abstände der beiden Segmente wird zu großen Abständen hin exponentiell abfallen. Durch die Simulation waren diese Abstände bekannt, die Verteilungen sind in Abbildung 34 dargestellt. Somit konnte ein kritischer Segmentabstand von 5,3 cm bestimmt werden. Im Rahmen einer einfachen Abschätzung wird man davon ausgehen, daß bei einem kleineren Abstand ein Einfach-, bei einem größeren Abstand ein Doppeltreffer vorlag. Diese Abschätzung wird später als Referenz verwendet. Für die Fälle größerer Segmentmultiplizitäten wurden bisher noch keine derartigen Abschätzungen entwickelt.

Es wurden die drei Netze 16-16-2, 16-50-2 und 16-50-50-2 untersucht. Bei der 16-50-2 Struktur wurde als Ausgabefunktion die Fermifunktion, bei den anderen beiden MSPs der Tangens Hyperbolicus verwendet.

Abbildung 35 zeigt exemplarisch die Abnahme des Fehlers auf Lern- und Kontrolldatensatz für fünf Fehlerminimierungen des 16-16-2 und ein Trainingslauf des 16-50-50-2 Netzes. Da im Falle des 16-50-2 Netzes die Fermifunktion als Ausgabefunktion verwendet wurde, ist der durch sie gegebene Fehler nicht ohne weiteres mit den Fehlern der anderen beiden Netze vergleichbar, daher wird diese Fehlerentwicklung nicht abgebildet. Es zeigt sich, daß Fehler auf Lern- und Kontrolldatensatz ab etwa 400 Epochen auseinanderlaufen; der Fehler auf dem Kontrolldatensatz bleibt etwa konstant bei 0,2250 der des Trainingsdatensatzes fällt dagegen deutlich weiter auf etwa 0,2113 nach 10^4 Epochen.

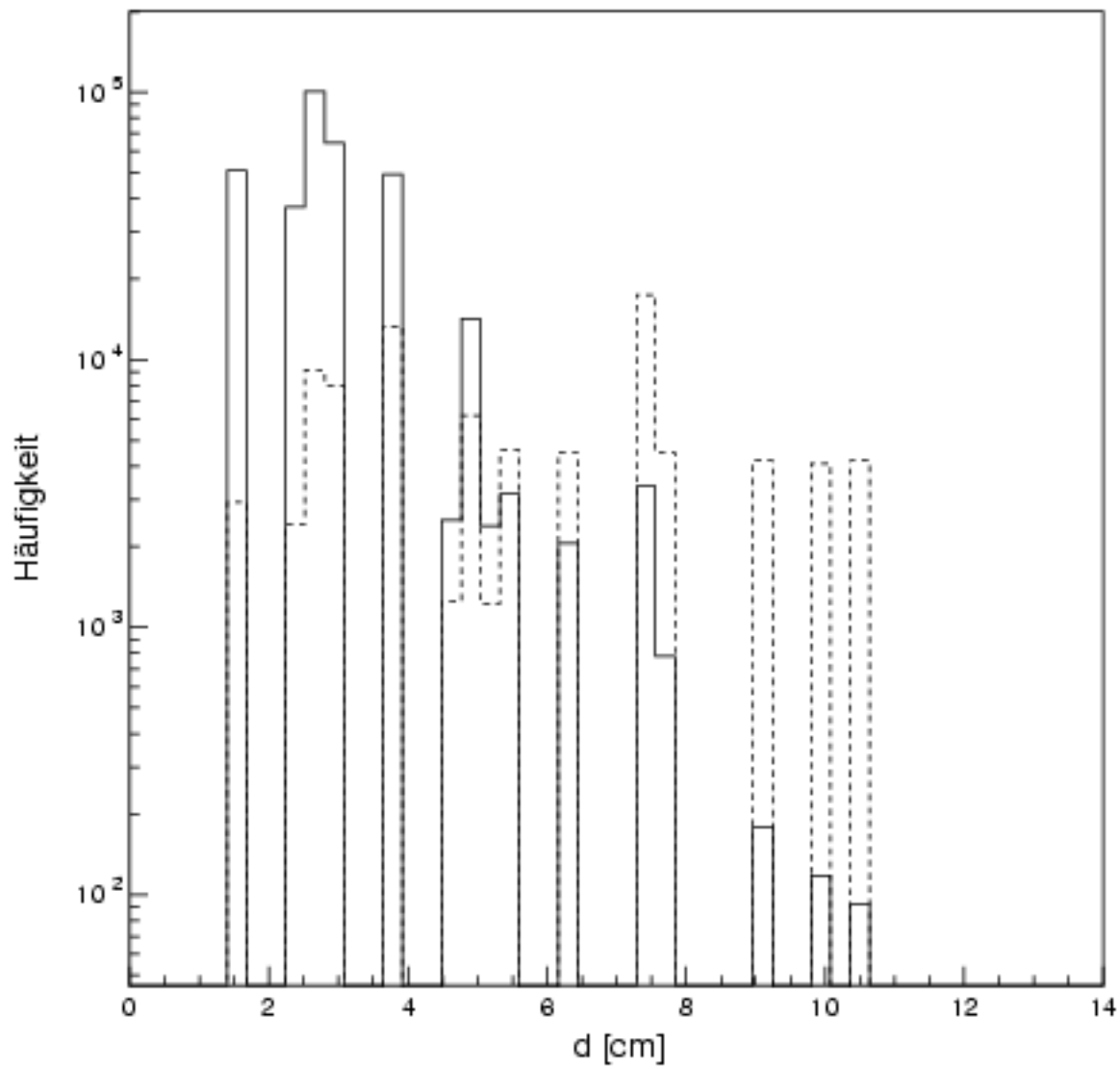


Abbildung 34: Verteilung der Abstände der angesprochenen Segmente im Falle einer Segmentmultiplizität von zwei für Einfachtreffer (durchgezogen) und Doppeltreffer (gebrochen).

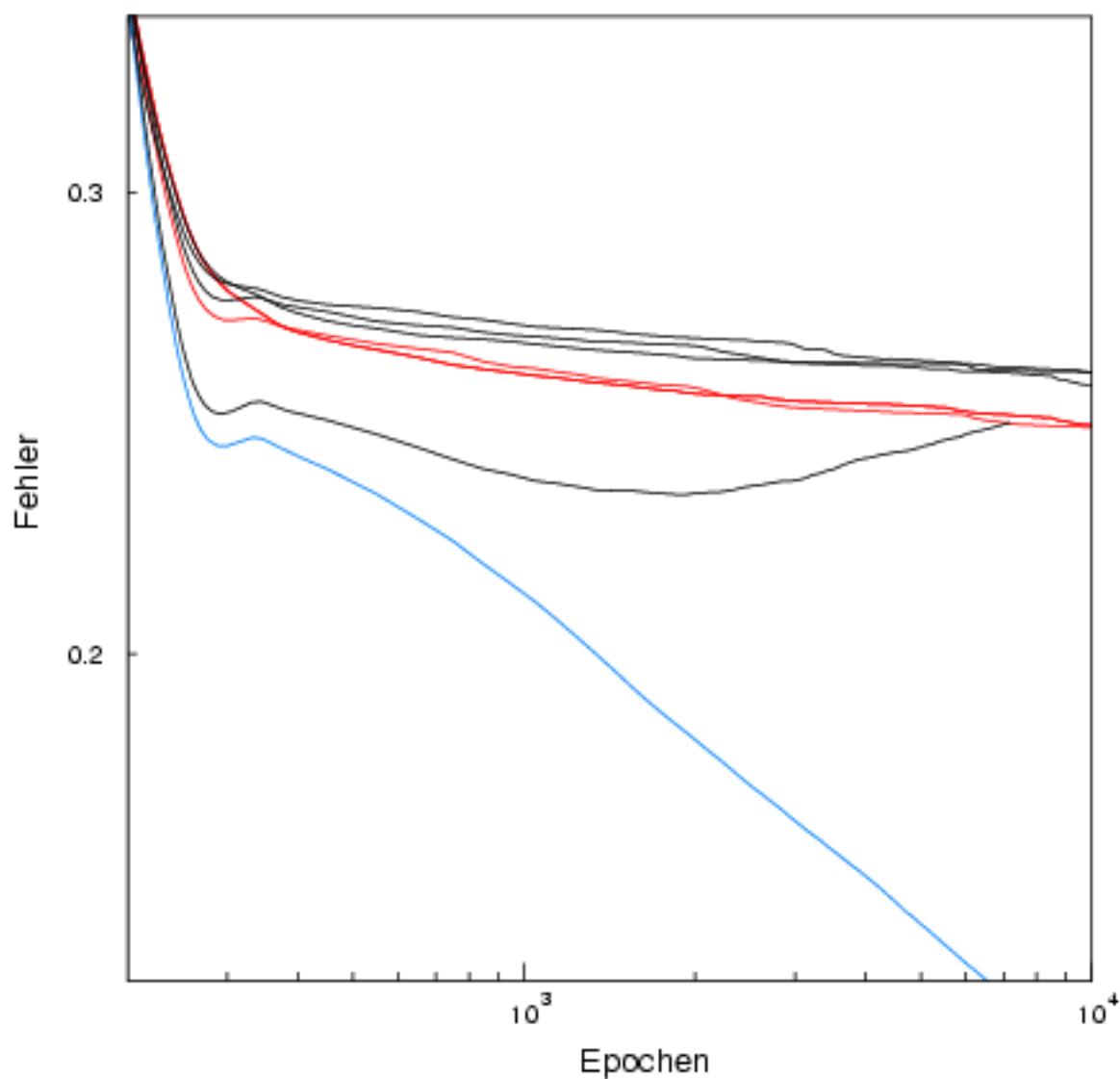


Abbildung 35: Abnahme des Fehlers auf Kontroll- (schwarz) und Lerndatensatz für das 16-16-2 (rot) und das 16-50-50-2 Netz (blau).

Netz	γ -Multiplizität	Segmentmultiplizität 2	Segmentmultiplizität 3	Segmentmultiplizität 4	Summe über alle Segmentmultiplizitäten
16-16-2	1	96,3 (0,3)	76,1 (0,3)	36,5 (0,4)	86,6 (0,2)
	2	54,3 (0,3)	75,2 (0,3)	85,1 (0,5)	77,8 (0,2)
16-50-2	1	96,8 (0,3)	77,8 (0,3)	41,4 (0,5)	87,4 (0,3)
	2	56,6 (0,3)	74,9 (0,3)	83,6 (0,5)	78,0 (0,3)
16-50-50-2	1	96,2 (0,3)	76,5 (0,4)	39,8 (0,5)	86,6 (0,3)
	2	57,8 (0,3)	74,4 (0,3)	90,0 (0,4)	78,6 (0,2)

Tabelle 1: Prozentuale Wahrscheinlichkeiten der richtigen Erkennung von Einfach- und Doppeltrefferereignissen. Statistische Fehler geklammert.

Die Fehler des 16-16-2 Netzes zeigen noch keine Anzeichen für ein Auswendiglernen, da sie - obgleich ihre Absolutwerte deutlich differieren - beide bei fortschreitendem Training weiter fallen. Dagegen zeigt das - bewußt überdimensionierte - 16-50-50-2 Netz sehr schön, daß zu viele freie Parameter des Systems in Verbindung mit einer übermäßig langen Lernphase dazu führen, daß zwar der Fehler auf dem Trainingsdatensatz weiter fällt, der Fehler auf dem Kontrolldatensatz jedoch ein Minimum erreicht und bei weiterem Training sogar wieder ansteigen kann.

Um die Leistungsfähigkeit der Netze zu untersuchen, werden nun die durch sie bestimmten Quoten richtiger Unterscheidung von Einfach- und Doppeltreffern in Tabelle 1 aufgegliedert. Es wird dabei zwischen den drei untersuchten Netzen, zwischen Einfach- und Doppeltreffern und unterschiedlichen Segmentmultiplizitäten unterschieden. Für den Fall einer Segmentmultiplizität von zwei werden durch die Abschätzung 93,3 Prozent aller Eintreffern und 50,8 Prozent aller Doppeltreffern als solche erkannt. Die Netze zeigen alle ein ähnliches Verhalten. Summiert über alle Segmentmultiplizitäten werden Eintreffern zu etwa 87 Prozent als solche erkannt, bei Doppeltreffern ergibt sich eine Quote von 78 Prozent richtiger Zuordnungen.

Legt man ein Verhältnis von Einfach- zu Doppeltreffern zugrunde, wie es aus Abbildung 32 abzulesen ist, so ergibt sich mit der Trefferquote des 16-50-2 Netzes nach Tabelle 1 eine ereignismultiplizitätsabhängige Quote richtiger Unterscheidungen zwischen Einfach- und Doppeltreffern nach Abbildung 36. Ebenfalls eingetragen ist zum Vergleich die Wahrscheinlichkeit, daß nur ein Quant in den Detektor gelangt. Man erkennt, daß es bis zu einer Ereignismultiplizität von etwa zwölf günstiger ist, generell anzunehmen, es liege ein Eintreffer vor. Bei höheren Ereignismultiplizitäten wird es zunehmend wichtiger, ein Verfahren zur Unterscheidung zwischen Einfach- und Doppeltreffern einzusetzen.

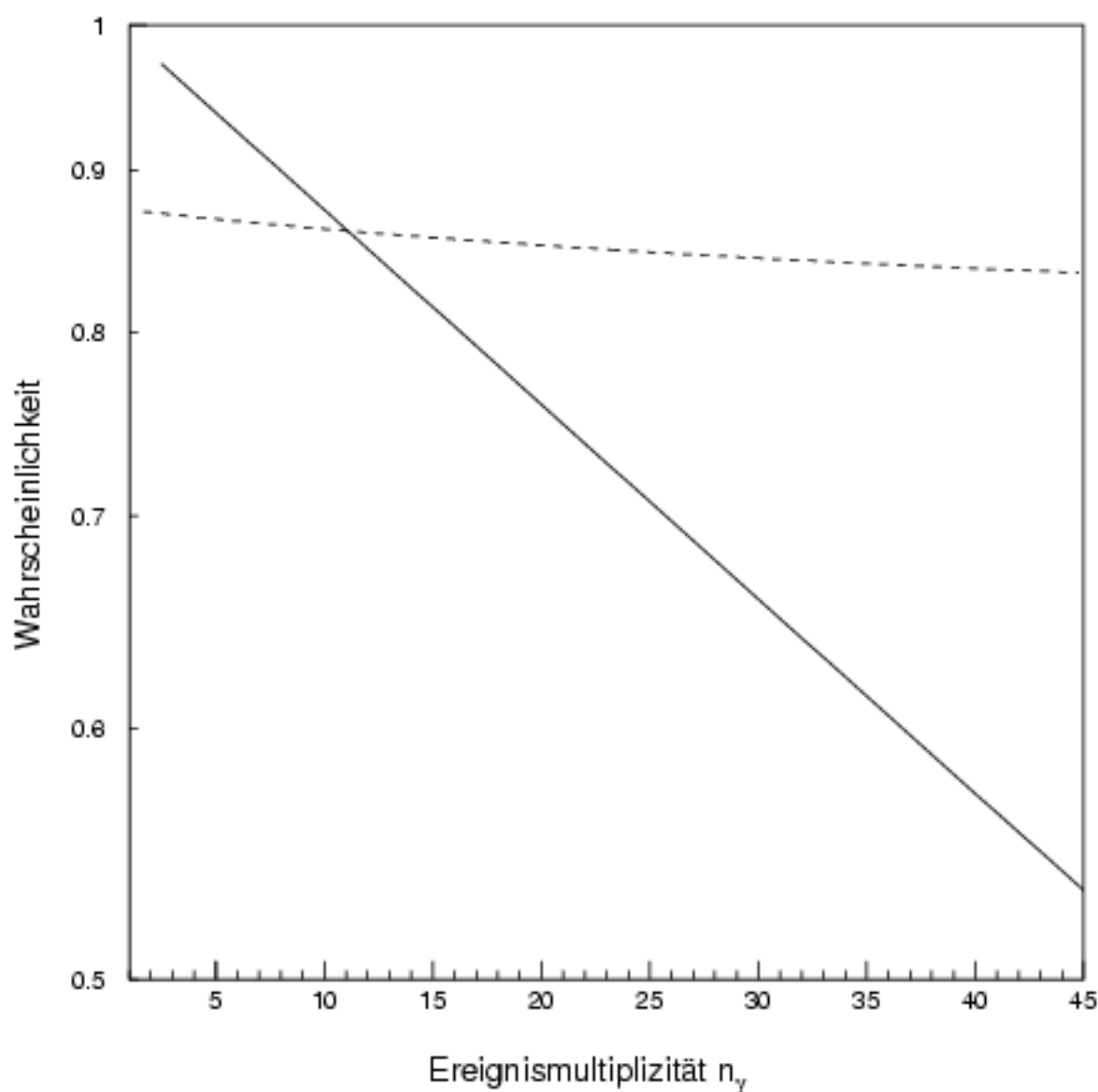


Abbildung 36: Wahrscheinlichkeit der richtigen Unterscheidung zwischen Einfach- und Doppeltreffern in Abhängigkeit von der Ereignismultiplizität bei genereller Annahme eines Eintreffers (durchgezogen) und Anwendung eines MSPs (gestrichelt).

5 Zusammenfassung und Ausblick

Ziel dieser Arbeit war es, die Leistungsfähigkeit künstlicher neuronaler Netze zur Verarbeitung von Detektorsignalen kernphysikalischer Experimente zu untersuchen. Dazu wurden exemplarisch mehrschichtige Perzeptrons (MSPs) betrachtet um sowohl eine unbekannte Funktion zu approximieren als auch um Energiemuster zu klassifizieren. Es konnte gezeigt werden, daß mehrschichtige Perzeptrons erfolgreich zur Verarbeitung aufgezeichneter Meßdaten nach Abschluß eines Experiments eingesetzt werden können.

Im Fall des Parallelplattenzählers PYRAMIDE [Vet94] sind MSPs mit zehn bis fünfzehn versteckten Knoten in der Lage, eine unbekannte Abbildung zur Umrechnung von vier gemessenen Ladungen in zweidimensionale Ortskoordinaten zu approximieren. Die verbleibende Ungenauigkeit in der Ortsbestimmung durch das Netz beträgt ein Prozent, so daß die an den Detektor gestellten Anforderungen erfüllt sind.

Es wurde gezeigt, daß das MSP zwischen im Lerndatensatz enthaltenen Punkten sinnvoll zu interpolieren vermag, daß seine Extrapolationsfähigkeiten jedoch begrenzt sind. Da durch eine Eichung des Zählers Meßpunkte der gesamten aktiven Fläche vorliegen, stellt dieses jedoch keine Einschränkung dar. Die Rechenzeiten zum Training der verwendeten Netze halten sich bei einem Trainingsdatensatz von 10^3 Mustern und einer anfänglichen Beschränkung der Lernphase auf 10^4 Epochen in Grenzen - siehe Anhang A -, so daß während weniger Tage bequem mehrere Architekturen, Trainingsalgorithmen und Parameterkombinationen untersucht sowie gefundene Netze durch den Optimal Brain Surgeon-Algorithmus beschnitten werden konnten.

Die Energiemuster segmentierter Germaniumdetektoren können mit Hilfe von MSPs bezüglich des Segments, in dem die erste Wechselwirkung eines γ -Quants mit dem Detektormaterial stattfand, erfolgreich klassifiziert werden.

Für zweifach und vierfach segmentierte Detektoren entspricht die Ausgabe der Netze einer Bayesschen Wahrscheinlichkeit. Im Falle einer Segmentmultiplizität von zwei erreicht die Vorhersagekraft der MSPs im Rahmen der statistischen Fehler die physikalisch bedingte Grenze. Im Falle einer Segmentmultiplizität von drei ist diese Grenze noch nicht für alle γ -Energien erreicht. Generell ist jedoch eine Überlegenheit der MSPs gegenüber herkömmlichen Abschätzungen zur Klassifikation der Energiemuster festzustellen. Ab einer Segmentmultiplizität von drei könnte die durch das Netz erzielte Vorhersagekraft nur unter unverhältnismäßig hohem Aufwand durch konventionelle Verfahren erreicht werden. Wie im Falle des Parallelplattenzählers sind die für die Lernphase benötigten Rechenzeiten gering, Details sind in Anhang A zu finden. Im Fall des zweifach segmentierten Detektors erweisen sich vier bis zehn versteckte Knoten als ausreichend, im Fall des vierfach segmentierten Detektors zwanzig versteckte Knoten.

Im Falle des SEGMENTED CLOVER-Detektors [Ger92] ist bei einer Segmentmultiplizität von zwei ein zu den geringer segmentierten Detektoren analoges Verhalten zu beobachten, so daß auch hier die physikalisch mögliche Grenze der Vorhersagewahrscheinlichkeit erreicht wurde. Bei einer Segmentmultiplizität von drei ist die Leistung des Netzes noch nicht zufriedenstellend, jedoch wurde beim vierfach segmentierten Detektor gezeigt, daß durch stärkere Wichtung dieser Ereignisse im Lerndatensatz auch für diese Segmentmultiplizität gute Ergebnisse erreicht werden. Die Rechenzeiten zum Training der untersuchten Strukturen sind mit mehreren Tagen allerdings zu lang, um in ausreichendem

Maße unterschiedliche Architekturen oder Parameterkombinationen beurteilen zu können. Es bleibt zu untersuchen, inwieweit die vorliegenden Symmetrien des Detektors ausgenutzt werden können, um durch eine Reduktion der Eingabegrößen im Rahmen einer geeigneten Vorverarbeitung der Segmentenergien eine Verkleinerung der Netze und damit eine Beschleunigung des Trainingsvorgangs erreicht werden kann. Auch könnten Sigma-Pi Knoten betrachtet werden um mit insgesamt weniger versteckten Knoten auszukommen. Da diese Knoten höherer Ordnung jedoch über deutlich mehr Parameter als einfache Sigma-Knoten verfügen, bleibt abzuwarten, inwieweit sich bei diesem Ansatz eine Beschleunigung der Lernphase erreichen läßt.

Da im Falle einer Segmentmultiplizität von zwei die Verteilungen der zu klassifizierenden Muster sehr einfach strukturiert sind, kann mit konventionellen Ansätzen eine ähnlich hohe Vorhersagekraft wie durch die untersuchten Netze erreicht werden. Der Vorteil des künstlichen neuronalen Netzes tritt ab einer Segmentmultiplizität von drei deutlich hervor, da in dieser Situation eine manuelle Analyse der Energieverteilung über die Segmente - insbesondere im Falle des SEGMENTED CLOVER-Detektors - nicht mehr praktikabel ist. Die Einengung des Ortes der ersten Wechselwirkung zum Zwecke einer Dopplerkorrektur erfolgt in zu konventionellen Methoden vergleichbarem Maße.

Die Unterscheidung zwischen Einfach- und Doppeltreffern bei Ansprechen mehrerer Segmente des SEGMENTED CLOVER-Detektors ist durch mehrschichtige Perzeptrons bei γ -Energien bis 2.5 MeV mit zwischen 70 und 85 Prozent liegender Wahrscheinlichkeit möglich. Unter Vernachlässigung der Nachweiswahrscheinlichkeit der Detektoren ergibt sich dabei für die gewählte Geometrie ab einer Ereignismultiplizität von zwölf ein Vorteil gegenüber der generellen Annahme eines Einfachtreffers. Da auch große Netze mit zwei versteckten Schichten zu je 50 Knoten keine wesentlich bessere Leistung als ein relativ kleines Netz mit einer versteckten Schicht zu 16 Knoten zeigten, bleiben die zum Training erforderlichen Rechenzeiten in einem akzeptablen Rahmen. Für den Fall einer Segmentmultiplizität von zwei wurde eine einfache Abschätzung zur Unterscheidung zwischen Einfach- und Doppeltreffern erstellt, welche die Leistung des MSPs erreicht. Zukünftig ist nach geeigneten Abschätzungen für höhere Segmentmultiplizitäten zu suchen, um auch in diesen Fällen die Leistung der Netze bewerten zu können. Darüberhinaus sind Ereignismultiplizitäten größer zwei zu betrachten.

Mit den heute vorliegenden Modifikationen des Backpropagation-Algorithmus stehen eine Vielzahl interessanter Verfahren zur Minimierung der Fehlerfunktion eines mehrschichtigen Perzeptrons zur Verfügung, deren Implementation keine Schwierigkeiten bereitet. Mit Hilfe heutiger Workstations können nach den in dieser Arbeit gewonnenen Erfahrungen MSPs mit bis zu etwa 10^8 Synapsen auf Lerndatensätze von 10^4 Mustern bei bis zu 10^4 Epochen dauernden Lernphasen zu akzeptablen Rechenzeiten von bis zu einem Tag trainiert werden, so daß innerhalb von zwei bis drei Wochen verschiedene Architekturen und Fehlerminimierungsverfahren untersucht und bestehende Netze durch Algorithmen beschnitten werden können. Die Rechenzeiten bei Netzen ab 2×10^8 bis 3×10^8 Synapsen und den dann sinnvollerweise zu verwendenden Lerndatensätzen mit mindestens 2×10^4 bis 3×10^4 Mustern betragen jedoch bei Lernphasen von 10^4 Epochen leicht eine Woche, so daß die erforderlichen Untersuchungen zur Architektur oder Vergleiche verschiedener Minimierungsverfahren nicht mehr bzw. nur bei gleichzeitiger Nutzung einer Vielzahl von Workstations in befriedigendem Maße durchgeführt werden können.

A Software

Zur Untersuchung mehrschichtiger Perzeptrons wurde im Rahmen dieser Arbeit ein Simulationsprogramm entwickelt. Mit seiner Hilfe können MSPs beliebiger Topologie, also einer beliebigen Zahl versteckter Schichten variabler Breite, konfiguriert werden. Beliebige Datensätze aus Eingabe- und dazu zu assoziierenden Sollausgabevektoren an das jeweilige MSP angepasster Dimensionen können zum Training verwendet werden; sie lassen sich in Lern- und Kontrolldatensatz aufspalten. Es kommen generell Sigma-Knoten zum Einsatz, als Ausgabefunktion ist der Tangens Hyperbolicus vorgesehen. Er kann durch Stauchung und Verschiebung - beispielsweise zur Fermifunktion - umgeformt werden. In der Ausgabeschicht kann auch die Identität verwendet werden, um den Wertebereich der Netzausgabe nicht prinzipiell einzuschränken. Die Berechnung der zum Training erforderlichen Gradienten erfolgt mittels Backpropagation, die Fehlerminimierung selbst kann durch einen stochastischen Gradientenabstieg im Online-Verfahren, im Rahmen des Batch-Modus oder durch den RPROP-Algorithmus erfolgen. Die Entwicklung von Fehler auf Lern- und Kontrolldatensatz, die Entwicklung aller Synapsen im Netz sowie die Entwicklung der im Rahmen des RPROP-Algorithmus verwendeten Parameter während der Lernphase kann in regelmäßigen zeitlichen Abständen dokumentiert werden. Im Batch-Modus kann die Lernrate je nach Fehlerentwicklung automatisch verändert werden. Ein Impulsterm kann verwendet werden, Synapsen können während der Lernphase automatisch zerfallen. Die Größe der zufällig gewählten Startsynapsen kann beliebig vorgegeben werden, ebenso die Zahl der jeweils durchzuführenden Fehlerminimierungen. Die Steilheit des Tangens Hyperbolicus, die Größe der Lernrate, der Impulsterm, die Parameter zur Anpassung der variablen Lernrate im Batch-Modus sowie die Parameter zur Steuerung des Zerfalls von Synapsen können systematisch variiert werden, um deren Einfluß auf die Lernphase untersuchen zu können. Trainierte Netze können mit Hilfe des OBS-Algorithmus beschnitten werden, die Bestimmung der erforderlichen Gradienten erfolgt wiederum mittels Backpropagation, die der inversen Hessematrix durch das in [Has93] vorgeschlagene Iterationsverfahren. Vorläufige Routinen zur Simulation der Cascade-Correlation Learning Architecture [Fah91] sind vorhanden.

Das Programm umfasst mittlerweile etwa 12000 Zeilen C-Quelltext und läuft zur Zeit unter OpenVMS 6.1 auf dem AXP-Cluster der GSI. Die unterschiedlichen, durch den DEC-C Compiler vorgegebenen Möglichkeiten der Optimierung wurden untersucht, jedoch keine signifikanten Unterschiede bezüglich der jeweils benötigten Rechenzeiten festgestellt. Tabellarisch seien auszugsweise Rechenzeiten zum Training der Netze für jeweils 10^5 Epochen - im Falle des stochastischen Gradientenabstiegs Zyklen - genannt. Dazu ist anzumerken, daß es nicht eigentliches Ziel dieser Arbeit war, das verwendete Programm zu optimieren. Sofern der zeitliche Rahmen es zuließ, wurde das Programm mit Hilfe des Performance Coverage Analyzers von DEC auf Flaschenhalse untersucht und diese beseitigt. Es ist jedoch nicht auszuschließen, daß weitere Verminderungen der Rechenzeiten denkbar wären. Es wäre zu prüfen, ob durch Umschreiben der zeitkritischen Routinen - Propagation eines Musters durch das MSP bzw. rückwärtige Propagation zwecks Gradientenbestimmung - in Assembler ein weiterer Geschwindigkeitsvorteil möglich wäre. Ebenfalls zu untersuchen bleiben die Rechenzeiten auf dem AIX Cluster der GSI oder - nach Umstellen wesentlicher Routinen - eine mögliche Nutzung des Vektorrechners GOETHE des Rechenzentrums der Johann Wolfgang Goethe-Universität Frankfurt.

Lern- und Kontrolldatensätze zur Eichung der PYRAMIDE stammen aus einer Eichmessung mit Hilfe einer ^{252}Cf -Spaltquelle; die zur Untersuchung der segmentierten Germaniumdetektoren aus Simulationen durch GEANT 3.21. Die dortige Implementation der dabei verwendeten Wechselwirkungen wie Photo-, Compton- und Paareffekt sind in dem Rahmen, in dem sie zur Erzeugung der in dieser

Netz	Zahl der Muster	Verfahren	CPU-Zeit [min]	Rechner
2-4-2	10^4	RPROP	3.5	DEC 3000 300
2-4-4-2	10^4	RPROP	6.5	DEC 3000 300
2-5-2	10^4	RPROP	4.1	DEC 3000 400
2-7-2	10^4	RPROP	6.1	DEC 3000 400
2-10-2	10^4	RPROP	7.2	DEC 3000 300
2-20-2	10^4	RPROP	12.7	DEC 3000 300
4-4-4	2×10^4	RPROP	8.6	DEC 3000 300
4-4-4-4	2×10^4	RPROP	17.9	DEC 3000 400
4-10-4	2×10^4	RPROP	20.8	DEC 3000 300
4-20-4	2×10^4	RPROP	42.3	DEC 3000 300
4-10-2	980	RPROP	0.9	DEC 3000 300LX
4-10-10-2	980	Batch	3.7	DEC 3000 300
	980	Online	1.2	DEC 2100 A500MP
	980	RPROP	2.3	DEC 3000 300LX
16-16-2	3×10^4	RPROP	58.6	DEC 3000 600
	3×10^4	RPROP	68.8	DEC 3000 300
16-50-2	3×10^4	RPROP	176.5	DEC 3000 600
16-50-16	3×10^4	RPROP	380.7	DEC 2100 A500MP
16-80-16	3×10^4	RPROP	772.4	DEC 3000 300
	3×10^4	RPROP	634.7	DEC 2100 A500MP

Tabelle 2: Beispiele für eine Lernphase von 10^5 Epochen benötigter Rechenzeiten.

Arbeit verwendeten Daten zur Anwendung kamen, als zuverlässig anzusehen. Vielfach durchgeführte Vergleiche des Ansprechverhaltens durch GEANT simulierter und realer Germaniumdetektoren zeigen auf einige Prozent genaue Übereinstimmungen.

B Bond-Diagramme

Zur Visualisierung der durch den Fehlerminimierungsprozeß gefundenen Synapsen folgen nun abschließend exemplarisch vier Bond-Diagramme [Hay94] trainierter Netze. Die formalen Knoten der MSPs sind durch leere Kreise gekennzeichnet, Schwellwerte durch Synapsen zu in den jeweils vorausgehenden Schichten künstlich eingeführten Knoten mit einer konstanten Aktivität von -1, welche durch schraffierte Kreise dargestellt sind. Der Betrag der Synapsen wird durch die Größe zweier Dreiecke dargestellt, welche den post- mit dem präsynaptischen Knoten verbinden. Positive Synapsen werden durch ausgefüllte, negative Synapsen durch leere Dreiecke charakterisiert. Die Eingabeschicht der MSPs ist im Bild jeweils oben, die Ausgabeschicht jeweils unten zu sehen.

Zum Zwecke der Darstellung werden nicht die Werte der Synapsen selbst, sondern deren Quadratwurzeln dargestellt, so daß auch Synapsen kleinerer Beträge noch gut zu erkennen sind. Da oftmals einige wenige Synapsen Werte annahmen, welche um bis zu drei Größenordnungen über denen der restlichen Synapsen liegen, wurden - erneut zum Zwecke der Darstellung - solche extremen Werte auf eine jeweils angegebene Obergrenze reduziert.

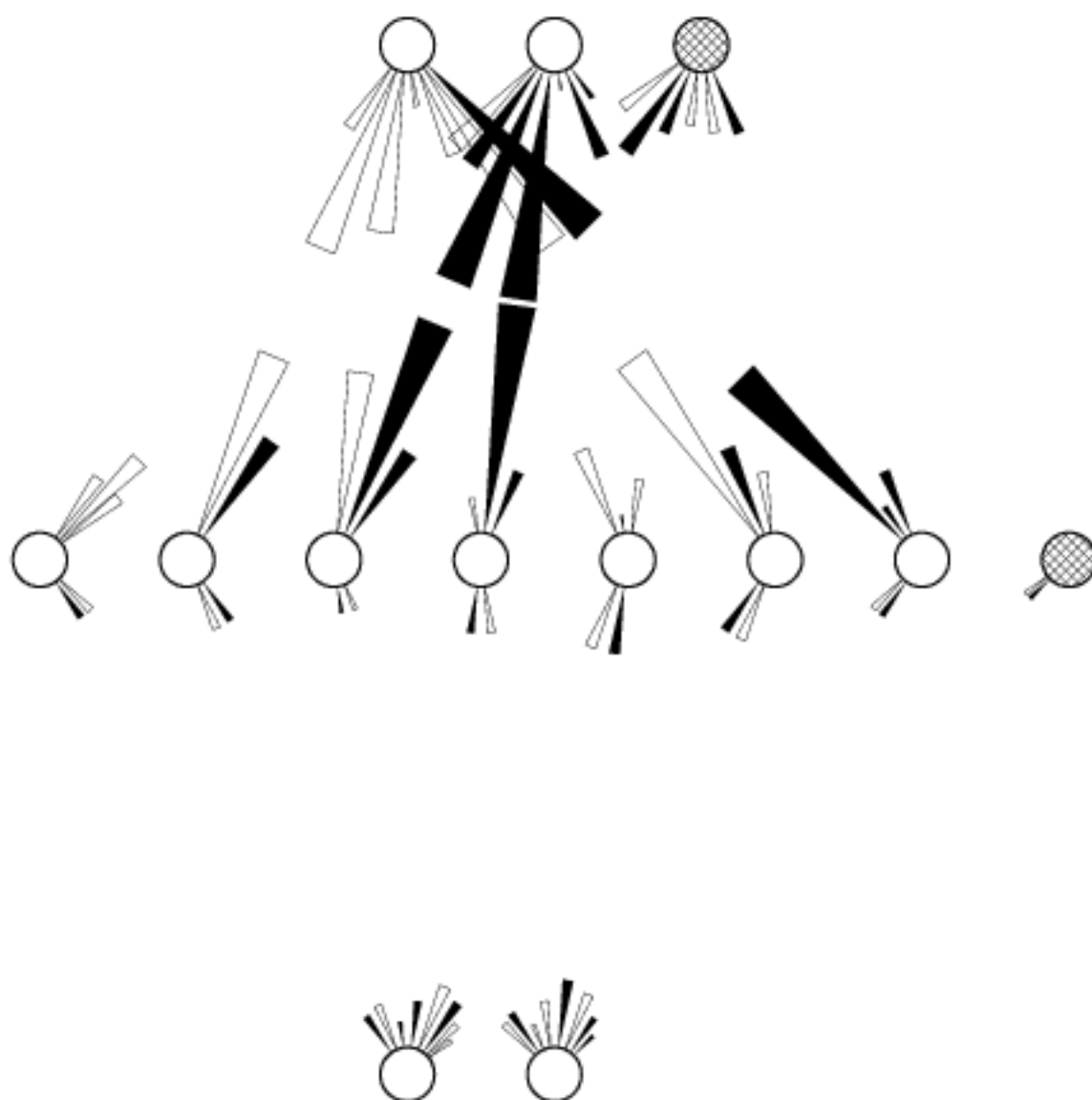


Abbildung 37: Bond-Diagramm des in Kapitel 4.1.2 verwendeten 2-7-2 Netztes. Zwei Synapsen mit Werten von 72 und 111 wurden auf 50 reduziert.

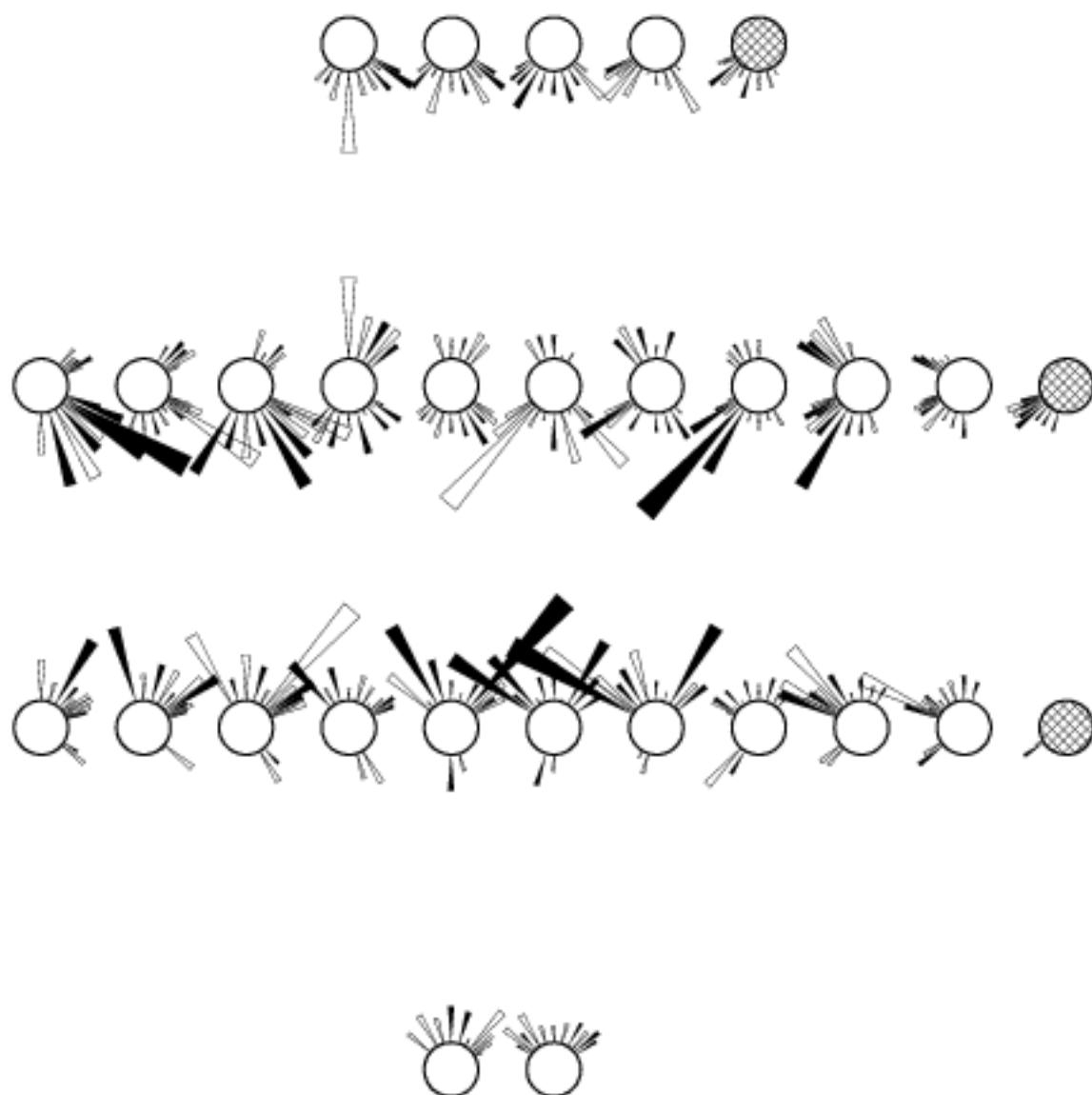


Abbildung 38: Bond-Diagramm des in Kapitel 3.3 diskutierten 4-10-10-2 Netzes. Die maximale Synapse beträgt 22.

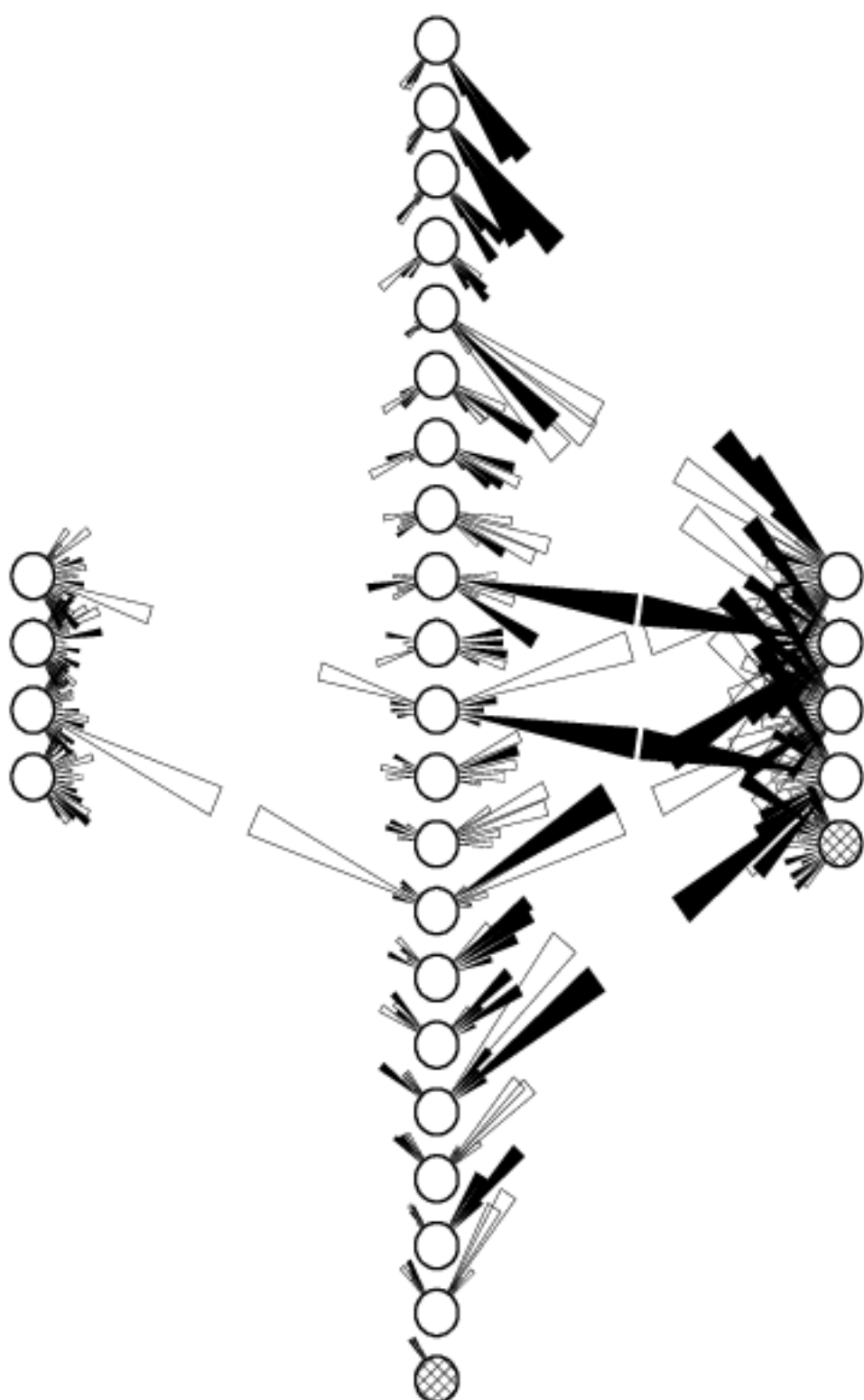


Abbildung 39: Bond-Diagramm des in Kapitel 4.1.3 diskutierten 4-20-4-Netzes. Sechs Synapsen mit Beträgen zwischen 10^3 und 2.8×10^4 wurden auf 90 reduziert.

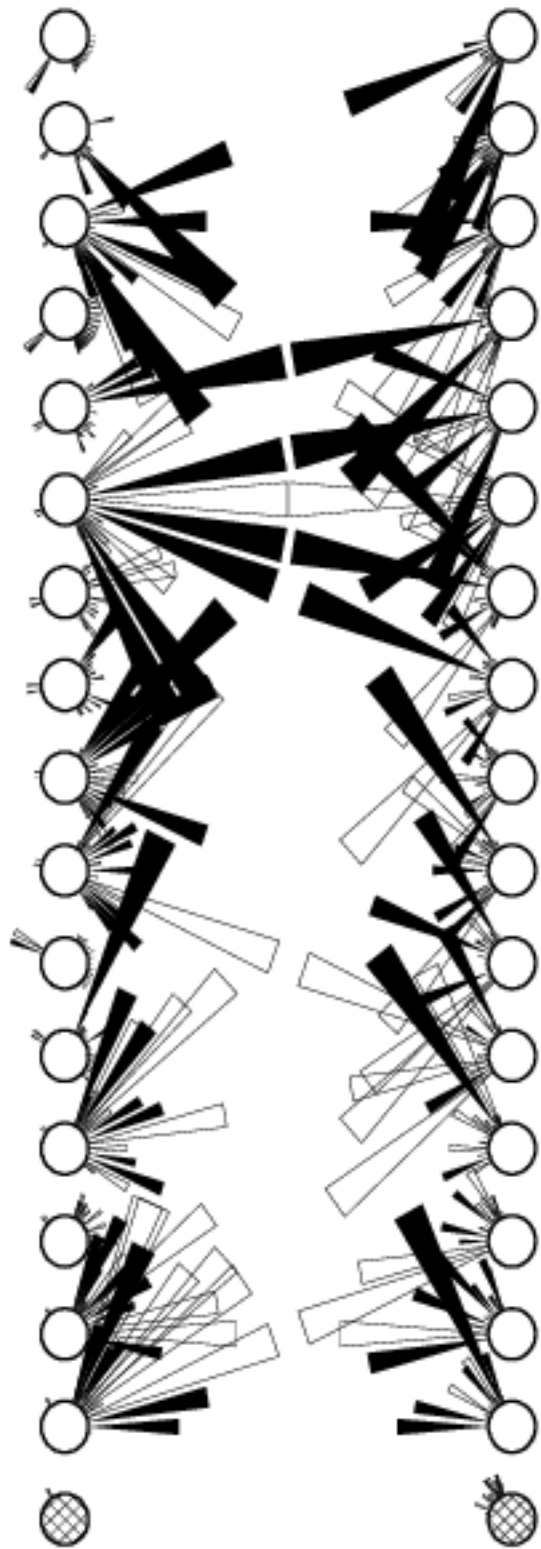


Abbildung 40: Bond-Diagramm des in Kapitel 4.2 diskutierten 16-16-2-Netz. Zahn-Synapsen mit Beträgen 500 und 10^5 wurden auf 500 reduziert.

Literatur

- [Ami92] Daniel J. Amit. *Modeling Brain Function*. Cambridge University Press, Cambridge, 1992.
- [Ang94] Peter J. Angeline, Gregory M. Saunders and Jordan B. Pollack. An evolutionary algorithm that constructs recurrent neural networks. *IEEE Transactions on Neural Networks*, 5 No. 1:54–65, January 1994.
- [Bar92] Etienne Barnard. Optimization for training neural nets. *IEEE Transactions on Neural Networks*, 3 No. 2:232–240, March 1992.
- [Bau89] Eric B. Baum. What size net gives valid generalization? *Neural Computation*, January 1989.
- [Bra91] Rüdiger Brause. *Neuronale Netze*. B. G. Teubner, Stuttgart, 1991.
- [Bro89] I. N. Bronstein and K. A. Semendjajew. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Thun und Frankfurt am Main, 1989.
- [Duc90] Gunther Duck and Tobias Scheuer. Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing. *Journal of Computational Physics*, 90:161–175, 1990.
- [Fah88] Scott E. Fahlman. An empirical study of learning speed in back-propagation networks. Technical Report CMU-CS-88-162, School of Computer Science, Carnegie Mellon University, Pittsburgh, 1988.
- [Fah91] Scott E. Fahlman and Christian Lebiere. The cascade-correlation learning architecture. Technical Report CMU-CS-90-100, School of Computer Science, Carnegie Mellon University, Pittsburgh, 1991.
- [Fie94] E. Fiesler. Neural network classification and formalization. *Computer Standards and Interfaces*, 16, 1994.
- [Fl⁹³] Thomas Flügel. *Untersuchungen zur Entwicklung eines aktiven Kollimators für Ge-Detektoren*. Diplomarbeit, Institut für Kernphysik der Johann Wolfgang Goethe-Universität Frankfurt am Main, August 1993.
- [Fun89] Ken-Ichi Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2:183–192, 1989.
- [GEA93] CERN, Geneva, Switzerland. *GEANT Detector Description and Simulation Tool*, 1993.
- [Ger92] Jürgen Gerl and R. M. Lieder, editors. *EUROBALL III European γ -ray facility*. Gesellschaft für Schwerionenforschung GSI, Darmstadt, 1992.
- [Hal92] Zach W. Hall. *An Introduction to Molecular Neurobiology*. Sinauer Associates, Inc., Sunderland, Massachusetts, 1992.
- [Has93] Babak Hassibi and David G. Stork. Second order derivatives for network pruning: Optimal brain surgeon. *Advances in Neural Information Processing Systems*, 5:164–171, 1993.
- [Hay94] Simon Haykin. *Neural Networks - A Comprehensive Foundation*. Macmillan College Publishing Company, Inc., New York, 1994.

- [Her91] John Hertz, Anders Krogh and Richard G. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, Reading, Massachusetts, 1991.
- [Jac88] Robert A. Jacobs. Increased rates of convergence through learning rate adaptation. *Neural Networks*, 1:295–307, 1988.
- [LeC89] Y. LeCun, J. S. Denker and S. A. Solla. Optimal brain damage. *Advances in Neural Information Processing Systems*, 2:598–605, 1989.
- [Lee93] Youngjik Lee, Sang-Hoon Oh and Myung Won Kim. An analysis of premature saturation in back propagation learning. *Neural Networks*, 6:719–728, 1993.
- [McC86] James L. McClelland and David E. Rummelhart. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*. MIT Press, Cambridge, Massachusetts, 1986.
- [Min69] M. L. Minsky and S. A. Papert. *Perceptrons*. MIT Press, Cambridge, 1969.
- [Mö94] Jürgen Möck. *Einsatz neuronaler Netze als intelligente Trigger im HI-Experiment*. Diplomarbeit, Fakultät für Physik der Technischen Universität München, März 1994.
- [Mü91] B. Müller and J. Reinhardt. *Neural Networks - An Introduction*. Springer-Verlag, Berlin, Heidelberg, New York, 1991.
- [Nic92] John G. Nichols, A. Robert Martin and Bruce G. Wallace. *From Neuron to Brain - A Cellular and Molecular Approach to the Function of the Nervous System*. Sinauer Associates, Inc., Sunderland, Massachusetts, 1992.
- [Per92] Pierre Peretto. *An Introduction to the Modeling of Neural Networks*. Cambridge University Press, Cambridge, 1992.
- [Pet94] Ingo Peter. *Analyse von digitalisierten Germanium-Detektorsignalen*. Diplomarbeit, Institut für Kernphysik der Johann Wolfgang Goethe-Universität Frankfurt am Main, Oktober 1994.
- [Ree93] Russel Reed. Pruning algorithms - a survey. *IEEE Transactions on Neural Networks*, 4 No. 5:740–747, September 1993.
- [Ric91] M. Richard and R. Lippmann. Neural network classifiers estimate bayesian a posteriori probabilities. *Neural Computation*, 3:461–483, 1991.
- [Rie92] Martin Riedmiller. *Schnelle adaptive Lernverfahren für mehrschichtige Feedforward-Netzwerke - Vergleich und Weiterentwicklung*. Diplomarbeit, Fakultät für Informatik, Universität Karlsruhe, 1992.
- [Rie94] Martin Riedmiller. Advanced supervised learning in multi-layer perceptrons - from back-propagation to adaptive learning algorithms. *Computer Standards and Interfaces*, Special Issue on Neural Networks 5, 1994.
- [Rum86] David E. Rummelhart and James L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. MIT Press, Cambridge, Massachusetts, 1986.
- [SNN94] Universität Stuttgart. *SNN Stuttgart Neural Network Simulator*, 1994.

- [Sun90] Junping Sun, William I. Grosky and Mohamad H. Hassoun. A fast algorithm for finding global minima of error functions in layered neural networks. *International Joint Conference on Neural Networks*, 1:715–720, 1990.
- [Vet94] K. Vetter, Th.W. Elze, J. Gerl, M. Kaspar and T. Kröll. A position sensitive gas counter with resistive layer cathode decoded by a neural network algorithm. *Nuclear Instruments and Methods in Physics Research, A* 334:607–610, 1994.
- [Whi90] D. Whitley, T. Starkweather and C. Bogart. Genetic algorithms and neural networks: optimizing connections and connectivity. *Parallel Computing*, 14:347–361, 1990.
- [Wid63] B. Widrow. Adaline and madaline. *IEEE 1st International Joint Conference on Neural Networks*, 1:143–158, 1963.

Danksagung

Dem geschäftsführenden Direktor des Instituts für Kernphysik der Johann Wolfgang Goethe-Universität, Herrn Prof. Dr. K. Bethge, bin ich für die Aufnahme in das Institut und die damit gegebene Voraussetzung zur Durchführung dieser Arbeit zu Dank verpflichtet.

Ich danke Herrn Prof. Dr. Th.W. Eke für die interessante Themenstellung, für die Möglichkeit, diese Arbeit im Rahmen der Zusammenarbeit zwischen dem Institut für Kernphysik und der Gesellschaft für Schwerionenforschung durchführen zu können, sowie für seine Betreuung und stete Hilfsbereitschaft.

Bei Herrn Dr. J. Gerl bedanke ich mich für die Aufnahme in die EUROBALL-Gruppe der GSI in Darmstadt und für seine dortige Betreuung der Arbeit.

Dank für vielfältige Diskussionen und eine angenehme Arbeitsatmosphäre gilt den Mitarbeitern der EUROBALL-Gruppe der GSI Dr. Th. Happ, Th. Kröll, I. Peter, H. Schaffner, S. Schremmer, R. Schubert sowie insbesondere Herrn K. Vetter für wertvolle Anregungen und Vorschläge, welche maßgeblich zum Gelingen dieser Arbeit beigetragen haben.

Weiterhin danke ich den Herren Dr. W. Schiebel, Dr. H. Essel und Dr. P. Koczon für ihre Hilfe bei den alltäglichen Soft- und Hardwareproblemen des VAX- und AXP-Clusters der GSI, sowie all den Arbeitsgruppen, die mir die Nutzung ihrer Workstations zur Durchführung der - teilweise sehr langen - Rechnungen gestatteten. Dank an K. Cowan, S. Hobbs und S. Lionel von DEC für prompte Hilfe bei Problemen mit VAX-C, DEC-C und Fortran Compilern.

Dank an S. Giani von CERN für die umgehende Beseitigung eines Fehlers in GEANT 3.21.

Herrn Dr. J. Cub sei für zahlreiche Anregungen zu dem zur Simulation der mehrschichtigen Perzeptons erstellten Programm gedankt.

Dank den Damen A. Dietz, I. Malwitz und R. Montay der Bibliothek der GSI für die Beschaffung zahlreicher Veröffentlichungen aus Fachzeitschriften über neuronale Netze, welche nicht von der GSI bezogen werden und daher aus entfernten Bibliotheken angefordert werden mußten. Für die Übersendung von Veröffentlichungen danke ich den Herrn D. McCormack, J. Ruhland und D. Stork.

Schließlich danke ich meinen Eltern, daß sie mir das Studium ermöglicht haben.