Einsatz neuronaler Netze als intelligente Trigger im H1–Experiment

Diplomarbeit an der Fakultät für Physik der
Technischen Universität München

vorgelegt von

Jürgen Möck

angefertigt am

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

März 1994

Inhaltsverzeichnis

Kapitel 1

Einleitung

Die Frage nach der Struktur der Materie, aus der unsere Welt besteht, hat die Menschheit seit ihren Anfängen beschäftigt. Streuexperimente bilden seit der Entdeckung des Atomkerns durch Rutherford (1911) das wichtigste Instrument, um der Natur einen tieferen Einblick in ihren fundamentalen Aufbau abzuringen [1] . Die quantenmechanische Welleneigenschaft der Teilchen verknüpft jedoch die erreichbare räumliche Auflösung λ mit der Energie E des streuenden Teilchens (de Broglie 1923),

$$\lambda \approx \frac{2\pi\hbar c}{E} \qquad \text{im relativistischen Limes .} \tag{1.1}$$

Dieser reziproke Zusammenhang zwischen Energie und Auflösungsvermögen gab den Anstoß zum Bau immer größerer Beschleuniger. In diesen werden Elektronen, Protonen und ihre Antiteilchen wie Strahlen in einem Lichtmikroskop dazu verwendet, die kleinsten Bausteine der Materie und die Kräfte, die sie zusammenhalten, zu erforschen. Im Jahre 1966 wurde am SLAC ein Experiment durchgeführt, in dem Elektronen mit bis zu $16\,GeV$ auf ruhende Wasserstoffatome geschossen wurden (Friedman, Kendall, Taylor [2]). Aus den Ergebnissen dieses Experiments konnten J. Bjorken und R. Feynman 1968 ableiten, daß die Elektronen nicht nur am Proton als ganzes gestreut werden, sondern gelegentlich auch an nahezu punktförmigen Konstituenten(Partonen) im Innern des Protons, die nur einen Bruchteil von dessen Energie tragen [3, 4]. Gell-Mann hatte schon vier Jahre zuvor mittels Symmetrieüberlegungen diese Teilchen vorhergesagt und sie Quarks getauft [5].
Da die Schwerpunktsenergie bei „Fixed Target"-Experimenten mit $\sqrt{E_{acc}}$ skaliert, war es bisher für ep-Kollisionen nur möglich, Schwerpunktsenergien von $\sqrt{s} \approx 10 - 15\,GeV$ zu erreichen. Mit dem Speicherring HERA wurde erstmals die Idee verwirklicht, sowohl die Elektronen als auch die Protonen zu beschleunigen und dann zur Kollision zu bringen. Bei einer Elektronenenergie von $30\,GeV$ und einer Protonenenergie von $820\,GeV$ wird so eine Schwerpunktsenergie von $\sqrt{s} = \sqrt{4E_pE_e} = 314\,GeV$ erreicht, entsprechend einem räumlichen Auflösungsvermögen von ca. $10^{-18}cm$. Der Bestimmung der Protonstrukturfunktion kommt durch die Erschließung dieses neuen kinematischen Bereiches bei den beiden Experimenten H1[6] und ZEUS[7] eine besondere Bedeutung zu [8].
Darüber hinaus sind bei der genannten Schwerpunktsenergie elektroschwache Reaktionen nicht selten, die verbunden sind mit dem Austausch von intermediären

Vektorbosonen($W^{\pm}, Z^0$). Das ermöglicht eine genaue Bestimmung von Parametern des Standardmodells [10]. Auch für die Suche nach einer bestimmten Klasse neuer Teilchen, wie Leptoquarks und supersymmetrische Teilchen, ist ein ep-Speicherring wie HERA gut geeignet [11]. In den ersten beiden Jahren wurde mit H1 eine integrierte Luminosität von ca. $25nb^{-1}$(1992) und ca. $500nb^{-1}$(1993) an analysierbaren Daten aufgezeichnet. Erste Ergebnisse aus dieser frühen Anfangsphase liegen bereits vor [12]. Die Designluminosität von $1.5 \cdot 10^{31} cm^{-2} s^{-1} (100pb^{-1}/a)$ soll innerhalb der nächsten Jahre schrittweise erreicht werden. Um diese hohe Luminosität zu erzielen, müssen sich bei HERA alle $96ns$ ein Elektron- und Protonbündel durchdringen.
Die erforderlichen hohen Ströme haben aber auch einen unerwünschten Effekt. Wechselwirkungen der Teilchen mit dem Restgas im Strahlrohr und dem Strahlrohr selbst führen zu einer hohen Rate($\sim 0.5MHz$) von Untergrundereignissen im Detektor. Die Rate der interessanten Physikereignisse liegt jedoch um 5 Größenordnungen niedriger. Das stellt hohe technische Anforderungen an das Triggersystem, dessen Aufgabe es ist, die wenigen Physikereignisse möglichst effizient aus der Gesamtzahl der Ereignisse zu selektieren. Die Randbedingungen sind gegeben durch die Datennahmerate ($\mathcal{O}(10Hz)$) und durch die Totzeit ($< 10\%$). Das ist die Zeit, in der Daten eines Ereignisses ausgelesen werden, und kein weiteres Ereignis getriggert werden kann.
Um diesen Bedingungen gerecht zu werden, wurde ein vierstufiges Triggerkonzept entwickelt [13]. Die erste Stufe(L1) ist eine Kombination von Schwellentriggern verschiedener Subkomponenten des H1-Detektors. Für die zweite Stufe(L2) befindet sich ein *Neuronaler–Netz–Trigger* auf der Basis eines Parallelcomputers in der Entwicklung. Sein Einsatz ist für Ende 1994/Anfang 1995 geplant. Insbesondere sollen auf der zweiten Stufe Korrelationen der Triggerdaten und ihre topologische Information ausgenutzt werden [14, 15]. An die beiden Hardwarestufen schließen sich zwei Softwarestufen(L3, L4) an. 1992/1993 waren nur die Stufen L1 und L4 realisiert. Im Laufe des Jahres 1994 wird L3 zur Verfügung stehen und es erlauben, Algorithmen der zweiten Stufe zu simulieren oder bei Bedarf einzusetzen.
Diese Arbeit untersucht den Einsatz neuronaler Netze für die zweite Triggerstufe im H1-Experiment insbesondere auf der Grundlage von realen Daten aus dem Jahr 1993. Für die erste Realisierung (1994/95) werden Reduktionsfaktoren von $\mathcal{O}(10)$ angestrebt. Entsprechend der zu erwartenden Luminosität und der schrittweisen Erweiterung der Triggerhardware müssen die Reduktionsraten dynamisch angepaßt werden. Das Ziel ist es, bei voller Luminosität eine Verminderung der Untergrundrate um den Faktor 50 zu erreichen.
Die Optimierung der Effizienz und Ratenreduktion eines als Trigger verwendeten neuronalen Netzes ist abhängig von folgendem:

- Auswahl der Ereignisse für das Training
- Auswahl der Eingabegrößen
- Informationsgehalt, Qualität und Stabilität der gewählten Eingabegrößen
- Parameter des Lernalgorithmus

Gegenstand dieser Arbeit ist die detaillierte Untersuchung dieser Punkte im Hinblick auf eine erste Realisierung. Die geplante Hardware-Implementierung wird ebenfalls beschrieben.

Kapitel 2

Das H1-Experiment

2.1 Der Speicherring HERA

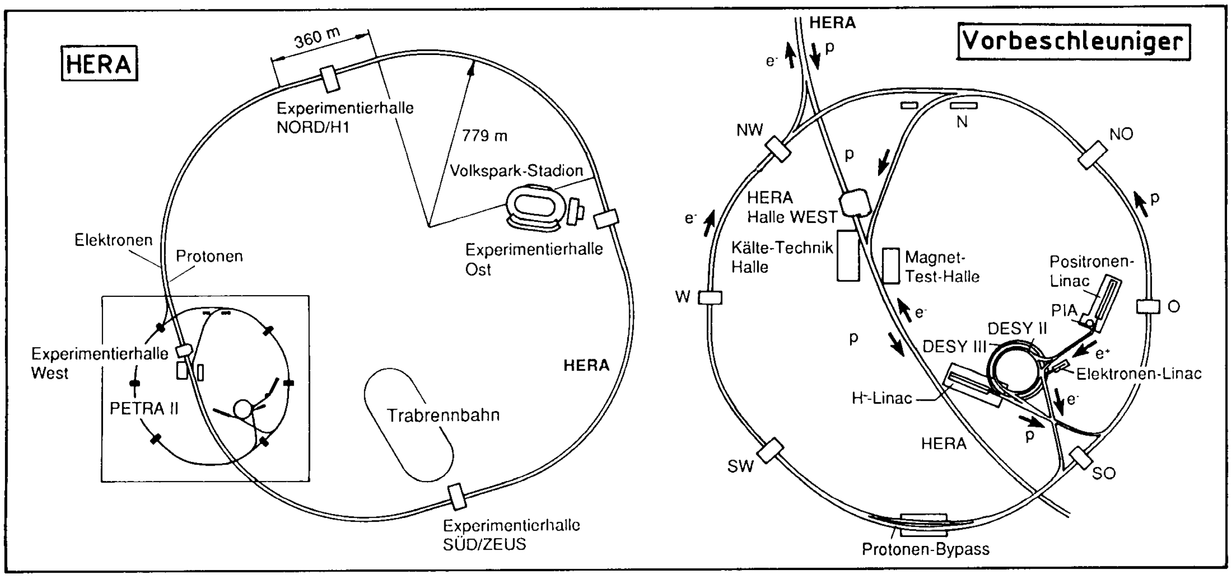

Abbildung 2.1: **Links**: HERA-Speicherringe. **Rechts**: Vorbeschleuniger (Endenergie in Klammern). Elektronen: Linac($200\,MeV$); Synchrotron DESY II($7\,GeV$); Synchrotron PETRA II($14\,GeV$). Protonen: HF-Quadrupol($750\,keV$) und H^--LINAC($50\,MeV$) für H^--Ionen; Synchrotron DESY III($7.5\,GeV$), beim Einschuß werden die beiden Elektronen von den H^--Ionen abgestreift; Synchrotron PETRA II($40\,GeV$).

Die Hadron–Elektron–Ring–Anlage HERA am DESY in Hamburg besteht aus zwei getrennten Speicherringen von $6.3\,km$ Länge, in denen an zwei(max. 4) Wechselwirkungszonen Elektronen und Protonen zur Kollision gebracht werden. In einem System von Vorbeschleunigern werden beide Teilchensorten über mehrere Stufen auf ihre Einschußenergien von $14\,GeV(e)$ und $40\,GeV(p)$ beschleunigt, bevor sie jeweils separat in den HERA-Ring injiziert werden. Dort werden die Elektronen auf ihre nominale

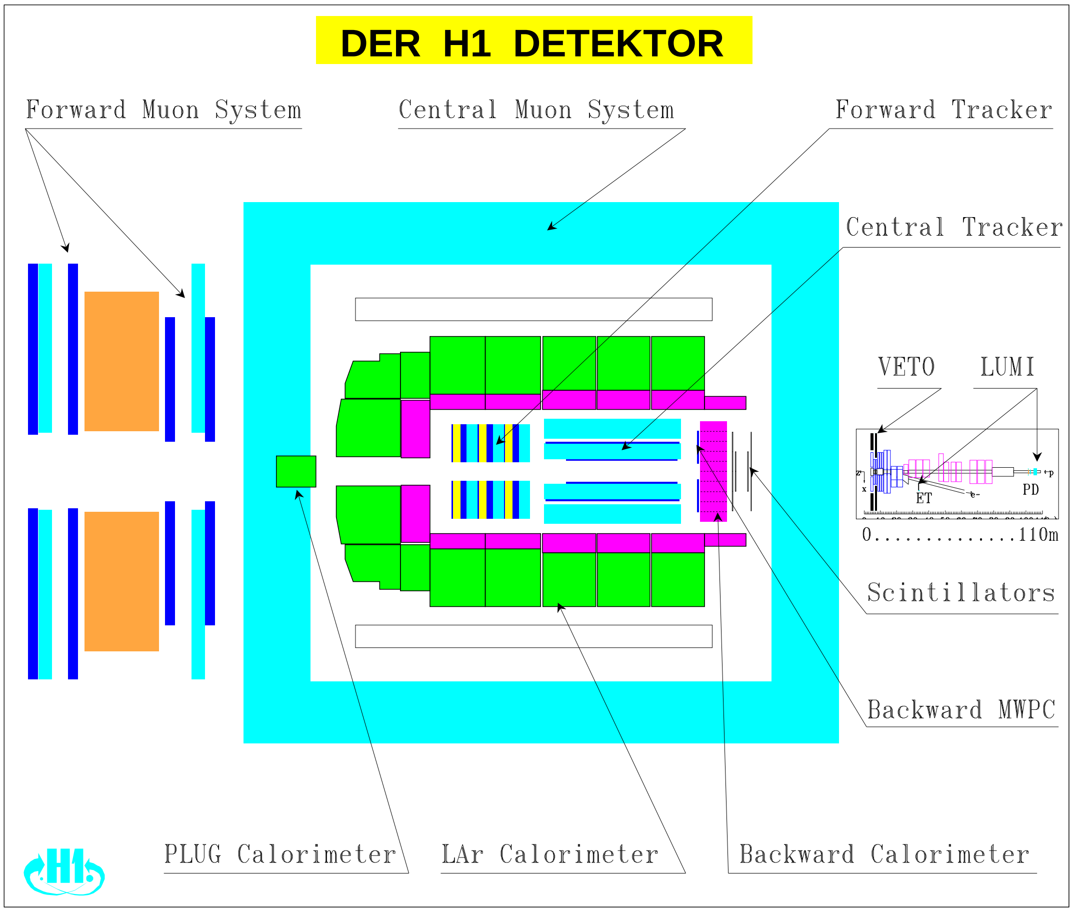

Abbildung 2.2: **Seitenansicht des H1-Detektors.** *Die Elektronen treten von links und die Protonen von rechts in den Detektor ein.*

Energie von $30\,GeV$ (1993 $26.7\,GeV$) und die Protonen auf $820\,GeV$ beschleunigt. Die begrenzenden Faktoren sind für die Protonenenergie das zur Verfügung stehende Feld der supraleitenden Ablenkmagnete von $4.7\,T$ und für die Elektronen die Synchrotronstrahlung, die zu einem raschen Energieverlust führt, der ständig von Hohlraumresonatoren ausgeglichen werden muß. Jeweils 210 Teilchenpakete mit je etwa 10^{11} Teilchen kollidieren mit einer Frequenz von $10.4\,MHz$. Im Herbst 1992 ist ein Umlauf mit 10 Strahlpaketen und 1993 ein Umlauf mit 90 Strahlpaketen verwirklicht worden. Lebensdauern von über 24 Stunden wurden für den Protonstrahl erreicht, während die Lebensdauer für den Elektronstrahl auf 6 Stunden limitiert war.

2.2 Der H1-Detektor

Die Eigenschaften des H1–Detektors bestimmen sich aus den physikalischen Anforderungen. Das Erreichen einer exzellenten Leptonenidentifikation war eines der wichtigsten Entwicklungsziele. Deshalb besitzt der Detektor eine hohe Granularität und Energieauflösung, verbunden mit einer möglichst guten Hermizität, um auch den fehlenden Transversalimpuls einer ep-Kollision bestimmen zu können. Ähnliche Zielsetzungen

mit Schwerpunkt auf die hadronische Energieauflösung wurden bei ZEUS verfolgt.
Entsprechend der Ereigniskinematik tiefinelastischer Lepton–Nukleon–Streuung (DIS; *Deep Inelastic Scattering*) bei HERA-Energien ist der Detektor asymmetrisch gestaltet. Seine wesentlichen Elemente sind ein Spurkammernsystem, verschiedene Kalorimeterkomponenten, die den Winkelbereich von $0.6^0 - 175^0$ † abdecken, ein Myondetektor und ein Luminositätsmeßsystem. [16]
Ausgehend vom Wechselwirkungspunkt($\pm 50cm$) treffen die gestreuten Teilchen innerhalb eines Winkelbereichs von $25^0 - 155^0$ auf die zentralen Spurkammern, die der Bestimmung des Vertex, sowie zur Ladungs- und Impulsmessung dienen. Vorwärtsspurkammern decken den Winkelbereich von 5^0 bis 25^0 ab. Eine Rückwärtsproportionalkammer zur verbesserten Messung des Streuwinkels des Elektrons überstreicht den Bereich von $155^0 - 175^0$. Die maßgebende Komponente von H1, insbesondere für DIS-Ereignisse, ist ein Flüssigargonkalorimeter (LAr), dessen Kryostat ringförmig die Spurkammern umgibt und den Winkelbereich von $4^0 - 150^0$ abdeckt. Es ist unterteilt in ein inneres elektromagnetisches Kalorimeter mit Blei als Absorbermaterial und ein hadronisches Kalorimeter, das Stahlplatten als Absorber verwendet. Die Energieauflösung beträgt $12\%/\sqrt{E_e} \oplus 1\%$ für das elektromagnetische und $50\%/\sqrt{E_h} \oplus 2\%$ für das hadronische Kalorimeter [17]. Der rückwärtige Bereich ($150^0 - 175^0$) wird von einem konventionellen Blei-Szintillations-Sandwichkalorimeter eingenommen, das eine dem LAr vergleichbare elektromagnetische Energieauflösung besitzt, jedoch wegen der geringen hadronischen Wechselwirkungslänge $\lambda \approx 1$ weniger für die hadronische Energiemessung geeignet ist. Eine supraleitende Spule mit einem Durchmesser von $6m$ und einer Länge von $5.75m$ liefert ein Magnetfeld von $1.15\,T$. Mehrere Streamerkammern in und um das Eisenjoch dienen der Myonidentifikation und der Verbesserung der hadronischen Energiemessung von Schauern, die das Kalorimeter durchdringen. Die besonders steifen Myonspuren in Vorwärtsrichtung werden mit äußeren Driftkammern, die einen Toroidalmagneten($1.5 - 1.75\,T$) umschließen, analysiert. Ein kleines, warmes Si–Cu Kalorimeter(PLUG) schließt die Lücke zwischen $0.6^0 - 4^0$, die das LAr-Kalorimeter nach vorne offen läßt. In Rückwärtsrichtung registriert ein Elektrontagger ($z = -33m$) Photoproduktionsereignisse und in Koinzidenz mit einem Photontagger ($z = -103m$) wird die Luminosität über den Bremsstrahlungsprozeß abgeleitet. Zwei Szintillatorwände in Rückwärtsrichtung dienen dem schnellen Verwerfen von Untergrundereignissen. Eine detaillierte Beschreibung der für das Triggersystem wichtigen Detektorkomponenten erfolgt in Kapitel 5.

†Winkel- und Längenangaben beziehen sich auf die Protonstrahlrichtung, für Energien wurde die Einheit GeV verwendet.

Kapitel 3

Physik im H1-Experiment

3.1 Ereigniskinematik

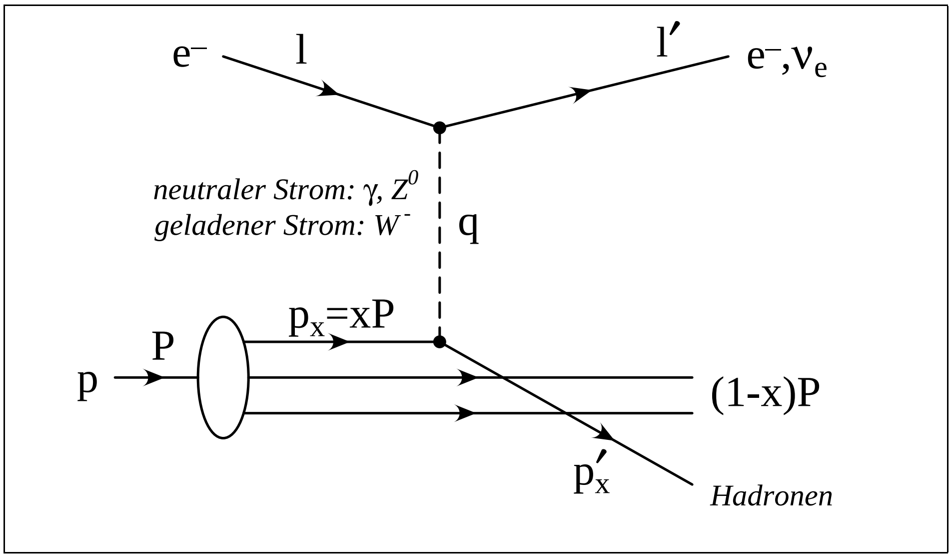

Abbildung 3.1: **Feynman-Graph einer Lepton-Quark-Streuung für neutrale und geladene Ströme**

Aus der klassischen Mechanik ist bekannt, daß ein elastischer Streuprozeß bei gegebener Schwerpunktsenergie s wegen der Energie– und Impulserhaltung bereits vollständig durch Messung des Streuwinkels Θ beschrieben werden kann. Da aber Winkel nicht lorentzinvariant sind, ist es im Falle hochenergetischer, ultrarelativistischer Projektile günstiger, den Streuprozeß durch einen Lorentzskalar, z. B. dem negativen Quadrat des Impulsübertrag Q^2 mit

$$q_\mu = l'_\mu - l_\mu = -(p^x_\mu{}' - p^x_\mu) \quad (3.1)$$

$$Q^2 = -q^2 \quad (3.2)$$

$$s = (l+p)^2\,, \quad (3.3)$$

zu beschreiben. l_μ ist der 4er–Impuls des einlaufenden Elektrons, p^x_μ bezeichnet den 4er–Impuls des Streupartners (Quark) und die gestrichenen Größen beziehen sich auf die auslaufenden Teilchen. Im elastischen Fall können wir für Q^2 auch schreiben

$$Q^2 = 2M\nu \quad \text{mit} \quad \nu = E_{p'} - M. \tag{3.4}$$

M bezeichnet die Ruhemasse des hadronischen Streuteilchens und ν die durch den Stoß übertragenen Energie in dessen Ruhesystem. Der inelastische Fall ist jetzt gekennzeichnet durch

$$Q^2 < 2M\nu. \tag{3.5}$$

Um den inelastischen Fall vollständig beschreiben zu können, bedarf es also eines weiteren Lorentzskalars. Als günstige Wahl erweist sich

$$x = \frac{-q^2}{2M\nu} \qquad (0 \leq x \leq 1); \tag{3.6}$$

wenn die inelastische Streuung eines Leptons an einem Proton als elastischer Stoß an einem seiner Konstituenten(Partonen) interpretiert wird, gibt x genau den Impulsanteil des wechselwirkenden Partons am Gesamtimpuls des Protons wieder. Alternativ wäre auch die Wahl eines anderen Lorentzskalar möglich gewesen, z. B. der normierte Energieübertrag im Ruhesystem des Protons

$$y = \frac{2M\nu}{s} = \frac{2p^\mu q_\mu}{s}. \tag{3.7}$$

Aus den obigen Definitionen ergibt sich die Beziehung

$$Q^2 = sxy, \tag{3.8}$$

die die einzelnen Größen miteinander verknüpft.
Der Streuprozeß kann in niedrigster Ordnung (Bornstreuung) durch den Austausch von elektroschwachen Eichbosonen zwischen den Streupartnern beschrieben werden. Abb. 3.1 illustriert diesen Sachverhalt. Im Falle neutraler Ströme (NC; *Neutral Current*) wird ein γ oder ein Z^0 ausgetauscht, bei geladenen Strömen (CC; *Charged Current*) ein W^+ oder W^-, wobei das gestreute Lepton als Neutrino nicht mehr beobachtet werden kann.
Um die beiden unabhängigen Lorentzskalare zu bestimmen, müssen auch im Experiment gleichzeitig zwei unabhängige Größen gemessen werden. Da der Streuprozeß für unpolarisierte oder longitudinal polarisierte Teilchen unabhängig vom Azimutwinkel ϕ ist, bieten sich die Polarwinkel $\Theta_{e'}, \Theta_{h'}$, sowie die Energien $E_{e'}, E_{h'}$ an. Weil nur zwei Variable gebraucht werden, ist für neutrale Ströme die Kinematik überbestimmt. Wird z. B. Q^2 und y über das Elektron alleine ermittelt, so ist

$$Q^2_l = 4E_e E_{l'} \sin^2 \frac{\Theta_{l'}}{2} \tag{3.9}$$

$$y_l = 1 - \frac{E'_l}{E_e} \cos^2 \frac{\Theta_{l'}}{2}. \tag{3.10}$$

Im Falle der hadronischen Variablen kommt die Schwierigkeit hinzu, daß nicht das streuende Quark selbst beobachtet wird, sondern ein breiter Jet fragmentierter Hadronen. Der in diesem Fall fast unvermeidliche Verlust von Energie im Strahlrohr erfordert, y und Q^2 auf andere Weise über den Transversalimpuls und $\sum_{had} E_i - p_{z_i}$ zu berechnen [18]. Wegen der fehlenden Leptonsignatur kann die Kinematik eines CC-Ereignisses nur auf diesem Wege gemessen werden.

3.2 QCD bei kleinem x

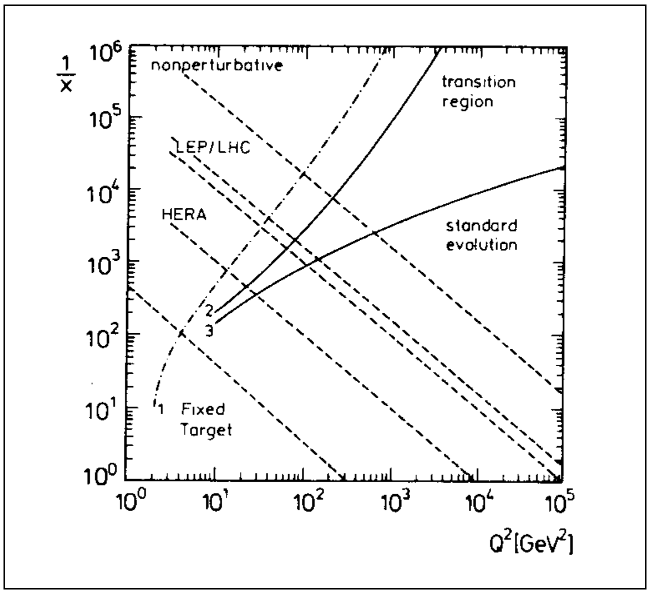

Abbildung 3.2: **Kinematische Region von HERA und einem möglichen *LEP/LHC* ep–Beschleuniger.** HERA überstreicht den Bereich zwischen der zweiten und dritten gestrichelten Diagonale.

Eines der wesentlichen Ziele im H1-Experiment ist, die QCD im Gebiet kleiner x ($< 10^{-2}$) zu erforschen. Einen Überblick über die verschiedenen kinematischen Bereiche, ausgedrückt durch $1/x$ und Q^2 gibt Abb. 3.2 . Unterhalb der ersten gestrichelten Linie liegt der Bereich der durch „*Fixed Target* "-Experimente bekannt ist. Die zweite und dritte Diagonale begrenzen den bei HERA zugänglichen Bereich, die dazwischenliegende Lücke kann geschlossen werden, indem HERA mit niedrigeren Strahlenergien betrieben wird. Im rechten unteren Teil liegt der Bereich, in dem die störungstheoretische Behandlung der QCD gültig ist ($x, Q^2 \to \infty$; Bjorken Limit). Unter Benutzung feldtheoretischer Verfahren (Renormalisierungstheorie, Operatorexpansion, etc.) haben Gribov, Lipatov, Altarelli und Parisi [19, 20] Entwicklungsgleichungen aufgestellt, die diesen Bereich theoretisch gut beschreiben. Geht man jedoch in dem Diagramm von rechts unten nach links oben ($x^{-1}, s \to \infty$; Regge Limit), wird ein Gebiet erreicht, in dem die Entwicklungsgleichungen ein starkes Anwachsen der Dichte weicher Gluonen voraussagen. Dies führt zur Verletzung der Unitarität, was bedeutet, daß die Störungstheorie nicht mehr anwendbar ist. HERA wird also einen Einblick

in den Übergang von störungstheoretischer QCD in ihren nicht störungstheoretischen Bereich geben können.
Wichtiger Bestandteil dieses Programms ist die Bestimmung der Protonstrukturfunktion über den differentiellen (NC-)Wirkungsquerschnitt

$$\frac{d^2\sigma_{NC}(e^{\mp})}{dx\ \ dQ^2} = \frac{4\pi\alpha^2}{xQ^4}\left[(1-y+\frac{y^2}{2})F_2(x,Q^2) - \frac{y^2}{2}F_L(x,Q^2) \pm (y-\frac{y^2}{2})xF_3(x,Q^2)\right].$$

Für kleines x kann der Beitrag von F_3 vernachlässigt werden. Der Ausdruck vor der Klammer kommt vom Photonpropagator. Das Analyseziel ist es, durch Auffalten des Bornschen Wirkungsquerschnittes die Strukturfunktion F_2 zu bestimmen ($F_L = F_2 - 2xF_1$ nennt man die longitudinale Strukturfunktion; sie kann mit Einschränkungen [21] bei HERA über radiative Ereignisse abgeleitet werden). Die nicht einfache Rekonstruktion bedingt, daß Triggereffizienzen hier für die Fehler nur eine untergeordnete Rolle spielen. Bezüglich des neuronalen L2-Triggers ist es in erster Linie wichtig, für diese Ereignisklasse eine hohe Unterdrückung des Maschinenuntergrundes bei gleichzeitig sehr guter Stabilität sicherzustellen.
Erste noch mit großen Fehlern behaftete Ergebnisse liegen bereits vor (Daten von 1992; $25nb^{-1}$). Als Beispiel ist in Abb 3.3 die Strukturfunktion F_2 bei $Q^2 = 15\,GeV^2$ aufgetragen (globaler Normierungsfehler $\pm 12\%$). Tendenziell passen Modelle mit starkem Anstieg des Gluonanteils bei kleinem x eher zu den Daten [12].

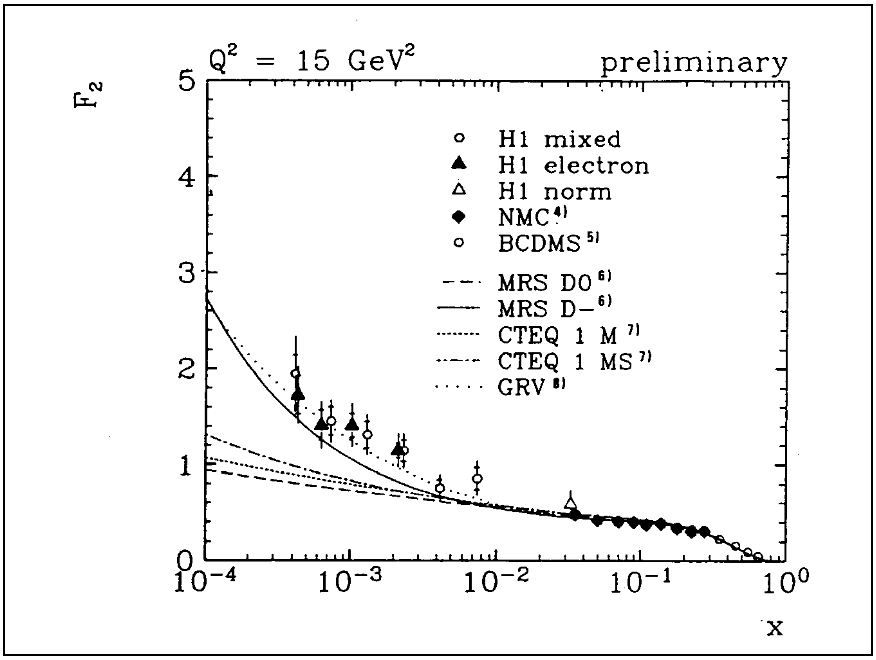

Abbildung 3.3: **F2 aus den H1-Daten von 1992 im Vergleich mit verschiedenen Partonparametrisierungen.**

3.3 Geladene Stromereignisse

Der theoretische Wirkungsquerschnitt dieser Ereignisklasse im Quark-Partonmodell ist gegeben durch

$$\frac{d^2\sigma}{dx\ \ dQ^2} = \frac{\pi^3}{xs} \sum_f |M_f|^2\, q_f(x, Q^2). \tag{3.11}$$

q_f gibt die Wahrscheinlichkeitsdichte einer bestimmten Quarkflavor bei festem x und Q^2 an. M_f ist das Matrixelement der Wechselwirkung. Dieses ist proportional dem Propagator des ausgetauschten Eichbosons

$$M_f \propto \frac{1}{M_{boson}^2 + Q^2}. \tag{3.12}$$

Daraus folgt, daß der Wirkungsquerschnitt mit zunehmendem Impulsübertrag proportional zum Quadrat des Photonpropagators mit $1/Q^4$ stark abfällt. Bei kleinem Q^2 ($< 10^3\,GeV^2$) sind die geladenen Stromereignisse wegen $M_W = 81.2\,GeV$ stark unterdrückt gegenüber den NC-Ereignissen. Für große Impulsüberträge hingegen werden die Wirkungsquerschnitte vergleichbar. Bei HERA–Energien sind geladene Stromereignisse um den Faktor 10^3 seltener als neutrale Stromereignisse mit $x > 10^{-3}$. Aus einer Datenmenge, die $370nb^{-1}$ entspricht, konnten 1993 insgesamt 14 solcher Ereignisse identifiziert werden [22]. Abb. 3.4 zeigt eines dieser Ereignisse. Aus σ_{cc} kann die Masse des W-Bosons abgeleitet werden, allerdings mit geringerer Genauigkeit als bei LEP. Außerdem können Rückschlüsse über Quarkdichten im Proton abgeleitet werden [23]

$$d\sigma_{cc}(e^-) \sim x[u + c + (1-y)^2(\bar{d} + \bar{s})]. \tag{3.13}$$

Auf Triggerebene ist das Erkennen von CC-Ereignissen wegen des nicht detektierbaren Neutrinos mit größeren Schwierigkeiten als für NC–Ereignisse verknüpft. In einem idealen hermetischen Detektor lassen sich solche Ereignisse am leichtesten über ihren fehlenden Transversalimpuls identifizieren. Ist dieser jedoch klein, und geht ein Teil der Energie im Strahlrohr in Vorwärtsrichtung verloren, dann wird die Signatur dem eines Untergrundereignisses sehr ähnlich. Da aber auch auf Rekonstruktionsebene in diesem Fall eine Untergrundabtrennung kaum möglich ist, wird meist ein relativ hoher fehlender Transversalimpuls als Selektionskriterium verlangt. Untergrundereignisse mit dieser Signatur (Halo, kosmische Strahlung) besitzen eine beherrschbare Rate, so daß auch hier sehr hohe Triggereffizienzen möglich sind. Die Kombination von Spurkammer- mit Kalorimeterinformation steigert für diese Ereignisklasse die Triggereffizienz [24], was dem Einsatz neuronaler Netze entgegenkommt.

3.4 Andere Forschungsgebiete

Da die Schwerpunktsenergie von $\sqrt{s} = 314\,GeV$ für ep–Kollisionen auf der Welt einzigartig ist, ist auch das übrige physikalische Programm breit gefächert. Wichtige Gebiete abseits der DIS–Physik sind Suche nach neuen Teilchen, Photoproduktion und Erzeugung schwerer Quarks.

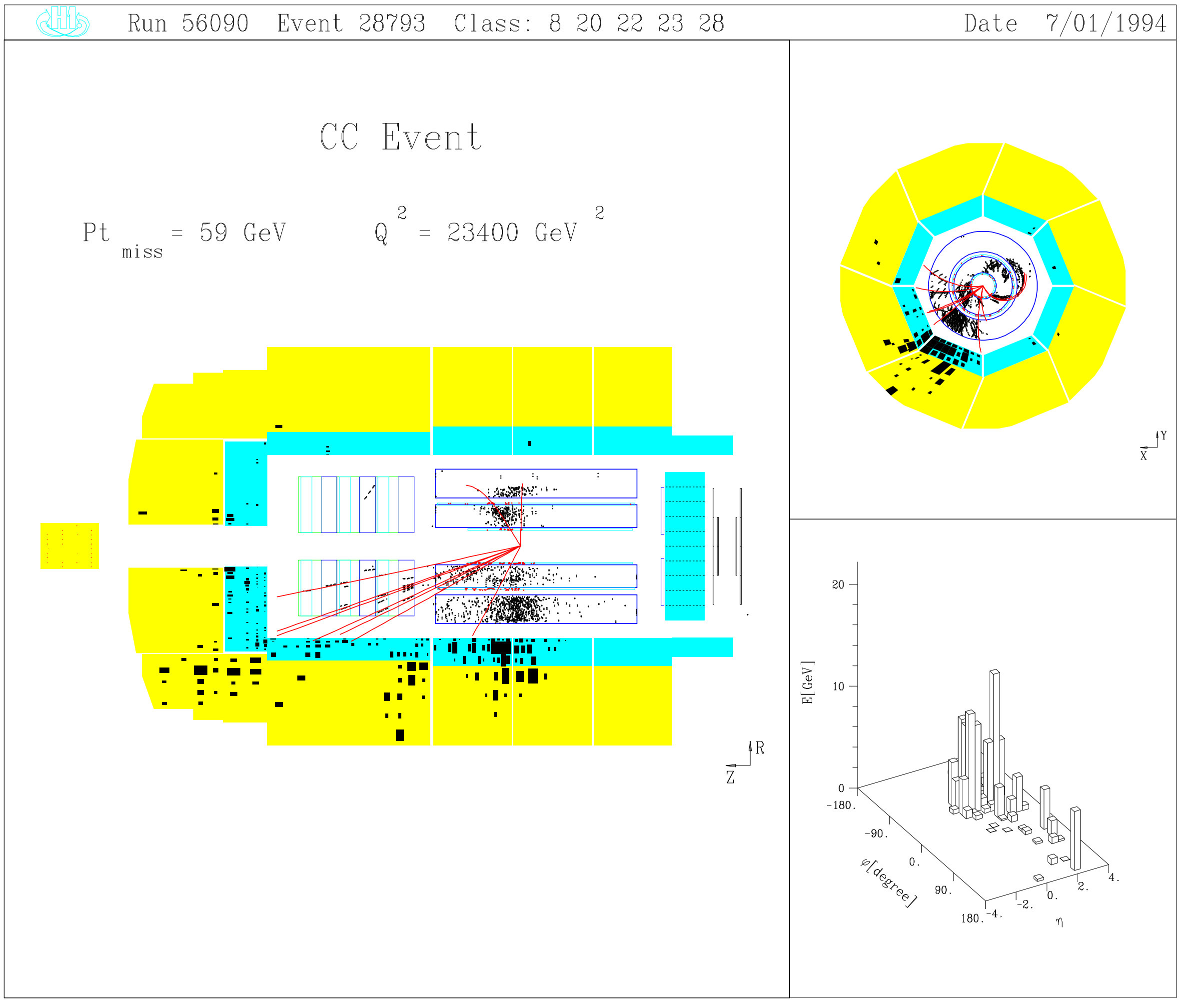

Abbildung 3.4: **Ein geladenes Stromereignis im H1–Detektor.**

Neue Teilchen

Um neue Teilchen wie Leptoquarks zu identifizieren, sind besonders deutliche Signaturen gefordert. Ereignisse, mit denen neue Teilchen gesichert nachgewiesen werden können, sind nicht zuletzt wegen deren hohen Massen meist mit großen Energiedepositionen im Kalorimeter oder mit hohem fehlenden Transversalimpuls verknüpft. Derartige Signaturen sind für das Triggersystem von H1 unproblematisch. Die eigentliche Schwierigkeit lastet hier auf dem zu führenden Beweis, daß diese Ereignisse nicht mit dem Standardmodell verträglich sind.

Photoproduktion

Mit dem nächsthöheren Wirkungsquerschnitt nach den rein elektromagnetischen Reaktionen wie $ep \rightarrow ep\gamma$ stellt die Photoproduktion für die DIS-Physiker einen lästigen Untergrund dar, verdient aber eine eigene Untersuchung. Das Photon wird meist als punktartiges Eichboson betrachtet, kann aber trotzdem sehr komplexe Reaktionen höherer Ordnung eingehen, anstatt direkt wechselzuwirken. γp-Physik gibt

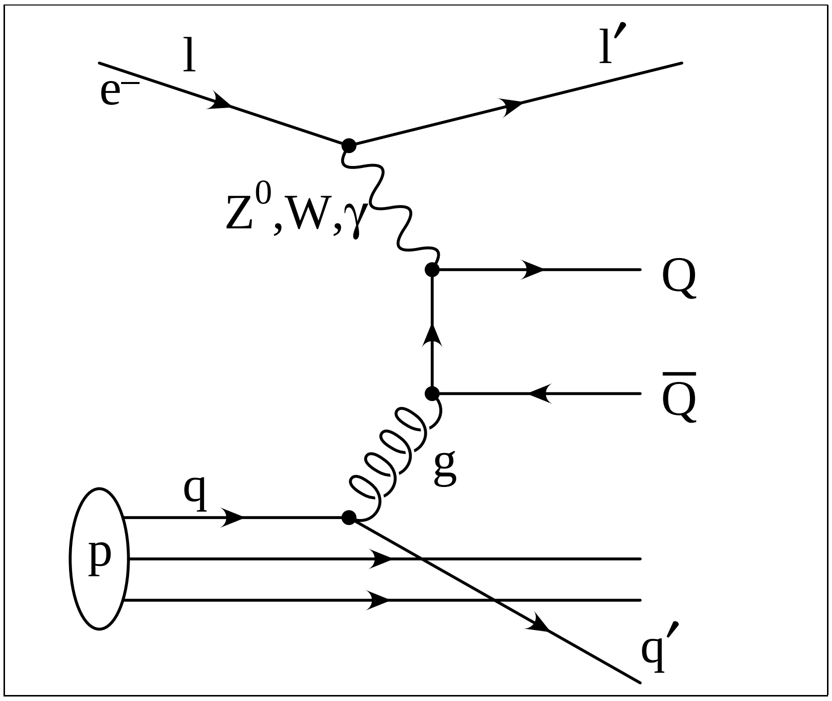

Abbildung 3.5: **Feynman-Graph einer $Q\bar{Q}$-Erzeugung durch Boson-Gluon-Fusion.**

deshalb einen tieferen Einblick in die Struktur des Photons. Bei hohem Transversalimpuls $p_\perp$ können diese Ereignisse störungstheoretisch behandelt werden, für niedriges $p_\perp$ existieren phänomenologische Modelle. Im H1–Experiment kann bei kleinen Ablenkungswinkeln des Elektrons ein γp–Ereignis mit Hilfe des Elektrontaggers ($z = -33m$) getriggert werden. Das stellt jedoch nur einen Bruchteil der Gesamtereignisrate dar. Für die nicht getaggten Ereignisse sind die Triggereffizienzen stark von den angewandten kinematischen Schnitten abhängig.

Erzeugung schwerer Quarks

Schwere Quarks *(charm, beauty)* werden bei HERA überwiegend via Photoproduktion der Ordnung $\alpha^2\alpha_s$ wie der in Abb. 3.5 gezeigten Boson–Gluon–Fusion erzeugt. Neben der Bestimmung von Verzweigungsverhältnissen verschiedener Zerfallskanäle von schweren Quark-Teilchen kann auch auf die Gluondichte geschlossen werden. Die inklusiven Wirkungsquerschnitte für die Produktion schwerer Quarks bei HERA können abgeschätzt werden zu [25, 26]

$$\begin{aligned} \sigma(ep \to cX) &\approx \mathcal{O}(1\mu b) \\ \sigma(ep \to bX) &\approx \mathcal{O}(10nb). \end{aligned} \tag{3.14}$$

Das Top-Quark kann bei einer unteren Massengrenze von $113\,GeV$ [27] für die gegebenen HERA-Energien nicht gefunden werden. Unter der Annahme einer integrierten Luminosität von $100pb^{-1}/Jahr$ erhält man $\mathcal{O}(10^8)$ *Charm*– und $\mathcal{O}(10^6)$ *Beauty*–Hadronen. Eine Zahl von 10^8 Ereignissen ist durchaus kritisch im Hinblick auf Speicherkapazität und Triggerrate. Um die Gluondichte zu ermitteln, untersucht man ausgesuchte Zerfallskanäle mit günstiger Signatur, aber mit entsprechend niedrigerem Wirkungsquerschnitt, z. B.

$$\begin{aligned} \sigma(D^{*\pm} \to D^0\pi^\pm \to (K^\mp\pi^\pm)\pi^\pm) &\approx 1nb \\ \sigma(ep \to c\bar{c}X \to e^+e^-X) &\approx 60pb \\ \sigma(ep \to J/\psi X) &\approx (10-20)nb\ . \end{aligned} \tag{3.15}$$

Zu beachten ist, daß die augenblickliche Luminosität (1993) von $0.5pb^{-1}$ keine ausreichende Statistik für die Analyse dieser Reaktionen darstellt. Die Vielfalt der Endzustände und die involvierten geringen Energien ($\sim 1\,GeV$) erschweren das Triggern solcher Ereignisse. Verglichen mit anderen Reaktionen sind hier die Effizienzen auch am geringsten. Deshalb gilt es auch aus Sicht eines L2–Neurotriggers, auf diese Klasse ein besonderes Augenmerk zu richten.

3.5 Untergrundereignisse

Als Untergrundereignis sei im folgenden alles bezeichnet, was ein Triggersignal im Detektor auslösen kann, das nicht von einer Wechselwirkung zwischen Elektron und Proton herrührt.

3.5.1 Strahl-Gas-Reaktionen

Eine prominente Untergrundquelle bei H1 sind Kollisionen von Protonen mit den Restgasatomen im Vakuum des Strahlrohrs ($p \sim 10^{-9}mbar$). Für Elektronen spielt die analoge Reaktion dagegen eine untergeordnete Rolle. Die Rate von Strahl-Gas-Ereignissen berechnet sich als das Produkt des geometrischen Wirkungsquerschnitts $\mathcal{F}$ der Reaktion mit der Flächendichte η des Gases und dem Strom I der einfallenden Protonen zu

$$\mathcal{R}_{BG} = \underbrace{\sigma_{pp}A^{2/3}}_{\mathcal{F}}\ \underbrace{\rho l}_{\eta}\ \underbrace{n_p f}_{I}\ . \tag{3.16}$$

σ_{pp}	totaler $p-p$ Wirkungsquerschnitt
A	Massenzahl des Kerns
ρ, l	Teilchendichte, Länge des betrachteten Volumens
n_p, f	Anzahl der Protonen im Ring, Umlauffrequenz

Die Schwerpunktsenergie einer p-p-Kollision mit $E_p = 820\,GeV$ und einem ruhenden Target ist $39\,GeV$. Durch Experimente bei vergleichbaren Energien am ISR(CERN) weiß man, daß der totale Wirkungsquerschnitt bei $39\,GeV$ etwa $40\,mb$ beträgt, wobei der elastische Kanal nur etwa $7\,mb$ beiträgt. pN-Stöße können bei diesen Energien als Nukleon-Nukleon-Reaktion mit einer sich anschließenden intranuklearen Kaskadenreaktion verstanden werden. Die entstehenden Fragmente werden entweder prompt aus dem Kern herausgeschlagen oder *dampfen* mit einer gewissen Verzögerung vom Restkern *ab*. Die mittleren geladenen Multiplizitäten$(10-20)$ unterscheiden sich allerdings für die pp- und pN-Reaktionen nur um den Faktor 2, da auch die inelastische

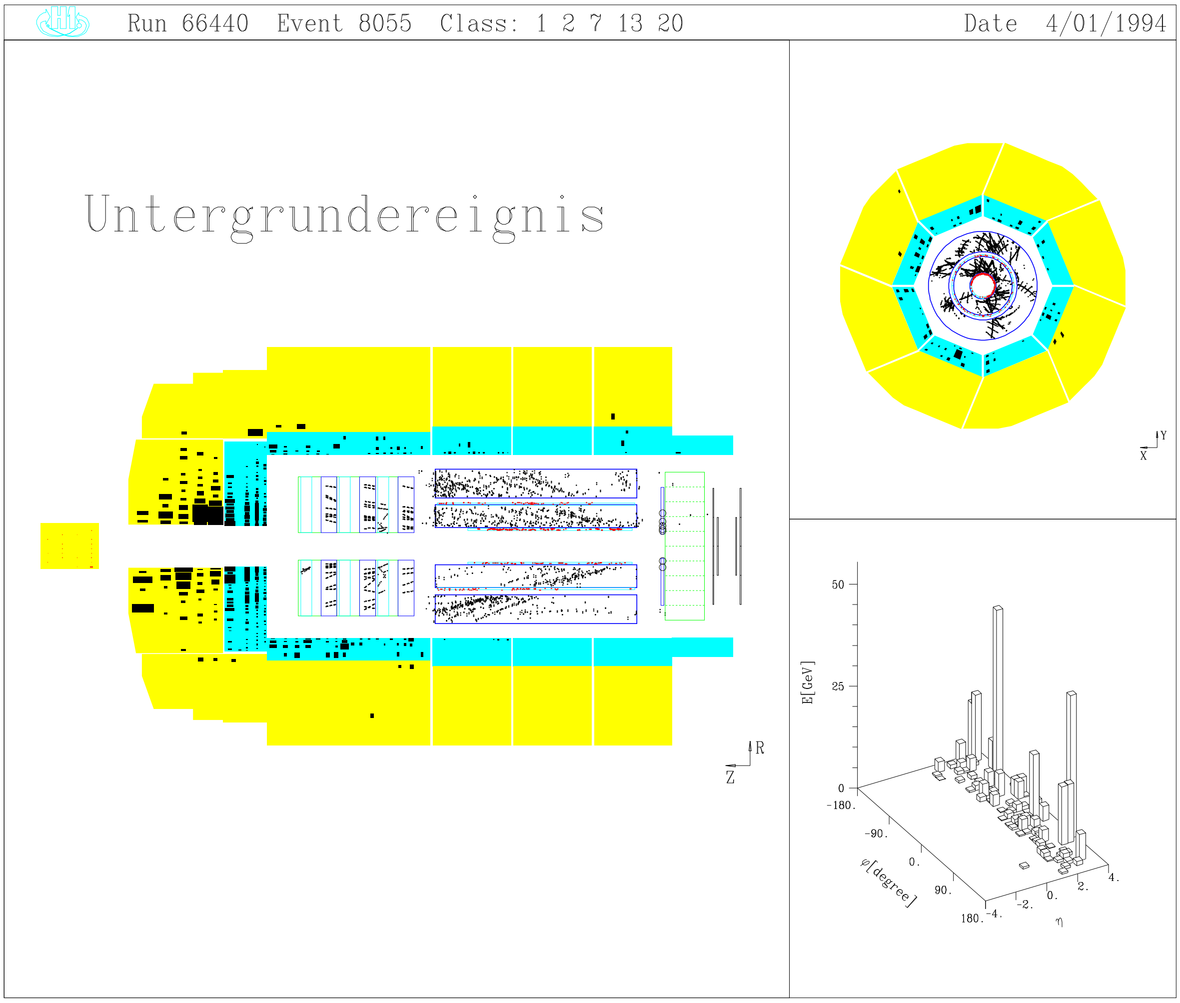

Abbildung 3.6: **Eines der zahlreichen Untergrundereignisse aus einem Run im Herbst 1993.** *Es handelt sich sehr wahrscheinlich um eine Sekundärwechselwirkung, die sowohl von einer Strahl–Gas– als auch von einer Strahl–Wand–Reaktion herrühren kann.*

pp-Reaktion in einer hadronischen Aufschauerung endet. Aus der Gleichung (3.16) läßt sich die Rate bei einem nominalen Strom von $480\,mA$ $(2 \cdot 10^{13} p)$ abschätzen zu $\sim 5\,kHz/m$, auch wenn die mittlere freie Weglänge eines Protons dabei $10^{12}km$ entspricht! Die chemische Zusammensetzung des Restgases ist nicht genau bekannt. Wasserstoff, Stickstoff und Kohlenmonoxid sind mögliche Komponenten. Daraus resultiert bereits ein Unsicherheitsfaktor von ungefähr 2.
Wegen der stark in Vorwärtsrichtung ausgeprägten Kinematik können Kollisionen mit bis zu $100\,m$ Entfernung vom Detektor in diesem ein Signal hinterlassen. Bei Entfernungen über $10\,m$ spielt die primäre Wechselwirkung jedoch eine untergeordnete Rolle. Kollimatoren an der Detektorfront fangen zwar langreichweitige Strahl-Gas-Teilchen ab, sind dadurch aber auch die dominierende Quelle von Sekundärwechselwirkungen, die den Detektor erreichen. Aus Simulationen ergibt sich eine Obergrenze der Rate für Ereignisse mit signifikanten Signalen im Detektor von $45\,kHz$. Es besteht eine Un-

sicherheit der Rate von etwa einem Faktor 5 [28]. Typisch für Strahl-Gas-Reaktionen sind vom nominalen Vertex abweichende Spuren mit einer geringen Energiedeposition im Kalorimeter, meist nahe des Strahlrohrs. Die Rate von Strahl-Gas-Reaktionen mit gültigem Vertex wird durch eine Ionen-Getter-Pumpe in unmittelbarer Nähe des Vertex herabgesetzt.
Ein typisches Untergrundereignis zeigt Abb. 3.6 . Es ist deutlich zu sehen, daß die Spuren auf einen Vertex in der Nähe des rückwärtigen Kalorimeter zeigen, etwa 1 m von der eigentlichen Wechselwirkungszone entfernt. Die Energiedepositionen im Vorderteil des Flüssigargonkalorimeters sind ebenfalls nicht untypisch.

3.5.2 Strahl-Wand-Reaktionen

Im Gegensatz zu Elektronen werden bei Protonen Abweichungen von der Sollbahn nicht durch Synchrotronstrahlung gedämpft. Daher weiten sich die Strahlpakete mit der Zeit auf, und die Teilchendichte im Randbereich des Protonstrahls nimmt zu. Die mittlere Protonenverlustrate beträgt $10^5 s^{-1} m^{-1}$ bei einer Lebensdauer des Strahls von 20 Stunden. Der Löwenanteil dieses Halos wird mittels Kollimatoren(*Scraper*) in einem geraden Teilstück von HERA aus dem Ring entfernt. Dennoch sind aber vor allem Stellen mit verringerter Apertur, wie die Halbquadrupole im Bereich $z = -40$ bis $-27\,m$ vor dem Detektor, eine Quelle von primären Strahl–Wand–Ereignissen.
Elastisch gestreute Protonen durchqueren aufgrund ihrer sehr kleinen Polarwinkel einfach das Strahlrohr des Experiments, ohne ein Signal zu hinterlassen. Führende Sekundärteilchen aus multihadronischen Reaktionen treffen jedoch in großer Zahl auf den Detektor [28]. Insbesondere Myonen können die Abschirmungen leicht durchdringen. Die signifikante Rate liegt bei $\mathcal{O}(100\,kHz)$. Sie ist vergleichbar mit der Rate von Strahl-Gas-Ereignissen, aber schwieriger abzuschätzen wegen der starken Abhängigkeit von Strahllebensdauer, Kollimatoreffizienz und Aperturgeometrie.

3.5.3 Kosmische Strahlung

Protonen und α-Teilchen aus der kosmischen Strahlung erzeugen beim Eintritt in die Erdatmosphäre eine große Anzahl von hochenergetischen Myonen durch pN-Reaktionen. Die relativ große mittlere Lebensdauer von $2.2 \cdot 10^{-6} s$ und die große mittlere Reichweite von Myonen in Materie lassen diese mit einer primären Rate von $10^3\,Hz$ auf den Detektor niederschauern. Diese hohe Rate kommt in erster Linie von der Größe des Detektors, wird ein Vertex aus der Wechselwirkungszone verlangt, so reduziert sich die Rate auf $0.1\,Hz$. Im Prinzip ließe sich auch ein Großteil dieser Ereignisse auf L2 verwerfen. Da aber die Topologie eines NC-Ereignisses mit hohem Q^2 dem eines kosmischen Treffers ähneln kann, und von den NC-Ereignissen nach Möglichkeit keines verloren gehen sollte, ist es ratsamer diesen Untergrund auf den Softwarestufen L3 und L4 zu verwerfen, wo detailliertere Informationen zur Verfügung stehen. Zudem ist diese Art von Untergrund unabhängig von der Luminosität.

3.5.4 Rauschen

Die Triggersignale des Flüssigargonkalorimeters werden über sogenannte *Triggertower* (TT) ausgelesen, die Signale aus Kalorimeterzellen mit feinerer Granularität aufsummieren und umformen. Bis zu 4 solcher Triggertower werden zu jeweils 240 elektromagnetischen und hadronischen *Bigtowern* (BT) zusammengefaßt. Die gegebene Zellarchitektur und Ausleseelektronik mit ihren Vorverstärkern erzeugt ein weißes Rauschen, wobei $1\,\sigma$ einer Energiedeposition von $250\,MeV$ entspricht. Kohärentes Rauschen kann ebenfalls zum Untergrund beitragen. Wird nun die Triggerschwelle entsprechend hoch gewählt (5σ), so ist die Rate der durch Rauschen zufällig ausgelösten L1-Trigger zu vernachlässigen. Wird diese Schwelle erniedrigt, oder treten kohärente Rauschquellen hinzu, so können zufällig getriggerte Ereignisse auf L2-Ebene massiv zunehmen. Da jedoch kaum weitere Signaturen im Detektor auftreten, sind solche Ereignisse auf höheren Triggerebenen sehr leicht zu verwerfen. Es darf aber nicht übersehen werden, daß die höheren Triggerstufen nicht totzeitfrei sind, so daß derartiger Untergrund bevorzugt auf L1 zu unterdrücken ist.
Eine weitere Untergrundquelle ist die Synchrotronstrahlung des Elektronenstrahls vom letzten Ablenkmagneten vor der Wechselwirkungszone. Die Strahlungsleistung in Detektorrichtung beträgt immerhin $3.6\,kW$ mit einer kritischen Energie von $42.4\,keV$. Masken mit einer geeigneten Beschichtung zur Senkung der Albedo verringern den Strahlungsfluß jedoch ausreichend, so daß auch bei Designluminosität die durch Synchrotronstrahlung ausgelösten Trigger in den inneren Spurkammern eine beherrschbare Rate besitzen sollten.

Kapitel 4

Neuronale Netze

Dieses Kapitel stellt keine umfassende Beschreibung des Themengebiets der neuronalen Netze dar. Die Popularität des Themas hat in den letzten Jahren eine ganze Reihe hervorragender Einführungen entstehen lassen, die diesen Bedarf abdecken [29, 30, 31]. Es ist vielmehr beabsichtigt, die Grundideen und -konzepte zu vermitteln, auf die diese Arbeit im weiteren aufbaut.

4.1 Neuronale Netze als Trigger

Ein Ereignis, das die Triggerebene L1 im Experiment passiert hat, gilt es mittels der zur Verfügung stehenden Daten entweder als Kandidat einer ep–Kollision(Physik) oder als Untergrund zu klassifizieren. Für Erzeugung und Transfer der Daten stehen nur wenige Mikrosekunden zur Verfügung. Die Präzision und der Informationsgehalt der Daten ist deshalb begrenzt. Erzeugung und Verarbeitung zusammen dürfen insgesamt nur etwa $20\mu s$ an Zeit verbrauchen.
Ein frühzeitig als Untergrund klassifiziertes Ereignis hilft Totzeit einzusparen, da die Datenauslese abgebrochen und das Experiment sofort wieder gestartet werden kann. Serielle Rechner können in $20\mu s$ nur einige 100 elementare Operationen ausführen. Das reicht wegen der Datenmenge für einen effizienten Klassifizierungsalgorithmus nicht aus. Analog zu L1 ließen sich auch konservative komplexere Schwellentrigger in Hardware implementieren. Da aber auf L2 bereits einige *kByte* an Daten verfügbar sind, besteht die prinzipielle Schwierigkeit, sowohl der Komplexität habhaft zu werden, als auch gleichzeitig Effizienz und Fehlertoleranz zu optimieren. Die Idee, durch Parallelverarbeitung Selektionsprozesse drastisch zu beschleunigen, ist nicht neu [32], hat aber durch die Verfügbarkeit von Parallelrechnern eine neue Qualität bekommen. Dies motiviert den Einsatz von Algorithmen auf der Basis neuronaler Netzwerke, da diese inhärent versuchen, eine optimale Klassifikation zu approximieren [14, 33]. Der auf L2 eingesetzte Parallelrechner erlaubt zudem in Grenzen die Implementierung anderer parallelisierbarer Algorithmen und erhöht so die Flexibilität des Systems.

4.2 Welche Probleme kann ein neuronales Netz lösen?

In den letzten Jahren sind neuronale Netzwerke für eine Vielzahl von Anwendungen vorgeschlagen und eingesetzt worden. Mustererkennung, Prognosen, Funktionsapproximierung sind allgemeine Beispiele. Da sich diese Arbeit konkret mit einer Anwendung auseinandersetzt und nicht die Theorie neuronaler Netzwerke diskutiert, wollen wir den Blickwinkel etwas einschränken und die Diskussion neuronaler Netzwerke mit einer formalen Problembeschreibung beginnen.

Gegeben sei eine endliche Anzahl N von Ereignissen, denen jeweils ein Datenvektor $\vec{x}_j$ mit dem Ereignisindex j zugeordnet ist. Desweiteren sei a priori eine partionierte Klasseneinteilung $\mathbb{C}$ mit einer fixierten Klassenanzahl K (= 2) gegeben, wobei

$$\mathbb{C}(k) \cap \mathbb{C}(k') = \emptyset \qquad \forall k, k' \ mit \ k \neq k' \tag{4.1}$$

$$\bigcup_{k=0}^{K} \mathbb{C}(k) = \{1, 2, \ldots, N\}. \tag{4.2}$$

Letzeres besagt lediglich, daß jedes Ereignis ein eineindeutiges Klassifikationsmerkmal besitzt. Die umsichtige Betrachtung solcher formaler Definitionen ist nicht unwichtig, wie leicht daran ermessen werden kann, daß eine a priori gegebene Klasseneinteilung im Widerspruch zur Natur eines Experiments an sich steht. Das Gesamtvolumen der Information kann jedoch ohne weiteres genutzt werden, um aus der a posteriori ermittelten Klassifikationswahrscheinlichkeit eine a priori Klasseneinteilung im Hinblick auf die reduzierte Information der Triggerdaten von L2 zu bilden. Einfacher ausgedrückt: ein Ereignis, das in der Analyse mit einer Sicherheit von 99.8% als NC-Ereignis klassifiziert wurde, darf für unsere Zwecke ohne besondere Einschränkungen als gegebenes NC-Ereignis betrachtet werden.

Nun sei $\Re(\mathbb{C})$ eine beliebige eindeutige Abbildung der Klasseneinteilung auf die Menge der reellen Zahlen, z. B. der Klassenindex k selbst ('0' für Untergrund und '1' für Physik).

Die Aufgabe besteht nun darin, für einen gegebenen Datenvektor $\vec{x}$, dessen Klassifikation unbekannt ist, eine optimale Klassifikationsprognose abzugeben. Wenn mit $p(\mathbb{C}(k)|\vec{x})$ die a posteriori Wahrscheinlichkeit bezeichnet wird, daß $\vec{x}$ zur Klasse $\mathbb{C}(k)$ gehört, dann ist das Problem mit Auffinden der Funktion F,

$$F(\vec{x}) = \Re \left(\mathbb{C}(M) \middle| \ p(\mathbb{C}(M)|\vec{x}) \geq p(\mathbb{C}(k)|\vec{x}) \ \ \forall k \neq M \right) , \tag{4.3}$$

gelöst. $p(\mathbb{C}(k)|\vec{x})$ ist im allgemeinen aber unbekannt. Es ist aber möglich, über den Bayes'schen Klassifizierer p auf elegante Weise durch die *a priori* Wahrscheinlichkeiten P auszudrücken

$$p(\mathbb{C}(k)|\vec{x}) = \frac{P(k) \cdot P(\vec{x}|k)}{\sum^K P(k) \cdot P(\vec{x}|k)} \,. \tag{4.4}$$

Kennt man die a priori Wahrscheinlichkeiten, so kann das Klassifikationsproblem analytisch gelöst werden. Für neuronale Netze sind deshalb besonders solche Probleme interessant, in denen die Wahrscheinlichkeitsverteilungen unbekannt sind.

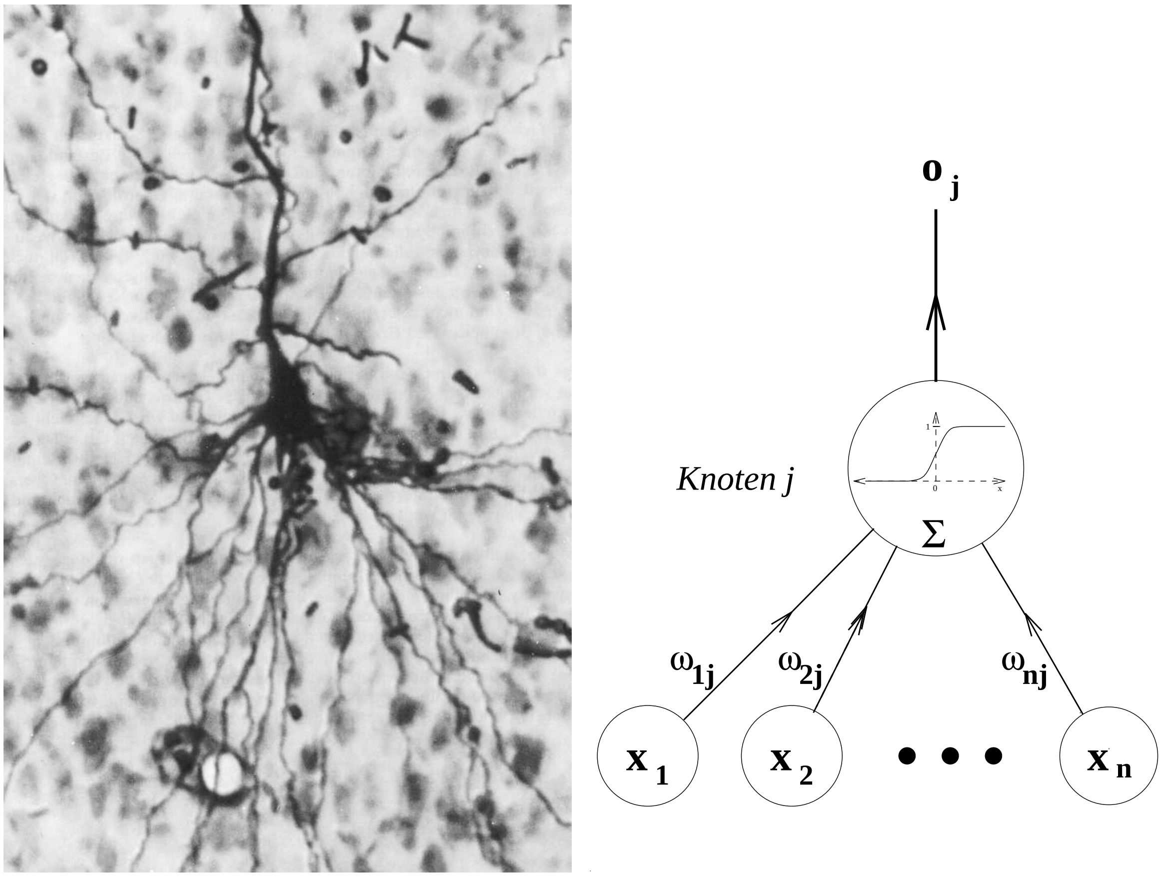

Abbildung 4.1: **Links:** *Ein Neuron aus der Sehrinde einer Katze, das durch Anfärben sichtbar gemacht wurde. Das Axon ist der etwas dickere, sich nach oben windende Fortsatz.* **Rechts:** *Das mathematische Modell eines Neurons. Die Eingabewerte* $\mathrm{x_i}$ *werden mit* ω_{ij} *individuell gewichtet und im Knoten* j *aufsummiert. Über die Aktivierungsfunktion wird aus der Summe das Ausgabesignal* $\mathrm{o_j}$ *gebildet.*

Menschen und Tiere lösen ständig selbst die komplexesten solcher Aufgaben innerhalb von Bruchteilen von Sekunden, z. B. die Erkennung eines Gesichts. Es liegt deshalb nahe, zu versuchen, es der Natur nachzumachen, und neuronale Netzwerke zur Lösung derartiger Probleme einzusetzen.

4.3 Der Zusammenhang mit der Biologie

Das menschliche Gehirn ist eine der kompliziertesten Strukturen, die uns bekannt ist. Um so beachtenswerter ist, daß im Gehirn nur ein Grundtypus von Zelle existiert, der Information übertragen und diese in gewisser Weise auch speichern kann. Dieser Grundtypus wird mit *Nervenzelle* oder *Neuron* bezeichnet.
Jede Nervenzelle besteht aus einem Zellkörper und einer Reihe von faserartigen Fortsätzen. Doch nur eine einzige dieser Fasern, das Axon, dient der Weitergabe von Information, alle anderen Fortsätze, die sogenannten Dendriten, empfangen Signale von anderen Neuronen. Die Länge von Axonen variiert beträchtlich ($10\mu m - 1m$), während

Dendriten etwa $10\mu m - 100\mu m$ lang sind. Das Axon spaltet sich an seinem Ende in bis zu 1000 dünne Äste auf, die in kleinen Verdickungen enden. Dort findet die Informationsübertragung auf die benachbarte Zelle über eine *Synapse* statt, einem schmalen Spalt ($20nm$) zwischen Axonendigung und dem benachbarten Dendriten. Das Axon ist bei Wirbeltieren mit fetthaltigem Myelin überzogen. Diese kleine technische Neuerung gegenüber niederen Tiergattungen steigert die Übertragungsgeschwindigkeit um ein Vielfaches, und begründet die evolutionsgeschichtliche Sonderstellung der Wirbeltiere im Tierreich.

Innerhalb der Nervenzelle pflanzt sich ein Signal auf elektrochemischem Wege über die Steuerung von Membranpotentialen fort. An den Synapsen werden die Signale chemisch durch Ausschüttung von Neurotransmittern übertragen, die sowohl erregend als auch hemmend wirken können. Diese binden sich an Rezeptoren auf der postsynaptischen Membran der Dendriten und verändern so deren Membranpotential. Im Zellkörper der Nervenzelle summieren sich die Potentiale der einzelnen Dendriten auf. Wird ein bestimmter Schwellwert überschritten, dann feuert das Neuron, d. h. es sendet ein Impulssignal über das Axon zu seinen Nachbarn, andernfalls bleibt es ruhig.

Technisch gesehen ist also eine Nervenzelle verblüffend einfach. Die kalorimetrischen Schwellentrigger von L1 im H1-Experiment arbeiten auf analoge Weise. Signale mehrerer Zellen werden summiert, verstärkt, und ab einer bestimmten Schwelle feuert der Trigger. Es bleibt jedoch die überraschende Frage zurück, wo die *Information* selbst eigentlich zu finden ist. Da die Nervenzelle nur Signale überträgt, kann sie selbst keine Information speichern. Also muß die Stärke der synaptischen Verbindung und die Topologie des Gehirns selbst das sein, was die Information enthält (für Details siehe [34, 35, 36]).

Der mathematische Modellansatz, ein Neuron als einen Summierverstärker mit individueller Gewichtung seiner Eingänge zu verstehen, ist durchaus naheliegend. Erste Ansätze zur Beschreibung neuronaler Netze wurden bereits 1943 von McCulloch und Pitts gemacht [37]. Im Modell werden Neuronen auch *Knoten* genannt. Das etwas komplizierte Übertragungsschema von Axonen und Dendriten wird durch Verbindungen mit variablen Gewichten ersetzt. Die Signale eines Eingabeknoten i zu einem Neuron j werden mit dem Gewicht ω_{ij} multipliziert. Als Aktivität o_j des Knotens j bezeichnet man

$$o_j = g\left(\sum_{i=1}^{N} \omega_{ij} x_i - t_j\right). \tag{4.5}$$

t_j entspricht dem Schwellwert, ab der ein Neuron feuern soll. McCulloch und Pitts wählten für die Aktivierungsfunktion g eine Stufenfunktion. Es ist aber sowohl biologisch plausibler, als auch für den im folgenden vorgestellten Algorithmus notwendig, eine sigmoide Aktivierungsfunktion zu wählen, wie z. B. die den Physikern wohlvertraute Fermifunktion

$$g(z) = \frac{1}{1 + \exp(-z)}. \tag{4.6}$$

Kennzeichnend für eine sigmoide Funktion ist ihr asymptotisches Sättigungsverhalten ($g(z) \rightarrow \{^1_0$ für $z \rightarrow \pm\infty$) und das Vorhandensein eines dynamischen, monoton ansteigenden Übergangsbereichs.

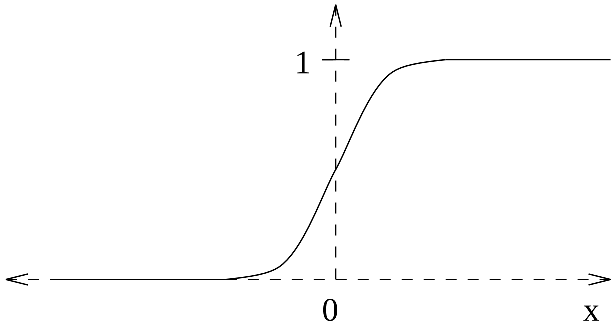

Abbildung 4.2: **Sigmoidfunktion.**

4.4 Welche Strukturen haben neuronale Netze?

Die Topologie oder Architektur hat entscheidenden Einfluß auf Funktionalität und Leistungsfähigkeit eines neuronalen Netzes. Auch im Gehirn sind die Neuronen in Schichten, Spalten oder zweidimensionalen topologischen Karten angeordnet, die z. B. die Hautoberfläche widerspiegeln.
Es gibt zwei grundsätzlich unterschiedliche Verknüpfungsverfahren von neuronalen Knoten. Bei vorwärtsgekoppelten *(Feedforward-)* Netzen empfängt ein Neuron weder direkt noch indirekt von den Neuronen Signale, an die es selbst Signale überträgt. Ein solches Netz besitzt immer zumindest eine Eingabeschicht *(Input-layer)*, das sind alle Knoten, die nur Signale weiterleiten, ohne von irgendwelchen anderen Knoten Signale zu empfangen, und eine Ausgabeschicht *(Output-layer)*, das sind alle Neuronen, die nur Signale empfangen, aber keine weiterleiten. Meistens werden die zwischen der Eingabe- und Ausgabeschicht liegenden Neuronen in einer oder mehreren versteckten Schichten angeordnet *(Hidden-layer)*. Wir werden uns aus Gründen, die später klar werden, bevorzugt auf die Topologie mit einer einzigen versteckten Schicht konzentrieren.
Die Natur bevorzugt eine andere Art von Architektur, das rückgekoppelte *(Feedback-)* Netz, bei dem das Ausgabesignal eines Neurons über mehrere Zwischenneuronen als Eingabesignal erneut auf das Ausgangsneuron rückwirken kann. Die Vorteile dieses Typs für die sensomotorische und vegetative Steuerung leuchten unmittelbar ein. Für Klassifikationsaufgaben eignen sich jedoch beide Netzwerktypen gleichermaßen, wobei Feedforward-Netze leichter technisch zu implementieren sind, und für das vorliegende Problem deshalb zu bevorzugen sind.

4.5 Wie wird ein Klassifikationsproblem implementiert?

Ein Neuron mit einer sigmoidalen Aktivierungsfunktion liefert als Ausgabe eine reelle Zahl im Intervall $[0, 1]$. Zur Darstellung von partionierten Klassen ist eine Diskretisierung erforderlich. Jeder Klasse kann z. B. ein eigenes Ausgabeneuron zugeordnet werden. Das Netzwerk kann in diesem Fall das in 4.2 gestellte Problem lösen, indem das Neuron mit dem höchsten Ausgabewert der gesuchten Klasse entspricht, die dem Eingabevektor $\vec{x_l}$ zugeordnet werden soll. Eine andere Möglichkeit besteht

darin, die Abbildung $\Re(\vec{x})$ aus Gleichung (4.3) mittels eines einzigen Ausgabeneurons zu realisieren. Letzteres ist für zwei Klassen sicher einfacher, wenn es nur um die Ja/Nein–Entscheidung geht. Für unser Problem wurde generell der Wert '0' als Zielwert *(target)* für die Klasse der Untergrundereignisse gewählt und '1' als Zielwert für die Physik. Der natürliche Schnitt für die Klassenprognose liegt bei 0.5, kann aber auch anders gewählt werden.
Klassifikationsprobleme sind eine Untermenge des allgemeineren Problems, eine beliebige Funktion $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit einer gegebenen Netzarchitektur durch geeignete Gewichte zu approximieren. A bezeichnet einen kompakten, n-dimensionalen Unterraum der Eingabedaten. In diesem ist eine endliche Anzahl N von Datenpaaren $(\vec{x_l}, \vec{y_l})$ mit $f(\vec{x_l}) = \vec{y_l}$ und einer festen Wahrscheinlichkeitsverteilung $\rho(\vec{x})$ gegeben. Aus der Gesamtmenge der Daten soll die Funktion f abgeleitet werden. Wollen wir dem Beispiel der Natur folgen, dann ist es unser Ziel, das neuronale Netz durch wiederholtes Präsentieren der Muster $(\vec{x_l}, \vec{y_l})$ dazu zu bringen, die Funktion f zu *erlernen*. Die nächsten Absätze gehen dieses Problem an.

4.6 Wie wird ein neuronales Netz bewertet?

Der Lernvorgang für ein neuronales Netz wird *Training* genannt. Die Begriffswahl ist suggestiv, hat aber eine engere Bedeutung als in der Umgangssprache. Während des Trainings wird das Netz mit Musterpaaren $(\vec{x_l}, \vec{y_l})$ gefüttert und daraus ein Klassifizierungsfehler errechnet. Anschließend wird versucht, die Gewichte so abzuändern, daß der Klassifizierungsfehler verkleinert wird. Dieser Prozeß wird solange iteriert, bis die gewünschte Appproximationsgenauigkeit erreicht oder die Geduld des Trainers erschöpft ist. Für unser Ausgangsproblem interessiert zunächst die relative Häufigkeit von Fehlklassifikationen. Für den im nachfolgenden Kapitel vorgestellten Backpropagation-Algorithmus(BPG) benötigt man aber ein kontinuierliches Fehlermaß wie

$$E(W) = \frac{1}{2} \sum_{l=1}^{N} |\vec{y}(\vec{x}_l) - \vec{o}(\vec{x}_l, W)|^2, \tag{4.7}$$

der Summe der quadratischen Abstände zwischen Zielvektor $\vec{y}(\vec{x}_l)$ und Ausgabevektor $\vec{o}(\vec{x}_l, W)$ über alle Musterpaare. W repräsentiert die Gesamtheit der Gewichte. Das gesuchte Netz ist dasjenige, das für $E(W)$ ein globales Minimum annimmt. Die Wahl anderer Fehlerfunktionen ist möglich und unter Umständen auch vorteilhafter.
Das obige Fehlermaß ist nicht völlig unproblematisch. Ein Sprichwort sagt, *non scholae, sed vitae discimus*. Das Analoge trifft auch auf ein neuronales Netz zu. Es interessiert nicht die Leistung im Training, sondern wie das Netz auf noch nie gesehene Eingabemuster reagiert. Präzise gesagt suchen wir eigentlich nach dem globalen Minimum der Funktion

$$E_\infty(W) = \frac{1}{2} \int_A |\vec{y}(\vec{x}) - \vec{o}(\vec{x}, W)|^2 \rho(\vec{x}) d\vec{x}, \tag{4.8}$$

welches erheblich von dem von $E(W)$ abweichen kann, wenn uns nicht genügend Eingabemuster zur Verfügung stehen. In diesem Fall wird das Netz zwar brav die

Trainingsmuster auswendig lernen, bleibt aber im unabhängigen Test hinter den Trainingsergebnissen zurück (*overfitting*; ein Problem, das auch bei multivariaten statistischen Regressionsmethoden auftritt).

4.7 Der Backpropagation-Lernalgorithmus

Obwohl dieser Algorithmus bereits 1974 von P.Werbos [38] vorgestellt wurde, verhalf ihm erst 1986 die Wiederentdeckung von McClelland und Rumelhart [39] zum Durchbruch. Erst die erheblich gestiegene und preisgünstigere Rechenkapazität gegenüber den 70er Jahren ermöglichte es, diesen Algorithmus erfolgreich auf eine Vielzahl sehr unterschiedlicher Probleme anzuwenden. Seine Überlegenheit gegenüber früheren Konzepten beruht auf der Möglichkeit, Gewichte in verborgenen Schichten mehrlagiger Netze zu adjustieren. Im folgenden sei der Algorithmus für ein dreilagiges Netz mit den Bezeichnungen aus Abb 4.3 beschrieben; eine Erweiterung auf mehrere Schichten ist leicht zu realisieren. Gehen wir von Gleichung (4.5) aus. Der Schwellwert t_j läßt sich

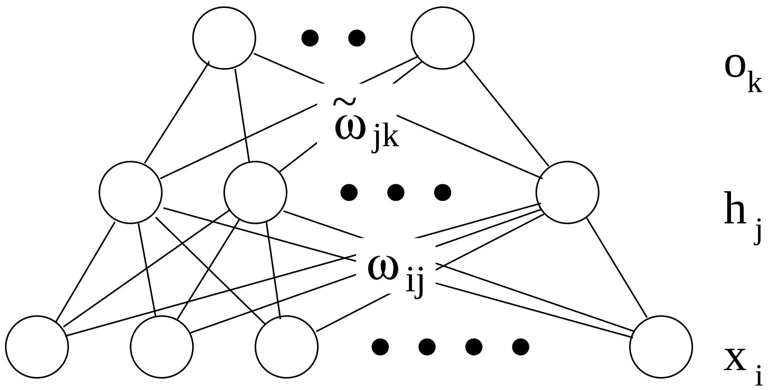

Abbildung 4.3: **Ein dreilagiges Feedforward-Netzwerk.** *$\vec{x}$ ist der Eingabevektor, $\vec{o}$ der Ausgabevektor und $\vec{h}$ der Ausgabevektor der Hidden-Layer. Die Gewichte zwischen den einzelnen Schichten können durch Matrizen ω_{ij} bzw. $\tilde{\omega}_{ij}$ beschrieben werden.*

in den Gewichtsvektor integrieren, indem wir einen zusätzlichen Eingabeknoten mit konstantem Wert $x_0 = -1\,(t_j = x_0\omega_{0j})$ hinzufügen. Damit berechnet sich die Aktivität eines Neurons aus der Hidden-Layer h_j und aus der Output-Layer o_k zu

$$\begin{aligned} h_j &= g\left(\sum_{i=0}^{n} \omega_{ij} x_i\right) \\ o_k &= g\left(\sum_{j=0}^{J} \tilde{\omega}_{jk} h_j\right), \end{aligned} \tag{4.9}$$

wobei J der Anzahl der versteckten Knoten und n der Dimension des Eingaberaums entspricht. g ist eine sigmoidale Aktivierungsfunktion, häufig wird die Fermifunktion benutzt wegen ihres einfachen Zusammenhangs mit ihrer Ableitung.
Ziel des Backpropagation-Lernalgorithmus(BPG) ist es, das globale Minimum der in (4.7) definierten Funktion E mittels eines Gradientenabstiegsverfahrens zu finden. Wie

bei jeder Gradientenmethode steht die Wahl des Startpunkts zur Diskussion. Es genügt meistens, die Gewichte W mit zufälligen Werten aus einem Intervall $[-\epsilon, \epsilon]$ zu initialisieren, um Symmetrien zwischen verschiedenen Gewichtsvektoren aufzuheben.
In einer Abänderung des Algorithmus *(Online-Updating)* kann die Veränderung der Gewichte auch Muster für Muster vorgenommen werden. Dann entfällt in (4.7) der Index l. Für die ursprüngliche Form des Algorithmus *(Offline-Updating)* ist im folgenden der Term $(o_k - y_k)$ um die Summation über alle Muster zu ergänzen. Beim Online-Updating wird in jedem Lernschritt ein Muster zufällig aus der Trainingsmenge herausgegriffen und $\vec{o}(\vec{x}, W)$ berechnet. Daraus ergibt sich ein Fehler von $E(W) = \sum(o_k - y_k)^2$. Jetzt wird der Fehlergradient Schicht für Schicht zurückverfolgt (daher der Name des Algorithmus). Zuerst werden die Gewichte zwischen Ausgabe- und versteckter Schicht in Richtung des stärksten Abstiegs auf der Oberfläche der Fehlerfunktion abgeändert:

$$\Delta\tilde{\omega}_{jk} = -\eta\frac{\partial E}{\partial\tilde{\omega}_{jk}} + \alpha\Delta\tilde{\omega}_{jk}^{alt} = -\eta\delta_k h_j + \alpha\Delta\tilde{\omega}_{jk}^{alt} \tag{4.10}$$

$$\text{mit } \delta_k = (o_k - y_k) \cdot g'(\widetilde{\vec{\omega}_k} \cdot \vec{h}). \tag{4.11}$$

Damit ist auch klar, warum eine sigmoidale Aktivierungsfunktion gebraucht wird; g' ist für solche Funktionen positiv und nur in einem bestimmten dynamischen Intervall wesentlich von 0 verschieden. Mit Hilfe der Kettenregel ergeben sich die nachfolgenden Gewichtsänderungen zwischen Eingabe- und versteckter Schicht

$$\Delta\omega_{ij} = -\eta\frac{\partial E}{\partial\omega_{ij}} + \alpha\Delta\omega_{ij}^{alt} = -\eta\delta_j x_i + \alpha\Delta\omega_{ij}^{alt} \tag{4.12}$$

$$\text{mit } \delta_j = \sum_{k=1}^{m} \tilde{\omega}_{jk}\delta_k g'(\vec{\omega}_j \cdot \vec{x}). \tag{4.13}$$

Eine Wiederholung der Prozedur für etwaige nachfolgende versteckte Schichten läuft nach demselben Schema ab. Die Schrittweite η wird als Lernparameter bezeichnet. Seine Größe ist nicht aus dem Algorithmus ableitbar und muß für jedes Problem neu ermittelt werden. Erfahrungsgemäß ist sein Einfluß zwar groß auf die Lerngeschwindigkeit, günstigerweise aber weniger auf die erreichbare Leistungsfähigkeit. Das Hinzuaddieren der Gewichtsänderungen $\Delta\omega_{ij}^{alt}$ und $\Delta\tilde{\omega}_{jk}^{alt}$ aus dem vorangegangenen Lernschritt versehen mit einem Dämpfungsterm α, auch *Momentumparameter* genannt, stellt eine Verbesserung des ursprünglichen Algorithmus dar. Regionen mit flachem Gradientenabfall werden schneller überwunden und Oszillationen vermieden. Wie für jedes Gradientenverfahren besteht das Risiko, in einem lokalen Minimum zu landen. Aufgrund der großen Anzahl von Gewichtsparametern ist dies in der Praxis selten der Fall. Dennoch ist es ratsam, das gleiche Netz mehrmals mit verschiedenen Startparametern zu trainieren.

4.8 Was macht die versteckte Schicht?

Wenn wir die Fermifunktion als Aktivierungsfunktion wählen, dann liegen alle Punkte des n-dimensionalen Raums der Eingangsvariablen, die für diese Funktion den Wert 0.5

annehmen, auf einer Hyperebene, die durch die Gleichung $\vec{\omega}_j\vec{x} = t_j$ beschrieben wird. Der gerichtete Abstand $a = \vec{\omega}_j\vec{x} - t_j$ eines beliebigen Punktes zur Hyperebene ist nichts anderes als das Argument der Aktivierungsfunktion selbst. Der dynamische Übergangsbereich, in dem die Funktion von 0 auf 1 anwächst, liegt in einem Streifen um die Hyperebene, dessen Breite mit $1/|\vec{\omega}|$ skaliert. Längere Gewichtsvektoren entsprechen also einer relativ scharfen Stufenfunktion entlang einer Hyperebene, während kleine Gewichte einen sanften Übergang erzeugen.
Eines der frühen Modelle eines neuronalen Netzes, das *Perzeptron*, unterschied sich von der Netzarchitektur in Abb. 4.3 durch das Fehlen einer versteckten Schicht und der Verwendung einer Heavyside-Funktion als Aktivierungsfunktion. Obige Betrachtung erlaubt, ein solches Perzeptron als multidimensionale *lineare* Schnitte zu verstehen, die den Eingaberaum entlang von Hyperebenen in zwei Hälften separieren. Es konnte zwar bewiesen werden, daß eine ähnliche δ-Lernregel wie in (4.10) für alle linear separierbaren Probleme zur Konvergenz führt, aber ebenso gelang der Gegenbeweis für sehr einfache nicht linear separierbare Probleme, wie dem exklusiven Oder aus der Prädikatenlogik [40]. Dieser Schwachpunkt hat der frühen Entwicklung neuronaler Netze schwer geschadet, erhellt aber auch die Bedeutung einer versteckten Schicht. Läßt man die Eingabeschicht unberücksichtigt, dann entsprechen Ausgabe- und versteckte Schicht zusammen der ursprünglichen Perzeptronarchitektur, die nur linear separierbare Probleme lösen kann. Also besteht die Aufgabe der versteckten Schicht darin, für jedes nicht linear separierbare Problem eine interne linear separierbare Darstellung zu finden.

4.9 Wie lernt ein neuronales Netz ?

Ein kleines Beispiel soll die Funktionsweise des Algorithmus näher erläutern. Dem in Abb. 4.4a dargestellten kleinen Netz wurden im Online-Updating jeweils drei verschiedene Punkte präsentiert. Die Hyperebene, die dem versteckten Knoten nach Ausführung eines Backpropagation-Schritts ($\eta = 6.0$, g aus (4.6)) entspricht, ist in b)-d) durch die durchgezogene Gerade dargestellt. Die ursprüngliche Lage ist durch die mittlere gestrichelte Linie angedeutet. Der Betrag der Lernrate wurde unrealistisch hoch gewählt, um die Gewichtsänderungen auch visualisieren zu können. Die ursprünglichen Gewichte sind so gewählt, daß der Bereich unterhalb der Geraden der Physik($y = 1$) und darüber dem Untergrund($y = 0$) entspricht. Die die Hyperebenen umhüllenden Streifen entsprechen dem Bereich, für den $|h - 0.5| < 0.23$ gilt. Er soll insbesondere Breite und Lage des Übergangsbereichs andeuten. Gestrichelte Linien deuten die ursprüngliche Lage dieses Bereichs an.
In b) wurde das Netz mit einem Punkt auf der *richtigen* Seite der Hyperebene trainiert. Nicht unerwartet trägt der versteckte Knoten nach Durchführung des Lernschritts mehr zum Ausgabewert bei. Die Änderungen sind jedoch relativ klein, weil der Fehler klein ist. Dadurch konvergiert die Annäherung an steile Übergänge auch sehr langsam. In c) wurde dem Netz derselbe Punkt, aber nun als Untergrundereignis, präsentiert. Alle Änderungen der Gewichte sind entgegengesetzt zu b) und zum Ausgabefehler proportional größer. In d) wurde jetzt ein Punkt dem Netz präsentiert, der noch weiter auf

a) b) c) d)

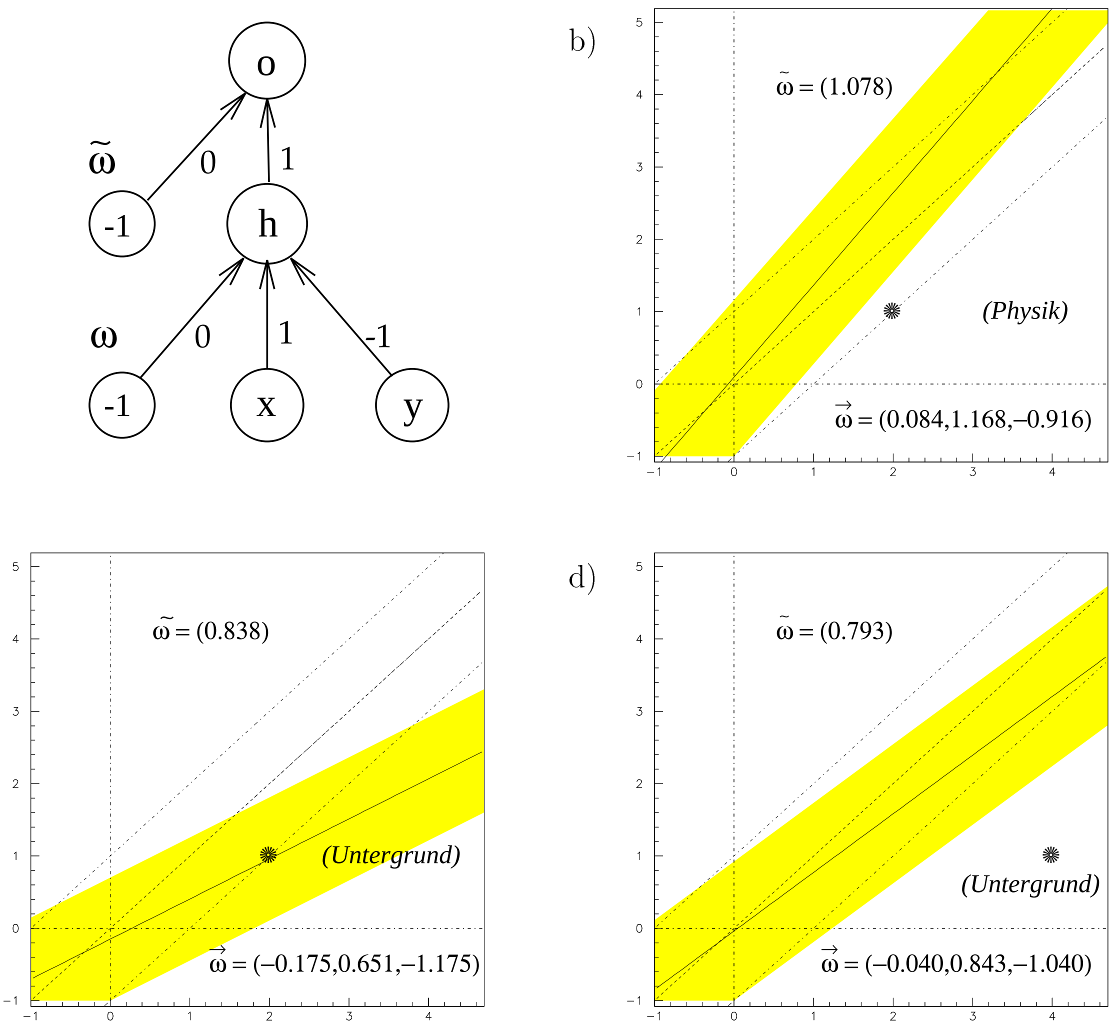

Abbildung 4.4: **Online-Updating eines kleinen BP-Netzes(a).** *In b) – d) sind jeweils die Änderungen der Gewichte ω_{ij} graphisch durch die entsprechenden Hyperebenen und ihre umliegenden Streifen dargestellt. Der präsentierte Punkt ist durch ein ($\star$) markiert.*

der falschen Seite des Schnittes liegt. Überraschenderweise sind jetzt aber die Änderungen in ω kleiner ausgefallen als in c). Um das zu verstehen, betrachten wir nochmals Gleichung (4.12)

$$\Delta\omega_{ij} = -\eta\delta_j x_i \text{ mit } \delta_j \sim g'(\vec{\omega_j} \cdot \vec{x}).$$

Es sind zwei unabhängige Größen, die $\Delta\omega_{ij}$ beeinflussen. Zum einen der Betrag von x_i selbst, zum anderen der gewichtete Abstand zur Hyperebene über das Argument von g'. Wegen $\lim_{z\to\pm\infty} g'(z) = 0$ und einem Maximum von g' bei $z = 0$ beeinflussen Punkte, die nahe bei der Hyperebene liegen, die Gewichte mehr als ferne Punkte. Bereiche außerhalb solcher Streifen werden deshalb sehr schlecht gelernt. Eine ungeschickte Initialisierung kann z. B. dazu führen, daß für das Netz keine Konvergenz beobachtet werden kann, wenn alle Datenpunkte außerhalb der dynamischen Übergangsbereiche

der versteckten Knoten liegen.
Die Abstandsabhängigkeit hat aber auch einen sehr großen Vorteil im Hinblick auf das Zusammenwirken mehrerer versteckter Knoten. Um konkave oder ringförmige Strukturen auf Feedforward-Netze abzubilden, werden versteckte Knoten mit verschobenen, diametral entgegengesetzten Gewichtsvektoren benötigt. Das bedeutet für viele Punkte, daß sie einmal auf der falschen und einmal auf der richtigen Seite einer Hyperebene liegen. Damit das Netz während des Lernens diesen Konflikt lösen kann, ist es entscheidend, daß jeder Knoten nur auf Punkte in seinem eigenen Streifen sensitiv ist.
Um die Leistungsfähigkeit des Algorithmus zu demonstrieren, sind in Abb. 4.5 und Abb. 4.6 die Ergebnisse für zwei komplexe Kartographierungsprobleme gezeigt. Dem Netz wurden in zufälliger Reihenfolge zwei Koordinaten x und y als Eingabewerte präsentiert und gesagt, ob der Punkt sich im Gebiet '1' oder '0' einer 2-dimensionalen Karte befindet. Buchstaben mit ihren Kanten, Einbuchtungen und Löchern sind sehr schwierig zu erlernende Karten, eignen sich aber gut, um die Ausgabefunktion $o(x, y, W)$ optisch wiederzugeben.
Bei den Abbildungen handelt es sich um Konturgraphen dieser Ausgabefunktionen, wobei dunklere Schattierungen größeren Werten der Ausgabefunktion entsprechen. Es ist offensichtlich, daß die Netze die gestellten Probleme hervorragend lösen können. Für das *H1*-Problem wurde eine Architektur von $2 \times 40 \times 1$ Knoten gewählt und im Falle des *Neuron*-Problems $2 \times 89 \times 10 \times 1$ Knoten. Letzteres Problem konnte auch mit einem dreilagigen Netz gelöst werden. Es wurden jedoch etwa $\mathcal{O}(10^8)$ Lernschritte benötigt, um das Netz auszutrainieren. Die Konvergenz solcher komplexen Probleme ist zudem sehr sensitiv gegen die Wahl von Trainingsparametern, u. a. von der Größe des Intervalls, in dem die Gewichte zufällig initialisiert werden. In Abb 4.6 ist gut zu erkennen, wie das Netz zunächst die Umrisse der konvexen Hülle der Strukturen erlernt und sich dann in die konkaven Einbuchtungen und Löcher hineinarbeitet. Diese Eigenschaft der allmählich fortschreitenden Interpolation erlaubt eine Approximation ohne a priori Annahmen über Wahrscheinlichkeitsverteilungen zu machen, im Gegensatz zu den Methoden der multivariaten Regressionsanalyse.

4.10 Was kann ein neuronales Netz leisten?

Darstellbarkeit von Abbildungen

Eine wichtige Frage ist, ob jede beliebige Funktion durch ein dreilagiges Feedforward-Netz darstellbar ist. Es ist nicht leicht, darauf eine Antwort zu finden, aber ein 1957 von Kolmogorow [41] bewiesenes Theorem erlaubt es, eine begrenzte Existenzaussage zu formulieren.
Bezogen auf ein dreilagiges Feedforward-Netz besagt es, daß jede kontinuierliche Funktion $f : [0,1]^n \to \mathbb{R}^m, \ \mathbf{f}(\vec{x}) = \vec{y}$ durch ein neuronales Netz mit n Eingabe-, $2n+1$ versteckten und m Ausgabeknoten exakt dargestellt werden kann [42, 29].
Der Beweis macht gewisse einschränkende Annahmen über die zu verwendenden Aktivierungsfunktionen, bis heute existiert aber kein konstruktives Verfahren zur Bestimmung dieser Aktivierungsfunktionen. Es ist nicht einmal ein einziges nichttriviales

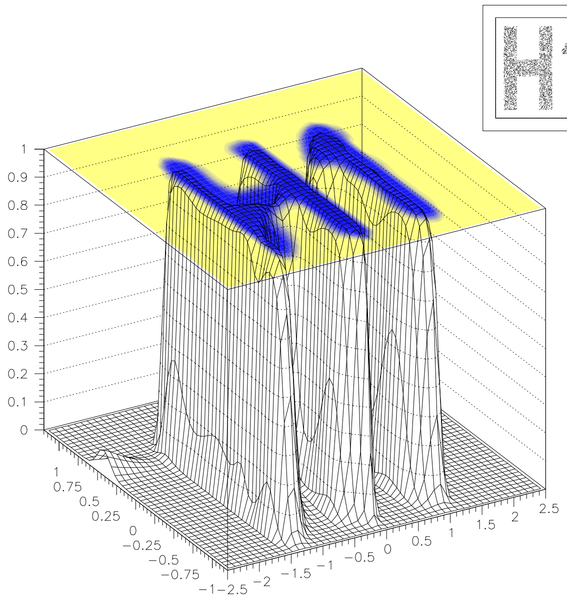

Abbildung 4.5: *Ausgabefunktion $o(x, y, W)$ eines neuronales Netzes, das auf die Vorlage in der rechten oberen Ecke trainiert wurde.*

Beispiel eines solchen Netzes bekannt. Obwohl durchaus bemerkenswert, ist das Theorem in praktischer Hinsicht von geringer Bedeutung.
Eine sehr viel wichtigere Frage ist, ob es möglich ist, jede Funktion konsistent mit unserem Lernalgorithmus beliebig genau zu approximieren. Diese Frage konnte positiv beantwortet werden [43]. Ein Beweis dafür, daß für jede quadratintegrable Funktion $f \in L_2$; $f : [0,1]^n \to \mathbb{R}^m$ mit jedem $\epsilon > 0$ ein Backpropagation-Netzwerk existiert, das f mit einem kleineren quadratischen Fehler als ϵ approximiert, findet sich in [44].
Diese Aussage gibt aber keinen Hinweis, ob und wie wir zu den Gewichten dieses Netzes gelangen. Die Konvergenzeigenschaften von Lernalgorithmen sind ein aktuelles Forschungsgebiet der Theorie.
Auch wenn diese Ergebnisse nicht unwichtig sind, so gehen doch immer ideale Annahmen ein, die in der Praxis nicht leicht zu verwirklichen sind. Limitationen in der Zeit, Rechengenauigkeit und -kapazität sind Randbedingungen, die wesentlich für eine erfolgreiche Anwendung miteinbezogen werden müssen.

Generalisierung

Jede Trainingsmenge besteht aus einer endlichen Anzahl von Elementen, die die wahre Wahrscheinlichkeitsdichte $\rho(x)$ nur approximiert. Zwischen den einzelnen Punkten aus der Trainingsmenge muß das Netz also interpolieren. Der Backpropagation-Algorithmus kann aber nicht garantieren, die optimale Interpolation zu finden. Das ist der Preis, der zu bezahlen ist, wenn man einen völlig allgemeinen Algorithmus ohne Annahmen über Wahrscheinlichkeitsverteilungen haben möchte.
Das Generalisierungsverhalten ist zwar meist sehr gut, zeigt jedoch eine charakteristische Besonderheit. Prüft man während des Trainings mit einer unabhängigen

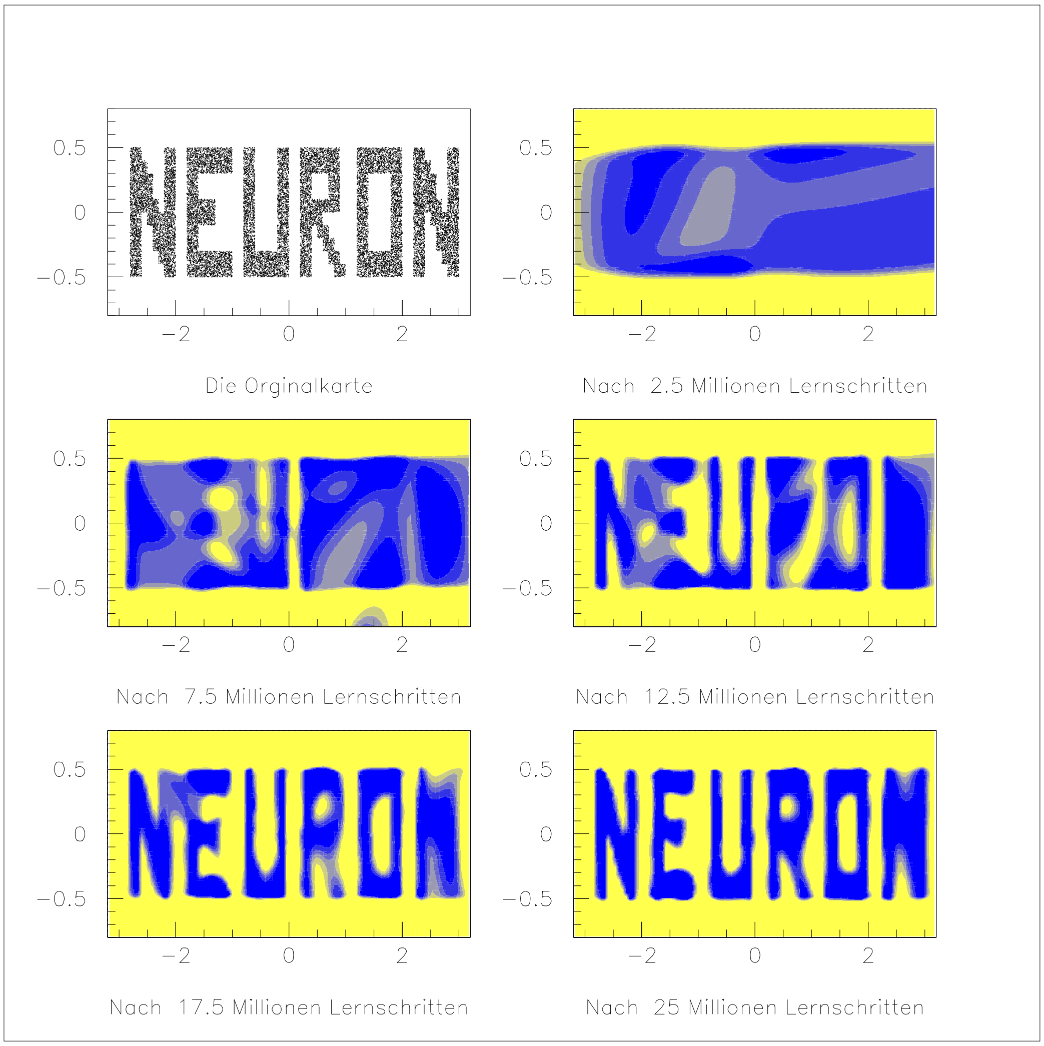

Abbildung 4.6: *Entwicklung der Ausgabefunktion mit fortschreitender Lerndauer für das in der linken oberen Ecke dargestellte Kartographierungsproblem des Wortes* **NEURON**.

Testmenge den Generalisierungsfehler, so wird irgendwann ein Minimum erreicht, in welchem das Netz am besten generalisiert, danach steigt der Fehler für die unabhängige Testmenge wieder, obwohl sich der Fehler für die Trainingsmenge weiter verringert. Dieses Verhalten wird *Overtraining* genannt. Es hängt entscheidend vom Verhältnis der Gesamtanzahl an Gewichtsparametern zu der Anzahl an Trainingsmustern N_W/N ab [45]. Als eine Erfahrungsregel hat sich bewährt, mindestens zehnmal mehr Eingabemuster zu benutzen, als das Netz Gewichte hat, um zu einem vernachlässigbaren Generalisierungsfehler zu gelangen.

Theoretisch ist das Generalisierungsverhalten noch nicht gut verstanden, eine zwanglose Erklärung ergibt sich jedoch im Zusammenhang mit den Betrachtungen aus dem letzten Abschnitt. Die Konturen statistischer Verteilungen haben in mehrdimensionalen Räumen meistens eine konvexe Form (das gilt insbesondere für Gaußverteilungen), während die statistischen Fluktuationen in diesen Konturen kleinere Inseln und Einbuchtungen bilden. Das Netz beginnt zunächst damit, die konvexen

Konturen der Verteilung zu erlernen und leistet dabei die geforderte Interpolation. Mit fortschreitender Lerndauer werden aber auch die statistischen Fluktuationen erlernt, und die Generalisierung beginnt sich zu verschlechtern. Daraus ergibt sich die Konsequenz, die statistischen Fluktuationen durch Verwendung ausreichender Statistik für die Trainingsmenge auf ein vertretbares Maß zu reduzieren.

Kapitel 5

Die Daten für das neuronale Netz

Ein Triggersystem soll überschaubar sein und eine kontinuierliche Qualitätskontrolle ermöglichen. Für einen Neuronalen-Netz-Trigger heißt das, aus dem Gesamtstrom der Daten möglichst signifikante Eingabegrößen zu bilden, um die Dimension n des Eingabevektors $\vec{x}$ auf ein Minimum zu reduzieren. Das folgende Kapitel beschreibt die experimentelle Datenerzeugung und -extraktion für einen neuronalen L2-Trigger beginnend mit einem Überblick über das Triggersystem von H1.

5.1 Das Triggersystem von H1

Die Elektronen- und Protonenbündel durchdringen sich bei HERA mit einer Frequenz von $10.46\,MHz$, das entspricht einem Zeitintervall von $96ns (= 1BC$; *bunch cross).* Dieses Zeitintervall ist unzureichend für eine komplette Auslese und Speicherung der experimentellen Daten. Es ist aber nicht notwendig, für jede Bündeldurchdringung Daten auszulesen, da auch bei Designluminosität die Rate an interessanten ep-Kollisionen bei $\mathcal{O}(10Hz)$ liegt. Die Datenaufzeichnungsrate des H1-Experiments ist etwa auf diese Größenordnung ausgelegt. Die komprimierte Gesamtdatenmenge pro Ereignis beträgt etwa $70KByte$.
Die Aufgabe des Triggersystems ist, die etwa 10 interessanten Ereignisse aus der Menge von 10^7 Ereigniskandidaten pro Sekunde zu erkennen. Zunächst ist zu bemerken, daß die $96ns$ zwischen zwei Bündeldurchdringungen nicht ausreichen, um eine Triggerentscheidung zu fällen. Die Driftzeiten in den Spurkammern betragen $\sim 1\mu s$. In derselben Größenordnung liegen die Integrationszeiten der kalorimetrischen Vorverstärker. Da aber die Untergrundrate mit $\sim 100\,kHz$ wesentlich kleiner als die Durchdringungsrate von $10.46\,MHz$ ist, erlaubt ein Pipeline-Verfahren eine totzeitfreie erste Triggerstufe mit einer verzögerten Triggerentscheidung nach $2.2\mu s$. Die Untergrundrate ist jedoch zu hoch, um mit einer einzigen Triggerstufe die geforderte Reduzierung der Rate von 10^6 zu erreichen. Deshalb wurde ein vierstufiges Triggerkonzept mit zwei Hardwarestufen(L1,L2) und zwei Softwarestufen(L3,L4) entwikkelt. Die Aufgliederung in mehrere Stufen macht Sinn, da es mehrere ms dauert, bis alle Rohdaten eines Ereignisses ausgelesen sind, vorher aber schon zu verschiedenen Zeitpunkten Daten verfügbar gemacht werden können, auf deren Basis Triggerentscheidungen getroffen werden können. Je früher ein Untergrundereignis verworfen wird, desto

weniger Totzeit muß im Experiment in Kauf genommen werden. Die Totzeit selbst errechnet sich aus dem Produkt von Ereignisrate und Verarbeitungszeit. Tab 5.1 gibt einen Überblick über die einzelnen Triggerstufen [13]. Die maximalen Raten errechnen sich aus der Randbedingung, daß die L1 nachfolgenden Triggerstufen weniger als 10% Totzeitanteil erreichen sollen.

Stufe	Verarbei- -tungszeit	max. Rate	Information
L1	$2\,\mu s$	$5 \cdot 10^3\,Hz$	8-Bit Energiesummen, festverdrahtete Logik für Spurmuster & Koinzidenzen
L2	$20\,\mu s$	$100\,Hz$	Triggerdaten verschiedener Subdetektoren, z. B. 8-Bit Energiesummen, Bitfelder für Spurrichtungen, etc.
L3	$1\,ms$	$50\,Hz$	dito, mit einigen zusätzlichen Daten
L4	$30\,ms$	$\sim 5\,Hz$	alle Detektordaten (im Rohformat)

Tabelle 5.1: **Unterteilung des H1-Triggersystems.**

- **Die Triggerstufe L1**

Die einzelnen Subkomponenten des Detektors liefern an die zentrale Triggerlogik(CTL1) Triggersignale, die sogenannten *Triggerelemente*. Jedes entspricht einem von 128 Bits in der CTL1. Über frei adressierbare Speicher werden aus diesen 128 Triggerelementen logische Verknüpfungen gebildet, die *Subtrigger* genannt werden. L1 gibt ein Ereignis an die nächsthöhere Triggerstufe weiter (*L1-Keep*), sobald einer dieser Subtrigger anspricht. Diese können auch herunterskaliert werden, so daß empfindliche Subtrigger mit eigentlich sehr hohen Ereignisraten verwirklicht werden können. Die Zeit von der Wechselwirkung bis zur L1-Entscheidung beträgt $\sim 2.2\mu s\,(22\,BC)$. Solange werden die Daten in einer Pipeline gehalten, die je nach Subdetektor 27 bis 35 BC lang ist. Bei einer positiven L1-Entscheidung wird die Pipeline angehalten und die Totzeit beginnt. Das Pipeline–Verfahren hat auch den Vorteil, daß die Historie eines Ereignisses miterfaßt werden kann. Dadurch können eventuell auftretende Timingprobleme sehr schnell erkannt und behoben werden.

- **Die Triggerstufe L2**

Unmittelbar $(1\,BC)$ nach L1-Keep beginnt die Auslese der L2–Daten und die Weitergabe an den L2–Datenbus. Deshalb handelt es sich bei den L2–Daten zu einem guten Teil um die Daten, die Grundlage für die L1-Entscheidung waren. Da aber diese Daten in ihrer Gesamtheit und mit der entsprechenden Ortsinformation vorliegen, gehen die Möglichkeiten weit über die logische Verknüpfung einzelner Bits hinaus. Topologische Koinzidenzen und komplexe Korrelationen können auf dieser Triggerstufe ausgenutzt werden. Im Falle einer positiven L2-Entscheidung wird die Auslese aller Rohdaten eines

Ereignisses gestartet. Das nachfolgende Unterkapitel geht näher auf diese Triggerstufe ein.

- **Die Triggerstufe L3**

Auf L3 erlaubt ein AM29k RISC Prozessor die erste Anwendung softwaremäßig implementierter Algorithmen [46]. Im Falle einer negativen Entscheidung wird die mit L2-Keep eingeleitete Datenauslese abgebrochen und das Experiment innerhalb weniger Mikrosekunden neu gestartet. Da hier alle L2-Daten zur Verfügung stehen, eignet sich diese Triggerstufe auch für die Simulation von L2-Algorithmen. Im Jahr 1993 war die Datenauslese für L3 erst im Entstehen, so daß dieses Vorhaben für 1994 anvisiert wird.

- **Die Triggerstufe L4**

Im Gegensatz zu L2 und L3 arbeitet L4 asynchron, da ein Pufferspeicher (*Multi-Event-Buffer*) vorgeschaltet ist. L4 stehen die gesamten Rohdaten eines Ereignisses zur Verfügung. Diese Triggerstufe ist in das allgemeine Datennahmesystem integriert und entscheidet, ob ein Ereignis auf Band geschrieben wird oder nicht. Realisiert ist diese Stufe durch eine Farm schneller *mips* R3000 Prozessoren. Um die Entscheidungszeit zu minimieren, werden die Farmalgorithmen hierarchisch modularisiert, so daß zu rekonstruierende Größen nur errechnet werden, wenn sie von einem Modul angefordert werden. Das erlaubt, Teile der *Off line*-Standardanalyse in die Filteralgorithmen zu integrieren. 1992 und 1993 wurden in erster Linie Ereignisse mit Spuren aus dem rückwärtigen Bereich verworfen (ca. $(70 - 80)\%$ der Daten). Etwa 1% dieser zurückgewiesenen Ereignisse werden dennoch für Monitorzwecke aufgezeichnet. Diese Untergrundereignisse können für das Training neuronaler Netze herangezogen werden.

5.2 Die L2–Daten

5.2.1 Datenfluß

Das L2–Datennahmesystem ist in die Standard VME STC Umgebung von H1 integriert. Ein Crate des PQZP-Sytems (*Parallel Quickbus Zero-suppressed Processor*) umfaßt bis zu 8 Speicherkarten und eine L2L3–Karte. Die Speicherkarten spielen eine Übermittlerrolle. Solange L1 nicht triggert, empfangen die Speicherkarten Daten($4 \times 16\,Bit$) von L1–Triggersystemen und füllen sie in eine Pipeline mit der HeraClock-Rate von $10.46\,MHz$. Bei einem L1-Keep–Signal werden die Speicherkarten sequentiell von der L2L3–Karte mit derselben Frequenz ausgelesen. Diese Karte beschickt den L2– und L3–Bus mit den Daten. *Crossbar Switches* zwischen L2–Bus und L2L3–Karte erlauben eine blockweise Umsortierung. Um 8×4 16Bit-Worte auszulesen werden 32 Zyklen benötigt, also ca. $3.2\,\mu s$. Zwei PQZP–Systeme können im Gegentakt so betrieben werden, daß sich ihre Datenströme mit doppelter HERA–Frequenz auf dem L2–Bus überlagern. Insgesamt 16 PQZP–Busse können den $128Bit$ breiten L2–Datenbus gleichzeitig beliefern [47]. In den $3.2\,\mu s$ können also maximal $128 \times 64Bit = 1KByte$ übertragen werden. Die Flüssigargon-Triggerdaten werden

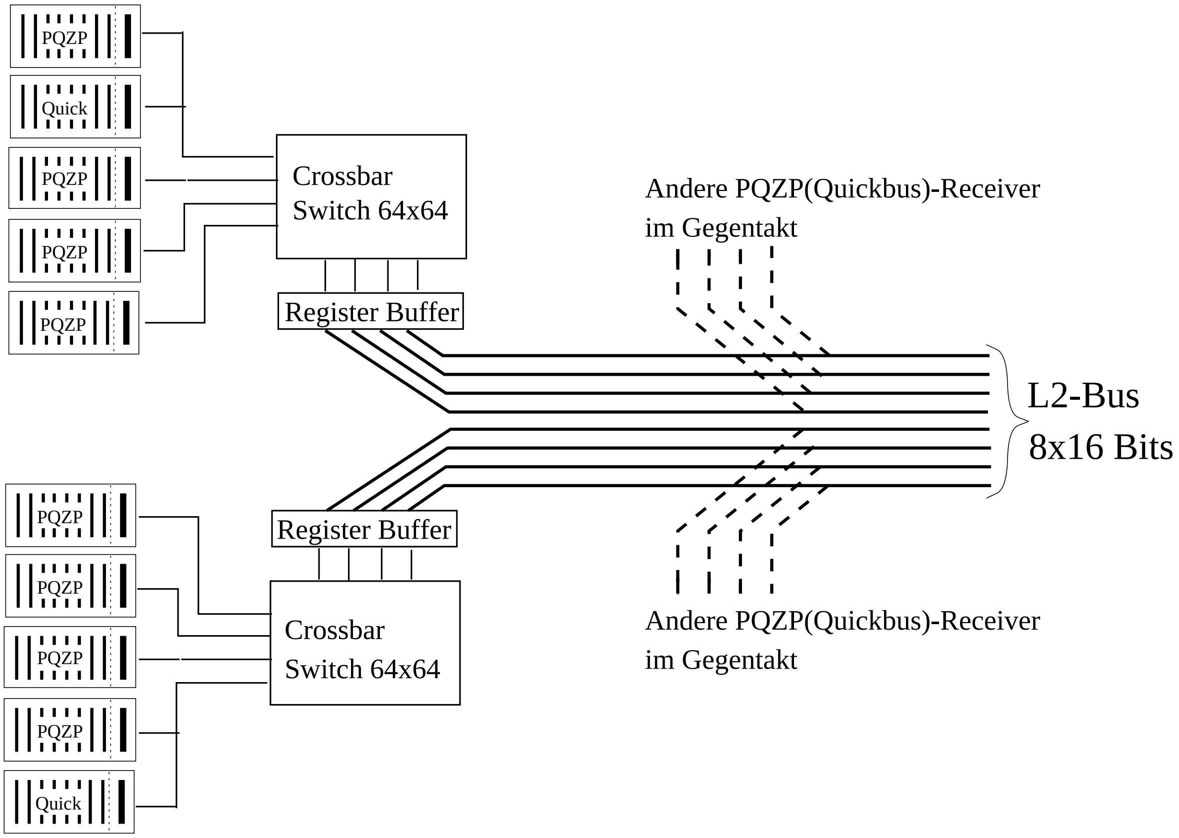

Abbildung 5.1: **Das Schema des L2-Datennahmesystems.**

adresskodiert über ein eigenes *Quickbus*-System im Anschluß an die PQZP–Daten ausgelesen. Hier kann die Datenmenge variieren. Maximal kann es etwa noch einmal $\sim 1.6\mu s$ dauern, bis die Datenauslese komplett abgeschlossen ist. Der L2–Neurotrigger selbst besteht aus zwei Teilen, einer *DataBox*, die den L2–Bus ausliest und daraus die Netzeingabegrößen erzeugt, und einem Parallelcomputer (CNAPS), der ein neuronales Netz im Feed-Forward-Modus betreibt.

5.2.2 Spurkammern

Die Vieldrahtproportionalkammern (MWPC's)

Das zentrale Spurkammernsystem umgibt schalenförmig die Wechselwirkungszone (s. Abb. 5.2). Der innerste und zweitäußerste Zylinder sind mit Vieldrahtproportionalkammern ausgestattet (CIP; *Central Inner Proportional Chamber* und COP; *Central Outer Proportional Chamber*). Die Anodendrähte aus *Al* oder *Cu* sind parallel zur Strahlachse montiert. Beide Kammern bestehen aus 3 Kathodenebenen, die zwei Drahtebenen umgeben. Die Kathodenebenen sind für die CIP in $8\Phi \times 60z$ und für die COP in $32\Phi \times 60z$ sogenannte *Pads* unterteilt. Die zwei Teilsysteme der CIP sind um $\pi/16$ verschoben, so daß sich für die innere Driftkammer insgesamt eine 16fache Segmentierung in Φ ergibt.
Diese Kammern dienen der Bestimmung der z-Komponente des Wechselwirkungsver-

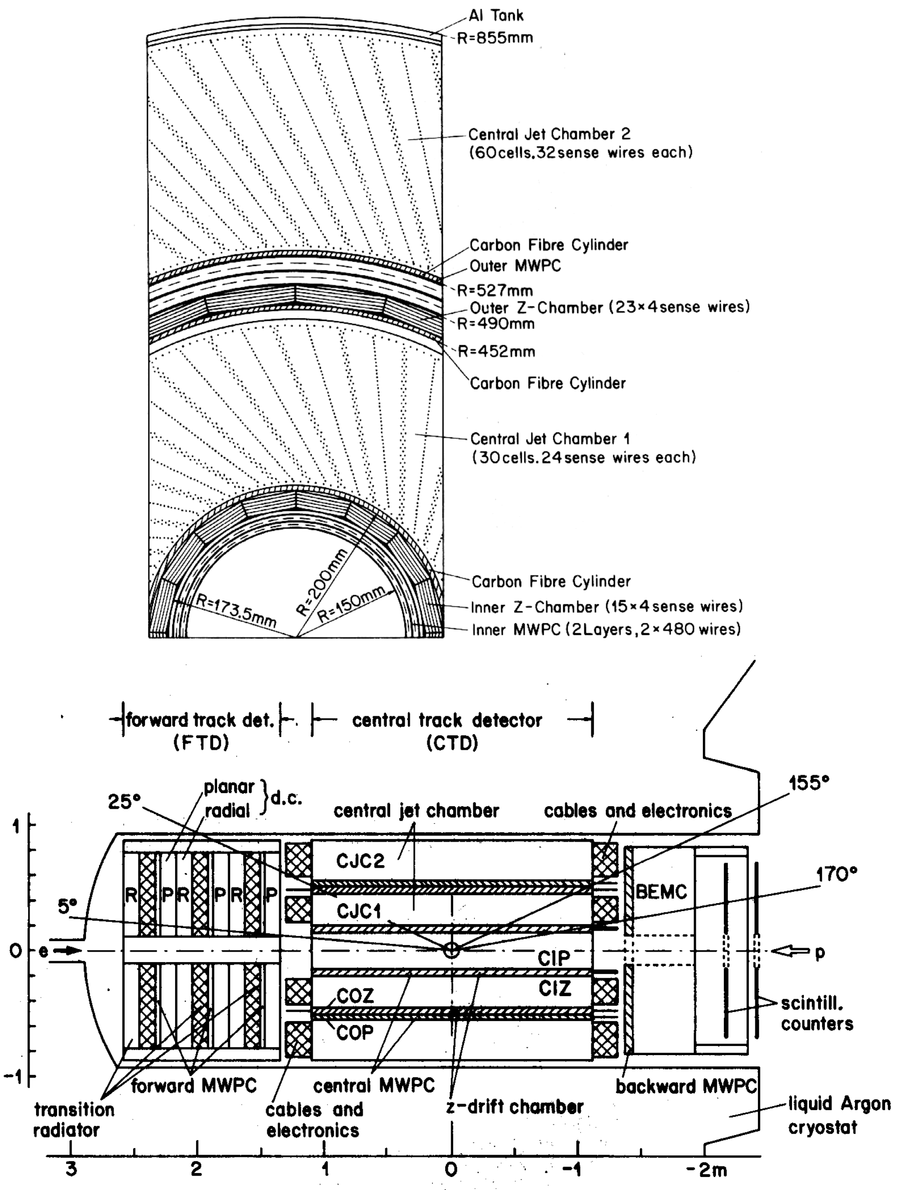

Abbildung 5.2: **Das Spurkammernsystem von H1.** *Oben ist die r-Φ–Ansicht dargestellt, unten ein Querschnitt durch die r-z–Ebene.*

tex zu Triggerzwecken. Aus der Koinzidenz von Pad-Signalen in der CIP und COP lassen sich grob Spuren rekonstruieren, die die Strahlachse schneiden. Die errechneten Schnittpunkte werden in ein Histogramm eingetragen, dessen 16 Bins je einen Abschnitt von $5.2\,cm$ im Intervall $\pm 40\,cm$ um den Wechselwirkungspunkt überstreichen. Um einen größeren Winkelbereich abzudecken ($9^0 - 155^0$), wird die Proportionalkammer des ersten Moduls der Vorwärtsspurkammern in die Histogrammbildung miteinbezogen. Das Histogramm steht auch auf L2 zur Verfügung. Als Netzeingabegrößen werden der maximale Bin-Eintrag, die Summe, sowie die Lage des Maximums selbst herangezogen. Diese Größen werden bereits in L1 bestimmt, und müssen nicht mehr auf L2 errechnet werden. Eine aus dem Histogramm extrahierte Signifikanz, definiert als

(Maximum+nächste Nachbarn)/Summe, wurde ebenfalls getestet. Die geometrische Spurinformation ist in einem Bitfeld kodiert, das in 14 Θ– und 16 Φ–Sektoren unterteilt ist. Damit kann überprüft werden, ob eine Spur in eine Region zeigt, in der eine Energiedeposition im Kalorimeter stattgefunden hat (*Bigray*). In dieses Bitmuster sind alle drei MWPC's *(Multi Wire Proportional Chambers)* der Vorwärtsspurkammern miteinbezogen, so daß der Bereich von $5^0 - 155^0$ abgedeckt ist.

Beschreibung	L2-Format	$\sum Bit$	Netzeingabegröße	Abk.
CIP-COP-z-Vertex-Histogramm	$16Bin \times 8Bit$	128	max. Bin-Eintrag [a]	cpvmax [b]
			Pos. des Maximum	cpvpos
			Summe über alle Bins	cpvsum
			Signifikanz (siehe Text)	cpvsig
Bigray–Topologie	$14\Theta \times 16\Phi$-Bins	(256)	(Koinzidenzen)	

[a] Maximum, Summe und Pos. des Maximum sind im L2-Datenstrom enthalten.
[b] Die Abkürzungen werden im folgenden in Diagrammbeschriftungen verwendet.

Tabelle 5.2: **L2-Daten der inneren Vieldrahtproportionalkammern.**

z-Kammern

Der zweite und der vierte Zylinder der inneren Spurkammern beherbergen schmale Driftkammern, deren Anodendrähte konzentrisch zur Strahlachse aufgespannt sind. Sie werden innere(CIZ, *Central Inner Z-Chamber*) und äußere(COZ, *Central Outer Z-Chamber*) z-Kammern genannt. Die innere besteht aus 15, die äußere aus 24 Auslesezellen mit je vier Anodendrähten. Die Auflösung in z ist für die Driftkammern besser als für die Proportionalkammern. Das entsprechende z-Vertex-Histogramm hat 96 Bins mit einer Auflösung von $5\,mm$ je Bin. Zusätzlich werden sogenannte Vertex-Zell-Histogramme und Untergrund-Zell-Histogramme angelegt, deren Bins den Auslesezellen der z-Kammern zugeordnet sind. Für jede CIZ-COZ-Koinzidenz, die in die Wechselwirkungszone zeigt, findet ein Eintrag in die entsprechenden zwei Bins für die jeweiligen Auslesezellen in den Vertex-Zell-Histogrammen statt. Entsprechendes gilt für die Untergrund-Zell-Histogramme, falls der Vertex in den rückwärtigen Bereich der Wechselwirkungszone zeigt. Das Triggersystem für die z-Kammern befindet sich in der Entwicklung und wird Ende 1994/Anfang 1995 zur Verfügung stehen. Es wurde in MC-Simulationen bereits in das Training von neuronalen L2-Triggern miteinbezogen [48, 49].

Die zentralen Jetkammern

Den größten Raum der inneren Spurkammern nehmen zwei Driftkammern ein, die zentralen Jetkammern CJC1 und CJC2. Die einzelnen Drähte verlaufen parallel zur Strahlachse, während die Drahtebenen spiralförmig angeordnet sind. Mit diesem

Beschreibung	L2-Format	Σ_{Bit}	Netzeingabegröße	Abk.
Spursignal in 45 Sektoren mit Ladungsvorzeichen und Impuls-klassifizierung	$45 \times 4\,\mathrm{Bit}$	180	pos. Ladung&niedr. Impuls [a]	trlopos
			pos. Ladung&hoher Impuls	trhipos
			neg. Ladung&niedr. Impuls	trloneg
			neg. Ladung&hoher Impuls	trloneg
			totale Sektormultiplizität	trtot

[a]Sektormultiplizitäten

Tabelle 5.3: **Verwendete L2-Daten der inneren Jetkammern(CJC).**

Driftkammernsystem ist sowohl die Bestimmung der r– und Φ–Koordinate einer Ionisationsquelle, als auch über Ladungsteilung eine etwas ungenauere Bestimmung der z–Koordinate möglich. Damit können Spuren vollständig räumlich rekonstruiert werden. Auf Triggerebene ist nur eine rudimentäre Spuranalyse möglich. Dazu werden 10 der 56 Signaldrähte einer Drahtebene verwendet. Nach einer Digitalisierung werden ungefähr 10000 Spurmasken mit den Signalen verglichen. Daraus wird extrahiert, ob in einem von 45 Φ–Sektoren eine Spur aufgetreten ist. Die Bestimmung des Ladungsvorzeichens des Teilchens und eine grobe Unterteilung in niedrigen und hohen Impuls wird ebenfalls durchgeführt. Als Netzeingangsgrößen dienen die Sektormultiplizitäten mit Impuls- und Ladungsunterscheidung.

5.2.3 Myonkammern

Das Myonsystem ist vierfach unterteilt in Front- und Endkappe, sowie vorderern und hinterern Rumpfteil (siehe Abb. 2.2). Jedes dieser Teile ist in 16 Module untergliedert. Die Module bestehen aus 16 Lagen von Streamerkammern, die in das Eisenrückflußjoch der supraleitenden Spule integriert sind. Myonen durchdringen wegen ihrer geringeren Ionisationswahrscheinlichkeit die kalorimetrischen Komponenten des Detektors im Gegensatz zu Elektronen und Hadronen und hinterlassen deshalb in diesen Kammern ein spezifisches Signal. Das gilt auch für kosmische Myonen. Die Drahtsignale der Lagen 3, 4, 5, 8 und 12 jedes Moduls werden ausgelesen, um über eine Triggerlogik eine Triggerbedingung (z. B. 3 aus 5) zu bilden. Diese Triggerinformation ist auf

Beschreibung	L2-Format	$\sum$ Bit	Netzeingabegröße	Abk.
Anzahl getroffener Module der Myonkammern im Eisen	$4 \times 16\,\mathrm{Module}$	64	Frontkappe(Modulanzahl)	ironfe
			vorderer Rumpfteil	ironfb
			hinterer Rumpfteil	ironbb
			Endkappe	ironbe
			tot. Modulanzahl	irontot

Tabelle 5.4: **Verwendete L2-Daten der Myonkammern.**

L2 zugänglich und wird verwendet. Eine ähnliche Triggergröße, die alle Modullagen miteinbezieht, aber nur auf L3 zur Verfügung stehen wird, wurde ebenfalls als Netzeingangsgröße verwendet (Abk. ironwire).
Energiereiche Myonen in Vorwärtsrichtung werden mittels Driftkammern vor und hinter einem Toroidal-Magneten detektiert. Es werden Triggersignale in einer r-Φ–Segmentierung erzeugt. Diese Information wurde bisher für das Netzwerktraining nicht genutzt.

5.2.4 Der rückwärtige Bereich des Detektors

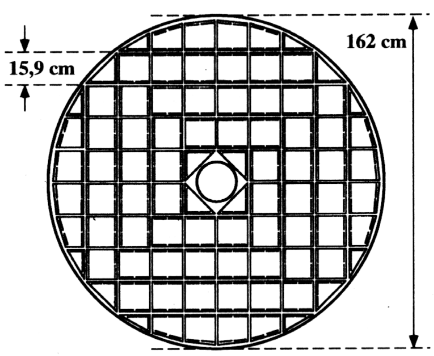

Abbildung 5.3: **Das BEMC in einer r-Φ–Projektion.**

Der rückwärtige Bereich wird von einem elektromagnetischen Kalorimeter(BEMC; *Backward ElectroMagnetic Calorimeter*), einer MWPC(BPC; *Backward Proportional Chamber*) und zwei Szintillatorwänden(ToF; *Time of Flight–System*) eingenommen.

Das rückwärtige elektromagnetische Kalorimeter(BEMC)

Für DIS-Ereignisse mit $Q^2 < 100\,GeV^2$ trifft das Elektron auf das BEMC, einem Blei-Szintillations-Sandwich-Kalorimeter. Die Zellgröße ist an der leptonischen Schauerausdehnung orientiert. Die Auslese der Energie erfolgt mittels Photomultipliern über Wellenlängenschieber an den Seitenflächen der Kalorimeterzellen. Aufgabe des BEMC-Triggers ist die Erkennung von Energie-Clustern (*engl.* für Agglomeration). Eine Kalorimeterzelle, in der eine bestimmte Energiedeposition überschritten wird, dient als Clusterzentrum. Die Clusterenergie ergibt sich aus der Summierung über die Nachbarzellen. Auf L2 stehen die Clusterenergie, die Gesamtenergie, sowie für jede Zelle zwei Schwellenbits zur Verfügung, die anzeigen, ob eine niedrige oder hohe Energieschwelle überwunden wurde. Das BEMC wird auch von Hadronen aus dem Strahluntergrund getroffen, diese deponieren typischerweise 30% ihrer Energie im BEMC.

Die rückwärtige Proportionalkammer (BPC)

Eine Rückwärtsproportionalkammer verbessert die r-Φ–Auflösung des BEMC. Sie ist in 4 Ebenen á 20 Streifen angeordnet. Im Gegensatz zu den anderen Proportionalkammern sind die Kathodenebenen nicht segmentiert und die Anodensignale werden ausgelesen. Auf L2 steht die Information zur Verfügung, welche der Streifen getroffen wurden.

Flugzeitmeßsystem (ToF)

Das ToF ist ein Hodoskop, bestehend aus zwei $3\,cm$ dicken Szintillatorwänden im Abstand von $30\,cm$, die mit Blei gegen die Synchrotronstrahlung abgeschirmt sind. Das Zeitintervall zwischen Ereignissen aus der Wechselwirkungszone und protoninduzierten Untergrundereignissen aus dem Bereich hinter den ToF-Wänden beträgt $\sim 13ns$. Deshalb werden relativ zum HeraClock-Signal zwei Zeitfenster definiert, ein Wechselwirkungsfenster und ein Untergrundfenster. Das von Treffern im Untergrundfenster abgeleitete Triggerelement „TOF-BG“ dient als effizienter und einfacher Veto-Trigger zur Unterdrückung von protoninduziertem Untergrund auf L1. Auch auf L2 kann die Information aus dem ToF-System in Korrelation mit anderen Größen weiter ausgenutzt werden. Alle für das Training verwendeten Variablen aus dem rückwärtigen Detektorbereich sind in Tab. 5.5 zusammengefaßt.

Beschreibung	L2-Format	Σ_{Bit}	Netzeingabegröße	Abk.
BEMC				
Zellen mit $E_{Zelle} > E_{niedrig}$	88 Zellen × 1Bit	88	Anzahl der Zellen	bemclow
Zellen mit $E_{Zelle} > E_{hoch}$	88 Zellen × 1Bit	88	Anzahl der Zellen	bemchig
Cluster-Energie	8Bit-FADC[a]	8	idem [b]	bemcclu
tot. Energie im BEMC		8		bemce
BPC				
getroffene Streifen in BPC-Ebene a-d	4 × 20 Streifen × 1Bit	20	Streifen in Ebene a	bpchita
		20	Streifen in Ebene b	bpchitb
		20	Streifen in Ebene c	bpchitc
		20	Streifen in Ebene d	bpchitd
			tot. Anzahl Streifen	bpctot
ToF				
Wand 0, Untergrundzeitfenster	8 Zellen × 1Bit	8	getroffene Zellen im Zeitfenster	bgwin0
Wand 0, Physikzeitfenster		8	getroffene Zellen im Zeitfenster	intwin0
Wand 1, Untergrundzeitfenster	16 Zellen × 1Bit	16	getroffene Zellen im Zeitfenster	bgwin1
Wand 1, Physikzeitfenster		16	getroffene Zellen im Zeitfenster	intwin1

[a] *Flash Analog to Digital Converter*
[b] Die Größe wurde unverändert als Netzeingabevariable benutzt.

Tabelle 5.5: **Verwendete L2-Daten aus dem rückwärtigen Detektorbereich.**

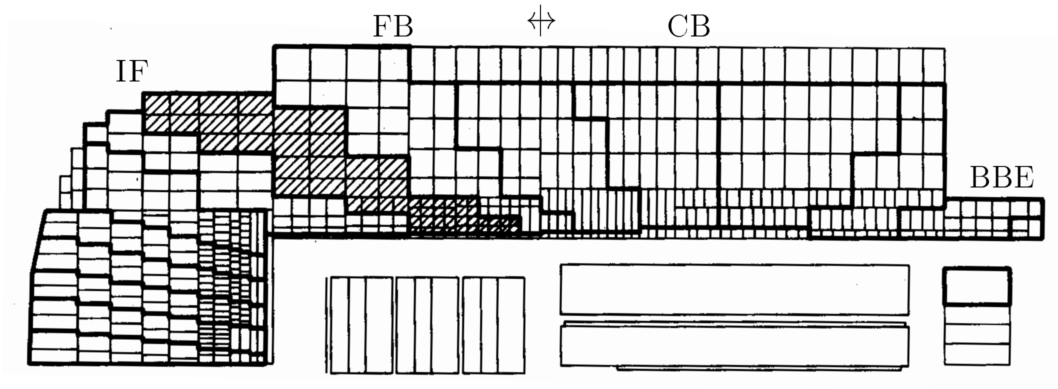

Abbildung 5.4: **Ein Querschnitt durch das halbe Flüssigargonkalorimeter.** *Der gestrichene Bereich markiert einen der 252 Bigtower.*

5.2.5 Das Flüssigargonkalorimeter

Dieses Kalorimeter ist ein Schichtkalorimeter, in dem Absorberplatten aus Blei oder Stahl abwechseln mit aktivem Ionisationsmaterial, in diesem Fall flüssigem Argon (LAr; *Liquid Argon*). Es ist nicht kompensierend. Hadronische Schauer müssen durch ihre Struktur (größere Schauertiefe und -winkel) von den elektromagnetischen Schauern unterschieden werden, bevor die deponierte Energie errechnet werden kann. Entsprechend dieser Topologie ist das Kalorimeter unterteilt in einen inneren elektromagnetischen und einen äußeren hadronischen Teil. Das Kalorimeter besitzt 45000 Auslesezellen. Eine analoge Elektronik summiert über mehrere Zwischenstufen die Signale der Auslesezellen zu 240 Triggersignalen auf, die *Bigtower* genannt werden. Die Struktur der Bigtower orientiert sich an vom Wechselwirkungspunkt wegzeigenden Strahlen (siehe Abb. 5.4). Die Bigtower sind 16-fach im Θ– und 8- bis 32-fach im Φ–Winkel unterteilt. Das warme Silizium-Kupfer-Sandwich-Kalorimeter(PLUG) in Vorwärtsrichtung ist in Form von 4 Bigtowern in die Ausleseelektronik des LAr miteinbezogen, ebenso das BEMC mit 8 Bigtowern. Die digitalisierten Bigtowersignale werden auf L1 zu topologischen Triggergrößen weiterverarbeitet. Es werden Energiesummen für den Vorwärtsbereich(IF; *Inner Forward*), den vorderen Rumpfbereich(FB; *Forward Barrel*) und den zentralen Rumpfbereich(CB; *Central Barrel*) gebildet. Desweiteren werden über Wichtungstabellen die transversale Energie($\sin\Theta$), sowie die x($\sin\Theta\sin\Phi$)- und y($\sin\Theta\cos\Phi$)-Koordinate und weitere Größen errechnet. Die x- und y-Koordinate sind essentiell zur Bestimmung der fehlenden Transversalenergie

$$E_{tmiss} = \sqrt{E_x^2 + E_y^2}\,. \tag{5.1}$$

Die Energiesummen dienen als Netzeingabegrößen auf L2. Da die Bigtower-Energien aber auch selbst zur Verfügung stehen, können neue Größen gebildet werden, z. B. *Jets* aus der Summierung über die Nachbarzellen der „heißesten“ Bigtower. Ebenso können

Beschreibung	L2-Format	$\sum$ Bit	Netzeingabegröße	Abk.
Plugenergie(Cu-Si!)	8Bit–FADC	8	idem	pluge
IF(Fl. Argon)-Energie		8		larife
FB- Energie		8		larfbe
CB-Energie		8		larcbe
Erg. aus Vowärtsbereich		8	$\sum E_{PL} + E_{IF}$	calfwde
Erg. aus Rumpfbereich		8	$\sum E_{FB} + E_{CB}$	calbare
Erg. aus Rückw. bereich		8	E_{BEMC}	calbwde
x-Komponente Energie	8Bit–FADC mit Vorzeichen!	8	$\|E_x\|$	larex
y-Komponente Energie		8	$\|E_y\|$	larey
transversale Energie	8Bit–FADC	8	idem	laret
gewichtete Totalenergie		8		laretot
Anzahl e^--Kandidaten		8		larele1
Quadranten in PLUG, IF,FB,CB,BEMC	8Bit–FADC	8*20	idem	eq(0-19)
Bigtower(BT) unterteilt in elektromagnetischen und hadronischen Anteil	zero-suppressed 8Bit-Adresse & 8Bit-FADC	var.	Anz. $\mathrm{BT}_{E>E_{min}}$	bigtnum
			Anz. $\mathrm{BT}_{zeitvalidiert}$	calbt
			$\sum_{BT}$Energie	bigterg
			Jetenergie(3)	jet(1-3)
			$\Theta_{Jet}(3)$	jtheta()
			$\Phi_{Jet}(3)$	jphi()

Tabelle 5.6: **Verwendete L2-Größen des Flüssigargonkalorimeters.**

über *Lookup-Tables* die Winkel des Bigtowers, der das Jetzentrum bildet, als Netzeingabegröße benutzt werden. Bigtower werden nicht über das PQZP-System, sondern über ein eigenes Quickbus-System übertragen, das die Daten in einer adresskodierten Form auf den L2-Bus gibt. Tab. 5.6 faßt die für das Training verwendeten Größen zusammen.

5.2.6 Weitere Netzeingabegrößen

Da Datenformate und -inhalt für die L2-Daten noch nicht endgültig festgelegt sind, wurden auch Größen als Netzeingabevariable getestet, die in der derzeitigen Planung nicht auf dem L2-Bus zur Verfügung stehen. Herauszugreifen sind hier das Luminositätssystem mit den Energien im Elektrontagger(etag) und Photontagger(photag).

5.3 Die Trainingsdaten

5.3.1 Die Extraktion der Trainingsdaten

Die von L1 erzeugten Triggergrößen gehen als primäre Selektionskriterien mit in die physikalische Datenanalyse ein. Um Triggereffizienzen berechnen zu können, müssen deshalb alle relevanten Triggerdaten eines Ereignisses auf Band aufgezeichnet werden. Das ermöglicht die Simulation eines neuronalen L2-Triggers mit realen experimentellen Daten. Zu diesem Zwecke wurde ein Datenextraktionsprogramm erstellt, das die Eingangsvariablen für das neuronale Netz aus den einzelnen Ereignis-Datenbanken bestimmt, und in eine eigene Datei überträgt. Aus dieser Datei werden dann die Trainings- und Testdatensätze erstellt. Soweit irgend möglich wurden nur Daten des L1-Triggersystems als Netzeingabegrößen verwendet. Für das Luminositätsmeßsystem war es erforderlich, auf rekonstruierte Daten zurückzugreifen. Zu Vergleichszwecken wurden auch rekonstruierte Energien als Netzeingabegrößen benutzt.

5.3.2 Vergleichbarkeit der Trainingsdaten

Ein kritischer Punkt beim Training von neuronalen Netzen ist die Vergleichbarkeit der Eingabedaten für verschiedene Klassen. Es ist wichtig, daß **alle** Eingabegrößen sowohl für die Untergrundklasse als auch für die Physikklasse in ihrer Bedeutung übereinstimmen müssen. Auch Korrelationen mit dem Klassifkationsmerkmal, die nicht der gegebenen experimentellen Situation entsprechen, werden von dem Lernalgorithmus ausgenutzt, und können falsche Effizienzen vortäuschen.

Bei einer größeren Anzahl sehr verschiedener Eingabedaten kann nicht völlig ausgeschlossen werden, daß simulierte Eingabedaten fehlerhaft sind, z. B. durch falsche und veraltete Datenbankbeschreibungen. Solange aber Untergrund- und Physikdaten mit demselben Programm extrahiert werden, bleiben die Daten vergleichbar, da der Fehler in beiden Klassen gleichermaßen auftritt, und keine statistische Signifikanz hat.

Zeitabhängige Effekte können ebenfalls einen erheblichen Einfluß auf die Vergleichbarkeit von Daten haben. Die Reproduzierbarkeit der Resultate ist maßgebend für die Qualität eines Experiments. In diesem Sinne ist auch auf die Zeitstabilität der L1-Triggerelemente geachtet worden. Die auf L2 verfügbaren Netzeingabegrößen sind jedoch keine physikalischen Größen, die in GeV oder sonst einer Einheit ausgedrückt werden können, sondern es handelt sich um (8Bit-) Maschinenzahlen, die über eine elektronische Verarbeitungskette mit den wahren physikalischen Größen korreliert sind.

Ein typischer zeitabhängiger Effekt ist z. B. die Nullpunktsverschiebung eines FADC-Wertes. Die analoge Eingangsgröße '0' kann beispielsweise in verschiedenen Zeiträumen einmal der digitalen Ausgangsgröße `00` und ein anderesmal `01` entsprechen. Strahlbedingte Einflüße wie Ströme und Vakuumqualität sind ebenfalls zeitlichen Änderungen unterworfen. Diese Randbedingungen haben aber keinen Einfluß auf die Ereignistopologie sondern nur auf die statistische Verteilung verschiedener Arten von Untergrundereignissen. Das gleiche Argument trifft zu für unterschiedliche Einstellungen der Schwellentrigger von L1, insofern sie nicht mit einer Änderung der L2-Datenstruktur einhergehen.

Ein typisches Beispiel ist in Abb. 5.5 zu sehen. Es zeigt die Verteilung des FADC-Wertes der totalen BEMC-Energie für im August und Oktober 1994 extrahierte Untergrunddaten. Die Unterschiede der Verteilungen nahe des Nullpunkts sind deutlich zu erkennen. Trainiert man diese Daten gegeneinander, so werden beachtliche Trenneffizienzen erreicht. Das Ergebnis ist jedoch wertlos, weil die Daten aufgrund des FADC-Offsets(u. ä.) nicht vergleichbar sind. Wird aber dafür Sorge getragen, daß Physik- und Untergrundereignisse aus demselben Zeitraum stammen, dann verlieren

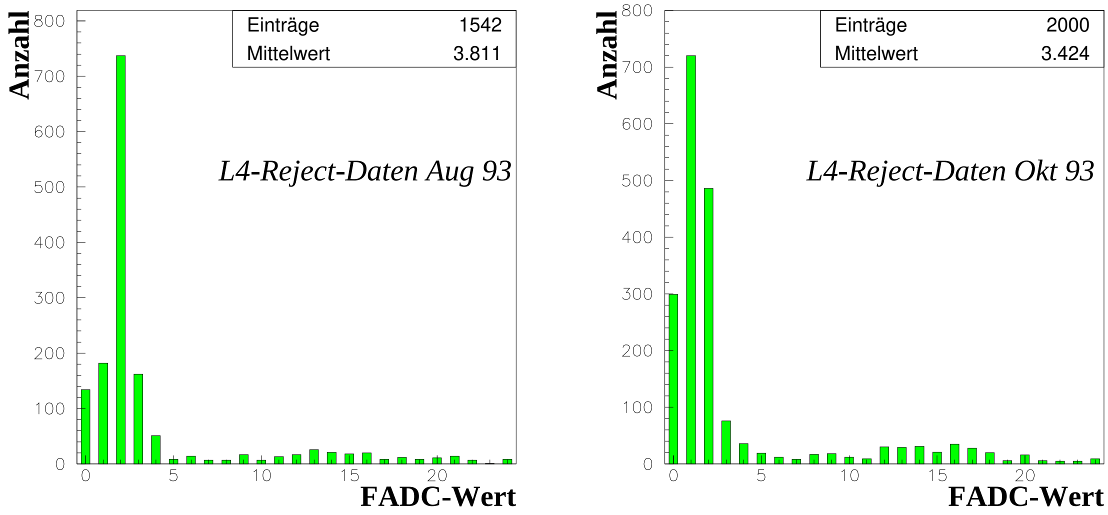

Abbildung 5.5: **Verteilung des FADC-Werts der totalen BEMC-Energie im Vergleich für Untergrunddaten aus August 1993(links) und Oktober 1993(rechts).**

zeitabhängige Effekte ihre statistische Signifikanz, da sie nicht mehr mit der Klassifizierung korreliert sind. Da grundsätzlich für jeden Zeitraum Untergrundereignisse (ca. 1%) zu Monitorzwecken aufgezeichnet werden, stehen für den Vergleich mit einer Physikklasse immer Untergrundereignisse zur Verfügung.
Aus den genannten Gründen wurden nur Physik- und Untergrunddaten gegeneinander trainiert, die aus vergleichbaren Zeiträumen stammen und mit demselben Datenextraktionsprogramm erzeugt wurden. Wo sich das nicht streng einhalten läßt (z. B. bei Runs, in denen ein spezifischer Untergrund getriggert wird, der später als Trainingsklasse benutzt wird), müssen Vergleichbarkeitsbetrachtungen angestellt werden.
Anzumerken ist, daß die obige Argumentation nicht für den Einsatz eines austrainierten Netzes in der Ausführungsphase zutrifft. Hat ein Netz durch ein Training mit vergleichbaren Daten einmal gelernt, sich bei der Selektion auf die physikalisch relevanten Größen zu stützen, dann werden Nullpunktsschwankungen eines FADC-Wertes nur geringen Einfluß auf die Effizienz des Netzes haben. Der quantitative Fehler ist zudem durch eine Simulation ermittelbar.

5.3.3 Statistische Analysemethoden vor dem Training

Statistische Analysen helfen in mehrfacher Hinsicht, die Trainingsergebnisse zu verbessern und zu bewerten. Eine routinemäßige Kontrolle eindimensionaler Verteilungen wie in Abb. 5.5 hilft, die korrekte Extraktion sicherzustellen. Eine wichtige statistische Größe in der linearen Regressionsanalyse ist der Korrelationskoeffizent zwischen den Zufallsgrößen x_i und x_j

$$\rho_{ij} = \sum_{n=1}^{N} \frac{(x_i^{(n)} - \bar{x_i})(x_j^{(n)} - \bar{x_j})}{N^2(\sigma_i \cdot \sigma_j)} \,. \tag{5.2}$$

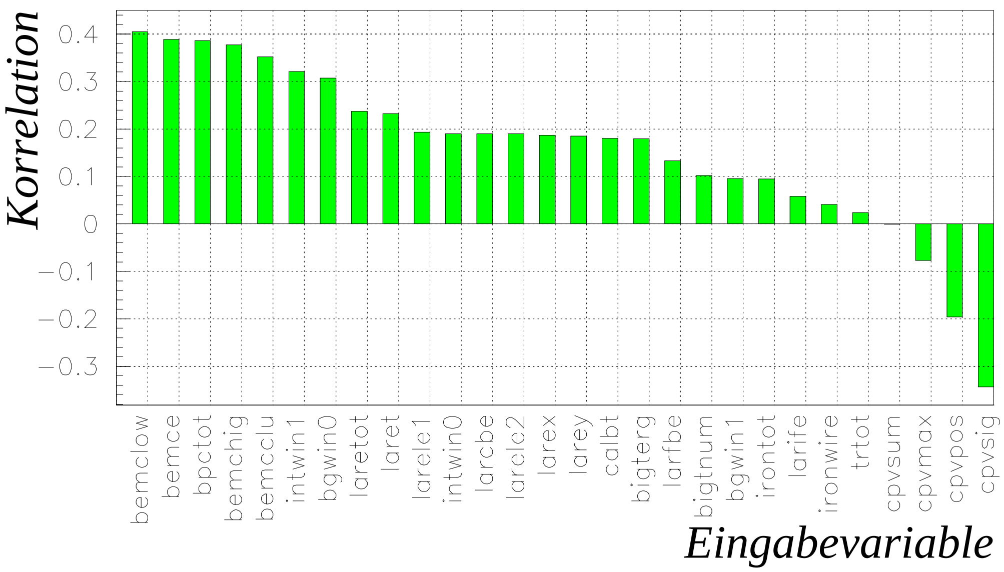

Abbildung 5.6: **Korrelation einzelner Eingabevariablen mit dem Klassifikationsindex.** *'1' ist der Index für ein nicht-Strahl-Gas-Ereignis, und '0' für ein Strahl-Gas-Ereignis.*

σ_i bezeichnet die Standardabweichung der i-ten Eingangsvariable und N die Anzahl der verfügbaren Ereignisse. Der Korrelationskoeffizient liegt immer zwischen -1 und $+1$. Ein Absolutwert nahe 1 entspricht stark korrelierten Zufallsvariablen, nahe 0 unkorrelierten. Die Korrelation zwischen den Eingabegrößen und dem Klassifikationsmerkmal ist in Abb. 5.6 aufgetragen für das Problem, Strahl-Gas-klassifizierte Ereignisse von nicht-Strahl-Gas-klassifizierten Ereignissen zu trennen. An einem solchen Korrelationsdiagramm läßt sich erkennen, ob eine Größe unphysikalisch hoch mit der Klassifizierung korreliert ist. Ist letzteres der Fall, muß die Vergleichbarkeit im Sinne des vorigen Abschnitts angezweifelt werden. Korrelationsberechnungen sind somit ein einfacher Test für die Überprüfung der Vergleichbarkeit. Ist diese hinreichend sichergestellt und eine Größe mit dem Klassifikationsmerkmal dennoch hochkorreliert, dann ist das Problem trivial durch einen Schnitt auf diese eindimensionale Größe lösbar. Das neuronale Netz wird zwar ebenfalls eine hohe Effizienz erreichen, aber die Anwendung von Lernalgorithmen ist in diesem Fall überflüssig. Abb. 5.6 zeigt, daß für das Strahl-Gas-Klassifikationsproblem keine Größe hochkorreliert mit der Klasseneinteilung ist, aber viele Größen eine leichte bis mittlere Korrelation aufweisen. Das stimmt mit der Erwartung überein.

Kapitel 6

Das Training mit H1-Ereignisdaten

6.1 Klassifizierung der Ereignisse

6.1.1 Physikklassen

Experimentelle Daten lassen sich nur a posteriori klassifizieren. Diese Klassifizierung ist somit immer mit Unsicherheiten behaftet. Die Minimierung des Klassifizierungsfehlers ist wesentlicher Bestandteil der experimentellen Ermittlung von Wirkungsquerschnitten. Die Physikgruppen bei H1 beschäftigen sich intensiv damit, kinematische Schnitte zu finden, die sehr selektiv sind, um bestimmte Physikreaktionen aus der Gesamtmenge der experimentellen Daten zu filtern. Dieses Wissen kann indirekt für den neuronalen L2-Trigger ausgenutzt werden, indem die von den Physikgruppen selektierten Ereignisse als Eingabemuster für ein Backpropagation-Training verwendet werden. Für Ereignisse mit geringer Rate und besonderer Bedeutung (NC,CC) ist es jedoch erforderlich, nicht die endgültigen Analysekriterien für die Auswahl der Trainingsmenge zu benutzen, sondern weichere Selektionskriterien, die nur einer Vorauswahl dienen. So wird erreicht, daß die L2-Triggereffizienz möglichst wenig variiert in Abhängigkeit von verschiedenen kinematischen Schnitten, die in der Physikanalyse Anwendung finden. Eine solche Vorselektion wird bei H1 automatisch durchgeführt. Verschiedene Klassifizierungsalgorithmen werden auf die von L4 akzeptierten Ereignisse angewendet. Bei diesem auch als „L5“ bezeichneten Schritt werden aus den Rohdaten die rekonstruierten Ereignisse auf den sogenannten *Production Output Tapes* (POT) gespeichert und die Kandidaten für die Physikanalyse herausgefiltert und auf *Data Summary Tapes* (DST) geschrieben. In Tabelle 6.1 sind die im weiteren wichtigen Selektionsklassen aufgeführt. Insgesamt gibt es 32 Selektionsklassen.

6.1.2 Untergrundereignisse

Um Untergrundereignisse aus den experimentellen Daten zu extrahieren, gibt es mehrere Möglichkeiten. Eine davon ist, die von L4 zurückgewiesenen und zur Überwachung dennoch aufgeschriebenen Ereignisse zu verwenden. In Kombination mit den L5-Untergrundklassen (s. Tab. 6.2) erhält man so eine sehr reine und große Menge an Untergrunddaten. Vorteilhaft ist, daß die Vergleichbarkeit dadurch sichergestellt

Klasse	Erklärung & Selektionskriterien
NCHQSQ	NC-Ereignis mit hohem Q^2 ($> 100\,GeV^2$) • Elektronkandidat im Barrel- oder Vorwärtsbereich • $p_{tmiss} < 40\,GeV$, kein fehlender Transversalimpuls • eine gute Spur (Vorwärts- oder Zentralkammern) • Cosmics- und Myon-Veto
NCLQSQ	NC-Ereignis mit niedrigem Q^2 ($< 100\,GeV^2$) • kein Untergrundflag • eindeutige ToF-Signatur • Cluster nicht in inneren dreieckigen Zellen (s. Abb. 5.3) • 1 BPC-Treffer oder d(Cluster–Strahlachse) $> 68cm$ • ein Vertex oder $22\,GeV \leq E_{Clu} \leq 32\,GeV$ • $4\,GeV \leq E_{Clu} \leq 100\,GeV$ (BEMC)
EWKCC	CC-Ereignisse (Vorselektion) • $E_{tmiss} > 10\,GeV$ • E_{tmiss}(ohne e^-) $> 10\,GeV$, zum Vergleich • eine gute Spur
JETJET	Kandidaten für Jet-Jet-Ereignisse • $E_t > 10\,GeV$ • $E_{etagger} > 4\,GeV$ oder $E_t(25^0 \leq \Theta \leq 180^0) > 20\,GeV$ • Cosmics- und Myon-Veto
PGAMMA	Kandidaten für prompte Photon-Ereignisse im LAr • E_t(elmag. Cluster) $> 3\,GeV$ • $E_t(\Theta > 10^0) > 5\,GeV$ • eine gute Spur
CCBAR	$c\bar{c}$-Kandidaten ($D^*, \Lambda, etc.$) • kein Untergrundflag • 2 gute Spuren • $p_{t,Spur(1)} > 0.6\,GeV$ und $p_{t,Spur(2)} > 0.45\,GeV$ • weitere Spurkriterien (Ladung, Winkel)

Tabelle 6.1: **L5-Physikklassen.** *Aufgeführt sind 6 der insgesamt 27 Vorselektionsklassen der Runperiode 1993, die für das Training neuronaler Netze herangezogen wurden.*

wird, daß die Untergrunddaten gleichzeitig mit den Physikdaten aufgezeichnet werden. Eine andere Möglichkeit ist, Daten von speziellen Runs zu verwenden, bei denen bestimmte L1-Triggerschwellen auf ein Minimum gesetzt wurden. Durch Herabsetzen der Energieschwellen des Flüssigargonkalorimeters kann beispielsweise eine fast reine Menge an durch Rauschen getriggerten Ereignissen erhalten werden. Hier ist die Vergleichbarkeit zu überprüfen.
Eine dritte Möglichkeit besteht darin, sogenannte Pilotbündel-Ereignisse aus den Daten zu extrahieren. Pilotbündel sind Strahlbündel, denen das Partnerbündel am Wechselwirkungspunkt fehlt. Ein Ereignis, das getriggert wurde, während ein Proton-Pilotbündel die Wechselwirkungszone durchquert, kann also nur ein Untergrundereignis sein. In jedem Durchlauf werden eine Reihe solcher Pilotbündel mit Teilchen gefüllt. Es stehen weniger Ereignisse als für die erste Möglichkeit zur Verfügung. Alle drei Möglichkeiten wurden angewendet, eine Kombination der ersten und dritten Möglichkeit in ihrer natürlichen statistischen Verteilung wurde als Standarduntergrund bevorzugt.

Klasse	**Erklärung & Selektionskriterien**
COSMICS	Kandidaten für kosmische Strahlung • komplexer topologischer Algorithmus
BGAS	Kandidaten für Strahl-Gas-Ereignisse • komplexer topologischer Algorithmus; benutzt den Energiefluß im LAr in Kombination mit Spurinformationen und zusätzlich eine Reihe von Nebenbedingungen wie *Anzahl rückwärtszeigender Spuren* ≤ 2
ELPIL	Ein Elektron-Pilotbündel-Ereignis
PRPIL	Ein Proton-Pilotbündel-Ereignis
NOT_CL	Das Ereignis konnte keiner Klasse zugeordnet werden

Tabelle 6.2: **L5-Untergrundklassen.** *Diese für die Runperiode 1993 gültigen Untergrundklassen waren Grundlage für die Extraktion von Untergrunddaten für das Netztraining.*

6.2 Durchführung des Trainings

Eine Reihe von Anwenderprogrammen steht zur Verfügung, um ein Backpropagation-Training durchzuführen. Das anfänglich benutzte Programmpaket JETNET2.0 [50] erwies sich in Geschwindigkeitstests dem Programmpaket Aspirin/MIGRAINES V6.0 [51] etwa um den Faktor 5 unterlegen. Bei typischen Trainingszeiten von einigen Stunden auf einer IBM RS/6000 Workstation bei Verwendung des Aspirin-Paketes wurde deshalb letzterem der Vorzug gegeben. Seit Ende 1993 steht der für den L2-Trigger vorgesehene CNAPS-Parallelrechner u.a. für Trainingszwecke zur Verfügung. Der Zeitvorteil ist proportional zur Anzahl der versteckten Knoten; ein typischer Trainingsdurchlauf (1 Mio. Vektor-Lernschritte) dauert einige Minuten und erlaubt so eine rasche Durchführung systematischer Studien.

6.2.1 Das Training mit Aspirin/MIGRAINES

Das von der *MITRE Corporation* entwickelte Softwarepaket Aspirin/MIGRAINES V6.0 ermöglicht die Definition einer Netzwerkarchitektur über eine eigene Macro-Sprache. Für diese Architektur wird ein optimierter C-Kode erstellt, mit dem sich für unterschiedliche Parameter und Daten ein Backpropagation-Training und nachfolgende Analysen durchführen lassen. Die Problemabhängigkeit der optimalen Wahl von Lern- und Momentumparameter kann durch eine Transformation der Eingangsdaten gemildert werden. Dabei werden die Daten so skaliert, daß der Mittelwert jeder Eingabegröße '0' entspricht, und die dreifache Standardabweichung $3\sigma_i$ auf '1' normiert wird. Da es sich um eine affine Transformation mit einer Diagonalmatrix Λ,

$$\vec{x}' = \Lambda\vec{x} + \vec{m} \ \text{ mit } \ \lambda_{ii} = 1/3\sigma_i, \tag{6.1}$$

handelt, kann aus dem Netz für die skalierten Eingabegrößen x'_i ein Netz für die unskalierten über

$$\vec{\omega}_j = \Lambda^T\vec{\omega}'_j \text{ und } t_j = \vec{m}^T\vec{\omega}'_j + t'_j \tag{6.2}$$

konstruiert werden. $\vec{\omega}_j$ ist der Gewichtsvektor zur Eingabeschicht für den versteckten Knoten j und t_j der Schwellwert. Die Möglichkeit einer Rücktransformation ist unabdingbar, da für den Einsatz des neuronalen L2-Triggers keine Zeit für die Durchführung derartiger Transformationen zur Verfügung steht. In der Regel beschleunigt die Nivellierung des dynamischen Bereichs der Eingangsvariablen auch die Trainingsdauer.

6.2.2 Das Training mit dem CNAPS–Parallelcomputer

Das Design des CNAPS®-Parallelcomputers ist speziell auf die Anwendung für neuronale Netze ausgerichtet. Es handelt sich um Chips in SIMD[1]-Architektur, bei denen ein *Sequencer* bis zu 256 Prozessorknoten (PN; *processor nodes*) im $20\,MHz$-Doppeltakt mit Steuerbefehlen speist. Ein Chip hat 64 PN's, bis zu 4 Chips können zusammengefaßt werden. Jeder PN hat einen lokalen Speicher von $4\,KByte$, in dem die Gewichtsvektoren von einem oder mehreren Netzknoten abgelegt werden. Eine genauere Beschreibung erfolgt im Zusammenhang mit der L2-Hardware (Abschnitt 8.3). Von Bedeutung für das Training sind die Limitierung der Rechengenauigkeit (siehe Tab. 6.3) und die Verwendung von Festkomma-Arithmetik. Rundungsfehler sind damit abhängig vom Absolutbetrag einer Zahl, kleine Zahlen bedeuten große Fehler. Die damit einhergehende Problematik wurde in einer Parameterstudie untersucht.

[1] single-instruction multiple-data

	BP(normal)	BP(erweitert)
Gewichtsänderung	16 bit	32 bit
Datenformat	8 bit integer	16 bit signed integer
Gewichte	16 bit sign. integer, skal. 12[a]	16 bit sign. integer, skal. 12
Transferfunktion	logist. Sigmoidfunktion	tanh()
Trainingsgeschw.	BP(erweitert) = 4∗BP(normal)	
Parameter	Binärbruch ; 8 bit integer mit versch. Skalierung	

[a]Der Binärbruch hat 12 Stellen hinter dem Komma.

Tabelle 6.3: **Daten- und Parameterformate für CNAPS.**

6.3 Parameterstudien für den CNAPS–Rechner

Die für das Training mit Fließkommazahlen benutzten Parameter lassen sich nicht für die Festkomma-Arithmetik des CNAPS-Rechners übernehmen. Für kleine Lernparameter wirkt sich Festkomma-Arithmetik sehr ungünstig aus, da Gewichtsänderungen unterhalb der kleinsten darstellbaren Zahl nicht durchführbar sind. In flachen Regionen der Fehlerfunktion kann der Lernprozeß steckenbleiben oder sich dramatisch verlangsamen. Um den Preis der Einführung eines neuen freien Parameters kann man versuchen, diesen Defekt dadurch zu beheben, daß zur Ableitung g' in der δ-Lernregel (4.10) ein kleiner konstanter Term d_{off} hinzuaddiert wird. Dieser Term wird im weiteren *Ableitungskonstante (derivation offset)* genannt.
Für die Parameterstudie wurde das in Abschnitt 5.3.3 bereits vorgestellte Klassifizierungsproblem von nicht-Strahl-Gas-Reaktionen gegen Strahl-Gas-Reaktionen verwendet. Diese auf L5 durchgeführte Klassifizierung ist sehr unspezifisch, und das Nachvollziehen dieser Separation auf L2 stellt ein schwieriges Problem dar. Von den zur Verfügung stehenden zwei Modi der Bitauflösung (s. Tab.6.3) wurde der Modus mit 8 Bit für die Eingangsdaten und 16 Bit für die Gewichte gewählt, da die Eingangsdaten ohnehin nur eine Auflösung von 8 Bit besitzen. Auf eine Skalierung wurde vorerst verzichtet.
Das Generalisierungsverhalten wurde untersucht, indem aus der ursprünglichen Menge von 12467 Ereignissen zufällig je 1000 Test- und Trainingsvektoren ausgewählt wurden. Als Eingabegrößen dienen die 28 Größen, die auch auf der Abszisse von Abb. 5.6 aufgetragen sind. Die Architektur des Netzes ($28 \times 20 \times 1$) wurde mit Absicht so gewählt, daß die Generalisierung kritisch ist. In Abb. 6.1 ist der gemittelte quadratische Fehler über 500 Lernepochen für verschiedene Lernparameter aufgetragen. Eine Epoche entspricht der einmaligen Präsentation aller 1000 Trainingsmuster im Online-Updating. Das Netz wurde gleichzeitig mit unbekannten Testdaten überprüft, ohne daß die Gewichte verändert wurden. Das Minimum der Lernkurve für die Testdaten markiert das Netz, das letztendlich für die spätere Anwendung ausgewählt wird.
Abb.6.1a zeigt die Lernkurve für ein Training ohne Momentumfaktor mit $\eta = 0.25$. Auch nach 500 Epochen scheint kein Generalisierungsminimum erreicht worden zu sein, nach 200 Epochen bleibt der Fehler für die Testmenge nahezu konstant. Durch Hinzufügen einer kleinen Ableitungskonstante d_{off} in Abb. 6.1b wird der Lernprozeß

a) b)

c) d)

Abbildung 6.1: **Lernkurven mit verschiedenen Parametern für das Strahl-Gas-Klassifizierungsproblem mit reduzierter Datenmenge.**

sichtbar beschleunigt, nach etwa 200 Epochen ist aber das Generalisierungsminimum erreicht und es kommt zu einem schwach ausgeprägten Overtraining. Das Training mit einem Momentumfaktor $\alpha = 0.668$ in c) führt zu einer dramatischen Beschleunigung des Lernens. Nach bereits 38 Epochen ist das Generalisierungsminimum erreicht. Die an diesem Punkt erreichte Generalisierung wurde von einem BP-Netz ohne Momentumfaktor auch nach 2000 Epochen nicht unterboten. Das danach auftretende Overlearning führt jedoch zu einer massiven Verschlechterung der Generalisierung. Die Wahl eines noch höheren Momentumfaktors $\alpha = 0.875$ vergrößert zwar die anfängliche Lerngeschwindigkeit, führt aber im weiteren Verlauf zu Instabilitäten im Lernverhalten. Auch für Fließkomma-Arithmetik wird dieses Verhalten beobachtet, weshalb hohe Momentumparameter gewöhnlich mit kleinen Lernparametern kombiniert werden. Der kleinstmögliche darstellbare Lernparameter für den 8/16-Bit-Modus ist mit 2^{-6} bereits zu hoch, um mit Momentumparametern > 0.9 ein stabiles Lernverhalten zu erhalten.

a) b)

Abbildung 6.2: **Systematische Variation von Lernparameter und Anzahl der versteckten Knoten.** *Die dynamische Skalierung der Ordinate ist zu beachten.*

In einer weiteren Studie mit der Gesamtmenge von 12467 Ereignissen wurden Lernparameter, Momentumparameter, Ableitungskonstante, Anzahl versteckter Knoten und Initialisierungsintervall der Gewichte systematisch variiert. Die Ereignisse wurden gemäß der Zweier-Parität ihrer Reihenfolge in eine Trainings- und Testmenge unterteilt. Das Training erfolgte über 1000 Epochen, das entspricht etwa $6 \cdot 10^6$ Lernschritten.

Lernparameter η wurde in dem Bereich $[0.05, 0.95]$ systematisch für ein $(28\times20\times1)$-Netz variiert. In Abb. 6.2a ist der mittlere quadratische Fehler der besten Epoche gegen den Lernparameter aufgetragen. Es ist zu beachten, daß das Fehlerminimum für die Test- und Trainingsmenge in der Regel nicht in dieselbe Epoche fallen, wie die Lernkurven im vorigen Beispiel auch deutlich zeigen. Für den Lernparameter 0.2 wurden 10 Durchläufe mit verschiedenen Startwerten des Zufallsgenerators durchgeführt. Mittelwert und Standardabweichung dieses Tests sind in die Abbildung eingetragen. Um die Schwankungen zu glätten, wurde die Variation des Lernparameters dreimal mit verschiedenen Startwerten des Zufallsgenerators durchgeführt. In der Abbildung ist das gemittelte Ergebnis über alle drei Durchläufe aufgetragen. Für das Lernen ohne Momentumparameter favorisiert die Studie für das vorliegende Problem eindeutig Lernparameter im Bereich von $0.2 - 0.3$. Kleinere Lernparameter benötigen unverhältnismäßig große Trainingszeiten, während das Ergebnis bei höheren Lernparametern durch zu große Schrittweiten verschlechtert wird.

Anzahl der versteckten Knoten Für $\eta = 0.20$ wurde die Zahl der versteckten Knoten im Bereich $[1, 20]$ untersucht. Es wurden wieder drei Durchläufe wie für den Lernparameter durchgeführt und das Ergebnis gemittelt (s. Abb. 6.2b). Die Differenz zwischen bestem Test– und Trainingsresultat nimmt erwartungsgemäß mit der Anzahl

a) b)

Abbildung 6.3: **Systematische Variation von Momentumfaktor und Ableitungskonstante.**

von Gewichtsparametern zu. Die Generalisierung erreicht jedoch kein klar identifizierbares Minimum. Zwischen 10 und 20 Knoten ist die Variation des Testergebnisses sehr gering. Es wurden auch Netze mit bis zu 200 versteckten Knoten trainiert, erst oberhalb von 100 versteckten Knoten begann sich die Generalisierung bei Zunahme der Variationsbreite signifikant zu verschlechtern.

Momentumparameter Der Variationsbereich lag für $\eta = 0.20$ zwischen 0.0 und 0.95. Der Testdurchlauf bestätigt das in Abb. 6.1c angedeutete Verhalten. Mit zunehmendem Momentumparameter tritt verstärkt Overlearning auf. Das ist an der Zunahme der Differenz zwischen dem Testergebnis der besten Epoche und dem der letzten Epoche abzulesen. Die Studie zeigt auch, daß Momentumparameter zwischen 0.5 und 0.7 für $\eta = 0.20$ zu den Netzen mit dem besten Generalisierungsverhalten führen. Größere Momentumparameter zeigen das bereits besprochene Instabilitätsverhalten.

Ableitungskonstante Abgesehen von der Beschleunigung des Lernens konnte ein systematischer Einfluß auf das Generalisierungsverhalten bei einer Variation der Ableitungskonstante im Intervall $[0.0, 0.18]$ nicht beobachtet werden (s. Abb. 6.3). Für die Variation der Intervallgrenzen der Initialisierungsgewichte ließ sich für das vorliegende Problem weder ein Einfluß auf die Lerngeschwindigkeit noch auf die Generalisierung nachweisen.

Im ganzen bestätigt die Studie die Stabilität des Algorithmus. Die Variation der Effizienz der Netze bewegt sich in einem sehr engen Rahmen. Im Vergleich mit den Ergebnissen, die unter Verwendung des ASPIRIN-Paketes erzielt wurden, zeigt sich, daß die Generalisierung signifikant schlechter ist. Das beste ASPIRIN-Netz wurde von keinem

Netz der Studie in seiner Effizienz übertroffen. Die Trenneffizienzen liegen aber nahe beieinander (Bei einem Schnitt bei 0.5 beträgt die Physikeffizienz für das CNAPS-Netz 81.1%(A. 83.0%) und die Akzeptanz für den Untergrund 10.8%(A. 8.8%)). Eine Reihe von Differenzen kann für die schlechtere Generalisierung verantwortlich sein. Neben den genannten Unterschieden (Skalierung, Festkomma-Arithmetik) ist auch die Reihenfolge der präsentierten Ereignisse von Epoche zu Epoche durch die CNAPS-Software starr festgelegt, während sie für ASPIRIN variiert wird. Generell kann gesagt werden, daß sich die Wahl von geeigneten Parametern unter Verwendung von Festkomma-Arithmetik schwieriger gestaltet als für Fließkomma-Arithmetik. Hier werden weitere Untersuchungen erforderlich sein.

6.4 Trainingsergebnisse mit vorselektierten Daten

6.4.1 Verwendete Netzwerkkonfigurationen

In diesem Abschnitt wird untersucht, inwiefern die auf L5 durchgeführte Vorselektion experimenteller H1-Daten als Klassifikationskriterium für ein Training mit L2-Daten geeignet ist. Um die einzelnen Physikklassen miteinander vergleichen zu können, wurden feste Eingabevektoren für das Training verwendet. Drei verschiedene Netzwerkkonfigurationen (s. Tab. 6.4) wurden für jedes Problem mit unterschiedlichen Parametern ($\alpha = 0.5 \ldots 0.99$; $\eta = 0.01 \ldots 0.001$) unter Verwendung des ASPIRIN-Paketes trainiert, und das Netz mit den besten Trenneffizienzen ausgewählt. Das Training wurde auf dem CNAPS-Rechner mit entsprechend angepaßten Werten im 8/16-Bit-Modus wiederholt. Dabei wurden ähnliche Resultate mit leichten Präferenzen für die mit ASPIRIN trainierten Netze erzielt. Die Mengen an Untergrund- und Physikereignissen, die gegeneinander trainiert wurden, enthielten immer ungefähr die gleiche Anzahl von Ereignissen, die mit jeweils gleicher Wahrscheinlichkeit dem Netz präsentiert wurden. Es wurde für die Bestimmung der Trenneffizienz ein Schnitt bei 0.5 durchgeführt, einer natürlichen Wahl bei einem Zielwert von '0' für Untergrund und '1' für Physik. Von der Möglichkeit, einen anderen Wert für die Triggerschwelle zu wählen, wurde gegebenenfalls zu Vergleichszwecken Gebrauch gemacht.

Der kleinste Eingabevektor mit 28 Größen benutzt vorwiegend topologische L1–Summen, deren Korrelation das Netz ausnutzen soll. Das zweite Netz mit 49 Größen benutzt dieselben Variablen in etwas feinerer Auflösung. In einem dritten Netz mit 66 Größen wurden alle verfügbaren L2-Größen zusammengefaßt, und einige voraussichtlich nicht verfügbare, rekonstruierte Größen hinzugenommen, unter anderem die Energie des Elektrontaggers.

Für die Bewertung eines Netzes ist allein der Test mit im Training nicht verwendeten Daten relevant. Deshalb wurden die zur Verfügung stehenden Daten in zwei gleichgroße Mengen an Test- und Trainingsdaten aufgeteilt. Das nur mit den Trainingsdaten erzielte Resultat ist zwar wichtig, um während des Trainings das Lernverhalten zu überprüfen, hat aber für die abschließende Bewertung keine eigenständige Bedeutung. Aus Gründen der Übersichtlichkeit werden deshalb nur die Testresultate aufgeführt. Die Namen der verwendeten Datenfiles, sowie dazugehörige statistische Angaben finden sich im Anhang.

Netz I	$28 \times 14 \times 1$
LAr(12)	laretot, laret, larex, larey, larife, larfbe, larcbe,
	larele1, larele2, bigterg, bigtnum
Back(9)	bemce, bemcclu, bemclow, bemchig, bpctot, intwin0, intwin1,
	bgwin0,bgwin1
μ(2)	irontot, ironwire
track(5)	cpvsum, cpvmax, cpvpos, cpvsig, trtot
Netz II	$49 \times 25 \times 1$
LAr(+12–3)	Quadranten für Inner Forward, Forward Barrel, Central Barrel
Back(+4–1)	bpctot $\rightarrow$ bpchita..d
μ(+8–2)	irontot, ironwire $\rightarrow$ vierfache Segmentierung
track(+4–1)	trtot $\rightarrow$ niedr.,hoh. Impuls; Ladungsvorzeichen
Netz III	$66 \times 35 \times 1$
lumi(+2)	eletag, photag
LAr(+2)	calfwde, calbare
z-Kammern(+2)	innere und äußere Pad-Koinzidenzen
Back(+4)	Erg. heißeste BEMC-Zelle, Clusteranzahl, rek. (Cluster-)Erg.
(+7)	Summen von Netz **I**

Tabelle 6.4: **Verwendete Eingangsgrößen der Netzkonfigurationen.** *Zur Beschreibung der Abkürzungen siehe Kapitel 5 . Die Zahlen in Klammern bilanzieren die Anzahl der verwendeten Größen.*

6.4.2 Strahl-Gas-(L5)-Klassifikation

In einem ersten Versuch wurden L4-Keep-Ereignisse gegen L4-Reject-Ereignisse trainiert. Der L4-Algorithmus verwendet detaillierte Spurinformation, die auf L2 nicht verfügbar ist. Es erstaunt deshalb nicht, daß es einem neuronalen Netz nur begrenzt möglich war, die L4-Entscheidung mit L2-Daten nachzuvollziehen (s. Tab. 6.5). Eine Ursache für dieses Ergebnis ist, daß ein erheblicher Anteil der Ereignisse, die L4 passieren (ca. 1/6 aller L1-Keep-Ereignisse), immer noch Untergrundereignisse sind ($\sim 30\,\%$ fallen in eine L5-Untergrundklasse). Deshalb ist der Überlapp zwischen den beiden Trainingsklassen sehr hoch.
Ein vernünftigeres Selektionskriterium für das Training ist die auf L5 durchgeführte Strahl-Gas-Klassifizierung. Der zugrunde liegende Algorithmus zielt darauf ab, klare Untergrundereignisse als solche zu kennzeichnen, um sich weitere Analysen zu ersparen. Wird die L4-Reject-Entscheidung mit der L5-Strahl-Gas-Klassifizierung kombiniert, erhält man mit sehr hoher Wahrscheinlichkeit ein Untergrundereignis. Etwa 2/3 der L1-Keep-Ereignisse erfüllten in der Runperiode 1993 diese Bedingung. Diese Menge bildete den Standard–„Untergrund“ für das Netztraining.
Es wurde versucht, die L5-Strahl-Gas-Klassifizierung mit einem mit L2-Daten trai-

L4-Keep(5566) ⇔ L4-Reject(4585)			
Testklasse	Netz **I**	Netz **II**	Netz**III**
L4-Keep(5563)	0.821	0.754	0.813
L4-Reject(4583)	0.250	0.270	0.223
nicht-Strahl-Gas(L5)(3371) ⇔ Strahl-Gas(L5) (2865)			
nicht-Strahl-Gas(L5)(3368)	0.844	0.784	0.915
Strahl-Gas(L5) (2863)	0.081	0.066	0.102
NC(L5)($Q^2 < 100 GeV^2$)(7145)	0.983	0.969	1.000
γp(L5)(4380)	0.526	0.427	0.789

Tabelle 6.5: **Testergebnisse für die auf die L4- und die L5-Strahl-Gas-Selektion trainierten Netze.** *Die Zahlen in Klammern geben die Statistik an.*

NC-(L5)($Q^2 < 100\,GeV^2$)(3573) ⇔ Untergrund(2865)			
Testklasse	Netz **I**	Netz **II**	Netz**III**
NC-(L5)($Q^2 < 100\,GeV^2$)(3572)	0.994	0.990	0.998
Untergrund(2863)	0.008	0.011	0.008
L4-Reject(4583)	0.102	0.095	0.079
Ratenreduktion(L1 $\to$ L2)	7.5	8.2	11.2

Tabelle 6.6: **Testergebnisse für auf die NC-(L5)($Q^2 < 100\,GeV^2$)-Klassifikation trainierte Netze.**

nierten Netz nachzuvollziehen. Die Ergebnisse finden sich in Tab. 6.5 . Das Physik-Untergrund-Verhältnis von etwa 10 zu 1 zeigt, daß bereits auf L2 eine Strahl-Gas-Klassifizierung möglich ist. Die Verteilung der Ausgabefunktion ist in Abb. 6.4 zu finden. Es wurden ebenfalls $NC_{Q^2<100\,GeV^2}$(L5)- und γp(L5)-Daten mit diesem Netz getestet. Während die NC-Ereignisse nahezu alle getriggert werden, sind es nur um die 50 % für die γp-Klasse. Dies bringt die größere Ähnlichkeit zwischen γp- und Strahl-Gas-Ereignissen zum Ausdruck.

6.4.3 NC-(L5)-Vorselektion ($Q^2 < 100\,GeV^2$)

Trigger und Detektor sind effizient für diese Ereignisklasse ausgelegt, weshalb die Vorselektionskriterien auf L5 (s. Tab. 6.1) auch eher harten kinematischen Schnitten entsprechen. Wie die Trainingsergebnisse in Tabelle 6.6 zeigen, genügen dem neuronalen Netz diese Vorselektionskriterien, um nahezu eine vollständige Separation der NC-Kandidaten für $Q^2 < 100\,GeV^2$ vom Untergrund zu erlernen. Das bestätigt die mit MC-Ereignissen erzielten hohen Effizienzen [33]. Der für das Training verwendete Untergrund wurde durch eine Kombination von L4-Reject- und L5-Untergrund-

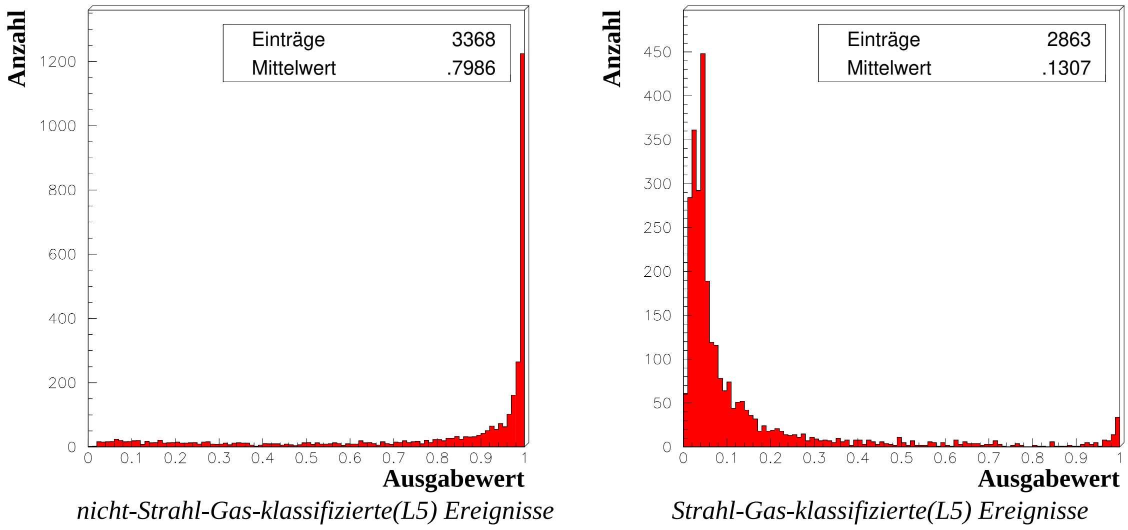

Abbildung 6.4: **Verteilung der Ausgabefunktion für ein auf die Strahl-Gas-(L5)-Klassifikation trainiertes Netz.** *Links ist die Verteilung für die Physikkandidaten, rechts die für den Untergrund zu sehen.*

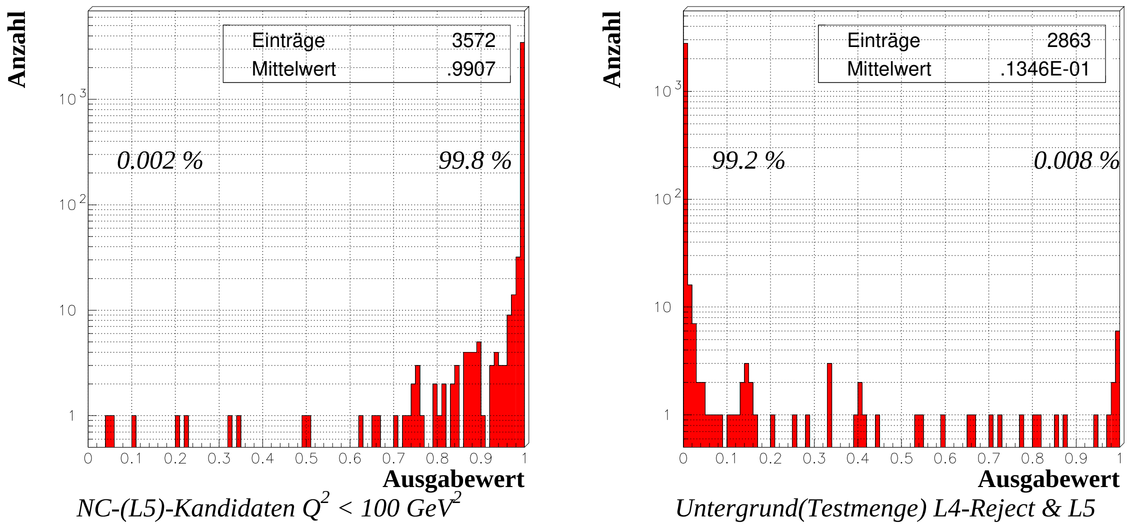

Abbildung 6.5: **Verteilung der Ausgabefunktion für ein auf die NC-(L5)-Vorselektion ($Q^2 < 100\,GeV^2$) trainiertes Netz.** *Links ist die Verteilung der Testmenge für die NC-Kandidaten, rechts die für den Untergrund zu sehen. Die logarithmische Skala ist zu beachten.*

NC-(L5)($Q^2 > 100\,GeV^2$)(3902,anger.) $\Leftrightarrow$ Untergrund(2865)			
Testklasse	Netz **I**	Netz **II**	Netz**III**
NC-(L5)($Q^2 < 100\,GeV^2$)(3899)	0.897(0.914)	0.905(0.910)	0.930(0.932)
Analyseselektion(866)	0.992(0.988)	0.986(0.923)	0.986(0.916)
Untergrund(2863)	0.065(0.065)	0.055(0.060)	0.057(0.048)
L4-Reject(4583)	0.226(0.146)	0.113(0.134)	0.137(0.109)
Ratenreduktion(L1 $\rightarrow$ L2)	3.9(5.4)	6.7(5.6)	5.7(6.9)

Tabelle 6.7: **Testergebnisse für auf die NC-(L5)($Q^2 > 100\,GeV^2$)-Klassifikation trainierte Netze.** *Die Zahlen in den Klammern gelten für ein nur auf die L5-Klasse trainiertes Netz, die Zahlen davor für ein um 867 Analyseselektionsereignisse angereichertes Netz.*

Klassifikation (Strahl-Gas, Pilotbündel) ermittelt. Dies entspricht nur $\approx 2/3$ der gesamten L1-Datenmenge. Die NC(L5)-Menge entspricht aber nur einem Bruchteil des restlichen Drittels. Der Unterschied in der Effizienz zwischen der mit „Untergrund“ bezeichneten Testmenge in Tab. 6.6 und der L4-Reject-Menge erklärt sich aus der Nichtverwendung eines Drittels der Daten im Training. Der Verzicht auf diese Teilmenge hat aber einen zweifachen Vorteil. Zum einen wird Ineffizienzen durch L4- oder L5-Klassifizierungsfehler aus dem Weg gegangen, zum anderen kann die Verbesserung in der Physikeffizienz dazu benutzt werden, eine härtere Triggerschwelle für die Ausgabefunktion zu wählen. Wie in Abb. 6.5 zu ersehen, ist auch ein Schnitt bei 0.8 ohne weiteres zu vertreten. Damit ergibt sich auch ein höherer Ratenreduktionsfaktor. Dieses aus der Gesamtmenge von L1-Keep-Daten errechnete Verhältnis der getriggerten L1- zu den getriggerten L2-Ereignissen ist in den Tabellen aufgeführt.

6.4.4 NC-(L5)-Vorselektion($Q^2 > 100\,GeV^2$)

Für $Q^2 > 100\,GeV^2$ trifft das gestreute Elektron nicht mehr auf das BEMC, sondern auf das Flüssigargonkalorimeter. Für 1993 gab es etwa 1500 solcher Ereignisse, und der eindeutigen Signatur wegen war die Erwartung, daß es sich um eine unkritische Klasse handelt. Die entsprechende L5-Selektion ist aber noch so weich, daß weniger als 5% der selektierten Ereignisse wirkliche NC-Ereignisse sind. Unter diesen Umständen war die Effizienz von ca. 90% (s. Tab. 6.7) für die L5-Klasse auch befriedigend. Nicht so gut verlief ein Test mit der Analyseselektionsmenge dieser Klasse (1728 Ereignisse; Elektronfinderalgorithmus, $E_t(e^-) > 10\,GeV$ oder $y < 0.85$, kein Cosmic-finder, Reinheit ca. 96%), abgesehen von dem Ergebnis für das Netz **I**. Um die Effizienz zu verbessern, wurde die (L5)-Vorselektion um die Hälfte der Analyseselektionsmenge angereichert und die zweite Hälfte zum Test verwendet. Das Ergebnis war zwar deutlich verbessert, lag aber nur knapp bei den angestrebten 99%. Die verbleibenden nicht getriggerten Ereignisse wurden daher einzeln untersucht. Es stellte sich heraus, daß es immer dieselben ca. 17 Ereignisse (von

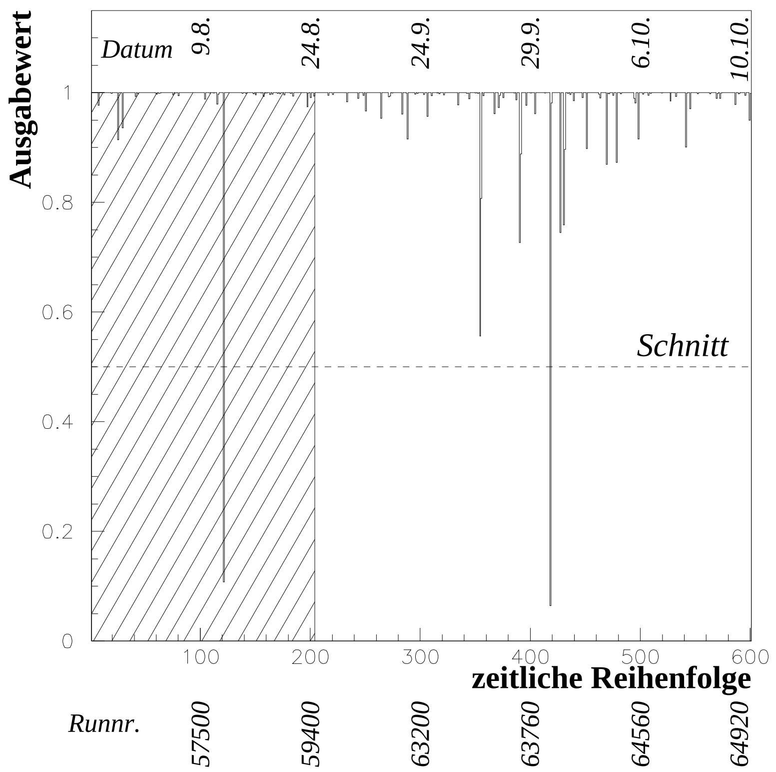

Abbildung 6.6: **Ausgabefunktion des auf (L5)-NC$_{highQ^2}$ trainierten Netzes für die Testmenge der Analyseselektion aufgetragen in der zeitlichen Reihenfolge der Ereignisse.** *Der Zeitraum aus dem die Trainingsdaten stammen, ist schraffiert.*

866) waren, die von den Netzen zurückgewiesen wurden. Im einzelnen waren das:

- 8-10 Ereignisse waren Untergrund (kosm. Strahlung, etc.)
- 3-4 Ereignisse konnten nicht eindeutig klassifiziert werden
- 4-5 Ereignisse hatten eine eindeutige NC-Signatur, trafen aber einen Bigtower im BBE (Fortsetzung des LAr im Übergangsbereich zum BEMC), der wegen Hochspannungsproblemen abgeschaltet war. Deshalb wurde auf Triggerebene nur wenig Energie im LAr registriert. Typischerweise sind diese Ereignisse von Spurtriggern getriggert.
- 1-2 Ereignisse betrafen andere L1-Triggerprobleme

Allen diesen Ereignissen war eine geringe Energiedeposition im LAr auf Triggerebene gemeinsam. Damit waren die Ineffizienzen im wesentlichen auf technische Probleme des L1-Triggers zurückzuführen. Auch unter Einbeziehung der technischen Probleme betrug die Effizienz $\sim 99\%$.
Da sowohl die für diese Analyse verwendeten L5-Daten, als auch die Untergrunddaten aus dem ersten Drittel der Runperiode 1993 entstammen, die Analysedaten hinge-

CC-(L5)-Vorselektion(2475) ⇔ Untergrund(2865)			
Testklasse	Netz **I**	Netz **II**	Netz**III**
CC-(L5)-Vorselektion(2473)	0.898	0.897	0.901
Untergrund(2863)	0.111	0.103	0.135
11 „heiße Kandidaten“	1.000	1.000	1.000
L4-Reject(4583)	0.245	0.174	0.178
Ratenreduktion(L1 $\rightarrow$ L2)	3.7	4.6	4.5

Tabelle 6.8: **Testergebnisse für auf die CC-(L5)-Klassifikation trainierte Netze.**

gen sich über die ganze Runperiode erstrecken, bot sich an, das Zeitstabilitätsverhalten des neuronalen Netzes mit Hilfe der Analysedaten zu überprüfen. Dazu ist der Ausgangswert des Netzes für die Testmenge der Analysedaten in Abb. 6.6 in der zeitlichen Reihenfolge des Auftretens der Ereignisse aufgetragen. Der Zeitraum, aus dem Untergrund- und Vorselektionsdaten für das Training stammen, ist schraffiert. Es ist zu erkennen, daß der Ausgabewert für die meisten Ereignisse sehr nahe am Zielwert von 1.0 liegt. Desweiteren ist das stabile Zeitverhalten des neuronalen Netzes bezüglich der Physikeffizienz auch über den Zeitraum für die Trainingsdaten hinaus gut zu erkennen.
Die in der Tabelle aufgeführten Reduktionsraten von ca. 5 sind für diese Klasse eigentlich zu gering, Abb. 6.6 zeigt aber, daß sich der Schnitt zu höheren Werten als 0.5 verschieben läßt, ohne daß es zu größeren Verlusten in der Physikeffizienz kommt. Der Einfluß dieser Verschiebung auf die Triggerrate wird in 7.1 quantifiziert.

6.4.5 CC-(L5)-Vorselektion

Wegen der geringen zur Verfügung stehenden Statistik haben für diese Ereignisklasse MC-Untersuchungen größeres Gewicht. Das Training von CC-(L5)-Vorselektionsdaten demonstriert mehr die Fähigkeit, CC-ähnlichen Untergrund von anderem Untergrund zu trennen, da wirkliche CC-Ereignisse nur einen Anteil von 0.1% an der CC-(L5)-Datenmenge haben. Die Ergebnisse sind in Tab. 6.8 zusammengefaßt. Der Test mit 11 „heißen Kandidaten“ zeigt zumindestens, daß das auf die Vorselektion trainierte Netz in der Lage ist, diese Ereignisse zu triggern. Wegen des hohen Schnittes auf den fehlenden Transversalimpuls von $25\,GeV$ können daraus aber keine Schlüsse abgeleitet werden. MC-Untersuchungen mit der kinematisch natürlichen Verteilung ergaben befriedigende Resultate [49].

6.4.6 γp(L5)-Vorselektion

Es wurden L5-vorselektierte γp-Daten(JETJET,PGAMMA) untersucht, die überwiegend mit dem e^--Tagger getriggert wurden. Obwohl die Resultate allein für die L2–Daten ohne Einbeziehung des Elektrontaggers als gut zu bezeichnen sind,

γp-(L5)-Vorselektion(2190) $\Leftrightarrow$ Untergrund(2865)			
Testklasse	Netz **I**	Netz **II**	Netz**III**
γp-(L5)-Vorselektion(2190)	0.838	0.896	0.994
Untergrund(2863)	0.106	0.088	0.003
L4-Reject(4583)	0.163	0.116	0.015
Ratenreduktion(L1 $\rightarrow$ L2)	4.8	6.2	40

Tabelle 6.9: **Testergebnisse für auf die γp-(L5)-Klassifikation trainierte Netze.**

	Testmengen					
Training	NC$_{lowQ^2}$(L5)	NC$_{highQ^2}$(L5)	CC(L5)	γp(L5)	Untergr.	Rat.red.
NC$_{lowQ^2}$	**0.998**	0.449	0.225	0.391	0.008	11.2
NC$_{highQ^2}$	0.961	**0.932**	0.792	0.987	0.048	6.9
CC	0.948	0.866	**0.901**	0.586	0.135	4.5
γp	0.150	0.142	0.199	**0.994**	0.003	40
verodert	1.000	0.960	0.921	0.999	0.166	4.0

Tabelle 6.10: **Kombinierte Testergebnisse für die auf die (L5)-Vorselektion trainierten Netze.** *In der letzen Reihe sind die Effizienzen der durch ein logisches Oder verknüpften Triggerentscheidungen aufgeführt.*

ergibt sich unter Hinzunahme der Elektrontagger-Energie eine erhebliche Verbesserung von Physikeffizienz und Ratenreduktion für das Netz **III** (s. Tab. 6.9). Insgesamt sind getaggte γp-Ereignisse eine unkritische Ereignisklasse. Interessanter, aber auch schwieriger ist deshalb die Untersuchung von nicht getaggten 1- und 2-Jet-γp–Ereignissen in Abschnitt 6.5.1 .

6.4.7 Zusammenfassung

Die vorigen Abschnitte haben gezeigt, daß Vorselektionskriterien durchaus geeignet sind, Trainingsmaterial für einen L2-Trigger zu liefern. Die eher geringen Ratenreduktionsfaktoren weisen aber auch darauf hin, daß bei Designluminosität Verbesserungen durch schärfere Schnitte, höhere Qualität der Eingangsdaten und die Verwendung detaillierter Informationen (z. B. z-Kammern) notwendig werden. Unter den gegebenen Randbedingungen, daß sich H1 in einer frühen Betriebsphase befindet und das gesamte Triggersystem noch nicht vollständig ausgebaut und funktionsfähig ist, sind die erzielten Resultate aber zufriedenstellend. In Tabelle 6.10 sind die gemeinsamen Effizienzen der kombinierten (L5)-trainierten Netze aufgeführt. Da die Triggerentscheidungen der auf verschiedene Physikklassen trainierten Netze verodert werden, sind auch die sich

überkreuzenden Effizienzen zu berücksichtigen. Es ist z. B. sehr natürlich, daß ein auf die CC-Vorselektion trainiertes Netz auch auf NC-Ereignisse sensitiv ist, da die Topologie bezüglich des hadronischen Jets identisch ist. Die Veroderung aller Triggermodule führt zu einer Verbesserung der Physikeffizienz eines einzelnen Moduls, die mögliche Verschlechterung der Gesamtreduktionsrate muß aber im Auge behalten werden.
Abschließend sei noch auf die Fehlerquellen für die in den Tabellen aufgeführten Effizienzen eingegangen. Diese Zahlen sind als Ergebnis eines deterministischen Algorithmus (bei fixen Startzahlen für den Zufallsgenerator) zwar exakt reproduzierbar, die Verwendung experimenteller Daten erfordert es aber, auf die in ihnen enthaltenen Fehler Rücksicht zu nehmen. Im einzelnen sind zu nennen:

- Limitierung in der Statistik (siehe Tabelle im Anhang)
- technische Probleme von Unterkomponenten des L1-Triggers
- Meßfehler
- wechselnde Maschinenbedingungen (Ströme, Vakuum)

Der letzte Punkt kann durch die Schwankungen der Effizienzen für zeitlich getrennte Mengen an L4-Reject-Ereignissen ermittelt werden. Die Reduktionsrate weist größenordnungsmäßig Schwankungen von $\pm 25\%$ auf. Der Ausfall einzelner Subkomponenten ist ein sehr ernster Punkt, da er die Vergleichbarkeit von Daten erheblich beeinflußt. Das Datenextraktionsprogramm wurde deshalb auch dahingehend geändert, Ausfälle einzelner Subdetektoren automatisch abzufragen. Begrenzte technische Defekte wie der Ausfall eines einzelnen Bigtowers können bei der Komplexität des Gesamtdetektors nicht im einzelnen berücksichtigt werden. Ein funktionsfähiger L2-Trigger muß mit solchen Bedingungen zu recht kommen können. Die Untersuchung zu den NC-Ereignissen mit $Q^2 > 100\,GeV^2$ zeigt, daß der Einfluß auf die L2-Triggereffizienz merklich werden kann. Systematische Meßfehler beeinflußen die Effizienz dann erheblich, wenn sie die Vergleichbarkeit der Trainingsdaten durch zeitliche Schwankungen herabsetzen. Der eigentliche Meßfehler ist implizit durch Angabe einer Effizienz bereits ausgedrückt. Eine verbesserte Energiemessung des LAr-Triggers wird sich z. B. auch in besseren Effizienzen niederschlagen. Die Limitierung in der Statistik geht auch über den ansteigenden Generalisierungsfehler für komplexere Netze ein. Dieser Effekt ist zwar klein gegenüber den bisher erwähnten, führt aber in Fällen, in denen die feinere Granularisierung keine Verbesserungen mit sich bringt, zu etwas schlechteren Effizienzen für die Netze **II** und **III** gegenüber dem kleinsten Netz **I** mit 28 Eingangsvariablen.

6.5 Kritische Physikklassen

6.5.1 1–Jet- und 2–Jet-γp-Ereignisse

Für kritische Physikklassen müssen die härteren Selektionskriterien einer Physikanalyse verwendet werden, um eine brauchbare Trainingsmenge zu erhalten, oder ein Training mit einer natürlichen Verteilung in MC-Daten durchgeführt werden. Die zur Verfügung stehenden 1–Jet- und 2–Jet-γp-Ereignisse entstammen dem zweiten Teil der Runperiode 1993. Häufige und unregelmäßige Ausfälle des Myon-Triggers machten es notwendig, die entsprechenden Größen aus dem Eingabevektor zu entfernen, so daß nicht mehr die Netzkonfigurationen des vorigen Abschnitts verwendet werden konnten. Die kinematischen Schnitte für die 1–Jet-Ereignisse sind [52]

- Kegelalgorithmus zur Eingrenzung der Jets
- $p_t(jet) > 7GeV$; $-1 < \eta < 1.5$
- L1-Triggersignal im LAr-Barrel & r-Φ–Driftkammer-Triggersignal.

Folgende kinematischen Schnitte gelten für die 2–Jet-Ereignisse [53]

- $E_t(jet) > 20\,GeV$ für $\Theta > 10^0$
- 2 Jets, jeweils $p_t > 8\,GeV$
- kein Elektronkandidat
- keine Spuren aus dem rückwärtigen Raum
- topologischer Schnitt gegen Strahl-Gas-Ereignisse.

Beide Jet-Klassen wurden spezifisch gegen Untergrunddaten aus dem jeweiligen Zeitraum der Datennahme trainiert. Von besonderem Interesse ist für diese Klasse die Benutzung von Jet-Energien und Jet-Winkel auf Bigtower-Ebene. Diese Größen müssen mit entsprechendem Hardware-Aufwand erst in der dem neuronalen Netz vorgeschalteten DataBox berechnet werden, und es wurde deshalb ermittelt, welche Trenneffizienzen durch die Verwendung dieser Größen erreicht werden kann. Anstelle der genauen Winkel wurden interne Indizes für das Training verwendet. Die mittleren Winkel eines Bigtowers stellen nur eine Näherung der wirklichen Jetwinkel dar (max. Abweichungen liegen für den Θ-Winkel zwischen 1^0 und 15^0, s. Abb. 5.4). Die Trainingsresultate sind in Tabelle 6.11 aufgelistet. Wegen der geringen zur Verfügung stehenden Statistik für die 2–Jet-Ereignisse wurde die Testmenge auf 408 Ereignisse gegenüber der Trainingsmenge mit 1225 Ereignissen verkleinert.
Das erste Netz wurde nur mit den Jetgrößen (Energie, Θ– und Φ–Index der drei Bigtower-Jets mit der größten Energie) trainiert, ein zweites um die Totalenergie ergänzt. Die Ergebnisse zeigen, daß die in dieser Art definierten Jetgrößen als Klassifikationskriterium herangezogen werden, die Effizienzen sind aber mit etwa 80% gegen 20% nicht besonders hoch ausgefallen. Durch die Hinzunahme der gewichteten totalen kalorimetrischen Energie läßt sich das Ergebnis deutlich verbessern. Das ist auch verständlich, da sich der Untergrund insbesondere durch eine andere Verteilung des Gesamtenergieflusses im LAr von diesen γp-Jet-Ereignissen unterscheidet. Unter Verwendung der Größen des Netz **II**(ohne μ) aus dem vorigen Abschnitt werden jedoch bessere Resultate erzielt. Die Korrelationen verschiedener Subdetektor-Informationen besitzen in diesem Fall einen höheren Informationsgehalt gegenüber den Jetgrößen. Eine der Ursachen liegt sicherlich in den Ungenauigkeiten der Energiemessung auf Triggerebene (Die Triggerschwelle entsprach für 1993 ca. $1.2\,GeV$). An einer Verbesserung dieser Energiemessung wird gearbeitet. Die Hinzunahme der Jets zu den bisherigen Größen führt für die 1–Jet-γp-Ereignisse nur zu einer geringfügigen Verbesserung der Effizienzen. Bei den 2–Jet-γp-Ereignissen kann aufgrund der geringen Statistik der Test- und Trainingsmenge keine verläßliche Aussage über diesen letzten Punkt gemacht werden. Dafür ist eine größere Datenmenge erforderlich.

1–Jet-γp-Klasse(3915) $\Leftrightarrow$ Untergrund(4086)				
Testklasse	Jet-Netz(9)	Jet+E_{tot}(10)	Netz**II**(o. μ)(41)	Jet&Netz**II**(50)
1–Jet-γp(1960)	0.813	0.889	0.945	0.869
Untergrund(2040)	0.243	0.151	0.079	0.042
2–Jet-γp-Klasse(1225) $\Leftrightarrow$ Untergrund(1276)				
1–Jet-γp(408)	0.797	0.838	0.939	0.953
Untergrund(1275)	0.202	0.154	0.035	0.038

Tabelle 6.11: **Testergebnisse für auf 1–Jet-γp- und 2–Jet-γp-Ereignisse trainierte Netze.** *Die Zahlen in Klammern hinter den Ereignisklassen geben die Statistik an und die Zahlen in Klammern hinter den Netzbezeichnungen die Dimension des Eingabevektors.*

D^*–Kandidaten(2942) $\Leftrightarrow$ Untergrund(3058)				
Testklasse	Jet-Netz(9)	Jet+E_{tot}(10)	Netz**II**(o. μ)(41)	Jet&Netz**II**
D^*–Kandidaten(1961)	0.741	0.712	0.887	0.857
Untergrund(2039)	0.455	0.438	0.142	0.124
K^0-Kandidaten(2942) $\Leftrightarrow$ Untergrund(3058)				
K^0-Kandidaten(1961)	0.738	0.725	0.888	0.895
Untergrund(2039)	0.467	0.442	0.171	0.179

Tabelle 6.12: **Testergebnisse für die auf D^*-Kandidaten und K^0-Kandidaten trainierten Netze.**

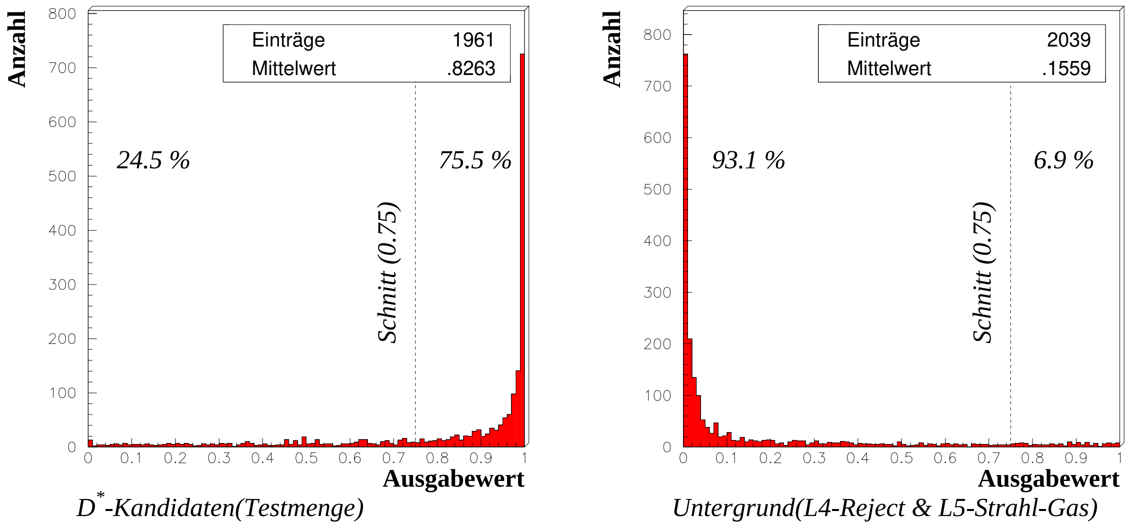

Abbildung 6.7: **Verteilung der Ausgabefunktion für das auf D^*-Kandidaten trainierte Netz** Jet&Netz **II.** *Die eingetragenen Effizienzen beziehen sich auf einen Schnitt bei 0.75 .*

6.5.2 Kandidaten der Erzeugung schwerer Quarks

Diese Ereignisklasse bereitete im Training die meisten Probleme. Die derzeitige Genauigkeit der kalorimetrischen Energiemessung auf Triggerebene ist nicht effizient genug für Energiedepositionen in der Größenordnung von 1 GeV, wie sie für die Erzeugung schwerer Quarks (als Quark-Antiquark–Paare) typisch ist. Auf Informationen der Spurtrigger kann für diese Klasse nicht verzichtet werden. Desweiteren ist die Separation von Untergrundereignissen auch bei exakter Energiemessung nur mit statistischen Methoden möglich, weshalb hier auch nur von „Kandidaten“ die Rede ist. Für das Training wurden D^*- und K^0-Kandidaten verwendet, diese wurden aus der entsprechenden (L5)-Vorselektionsklasse durch Schnitte auf die aus den Spurenergien ermittelte invariante Masse extrahiert. Die Bezeichnung K^0-Kandidaten bezieht sich auf Zerfallskanäle von D-Mesonen über dieses Teilchen. Es wurden dieselben Netzarchitekturen wie für die γp–Jet-Klassen verwendet. Die Jetgrößen waren hier weniger erfolgreich als in den γp-Klassen, die Hinzunahme der gesamten Kalorimeterenergie führt ebenfalls zu keiner signifikanten Verbesserung. Die Trenneffizienzen für das Netz **II** mit den in ihnen enthaltenen Spurtriggerinformationen sind deutlich besser. Für eine ernsthafte Triggeranwendung ist die Effizienz der Untergrundunterdrückung etwa um den Faktor 2 zu schlecht. In den in Abb. 6.7 aufgetragenen Verteilungen für das auf D^*-Kandidaten trainierte Netz „Jet & Netz **II**“ wurde deshalb exemplarisch ein Schnitt bei 0.75 durchgeführt und die entsprechenden Effizienzen angegeben.

Kapitel 7

Analysen der Trainingsergebnisse

7.1 Triggerraten

Die Untergrundrate läßt sich durch einen Schnitt auf den Ausgabewert variieren. Eine Erhöhung der Triggerschwelle ist aber immer mit einer Einbuße an Physikeffizienz verbunden. Deshalb ist es wichtig, die Triggerrate in Abhängigkeit von der erreichten Physikeffizienz zu ermitteln. Die Triggerrate kann durch einen Test über die Menge der L1–Keep–Daten errechnet werden. Eine L1–Keep–Datenmenge kann unter Berücksichtigung der entsprechenden L4–Reduktionsrate für die Runperiode 1993 aus L4–Keep– und L4–Reject–Monitor–Ereignissen rekonstruiert werden.
Entsprechend der dynamischen Steigerung der Luminosität in den nächsten Jahren ist eine Angabe der Rate in Hertz ungünstig, da diese sich nur auf einen festzulegenden Maschinenzustand bezieht. Ein allgemeineres Bewertungskriterium ist die Reduktionsrate eines L2–Triggers relativ zur L1–Rate in Abhängigkeit von der erreichten Physikeffizienz. Diese Größe wurde aus dem Verhältnis von L2–Keep– zu L1–Keep–Daten für die Einstellungen des L1-Triggers aus der Runperiode 1993 errechnet. Dabei entsprach etwa 1% der geplanten Luminosität einer L1-Rate von $30 - 50\,Hz$. Extrapoliert auf die volle Luminosität ergibt sich die vorgesehene Rate von $5\,kHz$ (s. Tab. 5.1). Gravierende Änderungen der L1-Triggereinstellungen sind deshalb nicht mehr zu erwarten.
In Abb. 7.1 ist die Ratenreduktion in Abhängigkeit von der Physikeffizienz für verschiedene auf L5-Klassen trainierte Netze aufgetragen. In den Diagrammen sind jeweils die mit den in 6.4 definierten Netzarchitekturen Netz **I, II** und **III** erzielten Ergebnisse dargestellt. Die Ratenreduktion wurde durch einen Test mit einer rekonstruierten L1–Keep–Datenmenge mit einer Anzahl von $\mathcal{O}(10^4)$ Ereignissen bestimmt. Die Größe *Physikeffizienz* bezieht sich auf die unabhängige Testmenge der jeweiligen L5-Vorselektionsklasse mit Ausnahme der Netze für $\mathrm{NC}_{Q^2>100GeV^2}$. Diese Ausnahme wurde gemacht, da bei Testdurchläufen für die Analyseselektionsmenge Verluste in der Effizienz zu beobachten waren, und die wirklichen NC-Ereignisse in dieser Klasse mit ca. 5% deutlich unterrepräsentiert sind. Die Verunreinigung dieser Analyseselektionsmenge beträgt noch etwa 4%.
Für beide NC-L5-Klassen ($Q^2 < 100\,GeV^2$ und $Q^2 > 100\,GeV^2$) ergeben sich akzeptable Physikeffizienzen bis hin zu Ratenreduktionsfaktoren von 15-20. Für höhere Ratenreduktionsfaktoren sind die auf die Vorselektion trainierten Netze allerdings nicht

mehr geeignet. Das Raten-Effizienz-Diagramm für die $NC_{Q^2>100GeV^2}$-Klasse weist die Besonderheit auf, daß das kleinste Netz **I** mit den wenigsten Größen bei einer Ratenreduktion > 10 die höchste Effizienz aufweist, während für kleinere Reduktionsfaktoren die Netze **II** und **III** bessere Resultate erzielen. Für die CC-Vorselektionsklasse ergibt sich eine Effizienz von 80% bei einer Ratenreduktion von 10. Die Resultate der verschiedenen Netzkonfigurationen sind nahezu identisch. Das Diagramm für die getaggte γp-Klasse bestätigt, daß diese Klasse hinsichtlich der Reduktionsrate als unkritisch zu bezeichnen ist. Das Netz **III** übertrifft die Netze **I** und **II** in der Ratenreduktion jedoch um zwei Größenordnungen. Das ist insbesondere auf die Hinzunahme der Elektrontagger-Energie zurückzuführen.
Abb. 7.2 zeigt die entsprechenden Raten-Effizienz-Diagramme für die kritischen Physikklassen. Die Physikeffizienzen beziehen sich wiederum auf die für das Training nicht verwendeten Testmengen. Es sind die Ergebnisse dargestellt, die mit dem Netz **II**(ohne μ) und unter Hinzunahme der Bigtower-Jetgrößen erzielt wurden. Für die 1–Jet–γp-Klasse ergab sich eine Ratenreduktion von 10-15 bei Effizienzen von $(90-95)\%$. Die Resultate für das mit den Jetgrößen trainierte Netz sind nahezu identisch mit dem Netz ohne diese Größen. Für die 2–Jet–γp-Klasse ergeben sich Effizienzen von $(93-97)\%$ bei einer Reduktionsrate von 10-15. Da die Testmenge der Physikklasse hier nur aus 408 Ereignissen besteht, beruht das Ergebnis für eine Physikeffizienz von 95% auf einer Zurückweisung von ca. 20 Ereignissen. Die entsprechende statistische Unsicherheit liegt bei ± 5 Ereignissen. Für diese Klasse zeigt das um die Jetgrößen erweiterte Netz bessere Resultate.
Für die Kandidaten der schweren Quarks ergibt sich bei einer Ratenreduktion um den Faktor 10 eine Physikeffizienz im Bereich von $(70-80)\%$. Bei einer Ratenreduktion nur um den Faktor 5 lassen sich Effizienzen um 90% erzielen. Da diese Ratenreduktion bezogen auf die ganze L1-Keep-Menge gering ist, ist es hier denkbar, nur von bestimmten L1-Subtriggern getriggerte Ereignisse auf L2 für diese Klasse zu berücksichtigen.

7.2 Bedeutung einzelner Eingabegrößen

Es existieren keine allgemeingültigen Kriterien für die Auswahl von Eingabegrößen eines Backpropagation-Netzes. Lineare statistische Methoden sind fragwürdig, da neuronale Netze auch in der Lage sind, nichtlineare Korrelationen auszunutzen. Da die Anzahl der Eingabevariablen aber unmittelbar in den Generalisierungsfehler eingeht, ist es notwendig, herauszufinden, welche Eingabegrößen von einem neuronalen Netz zur Klassifikation benutzt werden. Um den quantitativen Einfluß einer Variation einzelner Eingabegrößen auf Änderungen im Ausgabewert zu ermitteln, wurde folgende Größe definiert:

$$r_i \equiv \sum_{j=1}^{n_{train}} \left(o(\vec{x}_j) - o(\vec{x}_j | x_{j,i} = \bar{x_i})\right)^2 / n_{train}. \qquad (7.1)$$

Für die i-te Eingangsgröße wird der Wert x_i auf seinen Mittelwert gesetzt und die sich ergebende Variation des Ausgangswerts ermittelt. Der Wert von r_i wird durch Summierung über alle Trainingsmuster und Mittelung über deren Anzahl bestimmt. Diese Größe wird im folgenden *Relevanz* genannt. Sie ist ein Maß dafür, welchen Anteil eine

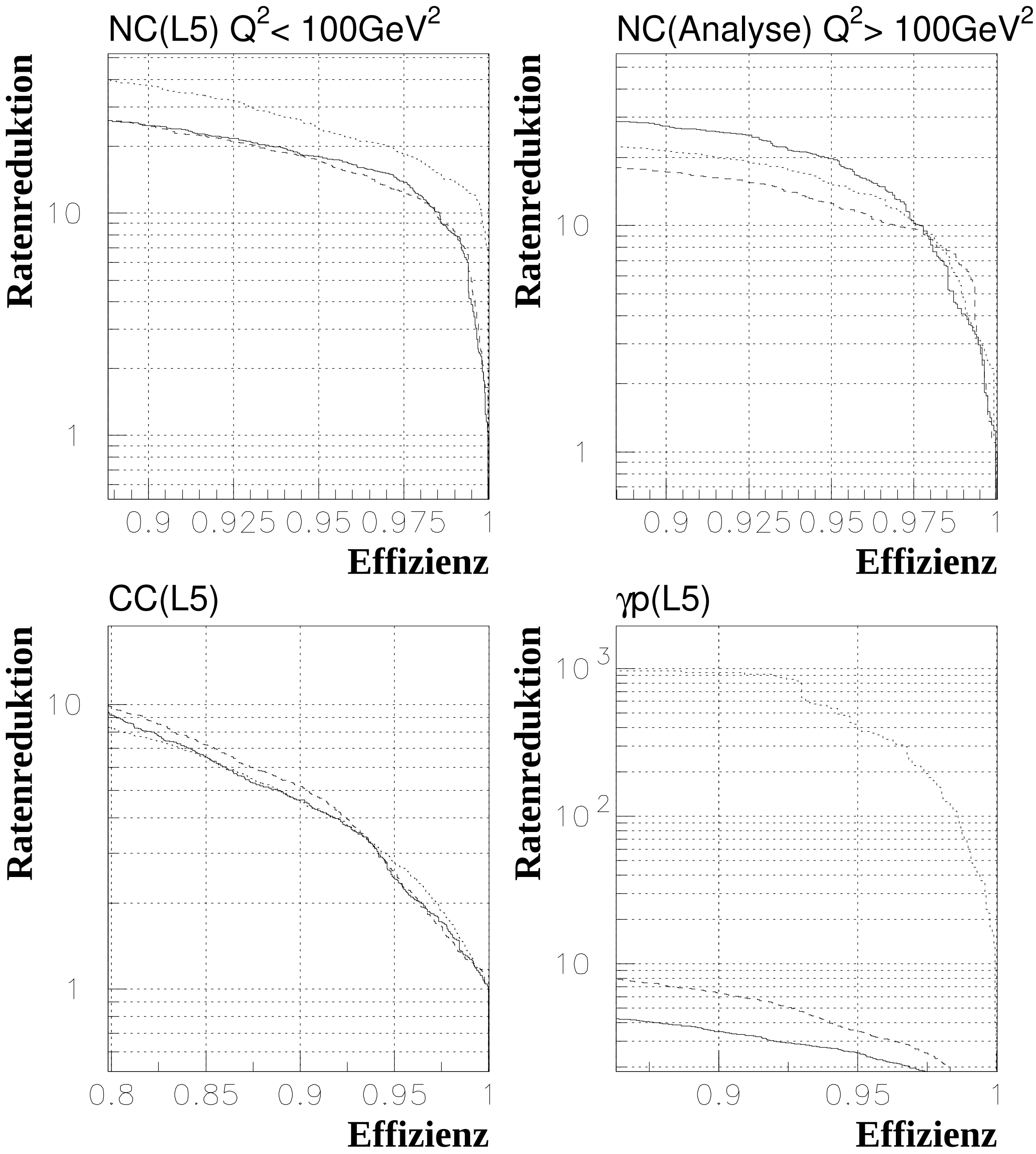

Abbildung 7.1: **Effizienz-Raten-Diagramme für die auf L5-Vorselektionsklassen trainierten Netze.** *Die durchgezogene Kurve entspricht Netz* **I**, *die gestrichelte Kurve Netz* **II** *und die gepunktete Kurve Netz* **III.**

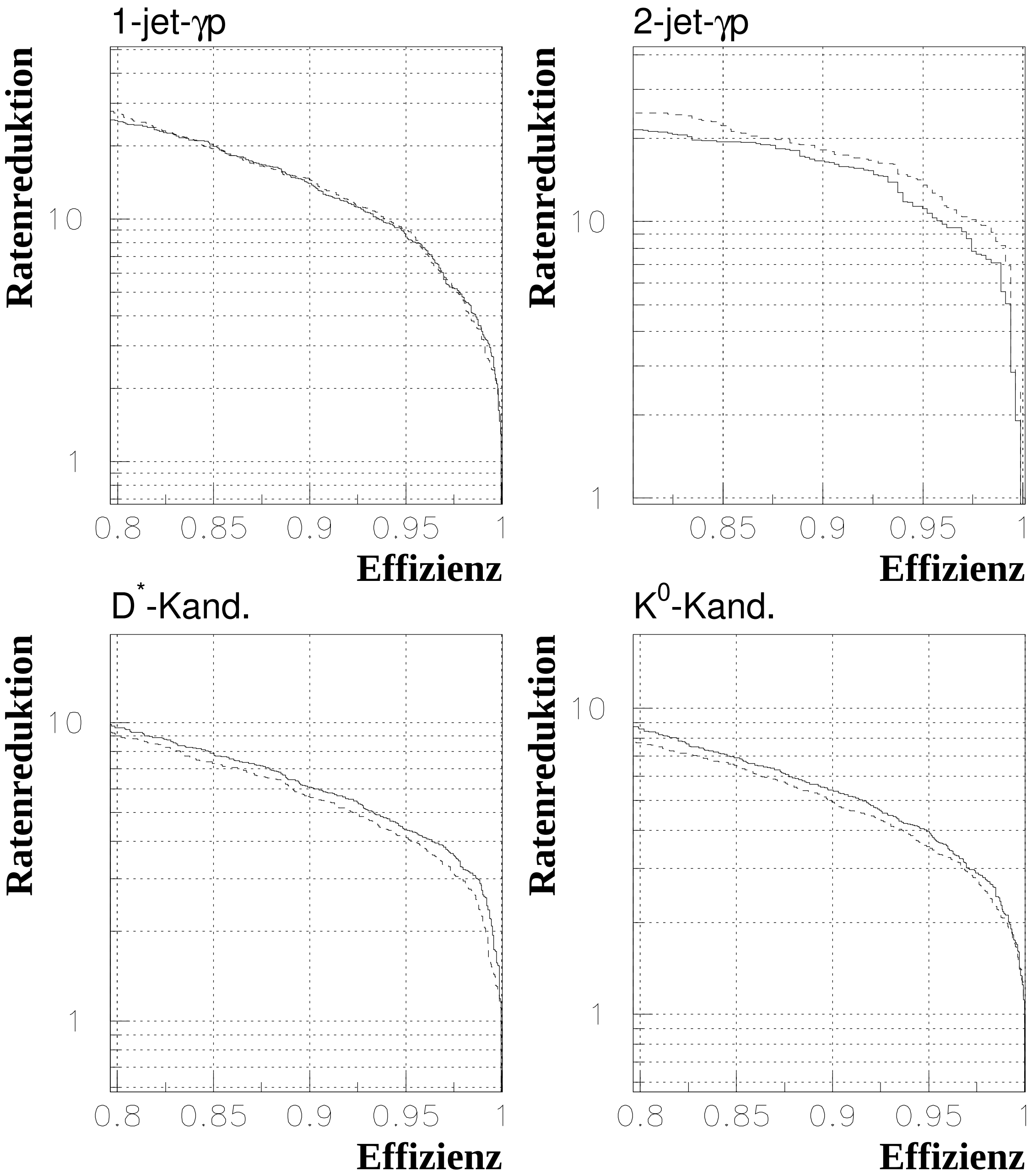

Abbildung 7.2: **Effizienz-Raten-Diagramme für die auf kritische Physikklassen trainierten Netze.** *Die durchgezogene Kurve entspricht einem Netz ohne Bigtower-Jetgrößen. Die gestrichelte Kurve entspricht demselben Netz unter Hinzunahme der Bigtower-Jetgrößen.*

einzelne Größe an der Klassifikationsentscheidung eines neuronalen Netzes hat. Korrelationen mehrerer Größen werden nicht erfaßt. Ein Vergleich des Absolutwerts von r_i zwischen zwei unterschiedlichen Netzen ist schwierig. Dieselbe Information kann redundant in zwei Variablen vorhanden sein. Damit können sich sehr unterschiedliche Werte für r_i bei gleicher Effizienz ergeben. Die Aussagekraft beschränkt sich daher auf den Vergleich verschiedener Eingabegrößen innerhalb eines Netzes. In den Abbildungen wurde deshalb die Relevanz r_i aller Eingabegrößen eines Netzes jeweils nach der Größe ihres Absolutwertes sortiert.
Im oberen Teil der Abb. 7.3 sind die Resultate der Konfiguration Netz **I** für die NC(L5)-Vorselektionsklassen dargestellt. Bezüglich der Klasse NC(L5)$_{Q^2<100\,GeV^2}$ haben die BEMC-Gesamtenergie(bemce) und die Anzahl getroffener Streifen in der BPC(bpctot) die größte Relevanz. Deutlich ist die Dominanz der Größen, die den rückwärtigen Detektorraum beschreiben, gegenüber den restlichen Größen zu erkennen. Das entspricht der physikalischen Erwartung, da das Elektron für Impulsüberträge $Q^2 < 100\,GeV^2$ in diesen Bereich gestreut wird. Indirekt ist dies auch ein Indiz dafür, daß die Vergleichbarkeit zwischen Physik- und Untergrunddaten gegeben ist, und keine unphysikalischen Korrelationen das Ergebnis verfälschen.
Für die NC(L5)$_{Q^2>100\,GeV^2}$-Klasse stützt sich das Netz ebenfalls auf die erwarteten Größen. Die topologischen Energieinformationen des Flüssigargonkalorimeters weisen die größte Relevanz auf (Transversalenergie(laret), Anzahl triggernder Bigtower(bigtnum), gewichtete Gesamtenergie(laretot), etc.). Für den Übergangsbereich um $Q^2 \approx 100 GeV^2$ ist auch die Clusterenergie des BEMC(bemcclu) eine wichtige Größe.
Die Relevanz-Diagramme für die γp(L5)-Netze zeigen, daß dieses Klassifikationsproblem nahezu alleine durch einen Schnitt auf die Elektrontagger-Energie lösbar ist. Daraus muß aber auch die Schlußfolgerung gezogen werden, für diese Klasse auf ein neuronales L2-Modul zu verzichten, da ein entsprechend effizientes L1-Subtriggerelement mit geringer Rate definiert werden kann. Die durch den Elektrontagger erfaßten Ereignisse stellen aber nur einen Bruchteil aller Photoproduktionsereignisse dar. Die Effizienzen der ungetaggten γp-Ereignisse haben aus Sicht eines L2-Triggers deshalb einen höheren Stellenwert. Ohne den Elektrontagger stützt sich das Netz **I** für die Klassifikation auf mehrere Subdetektorkomponenten.
Für die Photoproduktions–Jetklassen sind die Relevanz-Diagramme in Abb. 7.4 jeweils für das auf die Bigtower-Jetgrößen trainierte Netz und für die Konfiguration „Netz **II** & Jet“ dargestellt. Bezüglich der BT-Jetgrößen stützt sich das Netz für die 1-Jet–γp–Klasse vorwiegend auf die Energien(jet(..)), während für die 2-Jet–γp–Klasse die Θ–Indizes(jtheta(..)) bevorzugt werden. Für die 1-Jet–γp–Klasse ist die Relevanz für den dritt- und zweit-energiereichsten BT-Jets größer als für den Jet mit der höchsten Energie, was auf die Bedeutung des Gesamtenergieflusses zur Unterscheidung von Untergrundereignissen hinweist. Der Detektorsymmetrie entsprechend ist die Bedeutung des Φ-Index am geringsten. Unter Einbeziehung aller Größen stützt sich das Netz aber bevorzugt auf eine Kombination des maximalen Eintrages in das CIP-COP-z-Vertex-Histogramm(cpvmax) und der gewichteten(laretot) oder ungewichteten(bigterg) Gesamtenergie. Die Jetgrößen haben jedoch eine Nebenbedeutung. Für die Klasse der D^*-Kandidaten in Abb. 7.5 zeigt sich, daß von den BT-

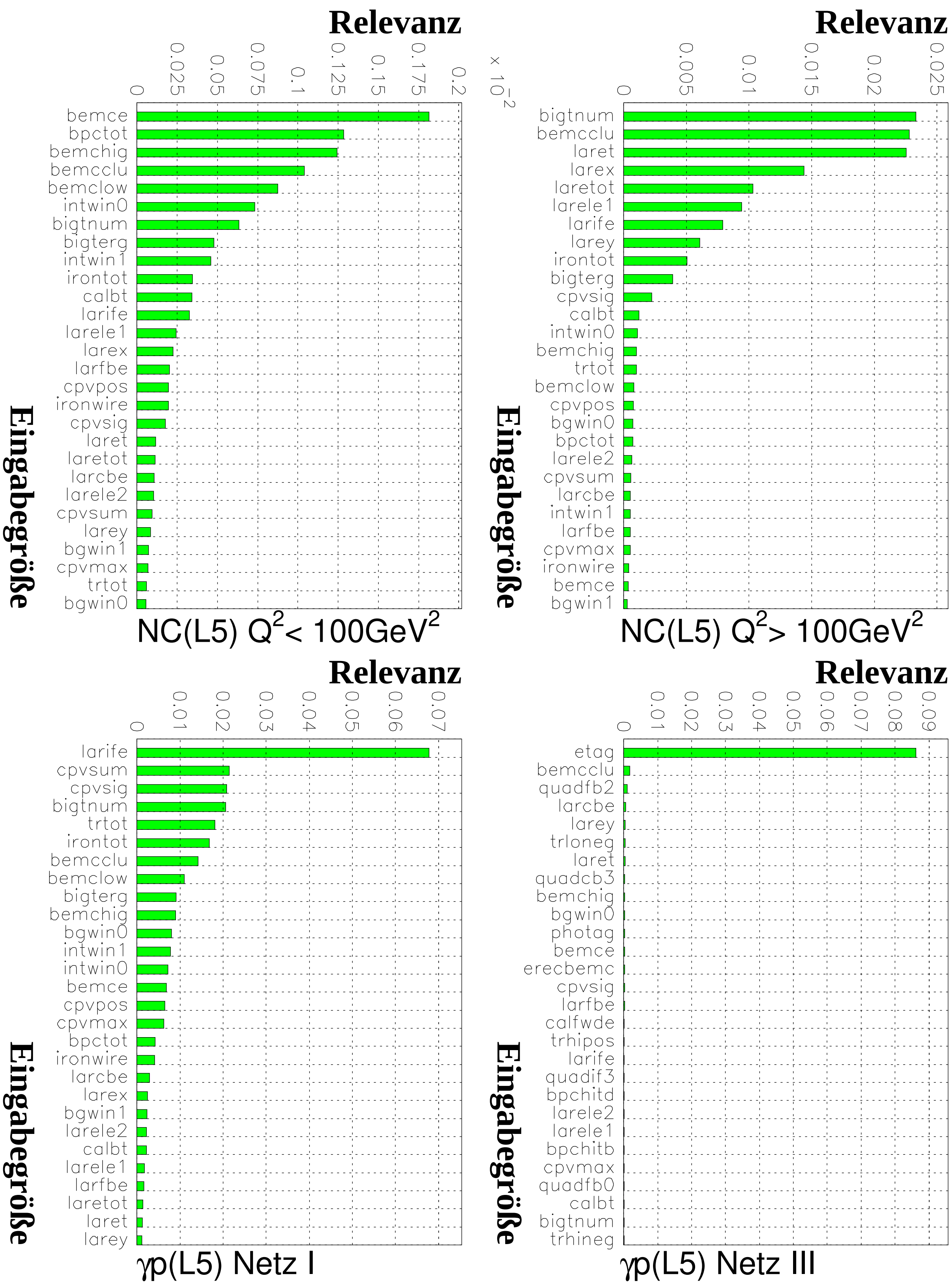

Abbildung 7.3: **Relevanz-Diagramme für die L5-Vorselektionsklassen.** *Oben: NC-(L5)-Klassen für das Netz* **I**. *Unten: γp-(L5)-Klasse für das Netz* **I** *und* **III**. *Für das Netz* **III** *sind nur die 28 Eingabegrößen mit der größten Relevanz dargestellt.*

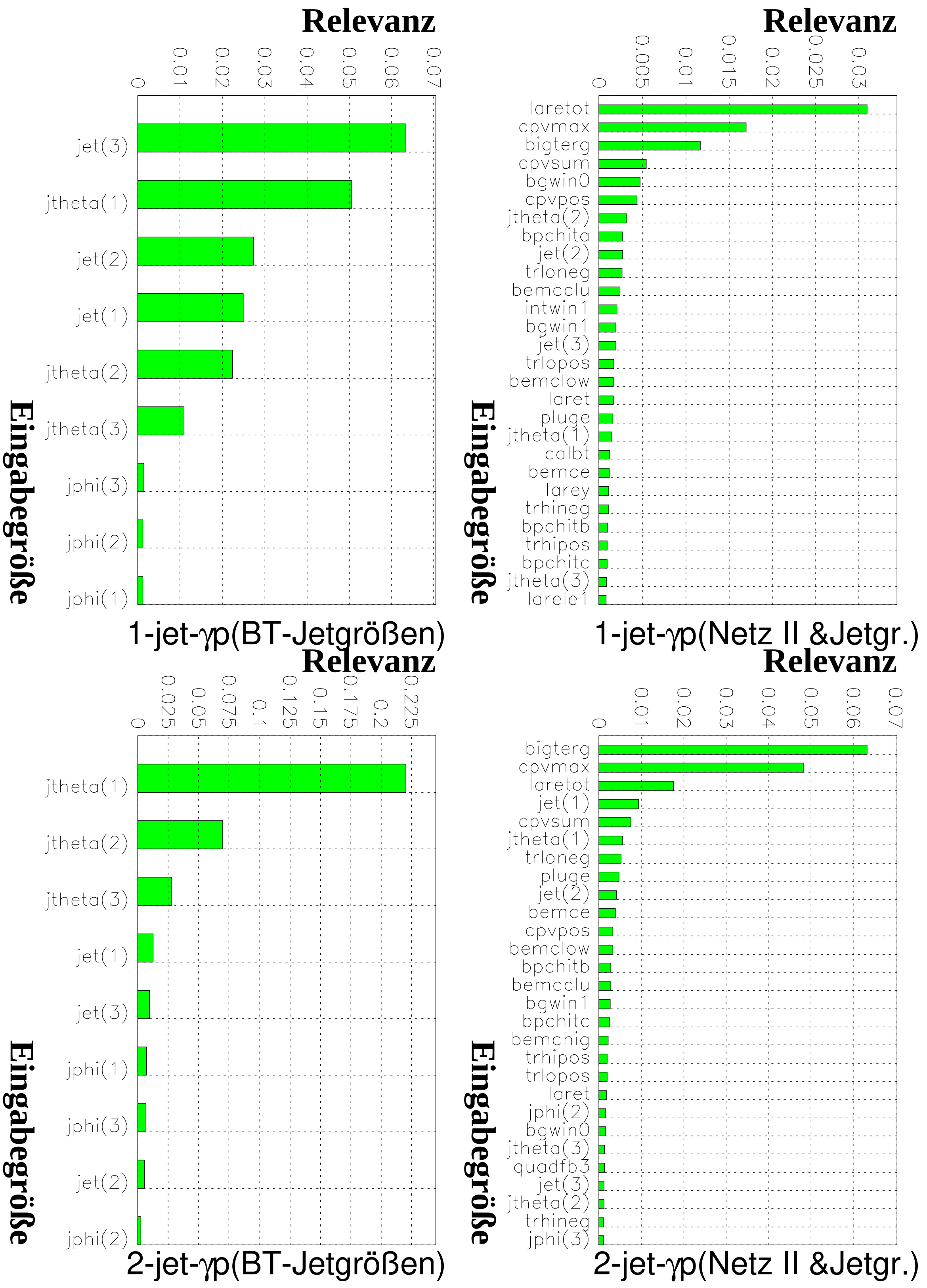

Abbildung 7.4: **Relevanz-Diagramme für die Photoproduktions-Jetklassen.** *Für die Konfiguration* Netz **II**&Jet *sind nur die 28 Eingabegrößen mit der größten Relevanz dargestellt.*

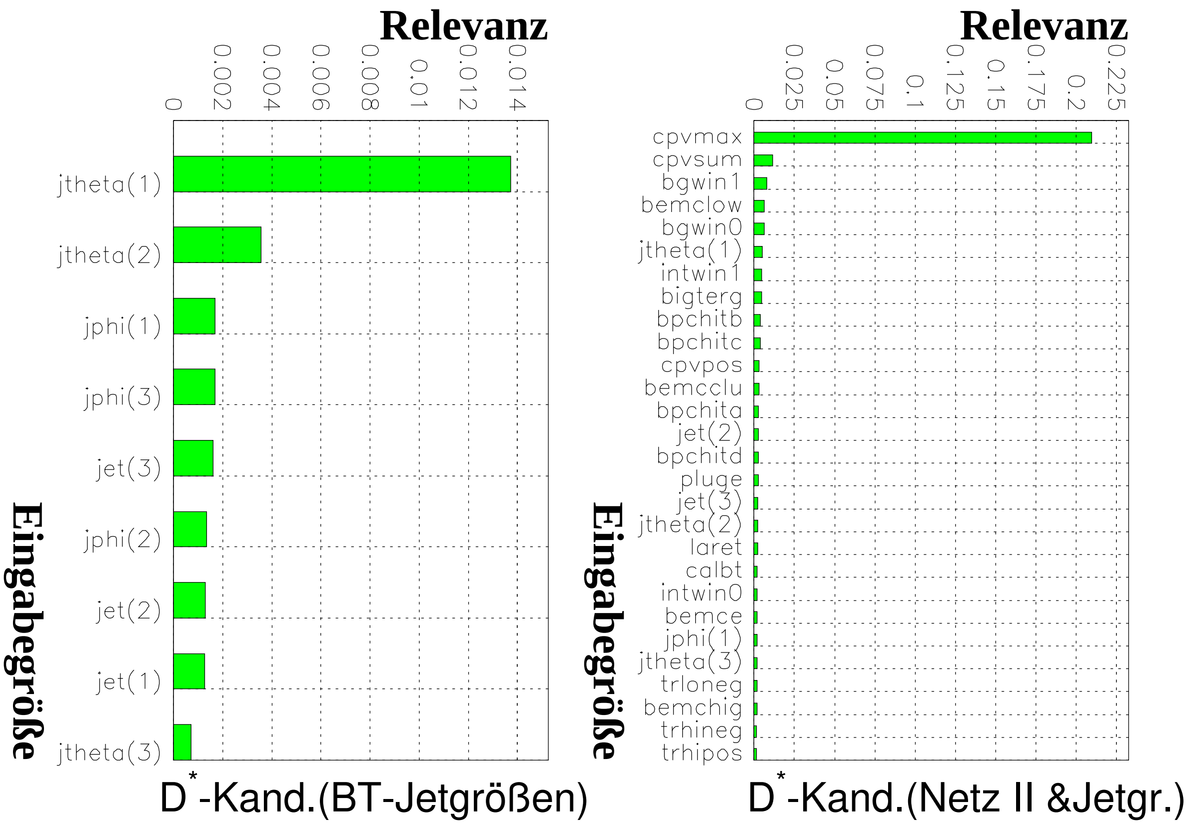

Abbildung 7.5: **Relevanz-Diagramme für die Klasse der D^*-Kandidaten.**

Jetgrößen nur die Winkelindizes, insbesondere der Θ-Index des energiereichsten BT-Jets, verwendet werden, während die Energien selbst keine Rolle spielen. Das entspricht wegen der geringen Energiedeposition im Flüssigargonkalorimeter auch den Erwartungen. Die größte Relevanz für das Netz **II** weist der maximale Bin-Eintrag des CIP-COP-z-Vertex-Histogramms auf. Die topologischen Energiesummen des Flüssigargonkalorimeters haben für diese Klasse keine Bedeutung. Die Klasse der K^0-Kandidaten ist nicht abgebildet, da sich die Diagramme wenig von denen der D^*-Kandidaten unterscheiden.

7.3 Untersuchungen zur Netzarchitektur

Die Relevanz-Diagramme sind zwar sehr hilfreich, um zu erkennen, welche Eingabegrößen überwiegend zur Klassifikation benutzt werden, aber die quantitative Aussagekraft ist zu gering, um entscheiden zu können, ob eine Netzeingabegröße ohne Schaden für die Effizienz weggelassen werden kann. Der naheliegende Versuch, systematisch Netze unter Auslassung jeweils einer Größe zu trainieren, ist problematisch, da z. B. Redundanzen zwischen verschiedenen Größen nicht erkannt werden, weil in diesem Fall die Variablen gegeneinander austauschbar sind. Da die Zahl an möglichen Kombinationen von Eingangsgrößen bei 60-70 gegebenen Variablen astronomisch hoch ist, können aber auch nicht alle getestet werden.

Aus diesem Grunde wurde die Reihenfolge der Eingangsvariablen in den Relevanz-Diagrammen des vorigen Abschnitts dazu benutzt, die Dimension des Eingangsvektors systematisch um jeweils eine Größe zu erweitern, beginnend mit einem einzigen Knoten. Die Zahl der versteckten Knoten wurde mit 64 sehr hoch gewählt, um dem Netz auch bei sehr wenigen Eingabeknoten die Abbildung komplizierter Korrelationen zu ermöglichen.

Dieses Verfahren wurde angewandt auf die Klassifikation der $NC(L5)_{Q^2>100\,GeV^2}$-vorselektierten Daten und der D^*-Kandidaten. Die 28 Eingabegrößen des Netzes **I**

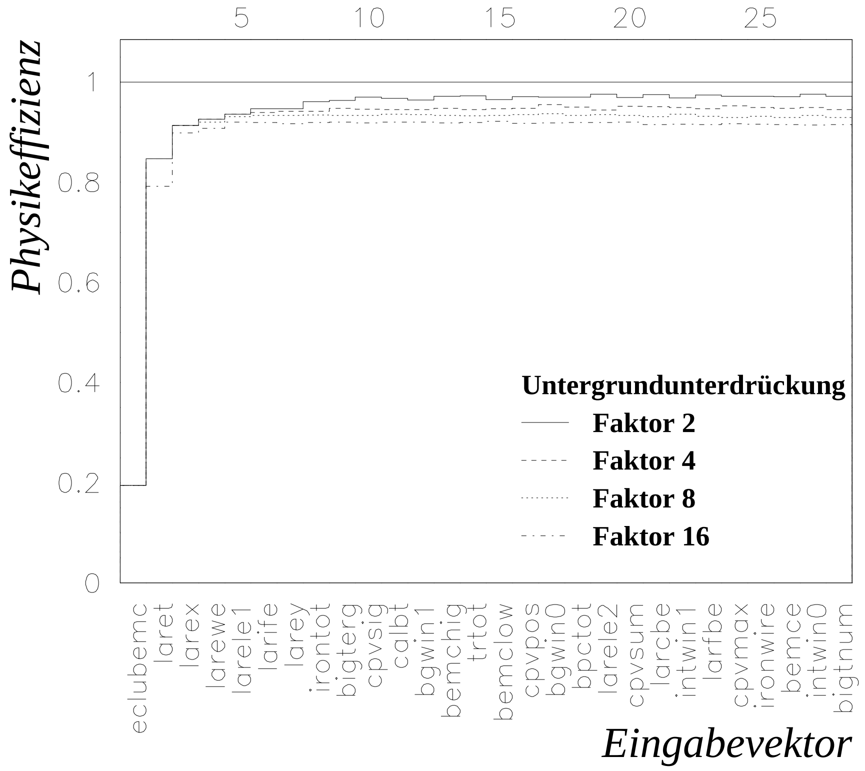

Abbildung 7.6: **Physikeffizienz aufgetragen bei festgehaltener Untergrundeffizienz für systematisch erweiterte Netze trainiert mit $\mathbf{NC(L5)}_{Q^2>100\,GeV^2}$.** *Die verwendeten Eingangsvariablen eines Netzes sind diejenigen, die bis zur angegeben Anzahl von Eingabevariablen in der Achsenbeschriftung aufgeführt sind.*

definieren den Eingabevektor mit der maximalen Dimension. Um die Netze miteinander vergleichen zu können, wurde die Physikeffizienz bei einer festgelegten Untergrundeffizienz für die jeweiligen Testmengen ermittelt. Diese Untergrundeffizienz ist in den Diagrammen durch ihre Inverse, den Reduktionsfaktor, ausgedrückt.

Für die $NC(L5)_{Q^2>100\,GeV^2}$-Klasse in Abb. 7.6 wurde die erste Größe, die Anzahl der triggernden Bigtower(bigtnum), an die letzte Stelle gesetzt, da zum Zeitpunkt der Analyse nicht klar war, ob diese Größe überhaupt zur Verfügung stehen wird. Da der Beitrag der BEMC-Clusterenergie(bemcclu) nur für eine Untermenge dieser Physikklasse von Bedeutung ist, erstaunt die geringe Effizienz mit einem Knoten nicht.

Im weiteren Verlauf tritt aber für alle Netze mit mehr als zehn Eingabeknoten eine Nivellierung der Physikeffizienz für alle aufgeführten Untergrundreduktionsfaktoren ein. Ein kleines Netz bestehend aus den topologischen LAr-Energiesummen, ergänzt um 3-4 andere Größen, ist folglich für diese Klasse zum Erreichen der besten Effizienz ausreichend. Der Vorteil kleiner Netze besteht unter anderem auch darin, daß Effizienz- und Ratenberechnungen verläßlicher sind. Desweiteren konnte gezeigt werden, daß auf die strittige Größe, die Anzahl der triggernden Bigtower(bigtnum) ohne merkliche Einbuße in der Effizienz verzichtet werden kann.
Für das zweite Beispiel in Abb. 7.7 wurden 28 Netze mit D^*-Kandidaten gegen den

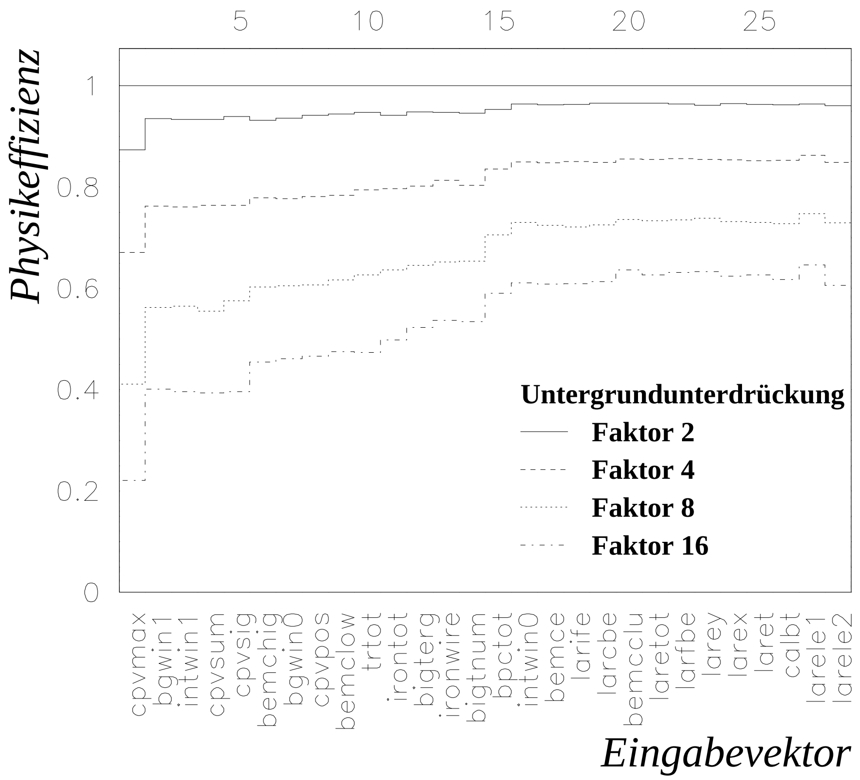

Abbildung 7.7: **Physikeffizienz aufgetragen bei festgehaltener Untergrundeffizienz für systematisch erweiterte Netze trainiert mit D^*-Kandidaten.** *Die verwendeten Eingangsvariablen eines Netzes sind diejenigen, die bis zur angegeben Anzahl von Eingabevariablen in der Achsenbeschriftung aufgeführt sind.*

Untergrund trainiert. Der Wert der Spurgrößen für diese Klasse ist deutlich zu erkennen, bis zur 20. Größe ergibt sich allerdings für fast alle hinzugefügten Größen eine Steigerung in der Effizienz. Erst für die topologischen Energiesummen des Flüssigargonkalorimeters am Ende der Liste ergibt sich kein weiterer Anstieg der Effizienz mehr. Für die D^*-Kandidaten werden also mehr Eingabegrößen benötigt als im ersten Fall, um zur besten Effizienz zu gelangen. Im übrigen ist der deutliche Effizienzverlust bei höheren Untergrundunterdrückungsfaktoren zu erkennen.

7.4 Ergänzung durch einen weiteren Algorithmus

Um neuronale Netze als Trigger verwenden zu können, wird sowohl für die Physik– als auch für die Untergrundereignisse eine Stichprobe mit ausreichender Statistik benötigt. Für eine besondere Klasse – unbekannte Physik – ist diese Situation weder in den Realdaten noch in den MC-Daten gegeben. Da es sich hier mit Abstand um die interessanteste Klasse handelt, wurde das neuronale L2-Triggerkonzept um einen weiteren Algorithmus ergänzt, der sich nur auf Untergrunddaten stützt. Eine Beschreibung dieses Anfang 1993 von P. Ribarics entwickelten Algorithmus findet sich in [54]. Auf Basis eines Clusteralgorithmus wird versucht, den Untergrund in ein möglichst kleines Volumen[1] im Raum der Eingabegrößen einzuschließen(→ *Background Encapsulation*). Dies geschieht mittels Quadern, die entlang den Hauptachsen der Kovarianzmatrix *eines* Clusters orientiert sind. Die Lage der Hyperebenen, die den Quader im N-dimensionalen Raum der Eingabegrößen definieren, wird über ein σ-Kriterium bestimmt. Die Quader lassen sich leicht in einer (neuronalen) Netzarchitektur implementieren, indem eine Intervallfunktion

$$f(x) = \begin{cases} 1 & \forall\, x \ \in]-1,1[\\ 0 & \text{sonst} \end{cases}$$

als Aktivierungsfunktion g verwendet wird. Für jede Dimension wird ein versteckter Knoten zur Definition zweier paralleler Hyperebenen benötigt. Die Architektur des Netzes ist demzufolge $N \times N \times 1$. Die mit diesem Algorithmus erzielten Resultate sind erstaunlich gut. Bei geschickter Definition dessen, was als Untergrundereignis zu bezeichnen ist, waren die erzielten Effizienzen für die L5-Vorselektionsklassen bei ähnlichen Ratenreduktionen vergleichbar mit denen der neuronalen Netze. Im Falle der kritischen Physikklassen ergaben sich Vorteile für die neuronalen Netze, hier ist auch der Überlapp mit dem Untergrund größer. Da dieser Algorithmus a priori keine Annahmen über die Physik macht, hängen die Effizienzen empfindlich davon ab, was als Untergrund definiert wird. Gute Resultate wurden insbesondere erzielt mit der auch in dieser Arbeit benutzten Eingrenzung des Untergrunds durch die Bedingung ‚L4-Reject&L5-Untergrundklasse'. Durch die andere Philosophie, die hinter diesem Algorithmus steht, ergänzen sich die neuronalen Netze und der „Background Encapsulation"–Algorithmus mehr, als daß sie in Konkurrenz zueinander stehen.

[1] Da Triggervariable willkürliche Einheiten sind, bezieht sich dieses Wort auf normalisierte Daten. Die Metrik ist durch die Inverse der Kovarianzmatrix festgelegt.

Kapitel 8

Die L2-Hardware-Architektur

8.1 Das Gesamtdesign des neuronalen L2-Triggers

Das breite Spektrum an verschiedenen Physikklassen mit sehr unterschiedlichen Signaturen macht es sinnvoll, das L2-System zu modularisieren. Ein speziell auf eine bestimmte Physikklasse trainiertes Netz entspricht einem solchen Modul. Alternativ kann aber auch ein konstruiertes Netz wie für den *„Background Encapsulation"*-Algorithmus in einem Modul implementiert werden. Der Wert des Ausgabeneurons jedes Moduls wird diskriminiert, und das Resultat der verschiedenen Module durch ein logisches Oder verknüpft. Das Hinzufügen eines weiteren Moduls hat somit keinen Einfluß auf die Physikeffizienzen der anderen Module, sondern kann lediglich die Gesamttriggerrate erhöhen. Das erlaubt, sehr flexibel auf neue Anforderungen seitens der H1-Physikgruppen zu reagieren, ohne bestehende Systemteile wegen sich störender Interferenzen verändern zu müssen.
Das Schema des Gesamtsystems ist in Abb. 8.1 und der Aufbau eines Moduls in Abb. 8.2 wiedergegeben. Jedes Modul besteht aus zwei Teilen. Einem *Data Distribution Board* (DDB), das aus dem Datenstrom des L2-Busses die Netzeingabegrößen ermittelt, und einer CNAPS®/VME-Karte, auf der das neuronale Netz realisiert ist. Diese Karte gibt das Resultat an das DDB zurück, das diese Entscheidung an den *Final Decider* weiterleitet. Ein vom neuronalen Netzkonzept unabhängiger topologischer L2-Trigger sendet seine Entscheidung über den VMVBus ebenfalls an den *Final Decider*. Der VMVBus dient auch der Übermittlung des Endergebnisses an die zentrale Triggerlogik von H1, die entsprechend dem Ergebnis entweder die L1-Pipeline freigibt, oder die Datenauslese einleitet.

8.2 Das Data Distribution Board

Aufgabe des *Data Distribution Boards* (DDB oder DataBox) ist es, innerhalb weniger Mikrosekunden die Netzeingabegrößen aus dem L2-Datenstrom bereitzustellen. Da jedes Netzwerk verschiedene Eingabegrößen entsprechend der Signatur der jeweiligen Physikklasse benötigt, muß eine flexible Programmierung der DDB möglich sein. Das DDB befindet sich zur Zeit(Feb. 94) in der Designphase. Ein Prototyp soll bis Som-

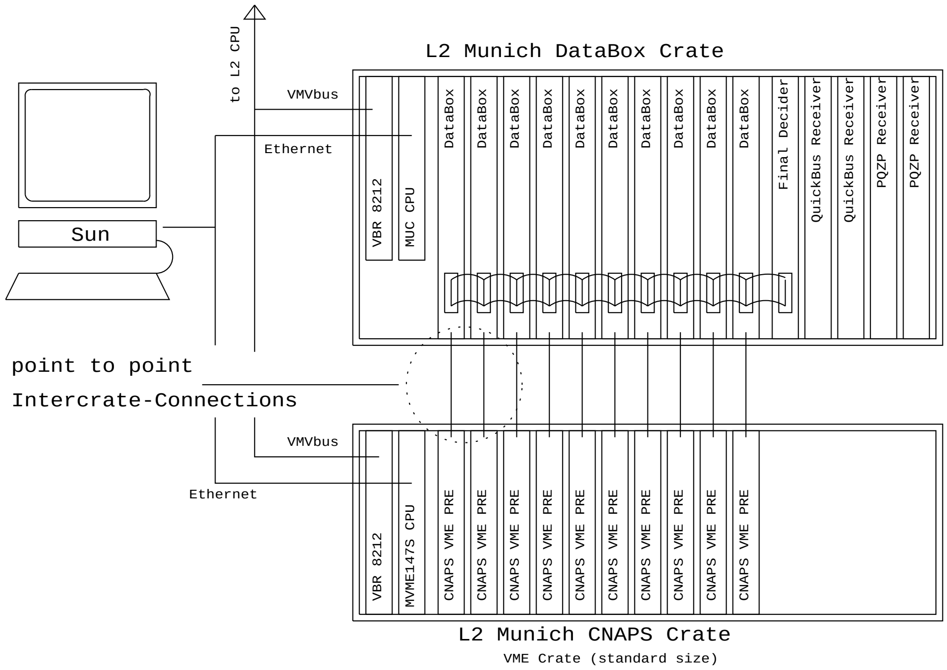

Abbildung 8.1: **Schema der neuronalen L2-Triggerumgebung.** *(vorläufiges Design).*

mer/Herbst 94 verfügbar sein. Dementsprechend dienen die Ausführungen dieses Abschnitts mehr der Orientierung, als daß sie feststehende Tatsachen beschreiben.
Die L2-Datenstruktur macht es erforderlich, einige elementare Rechenoperationen mit den angelieferten Datenwörtern durchzuführen. Die 128 Bits des L2-Busses sind in 8×16-Bit-Wörtern unterteilt, die in unterschiedlichen Formaten angeliefert werden. Neben den in Tabelle 8.1 aufgeführten Grundtypen werden auch subdetektorspezifisch

L2-Wort (16bit)				**Operation DDB**
Wert		Adresse		spezielle Algorithmen
Wert		Wert		Σ, direkt oder f(Wert)
Wert	Wert	Wert	Wert	$\sum_{1+2}, \sum_{3+4}$, dir. oder f(Wert)
Bitmuster				$\sum_{Bits}$
Bitmuster				Selektion von Bits

Tabelle 8.1: **L2-Datenformate und notwendige DDB-Operationen.** *„$\sum$“ bedeutet Summierung der Werte, „*direkt*“ heißt direkte Weiterleitung an das neuronale Netz und „*f(Wert)*“ Berechnung einer beliebigen Funktion des Wertes.*

einige kompliziertere Verarbeitungsschritte realisiert. Die Selektion der Daten geschieht über XILINX (frei programmierbare *logic cell arrays*), und die Durchführung der Operationen wird überwiegend mittels Lookup-Tables und Addierbäumen verwirklicht. Dieses flexible Konzept gestattet u. a. die Bildung signifikanter 8-*Bit*–Eingangsgrößen

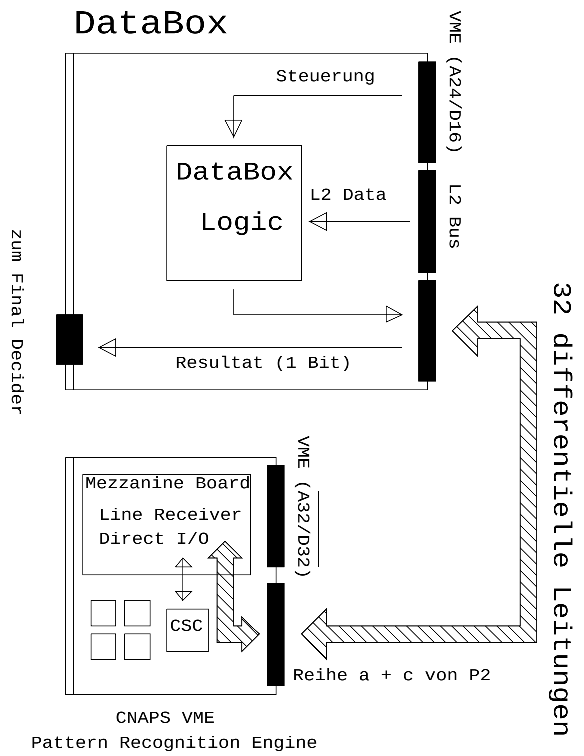

Abbildung 8.2: **Aufbau eines neuronalen L2-Modules.** *(vorläufiges Design).*

aus verkodierten größeren Bitmustern.
Darüberhinaus ist geplant, aus der detaillierten topologischen Kalorimeterinformation neue Netzeingabegrößen mit Hilfe spezieller Algorithmen zu berechnen. Im einzelnen sind vorgesehen:

- Zusammenfassung der maximal 512 Bigtower zu 64 *Supertowern* (mit möglicher Separation in elektromagnetischen und hadronischen Anteil)
- Bildung von Energiehistogrammen mit vom Maximum abhängiger Energieskalierung (Maximum u. Skalierung können ebenfalls als Netzeingangsvariable dienen)
- Bildung eines weiteren Histogramms für solche Bigtower, die mit einem Triggersignal der Proportionalkammern in Koinzidenz stehen (Bigrays)
- Ein *Jet-finder* Algorithmus, der die drei Bigtower-Jets mit den höchsten Energien ermittelt [55]. Als Netzeingabegröße werden neben der Jet-Energie auch die Winkel Θ und Φ des Jets zur Verfügung gestellt

Für jedes Modul bestimmt ein programmierbares *„Selection RAM“* , welche der Netzeingabegrößen für ein einzelnes Modul in Register zwischengespeichert werden. Diese Register werden seriell ausgelesen und die Daten über eine direkte I/O-Leitung an die CNAPS/VME-Karte weitergeleitet.

Abbildung 8.3: **Innerer Aufbau eines CNAPS-Prozesssorknotens.**

8.3 Die CNAPS/VME-Karte

CNAPS® (*Connected Network of Adaptive ProcessorS*) ist ein von *Adaptive Solutions, Inc.* entwickeltes Hardwarekonzept in SIMD-Technologie (*Single Instruction Multiple Data).* Eine Matrix mit 64 bis 256 Prozessorknoten (PN) wird von einem Sequencer (CSC; *CNAPS Sequencer Chip*) über einen gemeinsamen Bus (PNCMD) mit Instruktionen gespeist. Das Innenleben eines PN ist in 8.3 abgebildet. Jeder PN verfügt über einen lokalen Speicher von $4\,KByte$ und stellt einen kompletten Rechner für Festkomma-Arithmetik dar (ohne Befehlseinheit) [56].
Die Abbildung eines neuronalen Netzes auf die PN-Matrix ist intuitiv (Abb. 8.4). Die Gewichte eines Knotens zur darunterliegenden Schicht werden in den lokalen Speicher eines PN geschrieben; die Eingabedaten werden seriell über den IN-Bus(8-Bit) an die PN's weitergeleitet und dort mit den Gewichten multipliziert. Damit lassen sich die Ausgabewerte der einzelnen Knoten einer Schicht parallel berechnen. Diese Werte können über den OUT-Bus(8-Bit) für die Berechnungen in der nachfolgenden Schicht auf den IN-Bus umgeleitet werden. Der Rechenaufwand ist somit nur noch etwa proportional zur Anzahl der Knoten, anstatt zur Anzahl der Gewichtsparameter.
Die Taktfrequenz des CNAPS-Rechners beträgt $20\,MHz$. Innerhalb eines Taktzyklus läßt sich ein Multiplikations-Addier-Schritt komplett durchführen. Für die angestrebte maximale Netzgröße von $64 \times 64 \times 1$ werden somit mindestens 129 Taktzyklen benötigt. Das macht $6.5\,\mu s$ bei $20\,MHz$. Mit einem unvermeidlichen Überhang an Steuerbefehlen von ca. 50 Zyklen ergibt sich unter Verwendung der etwas höheren HeraClock-Rate von $20.8\,MHz$ [1] eine Zeitspanne von $8.9\,\mu s$.

[1] Diese Frequenz liegt geringfügig über der des eingebauten Quarzes. Die Verwendbarkeit der höheren Frequenz ist von Adaptive Solutions zugesichert worden.

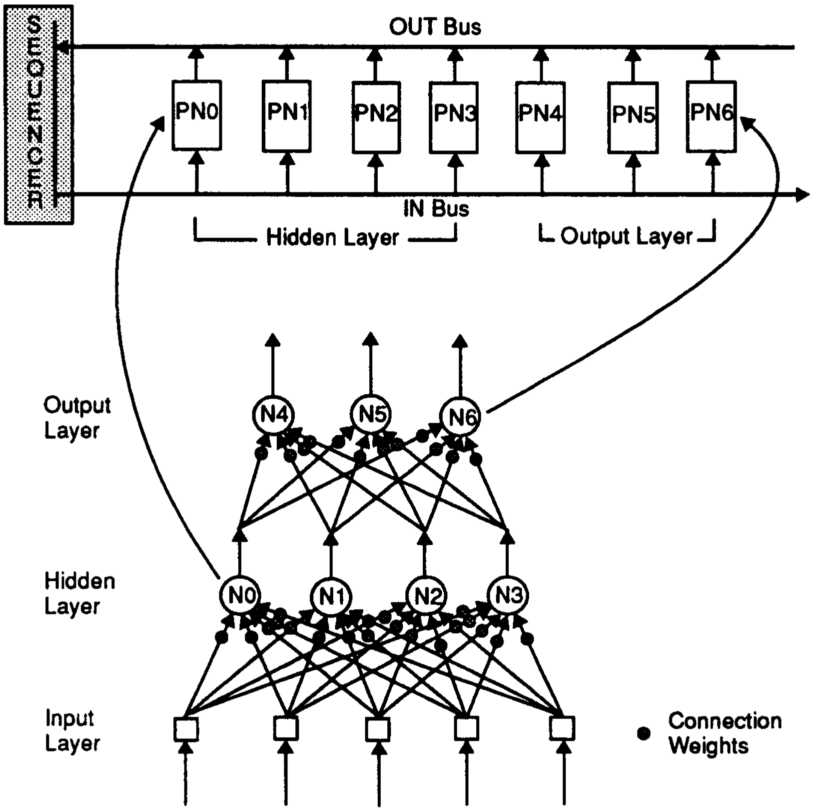

Abbildung 8.4: **Abbildung eines neuronalen Netzes auf die PN-Matrix.** *Die untere Figur kann auch so interpretiert werden, daß jeder senkrechte Pfeil einem Multiplikations-Addier-Taktzyklus des CNAPS-Rechners, und jeder diagonale Pfeil einem Multiplikations-Addier-Taktzyklus eines Seriellrechners entspricht.*

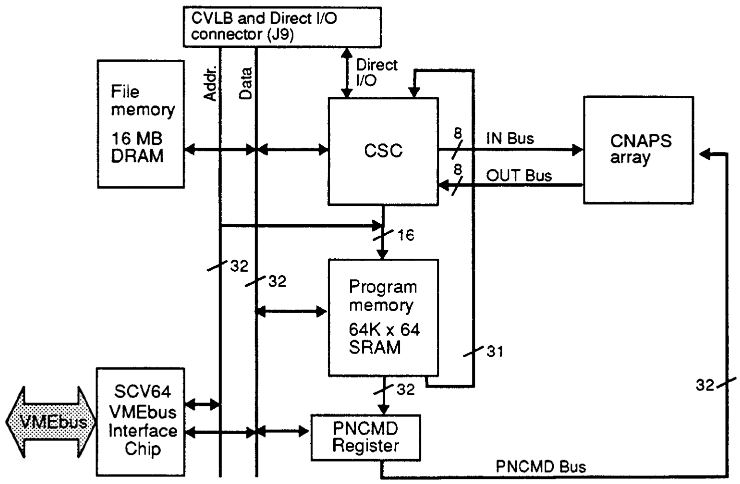

Abbildung 8.5: **Blockdiagramm der CNAPS/VME-Karte.** *Das Interface zum lokalen VME-Bus (CLVB) teilt sich den Ausgangsport J9 mit einem direkten I/O–Interface von zwei 8-Bit-Leitungen, in die Daten mit der HeraClock-Frequenz von* 20.8 *MHz eingespeist werden können. CSC ist die Abk. für den Sequencer.*

Für die Echtzeitanwendung eines L2-Triggers ist von entscheidender Bedeutung, daß CNAPS auch eine VME-Karte entwickelt hat, die es erlaubt, Daten über eine Direktleitung extern einzuspeisen (Abb. 8.5) [57]. Das direkte I/O–Interface mit zwei 8-Bit-Kanälen gestattet, den Feed-Forward-Schritt eines neuronalen Netzes mit Eintreffen des ersten Datenwortes auf dem IN-Bus zu starten. Der diskontinuierliche L2-Datenstrom läßt sich auf diese Weise bearbeiten.
Es wurde ein Zeittest für ein $64 \times 64 \times 1$-Netz mit Einspeisung über das I/O–Interface durchgeführt. Dieser Test bestätigte erfolgreich die errechnete Zeitspanne von $8.9\,\mu s$.

Kapitel 9

Zusammenfassung

Diese Arbeit befaßt sich mit dem zukünftigen Einsatz neuronaler Netze als L2-Trigger im H1-Experiment. Die Aufgabe der Triggerstufe L2 ist es, innerhalb von $20\mu s$ ein Ereignis als Untergrund oder Physikkandidat zu klassifizieren.
Es wurde dargestellt, wie sich ein solches Problem in einer parallelen Netzarchitektur implementieren läßt und ein entsprechender Optimierungsalgorithmus erläutert. Der wesentliche Schritt dieses Algorithmus ist das *Training*. Dabei werden dem Netz Muster einer statistischen Stichprobe präsentiert und ein Klassifizierungsfehler errechnet. Über ein Gradientenabstiegsverfahren werden damit die inneren Parameter des Netzes optimiert. In einem unabhängigen Test werden dem Netz unbekannte Eingabemuster präsentiert, und der für die Triggeranwendung entscheidende Generalisierungsfehler gemessen.
Im Hinblick auf die angestrebte Realisierungszeit von 1-2 Jahren waren die Untersuchungen so nah wie möglich an den experimentellen Gegebenheiten orientiert. Aus diesem Grunde wurden fast ausschließlich Triggerdaten von Ereignissen der Runperiode 1993 des H1-Experiments als Trainings- und Testdaten verwendet.
Als eine gefährliche Fehlerquelle wurde das Problem der Vergleichbarkeit von Untergrund- und Physikdaten erkannt. Die Vergleichbarkeit ist verletzt, wenn statistisch signifikante Merkmale der Stichprobenmenge nicht mit der Gesamtmenge übereinstimmen. Um auch bei einer größeren Anzahl von Eingabegrößen die Vergleichbarkeit zu garantieren, wurden für das Training Untergrund- und Physikdaten aus übereinstimmenden Zeiträumen verwendet, sowie Signifikanztests vor dem Training und Relevanztests nach dem Training durchgeführt. Im Rahmen dieser Arbeit wurde untersucht, inwiefern ein Parallelcomputer(CNAPS) mit eingeschränkter Rechengenauigkeit($8\,Bit$ Eingabe, $16\,Bit$ Parameter) unter Benutzung von Festkomma-Arithmetik für das Training geeignet ist. Es wurde eine leichte aber signifikante Verschlechterung der Trenneffizienzen festgestellt.
Es wurde ein Training mit verschiedenen von Physikgruppen in H1 benutzten Vorselektionsklassen durchgeführt. Es konnte gezeigt werden, daß für die Vorselektionsklassen für neutrale Stromereignisse auf L2 eine Reduktion der Triggerrate von $\mathcal{O}(10)$ erreicht werden kann. Für Impulsüberträge $Q^2 < 100\,GeV^2$ war die Physikeffizienz der Vorselektion bereits ausreichend ($> 99\%$), für $Q^2 > 100\,GeV^2$ ergibt sich unter Berücksichtigung technischer Probleme des L1-Triggers eine Physikeffizienz von $(98 - 99)\%$.

Die durch den Elektrontagger getriggerten Photoproduktionsereignisse wurden als unkritische Ereignisklasse identifiziert. Für die problematischen Physikklassen konnten für Photoproduktions–Jet–Ereignisse ebenfalls befriedigende Reduktionsraten im Bereich $10 - 15$ bei Effizienzen von $(90 - 95)\%$ erzielt werden. Als kritisch erwiesen sich die Kandidaten für die Erzeugung schwerer Quarks. Hier wurden nur Reduktionsraten im Bereich $5-10$ bei einer Effizienz von $(80-90)\%$ erreicht. Hinsichtlich der bestehenden Probleme, diese Ereignisse überhaupt effizient zu triggern, ist das Ergebnis aber positiv zu bewerten.
Um einen Einblick zu gewinnen, welche Eingabegrößen von einem Netz zur Klassifikation verwendet werden, wurde eine Größe definiert, die die Relevanz einer Eingabegröße für die Triggerentscheidung quantifiziert. Mit dieser Größe konnte gezeigt werden, daß die Netze jeweils die Eingabegrößen für die Klassifikation benutzen, die aus physikalischen Erwägungen heraus auch sinnvoll sind. Unter Ausnutzung der Ergebnisse dieser Untersuchung wurden Trainingsdurchläufe mit systematisch vergrößerten Eingabevektoren durchgeführt. Dabei konnte gezeigt werden, daß für die Vorselektion der neutralen Stromereignisse mit hohem Impulsübertrag die Effizienz für Netze mit mehr als 10 Eingabegrößen nicht mehr anstieg. Ein ähnliches Experiment mit den Kandidaten für schwere Quarks ergab eine kontinuierliche Verbesserung für die Hinzunahme weiterer Eingabegrößen mit Ausnahme der topologischen Energiesummen.
Es wurde das in der Designphase befindliche Hardware–Konzept dargestellt und gezeigt, daß die zeitliche Vorgabe, die L2-Entscheidung innerhalb von $20\mu s$ zu fällen, erfüllt werden kann.

Anhang A

Verwendete Datenfiles

Klasse	Filename	#	Runs
L4-Reject	HERA.H01POT2.L4RJT.C9301031-37	7145	58790-58973
	HERA.H01POT2.L4RJT.C9301130-43	10208	63983-64407
	HERA.H01POT4.L4RJT.C9305020-080(Ausw.)	9690	63286-64945
	HERA.H01POT4.L4RJT.C9305066-078(Ausw.)	6617	64571-64974
L4-Keep	HERA03.H1POT2.C9301586-1601(Ausw.)	9168	58790-58973
NC(L5)$_{Q^2<100GeV^2}$	F21HHK.H1DST1.CL11.HE4.Axx	7145	51017-60023
NC(L5)$_{Q^2>100GeV^2}$	F21LEV.H1DST1.C93CL9.A00..A03	4959	51018-61218
Analyseselektion	F21LEV.H1DST1.C93CL9.MINIDST6.FAST.A00	1739	51018-66609
CC(L5)	H1KHAP.H01.MDST1.V2.R057434..R059224	6072	57434-59224
γp(L5)(tagged)	H01ERD.GAMMAP.MINIDST12.RUN93.Axx	4380	51018-57508
γp-(2-Jet,Analyse)	H1KRIK.DST1.JET8.CL20.FINAL.A20/A21	1633	63125-65148
γp-(1-Jet,Analyse)	H1KBUN.JAN06.JETS.TR7879.AFTCLEAN.A00..A03	6612	62577-66609
Schwere Quarks	H1KGER.HFLMDST.Axx/Bxx		
D^*-Kandidaten	FPACK-Klasse 25	7538	63417-64133
D^*-Kandidaten	FPACK-Klasse 25	5156	64133-64877
K^0-Kandidaten	FPACK-Klasse 29	10803	63417-64133
K^0-Kandidaten	FPACK-Klasse 29	5395	64576-64901

Tabelle A.1: **Für das Training verwendete Dateien mit experimentellen H1-Daten der Runperiode 1993.** *Diese Dateien sind auf Bändern des DESY-IBM-Zentralrechners gespeichert. Auf die Selektionskriterien für die jeweiligen Dateien wird im Text eingegangen.*

Literaturverzeichnis

[1] E. Rutherford, Phil. Mag. Vol. **21**, (1911)p669.

[2] J. I. Friedman, H. W. Kendall, R. E. Taylor et al., Phys. Rev. Lett. **20**, (1968)p292-5.

[3] J. D. Bjorken, E. A. Paschos, Phys. Rev. **179** (1969) p1547.

[4] R. P. Feynman, Phys. Rev. Lett. **23**, (1969)p1415.

[5] M. Gell–Mann, Phys. Lett. **8**, (1964)p214.

[6] H1–Kollaboration, *Technical Proposal for the H1 Detector*, März 1986.

[7] Technical Proposal of the ZEUS Collaboration, *The ZEUS Detector*, März 1986.

[8] Physics at HERA, Proceedings of the HERA Workshop, Vol. 1, DESY, Hamburg, 29.-30. Oktober 1991, p19-p130.

[9] Physics at HERA, Proceedings of the HERA Workshop, Vol. 2, DESY, Hamburg, 29.-30. Oktober 1991.

[10] V.Brisson et al., *The measurement of electroweak parameters at HERA*, [9] op. cit., p947.

[11] [9] op. cit., p1029-1151.

[12] M. Erdmann, H. Kuester, S. Levonian, Y. Sirois and C. Vallee, *New Results from the H1 Experiment at HERA on Photoproduction, Deep Inelastic Scattering and Searches for New Particles*, Proceedings of XXVIIth Rencontres de Moriond les Arcs, France, March 1993, (DESY 93-077, Juni 1993).

[13] R. J. Ellison et al., *The H1 Trigger*, H1–Report, H1–11/90-157.

[14] J. Fent, C. Kiesling, P. Ribarics, *A Level 2 Calorimeter Trigger using Neural Networks*, H1–Report H1-04/91 - 172, 1991.

[15] J.C. Bizot et. al., *Hardware study for a topological level 2 trigger*, H1-report 09/92-212, DESY, Hamburg(1992).

[16] H1-Kollaboration, *The H1 Detector at HERA*, edd. P. Truöl, G. Cozzika et al., DESY-93-103, Juli 1993, wird in verkürzter Form in Nucl. Instr. Meth. publiziert.

[17] H1-Kalorimetergruppe, *The H1 Liquid Argon Calorimter System*, Nucl. Instr. Meth. A 336, (1993)p460-509.

[18] F. Jaquet und A. Blondel, Proceedings of the study of an ep facility for Europe 79/48, U. Amaldi ed., (1979)p391-394.

[19] V. N. Gribov und L. N. Lipatov, Sov. Journ. Nucl. Phys. **15**, (1972) p438 und p675.

[20] G. Altarelli und G. Parisi, Nucl. Phys. **126**, (1977)297.

[21] J. Bartels und J. Feltesse, *QCD at low x* , in [8] Vol. 1, (1991)p131.

[22] H1 Kollaboration, *First Measurement of the Charged Current Cross Section at HERA*, subm. Physics Letters B(1994).

[23] R. Rückl, Introduction to QCD, MPI-Seminar Sommer 93.

[24] D. Handschuh, *Studien zum Protoninduzierten Untergrund am H1 Detektor und Entwicklung von Filteralgorithmen zu dessen Unterdrückung*, Dissertation Universität Hamburg, (1993)p67.

[25] A. Ali, D. Wyler, [11] op. cit. Vol. 2, (1991)p669.

[26] A. Ali, *Heavy Quark Physics(...)*, Lectures given at 21^{st} International Meeting on Fundamental Physics: Physics at HERA, Madrid, Spanien, 9-15.Mai 1993, DESY-93-105, Juli 1993.

[27] J. Ellis et al., *Precision Electroweak Data, Supersymmetry and GUTS*, CERN-Preprint CERN-TH. 7116/93 (auch BARI-TH/164-93), Genf, Dez. 1993.

[28] [24] op. cit., Kapitel 4 und 5.

[29] R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley, 1991.

[30] H. Ritter et al., *Neuronale Netze*, Addison-Wesley, 1990.

[31] E. Schöneburg et al., *Neuronale Netzwerke:Einführung, Überblick und Anwendungsmöglichkeiten*, Markt & Technik-Verlag,1990.

[32] Gebrüder Grimm, *Aschenputtel*, Kinder- und Hausmärchen, Band I, Große Ausgabe, 7. Aufl., Göttingen, Dietrich'sche Buchhandlung, (1857) 21. Märchen.

[33] A. Gruber, *Neuronale Netzwerke als L2-Trigger im H1-Experiment*, Diplomarbeit LMU München, Dez. 1992.

[34] R. F. Thompson, *Das Gehirn*, Spektrum.d.Wiss.-Verlagsges. , Heidelberg 1990.

[35] J. W. Cohen, *Funktionelle Annatomie des Nervensystems*, 4. Aufl. , Schattauer, Stuttgart/New York 1985.

[36] J. C. Eccles, *The Understanding of the Brain*, 2. Aufl. , McGraw-Hill, New York 1977.

[37] W. S. McCulloch und W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bulletin of Math. Bio. **5**, (1943)p115-133.

[38] P. J. Werbos, *Beyond Regression*, Doctoral Dissertation, Appl. Math. , Harvard University, November 1974.

[39] D. E. Rumelhart und J. L. McClelland, Parallel Distributed Processing, Vol. 1 & 2, insbesondere **1**, p318-362, MIT Press Cambridge, 1986.

[40] M. Minsky, S. Papert, *Perceptrons*, MIT Press, Cambridge MA,1969.

[41] A. N. Kolmogorov, Dokl. Akad. Nauk. USSR, **114**, (1957)p953-956 (in Russisch).

[42] R. Hecht-Nielsen, Proc. of the Int. Conf. on Neural Networks,**II**, 19-32, IEEE Press, New York, June 1987.

[43] K. Hornik, M. Stinchcombe, H. White, *Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks*, Neural Networks, 3 (1990), pp551.

[44] R. Hecht-Nielsen, Proc. of the Int. Joint Conf. on Neural Networks, **I**, IEEE Press, New York, June 1989, p593-611.

[45] E. Baum, D. Haussler, *What size net gives valid generalization?*, Neural Computation **1**, (1989)p151-160.

[46] E. Barrelet et al. , H1-report 12/88-100, DESY Hamburg, 1988.

[47] H. Krehbiel, *The L2 Crates of H1*, H1 Trigger Note **1**, Hamburg 1993.

[48] T. Krämerkämper et al., *ANN as Level 2 Trigger for the H1 Experiment, Learning Strategies and Visualization*, Proceedings of the 3^{rd} International Workshop on Software Engeneering, Artificial Intelligence and Expert Systems in High Energy and Nuclear Physics, Oberammergau, 4.-8. Oktober 1993, wird veröffentlicht.

[49] A. Gruber, J. Möck et al., *Performance of Backpropagation Networks in the Second Level Trigger of the H1-Experiment*, Proceedings of the 3^{rd} International Workshop on Software Engeneering, Artificial Intelligence and Expert Systems in High Energy and Nuclear Physics, Oberammergau, 4.-8. Oktober 1993, wird veröffentlicht.

[50] L. Lönnblad, C. Peterson, T. Rögnvaldsson, *Pattern recognition in high energy physics with artificial neural networks - JETNET 2.0*, Computer Physics Communication **1**, 70(1992).

[51] R. R. Leighton, *The Aspirin/MIGRAINES Neural Network Software*, User's Manual V6.0, MITRE Corporation, Oct. 1992.

[52] A. Buniatian, persönliche Mitteilung.

[53] H. Rick, persönliche Mitteilung.

[54] P. Ribarics, *L2-L3 Background Rejector Based on the Encapsulation of Background Clusters*, H1-Report H1-08/93-310, Hamburg, 1993.

[55] Idee von D. Goldner, Uni Dortmund.

[56] CNAPS System Architecture Manual 3.0, Adaptive Solutions, Inc. ; Beaverton OR (1993).

[57] CNAPS/VME Board Reference Manual 1.0, Adaptive Solutions, Inc. ; Beaverton OR (1993).