

~~Interner Bericht~~  
~~DESY F35D-94-11~~  
Oktober 1994

# Identifikation von Elektronen im ZEUS-Hadron-Elektron-Separator mit neuronalen Netzwerken

von

J. O. Carstens

Eigentum der	<b>DESY</b>	Bibliothek
Property of		
Zugang:	20. OKT. 1994	
Accession:		
Leih:		
Loan period:	1	days

**DESY behält sich alle Rechte für den Fall der Schutzrechtserteilung und für die wirtschaftliche Verwertung der in diesem Bericht enthaltenen Informationen vor.**

**DESY reserves all rights for commercial use of information included in this report, especially in case of filing application for or grant of patents.**

**"Die Verantwortung für den Inhalt dieses  
Internen Berichtes liegt ausschließlich beim Verfasser"**

# Identifikation von Elektronen im ZEUS-Hadron-Elektron-Separator mit neuronalen Netzwerken

Experimentelle Diplomarbeit  
am Fachbereich Physik  
der Universität Hamburg

Jan Ole Carstens ✓

Oktober 1994



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>ZEUS</b>	<b>3</b>
2.1	Der ZEUS-Detektor . . . . .	3
2.2	Elektronenidentifikation mit dem ZEUS-Detektor . . . . .	5
2.2.1	Einleitung . . . . .	5
2.2.2	Die Spurkammer . . . . .	5
2.2.3	Das Kalorimeter . . . . .	5
2.2.4	Der Hadron-Elektron-Separator . . . . .	7
2.2.5	Korrelation der Schauerentwicklung im Kalorimeter und im HES . . . . .	7
<b>3</b>	<b>Neuronale Netzwerke</b>	<b>11</b>
3.1	Einleitung . . . . .	11
3.2	Begriffsbestimmung . . . . .	11
3.3	Theorie der überwachten <i>feed forward</i> -Netze . . . . .	12
3.3.1	Nomenklatur und mathematische Beschreibung . . . . .	12
3.3.2	Trainingsvorgang . . . . .	14
3.3.3	Geometrische Deutung . . . . .	15
3.3.4	Zur Bedeutung der „Temperatur“ . . . . .	17
3.3.5	Problemquellen . . . . .	17
3.4	JETNET 1.0 . . . . .	18
3.5	Niedrigdimensionale Anwendungen . . . . .	19
3.5.1	Das minimale Netz . . . . .	19
3.5.2	Zweidimensionale Anwendungen . . . . .	27
3.6	Problem der redundanten Schnitte . . . . .	29
3.6.1	Einleitung . . . . .	29
3.6.2	<i>Pruning</i> . . . . .	29
3.6.3	Die Methode des künstlichen Rauschens . . . . .	31
3.6.4	Die Methode der kritischen Datenpunkte . . . . .	32
3.6.5	Zusammenfassung . . . . .	33
<b>4</b>	<b>Identifikation niederenergetischer Elektronen</b>	<b>37</b>
4.1	Einleitung . . . . .	37
4.2	Der Trainings- und Testdatensatz . . . . .	37
4.2.1	Datenquellen . . . . .	37
4.2.2	Elektronen aus Photonkonversion . . . . .	38

4.2.3	Datenaufbereitung im Überblick . . . . .	39
4.2.4	Gruppenbildung: Isles und Islands . . . . .	40
4.2.5	Energie- und Positionsbestimmung . . . . .	40
4.2.6	Isle-Island- und Spur-Island-Zuordnung . . . . .	40
4.2.7	Identifikation der Konversionselektronen . . . . .	41
4.2.8	Spur-Isle-Zuordnung . . . . .	42
4.2.9	Spurselektion . . . . .	44
4.3	Die Eingabevariablen . . . . .	46
4.4	Trainingsergebnisse . . . . .	50
4.4.1	Einleitung . . . . .	50
4.4.2	Definition der Effizienz, Diskriminanzkraft und Signalverstärkung . . . . .	50
4.4.3	Der einfache HES-Schnitt . . . . .	51
4.4.4	Verbesserung durch die Variable HAC-Turm . . . . .	52
4.4.5	Alle HES-Variablen . . . . .	54
4.4.6	Alle Kalorimetervariablen . . . . .	54
4.4.7	Alle Eingabevariablen . . . . .	54
4.4.8	hesnn . . . . .	55
4.4.9	Höhere Impulse . . . . .	60
4.4.10	Berücksichtigung der Spur-Isle-Zuordnung . . . . .	62
4.4.11	Zusammenfassung . . . . .	63
4.5	Korrekturen . . . . .	63
4.5.1	Einleitung . . . . .	63
4.5.2	Der Konversionsfinder . . . . .	64
4.6	Zusammenfassung . . . . .	65
<b>5</b>	<b>Zusammenfassung</b>	<b>67</b>
<b>A</b>	<b>Anhang zum Kapitel „Neuronale Netzwerke“</b>	<b>69</b>
A.1	Der <i>back propagation</i> -Algorithmus . . . . .	69
A.2	Detaillierte Darstellung eines Trainingsverlaufs . . . . .	72
<b>B</b>	<b>Anhang zum Kapitel „Niederenergetische Elektronen“</b>	<b>77</b>
B.1	Der Isle- und Island-Algorithmus . . . . .	77
B.2	Signifikanz unterschiedlicher Diskriminanz zweier Elektronenfinder . . . . .	80
B.3	Entfaltung von Effizienz und Diskriminanz . . . . .	82

# Kapitel 1

## Einleitung

Der Speicherring HERA, seit 1992 in Betrieb, dient der Grundlagenforschung. Er bringt  $26,7 \text{ GeV}$ -Elektronen und  $820 \text{ GeV}$ -Protonen zur Kollision. Die beiden Universaldetektoren H1 und ZEUS beobachten diese Kollisionsereignisse. Sie erlauben die Untersuchung der inneren Struktur des Protons in bisher unerreichbaren kinematischen Bereichen, insbesondere bei kleinen  $x$ , d. h. bei Kollisionen des Elektrons mit Protonbestandteilen, die nur einen kleinen Teil (bis hinunter zu  $10^{-4}$ ) des Protonimpulses tragen. Daneben sind Präzisionstests der Vorhersagen der Quantenchromodynamik und die Produktion zahlreicher Teilchen mit schweren *quarks* möglich.

Ein wichtiger Bestandteil aller Untersuchungen ist die Identifikation von Elektronen unter den bei der Kollision erzeugten Teilchen (Reaktionsprodukten). In der tiefunelastischen Streuung dient das gestreute Elektron zur Berechnung der kinematischen Variablen  $x$  und  $Q^2$  (Quadrat des Impulsübertrages) der Streuung. Weitere Elektronen dienen als Nachweis von *charm* und *bottom quarks*.

Das ZEUS-Uran-Kalorimeter erlaubt die Identifikation von Elektronen und die Bestimmung ihrer Energie. Es existieren bereits eine Reihe von Algorithmen, die auf der Basis der Kalorimeterdaten die Elektronenerkennung durchführen. Nicht berücksichtigt wurde dabei bislang der in das Kalorimeter eingebaute Hadron-Elektron-Separator (HES), ein Siliziumdetektor, der im wesentlichen ermittelt, wie viele Teilchen jedes seiner ca.  $3 \times 3 \text{ cm}^2$  großen Segmente durchlaufen. Er prüft damit die Schauerentwicklung in einer Tiefe von  $3,3$  Strahlungslängen.

Die Zahl der zur Elektronenerkennung benötigten Variablen kann je nach Teilchenenergie sehr groß sein. Dabei sind Korrelationen zwischen den einzelnen Variablen zu berücksichtigen. Dies macht die Elektronenidentifikation zu einem hochdimensionalen Problem. Hier bieten die neuronalen Netzwerke einen neuen Zugang. Sie wurden bereits erfolgreich zur Elektronenerkennung mit dem Kalorimeter eingesetzt.

In dieser Arbeit werden Möglichkeiten und Grenzen der Elektronenerkennung unter Berücksichtigung von Kalorimeter und HES mit Hilfe neuronaler Netzwerke untersucht. Es wird zunächst der ZEUS-Detektor vorgestellt, insbesondere der Hadron-Elektron-Separator und das Kalorimeter. In einem weiteren Kapitel wird ein Einblick in die Wirkungsweise neuronaler Netzwerke gegeben. An einfachen Beispielen wird der Trainingsvorgang solcher Netzwerke diskutiert. Schließlich werden verschiedene Elektronenfinder an besonders schwierig zu identifizierenden Elektronen im Impulsbereich von  $0,5$  bis  $3 \text{ GeV}/c$  erprobt.



## Kapitel 2

# ZEUS

### 2.1 Der ZEUS-Detektor

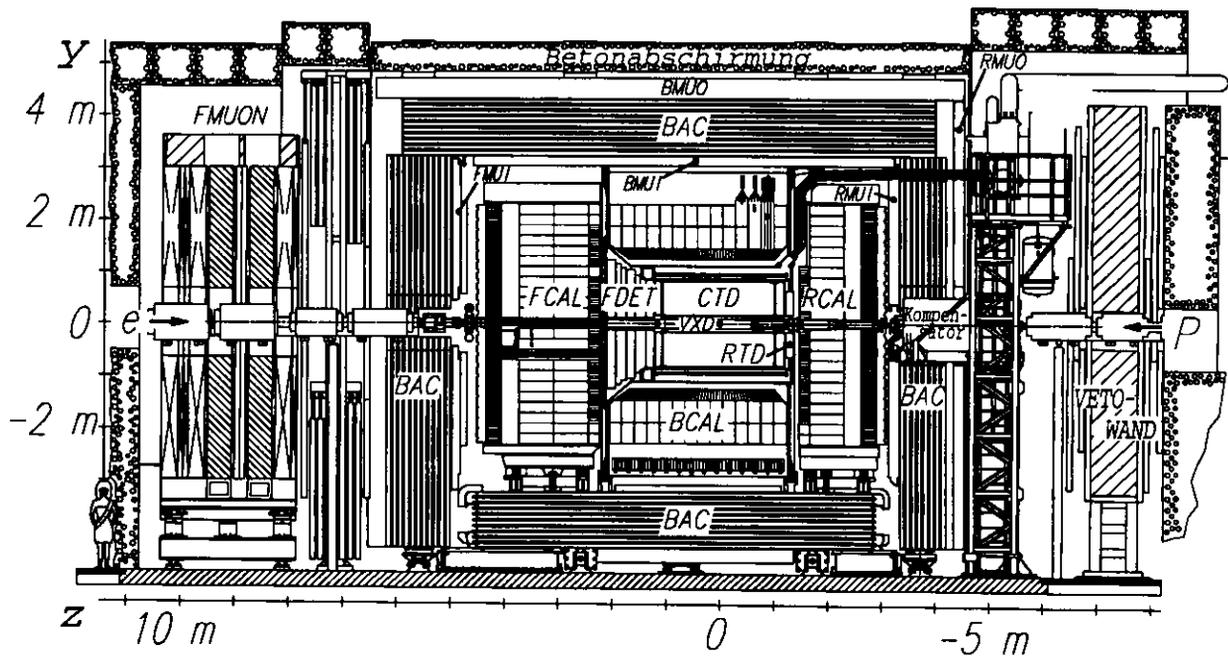


Abbildung 2.1: Längsschnitt durch den ZEUS-Detektor. Die Bezeichnungen werden im Text erklärt.

Die Abb. 2.1 zeigt den ZEUS-Detektor im Längsschnitt. Von der rechten Seite kommen die Protonenbündel im evakuierten Strahlrohr. Sie kollidieren im Zentrum des Detektors mit den entgegenkommenden Elektronenbündeln. Da das Proton einen wesentlich höheren Impuls trägt als das Elektron, werden die energiereichsten Produkte der Kollision in der Flugrichtung des Protons erwartet. Diese Richtung wird als Vorwärtsrichtung oder positive  $z$ -Richtung bezeichnet.

Der nominelle Kollisionspunkt heißt Wechselwirkungspunkt. Er bildet den Ursprung des ZEUS-Koordinatensystems. Tatsächlich begegnen sich Elektronen und Protonen aber auf einer Länge von etwa 50 cm entlang der  $z$ -Achse. Der Ort der eigentlichen Wechselwirkung wird als

Hauptvertex des Ereignisses bezeichnet. Zur genauen Bestimmung des Vertexortes dient der Vertexdetektor (VXD), eine kleine zylinderförmige Driftkammer, die das Strahlrohr umschließt.

Nach dem Vertexdetektor durchqueren die in der Wechselwirkung entstandenen Teilchen die zentrale Spurkammer (CTD). Diese ist von einer supraleitenden Spule umschlossen, die in der Kammer ein 1,43 T starkes, recht homogenes Magnetfeld erzeugt, dessen Feldlinien parallel zur Strahlachse verlaufen. Das Feld krümmt die Bahn der Teilchen. Krümmungsradius und Richtung lassen Impuls und Ladung des Teilchens erkennen. Die weiteren Spurkammern (FDET, RTD) waren 1993 noch nicht an der Datennahme beteiligt.

Im Kalorimeter wird die Energie der Teilchen gemessen. Das ZEUS-Kalorimeter besteht aus abwechselnden Lagen von abgereichertem Uran (3,3 mm) und Szintillator (2,6 mm). Tritt ein Elektron oder Positron in das Kalorimeter ein, so wird es gestreut und erzeugt Bremsstrahlung. Diese konvertiert wiederum in ein Elektron-Positron-Paar usw. Auf diese Weise wird die Energie des Ursprungsteilchens auf viele Teilchen verteilt, das Teilchen „schauert auf“. Unterschreitet die Energie der einzelnen Schauerteilchen einen kritischen Wert, so tritt die Paarbildung hinter anderen Prozessen zurück, und die Zahl der Schauerteilchen nimmt wieder ab. Im Endeffekt wird die gesamte Energie durch Ionisation verbraucht. Bei diesem Prozeß entstehen im Szintillator optische Photonen, die gesammelt und an seitlich angebrachte Wellenlängenschieber weitergegeben werden. Diese absorbieren das Licht und strahlen es mit einer größeren Wellenlänge wieder aus. Durch Lichtleiter wird es zu den Photoverstärkern geleitet und dort ausgelesen. Hadronen (etwa Protonen und Pionen) reagieren elektromagnetisch *und* durch die starke Kernkraft. Im Verlauf der Schauerentwicklung treten Prozesse unterschiedlichster Art auf: Zerfallende neutrale Pionen erzeugen elektromagnetische Schaueranteile, Kerne werden angeregt, Neutronen freigesetzt. Die Unterscheidungsmöglichkeiten elektromagnetischer und hadronischer Schauer im Kalorimeter werden ausführlich in Abschnitt 2.2 beschrieben.

Das Energie-Auflösungsvermögen des Kalorimeters beträgt für Elektronen bzw. Hadronen (in GeV)

$$\frac{\sigma_{E,e}}{E} \approx \frac{18\%}{\sqrt{E}} \quad \text{bzw.} \quad \frac{\sigma_{E,h}}{E} \approx \frac{35\%}{\sqrt{E}}$$

Das Kalorimeter ist unterteilt in das Vorwärtskalorimeter (FCAL), das zylindrisch die Spule umgebende Ringkalorimeter (BCAL) und das Rückwärtskalorimeter (RCAL). Aufgrund der unterschiedlichen Impulse von Elektronen und Protonen hat das FCAL eine wesentlich größere Tiefe als das RCAL. In das Rückwärtskalorimeter ist der Hadron-Elektron-Separator eingebaut. Er wird im nächsten Abschnitt im Zusammenhang mit der Elektronenidentifikation ausführlich beschrieben. Die drei Teile umschließen, vom Wechselwirkungspunkt aus gesehen, 99,7% des Raumwinkels.

Um das Kalorimeter herum sind Myonkammern und das eiserne Joch gebaut. Das Joch dient der Rückführung des Magnetfeldes. Ein eingebautes Außenkalorimeter (BAC) stellt fest, ob aus dem Hauptkalorimeter Energie „herausgeleckt“ ist.

In Vorwärtsrichtung befindet sich ein Myonspektrometer (FMUON), und eine Reihe kleinerer Zähler, die nicht mit abgebildet sind. Sie finden elastisch gestreute Protonen, Neutronen und im Falle unelastischer Reaktionen den Protonrest. Großflächige Szintillatoren kontrollieren zu beiden Seiten des Kalorimeters eintretende Teilchen aus Kollisionen von Elektronen oder Protonen mit Restgasatomen. In Rückwärtsrichtung befinden sich zwei Luminositätsdetektoren, die unter kleinem Winkel gestreute Elektronen und Photonen der Bremsstrahlung nachweisen.

Zwischen dem Rückwärtskalorimeter und dem Kompensatormagneten wird ein kleines Wolfram-Silizium-Kalorimeter erprobt.

## 2.2 Elektronenidentifikation mit dem ZEUS-Detektor

### 2.2.1 Einleitung

Mesonen mit *charm* und *bottom quarks* besitzen neben rein hadronischen Zerfallskanälen auch solche, die Elektronen oder Myonen enthalten. Mit Hilfe eines Elektronenfinders lassen sich diese Mesonen identifizieren. Aus den bekannten Verzweigungsverhältnissen kann man Wirkungsquerschnitte für die Produktion dieser Teilchen (und ggf. ihrer Mutterteilchen) berechnen. Die Genauigkeit solcher Rechnungen wird aber beeinträchtigt durch eine Reihe von Fehlerquellen:

1. Fälschliche Zuordnung von Zerfallsteilchen
2. Fehlerhafte Impuls- oder Energiemessung
3. Misidentifikation von Hadronen als Elektronen und umgekehrt

Aufgabe eines Elektronenfinders ist es, bei nicht zu geringer Effizienz<sup>1</sup> die Misidentifikationswahrscheinlichkeit zu minimieren. Gleichzeitig sollen Effizienz und Misidentifikationswahrscheinlichkeit bzw. ihr Komplement, die Diskriminanzkraft, möglichst genau bekannt sein. Darüber hinaus wäre es wünschenswert, in gewissem Rahmen das Verhältnis von Effizienz und Diskriminanzkraft der Aufgabenstellung anpassen zu können. Wird ein seltenes Signal von einem großen Untergrund überdeckt, so möchte man eine hohe Diskriminanzkraft anwenden, auch auf Kosten der Effizienz. Läßt sich dagegen der Untergrund auf andere Weise ausreichend reduzieren, so wird vielleicht zum Zwecke höherer Statistik eine recht hohe Effizienz gewünscht.

Der ZEUS-Detektor ermöglicht die Identifikation von niederenergetischen Elektronen bzw. Positronen auf mehrfache Weise.

### 2.2.2 Die Spurkammer

Teilchen mit nicht zu hohen Impulsen lassen sich durch die Dichte der auf ihrer Bahn erzeugten Gasionen unterscheiden. In bestimmten Impulsbereichen werden dadurch Protonen, Kaonen, Myonen und selbst Pionen gegenüber Elektronen diskriminiert.

### 2.2.3 Das Kalorimeter

Im Kalorimeter geben sich die verschiedenen Teilchensorten durch die Formen der von ihnen erzeugten Teilchenschauer zu erkennen. Myonen bilden i. a. keine Schauer aus. Sie passieren meistens das Kalorimeter als minimal ionisierende Teilchen. Die Schauer von Elektronen, Positronen und Photonen unterscheiden sich von denen der Hadronen vor allem in ihrer longitudinalen Energieverteilung bzw. Lage im Kalorimeter. Die elektromagnetische Strahlungslänge des ZEUS-Kalorimeters beträgt

$$X_0 = 0,74 \text{ cm.}$$

Die mittlere freie Weglänge eines Hadrons liegt dagegen bei

$$\lambda = 21,0 \text{ cm}$$

---

<sup>1</sup>Die Begriffe Effizienz, Diskriminanzkraft und Misidentifikationswahrscheinlichkeit werden in einem eigenen Abschnitt des Kapitels 4 definiert.

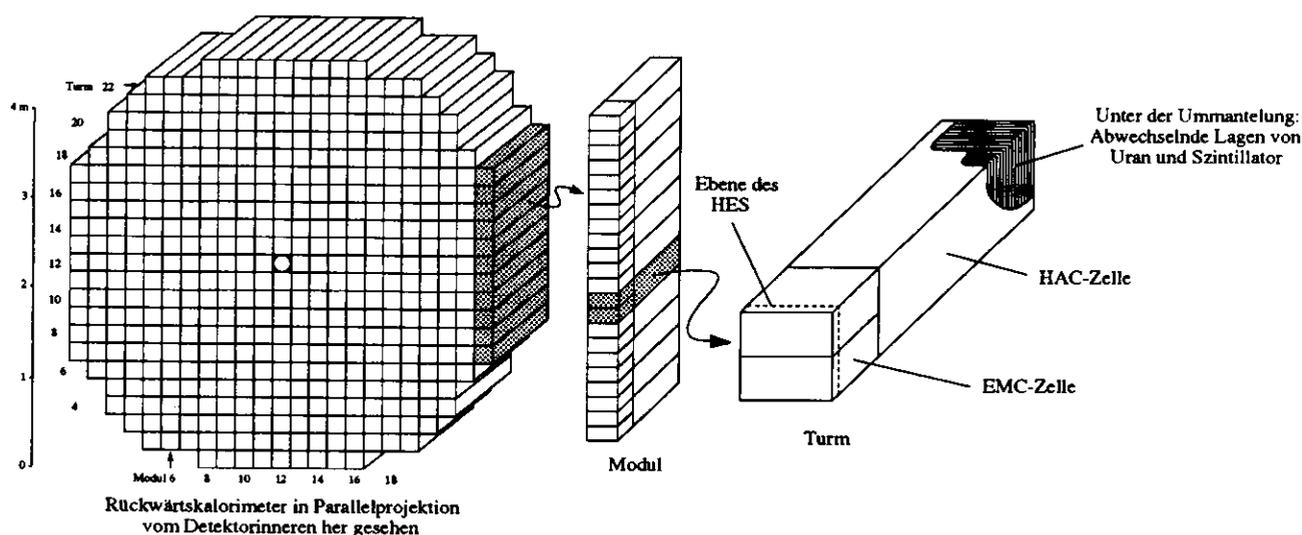


Abbildung 2.2: Aufbau des Rückwärtskalorimeters. Das Rückwärtskalorimeter (links) ist unterteilt in 23 Module, von denen jedes aus 11 bis 23 Türmen besteht (Mitte). Jeder Turm setzt sich zusammen aus zwei EMC-Zellen für elektromagnetische Schauer und einer HAC-Zelle. Jede Zelle ist wiederum aufgebaut aus einer Schichtung von Uran und Szintillator. Nicht abgebildet sind die Wellenlängenschieber, die Photoverstärker und die Elektronik sowie die Kühlung und die Eisenträger.

Die Schauer von Elektronen, Positronen und Photonen deponieren ihre Energie in den ersten 20 cm des Kalorimeters. Hadronische Schauer entwickeln sich meist erst tiefer im Kalorimeter. Das Rückwärtskalorimeter ist daher segmentiert in ein 23,6 cm tiefes elektromagnetisches Kalorimeter (EMC) und ein dahinter liegendes hadronisches Kalorimeter (HAC). Die beiden Sektionen werden getrennt ausgelesen. Das Verhältnis der in den beiden Sektionen deponierten Energien eines Schauers kann als Kriterium zur Unterscheidung zwischen Elektronen, Positronen und Photonen einerseits und Hadronen andererseits beitragen. Ein nicht geringer Teil vor allem der niederenergetischen Hadronen schauert jedoch bereits im EMC. Dies führt zu Misidentifikationen, wenn man nur die Kalorimeterinformation benutzt.

Neben der longitudinalen Energieverteilung unterscheiden sich die elektromagnetischen und hadronischen Schauer auch durch ihre laterale Ausdehnung. Hadronische Schauer sind i. a. breiter als elektromagnetische gleicher Energie. Problematisch sind hier jedoch die Überlagerungen von Schauern mehrerer Teilchen mit oder ohne Elektron.

Die Abbildung 2.2 zeigt genauer die Segmentierung des Rückwärtskalorimeters. Es setzt sich zusammen aus 23 Modulen, die ihrerseits bis zu 23 Türme haben. Jeder Turm<sup>2</sup> besteht aus einer hadronischen (HAC-) und zwei elektromagnetischen (EMC-) Zellen. Jede dieser Zellen ist über Wellenlängenschieber an beiden Seiten des Moduls mit Photoverstärkern verbunden. Die  $x$ -Koordinate der Schauerposition kann daher aus dem Verhältnis der Signale der beiden Photoverstärker jeder Zelle ermittelt werden.

Die elektromagnetischen Zellen sind 20 cm breit und 10 cm hoch. Ihre Tiefe beträgt 23,6 cm.

<sup>2</sup>Die Bezeichnung Turm entstand durch die senkrechte Lage dieser Kalorimeterteile bei der Schichtung der Uran- und Szintillatorlagen.

Die hadronischen Zellen haben eine Größe von  $20 \times 20 \times 64,4 \text{ cm}^3$ . Die Mitten zweier benachbarter Türme und Module haben einen Abstand von  $20,33 \text{ cm}$ . Die genaue Position des Gesamtkalorimeters im ZEUS-Koordinatensystem ist mit den Daten festgehalten.

### 2.2.4 Der Hadron-Elektron-Separator

Der Hadron-Elektron-Separator (HES) besteht aus  $2,97 \times 3,32 \text{ cm}^2$  großen Siliziumdioden, angeordnet in einer Ebene,  $3,3$  Strahlungslängen hinter der Oberfläche des Rückwärtskalorimeters (vgl. Abb. 2.3 und 2.4). Sie sind auf bis zu  $4 \text{ m}$  langen Kunststoffstreifen (Skis) montiert, die senkrecht, zu dritt nebeneinander in die Module des Rückwärtskalorimeters eingelassen wurden. Die Skis tragen bis zu 113 Zeilen aus je zwei Dioden. Die Diodenzeilen sind abwechselnd auf der Vorder- und Rückseite der Skis montiert. In den Zwischenräumen zu den übernächsten Zeilen sind Vorverstärker untergebracht. Die weitere Ausleseelektronik befindet sich außerhalb des Kalorimetervolumens.

Die Dioden sind projektiv angeordnet, d. h., die Dioden auf der Rückseite der Skis bedecken vom Wechselwirkungspunkt her gesehen gerade die Lücken zwischen den Dioden auf der Vorderseite der Skis. Die hinteren Zeilen sind zu diesem Zweck wegen ihrer größeren Entfernung vom Wechselwirkungspunkt in größeren Abständen angeordnet (vgl. Abb. 2.4 links). Die Verschiebung beträgt aber selbst an den Rändern der HES-Fläche weniger als  $5 \text{ mm}$ .

Die Lücken zwischen Dioden auf einem Ski sind kleiner als  $1 \text{ mm}$ . Benachbarte Dioden auf verschiedenen Skis sind etwa  $3 \text{ mm}$ , solche in verschiedenen Modulen  $16 \text{ mm}$  voneinander getrennt. Es kann daher vorkommen, daß Teilchen, die die Modulgrenze treffen, vom HES nicht registriert werden. In  $y$ -Richtung ist die Bedeckung (vom Wechselwirkungspunkt her gesehen) lückenlos.

Teilchen, die in Sperrichtung betriebenen Siliziumdioden durchqueren, erzeugen in ihnen Elektron-Loch-Paare. Der entstehende kleine Strom wird aufgezeichnet. Unterscheiden lassen sich die Elektronen, Positronen und Photonen von den Hadronen durch die Menge der im HES erzeugten Ladung. Die meisten Hadronen schauern erst hinter der HES-Ebene (vgl. Abb. 2.5). Sie passieren den HES also als einzelne minimal ionisierende Teilchen. Die im Mittel von einem minimal ionisierenden Teilchen im Detektor hinterlassene Energie definiert die Einheit  $1 \text{ mip}$ .

Elektromagnetische Schauer dagegen sind in der HES-Ebene bereits voll entwickelt, da sie schon  $3,3$  Strahlungslängen Kalorimetermaterial passiert haben. Hinzu kommen je nach Nähe zum Strahlrohr noch ein bis drei Strahlungslängen totes Material.

Trotzdem identifiziert auch der HES nicht fehlerfrei. Es kann vorkommen, daß Hadronen bereits vor der HES-Ebene schauern. Andererseits kann es niederenergetische Elektronen geben, deren Schauer bereits stark abgeklungen ist, oder solche, die die HES-Ebene im Bereich einer Lücke zwischen den Kalorimeter-Modulen passierten.

### 2.2.5 Korrelation der Schauerentwicklung im Kalorimeter und im HES

Um die Anzahl der Fehler erster oder zweiter Art, d. h., die Misidentifikation eines Elektrons als Hadron oder die Misidentifikation eines Hadrons als Elektron, möglichst gering zu halten, sollen das Kalorimeter und der Hadron-Elektron-Separator gemeinsam eingesetzt werden. Neuronale Netze ermöglichen eine gleichzeitige, (in gewissen Grenzen) optimale Nutzung der gegebenen Informationen unter Berücksichtigung aller Korrelationen. Gerade die hochdimensionalen Korrelationen stellen den Vorteil neuronaler Netze gegenüber herkömmlichen einfachen Schnitten in niedrigdimensionalen Räumen dar. Dies ist allerdings nur dann von Nutzen, wenn die Informatio-

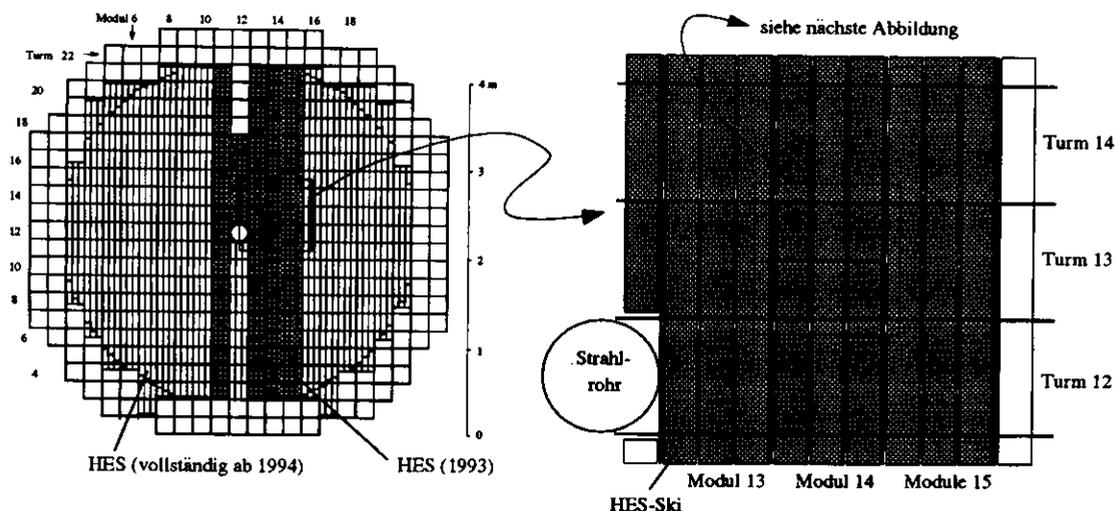


Abbildung 2.3: Der gesamte HES bedeckt eine nahezu kreisförmige Fläche (links). Jedes Kalorimetermodul enthält drei Streifen (Skis), die mit jeweils zwei Kolonnen von Dioden bestückt sind (Dioden nicht abgebildet).

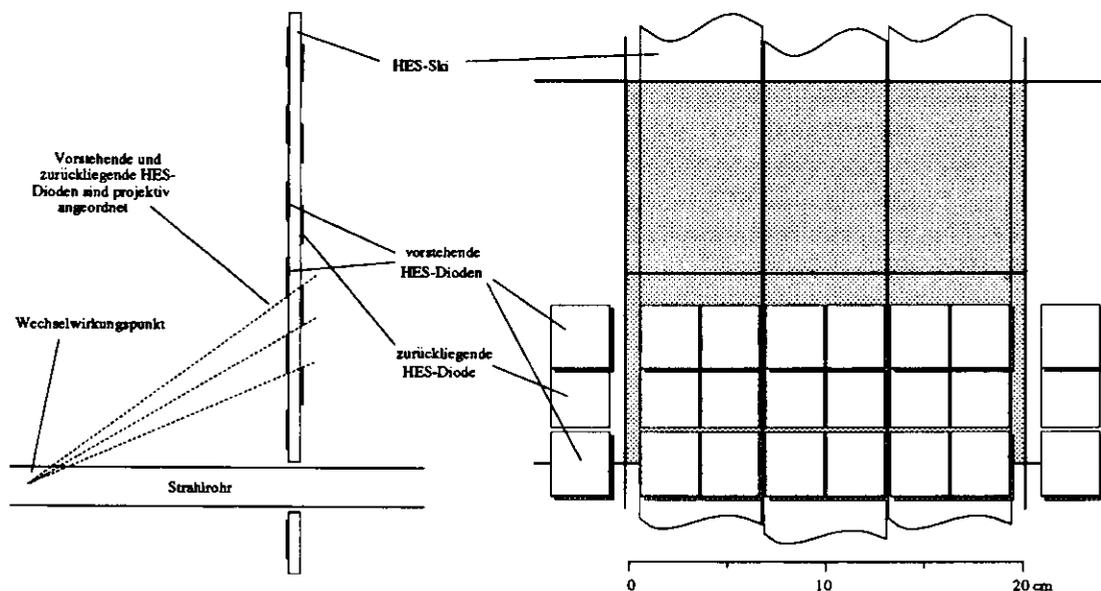


Abbildung 2.4: Detailansicht der Lage der HES-Dioden. Um auf den HES-Skis den Vorverstärkern Raum zu geben, wurde jede zweite Diodenzeile auf der Rückseite der Skis angebracht. Die vom Wechselwirkungspunkt her gesehen zurückliegenden Dioden wurden mit größerem Abstand montiert, um eine vom Wechselwirkungspunkt her gesehen vollständige Flächenbedeckung zu erreichen. Diese projektive Anordnung ist links schematisch dargestellt. Die rechte Abbildung zeigt maßstabgetreu die drei Diodenzeilen der unteren EMC-Zelle des Turmes 13 im Kalorimetermodul 14. Die vorstehenden HES-Dioden sind mit einem Schatten versehen. Man erkennt deutlich die Verschiebung der mittleren, zurückliegenden Zeile nach oben.

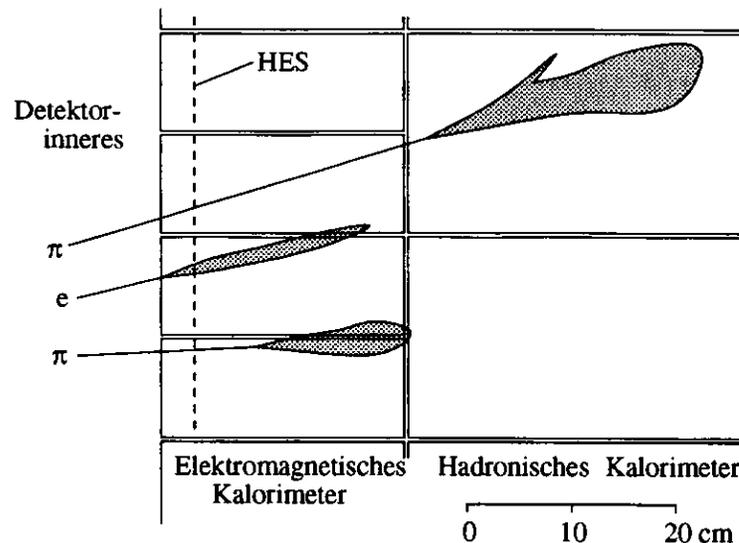


Abbildung 2.5: Prinzip der Elektronenidentifikation mit dem HES. Die Abbildung zeigt eine Seitenansicht des Rückwärtskalorimeters. Nicht alle Pionen deponieren den Hauptteil ihrer Energie im hadronischen Kalorimeter. Der HES kann früh schauernde Pionen erkennen, wenn ihr Schauer sich nicht vor der HES-Ebene ( $\lambda = 0.1$ ) ausbildet.

nen voneinander unabhängig sind, d. h., wenn das Versagen einer Komponente bei der Beurteilung des Teilchentyps nicht mit dem Versagen der anderen gekoppelt ist. Das trifft leider teilweise zu bei dem Kalorimeter und dem Hadron-Elektron-Separator. Es bestehen die folgenden einseitigen Abhängigkeiten:

1. Entwickelt sich ein hadronischer Schauer untypisch früh, so daß er bereits im HES ein hohes Signal gibt, so sieht auch das Kalorimeter den Großteil der deponierten Energie in der elektromagnetischen Sektion.
2. Versagt das elektromagnetische Kalorimeter bei der Identifikation eines Elektrons (z. B. weil die Modulgrenze getroffen wurde), so wird auch im HES dieses Elektron nicht zu sehen sein.

Umgekehrt kann aber ein vom Kalorimeter als Elektron gesehenes Pion vom HES als Hadron identifiziert werden, wenn es erst in der Mitte oder im hinteren Teil des EMC aufgeschauert ist. Ebenso gibt es Fälle, in denen das Kalorimeter Elektronen erkennt, die vom HES nicht als solche identifiziert werden.

Tatsächlich konnten diese Zusammenhänge in den in Kap. 4 beschriebenen Daten festgestellt werden. Von den Hadronen, die im HES mehr als 7 *mip* deponierten, überschritten nur etwa 15 % die Energieschwelle der HAC-Zelle (100 *MeV*), während die Hadronen mit einem HES-Signal von weniger als 2 *mip* zu 36 % das hadronische Kalorimeter erreichten.

Noch deutlicher trat die zweite erwartete Abhängigkeit zutage: Während die Elektronen ohne hadronische Energie zu 82 % die (willkürlich gewählte) Grenze von 3 *mip* überschritten, trat dies bei denen mit gleicher Energie in EMC- und HAC-Sektion nur zu etwa 40 % auf. Elektronen, die mehr als 75 % ihrer Energie im HAC deponierten, hinterließen grundsätzlich weniger als 3 *mip* im HES.



# Kapitel 3

## Neuronale Netzwerke

### 3.1 Einleitung

Neuronale Netzwerke haben im letzten Jahrzehnt bei der Behandlung multidimensionaler Parameteranpassungen rasch an Popularität gewonnen. Sie klassifizieren Daten nach ihren Eigenschaften, ohne daß der Benutzer diese kennen muß. Sie lernen von Beispielen. Sie verkraften Fehler. Sie unterscheiden Wichtiges von Unwichtigem. Sie sind universell einsetzbar, z. B. in der Schrift- und Spracherkennung, in der Wetter- und Börsenprognose, zur Prozeßkontrolle und als assoziatives Gedächtnis. Außerdem sind sie schnell. Das gilt besonders für die als elektronische Schaltung realisierten, eigentlichen neuronalen Netzwerke. Solche fest verdrahteten neuronalen Netzwerke eignen sich z. B. zur Vorselektion von Ereignissen in elektronischen Detektoren der Hochenergiephysik [11]. Aber auch die simulierten neuronalen Netze lösen viele Klassifikations- und Mustererkennungsaufgaben schneller und mit mehr Erfolg als konventionelle Verfahren.

Dennoch sind neuronale Netze keine Wunderalgorithmen. Sie zu trainieren bedeutet nicht-lineare Minimierungsaufgaben zu lösen, die mathematisch nicht einfach zu kontrollieren sind. Dieses Kapitel dient der Einführung in die Theorie der neuronalen Netzwerke und der Beschreibung der in dieser Arbeit benutzten Netze.

### 3.2 Begriffsbestimmung

Der Ausdruck „neuronales Netzwerk“ bezeichnet ein elektronisch oder als Computersimulation realisiertes Netzwerk von Funktionen, die die Neuronen des Gehirns modellieren. Im folgenden soll nur von simulierten Netzwerken die Rede sein. Ihre „Neuronen“ oder „Knoten“ bestehen im Kern aus Schwellfunktionen, das sind i. a. sigmoide Funktionen wie etwa die Signum- oder Tangens-Hyperbolicus-Funktion. Deren Argument ist im einfachsten Falle die Summe der gewichteten einlaufenden Signale (vgl. Abb. 3.1). Ihr Wert wird an die nachfolgenden „Neuronen“ weitergegeben. Die daran beteiligten Parameter werden bzgl. eines vorgegebenen Problems optimiert. Man unterscheidet entsprechend zwischen

1. linearen und nichtlinearen,
2. rekursiven und nichtrekursiven und
3. überwachten und nichtüberwachten neuronalen Netzen.

Ferner lassen sich neuronale Netzwerke nach ihren Eingabeverknüpfungen und Schwellfunktionen, der Zahl ihrer Netzebenen und (bei rekursiven Netzen) der Art der Rückkopplung klassifizieren.

Lineare neuronale Netze werden nur einschichtig rekursiv konstruiert, da Linearkombinationen linearer Funktionen wieder nur lineare Funktionen der Eingangsvariablen sind. Ein solches lineares neuronales Netzwerk ist aber nichts anderes als die lineare Diskriminanzanalyse, die als Eigenwertproblem dargestellt werden kann.

Nichtlineare rekursive neuronale Netzwerke zeichnen sich durch rückkoppelnde Verbindungen aus, die die Auswertung des Netzes zu einem Konvergenzprozeß werden lassen. Sie werden z.B. dazu benutzt, die beschädigte Pixelstruktur eines Zeichens wieder in ihre Standardform zurückkonvergieren zu lassen (Schrifterkennung). Als Beispiel seien hier die Hopfield- und die Boltzmann-Netze genannt [7].

Überwachte und nicht überwachte neuronale Netze unterscheiden sich in der Methode ihres Trainings. Überwachtes Lernen bedeutet, daß das Netz zu jedem Trainingsmuster die erwünschte Ausgabe gezeigt bekommt. Aus der Differenz der erbrachten zur erwünschten Ausgabe berechnet es, in welcher Weise die Netzparameter zu korrigieren sind.

Nichtüberwachte neuronale Netze kommen mit Trainingsbeispielen *ohne* vorgegebene Klassifikation aus. Sie benötigen aber alternative Kriterien zur Korrektur ihrer Gewichte. Der bekannteste Algorithmus ist hier der des konkurrierenden Lernens. Die Ausgabeknoten sind in einem ein-, zwei- oder höherdimensionalen Gitter angeordnet, auf das die Strukturen des Eingaberaumes projiziert werden sollen. Benachbarte Knoten kämpfen nun um die stärkste Reaktion auf jedes Eingabemuster. Der Gewinner wird auf das Eingabemuster hin korrigiert, während die benachbarten Knoten davon weg korrigiert werden. Im Verlauf dieses Lernprozesses entwickeln sich die Ausgabeknoten zu Vektoren, die auf Verdichtungen in den Strukturen des Eingaberaumes zeigen und diesen damit in Klassen zerlegen. Diese Klassen werden zum Schluß durch den Anwender interpretiert. Beispiele für nichtrekursive neuronale Netzwerke dieser Art finden sich bei Kohonen [10], solche für rekursive bei Grossberg [5].

In dieser Arbeit werden nur *nichtlineare nichtrekursive überwachte* neuronale Netze verwendet, denn die linearen Netze vermögen die erwarteten Ereignisse nicht zu klassifizieren. Nicht überwachte Netze können in der Elektronenidentifikation nicht eingesetzt werden, da die Teilchenklassen im Ereignisraum zu stark überlappen. Von den nichtlinearen neuronalen Netzen mit überwachtem Training wurden für diese Arbeit die nichtrekursiven gewählt, weil sie mathematisch am einfachsten sind. Diese sogenannten *feed forward*-Netze sollen im folgenden Abschnitt näher beschrieben werden.

### 3.3 Theorie der überwachten *feed forward*-Netze

#### 3.3.1 Nomenklatur und mathematische Beschreibung

Die *feed forward*-Netze setzen sich zusammen aus Knoten, die in Ebenen angeordnet sind, und Gewichten, die die Knoten benachbarter Ebenen verbinden, d. h., ihre Eingabeverknüpfungen sind gewichtete Summen (vgl. Abb. 3.1).

Die durch die Eingabeebene ankommenden Daten werden an die verborgene Ebene weitergeleitet. Dabei werden sie an jeder Verbindung gewichtet und vor dem Knoten summiert. Der Summe wird ein Schwellwert  $t$  beigefügt. Im Knoten wird auf die Summe eine Schwellfunktion

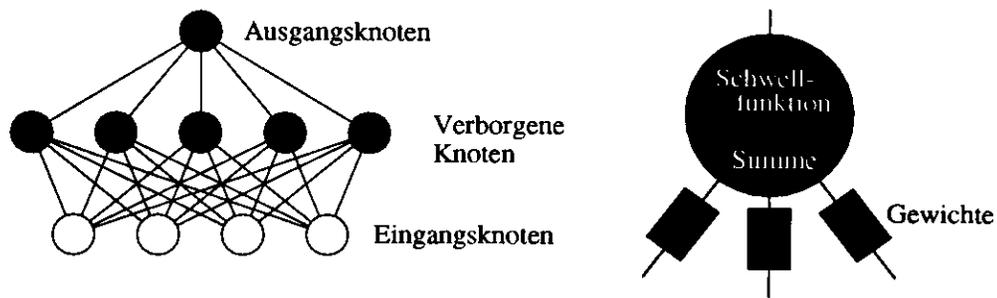


Abbildung 3.1: Beispiel der Geometrie eines *feed forward*-Netztes und Detaildarstellung eines Knotens.

angewandt. Als Schwellfunktion dient eine sigmoide Funktion, etwa

$$a \mapsto g(a) = \tanh a.$$

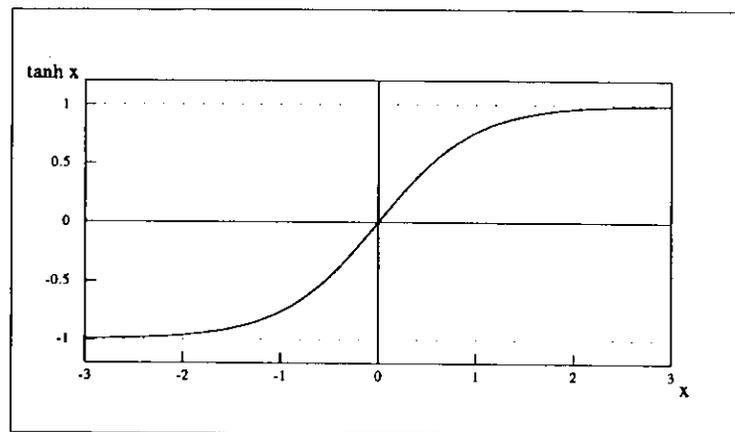


Abbildung 3.2: Graph der Funktion  $\tanh x$ .

Der Funktionswert wird an die nächste Ebene weitergegeben. Im Beispiel der Abb. 3.1 handelt es sich dabei um die Ausgabeebene, allgemein werden jedoch auch Netze mit keiner oder mehr als nur einer verborgenen Ebene benutzt. In diesen verborgenen Ebenen wird die Information von den (passiven) Eingangsknoten vorverarbeitet, bevor der oder die Ausgangsknoten die endgültigen Funktionswerte berechnen. Von Ebene zu Ebene werden dabei abstraktere Eigenschaften der Eingangsdaten erkannt. Dies ist genauer in Abschnitt 3.3.3 beschrieben.

Das in Abb. 3.1 gezeigte Netz steht für die folgende Funktion:

$$f_k(\vec{x}) = \underbrace{g\left(t_{2k} + \sum_{j=1}^5 w_{2kj} \times \underbrace{g\left(t_{1j} + \sum_{i=1}^4 w_{1ji} \underbrace{x_i}_{\text{Eingabeebene}}\right)}_{\text{Verborgene Ebene}}\right)}_{\text{Ausgabeebene}} \quad (3.1)$$

Die vier Eingangswerte  $x_i$  werden mit  $w_{1ji}$  gewichtet summiert und an den jeweiligen Knoten der verborgenen Ebene weitergegeben. Dort wird der Schwellwert  $t_{1j}$  addiert und die Schwellfunktion angewandt. Dies geschieht an allen fünf Knoten der verborgenen Ebene. Die Ergebnisse

werden wieder gewichtet, summiert, um den Schwellwert  $t_{2k}$  vermehrt und als Argument der Schwellfunktion des einzigen Ausgangsknotens ( $k = 1$ ) übergeben.

Das in Abb. 3.1 dargestellte Netz liefert für jedes fünfkompontige Eingabemuster einen einzelnen Wert im Bereich  $] - 1, 1[$ . Ziel des Trainings ist es nun, die Parameter  $w_{ij}$  und  $t_{ij}$  aller Ebenen so einzustellen, daß die Gesamtfunktion  $f$  die Eingabemuster klassifiziert. Im Falle der Elektronenfinder kann das bedeuten, daß Signaturen für Elektronen ein Wert nahe 1 und solchen für Nichtelektronen einer nahe  $-1$  zugewiesen wird.

In dieser Arbeit soll für die Variablen und ihre Indizes eine einheitliche Nomenklatur verwandt werden, die in Tab. 3.1 eingeführt wird.

Kurzbezeichnung	Gegenstand	Bedeutung der Indizes
$x_i$	Eingabe	Knoten
$f_k$	Ausgabe	Knoten
$f_{soll,k}$	Erwünschte Ausgabe	Knoten
$w_{ij}$	Gewicht	Obere Ebene, oberer,unterer Knoten
$g$	Schwellfunktion	
$t_{ij}$	Schwellwert	Ebene,Knoten
$a_{ij}$	Argument	Ebene,Knoten
$o_{ij}$	Ausgabe aller Knoten	Ebene,Knoten
$l, p = 0..m$	Ebenenindex	
$i, j, k, q = 0..n_l$	Knotenindizes	ggf. Ebene
$r = 1..R$	Eingabemuster	
$s = 1..S$	Trainingsschritte	

Tabelle 3.1: Kurzbezeichnungen und Indizes.

### 3.3.2 Trainingsvorgang

Zu Beginn des Trainings werden die Anfangswerte der Parameter festgelegt. Wichtig ist hierbei, daß keine zwei Knoten paarweise gleiche Gewichte auf den Verbindungen zu allen Eingabeknoten und der Ausgabe sowie gleiche Schwellwerte besitzen. Eine solche Symmetrie kann im Verlauf eines Trainings nicht aufgehoben werden.

Dann werden dem Netz nacheinander verschiedene Trainingsmuster und die dafür erwünschten Ausgaben vorgeführt. Es berechnet daraus mit den jeweiligen Parameterwerten  $w_{ij}^{(s)}$  und  $t_{ij}^{(s)}$  eine Ausgabe. Aus dem Ausgabefehler  $E$ , das ist die halbe<sup>1</sup> quadratische Differenz zwischen der vom Netz berechneten und der erwünschten Ausgabe, und den partiellen Ableitungen dieses Fehlers nach den Parametern wird eine Korrektur für die Parameter berechnet. Auf diese Weise werden die Gewichte und Schwellwerte iterativ optimiert. Die einfachste Form der Korrektur ist ein

<sup>1</sup>Der Faktor  $\frac{1}{2}$  dient lediglich der Vereinfachung der partiellen Ableitungen der Fehlerfunktion. Er wird daher in den Anwendungen weggelassen.

Schritt der Schrittweite  $\eta$ :

$$E(\vec{t}, \vec{w}) = \frac{1}{2} \sum_{i_m=1}^{n_m} (f_{i_m}(\vec{t}, \vec{w}) - f_{\text{sol}, i_m}(\vec{t}, \vec{w}))^2$$

$$t_{ij}^{(s+1)} = t_{ij}^{(s)} - \eta \frac{\partial E}{\partial t_{ij}^{(s)}}$$

$$w_{iji}^{(s+1)} = w_{iji}^{(s)} - \eta \frac{\partial E}{\partial w_{iji}^{(s)}}$$

Zur Vermeidung starker Schwankungen der Parameterwerte kann man dem Netz die Daten auch blockweise vorführen, d. h. die Korrekturen zunächst über einen Block von Daten mitteln, bevor sie an den Parametern durchgeführt werden. Weitere Methoden zur Verbesserung des Konvergenzverhaltens sind in Abschnitt 3.4 beschrieben.

Ein großer Vorteil der Minimierung der Fehlerfunktion von *feed forward*-Netzen gegenüber allgemeinen nichtlinearen Funktionen ist die Abhängigkeit vieler partieller Ableitungen voneinander. So ist der Aufwand der Berechnung der partiellen Ableitungen eines  $N$ -parametrischen Netzes statt  $(N + 1)$ -mal nur etwa zwei- bis dreimal so groß wie die Berechnung des Funktionswertes selbst. Das dabei benutzte Verfahren nennt man *back propagation*-Algorithmus, da sich in ihm der Fehler rückwärts durch das Netz fortpflanzt.<sup>2</sup>

### 3.3.3 Geometrische Deutung

Eine der besonderen Stärken der *feed forward*-Netze ist ihre einfache geometrische Interpretierbarkeit. Sie erleichtert das Verständnis der Mathematik neuronaler Netzwerke und macht so die kreative, gezielte Arbeit mit einem derartig hochdimensionalen nichtlinearen Optimierungsproblem erst möglich.

Man stelle sich einen  $n$ -dimensionalen Eingaberaum vor. In diesem seien die zu klassifizierenden Daten gleich Wolken beliebiger, aber nicht zu komplizierter Form verteilt. Die gewichtete Summe der Eingabedaten am ersten Knoten der (ersten) verborgenen Ebene ist eine lineare Funktion, deren Äquipotentialhyperfläche

$$\sum_{i=1}^n w_{11i} x_i = -t_{11}$$

den Raum in zwei Halbräume teilt (vgl. Abb. 3.3 (b)). Im Falle der sigmoiden Schwellfunktion  $g(x) = \tanh x$  erhält der eine Halbraum Werte nahe  $-1$ , der andere solche nahe  $1$ . An der Grenze ergibt sich ein weicher Übergang, dessen Ausdehnung umgekehrt proportional zum Betrag des Gewichtsvektors ist.

Jeder weitere Knoten der ersten verborgenen Ebene definiert eine weitere derartige Teilung des Eingangsraumes.

Der Übergang von der (ersten) verborgenen zur nächsthöheren Ebene ist grundsätzlich anders zu interpretieren, da die nun vorliegenden Daten nahezu binär sind. Man sollte sich daher diesen und ggf. die weiteren Übergänge als Operationen auf Mengen, nämlich den definierten Halbräumen, oder als logische Operationen vorstellen.

<sup>2</sup>Eine ausführliche Berechnung der partiellen Ableitungen des *feed forward*-Netzes wird in Anhang A.1 durchgeführt.

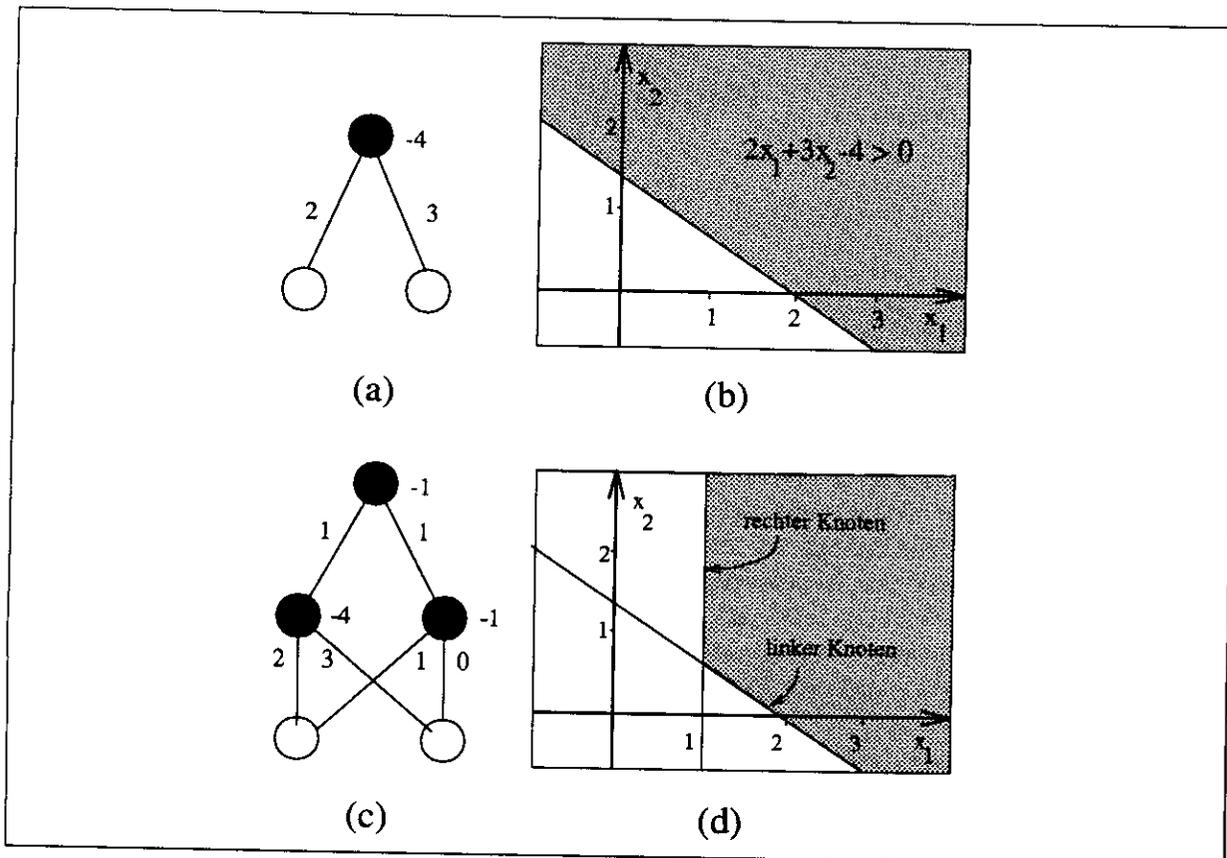


Abbildung 3.3: Das Netz (a) vermag den zweidimensionalen Eingaberaum (b) durch eine Gerade in zwei Klassen zu teilen. Die zweite Stufe des Netzes (c) realisiert eine UND-Funktion der Entscheidungen des linken und rechten Knotens der verborgenen Ebene. (d) zeigt das durch dieses Netz hervorgehobene offene Polygon. Der Einfachheit halber wurde als Schwellfunktion die Sigmoid-Funktion verwendet.

Am (ersten) Knoten der nächsthöheren Ebene wird eine Schnitt- und Vereinigungsmenge der Halbräume gebildet (Abb. 3.3 (d)). Sie definiert die den Eingaberaum klassifizierende Trennfläche, die i. a. keine Hyperebene mehr ist, sondern ein nicht notwendig geschlossener Polyeder, der im Falle stetig differenzierbarer Schwellfunktionen abgerundete Ecken und Kanten besitzt. Es können sogar mehrere Polyeder sein, die ein nicht einfach oder auch gar nicht zusammenhängendes Gebiet vom Restraum trennen. Jede Seite des Polyeders entspricht dabei einem Knoten der (ersten) verborgenen Ebene.

Auf dieselbe Weise kann man an den weiteren Knoten der Ebene zusätzliche Polyeder herausheben und diese in einer höheren Ebene wieder kombinieren usw.

Im Einzelfall stellen diese Mengenverknüpfungen logische Verknüpfungen dar. Für stetige Schwellfunktionen handelt es sich dabei um „weiche“ logische Verknüpfungen.<sup>3</sup> Jeder Knoten der (ersten) verborgenen Ebene trifft eine Entscheidung über die Eingangsdaten. Diese Entscheidungen werden in der bzw. den weiteren Ebenen logisch verknüpft.

<sup>3</sup>Mit derartigen Verknüpfungen beschäftigt sich die *fuzzy logic*.

Im Prinzip läßt sich bereits mit *einer* verborgenen Ebene jede Trennung des Eingaberaumes beliebig genau approximieren ([6], Fußnote 7, S. 182). Man könnte also auf die weiteren Ebenen verzichten. Falls aber tatsächlich mehrere Polyeder zur Klassifizierung des Eingaberaumes benötigt werden, reduziert die Einführung einer weiteren verborgenen Ebene den Lernaufwand erheblich ([8], S. 102).

Diese Arbeit beschränkt sich auf Netze mit *einer* verborgenen Ebene. Die nächsthöhere Ebene ist also bereits die Ausgangsebene und hat nur einen Knoten.

### 3.3.4 Zur Bedeutung der „Temperatur“

Im Zusammenhang mit dem Training neuronaler Netze wird häufig über die „inverse Temperatur“  $\beta$  gesprochen, die die Ausdehnung des Überganges zwischen den klassifizierten Bereichen kontrolliert:

$$g(a) = \tanh \beta a$$

Sie soll wie die Gewichte und Schwellwerte optimiert werden, am Anfang klein sein (weiche Übergänge) und am Ende des Trainings von selbst groß werden, falls der Eingaberaum klassifizierbar ist. In einigen Modellen ist eine universale, in anderen für jede Ebene eine eigene Temperatur vorgesehen.

Für die hier beschriebenen *feed forward*-Netze ist die Temperatur jedoch kein sinnvoller Parameter, da sie den Raum der Lösungsfunktionen nicht erweitert. Es kann vielmehr

$$\begin{aligned} \tanh(\beta(t + \bar{w} \cdot \bar{x})) &= \tanh(\beta t + \beta \bar{w} \cdot \bar{x}) \\ &= \tanh(\tilde{t} + \tilde{w} \cdot \bar{x}) \end{aligned}$$

auch ohne  $\beta$  ausgedrückt werden. Über den Einfluß dieses redundanten Parameters auf die Konvergenz kann nichts Allgemeingültiges ausgesagt werden. Vergleichende Trainingsdurchläufe ergaben, daß die Einführung der inversen Temperatur die Konvergenz unwesentlich beschleunigte.

Ein anderes Konzept neuronaler *feed forward*-Netze benutzt normierte Eingabedaten und Gewichtsvektoren. In derartigen Netzen wird die Temperatur als weiterer Freiheitsgrad benötigt. Sie haben an jedem Knoten (außer in der Eingangsebene) einen Freiheitsgrad weniger, dafür aber den allen Knoten gemeinsamen Freiheitsgrad  $\beta$ . Veranschaulicht bedeutet das, daß alle Schnitte dieser Netze im Raum gleiche Übergangsbreite haben. Damit ist auch die Menge der möglichen Polyederkanten-Abrundungen eingeschränkt. Ihr Vorteil gegenüber den in dieser Arbeit beschriebenen Netzen ist vermutlich eine bessere Konvergenz, ein direkter Vergleich steht noch aus.

### 3.3.5 Problemquellen

Das Training eines neuronalen Netzes, d. h. die Minimierung seiner nichtlinearen Fehlerfunktion, die eine Verkettung von Summen transzendenter Funktionen ist, bereitet erwartungsgemäß eine Reihe von Schwierigkeiten.

1. Die große Zahl der Dimensionen macht eine Visualisierung der zu minimierenden Funktion unmöglich.
2. Der Minimierungsalgorithmus kann zu einem lokalen Minimum der Fehlerfunktion führen, so daß das absolute Minimum, die optimale Lösung, nicht gefunden wird.

3. Auch wenn das absolute Minimum grundsätzlich erreichbar ist, kann die Form der zu minimierenden Funktion den Konvergenzprozeß derart verlangsamen, daß der Algorithmus das Minimum nicht erreicht.

Hinzu kommen Probleme, die für das Training neuronaler Netze spezifisch sind.

4. Die Geometrie des Netzes ist der Topologie des Eingaberaumes anzupassen, die aber *a priori* unbekannt ist.
5. Die Zahl der erforderlichen Trainingsmuster ist ebenfalls von der Topologie des Eingaberaumes abhängig.
6. Es ist zu vermeiden, daß das Netzwerk Klassifikationsmerkmale lernt, die zwar für die Trainingsmuster, nicht aber für die Gesamtheit der Daten gelten, daß es also statt die allgemeinen Regeln zu erkennen, das Trainingsbeispiel „auswendig lernt“ (Übertraining).

In den weiteren Abschnitten dieses Kapitels wird näher auf diese Probleme eingegangen.

### 3.4 JETNET 1.0

Das Programmpaket JETNET<sup>4</sup> implementiert die in den Abschnitten 3.2 und 3.3 beschriebenen *feed forward*-Netze.

Es erlaubt dem Benutzer die Konstruktion derartiger neuronaler Netzwerke mit bis zu zehn Ebenen, 100 Eingabevariablen, 200 Knoten und 10 000 Gewichten. JETNET berechnet die partiellen Ableitungen der Fehlerfunktion (vgl. Tab. 3.1 für die Bedeutung der Variablen)

$$E(\vec{t}, \vec{w}) = \frac{1}{2} \sum_{i=1}^n (f_i(\vec{t}, \vec{w}) - f_{\text{sol},i}(\vec{t}, \vec{w}))^2,$$

für jedes Trainingsmuster und korrigiert die Gewichte und Schwellwerte wahlweise sofort oder zusammengefaßt für mehrere Muster. Korrigiert wird um den negativen Gradienten der Fehlerfunktion multipliziert mit einer Schrittweite  $\eta$ . Alternativ kann man statt des Gradienten auch einfach die Vorzeichen der partiellen Ableitungen benutzen (*Manhattan-updating*).

Zur Glättung der Parameterkurve läßt sich zu jeder Korrektur noch das  $\alpha$ -fache der letzten Korrektur hinzufügen. Der Trägheitsfaktor  $\alpha$  wird üblicherweise dicht unterhalb von 1 gewählt. Da jede Korrektur dabei

$$\sum_{n=0}^{\infty} \alpha^n = \frac{1}{1 - \alpha}$$

Male zur Anwendung kommt, sollte die Schrittweite entsprechend verkleinert werden.

Als sigmoide Funktionen bietet JETNET zwei verschieden skalierte tanh-Funktionen, eine lineare Funktion und die Exponentialfunktion an. Der Code läßt sich aber leicht auf weitere differenzierbare Schwellfunktionen erweitern.

Um redundante Information (zwei Knoten finden denselben Schnitt) auszumerzen, ist in JETNET eine *pruning*-Option gegeben. Diese wird näher in Abschnitt 3.6.2 beschrieben.

Weitere Details zu dem Programmpaket und die Weiterentwicklung zu JETNET 2.0 lassen sich nachlesen in [13] und [12].

<sup>4</sup>L. Lönnblad, C. Peterson und T. Peterson, Universität Lund, Schweden.

## 3.5 Niedrigdimensionale Anwendungen

### 3.5.1 Das minimale Netz

Um das Verhalten des neuronalen Netzes verstehen zu lernen, wurden zunächst einfache Probeanwendungen genauer untersucht.

Das einfachste neuronale „Netz“, das man sich vorstellen kann, hat zwei Parameter. Es besteht aus einem Eingabeknoten, der über ein Gewicht an einen Ausgabeknoten mit einem Schwellwert angekoppelt ist.

In Abb. 3.4 (a) ist die Geometrie dieses Netzes dargestellt. Sie entspricht der Funktion

$$f(x) = g(t + wx).$$

Als Schwellfunktion  $g$  wurde

$$g(a) = \frac{1}{2}(1 + \tanh a)$$

gewählt. Sie liefert einen weichen Übergang von Werten nahe 0 für  $a < 0$  zu solchen nahe 1 für  $a > 0$ . Die Gesamtfunktion des minimalen neuronalen Netzes lautet dann

$$f(x) = \frac{1}{2}((1 + \tanh(t + wx))). \quad (3.2)$$

Sie hat zwei Parameter, nämlich das Gewicht  $w$  und den Schwellwert  $t$ .

Zu diesem Netz wird ein sehr einfaches „Trainingsmuster“ gewählt. Es soll eine 1 zu 1 klassifizieren. Für zufällig gewählte Anfangswerte  $t = -3$  und  $w = -6$  gelingt dies nicht, denn

$$f(1) = \frac{1}{2}(1 + \tanh(-3 - 6 \cdot 1)) \approx 0.$$

Der Fehler, den das Netz hierbei macht, ist

$$E = (f(1) - f_{\text{soll}}(1))^2 \approx (0 - 1)^2 = 1.$$

In Abb. 3.4 (b) ist der Ausgabefehler des Netzes für den Bereich

$$(w, t) \in [-7, 7] \times [-7, 7]$$

gezeigt. Die Aufgabe, eine 1 zu 1 zu klassifizieren, ist trivial. Es soll deshalb ein zweites „Muster“ hinzukommen, nämlich eine -1, die zu 0 klassifiziert werden soll. Das neue Muster allein würde nun eine Ausgabefehlerfunktion erzeugen, wie sie in Abb. 3.4 (c) zu sehen ist. Für beide Muster zusammen überlagern sich die Fehler, wie in Abb. 3.4 (d) gezeigt. Tabelle 3.2 zeigt die Trainingsbedingungen im Überblick.

An diesem Beispiel wurde nun trainiert. Abb. 3.4 (e) zeigt das Ergebnis. Hier sind die Äquipotentiallinien der Fehlerfunktion, also die Linien gleichen Fehlers bzw. gleicher Höhe der Funktion aus Abb. 3.4 (d), dargestellt. Vor diesem Hintergrund sind die Parameterpunkte des Netzes im Verlauf des Trainings eingetragen. Am bewußt ungünstig gewählten Startpunkt  $(w, t) = (-1, 5, 0, 5)$  (Kreis) ist der Fehler nahe bei 2. Im Verlauf der Minimierung kreuzt die Parameterkurve alle Äquipotentiallinien und erreicht nach ca. 25 Schritten die Ebene mit  $E \approx 0$ . Nach 50 Schritten ist  $(w, t) \approx (1, 5, 0)$ . Der Fehler für das gesamte Beispiel ist dann

$$\begin{aligned} E &= (f(1) - f_{\text{soll}}(1))^2 + (f(-1) - f_{\text{soll}}(-1))^2 \\ &= \left(\frac{1}{2}(1 + \tanh(0 + 1,5 \cdot 1)) - 1\right)^2 + \left(\frac{1}{2}(1 + \tanh(0 + 1,5 \cdot (-1))) - 0\right)^2 \\ &\approx 0,095 \end{aligned}$$

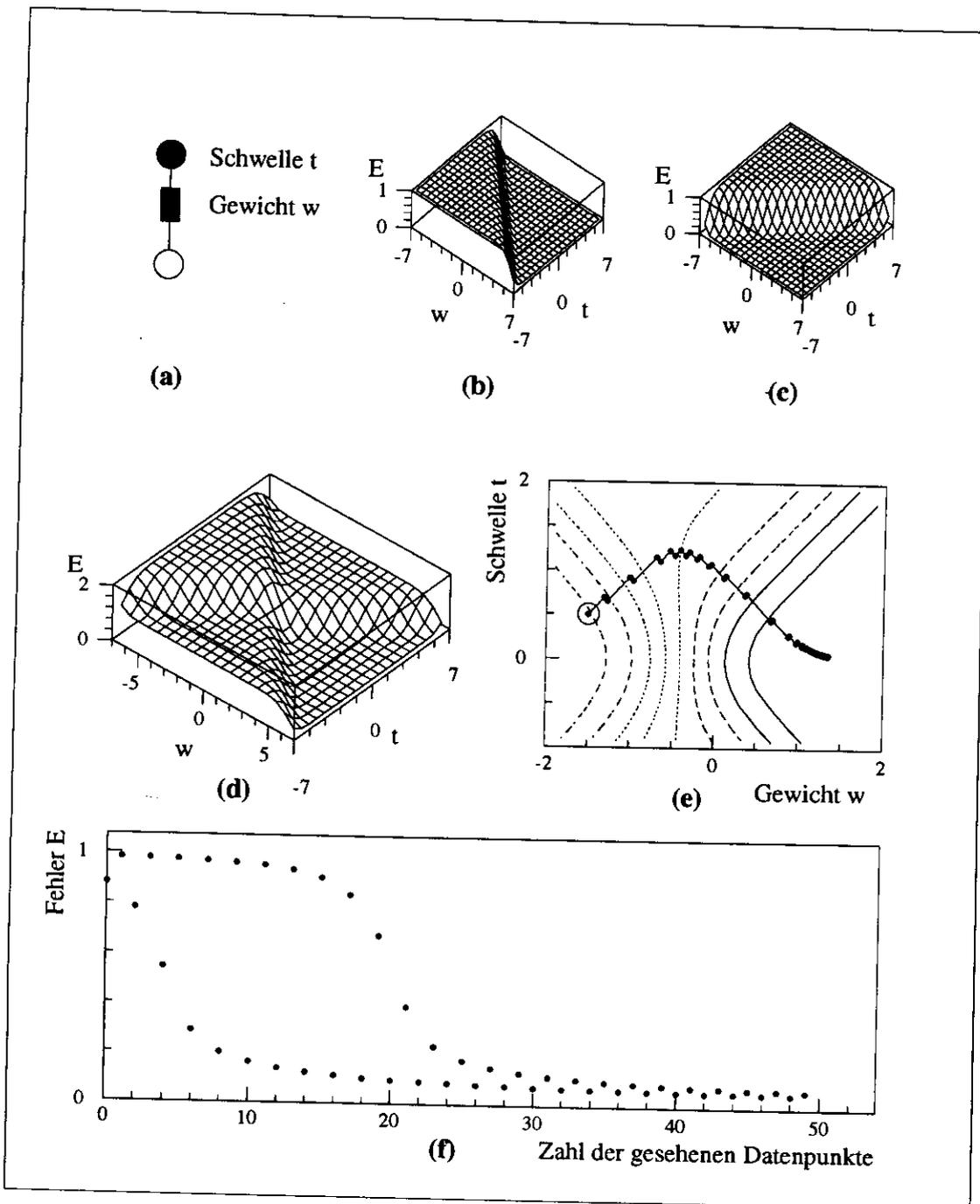


Abbildung 3.4: (a) Geometrie des minimalen neuronalen Netzes. (b) Fehler, den das Netz bei der Klassifikation einer 1 macht, für die eine 1 als Ausgabe gewünscht wird. Der Fehler ist als Funktion des Gewichts  $w$  und des Schwellwertes  $t$  aufgetragen. Die Schwellfunktion ist  $g(a) = \frac{1}{2}(1 + \tanh a)$ . (c) Wie (b) für eine -1 die auf 0 abgebildet werden soll. (d) Summe aus (b) und (c). (e) Äquipotentiallinien der Funktion  $E(w, t)$  aus Abb. (d), und Entwicklung der Netzparameter  $w$  und  $t$  im Verlauf des Trainings (Beginn im Kreis). (f) Fehler  $E$ , den das Netz für jeden Datenpunkt macht. (Nähere Beschreibung: Siehe Text.)

	Eingabe	Erwünschte Ausgabe	Fehler $E(w,t)$
Erstes Muster	$x = 1$	$f(1) = 1$	Abb. 3.4 (b)
Zweites Muster	$x = -1$	$f(-1) = 0$	Abb. 3.4 (c)

Tabelle 3.2: Trainingsbeispiel für das minimale Netz.

Im Vergleich zu einem Fehler von  $E = 0,5$  für zwei halbrichtig gelernte Muster und  $E = 1,0$  für ein richtig und ein falsch gelerntes Muster ist dieser Fehler klein.

Daß die Parameterkurve in Abb. 3.4 (e) so zackig ist, läßt sich folgendermaßen erklären: Das Netz trainiert *abwechselnd* an beiden Mustern 1 und  $-1$ , d. h., es minimiert nicht direkt die Fehlerfunktion  $E(w,t)$  der Abbildungen 3.4 (d) bzw (e), sondern es korrigiert abwechselnd, den Gradienten der Fehlerfunktionen aus Abb. 3.4 (b) und (c) folgend. Die Korrekturen sind orthogonal.

Auch der globale Verlauf der Parameterfunktion läßt sich im Blick auf die Fehlerfunktionen bzgl. der einzelnen Muster erklären. Der Trainingsstartpunkt  $(w,t) = (-1,5,0,5)$  befindet sich nahe der Stufe der Fehlerfunktion zum ersten Muster (Abb. (b)), jedoch weiter entfernt von der Stufe derjenigen zum zweiten Muster (Abb. (c)). Der in Abb. (b) vom Netz gesehene Gradient ist daher größer. Entsprechend sind die von diesem Muster verursachten Korrekturen zunächst größer. Die Parameterkurve in Abb. (e) läuft nach schräg rechts oben. Das erste Muster wird zuerst gelernt. Ist die Stufe aber überwunden, so nimmt der Gradient in dieser Richtung wieder ab, und mit ihm die „Motivation“ des Netzes, den nun sehr kleinen Fehler, den es bei der Klassifizierung *dieses* Datenpunktes macht, weiter zu minimieren. Inzwischen ist die Parameterkurve aber – wenn auch in kleinen Schritten – in die orthogonale Richtung (nach schräg rechts unten) korrigiert worden. Sie hat damit die Kante der Stufe der Fehlerfunktion bzgl. des zweiten Musters erreicht. Jetzt werden die Korrekturen in *dieser* Richtung dominant. Auch dieser Punkt wird gelernt.

In Abb. 3.4 (f) spiegelt sich der Trainingsverlauf wider. Es sind die Fehler dargestellt, die das Netz bei jedem Schritt an dem Muster macht, das gerade gelernt wird. Da die Muster (=Datenpunkte) abwechselnd gelernt werden, repräsentieren die Punkte abwechselnd die Fehler bzgl. der einzelnen Muster, ihre Summe ist dann der Gesamtfehler  $E$ . Man sieht deutlich, daß eins der Muster bereits nach etwa zehn Schritten, das zweite aber erst nach Schritt 25 gelernt wird. Dies läßt sich vermeiden, indem man den Anfangswert in den Übergangsbereich *beider* Trainingsmuster legt. Dazu sind die Gewichte und Schwellen genügend klein zu wählen:

$$\begin{aligned} \sum_j w_{1ij} x_j &\ll 1 \\ \sum_j w_{lij} &\ll 1, & l > 1 \\ t_{li} &\ll 1 \end{aligned}$$

Durch Einsetzen in Gl. 3.2 kann man sich leicht davon überzeugen, daß diese Wahl der Gewichte und Schwellen an jedem Knotenausgang Werte um 0,5, also im Übergangsbereich der Fehlerfunktionen, garantiert. Anders interpretiert erzeugen die kleinen Gewichte eine hohe „Temperatur“, d. h. einen breiten Übergang der sigmoiden Funktion im Eingaberaum, so daß alle Datenpunkte in diesem Übergangsbereich liegen.

Neben der Wahl des richtigen Startpunktes mag die Form der Parameterkurve für die Konvergenz des Trainingsvorgangs kritisch sein. Zur Vermeidung zackiger Parameterkurven werden i. a. die folgenden zwei Methoden eingesetzt:

1. Die Korrektur der Parameter wird jeweils erst nach der Sichtung *beider* Datenpunkte (oder allgemein einer repräsentativen Gruppe von Datenpunkten) vorgenommen.
2. Es wird mit kleinerer Schrittweite korrigiert, dafür wird die Korrektur in allen folgenden Schritten, jeweils um einen Faktor  $\alpha < 1$  gedämpft, noch einmal angewandt.

Abb. 3.5 zeigt die drei Trainingsverläufe im Vergleich. Das einfache Training wurde mit  $\eta = 2$

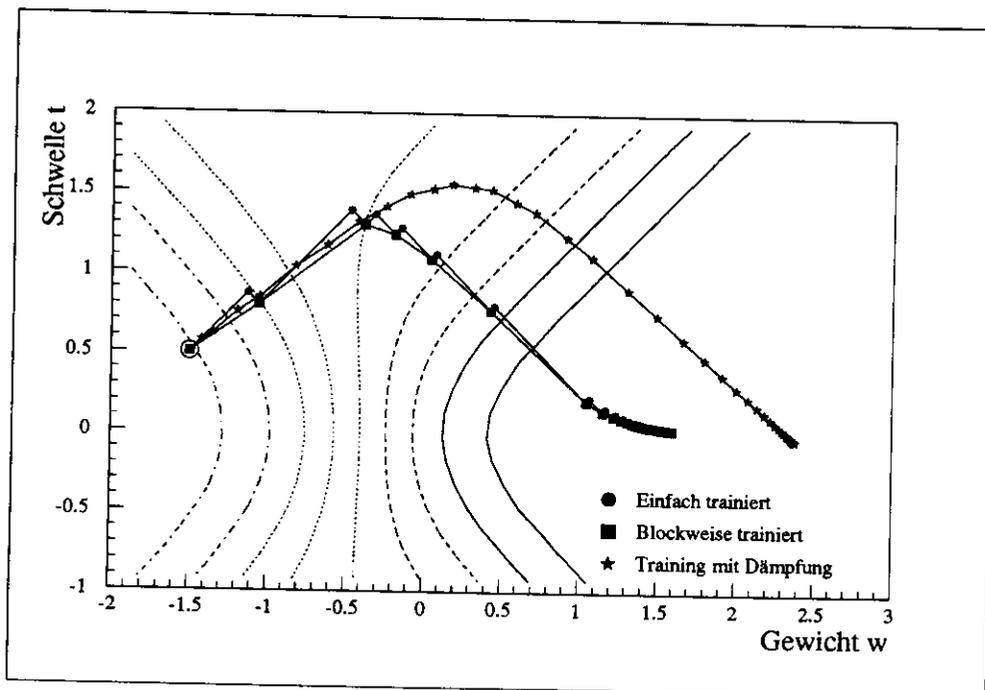


Abbildung 3.5: Parameterkurven für einfaches Training, Training in Blöcken und Training mit Dämpfung.

durchgeführt. Mit derselben Schrittweite wurde auch blockweise trainiert. In diesem Beispiel erzielen das Training in Blöcken und das einfache Training dasselbe Ergebnis. Das liegt daran, daß die Korrekturen, die die beiden Trainingsmuster erzeugen, orthogonal sind: Der Gradient der Fehlerfunktion bzgl. des einen Musters ist stets parallel zur Äquipotentiallinie der Fehlerfunktion bzgl. des anderen Musters. Normalerweise hat das Blocktraining jedoch den Vorteil, daß es ruhiger verläuft, daß die Parameterkurve daher nicht so leicht völlig „aus der Bahn geschossen“ werden kann. Dafür hat das unruhigere, einfache Training den Vorteil, daß es sich gelegentlich aus lokalen Minima befreien kann.

Das Training mit Dämpfung ( $\alpha = 0,8$ ) wurde mit der Schrittweite  $\eta = 0,4$  durchgeführt. Daher beginnt es mit kleineren Schritten. Auf den zweiten Schritt wurde aber bereits das  $\alpha$ -fache des ersten Schrittes addiert, auf den dritten das des zweiten usw. Daher nähert sich die Schrittweite asymptotisch der der anderen Verfahren. In derselben Weise geht das Verfahren mit

Dämpfung auch verspätet in die Kurve und rollt tiefer in das Minimumtal hinein, bevor es wegen des geringen Gradienten zum Stehen kommt. Deshalb nennt man  $\alpha$  auch den „Impulsterm“. Durch das Verschleppen der Wirkung einer Korrektur wird die Parameterkurve geglättet, wenn man

$$\alpha \approx \frac{1}{1 - R}$$

wählt, wobei  $R$  die Anzahl der Muster im Trainingsbeispiel ist. Die Dämpfung hat daneben den Vorteil, das Training gegen kleine Nebenminima robuster zu machen. Dies gilt vor allem bei gleichzeitigem Training in Blöcken. Problematisch ist jedoch, daß man *a priori* nicht weiß, wie  $\alpha$  zu wählen ist, um die Nebenminima zu durchlaufen, ohne gleichzeitig das Hauptminimum zu verfehlen oder zu überrennen.

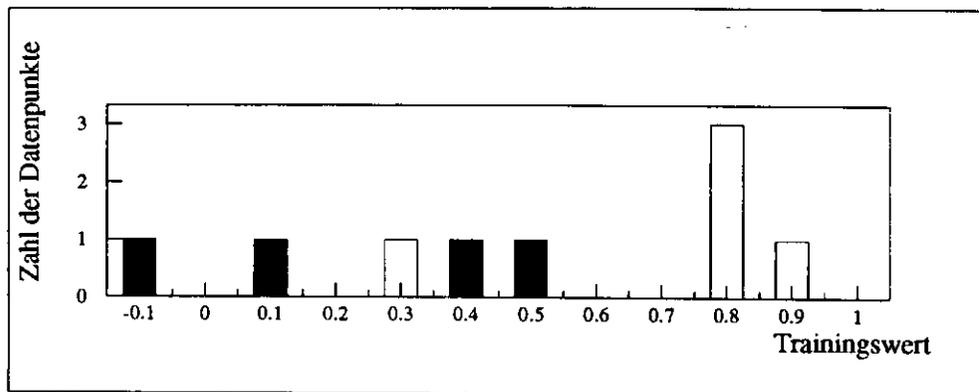


Abbildung 3.6: Trainingsbeispiel: Die schwarz dargestellten Datenpunkte sollen mit einer 1, die weiß dargestellten mit einer 0 klassifiziert werden. Nähere Erklärungen im Text.

Als nächste Stufe wurde dem Netz ein nicht perfekt klassifizierbares Trainingsbeispiel gegeben, das in Abb. 3.6 dargestellt ist. Der Trainingswert (einkomponentiges „Trainingsmuster“) ist die Zahl, die dem Netz als Eingabe gegeben wird. Er liegt in diesem Beispiel zwischen -0,1 und 0,9. Der Trainingswert 0,8 ist dreimal vorhanden. Das Netz soll dahingehend trainiert werden, die schwarz dargestellten Werte mit 1, die weiß dargestellten mit 0 zu klassifizieren.

Da das Netz nur einen Knoten besitzt, kann es nur eine Klassifikationsgrenze setzen. Zieht es die Klassifikationsgrenze bei  $x = 0,2$ , so werden die Datenpunkte bei 0,4 und 0,5 falsch klassifiziert. Zieht es sie jedoch bei 0,65, so wird der Datenpunkt bei 0,3 falsch klassifiziert. Das Problem ist also nicht vollkommen lösbar. Es ist wünschenswert, daß das Netz sich für den Schnitt bei 0,65 entscheidet, da in diesem Fall nur ein Trainingswert falsch klassifiziert wird.

Die drei beschriebenen sowie ein weiteres Verfahren, das blockweise korrigiert und zusätzlich dämpft, wurden mit geeigneten Werten  $\eta$  und  $\alpha$  an dem Problem erprobt.

Aus den Abbn. 3.7 und 3.8 lassen sich die folgenden Schlüsse ziehen:

1. Die Minimumtäler von Trainingsbeispielen mit vielen Mustern können lange, tiefe Rinnen mit steilen Flanken sein. Derartige Funktionen bereiten Minimierungsalgorithmen, die proportional zu und in Richtung auf den negativen Gradienten minimieren, in der Regel große Schwierigkeiten.
2. Erwartungsgemäß ist die Parameterkurve für einfaches Training sehr zackig. Sie läuft überdies fast in die Nebenrinne. Beides würde durch die Wahl einer kleineren Schrittweite zwar

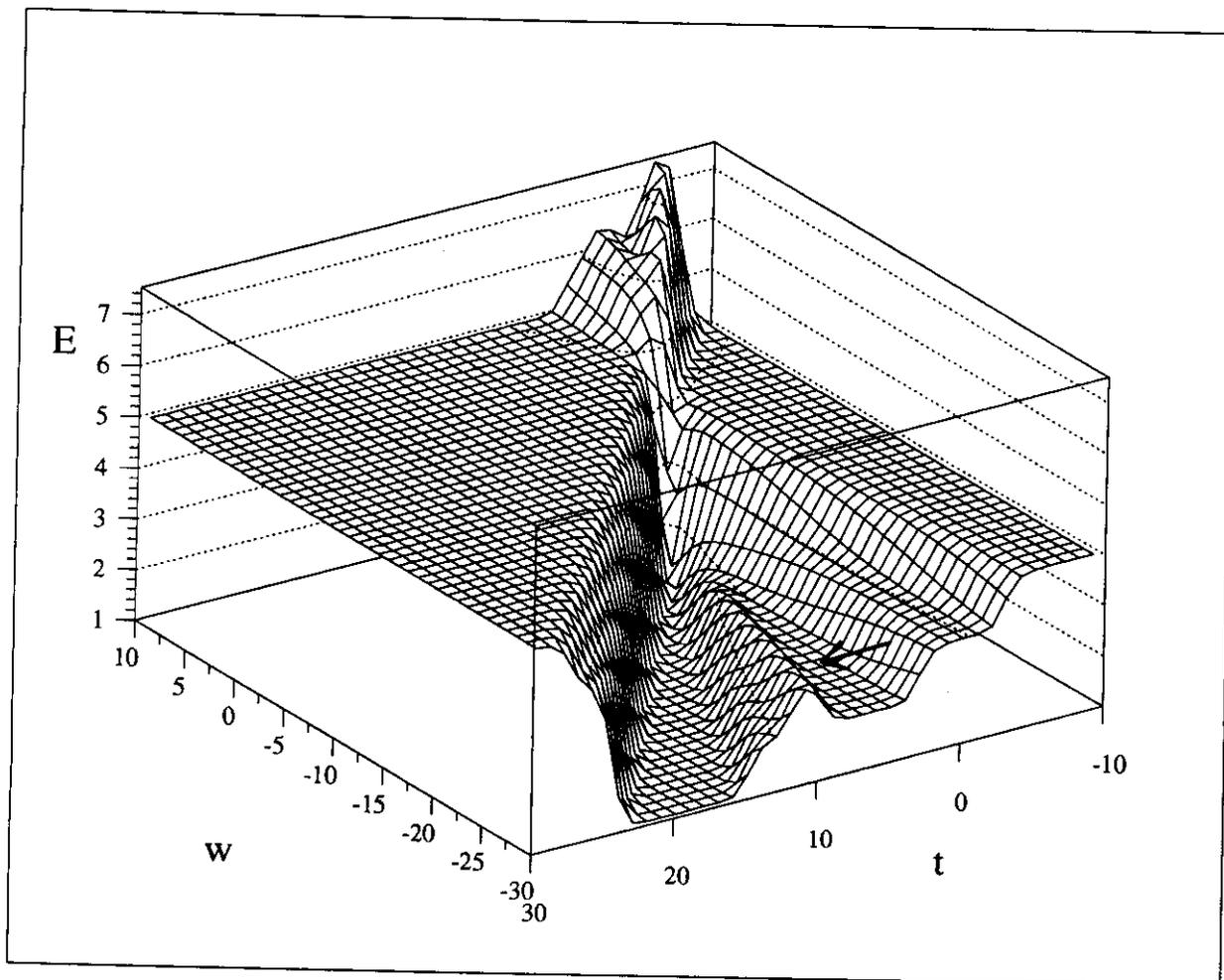


Abbildung 3.7: Gesamtfehlerfunktion. Der Schnitt  $w = -30$  spiegelt die Struktur des Trainingsbeispiels wider, insbesondere die durch den Datenpunkt bei  $x = 0,3$  erzeugte Nebenrinne (Pfeil).

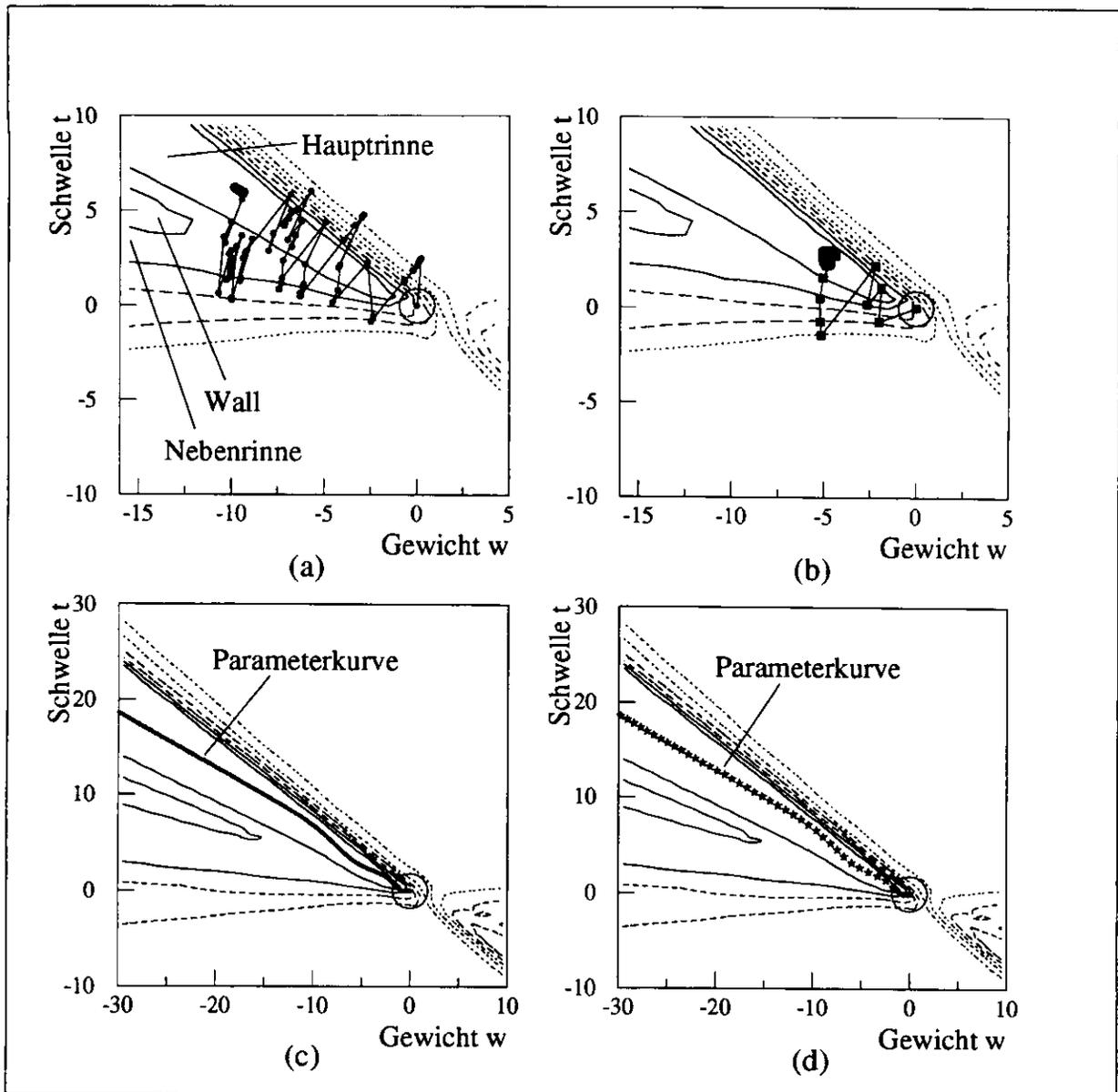


Abbildung 3.8: Parameterkurven unter Benutzung verschiedener Verfahren. Unterlegt sind die Äquipotentiallinien der Gesamtfehlerfunktion aus Abb. 3.7. Die Schrittweite  $\eta$  und ggf. die Dämpfung  $\alpha$  sind jeweils für das Verfahren optimiert. Startpunkt ist der Ursprung:  $w = 0, t = 0$ . (a) Einfaches Training.  $\eta = 10$ . (b) Blocktraining. Das gesamte Trainingsbeispiel wurde in jede Korrektur einbezogen.  $\eta = 3$ . (c) Einfaches Training mit Dämpfung.  $\eta = 0,1, \alpha = 0,999$ . (d) Blocktraining mit Dämpfung.  $\eta = 1, \alpha = 0,99$ .

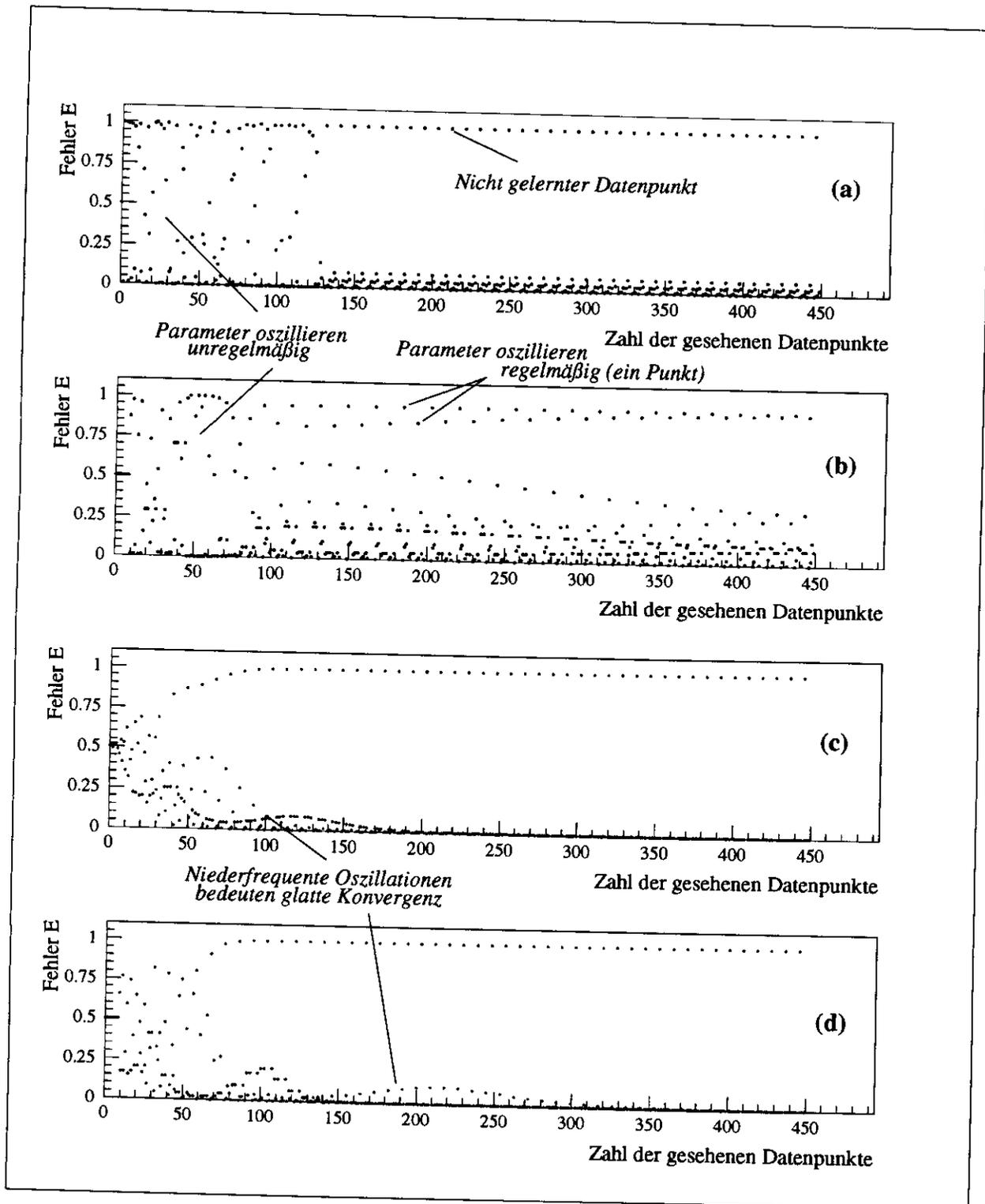


Abbildung 3.9: Fehler der Netzausgabe für jeden Datenpunkt im Verlauf des (a) einfachen Trainings, (b) des Trainings in Blöcken, (c) des Trainings mit Dämpfung und (d) des Blocktrainings mit Dämpfung.

entschärft, die Konvergenzgeschwindigkeit würde sich aber stark verringern. Sobald das Verfahren den Boden der Rinne trifft, wird es derart langsam, daß man die letzten 300 der 450 Parameterpunkte nur als einen sieht.

3. Das Blockverfahren resultiert ebenfalls in einer zackigen Parameterkurve, die wegen der geringeren Anzahl der Korrekturen zudem noch weniger weit konvergiert als die des einfachen Trainings. Warum ist diese Parameterkurve so zackig, obwohl doch über das ganze Trainingsbeispiel gemittelt wird? In den Abbildungen ist die Fehlerfunktion so dargestellt, wie sie von den Blocktrainingsverfahren gesehen wird. Die steilen Flanken der Hauptrinne, besonders die, die durch die drei Datenpunkte bei  $x = 0,8$  erzeugt wird, haben einen großen Gradienten in Querrichtung, der dann aktiv wird, wenn der Parameterpunkt auf die jeweilige Flanke fällt. Korrigiert man mit dem ganzen Trainingsbeispiel, so fällt der Datenpunkt bei  $0,8$  z. B. dreimal ins Gewicht, weil er in dem Beispiel dreimal vorhanden ist. Bei einfachem Training kann er jedoch höchstens einmal pro Schritt aktiv werden. Tut er das, so liegt der nächste Parameterpunkt nicht mehr auf der Flanke. Daher konnte die Schrittweite für einfaches Training ca. dreimal so hoch gewählt werden wie die für das Blocktraining.
4. Mit starker Dämpfung, also kleiner Schrittweite und großem Impulsterm, ließ sich das Problem sowohl mit einfachem Training als auch mit Blocktraining sehr erfolgreich lösen. Das hängt damit zusammen, daß das Minimumtal völlig gerade und unendlich lang ist. Man kann sich leicht überlegen, daß die Fehlerfunktionen aller *feed forward*-Netze ohne verborgene Ebene, die mit der Signumfunktion als Schwellfunktion betrieben werden, strahlenförmig nach außen verlaufende Konturen haben, da eine gleichzeitige Skalierung aller Gewichte und Schwellen mit demselben Faktor den Ausgangswert unverändert läßt. Daher haben alle einlagigen *feed forward*-Netze gerade Minimumtäler, die allenfalls noch durch Nebenrinnen mit wachsendem Abstand zum Ursprung ausdifferenziert werden können.

Die Parameterkurven größerer Netzwerke lassen sich nicht bildlich darstellen. Was man aber darstellen kann, ist der Fehler entlang der Parameterkurve, d. h. der Fehler für jeden Datenpunkt aufgetragen gegen die Trainingsdurchgänge.

Aus diesen läßt sich in begrenztem Maße ebenfalls auf die Parameterkurve schließen. In Abb. 3.9 wird dies für die vier Verfahren vorgeführt. Nach 130 Durchgängen des einfachen Trainings, etwa 300 Durchgängen des Blocktrainings bzw. 90 Durchgängen der Verfahren mit Dämpfung wurden alle Fehler, abgesehen von dem unvermeidlichen, auf unter 0,5 gedrückt.

### 3.5.2 Zweidimensionale Anwendungen

Vor der Anwendung auf die hochdimensionalen Aufgaben der Elektronenidentifikation wurde das Verhalten neuronaler Netze noch an weiteren, in gewissem Maße überschaubaren Übungsproblemen untersucht. Dabei lag das besondere Interesse auf dem Zusammenwirken mehrerer Schnitte bei Netzen mit einer verborgenen Ebene. Die Zahl ihrer Parameter erlaubt keine bildliche Darstellung der Fehlerfunktion mehr. Statt dessen wurde der Verlauf des Trainings anhand der Klassifikationsgrenzen im zweidimensionalen Eingaberaum untersucht.

In Abb. 3.10 sind die Geometrie des Netzes und drei Klassifikationsprobleme dargestellt. Die starken Punkte dieser Binormalverteilungen sollen zu 1, die schwachen zu 0 klassifiziert werden. Im ersten Beispiel sind zwei Schnitte notwendig, um die mittlere Gruppe aus dem Restraum herauszuschälen. Beim zweiten Beispiel könnte dies fast mit einem Schnitt geschehen. Das dritte

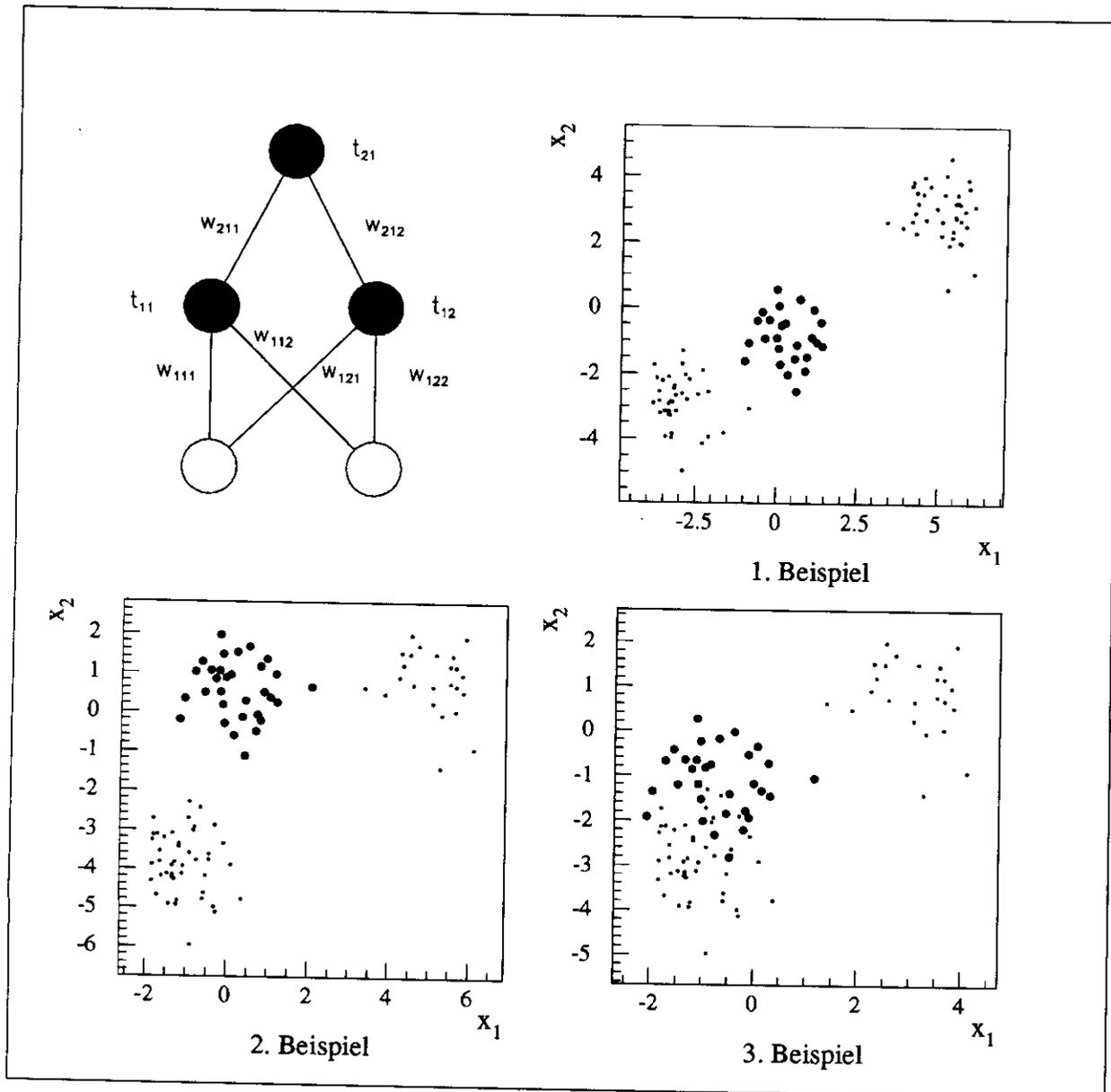


Abbildung 3.10: Neuronales Netz mit verborgener Ebene und drei Beispiele zur Untersuchung seiner Konvergenzeigenschaften.

Beispiel ist wieder eine nicht perfekt lösbare Aufgabe, da sich die Gruppe starker Punkte mit einer Gruppe schwacher Punkte überschneidet.

Die erste dieser Klassifikationsaufgaben wurde durch Training mit Dämpfung in 2000 Schritten gelöst. Abb. 3.11 zeigt die Entwicklung der Fehler, die resultierenden Klassifikationsgrenzen und die Ausgabe des Netzes für die verschiedenen Punktgruppen. Die Sammlung der Punkte Fehlerdiagramm (obere Abbildung) zu drei schwingenden Linien rührt daher, daß das Netz den einzelnen Punkten jeder Gruppe jeweils fast gleiche Werte zuordnet.<sup>5</sup>

Für das zweite Beispiel benötigte das Training mit Dämpfung etwa 25 000 Schritte, also wesentlich länger. Aus der Entwicklung der Fehler (Abb. 3.12) ersieht man, daß die Parameter über einen großen Teil des Trainings in einer festen Position verharren. Abb. 3.12 (b) zeigt die Lage der Klassifikationsgrenzen in diesem Zustand. Beide Grenzen stellen sich so ein, als sollten sie die Aufgabe jeweils allein lösen. Sie enthalten beide dieselbe Information. Dieses Problem wird im nächsten Abschnitt eingehender untersucht.

Das dritte Beispiel stellt ebenfalls hohe Anforderungen. Wiederum ist es mit einem Schnitt fast so gut lösbar wie mit zweien, beide Lösungen klassifizieren aber nicht vollkommen, weil sich in diesem Beispiel Gruppen verschieden zu klassifizierender Daten stark überlappen. Die im nächsten Abschnitt untersuchten Methoden zum Abbau redundanter Schnitte zugunsten kooperativer Schnitte wurden auch auf dieses Beispiel angewandt. In der Elektronenidentifikation werden Daten wie diese erwartet. Das Training am dritten Beispiel und seine Lösung sind in Abb. 3.16 dargestellt.

## 3.6 Problem der redundanten Schnitte

### 3.6.1 Einleitung

Die *feed forward*-Netze, trainiert mit dem *back propagation*-Algorithmus, verlangsamen ihre Konvergenz stark mit dem kleiner werdenden Fehler. Dies ist zwar einerseits ein Vorteil, andererseits wird dadurch in einigen Problemen das absolute Minimum der Fehlerfunktion einfach deshalb nicht erreicht, weil man nicht lange genug trainieren kann. Im folgenden werden einige Methoden vorgestellt, die das Konvergenzverhalten des Trainings in seiner letzten Phase verbessern sollen. Als konkreter Gegenstand der Untersuchung dient das zweite Beispiel aus dem vorhergehenden Abschnitt.

### 3.6.2 Pruning

Die im Programmpaket JETNET 1.0 für die Unterdrückung von Gewichten und Schwellen überflüssiger Knoten vorgesehene Methode ist das *pruning*. Dem Fehler zur Berechnung der Korrekturen wird ein Term hinzugefügt, der für große Gewichte gleich einem vorgegebenen Parameter  $\lambda$  ist und für kleine Gewichte verschwindet. Die Bedeutung von groß und klein wird durch einen weiteren Parameter  $w_0$  definiert.

$$E \rightarrow E + \lambda \left( \sum_{ij} \frac{(w_{ij}/w_0)^2}{1 + (w_{ij}/w_0)^2} + \sum_i \frac{(t_i/t_0)^2}{1 + (t_i/t_0)^2} \right) \quad (3.3)$$

<sup>5</sup>Eine detaillierte Beschreibung des Trainingsvorgangs und ausführliche Interpretation des Fehlerdiagramms findet sich in Anhang A.2.

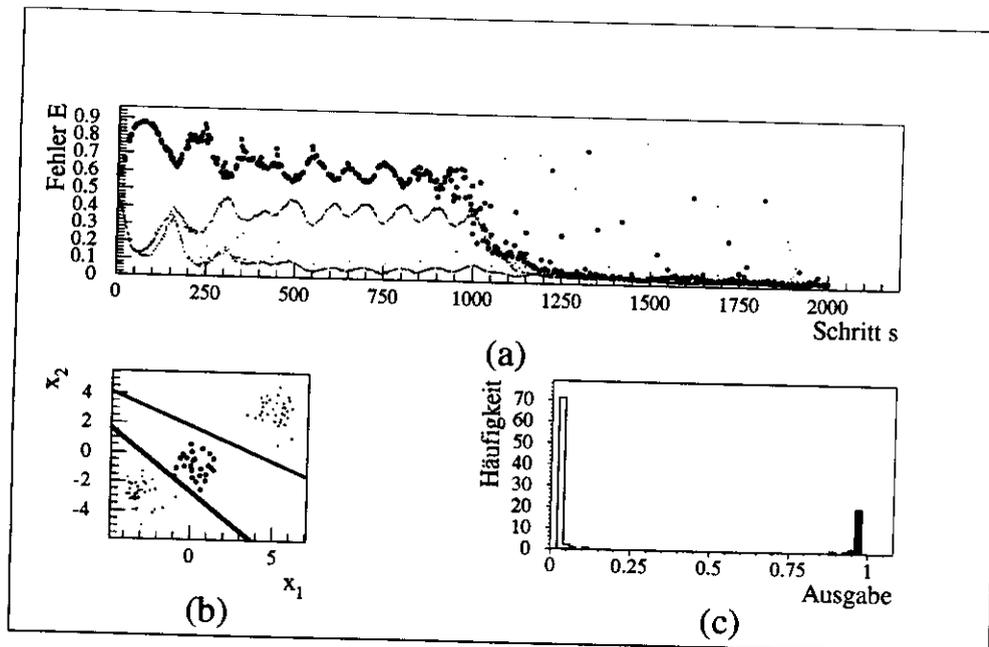


Abbildung 3.11: Beispiel 1. (a) Fehler des Netzes für jeden Datenpunkt im Verlauf des Trainings. (b) Klassifikationsgrenzen im Eingaberaum nach 2000 Schritten mit  $\eta = 0,01$  und  $\alpha = 0,99$ . (c) Ausgabewerte des Netzes für Eingabemuster der starken Punkte (schwarz) und der schwachen (weiß).

Dieser zusätzliche Fehlerterm bewirkt eine Tendenz zu kleinen Gewichten. Dem wirkt i. a. der eigentliche Fehler des Trainings entgegen. Es stellt sich ein Gleichgewicht ein. Hat man jedoch zwei Knoten, die dieselbe Information erhalten, so können die Gewichte eines dieser Knoten abgebaut werden, ohne daß der Fehler, den das Netz am Trainingsbeispiel macht, dabei größer wird. Der zusätzliche Fehlerterm gibt nun dem Netz den nötigen Antrieb, dies tatsächlich zu vollziehen. Der von der Klassifikationsaufgabe entbundene Knoten kann eine neue Aufgabe übernehmen. Falls die Klassifikation noch nicht perfekt ist, wird er zu einer den anderen Knoten ergänzenden Funktion streben. Dadurch entsteht die Zusammenarbeit der beiden Knoten.

Ein Nachteil des *prunings* ist, daß nicht mehr die reine Fehlerfunktion optimiert wird, sondern die modifizierte aus Gl. 3.3. Daher sind die Lösungen dieses Verfahrens niemals optimal, sondern die Gewichte sind etwas zu klein, die Übergänge also zu breit, was aber in der Praxis nicht von wesentlicher Bedeutung ist. In Abb. 3.13 (a) erkennt man einen solchen Effekt ganz am Ende des Trainings. Der Fehler steigt zum Schluß leicht an.

Das Beispiel 2 aus Abschnitt 3.5.2 wurde mit denselben Parametern wie oben erneut trainiert, diesmal jedoch mit *pruning*, und zwar mit  $\lambda = 10^{-3}$  und  $w_0 = 1$ . In Abb. 3.13 ist der Verlauf der Fehlerfunktion dargestellt. In der Phase des Trainings, in der die beiden Schnitte sich auf dieselbe Position eingestellt haben (Abb. 3.12 (b)), also etwa ab Schritt 12 000 kommt das *pruning* zur Wirkung und kürzt die Konvergenzdauer von weiteren 12 000 auf etwa 3000 ab. Die Gesamtkonvergenzdauer verkürzt sich damit auf die Hälfte. Am Ende des Trainings wächst der Fehler noch ein wenig. Hier zeigt sich, daß der Gleichgewichtszustand nach erfolgtem Training mit *pruning* nicht im Minimum der Fehlerfunktion liegt, wie bereits oben erklärt.

Das Training an Beispiel 3 ließ sich mit dem *pruning* nicht verbessern, da die Trennung der

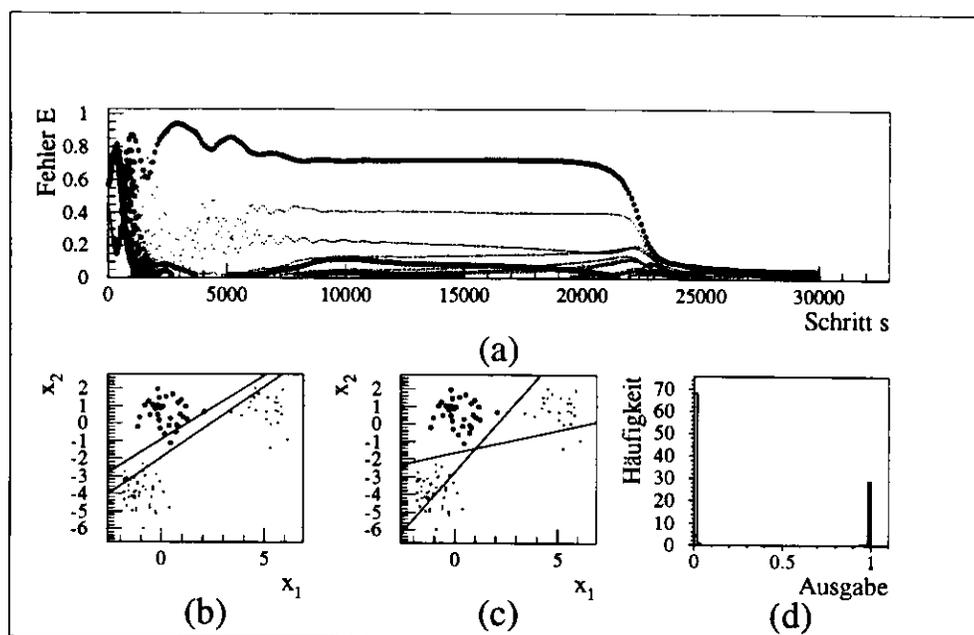


Abbildung 3.12: Beispiel 2. (a) Fehler des Netzes für jeden Datenpunkt im Verlauf des Trainings. (b) Klassifikationsgrenzen im Eingaberaum nach 15 000 und (c) 30 000 Schritten mit  $\eta = 3 \cdot 10^{-4}$  und  $\alpha = 0,999$ . (d) Ausgabewerte des Netzes für Eingabemuster der starken Punkte (schwarz) und der schwachen (weiß).

Schnitte in diesem Fall bereits gelernt wurde, bevor die Schnitte wirklich aufeinanderlagen, also dieselbe Information lieferten.

### 3.6.3 Die Methode des künstlichen Rauschens

Die Phase schlechter Konvergenz im Beispiel 2 des Abschnitts 3.5.2 läßt sich durch Einfügen eines künstlichen Rauschens ebenfalls verkürzen. Die Idee dabei ist, daß das Rauschen in den Parametern zufällig zu einer Konstellation in den Gewichten zwischen der verborgenen Ebene und der Ausgangsebene führt, die eine UND-Funktion der Entscheidungen der beiden Schnitte in der verborgenen Ebene realisiert. Eine solche Konstellation ermöglicht den beiden Knoten eine Zusammenarbeit, d. h., Punkte, die von einem Knoten fälschlich mit 1 bewertet werden, führen nicht zu einer Korrektur der Eingangsgewichte dieses Knotens, wenn der andere Knoten sie korrekt mit 0 bewertet hat, da am Ausgang in diesem Falle auch eine (korrekte)  $1 \wedge 0 = 0$  anliegt. Die Zusammenarbeit der Knoten bestärkt dann die Notwendigkeit der UND-Funktion am Ausgang, so daß der Zustand sich gegen das Rauschen stabilisiert.

Diese Methode garantiert natürlich nicht den Erfolg. Besonders in hochdimensionalen Problemen ist es unwahrscheinlich, zufällig eine bessere Konstellation zu finden und diese nicht durch weiteres Rauschen wieder zu verlieren.

In neuere, hier nicht untersuchte Versionen von JETNET wurde diese Methode unter dem Namen „Langevin-Rauschen“ aufgenommen. In dieser einfachen Anwendung wurde nicht das Rauschen durch einen Zufallsgenerator erzeugt, sondern die Parameterkurve einfach durch aggressiveres Training zackiger gemacht ( $\eta = 0,03$  und  $\alpha = 0,98$ ).

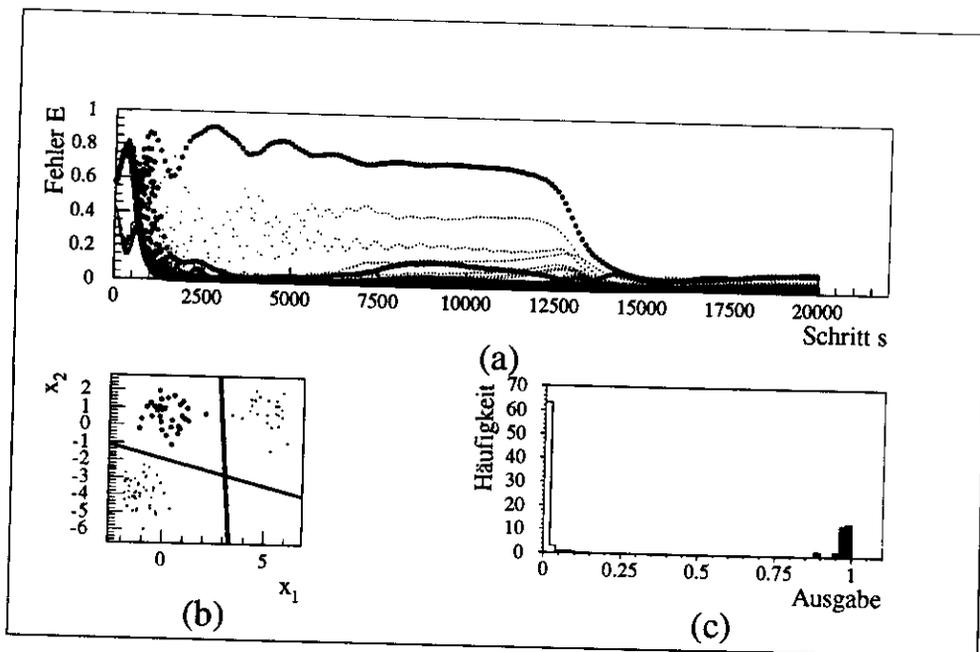


Abbildung 3.13: Beispiel 2. (a) Fehler für jeden einzelnen Datenpunkt im Verlauf des Trainings mit *pruning*. (b) Klassifikationsgrenzen im Eingaberaum nach 20 000 Schritten mit  $\eta = 3 \cdot 10^{-4}$ ,  $\alpha = 0,999$ ,  $\lambda = 10^{-3}$  und  $w_0 = 1$ . (c) Ausgabewerte des Netzes für Eingabemuster der starken Punkte (schwarz) und der schwachen (weiß).

Wie Abb. 3.14 zeigt, ist dieses Verfahren in unserem Fall extrem erfolgreich. Es hat aber i. a. die Nachteile, daß man im Fehlerdiagramm keine Strukturen mehr erkennen kann, also blind trainiert, und daß die Parameter  $\eta$  und  $\alpha$  geeignet erraten werden müssen, damit die Parameterkurve nicht völlig aus der Bahn schießt.

In Beispiel 3 ist die Schrittweite bereits voll ausgereizt. Es wurden größere Schrittweiten, auch bei anderem  $\alpha$ , ausprobiert. Sie führten aber nur zu einer Verschlechterung der Konvergenzeigenschaften. Es kann jedoch nicht ausgeschlossen werden, daß ganz bestimmte Schrittweiten zufällig zum richtigen Ergebnis führen.

### 3.6.4 Die Methode der kritischen Datenpunkte

Das Problem des Beispiels 2 im Vergleich zum Beispiel 1 (beide aus Abschnitt 3.5.2) ist, daß es schon mit *einem* Schnitt recht gut klassifiziert werden kann. Die Fehlerfunktion hat dementsprechend etwa nach Schritt 10 000 des Trainings (vgl. Abb. 3.12 (a)), wo die Knoten der verborgenen Ebene dieselbe Information tragen, einen sehr geringen Gradienten. Dies kam im ersten Beispiel mit kollinear verteilter Gruppen nicht vor. Die Idee der hier beschriebenen Methode besteht nun darin, daß man aus dem Beispiel alle Datenpunkte ausblendet, die bereits gut klassifiziert werden. Die übrigen, „kritischen Datenpunkte“ liegen dann einigermaßen kollinear, was zu einer Verbesserung der Konvergenz führen sollte.

Es wurde noch einmal mit den Parametern des vorhergehenden Abschnitts trainiert, aber diesmal nur an Punkten mit einem Fehler von über 0,1. Das neuronale Netz wurde insgesamt 964mal ausgewertet, dabei wurden jedoch nur 288 Korrekturen durchgeführt. Danach waren

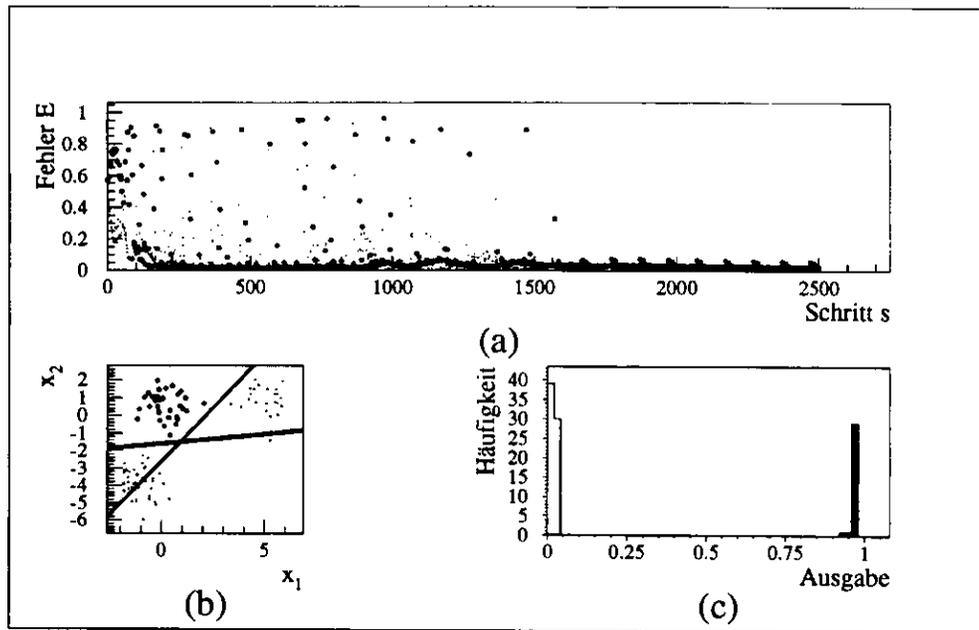


Abbildung 3.14: Beispiel 2. (a) Fehler für jeden einzelnen Datenpunkt im Verlauf des Trainings mit großer Schrittweite. (b) Klassifikationsgrenzen im Eingaberaum nach 2500 Schritten mit  $\eta = 0,03$  und  $\alpha = 0,98$ . (c) Ausgabewerte des Netzes für Eingabemuster der starken Punkte (schwarz) und der schwachen (weiß).

alle Punkte besser als 0,1 klassifiziert. Eine Auswertung entspricht etwa dem halben Aufwand einer Korrektur. Damit ist der Aufwand gegenüber dem ersten Trainingsdurchgang um den Faktor 50 verbessert worden. In Abb. 3.15 sind der Verlauf des Fehlers für die in der Korrektur berücksichtigten Datenpunkte und einige Momentaufnahmen des Trainings festgehalten.

Im Beispiel 3 konnte mit dieser Methode eine Verbesserung von etwa 35% erzielt werden.

### 3.6.5 Zusammenfassung

Das beschriebene *feed forward*-Netz wurde auf verschiedene zweidimensionale Beispielprobleme angewandt. Dabei wurde besonders auf das Problem redundanter Schnitte eingegangen. Beispiel 3 enthält als zusätzliche Schwierigkeit sich überlappende Verteilungen, die nicht vollständig trennbar sind. Dies ist auch in den Daten des Kap. 4 der Fall. Alle drei Aufgaben wurden gelöst. Verschiedene Trainingsmethoden kamen dabei zu demselben Ergebnis, jedoch mit stark unterschiedlichem Aufwand.

Die Tabelle 3.3 zeigt einen Vergleich der Verfahren bzgl. der Zahl der benötigten Lernschritte.

Mit Hilfe der zweidimensionalen Beispielaufgaben der Abb. 3.10 konnte die Wirkungsweise der *feed forward*-Netze trotz ihrer großen Parameterzahl auf anschauliche Weise untersucht werden. Dabei stellte sich heraus, daß das einfache Training mit Dämpfung vielfältige Aufgaben zu lösen vermag, daß jedoch in speziellen Fällen (abhängig von der Topologie des Problems) angepaßte Methoden die Konvergenz des Trainingsvorganges extrem stark beschleunigen können.

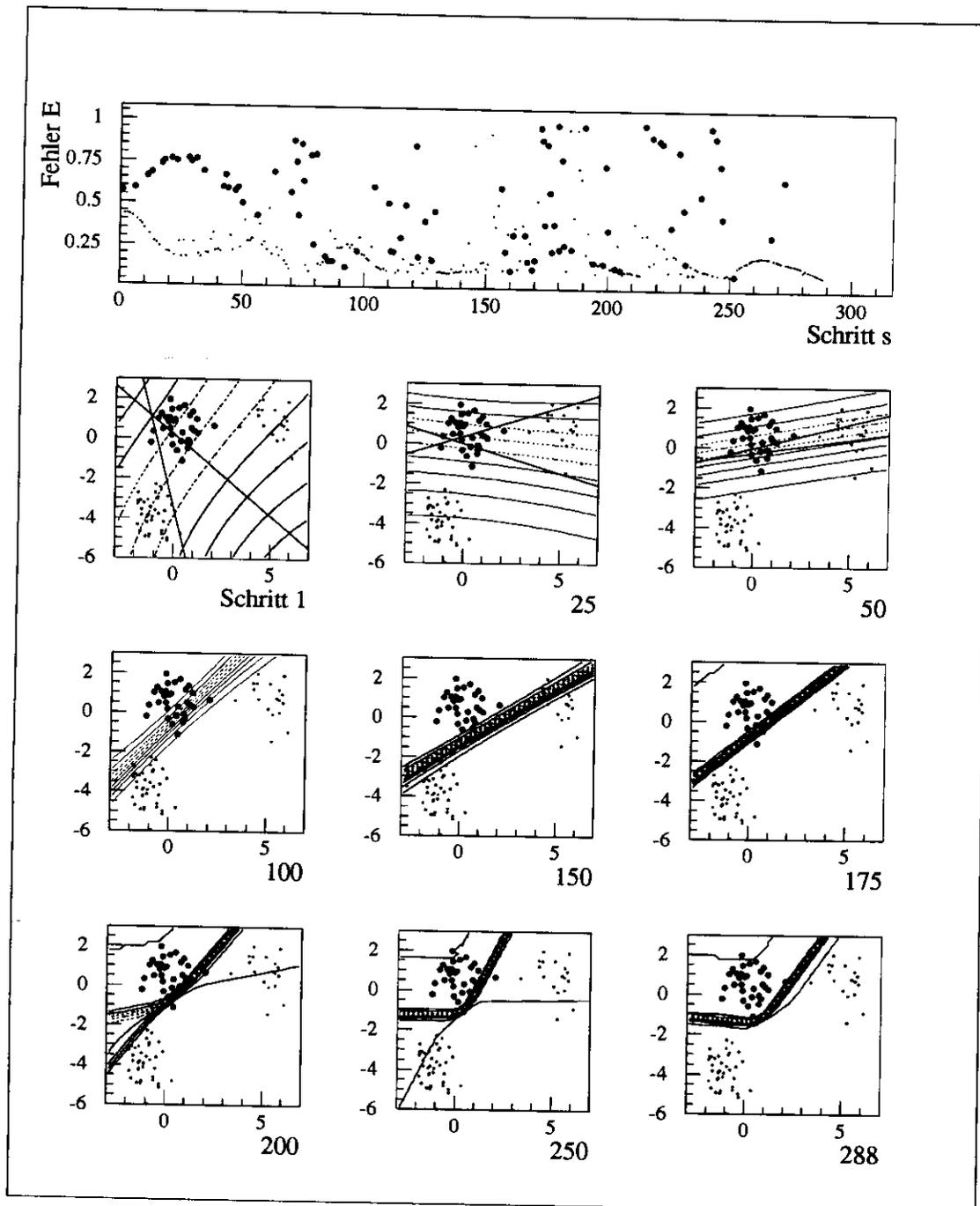


Abbildung 3.15: Beispiel 2. Fehler für jeden einzelnen Datenpunkt im Verlauf des Trainings an kritischen Datenpunkten (oben) und Momentaufnahmen der Klassifikationsfunktion im Eingangsraum. Die ersten drei Aufnahmen zeigen neben den Äquipotentiallinien der tatsächlichen Ausgangsfunktion die Klassifikationsgrenzen der beiden verborgenen Knoten. Parameter:  $\eta = 0,03$ ,  $\alpha = 0,98$ ,  $E_{krit} = 0,1$ .

In dieser Bildfolge erkennt man deutlich, wie der Übergang in der Ausgangsfunktion sich mit fortschreitendem Training (infolge der wachsenden Gewichte) zusammenzieht, also schärfer wird.

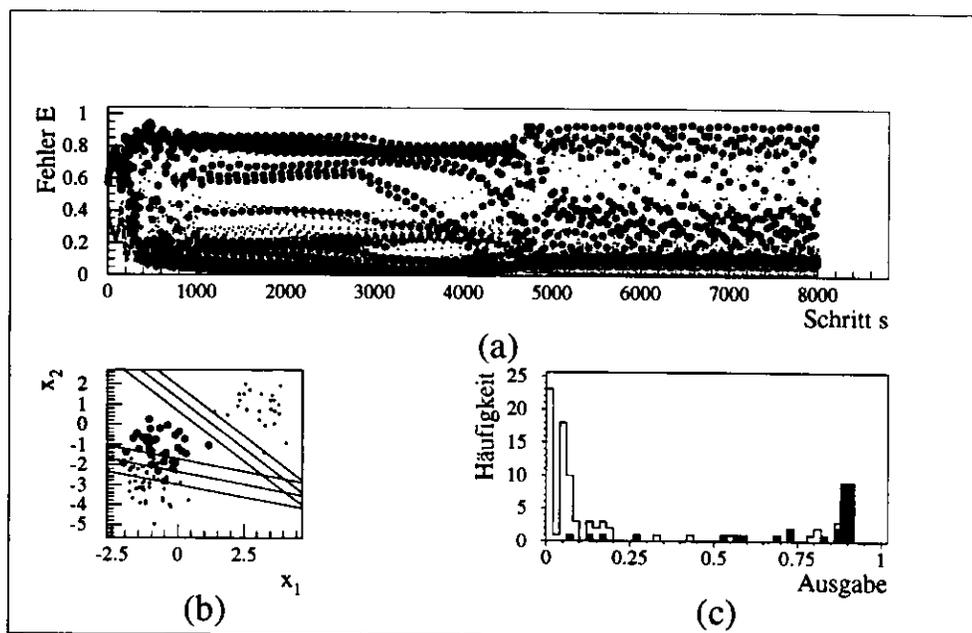


Abbildung 3.16: Beispiel 3. (a) Fehler des Netzes für jeden Datenpunkt im Verlauf des Trainings. (b) Klassifikationsgrenzen im Eingaberaum nach 8000 Schritten mit  $\eta = 0,008$  und  $\alpha = 0,99$ . Die Breite der Übergänge wird durch die eingezeichneten Parallelen verdeutlicht. (c) Ausgabewerte des Netzes für Eingabemuster der starken Punkte (schwarz) und der schwachen (weiß).

Verfahren	Problem	Parameter		Zahl der Schritte
Dämpfung	Beispiel 1	$\eta = 0,01$	$\alpha = 0,99$	2000
Dämpfung	Beispiel 2	$\eta = 3 \cdot 10^{-4}$	$\alpha = 0,999$	25000
<i>pruning</i>	Beispiel 2	$\eta = 3 \cdot 10^{-4}$	$\alpha = 0,999$	15000
Rauschen	Beispiel 2	$\eta = 0,03$	$\alpha = 0,98$	1750
Kritische DP	Beispiel 2	$\eta = 0,03$	$\alpha = 0,98$	600
Dämpfung	Beispiel 3	$\eta = 0,008$	$\alpha = 0,99$	8000
Kritische DP	Beispiel 3	$\eta = 0,008$	$\alpha = 0,99$	5000

Tabelle 3.3: Zahl der benötigten Schritte der verschiedenen Verfahren zur Lösung der drei Beispielaufgaben.



## Kapitel 4

# Identifikation niederenergetischer Elektronen

### 4.1 Einleitung

Die existierenden ZEUS-Elektronenfinder zeigen Schwächen bei der Identifikation von Teilchen mit Impulsen unterhalb von  $5 \text{ GeV}/c$ . Der Hadron-Elektron-Separator soll in Verbindung mit dem Kalorimeter dazu eingesetzt werden, Elektronen mit Impulsen unterhalb dieser Schwelle zu identifizieren.

Dazu wird zunächst ein Datensatz ausgewählt, der außer den HES- und Kalorimeterinformationen auch Angaben über die Identität der Teilchen (Elektron oder Hadron) enthält. Er wird aufgeteilt in einen Trainingsdatensatz, mit dem das Netzwerk trainiert werden soll und einen Testdatensatz, an dem es hinterher geprüft werden soll. Danach werden die Eingabevariablen für das neuronale Netzwerk bestimmt und berechnet. Schließlich wird das Netzwerk trainiert und anhand des Testdatensatzes beurteilt.

Die zur Beurteilung der Leistung eines Elektronenfinders benötigten Informationen über die wahre Identität der Teilchen sind von den Informationen, die dem Elektronenfinder zur Verfügung gestellt werden, streng zu trennen.

### 4.2 Der Trainings- und Testdatensatz

#### 4.2.1 Datenquellen

Zum Training und Test neuronaler Netzwerke wird ein Trainings- und Testdatensatz benötigt. Dieser Datensatz besteht aus einer Anzahl von Teilchen, deren Signale und wahre Identitäten bekannt sein müssen. Das Hauptproblem des Trainings neuronaler Netzwerke besteht darin, ein solches Trainingsbeispiel zu konstruieren. Im folgenden sind einige mögliche Quellen für Trainingsmuster und ihre Eigenschaften aufgeführt. In dieser Arbeit wird aber nur *ein* Datensatz benutzt.

**Testdaten.** Für das ZEUS-Kalorimeter existieren Daten aus Teststrahlexperimenten am CERN. Die Identität der einzelnen Teilchen kann zuverlässig bestimmt werden. Das letzte Prototypkalorimeter, bestehend aus vier FCAL-Modulen, die jeweils aus vier Türmen aufgebaut waren, stellt ein sehr gutes Modell des in ZEUS eingebauten Kalorimeters dar. Die Energien, Einfallswinkel

und Einfallspalten lassen sich aber nur in Schritten variieren, die tote Materie kann nicht im Detail modelliert werden, und es wird nur der Einfluß einzelner Teilchen auf den Detektor untersucht. Elektronen in Teilchenbündeln (*jets*) treten nicht auf.<sup>1</sup>

Monte-Carlo-Simulation. Der ZEUS-Detektor wird in allen Einzelheiten durch eine Simulation beschrieben. Mit Hilfe simulierter Ereignisse läßt sich relativ einfach ein Datensatz zum Training und Testen eines neuronalen Netzwerkes konstruieren. Das Verhalten eines Monte-Carlo-trainierten Netzwerkes ist empfindlich abhängig von der Qualität der Simulation. Ist die Software zum Training eines Netzwerkes vorhanden, so läßt sich das Training eines neuronalen Netzes sehr schnell durchführen. Es ist daher ohne viel Aufwand möglich, einen Elektronenfinder dieser Art mit jeder Verbesserung der Simulation auf dem neuesten Stand zu halten.

Daten. Auch aus den Daten selbst lassen sich Trainingsbeispiele konstruieren. Dazu die folgenden Ideen:

1. Für Elektronen der tiefunelastischen Streuung läßt sich Ort und Energie in ausgesuchten Ereignissen aus den hadronischen Variablen bestimmen. Wird dadurch ein isoliertes Teilchen klar als Elektron ausgezeichnet, so kann dieses zum Training eines Elektronenfinders für isolierte Teilchen benutzt werden.
2. Mit einem existierenden Elektronenfinder für isolierte Teilchen kann man Daten für das Training eines Elektronenfinders konstruieren, der auch in *jets* arbeitet, indem man verschiedene Ereignisse überlagert. Diese Methode beruht auf der Linearität des HES und des Kalorimeters, die gut erfüllt ist.
3. Eine Detektorkomponente, deren Information nicht von dem Elektronenfinder benutzt wird, kann als Referenz für den Teilchentyp genutzt werden, falls ihre Effizienz und Diskriminanz ausreicht und genügend genau bekannt ist.

Die Datenverfahren besitzen den großen Vorteil, daß sie den Detektor perfekt „simulieren“. Ihr Problem ist jedoch die Bestimmung des Anteils von Untergrund bei den Elektronen und umgekehrt des Anteils von Elektronen bei den Hadronen.

In dieser Arbeit wird das Datenverfahren unter Punkt 3 benutzt. Die zentrale Spurkammer (CTD) zeichnet gewisse Spuren als Elektronen aus. Es handelt sich um Elektron-Positron-Paare aus der Photonkonversion.

#### 4.2.2 Elektronen aus Photonkonversion

Für das Training an niederenergetischen Elektronen bietet sich der vor kurzem von L. Koepke und R. van Woudenberg geschriebene ZEUS-Photonkonversions-Finder CONVERT2 an [9]. Er benutzt ausschließlich die Spurinformatoren der CTD und liefert Elektronen und Positronen mit relativ hoher Reinheit. Die Konversionsspuren werden erkannt als Paare gegensätzlich gekrümmter Spuren, die sich in einem Punkt berühren und darin dieselbe Richtung aufweisen. Dabei werden zunächst die folgenden Schritte angewandt:

- Der Polarwinkel der Spuren darf sich um maximal  $5,7^\circ$  unterscheiden.
- Am „Berührungspunkt“ der beiden extrapolierten Spuren darf ihr Abstand nicht größer sein als  $4\text{ mm}$ .

<sup>1</sup>Mir ist nicht bekannt, ob Datensätze mit Hadronen und Elektronen von weniger als  $5\text{ GeV}$  existieren, für die im Kalorimeterprototyp auch der HES aktiv war.

Spuren, die diese Schnitte passieren werden zu Paaren zusammengefaßt. Die Prioritäten werden dabei gemäß einer dimensionslosen Variablen verteilt, die sich wie folgt berechnet:

$$DISTB = \sqrt{\left(\frac{XYDIF}{\sigma_{xy}}\right)^2 + \left(\frac{THDIF}{\sigma_{\theta}}\right)^2}$$

XYDIF und THDIF sind der Abstand der „Berührungspunkte“ und die Differenz der Polarwinkel der beiden Spuren,  $\sigma_{xy}$  und  $\sigma_{\theta}$  sind Auflösungen, die durch Anpassungen von Gaußfunktionen an die Verteilungen von XYDIF und THDIF gewonnen wurden. Die Anpassungen wurden für Spuren mit und ohne VXD-Treffer, für Spuren mit und ohne Stereo-Treffer (zweite Superlage in der CTD) sowie für Daten und MC getrennt durchgeführt. Ein kleines DISTB entspricht einer guten Zuordnung.

An die gefundenen Spurpaare und die zugehörigen Ereignisse wurden schließlich weitere Schnitte angelegt:

- Der Hauptvertex darf in  $z$  nicht weiter als 50 cm vom nominellen Wechselwirkungspunkt entfernt sein.
- Der Winkel zwischen der Richtung des rekonstruierten Photons und der Verbindungslinie zum Vertex darf nicht größer sein als  $18,2^\circ$ . In der  $xy$ -Ebene soll er  $11,5^\circ$  nicht übersteigen.
- Die Spuren dürfen keine Treffer in der Vertexkammer (VXD) haben, falls der Berührungspunkt außerhalb der Kammer rekonstruiert wurde.

Der vorliegende Datensatz enthält 5290 Photoproduktionsereignisse mit Konversionselektronen, von denen jeweils mindestens eines einen Impuls von über 700 MeV besitzt.

Für weitere Informationen über den Konversionsfinder sei auf [9] verwiesen.

Zum Training werden neben Elektronen auch Hadronen benötigt. Analog zu den Konversionselektronen existieren auch saubere Untergrund-Datensätze. So würden sich beispielsweise sichere Pionsignale aus dem bereits bei ZEUS existierenden Datensatz elastischer  $\rho$ -Ereignisse extrahieren lassen. Die geringe Teilchendichte könnte erhöht werden durch Überlagerung mehrerer Ereignisse. Dieses Verfahren wurde jedoch zunächst nicht durchgeführt. Statt dessen wurden einfach alle diejenigen Spuren, die vom Konversionsfinder abgelehnt wurden, als „Hadronen“ in den Datensatz aufgenommen. Unter diesen Spuren befindet sich ein (wenn auch geringer) Teil Elektronen, auf den korrigiert werden muß.

### 4.2.3 Datenaufbereitung im Überblick

Aus dem im letzten Kapitel beschriebenen Photoproduktionsdatensatz mit Konversionselektronen wurde der Trainings- und Testdatensatz extrahiert. Die im folgenden aufgezählten Schritte dazu werden später ausführlich erläutert.

1. Benachbarte Kalorimeterzellen mit einer Energie von über 100 MeV wurden zu Gruppen zusammengefaßt. Diese Gruppen nennen sich *Islands*. Benachbarte HES-Dioden mit Signalen über 0,6 mip wurden zu sogenannten *Isles* formiert.
2. Die Energie und Position der *Isles* und *Islands* wurden bestimmt.
3. Mit Hilfe des Konversionsfinders wurde die Identität der Spuren definiert: Elektron oder Untergrund.

4. *Isles*, Spuren und *Islands* wurden einander zugeordnet.
5. Die in Abschn. 4.2.9 aufgelisteten Schnitte wurden ausgeführt.
6. Die Netz-Eingabevariablen wurden berechnet (vgl. Abschnitt 4.3).

#### 4.2.4 Gruppenbildung: Isles und Islands

Die genaue Beschreibung des *Island*- und *Isle*-Algorithmus ist sehr technisch und wurde daher in den Anhang B.1 verlegt. Im wesentlichen handelt es sich dabei um eine Zusammenfassung von Kalorimeter- bzw. HES-Signalen um Zellen mit lokal höchstem Signal.

#### 4.2.5 Energie- und Positionsbestimmung

Die Energie der *Isles* und *Islands* ist einfach die Summe der in ihren Dioden bzw. Zellen deponierten Energien. Für *Isles* wird die Energie in *mip* angegeben<sup>2</sup>, für *Islands* in *GeV*.

Die Positionsbestimmung wurde für *Islands* mit der Routine ELECPO durchgeführt. Darin wird die Position als energiegewichtetes Mittel der Schauerpositionen in den einzelnen Zellen berechnet. Die *x*-Positionen in den einzelnen Zellen werden aus dem Verhältnis der Signale in den beiden Photoverstärkern, die jede Zelle von rechts und links her auslesen, berechnet. Das etwas aufwendige Verfahren wurde durch Vergleich mit dem HES gestimmt. Für die *y*-Positionen stehen nur die Mitten der Zellen zur Verfügung. Zur Vermeidung systematischer Strukturen wurden die *y*-Positionen über die Zellbreite zufallsverteilt. Die Routine ELECPO wurde eigentlich für Elektronen mit Energien über 2 *GeV* geschrieben. Tatsächlich ist eine Ortsbestimmung bei oftmals nur einer angeschlagenen Zelle sehr ungenau. Sie wurde in dieser Untersuchung daher nicht benutzt.<sup>3</sup>

Der Algorithmus zur Positionsbestimmung für *Isles* stammt aus dem Rekonstruktionsprogramm *hesrecon*. Die Position wird als Mittel der Positionen der einzelnen Dioden berechnet, gewichtet mit dem Logarithmus ihres Signals. Dieses Verfahren hat sich gegenüber der linearen Gewichtung als genauer herausgestellt, ist aber in gewissem Sinne unbefriedigend, weil der Logarithmus auf eine willkürlich gewählte Einheit bezogen werden muß. Zusätzlich müssen zu kleine Signalbeiträge weggelassen werden, um negative Gewichte zu vermeiden. Einheit und gleichzeitig untere Signalgrenze ist das  $e^{-5}$ -fache des *Isle*-Signals.

#### 4.2.6 Isle-Island- und Spur-Island-Zuordnung

*Isles* wurden dem Kalorimeterturm zugeordnet, in dem ihre Diode maximalen Signals liegt. Einige Dioden liegen auf der Grenze zwischen zwei Türmen. In diesem Fall wurde die *Isle* dem Turm größerer Energie zugerechnet. Enthielt der Turm einen Energieeintrag, so ist er Teil eines *Islands*. Diesem *Island* wurde die *Isle* zugeordnet. Enthielt der Turm keine Energie, so blieb die *Isle* unzugeordnet.

Spuren wurden ebenfalls zunächst dem Turm zugeordnet, auf den sie zeigten. Enthielt dieser keine Energie, so wurden sie dem nächstliegenden Turm mit Energieeintrag zugeordnet, sofern

<sup>2</sup>Ein *mip* ist die mittlere Energie, die ein minimal ionisierendes Teilchen beim Durchgang durch die Diode hinterläßt.

<sup>3</sup>Eine Ausnahme bildet hier allerdings die Variable *Island*-Abstand (siehe Abschnitt 4.3), die sich aber als unwichtig herausgestellt hat.

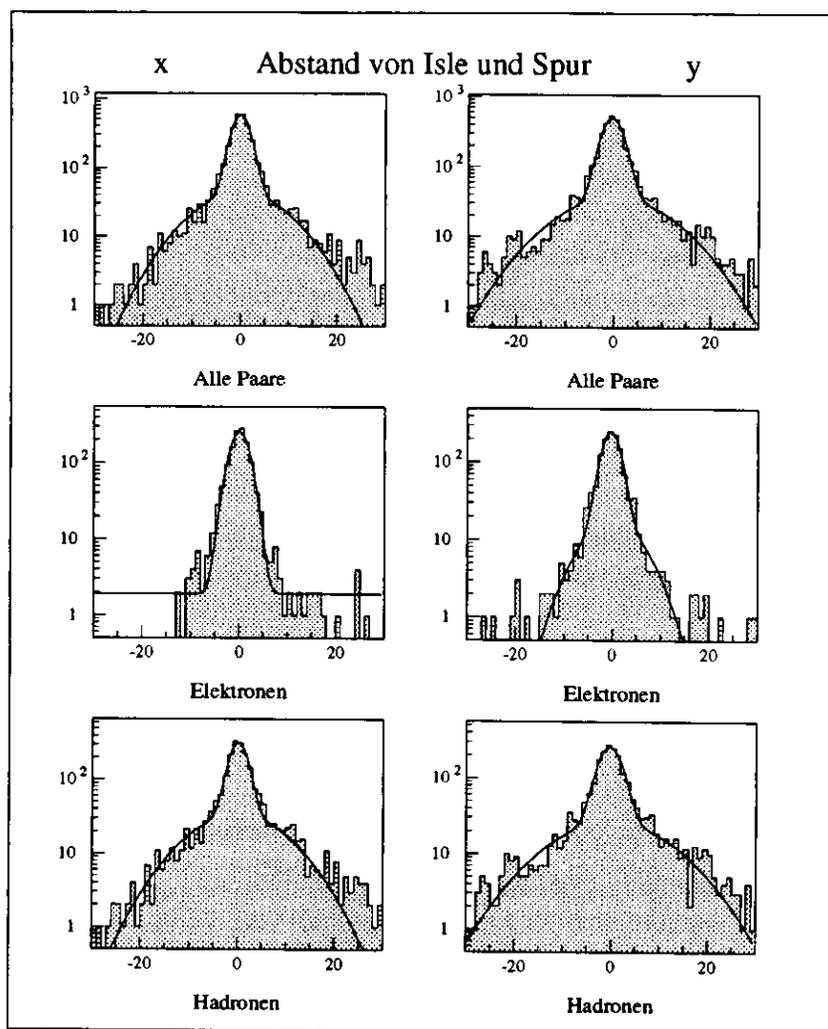


Abbildung 4.1: Anpassung von Gaußfunktionen an die Abstandsverteilungen von Spuren und Isles in den Koordinaten  $x$  und  $y$  für alle Daten (oben), Elektronen (Mitte) und Hadronen (unten).

der Schnittpunkt ihrer Extrapolation mit dem HES vom Rande dieses Turmes nicht mehr als 5 cm entfernt war. Sie wurden wie die Isles dem Island zugeordnet, zu dem auch der Turm gehörte.

#### 4.2.7 Identifikation der Konversionselektronen

Mit Hilfe des Konversionsfinders wurden in den Daten die Konversionselektronen identifiziert. Die durch die in Abschnitt 4.2.2 aufgeführten Bedingungen definierten Konversionskandidaten wurden noch durch die folgenden Kriterien in drei Klassen eingeteilt: Elektron, Hadron, Unbekannt.

1. Spurpaare, für die DISTB (vgl. Abschn. 4.2.2) größer ist als 25, werden als Hadronen angesehen. Diese Bedingung trifft allerdings nur ganz wenige Paare.
2. Für Spurpaare, deren Berührungspunkt klar außerhalb von festem Material liegen, also im Strahlrohr ( $R < 7,5\text{cm}$ ) oder in der CTD ( $R > 18\text{cm}$ ), wurde die Identität offengelassen.

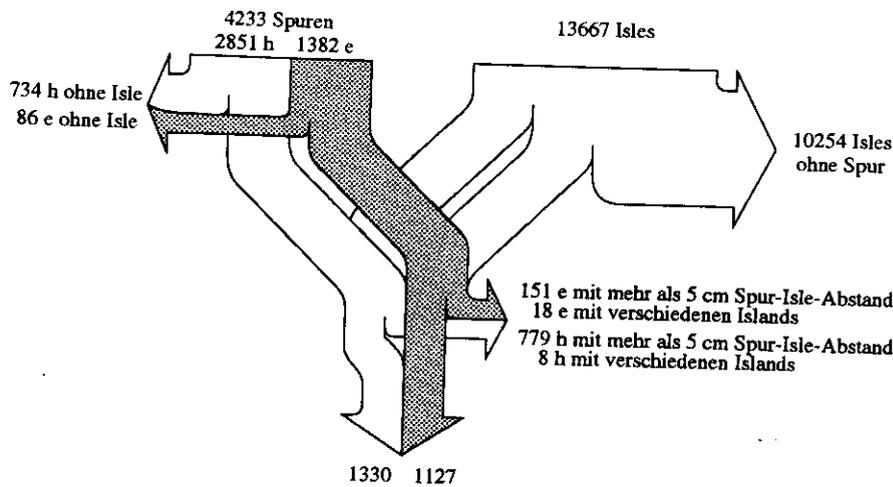


Abbildung 4.2: Wirkung der Zuordnungsschnitte auf die Zahl der Spuren und Isles. Es wurden nur Spuren und Isles berücksichtigt, die die Schnitte aus Abschnitt 4.2.9 passierten.

Vor allem der Schnitt gegen Paare mit Berührungspunkt im Strahlrohr schneidet viel Untergrund weg, da hier die Dichte der (extrapolierten) Spuren besonders hoch ist.

#### 4.2.8 Spur-Isle-Zuordnung

Spuren und Isles wurden auf die folgende Weise einander zugeordnet.

1. Zu jedem Spur-HES-Schnittpunkt wurden alle Isles im Umkreis von 30 cm gesucht und nach ihrer Nähe zur Spur sortiert. Die fünf nächstliegenden Isles wurden samt ihren Abständen in einer Liste festgehalten.
2. Beginnend mit dem kleinsten Spur-Isle-Abstand des gesamten Ereignisses wurden die Spuren und Isles „locker“ zugeordnet. Dabei kann es vorkommen, daß eine Spur nicht der nächstliegenden Isle zugeordnet wird, weil diese bereits einer anderen, ihr näheren Spur zugeordnet wurde.
3. Die so gewonnenen Spur-Isle-Paare wurden endgültig zugeordnet, falls sie die folgenden Bedingungen erfüllten:
  - Die HES-Isle ist horizontal und vertikal maximal 5 cm von der Spur entfernt.
  - Spur und Isle wurden demselben Island zugeordnet.

Diese Zuordnung wurde nicht in einem Stück vorgenommen. Zunächst wurde nur locker zugeordnet (Punkt 1 und 2), dann wurden die Schnitte angelegt, die im nächsten Abschnitt beschrieben werden, und erst danach wurde die endgültige Zuordnung vorgenommen. Dadurch haben nur die ausgewählten Spuren und Isles Einfluß auf die Kriterien unter Punkt 3 genommen.

Zur Bestimmung der maximalen Abstände in  $x$  und  $y$  wurden die Verteilungen dieser Abstände parametrisiert. Es wurde davon ausgegangen, daß die Abstände von Isles, die durch geladene Teilchen verursacht wurden, zu ihren Spuren gaußverteilt sind. Zusätzlich wurde ein gleichverteilter Untergrund angenommen. Aufgrund der Zuordnung der nächstliegenden Isle zur Spur sind die

Abstände zwischen Spur und locker zugeordneter Untergrund-*Isle* in den Koordinaten  $x$  und  $y$  ebenfalls gaußverteilt, und zwar zentriert um 0. Die anzupassenden Funktionen sind also Summen aus einer Gaußverteilung mit freiem Erwartungswert und einer unverschobenen Gaußverteilung. Durch geeignete Startwerte für die Fehlerminimierung kann man erreichen, daß die unverschobene Gaußfunktion an den Untergrund angepaßt wird.

Die Ergebnisse der Anpassung sind in Tab. 4.1 zusammengefaßt. Es ist die Auflösung der Zuordnung, die mittlere Verschiebung der extrapolierten Spur gegen den HES und die Standardabweichung der Verteilung der Abstände Spur/Untergrund-*Isle* aufgeführt.

Art der Spur	$\sigma_{Signal}(cm)$	$\mu_{Signal}(cm)$	$\sigma_{Untergrund}(cm)$
Alle ( $x$ )	$1,78 \pm 0,05$	$0,09 \pm 0,05$	$8,6 \pm 0,4$
Alle ( $y$ )	$2,07 \pm 0,06$	$-0,29 \pm 0,05$	$10,3 \pm 0,6$
Elektronen ( $x$ )	$1,90 \pm 0,05$	$0,02 \pm 0,06$	—
Elektronen ( $y$ )	$1,77 \pm 0,10$	$-0,36 \pm 0,06$	$5,5 \pm 0,9$
Hadronen ( $x$ )	$1,78 \pm 0,08$	$0,13 \pm 0,07$	$8,9 \pm 0,5$
Hadronen ( $y$ )	$2,24 \pm 0,09$	$-0,23 \pm 0,08$	$10,9 \pm 0,7$

Tabelle 4.1: Doppel-Gauß-Anpassung an die Abstände zwischen Spur und *Isle*

Aus der Standardabweichung der Verteilung der Abstände Spur/Untergrund-*Isle* läßt sich eine Untergrund-*Isle*-Dichte berechnen. Tut man dies, so erhält man eine Dichte, die um einen Faktor vier zu groß ist. Der Untergrund ist also nicht gleichverteilt.

Das hängt damit zusammen, daß *Isles* ohne zugehöriges *Island* nicht mitbetrachtet werden (siehe nächster Abschnitt). Die Verteilung wird damit schmaler und täuscht eine zu große Dichte vor. Trotzdem sind die Anpassungen gut genug, um den Schnitt bei 5 cm zu setzen und den Anteil fälschlich zugeordneter Spur-*Isle*-Paare und den von wahren Paaren, die nicht zugeordnet wurden, zu berechnen. Diese Rechnung wurde zwar durchgeführt, ist aber für diese Arbeit nicht von Bedeutung.

Die Wirkung der Zuordnung auf die Zahl der Spuren und *Isles* ist in Abb. 4.2 dargestellt. Die Effizienz der Zuordnung ist das Verhältnis der Zahl der Elektronenspuren vorher und nachher. Die Diskriminanz ist der Anteil der Hadronenspuren, die durch die Zuordnung ausgeschieden wurden. Die Werte von Effizienz und Diskriminanz betragen

$$\begin{aligned}
 E_Z &= \frac{1127}{1382} = (81,5 \pm 1,8)\% \\
 D_Z &= \frac{2851 - 1330}{2851} = (53,3 \pm 1,6)\%
 \end{aligned}
 \tag{4.1}$$

Da durch die Spur-*Isle*-Zuordnung bereits einige Elektronen und ein Großteil der Hadronen ausscheiden, beeinflußt sie die Effizienz und Diskriminanz des Elektronenfinders. Der Spur-*Isle*-Zuordnungsalgorithmus ist daher als *Teil des Elektronenfinders* anzusehen. Die Effizienz  $E_Z$  stellt eine Grundeffizienz des Elektronenfinders dar. Sie kann nicht mehr überschritten werden, da die durch die Zuordnung abgelehnten Elektronen nicht mehr zurückgeholt werden können. Andererseits können die mangels einer zugeordneten *Isle* abgelehnten Hadronen nicht wieder als Elektronen mißinterpretiert werden. Damit sind 53,3% der Hadronen bereits durch die Forderung

einer zugeordneten *Isle* diskriminiert. Der Elektronenfinder kann also in der Diskriminanz nur noch besser sein als  $D_z$ .

#### 4.2.9 Spurselektion

An die Spuren aus dem in Abschnitt 4.2.2 beschriebenen Datensatz wurden die folgenden Schnitte angelegt:

**Geometrischer Schnitt:** Die Extrapolation der Spur muß auf die Fiduzialfläche zeigen. Die Fiduzialfläche ist definiert als der Teil der HES-Ebene, dessen Dioden 1993 erfolgreich ausgelesen werden konnten, abzüglich eines Sicherheitsrandes von 5 cm. Sie ist abgebildet in Abb. 4.3

**Impulsschnitt:** Für die Spuren wurde ein Impuls von  $p > 500 \text{ MeV}/c$  verlangt.

**Qualitätsschnitte:** Die Spur mußte vom Hauptvertex<sup>4</sup> stammen (außer Konversionselektronen) und mindestens einen Treffer in der zweiten Lage der Spurkammer haben (Stereo-Information).

**Kalorimeterschnitt:** Die Spur soll sich einem *Island* zuordnen lassen (vgl. Abschnitt 4.2.6).

**Identitätsschnitt:** Es muß bekannt sein, ob die Spur aus einer Photonkonversion stammt oder nicht (vgl. Abschn. 4.2.7).

Die genannten Schnitte beeinflussen natürlich die Eigenschaften des Datensatzes. So werden beispielsweise Teilchen mit sehr kleinem und sehr großem Transversalimpuls unterdrückt. Alle betrachteten Spuren liegen im Winkelbereich  $135^\circ < \theta < 165^\circ$ , wobei  $\theta$  der Polarwinkel des zylindrischen ZEUS-Koordinatensystems ist.

An die *Isles* wurden die folgenden zwei Schnitte gelegt:

**Geometrischer Schnitt:** Ihre Position muß sich in der Fiduzialfläche befinden.

**Kalorimeterschnitt:** Die *Isle* läßt sich einem *Island* zuordnen.

Der zweite Schnitt eliminiert den größten Teil aller *Isles* (Rauschen).

In der Phase der Spurselektion waren den Spuren bereits *Isles* „locker“ zugeordnet (vgl. Abschn. 4.2.8). Mußte ein Partner eines Spur-*Isle*-Paares verworfen werden, so wurde der andere ebenfalls weggelassen.

Abb. 4.4 zeigt die Zahl der Spuren und *Isles*, die durch die einzelnen Schnitte ausgesondert wurden. Selbstverständlich wirken der Impuls-, der Qualitäts- und der Identitätsschnitt zunächst auf *Spuren*. Mit ihnen wurden aber ihre „locker“ zugeordneten Partner-*Isles* verworfen. Der Kalorimeterschnitt wirkt dagegen auch auf *Isles* ohne Spur.

Das Impulsspektrum derjenigen Spuren, die den geometrischen Schnitt passierten, vor und nach den Qualitätsschnitten sowie nach dem Impulsschnitt ist in Abb. 4.5 dargestellt.

Die nach diesen Schnitten verbleibenden Spuren und *Isles* stellen die Grundmenge dar, auf die sich die Effizienzen und Diskriminanzen der weiteren Algorithmen beziehen. Die Abbildung 4.3 zeigt die Schnittpunkte aller selektierten Spuren mit der HES-Ebene und die Positionen der ihnen zugeordneten *Isles*.

<sup>4</sup>Für den Datensatz existieren allerdings keine Informationen über die Vertexanbindung, daher wurde in der Variablen  $\cot \theta$  geschnitten, die festhält, ob sich die Spur an ihrem Entstehungspunkt von der Strahlachse entfernte oder sich ihr näherte.

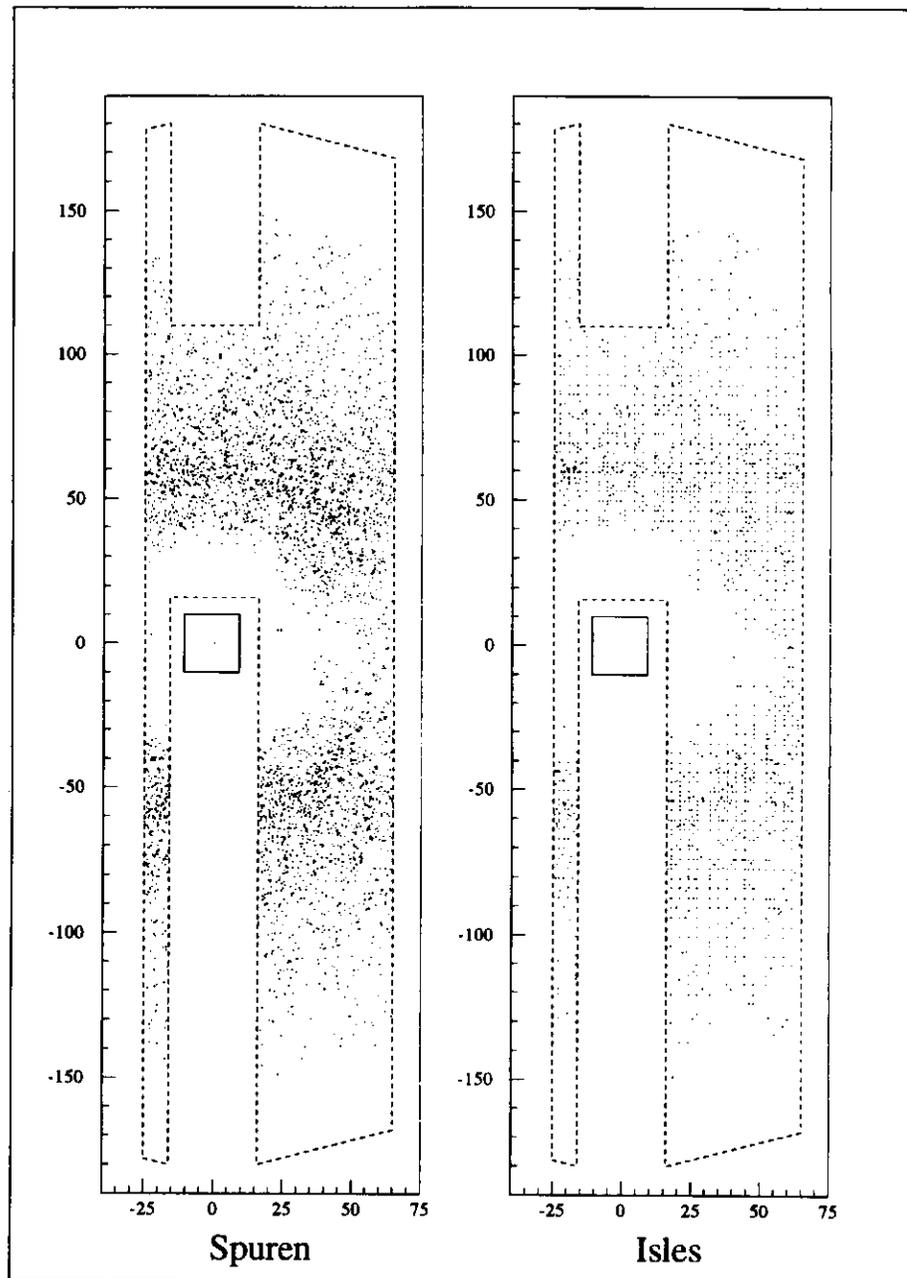


Abbildung 4.3: Schnittpunkte aller selektierten Spuren mit der HES-Ebene und Positionen der ihnen zugeordneten Isles. Die Struktur der Isle-Positionen spiegelt die Anordnung der HES-Dioden wider.

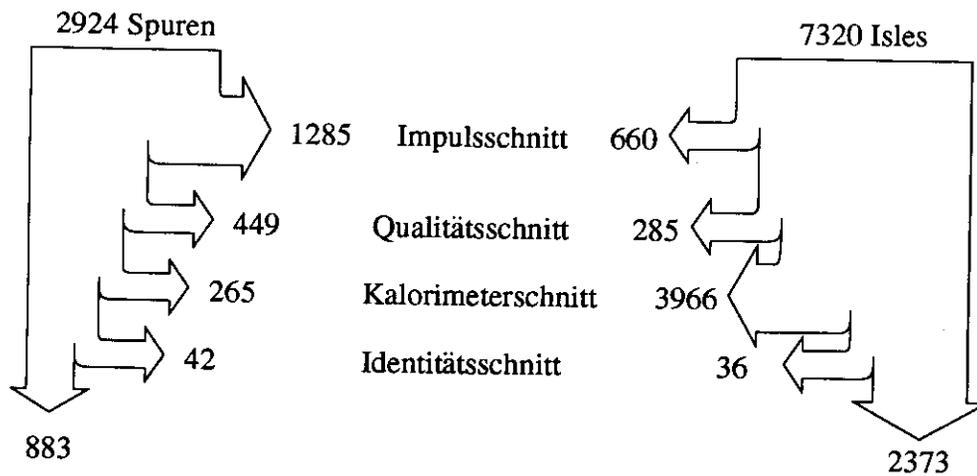


Abbildung 4.4: Wirkung der verschiedenen Schnitte auf die Zahl der Spuren und Isles. Die Grundmenge enthält alle diejenigen Spuren aus 1000 Ereignissen des Datensatzes mit Konversionselektronen, die den geometrischen Schnitt passiert haben. Die einzelnen Schnitte werden im Text genauer beschrieben.

### 4.3 Die Eingabevariablen

Unter etwa 50 plausiblen Eingabevariablen wurden zehn durch Vergleich ihrer Verteilungen für Elektronen und Hadronen in die engere Wahl gezogen. Die Abbildung 4.6 zeigt einige Verteilungen der Eingabevariablen für Elektronen (schraffiert) und Hadronen (grau unterlegt) überlagert. Es liegen gleiche Zahlen von Teilchen vor. Anhand dieser Verteilungen lassen sich bereits erste Vermutungen über ihre Brauchbarkeit zur Elektronenidentifikation anstellen:

**Isle-Signal:** Das Isle-Signal wird zweifellos die wichtigste Variable darstellen. Ein Schnitt wird bei etwa 3 bis 5 *mip* erwartet. Das Isle-Signal allein kann aber nicht perfekt klassifizieren.

**Relatives Isle-Signal:** Das Verhältnis von Isle-Signal und *Island*-Energie wirkt ebenfalls noch recht stark diskriminierend. Vermutlich ist es aber mit der ersten Variable korreliert, so daß es keine neuen Informationen hinzufügen kann. Um dies zu testen, wird es dennoch aufgenommen.

**Isle-Konzentration:** Anteil des Isle-Signals in der zentralen Diode. Unter den Isles mit mehr als einer Diode (*Isle*-Konzentration kleiner als 100%) sind mehr Elektronen als Hadronen. Auch diese Variable ist mit dem Isle-Signal korreliert, jedoch nicht vollkommen.

**EMC-Zelle:** Energiedeposition in der getroffenen EMC-Zelle. Die Verteilung der Energie in der getroffenen Zelle des elektromagnetischen Kalorimeters ist für Hadronen stark bei niedrigen Werten konzentriert. Das hängt damit zusammen, daß vor allem die höherenergetischen Hadronen einen Teil ihrer Energie erst im hadronischen Kalorimeter deponieren. Die Elektronen dagegen deponieren ihre volle Energie im EMC. Daher ist ihre Energieverteilung zu höheren Werten verschoben. Von dieser Variablen ist zu erwarten, daß sie zusätzliche Information liefern wird. Ihr Problem ist aber: Unterschiede in der Energieverteilung der Elektronen und Hadronen werden vom Netzwerk mitgelernt. Im vorliegenden Trainingsbeispiel ist das Spektrum für beide Teilchensorten jedoch recht gleich. Das Verhältnis von

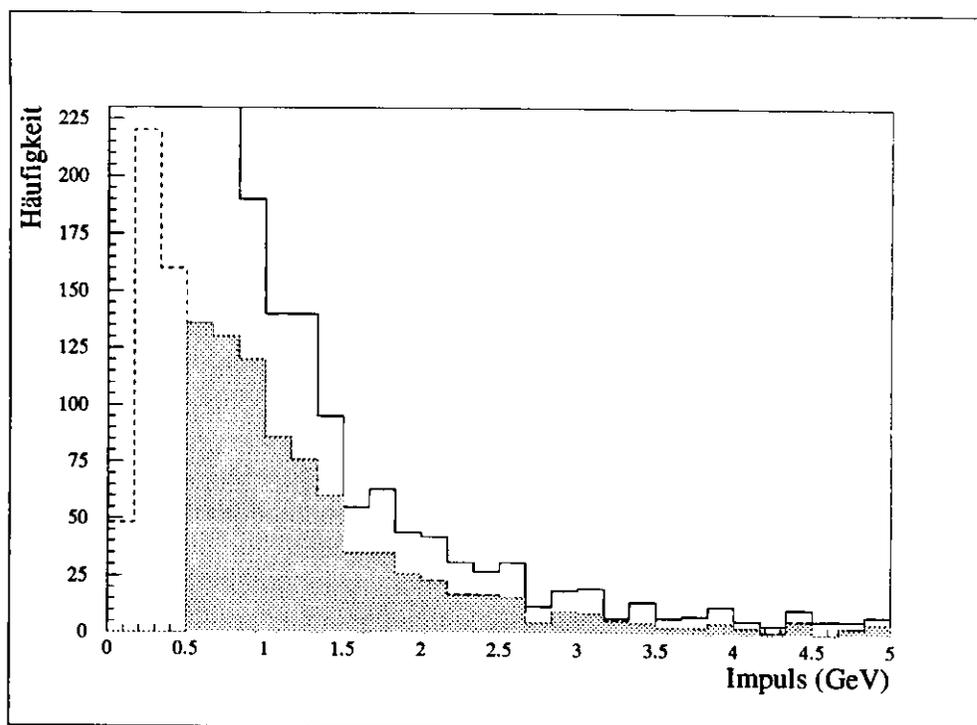


Abbildung 4.5: Impulsspektrum aller Spuren im Datensatz, die den geometrischen Schnitt passieren (durchgezogene Linie), die zusätzlich die Qualitätsanforderungen erfüllen (gestrichelte Linie) und die einen Impuls größer als  $500 \text{ MeV}/c$  besitzen (schattiert). Grundlage sind 1000 Ereignisse des Datensatzes mit Konversionselektronen.

elektromagnetischer und hadronischer Energie kann diese Variable nicht ersetzen, da nur etwa 30% der Teilchen überhaupt das hadronische Kalorimeter erreichen.

**HAC-Turm:** Energiedeposition in der HAC-Zelle des getroffenen Turms. Einige Teilchen erreichen das hadronische Kalorimeter. Die allermeisten von ihnen sind tatsächlich Hadronen. Für diese wenigen Teilchen wird das neuronale Netz vermutlich großes Gewicht auf diese Variable legen und sie als Hadronen diskriminieren.

**EMC-Anteil:** Der Anteil elektromagnetischer Energie in der Umgebung des eingetroffenen Teilchens an der gesamten *Island*-Energie ist im Gegensatz zur Variablen „EMC-Zelle“ nicht empfindlich gegen Unterschiede im Spektrum der Elektronen und Hadronen. Da ein größerer Bereich betrachtet wird, kommt es öfter als im Falle eines einzelnen Turmes zu einer Aufteilung der Energie im EMC und HAC. Entsprechend verwischer sind jedoch auch die beiden Verteilungen (nicht abgebildet). Vom neuronalen Netzwerk ist zu erwarten, daß es dieser Variablen relativ wenig Bedeutung beimißt.

**Energiekonzentration:** Anteil der *Island*-Energie im getroffenen Turm. Elektromagnetische Schauer sind stärker gebündelt als hadronische gleicher Energie. Bei einer Segmentierung des EMC in  $20 \times 10 \text{ cm}^2$  große Zellen gehen diese Unterschiede für Teilchen von einigen hundert  $\text{MeV}$  jedoch fast unter. Die Verteilung dieser Variablen zeigt aber noch die Tendenz an: Elektronen hinterlassen öfter höhere Anteile ihrer Energie in einem einzigen Turm.

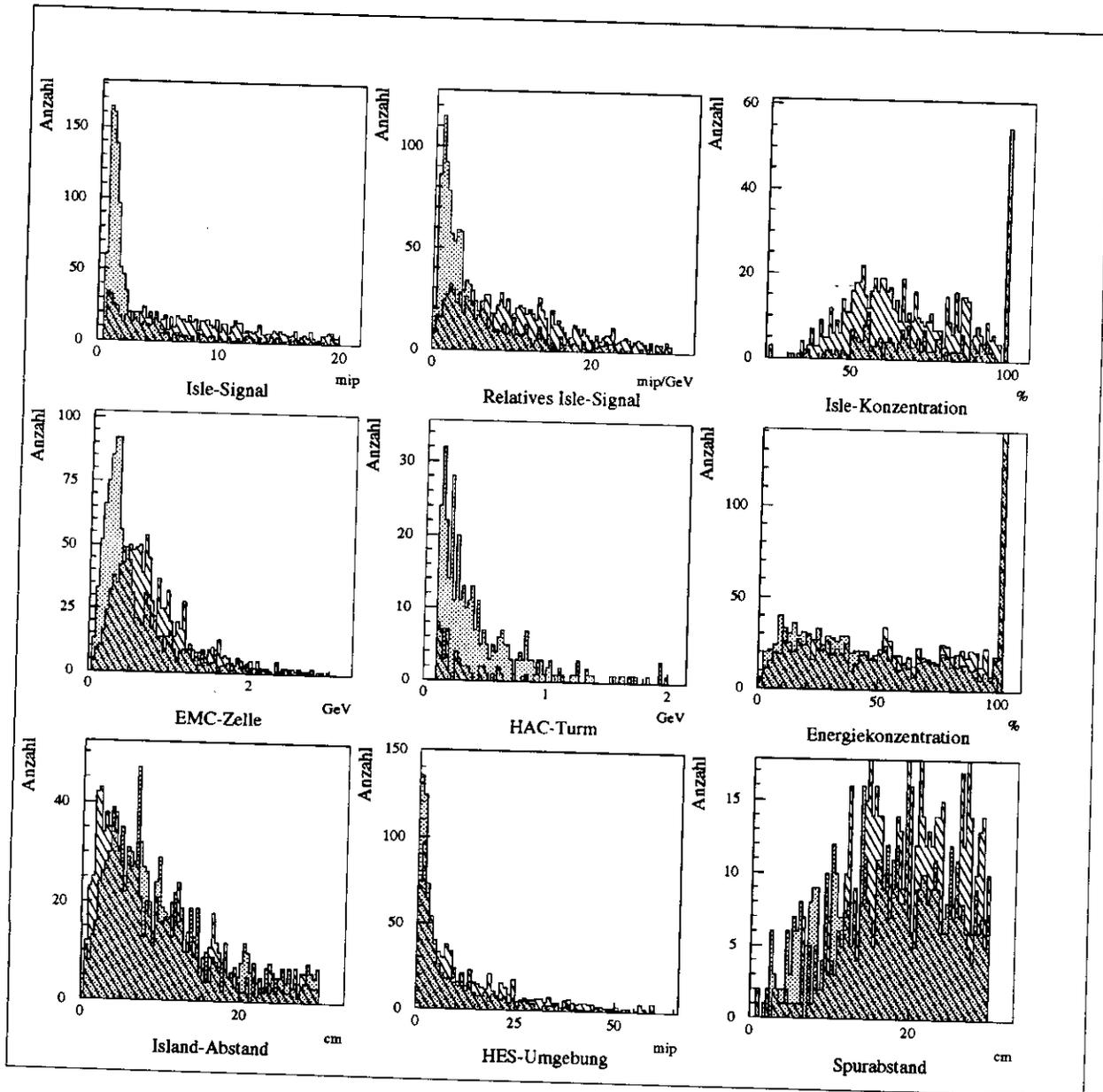


Abbildung 4.6: Verteilung von Elektronen (schraffiert) und Hadronen (grau unterlegt) in verschiedenen Variablen. Unterschiede in den beiden Verteilungen einer Variablen geben einen Hinweis auf deren Brauchbarkeit. Streng genommen stellen sie aber weder eine notwendige noch hinreichende Bedingung dar. Die einzelnen Variablen sind im Text erklärt.

**Island-Abstand:** Abstand von *Isle* und *Island*. Die Übereinstimmung der Positionen von *Isle* und *Island* ist bei Elektronen tendenziell besser als bei Hadronen, da die Elektronenschauer enger sind als die der Hadronen. Hinzu kommt i. a., daß Elektronen aus semielektronischen Zerfällen isoliert auftreten, während die meisten Hadronen in Bündeln (*jets*) auftreten. In diesen *jets* sind die Spurdichten so hoch, daß häufig mehrere Spuren auf ein *Island* zeigen, welches dann eine Position zwischen den Spuren einnimmt. In diesem Datensatz haben jedoch nur 40% der Elektronen ein *Island*, auf das keine weiteren Spuren zeigen (Hadronen 49%). Das hängt damit zusammen, daß die Konversionselektronen paarweise auftreten. Hinzu kommt, daß ihre Mutterteilchen (Photonen) zum großen Teil aus Zerfällen neutraler Pionen stammen, die ihrerseits genauso Teil eines *jets* sein können wie andere Pionen auch.

**HES-Umgebung:** Summe über die Signale aller HES-Dioden, des getroffenen Kalorimeterturmes und der acht benachbarten Türme, jedoch ohne Berücksichtigung der gerade betrachteten *Isle*. In der Umgebung eines Elektrons finden sich häufiger als bei Hadronen weitere *Isles* mit einem Gesamtsignal von mehr als 5 *mip*. Dies mag damit zusammenhängen, daß der Konversionspartner in der Nähe aufgetroffen ist oder daß das konvertierte Photon aus einem  $\pi^0$ -Zerfall stammte und das zweite Zerfallsphton nahebei eintraf. In diesem Fall ist es nicht ratsam, diese Variable zu verwenden. Es kann aber auch auf begleitende Photonen hinweisen, die von dem Elektron beim Durchgang durch Materie ausgestrahlt wurden. Welche Bedeutung den einzelnen Prozessen beizumessen ist, wurde nicht untersucht.

**Spurabstand:** Der Abstand zur zweitnächsten Spur ist ein weiteres Isolationskriterium. Er ist für Hadronen tendenziell geringer als für Elektronen. Für Elektronen aus semileptonischen Zerfällen sollte dieser Effekt noch stärker sein.

	Name	Bedeutung
1.	Isle-Signal	Summe der Ladungen aller Dioden einer <i>Isle</i> in <i>mip</i> .
2.	Relatives Isle-Signal	Isle-Signal relativ zur Island-Energie ( <i>mip/GeV</i> ).
3.	Isle-Konzentration	Anteil der zentralen Diode am Isle-Signal.
4.	EMC-Zelle	Energie in der getroffenen Zelle des EMC.
5.	HAC-Turm	Energie im getroffenen Turm des HAC.
6.	EMC-Anteil	Energie in sechs EMC-Zellen um das Teilchen relativ zur Island-Energie.
7.	Energiekonzentration	Energie im zentralen Turm relativ zur Island-Energie.
8.	Island-Abstand	Abstand der Islandposition von der <i>Isle</i> .
9.	HES-Umgebung	Summe der Ladungen aller Dioden in $3 \times 3$ Türmen um die betrachtete <i>Isle</i> herum.
10.	Spurabstand	Abstand der zweitnächsten Spur der <i>Isle</i> .

Tabelle 4.2: Eingabevariable für das neuronale Netzwerk.

Netzgeometrie	1-5-1...10-5-1 <i>feed forward</i> , je nach Zahl der Eingabevariablen.
Schrittweite	$\eta = 5 \cdot 10^{-3} \dots 5 \cdot 10^{-5}$ kontinuierlich abnehmend.
Dämpfung	$\alpha = 0,98$
Training nicht in Blöcken.	
Kein <i>pruning</i> .	
Zahl der Schritte	$10^5$
Kodierung	$\approx 1$ für Elektronen, $\approx 0$ für Hadronen.
Variablenskalierung	0,4, 0,1, 2, 1, 3, 2, 0,5, 0,1, 0,1, 0,05
Trainingsbeispiel:	
Impulsverteilung	$p = 500 \text{ MeV}/c \dots 5 \text{ GeV}/c$ , exponentiell abfallende Verteilung.
Anzahl	540 Elektronen, 660 Hadronen
Testbeispiel	587 Elektronen, 670 Hadronen, unabhängig vom Trainingsbeispiel.

Tabelle 4.3: Einige technische Details des Netzwerk-Trainings.

## 4.4 Trainingsergebnisse

### 4.4.1 Einleitung

Die Tabelle 4.2 zeigt die zehn Eingabevariablen der engeren Wahl. Verschiedene Kombinationen dieser Variablen führten zu leicht unterschiedlichen Ergebnissen, die in den folgenden Abschnitten diskutiert werden. Einige technische Details des Trainings sind in Tab. 4.3 zusammengefasst.

Alle Ergebnisse wurden auf eine Effizienz von 75% standardisiert. Bei niedrigeren Effizienzen erreicht man zwar noch bessere Verstärkung des Signals relativ zum Untergrund, die statistischen Fehler wirken sich aber stärker aus, da die Misidentifikationswahrscheinlichkeit dann zu klein wird.

### 4.4.2 Definition der Effizienz, Diskriminanzkraft und Signalverstärkung

Zum Vergleich der Leistungen verschiedener Elektronenfinder sollen die Größen Effizienz, Diskriminanzkraft und Signalverstärkung benutzt werden.

Die *Effizienz*  $E$  gibt an, welcher Anteil der Elektronen im Testdatensatz durch den Elektronenfinder nachgewiesen werden konnte.

Die *Diskriminanzkraft* oder kürzer *Diskriminanz*  $D$  sagt aus, welcher Anteil der Hadronen im Testdatensatz zurückgewiesen (diskriminiert) werden konnte. Ihr Komplement ist die *Misidentifikationswahrscheinlichkeit*  $\bar{D} = 1 - D$ .

Die *Signalverstärkung*  $S$  gibt an, um welchen Faktor ein Elektronensignal im Vergleich zum Untergrund durch Anwendung des Elektronenfinders verstärkt wird. Berechnet wird sie als Quotient aus der Effizienz und der Misidentifikationswahrscheinlichkeit.

Des weiteren wird häufig die *Reinheit* eines Elektronenfinders angegeben. Diese Größe soll hier nicht benutzt werden, da sie von der Zusammensetzung des Testdatensatzes abhängig ist. Man kann sie aber aus der Effizienz  $E$ , der Diskriminanz  $D$  und dem Verhältnis

$$r := \frac{n_e}{n_e + n_u}$$

der Zahl der Elektronen zu der Gesamtzahl der Testdaten errechnen (der Querstrich meint jeweils

das Komplement):

$$R = \frac{rE}{rE + \bar{r}\bar{D}}$$

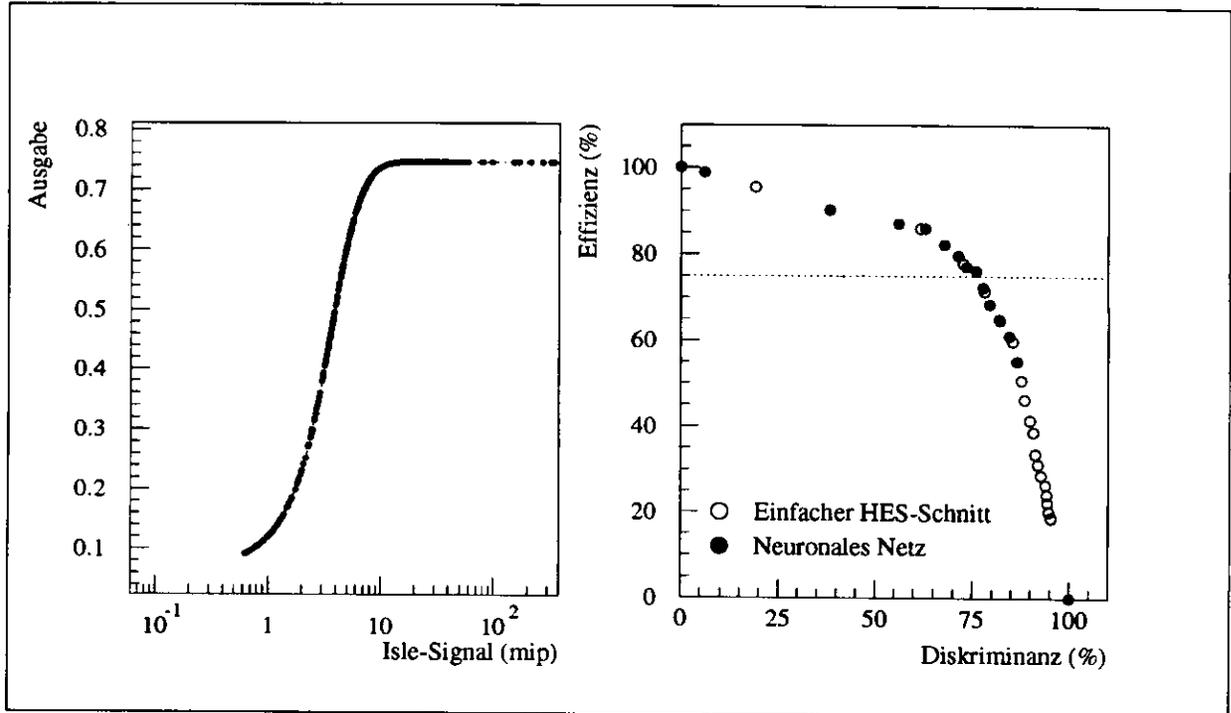


Abbildung 4.7: Der einfache HES-Schnitt. Die linke Abbildung zeigt die Netzausgabe in Abhängigkeit vom *Isle*-Signal. Rechts ist die Effizienz des neuronalen Netzes und des einfachen HES-Schnittes gegen die Diskriminanz aufgetragen. Für das HES-Signal wird der Schnitt von Punkt zu Punkt 1 *mip* heraufgesetzt, in der Netzausgabe unterscheiden sich benachbarte Schnitte um 0,05. Die gepunktete Linie zeigt 75% Effizienz an.

#### 4.4.3 Der einfache HES-Schnitt

Zunächst wurde das Netz mit einer einzigen Variablen trainiert, dem *Isle*-Signal. Der vorgegebene Zielwert für Elektronen war 1, für Hadronen 0. (75,0 ± 1,8)% Effizienz erreicht man bei einem Schnitt von 0,468 in der Ausgabevariablen. Das entspricht einem Schnitt bei 3,64 *mip*. Die Diskriminanz beträgt (76,2 ± 1,7)%. Damit erzielt man eine Signalverstärkung von

$$S = \frac{E}{1 - D} = 3,15 \pm 0,20$$

ohne Berücksichtigung der Spur-*Isle*-Zuordnung.

Die Abbildung 4.7 zeigt links die Netzausgabe als Funktion des *Isle*-Signals. Zur besseren Proportionierung der Kurve ist das *Isle*-Signal logarithmisch aufgetragen. Kaum zu sehen ist wegen der großen Punktdichte, daß die Elektronen als starke, die Hadronen als schwache Punkte eingetragen sind. Rechts ist für zwanzig verschiedene Schnitte in der Netzausgabe (ausgefüllte

Kreise) und im *Isle*-Signal (hohle Kreise) die Effizienz gegen die Diskriminanzkraft dargestellt. Wie zu erwarten folgen sie derselben durch die Daten vorgegebenen Linie. Je näher diese Linie der rechten oberen Ecke des Diagramms kommt (perfekte Diskriminanz bei voller Effizienz), desto besser ist der Elektronenfinder zu bewerten. Eine zufällige Trennung von Elektronen und Hadronen würde zu einer Diagonalen führen. Der einfache HES-Schnitt erweist sich also als wirkungsvoll, aber lange nicht perfekt.

#### 4.4.4 Verbesserung durch die Variable HAC-Turm

Bei zwei Variablen bleibt die Arbeit des neuronalen Netzwerkes noch durchschaubar. Abb. 4.8 zeigt die Eigenschaften des Elektronenfinders mit den Variablen *Isle-Signal* und HAC-Turm.

Die Gewichtsmatrix wird folgendermaßen gelesen: Die oberste Zeile zeigt die Gewichte, die die einzelnen verborgenen Knoten mit dem Ausgangsknoten verbinden. Sie bestimmen, welche Bedeutung den Entscheidungen der einzelnen verborgenen Knoten zukommen soll. In diesem Fall ist die Bedeutung des ersten verborgenen Knotens größer als die aller anderen zusammen. Aus den weiteren Zeilen des Diagramms geht hervor, welche Linearkombinationen der Eingangsvariablen der Entscheidung der einzelnen verborgenen Knoten zugrunde liegt. Der erste und wichtigste verborgene Knoten sieht als Argument seiner Schwellfunktion eine Linearkombination der beiden benutzten Eingangsvariablen. Der HAC-Turm erhält ein großes, negatives Gewicht (weiß), das *Isle-Signal* ein kleines, positives (grau). Energie im HAC-Turm führt also zu kleinen Ausgabewerten (negatives Gewicht), was einer Klassifikation zum Hadron entspricht. Große *Isle-Signale* führen wegen des positiven Gewichtes zu großen Ausgabewerten: Elektron. Das Verhältnis der Größen der beiden Gewichte sagt aus, daß 1 GeV hadronischer Energie genauso stark für ein Hadron sprechen wie 16 *mip* *Isle*-Energie für ein Elektron. Das Gewicht der Variablen HAC-Turm ist 16mal so groß wie das der Variablen *Isle-Signal*. Letzteres könnte man in dem Diagramm kaum noch sehen. Deshalb wurden die Gewichte dieser Variable zehnfach vergrößert dargestellt. Wie aus dieser Diskussion bereits deutlich wird, kann die Bedeutung einer Variablen für die Entscheidungsfindung nicht streng quantitativ angegeben werden. Das liegt zum einen daran, daß die Skalen der einzelnen Variablen nicht normierbar sind (die Verteilungen sind zu unterschiedlich), zum anderen daran, daß die relative Größe der Gewichte der einzelnen Eingabevariablen für die verschiedenen verborgenen Knoten i. a. unterschiedlich ausfällt.

Die Ausgangsfunktion, die durch diese Gewichtsmatrix repräsentiert wird, ist ein Bild weiter durch ihre Höhenlinien in der Ebene der beiden Eingangsvariablen dargestellt. Die Linearkombination des ersten verborgenen Knotens erzeugt eine diagonal durch das Bild verlaufende Schwelle. Links oberhalb entscheidet das Netz zu 0 (Hadron), rechts unterhalb zu 1 (Elektron). Durch die Variable HAC-Turm werden also einige Hadronen bis zu 10 *mip* noch diskriminiert, ein Elektron bei 10 *mip* mit 0,9 GeV HAC-Energie geht aber dafür verloren, einige weitere auf der Schwelle eventuell auch. Die meisten Teilchen befinden sich in dem „Balken“ bei 0 GeV HAC-Energie.

Die durch ihre HAC-Energie besser als Hadronen zu erkennenden Teilchen sondern sich im rechten unteren Diagramm (Netzausgabe gegen *Isle-Signal*) deutlich von der bekannten *Isle-Signal*-Kurve ab. Das eben erwähnte Elektron findet sich in dieser Darstellung bei etwa 10 *mip* unten rechts. Die senkrechte Linie zeigt den vom einfachen HES-Signal vorgeschlagenen Schnitt, die waagerechte den Schnitt in der Netzausgabevariablen dieses Netzes. Beide Schnitte sind auf nahezu 75% Effizienz eingestellt. Damit sind rechts unten die Teilchen, die durch die neue Variable zusätzlich abgelehnt werden konnten, zehn Hadronen (von 671) und mit ihnen vier Elektronen (von 587). Links oberhalb sind die zusätzlich akzeptierten Teilchen, vier Elektronen

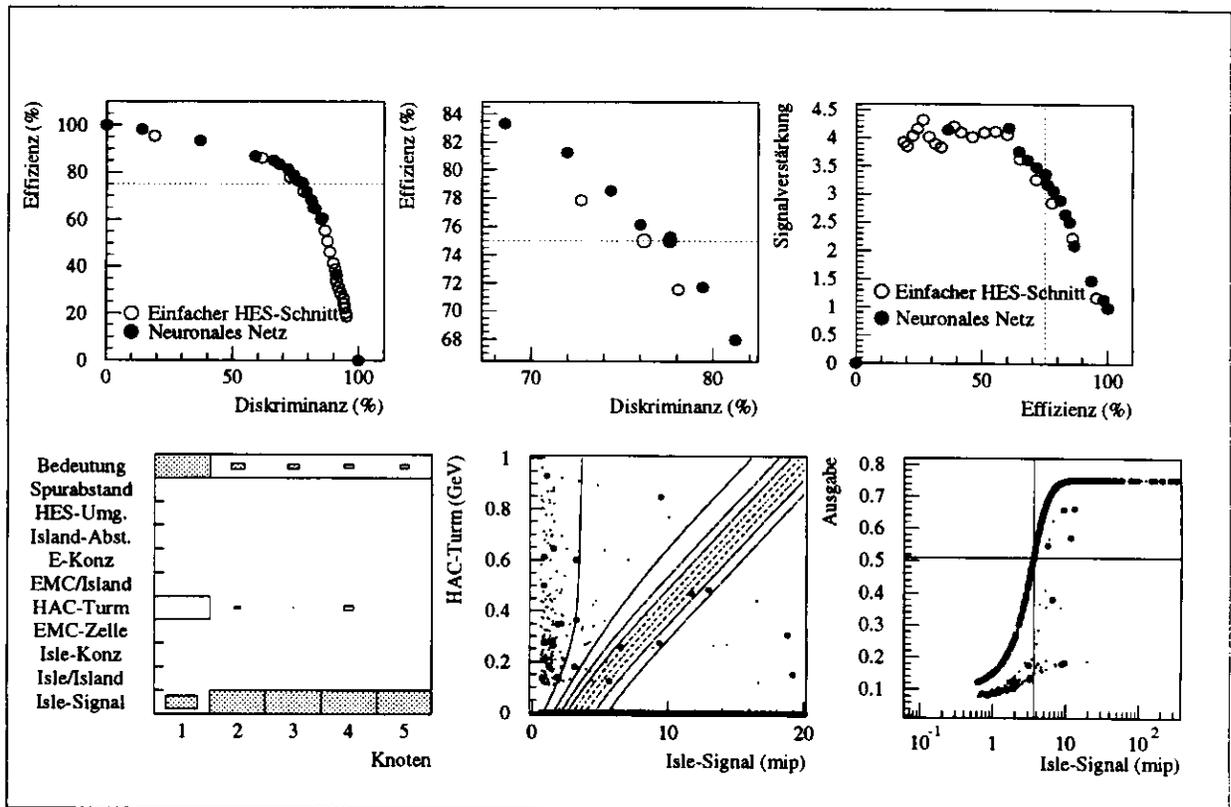


Abbildung 4.8: Isle-Signal und HAC-Turm. Obere Reihe: Effizienz gegen Diskriminanzkraft für verschiedene Schnittwerte in der Netzausgabe im Vergleich zum einfachen HES-Schnitt (links), Detaildarstellung desselben Diagramms um die Vergleichseffizienz von 75% herum (Mitte) und Signalverstärkung gegen Effizienz (rechts). Unten links: Gewichtsmatrix. Abgebildet sind die Größen der Gewichte zwischen den Eingangsvariablen (Zeilen außer der obersten) und den verborgenen Knoten (Spalten). Negative Gewichte sind weiß, positive grau dargestellt. Die Spalten sind getrennt normiert. Die unteren 10 Zeilen zeigen also die Linearkombination der Eingangsvariablen, die jedem verborgenen Knoten als Argument dient. Die oberste Zeile zeigt die Gewichte der verborgenen Knoten in der Ausgangsentscheidung. Interpretation der Gewichtsmatrix im Text. Unten Mitte: Netzausgabe als Funktion der beiden Variablen Isle-Signal und HAC-Turm, dargestellt durch Höhenlinien. Teilchen rechts unterhalb der Schwelle werden als Elektronen, links oberhalb als Hadronen bezeichnet. Die wahren Elektronen sind durch starke, die Hadronen durch schwache Punkte gekennzeichnet. Unten rechts: Netzausgabe gegen Isle-Signal. Die Linien markieren den einfachen HES-Schnitt (senkrecht) und den Schnitt dieses Elektronenfinders, beide mit 75% Effizienz.

(in der Hauptkurve).

Insgesamt erzielte das Training mit *Isle*-Signal und HAC-Turm bei einem Schnitt von 0,51 eine Effizienz von  $(75,0 \pm 1,8)\%$  bei einer Diskriminanz von  $(77,6 \pm 1,7)\%$ . Daraus läßt sich eine Signalverstärkung von  $S = 3,3 \pm 0,3$  berechnen. Es übertrifft damit den einfachen HES-Schnitt um 6%. Bei diesem Vergleich ist zu berücksichtigen, daß beide Elektronenfinder sich hauptsächlich nach dem Wert des *Isle*-Signals richten. Sie sind daher in ihren Entscheidungen stark korreliert. Aus diesem Grund kann die statistische Signifikanz des Unterschiedes in der Signalverstärkung nicht aus der Größe der Vertrauensintervalle ersehen werden. Man muß vielmehr die Abhängigkeit berücksichtigen.

Ausgehend von der Zahl der Hadronen, die vom HES-und-HAC-Elektronenfinder gefunden wurden, vom einfachen HES-Schnitt jedoch nicht (10) und der Zahl der Hadronen, die durch die Hinzunahme der HAC-Information verloren gingen (0), wurde für die Ungleichheit dieser Zahlen getestet<sup>5</sup>. Die Hinzunahme der Variablen HAC-Turm verbessert demnach die Diskriminanz des einfachen HES-Schnitts (und damit auch ihre Signalverstärkung) signifikant.

#### 4.4.5 Alle HES-Variablen

Das dritte Training wurde mit den drei reinen HES-Variablen vorgenommen: *Isle*-Signal, *Isle*-Konzentration und HES-Umgebung. Hier zeigt sich die HES-Umgebung als stärkste Korrektur zu der unumstritten wichtigsten Variablen *Isle*-Signal. Es wurden ähnliche Untersuchungen angestellt wie im letzten Abschnitt, sie sollen aber nicht einzeln diskutiert werden.

Insgesamt erzielt dieses Netz bei einer Effizienz von  $(75,0 \pm 1,8)\%$  eine Diskriminanzkraft von  $(77,2 \pm 1,7)\%$ . Mit einer Signalverstärkung von  $S = 3,3 \pm 0,3$  übertrifft es den einfachen HES-Schnitt um 4%. Die Verlässlichkeit des Vergleichswertes ist schlechter als beim letzten Vergleich. Es gingen nämlich 19 Elektronen und 16 Hadronen verloren, andere 19 Elektronen und 23 Hadronen wurden gewonnen. Der Unterschied ist damit nicht signifikant.

#### 4.4.6 Alle Kalorimetervariablen

Ein wirklicher Unterschied in der Leistung des Elektronenfinders ist zu erwarten, wenn man auf das *Isle*-Signal verzichtet. Es bleiben dann die reinen Kalorimeter-Variablen: EMC-Zelle, HAC-Turm, EMC-Anteil, Energiekonzentration. Aus der Gewichtsmatrix des trainierten Netzes geht hervor, daß die EMC-Zelle und der HAC-Turm die Entscheidung des Netzwerkes dominieren. Die klarste Entscheidung trifft die Variable HAC-Turm, jedoch nur für die wenigen Ereignisse mit HAC-Energie.

Die gemessene Effizienz beträgt  $(75,0 \pm 1,8)\%$  bei einer Diskriminanz von  $(64,7 \pm 1,9)\%$ . Die Signalverstärkung ist auf  $S = 2,1 \pm 0,2$  abgesunken. Bezieht man die Spur-*Isle*-Zuordnung in den Vergleich mit ein, so wird der Unterschied noch größer, da das Kalorimeter allein diese Zuordnung gar nicht ausführen könnte. Viele Hadronen, die mangels einer *Isle* schon im Vorfeld diskriminiert wurden, muß der Kalorimeter-Elektronenfinder noch untersuchen.

#### 4.4.7 Alle Eingabevariablen

Das optimale Ergebnis erwartet man für die Verwendung aller Variablen. Das mit allen Variablen trainierte Netzwerk zeigte bei einer Effizienz  $(75,0 \pm 1,8)\%$  eine Diskriminanzkraft von  $(79,9 \pm$

<sup>5</sup>Die genaue Rechnung ist in Anhang B.2 wiedergegeben. Alle Vergleiche zwischen Elektronenfindern beziehen sich auf ein Signifikanzniveau von 5%.

1,6)%. Dies resultiert in einer Signalverstärkung von  $S = 3,7 \pm 0,3$ . Damit wird der einfache HES-Schnitt um 18% (signifikant) übertroffen. Ein Nachteil dieses Netzes ist, daß es Variable enthält, die auf das größere Umfeld der Teilchen sensibel sind.

#### 4.4.8 hesnn

Ein weiterer untersuchter Elektronenfinder wurde aus vier Variablen konstruiert, die nur dem zentralen Kalorimeterturm und der HES-Isle entnommen sind. Diese Variablen werden auch „lokale Variable“ genannt, da sie nur der nächsten Umgebung des Ortes entstammen, an dem das Teilchen die HES-Ebene passierte. Dieser Elektronenfinder ist daher nicht sensibel auf den größeren Kontext, etwa den Konversionspartner oder die Spurdichte. Da mit ihm ähnlich gute Ergebnisse erzielt werden konnten wie mit allen Variablen, wurde diese Version als *der* Elektronenfinder erkoren und bekam den Namen *hesnn*. Seine gegen „Umfeldeinflüsse“ unempfindlichen Variablen sind das *Isle*-Signal, die *Isle*-Konzentration, die EMC-Zelle und der HAC-Turm.

Die Abbildung 4.9 zeigt die Entwicklung der Gewichte im Verlauf der 100 000 Trainings-schritte. Im Gegensatz zu Kap. 3 wurde hier die Schrittweite während des Trainings von Schritt zu Schritt um einen festen Faktor verkleinert. Das ist nötig, weil die zu trennenden Gruppen im Raum der Eingangsvariablen stark überlappen. Es bleiben auch im optimalen Zustand noch starke Korrekturen. Zudem wurden Skalierungsfaktoren für die einzelnen Variablen eingefügt, um dem Netzwerk für jede Variable etwa dieselbe Skala zu bieten. Auch dies war in den Verteilungen des Kapitels 3 nicht nötig. Diese Techniken wurden übrigens in allen hier verglichenen Trainingsdurchgängen gleichermaßen angewandt.

Jedes Einzelbild der Abb. 4.9 zeigt alle fünf Gewichte *einer* Eingangsvariablen, d.h. die fünf Gewichte, die die entsprechende Eingangsvariable mit den fünf verborgenen Knoten verbindet. Die nichtlokalen Variablen sind vom Training ausgeschlossen, ihre Gewichte sind daher stets 0. Die Gewichte der lokalen Variablen variieren zu Beginn des Trainings stark infolge der großen Schrittweite und der höheren Gradienten. Schließlich verharren sie aber in ihrer Endposition.

Die Abbildung 4.10 zeigt die Netzausgabe für die Elektronen (erste Zeile) und Hadronen (zweite Zeile). Aus Mangel an Speicherplatz ist jeweils nur jeder 39. Schritt aufgeführt. Nach dem Training wurde zur Bestimmung von Effizienz und Diskriminanz noch einmal das volle Trainings- und Testbeispiel durchlaufen. Hier sind alle Punkte wiedergegeben, was zu den beiden Anhäufungen auf der rechten Seite jedes der beiden Diagramme führt.

Zusätzlich zu dieser genauen Bestimmung der Effizienz und Diskriminanzkraft wurden die beiden Größen wenigstens für das Trainingsbeispiel auch während des Trainings beobachtet. Sie wurden mit 0 initialisiert und dann bei jedem erfolgreichen Schritt mit

$$\begin{aligned} E^{(s+1)} &= 0,999 \cdot E^{(s)} + 0,001 \\ D^{(s+1)} &= 0,999 \cdot D^{(s)} + 0,001 \end{aligned}$$

aktualisiert. Bei erfolglosen Schritten wurde der zusätzliche Summand weggelassen. Die Entwicklung dieser beiden Größen ist in den beiden unteren Zeilen der Abb. 4.10 zu sehen. Zu Beginn müssen sie sich zwar erst einpendeln und spiegeln daher nicht die wahre Entwicklung der Effizienz und Diskriminanz wider, nach etwa 4600 Schritten jedoch ist ihre Abweichung von der wahren Momentaneffizienz und -diskriminanz kleiner als 1%. (Man beachte, daß die Abszisse mit 10 skaliert ist!) Beide Größen erreichen sehr bald Werte nahe dem Endwert und ändern sich danach nur noch geringfügig.

In Abb. 4.11 sind die Eigenschaften des Elektronenfinders *hesnn* dargestellt.

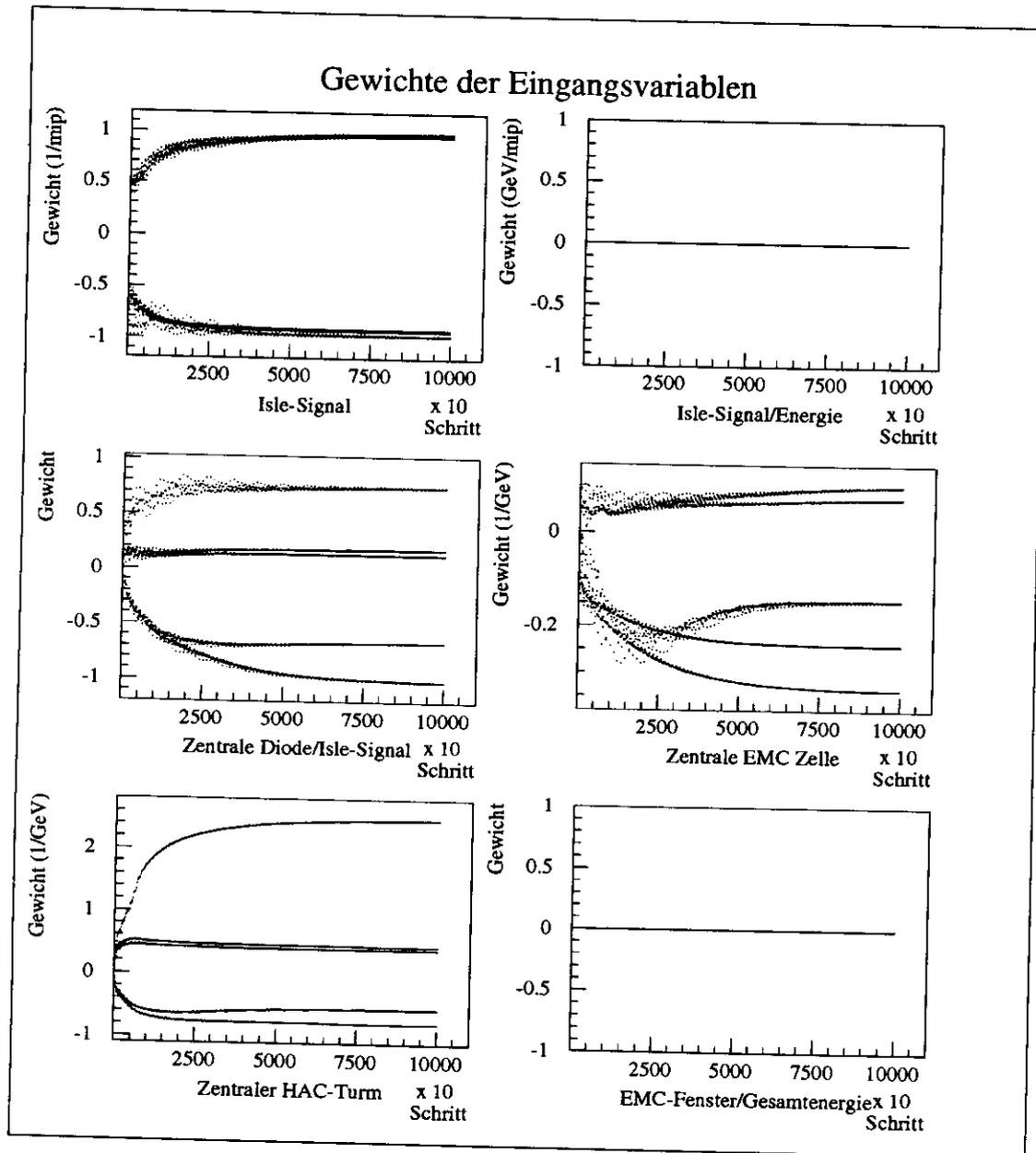


Abbildung 4.9: Entwicklung der Gewichte beim Training von *hesnn*. In jedem Bild sind die Gewichte einer Variablen zu den fünf verborgenen Knoten gezeigt.

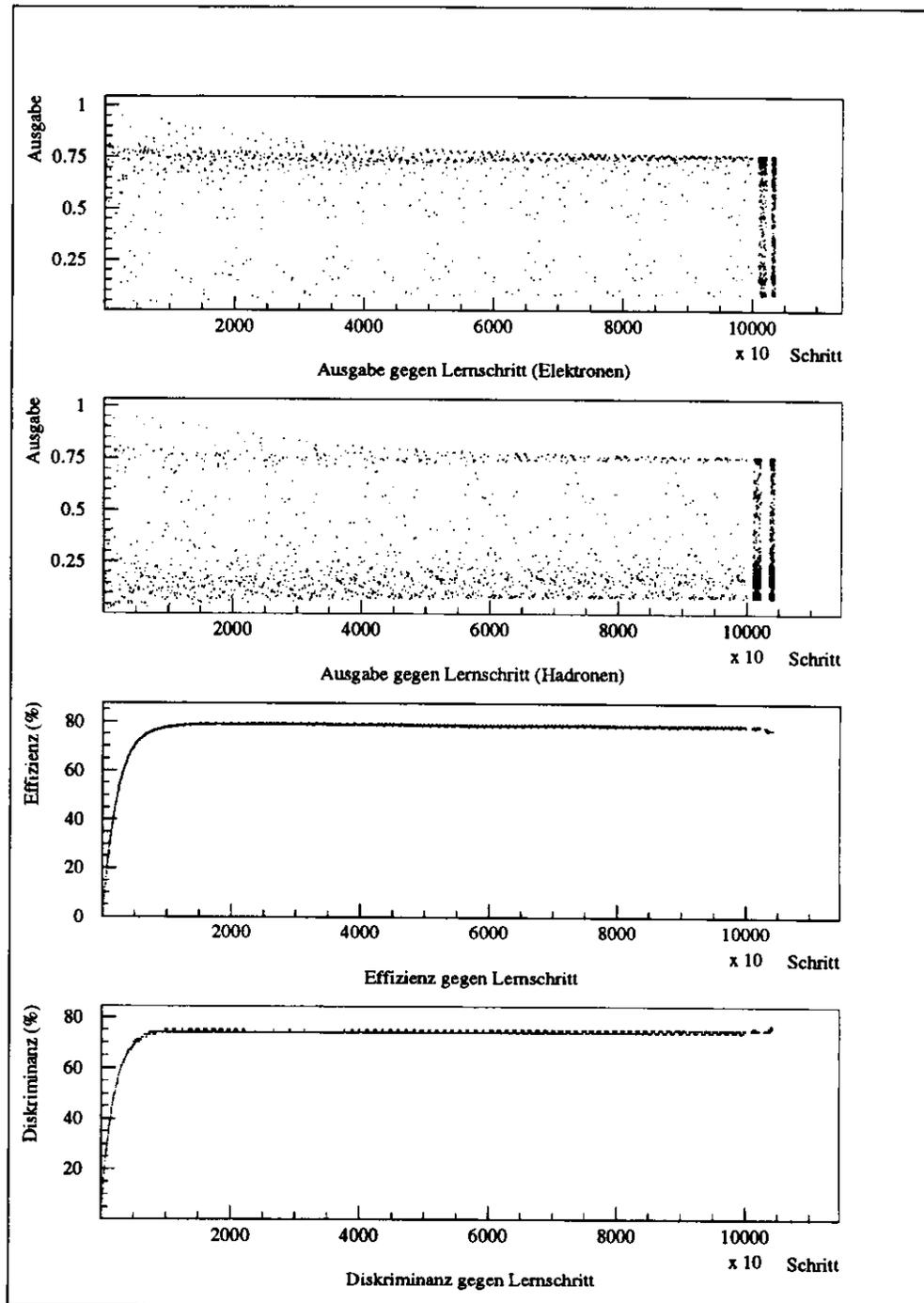


Abbildung 4.10: Netzausgabe für die Elektronen (oben) und Hadronen im Verlauf des Trainings von *hesnn*. Die Punktansammlungen ganz rechts rühren von den abschließenden Testdurchgängen her. Entwicklung der Effizienz und Diskriminanz im Verlauf des Trainings.

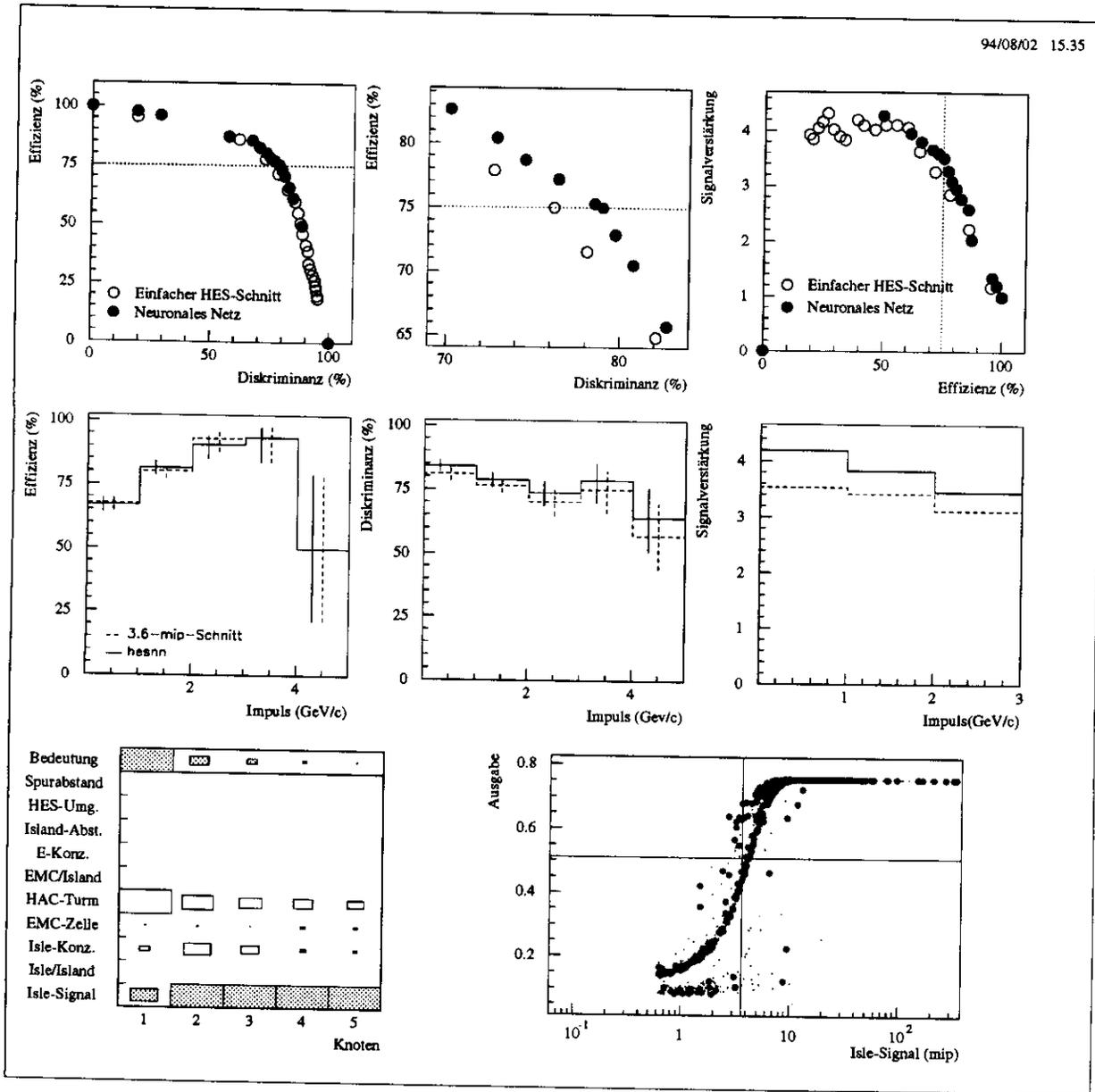


Abbildung 4.11: Der Elektronenfinder *hesnn*. Die obere Reihe zeigt wieder die Abhängigkeiten von Effizienz, Diskriminanzkraft und Signalverstärkung im Vergleich zum einfachen HES-Schnitt. In der Mitte ist die Effizienz, Diskriminanzkraft und Signalverstärkung in Abhängigkeit vom Impuls aufgetragen. Bei der rechten Abbildung wurden die Klassen oberhalb von 3 GeV wegen zu geringer Statistik weggelassen. Unten sieht man die Gewichtsmatrix und die *Isle*-Netzausgangskurve, diesmal mit zusätzlichen Korrekturen nach oben durch die Variable *Isle*-Konzentration.

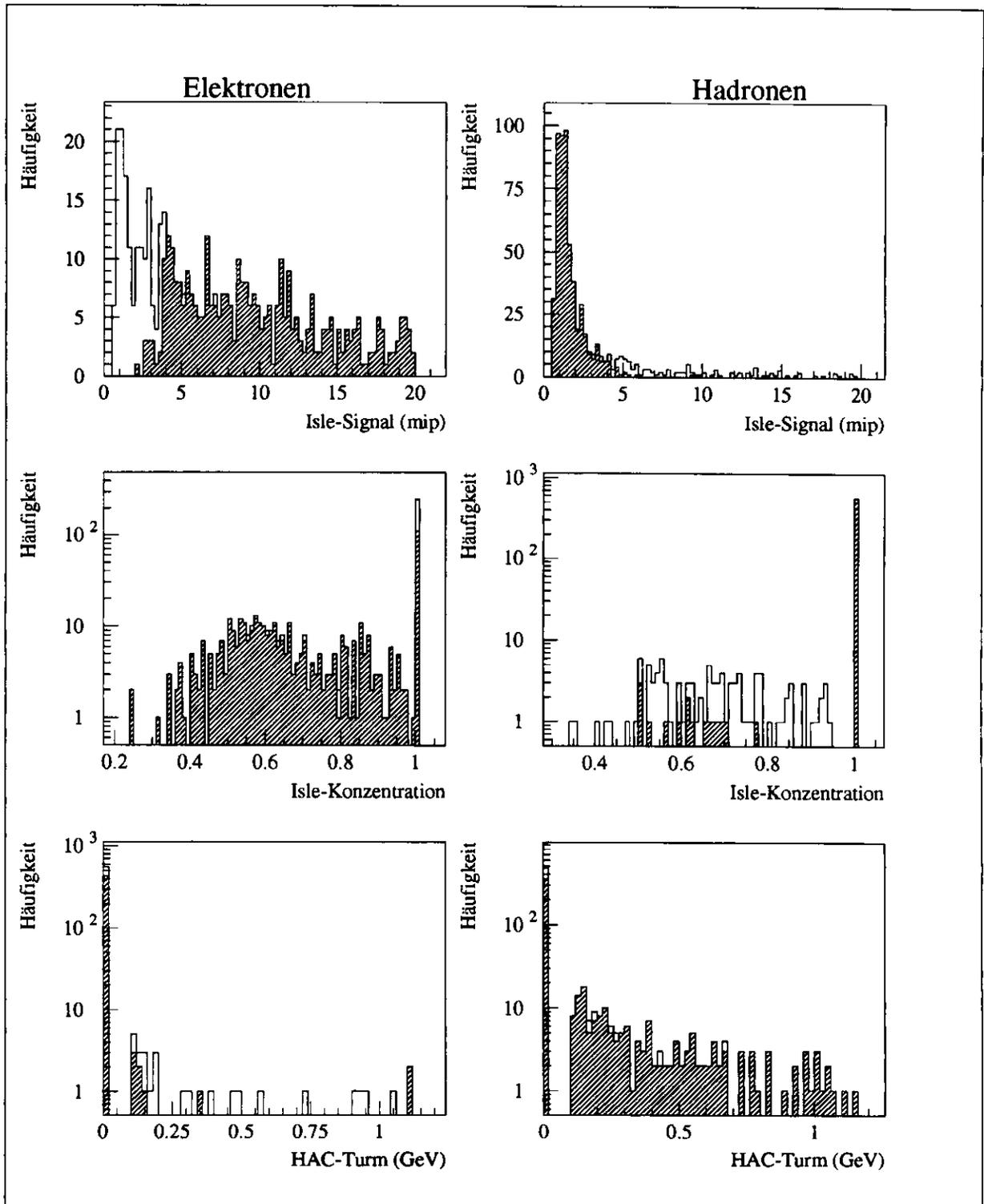


Abbildung 4.12: Der Elektronenfinder *hesnn*. Korrelation der wichtigsten Eingangsvariablen mit dem Netzausgang, jeweils links für die Elektronen und rechts für die Hadronen. Die vom Netz gefundenen Teilchen sind schraffiert eingetragen.

Das Effizienz-Diskriminanz-Diagramm zeigt einen leichten Vorteil dieses Elektronenfinders gegenüber dem einfachen HES-Schnitt. Mit  $E = (75,1 \pm 1,8)\%$  und  $D = (79,0 \pm 1,6)\%$  verbesserte sich die Signalverstärkung um etwa 14% auf  $D = 3,6 \pm 0,3$ . Insgesamt fand der Elektronenfinder 28 Hadronen, die vom einfachen HES-Schnitt als Elektronen angesehen wurden, 9 andere, die der einfache HES-Schnitt zu identifizieren vermochte, wurden jedoch von *hesnn* als Elektronen angesehen. Die Elektronenfinder unterscheiden sich auf einem Signifikanzniveau von 5%.

Abb. 4.11 Mitte vergleicht die Effizienz und Diskriminanzkraft des Elektronenfinders *hesnn* mit denen des einfachen HES-Schnittes. Die beiden Abbildungen links und in der Mitte zeigen die Effizienz und Diskriminanzkraft von *hesnn* und einfachem HES-Schnitt in verschiedenen Klassen (*bins*) des Teilchenimpulses. Mit zunehmendem Impuls nimmt die Statistik allerdings exponentiell ab, so daß die Fehler in den letzten beiden Klassen sehr groß werden. Die Schnitte sind derart gewählt, daß beide Elektronenfinder eine mittlere Effizienz von ca. 75% haben. Die rechte Abbildung zeigt schließlich die daraus errechnete Signalverstärkung in Abhängigkeit vom Impuls. Wegen der großen Fehler oberhalb von  $3 \text{ GeV}/c$  wurden nur die ersten drei Klassen berechnet.

Unten rechts ist die Netzausgabe gegen das *Isle*-Signal aufgetragen. Man erkennt sofort die bekannte Hauptkurve und die nach unten korrigierten Teilchen mit HAC-Energie. Zusätzlich gibt es offensichtlich eine Korrektur nach oben. Überraschenderweise wird diese Korrektur nicht durch die EMC-Zelle hervorgerufen, wie die Gewichtsmatrix deutlich zeigt (Zeile 4 hat fast keine Gewichte), sondern durch die *Isle*-Konzentration. Das beweist, daß die *Isle*-Konzentration eine bessere Ergänzung des *Isle*-Signals darstellt als die EMC-Zelle. Die Informationen der EMC-Zelle werden von dem Netzwerk als überflüssig angesehen.

Die Abb. 4.12 zeigt die Korrelation der wichtigsten Eingabevariablen mit der Entscheidung des Elektronenfinders *hesnn*. In der ersten Variablen ist ein klarer Schnitt bei  $3 \text{ mip}$  zu erkennen. Unterhalb von  $2 \text{ mip}$  werden praktisch alle Elektronen abgelehnt, oberhalb von  $4 \text{ mip}$  werden nahezu 100% der Elektronen akzeptiert. Im schmalen Übergangsbereich werden die Beiträge der anderen Eingabevariablen bedeutsam. Sie können tatsächlich bis zur Grenze von  $4 \text{ mip}$  fast alle Hadronen diskriminieren.

Auch die *Isle*-Konzentration zeigt Struktur: Während beim Wert von 1 nur etwa die Hälfte der Elektronen identifiziert werden kann (logarithmische Darstellung!), werden die Elektronen mit mehr als einer angeschlagenen Diode praktisch alle gefunden. Bei den Hadronen ist das Verhalten gerade umgekehrt. Aus diesen Darstellungen ist aber nicht ersichtlich, ob die *Isle*-Konzentration Informationen beiträgt, die von denen des *Isle*-Signals unabhängig sind.

Den Verteilungen der Energie im HAC-Turm sieht man einen einseitigen Beitrag an: Energien von mehr als  $200 \text{ MeV}$  im HAC-Turm führen zur Einschätzung des Teilchens als Hadron. Daher werden fast alle Hadronen gefunden, fast alle Elektronen mit derartigen HAC-Energien nicht als solche identifiziert. Die überwiegende Zahl der Teilchen ist jedoch in der Klasse bei  $0 \text{ MeV}$ . Für sie liegt von dieser Variablen keine Entscheidung vor.

#### 4.4.9 Höhere Impulse

Die letzten Abschnitte haben gezeigt, daß der Hadron-Elektron-Separator das mit Abstand stärkste Instrument zur Identifikation von (Konversions-)Elektronen ist. Das hängt damit zusammen, daß bei den niedrigen Impulsen der Teilchen ihre Schauer im Kalorimeter zu klein sind, um bei der gegebenen Segmentierung ihre Energie auf mehrere Segmente zu verteilen. Daher sind nur minimale Informationen über die Schauerform vorhanden. 70% aller Hadronen zeigen zudem keine Energie im zentralen HAC-Turm.

Andererseits ist bekannt, daß neuronale Netze unter Benutzung des Kalorimeters allein bei Energien von mehr als 10 GeV mit hoher Effizienz und Diskriminanz Elektronen zu identifizieren vermögen [1]. In dem Zwischenbereich von 2 bis 10 GeV sollte es aber einen Übergang geben, bei dem es sich lohnt, ein neuronales Netz wie *hesnn* mit Kalorimeter *und* HES zu trainieren. Will man sich nicht auf die Monte-Carlo-Simulation verlassen, so muß allerdings zunächst ein geeignetes Trainingsbeispiel für derartige Elektronen und Hadronen gefunden werden (vgl. Abschnitt 4.2.1).

Tendenziell läßt sich der mit dem Impuls wachsende Vorteil des *hesnn* gegenüber dem einfachen HES-Schnitt bereits in dem vorliegenden Datensatz mit Konversionselektronen erkennen. Dazu wurde das neuronale Netz dreimal trainiert, und zwar einmal mit allen Trainingsdaten (wie in Abschn. 4.4.8 beschrieben), zum zweiten mit allen Teilchen, deren Spuren einen Impuls von mindestens 1,0 GeV/c aufwiesen, und schließlich nur noch mit den Teilchen eines Impulses von mehr als 1,5 GeV/c (vgl. Abb. 4.13).

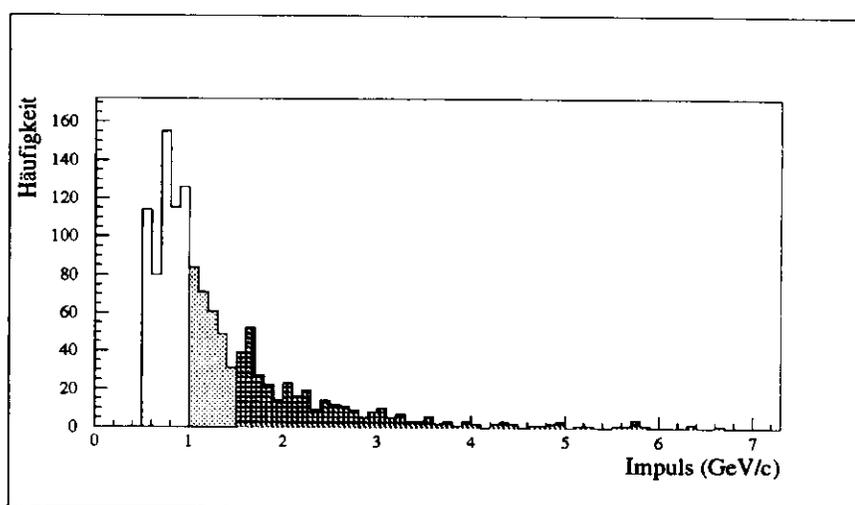


Abbildung 4.13: Impulsverteilung der (Trainings- und) Testdaten für drei ansonsten gleiche Elektronenfinder mit neuronalen Netzwerken.

Die entstandenen Elektronenfinder erreichten auf entsprechenden Teilmengen des Testdatensatzes die Effizienzen und Diskriminanz

$$E_{0,5} = (75,0 \pm 1,8)\% \quad D_{0,5} = (76,2 \pm 1,7)\%$$

$$E_{1,0} = (71,6 \pm 2,7)\% \quad D_{1,0} = (82,8 \pm 2,0)\%$$

$$E_{1,5} = (70,3 \pm 3,9)\% \quad D_{1,5} = (86,8 \pm 2,3)\%$$

Die Effizienzen wurden derart gewählt, daß sie bei Berücksichtigung der Spur-Isle-Zuordnung gleich werden. Wie diese Umrechnung durchgeführt wird, ist im nächsten Abschnitt beschrieben. Hier nur die Ergebnisse: Mit den Zuordnungseffizienzen und Diskriminanz

$$E_{Z;0,5} = (81,5 \pm 1,1)\% \quad D_{Z;0,5} = (53,3 \pm 1,0)\%$$

$$E_{Z;1,0} = (85,4 \pm 1,4)\% \quad D_{Z;1,0} = (52,4 \pm 1,3)\%$$

$$E_{Z;1,5} = (86,8 \pm 2,0)\% \quad D_{Z;1,5} = (53,3 \pm 1,6)\%$$

erhält man die vergleichbaren Effizienzen und Diskriminanz

$$E_{0,5} = (61,2 \pm 1,7)\% \quad D_{0,5} = (90,2 \pm 0,8)\%$$

$$\begin{aligned}
 E_{1,0} &= (61,1 \pm 2,6)\% & D_{1,0} &= (91,8 \pm 1,0)\% \\
 E_{1,5} &= (61,0 \pm 3,7)\% & D_{1,5} &= (93,8 \pm 1,1)\%
 \end{aligned}$$

Diese Werte ergeben die Signalverstärkungen  $S_{0,5} = 6,2 \pm 0,5$ ,  $S_{1,0} = 7,5 \pm 0,9$  und  $S_{1,5} = 9,9 \pm 1,9$ . Sie sind im Vergleich zu den entsprechenden Verstärkungen durch den einfachen HES-Schnitt in Abb. 4.14 aufgetragen. Die Abszisse bezeichnet den mittlere Impuls der am Trai-

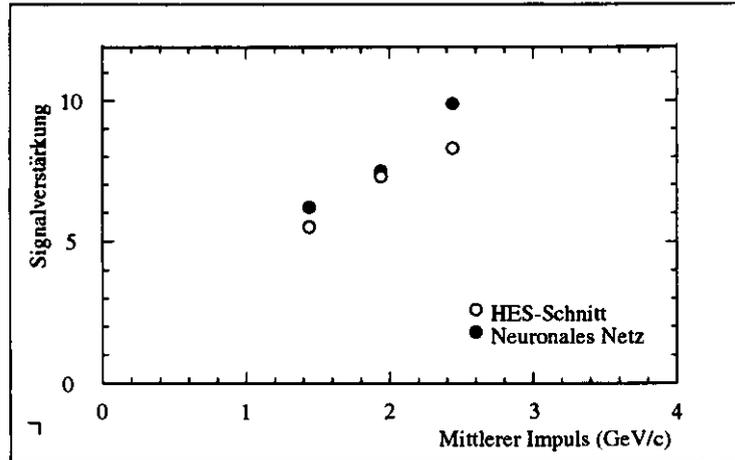


Abbildung 4.14: Signalverstärkung des einfachen HES-Schnittes und des Elektronenfinders *hesnn*, trainiert und getestet auf Teilmengen des Datensatzes (vgl. letzte Abbildung), aufgetragen gegen den mittleren Impuls der Trainings- und Testdaten.

ning und Test des jeweiligen Elektronenfinders beteiligten Spuren. Die Unterschiede zwischen den Elektronenfindern mit neuronalem Netz und den einfachen HES-Schnitten sind außer im Fall  $p > 1,0 \text{ GeV}/c$  signifikant. Betrachtet man das Effizienz-Diskriminanz-Diagramm des Elektronenfinders für Teilchen ab  $1,0 \text{ GeV}/c$  (nicht abgebildet), so fällt auf, daß bei der geforderten Effizienz die Diskriminanz dieses Elektronenfinders untypisch klein ist. Seine geringe Signalverstärkung ist also ein statistischer Effekt.

Die Verbesserungen gegenüber dem einfachen HES-Schnitt betragen für die drei Impulsbereiche  $(\Delta S/S)_{0,5} = 13,5\%$ ,  $(\Delta S/S)_{1,0} = 1,7\%$  und  $(\Delta S/S)_{1,5} = 19,7\%$ . Diese Entwicklung deutet an, daß die relative Verbesserung des Elektronenfinders mit neuronalem Netzwerk gegenüber dem einfachen HES-Schnitt mit wachsendem Impuls leicht steigt. Die Statistik reicht jedoch für einen strengen Beleg nicht aus.

#### 4.4.10 Berücksichtigung der Spur-Isle-Zuordnung

Die in den letzten Abschnitten genannten Effizienzen und Diskriminanzen bezogen sich auf die Spuren mit zugeordneter *Isle*. 75,1% Effizienz bedeutete also, daß 75,1% aller Elektronenspuren, denen sich eine HES-*Isle* zuordnen ließ, als Elektronen identifiziert werden konnten. Entsprechend erkannte beispielsweise *hesnn* 79% aller Hadronen *mit* zugeordneter *Isle* als solche. Interessanter ist aber, welche Anteile *aller* Elektronenspuren und Hadronenspuren erkannt wurden. Hier kommt es auf die Spur-*Isle*-Zuordnung an. Sie ist daher als Teil des Elektronenfinders anzusehen. Da nur  $E_Z = 81,5\%$  aller Elektronenspuren eine *Isle* verursachten und von diesen nur 75,1% als

Elektronenspür erkannt wurden, bleibt als Gesamteffizienz nur

$$E_{mit Z} = 75\% \cdot 81,5\% = 61,2\%$$

Die Diskriminanz erhöht sich dagegen, da durch die Zuordnung bereits über die Hälfte aller Hadronen abgelehnt wurden. Im nächsten Abschnitt werden die Effizienzen und Diskriminanz aller Elektronenfinder unter Berücksichtigung der Zuordnung miteinander verglichen. Sie wurden berechnet gemäß

$$\begin{aligned} E_{mit Z} &= E_{ohne Z} \cdot E_Z \\ D_{mit Z} &= D_{ohne Z} \cdot \bar{D}_Z + D_Z \end{aligned} \quad (4.2)$$

#### 4.4.11 Zusammenfassung

Elektronenfinder	Effizienz (%)	Diskriminanz (%)	Signalverstärkung
HES-Schnitt	$61,1 \pm 1,7$	$88,9 \pm 0,9$	$5,5 \pm 0,4$
$p > 1,0 \text{ GeV}/c$	$61,1 \pm 2,6$	$91,7 \pm 1,0$	$7,3 \pm 0,9$
$p > 1,5 \text{ GeV}/c$	$61,0 \pm 3,7$	$92,6 \pm 1,2$	$8,3 \pm 1,4$
HES u. HAC	$61,1 \pm 1,7$	$89,5 \pm 0,9$	$5,8 \pm 0,5$
Alle HES-Variablen	$61,1 \pm 1,7$	$89,4 \pm 0,9$	$5,7 \pm 0,5$
Alle Kalorimeter-V.	$61,3 \pm 2,1$	$73,5 \pm 1,8$	$2,3 \pm 0,2$
Alle Variablen	$61,1 \pm 1,7$	$90,6 \pm 0,8$	$6,5 \pm 0,6$
<i>hesnn</i>	$61,2 \pm 1,7$	$90,2 \pm 0,8$	$6,2 \pm 0,5$
$p > 1,0 \text{ GeV}/c$	$61,1 \pm 2,6$	$91,8 \pm 1,0$	$7,5 \pm 0,9$
$p > 1,5 \text{ GeV}/c$	$61,0 \pm 3,7$	$93,8 \pm 1,1$	$9,9 \pm 1,9$

Tabelle 4.4: Unkorrigierte Eigenschaften einiger Elektronenfinder. Der Elektronenfinder „Alle Kalorimetervariablen“ wurde mit allen Spuren (auch denen ohne zugeordnete *Isle*) neu trainiert. Für die übrigen Elektronenfinder wurden die Größen gemäß Abschnitt 4.4.10 berechnet.

Die Tabelle 4.4 zeigt die Trainingsergebnisse für die verschiedenen getesteten Elektronenfinder im Vergleich. Die Zahlen beziehen sich auf die Gesamtzahl der Hadronen- und Elektronenspuren, die für diese Untersuchung ausgewählt worden sind. Die Wirkung der Spur-*Isle*-Zuordnung ist also berücksichtigt. Der Elektronenfinder „Alle Kalorimetervariablen“ enthält keine Spur-*Isle*-Zuordnung, da die *Isles* Teil der HES-Informationen sind, die hier nicht benutzt werden sollten.

Es wird besonders darauf hingewiesen, daß die Signalverstärkung für *hesnn* 14% (oberhalb 1,5 *GeV* 20%) größer ist als für den einfachen HES-Schnitt.

## 4.5 Korrekturen

### 4.5.1 Einleitung

Jeder Elektronenfinder bezeichnet einen bestimmten Teil der Spuren als Elektronen, einen anderen als Untergrund. Diese Aussagen werden mit der Identität der Teilchen verglichen, die durch die

Art der Spuren definiert ist. Diese Identität kann aber fehlerhaft bestimmt worden sein. Die Hauptfehlerquellen sind dabei:

1. Der Konversionsfinder hat eine begrenzte Reinheit. Unter den „Konversionsspuren“ befinden sich daher auch Hadronen. Auf der anderen Seite gibt es unter den als Hadronspuren angesehenen restlichen Spuren noch Elektronen, nämlich Konversionselektronen und -positronen, die nicht als solche identifiziert wurden, und Elektronen aus Zerfallsprozessen oder das Streuelektron aus Untergrundereignissen der tiefunelastischen Streuung.
2. Bei der Spur-*Isle*-Zuordnung treten Fehler auf, wenn eine elektronische Spur fälschlich einer hadronischen *Isle* zugeordnet wird oder umgekehrt.

Auf den zweiten Fehler soll hier nicht korrigiert werden, da er als Eigenschaft des Elektronenfinders angesehen wird.

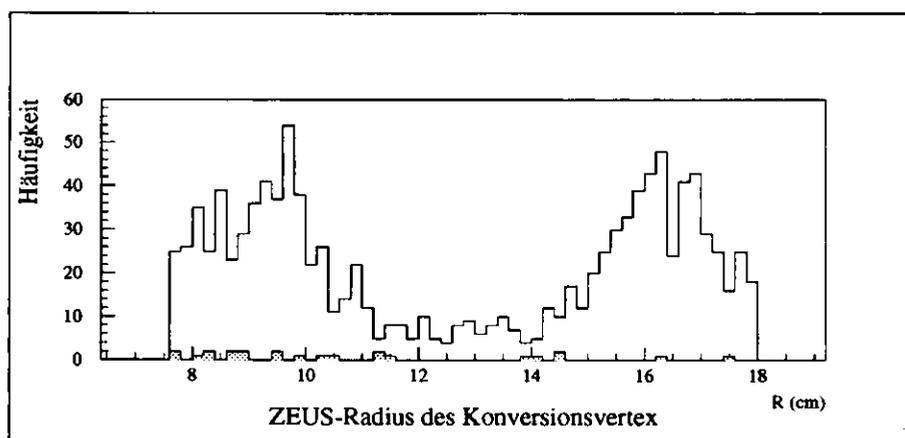


Abbildung 4.15: Entfernung des Konversionsvertex von der Strahlachse für ungleichnamig und gleichnamig (grau) geladene Spuren. Das innere Maximum wird erzeugt durch die Strahlrohrwand und die innere Wand des Vertexdetektors, das äußere durch die äußere Vertexdetektorwand und die innere Wand der Spurkammer. Konversionsspuren außerhalb eines Fensters von  $R = 7,5 \dots 18 \text{ cm}$  wurden aus dem Datensatz entfernt. Die gleichnamig geladenen Spuren dienen der Berechnung des Hadron-Untergrundes im Datensatz der Konversionselektronen.

#### 4.5.2 Der Konversionsfinder

Unter den vom Konversionsfinder selektierten Spurpaaren gegensätzlicher Ladung sind einige, die sich zufällig in einem Punkt berühren. Ihre Zahl wurde abgeschätzt aus der Zahl gleichnamig geladener Spuren, die dieselben Berührungskriterien erfüllen (vgl. Abb. 4.15). Es sind 25 im Vergleich zu 1127 Elektronenspuren. Nach Monte-Carlo-Untersuchungen von R. van Woudenberg wird der Untergrund dadurch um 30% unterschätzt ([9], S. 13). Korrigiert man darauf und setzt den systematischen Fehler mit 30% an, so erhält man im Elektronendatensatz einen Untergrund

von<sup>6</sup>

$$U_e = (3,2 \pm 0,9 \pm 1,0)\%. \quad (4.3)$$

Der Hadronendatensatz ist einfach aus den Restspuren zusammengesetzt. Wiederum nach einer Monte-Carlo-Untersuchung von R. van Woudenberg<sup>7</sup> enthält ein Datensatz aus der Photoproduktion (*Resolved photoproduction MC, 5 flavours*) 0,007% Elektronen aus semielektronischen Zerfällen und  $(4,1 \pm 0,2)\%$  Elektronen aus Konversionen. Der vorliegende Datensatz ist allerdings zusammengesetzt aus mehreren Datensätzen, an die verschiedene Schnitte angelegt wurden. L. Koepke gibt den Anteil der Elektronspuren semielektronischer Zerfälle an allen Spuren mit  $(0,5 \pm 0,5)\%$ , den Anteil der Elektronspuren aus der Photonkonversion mit  $(4 \pm 1)\%$  an<sup>8</sup>. Die Effizienz des Konversionsfinders CONVERT2 wird in [9] mit  $E = 541/1070 = (50,6 \pm 0,4)\%$  für Konversionspaare mit mindestens einem Konversionspartner mit  $p > 500 \text{ MeV}/c$  angegeben (Tab. 7, S. 17). Diese Zahl steigt auf  $E = 210/315 = (66,7 \pm 1,2)\%$  für Paare mit mindestens einer Spur von  $p > 1 \text{ GeV}/c$ . Der vorliegende Datensatz verlangt einen Partner mit mehr als  $700 \text{ MeV}/c$ . Die Effizienz soll daher angenommen werden zu

$$E_{\text{CONVERT2}} = (55 \pm 1,2 \pm 20)\%.$$

Der Anteil nicht von CONVERT2 gefundener Konversionselektronen unter den Restspuren beläuft sich damit auf  $U_{\text{Konv}} = 45\% \cdot 4\% = (1,8 \pm 0,1 \pm 1,3)\%$ . Zusammen mit den Elektronen aus semielektronischen Zerfällen ergibt sich ein elektronischer Untergrund im hadronischen Datensatz von

$$U_h = (2,3 \pm 1,8)\%. \quad (4.4)$$

Der angegebene Fehler ist systematisch.

Die korrigierten Werte für Effizienz, Diskriminanz und Signalverstärkung sind für einige Elektronenfinder in Tab. 4.5 eingetragen<sup>9</sup>.

Bei der Rechnung wurde angenommen, daß Konversionsfinder und Elektronenfinder unabhängig sind, d. h., daß die Effizienz des Elektronenfinders für Elektronen, die als Konversionselektronen erkannt wurden, dieselbe ist wie für Elektronen unter den Restspuren.

## 4.6 Zusammenfassung

Mit Hilfe einer Datei mit Konversionselektronen konnte ein neuronales Netz darauf trainiert werden, Elektronen im Impulsbereich von  $0,5$  bis  $3 \text{ GeV}/c$  zu identifizieren. Der so konstruierte Elektronenfinder *hesnn* erreichte eine Effizienz und Diskriminanz von

$$E_{\text{hesnn}} = (62,9 \pm 1,9 \pm 0,7)\% \quad \text{und} \quad D_{\text{hesnn}} = (91,4 \pm 0,8 \pm 1,1)\%$$

für einen Testdatensatz der Photoproduktion mit Konversionselektronen. Das Impulsspektrum ist noch einmal in Abb. 4.16 dargestellt. Der Elektronenfinder *hesnn* ist damit in der Lage,

<sup>6</sup> Der statistische Fehler wird zuerst angeführt. Er wurde mit berechnet gemäß  $\sigma = \sqrt{r(1-r)/n}$ , wobei  $r$  die relative Häufigkeit der jeweils gezählten Teilchen und  $n$  die Größe der Stichprobe ist.

<sup>7</sup> Persönliche Notiz.

<sup>8</sup> Persönliche Notiz.

<sup>9</sup> Ausführliche Rechnung in Anhang B.3.

Elektronenfinder	Effizienz (%)	Diskriminanz (%)	Signalverstärkung
HES-Schnitt	$62,8 \pm 1,9 \pm 0,6$	$90,1 \pm 0,9 \pm 1,1$	$6,3 \pm 0,6^{+0,8}_{-0,6}$
$p > 1,0 \text{ GeV}/c$	$62,9 \pm 2,7 \pm 0,7$	$93,0 \pm 1,1 \pm 1,1$	$8,9 \pm 1,3^{+1,7}_{-1,2}$
$p > 1,5 \text{ GeV}/c$	$62,8 \pm 3,9 \pm 0,7$	$93,9 \pm 1,2 \pm 1,1$	$10,3 \pm 2,1^{+2,4}_{-1,6}$
HES u. HAC	$62,8 \pm 1,9 \pm 0,7$	$90,8 \pm 0,9 \pm 1,1$	$6,8 \pm 0,7^{+0,9}_{-0,7}$
Alle HES-Variablen	$62,8 \pm 1,9 \pm 0,7$	$90,6 \pm 0,9 \pm 1,1$	$6,7 \pm 0,6^{+0,9}_{-0,7}$
Alle Kalorimeter-V.	$62,5 \pm 2,2 \pm 0,5$	$74,3 \pm 1,8 \pm 0,8$	$2,4 \pm 0,2^{+0,1}_{-0,1}$
Alle Variablen	$62,9 \pm 1,9 \pm 0,7$	$91,9 \pm 0,8 \pm 1,1$	$7,7 \pm 0,8^{+1,2}_{-0,9}$
<i>hesnn</i>	$62,9 \pm 1,9 \pm 0,7$	$91,4 \pm 0,8 \pm 1,1$	$7,4 \pm 0,7^{+1,1}_{-0,8}$
$p > 1,0 \text{ GeV}/c$	$62,9 \pm 2,7 \pm 0,7$	$93,1 \pm 1,1 \pm 1,1$	$9,1 \pm 1,4^{+1,8}_{-1,2}$
$p > 1,5 \text{ GeV}/c$	$62,9 \pm 3,9 \pm 0,7$	$95,2 \pm 1,2 \pm 1,2$	$13,0 \pm 3,1^{+4,1}_{-2,4}$

Tabelle 4.5: Korrigierte Eigenschaften einiger Elektronenfinder.

Elektronensignale mit Impulsen von 0,5 bis 3  $\text{GeV}$  bei hoher Effizienz auf das  $(7,4 \pm 0,7 \pm 1,1)$ -fache zu verstärken.

Wie erwartet erbringt der Hadron-Elektron-Separator den Hauptanteil der Information. Dennoch lohnt sich der gemeinsame Einsatz von HES und Kalorimeter in *hesnn*. Der Elektronenfinder *hesnn* übertrifft die Entscheidung des HES-Signals allein um 17%. Der Unterschied ist signifikant. Das Kalorimeter bleibt mit einer Signalverstärkung von etwa 2,4 weit hinter den Leistungen des HES mit und ohne neuronales Netz zurück.

Die Daten deuten darauf hin, daß neuronale Netze schon bei Impulsen ab 2,0  $\text{GeV}/c$  genügend Informationen vom Kalorimeter bekommen, um die HES-Entscheidung um 25% zu übertreffen.

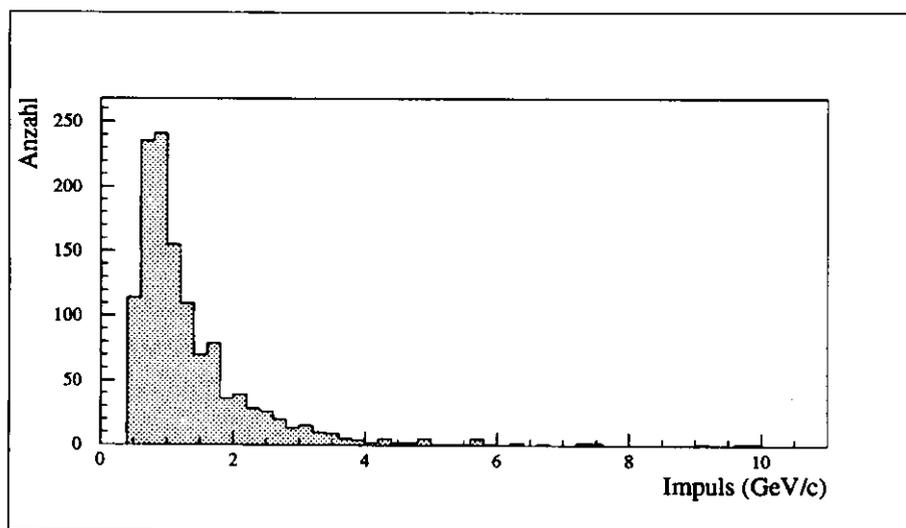


Abbildung 4.16: Impulsverteilung der an der Auswertung beteiligten Spuren.

## Kapitel 5

# Zusammenfassung

Es wurde ein Elektronenfinder für das ZEUS-Experiment konstruiert, der auf Elektronen im Impulsbereich 0,5 bis 3 GeV/c spezialisiert ist. Erstmals verknüpft dieser Elektronenfinder die Informationen des Kalorimeters mit denen des Hadron-Elektron-Separators (HES). Zu diesem Zweck wurde der Elektronenfinder mit einem neuronalen Netzwerk (NN) ausgerüstet. Der Elektronenfinder – sein Name ist *hesnn* – erzielte auf einem Datensatz von Photoproduktionsereignissen mit Konversionselektronen eine Effizienz und Diskriminanz<sup>1</sup> von

$$E_{hesnn} = (62,9 \pm 2,2 \pm 0,6)\% \quad \text{und} \quad D_{hesnn} = (91,4 \pm 0,8 \pm 1,1)\%.$$

Aus diesen beiden Größen läßt sich berechnen, daß der Elektronenfinder *hesnn* das Verhältnis Elektronen zu Untergrund um den Faktor  $\frac{E}{1-D} = 7,4$  erhöht (Signalverstärkung). Zum Vergleich: Ein neuronales Netz, dem nur Kalorimeterinformationen zur Verfügung gestellt wurden, erzielte bei gleicher Effizienz eine Signalverstärkung von 2,4. Ein einfacher Schnitt in der Variablen HES-Signal erreichte eine Signalverstärkung von 6,3. Der Elektronenfinder *hesnn* übertrifft den HES-Schnitt also um 17%, den Kalorimeter-Elektronenfinder um ein Vielfaches. Die Unterschiede sind signifikant ( $\alpha = 5\%$ ). Es wurden Hinweise gegeben, wie Trainingsdatensätze mit Elektronen und Hadronen höherer Energien gewonnen werden können (vgl. Abschn. 4.2.1). Mit solchen Datensätzen läßt sich der Arbeitsbereich des Elektronenfinders *hesnn* problemlos auf höhere Impulse ausdehnen.

Als Vorbereitung zur Konstruktion des Elektronenfinders wurde eine Einführung in die Grundlagen der Mathematik und der Anwendung neuronaler Netzwerke gegeben. Anhand von Beispielen wurden verschiedene Methoden zur Konvergenzverbesserung erprobt. Zahlreiche Darstellungen vermitteln anschauliche Vorstellungen über den mathematischen Prozeß des Trainings neuronaler Netzwerke.

---

<sup>1</sup>Die Diskriminanz bezeichnet die Wahrscheinlichkeit, ein Hadron als solches zu erkennen und als Nicht-Elektron zu diskriminieren.



## Anhang A

# Anhang zum Kapitel „Neuronale Netzwerke“

### A.1 Der *back propagation*-Algorithmus

Der *back propagation*-Algorithmus berechnet die partiellen Ableitungen der Fehlerfunktion eines *feed forward*-Netzes, indem es den Fehler rückwärts durch das Netz fortführt. Dies soll hier im Detail beschrieben werden. Das zu betrachtende *feed forward*-Netz sei rekursiv definiert durch

$$\begin{aligned} f_{i_m}(\vec{x}) &= o_{mi_m}(\vec{x}) \quad i_m = 1 \dots n_m, \\ o_{li_l}(\vec{o}_{l-1}) &= g \left( \underbrace{t_{li_l} + \sum_{i_{l-1}=1}^{n_{l-1}} w_{li_l i_{l-1}} o_{l-1, i_{l-1}}}_{a_{li_l}} \right), \quad (A.1) \\ &\quad i_l = 1 \dots n_l, l = 2 \dots m, \\ o_{1, i_1}(\vec{x}) &= g \left( \underbrace{t_{1, i_1} + \sum_{i_0=1}^{n_0} w_{1, i_1, i_0} x_{i_0}}_{a_{1, i_1}} \right), \\ &\quad i_1 = 1 \dots n_1. \end{aligned}$$

Hier bedeutet  $f$  die Ausgabefunktion des neuronalen Netzes. Sie kann zusammengesetzt sein aus  $n_m$  Komponenten (Ausgabeknoten). Der Index  $m$  bedeutet hier, daß das Netzwerk  $m$  Lagen (Ebenen) hat, die Eingabelage nicht mitgerechnet. Die Komponenten der Ausgabefunktion werden durch den Index  $i_m = 1 \dots n_m$  durchgezählt.

Der Buchstabe  $o$  bezeichnet die Ausgabe eines Knotens. Seine beiden Indizes stehen für die Lage  $l$  und die Knotennummer  $i_l = 1 \dots n_l$  innerhalb der Lage. In der  $m$ -ten Lage sind die Knotenausgaben  $o_{mi_m}$  identisch mit den Komponenten der Ausgabefunktion  $f_{i_m}$ . In der 0-ten Lage (Eingabeebene) sind die  $o_{0i_0}$  einfach die Eingabewerte  $x_{i_0}$ .

Die Schwellwerte  $t$  haben dieselben Indizes wie die Ausgaben. Das Gewicht  $w_{li_l i_{l-1}}$  verbindet den  $i_{l-1}$ -ten Knoten der Lage  $l-1$  mit dem  $i_l$ -ten Knoten der Lage  $l$ . Das Gewicht selbst wird

der Lage  $l$  zugerechnet.

Im folgenden soll die Ausgabefunktion des Netzwerkes nicht mehr als Funktion der Eingabe  $\vec{x}$ , sondern als Funktion der Schwellwerte und Gewichte betrachtet werden.

Ist die Fehlerfunktion

$$E(\vec{t}, \vec{w}) = \frac{1}{2} \sum_{i_m=1}^{n_m} \left( f_{i_m}(\vec{t}, \vec{w}) - f_{soll, i_m}(\vec{t}, \vec{w}) \right)^2, \quad (\text{A.2})$$

so lassen sich ihre partiellen Ableitungen folgendermaßen berechnen:

$$\begin{aligned} \frac{\partial E}{\partial t_{qpq}} &= \sum_{i_m=1}^{n_m} \left( f_{i_m}(\vec{t}, \vec{w}) - f_{soll, i_m}(\vec{t}, \vec{w}) \right) \times \frac{\partial f_{i_m}}{\partial t_{qpq}} \\ &= \sum_{i_m=1}^{n_m} \underbrace{\left( o_{i_m}(\vec{t}, \vec{w}) - o_{soll, i_m}(\vec{t}, \vec{w}) \right)}_{= d_{mi_m}} \times g'(a_{mi_m}) \times \frac{\partial a_{mi_m}}{\partial t_{qpq}} \\ &= \sum_{i_m=1}^{n_m} d_{mi_m} \times \frac{\partial}{\partial t_{qpq}} \left( t_{mi_m} + \sum_{i_{m-1}=1}^{n_{m-1}} w_{mi_m i_{m-1}} o_{m-1, i_{m-1}} \right) \\ &= \sum_{i_m=1}^{n_m} d_{mi_m} \times \left( \delta_{qm} \delta_{pq i_m} + \sum_{i_{m-1}=1}^{n_{m-1}} w_{mi_m i_{m-1}} \frac{\partial o_{m-1, i_{m-1}}}{\partial t_{qpq}} \right) \\ &= \sum_{i_m=1}^{n_m} d_{mi_m} \delta_{qm} \delta_{pq i_m} \\ &\quad + \sum_{i_m=1}^{n_m} d_{mi_m} \sum_{i_{m-1}=1}^{n_{m-1}} w_{mi_m i_{m-1}} \times g'(a_{m-1, i_{m-1}}) \times \frac{\partial a_{m-1, i_{m-1}}}{\partial t_{qpq}} \\ &= \sum_{i_m=1}^{n_m} d_{mi_m} \delta_{qm} \delta_{pq i_m} \\ &\quad + \underbrace{\sum_{i_{m-1}=1}^{n_{m-1}} \sum_{i_m=1}^{n_m} d_{mi_m} \times w_{mi_m i_{m-1}} \times g'(a_{m-1, i_{m-1}})}_{d_{m-1, i_{m-1}}} \times \frac{\partial a_{m-1, i_{m-1}}}{\partial t_{qpq}} \\ &= \sum_{i_m=1}^{n_m} d_{mi_m} \delta_{qm} \delta_{pq i_m} \\ &\quad + \sum_{i_{m-1}=1}^{n_{m-1}} d_{m-1, i_{m-1}} \times \frac{\partial}{\partial t_{qpq}} \left( t_{m-1, i_{m-1}} + \sum_{i_{m-2}=1}^{n_{m-2}} w_{m-1, i_{m-1}, i_{m-2}} o_{m-2, i_{m-2}} \right) \\ &= \sum_{i_m=1}^{n_m} d_{mi_m} \delta_{qm} \delta_{pq i_m} \\ &\quad + \sum_{i_{m-1}=1}^{n_{m-1}} d_{m-1, i_{m-1}} \delta_{qm-1} \delta_{pq i_{m-1}} \\ &\quad + \sum_{i_{m-1}=1}^{n_{m-1}} d_{m-1, i_{m-1}} \times \sum_{i_{m-2}=1}^{n_{m-2}} w_{m-1, i_{m-1}, i_{m-2}} \times g'(a_{m-2, i_{m-2}}) \times \frac{\partial a_{m-2, i_{m-2}}}{\partial t_{qpq}} \end{aligned} \quad (\text{A.3})$$

usw.

$$\begin{aligned}
&= \sum_{l=1}^m \sum_{i_l=1}^{n_l} d_{li} \delta_{q_l} \delta_{p_q i_l} \\
&= d_{qp}
\end{aligned} \tag{A.4}$$

$$\begin{aligned}
\frac{\partial E}{\partial w_{qpq p_{q-1}}} &= \sum_{i_m=1}^{n_m} \left( f_{i_m}(\vec{t}, \vec{w}) - f_{\text{oll}, i_m}(\vec{t}, \vec{w}) \right) \times \frac{\partial f_{i_m}}{\partial w_{qpq p_{q-1}}} \\
&= \sum_{i_m=1}^{n_m} d_{mi_m} \times \frac{\partial}{\partial w_{qpq p_{q-1}}} \left( t_{mi_m} + \sum_{i_{m-1}=1}^{n_{m-1}} w_{mi_m i_{m-1}} o_{m-1, i_{m-1}} \right) \\
&= \sum_{i_m=1}^{n_m} d_{mi_m} \sum_{i_{m-1}=1}^{n_{m-1}} o_{m-1, i_{m-1}} \delta_{qm} \delta_{p_q i_m} \\
&\quad + \underbrace{\sum_{i_{m-1}=1}^{n_{m-1}} \sum_{i_m=1}^{n_m} d_{mi_m} \times w_{mi_m i_{m-1}} \times g'(a_{m-1, i_{m-1}})}_{d_{m-1, i_{m-1}}} \times \frac{\partial a_{m-1, i_{m-1}}}{\partial w_{qpq p_{q-1}}} \\
&= \sum_{i_m=1}^{n_m} d_{mi_m} \sum_{i_{m-1}=1}^{n_{m-1}} o_{m-1, i_{m-1}} \delta_{qm} \delta_{p_q i_m} \\
&\quad + \sum_{i_{m-1}=1}^{n_{m-1}} d_{m-1, i_{m-1}} \sum_{i_{m-2}=1}^{n_{m-2}} o_{m-2, i_{m-2}} \delta_{qm-1} \delta_{p_q i_{m-1}} \\
&\quad + \sum_{i_{m-1}=1}^{n_{m-1}} d_{m-1, i_{m-1}} \times \sum_{i_{m-2}=1}^{n_{m-2}} w_{m-1, i_{m-1}, i_{m-2}} \times g'(a_{m-2, i_{m-2}}) \times \frac{\partial a_{m-2, i_{m-2}}}{\partial w_{qpq p_{q-1}}}
\end{aligned} \tag{A.5}$$

usw.

$$\begin{aligned}
&= \sum_{i_1=1}^{n_1} \sum_{i_0=1}^{n_0} d_{1, i_1} x_{i_0} \delta_{q_0} \delta_{p_q i_0} + \sum_{l=2}^m \sum_{i_l=1}^{n_l} \sum_{i_{l-1}=1}^{n_{l-1}} d_{li} o_{l-1, i_{l-1}} \delta_{q_l} \delta_{p_q i_l} \\
&= \begin{cases} d_{1p_1} x_{p_1} & \text{für } q = 1 \\ d_{qp_q} o_{q-1, p_{q-1}} & \text{sonst.} \end{cases}
\end{aligned}$$

Die partiellen Ableitungen lassen sich also rekursiv berechnen durch

$$\begin{aligned}
\frac{\partial E}{\partial t_{qp_q}} &= d_{qp_q} \\
\frac{\partial E}{\partial w_{qp_q p_{q-1}}} &= \left\{ \begin{array}{ll} d_{1p_1} x_{p_1} & \text{für } q = 1 \\ d_{qp_q} o_{q-1, p_{q-1}} & \text{sonst} \end{array} \right\} \text{ mit} \\
d_{li} &= g'(a_{l, i_l}) \sum_{i_{l+1}=1}^{n_{l+1}} d_{i_{l+1}, i_{l+1}} w_{l+1, i_{l+1}, i_l} \\
&\quad i_l = 1 \dots n_l, l = 1 \dots m-1 \\
d_{mi_m} &= \left( f_{i_m}(\vec{t}, \vec{w}) - f_{\text{oll}, i_m}(\vec{t}, \vec{w}) \right) g'(a_{mi_m}) \\
&\quad i_m = 1 \dots n_m
\end{aligned} \tag{A.6}$$

Zur Berechnung aller partiellen Ableitungen werden daher von der Ausgangsebene beginnend für alle Knoten die  $d_{ii}$  berechnet (*error back propagation*). Sie stellen bereits die Ableitungen nach den Schwellwerten dar. Die Ableitungen nach den Gewichten erhält man einfach durch Multiplikation mit den Ausgangswerten der entsprechenden Knoten der darunterliegenden Ebene.

## A.2 Detaillierte Darstellung eines Trainingsverlaufs

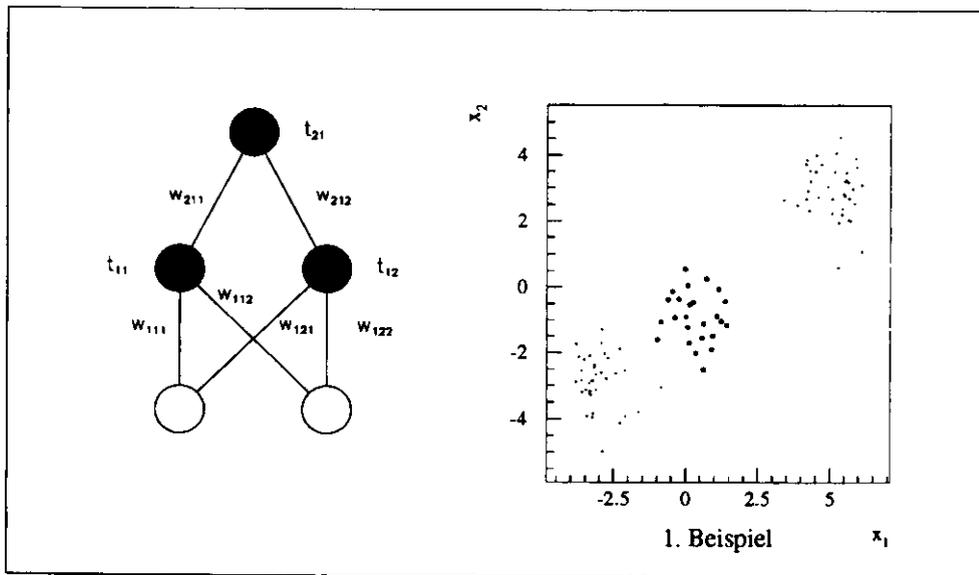


Abbildung A.1: Netzgeometrie und Übungsbeispiel zur Darstellung eines Trainingsvorgangs.

Das erste Beispiel aus Abschnitt 3.5.2 besteht aus drei kollinear angeordneten Gruppen von Datenpunkten, normalverteilt in einem zweidimensionalen Eingaberaum. Aufgabe des neuronalen Netzes ist es, durch zwei Schnitte die mittlere Gruppe (starke Punkte) von den anderen (schwache Punkte) zu trennen. Das Training mit Dämpfung  $\alpha = 0,99$  und einer Schrittweite von  $\eta = 0,01$  benötigte 2000 Korrekturschritte, sah das 100 Datenpunkte umfassende Trainingsbeispiel also zwanzigmal. Auf den folgenden Seiten ist die Entwicklung des Fehlers, den das Netz in jedem Schritt gemacht hat, die Einstellungen der Klassifikationsgrenzen in einigen Phasen des Trainings und die Entwicklung der Gewichte und Schwellwerte dargestellt.

Die kleinen Fenster der Abbildungen A.2 bis A.4 zeigen den Raum der Eingabevariablen mit den drei Gruppen von Datenpunkten, von denen eine (die starken Punkte in der Mitte) durch das neuronale Netzwerk von den anderen (die schwachen Punkte unten links und oben) getrennt werden soll. Das Netz bekommt in jedem Schritt einen der Datenpunkte zu sehen, klassifiziert diesen mit der momentanen Einstellung der Gewichte, macht dabei einen Fehler und wird diesem Fehler entsprechend korrigiert. Die Fehler sind auf jeder Seite oben als Funktion der Zahl der vorgenommenen Schritte dargestellt.

Die Position der Klassifikationsgrenzen ist zu Anfang zufällig gewählt. Ihre Lage im Raum der Eingangsvariablen ist in der ersten kleinen Darstellung (Schritt 1) durch zwei Geraden gekennzeichnet. Die stärkere dieser beiden Geraden repräsentiert den ersten, die schwächere den zweiten Knoten der verborgenen Ebene. Die letztendliche Entscheidung des Netzwerkes, also

die Funktion des Ausgangswertes in Abhängigkeit der beiden Eingabevariablen, ist durch ihre Äquipotentiallinien, also Linien gleichen Funktionswertes, verdeutlicht. Auf der ersten Seite liegt der gesamte Bildausschnitt im Übergangsbereich dieser Ausgangsfunktion, auf der zweiten Seite zieht sich dieser Übergang jedoch zusammen, und auf der dritten geht er in zwei distinkte Linien über.

In den Schritten 1 bis 60 ändern die Geraden ihre Lage nur geringfügig. Man erkennt daran, daß die Gewichte der verborgenen Ebene,  $w_{111}$ ,  $w_{112}$ ,  $w_{121}$  und  $w_{122}$  (Abb. A.1), deren Verhältnisse die Steigungen der Geraden definieren, und die Schwellwerte  $t_{11}$  und  $t_{12}$ , die den  $x$ - und  $y$ -Abschnitt der Geraden bestimmen, sich kaum ändern.

Die Äquipotentiallinien ändern ihre Gestalt dennoch stark. In Schritt 1 passen sie sich eher der schwächeren Geraden an, in Schritt 50 eher der stärkeren und in Schritt 60 wiederum der schwächeren. Das weist auf Schwankungen in den Gewichten der Ausgangsebene,  $w_{211}$  und  $w_{212}$ , hin. Hauptsächlich wird der Ausgangswert aber durch die Schwelle  $t_{21}$  bestimmt. Sie stellt diesen innerhalb der ersten 50 Schritte auf etwa  $f(x) = 0,2$  ein, was zu einem Fehler von 0,2 für die schwachen Punkte und 0,8 für die starken Punkte führt. Um die Summe der Fehler möglichst klein zu halten, werden die schwachen Punkte wegen ihrer größeren Zahl begünstigt.

Daß die Gewichte und Schwellen der verborgenen Ebene sich so geringfügig ändern im Vergleich zu denen der Ausgangsebene, hängt damit zusammen, daß sie wegen der zu Beginn des Trainings kleinen Gewichte der Ausgangsebene nur geringfügig den Ausgangswert beeinflussen können. Der Gradient der Ausgangsfunktion und damit auch der der Fehlerfunktion bzgl. dieser Parameter ist daher klein, was aufgrund der gradientengetriebenen Trainingsmethode zu kleinen Korrekturen führt.

Durch statistische Fluktuationen wird das Netz in den Schritten 25 bis 83 häufiger an den Punkten der oberen, rechten Gruppe trainiert als an denen der unteren, linken. Daneben ist sie noch räumlich etwas abgetrennt von den beiden anderen. Aus diesen Gründen stellen sich die Gewichte und Schwellen der verborgenen Ebene derart ein, daß sie die obere, rechte Gruppe von den übrigen trennen.

Im Fehlerdiagramm kann man ab Schritt 75 die Punkte der rechten, oberen Gruppe von denen der linken, unteren an ihrem etwas geringeren Fehler unterscheiden. Dieser Unterschied verstärkt sich infolge der wachsenden Gewichte bis zum Schritt 600.

Bei Schritt 600 wird der Eingangsraum ziemlich scharf in zwei Gebiete geteilt, ein oberes, für das das Netz einen Ausgangswert unter 0,1 liefert, und ein unteres, für das der Ausgangswert zwischen 0,35 und 0,45 schwankt. Für die Gruppe der schwachen Punkte oben rechts wird der Fehler damit kleiner als 0,1, für die anderen beiden Gruppen schwankt er zwischen 0,35 und 0,45 (schwache Punkte) bzw. 0,55 und 0,65 (starke Punkte). Die unteren schwachen Punkte sind gegenüber den starken immer noch begünstigt, da sie in der Überzahl sind (37 zu 26).

In dem Bild zu Schritt 600 erkennt man weiter, daß diese Entscheidung von *einem* Knoten der verborgenen Ebene getroffen wird: Die Äquipotentiallinien der Ausgangsfunktion sind parallel zu der schwachen Linie.

Ab Schritt 850 beginnt das Netz, die zweite Linie zur weiteren Differenzierung des Eingangsraumes zu nutzen. Dies ist ein kritischer Übergang. Hat das Netz erstmal einen leichten Nutzen aus der Linie gezogen, dann optimiert es die Lage der Linie schnell. Dies führt zu einer plötzlichen Verbesserung im Fehlerdiagramm. Die Korrekturen sind allerdings wegen der scharfen Grenzen und der damit verbundenen großen Gradienten so stark, daß die Linie überschlägt, so daß einige Punkte zwischenzeitlich sehr schlecht klassifiziert werden. Mit Hilfe der Dämpfung konvergiert das Netz aber dennoch in die gewünschte Lage.

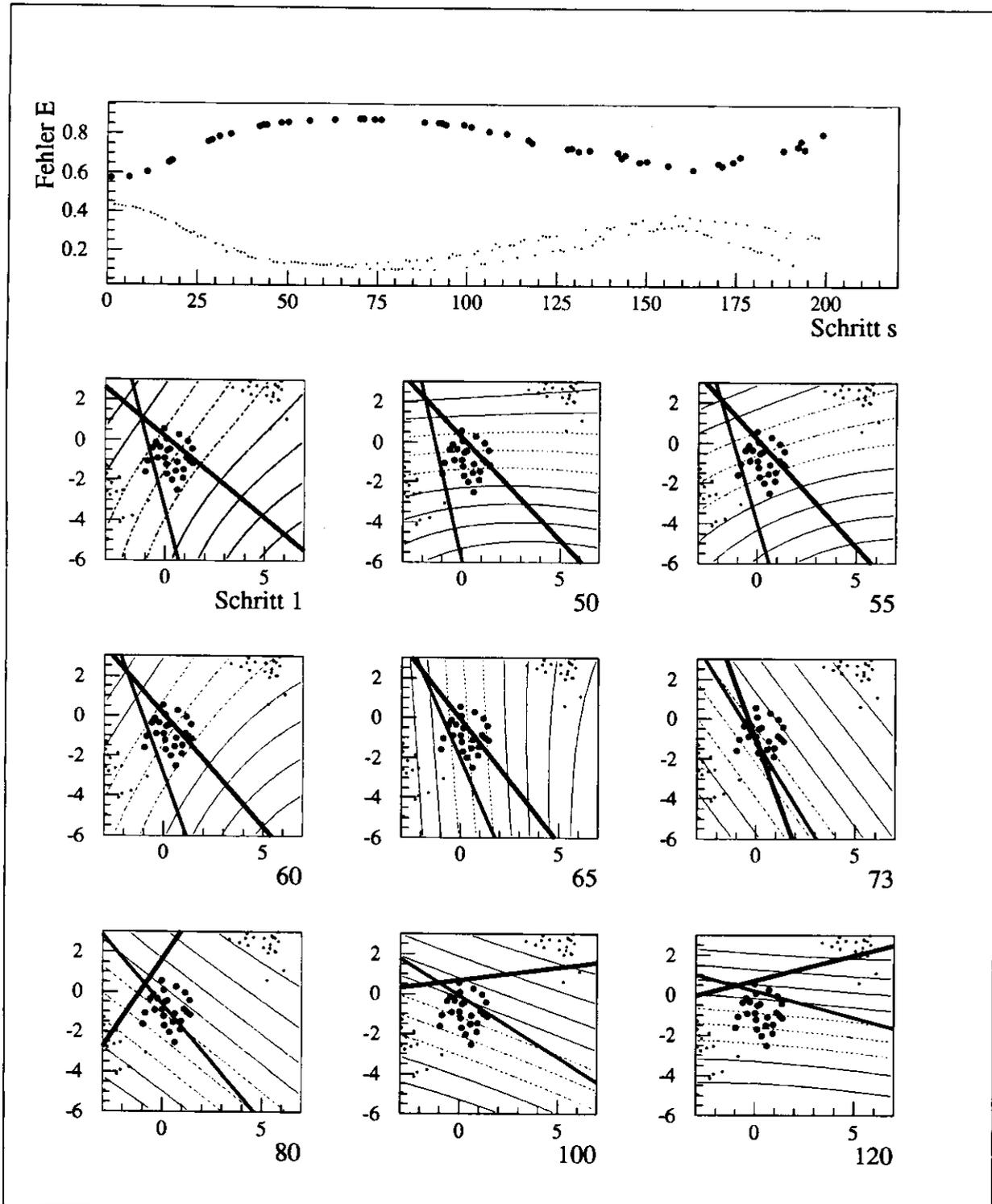


Abbildung A.2: Fehler des Netzes für jeden Datenpunkt und Momentaufnahmen der Klassifikationsgrenzen, wie sie von den Knoten der verborgenen Ebene definiert werden. Schritte 1 bis 120. Nähere Erklärungen im Text.

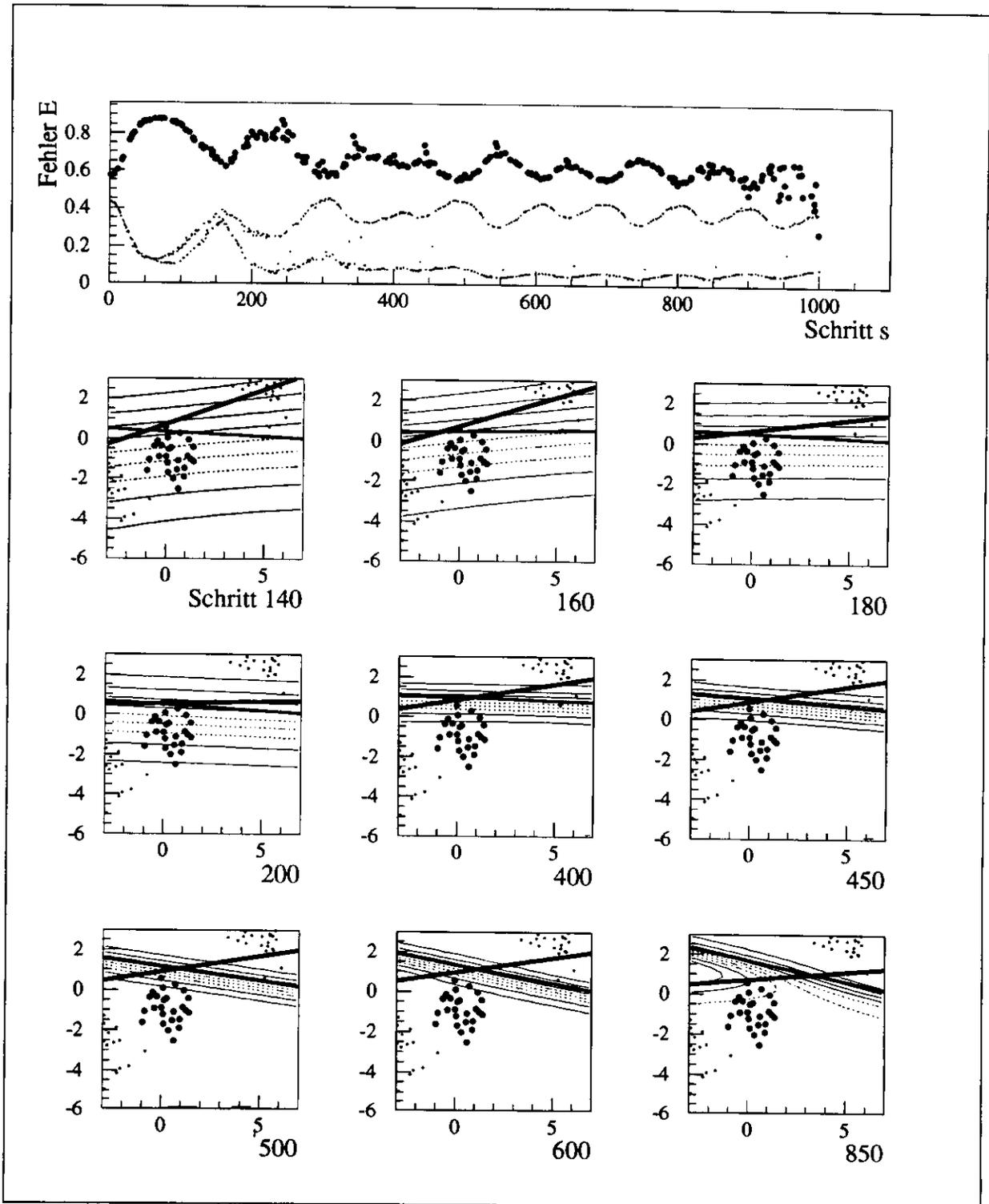


Abbildung A.3: Schritte 140 bis 850 des Trainings.

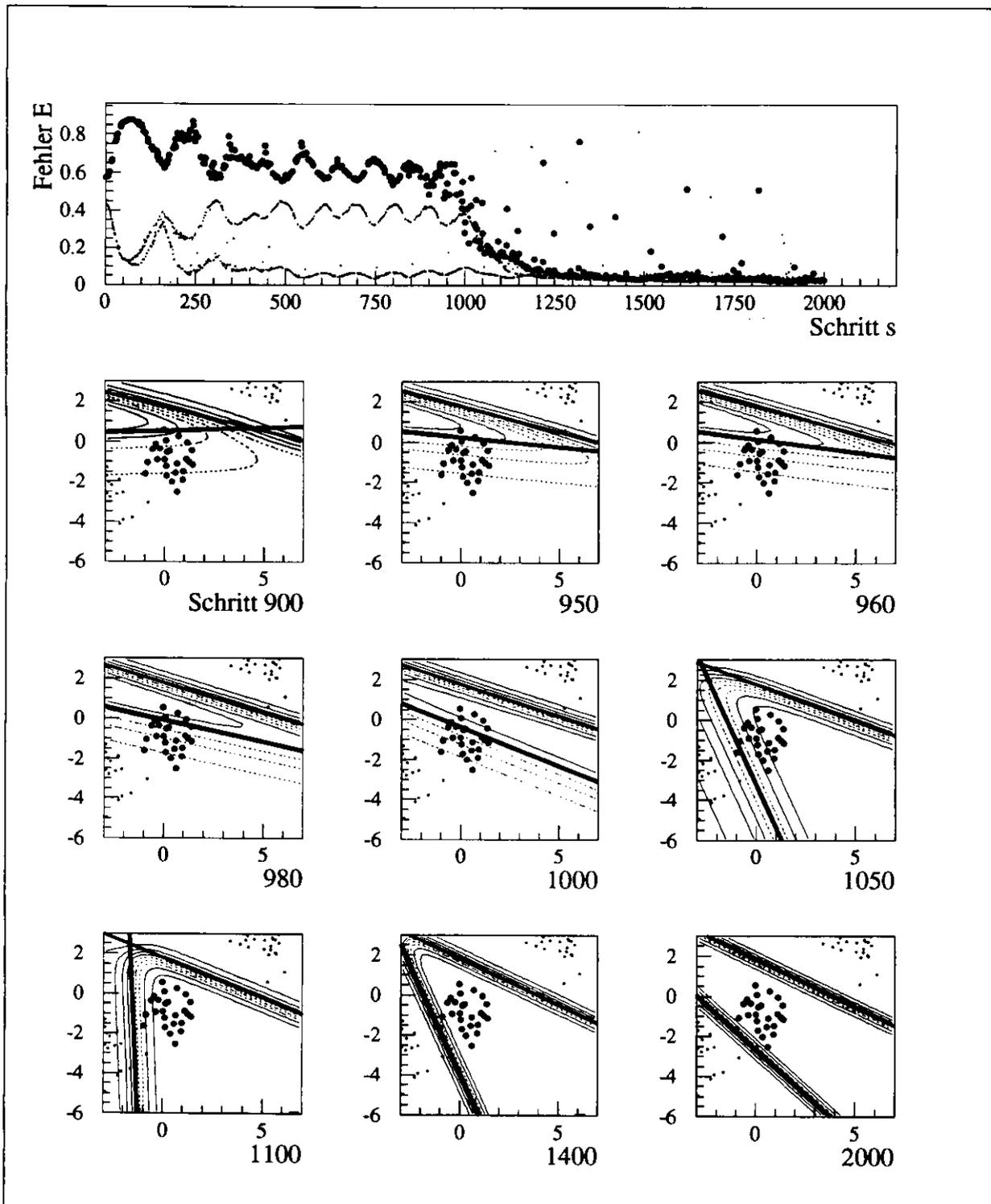


Abbildung A.4: Schritte 900 bis 200 des Trainings.

## Anhang B

# Anhang zum Kapitel „Niederenergetische Elektronen“

### B.1 Der Isle- und Island-Algorithmus

Der Hadron-Elektron-Separator ist segmentiert in rechteckige, das Rückwärtskalorimeter in quadratische Zellen. In der Regel erzeugt ein Teilchen in mehreren Zellen ein Signal. Aufgabe eines Gruppierungsalgorithmus ist es, die Signale aller von einem Teilchen erzeugten Zellen zusammenzufassen. Gleichzeitig soll möglichst vermieden werden, die von mehreren nahe beieinander eingetroffenen Teilchen erzeugten Signale zu einer Gruppe zusammenzufassen. Die geringste Entfernung zweier durch den Algorithmus trennbarer Teilchen stellt sein Auflösungsvermögen dar.

Der ZEUS-Standard-Algorithmus für die Kalorimeterdatenrekonstruktion arbeitet mit den dreidimensionalen Kalorimeterzellen. Er faßt alle angeschlagenen Zellen, die wenigstens einen Teil einer ihrer Flächen gemeinsam haben, zu einem sogenannten „Kondensat“ zusammen. Dieser Algorithmus vermeidet es, sicher die von einem Teilchen erzeugten Signale in zwei Gruppen aufzuspalten. Andererseits ist sein Auflösungsvermögen nicht besonders gut. Mitunter können selbst Teilchen, deren Schauerläufer sich kaum weniger als 40 cm nahe kommen und damit in benachbarten Zellen liegen in einem Kondensat zusammengefaßt werden.

Der *Island*-Algorithmus dagegen sucht nach lokalen Maxima in der deponierten Energie der Kalorimeterzellen und ordnet ihnen die umliegenden Signale zu. Eine Verschmelzung zweier Teilchen ist nur dann möglich, wenn sie den *Hauptteil* ihrer Energie in benachbarten Zellen deponieren.

In der ZEUS-Standard-Rekonstruktion der HES-Daten werden als Gruppen einfach die  $3 \times 3$  Diodensignale um lokale Maxima herum zusammengefaßt. Dieser Algorithmus hat zwar ein gutes Auflösungsvermögen, spaltet aber größere HES-Schauer in mehrere Gruppen.

Dies soll der *Isle*-Algorithmus vermindern. Wie der *Island*-Algorithmus, so sucht auch der *Isle*-Algorithmus nach lokalen Maxima und ordnet ihnen die Signale benachbarter Dioden zu. Benachbart sind zwei Dioden, wenn sie in dem vereinfachten Rechteckschema der HES-Geometrie mindestens einen gemeinsamen Punkt besitzen. Das Auflösungsvermögen des *Isle*-Algorithmus entspricht im Prinzip dem des *Island*-Algorithmus. Wegen der feineren Segmentierung des HES ist es allerdings weit besser. Zwei Teilchen, die ihre Hauptenergie in einer Entfernung von mehr als 8,5 cm (doppelte Diagonallänge einer Diode) deponieren, können (zumindest theoretisch) stets aufgelöst werden. Im günstigsten Falle kann man Teilchen trennen, deren Abstand 4,5 cm

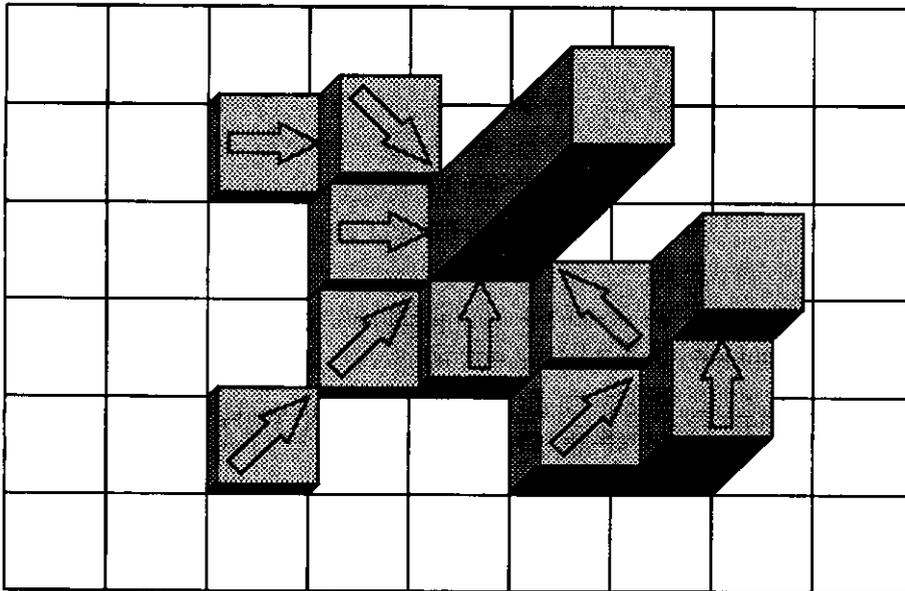


Abbildung B.1: Illustration des *Isle*-Algorithmus.

beträgt. Abb. B.2 zeigt die Wirkung der hier beschriebenen Gruppierungsalgorithmen an einem eindimensionalen Beispiel.

Technisch läuft die *Isle*-Bildung wie folgt ab (Zur Verfügung steht eine Tabelle von Diodennummern und dazugehörigen Signalen):

1. In der Tabelle werden zwei Spalten mit ganzzahligen Koordinaten in  $x$ - und  $y$ -Richtung angelegt. Die Koordinaten geben jeweils die Diodenzeile und -spalte an, in der sich die bezeichnete Diode befindet. Diese Koordinaten werden aus den Diodennummern gewonnen. Benachbart sind Dioden dann, wenn sie sich in den beiden Koordinaten um maximal 1 unterscheiden, also wenn sie sich in mindestens einem Punkt berühren.
2. Zu jeder Diode wird die benachbarte Diode mit größtem Signal gesucht. Auf diese, d. h. auf ihren Tabellenplatz, wird in einer weiteren Spalte ein Zeiger eingerichtet. Haben alle benachbarten Dioden ein kleineres Signal als die Diode selbst, so soll ihr Zeiger auf sie selbst zeigen. Solche Dioden stellen lokale Maxima dar.
3. Die lokalen Maxima werden durchnummeriert. Die Nummern sind bereits die *Isle*-Identifikationen und werden in einer weiteren Tabellenspalte vermerkt. Gleichzeitig wird eine *Isle*-Tabelle mit Zeigern auf die Dioden maximalen Signals eingerichtet.
4. Noch einmal werden alle Dioden durchgegangen. Zu jeder Diode wird den Zeigern auf den Nachbarn mit höchstem Signal folgend die *Isle*-Nummer ihres Maximums gefunden.
5. Nebenbei werden in einer weiteren Spalte die zu einer gemeinsamen *Isle* gehörigen Dioden durch Zeiger miteinander verkettet. Dies soll einen späteren Zugriff auf alle Dioden *einer* *Isle* beschleunigen.

Der *Island*-Algorithmus folgt demselben Prinzip, jedoch mit dreidimensionalen Nachbarschaftsbeziehungen.

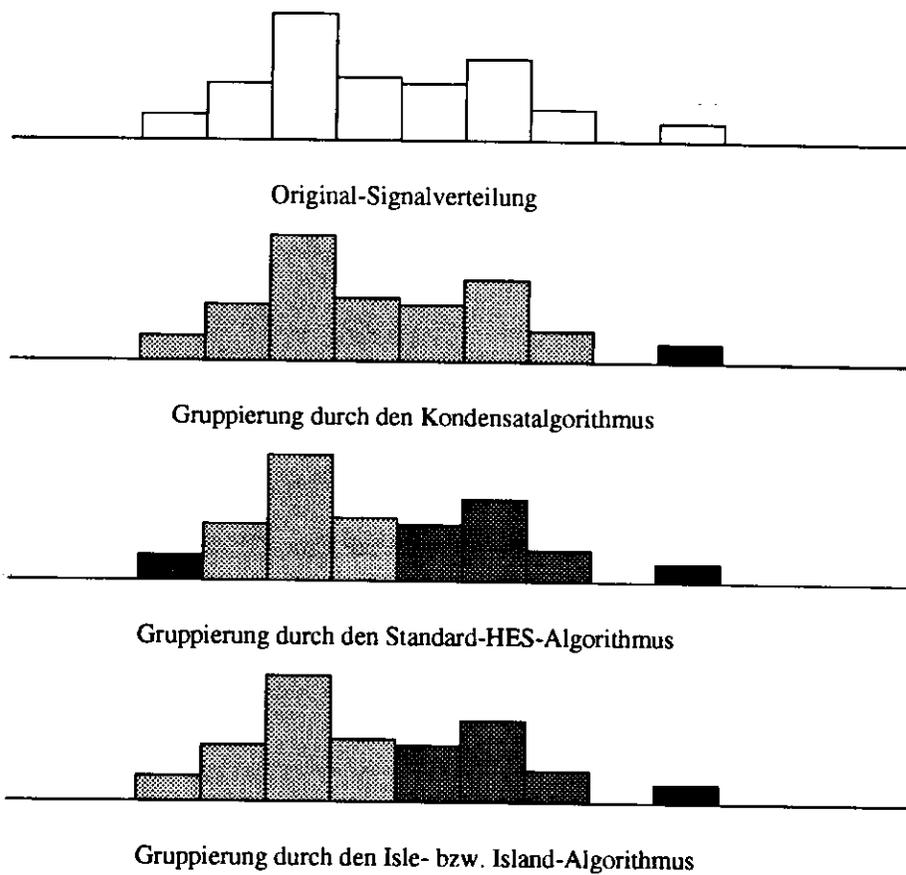


Abbildung B.2: Eindimensionale Darstellung zum Vergleich der Eigenschaften unterschiedlicher Gruppierungsalgorithmen.

## B.2 Signifikanz unterschiedlicher Diskriminanzen zweier Elektronenfinder

Dieser Abschnitt stellt dar, wie in Kapitel 4 entschieden wurde, ob der Unterschied der Diskriminanzen zweier Elektronenfinder signifikant ist, d. h., ob die Hypothese der Gleichheit der Diskriminanzen bei vorgegebener Irrtumswahrscheinlichkeit  $\alpha$  verworfen werden muß. Dabei liegt der spezielle Fall vor, daß die verschiedenen Elektronenfinder teilweise dieselben Eingabevariablen besitzen. Ihre Entscheidungen sind daher stark korreliert: Hadronen, die die eine Methode richtig identifiziert, werden meist auch von der anderen erkannt. Hadronen mit ungewöhnlicher Signatur (z. B. großem *Isle*-Signal) werden meist von beiden nicht erkannt. In einer relativ kleinen Anzahl von Fällen entscheiden die Elektronenfinder jedoch unterschiedlich, was zu einer unterschiedlichen Diskriminanz führt.

Üblicherweise wird der Student'sche t-Test angewendet, der jedoch unabhängige Verfahren voraussetzt. Diese Methode schätzt in unserem Fall die Signifikanz der Unterschiede falsch ab. So wurden beispielsweise für die Diskriminanzen des einfachen HES-Schnittes und des Elektronenfinders *hesnn* die folgenden Werte ermittelt (unkorrigiert und ohne Zuordnung):

$$D_{HES} = (76,2 \pm 1,7)\% = 74,5\% \dots 77,9\%$$

$$D_{hesnn} = (79,0 \pm 1,6)\% = 77,4\% \dots 80,6\%$$

Ihre  $1\sigma$ -Vertrauensintervalle überschneiden sich. Trotzdem kann die Hypothese der Gleichheit beider Diskriminanzen unter Berücksichtigung der Korrelationen mit einem Signifikanzniveau von  $\alpha < 5\%$  verworfen werden, wie im folgenden gezeigt wird.

Zur Prüfung der Nullhypothese, daß die Diskriminanzen der Elektronenfinder HES-Schnitt und *hesnn* gleich sind, soll ausgegangen werden von der Zahl  $k_1$  der Hadronen, die nur von *hesnn* erkannt wurden, von der Zahl  $k_2$  der Hadronen, die von beiden Elektronenfindern identifiziert wurden, von der Zahl  $k_3$  der Hadronen, die beide Elektronenfinder für Elektronen hielten, und von der Zahl  $k_4$  der Hadronen, die nur der HES-Schnitt richtig einstufte (vgl. Tab. B.1).

		HES sagt	
		e	h
<i>hesnn</i> sagt	h	$k_1=28$	$k_2=502$
	e	$k_3=132$	$k_4=9$

Tabelle B.1: Identifikation von Hadronen durch den einfachen HES-Schnitt und den Elektronenfinder *hesnn*.

Würde man dieses Experiment beliebig oft mit unabhängigen Datensätzen gleicher Größe wiederholen, so ergäbe sich für jedes einzelne Feld  $k_i$  eine Binomialverteilung mit der Gesamtzahl  $n = k_1 + k_2 + k_3 + k_4$  und den gesuchten Wahrscheinlichkeiten  $p_1, p_2, p_3$  und  $p_4$  (Indizierung wie in der Tabelle). Die vier Wahrscheinlichkeiten ergänzen sich zu 1. Nun sind die Verteilungen der  $k_i$  nicht unabhängig voneinander. Hat man beispielsweise  $k_1$  gewählt, so ist der Bereich möglicher Werte der übrigen  $k_i$  bereits eingeschränkt. Stehen  $k_1, k_2$  und  $k_3$  fest, so ist  $k_4$  wegen der Bedingung  $k_1 + k_2 + k_3 + k_4 = n$  sogar eindeutig bestimmt. Diese (Tetranomial-) Verteilung

mit den Parametern  $p_i$  ( $p_1 + p_2 + p_3 + p_4 = 1$ ) läßt sich beschreiben durch die Wahrscheinlichkeiten

$$P(k_1, k_2, k_3, k_4) = \begin{cases} \frac{n!}{k_1! k_2! k_3! k_4!} p_1^{k_1} p_2^{k_2} p_3^{k_3} p_4^{k_4} & \text{für } k_1 + k_2 + k_3 + k_4 = n \\ 0 & \text{sonst.} \end{cases} \quad (\text{B.1})$$

Strenggenommen ist jetzt die Aufgabe, die Kombination von Parametern ( $p_1, p_2, p_3, p_4$ ) mit

$$\begin{aligned} D_{hesnn} &= p_1 + p_2 = p_2 + p_4 = D_{HES}, & \text{also} \\ p_1 &= p_4, \end{aligned}$$

zu finden, für die die Wahrscheinlichkeit,  $k_1 \geq 28$  und  $k_4 \leq 9$  zu messen, maximal ist. Diese maximale Wahrscheinlichkeit gibt dann die Signifikanz an, mit der die Nullhypothese  $p_1 = p_4$  verworfen werden kann. Dies erscheint aber recht mühsam.

Es soll darum nur eine Mindestabschätzung für die gesuchte Signifikanz gegeben werden. Weiter soll angenommen werden, daß  $k_1$  und  $k_4$  unkorreliert poissonverteilt sind mit den (wahren) Mittelwerten  $\lambda_1$  und  $\lambda_2$ , was eine gute Näherung ist, falls  $p_1 + p_4 \ll 1$ .

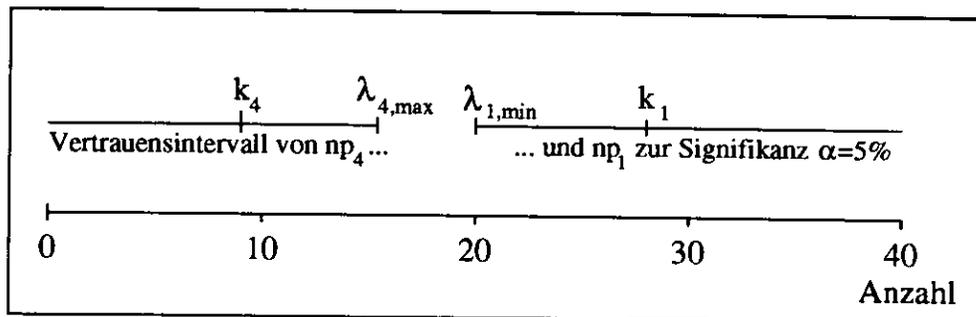


Abbildung B.3: Einseitige Konfidenzintervalle für zur Begrenzung von  $np_1$  und  $np_4$  für die Elektronenfinder HES-Schnitt und  $hesnn$ .

In diesem Fall kann man bei vorgegebenem Signifikanzniveau  $\alpha$  einen kleinsten Poissonparameter  $\lambda_{1,min}$  finden, der mit einer Wahrscheinlichkeit  $\alpha$  zu einer Messung von  $k_1 \geq 28$  führt (vgl. Abb. B.3).  $\alpha$  ist dann unter der Annahme  $np_1 \leq \lambda_{1,min}$  eine obere Schranke für die Wahrscheinlichkeit, die aktuelle Messung gemessen zu haben.

Ebenso läßt sich ein Parameter  $\lambda_{4,max}$  finden, der mit der Wahrscheinlichkeit  $\alpha$  zu einer Messung  $k_4 \leq 9$  führt.

Falls  $\lambda_{4,max} < \lambda_{1,min}$ , so ist die Nullhypothese  $p_1 = p_4$  mit der Signifikanz  $\alpha^* < \alpha$  zu verwerfen. Die Nullhypothese würde nämlich erfordern, daß entweder  $np_1 < \lambda_{1,min}$  oder  $np_4 > \lambda_{4,max}$  (oder beides), was zu einer Wahrscheinlichkeit kleiner als  $\alpha$  (im Falle beider Grenzverletzungen  $\alpha^2$ ) für das Eintreten der aktuellen Messung führen würde.

Die konkrete Berechnung von  $\lambda_{1,min}$  und  $\lambda_{4,max}$  ist in [2] sowie in [3] beschrieben. In dem vorliegenden Beispiel wurden zu einer Signifikanz von  $\alpha = 5\%$  die einseitigen Grenzwerte

$$\begin{aligned} \lambda_{1,min} &= 19,9 & \text{und} \\ \lambda_{4,max} &= 15,7 \end{aligned}$$

berechnet. Die Nullhypothese  $p_1 = p_4$  muß somit verworfen werden. Die Diskriminanzen der beiden Elektronenfinder unterscheiden sich signifikant.

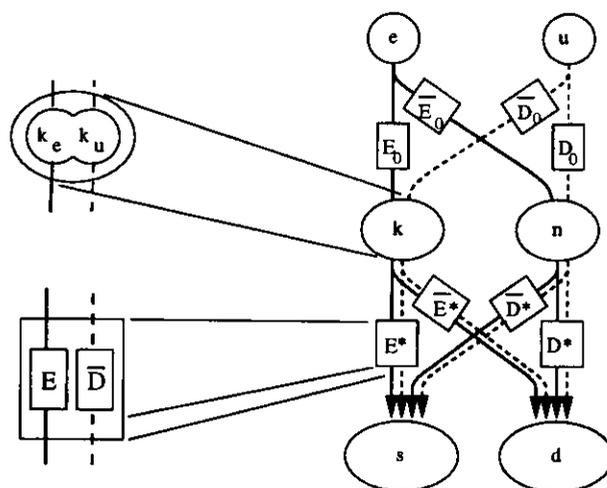


Abbildung B.4: Flußdiagramm eines zweistufigen Selektionsalgorithmus. (Erklärungen im folgenden Abschnitt.)

### B.3 Entfaltung von Effizienz und Diskriminanz

Der Entfaltung zweier nacheinander angewandter Selektionsalgorithmen, von denen die Effizienz und Diskriminanzkraft des zweiten gesucht sind, wird nach dem in Abb. B.4 vorgestellten Modell durchgeführt.

Ein Datensatz enthalte  $e$  Elektronen und  $u$  Untergrundteilchen (z. B. Hadronen). Ein erster Selektionsalgorithmus (in dieser Arbeit der Konversionsfinder) habe die Effizienz  $E_0$  und Diskriminanz  $D_0$ , d. h., er bezeichnet  $k_e = E_0 \cdot e$  Elektronen und  $k_u = \bar{D}_0 \cdot u = (1 - D_0)u$  Untergrundteilchen als (Konversions-)Elektronen. Weiter werden  $n_e = \bar{E}_0 \cdot e$  Elektronen und  $n_u = D_0 \cdot u$  Untergrundteilchen als Nicht-(Konversions-)Elektronen bezeichnet. Nach Anwendung des ersten Selektionsalgorithmus ist der Datensatz aufgeteilt in  $k = k_e + k_u$  als (Konversions-)Elektronen und in  $n = n_e + n_u$  als Nichtelektronen bezeichnete Teilchen.

Auf diesem Datensatz wird ein zweiter Selektionsalgorithmus (Elektronenfinder) erprobt. Dieser selektiert  $s_k$  der  $k$  als Elektronen bezeichneten Teilchen und  $s_n$  der  $n$  als Nichtelektronen bezeichneten Teilchen als Elektronen. Er diskriminiert  $d_k$  der  $k$  als Elektronen bezeichnete Teilchen und  $d_n$  der  $n$  Teilchen, die als Nichtelektronen bezeichnet worden waren. Daraus berechnet man die Effizienz  $E^* = s_k/k$  und die Diskriminanz  $D^* = d_n/n$ , bezogen auf den in Konversionselektronen und Nichtkonversionselektronen unterteilten Datensatz. Gesucht ist aber eigentlich die Effizienz  $E$ , bezogen auf die  $e$  wahren Elektronen (durchgezogene Linien) und die Diskriminanz  $D$ , bezogen auf die  $u$  wahren Untergrundteilchen (gestrichelte Linien).

Um diese Größen aus den bisher gegebenen zu berechnen, betrachten wir noch einmal genauer die zweite Selektion: Die Menge der  $k$  „Konversionselektronen“ setzt sich zusammen aus  $k_e$  Elektronen und  $k_u$  Untergrundteilchen. Von den  $k_e$  Elektronen hat der Elektronenfinder  $s_{k_e}$  gefunden. Das entspricht einer Effizienz von  $E_1 = s_{k_e}/k_e$ . Daneben hat er  $s_{k_u}$  Untergrundteilchen selektiert, die aber vom Konversionsfinder auch schon als Elektronen bezeichnet worden waren. Daraus erhalte man eine Misidentifikationswahrscheinlichkeit von  $\bar{D}_1 = s_{k_u}/k_u$ , vorausgesetzt, man würde die Größen  $k_u$  und vor allem  $s_{k_u}$  kennen. Sie sind aber nicht bekannt. Statt dessen hat man die von  $E_1$  und  $\bar{D}_1$  bestimmte Größe  $E^*$ .

Ebenso ist  $\bar{D}^*$  aus einer Effizienz  $E_2$  und Misidentifikationswahrscheinlichkeit  $\bar{D}_2$  zusammengesetzt, die sich auf die  $n_e$  Elektronen und  $n_u$  Untergrundteilchen beziehen, die zuvor von dem Konversionsfinder als Nichtelektronen bezeichnet worden waren.

Unter der Annahme, daß die beiden Selektionsstufen (Konversionsfinder und Elektronenfinder) unabhängig arbeiten, d. h. daß der Elektronenfinder gegenüber den Elektronen und Untergrundteilchen, die zuvor als Konversionselektronen bezeichnet worden sind, die gleichen Eigenschaften zeigt wie gegenüber den Elektronen und Untergrundteilchen, die zuvor als Nichtelektronen bezeichnet worden sind, ist  $E_1 = E_2 = E$  und  $D_1 = D_2 = D$ . In diesem Fall kann man  $E$  und  $D$  herausfalten:

Zu Beginn liegen eine unbekannt Anzahl  $e$  Elektronen und  $u$  Untergrundteilchen vor. Bekannt seien nur die Zahl der Konversionsspuren  $k$  und Nichtkonversionsspuren  $n$ , sowie die Effizienz  $E_0$  und Diskriminanzkraft  $D_0$  des Konversionsfinders und die gemessene scheinbare Effizienz  $E^*$  und Diskriminanz  $D^*$  des Elektronenfinders. Es wird weiter vorausgesetzt, daß die Selektionsstufen unabhängig voneinander arbeiten, d. h. daß die Effizienz des Elektronenfinders für Konversionselektronen dieselbe ist wie für nicht vom Konversionsfinder gefundene Elektronen und entsprechend für die Untergrundteilchen.

Die Zahl der Konversionsspuren und der Nichtkonversionsspuren setzt sich zusammen aus (vgl. Abb. B.4)

$$k = k_e + k_u = E_0 e + \bar{D}_0 u \quad (\text{B.2})$$

$$n = n_e + n_u = \bar{E}_0 e + D_0 u \quad (\text{B.3})$$

Löst man diese beiden Gleichungen nach  $e$  und  $u$  auf, so erhält man die Teilchenzahlen

$$e = \frac{kD_0 - n\bar{D}_0}{\alpha} \quad (\text{B.4})$$

$$u = \frac{nE_0 - k\bar{E}_0}{\alpha} \quad (\text{B.5})$$

$$\alpha = E_0 D_0 - \bar{E}_0 \bar{D}_0 \quad (\text{B.6})$$

Damit können wir  $e$  und  $u$ , sowie  $k_e$ ,  $k_u$ ,  $n_e$  und  $n_u$  durch  $k$ ,  $n$ ,  $E_0$  und  $D_0$  ausdrücken. In der zweiten Stufe ist

$$E^* k = E k_e + \bar{D} k_u \quad (\text{B.7})$$

$$\bar{D}^* n = E n_e + \bar{D} n_u \quad (\text{B.8})$$

Eliminiert man aus diesen Gleichungen  $D$ , so erhält man eine Gleichung für  $E$ . Kürzt man aus dieser  $u$  heraus und setzt für  $e$  ein, so ergibt sich schließlich

$$E = \frac{E^* D_0 r - \bar{D}^* \bar{D}_0 \bar{r}}{D_0 r - \bar{D}_0 \bar{r}} \quad (\text{B.9})$$

$$\text{mit } r := \frac{k}{k+n} \quad (\text{B.10})$$

Aus Symmetriegründen gilt entsprechend

$$D = \frac{D^* E_0 \bar{r} - \bar{E}^* \bar{E}_0 r}{E_0 \bar{r} - \bar{E}_0 r} \quad (\text{B.11})$$

Sind statt der Effizienz und Diskriminanz der ersten Selektionsstufe der Untergrund des Konversionsdatensatzes und der Nichtkonversionsspuren bzw. ihre Komplemente  $R_k = k_e/k$  und  $R_n = n_u/n$  direkt gegeben, so erhält man durch Auflösen der Gleichungen B.7 und B.8 nach  $E$  und Kürzen von  $k$  und  $n$

$$E = \frac{E^* R_n - \bar{R}_k \bar{D}^*}{R_k R_n - \bar{R}_k \bar{R}_n} \quad (\text{B.12})$$

Aus Symmetriegründen hat man sofort

$$D = \frac{R_k D^* - \bar{E}^* \bar{R}_n}{R_k R_n - \bar{R}_k \bar{R}_n} \quad (\text{B.13})$$

Die Rechnung wurde hier am Beispiel des Konversionsfinders durchgeführt. Sie läßt sich aber analog auf alle zweistufigen Selektionsalgorithmen übertragen, für die die Effizienz und Diskriminanzkraft der zweiten Stufe entfaltet werden soll.

# Literaturverzeichnis

- [1] H. Abramowicz, R. Sinkus, *Electron Identification with Neural Networks at ZEUS*, ZEUS note 93-117, November 1993.
- [2] V. Blobel, E. Lohrmann, *Statistische und numerische Methoden der Datenauswertung II*, Vorlesungs WS 1993/94 an der Universität Hamburg.
- [3] K. Bosch, *Statistik-Taschenbuch*, Zweite Aufl., Oldenbourg, 1993.
- [4] W. W. Cooley, P. R. Lohnes, *Multivariate Data Analysis*, John Wiley Publ. Comp.
- [5] S. Grossberg, *Competitive Learning: from interactive activation to adaptive resonance*, Cogn. Sci. 11 (1987),1,23-64.
- [6] S. J. Hanson, *What Connectionist Models Learn: Learning and Representation in Connectionist Networks*, Brain and Behavioral Sciences, 13(3), Seite 471-511.
- [7] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. USA, 79 (1982), Seite 2554-2558.
- [8] G. Jahnen, *Studium der Erzeugung schwerer Quarks mit Hilfe eines Siliziumzählers im ZEUS- Detektor*, Dissertation, Universität Hamburg, Juni 1992.
- [9] L. Koepke, R. van Woudenberg, *Photon Conversions in the ZEUS Detector*, ZEUS note 94-016, 1994.
- [10] T. Kohonen, *Self-organization and Associative Memory*, Springer-Verlag, Berlin, 2. Aufl., 1988.
- [11] L. Larsson, *Simulation neuronaler Netzwerke und Test des z-Kammer-Triggers am H1-Detektor*, Diplomarbeit.
- [12] L. Lönnblad, C. Peterson, T. Rönngvaldsson, *Pattern recognition in high energy physics with artificial neural networks - JETNET 2.0*, Comp. Phys. Comm. 70 (1992) 167-182.
- [13] L. Lönnblad, C. Peterson, T. Rönngvaldsson, *Using neural networks to identify jets*, Nucl. Phys. B 349 (1991) 675.

# Danksagung

Diese Diplomarbeit zu schreiben hat Spaß gebracht. Das habe ich vor allem Prof. Dr. E. Lohrmann zu verdanken, der mich als Diplomanden in die Kollaboration aufgenommen hat. Ihm verdanke ich auch die thematische Freiheit, die es mir ermöglicht hat, mich intensiv mit neuronalen Netzen auseinanderzusetzen.

Meinem Betreuer Dr. L. H. Koepke bin ich sehr dankbar, daß er *immer* Zeit für mich hatte, wenn ich seine Hilfe brauchte. Dank sei ihm auch ausgesprochen für seinen guten Rat und die Korrekturen an dieser Arbeit.

Ein herzliches *thanks* auch an M. Kuze und I. Fleck für alle Auskünfte über den Hadron-Elektron-Separator, an R. van Woudenberg für die MC-Untersuchungen und an N. Pavel für die netten Diskussionen über Fragen der Statistik. Vielen Dank auch den ZEUS-Diplomanden und Doktoranden H. Beier, O. Deppe, K. Deßler, J. Große-Knetter, M. Löwe, B. Micheel, S. Nickel, T. Plehn, B. Ruhm, G. Schmidt, T. Stange, B. Surrow und D. Westphal für die lustige Zeit, die Diskussionen über Physik und Unphysik, und den gemeinsamen Kampf gegen die kleinen Widrigkeiten des Physikeralltags.

Zuletzt — damit es besser in Erinnerung bleibt — danke ich meiner lieben Frau Eva und meinen Eltern für die enorme Rückenstärkung, die sie mir gewährt haben und ohne die diese Arbeit nicht möglich gewesen wäre.

# Erklärung

Hiermit versichere ich, die vorliegende Arbeit unter Angabe aller verwendeten Quellen und Hilfsmittel selbständig angefertigt zu haben.

Hamburg, im Oktober 1994

J. O. Carstens  
(Jan Ole Carstens)