

Olsson F11

**MAX-PLANCK-INSTITUT FÜR PHYSIK**  
WERNER-HEISENBERG-INSTITUT

45x

~~MPI-PhE/93-07~~

~~März 1993~~

Neuronale Netzwerke als L2-Trigger  
im H1-Experiment

Diplomarbeit an der Fakultät für Physik der  
Ludwig-Maximilians-Universität München

vorgelegt von

ANDREAS GRUBER

angefertigt am

~~Max-Planck-Institut für Physik~~  
~~- Werner-Heisenberg-Institut -~~

Dezember 1992

~~8000 München 40 · Föhringer Ring 6~~

Alle Rechte vorbehalten

Max-Planck-Institut für Physik, München.

# Neuronale Netzwerke als L2-Trigger im H1-Experiment

Diplomarbeit an der Fakultät für Physik der  
Ludwig-Maximilians-Universität München

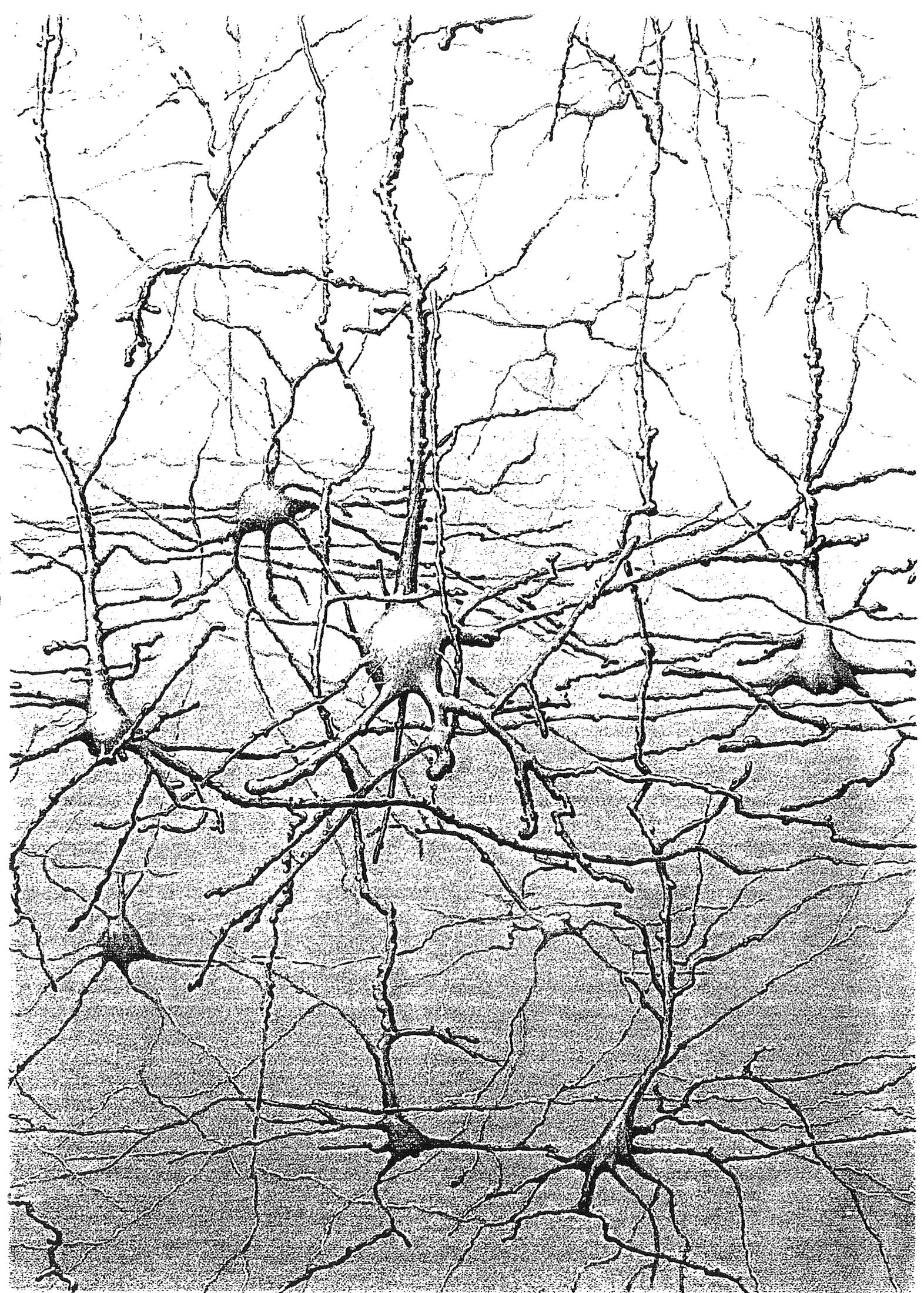
vorgelegt von

ANDREAS GRUBER ✓

angefertigt am

Max-Planck-Institut für Physik  
– Werner-Heisenberg-Institut –

Dezember 1992



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Physik am Speicherring HERA</b>	<b>3</b>
2.1	Der HERA-Speicherring . . . . .	3
2.2	Elektron-Proton-Streuung . . . . .	5
2.2.1	Neutrale Ströme . . . . .	6
2.2.2	Geladene Ströme . . . . .	8
2.2.3	Boson-Gluon-Fusion . . . . .	8
2.2.4	Untergrundereignisse . . . . .	10
2.2.4.1	Strahl-Gas-Reaktionen (Beam-Gas) . . . . .	10
2.2.4.2	Strahl-Wand-Reaktionen (Beam-Wall) . . . . .	11
2.2.4.3	Synchrotronstrahlung . . . . .	11
2.2.4.4	Kosmische Strahlung . . . . .	11
<b>3</b>	<b>Das Triggersystem des H1-Detektors</b>	<b>13</b>
3.1	Das mehrstufige H1-Triggerkonzept . . . . .	13
3.2	Triggerkomponenten des H1-Detektors . . . . .	16
3.2.1	Das Flüssig-Argon-Kalorimeter . . . . .	17
3.2.2	Das PLUG-Kalorimeter . . . . .	18
3.2.3	Das rückwärtige elektromagnetische Kalorimeter . . . . .	18
3.2.4	Die zentralen Jetkammern . . . . .	18
3.2.5	Die zentralen Proportionalkammern . . . . .	19
3.2.6	Die Z-Kammern . . . . .	20
3.2.7	Die rückwärtigen Proportionalkammern . . . . .	20
3.2.8	Die Spurkammern im Vorwärtsbereich . . . . .	21
3.2.9	Die Myon-Kammern . . . . .	21
3.2.10	Das Luminositäts-Meßsystem . . . . .	22
3.2.11	Das Flugzeit-Meßsystem (TOF) . . . . .	23
3.2.12	Die zentrale L1-Triggerlogik . . . . .	24
<b>4</b>	<b>Neuronale Netzwerke</b>	<b>25</b>
4.1	Motivation . . . . .	25
4.2	Grundlagen . . . . .	25
4.2.1	Von der Biologie zur Mathematik . . . . .	25
4.2.2	Das Multilayer-Feedforward-Netzwerk . . . . .	28

4.2.3	Lern- und Ausführungsphase . . . . .	30
4.2.4	Der Backpropagation-Lernalgorithmus . . . . .	30
4.3	Geometrische Interpretation der Netzwerkleistung . . . . .	32
4.4	Simulationsbeispiele in zwei Dimensionen . . . . .	33
4.5	Theoretischer Ausblick . . . . .	37
4.5.1	Der Bayes'sche Klassifizierer . . . . .	38
4.5.2	Realisierbarkeit von Abbildungen . . . . .	39
4.5.3	Lernbarkeit . . . . .	39
<b>5</b>	<b>Simulation Neuronaler Netze mit MC-Daten</b>	<b>40</b>
5.1	Monte-Carlo-Daten . . . . .	40
5.2	Der Standard-L1-Trigger . . . . .	41
5.3	Netzwerksimulationen . . . . .	45
5.3.1	Das JETNET-Programm . . . . .	45
5.3.2	Netzwerktopologien und Training . . . . .	45
5.4	Veranschaulichung von L2-Schnitten . . . . .	48
5.5	Ergebnisse des Neuronalen L2-Triggers . . . . .	49
5.6	Fehlertoleranz des Neuronalen Netzes . . . . .	55
<b>6</b>	<b>Einsatz Neuronaler Netze mit realen Daten</b>	<b>59</b>
6.1	Beschreibung der Daten und Triggergrößen . . . . .	59
6.1.1	Das H1-Datennahmesystem . . . . .	59
6.1.2	Untergrund-Daten . . . . .	61
6.1.3	Physik-Daten . . . . .	64
6.1.4	Triggergrößen . . . . .	64
6.2	Netzwerktraining und Resultate . . . . .	64
6.2.1	Netztopologie und Training . . . . .	64
6.2.2	Ergebnisse . . . . .	66
6.2.3	Eindimensionale Verteilungen . . . . .	69
6.2.4	Deutung der Netzwerkschnitte . . . . .	70
6.3	Vergleich mit Monte-Carlo-Daten . . . . .	75
6.4	Stabilitätsuntersuchungen . . . . .	76
6.4.1	Test der Untergrundstabilität . . . . .	76
6.4.2	Stabilität bei Subdetektorausfall . . . . .	78
<b>7</b>	<b>Hardware-Realisierung</b>	<b>81</b>
7.1	Konzept zur Hardware-Realisierung . . . . .	81
7.2	Chips für Neuronale Netzwerke . . . . .	81
7.3	Das Mustererkennungsmodul (PRM) . . . . .	83
7.4	Das Datenverteilungsmodul (DDM) . . . . .	85
	<b>Zusammenfassung</b>	<b>86</b>

# Kapitel 1

## Einleitung

Die Frage nach den elementaren Bausteinen der Materie und den Kräften, die sie zusammenhalten, beschäftigt den Menschen seit der Antike.

Der lange Weg von den ersten atomistischen Theorien Demokrits (460–370 v. Chr.), Platons (429–348 v. Chr.) und Aristoteles' (384–322 v. Chr.) führte über die Entwicklung des Periodensystems der Elemente (Meyer und Mendelejeff (1869)) schließlich zur Entdeckung des Atomkerns durch Streuexperimente mit  $\alpha$ -Teilchen an einer Goldfolie durch Rutherford (1911).

Hofstadter et al. [1] entdeckten durch Streuexperimente mit Elektronen, daß das Proton eine wohldefinierte räumliche Ausdehnung besitzt. Dies bildete – zusammen mit Beobachtungen bei Pion–Nukleon Streuexperimenten – die Grundlage für das Parton Modell der Nukleonen [2, 3]. In diesem werden Protonen und Neutronen als zusammengesetzte Objekte beschrieben, die aus Quarks bestehen und durch Gluonen zusammengehalten werden.

Der zur Beschreibung des Streuprozesses notwendige Wirkungsquerschnitt kann mit Hilfe von sogenannten Strukturfunktionen ausgedrückt werden, die vom übertragenen Impuls  $Q^2$  und der übertragenen Energie  $\nu$  abhängen.

Die "Skalenhypothese" [4] von Bjorken (1969) reduzierte die Anzahl der Variablen, die in der Strukturfunktion auftauchen, im Grenzfall hoher Energie- und Impulsüberträge auf nur eine Größe  $x \equiv \frac{Q^2}{2M\nu}$ , die als Impulsanteil des Partons am Gesamtimpuls interpretiert werden kann.

Die heutigen Experimente erlauben eine genauere Messung der Strukturfunktionen, so daß Abweichungen vom "Skalenverhalten" erfaßt werden können. Mit Hilfe der Theorie der Quantenchromodynamik kann innerhalb des Standardmodells diese Abweichung berechnet und mit dem Experiment verglichen werden. Somit werden weitere Tests der Theorie der starken Wechselwirkung und Messungen der starken Kopplungskonstante ermöglicht.

Der HERA–Beschleuniger am DESY in Hamburg bietet die Möglichkeit, Wechselwirkungen zwischen Elektronen mit 30 GeV Energie und Protonen mit 820 GeV Energie zu untersuchen [5]. Die Schwerpunktsenergie von 314 GeV und die möglichen Impulsüberträge von bis zu 96000 GeV<sup>2</sup> stellen einen stark erweiterten Meßbereich in der  $x$ - $Q^2$ -Ebene gegenüber den bisherigen Experimenten dar [6].

Außer der Gültigkeit des Standardmodells soll bei HERA aber auch die Existenz exo-

tischer neuer Physik untersucht werden, wie z.B. supersymmetrische Teilchen oder Leptoquarks. Vielleicht eröffnen diese Streuexperimente sogar einen Einblick in eine weitere Substruktur der Quarks oder Leptonen angeregte Quarks und Leptonen).

Zur Durchführung dieser Untersuchungen bei HERA wurden die Detektoren H1 [7] und ZEUS [8] gebaut. Beide Experimente sehen sich hohen experimentellen Anforderungen gegenüber. Eine der Hauptaufgaben ist es, die geplante extrem hohe Luminosität von  $1.5 \cdot 10^{31} \text{cm}^{-2}\text{s}^{-1}$  zu bewältigen, die notwendig ist, um Reaktionen mit niedrigem Wirkungsquerschnitt in ausreichender Anzahl zu beobachten. In den Wechselwirkungszone von HERA bedeutet dies, daß sich alle 96 nsec ein Elektronen- und ein Protonenbündel durchdringen.

Beim Umlauf der Teilchen kommt es aber auch zu unerwünschten Wechselwirkungen mit Strahlführungskomponenten oder Restgasatomen im Vakuum des Strahlrohrs, die Teilchenschauer produzieren und im Detektor Physikreaktionen vortäuschen können. Die Rate für diese Untergrundereignisse liegt bei etwa 0.5 MHz und ist damit um einen Faktor  $O(10^5)$  höher als die der erwarteten Physik. Dies stellt sehr hohe Ansprüche an ein Triggersystem, das die Aufgabe besitzt, die Untergrundereignisse zu verwerfen, ohne dabei Physikereignisse zu verlieren.

Unter der Randbedingung einer Datennahmerate von lediglich 5 Hz, bedingt durch das große Datenvolumen der einzelnen Ereignisse, muß eine Reduktion um 5 Größenordnungen erreicht werden. Das H1-Experiment plant, dies durch ein mehrstufiges Triggersystem aus zwei Hardwarestufen (L1,L2) und zwei Softwarestufen (L3,L4) [9] zu erreichen. Wenn ein Ereignis von einer Stufe akzeptiert wird, beginnt die nächsthöherer Stufe mit der ihr zur Verfügung stehenden Information eine Entscheidung zu fällen. Jede Stufe übernimmt somit einen Teil der Ratenreduktion.

Die erste Stufe (L1) ist ein gewöhnlicher Schwellentrigger, der Signale aus verschiedenen Subdetektoren diskriminiert und logisch miteinander verknüpft. Fehlt die zweite Stufe, so müßte L1 den Reduktionsfaktor von L2 mit übernehmen, was nur durch Anheben der Schwellen möglich ist und prinzipiell zu einem Verlust von Physikereignissen führt. Eine mögliche Realisierung eines L2-Triggers kann z.B. durch bessere Ausnutzung der bereits L1 zur Verfügung stehenden Information eine effizientere Untergrundunterdrückung erreichen.

Im Rahmen dieser Arbeit wird der Einsatz von Neuronalen Netzen zur verbesserten Informationsausnutzung durch Mustererkennung und damit Auffinden von Korrelationen zwischen den L1-Triggergrößen untersucht. Außerdem wird die mögliche Hardware-Implementierung massiv paralleler Rechnerarchitekturen diskutiert.

Es zeigt sich, daß mit Hilfe Neuronaler Netze die angestrebte Untergrundreduktion um einen Faktor 100 auf der zweiten Triggerstufe erreicht werden kann und gleichzeitig eine hohe Effizienz bei der Erkennung von Physikereignissen erzielt wird.

# Kapitel 2

## Physik am Speicherring HERA

### 2.1 Der HERA-Speicherring

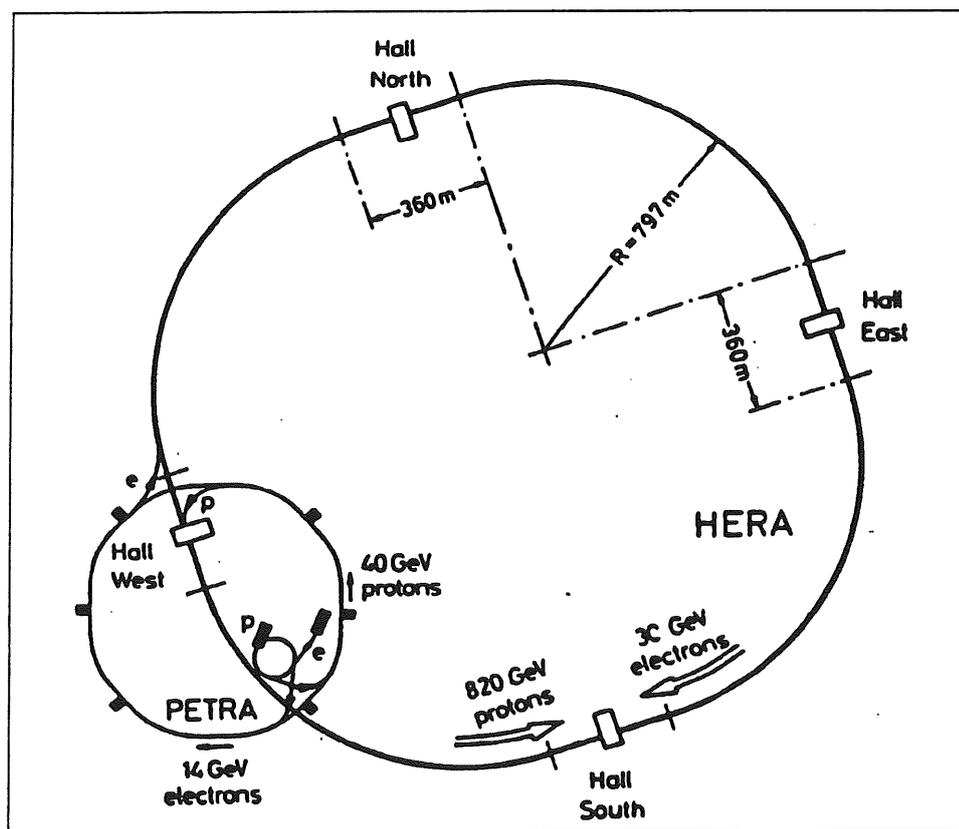


Abbildung 2.1: Schematische Darstellung des HERA-Beschleunigers

Beim Deutschen Elektronen-Synchrotron DESY in Hamburg bietet das Speicherringssystem HERA (Hadron- Elektron-Ring-Anlage) zum erstenmal die Möglichkeit, tiefinelastische Elektron-Proton-Stöße bei Schwerpunktsenergien von  $\sqrt{s} = 314 \text{ GeV}$  zu beobachten [10]. HERA besteht aus einem Ring, in dem Elektronen auf 30 GeV beschleunigt werden und einem zweiten Ring, der Protonen auf 820 GeV beschleunigt (siehe Abb. 2.1).

In der Endausbaustufe laufen Elektronen und Protonen in 210 Bündeln zu je  $10^{11}$  Teilchen in entgegengesetzter Richtung durch die 6.3 km langen Ringe und können in drei Wechselwirkungszone zur Kollision gebracht werden.

Um die Teilchen auf ihre Endenergie zu bringen, sind mehrere Vorbeschleunigerstufen erforderlich (siehe Abb. 2.2).

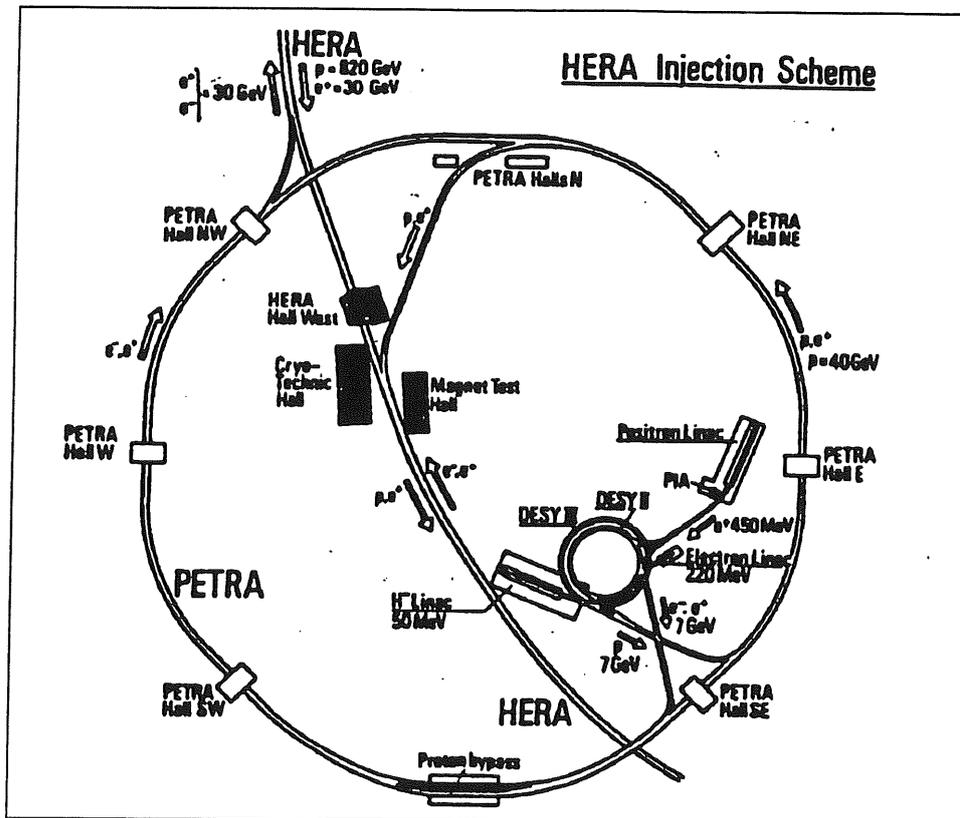


Abbildung 2.2: Schematische Darstellung der Vorbeschleuniger

- Elektronen

Die Elektronen werden in einem Linearbeschleuniger (LINAC) auf 200 MeV Energie gebracht, dann in das Synchrotron DESY II injiziert und auf 7 GeV beschleunigt. In einem weiteren Schritt wird der ehemalige  $e^+e^-$ -Speicherring PETRA zur Erreichung der 14 GeV Einschubenergie für den HERA Elektronenring verwendet. HERA benötigt normal- und supraleitende Hohlraumresonatoren mit einer Frequenz von 500 MHz, um die Elektronen auf ihre Endenergie zu bringen. Um die Elektronen auf ihrer Bahn zu halten, werden normaleitende Ablenkmagnete mit 0.165 Tesla Feldstärke verwendet. Die Energie der Elektronen ist hauptsächlich durch Synchrotronstrahlung limitiert die von den Resonatoren nachgeliefert werden muß. Durch zusätzliche supraleitende Resonatoren wird es in der Zukunft möglich sein, mehr als 35 GeV Elektronenenergie zu erreichen.

- Protonen

Die Protonen werden als  $H^-$ -Ionen ebenfalls zuerst in einem Linearbeschleuniger

(H<sup>-</sup>-LINAC) auf 50 MeV Energie gebracht und dann im Synchrotron DESY III auf 7.5 GeV beschleunigt. Während des Einschubvorgangs werden die überschüssigen Elektronen an einer 200  $\mu\text{m}$  dicken sogenannten Stripperfolie abgestreift. Im PETRA-Synchrotron wird die für HERA nötige Einschubenergie von 40 GeV erreicht. HERA verwendet Hohlraumresonatoren mit 52 MHz zur Beschleunigung, die auf 208 MHz hochgefahren werden, um die Länge der Bündel zu verkürzen. Zur Ablenkung der Protonen wurden erstmals industriell gefertigte supraleitende Magnete aus Niob-Titan-Filamenten eingesetzt, die das benötigte Feld von 4.7 Tesla erzeugen.

Die Füllzeit des Elektronenrings beträgt ca. 15 Minuten, die des Protonenrings etwa 20 Minuten.

## 2.2 Elektron-Proton-Streuung

Im folgenden werden die wichtigsten Physikprozesse bei tiefinelastischen Streuprozessen zwischen Elektronen und Protonen beschrieben [11, 12].

Das Elektron  $e^-$  kann über die bekannten elektroschwachen Austauschteilchen  $\gamma, W^\pm$

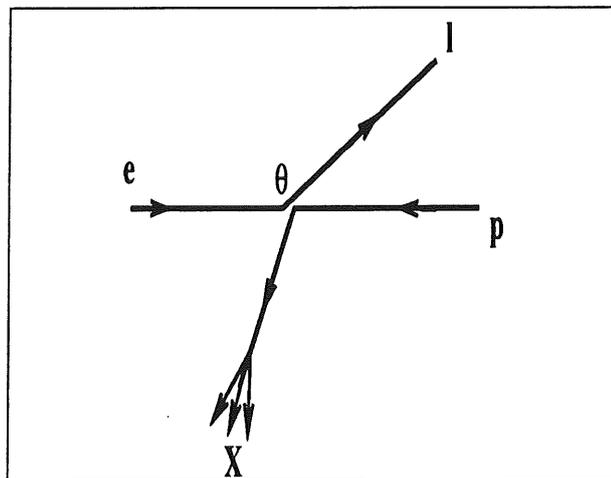


Abbildung 2.3: Elektron-Proton-Streuung

und  $Z^0$  mit dem Proton wechselwirken. Beim Austausch von Photonen ( $\gamma$ ) oder  $Z^0$ -Bosonen spricht man von neutralen Strömen (*engl. neutral current*  $\equiv$  NC), beim Austausch von  $W^+$  und  $W^-$  von geladenen Strömen (*engl. charged current*  $\equiv$  CC). Die Kinematik des Streuprozesses (siehe Abb. 2.3)

$$e + p \longrightarrow l + X \quad (2.1)$$

kann bei gegebener Schwerpunktsenergie  $\sqrt{s}$  durch nur zwei Variablen ausgedrückt werden, z.B. die Energie  $E_l$  und den Streuwinkel  $\Theta_l$  des gestreuten Leptons  $l$  ( $X$  kenn-

zeichnet das Hadronsystem des Endzustands). Die invariante Masse des Gesamtsystems ergibt sich mit den Viererimpulsen des Elektrons  $p_e$  und des Protons  $P$  zu

$$s \equiv (p_e + P)^2 \simeq 4E_e E_p. \quad (2.2)$$

Die Ausdrücke auf der rechten Seite sind in relativistischer Näherung bei Vernachlässigung der Massen gültig.  $E_e$  und  $E_p$  bezeichnen die Energien von Elektron und Proton. Die Vektorbosonen erhalten einen Viererimpuls  $q = p_e - p_l$ , der einem Impulsübertragsquadrat von

$$Q^2 \equiv -q^2 \simeq 4E_e E_l \sin^2 \frac{\Theta_l}{2} \quad (2.3)$$

und einem Energieübertrag von

$$\nu \equiv E_e - E_l = \frac{P \cdot q}{m_p} \quad (2.4)$$

entspricht ( $m_p$  bezeichnet die Masse des Protons). Die dimensionslosen Bjorken'schen Skalenvariablen  $x$  und  $y$  [4] werden als

$$x \equiv \frac{Q^2}{2 \cdot P \cdot q} = \frac{Q^2}{2 \cdot m_p \cdot \nu} \quad (2.5)$$

und

$$y \equiv \frac{E_e - E_l}{E_l} = \frac{P \cdot q}{P \cdot p_e} = \frac{2 \cdot P \cdot q}{s} \quad (2.6)$$

definiert. Dabei beschreibt  $x$  im Parton-Modell den Impulsanteil des wechselwirkenden Partons am Gesamtimpuls des Protons und  $y$  den relativen Energieübertrag des Elektrons auf das Parton. Es gilt die Beziehung

$$Q^2 = sxy. \quad (2.7)$$

Zur Beschreibung der Streuprozesse ist es üblich, die zwei unabhängigen Variablen  $x$  und  $Q^2$  zu wählen.

### 2.2.1 Neutrale Ströme

In Abbildung 2.4 ist der Feynmangraph für ein Ereignis mit Austausch eines Photons oder  $Z^0$ -Vektorbosons im Parton-Modell dargestellt. Der Endzustand ist gekennzeichnet durch ein Elektron, ein gestreutes Quark und ein Diquark-System.

Durch Hadronisierungsprozesse entstehen aus dem Quark ein *Stromjet* und aus dem Diquark-System ein *Targetjet*. Der Wirkungsquerschnitt der sich aus diesem Graphen ergibt kann im Rahmen des Standardmodells und des Quark-Parton-Modells berechnet werden, wenn dem wechselwirkenden Quark  $q_f$  ein Bruchteil  $x_q$  des Protonimpulses zugeschrieben wird. Die restlichen Partonen nehmen nicht am primären Streuprozess teil.

Zur Berechnung ist unter anderem die Kenntnis des Quarkimpulses nötig und somit die Wahrscheinlichkeitsverteilung für ein Quark  $q_f$  mit "flavour"  $f$  und Impulsanteil

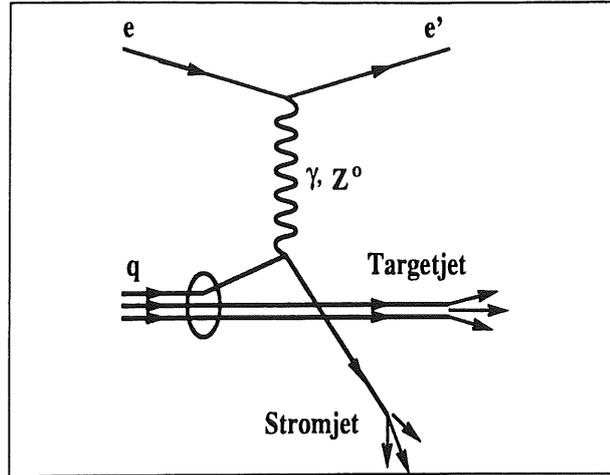


Abbildung 2.4: Feynman-Graph eines neutralen Stromereignisses

$x_q$  im Proton. Diese wird durch die Funktion  $q_f(x_q, Q^2)$  beschrieben.

Der differentielle Wirkungsquerschnitt ergibt sich dann als Funktion von  $x$  und  $Q^2$  zu

$$\frac{d^2\sigma_{NC}(e^\mp)}{dx dQ} = \frac{4\pi\alpha^2}{xQ^4} \left[ y^2 x F_1(x, Q^2) + (1-y) F_2(x, Q^2) \pm \left(y - \frac{y^2}{2}\right) x F_3(x, Q^2) \right] \quad (2.8)$$

[14]. Die Strukturfunktionen  $F_i$  beinhalten im wesentlichen die Impulsverteilungen der Quarks;  $\alpha$  bezeichnet die Feinstrukturkonstante. Im Parton-Modell sind  $F_1$  und  $F_2$  über die Callan-Gross-Relation

$$F_2 = 2x F_1 \quad (2.9)$$

miteinander verknüpft. Die Beziehung zur Impulsverteilung wird durch

$$F_2(x, Q^2) = \sum_f \left[ x q_f(x, Q^2) + x \bar{q}_f(x, Q^2) \right] A_f(Q^2) \quad (2.10)$$

$$F_3(x, Q^2) = \sum_f \left[ x q_f(x, Q^2) - x \bar{q}_f(x, Q^2) \right] B_f(Q^2) \quad (2.11)$$

hergestellt, in der über die Quarkflavours  $f$  summiert wird. Die Koeffizienten  $A_f$  und  $B_f$  sind durch

$$A_f(Q^2) = e_f^2 - 2e_f(v_e \pm a_e)v_f P_Z + (v_e \pm a_e)^2(v_f^2 + a_f^2)P_Z^2 \quad (2.12)$$

$$B_f(Q^2) = \mp 2e_f(v_e \pm a_e)a_f P_Z \pm 2(v_e \pm a_e)^2 v_f a_f P_Z^2 \quad (2.13)$$

gegeben. Dabei steht  $e_f$  für die elektrische Ladung ( $e_e = -1$ ),

$$v_f = (T_{3f} - 2e_f \sin^2 \Theta_W) / \sin 2\Theta_W \quad \text{und} \quad a_f = T_{3f} / \sin 2\Theta_W \quad (2.14)$$

stehen für die NC Vektor- und Axialvektorkopplungen, die mithilfe der dritten Komponente des schwachen Isospins ( $T_{3e} = -\frac{1}{2}$ ) und des Weinbergwinkels  $\Theta_W$  ausgedrückt wurden.  $P_Z$  beschreibt das Verhältnis von  $Z^0$ - und Photon-Propagator.

$$P_Z = \frac{Q^2}{Q^2 + M_Z^2} \quad (2.15)$$

Im Falle kleiner Impulsübertrage dominiert der Photonaustausch, d.h.  $F_3$  verschwindet. Der Koeffizient  $A_f$  ist dann gleich dem Quadrat der Fermionladungen.

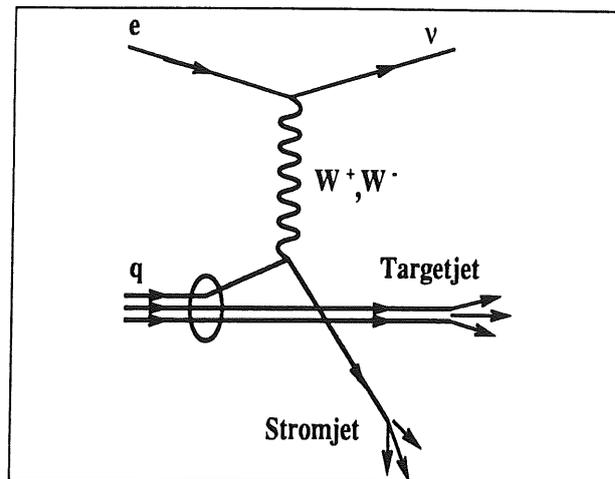


Abbildung 2.5: Feynman-Graph eines geladenen Stromereignisses

### 2.2.2 Geladene Ströme

Der Feynmangraph in Abbildung 2.5 unterscheidet sich nur geringfügig vom vorhergehenden. Diesmal wird jedoch ein geladenes Vektorboson  $W^\pm$  ausgetauscht, das heißt, der Endzustand ist gekennzeichnet durch ein auslaufendes Neutrino, ein Quark und ein Diquark, die wiederum in Hadronen fragmentieren.

Dies bedeutet experimentell einen großen Unterschied, denn das Neutrino kann nicht nachgewiesen, lediglich der fehlende Transversalimpuls festgestellt werden. Ein weiterer Unterschied besteht darin, daß diesmal nur massive Vektorbosonen ausgetauscht werden, und somit die Reaktionsrate gegenüber den neutralen Strömen, die vom  $\gamma$ -Austausch dominiert werden, stark unterdrückt ist.

Der differentielle Wirkungsquerschnitt für geladene Ströme ist gegeben durch

$$\frac{d^2\sigma_{CC}(e^-p)}{dx dQ^2} = \frac{\pi\alpha}{4 \sin^4 \Theta_W (Q^2 + M_W^2)^2} \sum_{i,j} \left[ |V_{u_i d_j}|^2 u_i(x, Q^2) + (1-y)^2 |V_{u_j d_i}|^2 \bar{d}_i(x, Q^2) \right],$$

wobei  $u$  und  $d$  die Impulsverteilungen der verschiedenen Quark-flavours bezeichnen [14].  $V_{ij}$  bezeichnet die Elemente der Cabibbo-Kobayashi-Maskawa-Matrix, welche die Mischung der Quarksorten beschreibt.  $M_W$  steht für die Masse des W-Bosons ( $M_W = 80.22$  GeV) und  $\Theta_W$  für den Weinberg-Winkel ( $\sin \Theta_W^2 = 0.23$ ) [15].

### 2.2.3 Boson-Gluon-Fusion

Zur Bildung von schweren Quarks ( $c - \bar{c}, b - \bar{b}$ ) in Quark-Antiquark-Paaren trägt hauptsächlich der Prozeß aus Abbildung 2.6 bei. Das Austauschboson wechselwirkt diesmal nicht mit einem Quark, sondern koppelt an ein Gluon aus dem Proton. Da

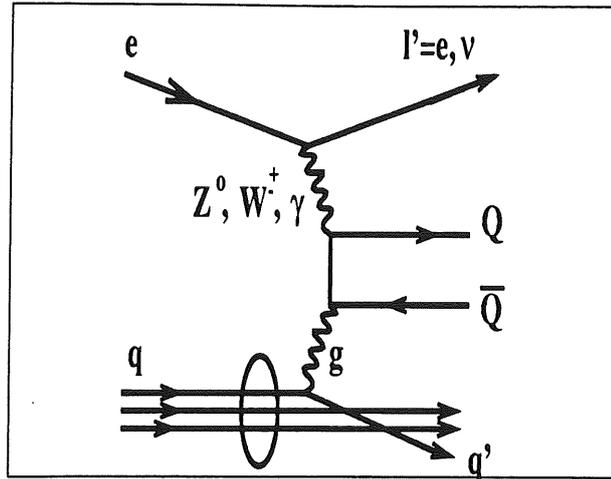


Abbildung 2.6: Feynman-Graph einer Boson-Gluon-Fusion

das Gluon nur der starken Wechselwirkung unterworfen ist, kann es nicht direkt an die schwachen Bosonen koppeln, sondern kann nur mit einem vom Boson erzeugten Quark-Antiquark-Paar wechselwirken. Auch in dieser Reaktion muß das Gluon als Parton behandelt und der Wirkungsquerschnitt durch eine Gluonstrukturfunktion  $G(x_g, M_g^2)$  beschrieben werden.

Der Endzustand dieser Reaktion ist vielfältiger als bei den Stromereignissen, da sowohl ein Neutrino als auch ein Elektron entstehen kann. Der Fragmentierungsprozeß der

Physikalischer Prozeß	$\sigma_{tot}$ (in nb)	Rate (in Hz)
Neutrale Ströme	152.200	2.283
Geladene Ströme	0.0688	$10^{-3}$
$c - \bar{c}$	439.800	6.6
$b - \bar{b}$	4.203	0.063

Tabelle 2.1: **Totale Wirkungsquerschnitte und erwartete Raten der wichtigsten Physikreaktionen bei einer Luminosität von  $1.5 \cdot 10^{31} \text{cm}^{-2} \text{s}^{-1}$ .** Bei den neutralen Stromereignissen wurde nur über  $x > 0.001$  integriert.

verschiedenen Quarks im hadronischen Endzustand kann durch Gluonemission zu mehreren Jets führen, die hauptsächlich in Protonrichtung weiterfliegen und schwer von Untergrundereignissen zu trennen sind.

Tabelle 2.1 faßt die totalen Wirkungsquerschnitte und die bei der Design-Luminosität erwarteten Raten für die beschriebenen Physikreaktionen zusammen [16].

### 2.2.4 Untergrundereignisse

Alle Ereignisse, die nicht von der Wechselwirkung zwischen Protonen und Elektronen herrühren, werden als Untergrundereignisse klassifiziert. Dazu gehören hauptsächlich Strahl-Gas-Reaktionen, Strahl-Wand-Reaktionen, kosmische Strahlung und Synchrotronstrahlung.

#### 2.2.4.1 Strahl-Gas-Reaktionen (Beam-Gas)

Kollisionen von Protonen mit den Restgasatomen im Vakuum des Strahlrohrs ( $p \approx 10^{-9}$  Torr), die Teilchenschauer hoher Multiplizität erzeugen [7], bezeichnet man als Strahl-Gas-Reaktionen.

Kollisionen von Elektronen mit Gas spielen im Vergleich zu  $p$ - $N$ -Reaktionen eine untergeordnete Rolle. Die Wirkungsquerschnitte der Proton-Kern-Reaktionen verhalten sich dabei proportional zu den elementaren Proton-Proton Wirkungsquerschnitten

$$\sigma_{p-N} \propto \sigma_{p-p} \cdot A^\alpha, \quad (2.16)$$

wobei  $A$  die Zahl der Nukleonen des Gasatoms beschreibt,  $\sigma_{p-p}$  den elementaren Wirkungsquerschnitt darstellt und  $\alpha \simeq 0.7$  die exponentielle Proportionalitätskonstante bezeichnet [17].

Die Teilchen des Endzustands können eingeteilt werden in

- Schauerteilchen, die aus einzelnen geladenen Hadronen bestehen und
- Kernfragmente, hauptsächlich Protonen.

Die mittlere Multiplizität der Schauerteilchen skaliert mit der Nukleonenzahl wie  $A^{0.14}$ , das bedeutet, daß die Multiplizitätsverteilungen von  $p$ - $p$ - und  $p$ - $Pb$ -Reaktionen sich nur um einen Faktor 2.1 unterscheiden. Das Impulsspektrum der Schauerteilchen zeigt eine beschränkte Verteilung des Transversalimpulses mit einem Mittelwert von  $\bar{p}_\perp \geq 400 \text{ MeV}/c$ .

Die Anzahl und Impulsverteilung der Kernfragmente ist bei hohen Energien nahezu unabhängig von der Energie der einfallenden Teilchen. Man unterscheidet herausgeschlagene und abgedampfte Teilchen.

Die herausgeschlagenen Teilchen zeigen ein Energiespektrum der Form

$$N(E)dE \propto E^{-\gamma}dE. \quad (2.17)$$

Dabei gibt  $N(E)$  die Zahl der Protonen pro Ereignis und Energieeinheit (MeV) an und  $\gamma$  den Wert  $1.09 \pm 0.02$  besitzt [18]. Die Winkelverteilung ist näherungsweise bestimmt durch

$$\frac{1}{\sigma} \frac{d\sigma}{d \cos \Theta} = e^{0.96 \cdot \cos \Theta} \quad (2.18)$$

Das Spektrum der abgedampften Teilchen folgt mit  $kT = m_\pi c^2$  aus

$$N(E) = e^{-\frac{E}{kT}} \quad (2.19)$$

und besitzt eine isotrope Winkelverteilung. Die mittlere Multiplizität von herausgeschlagenen und abgedampften Teilchen ist bis zum Wert 10 korreliert, danach geht die Multiplizität der abgedampften Teilchen beim Wert 13 in Sättigung über.

Die erwartete Rate für Proton-Kern-Reaktionen beträgt  $9000 \text{ Ereignisse m}^{-1}\text{s}^{-1}$  bei einem Protonenstrom von 480 mA (projektierter Strom der HERA-Maschine) und einem Druck von  $10^{-9}$  Torr. Der Detektor ist auf  $p$ - $N$ -Ereignisse aus bis zu 100 m Entfernung vor dem Wechselwirkungspunkt (WWP), gesehen aus der Protonrichtung, empfindlich.

#### 2.2.4.2 Strahl-Wand-Reaktionen (Beam-Wall)

Kollisionen von nicht auf der Sollbahn fliegenden Protonen mit Strahlführungskomponenten, bezeichnet man als Strahl-Wand-Reaktionen. Bei einer Protonenstrahl-Lebensdauer von 10 Stunden erwartet man eine Protonenverlustrate von  $3 \cdot 10^5 \text{ s}^{-1}\text{m}^{-1}$ , die bedeutend höher liegt als die Rate für Strahl-Gas-Reaktionen in Detektornähe. Ein Großteil dieser Protonen kann jedoch mit sogenannten *Scrapern* abgeschirmt werden. Trotz allem können hochenergetische Myonen aus großen Entfernungen alle Abschirmungen durchdringen oder Protonen nach dem letzten Scraper mit dem Quadrupolmagneten 27 m vor dem Detektor kollidieren und Teilenschauer erzeugen, deren Rate etwa  $1 \times 10^5 \text{ Hz}$  beträgt.

#### 2.2.4.3 Synchrotronstrahlung

Elektromagnetische Strahlung entsteht bei der Beschleunigung von Ladungen. Die Radialbeschleunigung, die in einem Speicherring die Teilchen auf der Bahn hält, veranlaßt somit die Teilchen, Photonen abzustrahlen. Um bei HERA die beiden separaten Ringe in den Wechselwirkungszonen zusammenzuführen, wird die Elektronenbahn zusätzlich gekrümmt; dies führt zu vermehrter Synchrotronstrahlung.

Mit Hilfe mehrerer Absorbermasken ist es möglich, die Photonrate um 10 Größenordnungen auf  $10^8 \text{ s}^{-1}$  in der zentralen Detektorregion ( $\pm 2.5 \text{ m}$  um den WWP) zu senken. Dies entspricht 10 Photonen bei der alle 96 ns stattfindenden Wechselwirkung zwischen Elektronen- und Protonenbündel. Da dies i.a. nicht zu Spuren im Zentraldetektor oder nennenswerten Energien im Kalorimeter führt, wird dieser Untergrund in den nachfolgenden Untersuchungen nicht berücksichtigt.

#### 2.2.4.4 Kosmische Strahlung

Die primäre kosmische Strahlung besteht zum größten Teil aus Protonen und  $\alpha$ -Teilchen, die beim Eintritt in die Erdatmosphäre mit den Gasatomen kollidieren und eine Vielzahl von sekundären Teilchen erzeugen.

Nur Myonen mit einer relativ großen mittleren Lebensdauer von  $2.2 \cdot 10^{-6} \text{ s}$  (Eigenzeit) erreichen die Erdoberfläche mit hohen Raten. Bei den HERA-Experimenten ist dieser Untergrund ernster zu nehmen als bei bisherigen; dies liegt daran, daß die Detektoren aufgrund der hohen Repetitionsrate von 10 MHz kein Zeitfenster für die Wechselwirkung besitzen. Die Rate, mit der kosmische Myonen das Kalorimetervolumen durchdringen, beträgt wegen der Größe des Detektors  $10^3 \text{ Hz}$ . Berücksichtigt

man jedoch den Abschirmungseffekt des Kalorimeters und die sehr kleine Wechselwirkungszone, die die Myonen treffen müssen, um für ein Physikereignis gehalten zu werden, dann reduziert sich die Rate auf 0.1 Hz. Sekundärteilchen der Myonstreuung in der Atmosphäre erreichen den Detektor mit einer Rate von 1 – 2 Hz und hinterlassen massive Energiedepositionen im Kalorimeter. Dies bedeutet aber erst ein Problem in den Software-Stufen des Triggersystems.

# Kapitel 3

## Das Triggersystem des H1-Detektors

In der nördlichen Wechselwirkungszone des HERA-Rings befindet sich das H1-Experiment. Um eine möglichst vollständige Information über die stattfindenden Kollisionsereignisse zu erhalten, umschließt der H1-Detektor hermetisch die Wechselwirkungszone. Aufgrund der unterschiedlichen Energien und Impulse von Elektronen und Protonen ist das Schwerpunktsystem nicht mit dem Laborsystem identisch, wie dies bei den meisten  $e^+e^-$ -Speicherringen der Fall ist. Dies hat zur Folge, daß der Detektor nicht symmetrisch bezüglich des Wechselwirkungspunktes aufgebaut ist, sondern der Ereignistopologie entsprechend in Richtung des Protons mehr und aufwendigere Detektorkomponenten aufweist.

Die Hauptkomponenten des H1-Detektors sind schalenförmig um die Wechselwirkungszone angeordnet (siehe Abb. 3.1), beginnend mit den Spurdetektoren aus Drift- und Proportionalkammern, gefolgt von elektromagnetischem und hadronischem Kalorimeter, die die Hauptarbeit bei der Energiemessung übernehmen. Instrumentiertes Eisen und Myonkammern zur Detektion minimalionisierender Teilchen bilden die äußerste Schicht des Detektors.

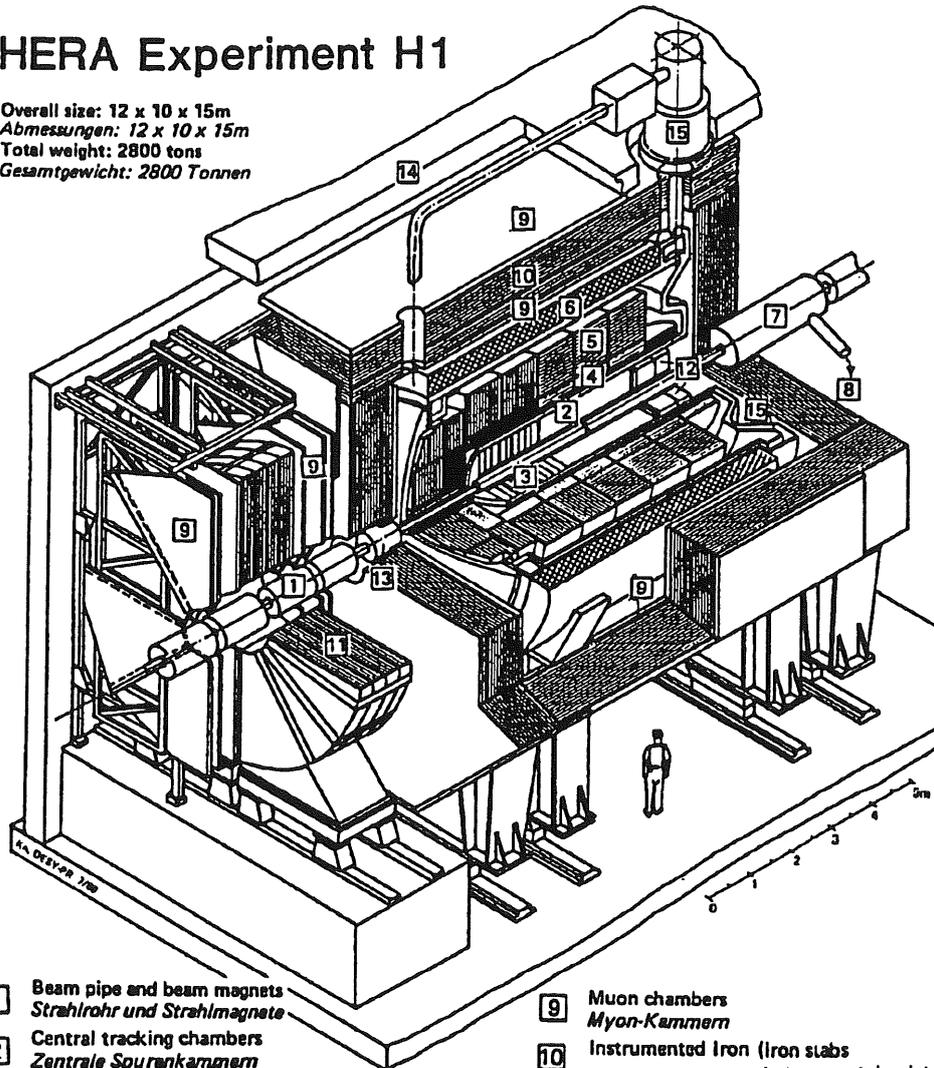
### 3.1 Das mehrstufige H1-Triggerkonzept

In den Wechselwirkungszone des HERA-Speicherrings durchdringen sich alle 96 nsec ein Elektronen- und ein Protonenbündel. Es ist nicht möglich, den Detektor bei jeder Strahlkreuzung auszulesen und die Daten aufzuzeichnen, da weder die Ausleseelektronik schnell genug ist, noch genügend Bandbreite zur Datenübertragung für 10 Millionen Ereignisse pro Sekunde zur Verfügung steht. Dies ist aber auch nicht erforderlich, da in der Mehrzahl der Fälle keine physikalisch interessanten Kollisionen stattfinden, und nur Untergrundereignisse aufgezeichnet werden würden.

Die erwarteten Untergrundraten liegen in der Größenordnung von mehreren 100 kHz, die Aufzeichnungsrate ist aber auf fünf bis zehn Ereignisse pro Sekunde limitiert. Deshalb besteht die Aufgabe des Triggers nicht nur darin, zu erkennen, ob ein Ereignis stattgefunden hat, sondern auch festzustellen, ob es sich um ein erwünschtes Physik-

## HERA Experiment H1

Overall size: 12 x 10 x 15m  
 Abmessungen: 12 x 10 x 15m  
 Total weight: 2800 tons  
 Gesamtgewicht: 2800 Tonnen



- |   |  |    |  |
|---|--|----|--|
| 1 | Beam pipe and beam magnets<br><i>Strahlrohr und Strahlmagnete</i>  | 9  | Muon chambers<br><i>Myon-Kammern</i>   |
| 2 | Central tracking chambers<br><i>Zentrale Spurenkammern</i>   | 10 | Instrumented Iron (iron slabs<br>+ streamer tube detectors)<br><i>Instrumentiertes Eisen (Eisenplatten +<br/>Streamerröhren-Detektoen)</i> |
| 3 | Forward tracking chambers and Transition radiators<br><i>Vorwärtsspurenkammern und Übergangstrahlungsmodul</i> | 11 | Muon toroid magnet<br><i>Myon-Toroid-Magnet</i>  |
| 4 | Electromagnetic Calorimeter (lead)<br><i>Elektromagnetisches Kalorimeter (Blei)</i>                            | 12 | Warm electromagnetic calorimeter<br><i>warmes elektromagnetisches Kalorimeter</i>  |
| 5 | Hadronic Calorimeter (stainless steel)<br><i>Hadronisches Kalorimeter (Edelstahl)</i>                          | 13 | Plug calorimeter (Cu, Si)<br><i>Vorwärts-Kalorimeter</i>   |
| 6 | Superconducting coil (1.2 T)<br><i>Supraleitende Spule (1,2 T)</i>   | 14 | Concrete shielding<br><i>Betonabschirmung</i>  |
| 7 | Compensating magnet<br><i>Kompensationsmagnet</i>  | 15 | Liquid Argon cryostat<br><i>Flüssig Argon Kryostat</i>   |
| 8 | Helium cryogenics<br><i>Helium Kälteanlage</i>   |    |  |

Abbildung 3.1: Perspektivische Darstellung des H1-Detektors .

Triggerstufe	Verarbeitungszeit	Rate	Information	Realisierung
L1	2 $\mu$ s	5 · 10 <sup>3</sup> Hz	8-bit Energiesummen Spursignale	Diskriminatoren Hardware-Logik
L2	20 $\mu$ s	100 Hz	8-bit Energiesummen Spursignale	komplexere Hardware zur Mustererkennung
L3	1 ms	50 Hz	Triggerdaten	RISC-Prozessoren
L4	20 ms	5 Hz	Detektordaten	RISC-Prozessoren

Tabelle 3.1: Unterteilung des H1-Trigger-Systems.

ereignis oder aber um Untergrund im Sinne der Diskussion von Abschnitt 2.2.4 handelt. Um diese Aufgabe zu bewältigen, erhält das H1-Trigger-System seine Informationen sowohl räumlich als auch zeitlich untergliedert [9]. Jeder eigenständige logisch-funktionale Teil des Detektors (Subdetektor) bildet ein eigenes Subtriggersignal aus seinen Meßdaten und liefert dieses an eine zentrale Triggerlogik oder eine Ausleseelektronik.

Die Auswertung der zeitlich unterschiedlich einlaufenden Information geschieht durch vier Stufen (*engl. Level*  $\Rightarrow$  L1, L2, L3, L4), die nacheinander ablaufen und unterschiedlich viel Zeit für ihre Entscheidung zur Verfügung haben. Sobald ein Ereignis von einer Stufe akzeptiert wird, beginnt die nächsthöhere damit, ihren Entscheidungsfindungsprozess abzuarbeiten <sup>1</sup>.

Die ersten beiden Stufen L1 und L2 arbeiten auf Hardware-Basis mit festverdrahteter Logik, die Stufen L3 und L4 sind Software-Trigger, die mit Hilfe schneller Mikroprozessoren realisiert werden. Die Tabelle 3.1 gibt einen Überblick über die veranschlagten maximalen Verarbeitungszeiten und die angestrebte Ausgangsrate zur nächsten Stufe. Die Zeiten und Raten sind dadurch bestimmt, daß auf jeder Triggerstufe nach L1 ein Totzeitanteil von maximal 10% erzeugt werden darf. Dies bedeutet, daß das Produkt aus Eingangsrate und Verarbeitungszeit kleiner als 0.1 sein muß. Nur der L1 Trigger bildet eine Ausnahme, da er totzeitfrei arbeitet.

- **Stufe 1 (L1)**

Die erste Triggerstufe ist als konventioneller Schwellentrigger aufgebaut. Jeder Subdetektor liefert an die zentrale L1-Triggerlogik eines oder mehrere Triggersignale, die dort diskriminiert und logisch verknüpft werden können.

Der L1-Trigger ist natürlich nicht in der Lage, schon nach 96 ns (ein HERA-Zyklus) eine Entscheidung zu fällen. Trotzdem gelingt es einen totzeitfreien Trigger zu konzipieren, indem man das *Pipelining*-Verfahren anwendet. Hierbei werden die Informationen der einzelnen Strahlkreuzungen solange nach dem "First-In-First-Out" Prinzip zwischengespeichert, daß die Triggerlogik genügend Zeit hat, eine Entscheidung zu fällen. Beim H1-Trigger werden ca. 2.2  $\mu$ s benötigt,

<sup>1</sup>Die Stufe L3 startet bereits, bevor L2 entschieden hat.

die einer Länge von 23 HERA-Zyklen in der Pipeline entspricht; danach fällt das Ereignis entweder heraus, wenn es nicht getriggert wurde, oder die Pipeline wird angehalten und die Auslese beginnt. Zu diesem Zeitpunkt beginnt auch die Totzeit.

- **Stufe 2 (L2)**

Alle Informationen, die dem L1-Trigger zur Verfügung stehen, sind auch dem L2-Trigger zugänglich. Hinzu kommt aber noch Information durch die Auslese von digitalen Signalprozessoren, die mehr Details und weitere Subtriggersignale liefern.

Die zweite Stufe kann also sowohl versuchen, die Level-1-Information besser auszuwerten, als auch, den neu hinzugekommenen Daten mehr Bedeutung beizumessen. Beide Möglichkeiten sind aber dadurch beschränkt, daß lediglich  $20 \mu\text{s}$  zur Entscheidungsfindung ausreichen müssen. Ein Weg zur Vereinigung beider Möglichkeiten besteht darin, Korrelationen zwischen verschiedenen Triggersignalen (Level-1 und Level-2 Informationen) zu finden, die eine deutliche Trennung von Untergrund- und Physikereignissen ermöglichen.

- **Stufe 3 (L3)**

Die dritte Stufe wird durch mehrere RISC-Prozessoren realisiert, denen sowohl L1- und L2-Informationen als auch ein Teil der zwischenzeitlich ausgelesenen detaillierteren Detektordaten vorliegen. Hier können per Software zum erstenmal Kombinationen und kompliziertere Verknüpfungen vorgenommen und zum Triggern verwendet werden.

Solange der L3-Trigger das Ereignis nicht verwirft, wird der Detektor weiter vollständig ausgelesen und die Daten zwischengespeichert; die Auslese benötigt etwa  $800 \mu\text{s}$ .

- **Stufe 4 (L4)**

Der auf einer Farm von RISC-Prozessoren basierende Level-4-Trigger hat die gesamten ausgelesenen Detektordaten zur Verfügung und kann zum erstenmal das Ereignis rekonstruieren. Aber auch hier kann auf Grund der beschränkten Zeit von 150 ms nur eine Teilrekonstruktion für wenige Detektorkomponenten durchgeführt werden.

Mit den auf Software-Ebene realisierbaren Schnitten ist eine flexible und effiziente Möglichkeit gegeben, Untergrundereignisse sehr selektiv zu unterdrücken und die geforderte Datennahmerate zu erreichen. Die Daten der von Level 4 akzeptierten Ereignisse werden dauerhaft auf Magnetband gespeichert und stehen der späteren Analyse zur Verfügung.

## 3.2 Triggerkomponenten des H1-Detektors

Die Subdetektoren von H1 stellen schnelle Signale für das Triggersystem zur Verfügung. Die Spurdetektoren versuchen mit Mustererkennungssystemen, die von der Wechselwirkungszone kommenden Teilchenbahnen zu rekonstruieren. Das Kalorimeter liefert Energiesummen aus verschiedenen Bereichen des Detektors und die Myonkammern

erkennen den Durchgang von Teilchen in verschiedenen Lagen des Myonsystems. Aufgrund der benötigten Schnelligkeit ( $< 22\mu\text{s}$ ), sind die Triggersignale nicht so exakt wie die später ausgelesenen Daten, so unterscheiden sich z.B. die Triggerenergien des Kalorimeters um einige Prozent von den rekonstruierten Werten. Im folgenden werden die wichtigsten Unterdetektoren vorgestellt und die von ihnen gelieferten Triggergrößen beschrieben.

### 3.2.1 Das Flüssig-Argon-Kalorimeter

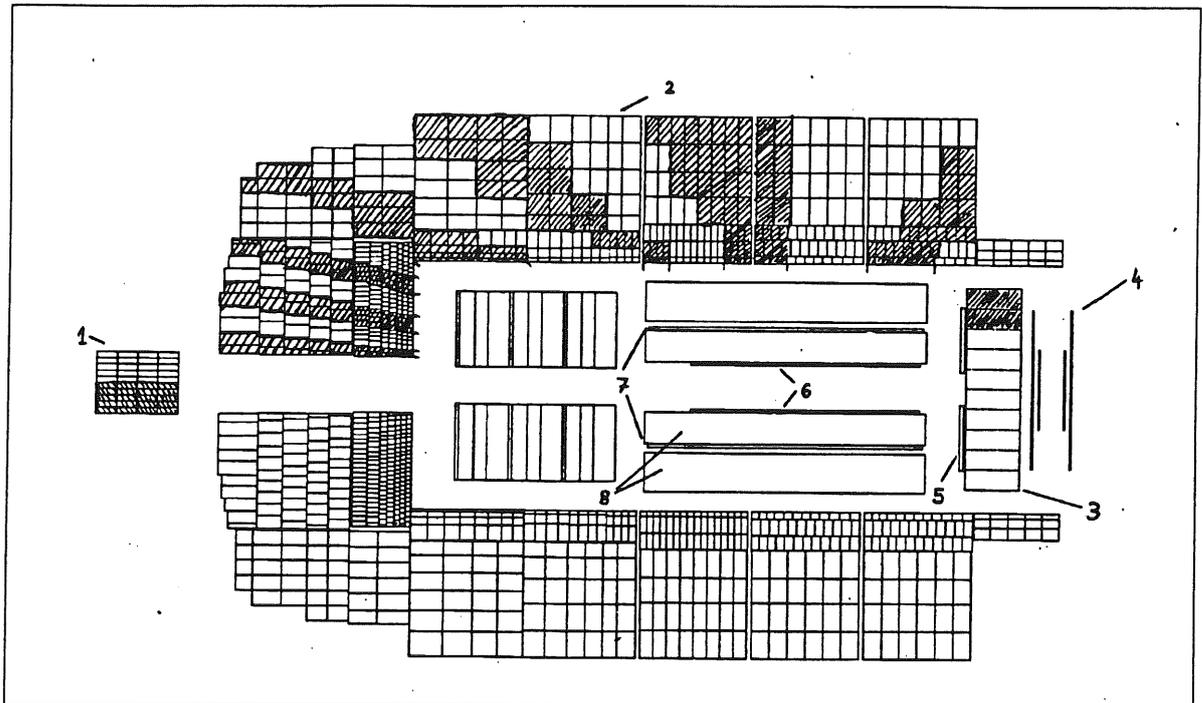


Abbildung 3.2: Schnitt durch den H1-Detektor (ohne Myonsystem). Im Kalorimeter (2) ist die Einteilung in Bigtowers schraffiert dargestellt. Außerdem sind das PLUG (1), das BEMC (3), die TOF-Wände (4), die BPC (5), die inneren Proportional- und Z-Kammern (6), die äußeren Proportional- und Z-Kammern (7) und Jetkammern (8) dargestellt.

Das Kalorimeter ist ein Schicht-("sampling")-Kalorimeter und ist unterteilt in einen (inneren) elektromagnetischen Bereich (siehe (4) in Abb. 3.1), der Blei als Absorbermaterial verwendet, und einen (äußeren) hadronischen Bereich (siehe (5) in Abb. 3.1), in dem Eisen als Absorber Verwendung findet. Als aktives Medium, in dem die in den Absorbern produzierten Teilchenschauer gemessen werden, dient in beiden Bereichen flüssiges Argon.

Die 40000 Zellen des Kalorimeters werden für die Auslese in sogenannte Türme ("Towers") zusammengefaßt, die auf den Wechselwirkungspunkt zeigen. Für das Triggersignal werden diese nochmals zu sogenannten Großtürmen (*engl. Bigtowers*) zusammengefaßt (siehe Abb. 3.2). Es besteht eine 15-fache Unterteilung im Azimutwinkel  $\Theta$  und eine, je nach Raumwinkel und Detektorgranularität, 8- bis 32-fache Unterteilung im

Polarwinkel  $\Phi$ , so daß sich eine Gesamtzahl von 252 Großtürmen ergibt [19].

Der Kalorimeter-Sub-Trigger liefert weitere Energiesummen in verschiedenen Regionen des Kalorimeters, die sich diesmal auf den Vorwärts-, Rückwärts- und Zentralbereich beziehen. Außerdem werden noch die transversale Energie

$$E_T = \sum_{BT} |E_T(BT)| \quad \text{wobei} \quad E_T(BT) = E(BT) \sin \Theta_{BT} \quad (3.1)$$

gemessen, die fehlende transversale Energie  $E_{T_{\text{miss}}} = \sum_{BT} E_T(BT)$  bestimmt, sowie die Gesamtenergie ausgerechnet und der zentralen Triggerlogik zur Verfügung gestellt.

### 3.2.2 Das PLUG-Kalorimeter

Das PLUG-Kalorimeter dient zur Messung der Energiedeposition im extremen Vorwärtsbereich ( $12.5 \text{ mrad} \leq \Theta \leq 60 \text{ mrad}$ ) (siehe (13) in Abb. 3.1 und (5) in Abb. 3.2). Es ist als silizium-instrumentiertes Sandwich-Kalorimeter mit Kupfer als Absorber konzipiert. Für den Trigger wird die PLUG-Energie in die Energiesummen des Flüssig-Argon-Kalorimeters in Form von Bigtowers miteinbezogen.

### 3.2.3 Das rückwärtige elektromagnetische Kalorimeter

Um den Winkelbereich von  $152^\circ \leq \Theta \leq 176^\circ$  im rückwärtigen Teil des Detektors in die Messungen miteinzuschließen, besitzt das H1-Experiment ein Blei-Szintillator-Kalorimeter, das aus mehreren Lagen aufgebaut ist (siehe (12) in Abb. 3.1 und (3) in Abb. 3.2). Das rückwärtige elektromagnetische Kalorimeter (Backward Electromagnetic Calorimeter, BEMC) ist in 88 Zellen unterteilt, deren Energien über Wellenlängenschieber ausgelesen werden können [20]. Als L1-Triggersignale liefert das BEMC vier Bits, die beschreiben, ob die Energie in einer der Zellen oberhalb dreier verschiedener Schwellen liegt und ob die Gesamtenergie eine gesetzte Schwelle überschritten hat. Für die höheren Triggerstufen stehen die Anzahl der getroffenen Zellen, die über einer Schwellenenergie liegen und die Energien selbst zur Auswertung bereit.

### 3.2.4 Die zentralen Jetkammern

Zur genauen Spurmessung steht im H1-Experiment ein leistungsfähiges Driftkammersystem zur Verfügung. Dies ist in zwei Bereiche aufgeteilt (Central Jet Chamber, CJC1 und CJC2) und sind mit insgesamt 2640 Driftdrähten bestückt. Die Länge der parallel zur Strahlachse verlaufenden Drähte beträgt 2.49 m. Mit diesem System ist sowohl eine Bestimmung der  $x$ - und  $y$ -Koordinaten als auch über Ladungsteilung eine Festlegung der  $z$ -Koordinate möglich.

Der Sub-Trigger hat allerdings nicht die Möglichkeit die volle Information der Jetkammern durch einen komplizierten Spurfindungs-Algorithmus auszuschöpfen, da dies zu viel Zeit benötigen würde. Deshalb werden lediglich die Muster der getroffenen Drähte mit vorgegebenen Mustern verglichen, welche Spuren mit fester Krümmung und Azimutwinkel entsprechen. Dadurch läßt sich feststellen, ob eine Spur in der  $r - \Phi$ -Projektion vom Wechselwirkungspunkt stammt.

Die zentrale Triggerlogik auf Level 1 erhält so Information über die Anzahl gefun-

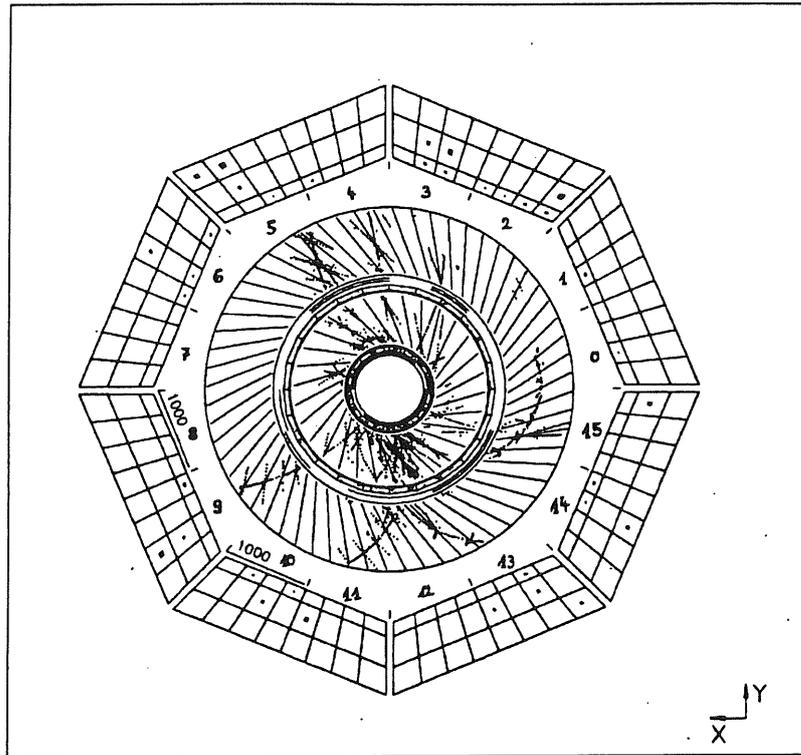


Abbildung 3.3: **Radialer Schnitt durch den Spurdetektor und das elektromagnetische Kalorimeter.** Von innen nach außen sind die sechs Komponenten des Spursystems (CIP, CIZ, CJC1, COZ, COP, CJC2) und die Treffer, die ein Ereignis hinterläßt, dargestellt. Die 16-fache Unterteilung der Bigrays, die der Unterteilung in Kalorimeter-Bigtowers angepaßt ist, wird zwischen CJC2 und dem elektromagnetischen Kalorimeter gezeigt.

dener Spuren mit positiver oder negativer Krümmung und hohem ( $> 800$  MeV) oder niedrigem Transversalimpuls ( $< 800$  MeV).

### 3.2.5 Die zentralen Proportionalkammern

Im zentralen Spurdetektor gibt es ein inneres und ein äußeres Proportionalkammersystem (Central Inner Proportional Chamber, CIP; Central Outer Proportional Chamber, COP). Die Anodendrähte sind konzentrisch um die Strahlröhre aufgespannt, wobei es in der CIP drei Lagen und in der COP zwei Lagen von Drähten gibt (siehe Abb. 3.3 und (6),(7) in Abb. 3.2). Die Kathodenebenen der Vieldrahtproportionalkammern (Multiwire Proportional Chambers, MWPC) sind in sogenannte *Pads* unterteilt,  $8\Phi \times 60z$ -Segmente in der CIP,  $32\Phi \times 60z$ -Segmente in der COP, die eine grobe Ortsinformation über Teilchendurchgänge liefern. Verlangt man Koinzidenz von Informationen aus CIP und COP bezüglich Spuren, die vom Wechselwirkungspunkt stammen, so erhält man Information über Teilchenrichtung und Ereignisursprung.

Zu Triggerzwecken wird ein Z-Vertex-Histogramm erstellt, das die Projektion der gefundenen Spuren (Koinzidenzen) auf die Strahlachse darstellt. Dieses Histogramm besteht aus 16 Bins von 5.2 cm Länge, die somit einen Bereich von  $\pm 40$  cm um den Wechselwirkungspunkt überstreichen. Physikereignisse sollten ein deutliches Maximum in diesem Histogramm hinterlassen, Untergrundereignisse jedoch keine oder stark verbreiterte Verteilungen erzeugen.

Für den Level-1-Trigger wird aus diesem Histogramm eine Z-Vertex-Signifikanz errechnet, die ein Maß für die Eindeutigkeit des Maximums darstellt und folgendermaßen definiert ist:

$$\sigma_z = \frac{n_{Max} - \frac{n_{Bgr}}{15}}{\sqrt{n_{Max}}} \quad (3.2)$$

Die Anzahl der Einträge im Maximum-Bin wird hierbei mit  $n_{Max}$ , die Gesamtzahl der Histogrammeinträge ohne den Maximum-Bin mit  $n_{Bgr}$  bezeichnet. Die Richtungsinformation der Spuren (Bigrays) kann zur Verifikation von Bigtower-Energien im Kalorimeter verwendet werden (siehe Abbildung 3.3, insbesondere die 16  $\Phi$ -Sektoren, in welche die Bigrays weisen).

### 3.2.6 Die Z-Kammern

In direkter Nachbarschaft zu den Proportionalkammern sind im H1-Detektor Driftkammern (Central Inner Z-Chamber, CIZ; Central Outer Z-Chamber, COZ) installiert (siehe Abbildung 3.4). In diesen Kammern sind pro Auslezelle vier Anodendrähte konzentrisch um die Strahlachse gespannt, das äußere Kammersystem besteht aus 24, das innere aus 15 Zellen.

Das Z-Kammersystem dient ebenfalls dem schnellen Nachweis von Teilchenspuren, erlaubt aber eine bessere Orts- und Richtungsauflösung als die Proportionalkammern. In Analogie zu den Proportionalkammern erzeugen auch die Z-Kammern ein Z-Vertex-Histogramm aus Koinzidenzen zwischen CIZ- und COZ-Spursegmenten. Dieses Histogramm besitzt aber 96 Bins mit einer Auflösung von 5 mm pro Bin. Zusätzlich werden sogenannte *Vertex-Zell-Histogramme* und *Untergrund-Zell-Histogramme* angelegt, deren Bins den Auslezellen der Z-Kammern zugeordnet sind. Das Vertex-Zell-Histogramm erhält einen Eintrag, wenn das Spursegment einer Zelle zum Vertex zeigt, andernfalls gibt es einen Eintrag in das Untergrund-Zell-Histogramm.

Die Analyse der Z-Vertex-Histogramme erfolgt für den Level-1-Trigger wie im Falle der Proportionalkammern durch eine Vertex-Signifikanz-Bestimmung. Durch eine komplexere Auswertung aller Histogramme können Signale für eine Level-2-Triggerlogik zur Verfügung gestellt werden [21].

### 3.2.7 Die rückwärtigen Proportionalkammern

Im rückwärtigen Teil des Detektors, direkt vor dem BEMC, befinden sich die rückwärtigen Proportionalkammern (Backward Proportional Chambers, BPC) zum Nachweis von geladenen Teilchen (siehe (5) in Abb. 3.2). Sie bestehen aus zwei Doppelkammern mit orthogonal gespannten Drähten, die in einer Ebene senkrecht zur Strahlachse stehen. Sie dienen zur Bestimmung der  $x$ - und  $y$ -Koordinaten der Teilchendurchgänge.

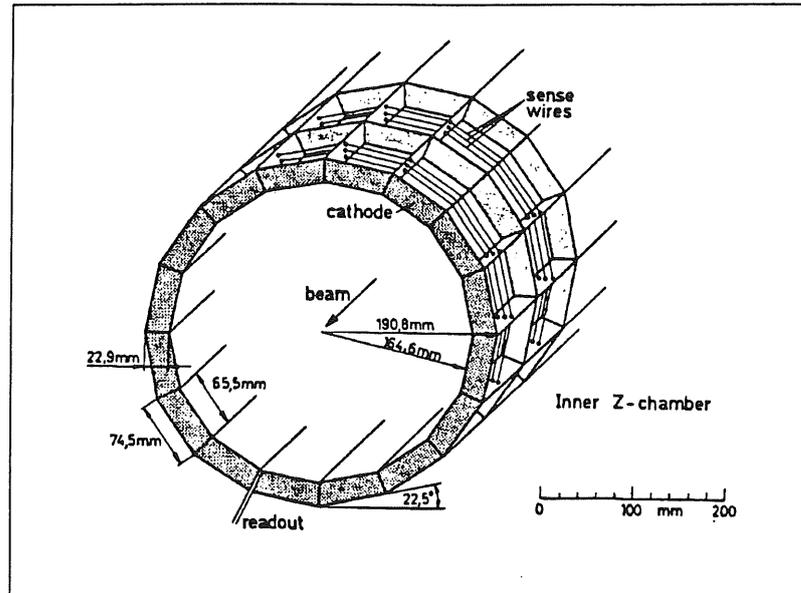


Abbildung 3.4: **Aufbau der inneren Z-Kammer.** Die CIZ ist in 15 Zellen unterteilt, die durch die Ebene der Nachweisdrähte nochmals in zwei Halbzellen separiert werden. Die Driftwege verlaufen parallel zur Strahlachse in der Zylinderebene. Die äußere Z-Kammer (COZ) ist prinzipiell gleich aufgebaut, unterscheidet sich aber in der Anzahl der Zellen ( $n_{COZ} = 24$ ).

Das Triggersystem erhält die Anzahl der getroffenen Drähte in der BPC als Signal.

### 3.2.8 Die Spurkammern im Vorwärtsbereich

Die Spurkammern im Vorwärtsbereich bestehen aus drei identischen *Supermodulen* (siehe (3) in Abb. 3.1). Jedes enthält ein planares Modul (drei Ebenen von parallel angeordneten Drähten, die im Winkel von  $60^\circ$  relativ zueinander stehen), gefolgt von einer Vieldrahtproportionalkammer, einer Schicht zur Erzeugung von Übergangsstrahlung und einem radialen Modul (eine Ebene mit radial gespannten Anodendrähten). Die Proportionalkammer liefert Informationen für den "Forward Ray"-Trigger, der die Anzahl von Spuren im Vorwärtsbereich bestimmt. Aufgrund von Ausfällen in der Hardware wurden in den Untersuchungen keine Informationen dieser Detektorteile verwendet.

### 3.2.9 Die Myon-Kammern

Zur Detektion minimal ionisierender Teilchen, insbesondere Myonen, die das Kalorimeter unbeeinflusst durchdringen können, dienen die Myonkammern, die als äußerster Detektorteil im Eisenjoch untergebracht sind. Dieses Eisenjoch besteht aus 10 Lagen von Eisenplatten mit 7.5 cm Dicke und hat die Aufgabe, den magnetischen Fluß der supraleitenden Spule zurückzuführen.

Insgesamt ist das Eisen mit 16 Lagen von Streamerkammern instrumentiert, wobei die Lagen 0/1, 6/7 und 14/15 als versetzte Doppellagen realisiert sind [22]. Das Myonsystem setzt sich aus vier Bereichen zusammen, der vorderen und hinteren Endkappe

und dem vorderen und hinteren Barrelteil. Jeder Bereich ist in 16 Module eingeteilt, [23], die Barrelbereiche in 16  $\Phi$ -Segmente, die Endkappen in ca. 90 cm hohe aufeinander gestapelte Module (siehe Abb. 3.5).

Als Triggersignale stellt der Myondetektor nur Informationen aus den Lagen 3, 4, 5,

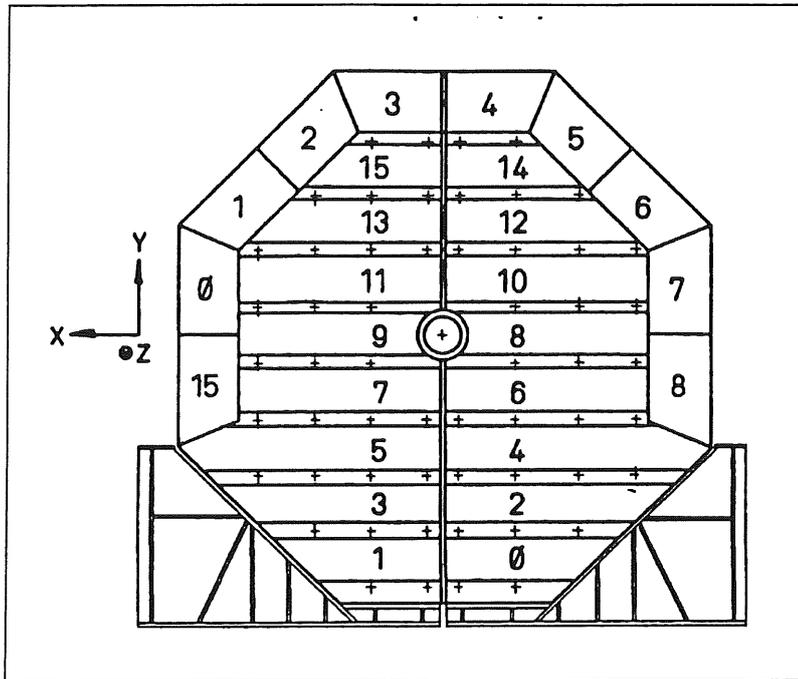


Abbildung 3.5: Frontansicht des Myon-Systems. Die Module der Endkappe sind von unten nach oben von 0 bis 15 numeriert. Ein Teil der Module des Barrelbereichs ist ebenfalls dargestellt und zylindrisch im Uhrzeigersinn beziffert.

8 und 12 bereit, dies aber aus allen Modulen [24]. Die zentrale Triggerlogik auf Level 1 erhält dabei nur die Anzahl der detektierten Myonen. Der Subtrigger zählt ein Teilchen, wenn es in mindestens drei Triggerlagen eines Moduls ein Signal ausgelöst hat. In höheren Triggerebenen ist eine feinere Granularität verwendbar, die z.B. topologische Eigenschaften des Ereignisses anzeigen kann.

### 3.2.10 Das Luminositäts-Meßsystem

Das Luminositäts-Meßsystem befindet sich weit außerhalb des eigentlichen Detektors 35 m bzw. 105 m vom WWP entfernt (Abb. 3.6). Die Hauptkomponenten sind ein Hodoskop aus  $7 \times 7$  KRS-6 Kristallen ( $22 \times 22 \times 200 \text{ mm}^3$ ), das Elektronen nach Kollisionen mit geringem Impulsübertrag und damit flachem Ablenkwinkel quasispektrometrisch nachweist, und ein weiteres Hodoskop aus  $5 \times 5$  KRS-6 Kristallen ( $20 \times 20 \times 200 \text{ mm}^3$ ), welches als Photon-Detektor flach abgestrahlte  $\gamma$ -Quanten erfaßt. Die Anordnung und Unterteilung der beiden Detektoren ist der Abbildung 3.6 zu entnehmen. Als Triggerinformation steht die registrierte Energie in den einzelnen Detektoren zur Verfügung.

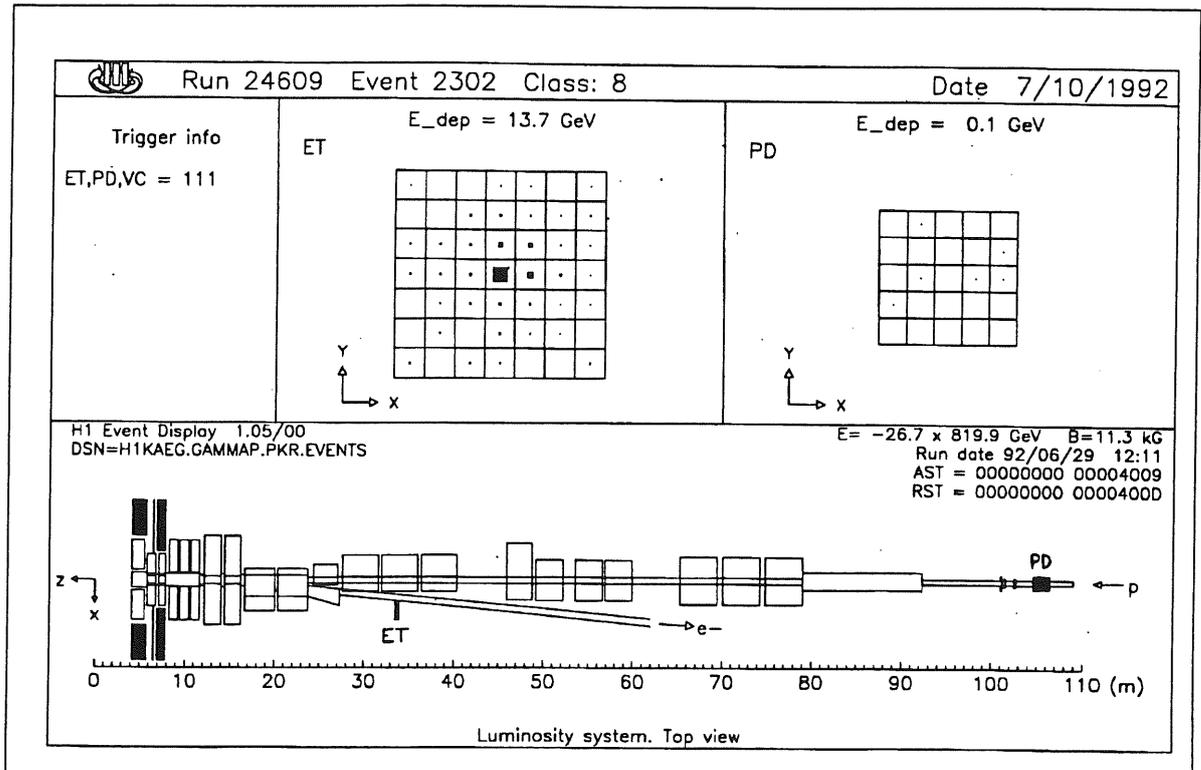


Abbildung 3.6: Das H1-Luminositäts-Meßsystem. Die untere Hälfte des Bildes zeigt einen Blick von oben auf das System, der Elektron-Detektor (ET) ist 35 m, der Photon-Detektor (PD) 105 m vom WWP entfernt. Oben links ist der Elektron-Detektor von vorn dargestellt. Die Einteilung in Module und die von einem Elektron stammenden Energiedepositionen sind deutlich erkennbar. Oben rechts ist der Photon-Detektor, in dem keine Energie registriert wurde, in der selben Ansicht zu sehen.

### 3.2.11 Das Flugzeit-Meßsystem (TOF)

Zur Unterdrückung von Haloteilchen des Protonenstrahls und anderen Reaktionen, die weit vor dem Detektor stattfinden, in diesem aber ein Signal hinterlassen, dient eine Flugzeitmessung. Diese Messung wird von zwei Szintillator-Ebenen durchgeführt, welche sich direkt hinter dem BEMC befinden (siehe (4) in Abb. 3.2). Mit einer Zeitauflösung von 3 ns besitzt das System außerdem die Fähigkeit, ein Interaktionszeitfenster zu definieren und zu messen.

Dieses Fenster definiert die Zeit, die verstreicht, bis ein Proton von der Szintillator-Wand zum Wechselwirkungspunkt gelangt, dort einen Teilchenschauer auslöst, und dieser zur Szintillator-Wand zurückgekehrt ist. Tritt solch ein Ereignis ein, so liefert das TOF-("Time-Of-Flight")-System ein positives Flugzeit-Signal (POSI-TOF) an die zentrale Triggerlogik. Bei koinzidenten Signalen in beiden Szintillatorebenen liefert der Sub-Trigger ein Flugzeit-Veto (TOF-VETO) als Triggersignal.

### 3.2.12 Die zentrale L1-Triggerlogik

Die zentrale Triggerlogik sammelt die Triggersignale der einzelnen Subdetektoren. Sie bildet eine "Summe von Produkten", d.h. sogenannte *Sub-Trigger* werden aus Koinzidenzen verschiedener Signale ("Trigger-Elemente") gebildet. Diese Sub-Trigger können dann mit einem logischen *OR* zu weiteren Sub-Triggeren verknüpft werden. Die Zusammenstellung logischer Verknüpfungen ist variabel und wird im Experiment durch die zentrale Datenerfassung kontrolliert.

Ist eine Triggerbedingung erfüllt, so stoppt die zentrale Triggerlogik die L1-Pipeline und aktiviert den L2-Trigger.

# Kapitel 4

## Neuronale Netzwerke

### 4.1 Motivation

Die große Menge an Triggerinformation, die von den einzelnen Subdetektoren erhältlich ist, kann aufgrund der Zeitbegrenzung von  $2.2 \mu\text{s}$  auf der ersten Triggerstufe nur zu einem Teil ausgenutzt werden. Ein Level-2-Trigger hat folglich die Aufgabe, in den ihm zur Verfügung stehenden  $20 \mu\text{s}$  eine "intelligentere" Analyse der Triggersignale durchzuführen. Dies bedeutet, daß die Information schnell und auf verbesserte Weise verarbeitet werden muß. Ein Optimierungsprozeß ist dabei zwingend, da der multidimensionale Charakter der Information die gebräuchliche Methode mit Triggerschwellen zu einem Ratespiel machen würde und somit die Effizienz und Anwendbarkeit in Frage gestellt wären.

Die grundlegende Idee besteht darin, dynamische Korrelationen zwischen den verschiedenen Triggergrößen zu erkennen und auszunutzen. Dies sollte als Weiterentwicklung des Schwellentriggers betrachtet werden, welcher feste (statische) Diskriminatoren auf Triggergrößen anwendet. Das Erkennen von Physikereignissen durch Schnitte im mehrdimensionalen Raum der Triggervariablen, kann folglich als Verallgemeinerung der auf Level 1 üblichen eindimensionalen Schnitte betrachtet werden [25].

Die hohen Geschwindigkeitsanforderungen auf der zweiten Triggerstufe erfordern massiv parallele Algorithmen zur Informationsverarbeitung. *Neuronale Netzwerke* stellen eine sehr effiziente Realisierung paralleler Algorithmen zum Auffinden hochdimensionaler Korrelationen dar und werden in dieser Arbeit hinsichtlich ihrer Einsetzbarkeit als L2-Trigger untersucht.

### 4.2 Grundlagen

#### 4.2.1 Von der Biologie zur Mathematik

*Künstliche Neuronale Netzwerke* stellen eine Abstraktion des biologischen Vorbildes dar. Um das mathematische Modell, das durch Künstliche Neuronale Netzwerke auf elektronischer Ebene realisiert wird, besser zu verstehen, ist es sinnvoll, das biologische Grundelement der Netzwerke näher zu betrachten.

Dieser elementare Baustein ist in der Biologie die Nervenzelle (Neuron). Es ist möglich,

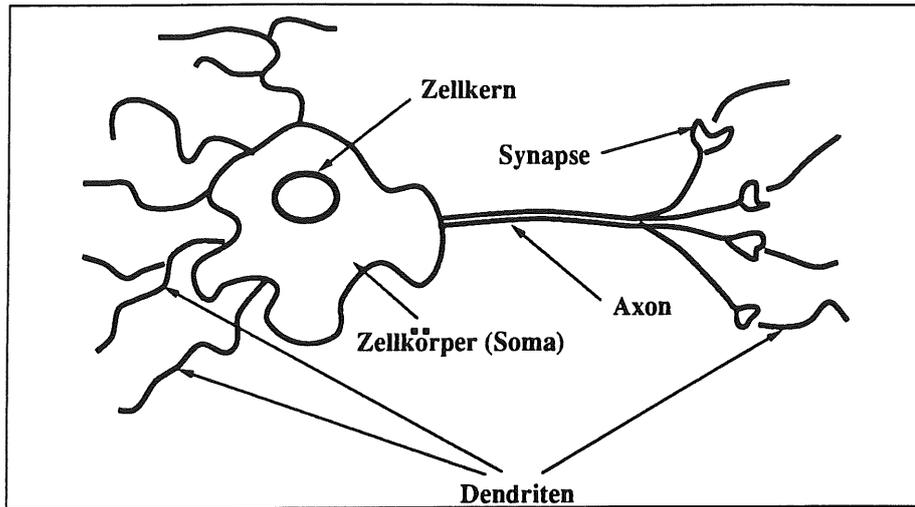


Abbildung 4.1: Darstellung einer Nervenzelle. Das Neuron erhält an den *Synapsen* Signale, die von *Dendriten* an den Zellkörper weitergeleitet und dort akkumuliert werden. Bei Übersteigen eines Grenzwertes wird das Signal mit Hilfe des *Axons* an andere Zellen weitergeleitet.

die Nervenzelle in weitere Funktionseinheiten aufzugliedern (siehe Abb. 4.1). Das Neuron bildet das eigentliche informationsverarbeitende Schaltelement, indem es durch Gewichtung und Summation von Eingangssignalen einen Ausgangsimpuls liefert, falls die Summe einen Schwellenwert überschritten hat.

Das Neuron empfängt seine Signale über *Dendriten*, die über *Synapsen* mit den Ausgängen (*Axonen*) der benachbarten Zellen verbunden sind. Die Gewichtung der Signale geschieht an den Synapsen durch sogenannte *Neurotransmitter*, welche auf chemischem Wege erregend (exzitatorisch) oder hemmend (inhibitorisch) wirken können.

Die im Zellkörper (*Soma*) von den Dendriten ankommenden elektrischen Signale bewirken eine Depolarisation der Zellmembran. Ab einem bestimmten Grenzwert (Schwelle) führt die Depolarisation zu einem elektrischen Impuls auf dem Axon, das diesen zu benachbarten Neuronen weiterleitet [26].

Diese Erkenntnisse aus der Biologie erlauben es nun, ein mathematisches Modell für ein Neuron zu entwickeln. Dabei ist es nicht beabsichtigt, eine realistische Beschreibung der biologisch-chemischen Vorgänge im Gehirn zu erhalten, sondern ein System zu konstruieren, das Eigenschaften hervorbringt, die auch das menschliche Gehirn auszeichnen, wie z.B. Parallelität, Mustererkennung, Robustheit und Lernfähigkeit.

Im Bild des mathematischen Modell-Neurons (Abb. 4.2), werden die Dendriten symbolisch als Verbindungslinien zu den Eingangswerten  $x_k$  dargestellt, denen ein Gewichtswert  $\omega_{jk}$  zugeordnet wird. Dieser Gewichtswert repräsentiert je nach Vorzeichen eine hemmende oder erregende synaptische Verbindung zum Neuron  $N_j$ . Die Aktivität  $y_j$  des Neurons  $j$  kann als Funktion der  $n$  gewichteten Eingangswerte als

$$y_j = g \left( \sum_{k=0}^n \omega_{jk} \cdot x_k + \Theta_j \right) \quad (4.1)$$

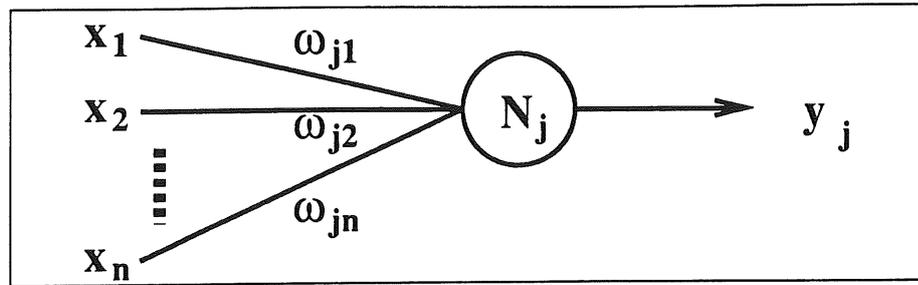


Abbildung 4.2: Das mathematische Modell-Neuron. Das Modell-Neuron funktioniert wie ein Summierverstärker, der Eingangswerte  $x_k$  mit Gewichten  $\omega_{jk}$  aufsummiert und nach einer Transformation durch eine Aktivierungsfunktion den Wert  $y_j$  weitergibt.

geschrieben werden. Dabei bezeichnet  $\Theta_j$  die Schwelle des Neurons und  $g(x)$  eine (meist nichtlineare) Funktion, die Aktivierungsfunktion genannt wird.

Im Modell werden i.a. nur beschränkte Funktionen zugelassen, deren Wertebereich zwischen  $y_j = -1 \dots +1$  oder  $y_j = 0 \dots +1$  liegen. Als Aktivierungsfunktion wird häufig eine sigmoide Funktion der Form

$$g(x) = \frac{1}{1 + e^{-\frac{x}{T}}} \quad \text{mit} \quad g(x) \in [0, 1] \quad (4.2)$$

gewählt. Die Wahl des Parameters  $T$  bestimmt die Steilheit der Kurve, für  $T \rightarrow 0$  ergibt sich eine Stufenfunktion (Heavyside-Funktion). Dies ist in Abbildung 4.3 dargestellt.

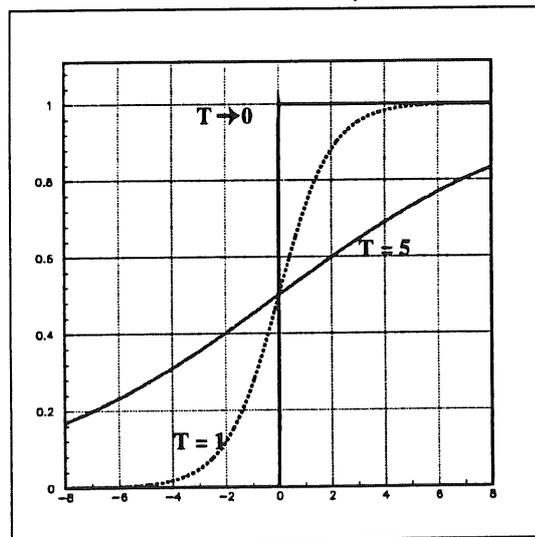


Abbildung 4.3: Sigmoide Aktivierungsfunktionen. Der Ausgabewert eines Neurons wird durch die Aktivierungsfunktion auf ein Intervall abgebildet. Die Steilheit dieser Funktion wird durch den Parameter  $T$  bestimmt.

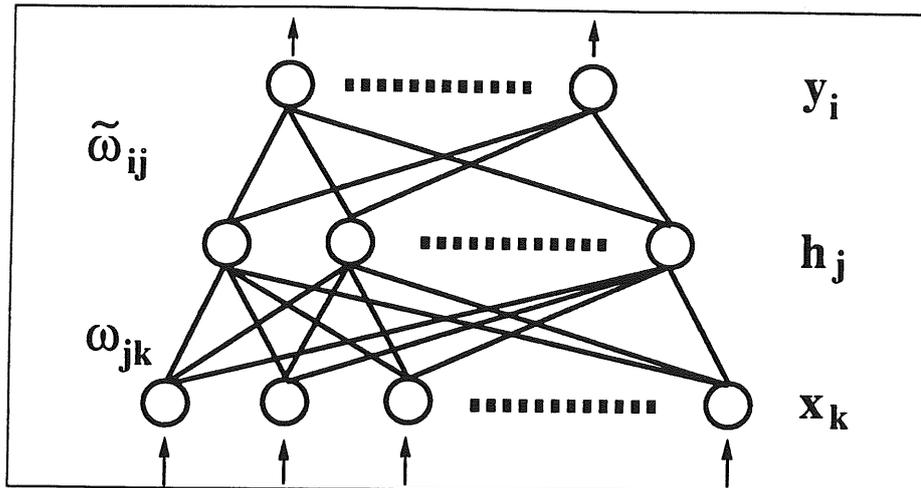


Abbildung 4.4: **Multilayer-Feedforward-Netzwerk.** Die Beschreibung eines dreischichtigen Netzes mit variabler Anzahl von Knoten kann durch die Variablen  $x_k$  für die Eingangswerte (“Input Layer”),  $h_j$  für die Aktivität der versteckten Neuronen (“Hidden Layer”) und  $y_i$  für die Ausgangswerte (“Output Layer”) erfolgen. Die Werte  $\omega_{jk}$  und  $\tilde{\omega}_{ij}$  beschreiben die Gewichte zwischen Eingabe- und versteckter Schicht bzw. zwischen versteckter und Ausgabeschicht.

#### 4.2.2 Das Multilayer-Feedforward-Netzwerk

Sowohl in der Biologie, als auch bei den Künstlichen Neuronalen Netzwerken bestimmt die Topologie, d.h. die Anordnung und Verknüpfungsstruktur der Neuronen, entscheidend die Funktion des Gesamtsystems. Im Rahmen dieser Arbeit wird nur auf das sogenannte *Multilayer-Feedforward-Netzwerk* eingegangen. Für eine umfassende Darstellung Neuronaler Netze wird auf Standardwerke verwiesen [27, 28, 29].

Die Anordnung von Neuronen in mehrere Schichten (*Layers*) nennt man *Multilayer-Feedforward-Netzwerk* oder *Multi-Layer-Perceptron* (MLP) (siehe Abb. 4.4). Dabei bestehen Verknüpfungen nur zwischen Neuronen aus verschiedenen benachbarten Schichten, und das Signal wird nur von der Eingabeschicht (*Input Layer*) in Richtung der Ausgabeschicht (*Output Layer*) weitergeleitet (*feedforward*). Die Schichten zwischen Ein- und Ausgabeschicht werden als verborgene Schichten (*Hidden Layers*), die Neuronen aufgrund ihrer Position in einem Netz häufiger auch als *Knoten* bezeichnet.

Das MLP eignet sich sehr gut zur Lösung von Klassifizierungsproblemen. Dabei verkörpern die Aktivitäten an den Neuronen in der Eingabeschicht die zur Verfügung stehenden Meßwerte (z.B. Triggersignale). Die Neuronen der Ausgabeschicht stellen eine beliebige Repräsentierung von Klassen dar. Im Falle eines Triggers könnte man zwei Ausgabeneuronen verwenden. Eines, das für die Zugehörigkeit zur Physikklasse den Wert 1 liefert, und ein anderes, das für die Untergrundklasse den Wert 0 annimmt. Da beide Klassen jedoch komplementär sind, genügt ein Ausgabeneuron zur Unterscheidung von Physik und Untergrund. In diesem Fall nimmt das Neuron z.B. den Wert 1 im Fall von Physikereignissen und den Wert 0 für Untergrundereignisse an.

Man kann die Gesamtheit der Neuronen abstrakt als “Black Box” betrachten, die eine

Abbildung  $F$  eines  $n$ -dimensionalen Vektors  $\vec{x}$  auf einen  $m$ -dimensionalen Vektor  $\vec{y}$  implementiert. Dabei beschreibt  $n$  die Anzahl der Neuronen in der Eingabeschicht und  $m$  die Anzahl der Neuronen in der Ausgabeschicht. Formal läßt sich dies durch

$$\vec{y} = F(\vec{x}) \quad (4.3)$$

mit

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{und} \quad \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \quad (4.4)$$

beschreiben. Die Abbildungsfunktion  $F$  hängt bei gegebener Aktivierungsfunktion  $g(x)$  nur von den Gewichten  $\omega_{jk}$  und den Schwellen  $\Theta_j$  ab. Die Bestimmung dieser Parameter ist die Aufgabe eines Optimierungsprozesses, der im Bereich der Künstlichen Neuronalen Netze *Lernen* genannt wird (siehe Abschnitt 4.2.3).

Um eine geschlossene kompakte Form zur Behandlung der Gewichte zu erhalten, ohne den Schwellen  $\Theta_j$  eine Sonderrolle zukommen zu lassen, kann man in jeder Schicht ein zusätzliches Neuron  $N_0$  mit konstanter Aktivität von 1 einführen, so daß  $\Theta_j$  durch  $\omega_{j0}$  beschrieben wird.

Damit ist eine genaue Beschreibung der Vorgänge innerhalb der "Black Box" möglich: Die Werte der Neuronen der Hidden-Layer  $h_j$  bzw. der Output-Layer  $y_i$  (zur Nomenklatur siehe Abb. 4.4) werden mit

$$h_j = g\left(\sum_{k=0}^n \omega_{jk} x_k\right) = g(\vec{\omega}_j \cdot \vec{x}) \quad (4.5)$$

$$y_i = g\left(\sum_{j=0}^l \tilde{\omega}_{ij} h_j\right) = g(\vec{\tilde{\omega}}_i \cdot \vec{h}) \quad (4.6)$$

berechnet, wobei  $l$  die Anzahl der Neuronen in der verborgenen Schicht beschreibt. Die gesamte Abbildung  $F$  läßt sich formal noch kompakter darstellen durch

$$F(\vec{x}) = g\left(\vec{W} \cdot \underbrace{[g(W \cdot \vec{x})]}_{\vec{h}}\right) \quad (4.7)$$

wenn man folgende Matrizen definiert

$$\vec{W} = \begin{pmatrix} \tilde{\omega}_{10} & \dots & \tilde{\omega}_{1l} \\ \vdots & \ddots & \vdots \\ \tilde{\omega}_{m0} & \dots & \tilde{\omega}_{ml} \end{pmatrix} \quad \text{und} \quad W = \begin{pmatrix} \omega_{10} & \dots & \omega_{1n} \\ \vdots & \ddots & \vdots \\ \omega_{l0} & \dots & \omega_{ln} \end{pmatrix}. \quad (4.8)$$

Sie beschreiben die Kopplungen der Synapsen zwischen Eingabeschicht, verborgener Schicht und Ausgabeschicht.

Hier sieht man deutlich den Einfluß der Aktivierungsfunktion und die Bedeutung ihrer Nicht-Linearität. Ohne die Aktivierungsfunktion wäre die versteckte Neuronenschicht sinnlos, denn die beiden Matrizen könnten durch eine einzige (ihr Produkt) ersetzt werden.

Die Nicht-Linearität ermöglicht das Lernen bei mehrschichtigen Netzwerken (siehe Abschnitt 4.2.4) und die Separierung von Klassen mit beliebigen Konturen (siehe Abschnitt 4.4).

### 4.2.3 Lern- und Ausführungsphase

Beim Einsatz von Neuronalen Netzen unterscheidet man zwischen zwei verschiedenen Zeitabschnitten, der *Lernphase* und der *Ausführungsphase*, die bei konventionellen Algorithmen der Programmierung und der Ausführung des Programms entsprechen.

Die Aufgabe in der Lernphase eines MLP's besteht darin, einen Satz von Gewichtsparametern zu finden, der das Klassifizierungsproblem optimal löst, d.h. eine Funktion  $F$  zu finden, die jedem Eingabevektor den gewünschten Ausgabevektor zuweist. Zu diesem Zweck werden dem neuronalen Netz wiederholt Eingabewerte (Lernmuster) präsentiert, zu denen der Ausgabevektor bekannt ist. Einem bestimmten Lernalgorithmus folgend werden dann die Parameter des Netzes so lange verändert (optimiert), bis der Ausgabevektor die gewünschten Werte approximiert. Diesen Vorgang nennt man *Lernen* oder *Training*.

Sind die Parameter einmal bestimmt, so kann das Netz im Ausführungsmodus betrieben werden, das heißt, bisher noch nicht gesehene unbekannte Muster werden an die Eingangsneuronen angelegt. Nach den im Abschnitt 4.2.2 beschriebenen Rechenregeln (Gleichung 4.7) produziert das Netz dann einen Ausgabevektor, der die richtige Klasse darstellen sollte.

Diese Fähigkeit, noch nie zuvor gesehene Eingabemuster richtig zu klassifizieren, wird *Generalisierung* genannt und ist stark vom Lernalgorithmus, der Repräsentativität der zum Lernen verwendeten Muster und von der Komplexität des Netzes abhängig.

### 4.2.4 Der Backpropagation-Lernalgorithmus

Im

Jahr 1986 stellten *McClelland* und *Rumelhart* den *Backpropagation-Lernalgorithmus* (BPG) vor, der in mehrlagigen Netzen zum erstenmal auch Gewichte in verborgenen Schichten verändern konnte [27]. Dies wurde dadurch erreicht, daß die Abweichungen der Ausgabevektoren von den angestrebten Werten, die Fehler, an die darunterliegenden Schichten weitergegeben wurden (Backpropagation). Essentiell war hierbei die Verwendung einer sigmoiden Aktivierungsfunktion.

*McClelland* und *Rumelhart* bewiesen auch, daß ihr Algorithmus ein Verfahren darstellt, mit dem das nichtlineare Optimierungsproblem zur Auffindung der Gewichte gelöst werden kann. Der Algorithmus entspricht einem Gradientenverfahren zur Lösung nichtlinearer Gleichungssysteme, bei dem die freien Parameter des Systems proportional zum Gradienten der Fehlerhyperfläche verändert werden.

Zu Beginn des Lernvorgangs werden die Gewichte  $\omega_{ij}$  und  $\omega_{jk}$  des Netzes mit zufälligen Werten aus einem Intervall  $[-a, a]$  initialisiert. Dies ist nötig, da der BPG-Algorithmus nicht in der Lage ist, Symmetrien in den Gewichten zu brechen. Danach werden Paare  $(\vec{x}^{(p)}, \vec{t}^{(p)})$  aus einem Satz zu lernender Muster (Trainingsmenge) an Eingangs- bzw. Ausgangsneuronen angelegt und die Gewichte so verändert, daß die globale Fehlerfunktion

$$E = \frac{1}{2} \sum_{p=1}^M \sum_{i=1}^m (y_i^{(p)} - t_i^{(p)})^2 \quad (4.9)$$

minimiert wird. Sie stellt die mittlere quadratische Abweichung vom gewünschten Ergebnis dar. Dabei bezeichnet  $y_i$  den Wert des  $i$ -ten Ausgabeneurons,  $t_i$  den gewünschten Wert und  $p$  die Nummer des Musters aus der Trainingsmenge, die  $M$  Paare enthält.

Diese von *Rumelhart* und *McClelland* vorgeschlagene Fehlerfunktion ist zwar in der Physik sehr gebräuchlich, aber im Fall von Neuronalen Netzen nicht unbedingt optimal, wenn man die Geschwindigkeit der Konvergenz beim Lernprozeß betrachtet [30]. Im folgenden wird das sogenannte *Online-Updating* dargestellt, bei dem nach jedem Muster eine Veränderung der Gewichte erfolgt, d.h. in Formel 4.9 wird nicht über  $p$  summiert. Ein Iterationsschritt des Algorithmus wird als *Lernzyklus* bezeichnet. Er ist dadurch gekennzeichnet, daß zuerst ein Eingabevektor an die Neuronen der Eingabeschicht angelegt und die Aktivitäten  $h_j$  und  $y_i$  in der Hidden- bzw. Output-Layer gemäß den Gleichungen 4.5 und 4.6 berechnet werden. Dann erfolgt die Veränderung der Gewichte in entgegengesetzter Richtung. Zuerst werden die Kopplungen  $\tilde{\omega}_{ij}$  zwischen Output- und Hidden-Layer proportional zum Fehlergradienten entsprechend

$$\Delta\tilde{\omega}_{ij} = -\eta \frac{\partial E}{\partial \tilde{\omega}_{ij}} + \alpha \Delta\tilde{\omega}_{ij}^{alt} = -\eta \delta_i h_j + \alpha \Delta\tilde{\omega}_{ij}^{alt} \quad (4.10)$$

verändert, wobei

$$\delta_i = (y_i - t_i) \cdot g'(\tilde{\omega}_i \cdot \vec{h}). \quad (4.11)$$

Daraufhin folgen die Gewichte  $\omega_{jk}$  zwischen Hidden- und Input-Layer, die analog nach

$$\Delta\omega_{jk} = -\eta \frac{\partial E}{\partial \omega_{jk}} + \alpha \Delta\omega_{jk}^{alt} = -\eta \delta_j x_k + \alpha \Delta\omega_{jk}^{alt} \quad (4.12)$$

mit

$$\delta_j = \sum_{i=1}^m \omega_{ij} \delta_i g'(\tilde{\omega}_j \cdot \vec{x}) \quad (4.13)$$

behandelt werden.  $g'$  beschreibt hierbei die erste Ableitung der Aktivierungsfunktion. Die Proportionalitätskonstante  $\eta$  wird als *Lernparameter* bezeichnet und muß empirisch dem Problem angepaßt werden. Sie definiert gewissermaßen die Schrittweite, die beim Suchprozeß im Parameterraum der Gewichte verwendet wird.

Die Größen  $\Delta\omega_{ij}^{alt}$  und  $\Delta\omega_{jk}^{alt}$  bezeichnen die im vorangegangenen Iterationschritt berechneten Änderungen und stellen zusammen mit dem *Momentumfaktor*  $\alpha$  bereits eine Variation des ursprünglichen BPG-Algorithmus dar. Sie dienen dazu, die Lernrate in Parameterregionen mit "flachem" Gradienten zu vergrößern oder in Bereichen mit "steilem" Gradienten mögliche Oszillationen zu dämpfen und die Bewegung des Suchprozesses in die richtige Richtung zu lenken. Der Wert für  $\alpha$  muß ebenfalls problem-spezifisch ermittelt werden, sinnvolle Werte liegen aber nur im Bereich  $\alpha = 0 \dots 1.0$  [26].

Die rekursive Berechnung des Fehlers  $\delta_j$  in der verborgenen Schicht – aus den  $\delta_i$ 's in der Ausgabeschicht – macht das Prinzip der Fehlerrückverfolgung (Error-Backpropagation) deutlich. Sie zeigt auch die Flexibilität bei der Erweiterung des Netzes zu mehr als drei Schichten.

Bei der Auswahl der Lernmuster aus der Trainingsmenge ist darauf zu achten, daß

die Muster dem Netz in zufälliger Reihenfolge präsentiert werden und die Muster aus verschiedenen Klassen mit gleicher Wahrscheinlichkeit gelernt werden. Wird dies nicht beachtet, so kann das Ergebnis des Optimierungsprozesses ein Nebenminimum liefern (siehe auch Abschnitt 4.4).

### 4.3 Geometrische Interpretation der Netzwerkleistung

Die mathematische Beschreibung der Prozesse in einem Neuronalen Netz eröffnet auch eine geometrische Interpretation des Klassifizierungsverfahrens. Der direkteste Weg besteht darin, die Arbeitsweise eines einzelnen Neurons zu betrachten. Die Gleichung 4.1 beschreibt zusammen mit der sigmoiden Funktion aus Gleichung 4.2 das Verhalten eines Neurons. Setzt man das Argument der Aktivierungsfunktion zu Null

$$\sum_{j=0}^n \omega_{jk} \cdot x_k + \Theta_j = 0 \quad \text{oder} \quad \vec{\omega}_j \cdot \vec{x} + \Theta_j = 0 \quad (4.14)$$

so erkennt man sofort die Gleichung einer Hyperebene im  $n$ -dimensionalen Raum der Eingangsvariablen  $x_k$ . Betrachtet man zuerst den Grenzfall einer Stufenfunktion, so führt das Neuron einen Schnitt durch, bei dem es für Muster (Punkte im Hyperraum), die auf der einen Seite der Ebene liegen, eine Aktivität von Eins und für Muster auf der anderen Seite eine Aktivität von Null am Ausgangsneuron produziert.

Die Ebene wird durch einen Satz von Gewichtswerten festgelegt, die die Komponenten eines Vektors darstellen, der zum Normalenvektor der Ebene parallel ist. Wenn die Ebenengleichung normiert wird, beschreibt die Schwelle  $\Theta_j$  den Abstand der Ebene vom Koordinatenursprung (siehe Abb. 4.5).

Ändert man die Stufenfunktion in eine sigmoide Aktivierungsfunktion, so bildet sich in der Umgebung der Trennebene eine Übergangszone. Für weit von der Ebene entfernte Punkte ist die Ausgabe noch immer nahe bei Eins oder nahe bei Null. Nähert man sich jedoch in senkrechter Richtung der Ebene, dann folgt der Wert des Ausgabeneurons dem Verlauf der Sigmoidfunktion.

Durch Normierung der Ebenengleichung kann man eine weitere Information aus dem Gewichtsvektor  $\vec{\omega}$  ablesen :

$$y = g(\vec{\omega} \cdot \vec{x} + \Theta) \quad (4.15)$$

$$= g\left(|\vec{\omega}| \cdot \left[\frac{\vec{\omega}}{|\vec{\omega}|} \cdot \vec{x} + \frac{\Theta}{|\vec{\omega}|}\right]\right) \quad (4.16)$$

$$= \frac{1}{1 + e^{-\frac{|\vec{\omega}|}{T} \cdot \left[\frac{\vec{\omega}}{|\vec{\omega}|} \cdot \vec{x} + \frac{\Theta}{|\vec{\omega}|}\right]}} \quad (4.17)$$

Die inverse Länge des Gewichtsvektors kann als Multiplikator für den Parameter  $T$  aufgefaßt werden: Ist der Gewichtsvektor lang, so beschreibt er eine schmale Übergangszone, ist er kurz, beschreibt er eine breite Zone.

Mit dieser Interpretation eines einzelnen Neurons ist es möglich, die Arbeit eines MLP's

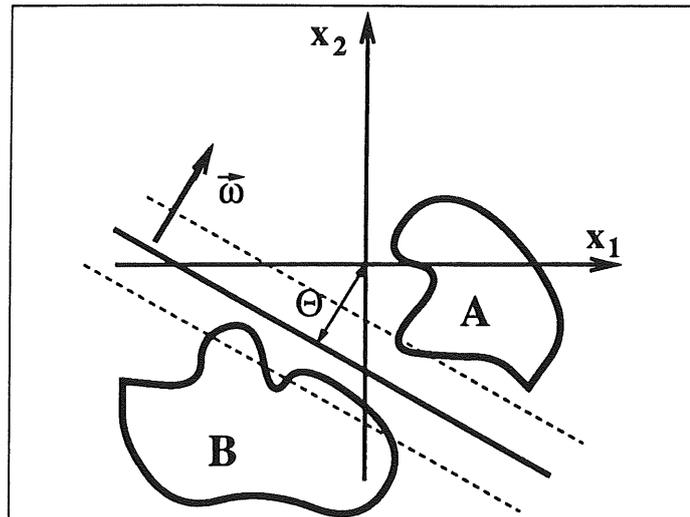


Abbildung 4.5: Trennung zweier Klassen A und B. Ein Neuron definiert mit seinem Gewichtsvektor  $\vec{w}$  und der Schwelle  $\Theta$  eine Hyperfläche im Raum der Eingangsvektoren. Die gestrichelten Linien geben den Bereich der Übergangszone im Fall einer sigmoiden Aktivierungsfunktion an.

mit einem Ausgabeneuron zu verstehen. Die Neuronen in der verborgenen Schicht arbeiten als *weiche* Schwellen in einem hochdimensionalen Raum. Das Ausgabeneuron kombiniert die Entscheidungen der tieferliegenden Neuronen, ähnlich dem Verfahren einer Diskriminanzanalyse. In geometrischer Sprache ausgedrückt, formt es Volumina aus den verschiedenen Ebenen und trennt damit die zwei Klassen, in unserem Fall Physik von Untergrund.

Obwohl es in den einzelnen Neuronen sanfte Übergänge gibt, existiert bezüglich der endgültigen Entscheidung keinerlei *Unschärfe*, da am Ende ein definitiver Schnitt beim Wert des Ausgabeneurons angebracht wird. Dieser Schnitt legt die wirkliche Kontur fest (siehe Abb. 4.6).

## 4.4 Simulationsbeispiele in zwei Dimensionen

Um die geometrischen Interpretationen des vorigen Abschnitts zu verifizieren und einen Einblick in den Prozeß des Lernens zu erhalten, wurden Simulationen in zwei Dimensionen durchgeführt. Als Netzwerktopologien wurden MLP's gewählt, die zwei Eingangsneuronen, ein Ausgabeneuron und eine je nach Problem verschiedene Anzahl von verborgenen Neuronen besaßen.

Das Training wurde so durchgeführt, daß nicht ein fester Satz von Lernmustern zur Verfügung stand, sondern zufällig Punkte aus dem zweidimensionalen Raum "gewürfelt" wurden, deren Zugehörigkeit zu einer Klasse vom Programm analytisch bestimmt wurde. Die Klassen bestanden aus geometrischen Figuren, deren Grenzen das Neuronale Netz lernen sollte. Punkte innerhalb der Figuren sollten als "Physik" mit einer Eins, Punkte außerhalb als "Untergrund" mit einem Ausgabewert von Null assoziiert werden.

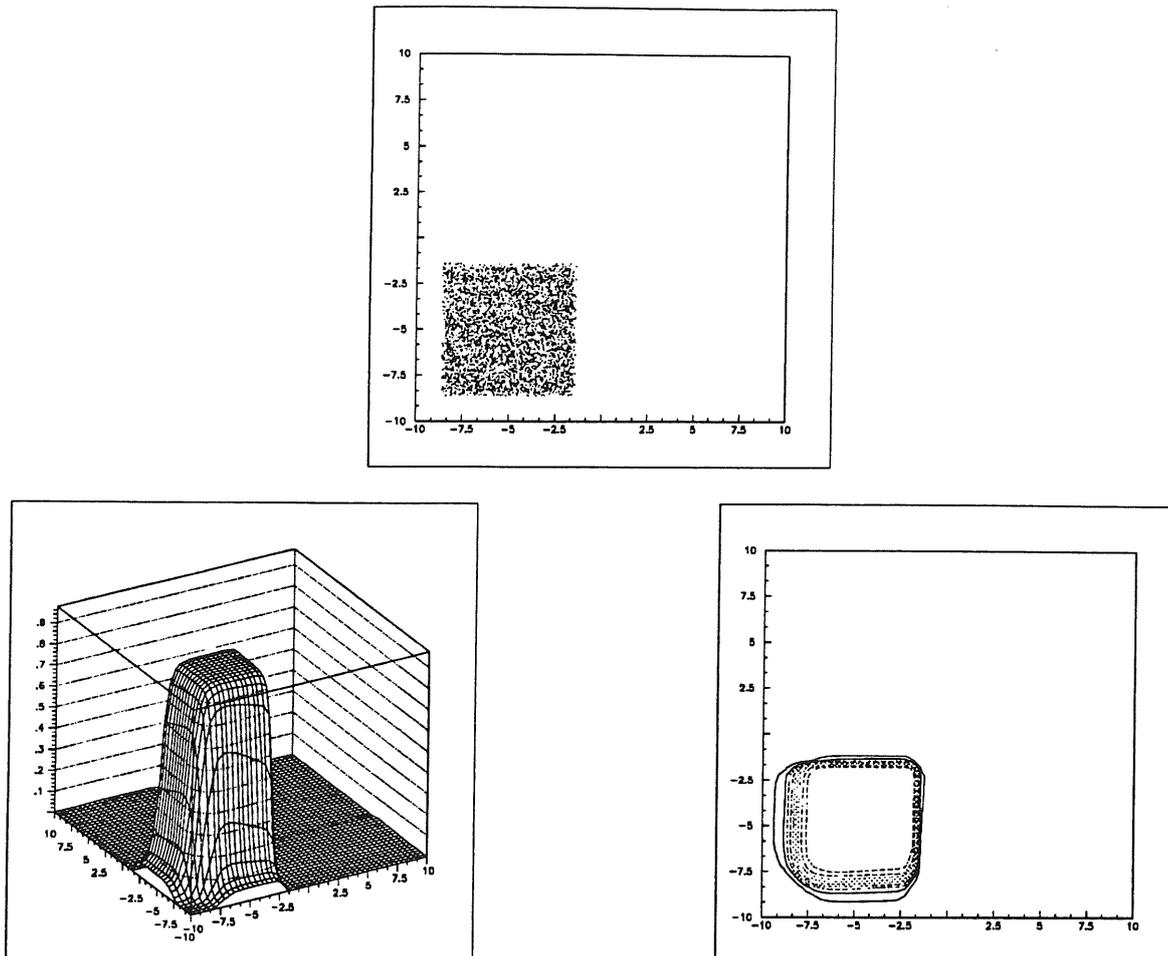


Abbildung 4.6: Mustererkennung durch ein Multilayer-Feedforward-Netz. Die Erkennung eines Quadrats mit zwei Eingabeneuronen, vier verborgenen Neuronen und einem Ausgabeneuron. In der oberen Hälfte ist das Lernmuster dargestellt, unten links die Aktivität des Ausgabeneurons und unten rechts ein *Contour-Plot* (Höhenlinien) des linken Bildes.

Im ersten Beispiel wurden dem Netzwerk vier verborgene Knoten zur Klassifizierung zur Verfügung gestellt, die nötig waren um das zu erkennende Quadrat abzugrenzen. In den folgenden Abbildungen ist das zu lernende Muster jeweils im oberen Teil dargestellt. Im unteren Teil sieht man eine dreidimensionale Darstellung der Aktivität  $y_j$  des Ausgabeneurons, die gegen den Raum der Eingabewerte in der  $x - y$ -Ebene aufgetragen ist. Um eine Klassifizierungsgrenze zu erhalten, ist es erforderlich, einen Schnitt im Ausgabewert, also der  $z$ -Koordinate, vorzunehmen. Dies wird im rechten unteren Teil als *Contour-Plot* dargestellt, in dem jede Linie einem unterschiedlichen Schnitt in  $z$  entspricht (Höhenlinien).

Im Laufe der Simulation wurde deutlich, daß mehr als 5 Millionen Muster und damit Iterationsschritte benötigt werden, um eine gute Lösung mit scharfen Grenzen zu erhalten (siehe Abb. 4.6). Dies war überraschend in Anbetracht dieses relativ einfachen Problems. Es stellte sich heraus, daß das Netz zwar schnell die richtigen Ebenen findet,

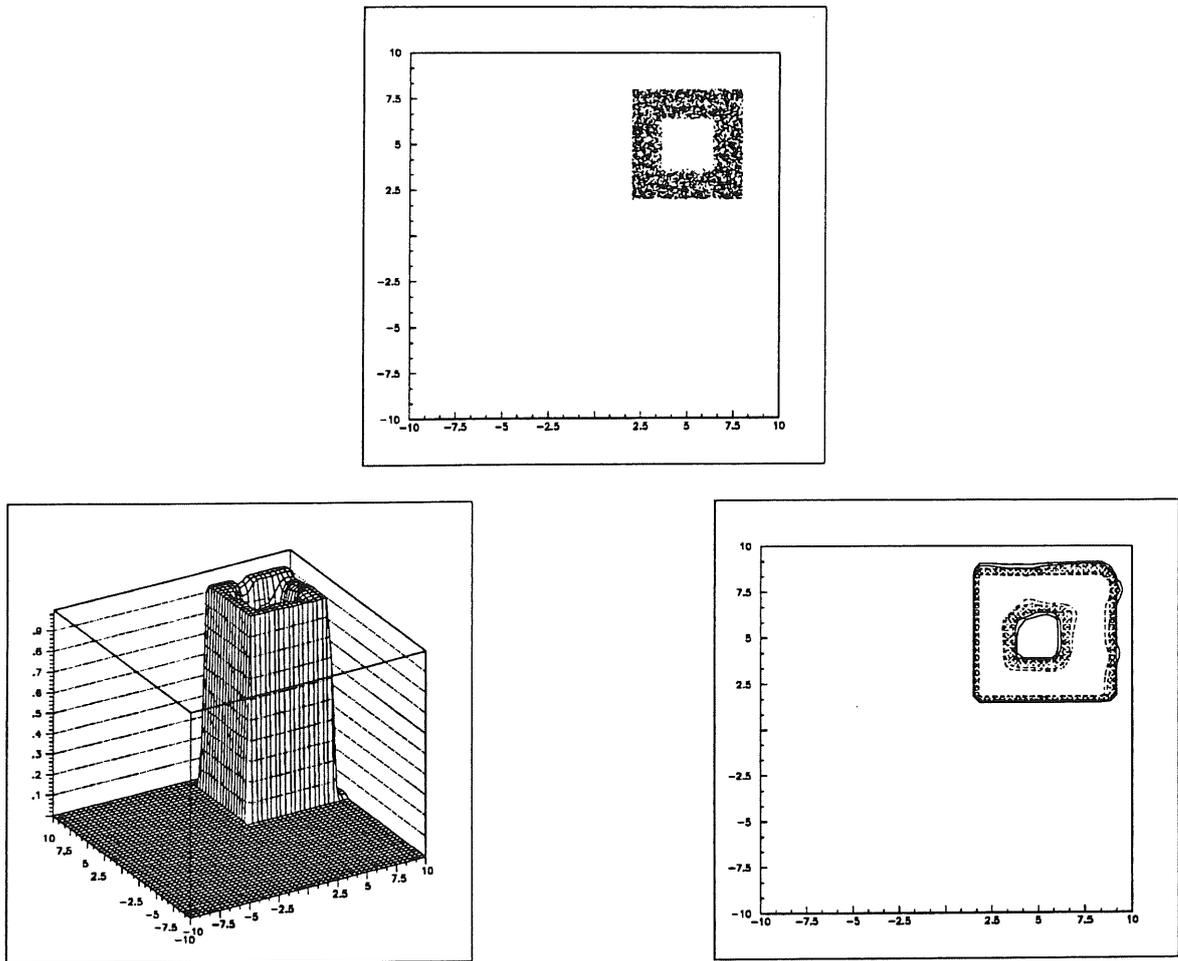


Abbildung 4.7: Erkennung eines konkaven ringförmigen Musters. Die Beschreibung eines quadratischen Lochs in einem Viereck ist auch mit nur einer verborgenen Schicht möglich. Es wurden nur die erforderlichen 8 Neuronen in der verborgenen Schicht benötigt, um die Trennebenen korrekt zu beschreiben.

es aber lange dauert, bis die Breiten der Übergangszonen (die Länge der Gewichtsvektoren), die weit vom Koordinatenursprung entfernt liegen, richtig adaptiert sind. Dies liegt darin begründet, daß die Sigmoidfunktion punktsymmetrisch um den Ursprung ist, das Beispiel aber einen Asymmetrie im  $x - y$ -Raum aufweist und deswegen das Lernen für weiter außen liegende Punkte langsamer verläuft. In vielen Lehrbüchern und Publikationen wird darauf hingewiesen, daß zur Klassifizierung von konkaven oder ringförmigen Mustern eine einzige versteckte Schicht nicht ausreicht. Dies widerspricht der geometrischen Interpretation des vorangegangenen Kapitels und wurde deshalb mit Hilfe einer weiteren Simulation überprüft. In diesem Beispiel hatte das Netz 8 Neuronen in der Hidden-Layer zur Verfügung um ein quadratisches Loch in einem Quadrat zu beschreiben. Die Darstellungen aus Abbildung 4.7 belegen deutlich die Richtigkeit der geometrischen Überlegungen.

In einem letzten Beispiel soll demonstriert werden, daß ein Neuronales Netz beliebige Umrisse approximieren kann, wie sie in der Realität auch weit häufiger vorkommen.

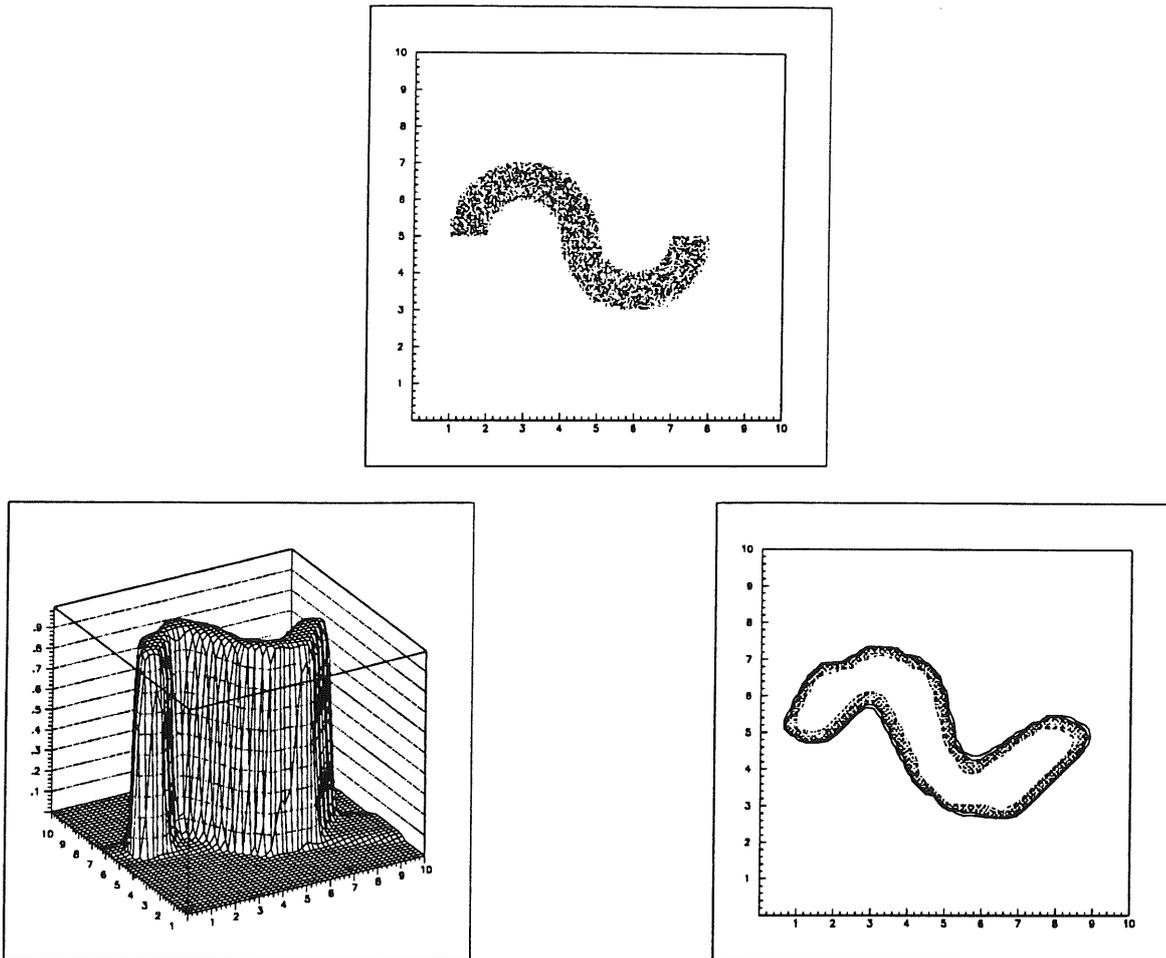


Abbildung 4.8: **Approximation von Mustern mit gekrümmten Konturen.** Das schlangenförmige Muster (oben) wurde mit 10 Neuronen in der versteckten Schicht approximiert. Dies zeigt die Fähigkeit von Neuronalen Netzen auch komplexe Muster mit Kombinationen von Trennebenen richtig zu beschreiben.

Das Muster bestand aus einer schlangenförmigen Figur, *Boa* genannt (siehe Abb. 4.8). Im *Boa*-Beispiel war es nicht offensichtlich, wieviele versteckte Neuronen gewählt werden müssen, um eine gute Lösung zu erhalten. In der Simulation wurden willkürlich 10 Neuronen verwendet. Die Abbildung zeigt die Fähigkeit des Netzes durch geeignete Kombination der Hyperflächen auch kompliziertere nicht-ebene Grenzen zu interpolieren. Die exakte Kontur der "Boa" wird durch den Schnitt am Ausgabeneuron definiert (siehe Abschnitt 4.3). Durch die künstliche Absenkung des Parameters  $T \rightarrow 0$  ist es möglich, die Trennebenen sichtbar zu machen; sie sind in Abbildung 4.9 zu sehen. Mit Ausnahme des *Boa*-Beispiels wurden nur Netztopologien mit minimaler Knotenzahl verwendet, d.h. die Anzahl der versteckten Knoten entsprach der Zahl der erforderlichen Trennebenen. In einigen Trainingsläufen für das Quadrat-Beispiel endete die Suche nach dem optimalen Satz von Gewichten in einem lokalen Minimum, das einem um  $45^\circ$  gedrehten Quadrat entsprach. Daraufhin wurde die Zahl der verborgenen Neuronen verdoppelt und eine schnellere und bessere Konvergenz beim Lernen

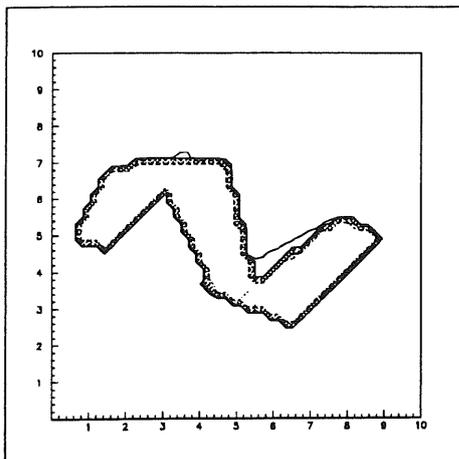


Abbildung 4.9: **Hervorhebung der Trennebenen des vorigen Beispiels.** Durch künstliches Verändern des Parameters  $T$  auf Null, werden die zehn Ebenen sichtbar, mit denen die "Boa" approximiert wird.

festgestellt. Es stellt sich die Frage, welche Ebenen die zusätzlichen Knoten repräsentieren. Diese Knoten besaßen Duplikate der Gewichtsvektoren anderer Neuronen und beschrieben somit die selben Ebenen. Das bedeutet, daß nur ein geringes Risiko für ein "over-fitting" der Daten besteht, solange die Menge der Muster groß ist gegenüber der Zahl der Gewichte im Netz.

Obwohl diese Simulationen nur in zwei Dimensionen stattfanden, ist zu erwarten, daß die gezogenen Schlüsse auch in höherdimensionalen Räumen zutreffen, mit einer Ausnahme: Sie sind nicht mehr so leicht anschaulich zu machen.

## 4.5 Theoretischer Ausblick

Im Rahmen dieser Arbeit ist es nicht möglich, die theoretischen Aspekte Neuronaler Netze eingehend zu diskutieren. Trotzdem sollen hier einige der neueren Ergebnisse [31, 32, 33, 34] dargestellt werden, die eine bislang fehlende theoretische Begründung für den Einsatz von Multilayer-Feedforward-Netzwerken zur Verfügung stellen.

In den vorausgegangenen Kapiteln war bereits zu sehen, daß ein MLP eine Abbildung  $F_{g,W}$  ( $g \hat{=}$  Aktivierungsfunktion,  $W \hat{=}$  Gewichte) vom Raum der Eingabevektoren in den Raum der Ausgabevektoren realisiert. Folgende Fragen sollen in den nächsten Abschnitten behandelt werden:

- Ist ein MLP in der Lage, das Gewünschte zu berechnen? Wie groß ist die Klasse von Funktionen  $F_{g,W}$ ?
- Was genau berechnet ein MLP?
- Ist ein MLP in der Lage zu lernen?
- Wie verhält sich ein MLP bezüglich der Generalisierung?

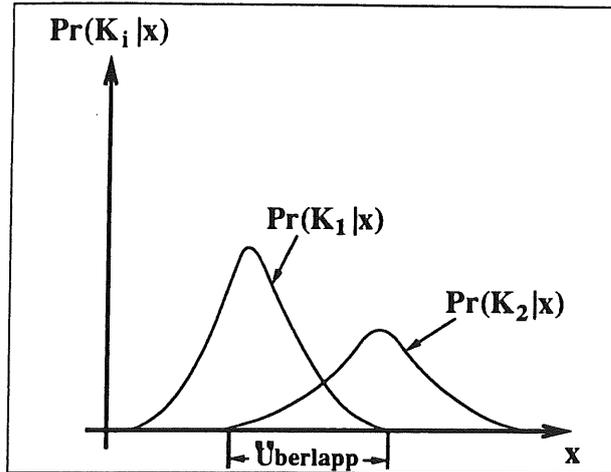


Abbildung 4.10: **Der Bayes'sche Limit.** Am Beispiel der Wahrscheinlichkeitsverteilungen für zwei Klassen  $K_1$  und  $K_2$  wird deutlich, daß im Überlappbereich nur eine statistische Zuordnung eines Vektors  $x$  mit einer bestimmten Fehlerwahrscheinlichkeit durchgeführt werden kann.

Die mathematischen Details werden soweit wie möglich vermieden, der interessierte Leser wird auf die Literaturstellen verwiesen.

#### 4.5.1 Der Bayes'sche Klassifizierer

Ein Neuronales Netz ist nicht immer in der Lage, jedes Muster richtig zu klassifizieren, denn es gibt eine theoretische Grenze, den *Bayes'schen Limit* [31]. Im Beispiel von Abb. 4.10 sieht man, daß ein Muster  $x$  aus dem Überlappbereich zu beiden Klassen gehören und nur mit Hilfe einer statistischen Aussage einer Klasse zugewiesen werden kann.

Die bezüglich einer Fehlklassifizierung optimale Entscheidung für solche Muster wird durch den Bayes'schen Klassifizierer gegeben. Im Fall von  $n$  möglichen Klassen  $K_i$  lautet dieser

$$Pr(K_i|x) = \frac{Pr(K_i) Pr(x|K_i)}{\sum_{i=1}^n Pr(K_i) Pr(x|K_i)} \quad (4.18)$$

Hierbei bedeuten

- $Pr(K_i|x)$  die *a posteriori* Wahrscheinlichkeit der Klassen nach Beobachtung von  $x$ , d.h. die Wahrscheinlichkeit, daß der gemessene Vektor  $x$  der Klasse  $K_i$  angehört hat,
- $Pr(K_i)$  die *a priori* Wahrscheinlichkeitsverteilung der Klassen,
- $Pr(x|K_i)$  die Wahrscheinlichkeitsverteilung der Vektoren  $x$  in jeder Klasse  $K_i$ .

Der Bayes'sche Klassifizierer weist also ein Muster  $x$  derjenigen Klasse  $K_i$  mit der höchsten *a posteriori* Wahrscheinlichkeit zu. Das Problem besteht nun darin, die verschiedenen Wahrscheinlichkeitsverteilungen zu berechnen, denn dazu müssen häufig

Annahmen über die Verteilung der Muster oder der Klassen gemacht werden.

Es ist von *Lippmann* und *Richard* [32] gezeigt worden, daß die Abbildung  $F_{g,W^*}$  die beste Approximation von  $Pr(K_i|x)$  bezüglich der Quadratnorm ist ( $\|F_{g,W^*}(x) - Pr(K_i|x)\|$ ).  $W^*$  stellt die Gewichte dar, die ein Neuronales Netz durch Minimierung der Fehlerquadrate am Ausgabeneuron berechnet ( $\|F_{g,W^*}(x) - t(x)\|$ ,  $t(x) \in (0, 1)$ ). Dies bedeutet, daß ein MLP eine Approximation des optimalen Bayes'schen Klassifizierers bestimmt, ohne *a priori* Annahmen über Verteilungen zu verwenden.

### 4.5.2 Realisierbarkeit von Abbildungen

Es sei  $F$  eine Abbildung, die nur durch Lernmuster  $(\vec{x}^{(p)}, \vec{t}^{(p)})$  definiert ist, d.h.  $\vec{t}^{(p)} = F(\vec{x}^{(p)})$ . Die Funktion  $F$  soll durch eine Netzwerkabbildung  $F_{g,W}$  approximiert werden. Es wurde von *Hornik et. al.* [33] gezeigt, daß MLP's mit nur einer Hidden-Layer und einer glatten Aktivierungsfunktion in der Lage sind, beliebig exakte Approximationen an meßbare Funktionen und ihre Ableitungen zu erreichen. Dies bedeutet auch, daß Netzwerke auch Funktionen approximieren können, die nicht differenzierbar sind (z.B. stückweise differenzierbare Funktionen).

### 4.5.3 Lernbarkeit

Ein weiterer wichtiger Punkt ist die Frage nach der Lernbarkeit, d.h. ob Methoden existieren, die es dem Netz ermöglichen, die richtigen Gewichte für die erforderliche Approximation aus der empirischen Beobachtung von Lernmustern zu gewinnen.

Es konnte von *White* [34] bewiesen werden, daß es Lernalgorithmen gibt, die die Wahrscheinlichkeit dafür, daß der Approximationsfehler des Netzwerkes eine beliebige Grenze überschreitet, gegen Null gehen lassen, wenn die Erfahrung des Netzwerkes, d.h. die Größe der Trainingsmenge, anwächst.

Das Hauptproblem ist in diesem Fall die Kontrolle der Netzwerk-Komplexität als Funktion der Erfahrung des Netzes. Beliebige exakte Approximationen an beliebige Funktionen benötigen beliebig komplexe Netzwerke. In der Realität sind aber nur Neuronale Netze endlicher Größe möglich, außerdem lernen zu komplexe Netze zuviel, d.h. sie lernen die Muster auswendig und liefern dadurch schlechte Ergebnisse bei der Generalisierung. Es ist nun möglich bei bekannter Anzahl von Trainingsbeispielen die Knotenzahl so zu bestimmen, daß die Gefahr von "over-fitting" vermieden wird.

# Kapitel 5

## Simulation Neuronaler Netze mit Monte-Carlo-Daten

### 5.1 Monte-Carlo-Daten

Die ersten Studien zur Einsatzfähigkeit Neuronaler Netze als L2-Trigger wurden mit Daten aus Monte-Carlo-Simulationen durchgeführt [35]. Dazu wurden Ereignisse verschiedener Physikklassen generiert, mit Hilfe des Programms H1SIM 1.07 [36] das Ansprechverhalten des H1-Detektors simuliert und die Triggergrößen rekonstruiert. Die unterschiedlichen Physikreaktionen und die entsprechende Anzahl von Ereignissen sind in Tabelle 5.1 zusammengefaßt.

Physikklasse	Bemerkungen	Statistik
Neutrale Ströme (NC1)	$Q^2 > 300 \text{ GeV}^2$	1853
Neutrale Ströme (NC2)	$Q^2 < 300 \text{ GeV}^2, x > 0.037$	1007
Neutrale Ströme (NC3)	$Q^2 < 300 \text{ GeV}^2, x < 0.037, y < 0.3$	1570
Neutrale Ströme (NC4)	$Q^2 < 300 \text{ GeV}^2, x < 0.037, y > 0.3$	2212
$c\bar{c}$ Produktion		3390
$b\bar{b}$ Produktion		785
Geladene Ströme (CC)		4975
Untergrundklassen		
Beam-Gas	3 Hz/Ereignis	51100
Beam-Wall	10 Hz/Ereignis	13389

Tabelle 5.1: Anzahl simulierter Ereignisse in den einzelnen Klassen.

In den Untersuchungen wird ein Vergleich zwischen einem reinen L1-Trigger und einer Kombination aus L1- und L2-Trigger durchgeführt. Das Ziel war, eine maximale Eingangsrate von 100 Hz zur dritten Triggerstufe zu erreichen. Dazu mußten sowohl die

L1-Triggerbedingungen als auch die Schwellenwerte zu verschiedenen Ausgangsraten bestimmt werden.

Sowohl für die Optimierung von L1-Triggerbedingungen, als auch für das Training von Neuronalen Netzen ist ein gutes Verständnis des Untergrunds erforderlich. Deshalb wurden auch detaillierte Monte-Carlo-Simulationen der wichtigsten Untergrundreaktionen (siehe Kap. 2.2.4) durchgeführt. Die Anzahl der generierten Ereignisse für Strahl-Gas- und Strahl-Wand-Reaktionen ist ebenfalls der Tab. 5.1 zu entnehmen. Für die Untergrundereignisse war es wichtig, die einem Ereignis entsprechende Rate zu kennen, um die Ausgangsraten zu kontrollieren. Ein Beam-Gas-Ereignis entspricht 3 Hz, ein Beam-Wall-Ereignis 10 Hz bei voller Luminosität des Beschleunigers. Da nur wenige Ereignisse den L1-Trigger passieren, bedeutet dies eine starke Beschränkung bezüglich des statistischen Fehlers.

## 5.2 Der Standard-L1-Trigger

1) $E_T > E_T(3) \text{ GeV}$
2) $E_T > E_T(2) \text{ GeV}$ and $\overline{TOF}$
3) $E_T > E_T(1) \text{ GeV}$ and $\overline{TOF}$ and $DC$
4) $E_{T_{miss}} > E_{T_{miss}}(3) \text{ GeV}$
5) $E_{T_{miss}} > E_{T_{miss}}(2) \text{ GeV}$ and $\overline{TOF}$
6) $E_{T_{miss}} > E_{T_{miss}}(2) \text{ GeV}$ and $DC$
7) $E_{T_{miss}} > E_{T_{miss}}(1) \text{ GeV}$ and $\overline{TOF}$ and $DC$
8) $E_{IF} < E_{IF}(1) \text{ GeV}$ and $E_{back} > E_{back}(3) \text{ GeV}$ and $TOF$
9) $E_{forw} < E_{forw}(1) \text{ GeV}$ and $E_{back} > E_{back}(3) \text{ GeV}$ and $\overline{TOF}$
10) $E_{barrel} > E_{barrel}(3) \text{ GeV}$
11) $E_{barrel} > E_{barrel}(2) \text{ GeV}$ and $\overline{TOF}$
12) $E_{barrel} > E_{barrel}(1) \text{ GeV}$ and $\overline{TOF}$ and $DC$
13) $E_{BEMC} > E_{BEMC}(1) \text{ GeV}$ and $DC$ and $\overline{TOF}$
14) $ELECT$ and $\overline{TOF}$
15) $CLUST$
16) $DC$ and $ZVTX$ and $\overline{TOF}$

Tabelle 5.2: Der L1-Trigger-Cocktail.

Die Monte-Carlo-Daten wurden benutzt, um den L1-Trigger zu optimieren, d.h. eine Anzahl von Bedingungen und Schwellenwerten (L1-Trigger-Cocktail) festzulegen, bei denen eine möglichst hohe Effizienz in der Erkennung von Physikereignissen besteht und gleichzeitig eine minimale Anzahl von Untergrundereignissen akzeptiert wird. Der

$E_T$	Transversale Energie (8 Bit)
$E_{T_{miss}}$	Fehlende transversale Energie (8 Bit)
$E_{IF}$	Energie im <i>Inner- und Outer Forward</i> -Bereich (8 Bit)
$E_{FB}$	Energie im <i>Forward Barrel</i> -Bereich (8 Bit)
$E_{CB}$	Energie im <i>Central Barrel</i> -Bereich (8 Bit)
$E_{BEMC}$	Energie im BEMC (8 Bit)
$E_{Plug}$	Energie im Plug (8 Bit)
$ELECT$	Identifiziertes Elektron (1 Bit)
$ZVTX$	Vertex Position entlang des Strahls (1 Bit)
$TOF$	<i>Time-of-flight</i> Veto (1 Bit)
$ZCH1$	Maximum im Z-Vertex Histogramm (1 Bit)
$ZCH2$	Signifikantes Maximum im Z-Vertex Histogramm (1 Bit)
$CLUST$	BEMC Cluster > 10 GeV (1 Bit)
$BM_{mul}$	Multiplizität in der BPC (8 Bit)
$DC_{mul}$	Anzahl der Spuren in den Jetkammern (8 Bit)
$DC_{lp}$	Spuren mit positiver Krümmung und einer Energie kleiner als 1 GeV (8 Bit)
$DC_{hp}$	Spuren mit positiver Krümmung und einer Energie größer als 1 GeV (8 Bit)
$DC_{ln}$	Spuren mit negativer Krümmung und einer Energie kleiner als 1 GeV (8 Bit)
$DC_{hn}$	Spuren mit negativer Krümmung und einer Energie größer als 1 GeV (8 Bit)
$E_{barrel}$	$E_{FB} + E_{CB}$
$E_{forward}$	$E_{PLUG} + E_{IF} + E_{FB}$
$E_{backward}$	$E_{CB} + E_{BEMC}$
$E_{tot}$	$E_{PLUG} + E_{IF} + E_{FB} + E_{CB} + E_{BEMC}$
$DC$	Spur im Driftkammertrigger (1 Bit)

Tabelle 5.3: **Bedeutung der Triggergrößen.** Die oberen 19 Variablen werden zum Trainieren der Neuronalen Netze verwendet, die unteren fünf Variablen bilden redundante Kombinationen der obigen Größen für den L1-Trigger.

L1-Trigger-Cocktail ist in Tabelle 5.2 dargestellt [37]. Die verwendeten Triggersignale sind in Tabelle 5.3 beschrieben.

Die auf Physikeffizienz optimierten Schwellen (siehe Tab. 5.4) liefern eine Untergrund-

Signal / Schwelle	L1-Raten		
	94 Hz	886 Hz	4355 Hz
$E_T$ (1)	13	12	5
$E_T$ (2)	28	20	6
$E_T$ (3)	44	35	10
$E_{T_{miss}}$ (1)	13	5	3
$E_{T_{miss}}$ (2)	28	12	5
$E_{T_{miss}}$ (3)	44	25	8
$E_{forw}$ (1)	5	5	2
$E_{IF}$ (1)	5	5	2
$E_{barrel}$ (1)	10	5	1
$E_{barrel}$ (2)	16	8	3
$E_{barrel}$ (3)	24	12	5
$E_{BEMC}$ (1)	20	5	1
$E_{back}$ (3)	25	25	8

Tabelle 5.4: Schwellen für den L1-Trigger-Cocktail. In der ersten Spalte (94 Hz) wurden die Triggerbedingungen (15) und (16) entfernt.

rate von 886 Hz, welche 72 Beam-Gas- und 67 Beam-Wall-Ereignissen entsprechen. Die Bestimmung restriktiver (entspannter) Triggerschwellen wurde dadurch ermöglicht, daß die Schwellen der Bedingungen mit den jeweils höchsten (niedrigsten) Raten angehoben (abgesenkt) wurden (siehe Abb. 5.1). In Tabelle 5.4 sind die Schwellen für die einzelnen Variablen und die daraus resultierenden verschiedenen Untergrundraten dargestellt.

Da der L2-Trigger nur die von L1 akzeptierten Untergrundeignisse sieht, ergibt sich durch die limitierte Statistik der Monte-Carlo-Daten ein Problem beim Training Neuronaler Netzwerke. Eine Abschätzung der erforderlichen Anzahl von Lernmustern bei gegebener Netzwerkkomplexität (vgl. Kap. 4.5.3) zeigt, daß je Gewicht etwa 10 Trainingsmuster zur Verfügung stehen sollten, um eine gute Generalisierung zu erreichen. Die Situation wurde dadurch entschärft, daß die L1-Triggerschwellen gesenkt wurden (3. Spalte in Tab. 5.4) und mit diesem "entspannten" L1-Trigger 691 Untergrundeignisse, einer Rate von 4355 Hz entsprechend, die Trainingsmenge bildeten. Durch Absenken der Kalorimeterschwellen erhöht sich die Rate jedoch, ohne entscheidende Gewinne in der Physikeffizienz zu erzielen (rechte Spalte in Tab. 5.5).

Um einen Vergleich zwischen einem reinen L1-Trigger und der Kombination aus L1 und L2 durchführen zu können, war es nötig, die angestrebten 100 Hz Ausgangsrate allein durch Verändern des L1-Cocktails zu erhalten und das Verhalten der Physikef-

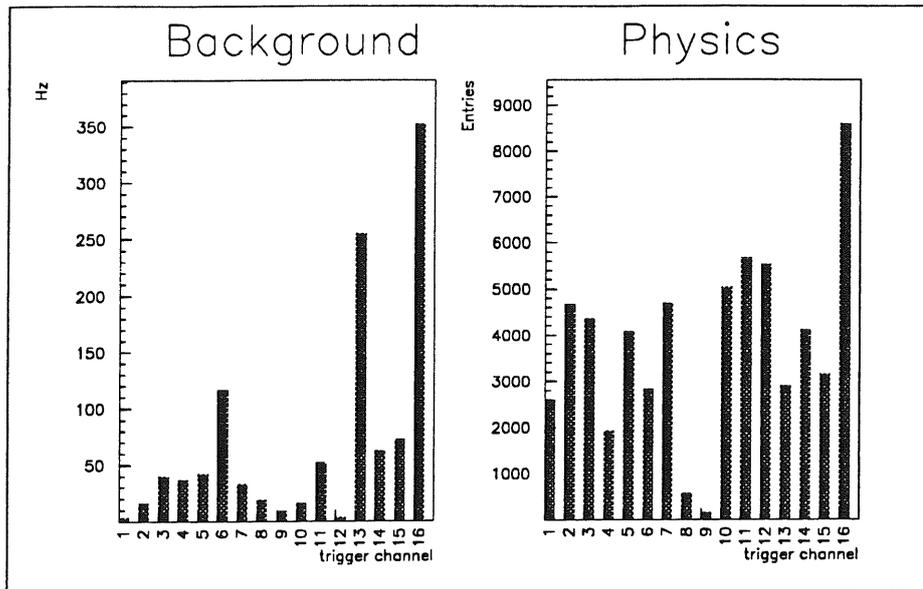


Abbildung 5.1: Häufigkeitsverteilungen der L1-Triggerkanäle beim Standard-L1-Trigger (886 Hz). Im linken Teilbild ist die Rate der Untergrundereignisse, im rechten die Anzahl der Physikereignisse aufgetragen, welche die Triggerbedingungen erfüllen.

fizienzen zu beobachten. Ein restriktiver L1-Trigger verwirft sehr viele Ereignisse der Schwer-Quark-Produktion und Ereignisse mit niedrigem  $Q^2$  (linke Spalte in Tab. 5.5). Dies hängt vor allem damit zusammen, daß die Bedingungen (15) und (16), die auf Spurinformatoren beruhen, aus dem Triggercocktail entfernt werden mußten, um die erforderliche Reduktion zu erreichen.

Physikklasse	Optimiert auf L1-Rate		
	94 Hz	886 Hz	4.4 kHz
Neutrale Ströme (NC1), $Q^2 > 300 \text{ GeV}^2$	99.4	99.9	99.9
NC2, $Q^2 < 300 \text{ GeV}^2, x > 0.037$	24.3	94.1	95.6
NC3, $Q^2 < 300 \text{ GeV}^2, x < 0.037, y < 0.3$	57.6	96.6	96.9
NC4, $Q^2 < 300 \text{ GeV}^2, x < 0.037, y > 0.3$	39.5	81.1	86.8
$c\bar{c}$ Produktion	3.5	47.2	53.3
$b\bar{b}$ Produktion	12.0	94.0	98.3
Geladene Ströme	72.9	92.2	97.2

Tabelle 5.5: Effizienzen des L1-Triggers bei verschiedenen Ausgangsraten.

## 5.3 Netzwerksimulationen

### 5.3.1 Das JETNET-Programm

Die Simulation Neuronaler Netze wurde mit Hilfe des Programmpakets JETNET (Version 2.0) [38] durchgeführt.

Die Routinen erlauben die Realisierung verschiedenster Netzwerktopologien und Trainingsalgorithmen. JETNET wurde eingesetzt, um ein Multilayer-Feedforward-Netz mit Backpropagation Lernalgorithmus zu implementieren.

Die JETNET-Programmmodule verwenden zur Steuerung und zum Datenaustausch common-Blöcke, die in der Programmdokumentation beschrieben werden [36]. Die wichtigsten Parameter der Simulationssoftware und die gewählten Werte sind in Tabelle 5.6 zusammengefaßt.

Die Simulationen wurden auf einer IBM 370 unter VM/CMS und auf einer IBM

Ausgangsneuronen	$m = 1$
Verborgenen Neuronen	$\ell = \text{variabel}$
Eingangsneuronen	$n = 19$
Aktivierungsfunktion	$g(x) = 1/(1 + e^{-\frac{x}{T}})$ $T = \frac{1}{2}$
Initialisierungsintervall der Gewichte	$\Delta\omega = -0.1 \dots +0.1$
Lernparameter	$\eta = 0.001$
Momentumfaktor	$\alpha = 0.5$
Änderung der Gewichte	nach jedem Lernmuster
Lernzyklen	$2 \cdot 10^6$

Tabelle 5.6: Simulationsparameter

RS/6000 Workstation unter AIX 3.1 durchgeführt.

### 5.3.2 Netzwerktopologien und Training

Um Untergrundereignisse von Physikereignissen zu trennen, wurden 19 Triggergrößen verwendet (vgl. Tab. 5.3), die bereits auf der ersten Triggerstufe zur Verfügung stehen. Die Energiesummen  $E_{forward}$ ,  $E_{barrel}$ ,  $E_{backward}$  und  $E_{tot}$  wurden nicht verwendet, da ihre Information redundant ist. Es wurde keinerlei Gebrauch von Daten gemacht, die erst auf Stufe 2 erhältlich sind (z.B. Bigtower-Energien). Damit sollte gezeigt werden, wie groß der Gewinn ist, wenn man nur die Korrelationen innerhalb der bestehenden Daten vollständig ausnutzt. Durch die Verwendung von detaillierterer Information ist zu erwarten, daß die Effizienz des Triggers noch weiter steigerbar ist.

Es wurden vier verschiedene Netztopologien untersucht, die alle mit den "L1-Trigger-Mustern" trainiert wurden:

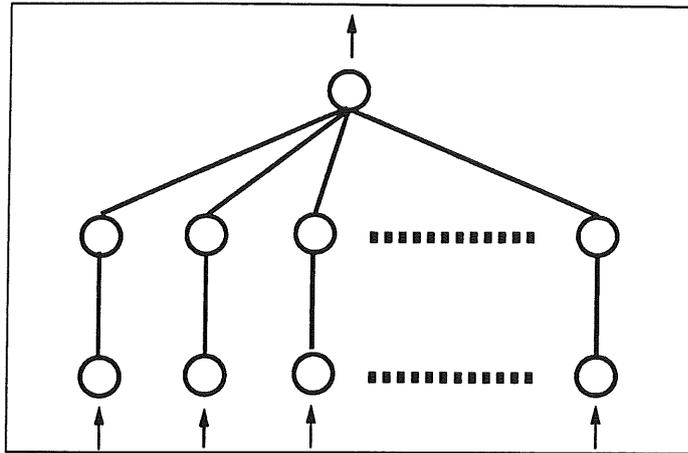


Abbildung 5.2: Topologie eines Single-Synapse-Netzes.

- **Das Linear-Separator-Netz**

Dieser Netzwerktyp weist lediglich einen versteckten Knoten auf und ist deshalb äquivalent zum *Linearen (Fischer) Diskriminator*, der nur eine einzige Ebene im Hyperraum definiert. Im Gegensatz dazu beschreibt das Single-Synapse-Netz einen Würfel im Raum der Triggervariablen. Für jede Physikklasse wurde ein eigenes Netz trainiert und diese Netze in einem Komitee zusammengefaßt, das heißt : Jedes der einzelnen Netze ist darauf trainiert, eine einzige Physikklasse von den beiden Untergrundklassen zu trennen. Diese Netze laufen parallel und bearbeiten alle dasselbe Ereignis (Trigger-Muster). Ein Ereignis wird dann akzeptiert, wenn mindestens eines der Netze dieses als Physik klassifiziert. Die Schwellen für das Ausgabeneuron sind bei allen Netzen gleich und werden auf eine Ausgangsrate von 100 Hz optimiert.

- **Das "Single-Synapse-Netz"**

Die Topologie eines solchen Netzes ist in Abb. 5.2 dargestellt. Die Zahl der Eingabeneuronen und der versteckten Neuronen ist gleich, und es besteht nur eine einzige synaptische Verbindung zwischen Eingabeneuron und verstecktem Neuron.

Durch diese Art der Verknüpfung wird das Netzwerk gezwungen, für jede Eingabegröße einen optimalen eindimensionalen Schnitt auszuwählen und am Ausgabeneuron eine Linearkombination der Einzelentscheidungen vorzunehmen. Aufgrund der nichtlinearen Aktivierungsfunktion in den versteckten Knoten ist dies nicht identisch mit einem Linearen Diskriminator (siehe unten). Die Single-Synapse-Netze wurden ebenfalls in Komitee-Form organisiert, d.h. jedes war auf eine spezielle Physikklasse trainiert.

- **Das "Merged-Netz"**

Hier handelt es sich um ein einziges, vollverknüpftes Netzwerk mit einer großen Anzahl von versteckten Neuronen (64). Diese Netz wurde darauf trainiert, alle Physikklassen von den Untergrundklassen unterscheiden zu können, indem die

Ereignisse aus den Physikdatensätzen mit gleicher Häufigkeit präsentiert wurden, aber immer Physik und Untergrund alternierend.

Die größere Anzahl von Knoten ist erforderlich, um die kombinierten Physikreaktionen vom Untergrund zu trennen. Eine weitere Vergrößerung des Netzes zeigte empirisch keine Verbesserung mehr in der Klassifizierungsfähigkeit, die Neuronen in der verborgenen Schicht begannen lediglich damit, die gleichen Gewichte anzunehmen und identische Schnitte durchzuführen (siehe Kap. 4.4).

- **Das Komitee-Netz**

Mehrere dreischichtige vollständig verknüpfte Netze mit 25 versteckten Knoten wurden als Komitee organisiert. Die Zahl der Knoten in der verborgenen Schicht wurde durch die Erkenntnisse aus Kapitel 4.4 auf mindestens  $n + 1$  festgelegt, um ein geschlossenes Volumen im  $n$ -dimensionalen Raum beschreiben zu können (wobei  $n$  die Anzahl der Eingabeneuronen beziffert). Die Zahl 25 wurde empirisch durch Variation bestimmt, bei der die Physikeffizienz als Kriterium diente. Die Verwendung von zwei versteckten Schichten brachte erwartungsgemäß keine signifikanten Verbesserungen.

Die Datensätze der Untergrundreaktionen (Beam-Gas und Beam-Wall) wurden während des Trainings gemischt und in gleichen Anteilen dem Netz präsentiert. Ebenso wurden auch Physik- und Untergrundereignisse zufällig, aber mit gleicher Häufigkeitsverteilung, trainiert.

Um die Generalisierungsfähigkeit der austrainierten Netze zu überprüfen, wurde die erste Hälfte der Ereignisse eines Datensatzes zum Training benutzt, die zweite Hälfte zum Testen der Generalisierung. Dies war trotz der eingeschränkten Untergrundstatistik notwendig, um eine Abschätzung der Leistungsfähigkeit unter realistischen Bedingungen, bei dem das Netz mit unbekanntem Mustern konfrontiert wird, vornehmen zu können.

Während des Lernvorgangs wurde mit FORTRAN-Fließkommagenauigkeit gearbeitet, um keine Konvergenzprobleme durch Rundungsfehler zu riskieren [39]. Im Ausführungsmodus der Test-Phase wurde dann auf 8-Bit Präzision bei Gewichten und Daten umgeschaltet, um eine realistische Situation, wie sie von einer L2-Hardware-Lösung dargestellt wird, abzubilden. Es wurde beim Übergang von Fließkomma- zu 8-Bit-Präzision kein Unterschied in der Leistungsfähigkeit des Systems festgestellt. Der tieferliegende Grund hierfür ist in der Form der sigmoiden Aktivierungsfunktion zu sehen, denn während des Lernens wird der sich stark ändernde mittlere Teil der Funktion auf den Wertebereich einer Eingangsvariablen abgebildet. Zur Beschreibung dieses Bereichs ist für das vorliegende Problem eine Dynamik von 8 Bit ausreichend. Die Anzahl der nötigen Lernzyklen stellte sich als weit höher heraus, als ursprünglich angenommen. In ersten Simulationen [25] schien nach bereits 50000 Mustern ein Plateau erreicht zu sein. In Wirklichkeit aber nahm die Lerngeschwindigkeit nur stark ab, der Klassifizierungsfehler verringerte sich aber doch stetig. Erst nach mehreren Millionen Zyklen ist wirklich ein Plateau erreicht. Dies wurde durch Langzeitsimulationen verifiziert.

Das Lernverhalten ist in Abbildung 5.3 als Verlauf des globalen Fehlers bzw. der Physikeffizienz gegen die Anzahl der Trainingsepochen dargestellt. Ein "Training Cycle"

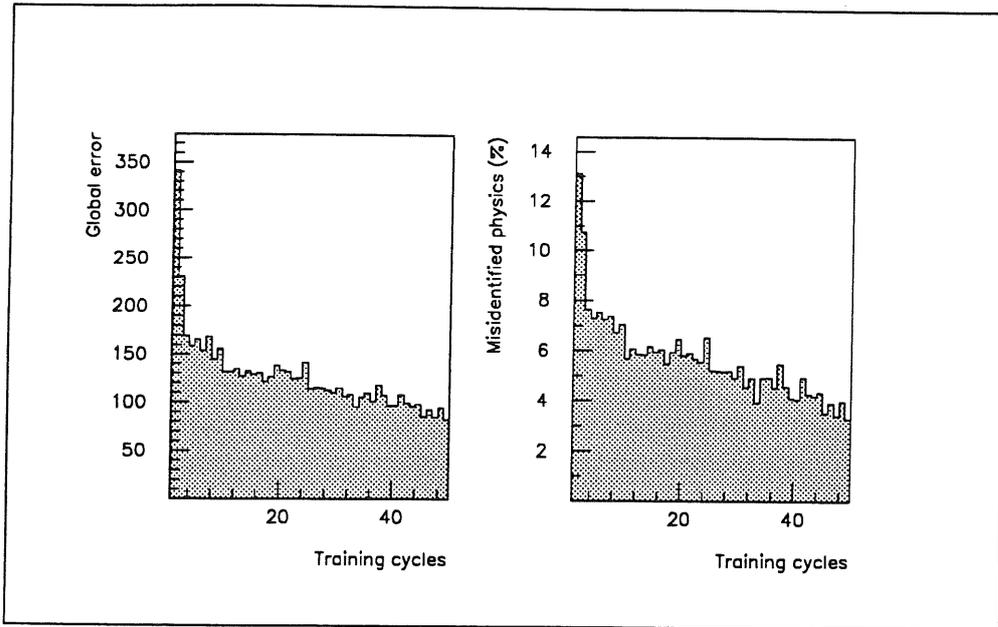


Abbildung 5.3: **Lernverhalten eines Neuronalen Netzes.** Im linken Bild ist der globale Fehler, gemittelt über alle Muster eines "Trainings Cycles" (50000 Muster) gegen die Anzahl der "Trainings Cycles" aufgetragen, im rechten Bild der Prozentsatz der falsch klassifizierten Physikereignisse. Man erkennt den schnellen Abfall des Fehlers während der ersten Epoche und die langsame aber stetige Verbesserung während der folgenden 2 Millionen Lernzyklen.

entspricht hier 50000 Lernzyklen. Man erkennt deutlich das schnelle Abfallen des Fehlers in der ersten Epoche und die Halbierung des Fehlers während der folgenden zwei Millionen Lernzyklen.

Aufgrund der Unsicherheiten in den Monte-Carlo-Simulationsdaten wurde kein Versuch unternommen, die Ergebnisse der Netzwerke bis an diese Grenzen zu steigern. Die Ergebnisse geben jedoch einen ersten Hinweis auf das bislang ungenutzte Informationspotential, das bereits in L1-Daten steckt.

## 5.4 Veranschaulichung von L2-Schnitten

Obwohl eine Visualisierung von Schnitten im 19-dimensionalen Raum eines vollständig verknüpften Netzes kaum möglich ist, erlaubt das "Single-Synapse-Netz" dennoch einen Weg aufgrund seiner minimalen Verknüpfung. Die optimalen eindimensionalen Schnitte, die das Netz findet, sind in Abbildung 5.4 zu betrachten.

Dazu ist in der linken Spalte jeweils die eindimensionale Verteilung einer Eingangsvariable der Untergrundereignisse dargestellt, in der mittleren Spalte die Verteilung dieser Variable für eine bestimmte Physikreaktion. Die gestrichelten Linien in beiden Spalten geben den Ort des Schnittes an, wenn man eine Stufenfunktion verwenden würde. Die rechte Spalte stellt die Breite der Übergangszone dar, die das zugehörige versteckte Neuron mit der Länge seines Gewichtsvektors festlegt.

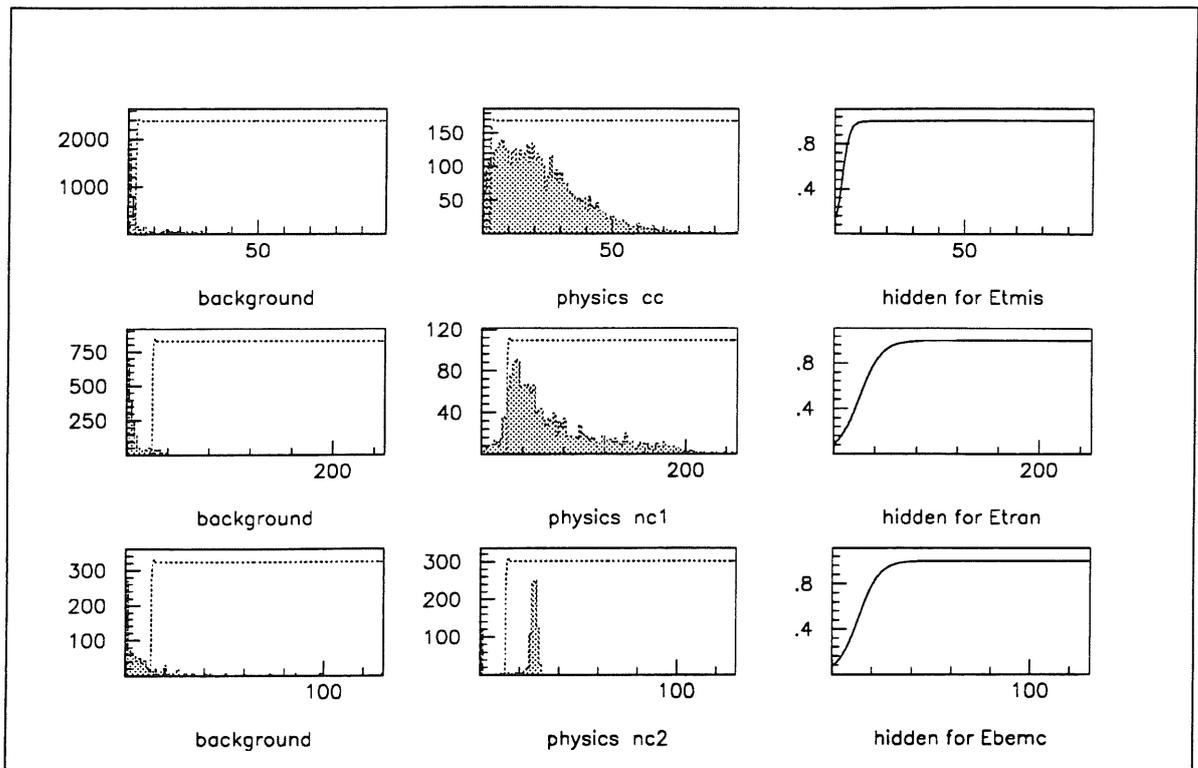


Abbildung 5.4: Illustration der eindimensionalen Schnitte des "Single-Synapse-Netzes". In der linken bzw. mittleren Spalte sind die Verteilungen der Eingangsvariablen für Untergrund bzw. Physik dargestellt. Die gestrichelten Linien zeigen den Ort des Schnittes an, die rechte Spalte beschreibt die Steilheit und Breite der Trennzone.

Man erkennt, daß die gefundenen Schnitte sinnvoll sind und der Anstiegsbereich der Sigmoidfunktion der Dynamik der Eingangsverteilung angepaßt ist, d.h. die Steilheit der Aktivierungsfunktion wird an den Überlapp von Untergrund und Physikverteilung adaptiert.

## 5.5 Ergebnisse des Neuronalen L2-Triggers

In Abbildung 5.5 sind zwei typische Ergebnisse eines Netzwerktrainings dargestellt. Anhand der relativ einfach zu klassifizierenden Physikreaktion NC1 (obere Bildhälfte) und anhand der  $c-\bar{c}$  Produktion (untere Bildhälfte) können die wesentlichen Merkmale der Netzwerke illustriert werden.

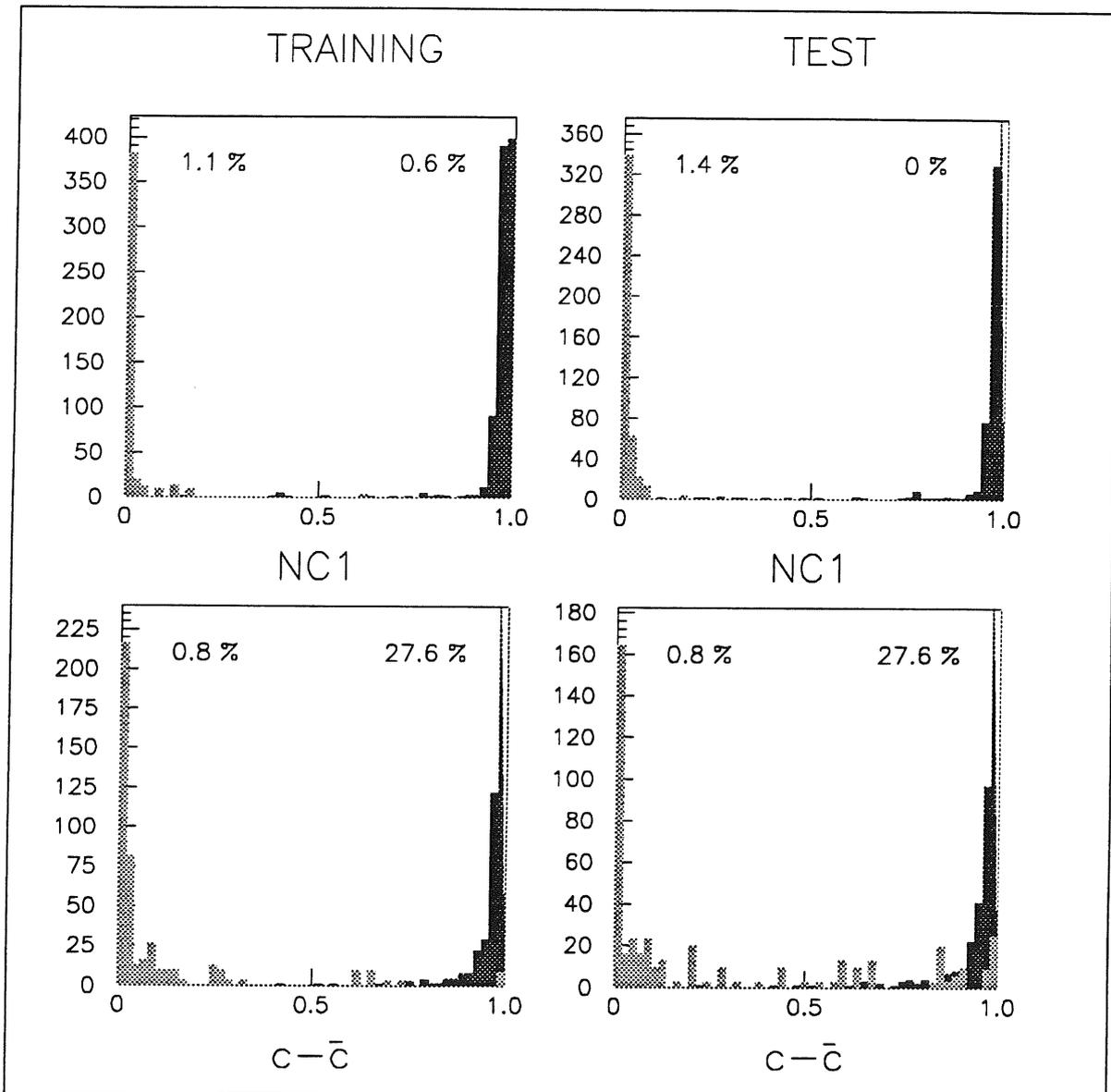


Abbildung 5.5: Ergebnisse eines Netzwerktrainings für die Reaktionen NC1 und  $c\bar{c}$ . Auf der Abszisse ist die Aktivität des Ausgabeneurons aufgetragen, gegen die die Häufigkeitsverteilungen von Untergrundereignissen (helle Schraffierung) und Physikereignissen (dunkle Schraffierung) aufgetragen sind. Die rechte Spalte beschreibt die Ergebnisse auf einer Testmenge und somit die Generalisierung. Die Prozentzahlen geben den Anteil der – bei einem Schnitt bei 0.5 – falsch klassifizierten Ereignisse an.

Dabei ist zu beachten, daß die Topologien der  $c - \bar{c}$ -Ereignisse sehr stark den Ereignistopologien der Untergrundreaktionen ähneln und die beiden Klassen somit schwer zu separieren sind. Auf der Abszisse ist die Aktivität des Ausgabeneurons aufgetragen, das darauf trainiert war, den Wert  $y = 0$  für Untergrund und den Wert  $y = 1$  für Physik zu liefern. Die hell schraffierten Flächen zeigen die Häufigkeit von Untergrundereignissen, die dunkel schraffierten die Häufigkeit von Physikereignissen mit einem bestimmten Ausgabewert. In der linken Hälfte des Bildes sind die Verteilungen für Ereignisse aus der Trainingsmenge gezeigt, rechts die Ergebnisse für die Testmenge, die das Netz nie zuvor gesehen hat.

Man erkennt, daß das Netz nicht in der Lage ist, eine perfekte Abbildung durchzuführen, d.h. nicht jedes Physikereignis liefert einen Wert von 1 und nicht jedes Untergrundereignis den Wert 0.

Es ist außerdem deutlich erkennbar, daß das Netz in der Lage ist, zu generalisieren, wohl aber auf Kosten einer etwas breiteren Verteilung und einer höheren Anzahl fehlklassifizierter Ereignisse (siehe Prozentangaben in Abb. 5.5).

Dies wird besonders bei der Klassifikation von  $c - \bar{c}$  Ereignissen ersichtlich, die aber in diesem Fall von der limitierten Untergrundstatistik beeinflußt ist und durch eine höhere Anzahl von Ereignissen eine bessere Generalisierung zeigen kann.

Die Untergrundrate läßt sich durch einen Schnitt auf dem Ausgangswert variieren. Dieser Schnitt verändert stetig die Größen *Reinheit* und *Effizienz*, denn ein Schnitt nahe bei  $y = 1$  ergibt eine hohe Reinheit (= niedrige Untergrundrate), gleichzeitig aber auch eine niedrige Physikeffizienz. Ein Schnitt nahe bei  $y = 0$  erreicht eine Physikeffizienz von 100%, bedeutet aber eine geringe Reinheit und somit keinerlei Untergrundreduktion. Dieser Schnitt ist als Analogon zu den Schwellen des L1-Triggers aufzufassen, mit deren Variation man ebenfalls die Ausgangsrate auf Kosten der Effizienz beeinflussen kann.

In den Abbildungen 5.6 und 5.7 sind Variationen des Schnittes durchgeführt und die Ausgangsraten des L2-Triggers als Funktion der Effizienz aufgetragen worden. Die Effizienz bezieht sich in diesen Darstellungen auf die Anzahl der Physikereignisse, die vom Neuronalen Netz akzeptiert wurden, relativ zur Zahl, die den L1-Trigger erfolgreich passieren konnten.

Den Abbildungen kann man entnehmen, wieviel Prozent an Physik geopfert werden muß, um einen bestimmten Untergrundreduktionsfaktor zu erzielen. Die drei Kurven beziehen sich auf das Komitee-Netz (durchgezogene Linie), das "Merged-Netz" (gestrichelte Linie) und das "Single-Synapse-Netz" (gepunktete Linie). Man erkennt, daß in allen drei Fällen nur wenige Physikereignisse verloren werden, wenn man die Untergrundrate von 886 Hz unter 100 Hz senken möchte, da der anfängliche Abfall der Kurven (von rechts nach links) sehr steil ist. (Man beachte die logarithmische Skala der Untergrundrate.)

Das Komitee-Netz und das "Merged-Netz" sind vergleichbar in ihrer Leistung. Das "Single-Synapse-Netz" zeigt Schwächen bei der Erkennung von Physik mit niedrigem  $Q^2$  und Ereignissen der Schwer-Quark-Produktion.

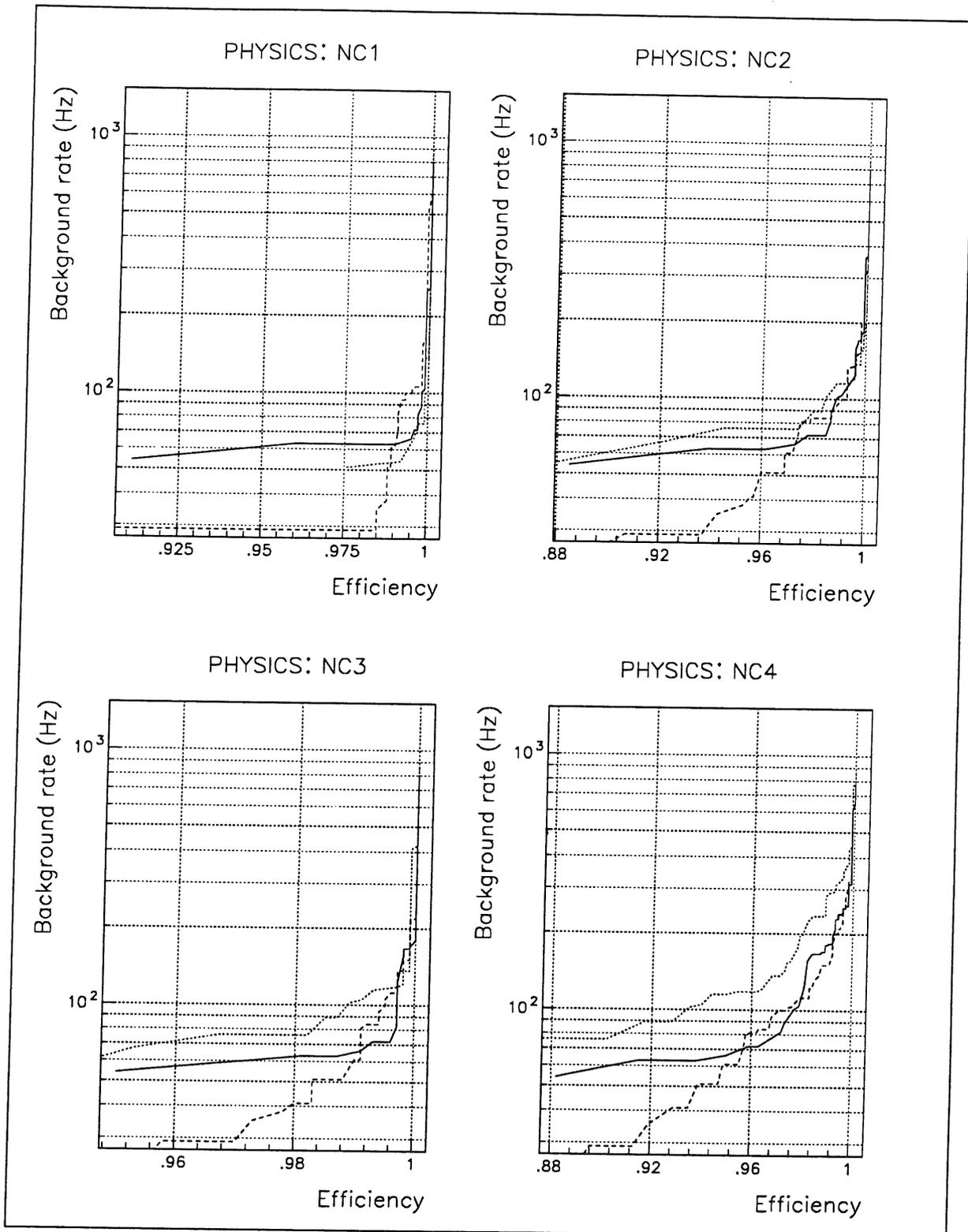


Abbildung 5.6: Untergrundraten als Funktion der Physikeffizienz. Die Physikeffizienzen relativ zur L1-Effizienz sind für das Komitee-Netz (durchgezogene Linie), das "Merged-Netz" (gestrichelte Linie) und das "Single-Synapse-Netz" (gepunktete Linie) bezüglich jeder der Physikklassen dargestellt.

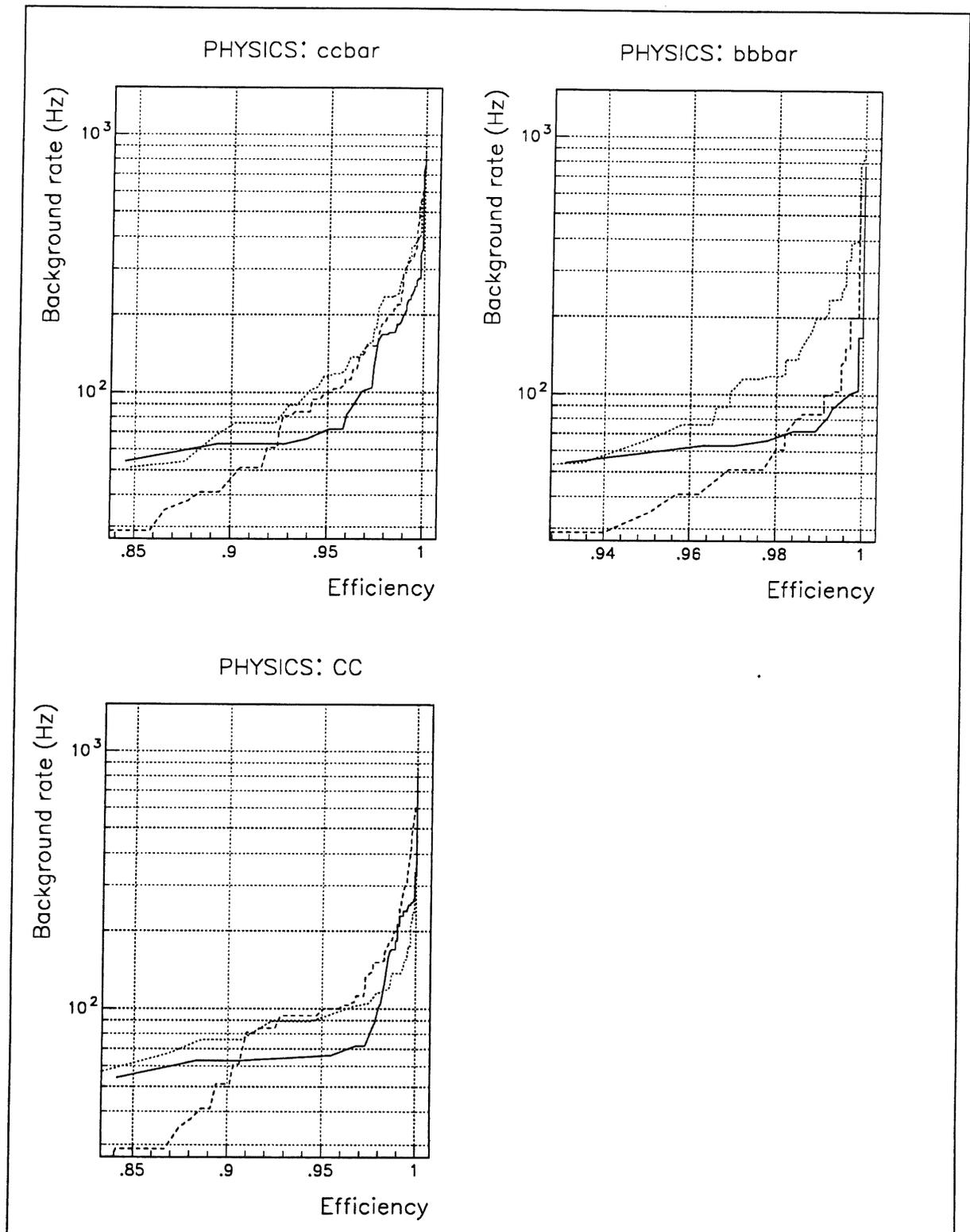


Abbildung 5.7: Untergrundraten als Funktion der Physikeffizienz. Die Physikeffizienzen relativ zur L1-Effizienz sind für das Komitee-Netz (durchgezogene Linie), das "Merged-Netz" (gestrichelte Linie) und das "Single-Synapse-Netz" (gepunktete Linie) bezüglich jeder der Physikklassen dargestellt.

In Tabelle 5.7 werden die Selektionseffizienzen der einzelnen Subnetze des Komitee-Netzwerks aufgelistet, wobei sich die Zahlen auf eine L2-Ausgangsrate von 100 Hz beziehen. Die Werte sind wiederum relativ zu L1-getriggerten Ereignissen zu verstehen.

Die Tabelle zeigt die Fähigkeit eines Subnetzes, das auf eine bestimmte Art von Phy-

Physik	Komitee-Netzwerk trainiert auf							gesamt	fremde	"Merged-Netz"
	NC1	NC2	NC3	NC4	$c\bar{c}$	$b\bar{b}$	CC			
NC1	96.6	55.8	74.9	49.6	20.5	26.4	88.3	99.8	93.6	99.9
NC2	10.9	98.4	75.9	10.8	30.4	1.5	12.1	98.7	81.8	98.7
NC3	6.1	64.0	98.4	46.2	78.1	33.1	8.8	99.7	97.6	99.6
NC4	24.6	6.4	52.4	95.3	69.5	55.8	31.4	97.3	80.7	97.4
$c\bar{c}$	1.2	5.9	21.0	19.1	94.8	53.0	17.7	96.3	62.2	96.4
$b\bar{b}$	12.1	0.5	44.6	55.4	95.8	96.2	52.8	99.3	97.4	99.3
CC	42.6	3.6	52.9	32.4	34.2	32.2	96.2	97.8	65.4	98.8

Tabelle 5.7: Effizienz des Neuronalen L2-Triggers relativ zu L1-getriggerten Ereignissen (in Prozent).

sik trainiert ist, auch Ereignisse anderer Physikreaktionen zu klassifizieren (Einträge außerhalb der Diagonalen). Dabei sieht man auch, daß ähnliche Physikreaktionen von den korrespondierenden Subnetzen ebenfalls sehr effizient erkannt werden (z.B.  $c - \bar{c}$  und  $b - \bar{b}$ ). Der starke Überlapp und die Redundanz der Entscheidungen resultieren in einer hohen kombinierten Identifikationseffizienz (drittletzte Spalte). Selbst wenn man das Subnetz, das für eine bestimmte Klasse optimiert war, inaktiviert, erhält man noch immer eine relativ hohe Effizienz für diese Klasse (vorletzte Spalte).

In der letzten Spalte sind zum Vergleich die Effizienzen des "Merged-Netzes" dargestellt, die die marginalen Unterschiede zum Komitee-Netz verdeutlichen.

Insgesamt erfüllen alle untersuchten Netzwerkarchitekturen das Kriterium, die Untergrundrate um den Faktor 10 zu senken, ohne dabei wesentlich Physik zu verlieren.

Ein Vergleich von "Single-Synapse-Netz" und Linearem Diskriminator ist in Tabelle

Netzwerk	Richtig klassifizierte Ereignisse (in %)								
	Beam-Gas	Beam-Wall	NC1	NC2	NC3	NC4	$c\bar{c}$	$b\bar{b}$	CC
Linearer Diskr.	90.1	85.2	99.1	97.0	98.2	82.9	61.8	97.4	73.1
Single-Synapse	89.3	94.5	99.6	99.2	99.0	90.6	77.8	96.8	94.5

Tabelle 5.8: Prozentsatz der korrekt klassifizierten Ereignisse von Linearem Diskriminator und "Single-Synapse-Netz".

5.8 dargestellt. Der prozentuale Anteil richtig klassifizierter Ereignisse wird sowohl für

die Physik- als auch die Untergrundklassen gezeigt. Es ist bemerkenswert, daß selbst ein so einfaches Netzwerk signifikant bessere Ergebnisse liefert als die Standardmethode des linearen Diskriminators..

## 5.6 Fehlertoleranz des Neuronalen Netzes

Um eine verlässliche Information über das Verhalten des Triggers in einer realen Umgebung zu erhalten, wurde die Robustheit der Netze bezüglich Fehlern in den Eingangsdaten untersucht. Diese Fehler können z.B. durch falsche Kalibration von Kalorimeterenergien oder Hardware-Ausfällen entstehen. Einen ersten Hinweis bzgl. der Sensitivität des Triggers lieferte ein künstlich eingeführter drastischer "Offset" von -20% in den Kalorimeterenergien. Dies betraf natürlich auch die Effizienzen des L1-Triggers, die zusammen mit den Verlusten des L2-Triggers in Tabelle 5.9 zusammengefaßt sind. Die Werte sind dieses Mal als absolute Prozentzahlen zu verstehen; als L2-Trigger fungierte hierbei das "Merged-Netz". Der Standard-L1-Cocktail für 886 Hz bleibt relativ stabil und verliert kaum Physik. Das Bild ändert sich jedoch drastisch im Fall des 100 Hz L1-Cocktails, der große Verluste bei verschiedenen Physikreaktionen produziert. Das Neuronale Netz in Kombination mit dem Standard-L1-Trigger liefert einen akzeptablen 93 Hz Untergrund und verliert in keiner der Physikklassen essentiell an Effizienz, insbesondere nicht in kritischen Klassen, die ohnehin bereits niedrige absolute Effizienzen aufweisen ( $c\bar{c}$ -Produktion).

Um die relative Wichtigkeit von kalorimetrischer Information und Spurinformaton zu

Ereignisklasse	Triggerstufe und Untergrundrate					
	L1 (101 Hz)	Verlust	L2 (93 Hz)	Verlust	L1 (757 Hz)	Verlust
NC1	99.2	-0.2	99.5	-0.4	99.9	0.0
NC2	15.9	-8.4	88.5	-4.4	93.8	-0.3
NC3	30.5	-27.1	95.4	-0.8	96.6	0.0
NC4	22.3	-17.2	76.0	-3.0	80.3	-0.8
$c\bar{c}$	1.6	-1.9	44.8	-0.7	46.7	-0.5
$b\bar{b}$	8.7	-3.3	92.6	-0.8	93.5	-0.5
CC	67.6	-5.3	88.5	-1.6	90.6	-1.6

Tabelle 5.9: Effizienzverluste nach Einführung eines -20% "Offsets" in den Kalorimeterenergien.

demonstrieren und den Ausfall eines der beiden Systeme zu simulieren, wurde jeweils eines von den Netzwerkeingängen entfernt, d.h. die entsprechende Eingangswerte auf Null gesetzt.

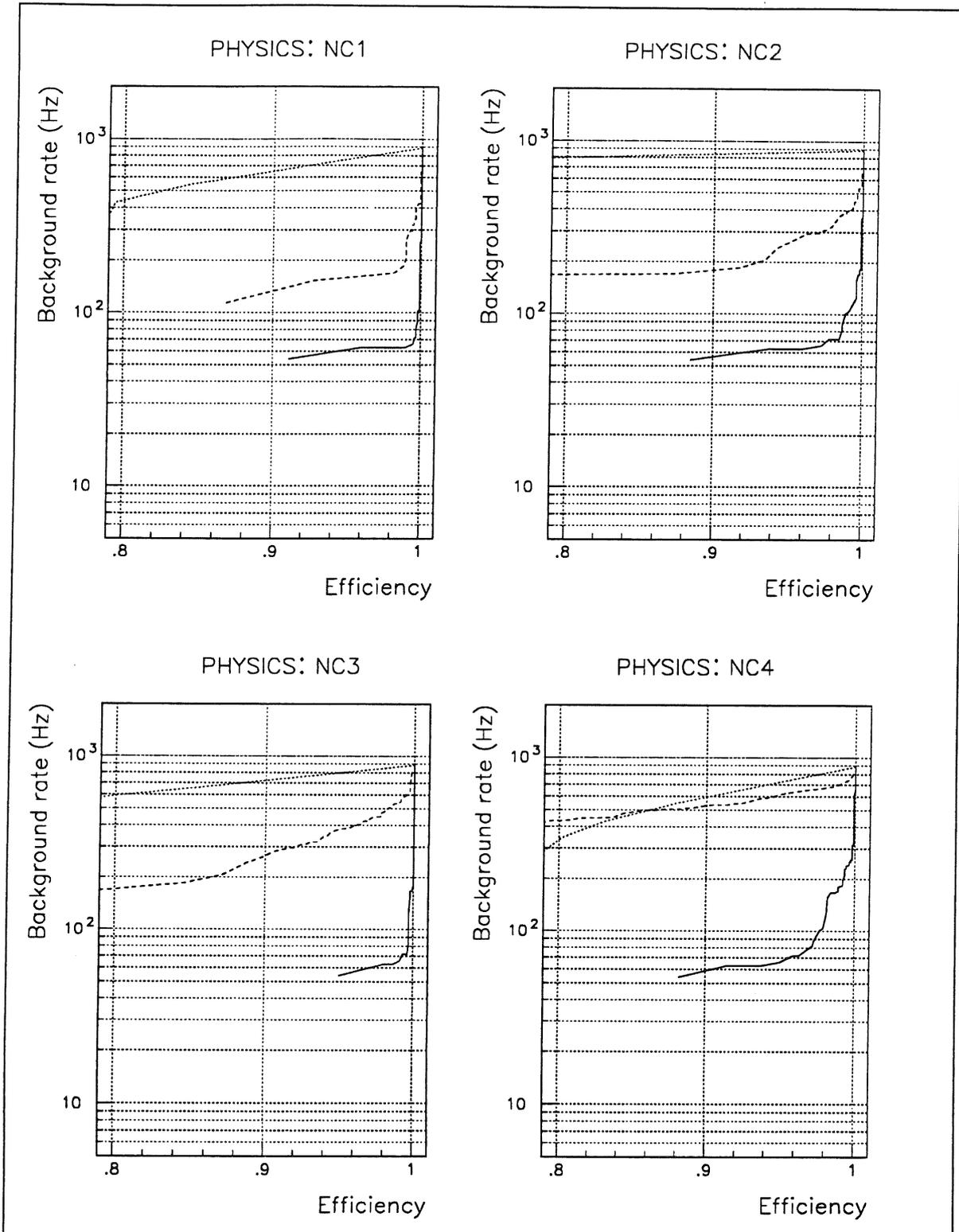


Abbildung 5.8: Untergrundraten als Funktion der Physikeffizienz bei Ausfällen von Subdetektoren. Die Physikeffizienzen relativ zur L1-Effizienz sind für den Ausfall der Spurinformationen (gestrichelte Linie), für den Ausfall der Energieinformationen (gepunktete Linie) im Falle des Komitee-Netzes (durchgezogene Linie) dargestellt.

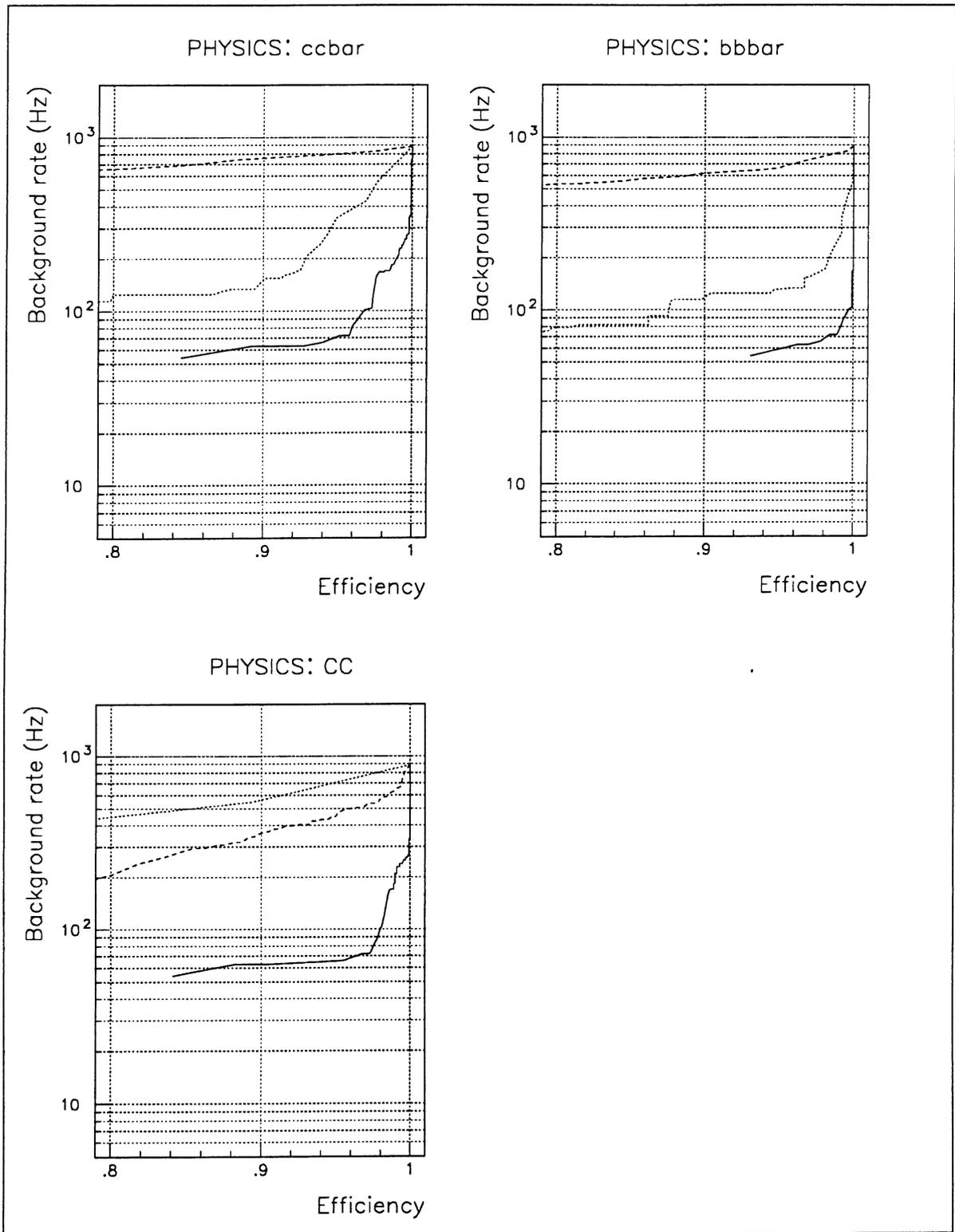


Abbildung 5.9: Untergrundraten als Funktion der Physikeffizienz bei Ausfällen von Subdetektoren. Die Physikeffizienzen relativ zur L1-Effizienz sind für den Ausfall der Spurinformatoren (gestrichelte Linie), für den Ausfall der Energieinformationen (gepunktete Linie) im Falle des Komitee-Netztes (durchgezogene Linie) dargestellt.

Die Effizienzen wurden wiederum durch Variation des Schnittes am Ausgabeneuron bestimmt und analog zu Abbildung 5.6 dargestellt. Die Resultate für fehlende Spurin-formation (gestrichelte Linie) und fehlende Energieinformation (gepunktete Linie) sind im Falle des Komitee-Netzes (durchgezogene Linie) in Abbildung 5.8 und 5.9 zu sehen. Wie erwartet, ist in Physikreaktionen mit niedriger Spurmultiplicität und niedrigem  $x$  bzw.  $Q^2$  ( $b\bar{b}$  und  $c\bar{c}$ ) die Energieinformation des Kalorimeters von geringerer Bedeutung, während bei tiefinelastischer Streuung mit hohem  $Q^2$  (NC1 - NC3) die Spurin-formation nur eine untergeordnete Rolle spielen.

# Kapitel 6

## Einsatz Neuronaler Netze mit realen Daten

### 6.1 Beschreibung der Daten und Triggergrößen

#### 6.1.1 Das H1-Datennahmesystem

Gegenwärtig besteht das H1-Triggersystem nur aus dem Level-1- und dem Level-4-Trigger. Der L1-Trigger im laufenden Experiment unterscheidet sich von dem im vorigen Kapitel dargestellten L1-Triggercocktail, da noch nicht alle Subdetektortrigger fertiggestellt sind. Außerdem ist noch nicht die volle Luminosität des Beschleunigers erreicht, auf die der Standard-L1-Triggercocktail optimiert wurde. Der für die folgenden Untersuchungen relevante L1-Trigger ist in Tabelle 6.1 dargestellt.

Für die in dieser Arbeit verwendeten Daten erreichte der L1-Trigger eine Rate von ca. 15–20 Hz, die durch den L4-Trigger nochmals um den Faktor zwei auf etwa 7–10 Hz gesenkt werden konnte. Dies geschah hauptsächlich durch Verwerfen von Ereignissen, die mit dem Drahttrigger getriggert wurden, aber keine Spuren im Zentraldetektor hatten. Nachdem ein Ereignis von L4 akzeptiert worden ist, werden die Daten auf Band geschrieben. Sie enthalten auch Ereignisse, die sowohl von L1- als auch von L4-Monitor-Triggerbedingungen akzeptiert wurden. Diese Triggerbedingungen dienen zur Überwachung und Akzeptanzanalyse der einzelnen Triggerstufen, und ihr Beitrag zur Triggerrate sollte 10% nicht übersteigen. Außerdem werden etwa 10% der von einem Triggerlevel verworfenen Ereignisse dennoch akzeptiert und aufgezeichnet, um die Effizienz und Richtigkeit der Entscheidungen des Triggers prüfen zu können.

An die Aufzeichnung der Rohdaten schließt sich eine detaillierte Rekonstruktion der Ereignisse an, die "offline" auf einem speziell dafür dedizierten Rechner erfolgt. Diese Rekonstruktion wird mit dem Programmpaket H1REC durchgeführt, das z.B. Programmodule zur Energiebestimmung in den Kalorimetern, zur Spurerkennung in den Jet- und Proportionalkammern und zur Erkennung von Spuren im Myonsystem enthält. Diesem Prozeß folgt eine erste grobe Klassifizierung der Ereignisse auf der Basis rekonstruierter Größen, wie z.B. Anzahl der gefundenen Spuren mit minimalem Transversalimpuls, transversale Energien oder fehlende transversale Energien. Die Daten von Ereignissen, die in eine oder mehrere der Physikklassen fallen, werden auf sogenannte

Name	Erklärung
eTAG	Trigger vom Elektrondetektor im Luminositätssystem
eTAG & Posi-Tof	Elektrontrigger in Koinzidenz mit dem Signal für das Physikzeitfenster
eTAG & BPC	Elektrontrigger in Koinzidenz mit der BPC
eTAG & ZVTX-sig1	Elektrontrigger in Koinzidenz mit dem Trigger für Z-Vertex Signifikanz
eTAG & CIP-4	Elektrontrigger und 4 Treffer in der CIP, davon mindestens einer in der oberen und unteren Hälfte der Kammer
BSET-Etot	BEMC-Trigger für Gesamtenergie
BSET-ECLU1	BEMC-Trigger für Cluster-Energie 1
BSET-ECLU2	BEMC-Trigger für Cluster-Energie 2
BSET-Etot & Posi-Tof	BEMC-Trigger für Gesamtenergie und Signal für das Physikzeitfenster in Koinzidenz
BSET-Etot & ZVTX-sig1	BEMC-Trigger für Gesamtenergie und Trigger für Z-Vertex Signifikanz 1
Mu-Multi & ZVTX-small	Mindestens zwei getroffene Myon-Module und mindestens ein Eintrag im Z-Vertex-Histogramm
Mu-GE1-FEC & ZVTX-small	Mindestens ein Myon in der vorderen Endkappe des Myonsystems in Koinzidenz mit Eintrag im Z-Vertex-Histogramm
Mu-GE1-barrel & ZVTX-small	Mindestens ein Myon im Barrel-Bereich des Myonsystems in Koinzidenz mit Eintrag im Z-Vertex-Histogramm
Mu-GE1-BEC & ZVTX-small	Mindestens ein Myon in der rückwärtigen Endkappe des Myonsystems in Koinzidenz mit Eintrag im Z-Vertex-Histogramm
Mu-Multi & ZVTX-sig1	wie oben, aber Koinzidenz mit Z-Vertex Signifikanz 1
Mu-GE1-FEC & ZVTX-sig1	wie oben, aber Koinzidenz mit Z-Vertex Signifikanz 1
Mu-GE1-barrel & ZVTX-sig1	wie oben, aber Koinzidenz mit Z-Vertex Signifikanz 1
Mu-GE1-BEC & ZVTX-sig1	wie oben, aber Koinzidenz mit Z-Vertex Signifikanz 1
LAr-Ebarrel-1	Energie im Barrel-Bereich des Kalorimeters oberhalb der Schwelle Nr.1
LAr-Etrans-0	Transversale Energie des Kalorimeters oberhalb der Schwelle Nr.0
LAr-Electron-1	Elektrontrigger des Kalorimeters oberhalb der Schwelle Nr.1
LAr-Barrel-0 & ZVTX-sig1	Energie im Barrel-Bereich des Kalorimeters oberhalb der Schwelle Nr.0 in Koinzidenz mit dem Z-Vertex-Signifikanztrigger
ZVTX-sig2 & ( <i>CIP - backwd</i> )	Trigger für Z-Vertex Signifikanz 2 und drei getroffene Sektoren im $-z$ - Bereich der CIP
ZVTX-sig1 & DC-R- $\Phi$ -Thr0	Trigger für Z-Vertex Signifikanz 2 koinzident mit dem Driftkammer Spurtrigger

Tabelle 6.1: Der L1-Triggercocktail für die in den Untersuchungen verwendeten Daten. Alle Trigger unterliegen der globalen Bedingung, daß kein TOF-Veto von den Veto-Wänden kommt.

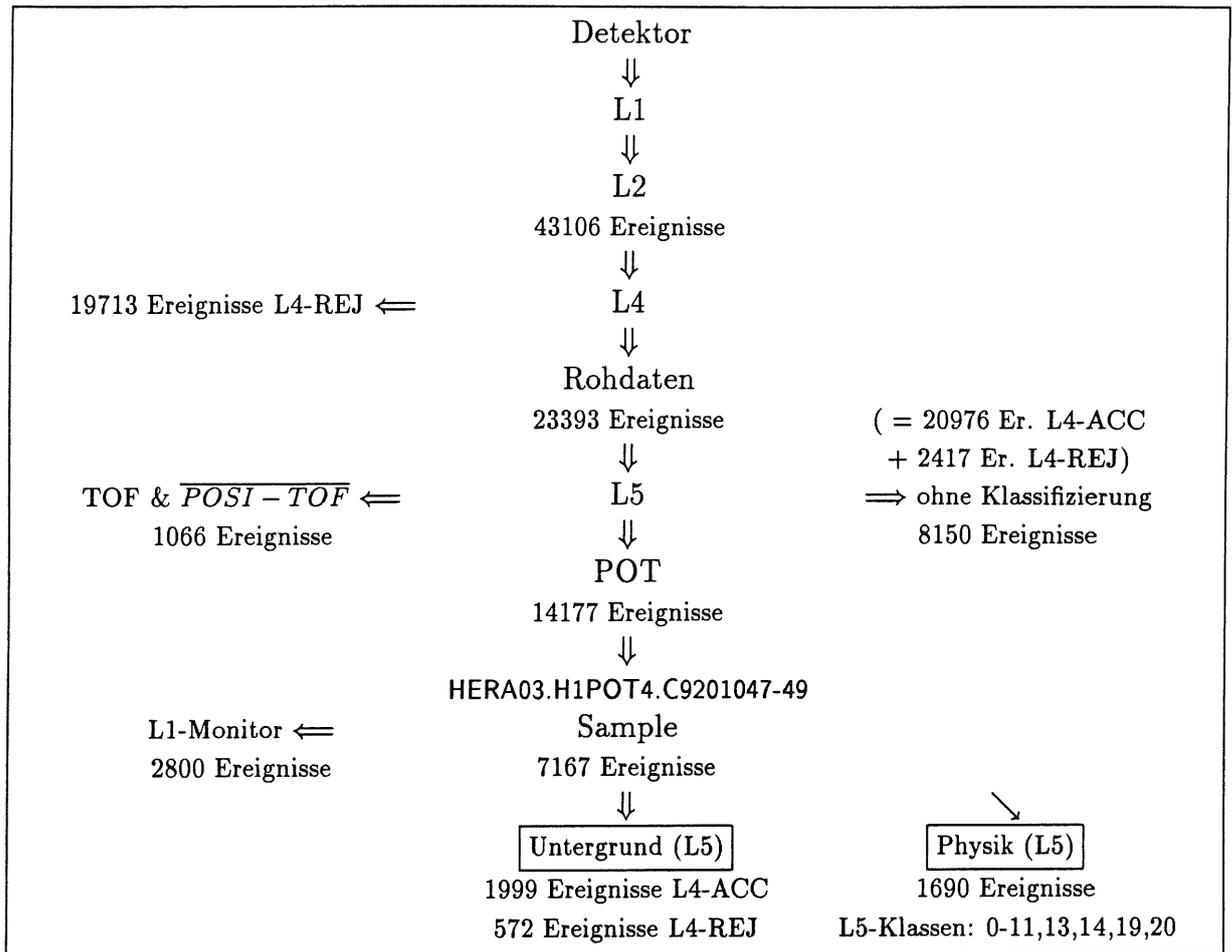


Abbildung 6.1: Datenselektion zur Extrahierung von Untergrund und Physik

*Production Output Tapes* (POT) geschrieben, falls sie nicht durch ein Veto der Veto-Wand gekennzeichnet sind und kein POSI-TOF-Signal aufweisen (siehe Kapitel 3.2.11). Der Klassifizierungsalgorithmus wird als “L5-Trigger” bezeichnet, obwohl dies im wörtlichen Sinn nicht korrekt ist, da keine Rohdaten verworfen werden.

Für Zwecke der Physikanalyse werden die POT-Daten etwas härteren Schnitten unterworfen und die selektierten Ereignisse auf ein sogenanntes *Data Summary Tape* (DST) geschrieben.

### 6.1.2 Untergrund-Daten

Die Auswahl von Untergrundereignissen ist in Abb. 6.1 exemplarisch anhand des *Run's* 27103 dargestellt. Dabei wurden wenig einschränkende Bedingungen an den Untergrund gestellt. Die Wahl beruhte hauptsächlich auf der L5-Klassifikation (siehe Tab. 6.3) von rekonstruierten Ereignissen, wobei die L5-Klassen 0-11, 13, 14, 19 und 20 für Physikkandidaten standen und alle anderen Klassen Untergrundereignisse bedeuteten. Zuvor wurden alle Ereignisse, die nur durch eine L1-Monitor-Triggerbedingung gekennzeichnet waren, herausgefiltert, da diese Ereignisse von keinem

höheren Trigger verworfen werden dürfen. Der mit dieser Selektion extrahierte “Untergrund (L5) Run 27103” wurde dem Neuronalen Netz als eine von zwei Untergrunddateien beim Training präsentiert. Die andere Art von Untergrund wurde aus den sogenannten Pilot-Bündeln gewonnen. Bei diesen handelt es sich um Teilchenbündel, die keine Partnerbündel der anderen Teilchensorte besitzen und somit keine Kollisionen, sondern nur Untergrundreaktionen erzeugen können. In dieser Arbeit wurden nur Proton-Pilotbündel verwendet, da sie eine höhere Untergrundrate erzeugen und somit den “gefährlicheren” Untergrund produzieren. Auch hier wurden L1-Monitor-Ereignisse herausgefiltert, um sich den realistischen Gegebenheiten eines L2-Triggers anzunähern.

Untergrund	Selektion	Statistik
(L5) Run 27103	<ul style="list-style-type: none"> <li>• Datei : HERA03.H1POT4.C9201047-49</li> <li>• siehe Abb. 6.1</li> </ul>	2571
Pilot-Bündel	<ul style="list-style-type: none"> <li>• Datei : H1TEMH.DST.PILOT.PROT.C9200738+46</li> <li>• Auswahl basierend auf Bündelnummer</li> <li>• nur Proton-Pilotbündel</li> </ul>	3974
Physik		
DIS (rein)	<ul style="list-style-type: none"> <li>• Datei : H1KADR.DIS.EVENTS</li> <li>• <math>20 \text{ GeV} &lt; E_e &lt; 32 \text{ GeV}</math></li> <li>• <math>4 \text{ GeV} &lt; E_e &lt; 20 \text{ GeV}</math> und 1 Barrel Spur</li> <li>• Koinzidenz von BEMC-Cluster-Zentrum mit BPC-Treffern</li> <li>• <math>E_{CL} &lt; 22 \text{ GeV}</math> und 1 Spur zur Vertexregion</li> </ul>	228
$\gamma p$ (rein)	<ul style="list-style-type: none"> <li>• Datei : H1KHUF.GAMMAP.NDST12+13</li> <li>• L5-Klasse 3 : JETJET <ul style="list-style-type: none"> <li>◦ <math>p_{t_{miss}} &lt; 100 \text{ GeV}</math> in <math>0^\circ &lt; \Theta &lt; 180^\circ</math> und 1 gute Spur</li> <li>◦ <math>E_t &gt; 10 \text{ GeV}</math> in <math>25^\circ &lt; \Theta &lt; 180^\circ</math> und 1 gute Spur</li> </ul> </li> <li>• L5-Klasse 19 : VDMETG <ul style="list-style-type: none"> <li>◦ <math>EETag &gt; 4 \text{ GeV}</math> und 1 gute Spur</li> </ul> </li> <li>• Schnitte in <math>y</math> : <ul style="list-style-type: none"> <li>◦ <math>0.35 &lt; y_{tag} &lt; 0.7</math></li> <li>◦ <math>0.3 &lt; y_{had} &lt; 1.0</math></li> <li>◦ <math>E_H(\Theta &lt; 25^\circ) &gt; 5 \text{ GeV}</math></li> </ul> </li> </ul>	289

Tabelle 6.2: Datensätze, die zum Training des Neuronalen Netzes verwendet wurden. Die Dateinamen beziehen sich auf Datensätze, die auf dem DESY IBM Hauptrechner abgelegt sind.

0. NCHQSQ	NC-DIS (hohes $Q^2$ ) : Elektronkandidat im Barrel- oder Vorwärtsbereich mit $p_{T_{miss}} < 40 \text{ GeV}$ , $E_T > 5 - 8 \text{ GeV}$
1. NCLQSQ	NC-DIS (niedriges $Q^2$ ) : Cluster im rückwärtigen Bereich mit $\Theta > 150^\circ$ und $E > 4 \text{ GeV}$
2. CCHQSQ	CC-DIS (hohes $Q^2$ ) : Transversale Energie $E_T > 20 \text{ GeV}$ und fehlender Transversalimpuls $p_{T_{miss}} > 20 \text{ GeV}$
3. JETJET	Jet-Jet-Ereignisse mit hohem $E_T$ : Im Bereich $25^\circ < \Theta < 180^\circ$ : $E_T > 10 \text{ GeV}$ und $p_{T_{miss}} < 100 \text{ GeV}$
4. SPSMU	Spezielle Signaturen mit Myonkandidaten : Eine rekonstruierte Spur eines Myonkandidaten mit $E_T > 5 \text{ GeV}$ des größten elektromagn. Clusters oder $p_{T_{miss}} > 10 \text{ GeV}$
5. SPSNMU	Spezielle Signaturen ohne Myonkandidaten : Zwei elektromagnetische Cluster mit $E_{T_1}, E_{T_2} > 5 \text{ GeV}$ und $E_{T_1} + E_{T_2} > 15 \text{ GeV}$ mit $\Delta\Phi_{1,2} > 90^\circ$ oder ein elektromagnetischer Cluster mit $E_T > 3 \text{ GeV}$ , $p_{T_{miss}} > 10 \text{ GeV}$ und $E_T + p_{T_{miss}} > 15 \text{ GeV}$
6. NCHAD	DIS-Kandidaten zur Untersuchung des Hadronflusses : eine gute Spur und NCHQSQ oder NCLQSQ
7. LETNSP	Ereignisse mit niedrigem $E_T$ ohne spezielle Signatur : $E_T > 5 \text{ GeV}$ im Bereich $25^\circ < \Theta < 180^\circ$
8. LETLEP	Ereignisse mit niedrigem $E_T$ und Leptonkandidat : wie LETNSP und ein Myonkandidat, zusätzlich eine gute Spur
9. LETTRK	Ereignisse mit niedrigem $E_T$ und wie LETNSP und $p_{T_{spur}} > 1 \text{ GeV}$ im Bereich $25^\circ < \Theta < 150^\circ$
10. LETETG	Ereignisse mit niedrigem $E_T$ und Elektron-Detektor (Tagger) : $E_T > 5 \text{ GeV}$ und $E_{T_{tagger}} > 5 \text{ GeV}$
11. LETSPS	Ereignisse mit niedrigem $E_T$ und spezieller Signatur : $J/\Psi$ -Selektion und $c-\bar{c}$ -Selektion auf Basis der Spurrekonstruktion
12. BCTRIG	Spezielle Klasse für BEMC-Studien : Mindestens ein BEMC-Cluster mit $E_{Cluster} > 4 \text{ GeV}$
13. BCELEC	siehe 13. mit $E_{Cluster} > 8 \text{ GeV}$
14. BCCALI	siehe 13. mit $E_{Cluster} > 15 \text{ GeV}$
15. TPCRBL	Klasse, die auf technische Probleme hinweist
16. MUCAND	Myonkandidaten : Ein Myonkandidat ist gefunden worden.
17. PTMISS	Ereignis mit großem fehlendem Transversalimpuls : $p_{T_{miss}} > 10 \text{ GeV}$
18. HIGHET	Ereignis mit großer transversaler Energie : $E_T > 20 \text{ GeV}$
19. VDMETG	Klasse für VDM-Ereignisse mit Signal im Elektron Tagger : $E_{T_{tagger}} > 4 \text{ GeV}$ und mindestens eine gute Spur
20. ZVTXS1	Klasse für $\gamma p$ -Studien : Mindestens eine gute Spur und ein ZVTX-SIG1-Triggersignal des L1-Triggers.
21. NOT_CL	Nicht klassifiziert : Ereignisse, die nicht in eine der Klassen 0-21 fallen.

Tabelle 6.3: Klassen der L5-Ereignisklassifizierung. Die Klassen 16, 17 und 18 werden nicht als Physikklassen verstanden. Sie sollen nur zusätzlich zu anderen Signaturen verwendet werden.

### 6.1.3 Physik–Daten

Die ersten Tests, Neuronale Netze mit realen Daten zu trainieren, bestanden darin, die zuvor dargestellten Untergrundklassen und die in Abb. 6.1 dargestellte “Physik (L5)”–Klasse zu verwenden. Trotz guter Erfolge zeigte sich, daß mit diesen Trainingsmengen die Rate nur etwa um den Faktor zwei gesenkt werden konnte, denn L5 klassifizierte etwa die Hälfte aller Ereignisse als Physikandidaten. Es ist offensichtlich, daß der “Physik (L5)”–Datensatz hochgradig mit Untergrundereignissen kontaminiert war, die ja eigentlich auch herausgefiltert werden sollten.

Aus diesem Grund wurde eine neue Strategie entwickelt, die nur hochreine Physikdatensätze als Trainingsbeispiele verwendete. Da bis August 1992 bei H1 erst zwei Physikklassen mit nennenswerter Statistik zur Verfügung standen, wurden diese für die Untersuchungen eingesetzt. Die erste Klasse bestand aus 228 Ereignissen der tiefinelastischen Streuung (Deep Inelastic Scattering, DIS), die zweite Klasse aus 289  $\gamma p$ –Ereignissen. Letztere sind Wechselwirkungen, bei denen das Elektron ein quasireelles Photon abstrahlt, das dann mit dem Proton wechselwirkt. In Tabelle 6.2 sind die wichtigsten Größen der Physik– und Untergrunddatensätze sowie die Selektionskriterien aufgeführt.

### 6.1.4 Triggergrößen

Durch die realen Gegebenheiten des Experiments war es nicht möglich, dieselben Triggergrößen wie bei den Monte-Carlo–Untersuchungen zu verwenden. Da aber gezeigt werden sollte, daß mit den momentan erhältlichen Triggerrohdaten eine Separation von Physik und Untergrund möglich ist, wurde auch keine Simulation anderer – noch nicht existierender – Subtriggergrößen durchgeführt. Insgesamt standen 24 Triggervariablen zur Verfügung, wovon sich 11 Größen auf Energiemessungen bezogen und die restlichen aus Spurinformatoren bestanden. Die beim Training verwendeten Triggergrößen sind in Tabelle 6.4 beschrieben.

## 6.2 Netzwerktraining und Resultate

### 6.2.1 Netztopologie und Training

Nach dem Vorbild des “Merged–Netzes” wurde eine einfache Netzwerkarchitektur gewählt, da nur wenige Klassen zu separieren waren. Ein dreischichtiges Multilayer–Feedforward–Netz mit 24 Eingangsneuronen, den Größen aus Tabelle 6.4 entsprechend, 30 verborgenen Neuronen und einem Ausgabeneuron wurde verwendet. Die Menge der Untergrundereignisse wurde durch die “Untergrund (L5)”– und “Pilot–Bündel”–Datensätze gebildet. Die Datensätze der tiefinelastischen Streuereignisse (DIS) und der  $\gamma p$ –Reaktionen standen als Physiktrainingsmenge zur Verfügung. Wie auch im Beispiel der Monte-Carlo–Untersuchungen wurden die Datensätze jeweils in Trainings– und Testmenge aufgeteilt, um die Generalisierungsfähigkeit zu testen. Dies war insbesondere dadurch gefordert, daß diesmal die Statistik der Physikereignisse limitiert war und so nicht a priori von einer guten Generalisierung ausgegangen werden konnte.

Name	Beschreibung
1. ZVTX-max	Zahl der Einträge im Maximum-Bin des Z-Vertex Histogramms
2. ZVTX-sum	Gesamtzahl der Einträge im Z-Vertex Histogramm
3. ZVTX-qua	= 1 , wenn keine Einträge außerhalb eines Fensterbereichs von 3 benachbarten Bins, ansonsten = 0
4. ZVTXsig1	“alte” Z-Vertex Signifikanz: $\frac{n_{Max}-n_{Bgr}}{\sqrt{n_{Bgr}}}$ mit $n_{Bgr} = \frac{n_{summe}-n_{Max}}{15}$
5. ZVTXsig2	“neue” Z-Vertex-Signifikanz: $\frac{\sqrt{n_{Max}-2.5}}{n_{Bgr} + \frac{1}{15}}$
6. Ncoinc-i	Anzahl der Pad-Koinzidenzen in 16 $\Phi$ Sektoren für die CIP, nur die letzten 20 Pads in $-z$ -Richtung werden gezählt
7. Ncoinc-o	dasselbe für die COP, nur die letzten 10 Pads in $-z$ - Richtung werden gezählt
8. Peak-pos	Position des Maximums im Z-Vertex Histogramm (0..15) -1 bei leerem Histogramm
9. Eele-tag	Deponierte Energie im Elektron-Detektor des Luminositätssystems
10. Epho-tag	Deponierte Energie im Photon-Detektor
11. Veto-tag	= 1 wenn der Veto-Zähler vor dem Photon-Detektor anspricht
12. Nbpclhit	Zahl der Nachweisebenen in der BPC mit mindestens einem Treffer
13. Nbpcl2hit	Zahl der Nachweisebenen in der BPC mit mindestens zwei Treffern
14. BSET-Etot	Gesamtenergie im BEMC oberhalb der zugehörigen Schwelle
15. BSET-ECLU1	Cluster-Energie im BEMC oberhalb der Cluster-Schwelle 1
16. BSET-ECLU2	Cluster-Energie im BEMC oberhalb der Cluster-Schwelle 2
17. BSET-ECLU3	Cluster-Energie im BEMC oberhalb der Cluster-Schwelle 3
18. Veto-Tof	= 1 bei Koinzidenz zwischen erster und zweiter Veto-Wand im Zeitfenster für Untergrund, Veto-Wände werden mit logischem “ODER” verknüpft, ohne geometrische Korrelation
19. Posi-Tof	wie 18. für das Physik-Zeitfenster
20. Et	Transversale Energie aus den Bigtower- $\Theta$ -Bins 0...13
21. Etmis	Fehlende transversale Energie aus den Bigtower- $\Theta$ -Bins 0...13
22. E-CB	Energie des Central-Barrel-Bereichs (Bigtower- $\Theta$ -Bins 9...13)
23. E-FB	Energie des Forward-Barrel-Bereichs (Bigtower- $\Theta$ -Bins 6...8)
24. E-IF	Energie des Inner-Forward-Bereichs (Bigtower- $\Theta$ -Bins 0...5)

Tabelle 6.4: Triggergrößen für das Training des Neuronalen Netzes.

Das Training des Neuronalen Netzes verlief analog zum Fall der Monte-Carlo-Daten, es wurden zufällig Physik- und Untergrundereignisse aus den beiden Trainingsmengen entnommen und dem Netz präsentiert. Für die Lernparameter  $\alpha$  und  $\eta$  wurden die in Tabelle 5.6 aufgeführten Werte verwendet. Als Aktivierungsfunktion wurde

$$g(x) = \tanh(x) \quad \text{mit} \quad g(x) \in [-1, +1] \quad (6.1)$$

eingesetzt. Als gewünschter Wert des Ausgabeneurons wurde für Untergrund  $y = -1$  und für Physik  $y = +1$  vorgegeben. Der Grund für die Wahl einer anderen Sigmoidfunktion liegt in einer damit erzielbaren Beschleunigung der Konvergenz des Lernprozesses [40].

### 6.2.2 Ergebnisse

In Abbildung 6.2 sind die Ergebnisse für die beiden untersuchten Untergrundklassen dargestellt. Wie in Abb. 5.5 sind hier die Werte  $y$  des Ausgabeneurons auf der Abszisse aufgetragen, und die Häufigkeit ihres Auftretens ist auf der Ordinate histogrammiert. Es ist deutlich erkennbar, daß das Neuronale Netz eine sehr gute Zuordnung der Ereignisse zum Wert  $y = -1$  realisiert. Weniger als 2% Ereignisse wird falsch klassifiziert (rechte Prozentangabe in den Teilbildern), bezogen auf einen Schnitt bei  $y = 0$ . Die linke Prozentzahl beziffert den Anteil der korrekt klassifizierten Ereignisse, die Differenz von 100% bei der Summe beider Werte ist auf Rundungsfehler zurückzuführen.

In der linken Hälfte von Abbildung 6.2 sind die Ergebnisse für die Trainingsmenge, in der rechten Hälfte die Resultate für die Testmenge zu sehen. Sie verdeutlichen, daß die Generalisierung des Netzes einen hohen Grad erreicht hat. Wie bei den Monte-Carlo-Ergebnissen ist der Effekt einer verbreiterten Verteilung zu beobachten, der sich aber nicht auf die Klassifizierungseffizienz auswirkt, denn diese wird durch den Schnitt bei  $y = 0$  bestimmt.

In Abbildung 6.3 werden die Ergebnisse für die beiden Physikklassen präsentiert, die einen Ausgabewert von  $y = +1$  generieren sollten. Im Fall der tiefinelastischen Ereignisse wurde eine sehr gute Abbildung erreicht, und man gelangt zu einer Selektionseffizienz von fast 100%.

Auch auf der Trainingsmenge der  $\gamma p$ -Ereignisse wird eine Effizienz von 100% erreicht. Größere Probleme traten jedoch bei der Generalisierung auf, die mit einer Effizienz von unter 90% hinter den sehr guten Trainingsresultaten zurückliegt. Dies ist sicher ein Indiz für die limitierte Statistik der Trainingsbeispiele. In Anbetracht der großen Ähnlichkeit dieser  $\gamma p$ -Physik mit Beam-Gas-Reaktionen in der Wechselwirkungszone ist das Ergebnis jedoch recht ermutigend.

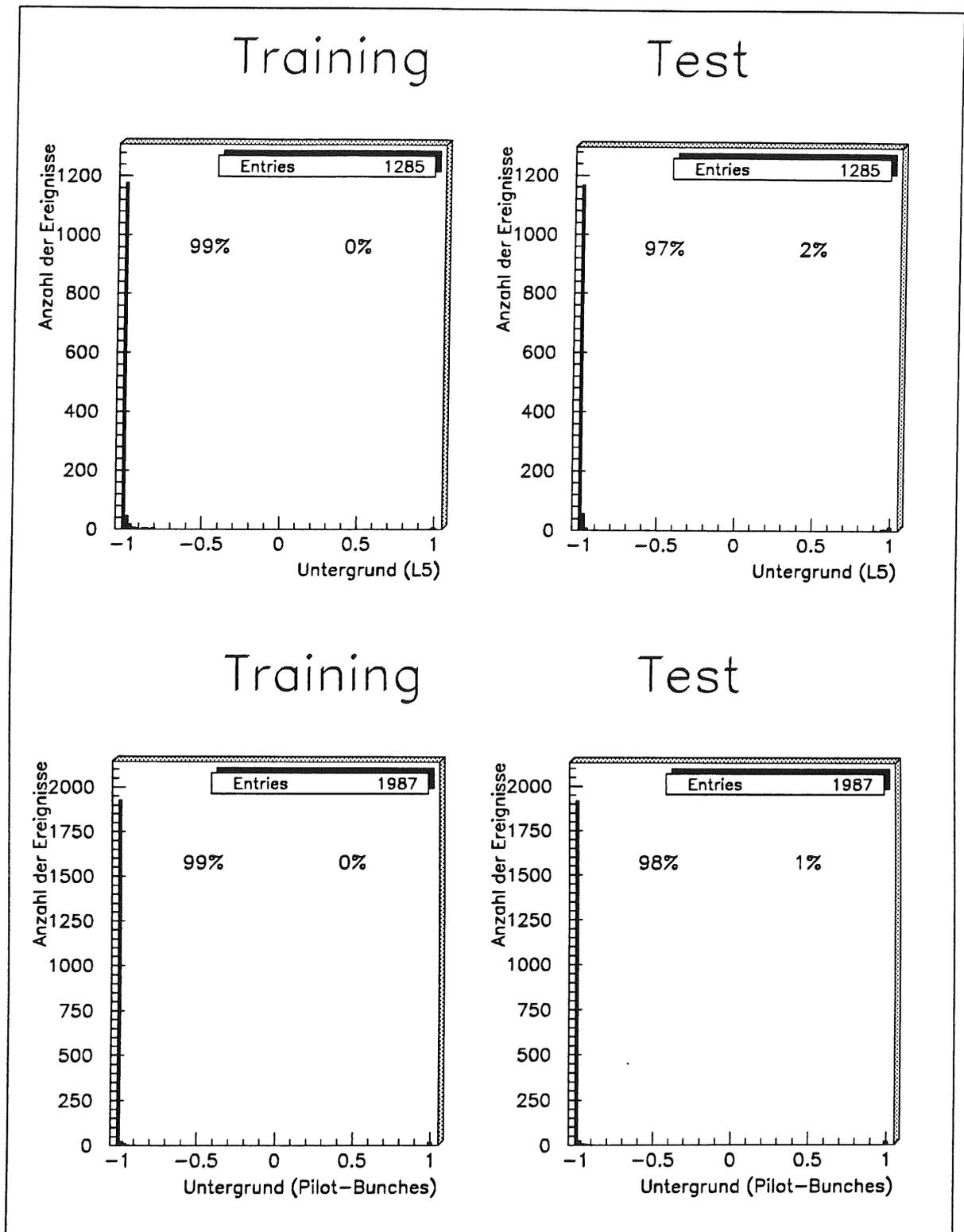


Abbildung 6.2: Ergebnisse des Netzwerktrainings für die Untergrundklassen (L5) und Pilot-Bündel. Die Abszisse beschreibt die Werte am Ausgabeneuron gegen die die Häufigkeit von Ereignissen, die diesen Wert produzieren, auf der Ordinate histogrammiert ist.

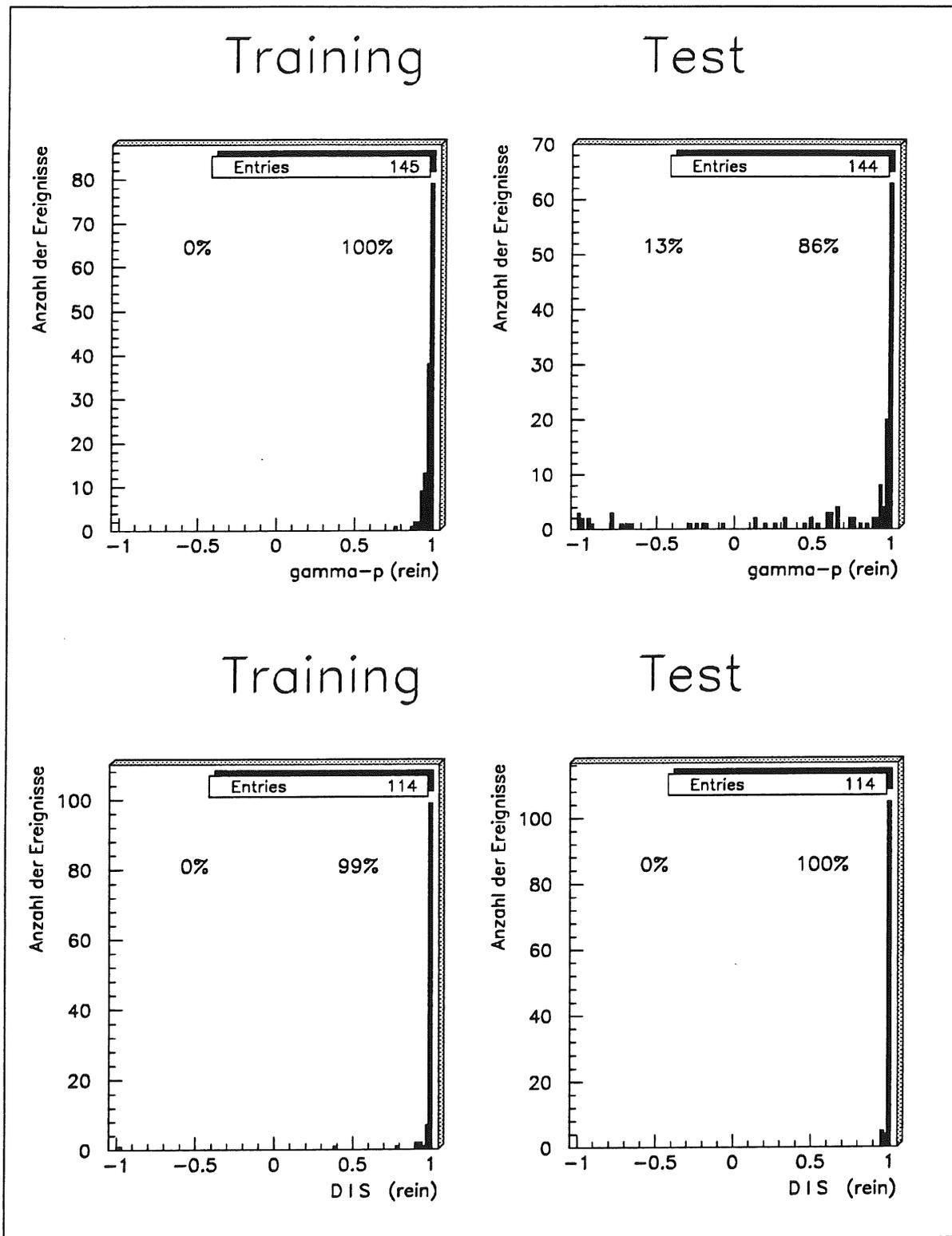


Abbildung 6.3: Ergebnisse des Netzwerktrainings für die Physikklassen DIS und  $\gamma p$ . Die Abszisse beschreibt die Werte am Ausgabeneuron gegen die die Häufigkeit von Ereignissen, die diesen Wert produzieren, auf der Ordinate histogrammiert ist.

### 6.2.3 Eindimensionale Verteilungen

Um ausschließen zu können, daß auch allein durch konventionelle eindimensionale Schnitte dieselben Klassifizierungen erreicht werden können, wurden die Verteilungen der einzelnen Triggervariablen untersucht. Es wurde versucht, die optimalen Schnitte in einer Variablen zu bestimmen, um eine bessere Physikeffizienz und gleichzeitig einen geringeren Untergrundanteil zu erhalten. Dies gelang in keiner der Variablen und wird anhand zweier signifikanter Triggergrößen in Abbildung 6.4 dargestellt. Die Prozentangaben beziehen sich auf den optimalen Schnitt, der 100% Physikeffizienz erreicht. Sie zeigen deutlich, daß bei diesem Schnitt eine große Anzahl von Untergrundereignisse (67%) akzeptiert werden müßte.

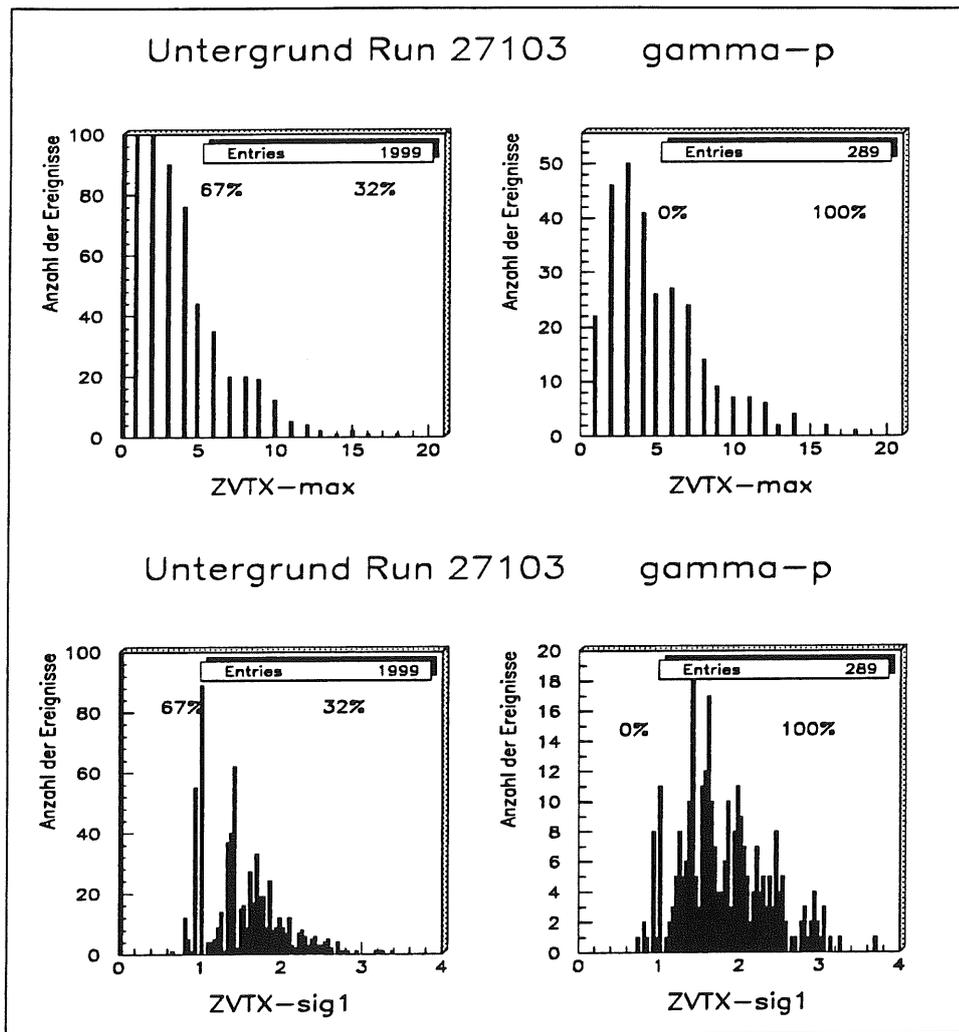


Abbildung 6.4: **Eindimensionale Verteilungen von Triggergrößen.** Eine Abtrennung des Untergrunds ist durch eindimensionale Schnitte nicht mit der gleichen Effizienz wie bei einem Neuronalen Netz erreichbar.

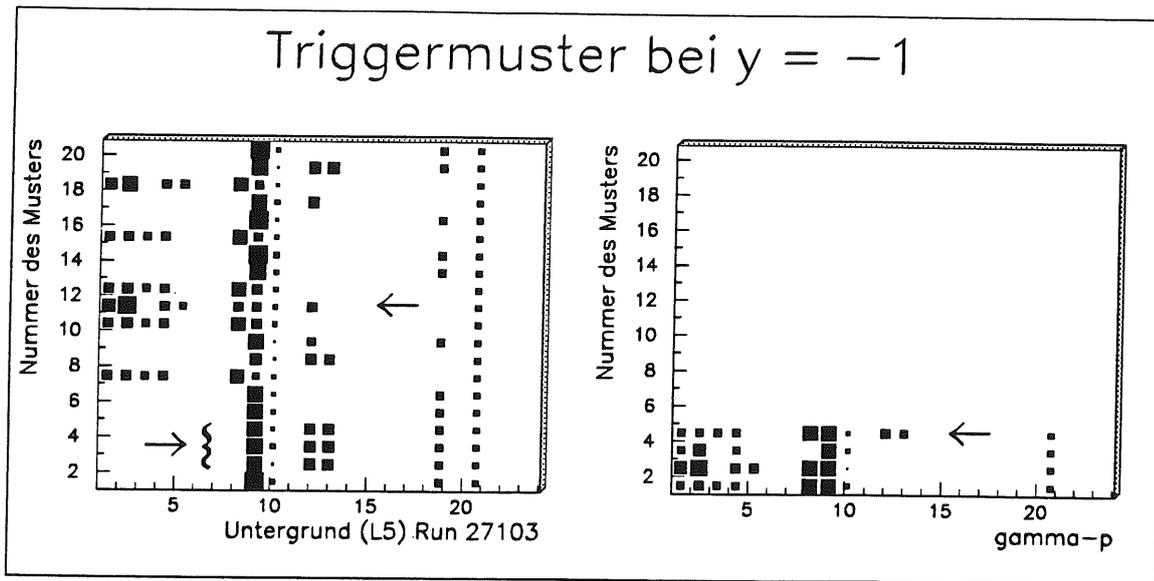


Abbildung 6.5: Triggermuster, die einen Wert von  $y = -1$  am Ausgabeneuron erzeugen. Eine Zeile entspricht dem Triggermuster eines Ereignisses. Die Werte auf der Abszisse beziffern die einzelnen Triggervariablen. Die Fläche der Quadrate gibt ein Maß für die Größe der jeweiligen Variable. Pfeile bezeichnen Ereignisse, auf die im Text verwiesen wird.

### 6.2.4 Deutung der Netzwerkschnitte

Die Aufklärung der Vorgänge innerhalb des Netzwerks liefert ein besseres Verständnis der typischen Merkmale von Untergrund- und Physikereignissen. Um zu klären, weshalb einige Physikereignisse falsch klassifiziert wurden, d.h. einen Ausgabewert von  $y = +1$  erhielten, wurden exemplarisch Triggermuster (Vektor aus Triggergrößen) von  $\gamma p$ -Ereignissen mit "Untergrund (L5)"-Ereignissen verglichen. In Abbildung 6.5 wurden auf der Abszisse die Nummern der Triggergrößen (siehe Tab. 6.4) dargestellt, auf der Ordinate die Nummer des Ereignisses. Eine Zeile entspricht somit dem Triggermuster eines Ereignisses. Die Visualisierung der Muster macht deutlich, daß es mehrere Arten typischer Untergrundereignisse gibt (linkes Teilbild). Zum Beispiel gehören die Muster 7, 10, 12 und 15 zu einer Art, die hauptsächlich Z-Vertex-Informationen (Triggergrößen Nr.1-5 und 8) tragen, und die Muster 2, 3, 4 und 8 zu einer Art, für die Treffer in der BPC (Triggergrößen Nr.12 und 13), Energie im Elektronendetektor des Luminositätssystems (Triggergröße Nr.9) und fehlende Z-Vertex-Information signifikant ist.

Ein Vergleich mit  $\gamma p$ -Ereignissen zeigt die Ähnlichkeit der Muster dieser falsch klassifizierten Ereignisse, wie z.B. das zweite  $\gamma p$ -Muster und das elfte Untergrundmuster (Pfeile in Abb. 6.5). Diese Ähnlichkeit auf der Ebene der Triggermuster bedeutet jedoch nicht, daß sich auch die Ereignisse ähneln, wenn man sie mit dem *H1-Ereignis-Display* betrachtet.

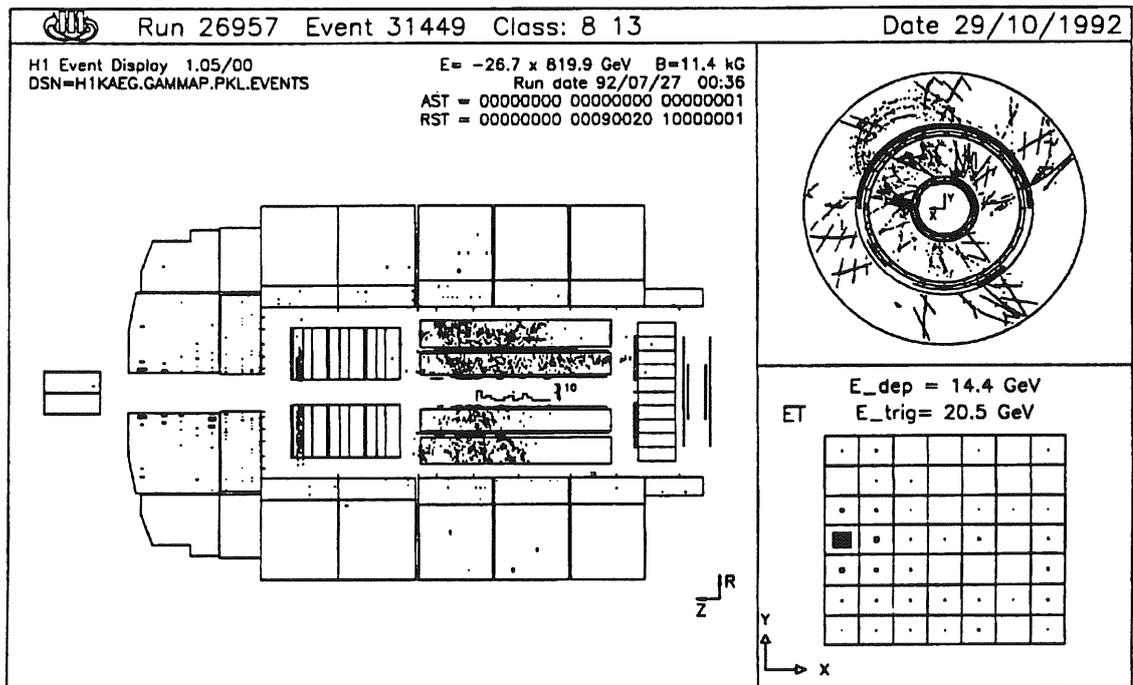
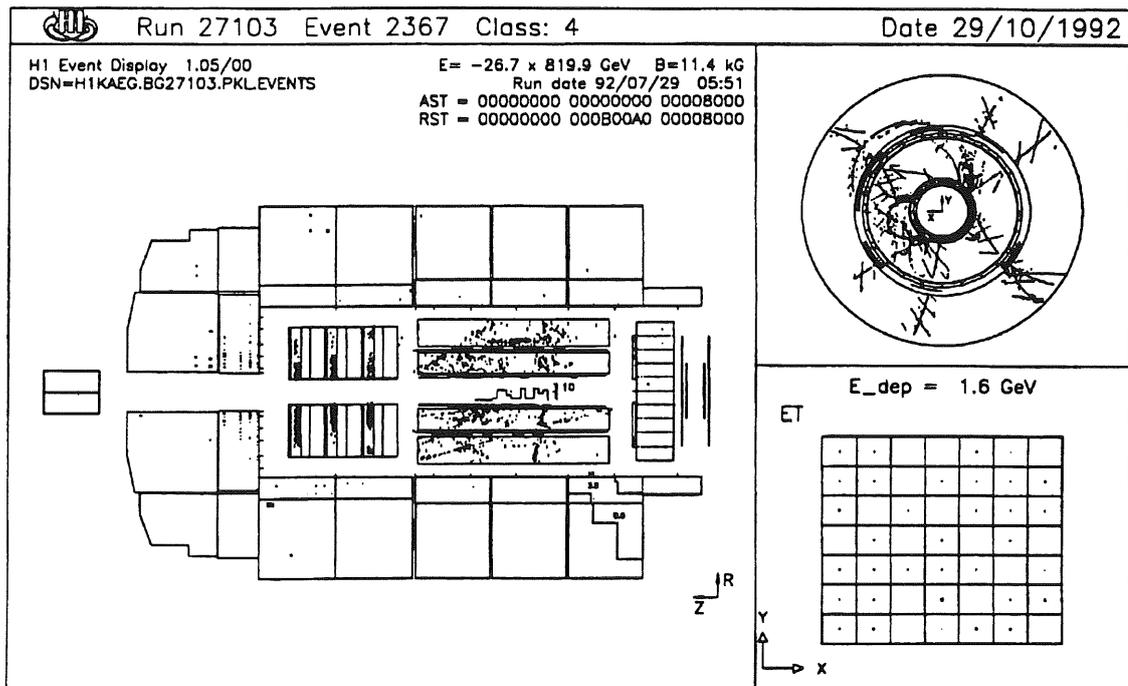


Abbildung 6.6: Vergleich von korrekt klassifiziertem “Untergrund (L5)”-Ereignis (oben) und einem  $\gamma p$ -Ereignis (unten), das vom Netz als “Untergrund” klassifiziert wurde.

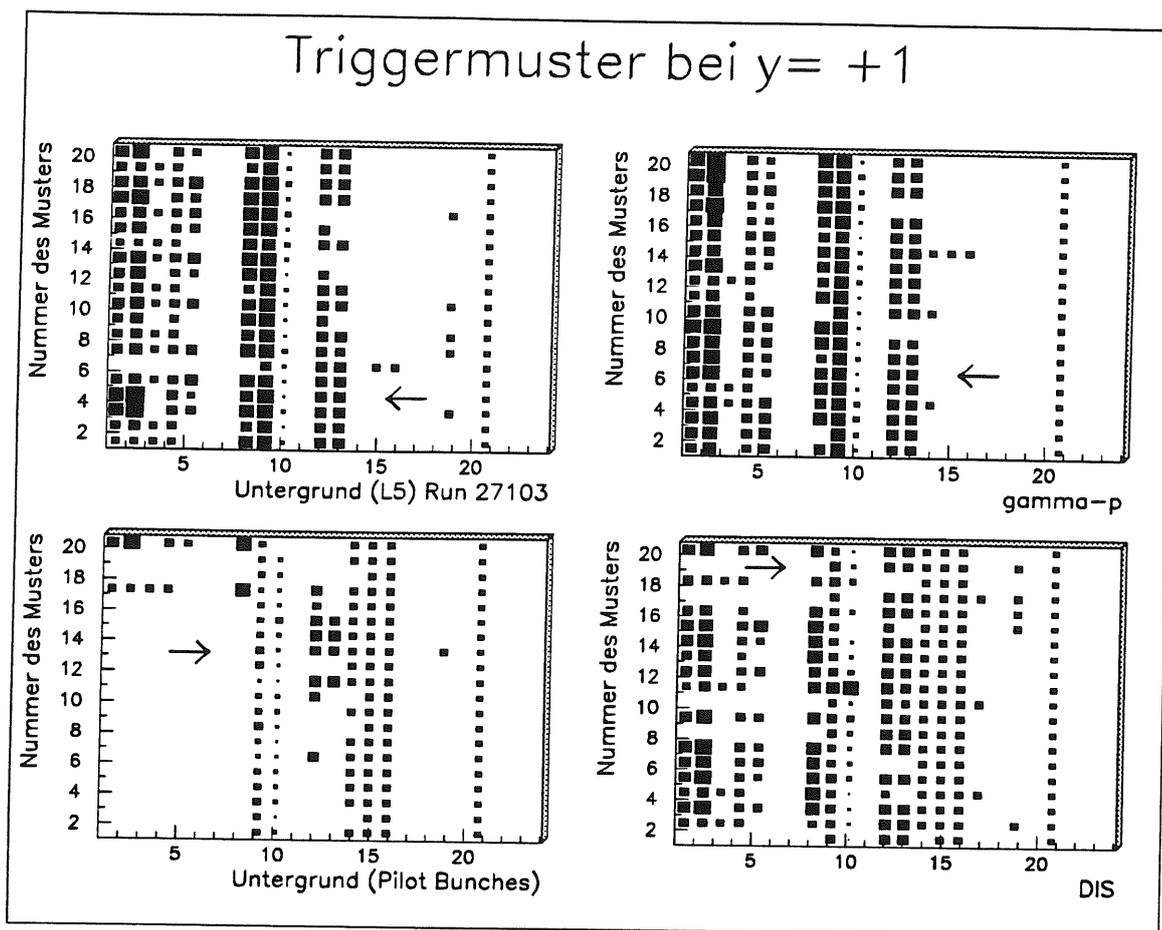


Abbildung 6.7: Triggermuster, die einen Wert von  $y = +1$  am Ausgabeneuron erzeugen. Eine Zeile entspricht dem Triggermuster eines Ereignisses. Die Werte auf der Abszisse beziffern die einzelnen Triggervariablen.

Im Fall des falsch klassifizierten  $\gamma p$ -Ereignisses erkennt man in Abbildung 6.6, daß die für die Erkennung dieser Physik wichtigen Triggergrößen, wie Treffer in der BPC, Energien im BEMC (Triggergrößen Nr. 14–16) oder signifikante Maxima im Z-Vertex-Histogramm, fehlen. Das heißt, daß durch die Reduktion auf Triggergrößen Information verloren gegangen ist und deshalb bei gewissen Ereignistypen eine starke Ähnlichkeit zu Untergrundereignissen eingetreten ist. Durch Hinzunahme von mehr Information, d.h. weiterer Triggergrößen, wäre eine Separation wieder möglich.

Auch die Untersuchung fehlklassifizierter Untergrundereignisse läßt die wichtigen Korrelationen erkennen, die vom Neuronalen Netz gefunden worden sind. In Abbildung 6.7 erkennt man die große Ähnlichkeit zwischen dem sechsten  $\gamma p$ -Muster und dem vierten "Untergrund (L5)"-Muster. Diese Ähnlichkeit erweist sich jedoch bei Betrachtung des gesamten Ereignisses im Ereignis-Display (Abb. 6.8) als Auswirkung der Reduktion auf Triggergrößen. Die typischen Kennzeichen von  $\gamma p$ -Physik liefert auch das Untergrundereignis, jedoch stammen die Einträge im Z-Vertex Histogramm von Treffern in CIP und COP, die durch ein Halo des Protonenstrahls erzeugt wurden.

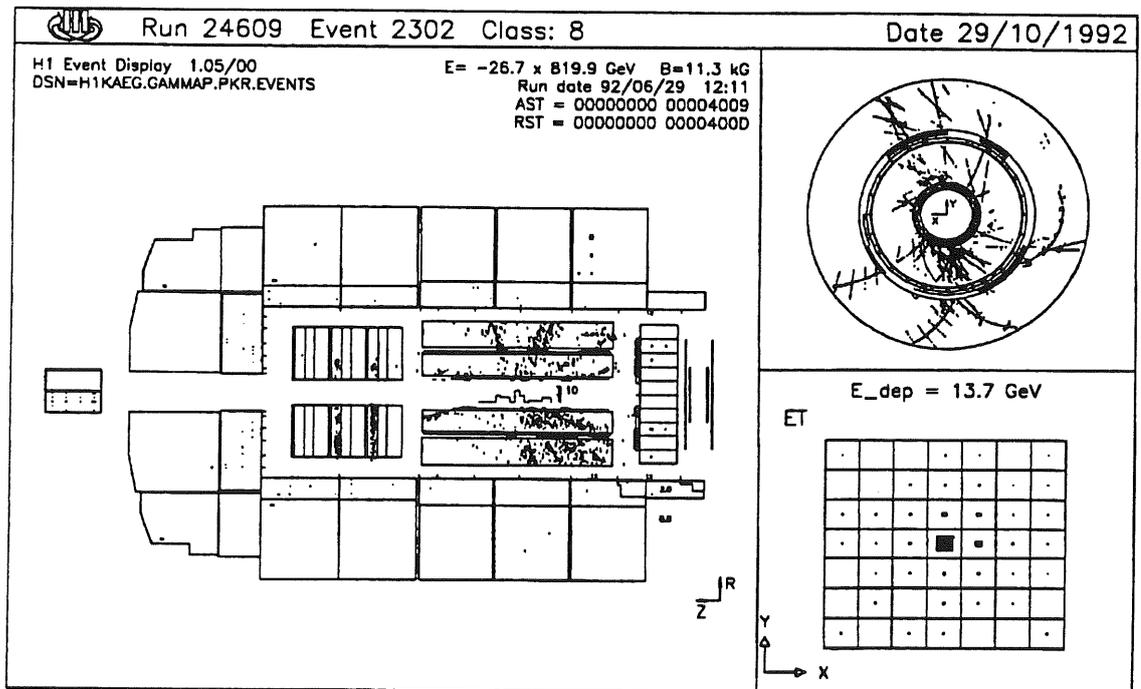
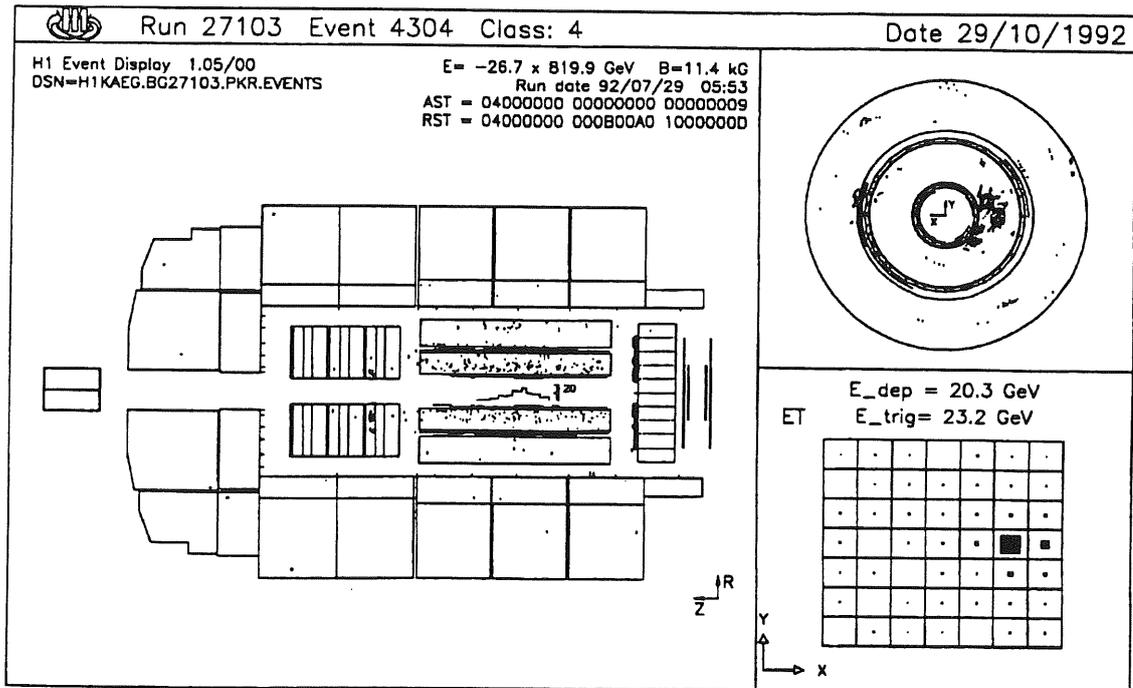


Abbildung 6.8: Vergleichende Darstellung eines vom Netz als  $\gamma p$ -Ereignis klassifizierten "Untergrund (L5)"-Ereignisses (oben) und eines korrekt klassifizierten  $\gamma p$ -Ereignisses (unten).

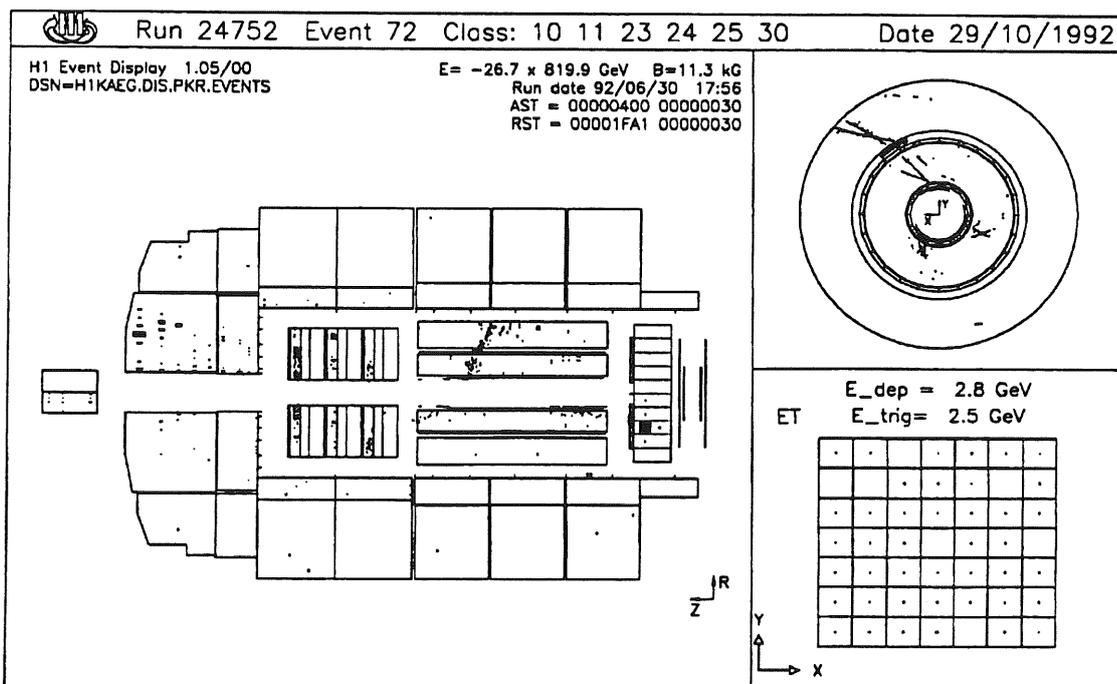
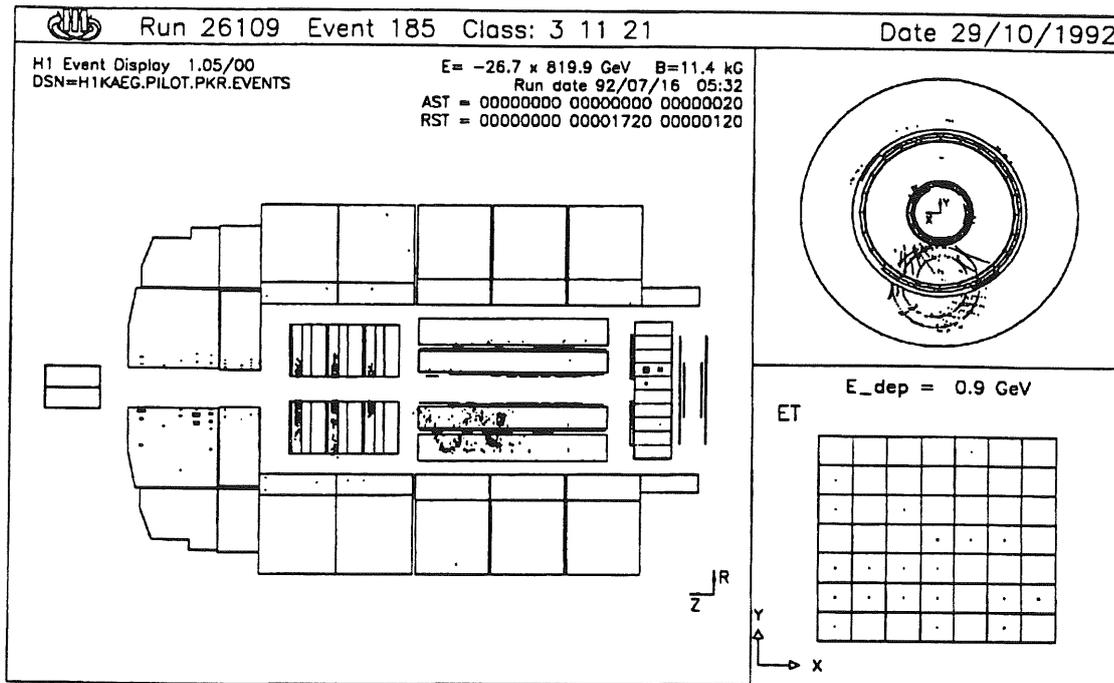


Abbildung 6.9: Vergleichende Darstellung eines vom Netz als “DIS” klassifizierten “Pilot-Bündel”-Untergrundereignisses (oben) und eines richtig klassifizierten DIS-Ereignisses (unten).

Dieses Halo erzeugte auch Treffer in der BPC, die koinzident mit einer wohl zufälligen Energiedetektion im Luminositätszähler waren. Durch die Verwendung eines  $r$ - $\Phi$ -Triggers zum Nachweis von Spuren in den Driftkammern wäre die richtige Klassifizierung des Untergrundereignisses vereinfacht worden. Dieser war jedoch zum Zeitpunkt der Datennahme noch nicht verfügbar.

In einem letzten Beispiel sollen das dreizehnte Muster aus dem Pilotuntergrund mit dem Muster 19 der DIS-Daten verglichen werden. Signifikant für tiefinelastische Streuereignisse sind Treffer in der BPC, Energiedepositionen im BEMC und Einträge im Z-Vertex-Histogramm. Wie man in Abbildung 6.9 sehen kann, war jedoch letztere Bedingung kein notwendiges Kriterium für DIS-Ereignisse, denn der Vertex konnte auch außerhalb der definierten Wechselwirkungszone liegen. Solch eine Ereignisstruktur kann von einem Pilot-Untergrundereignis vorgetäuscht werden, wenn Halo-Protonen sowohl das BEMC als auch die BPC aus der Protonrichtung treffen (siehe Abb. 6.9). Kollisionen, die nicht vom Z-Vertex kommen, sollten jedoch in Zukunft durch Erreichung der angestrebten kurzen Bündellängen nicht mehr auftreten und deshalb keine Gefahr mehr für eine Fehlklassifizierung darstellen.

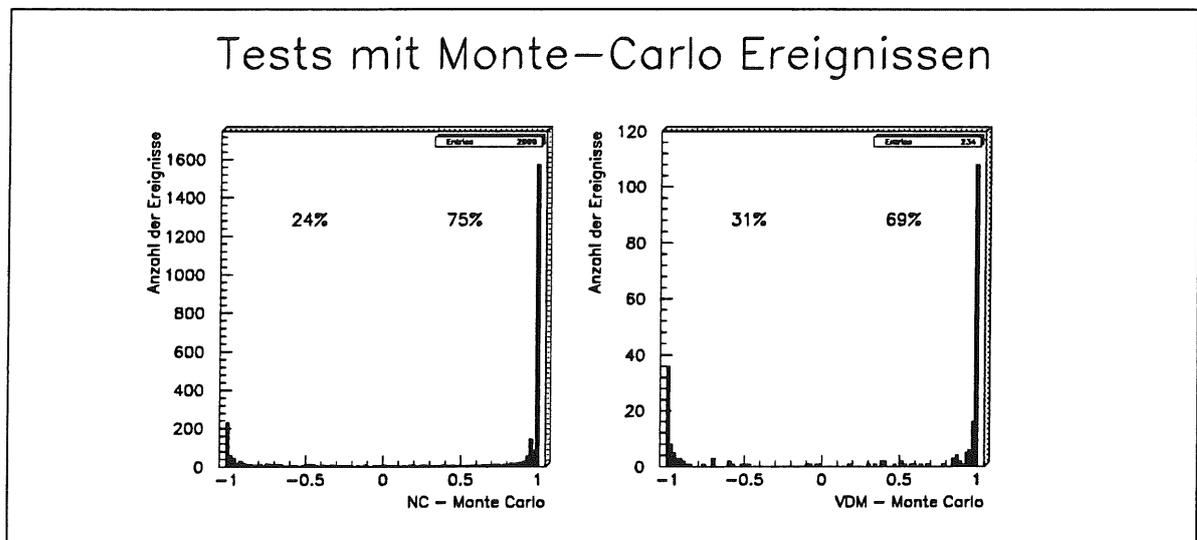


Abbildung 6.10: Test der Monte-Carlo-Daten mit dem auf reale Bedingungen trainierten Neuronalen Netz. Nur für die beiden Physikklassen, die in den realen Daten vertreten waren, d.h. NC-Ereignisse und  $\gamma p$ -Ereignisse nach dem Vektor-Dominanz-Modell (VDM), wurde ein Test durchgeführt.

### 6.3 Vergleich mit Monte-Carlo-Daten

Nachdem die Ergebnisse eines auf reale Daten trainierten Neuronalen Netzes in der Ausführungsphase eine gute Generalisierung aufgezeigt hatten, war es von grundlegendem Interesse, ob Ereignisse aus Monte-Carlo-Simulationen ebenfalls richtig klassifiziert werden, wenn man sie diesem Netzwerk präsentiert. Diese Situation stellte also

sowohl einen Test der Generalisierung des Netzes, als auch einen Test für die Realitätsnähe der Monte-Carlo-Daten dar.

Als Monte-Carlo-Datensätze fungierten 300 neutrale Stromereignisse, die durch das Programm LEPTO 5.2 [41] generiert wurden und 234  $\gamma p$ -Ereignisse, die von dem Monte-Carlo-Generator RAYPHOTON [42] erzeugt wurden. Die RAYPHOTON-Ereignisse basieren auf dem *Vektor-Dominanz-Modell* (VDM).

Alle Ereignisse wurden der Detektorsimulation H1SIM übergeben, welche das Ansprechen der Subdetektoren und die erzeugten Triggersignale simuliert. Zu den Signalen im Flüssig-Argon-Kalorimeter und im BEMC wurde elektronisches Rauschen addiert und anschließend ein  $2.5 \sigma$ -Schnitt durchgeführt. Die Ereignisse entstammen folgenden Datensätzen auf der DESY-IBM :

- NC : HERA02.LEPTO52.NC.MTB1.F001.H1SIM001.DISK
- VDM : HERA02.H1RAYVDM.C02.H1REC.B.A00

In Abbildung 6.10 sind die Ergebnisse für die beiden Physikklassen dargestellt. Im Vergleich zu den realen Daten werden geringere Effizienzen erreicht, aber eine gute Klassifizierung ist dennoch möglich ist. Zur Abnahme der Effizienz ist anzumerken, daß bei  $\gamma p$ -Reaktionen das Vektor-Dominanz-Modell nur einen Teil der  $\gamma p$ -Prozesse beschreibt. Dies bedeutet, daß das Monte-Carlo-Programm nicht hinreichend realistisch war.

Dieses Argument ist auch im Fall der neutralen Stromereignisse gültig, denn die bisherigen Daten enthielten nur Ereignisse mit kleinem  $Q^2$  ( $< 100 \text{ GeV}^2$ ) und kleinem  $x$  ( $< 0.001$ ). Die Monte-Carlo-Ereignisse enthalten auch sehr viele Ereignisse mit großem  $x$  und  $Q^2$  enthalten, welche eine deutlich andere Signatur im Detektor hinterlassen. Als kalorimetrische Information ist z.B. bisher nur Energie des BEMC verwendet worden, da die Impulsüberträge zu gering waren, um Energien im Flüssig-Argon-Kalorimeter zu deponieren. Dies führte zur besonderen Betonung von BEMC-Signalen, die aber in den Monte-Carlo-Ereignissen mit hohem  $Q^2$  nicht erzeugt wurden. In diesem Licht betrachtet sind diese ersten Vergleiche zufriedenstellend, da auch insbesondere die Monte-Carlo-Simulationsparameter noch nicht an die realen Daten angepaßt sind.

## 6.4 Stabilitätsuntersuchungen

### 6.4.1 Test der Untergrundstabilität

Um auszuschließen, daß das Neuronale Netz während der Trainingsphase eine "künstliche" Trennung einführt und unphysikalische Informationen verwendet, wurde der Untergrunddatensatz von Run 27103 in zwei Hälften aufgeteilt und versucht, diese von einem Neuronalen Netz separieren zu lassen. Dies bot gleichzeitig die Möglichkeit, einen Unterschied zwischen der ersten und der zweiten Hälfte des Runs zu erkennen. In Abbildung 6.11 wird deutlich, daß das Netz nicht in der Lage ist, die beiden Datensätze zu unterscheiden. Dies bedeutet, daß sich der Untergrund während dieses Runs nicht signifikant geändert hat.

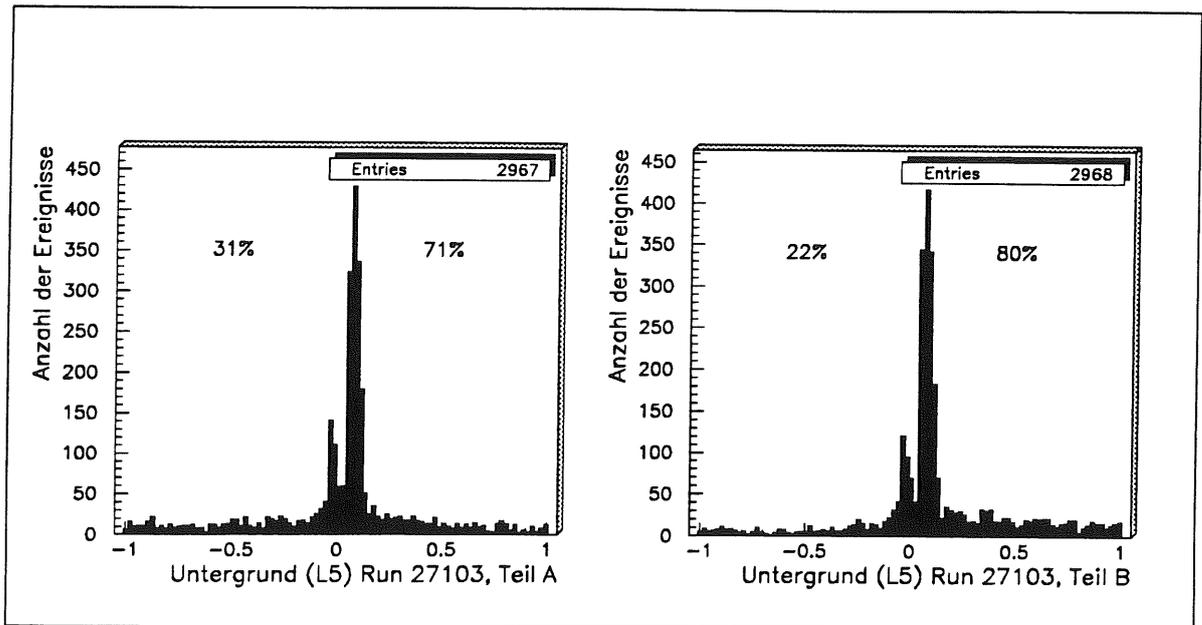


Abbildung 6.11: Ausgabe eines Neuronales Netzes, das Ereignisse aus der ersten Hälfte des Untergrund-Datensatzes von der zweiten Hälfte separieren sollte.

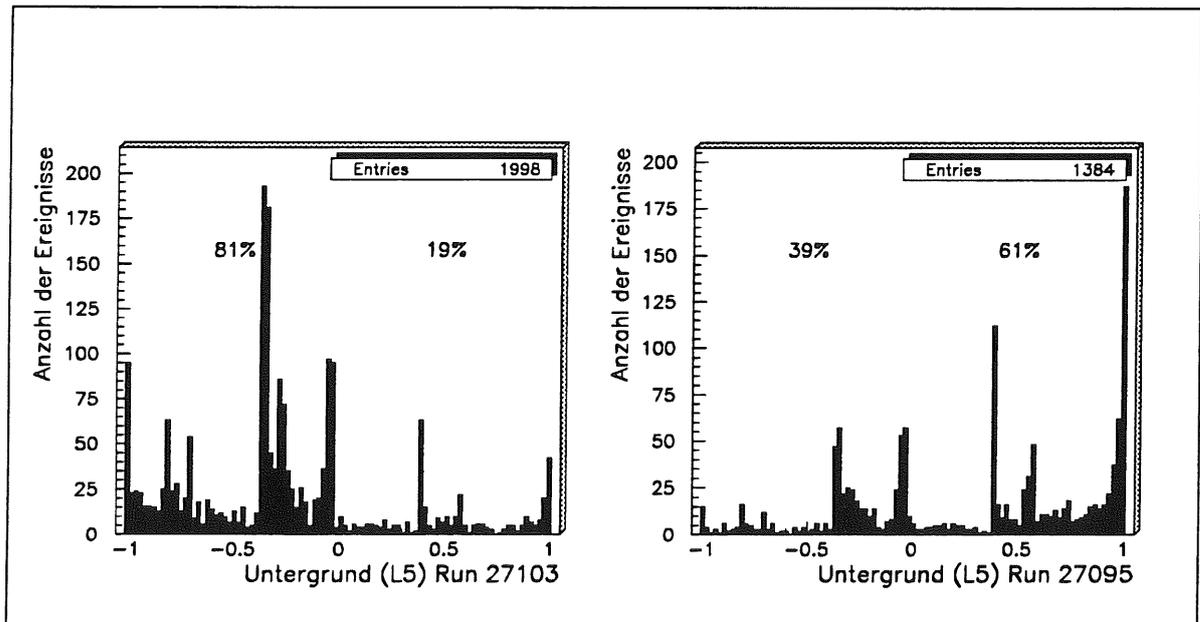


Abbildung 6.12: Ergebnis eines Neuronales Netzes, das auf Separation von Untergrund aus zwei verschiedenen Runs trainiert wurde. Es erfolgte keine wirkliche Trennung, lediglich eine Minimierung des globalen Fehlers durch Ausnutzung der verschiedenen Häufigkeiten bestimmter Untergrundarten.

Die Langzeitstabilität des Untergrunds, d.h. die Änderung von Run zu Run, wurde in einem zweiten Netzwerktraining untersucht, bei dem Untergrund von Run 27103 von Untergrund des Runs 27095 separiert werden sollte. In Abbildung 6.12 ist zu sehen, daß diesmal scheinbar eine Trennung möglich ist. Diese Trennung ist aber nicht auf einen wirklich unterschiedlichen Untergrund, sondern lediglich auf eine unterschiedliche Häufigkeit von bestimmten Arten des Untergrunds (siehe Abschnitt 6.2.4) zurückzuführen.

Dies wird durch die Maxima deutlich, die in beiden Runs an derselben Stelle aber mit unterschiedlicher Höhe auftreten. Das Neuronale Netz optimiert den globalen Fehler, indem es Maxima von Run 27103, die niedriger sind als in Run 27095, näher bei  $y = +1$  und die Maxima von Run 27095, die niedriger sind als in Run 27103, näher bei  $y = -1$  plaziert.

Ein auf solche Weise trainiertes Netz könnte im Experiment zur Überwachung des Untergrunds eingesetzt werden und wichtige Hinweise auf Veränderungen geben.

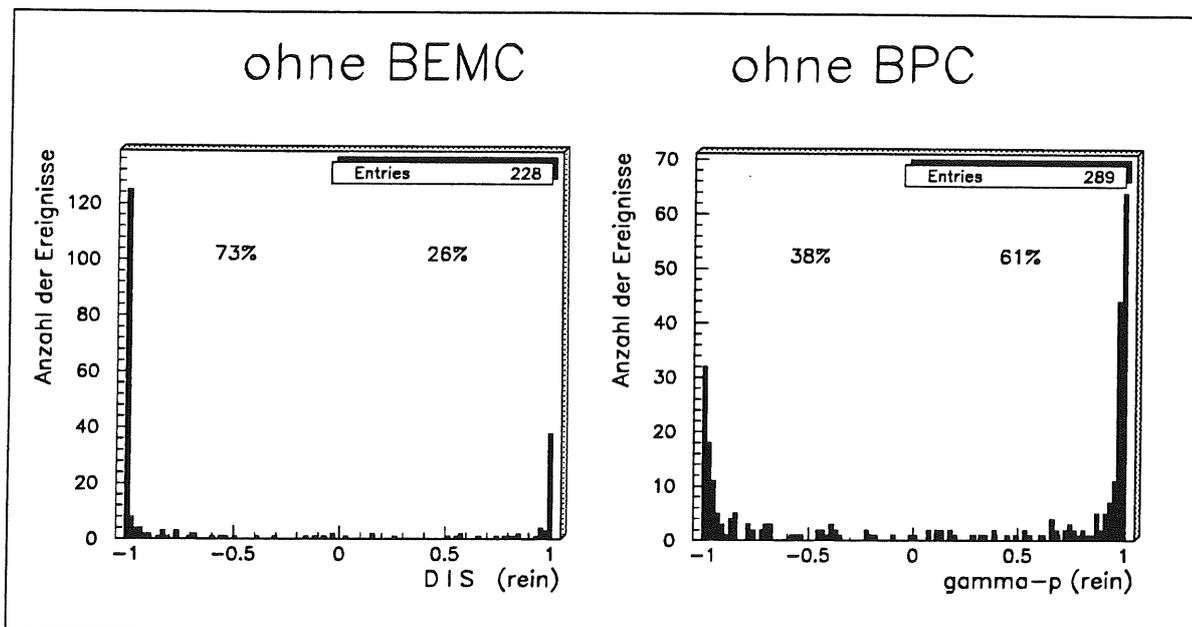


Abbildung 6.13: Ergebnisse für DIS-Physik bei abgeschaltetem BEMC (links) bzw. für  $\gamma p$ -Physik bei abgeschalteter BPC (rechts).

### 6.4.2 Stabilität bei Subdetektorausfall

Die Ergebnisse des Neuronalen Netzes wurden auf ihre Sensitivität bezüglich der Abschaltung oder des Ausfalls von Elektron- und Photodetektor, BEMC, zentralen Proportionalkammern (Z-Vertex) und rückwärtigen Proportionalkammern (BPC) untersucht. In Abbildung 6.13 sind die Einbußen bei Klassifizierung von  $\gamma p$ -Ereignissen nach Abschaltung der BPC-Information und von DIS-Ereignissen nach Abschaltung von BEMC-Informationen dargestellt.

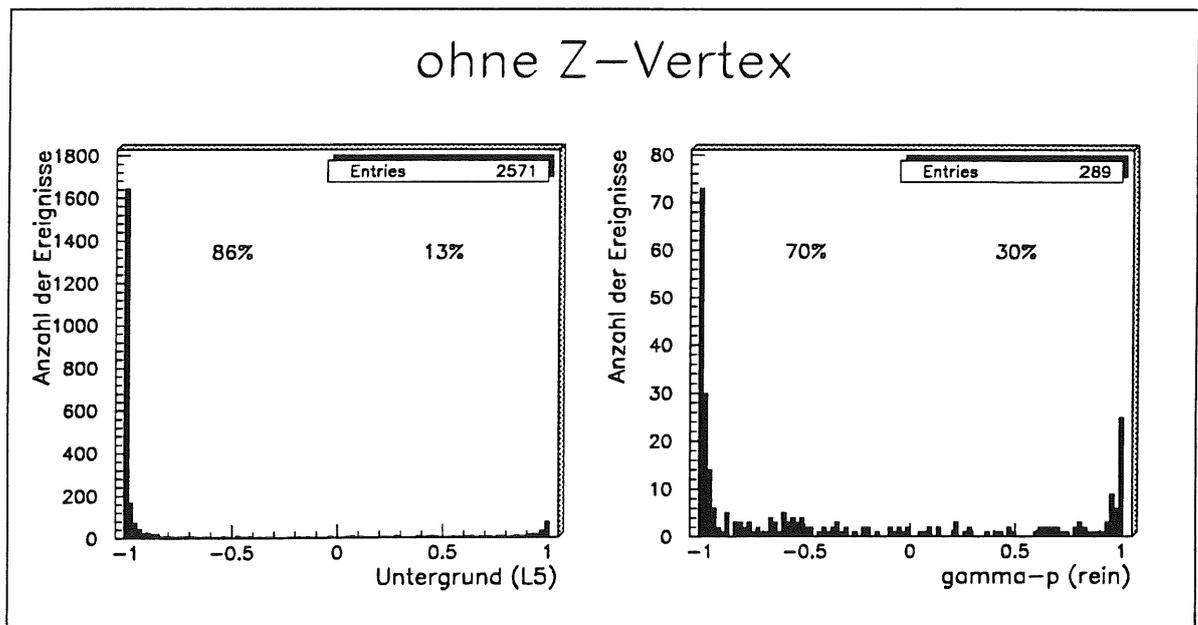


Abbildung 6.14: Ergebnisse für Untergrund (L5) und für  $\gamma p$ -Physik bei fehlender Z-Vertex-Information.

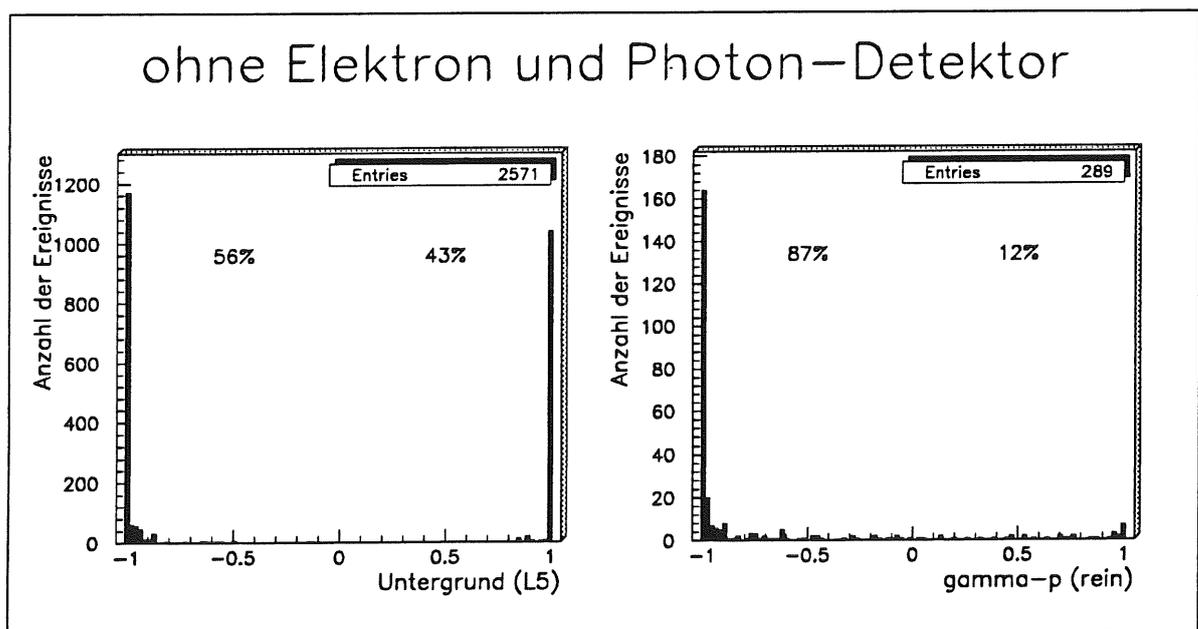


Abbildung 6.15: Ergebnisse für Untergrund (L5) und für  $\gamma p$ -Physik bei fehlendem Elektron- und Photondetektor des Luminositätssystems.

Abschaltung bedeutet in diesem Zusammenhang das Nullsetzen der Eingänge des Neuronalen Netzes bei den entsprechenden Variablen. Alle nicht dargestellten Klassen (Untergrund und Physik) blieben bei diesen Manipulationen unbeeinflusst. Die Effizienzverminderungen sind physikalisch erklärbar, da die fehlenden Detektorinformationen für die jeweilige Physikklasse ein signifikantes Merkmal darstellen.

In Abbildung 6.14 sind die Resultate für fehlende  $Z$ -Vertex-Information zu sehen. Man erkennt eine reduzierte Untergrundunterdrückung und einen Einbruch bei der Klassifizierungseffizienz für  $\gamma p$ -Physik. Dies betont die Wichtigkeit von Information über den Wechselwirkungsort bei der Erkennung von  $\gamma p$ -Ereignissen.

Deutlicher wirkt sich der Einfluß fehlender Information von Elektron- und Photondektektor aus (siehe Abb. 6.15). Die Effizienz bei der Erkennung sinkt von ursprünglich über 86% auf unter 13%, bei gleichzeitigem Anwachsen der fehlklassifizierten Untergrundereignisse (L5) von unter 4% auf über 40%. Der Effizienzverlust bei  $\gamma p$ -Physik ist verständlich, denn ein identifiziertes Elektron im Luminositätssystem ist essentiell für ein  $\gamma p$ -Ereignis. Die Vergrößerung der Zahl fehlklassifizierter Untergrundereignisse liegt darin begründet, daß diese Ereignisse nach dem Ausblenden der Luminositätsinformation nur noch BPC-Treffer aufweisen und im Physikzeitfenster der TOF-Wände liegen, somit einer typischen Form von  $\gamma p$ -Physikereignissen gleichen.

# Kapitel 7

## Hardware-Realisierung

### 7.1 Konzept zur Hardware-Realisierung

Die Ergebnisse der beiden vorangegangenen Kapitel haben gezeigt, daß die Anforderungen an Physikeffizienz und Untergrundunterdrückung von einem System parallel arbeitender Netzwerke, die auf eine bestimmte Physikklasse trainiert sind ("Komitee-Netzwerk"), erfüllt werden können. Diese modulare Architektur erlaubt eine effiziente Realisierung in Hardware, bei der für jede interessierende Physikklasse und die unterschiedlichen kinematischen Bereiche eigens trainierte Netzwerke eingesetzt werden können. Dadurch können auch problemlos Netze für neue Physikreaktionen hinzugefügt werden, ohne mit den bereits existierenden Netzen zu interferieren. Dieses Konzept der Modularität und Flexibilität führt zum grundlegenden Baustein des Neuronalen L2-Triggers, dem Mustererkennungsmodul (Pattern Recognition Module, PRM). Jedes PRM realisiert eines der Neuronalen Netze zur Trennung einer bestimmten Physikklasse von Untergrund.

Die PRM's erhalten ihre Daten von einem sogenannten Datenverteilungsmodul (Data Distribution Module, DDM), das die einlaufenden Informationen der einzelnen Subdetektoren sammelt und in geeigneter Weise auf die Eingänge der Neuronalen Netze verteilt (siehe Abb. 7.1). Die Werte der Ausgabeneuronen der einzelnen Netze werden nach angemessener Diskrimination mit einem Schwellenwert und möglicher Skalierung der Ausgangsrate durch ein logisches *ODER* verknüpft und formen das L2-Triggersignal.

Um genügend Flexibilität in Bezug auf Erweiterungen des Netzes zu erreichen, ist ein Netz mit 64 Eingangsneuronen, 64 versteckten Neuronen und einem Ausgabeneuron vorgesehen. Damit ist auch die Einbeziehung von Bigtower-Information realisierbar.

### 7.2 Chips für Neuronale Netzwerke

Das wichtigste Element eines PRM's ist ein Chip für schnelle Matrix-Vektor-Multiplikation, welche die Grundoperation eines Neuronalen Netzes bildet (siehe Glg. 4.7). Um ein Netz mit 64 Eingangsneuronen und 64 versteckten Neuronen in der verfügbaren Zeit von etwa 10  $\mu$ sec durchrechnen zu können, ist ein System er-

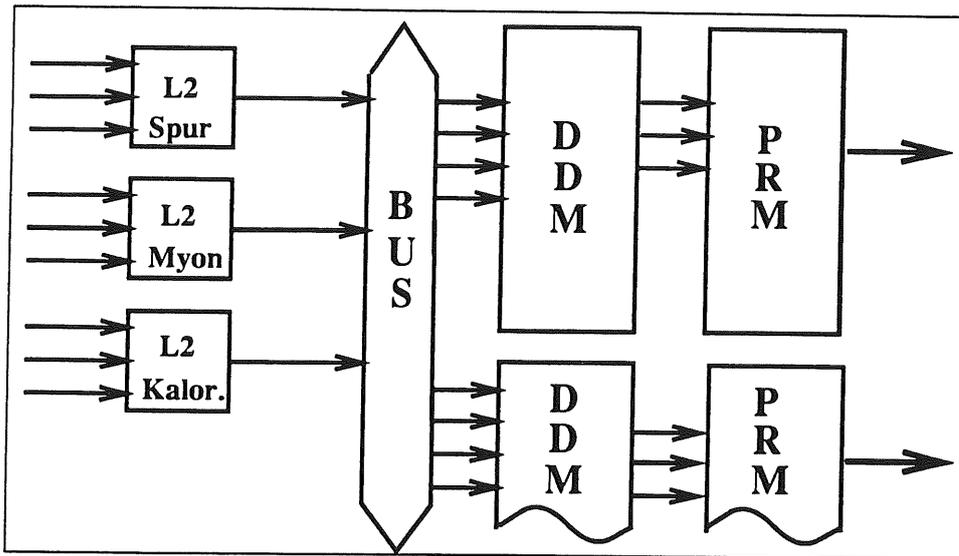


Abbildung 7.1: **Schema der Hardware-Realisierung.** Auf der linken Seite sind die L2-Triggerinformationen der Subdetektoren dargestellt, die mit zwei verschiedenen Taktfrequenzen an das DDM geliefert werden, welches die ausgewählten Daten über einen 10 MHz Bus an die PRM's verteilt.

forderlich, das eine Rechenleistung von mindestens 400 Millionen Verbindungen pro Sekunde besitzt. Aufgrund der Art der einlaufenden Information, sollte der Chip digitale Daten verarbeiten können. Eine Reihe von Bausteinen, die den Anforderungen genügen, sind bereits angekündigt, wie z.B.

- LNeuro 1.0 [43] (Philips)
- HANNIBAL [44] (British Telecom Laboratories)
- MA16 [45] (Siemens).

Alle Chips besitzen eine modulare Struktur und können so angeordnet werden, daß sie in der Lage sind, beliebig komplexe Netzwerke abzubilden. Der MA16-Chip von Siemens wurde bei gleicher Leistungsfähigkeit aller drei Chips aufgrund seiner Einfachheit und seines flexiblen Aufbaus gewählt.

Der Siemens MA16 ist ein 16 Bit Matrix-Vektor-Multiplizierer mit einer Taktfrequenz von bis zu 50 MHz (20 nsec pro Takt), auf 1  $\mu$ m CMOS Basis. Er beherbergt vier sogenannte Elementarketten, die parallel arbeiten. Jede Elementarkette repräsentiert vier Neuronen und ist mit internen "Dual-Port" Pufferspeichern ausgerüstet, die zwei 4x4 Matrizen (16er-Blöcke) aufnehmen, die miteinander multipliziert werden sollen. Eine dieser Matrizen beinhaltet die Gewichte, die andere repräsentiert vier Komponenten von vier verschiedenen Eingabevektoren. Diese Eigenschaft des Chips, das sogenannte *Pattern-Interleaving* kann jedoch in der Anwendung als Trigger nicht verwendet werden, da immer nur ein Muster (Ereignis) zu einer gegebenen Zeit vorliegt. Die zwei 4x4 Matrizen werden in 16 Taktzyklen miteinander multipliziert und den

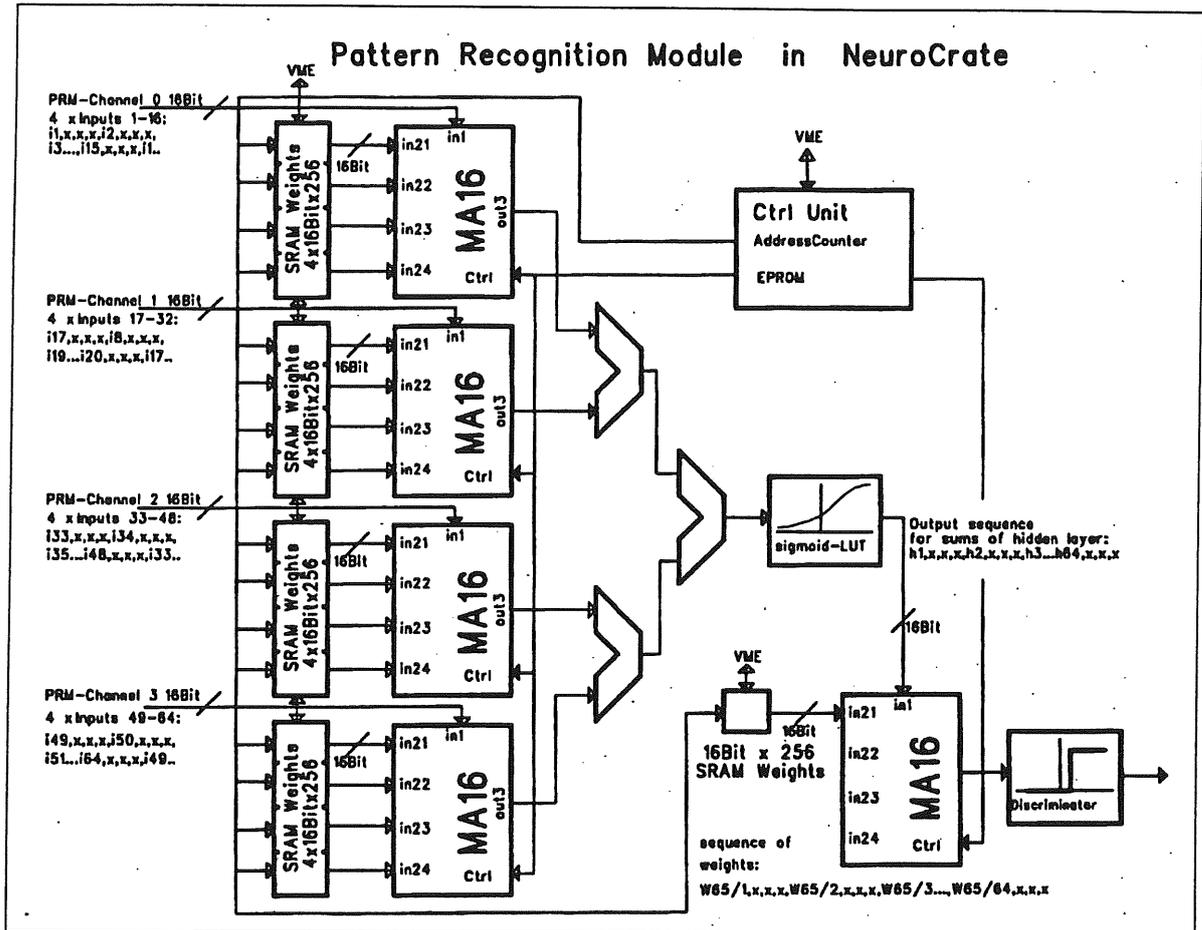


Abbildung 7.2: Schematische Darstellung des Mustererkennungsmoduls (PRM).

Ergebnissen der vorhergehenden Operationen addiert. Während dieses Vorgangs werden bereits die nächsten 16er-Blöcke in den internen Puffer geladen. Nach dem Laden der vier ersten 16er-Block-Paare benötigt eine Elementarkette nur noch  $16 \times (N/4)$  Taktzyklen für eine vollständige Berechnung der Aktivitäten von vier Neuronen mit  $N$  Eingabevektoren. Da vier Elementarketten auf einem MA16 untergebracht sind, kann ein Chip 16 Neuronen gleichzeitig berechnen.

### 7.3 Das Mustererkennungsmodul (PRM)

Da der MA16 im Grunde nur einen schnellen Matrix-Vektor-Multiplizierer darstellt, benötigt er eine externe Kontrolle, welche die Signale für die Gewichte und Eingabevektoren zur richtigen Zeit an die Eingänge des Chips anlegt. Die Gewichtsmatrizen werden in externen Speichern (SRAM) abgelegt. Zusätzlich wird eine "Look-Up-Table" benötigt, um die sigmoide Aktivierungsfunktion zu realisieren, mit der die Resultate der Multiplikation und Summation transformiert werden. Alle diese funktionalen Ele-

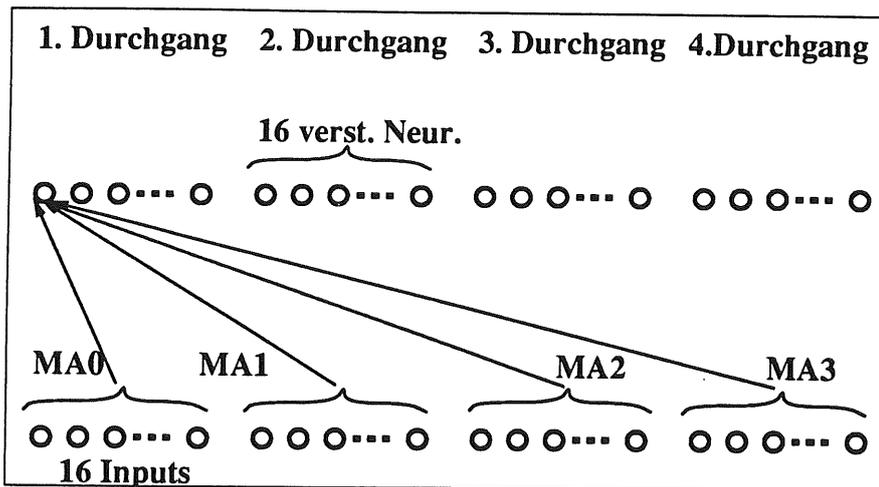


Abbildung 7.3: Ablauf der Berechnung des Neuronalen Netzes. Die vier MA16-Chips sind mit MA0 – MA3 bezeichnet. Jeder Chip ist an der Berechnung jedes versteckten Neurons beteiligt. In jedem Durchgang können 16 versteckte Neuronen berechnet werden.

mente, zusammen mit den MA16-Chips, bilden das *Pattern Recognition Module*. Ein schematisches Bild der logischen Elemente eines PRM sind in Abbildung 7.2 zu sehen. Es enthält die Hardware-Realisierung für ein Netz mit 64 Eingabewerten, 64 versteckten Neuronen und einem Ausgabeneuron. Mit dieser großzügig ausgelegten Topologie sind in einem Prototypen alle bisher simulierten Netzwerke eingeschlossen und es stehen noch Ressourcen für zukünftige Erweiterungen zur Verfügung.

Zur Erreichung der notwendigen Rechenkapazität sind vier MA16-Chips erforderlich, die jedoch nicht jeweils 16 Neuronen der versteckten Schicht vollständig berechnen, sondern je einer den vier 16er-Gruppen von Eingangssignalen zugeordnet ist. Dies bedeutet, daß jeder Chip an jedem versteckten Neuron "mitrechnet". Die Situation ist schematisch in Abbildung 7.3 dargestellt. Da auf jedem Chip gleichzeitig nur an 16 Neuronen mitgerechnet werden kann, werden zuerst die ersten 16 versteckten Neuronen von allen vier MA16-Chips berechnet (1. Durchgang), dann die Eingabemuster nochmals angelegt und die nächsten 16 versteckten Neuronen berechnet (2. Durchgang) und so fort.

Ein fünfter MA16 übernimmt die Berechnung des Ausgabeneurons. Eine detaillierte Berechnung der Latenzzeit bis zum Erhalt des Ergebnisses ergibt bei einer Taktfrequenz von 40 MHz, unter Berücksichtigung von 5 MA16-Chips, drei Addierern und zwei "Look-Up-Tables", etwa 11  $\mu\text{sec}$  (10.8  $\mu\text{sec}$  MA16 + 0.1 ... 0.5  $\mu\text{sec}$  für Addierer und "Look-Up-Tables"). Diese Geschwindigkeit ist ausreichend, da für das Anliefern der Daten eine Zeit von unter 10  $\mu\text{sec}$  vorgesehen ist und somit die Gesamtentscheidungszeit des L2-Triggers nicht überschritten wird.

## 7.4 Das Datenverteilungsmodul (DDM)

Die geplante Flexibilität des Neuronalen Netzes bezüglich der Verwendung von Triggerinformationen erfordert eine intelligente Lösung zur Auswahl und Aufteilung der angelieferten Subdetektorsignale. Die Daten für den L2-Trigger aus den sehr unterschiedlichen Subdetektoren werden auf zwei verschiedenen Bus-Systemen mit unterschiedlicher Taktfrequenz auf mehreren Datenkanälen zur Verfügung gestellt. Die Ankunftszeiten der Signale sind je nach Subdetektor und Triggerelektronik verschieden, so daß es von Vorteil ist, eine Verteilungsmatrix zu verwenden, welche die gewünschten Signale durch Auswahl der entsprechenden Eingangsleitungen und Zeitfenster festlegt. Die ausgewählten Signale müssen dann sequentiell in der richtigen Abfolge auf vier Kanäle verteilt werden, die den einzelnen MA16-Chips zugeführt werden. Durch Definition der Verteilungsmatrix wird eine extreme Flexibilität in der Bereitstellung der Daten ermöglicht. Die Zeit bis zum Eintreffen der Daten am DDM und die Zeit zur Verteilung summieren sich zu einer Zeitverzögerung von höchstens  $5 \mu\text{s}$ , so daß das Zeitlimit von  $9 \mu\text{s}$  eingehalten wird.

# Zusammenfassung

Im Rahmen dieser Arbeit wurde die Möglichkeit untersucht, ein Neuronales Netzwerk als L2-Trigger im H1-Experiment einzusetzen. Das Ziel war, eine Trennung physikalisch interessanter Ereignisse von Untergründereignissen durch Ausnutzung der Korrelationen zwischen L1-Triggergrößen zu erreichen.

Zum Verständnis der Arbeitsweise der hier verwendeten Multi-Layer-Feedforward-Netze wurden Simulationen in zwei Dimensionen, die zwei Triggervariablen entsprechen, durchgeführt. Netzwerke optimieren ihre Parameter, indem sie während des Lernvorgangs bekannte Eingangswerte auf bekannte Ausgabewerte abbilden. In der anschließenden Ausführungsphase berechnen sie die Ausgabe zu unbekanntem Eingabemustern (Generalisierung). Es wurde die Funktion eines versteckten Neurons, das eine Trennlinie in der Ebene der Eingangsgrößen festlegt, veranschaulicht und die Probleme dargestellt, die während des Optimierungsprozesses (Lernvorgangs) auftreten. Die Untersuchungen belegen, daß die Verwendung von Neuronalen Netzen mit drei Schichten für die beim Triggern auftretenden Klassifizierungsprobleme ausreichend ist.

Zur Untersuchung der Einsatzfähigkeit eines Neuronalen L2-Triggers wurden Studien sowohl mit Daten aus Monte-Carlo-Simulationen als auch mit realen Daten des Experiments durchgeführt. Für die Monte-Carlo-Studien wurden verschiedene Topologien und Architekturen von Netzwerken untersucht. Alle waren mit der Aufgabe konfrontiert, Physikereignisse aus Prozessen mit neutralem Stromaustausch, geladenem Stromaustausch und Boson-Gluon-Fusion zu erkennen und von Untergründereignissen aus Wechselwirkungen des Strahls mit Restgas und Elementen des Strahlführungssystems zu trennen.

Es konnte gezeigt werden, daß allein mit L1-Triggerinformation die Eingangsrate von etwa 1000 Hz, die der Standard-L1-Triggercocktail liefert, auf 100 Hz gesenkt und eine Effizienz relativ zu L1 von 96% - 99% bei der Erkennung der Physikereignisse erreicht werden kann. Ein L1-Trigger allein kann die Reduktion der Rate auf 100 Hz nur durch Anhebung der Schwellen und damit verbundenen größeren Verlusten von physikalisch interessanten Ereignissen erreichen.

Es wurden Tests zur Fehlertoleranz des Netzwerks durchgeführt, die darauf hinweisen, daß z.B. ein unerkannter systematischer Fehler in der Kalibration des Flüssig-Argon-Kalorimetertriggers nur geringe Auswirkungen auf die Klassifizierungseffizienz eines Systems aus L1 und L2 hat, jedoch zu starken Effizienzverlusten bei einem reinen L1-Trigger verursacht, der auf 100 Hz optimiert ist.

Für die Untersuchungen mit realen Daten wurde nur eine einzige Netzwerktopologie verwendet, die aus 24 Eingangsneuronen, 30 versteckten Neuronen und einem Ausgabeneuron bestand. Als Physikdatensätze dienten selektierte Ereignisse zur  $\gamma p$ -

Wechselwirkung und zur tiefinelastischen Streuung. Der zum Training des Neuronalen Netzes verwendete Untergrund bestand aus Ereignissen, die nach erfolgter Rekonstruktion von der verworfen wurden, und aus Ereignissen eines Protonenbündels, das kein Elektronbündel als Kollisionspartner besitzt (Pilot-Bündel). Die Eingabegrößen in dieser Simulation bestanden aus 24 L1-Triggersignalen.

Das Netz ermöglichte eine korrekte Klassifizierung für über 90% der  $\gamma p$ -Ereignisse und über 90% der tiefinelastischen Ereignisse, wobei gleichzeitig die Untergrundrate um den Faktor 20 reduziert wurde. Durch die Analyse ausgewählter falsch klassifizierter Ereignisse konnte die Arbeitsweise des Netzwerks erklärt und eine verbesserte Klassifizierung bei Hinzunahme detaillierterer Triggerinformation prognostiziert werden. Untersuchungen zur Stabilität des Netzes bzgl. des Ausfalls von Subdetektoren geben weitere Hinweise auf die Bedeutung und Korrelation einzelner Triggergrößen.

Ein Konzept zur Hardware-Realisierung eines Neuronalen L2-Triggers wurde vorgestellt, das auf der Verwendung eines Datenverteilungsmoduls und einer variablen Anzahl von Mustererkennungsmodulen, die ein "Komitee-Netzwerk" realisieren, basiert. Die Mustererkennungsmodule führen als grundlegende Operation Matrix-Vektor-Multiplikationen durch, die von einem speziellen Chip, dem MA16 von Siemens, übernommen werden. Die Geschwindigkeit des Gesamtsystems ist ausreichend, um innerhalb der vom Experiment geforderten Zeit von 20  $\mu\text{s}$  die L2-Triggerentscheidung zu fällen.

# Literaturverzeichnis

- [1] R. Hofstadter, Rev. Mod. Phys. 102, 1586 (1956).
- [2] M. Gell-Mann, Phys. Lett. 8 (1964) 214.
- [3] G. Zweig, CERN Report 8182/TH. 401 (1964) und CERN Report 8419/TH. 412 (1964)
- [4] J. D. Bjorken, E. A. Paschos, Phys. Rev. 179 (1969) 1547.
- [5] HERA – A Proposal for a Large Electron-Proton-Collider Beam Facility at DESY, DESY HERA 81-10 (1981).
- [6] J. Feltesse, Proc. of the 1989 Int. Symp. on Lepton and Photon Interactions at High Energies 13.
- [7] H1-Collaboration, Technical Proposal for the H1 Detector, H1 Collaboration, März 1986.
- [8] Technical Proposal of the ZEUS Collaboration, The ZEUS Detector, März 1986.
- [9] R. J. Ellison et al., The H1 Trigger, H1-Report, H1-11/90-157.
- [10] Deutsches Elektronen Synchrotron, Wissenschaftlicher Jahresbericht 1990, DESY, 1991.
- [11] Physics at HERA, Proceedings of the HERA Workshop, Vol. 1, DESY, Hamburg, 29.-30. Oktober 1991.
- [12] Physics at HERA, Proceedings of the HERA Workshop, Vol. 2, DESY, Hamburg, 29.-30. Oktober 1991.
- [13] Physics at HERA, Proceedings of the HERA Workshop, Vol. 3, DESY, Hamburg, 29.-30. Oktober 1991.
- [14] G. Ingelman et al., Deep Inelastic Physics and Simulation, in Proceedings of the HERA Workshop, Hamburg, Oktober 1987, Vol.1, S.3-18, 1987.
- [15] Particle Data Group, Review of Particle Properties, Physical Review D45, 1992.

- [16] K. P. Aicher, Möglichkeiten zur Abtrennung von Untergrundreaktionen bei der Untersuchung tiefinelastischer Elektron-Proton-Streuung am HERA-Speicherring, Diplomarbeit TU München, 1991.
- [17] Denisov et al., Nucl. Phys. B61 (1973), 62.
- [18] Tsai-Chu et al., Nuovo Cimento Lett. 20(1977), 257.
- [19] H1-Collaboration, Technical Progress Report 1990, H1 Collaboration, August 1990.
- [20] BEMC-Group, The H1 Backward Electromagnetic Calorimeter (BEMC), H1-Report, H1-08/92-233, 1992.
- [21] L. Larsson, Simulation neuronaler Netzwerke und Test des Z-Kammer-Triggers am H1-Detektor, Diplomarbeit, II. Institut für Exp.Physik, Uni Hamburg, Dezember 1991.
- [22] M. Atac et al., Self-Quenching Streamers, Nucl Instr. and Meth. 200 (1982) 345.
- [23] B. Koppitz, Geometrical Arrangement and Sizes of the H1 Chambers for Iron Instrumentation and Muon Detection, H1 Technical Report, H1 - TR 402, 1987.
- [24] J. Tutas, Persönliche Mitteilung.
- [25] J. Fent, C. Kiesling, P. Ribarics, A Level 2 Calorimeter Trigger using Neural Networks, H1-Report H1-04/91 - 172, 1991.
- [26] E. Schöneburg, N. Hansen, A. Gawelczyk, Neuronale Netzwerke - Einführung, Überblick und Anwendungsmöglichkeiten, Markt und Technik Verlag, 1990.
- [27] D. E. Rumelhart, J. L. McClelland, Parallel Distributed Processing, Volume 1: Foundations, MIT Press Cambridge, 1986.
- [28] D. E. Rumelhart, J. L. McClelland, Parallel Distributed Processing, Volume 2: Psychological and biological models, MIT Press Cambridge, 1986.
- [29] C. Peterson, T. Rögnvaldsson, Introduction to Artificial Neural Networks, Lund Universität Preprint, LU TP 91-23, 1991.
- [30] A. van Ooyen, B. Nienhuis, Improving the convergence of the Back-Propagation Algorithm, Neural Networks, Vol. 5, S.465-471, 1992.
- [31] R. O. Duda, P. E. Hart, Pattern Classification and Scene Analysis, John Wiley & Sons, 1973.
- [32] M. D. Richard, R. P. Lippman, Neural Network classifiers estimate bayesian a posteriori probabilities, Neural Computaion, S.461, 3 (1991).

- [33] K. Hornik, M. Stinchcombe, H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks*, S.551, 3 (1990).
- [34] H. White, Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings, *Neural Networks*, S.535, 3 (1990).
- [35] J. Fent, A. Gruber, C. Kiesling, H. Oberlack, P. Ribarics, Second Level Neural Network Trigger in the H1-Experiment, in "New Computing Techniques in Physics Research II, ed. D. Perret-Gallix, World Scientific, 1992.
- [36] H1-Collaboration, Dokumentationen zur H1 Simulations-Software H1SIM , Version 1.07.
- [37] A. De Roeck, Persönliche Mitteilungen.
- [38] L. Lönnblad, C. Peterson, T. Rönngvaldsson, Pattern recognition in high energy physics with artificial neural networks - JETNET 2.0, *Computer Physics Communications*, 70, 1 (1992).
- [39] M. Höhfeld, S. E. Fahlman, Probabilistic Rounding in Neural Network Learning with Limited Precision, in "Proc. of the 2. Int. Workshop on Microelectronics for Neural Networks, Munich, 1991.
- [40] C. Peterson, E. Hartmann, Explorations of the Mean Field Theory Learning Algorithm, *Neural Networks*, Vol. 2, S.475-494, 1989.
- [41] G. Ingelmann, LEPTO 5.2 Program Manual ,unpublished, 1988.
- [42] N. H. Brook et al., RAYPHOTON Version 2.0: An Interface for HERA Photo-production Physics, in [13], S.1453-1462.
- [43] M. Durantou, J. A. Sirat, Learning on VLSI: A general purpose digital neurochip, *Philips J. Res.* 45, 1 (1990).
- [44] D. J. Myers et al., Proceedings of the 2nd Int. Conf. on Microelectronics for Neural Networks, Munich, Germany, S.171, 1991.
- [45] U. Ramacher et al., Design of a 1st Generation Neurocomputer, in "VLSI design of Neural Networks", Kluwer Academic Publishers, 1991.

# Danksagung

Ich danke ganz besonders Herrn Priv. Doz. Dr. C. Kiesling für das faszinierende Thema dieser Arbeit, die Freiheit bei der Durchführung und die fruchtbaren Diskussionen, die ständig neue Probleme kreierten und lösten und damit zu einem tieferen Verständnis führten.

Ich danke Herrn Prof. Dr. G. Buschhorn für die Aufnahme am Max-Planck-Institut, die Ermöglichung von Dienstreisen zum Deutschen Elektronen Synchrotron (DESY) und die Förderung der Teilnahme an wichtigen Tagungen und Physikschiulen.

Mein ganz spezieller Dank gilt aber Dr. Pál Ribarics, der von Beginn an nicht nur intensiv mit mir zusammenarbeitete und jederzeit zu fachlichen Diskussionen bereit war, sondern auch immerzu mit unkonventionellen Ideen zu Arbeits- und Freizeitgestaltung aufwartete.

Schließlich möchte ich mich noch bei allen Mitgliedern der H1-Gruppe bedanken, die alle auf ihre Weise zu einer guten Arbeitsatmosphäre und damit zum Gelingen dieser Arbeit beigetragen haben.

Mein besonderer Dank gilt jedoch meinen Eltern, die mir dieses Studium erst ermöglicht haben.

# Erklärung

Hiermit erkläre ich, daß ich diese Arbeit unter Angabe aller wesentlichen Quellen und Hilfsmittel selbständig angefertigt habe.

München, Dezember 1992