Entwicklung neuronaler Netzwerke zur Erkennung von Ereignissen des geladenen Stromes bei H1

Diplomarbeit

am Fachbereich Physik der Universität Hamburg

vorgelegt von

Jürgen Rauschenberger

aus Toppenstedt

Hamburg

1998

Inhaltsverzeichnis

Kapitel 1

Einleitung

Will man sich nicht mit der phänomenologischen Beschreibung des Verhaltens der Materie, die uns umgibt und aus der wir bestehen, zufrieden geben, wird es unumgänglich ihre Struktur und die Wechselwirkungen in ihr näher zu analysieren. Die Zerlegung eines komplexen Systems in seine Bestandteile und deren Analyse hat nicht nur in der Physik, sondern in sämtlichen Naturwissenschaften zu einem tieferen Verständnis der Natur geführt.

Eine große Bedeutung bei der Analyse der Struktur der Materie haben die Streuexperimente. Grundlage dafür ist die Erkenntnis, daß ein Teilchen sich unter gewissen Bedingungen wie eine Welle verhält. Streut man zum Beispiel Elektronen an einem Doppelspalt, beobachtet man hinter diesem ein Interferenzmuster, welches nur durch die Beschreibung der Elektronen als Wellen erklärt werden kann. Aus der Analyse des Interferenzmusters ist es möglich die Struktur des Objektes an dem gestreut wurde zu rekonstruieren. So kann durch eine Fourieranalyse des Interferenzmusters der Doppelspalt rekonstruiert werden. Quantitativ wird diese *Welle-Teilchen-Dualität* durch die *de-Broglie-Wellenlänge* $\lambda = h/p$[1] eines Teilchens mit dem Impuls p beschrieben. Die Wellenlänge λ entspricht der maximalen Ortsauflösung, die bei einer Kollision von Teilchen mit einem Impuls p im Schwerpunktsystem erreichbar ist. Teilchen mit einem Impuls p werden an Strukturen gestreut, die mindestens die Größe $\lambda = h/p$ besitzen.

Durch die Streuung von α-Teilchen an einer Goldfolie konnte Ernest Rutherford in den Jahren 1911-13 die Existenz der Atomkerne nachweisen. 1956 zeigten die Experimente von Hofstadter, in denen Elektronen an Atomkernen gestreut wurden, daß es sich bei den Atomkernen um strukturierte Objekte handelt. Die Bestandteile der Atomkerne, die Nukleonen, wurden in den 60er Jahren am Stanford Linear Accelerator Center (SLAC) durch Elektron-Nukleon-Streuung näher analysiert. Es wurde gezeigt, daß die Nukleonen aus Partonen, die im Rahmen des Quark-Parton-Modells als Quarks identifiziert wurden, bestehen. Streuexperimente haben entscheidend zum Verständnis der schwachen Wechselwirkung beigetragen, so entdeckte man 1973 beim CERN durch die Streuung von Neutrinos an Elektronen und Nukleonen den Prozeß des neutralen Stroms.

---

[1] $h = 6{,}626 \cdot 10^{-34}$ Js ist die Planck-Konstante

Um die Struktur des Protons weiter zu analysieren wurde die Speicherringanlage HERA errichtet. Dort werden Elektronen bzw. Positronen an Protonen bei einer Schwerpunktsenergie von $\sqrt{s} = 300$ GeV gestreut. Mit HERA wird eine um eine Größenordnung höhere Auflösung des Protons erreicht (10^{-18} m) als bei Lepton-Nukleon-Streuexperimenten an einem stationären Target. Zwei neue kinematische Bereiche sind durch HERA zugänglich geworden:

- der Bereich sehr kleiner Impulsanteile der Partonen $x_{Bj} \simeq 5 \cdot 10^{-6}$ bei Viererimpulsüberträgen $Q^2 > 1{,}5$ GeV^2.
- der Bereich sehr großer Viererimpulsüberträge $Q^2 \simeq 10^5$ GeV^2.

Erstmals ist die Analyse der elektroschwachen Wechselwirkung in einem kinematischen Bereich möglich, in dem ein Einfluß der Massen der im Prozeß des neutralen- bzw. geladenen Stroms ausgetauschten Vektorbosonen Z^0 bzw. $W^\pm$ beobachtbar wird.

Eine grundlegende Voraussetzung zur Analyse der Prozesse des neutralen- und des geladenen Stroms ist eine effiziente Selektion der zugehörigen Ereignisse bei gleichzeitiger maximaler Untergrundreduktion.

Das Ziel der vorliegenden Arbeit ist die Entwicklung künstlicher neuronaler Netzwerke, welche in der Lage sind, Ereignisse des geladenen Stroms zu erkennen und von den Ereignissen der auftretenden Untergrundprozesse zu unterscheiden. Es soll untersucht werden, ob neuronale Netze bestehende Methoden der Selektion von Ereignissen des geladenen Stroms ersetzen oder ergänzen können. Den Schwerpunkt der Arbeit bildet die Entwicklung und der Test eines Netzwerkes zur Trennung der Ereignisse des geladenen Stroms von denen der Photoproduktion.

Im zweiten Kapitel werden die Begriffe der tiefunelastischen Positron-Proton-Streuung und der Photoproduktion erläutert. Es werden die theoretischen Vorhersagen der Wirkungsquerschnitte des neutralen- und des geladenen Stroms mit den am H1-Detektor erfolgten Messungen verglichen.

Das dritte Kapitel beinhaltet eine Beschreibung des HERA-Speicherings und die Darstellung der wichtigsten Komponenten des H1-Detektors.

Die Theorie der neuronalen Netzwerke wird im vierten Kapitel skizziert. Es beinhaltet die Darstellung der zum Training der Netze verwendeten Lernalgorithmen. Eine Einführung in die Grundlagen des zur Optimierung der Netzwerke verwendeten evolutionären Algorithmus wird gegeben.

Das fünfte Kapitel dient der Beschreibung der Signaturen von Ereignissen des geladenen Stroms im H1-Detektor und der Darstellung der einzelnen Untergrundprozesse. Die bisherigen Methoden der Untergrundreduktion werden dargestellt.

Im sechsten Kapitel wird die Entwicklung und der Test der Netzwerke dargestellt. Die verwendeten Programme werden erläutert und das Training der Netze wird beschrieben. Das Kapitel schließt mit dem Vergleich der Effizienzen der Netzwerke mit denen der bisherigen Methoden.

Kapitel 2

Theorie

2.1 Einleitung

Mit Hilfe des H1-Detektors am Speichering HERA wird die Struktur des Protons analysiert und mit den Vorhersagen des Standardmodells der Elementarteilchenphysik verglichen.
Das Standardmodell beschreibt die Wechselwirkung der elementaren Bausteine der Materie durch den Austausch von Feldquanten zwischen diesen. Dabei werden die Teilchen in zwei Klassen eingeteilt; die Leptonen (e,ν_e,μ,ν_μ,τ,ν_τ) und die Quarks (u,d,s,c,b,t), alle besitzen den Spin $\frac{1}{2}$ und sind somit Fermionen. Die Elementarteilchen wechselwirken über zwei [1] grundlegende Kräfte; die starke Wechselwirkung (QCD Quanten-Chromodynamik), mit den zugehörigen Feldquanten,den Gluonen,und die elektroschwache Wechselwirkung, mit den Eichbosonen γ, Z^0 und $W^{\pm}$.
Im Rahmen dieses Modells wird das Proton als ein aus Partonen bestehendes System betrachtet. Das einfachste Modell beschreibt das Proton als System von drei Valenzquarks, in der Konfiguration (uud). Im dynamischen Modell des Protons enthält dieses nicht nur die drei Valenzquarks, sondern zusätzlich Gluonen und weitere Quarks und Antiquarks, die sogenannten Seequarks.

Die bei HERA stattfindende Streuung von Positronen der Energie 27,5 GeV an Protonen der Energie 820 GeV, wird im Rahmen der Beschreibung durch theoretische Modelle in zwei kinematische Bereiche unterteilt:

- der Bereich tiefunelastischer Streuung (DIS, Deep Inelastic Scattering) mit großem Viererimpulsübertrag Q^2. In diesem Bereich wird die Wechselwirkung zwischen Positron und Proton durch den Austausch eines virtuellen punktförmigen Bosons zwischen einem der Partonen des Protons und dem Positron beschrieben. Je größer dabei der Viererimpulsübertrag Q^2 ist, um so größer ist die Ortsauflösung im

---

[1] Die Gravitation kann bei den in der Elementarteilchenphysik betrachteten kleinen Massen vernachlässigt werden

Proton. Der bei HERA erreichbare Viererimpulsübertrag von $Q^2 \simeq 10^5$ GeV^2 ermöglicht eine Auflösung der Strukturen des Protons mit 10^{-18} m.

- der Bereich der Photoproduktion mit $Q^2 \simeq 0$ GeV. Ein vom einlaufenden Positron abgestrahltes quasireelles Photon kann innerhalb der Wechselwirkungszeit in ein Vektormeson oder ein Quark-Antiquark Paar fluktuieren. Im Bereich der Photoproduktion wird die Streuung von Positronen an Protonen durch die Wechselwirkung der Partonen des Photons mit den Partonen des Protons beschrieben. Die Analyse der Ereignisse der Photoproduktion liefert dadurch Aufschlüsse über die Partonstruktur des Photons.

Die tiefunelastische Streuung wird im Abschnitt 2.2, die Photoproduktion im Abschnitt 2.3 behandelt.

2.2 Tiefunelastische Positron-Proton-Streuung

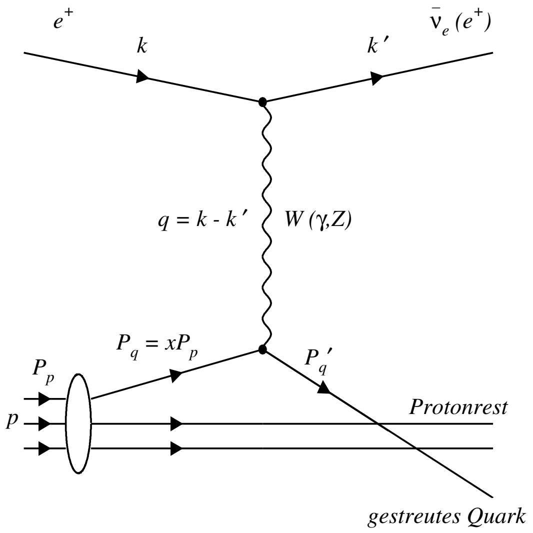


Abbildung 2.1: *Feynmandiagramm der tiefunelastischen Positron-Proton-Streuung in niedrigster Ordnung. In der Abbildung werden die Viererimpulse des ein- bzw. auslaufenden Leptons mit k bzw. k′ bezeichnet.* P_p *ist der Viererimpuls des einlaufenden Protons und* P_q *bzw.* P_q' *der des einlaufenden bzw. gestreuten Quarks. Der Viererimpulsübertrag wird mit q bezeichnet.*

Die tiefunelastische Positron-Proton-Streuung (DIS) kann in zwei Klassen unterteilt werden:
Handelt es sich bei dem gestreuten Lepton um ein Positron, spricht man vom Prozeß des neutralen Stroms (NC , neutral current)

$$e^+ + p \to e^+ + X,$$

ist das gestreute Lepton ein Antineutrino, wird der Prozeß als geladener Strom (CC , charged current) bezeichnet,

$$e^+ + p \to \bar{\nu}_e + X.$$

In Abbildung 2.1 ist das Feynmandiagramm der tiefunelastischen Positron-Proton--Streuung in Ordnung $\mathcal{O}(\alpha)$ der Kopplungskonstanten α der elektromagnetischen Wechselwirkung dargestellt.

Zwischen dem Positron und einem Parton des Protons wird beim neutralen Strom ein Photon γ oder ein neutrales Vektorboson Z^0 ausgetauscht, im Fall des geladenen Stroms handelt es sich bei dem ausgetauschten Teilchen um das geladene Vektorboson $W^\pm$.

Da beim Prozeß des neutralen Stroms das gestreute Positron im H1-Detektor nachgewiesen werden kann, das Neutrino im Fall des geladenen Stroms aber nicht nachweisbar ist, lassen sich die beiden Prozesse experimentell unterscheiden. Durch den Vergleich der gemessenen Wirkungsquerschnitte von neutralem- und geladenem Strom mit den theoretischen Vorhersagen, ist eine Überprfung der Theorie und eine Analyse der Struktur des Protons möglich. Zusätzlich ist die Bestimmung der freien elektroschwachen Parameter der Theorie α, G_μ und M_Z möglich (α ist die elektromagnetische Feinstrukturkonstante, G_μ ist die Zerfallskonstante des Myons und M_Z ist die Masse des Z^0-Bosons).

2.2.1 Kinematik

Zur Beschreibung des Streuprozesses der tiefunelastischen Positron-Proton Wechselwirkung, werden aus den Vierervektoren ($\vec{p}$,E) der ein- und auslaufenden Teilchen (vergl. Abbildung 2.1) folgende lorentzinvariante kinematische Größen berechnet:

$$Q^2 = -q^2 = -(k - k')^2 \tag{2.1}$$

$$y = \frac{P_p \cdot q}{P_p \cdot k} \tag{2.2}$$

$$x = \frac{Q^2}{2P_p \cdot q} \tag{2.3}$$

$$s = (k + P_p)^2 \tag{2.4}$$

Dabei ist q^2 das Quadrat des Viererimpulsübertrags am Lepton-Vertex, y der Bruchteil der Energie des einlaufenden Leptons welche auf den hadronischen Endzustand übertragen wird, x der Impulsanteil des Protons, der vom gestreuten Parton getragen wird (Bjorken Skalen Variable) und s die Schwerpunktsenergie des Positron-Proton Systems.

Zur Rekonstruktion der Kinematik des Streuprozesses aus den mit dem H1-Detektor gemessenen Impulsen und Energien existieren eine Reihe von Methoden. An dieser Stelle soll nur auf die Methode von Jacquet und Blondel eingegangen werden [Jac79]. Die Bestimmung der kinematischen Variablen basiert dabei auf den einzelnen Teilchen des hadronischen Endzustandes. Dieser besteht aus dem durch Fragmentation und Hadronisierung des gestreuten Partons entstehenden Stromjet und dem Protonrestjet [2]. Zusätzlich können weitere Jets durch Prozesse höherer Ordnung entstehen. So kann z.B. das gestreute Parton ein Gluon abstrahlen und dieses einen weiteren Jet durch Fragmentation und Hadronisierung erzeugen.

Da die Methode von Jacquet und Blondel eine Rekonstruktion der kinematischen Größen ohne das gestreute Lepton ermöglicht, ist sie beiden Ereignissen des geladenen Stroms in der Positron-Proton-Streuung bei HERA anwendbar. Die kinematischen Größen werden mit folgenden Formeln berechnet:

$$y_{JB} = \frac{\sum_i E_i(1-\cos\vartheta_i)}{2E_e} = \frac{E_{had} - p_{z,had}}{2E_e} \tag{2.5}$$

$$Q^2_{JB} = \frac{p^2_{\perp,had}}{1-y_{JB}} \tag{2.6}$$

$$x_{JB} = \frac{Q^2_{JB}}{s y_{JB}} \tag{2.7}$$

$$p_{\perp,had} = \sqrt{(\sum_i p_{x,i})^2 + (\sum_i p_{y,i})^2} \tag{2.8}$$

$$E_{\perp,had} = \sum_i \sqrt{p^2_{x,i} + p^2_{y,i}} \tag{2.9}$$

Der Viererimpuls eines Hadrons i sei $(\vec{p},E)_i$. Die Größe $E_{\perp,had}$ ist die skalare Summe der Transversalimpulse der Teilchen im hadronischen Endzustand eines Ereignisses. Im H1-Experiment kann nur ein geringer Anteil der Teilchen im Protonenrest nachgewiesen werden, da sie zum größten Teil im Strahlrohr entkommen. Mit der Jacquet-Blondel-Methode werden die kinematischen Variablen nur aus den Teilchen des hadronischen Endzustandes bestimmt. Diese Methode besitzt gleichzeitig eine geringe Sensitivität auf den Protonrest. Dies ist möglich, da die Teilchen, welche unter einem Polarwinkel $\vartheta_i \simeq 0$ gestreut werden, auch mit kleinem Gewicht $(1-\cos\vartheta_i)$ in die Bestimmung der Variablen eingehen.

---

[2] Eine Gruppe von Hadronen und Leptonen die sich im Rahmen einer Raumwinkelverteilung in eine gemeinsame Richtung bewegt wird als Jet bezeichnet.

2.2.2 Der neutrale Strom

Um mit der Theorie der elektroschwachen Wechselwirkung aus dem Wirkungsquerschnitt der Positron-Quark-Streuung den Wirkungsquerschnitt der Positron-Proton-Streuung in niedrigster Ordnung zu berechnen, beschreibt man das Proton im Rahmen des Quark-Parton Modells durch Strukturfunktionen, welche die Impulsverteilung der Quarks im Proton beschreiben.

Der Anteil des neutralen Stroms am differentiellen Wirkungsquerschnitt der tiefunelastischen Streuung von unpolarisierten Positronen an Protonen ist demnach [Spi97] :

$$\frac{d^2\sigma}{dxdQ^2}(e^+ + p \to e^+ + X) = \frac{2\pi\alpha^2}{xQ^4}\{F_2(1+(1-y)^2) + xF_3(1-(1-y)^2)\} \qquad (2.10)$$

Die Strukturfunktionen $F_2(x,Q^2)$,$F_3(x,Q^2)$ des Protons enthalten die Kopplungskonstanten und die Propagatorterme der elektroschwachen Wechselwirkung, sowie die Partonverteilungen:

$$F_2 = \sum_{f=u,d,...} A_f(Q^2)(q_f(x,Q^2) + q_{\bar f}(x,Q^2)) \qquad (2.11)$$

$$xF_3 = \sum_{f=u,d,...} B_f(Q^2)(q_f(x,Q^2) - q_{\bar f}(x,Q^2)) \qquad (2.12)$$

Dabei sind $q_f(x,Q^2)$,$q_{\bar f}(x,Q^2)$ die Verteilungen der Quarks und Antiquarks im Proton und $A_f(Q^2)$,$B_f(Q^2)$ die Terme der elektroschwachen Kopplung mit :

$$A_f = Q_f^2 - 2Q_f v_f v_e \chi_Z + (v_f^2 + a_f^2)(v_e^2 + a_e^2)\chi_Z^2 \qquad (2.13)$$

$$B_f = -2Q_f a_f a_e \chi_Z + 2v_f a_f 2v_e a_e \chi_Z^2 \qquad (2.14)$$

Die Ladung der Quarks f wird mit Q_f bezeichnet, die vektoriellen und axialen Kopplungskonstanten des Positrons v_e, a_e und der Quarks v_f, a_f an das Z^0-Boson sind folgendermaßen definiert:

$$v_{f,e} = I_3^{f,e} - 2Q_{f,e}^2 \sin^2(\theta_W), \quad a_{f,e} = I_3^{f,e} \text{ mit}$$

$$\sin^2(\theta_W) = 1 - \frac{M_W^2}{M_Z^2}. \qquad (2.15)$$

Dabei bezeichnet I_3 den schwachen Isospin des Teilchens und θ_W den Weinbergwinkel. Die experimentell bestimmte Masse des Z- bzw. W-Bosons ist M_Z = 91,2 GeV bzw. M_W = 80,4 GeV.

Da der Z-Propagator $\chi_Z(Q^2) \sim Q^2/(Q^2 + M_Z^2)$ ist, dominiert bei kleinem Q^2 der Photonaustausch. Erst bei Impulsüberträgen in der Größenordnung der Z^0-Masse M_Z erreicht der Z^0-Austausch die gleiche Größenordnung. QCD-Korrekturen höherer Ordnung (z.B. die Abstrahlung eines Gluons durch das gestreute Quark) werden in Monte-Carlo-Generatoren modelliert.

Bei H1 wurden aus den Daten der Jahre 1994/95 der differentielle und der totale Wirkungsquerschnitt des neutralen Stroms für Positron-Proton-Streuung gemessen [Mar97]. Der totale Wirkungsquerschnitt beträgt:

$$\sigma_{tot}^{NC}(p_{\perp} > 12{,}5\mathrm{GeV/c}\ ;\ 0{,}1 < y < 0{,}9) = (1289 \pm 24 \pm 95)\mathrm{pb}.$$

Der differentielle Wirkungsquerschnitt $d\sigma/dQ^2$ wird in der Abbildung 2.2 dargestellt.

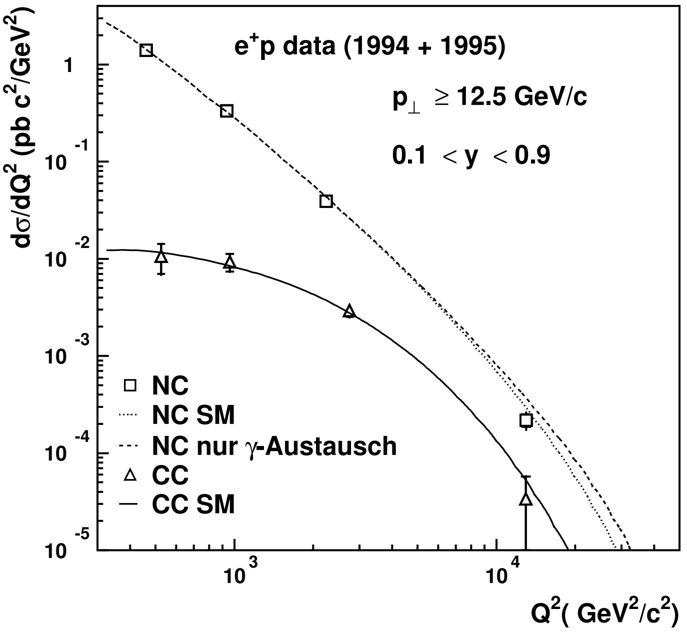


Abbildung 2.2: *Gezeigt ist ein Vergleich des gemessenen $d\sigma/dQ^2$-Wirkungsquerschnitts des geladenen Stroms und des neutralen Stroms mit der theoretischen Vorhersage des Generators HERACLES unter Verwendung der Partondichte MRSH [MRS93] für den geladenen Strom (durchgezogene Linie) und dem neutralen Strom (gepunktete Linie). Die unterbrochene Linie zeigt die theoretische Vorhersage für den Prozeß des neutralen Stroms unter der Annahme, daß nur virtuelle Photonen ausgetauscht würden. Aus [Mar97].*

2.2.3 Der geladene Strom

Beim Prozeß des geladenen Stroms handelt es sich um einen, im Gegensatz zum neutralen Strom, reinen Prozeß der schwachen Wechselwirkung.
Der Anteil des geladenen Stroms am differentiellen Wirkungsquerschnitt der Positron-Proton-Streuung, kann in niedrigster Ordnung im Rahmen des Quark-Parton Modells berechnet werden. Für unpolarisierte Positronen ist [Spi97]

$$\frac{d^2\sigma}{dxdQ^2}(e^+ + p \to \bar{\nu}_e + X) = \frac{G_\mu^2}{2\pi}\frac{M_W^4}{(Q^2+M_W^2)^2}\{q_{\bar{u}}(x,Q^2) + q_{\bar{c}}(x,Q^2) + \\ + (1-y)^2(q_d(x,Q^2) + q_s(x,Q^2))\} \tag{2.16}$$

Dabei ist G_μ die Fermi-Kopplungskonstante und M_W die Masse des W-Bosons. Der Einfluß von M_W im Propagatorterm auf den Wirkungsquerschnitt, der erst bei Viererimpulsüberträgen $Q^2 \simeq M_W$ meßbar wird, konnte erstmals bei HERA nachgewiesen werden [H194].

Die $(V-A)$-Kopplung der schwachen Wechselwirkung führt zu einer unterschiedlichen Abhängigkeit des Wirkungsquerschnitts von den Partondichten der Quarks und Antiquarks [Sch95]. Das W-Boson koppelt nur an linkshändige Teilchen und rechtshändige Antiteilchen. Aus der Drehimpulserhaltung folgt, daß das aus e^+q gebildete System den Drehimpuls eins (Spintriplett) und das $e^+\bar{q}$ System den Drehimpuls Null (Spinsingulett) trägt. Aufgrund dieser unterschiedlichen Kopplung ergibt sich die Abhängigkeit $(1-y)^2$ des Wirkungsquerschnitts von der Partondichte der Quarks, während keine y-Abhängigkeit von der Partondichte der Antiquarks besteht.

Analog der Vorgehensweise beim Prozeß des neutralen Stroms werden QCD-Korrekturen höherer Ordnung in Monte-Carlo-Generatoren modelliert. Durch die Korrekturen höherer Ordnungen werden z.B. die beobachteten $2+1$-Jet (zwei Jets und ein Protonrestjet) Ereignisse beschrieben.

Der totale Wirkungsquerschnitt des geladenen Stroms bei Positron-Proton-Streuung wurde bei H1 mit den Daten der Jahre 1994/95 bestimmt [Mar97]:

$$\sigma_{tot}^{CC}(p_\perp > 12{,}5\mathrm{GeV/c}\ ;\ 0{,}1 < y < 0{,}9) = (25{,}2 \pm 2{,}5 \pm 1{,}6)\mathrm{pb}.$$

Der Vergleich der totalen Wirkungsquerschnitte des neutralen- und geladenen Stroms zeigt, daß die Ereigniszahl N_{CC} des geladenen Stroms nur 1/50 der des neutralen Stroms beträgt.

Der in Abbildung 2.2 gezeigte Vergleich des gemessenen differentiellen Wirkungsquerschnitts $d\sigma/dQ^2$ mit der Vorhersage des Standardmodells zeigt, daß der Prozeß des geladenen Stroms durch den Austausch eines virtuellen $W^\pm$-Bosons der Masse M_W konsistent beschrieben wird.

2.3 Photoproduktion

Wird ein Positron, welches unter einem sehr kleinen Polarwinkel ϑ_e gestreut wurde im Endzustand eines Ereignisses beobachtet, wird dieses als ein Ereignis der Photoproduktion (γp-Ereignis) bezeichnet. In einem derartigen Ereignis beträgt der Viererimpulsübertrag $Q^2 \simeq 0$ GeV, denn

$$Q^2 = -(k - k')^2 = 4E_e E_{e'} \sin^2(\frac{\vartheta_e}{2}). \tag{2.17}$$

Im Prozeß der Photoproduktion strahlt das einlaufende Positron ein quasireelles Photon ab, welches mit dem Proton wechselwirkt

$$\gamma + p \rightarrow X.$$

Die Ereignisse der Photoproduktion können in zwei Klassen eingeteilt werden [Leb96] :

- Die direkten Prozesse (direct). Hier wechselwirkt das Photon direkt mit einem Parton des Protons, wobei die Fragmentation des Partons Teilchenbündel (Jets) mit hohen Transversalimpulsen erzeugt. Beispiele für direkte Prozesse sind die Boson-Gluon-Fusion und der QCD-Compton-Prozeß (Abbildung 2.3).

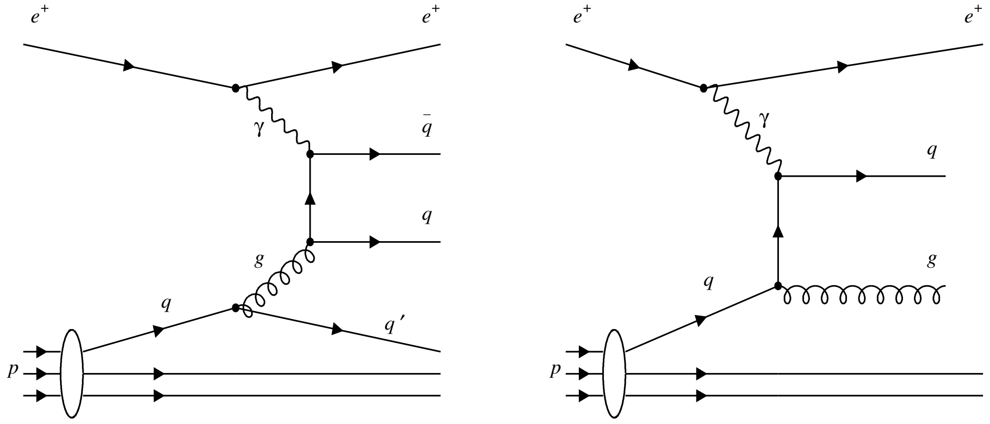


Abbildung 2.3: *Beispiele von Feynmandiagrammen der Boson-Gluon-Fusion (links) und des QCD-Compton-Prozeß*

- Die aufgelösten Prozesse (resolved). Es findet eine Wechselwirkung zwischen einem Parton des Photons und einem Parton des Protons statt. Durch die Fragmentation des Photonrestes entstehen im Vergleich zum direkten Prozeß weitere Teilchen. Ein Beispiel eines aufgelösten Prozeß ist in Abbildung 2.4 dargestellt.

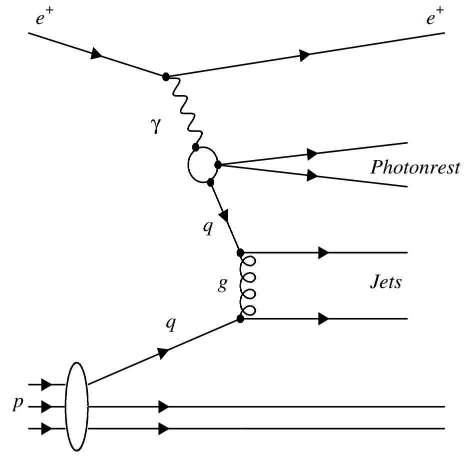


Abbildung 2.4: *Feynmandiagramm eines aufgelösten Prozesses*

Ereignisse der direkten Photoproduktion zeichnen sich in niedrigster Ordnung durch zwei Teilchenjets mit ausgeglichenem Transversalimpuls, die durch die Fragmentation der Quarks und Gluonen entstehen, aus. In Ereignissen der aufgelösten Photoproduktion entstehen durch die Fragmentation des Photonrestes weitere Teilchen. In höherer Ordnung führt die Abstrahlung von Gluonen durch die auslaufenden Partonen in den Ereignissen der direkten- oder aufgelösten Photoproduktion zur Erzeugung weiter Teilchen. Zusätzlich ist im Prozeß der aufgelösten Photoproduktion eine Wechselwirkung zwischen den Partonen des Photon- und des Protonrestes möglich durch die weitere Teilchen erzeugt werden.

In der Abbildung 2.5 ist der bei H1 gemessene differentielle Jet-Wirkungsquerschnitt $d\sigma/dE_{\perp}^{Jet}$ der Photoproduktion dargestellt. Es wurden Ereignisse mit mindestens einem Jet und einer transversalen Energie von $E_{\perp}^{Jet} > 7$ GeV sowie einem Viererimpulsübertrag von $Q^2 < 0{,}01\mathrm{GeV}^2$ ausgewählt.

Der Abfall des differentiellen Wirkungsquerschnittes $d\sigma/dE_{\perp}^{Jet}$ mit der transversalen Energie $E_{\perp}^{Jet}$ kann durch eine Funktion der Form $(E_{\perp}^{Jet})^{-n}$ beschrieben werden. Der Jet-Wirkungsquerschnitt fällt von $\sigma_{Jet} \simeq 10$ nb im Bereich 9 GeV $\leq E_{\perp}^{Jet} \leq$ 11 GeV auf $\sigma_{Jet} \simeq 0{,}2$ nb im Bereich 19 GeV $\leq E_{\perp}^{Jet} \leq$ 21 GeV ab. Bei kleinen transversalen Energien übersteigt der Jet-Wirkungsquerschnitt der Photoproduktion den totalen Wirkungsquerschnitt des geladenen Stoms von $\sigma_{tot}^{CC} \simeq 25$ pb bei weitem. Im folgenden wird sich zeigen, daß dieses Verhalten des γp-Wirkungsquerschnittes, bei kleiner transversaler Energie zu eienem Untergrund von γp-Ereignissen in der Menge der CC-Ereignisse führt.

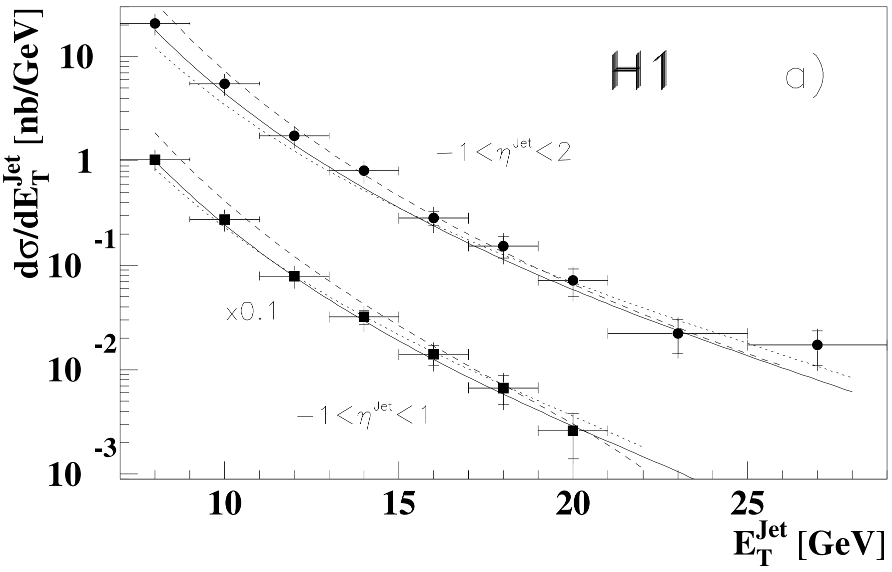


Abbildung 2.5: *Der bei H1 gemessene differentielle Jet Wirkungsquerschnitt der Photoproduktion. Die im Pseudo-Rapiditätsbereich von* $-1 < \eta^{Jet} < 2$ *gemessenen Werte sind als Kreise, die im Bereich von* $-1 < \eta^{Jet} < 1$ *gemessenen Werte als Quadrate dargestellt (um den Faktor 10 herabgesetzt). Die inneren Fehlerbalken bezeichnen den statistischen Fehler, die äußeren die quadratische Summe des statistischen und des systematischen Fehlers. Die eingezeichneten Linien entsprechen den Resultaten der Berechnungen des Wirkungsquerschnittes mit verschiedenen QCD Generatoren. Für die Generatoren PHOJET (durchgezogene Linie) und PYTHIA (unterbrochene Linie) unter Berücksichtigung der Wechselwirkung zwischen den Partonen des Photon- und des Protonrestes. Und für den Generator PYTHIA (gepunktete Linie) ohne die Berücksichtigung der Wechselwirkung zwischen Photon- und Protonrest. Aus [H195b].*

2.4 Ereignisgeneratoren

Zum Vergleich von Messungen der e^+p-Wechselwirkungen mit den Vorhersagen der theoretischen Modelle werden Ereignisgeneratoren verwendet. Mit Hilfe dieser *Monte-Carlo-Generatoren* können verschiedene physikalische Prozesse auf Vierervektorebene simuliert und die entstehenden Ereignisse konstruiert werden (vergl. Abschnitt 3.2.5) .
Eine Monte-Carlo-Simulation erfolgt schrittweise :

- Im ersten Schritt wird der elementare Streuprozeß basierend auf den theoretischen

Wirkungsquerschnitten und Parton-Dichte Verteilungen simuliert. Hierbei werden die Matrixelemente in Ordnung $\mathcal{O}(\alpha)$ und Strahlungskorrekturen berücksichtigt.

- Im zweiten Schritt werden QCD-Strahlungskorrekturen höherer Ordnung angebracht; zu deren Beschreibung stehen zwei unterschiedliche Modelle, das Parton-Schauer-Modell und das Farb-Dipol-Modell zur Verfügung [Hei96].
- Im dritten Schritt wird die Hadronisierung der Partonen durch phänomenologische Modelle (Lund-String- und Cluster-Fragmentierungs-Modell) simuliert.

Die Simulation von Ereignissen des neutralen und geladenen Stroms ist mit dem Programmpaket DJANGO 6 [Spi97] möglich. In diesem Programmpaket wird der elementare Streuprozeß unter Berücksichtigung der Strahlungskorrekturen durch den Ereignissgenerator HERACLES simuliert. Die QCD Effekte werden von ARIADNE, basierend auf dem Farb-Dipol-Modell, berücksichtigt.Der Algorithmus JETSET benutzt das Lund-String-Modell zur Beschreibung der Hadronisierung.
Ereignisse der direkten und der aufgelösten Photoproduktion können mit dem Programmpaket PYTHIA/JETSET [Sjo92] simuliert werden.

Kapitel 3

Das Experiment

3.1 Der HERA-Speicherring

Die 6,3 km lange Hadron-Elektron-Ring-Anlage (HERA) befindet sich beim Deutschen Elektronen Synchrotron (DESY) in Hamburg. Sie besteht aus zwei separaten Speicherringen für Protonen und Elektronen bzw. Positronen.

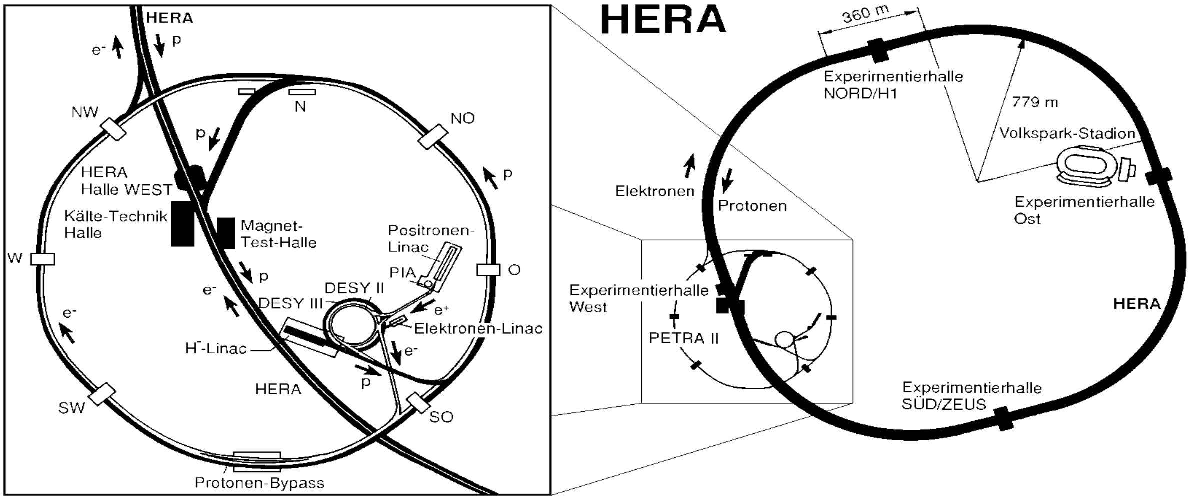


Abbildung 3.1: *Der Speicherring HERA und das Vorbeschleunigersystem*

In einem System aus Vorbeschleunigern werden die Teilchen zunächst auf Energien von 12 GeV (Positronen) bzw. 40 GeV (Protonen) vorbeschleunigt, bevor sie in den HERA-Ringen auf ihre Endenergien von 27,5 GeV bzw. 820 GeV beschleunigt und gespeichert werden (vgl. Abbildung 3.1). Damit ergibt sich eine Schwerpunktsenergie bei der Positron-Proton-Wechselwirkung von $\sqrt{s} \simeq 300$ GeV, so daß ein maximaler Viererimpulsübertrag Q^2 von ca. 10^5 GeV^2 erreichbar wird. Dieser Wert liegt zwei Größenordnungen über dem von Experimenten an ruhenden Nukleonen.

Um die Protonen auf der Sollbahn zu halten, ist der Speicherring mit supraleitenden Dipol- und Quadrupolmagneten ausgestattet, welche ein Magnetfeld von 4.68 T erzeugen. Die Positronen werden durch ein von normalleitenden Dipol- und Quadrupolmagneten erzeugtes Magnetfeld von 0.165 T auf ihrem Orbit gehalten. Positronen und Protonen werden in den Speicherringen in maximal je 220 Teilchenpaketen mit einem zeitlichen Abstand von 96 ns gespeichert. HERA wird zu Zeit mit 180 kollidierenden Teilchenpaketen betrieben. Die beiden Teilchenstrahlen werden in zwei Wechselwirkungszonen zur Kollision gebracht, an denen sich die Detektoren H1 (Halle Nord) und ZEUS (Halle Süd) befinden. Zusätzlich zu diesen Experimenten sind zwei Experimente mit stationärem Target am Speicherring HERA aufgebaut: HERMES untersucht unter Verwendung longitudinal polarisierter Positronen und polarisierter Protonen oder Neutronen eines internen Targets die Spinstruktur von Proton und Neutron. Noch im Aufbau befindlich ist das Experiment HERA-B, mit dem durch ein nahe an den Protonenstrahl herangeführtes Draht-Target B-Mesonen mit hoher Rate erzeugt werden und die CP-Verletzung beim Zerfall der B-Mesonen nachgewiesen werden soll.

3.2 Der H1-Detektor

Der Detektor der H1-Kollaboration wurde in der Experimentierhalle Nord des Speicherrings HERA aufgebaut. In Abbildung 3.2 ist ein Schema des Detektors dargestellt. Eine detaillierte Beschreibung der Detektorkomponenten findet man in [Abt96]. Hier soll nur ein kurzer Überblick über die wichtigsten Komponenten gegeben werden.

Der H1-Detektor umfaßt nahezu den vollständigen Raumwinkel um den Wechselwirkungsbereich und ermöglicht die Messung der Energie der Reaktionsteilchen bei einer guten Orts- und Impulsauflösung.

Das H1-Koordinatensystem ist ein rechtshändiges System mit dem Ursprung im Wechselwirkungspunkt. Die z-Achse entspricht der Bewegungsrichtung des einlaufenden Protons, die x-Achse zeigt zum Mittelpunkt des Speicherrings, die y-Achse weist nach oben. Der Bereich kleiner Polarwinkel θ wird als Vorwärtsrichtung bezeichnet.

Die einzelnen Detektorkomponenten sind zylindersymmetrisch um das Strahlrohr angeordnet, da man eine isotrope Verteilung der Ereignisse im Azimutwinkel ϕ erwartet. In Vorwärtsrichtung ist der Detektor aufwendiger instrumentiert, da sich das Ruhesystem der wechselwirkenden Teilchen aufgrund der höheren Energie der Protonen in Richtung des Protonenstrahls bewegt.

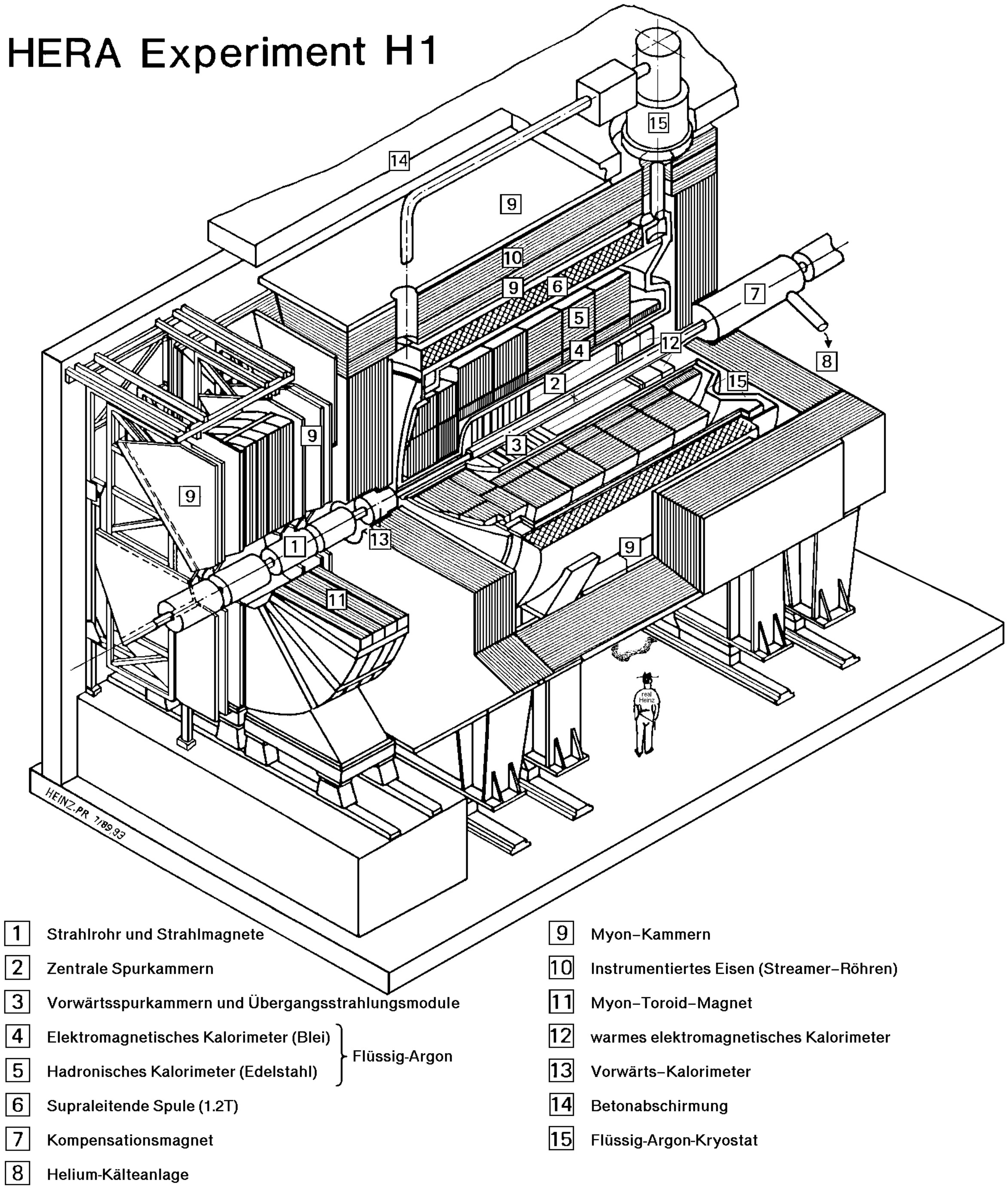


Abbildung 3.2: *Der H1-Detektor*

3.2.1 Das Spurkammersystem

Das gesamte Spurkammersystem (Abbildung 3.3) und das Flüssig-Argon-Kalorimeter des H1-Detektors befinden sich in einem Magnetfeld (1.2 T), das durch eine supraleitende Spule erzeugt wird. Die Spurkammern dienen zur Rekonstruktion der Spuren geladener Teilchen, zur Rekonstruktion des Ereignisvertex und als Spur-Trigger. Durch sie kann der Betrag und die Richtung des Impulses geladener Teilchen bestimmt werden. Das Spurkammersystem besteht aus Driftkammern zur präzisen Spurrekonstruktion und Impulsmessung geladener Teilchen sowie Proportionalkammern, die nur grobe Informationen über Spurverlauf und Ereignisvertex liefern, aber aufgrund ihrer höheren Zeitauflösung (~ 20ns) zu Triggerzwecken verwendet werden. Das Spurkammersystem setzt sich aus einem zentralen Spurdetektor, einem Vorwärtsspurdetektor und einem rückwärtigen Driftkammersystem zusammen.

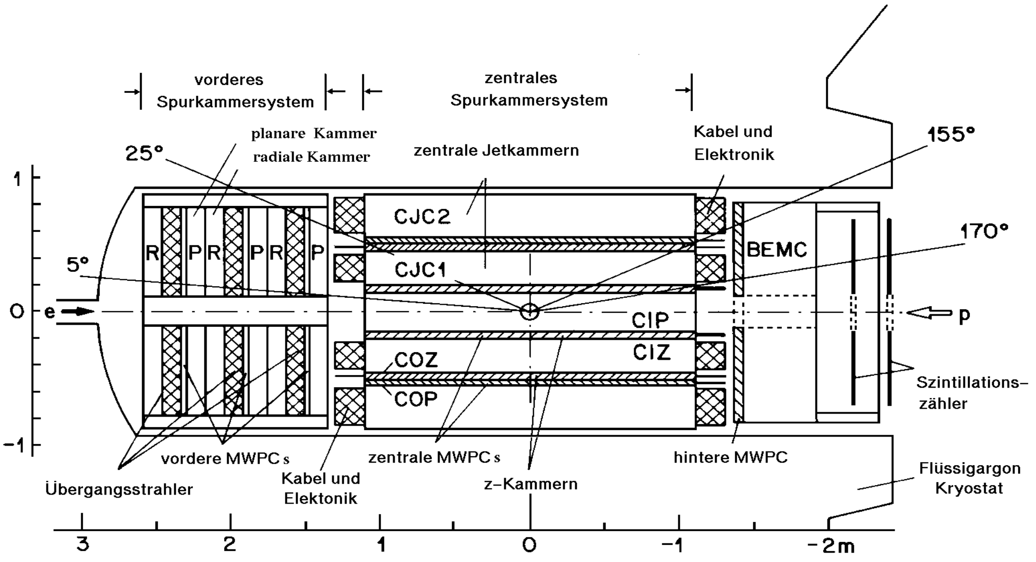


Abbildung 3.3: *Das Spurkammersystem des H1-Detektors*

Der zentrale Spurdetektor deckt einen Polarwinkelbereich von $25^\circ < \theta < 155^\circ$ ab. Mit ihm ist es möglich den Impulsbetrag und die Bewegungsrichtung von geladenen Teilchen mit einer Genauigkeit von $\sigma_p/p^2 \simeq 3 \times 10^{-3}\mathrm{GeV}^{-1}$ und $\sigma_\theta \simeq 1$ mrad zu bestimmen. Er setzt sich aus den zwei Jetkammern CJC1 und CJC2, zwei weiteren Driftkammern CIZ und COZ und den beiden Proportionalkammern CIP und COP zusammen.

Die Jetkammern CJC1, CJC2 bestehen aus 2640 parallel zur Strahlachse gespannten Drähten, welche in 30 bzw. 60 Driftzellen angeordnet sind. Geladene Teilchen ionisieren beim Durchqueren der Kammern das enthaltene Gas. Die Elektronen driften zu den Signaldrähten der Zellen und erzeugen in diesen Strompulse. Aus den gemessenen

Driftzeiten der Elektronen zu verschiedenen Signaldrähten der Kammer werden Spursegmente rekonstruiert. Entsprechen diese Parameter benachbarter Zellen einander, so werden sie zu einer Spur zusammengefaßt. Danach werden Verbindungen zwischen CJC1 und CJC2 und den Spuren im Vorwärtsspurdetektor gesucht. Mit den Jetkammern wird eine gute Auflösung $\sigma_{r\phi} = 170$ μm in der radialen Ebene erreicht, während die Genauigkeit in z gering ist, $\sigma_z = 22.0$ mm.
Eine höhere Präzision in z wird mit Hilfe der beiden z-Kammern CIZ und COZ erreicht, $\sigma_z \simeq 350$ μm.
Die führ Triggerzwecke verwendeten Proportionalkammern CIP und COP bestehen jeweils aus einer Doppelkammer mit je einer Drahtlage. Das Auslesen dieser Kammern wird mit rechteckigen Elektroden durchgeführt, wobei die Kammern in z-Streifen und φ-Segmente unterteilt ist.

Im Bereich $5° < \theta < 25°$ dient der Vorwärtsspurdetektor zur Spurrekonstruktion. Er besteht aus drei Supermodulen, die sich jeweils aus drei gegeneinander um 60° gedrehten planaren Driftkammern, einer Proportionalkammer, einer Übergangsstrahlungskammer und einer radialen Driftkammer zusammensetzen. Mit den planaren Kammern werden die x- und y-Koordinaten, mit einer Auflösung von $\sigma_{x,y} = 210\mu$m bestimmt. Die radiale Kammer besitzt die räumliche Auflösung $\sigma_{r,\phi} = 170\mu$m. Die Proportionalkammer dient zusammen mit den beiden Proportionalkammern des zentralen Spurkammersystems als Grundlage für den z-Vertex-Trigger.

Vervollständigt wird das Spurkammersystem seit 1995 durch acht im rückwärtigen Bereich des Detektors befindliche Driftkammerlagen BDC. In ihnen werden vorwiegend die Spuren der unter kleinem Winkel gestreuten Positronen ($155° \leq \theta \leq 175°$) nachgewiesen.

Die Spurkammer wurde 1995 im Vertexbereich durch Halbleiter-Spurdetektoren erweitert. Die Silizium-Spurdetektoren [Pre96] sind zwischen dem zentralen Spurdetektor und dem Strahlrohr installiert. Dabei umschließt der zentrale Silizium-Spurdetektor CST den Wechselwirkungspunkt, und dient als Vertexdetektor mit hoher Ortsauflösung. Der rückwärtige Silizium-Spurendetektor BST schließt in z-Richtung an den CST an. Er erlaubt eine präzise Messung des Polar- und Azimutwinkels der Spur geladener Teilchen im Winkelbereich $163° < \theta < 176°$. Seit 1997 sind Messungen mit den Silizium-Spurdetektoren möglich.

3.2.2 Das Flüssig-Argon-Kalorimeter

Das Flüssig-Argon-Kalorimeter setzt sich aus einem inneren elektromagnetischen und einem äußeren hadronischen Teil zusammen. Es ist entlang der Strahlachse in acht unabhängige Module unterteilt, wobei jedes Modul aus acht identischen Segmenten besteht (vergl. Abbildung 3.4).
Konstruktionsbedingte Spalte zwischen den Modulen werden zur Aufnahme der signalführenden Elektronik genutzt. Insgesamt besteht das Kalorimeter aus ca. 45000 einzeln auslesbaren Zellen, so daß eine gute Ortsauflösung erreicht wird. Der elektro-

Abbildung 3.4: *Das Flüssig-Argon-Kalorimeter des H1-Detektors*

magnetische Teil ist feiner segmentiert als der hadronische, um die kompaktere Struktur elektromagnetischer Schauer auflösen zu können. Das komplette System befindet sich in einem Flüssig-Argon-Kryostaten und deckt einen polaren Winkelbereich von $4^\circ < \theta < 155^\circ$ ab.

Im elektromagnetischen Teil des Kalorimeters wird Blei als Absorbermaterial verwendet. Dieser hat vom Wechselwirkungspunkt gesehen je nach z-Position eine Tiefe von ca. 20-30 Strahlungslängen X_0, wobei seine Energieauflösung $\sigma(E_e)/E_e \simeq 11\%/\sqrt{E_e} \oplus 1\%$ beträgt. Das passive Material des hadronischen Teils ist Edelstahl. Die Tiefe des hadronischen Teils beträgt abhängig vom Polarwinkel ca. 4.5 bis 8 nukleare Wechselwirkungslängen λ. Die Energieauflösung des hadronischen Teils beträgt $\sigma(E_h)/E_h \simeq 50\%/\sqrt{E_h} \oplus 2\%$.

Flüssiges Argon als aktives Medium eines Kalorimeters besitzt, trotz des erforderlichen Kühlungsaufwands (Siedetemperatur von Argon: 87.5 K), eine Reihe von Vorteilen. So ist Argon als Edelgas strahlungsresistent und über lange Zeiträume stabil. Zusätzlich friert durch die Kühlung ein Großteil der Verunreinigungen aus. Weitere Vorteile sind die hohe Elektronenbeweglichkeit (→ kleine Totzeiten), die hohe Ladungsausbeute und die Homogenität des flüssigen Argons.

3.2.3 Weitere Detektorkomponenten

Vor- und rückwärtige Kalorimeter

Im extremen Vorwärtsbereich ($0.7^\circ < \theta < 3.3^\circ$) dient das PLUG-Kalorimeter zum Nachweis von Hadronen. Die Energieauflösung dieses Kupfer-Silizium-Kalorimeters beträgt

$$\sigma(E)/E \simeq 150\%/\sqrt{E_h}.$$

Den rückwärtigen Bereich deckte in der Datennahmeperiode 1994 das elektromagnetische Kalorimeter BEMC ab. Dieses Blei-Szintillator- Kalorimeter mit 472 Auslesekanälen diente vor allem der Energiemessung des gestreuten Positrons im Winkelbereich von 151° bis 177° und erreichte dabei ein Energieauflösung $\sigma(E)/E$ von $\simeq 10\%/\sqrt{E_e} \oplus 1.7\%$.

Seit 1995 ist in diesem Bereich das SPACAL (SPAghetti CALorimeter) installiert, dieses Kalorimeter ist zum einen feiner segmentiert, und besitzt zum anderen einen eigenen hadronischen Teil (Die Tiefe des hadronischen Teil beträgt ca. 1λ). Die Energieauflösungen betragen für elektomagnetische Schauer im elektromagnetischen und hadronischen Teil $\sigma(E_e)/E_e \simeq 7.1\%/\sqrt{E_e} \oplus 0.6\%$ und $\sigma(E_e)/E_e \simeq 12.5\%/\sqrt{E_e} \oplus 3.8\%$ und für hadronische Schauer im hadronischen Teil $\sigma(E_h)/E_h \simeq 56\%$ [Swa97].

Myondetektoren

Durch die geringe Wahrscheinlichkeit für Bremsstrahlung und das Fehlen starker Wechselwirkungen haben Myonen ein hohes Durchdringungsvermögen. Im H1-Detektors werden sie in Streifen- und Drahtdetektoren nachgewiesen, die in das feldrückführende Eisenjoch der supraleitenden Spule implementiert sind. Das zentrale Myonsystem ist in vier Teile unterteilt: Den vorderen und hinteren Zentralbereich ($35^\circ < \theta < 130^\circ$), die hintere Endkappe ($130^\circ < \theta < 171^\circ$) und die vordere Endkappe ($4^\circ < \theta < 35^\circ$). Das instrumentierte Eisen ist im Zentralbereich in ϕ und z segmentiert, in den Endkappen in x und y. Dies führt zu 32 Modulen im Zentralbereich und 16 Modulen in jeder Endkappe. Zum Nachweis von Myonen sind in jedem Modul zwischen 7,5 cm dicken Eisenplatten Streamerrohrkammern angeordnet. Im Zentralbereich mit den Drähten parallel zur z-Richtung und in den Endkappen parallel zur x-Achse. Im Vorwärtsrichtung befindet sich ein Myondetektor, der aus Driftkammern besteht, die vor und hinter einem Toroidmagneten angeordnet sind. Dieser Detektor überdeckt den Winkelbereich von 3° bis 17°.

3.2.4 Reduktion der Datenmenge durch Trigger

Neben den Prozessen der Positron-Proton Wechselwirkung beobachtet man im Detektor eine Reihe von Untergrundereignissen, wie z.B. Strahl-Gas- oder Strahl-Wand-Ereignisse. Die dabei entstehende Ereignisrate von 100 kHz bis 1 MHz übersteigt die Verarbeitungsmöglichkeiten der Elektronik bei weitem. Die Aufgabe des dreistufigen

H1-Triggersystems (L1,L2,L4) ist es, diejenigen Ereignisse herauszufiltern, die der ep-Wechselwirkung zugeordnet werden können. Insbesondere dient das Triggersystem als Filter für Ereignisse mit hohem transversalen Impuls $p_\perp$.

3.2.5 Rekonstruktion und Simulation von Ereignissen

Rekonstruktion

Die in den einzelnen Detektorkomponenten aufgenommenen elektronischen Rohdaten, werden mit Hilfe des Programmpaketes H1REC [She93] rekonstruiert. Hierdurch erhält man Informationen, die zur Analyse der physikalischen Ereignisse notwendig sind. So werden z.B. die einzelnen Signale in den Spurkammern zu Spuren verbunden und parametrisiert. Die Parameter der so gefundenen Spuren werden in Datenbänken abgelegt. Im Kalorimeter werden die in den Zellen deponierten Enrgien nach Kalibration und Berücksichtigung von Rauschen rekonstruiert. Es wird versucht die einzelnen Zellenergien zu Gruppen,sogenannten Clustern, zusammenzufassen. Diese Cluster werden den Spuren in den Spurkammern zugeordnet.

Simulation

Mit dem Detektorsimulationsprogramm H1SIM wird das Verhalten einzelner Teilchen mit vorgegebenen Viererimpulsen im Detektor simuliert. Hierdurch erhält man aus Monte-Carlo-Ereignissen Daten, die mit den gemessenen Rohdaten verglichen werden können.
Die Detektorantwort wird in drei Stufen ermittelt:

- Das Programm GEANT simuliert die Wechselwirkung der Teilchen mit der Materie des Detektors. Dabei wird der Detektor in kleine Volumina, sogenannte Boxes und Tubes, unterteilt. Bei der schrittweisen Simulation der Teilchenspuren werden Volumengrenzen, Volumenmaterial (totes oder sensitives Material), Energieverlust und Ablenkung der Teilchen nach den bekannten physikalischen Gesetzen berechnet. Deponiert ein Teilchen Energie in einem sensitiven Bereich des Detektors, so werden Koordinaten und Energiedeposition dieses "Treffers" gespeichert. Die Simulation der elektromagnetischen und hadronischen Schauer erfolgt innerhalb GEANT durch die Programme H1FAST und GHEISHA.
- Im Programmpaket H1DIGI wird die Antwort der aktiven Detektorkomponenten ermittelt,wobei Signale erzeugt werden,die den Signalen der echten Datennahme entsprechen sollen.
- Durch das Programm H1TRIG werden die Triggerbedingungen simuliert.

Anschließend müssen die simulierten wie die gemessenen Rohdaten rekonstruiert werden (H1REC).

Kapitel 4

Neuronale Netze

4.1 Einleitung

Das menschliche Gehirn ist die komplexeste bekannte Struktur: Es besteht aus ca. einer Billion Zellen, davon sind 100 Milliarden Nervenzellen (Neuronen). Die in Netzwerken verknüpften Neuronen ermöglichen dem Menschen, Intelligenz, Kreativität, Gefühle, Gedächtnis und Bewußtsein zu entwickeln.

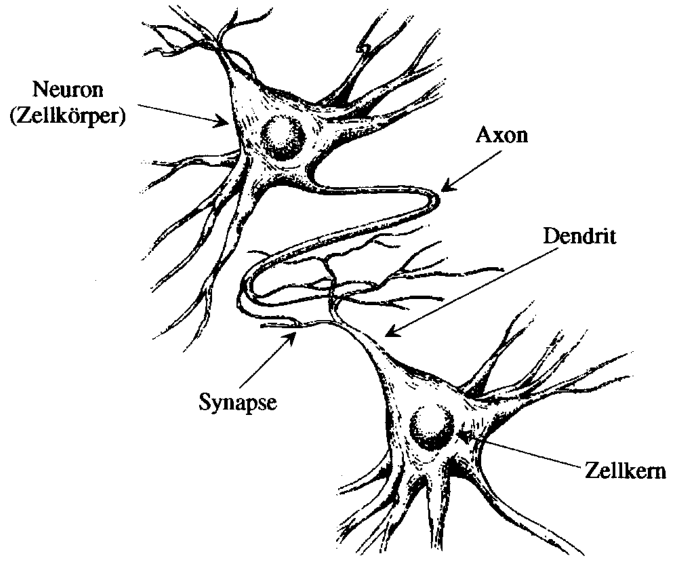


Abbildung 4.1: *Bild zweier biologischer Neuronen. Aus [Lue97].*

Ein typisches Neuron im menschlichen Gehirn (vgl. Abbildung 4.1) nimmt über seine zahlreichen Dendriten, baumartig verästelte Strukturen, Signale von anderen Neuronen auf. Über sein Axon, eine lange dünne Nervenfaser, die sich gleichfalls in Tausende von Zweigen verästelt, leitet es eigene Signale in Form elektrischer Entladungen weiter. Diese können ihrerseits die nachgeschalteten Neuronen erregen oder hemmen. Sind die erregenden Einflüsse auf ein Neuron im Verhältniss zu den hemmenden ausreichend groß, schickt es selbst einen elektrischen Impuls durch sein Axon ab. Wie stark sich ein Impuls in einem Neuron auf das nachgeschaltete auswirkt, hängt von der Effizienz der Übergangsstelle, der Synapse, ab. Indem sich diese synaptischen Übertragunsfaktoren verändern, findet Lernen statt.

Künstliche neuronale Netze sind mathematische Systeme zur stark vereinfachten Nachbildung der neuralen Netze des menschlichen Gehirns. Sie bestehen aus untereinander verbundenen Einheiten, wobei jede ein natürliches Neuron vertritt. In Analogie zu einem gewöhnlichem Netz spricht man von Knoten. Jeder Verbindung zwischen zwei Knoten ist eine Zahl, ein sogenanntes Gewicht, zugeordnet, die der synaptischen Effizienz entspricht. Künstliche Netze können allerdings bislang ein typisches Verknüpfungsmuster von Dendriten und Axonen weder in der Anzahl noch in der Komplexität der Verschaltungen nachbilden. Zudem gibt ein künstliches Neuron nicht wie das reale eine Folge von Impulsen aus, sondern eine Zahl, die seiner Aktivität (der Anzahl der Impulse pro Zeiteinheit) entspricht [Hin92].

Neuronale Netze haben sich u.a. bei Problemen der Mustererkennung, Funktionsanpassung, Robotik und Spracherkennung bewährt [Lis92]. In dieser Arbeit soll untersucht werden, ob sie es ermöglichen in einem von den kinematischen Variablen aufgespannten mehrdimensionalen Ereignissraum, Muster zu erkennen, die einzelnen Ereignissklassen bei der e^+p-Streuung (Photoproduktion, geladene Ströme usw.) zugeordnet werden können. Dieses Vermögen der Mustererkennung soll vom Netz durch ein Training mit bekannten Mustern "erlernt" werden. Dabei wird nach der Eingabe der Trainingsmuster aus dem Vergleich der Ausgabe des Netzes mit den vorgegebenen Zielausgaben ein Fehler bestimmt. Dieser wird verwandt um das Netz zu modifizieren und dadurch den Fehler zu minimieren. Dieser Vorgang wird als *error-back-propagating* bezeichnet, und das Netz als *Back-Propagation Netzwerk*.

In der Hochenergiephysik werden neuronale Netze u.a. zur Klassifizierung von Teilchen, bei der Rekonstruktion von Spuren in Drift- und Proportionalkammern, zur Rekonstruktion der invarianten Masse und als Ereignisstrigger eingesetzt. So wurden z.B. am FNAL[1] neuronale Netzwerke eingesetzt um das Top-Quark nachzuweisen [Ku94].

Bei H1 werden neuronale Netze in der Ereignisselektion [Nez95], in der Energierekonstruktion [Hop97] und in der L2 Triggerstufe angewandt. Der L2 Neuronale Netzwerktrigger [Koe97] setzt sich in der Hardware aus einer Reihe von CNAPS/VME-Karten (**C**onnectiv **N**etwork of **A**daptiv **P**rocessor**S**) zusammen,die mit jeweils einem **D**ata **D**istribution **B**oard (DDB) eine Triggerbox bilden, die einem Triggerelement entspricht (vgl. Abbildung 4.2).

Das DDB bereitet die Daten der Subdetektoren für die weitere Verwendung vor und leitet die Netzantwort an die zentrale Triggerlogik weiter. Die CNAPS/VME-Karten

---

[1] Fermi National Accelerator Laboratory

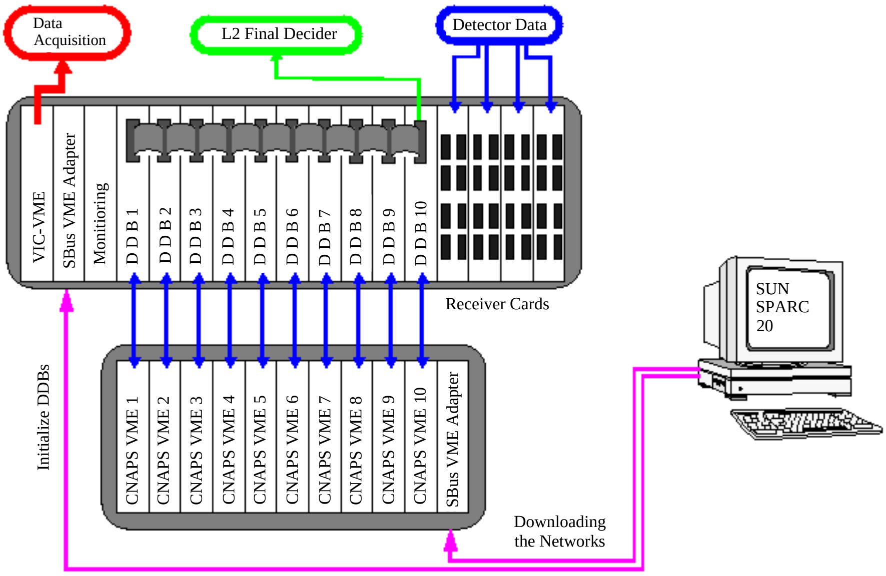


Abbildung 4.2: *Übersicht über das L2 Trigger System des H1-Detektors.*

bestehen aus einer Matrix von 64 bis 512 Prozessorknoten, die parallel aus den Detektorsignalen die Ausgabewerte der Knoten der ersten Lage des Netzes berechnen.Aus diesen Ergebnissen wird wiederum mit den Prozessorknoten der Wert am einzigen Knoten der Ausgabelage bestimmt. Das Triggerbit des L2NN-Triggerelements wird gesetzt,wenn das Ergebniss einen für das Netz spezifischen Schwellenwert überschreitet. Das Ereigniss wird in diesem Fall von der Triggerlogik akzeptiert und an folgende Triggerstufen weitergeleitet. Die Netze im L2NN wurden trainiert um z.B. Ereignisse vom Typ

$$\gamma p \rightarrow jets \qquad \text{oder} \qquad \gamma p \rightarrow p + J/\psi \rightarrow \mu\mu$$

zu erkennen und erreichen dabei Effizienzen von bis zu 90%.

4.2 Trennung der Ereignisklassen mit neuronalen Netzen

In dieser Arbeit soll untersucht werden, ob eine Unterscheidung der Ereignisse des geladenen Stroms von Untergrundprozessen mit Hilfe neuronaler Netze möglich ist.

Um interessante Ereignisse von Untergrundprozessen zu trennen, werden in der Analyse von Experimenten der Teilchenphysik die kinematischen und andere charakteristische Größen auf bestimmte Wertebereiche eingeschränkt.

Jedes Ereignis kann durch einen Punkt in dem, von den verfügbaren n Variablen aufgespannten n-dimensionalen Ereignisraum beschrieben werden. Alle Ereignisse einer Klasse können einem beschränkten Bereich dieses Ereignisraumes zugeordnet werden. Sind die Bereiche zweier Ereignisklassen separabel, so können sie durch eine Beschränkung des Ereignisraum vollständig getrennt werden. Durch die Beschränkung einzelner Variablen ist eine vollständige Trennung der Klassen nicht möglich. Als Beispiel betrachte man Abbildung 4.3.

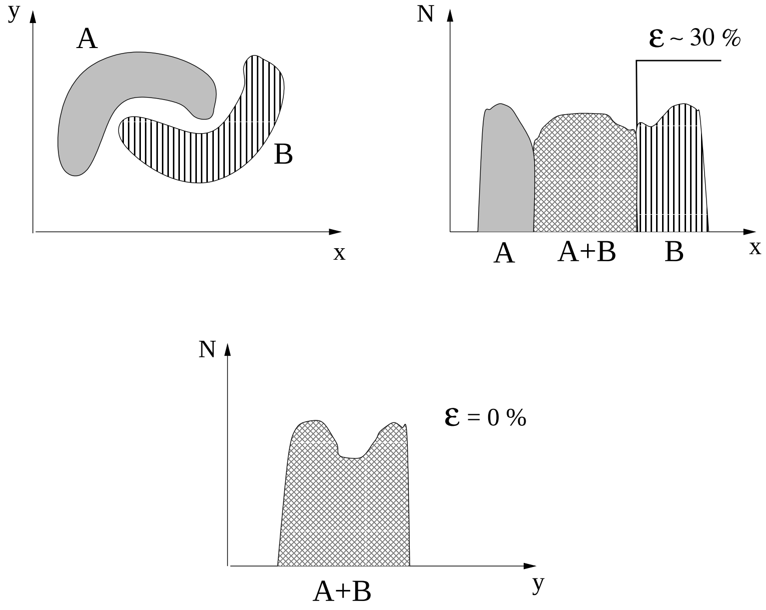


Abbildung 4.3: *Beispiel: Die Bereiche der beiden Klassen A und B sind im zweidimensionalen Raum separabel. Durch die Beschränkung der Varablen x bzw. y ist keine vollständige Trennung der Klassen möglich.*

Der n-dimensionale Ereignisraum wird hier durch zwei Dimensionen symbolisiert. Im zwei-dimensionalen Raum ist es möglich, die beiden Bereiche der Ereignisklassen A und B durch eine Linie zu trennen. Betrachtet man die Projektion auf die x-Achse, so erkennt man, daß die beiden Bereiche durch eine Beschränkung der Größe x nur mit geringer Effizienz ε getrennt werden können. Es ensteht eine Überlappung, welche die vollständige Trennung der Bereiche verhindert. Bei die Projektion auf die y-Achse ensteht eine vollständige Überlappung, so daß keine Trennung erzielt werden kann.

Durch die neuronalen Netze wird eine Methode zur vollständigen Trennung der Ereignisklassen genau dann zur Verfügung gestellt, wenn die notwendigen Eingabevariablen (diese spannen den Ereignisraum, auf in dem die Klassen separabel sind) identifiziert werden können. Jedes Neuron eines neuronalen Netzes trennt den n-dimensionalen Raum der Eingabevektoren durch eine zum Gewichtsvektor senkrecht stehende $(n-1)$-dimensionale Hyperebene. Ein neuronales Netz erzeugt im Trainingsprozeß ein System von Hyperebenen, welches die beiden Ereignisklassen trennt.

4.3 Theorie der neuronalen Netze

4.3.1 Das Modellneuron

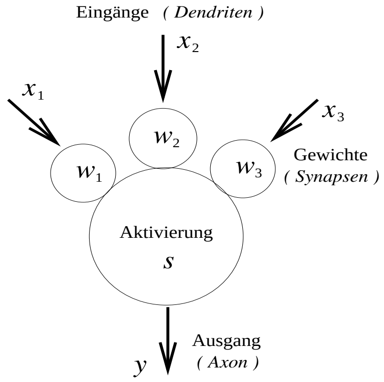


Abbildung 4.4: *Modell eines Neurons*

Das Grundmodell eines Neurons (siehe Abbildung 4.4) kann als eine Art Addierer, mit einem vorgegebenen Schwellenwert am Ausgang betrachtet werden. Die Verbindungen (Synapsen) eines Neurons nehmen eine Aktivierung x_i mit einer bestimmten Stärke w_i von anderen Neuronen auf, und summieren diese. Wenn die Summe der gewichteten Aktivierungen den Schwellenwert θ überschreitet, entsteht am Ausgang (Axon) des Neurons ein Signal y.
Faßt man die *Eingabeaktivitäten* $x_1,...,x_n$ zum *Eingabevektor* $\mathbf{x} = (x_1,...,x_n)$ und die *Gewichte* $w_1,...,w_n$ zum *Gewichtsvektor* $\mathbf{w} = (w_1,...,w_n)$ zusammen, so läßt sich die resultierende *Aktivität* s als Summe der gewichteten Eingaben und damit formal als Skalarprodukt beider Vektoren schreiben:

$$s(\mathbf{w},\mathbf{x}) = \sum_j w_j x_j = \mathbf{w} \cdot \mathbf{x} \tag{4.1}$$

Die Aktivität muß erst eine *Schwelle* θ überschreiten, bevor sie am Ausgang ein Signal erzeugt. Dies läßt sich durch die Minderung der Aktivität um den Schwellenwert darstellen:

$$s(\mathbf{w},\mathbf{x}) = \sum_{j=1}^{n} w_j x_j - \theta = \sum_{j=0}^{n} w_j x_j \tag{4.2}$$

mit $\theta = -w_0 x_0$.

Um den Wert y am Neuronausgang auf ein bestimmtes Intervall zu beschränken (z.B. [0,1]), wird der Ausgabewert y aus der internen Aktivität s des Neurons, mit einer Ausgabefunktion f, welche den Wertebereich [0,1] besitzt berechnet, $y = f(s)$. Die gesamte Reaktion des formalen Neurons kann durch eine Funktion,die Transferfunktion $y = F(\mathbf{x},\mathbf{w},s,f)$, dargestellt werden.

4.3.2 Neuronale Netzwerke

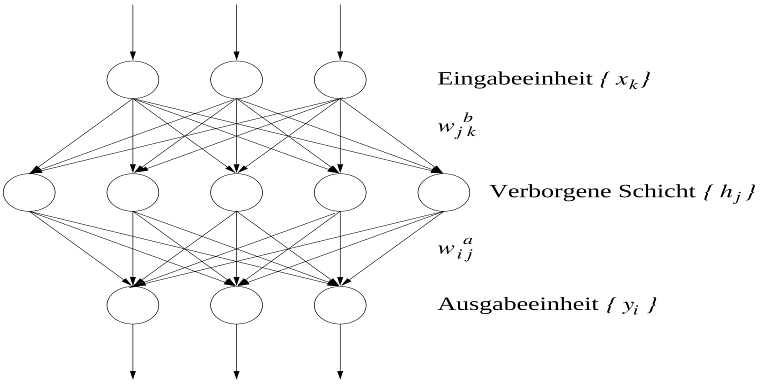


Abbildung 4.5: *Prinzip eines Feed-Forward Netzwerkes mit einer verborgenen Ebene. Die Richtung des Informationstransports ist durch die Pfeile angedeutet.*

Die Abbildung 4.5 zeigt den Aufbau eines typischen Netzwerks in einer Mehrschichten-Architektur. Dabei sind die einzelnen Neuronen in verschiedenen Ebenen angeordnet. Zwischen der Ein- und der Ausgabeebene befindet sich mindestens eine Ebene mit Neuronen, deren Ausgänge nicht beobachtet werden können. Diese Ebene wird daher als verborgene Ebene bezeichnet. Die Informationen werden von den einzelnen Knoten einer Ebene zu den Knoten der folgenden Ebene mit bestimmten Gewichten w_{ij},w_{jk} weitergegeben. Die Weitergabe von Signalen erfolgt dabei in nur einer Richtung: Von den Eingabeeinheiten zu den Knoten der verborgenen Ebene und weiter zu den Ausgabeeinheiten. Man nennt ein solches Netz ein *feed-forward Netz.*

Das Verhalten eines solchen Netzes wird durch eine Ausgabefunktion f beschrieben. Diese muß so gewählt werden, daß sie ein nichtlineares Verhalten des Netzes ermöglicht.

Um das im Folgenden beschriebene Lernen des Netzes mit einer Gradientenmethode zu ermöglichen, muß die Ausgabefunktion differenzierbar sein.

Eine mögliche Wahl der Ausgabefunktion f ist:

$$f(s) = \frac{1}{2}[1 + \tanh(\frac{s}{T})] \tag{4.3}$$

mit Werten im Bereich [0,1].

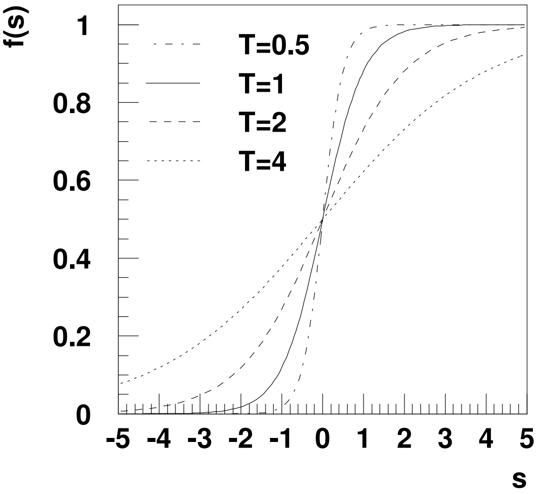


Abbildung 4.6: *Die Ausgabefunktion als Funktion der Aktivierung s, bei vier verschiedenen Werten der Temperatur. Für $T \to 0$ nähert sie sich der Heavyside-Funktion an.*

T ist ein Parameter, der die Steigung der Ausgabefunktion am Ursprung beschreibt.Man bezeichnet T als die Temperatur des Netzes. In Abbildung 4.6 ist die Ausgabefunktion für verschiedene Werte der Temperatur abgebildet. Es entstehen, bei Temperaturen $T > 0$, an den Ausgängen der einzelnen Knoten Werte zwischen Null und Eins.

Der Ausgabewert y_i des i-ten Knotens der Ausgabeschicht wird aus den Ausgabewerten der Knoten der vorhergehenden Schichten und den Eingabewerten berechnet:

$$y_i = f(s_i^a) \tag{4.4}$$

f ist die nichtlineare Ausgabefunktion, und s_i^a ist die Aktivität in der Ausgabeeinheit i

$$s_i^a = \sum_{j=1}^{M_h} w_{ij}^a h_j - \Theta_i^a = \sum_{j=0}^{M_h} w_{ij}^a h_j. \tag{4.5}$$

Dabei sind:

- M_h die Anzahl der Knoten in der verborgenen Ebene.
- w^a_{ij} die Gewichte der Verbindungen zwischen den Knoten der verborgenen und der Ausgabeebene.
- Θ^a_i die Schwellenwerte.
- h_j die Ausgabewerte der Knoten in der verborgenen Ebene :

$$h_j = f(s^b_j) \tag{4.6}$$

$$s^b_j = \sum_{k=1}^{M} w^b_{jk} x_k - \Theta^b_j = \sum_{k=0}^{M} w^b_{jk} x_k \tag{4.7}$$

Hierbei sind:

- s^b_j die Aktivitäten in den Knoten der verborgenen Ebene.
- M die Anzahl der Knoten in der Eingabeebene.
- w^b_{jk} die Gewichte der Verbindungen zwischen den Knoten der Eingabeebene und den Knoten der verborgenen Ebene.
- x_k die Eingabewerte.

Der Ausgabewert y_i des i-ten Neurons der Ausgabeschicht wird somit durch folgende Funktion aus den Eingabewerten x_k berechnet:

$$y_i = f\Big\{ \sum_{j=0}^{M_h} w^a_{ij} f(\sum_{k=0}^{M} w^b_{jk} x_k) \Big\} \tag{4.8}$$

4.3.3 Lernalgorithmen

Die Aufgabe eines Lernverfahrens besteht darin, die Gewichte des zu trainierenden Netzes so zu verändern, daß die Eingabegrößen in der gewünschten Weise auf die Ausgabegrößen abgebildet werden. Mit einem Lernalgorithmus werden durch ein Training mit Eingabemustern $\mathbf{x} = (x_1,...,x_n)$ und den dazu gewünschten Ausgabemustern $\mathbf{t}$ die Gewichte solange modifiziert, bis das Netzwerk die gewünschte Leistung (z.B. die Trennung zweier Ereignisklassen) mit genügender Genauigkeit erbringt. Dazu wird nach dem Durchlaufen des Eingabesignals $\mathbf{x}$ durch die Netzebenen der Fehler $\delta := \mathbf{y}(\mathbf{x}) - \mathbf{t}(\mathbf{x})$ des Ausgangssignals $\mathbf{y}$ bezüglich der gewünschten Ausgabe, durch alle Ebenen zurückgeführt.

Das Ziel ist die Minimierung des Fehlers bei der Mustererkennung für alle Trainingsmuster. Die im Folgenden verwendete, minimierende Fehlerfunktion $E(\mathbf{w})$ ist:

$$E(\mathbf{w}) = \frac{1}{2N_p}\sum_{p=1}^{N_p}\sum_i (y_i^{(p)} - t_i^{(p)})^2 = \frac{1}{2N_p}\sum_{p=1}^{N_p}(\mathbf{y}^{(p)} - \mathbf{t}^{(p)})^2 = \frac{1}{N_p}\sum_{p=1}^{N_p} E_p \tag{4.9}$$

Die Summation erstreckt sich über alle N_p Trainingsmuster $\mathbf{x}^{(p)}$.
Dabei ist:

- $\mathbf{y}^{(p)} = \mathbf{y}(\mathbf{x}^{(p)})$ die Netzausgabe bei Trainingsmuster p.
- $\mathbf{t}^{(p)} = \mathbf{t}(\mathbf{x}^{(p)})$ die Zielvorgabe bei Trainingsmuster p.
- $\mathbf{w}$ der Vektor der Gewichte im Netz

Um das Minimum aufzufinden, benutzt man eine Gradientenmethode. Dabei nimmt man an, die bekannte Fehlerfunktion $E(\mathbf{w})$ habe ein lokales Minimum. Ausgehend von einer initialen Wertematrix $\mathbf{w}_a$ in der Nähe des lokalen Minimums der Fehlerfunktion versucht man sich iterativ dem optimalen $\mathbf{w}_{min}$ anzunähern. Die Ableitung nach den Gewichten zwischen dem i-ten Neuron der Ausgabeebene und den Neuronen der verborgenen Ebene, kann in einem Vektor $\nabla_{\mathbf{w}_i} E(\mathbf{w}) := (\frac{\partial E(\mathbf{w})}{\partial w_{i1}}, \ldots, \frac{\partial E(\mathbf{w})}{\partial w_{ij}})$ zusammengefaßt werden.In der Nähe des Minimums, zeigt der so definierte Vektor der Richtungsableitung in die Richtung des stärksten Anstiegs der Funktion. Die Differenz von $\mathbf{w}(t-1)$ beim $(t-1)$-ten Schritt zu $\mathbf{w}(t)$ vom nächsten Schritt t ist proportional dem negativen Gradienten:

$$\mathbf{w}_i(t) = \mathbf{w}_i(t-1) - \gamma(t)\nabla_{\mathbf{w}_i} E(\mathbf{w}(t-1)) \tag{4.10}$$

Eine solche Iterationsgleichung zur Optimierung von Parametern einer Fehlerfunktion wird als *Lernregel*, die schrittabhängige Proportionalitätskonstante $\gamma(t)$ als *Lernrate* bezeichnet.
Der Gradientenalgorithmus für die Iteration des Gewichts w_{ij} von Neuron j zu Neuron i wird damit

$$w_{ij}(t) = w_{ij}(t-1) - \gamma\frac{1}{N_p}\sum_p \frac{\partial E_p}{\partial w_{ij}} \tag{4.11}$$

Nach Ausführung der partiellen Differentiation (durch mehrmalige Anwendung der Kettenregel) erhält man für die Ausgabeebene:

$$\Delta w_{ij}^{(a)} = w_{ij}(t) - w_{ij}(t-1) = \gamma\frac{1}{N_p}\sum_p \delta_i^{(a)} h_j^{(a)} \tag{4.12}$$

mit $\delta_i^{(a)} = -(y_i^{(a)} - t_i^{(a)}) f'(s_i^{(a)})$. Dabei ist:

- $(y_i^{(a)} - t_i^{(a)})$ der Fehler des Neurons i der Ausgabeebene

- $f'(s_i) = \frac{\partial y_i}{\partial s_i}$ die partielle Ableitung der Ausgabefunktion nach der Aktivität im Neuron i der Ausgabeebene.

Um die Änderung der Gewichte zwischen verborgener Ebene und Eingabeebene zu berechnen, wird der beobachtete Fehler im j-ten Neuron der verborgenen Ebene rekursiv durch die Fehler der Neuronen der Ausgabeebene dargestellt. Als Ergebnis erhält man:

$$\Delta w_{jk}^{(b)} = \gamma \frac{1}{N_p} \sum_p (- \sum_i \delta_i^{(a)} w_{ij}^{(a)} f'(s_j^{(b)})) x_k^{(b)} \tag{4.13}$$

Mit diesen Lernregeln wird für jedes Trainingsmuster die notwendige Korrektur der Gewichte berechnet. Aber erst nach dem letzten Trainingsmuster wird die tatsächliche Korrektur der Gewichte angebracht. Da die einzelnen Korrekturbeiträge auf der Basis der alten Gewichte ohne Kenntnis der bereits bestimmten Korrekturen errechnet wurden, wird dies als *offline Version* bezeichnet.

Es existieren eine ganze Reihe von Methoden die Gewichte innerhalb des Lernprozesses zu aktualisieren [Lon94]. Die Form

$$\Delta w_t = -\gamma \frac{\partial E_{t-1}}{\partial w} = -\gamma \nabla E_{t-1} \tag{4.14}$$

wird als *Standard-Back-Propagation* bezeichnet.

Oftmals wird ein Stabilisierungsterm addiert, welcher Ozillationen des Lernprozesses zwischen den gegenüberliegenden Randbereichen eines Minimums der Fehlerfunktion verhindern soll:

$$\Delta w_t = -\gamma \nabla E_{t-1} + \alpha \Delta w_{t-1} \tag{4.15}$$

dabei ist $\alpha < 1$ eine Konstante und w der gesamte Gewichtsvektor im Netz.

Eine weitere Möglichkeit, den Lernprozeß zu stabilisieren, ist durch Hinzufügen eines Rauschterms möglich. Dieser soll die Stagnation des Lernprozesses auf einem flachen Plateau der Fehlerfunktion verhindern:

$$\Delta w_t = -\gamma \nabla E_{t-1} + \sigma \tag{4.16}$$

wobei σ durch eine normalisierte Gaußfunktion mit Dimension(σ) = Dimension(w) erzeugt wird. Man bezeichnet diese Methode als *Langevin Aktualisierung.*

Eine Methode, welche nur die Richtung des Gradienten und nicht dessen Betrag berücksichtigt, ist die sogenannte *Manhattan Aktualisierung* :

$$\Delta w_t = -\gamma \; sgn(\frac{\partial E_{t-1}}{\partial w}) \tag{4.17}$$

Dabei bezeichnet w eine Komponente des Gewichtsvektors.

Die zur Zeit leistungsfähigste Aktualisierungsmethode ist der sogenannte *Resilient Backpropagation Algorithmus* (RPROP) [Ried93]. In ihm wird jedem Gewicht ein individueller Aktualisierungswert Δ_{ij} zugeordnet. Diese Aktualisierungsgrößen werden während des Lernprozesses, basierend auf dem lokalen Vorzeichen der Fehlerfunktion, in folgender Weise berechnet:

$$\Delta_{ij}^{(t)} = \begin{cases} \gamma^{+} * \Delta_{ij}^{(t-1)} & \text{wenn } \frac{\partial E^{(t-1)}}{\partial w_{ij}} * \frac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ \gamma^{-} * \Delta_{ij}^{(t-1)} & \text{wenn } \frac{\partial E^{(t-1)}}{\partial w_{ij}} * \frac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ \Delta_{ij}^{(t-1)} & \text{sonst} \end{cases} \tag{4.18}$$

wobei γ^{-},γ^{+} Konstanten sind, mit $0 < \gamma^{-} < 1 < \gamma^{+}$.
Bei jedem Vorzeichenwechsel der Partiellen Ableitung der Fehlerfunktion nach w_{ij} (Sprung über ein lokales Minimum), welcher anzeigt, daß die vorhergehende Aktualisierung zu groß war, wird der Aktualisierungswert Δ_{ij} um den Faktor γ^{-} verkleinert. Wechselt die Partielle Ableitung ihr Vorzeichen nicht, so wird der Aktualisierungswert, zur Beschleunigung der Konvergenz, um den Faktor γ^{+} erhöht. Die Aktualisierung der Gewichte wird nach folgender Vorschrift bestimmt: Ist die Ableitung positiv (wachsender Fehler), wird das Gewicht um Δ_{ij} verringert, ist die Ableitung negativ, wird Δ_{ij} zum Gewicht addiert:

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)} & \text{wenn } \frac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ +\Delta_{ij}^{(t)} & \text{wenn } \frac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ 0 & \text{sonst} \end{cases} \tag{4.19}$$

In dieser Arbeit wird der *Manhattan*- oder der *RPROP*-Algorithmus verwendet.

4.4 Optimierung neuronaler Netzwerke

Nicht nur die optimale Einstellung der Gewichte durch den Lernalgorithmus, sondern auch die Topologie (Anzahl der Ebenen und Anzahl der Knoten), die Parameter und die Auswahl der Eingabegrößen des Netzwerkes sind von entscheidender Relevanz für dessen Leistungsfähigkeit.

Die Optimierung der Topologie und Parameter des Netzes stellt Anforderungen die ein Gradientenverfahren nicht leisten kann, da z.B. keine differenzierbare Fehlerfunktion existiert, welche den Netzaufbau bewertet. Daher besteht die herkömmliche Vorgehensweise in langwierigen, wenig effizienten Reihenuntersuchungen, in denen Netzwerke mit verschiedenen Topologien, Parametern und Eingabegrößen trainiert und getestet werden. Alleine aus Zeitgründen ist eine vollständige Überprüfung des Testraumes nicht möglich. Eine Alternative zu dieser Methode besteht in der Anwendung evolutionärer Algorithmen.

4.4.1 Evolutionäre Algorithmen

Die Aufgabe neuronaler Netzen besteht darin eine vorgegebene Aufgabe zu lösen (z.B. Muster zu erkennen oder eine Funktion zu approximieren). Erreicht wird dies durch ihre flexible Architektur und die Einstellung der Gewichte im Netz mit einem Lernalgorithmus. Allgemein entspricht dies der Aufgabenstellung, eine vorgegebene Zielfunktion (z.B. die Effizienz der Mustererkennung) zu optimieren. Eine weitere erfolgreiche Methode, Optimierungsaufgaben zu lösen, besteht in der Anwendung biologischer Funktionsprinzipien, die der Evolution nachempfunden sind.

Die Grundidee evolutionärer Algorithmen besteht darin, die Parameter, für die man optimale Werte zur Lösung eines vorgegebenen Problems sucht, zu einem Tupel $\mathbf{g}$ zusammenzufassen und dieses Tupel als Bauplan (Chromosomen) eines Lebewesens zu betrachten [Rud94]. Genauso wie sich in der freien Natur die Lebewesen den Umweltbedingungen anpassen oder aussterben, sollen die Tupel Veränderungen unterworfen werden und nur die Besten als Ausgangspunkt für noch Bessere Verwendung finden. Die evolutionären Algorithmen versuchen ein solches Zahlentupel $\mathbf{g}_m$ zu finden, mit dem die vorgegebene Zielfunktion $R(\mathbf{g})$ maximiert bzw. minimiert wird.

Der Vorteil evolutionärer Algorithmen ist, daß sie zum einen universell anwendbar sind, und daß zum anderen stets mit einer Menge von Lösungen (Population von Tupeln) gearbeitet wird. Somit werden gleichzeitig mehrere Lösungsmöglichkeiten getestet. Aus den Tupeln mit den besten Resultaten werden durch *Rekombination* neue Tupel erzeugt. Durch Vererbung werden die Informationen über den besten Lösungsweg in der Population ausgetauscht.

Die evolutionären Algorithmen bieten die Möglichkeit, die neuronalen Netze zu optimieren. Neben der Optimierung (Lernen) der Gewichte, hat man die Möglichkeit die Topologie und Parameter (z.B. die Temperatur und die initiale Gewichte) der Netze zu optimieren.
In dieser Arbeit werden evolutionäre Algorithmen angewandt, um eine optimale Kombination möglicher Eingabevariablen zu finden und um die Architektur (Anzahl der Ebenen, Anzahl der Knoten usw.) der Netze zu optimieren.

Allgemein kann ein *Individuum* $\mathbf{a} \in \mathcal{A}$ durch das Tupel

$$\mathbf{a} = (\mathbf{g}, s_1, ..., s_r) \in \mathcal{M} \times \mathcal{S}_1 \times ... \times \mathcal{S}_r$$

mit $r \geq 0$ beschrieben werden, wobei $\mathbf{g} \in \mathcal{M}$ eine zulässige Lösung und $s_1, ..., s_r$ Strategieparameter (z.B. die Mutationsrate) repräsentieren. Jedem Individuum $\mathbf{a} \in \mathcal{A}$ wird über die Fitnessfunktion $F : \mathcal{A} \to R$ ein Fitnesswert (eine reelle Zahl) zugeordnet. Die Fitnessfunktion ist stets von der Zielfunktion abhängig, und in vielen Fällen ist sie mit ihr identisch.

Eine *Population* besteht aus einer Anzahl μ von Individuen und wird während jeder Iteration unter Verwendung der genetischen Operatoren verändert:

- Bei der *Mutation* wird jedes Individuum zufällig verändert.
- Die *Rekombination* erzeugt aus zwei zufällig gewählten Individuen ein neues Individuum durch Überkreuzen (cross-over) der Tupel.

- Die *Selektion* entscheidet anhand der Fitness, welche neu erzeugten Individuen in die nächste Iteration (Generation) übernommen werden.

4.4.2 Der kanonische genetische Algorithmus

Der sogenannte *kanonische genetische Algorithmus* benutzt zur Repräsentation der Individuen Binärstrings der Länge ℓ [Bec94]. Möchte man andere Problemklassen bearbeiten, so ist eine Umformulierung des Problems auf den Binärraum notwendig.

Jedes Bit im Binärstring eines Individuums kann bei der Mutation durch ein Zufallsexperiment invertiert werden. Dazu wird eine auf dem Intervall [0,1] gleichmäßig verteilte Zufallszahl erzeugt. Falls diese kleiner gleich einer fest vorgegebenen Mutationswahrscheinlichkeit p_m ist, wird das Bit invertiert. Durch die zufällige Mutation der Individuen wird eine Stagnation des Optimierungsprozesses verhindert und die Suche nach erfolgreichen Individuen auf den gesamten Testraum ausgeweitet.

Die Rekombination (cross-over) der Individuen wird nach folgender Vorschrift durchgeführt. Man wählt zufällig zwei Individuen aus der Population sowie eine Zufallszahl $c \in \{1,2,...,\ell - 1\}$. Ein neues Individuum erhält die ersten c Positionen seines Bitstrings von einem, die restlichen vom anderen Elternteil (vgl. Abbildung 4.7). Durch die Rekombination der Erbinformation ist es möglich, daß die erfolgreichen Eigenschaften zweier Individuen auf ein neues Individuum übertragen werden und dieses die beiden erfolgreichen Eigenschaften in sich vereint.

Aus den so erzeugten und anschließend mutierten Individuen wird die neue Population selektiert. Dabei werden die Individuen nach ihrer mit der Fitnessfunktion berechneten Fitness bewertet. Der gesamte Zyklus wird solange wiederholt, bis die Individuen eine genügend große Fitness besitzen.

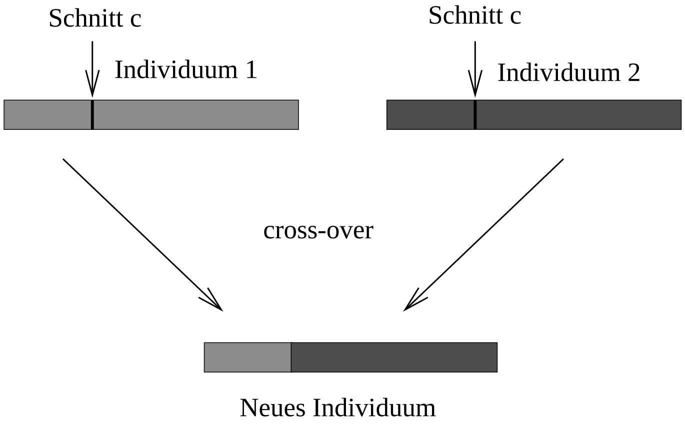


Abbildung 4.7: *Funktionsweise des Rekombinationoperators*

Kapitel 5

Geladener Strom und Untergrundprozesse

In dieser Arbeit soll untersucht werden ob eine Unterscheidung der Ereignisse des geladenen Stroms von den Ereignissen der Untergrundprozesse mit neuronalen Netzwerken möglich ist. Dazu werden in diesem Kapitel die Signaturen der Ereignisse des geladenen Stroms und der Untergrundereignisse dargestellt. Es werden die bisherigen Methoden zur Separation der Ereignisklassen, im Hinblick auf mögliche Eingabevariablen eines neuronalen Netzwerkes diskutiert.

5.1 Signatur der Ereignisse des geladenen Stroms

Ereignisse des geladenen Stroms in der e^+p-Wechselwirkung zeichnen sich im Endzustand durch ein Antineutrino und mindestens einen hadronischen Jet aus. Die Jets entstehen durch Hadronisierung des gestreuten Partons und des Protonrestes. Das Teilchenbündel des Protonrests besitzt im Allgemeinen einen kleinen transversalen Impuls und kann nicht vollständig nachgewiesen werden, da es den H1-Detektor durch das Strahlrohr verläßt.

Bezogen auf den Wechselwirkungspunkt läuft das Positron aus negativer z- und das Proton in positiver z-Richtung ein. Bei idealer Messung ist der Transversalimpuls der Ereignisse der e^+p-Wechselwirkungen bei HERA aufgrund der Impulserhaltung ausgeglichen. Da Neutrinos mit dem H1-Detektor nicht nachgewiesen werden, ist in Ereignissen des geladenen Stroms die gemessene Transversalimpulsbilanz nicht ausgeglichen. Der gemessene Betrag des Transversalimpulses des hadronischen Endzustandes entspricht dem des Neutrinos und wird mit V_{had} oder $p_{\perp,had}$ bezeichnet.

Bei den Ereignissen des geladenen Stroms werden im H1-Detektor folgende Signaturen erwartet:

- mindestens ein Teilchenjet.

- innerhalb des Jets Energiedepositionen im hadronischen Teil des Kalorimeters.
- eine Anzahl von Spuren in den Spurkammern die innerhalb des Jets von einem Vertex zu den Energiedepositionen im Kalorimeter gerichtet sind.
- eine asymmetrische Energiedeposition in der radialen Ebene des Detektors.
- keine isolierten Spuren und kein isolierter elektromagnetischer Cluster in der zum Jet entgegengesetzten Richtung.
- Energiedepositionen im instrumentierten Eisen nur in dem vom Jet aufgespannten Raumwinkelbereich.

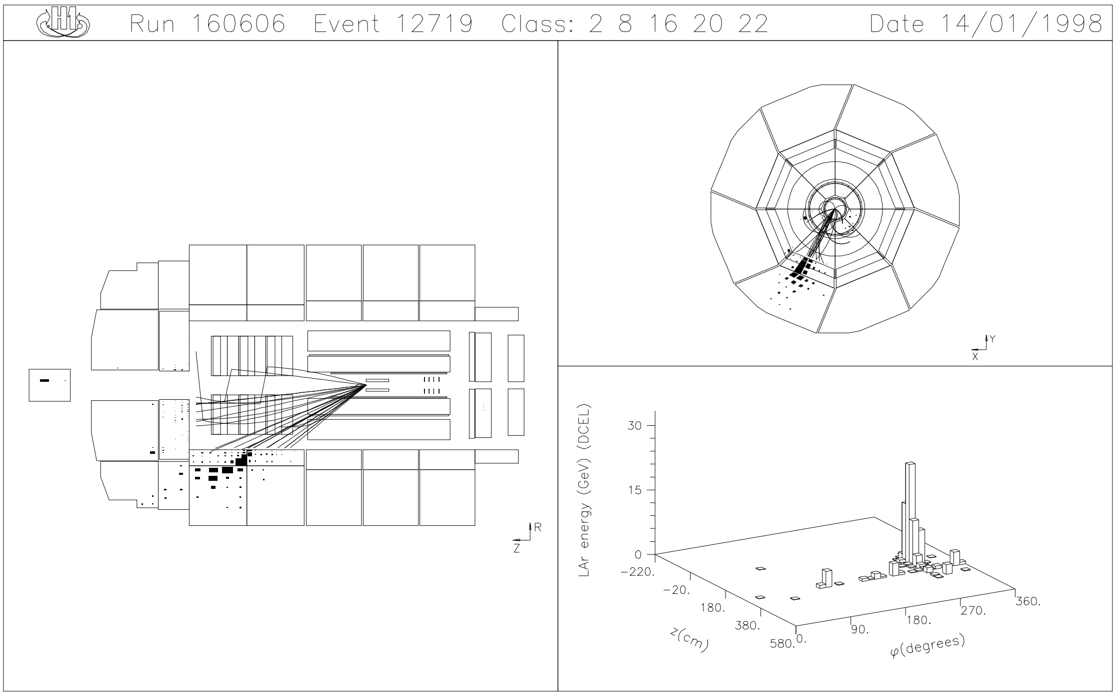


Abbildung 5.1: *Darstellung eines mit dem H1-Detektor beobachteten potentiellen Ereignisses des geladenen Stroms. In der radialen Projektion des Detektors erkennt man eine asymmetrische Energieverteilung.*

Bei der Selektion von Ereignissen des geladenen Stroms stellt der Transversalimpuls das wichtigste Selektionskriterium dar. In der Vorselektion werden Ereignisse die folgende Bedingungen erfüllen, als potentielle Ereignisse des geladenen Stroms, sogenannte *Klasse 8-Ereignisse*, bezeichnet:

- Der gemessene transversale Impuls beträgt $p_{\perp,had} > 10$ GeV.

- Mindestens eine rekonstruierte Spur in der zentralen Spurkammer muß existieren.
- Mindestens ein Trigger des Flüssig-Argon-Kalorimeters muß aktiviert sein.
- Der Vertex muß die Bedingung $|z_{vertex} - \bar{z}_{Vertex}^{Run}| < 50$ cm erfüllen.

Die Abbildung 5.1 stellt ein mit dem H1-Detektor beobachtetes Ereignis eines Kandidatens des geladenen Stroms dar. Man erkennt einen ausgeprägten Teilchenjet mit einer großen Anzahl von rekonstruierten Spuren in den Spurkammern und Energiedepositionen im Kalorimeter. An der asymmetrischen Energiedeposition in der radialen Ansicht des Detektors erkennt man, daß der Transversalimpuls nicht ausgeglichen ist. Die Abbildung 5.2 zeigt einen weiteren CC-Ereignis-Kandidaten mit zwei hadronischen Teilchenjets. In der radialen Sicht des Detektors ist zu erkennen, daß die beiden Jets nicht diametral angeordnet sind, und der Transversalimpuls somit nicht ausgeglichen ist.

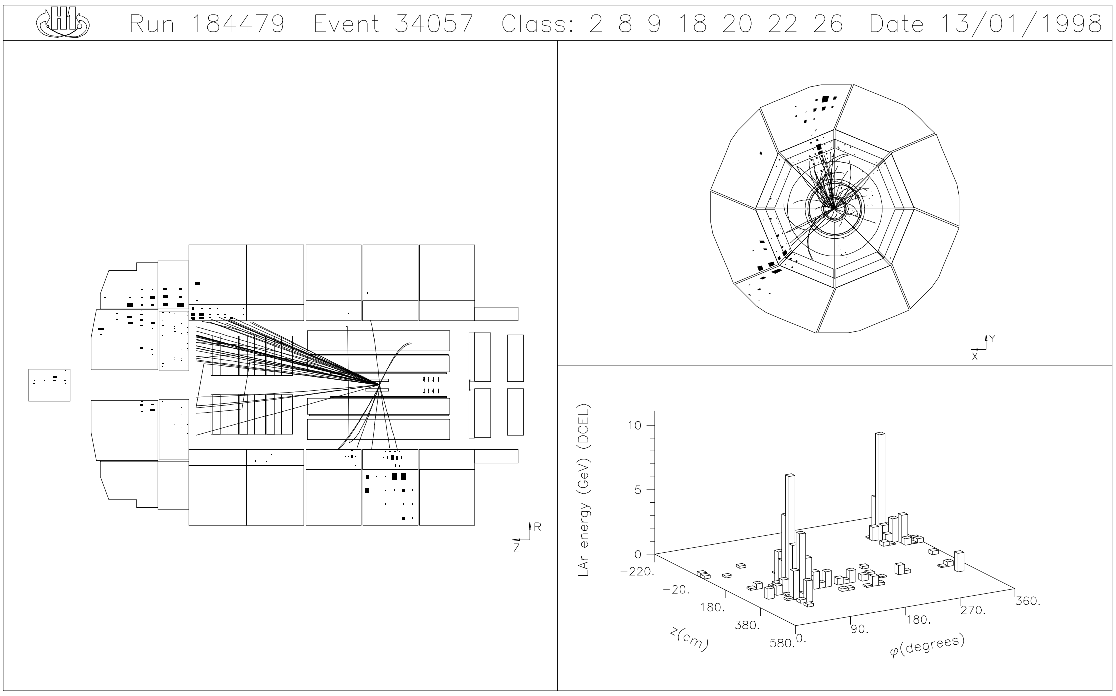


Abbildung 5.2: *Ein Ereigniskandidat des geladenen Stroms mit zwei Teilchenjets im H1-Detektor.*

5.2 Untergrundprozesse

Aufgrund von Fluktuationen in der Energiemessung oder durch inaktiven Bereichen des Kalorimeters treten auch bei anderen, balancierten Positron-Proton-Reaktionen, z.B. bei neutrale Strömen oder bei Ereignissen der Photoproduktion, unter Umständen erheblich von Null abweichende Werte des gemessenen Transversalimpulses auf. Ein weiterer wesentlicher Anteil der Untergrundereignisse entsteht durch die Wechselwirkung von Myonen der kosmischen Höhenstrahlung oder Myonen, die durch Reaktionen im Strahlrohr entstanden sind, im H1-Detektor.

5.2.1 Ereignisse mit Myonen

Myonen verlieren als geladene Leptonen beim Durchgang durch Materie einen Teil ihrer Bewegungsenergie durch elektromagnetische Wechselwirkung mit den Atomen. Der Energieverlust ist im wesentlichen auf Ionisation der Atome und auf Bremsstrahlung zurückzuführen. So kann z.B. ein Myon im elektromagnetischen Feld eines Atomkerns ein Photon emittieren, welches wiederum in einem Kernfeld ein $e^- e^+$-Paar bildet:

$$\begin{array}{l} \mu + N \rightarrow \gamma + \mu + N \\ \qquad\qquad\;\; \hookrightarrow e^- + e^+ \end{array}$$

Da sich der Wirkungsquerschnitt für Strahlungsprozesse proportional zum reziproken Massenquadrat verhält, ist der geringe Wirkungsquerschnitt der Myonen auf ihre hohe Masse ($m_\mu = 105.7$ MeV) zurückzuführen [Per90]. Damit ist die Produktion von Bremsstrahlungsphotonen bei Myonen um vier Größenordnungen gegenüber der bei Elektronen ($m_e = 0.51$ MeV) unterdrückt. Der Beitrag der Strahlungsverluste ist für Myonen mit Energien unterhalb einiger hundert GeV gegenüber den Ionisationsverlusten vernachlässigbar, erst bei höheren Energien (Kosmische Myonen) werden die Energieverluste durch Bremsstrahlung wichtig.

Halo-Myon-Ereignisse

Halo-Myonen entstehen durch die Wechselwirkung des Protonstrahls mit den Wänden des Strahlrohrs oder mit den Molekülen des Restgases im Strahlrohr. Dabei werden hochenergetische Pionen erzeugt, die in Myonen zerfallen. Der Entstehungsort der Halo-Myonen die Untergrundereignissen erzeugen können, konnte bei $z = -100$ m lokalisiert werden. Dort wechselwirken Protonen, die sich nicht auf der Sollbahn befinden mit der Strahlblende eines Magneten [Rie98]. Die entstehenden Myonen sind daher mit den Protonenpaketen zeitlich korreliert. Ein typisches Halo-Myon durchquert den H1-Detektor parallel zur z-Achse und deponiert auf seinem Weg Energie im Kalorimeter und im Myondetektor (vgl. Abbildung 5.3). Der Transversalimpuls eines solchen Ereignisses ist nicht ausgeglichen und es kann, wenn ein Vertex rekonstruiert wurde als ein Ereignis des geladenen Stroms fehlselektiert werden. Durch Halo-Myonen induzierte Ereignisse werden im folgenden als *Halos* bezeichnet.

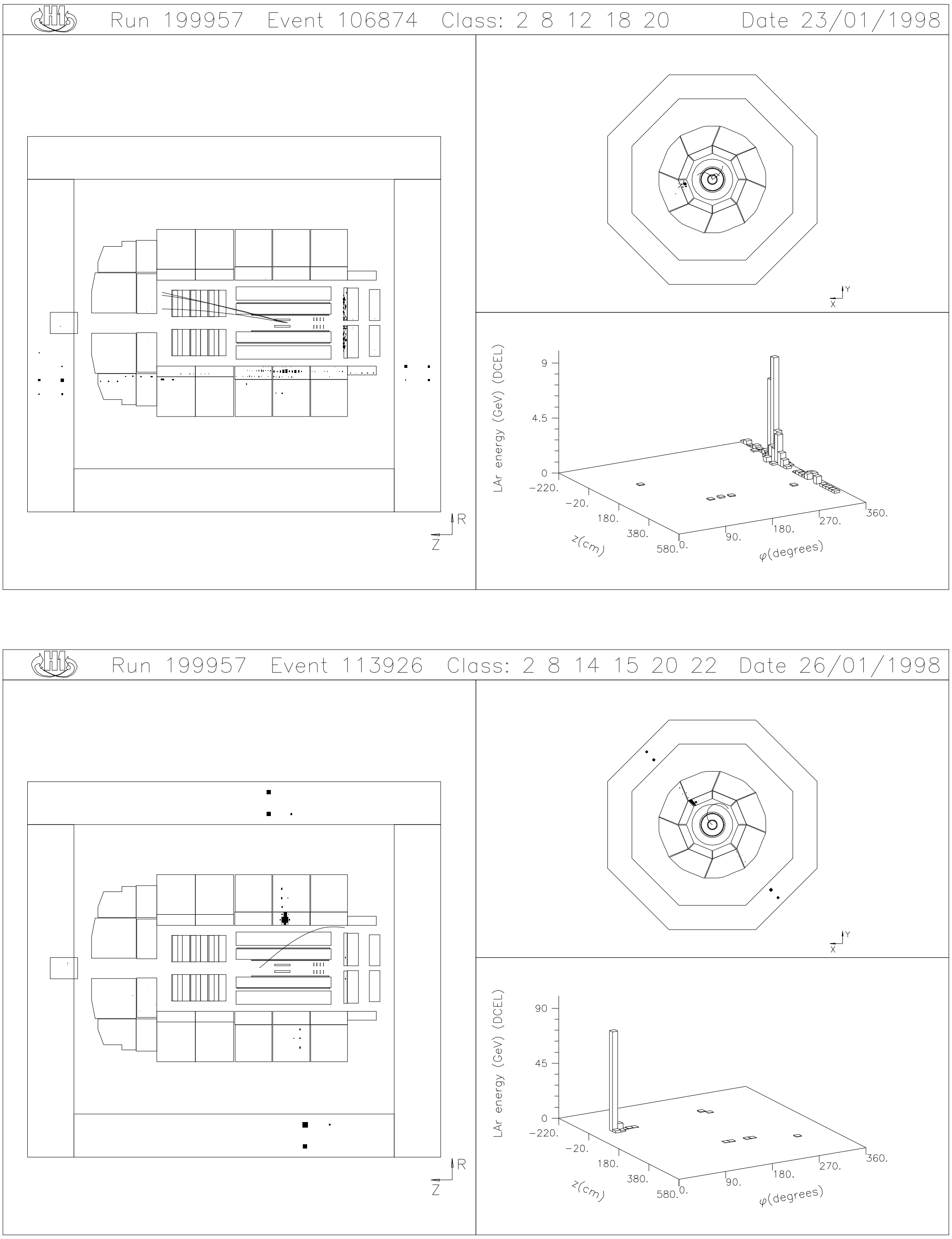


Abbildung 5.3: *Signatur eines Halo-Myons (oben) und eines Myons der kosmischen Strahlung (unten) im H1-Detektor. In der radialen Ebene erkennt man die asymmetrische Energiedeposition im Kalorimeter.*

Myonen der kosmischen Höhenstrahlung

Die zum größten Teil aus Protonen mit Energien bis zu 10^{12} GeV bestehende kosmische Strahlung erzeugt durch die Wechselwirkung mit Atomkernen in der Atmosphäre Pionen, welche in Myonen zerfallen. Diese können beim Durchqueren des H1-Detektors Energie deponieren. Kosmische Myonen erzeugen zumeist eine charakteristische Spur im instrumentierten Eisen und im Flüssig-Argon-Kalorimeter sowie in der zentralen Spurkammer, an der sie sich identifizieren lassen (vgl. Abbildung 5.3). Bei einer genügend asymmetrischen Energiedeposition können derartige Ereignisse als Reaktionen des geladenen Stroms fehlidentifiziert werden. Durch kosmische Myonen induzierte Ereignisse werden im folgenden als *Cosmics* bezeichnet.

Um Myon-Untergrund Ereignisse zu identifizieren, wurde eine Reihe von Algorithmen entwickelt, die sogenannten *Topologischen Filter* [Neg96]. Diese Programme nutzen die spezielle Topologie der Myon-Untergrund Ereignisse, z.B. gegenüberliegende Spuren im instrumentierten Eisen oder Ketten von Energiedepositionen in Kalorimeter aus, um sie von Ereignissen des geladenen Stroms zu trennen.

5.2.2 Ereignisse der Photoproduktion

Bei Ereignissen der direkten oder der aufgelösten Photoproduktion können Teilchenjets mit hohen transversalen Impulsen entstehen. Das Positron wird unter kleinem Winkel gestreut und kann für Streuwinkel von $\theta_l < 5$ mrad im Luminositätssystem des H1-Detektors nachgewiesen werden. Diese Einschränkung begrenzt den Bereich des Viererimpulsübertrags für γp-Ereignisse, die durch den Nachweis des gestreuten Positrons identifiziert werden können, auf Werte von $Q^2 < 0{,}01\text{GeV}^2$. Bei Ereignisse der Photoproduktion mit Viererimpulsüberträgen im Bereich $0{,}01\text{GeV}^2 < Q^2 < 4\text{GeV}^2$ ist das gestreute Positron im H1-Detektor nicht nachweisbar.

Ereignisse der direkten Photoproduktion zeichnen sich durch zwei Jets mit ausgeglichenem Transversalimpuls aus. Das gestreute Parton (im Fall der Boson-Gluon-Fusion) und der Protonrest tragen im allgemeinen einen kleinen Transversalimpuls und verlassen den Detektor durch das Strahlrohr (vgl. Abbildung 5.4).
Der Prozeß der aufgelösten Photoproduktion erzeugt durch die Hadronisierung des Photonrestes zusätzliche Teilchen.

Ereignisse der Photoproduktion können als solche des geladenen Stroms fehlerkannt werden, wenn ein nicht ausgeglichener Transversalimpuls gemessen wird und das gestreute Positron nicht nachgewiesen wird. Ein nicht-balancierter Transversalimpuls kann z.B. durch eine Fehlmessung der Energie im Kalorimeter oder durch den Verlust eines Teilchenjets im Strahlrohr entstehen.

Der differentielle Wirkungsquerschnitt $d\sigma/dE_\perp$ der Photoproduktion wurde im Bereich 7 GeV$< E_\perp <$ 30 GeV mit dem H1-Detektor gemessen. Es wurde folgende Abhängigkeit ermittelt [Leb96]:

$$\frac{d\sigma}{dE_{\perp,had}} \propto (E_{\perp,had})^{-n} \text{ mit } n = 6.1 \pm 0.5 \tag{5.1}$$

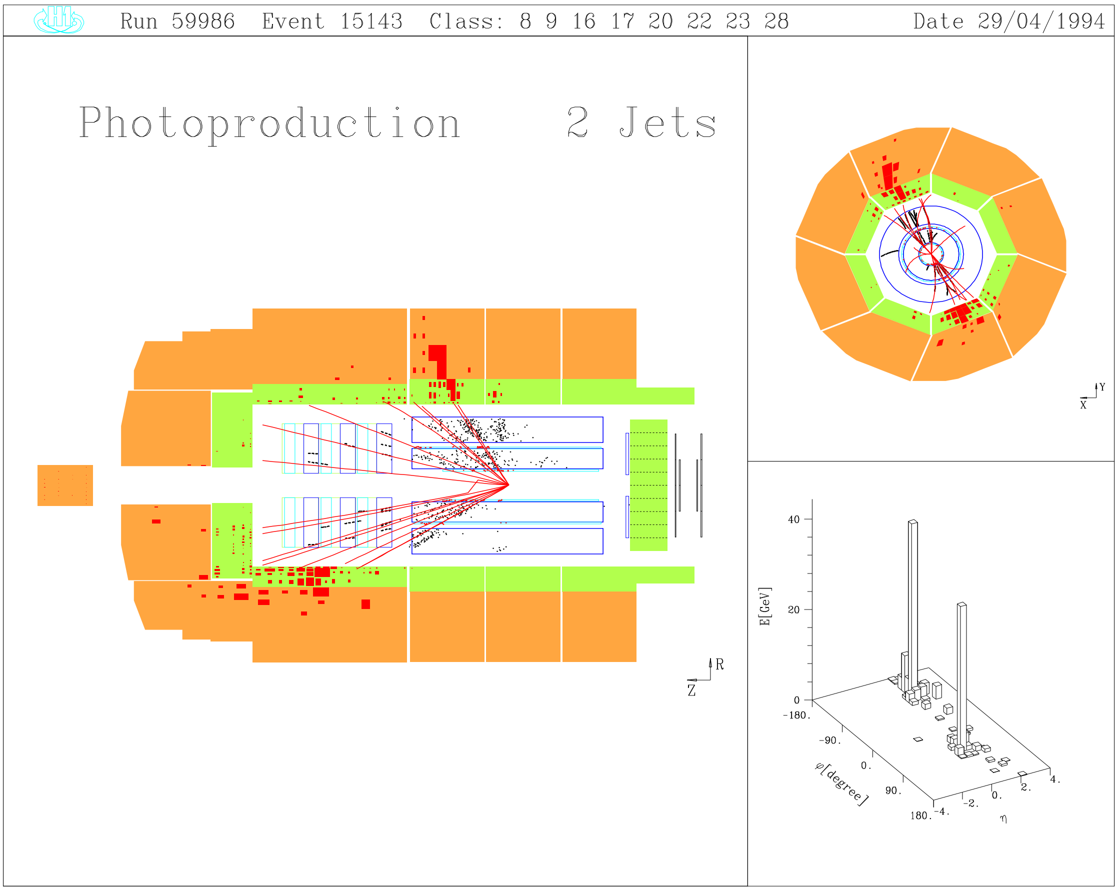


Abbildung 5.4: *Ein Ereignis der Photoproduktion mit zwei Jets. Man beachte die Symmetrie der Energieverteilung in der radialen Ebene.*

Da der Wirkungsquerschnitt der Photoproduktion mit abnehmenden $E_\perp$ stark anwächst, steigt gleichzeitig die Zahl der potentiellen Untergrundereignisse. Der Photoproduktionsuntergrund wird daher mit abnehmenden Transversalimpuls $p_\perp$ ansteigen.

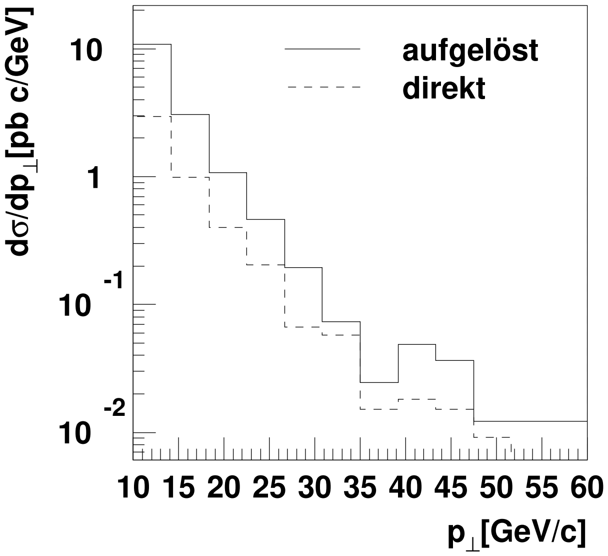


Abbildung 5.5: *Abhängigkeit des Wirkungsquerschnitts der direkten und der aufgelösten Photoproduktion vom hadronischen Transversalimpuls $p_\perp$, für Monte-Carlo-Datensätze.*

In Abbildung 5.5 ist die $p_\perp$-Abhängigkeit des Wirkungsquerschnitts der direkten und der aufgelösten Photoproduktion für Monte-Carlo-Datensätze dargestellt. Es wurden Ereignisse der Klasse 8 mit mindestens einem Jet mit $p_\perp > 10$ GeV ausgewählt. Der maximale Viererimpulsübertrag Q^2_{max} beträgt 4 GeV^2 und $y_{min} = 0.2$. Der Wirkungsquerschnitts wächst mit abnehmendem Transversalimpuls stark an.

Um den Untergrund durch Photoproduktion von Ereignissen des geladenen Stroms zu trennen, wurde bisher die Erfüllung bestimmter Bedingungen gefordert. So wurde bei den ersten Messungen des Wirkungsquerschnitts des geladenen Stroms am H1-Detektor ein Transversalimpuls von $p_\perp > 25$ GeV verlangt [H194][H195]. Abbildung 5.5 verdeutlicht, daß ein solcher Schnitt den größten Teil der Untergrundereignisse verwirft.
In folgenden Analysen wurde die untere Grenze im Transversalimpuls auf $p_\perp > 12.5$ GeV gesenkt [Mar97]. Um Ereignisse der Photoproduktion von denen des geladenen Stroms zu trennen, wurden spezielle Variable V_{ap} und V_p, welche die Art der Symmetrie der Energieverteilung in der radialen Ebene beschreiben eingeführt. Ereignisse des geladenen Stroms zeichnen sich durch eine asymmetrische Energieverteilung in der radialen Ebene aus (vgl. Abbildung 5.1). Im Gegensatz dazu, zeigen Photoproduktionsereignisse eine symmetrische radiale Energieverteilung (vgl. Abbildung 5.4). Eine ausgezeichnete Richtung in der radialen Ebene bei Ereignissen des geladenen Stroms, ist die Richtung

des Neutrinoimpulses, die entgegengesetzt zu $\mathbf{p}_\perp$ ist. Der hadronische Endzustand von Photoproduktionsereignissen besitzt keine ausgezeichnete Richtung. Diese Unterschiede werden zur Trennung der beiden Ereignissklassen ausgenutzt. Man definiert die Größen V_p (parallel) und V_{ap} (antiparallel):

$$\vec{V} = \sum_{Cluster\ i} \vec{V_i} = \sum_{Cluster\ i} \begin{pmatrix} E_i \sin(\vartheta_i)\sin(\phi_i) \\ E_i \sin(\vartheta_i)\cos(\phi_i) \end{pmatrix} \tag{5.2}$$

$$V_p = + \sum_{\substack{Cluster\ i \\ \vec{V}\cdot\vec{V_i}>0}} \frac{\vec{V}\cdot\vec{V_i}}{V} \tag{5.3}$$

$$V_{ap} = - \sum_{\substack{Cluster\ i \\ \vec{V}\cdot\vec{V_i}<0}} \frac{\vec{V}\cdot\vec{V_i}}{V} \tag{5.4}$$

Diese entsprechen einer Projektion der Cluster-Energie auf die durch $\vec{V}$ und den Vertex definierten Achse.

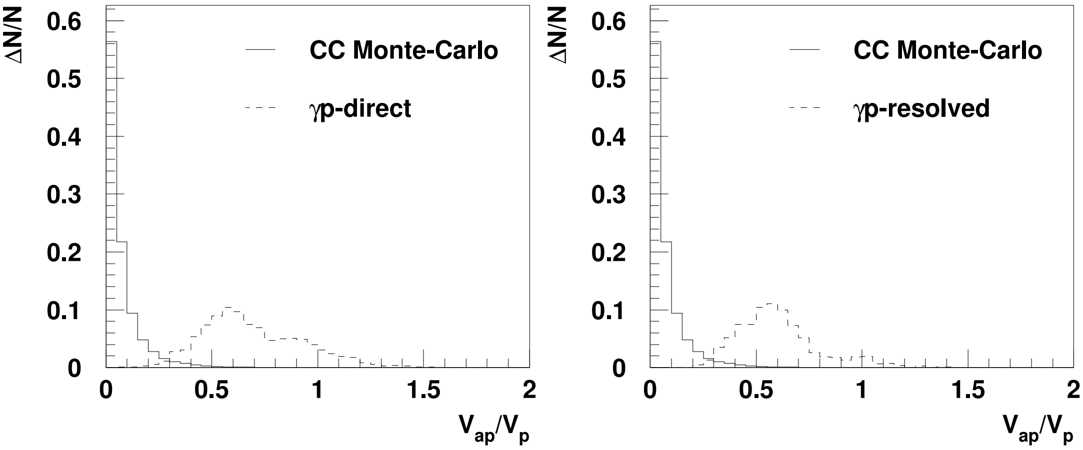


Abbildung 5.6: *Vergleich der relativen Zahl von Ereignissen der Photoproduktion und des geladenen Stroms in Abhängigkeit von* V_{ap}/V_p.

Die Abbildung 5.6 zeigt, daß anhand der V_{ap}/V_p-Verteilungen die Ereignisse der Photoproduktion zum größten Teil von denen des geladenen Stroms getrennt werden können. Eine vollständige Trennung ist aufgrund des Überlapps der Verteilungen nicht möglich. Bei der Messung des Wirkungsquerschnitts des geladenen Stroms mit den Daten der

Jahre 1994/95, wurde die Größe V_{ap}/V_p auf den Bereich $V_{ap}/V_p < 0.13$ beschränkt um den Untergrund durch Photoproduktion vollständig zu entfernen. Durch diese Einschränkung wird ein nicht zu vernachlässigender Teil der Ereignisse des geladenen Stroms verworfen (vgl. Abbildung 5.6). Der Anteil der verworfenen CC-Ereignisse konnte durch die Anwendung der V_{ap}/V_p-Einschränkung auf einen Monte Carlo Datensatz des geladenen Stroms abgeschätzt werden. Bei der Beschränkung von V_{ap}/V_p auf $V_{ap}/V_p < 0.13$ beträgt der Anteil der verworfenen CC-Ereignisse 13,6% und bei der Beschränkung auf $V_{ap}/V_p < 0.2$ beträgt er 6,3%.

5.2.3 Ereignisse des Neutralen Stroms

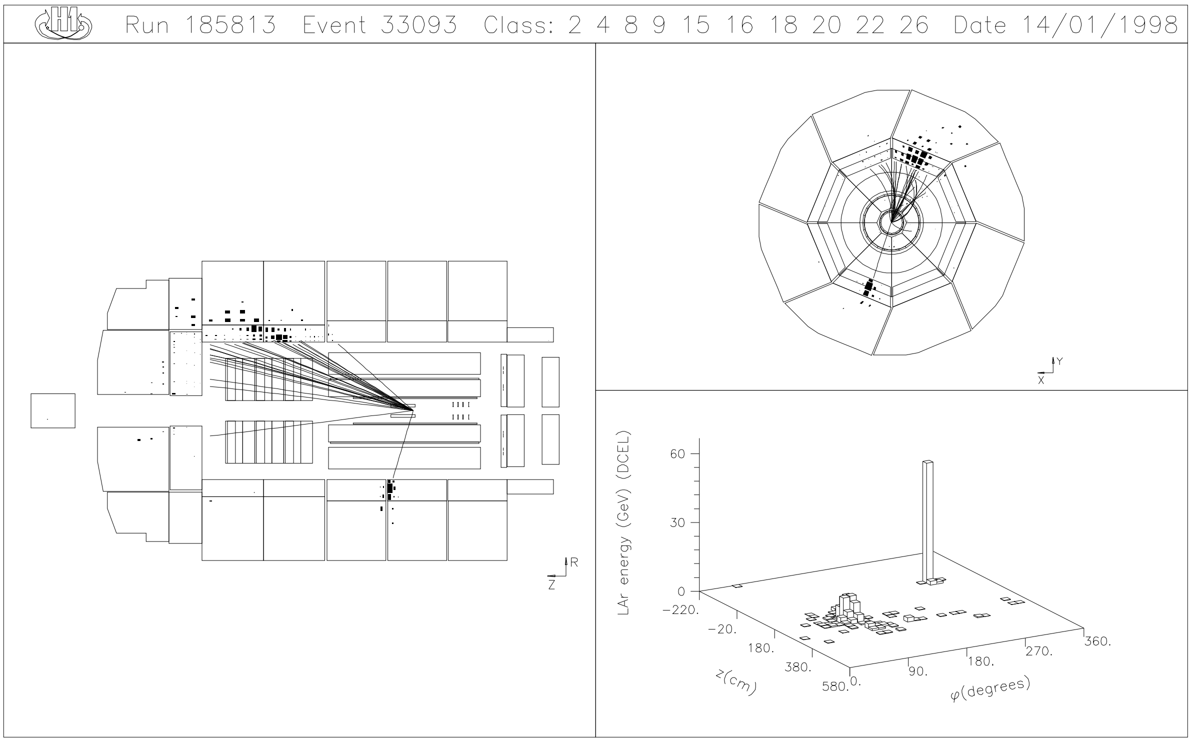


Abbildung 5.7: *Ein Ereigniskandidat des neutralen Stroms mit hadronischem Jet und gestreutem Positron. Zu erkennen ist die Symmetrie der Energieverteilung in der radialen Ebene des Kalorimeters.*

Ereignisse des neutralen Stroms zeichnen sich durch eine ausgeglichene Transversalimpulsbilanz und ein Positron im Endzustand aus. Abbildung 5.7 zeigt ein Ereignis des neutralen Stroms im H1-Detektor. Das Positron wird anhand einer isolierten Spur, die auf eine kompakte Energiedeposition im elektromagnetischen Teil des Kalorimeters weist, identifiziert. Der Transversalimpuls des gestreuten Positrons wird von dem des

hadronischen Endzustandes ausgeglichen.
Wird bei einem Ereignis des neutralen Stroms das gestreute Positron im Kalorimeter nicht nachgewiesen oder wird durch eine Fehlmessung der Energien ein unausgeglichener Transversalimpuls gemessen, kann es als ein Ereignis des geladenen Stroms fehlidentifiziert werden.
Ereignisse des neutralen Stroms bei denen kein Nachweis des gestreuten Positrons im Kalorimeter erfolgte, können durch die Analyse der Spuren identifiziert werden. Wird eine isolierte Spur in der zentralen Spurkammer rekonstruiert, und liegt diese dem hadronischen Endzustand in der radialen Ebene gegenüber, d.h. ist

$$|\varphi_{Spur} - \varphi_{had} - 180°| < 20°, \tag{5.5}$$

handelt es sich mit großer Wahrscheinlichkeit um ein Ereignis des neutralen Stroms.

5.2.4 Ereignisse weiterer Untergrundprozesse

- Bei H1 sind in den Jahren 1994-97 sechs Ereignisse mit einem isolierten Lepton (e^-,$\mu^\pm$) und hadronischen Teilchenjets mit unausgeglichenem Transversalimpuls im Endzustand beobachtet worden. Drei dieser Ereignisse sind innerhalb des Standardmodells mit der Produktion und anschließenden leptonischen Zerfall eines W Bosons verträglich. Entkommt das isolierte Lepton der Beobachtung im H1-Detektor, erscheint ein derartiges Ereignis als eines des geladenen Stroms. Bei der jetzigen Statistik kann dieser Untergrundprozeß vernachlässigt werden.

- Durch elektronisches Rauschen in den Zellen des Kalorimeters können Energiedepositionen, die die Triggerschwelle überschreiten vorgetäuscht werden. Kommt es zum Rauschen während eines Ereignisses, kann ein unausgeglichener gemessener Transversalimpuls erzeugt und die Bedingungen der Klasse 8 erfüllt werden. Es existieren Algorithmen zur Identifikation des Rauschens im Flüssig-Argon Kalorimeter [Alb95].

Zusammenfassend kann festgehalten werden: Bei der Selektion von Ereignissen des geladenen Stroms treten eine Reihe von Untergrundquellen auf. Mit Algorithmen (topologische Filter), oder durch die Beschränkung spezieller, den Prozeß charakterisierender Variablen (V_{ap}/V_p , $\varphi_{Spur} - \varphi_{had}$), ist es möglich, einen großen Teil der Untergundereignisse von den Ereignissen des geladenen Stroms zu trennen.

In dieser Arbeit soll untersucht werden, ob sich neuronale Netzwerke eignen, die einzelnen Untergrundklassen von der des geladenen Stroms zu trennen. Dazu werden in den folgenden Abschnitten eine Reihe von neuronalen Netzwerken entwickelt, getestet und ihre Effizienzen mit den bisherigen Methoden zur Separation der Untergrundprozesse verglichen.

Kapitel 6

Entwicklung der Netzwerke

6.1 Auswahl der Ereignisse für Training und Test

Um neuronale Netzwerke zu entwickeln, die in der Lage sind, Ereignisse des geladenen Stroms von Untergrundereignissen zu trennen, wird im ersten Schritt die Auswahl der Trainings- und Testdaten getroffen. Es werden vier verschiedene Netze aufgestellt um CC-Ereignisse von den Ereignissen jeweils einer Untergrundklasse (NC, γp, Halos, Cosmics) zu trennen. Zum Training der Netze wurden Monte-Carlo-Datensätze verwendet. Für das Training der Netze, die durch Halo-Myonen bzw. kosmische Myonen induzierte Ereignisse erkennen sollen, werden Daten verwendet, die mit den Halo- und Cosmic-Filtern selektiert wurden. Es existieren keine geeigneten Monte-Carlo-Generatoren, die in der Lage sind Myon-Ereignisse der Klasse 8 mit hoher Statistik zu generieren.

Da der CC-Trigger des H1-Detektors nur für Transversalimpulse von $p_{\perp,had} > 12{,}5$ GeV eine ausreichende Effizienz ($\epsilon_{CC} \geq 45\%$) erreicht, werden zum Training und Test der Netze Ereignisse ausgewählt, die diese Bedingung erfüllen.

Es wurden Ereignisse mit einem gemessenen Hadronwinkel θ_{had}, welcher innerhalb des vom Flüssig-Argon-Kalorimeter überdeckten Winkelbereichs $4^\circ < \theta_{had} < 126^\circ$ lag ausgewählt.

Die zum Training und Test ausgewählten CC- und NC-Datensätze sind in der Tabelle 6.1 aufgeführt. Um ein “Auswendiglernen” der mit dem H1-Detektor beobachteten Ereignissen des geladenen Stroms durch die Netze zu verhindern, wurden sie ausschließlich mit Monte-Carlo-CC-Ereignissen trainiert.

Eine von Monte-Carlo-Generatoren unabhängige Methode zum Test der Netze besteht in der Verwendung von sogenannten Pseudo-CC-Ereignissen. Diese Methode nutzt die Ähnlichkeit der Endzustände der Prozesse des geladenen- und des neutralen Stroms bei gleicher Ereigniskinematik aus. Die einzigen Unterschiede zwischen diesen Prozessen besteht im gestreuten Lepton und der relativen Häufigkeit der Ereignisse. Bei NC-Ereignissen kann das Positron im Detektor nachgewiesen werden. Entfernt man in den aufgezeichneten Daten diese Signale aus allen Detektorteilen, erhält man ein Pseudo-CC-Ereignis, welches nicht mehr von einem wirklichen Ereignis des geladenen Stroms zu unterscheiden ist.

Generator / Daten	Selektion	Luminosität	Ereignisse	Bezeichnung
Django 6 CC MRSH $Q^2 > 100$ GeV^2	$p_{\perp,had} > 12{,}5$ GeV $4^\circ < \theta_{had} < 126^\circ$ Klasse 8	1421,242 pb^{-1}	Training u. Test 32779	CCMC
Pseudo CC aus den Daten 1994/95	$p_{\perp,had} > 12{,}5$ GeV $4^\circ < \theta_{had} < 126^\circ$ Klasse 8	6,57 pb^{-1}	nur Test 7875	PSCC
H1-Daten CC 1995/96/97 $p_{\perp,had} > 12{,}0$ GeV $V_{ap}/V_p < 0{,}2$ $y > 0{,}05$	$p_{\perp,had} > 12{,}5$ $4^\circ < \theta_{had} < 126^\circ$	~ 35 pb^{-1}	nur Test 648	CCDATA
Django 6.1 NC $Q^2 > 100$ GeV^2	$p_{\perp,had} > 12{,}5$ GeV $4^\circ < \theta_{had} < 126^\circ$	23,45 pb^{-1}	Training u. Test 8397	NCMC

Tabelle 6.1: *Die für Training und Test der Netze verwendeten CC- und NC-Daten*

Zum Training und Test der neuronalen Netze mit Ereignissen der Photoproduktion als Untergrund werden Monte-Carlo-Daten verwendet, die mit dem Generator Pythia 5.7 erzeugt wurden. Mit diesem Generator ist es möglich Ereignisse der direkten- und aufgelösten Photoproduktion zu generieren. Die Statistik von γp-Ereignissen mit einem Transversalimpuls von $p_{\perp,had} > 12{,}5$ GeV, wurde durch die Verwendung mehrerer Monte-Carlo-Datensätze verbessert. Fordert man die Zugehörigkeit zur Ereignisklasse 8 für γp-Ereignisse, so nimmt die Statistik stark ab und die Anzahl von Trainingsmustern wird zu gering. Diese Bedingung an die Ereignisse wurde daher nur beim Test der Netze gefordert. Die Tabelle 6.2 gibt einen Überblick über die verwendeten Datensätze.

Die Selektion der Ereignisse, die durch kosmische- und Halo-Myonen induziert wurden, erfolgte mit den topologischen Untergrund-Filtern. Es wurden Ereignisse der Klasse 8 der Datennahmeperiode des Jahres 1996 analysiert. In der Tabelle 6.3 findet man die resultierenden Datensätze und die verwendeten Filter.

Generator	Selektion	Luminosität	Ereignisse	Bezeichnung
Pythia 5.7 aufgelöst, 2 Jets GRV-LO $Q^2 < 4$ GeV^2 $y > 0{,}1$ $p_{\perp,had} > 1{,}2$ GeV	$p_\perp > 12{,}5$ GeV $4^\circ < \theta_{had} < 126^\circ$ Klasse 8	11,10 pb^{-1}	Training u. Test 356	γpres1
Pythia 5.7 direkt GRV-LO $Q^2 < 4$ GeV^2 $y > 0{,}1$ $p_{\perp,had} > 1{,}2$ GeV	$p_\perp > 12{,}5$ GeV $4^\circ < \theta_{had} < 126^\circ$	11,10 pb^{-1}	Training u. Test 296	γpdir1
Pythia 5.7 aufgelöst, 2 Jets GRV-G LO $Q^2 < 4$ GeV^2 $y > 0{,}2$ $p_{\perp,had} > 4$ GeV	$p_\perp > 12{,}5$ GeV $4^\circ < \theta_{had} < 126^\circ$	50,0 pb^{-1}	Training u. Test 2462	γpres2
Pythia 5.7 direkt, 2 Jets GRV-G LO $Q^2 < 4$ GeV^2 $y > 0{,}2$ $p_{\perp,had} > 4$ GeV	$p_\perp > 12{,}5$ GeV $4^\circ < \theta_{had} < 126^\circ$	50,0 pb^{-1}	Training u. Test 22	γpdir2
Pythia 5.7 aufgelöst GRV-G LO $Q^2 < 4$ GeV^2 $y > 0{,}2$, $p_{\perp,Jet} > 10$ GeV $p_{\perp,had} > 7$ GeV	$p_\perp > 12{,}5$ GeV $4^\circ < \theta_{had} < 126^\circ$ Klasse 8	20,0 pb^{-1}	nur Test 400	γpres3
Pythia 5.7 direkt GRV-G LO $Q^2 < 4$ GeV^2 $y > 0{,}2$, $p_{\perp,Jet} > 10$ GeV $p_{\perp,had} > 12$ GeV	$p_\perp > 12{,}5$ GeV $4^\circ < \theta_{had} < 126^\circ$ Klasse 8	80,2 pb^{-1}	nur Test 527	γpdir3

Tabelle 6.2: *Tabelle der zum Training und Test verwendeten γp-Ereignisse*

Daten	Selektion	Luminosität	Ereignisse	Bezeichnung
1996 Halo Filter: HALAR HAMULAR [Neg96]	$p_{\perp,had} > 12{,}5$ GeV $4° < \theta_{had} < 126°$ Klasse 8	8,15 pb^{-1}	Training u. Test 7807	HALO96
1996 Cosmic Filter: COSMUMU COSMULAR COSTALAR COSTRACK [Neg96]	$p_{\perp,had} > 12{,}5$ GeV $4° < \theta_{had} < 126°$ Klasse 8	8,15 pb^{-1}	Training u. Test 8091	COSM96

Tabelle 6.3: *Die mit den topologischen Filtern selektierten Trainings und Testdaten*

6.2 Funktionsweise der Programme

In dieser Arbeit werden zwei Programme zur Entwicklung der neuronalen Netze verwendet:

- das Programm *JETNET 3.1* der Universität Lund [Lon94] zum Training und Test der Netze. Dieses in der Programmiersprache Fortran geschriebene Programm stellt in einer Bibliothek die benötigten Trainingsalgorithmen (s. Abschnitt 4.3.3) und Werkzeuge zur Definition der Netztopologie zur Verfügung.
- das Programm *CROSSOVER*, das im Rahmen dieser Arbeit zur Optimierung der Netzwerke entwickelt wurde. Dieses Programm soll unter Anwendung des in Abschnitt 4.4.2 beschriebenen genetischen Algorithmus die optimale Kombination der Eingabevariablen und Netzparameter finden.

In der Abbildung 6.1 ist als Beispiel der Ablauf beim Training eines Netzes zur Trennung von Ereignissen der Photoproduktion und des geladenen Stroms dargestellt. Nach der in Abschnitt 6.1 beschriebenen Auswahl der Datensätze, werden die Eingabevariablen berechnet. Die Menge aller Eingabevariablen eines Ereignisses werden als dessen Eingabemuster, und die Menge der Eingabemuster als Eingabemustersatz bezeichnet. Im folgenden Schritt werden die Eingabemustersätze der beiden Ereignissklassen (CC, γp), in jeweils einen Trainings- und einen Testmustersatz aufgeteilt.

Die zwei Trainingsmustersätze werden in einem Zufallsprozeß gemischt und zu einem Satz zusammengefaßt. Die Zugehörigkeit der einzelnen Eingabemuster zur jeweiligen Ereignisklasse bleibt dabei erhalten. Es hat sich gezeigt, daß durch die Mischung der Trainingsmuster ein effizienteres Lernen ermöglicht wird.

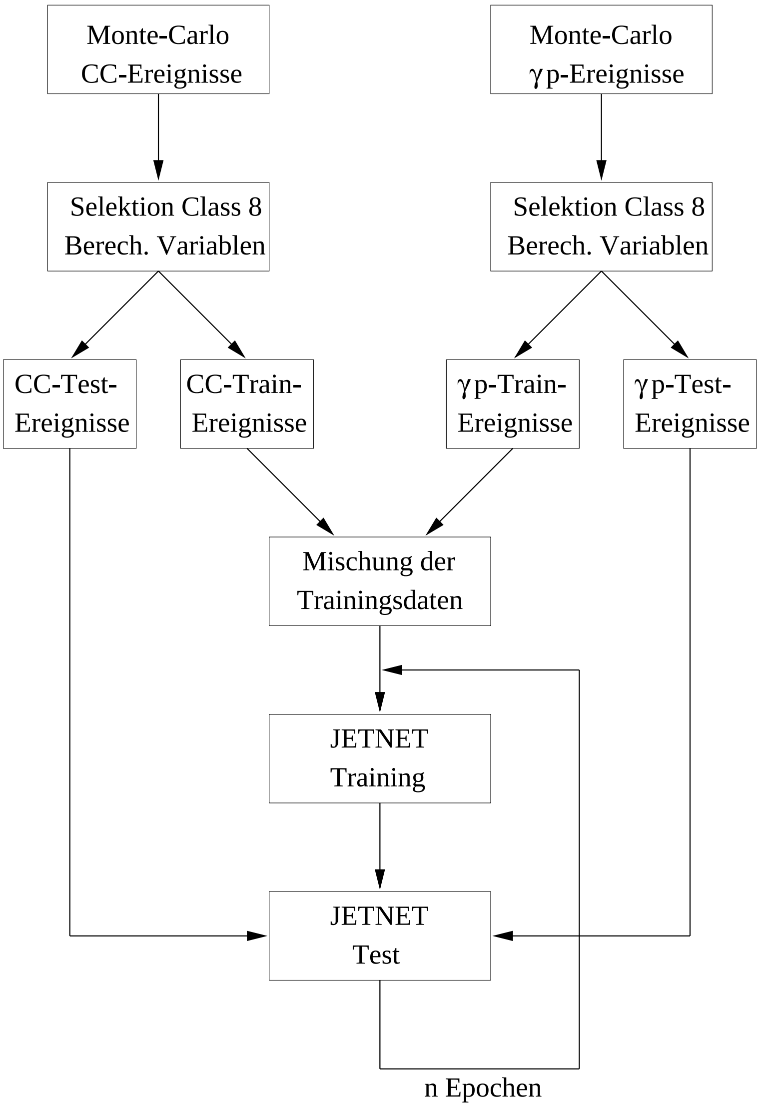


Abbildung 6.1: *Der Programmablauf beim Training, am Beispiel eines Netzes zur Trennung von Ereignissen der Photoproduktion und des geladenen Stroms.*

Mit dem entstandenen Trainingsmustersatz wird, nachdem die Topologie und der Trainingsalgorithmus festgelegt wurden, das neuronale Netz trainiert. Das Training selbst ist in zeitliche Abschnitte, die sogenannten *Epochen* unterteilt. Innerhalb einer Epoche wird dem Netz der gesamte Trainingsmustersatz präsentiert. Nach einer bestimmten Anzahl von Epochen (in dieser Analyse 10 Epochen) wird das Netz mit den beiden unabhängigen Testmustersätzen getestet und die mittlere Effizienz $\bar{\varepsilon}$ bestimmt:

$$\bar{\varepsilon} = \frac{\varepsilon_{CC} + \varepsilon_{\gamma p}}{2} \quad \text{mit} \quad \varepsilon_{CC} = \frac{N_{Netz(CC)}}{N_{CC}} \quad \text{und} \quad \varepsilon_{\gamma p} = \frac{N_{Netz(\gamma p)}}{N_{\gamma p}}. \tag{6.1}$$

Dabei sind $N_{Netz(\gamma p)}$, $N_{Netz(CC)}$ die Anzahl der vom Netz als γp bzw. CC erkannten Ereignisse.
Das Training wird abgebrochen, sobald keine weitere Steigerung der mit den unabhängigen Testmustersätzen bestimmten mittleren Effizienz erreicht werden kann. Als Resultat werden die Topologie, die Parameter und die Gewichte des trainierten Netzes gesichert.

Der erste Schritt der Realisation des evolutionären Algorithmus (Abschnitt 4.4.2) zur Optimierung der neuronalen Netze, besteht in der Auswahl der in den Chromosomen (Tupel $\mathbf{g}$) der Netze enthaltenen Informationen über die Netztopologie und Parameter:

- Die Auswahl der verwendeten Eingabevariablen
- Die Anzahl der verborgenen Ebenen
- Die Zahl der Knoten Kn_1,Kn_2 in den verborgenen Ebenen
- Die Temperatur T
- Das Intervall der initialen Gewichte $[-w_{init},w_{init}]$ (Die Gewichte werden vor Beginn des Trainings auf zufällige Werte $w_{jk} \in [-w_{init},w_{init}]$ gesetzt.)

Die Tupel $\mathbf{g}$ sollen als Binärstrings realisiert werden. Die ausgewählten Informationen werden daher ins Binärsystem übersetzt. Die Zahlen T, Kn_1, Kn_2, w_{init} werden im Binärcode dargestellt. Durch einleitende Untersuchungen konnten die Wertebereiche der Parameter T und w_{init} eingeschränkt werden. Es wurde beobachtet, daß Netze mit einer inversen Temperatur im Bereich $0 < T^{-1} < 3{,}1$ und einem initialen Gewicht $0 < w_{init} < 0{,}3$ die besten Resultate erzielten. Zur Darstellung der Parameter, innerhalb der ausgewählten Intervallgrenzen werden im Tupel für T^{-1} fünf und für w_{init} vier Bit benötigt.

Die maximale Anzahl der verborgenen Ebenen und die Zahl der Knoten in den Ebenen wird durch die benötigte Rechenzeit während des Lernprozesses begrenzt. Um zwei Klassen innerhalb eines n-dimensionalen Raumes durch geschlossene Grenzen zu trennen, werden mindestens $n + 1$ Knoten in den verborgenen Ebenen benötigt. Jeder Knoten in einer verborgenen Ebene grenzt die Klassen durch eine $n - 1$-dimensionale Hyperebene ab. Werden als Eingabegrößen des Netzwerkes n Variablen verwendet, und soll eine vollständige Trennung der Ereignisklassen erreicht werden, so muß die minimale Anzahl von verborgenen Knoten $n + 1$ betragen. Es hat sich gezeigt, daß ein Netz mit zwei verborgenen Ebenen weniger Knoten benötigt, als ein Netz mit nur einer verborgenen Ebene und gleicher Effizienz. Die maximale Anzahl verborgenen Ebenen wurde auf zwei und die maximale Knotenzahl pro Ebene auf 31 begrenzt (es werden maximal 25 Eingabegrößen verwendet). Zur Kodierung der Knotenanzahl werden jeweils fünf Bit benötigt. Die Anzahl der verborgenen Ebenen wird nicht separat kodiert, da ein Netz mit einer verborgen Schicht durch ein Netz mit zwei verborgenen Schichten ersetzt werden kann, wobei die zweite Schicht aus nur einem Knoten besteht. Der einzelne Knoten dient in diesem Fall als Transfereinheit.

Zur Kodierung der Auswahl der verwendeten Eingabevariablen wird pro mögliche Variable ein Bit benötigt. Ist das zur Variablen gehörige Bit gesetzt, wird die Variable im Netz verwendet, andernfalls nicht.
Das gesamte binäre Tupel $\mathbf{g}$ stellt sich folgendermaßen dar:

$$\mathbf{g} = \underbrace{5\ Bit}_{Kn_1}\underbrace{5\ Bit}_{Kn_2}\ \underbrace{25\ Bit}_{Eingabevariablen}\ \underbrace{5\ Bit}_{T^{-1}}\underbrace{4\ Bit}_{w_{init}}$$

Mit den so definierten Tupeln der neuronalen Netze kann der kanonische Genetische Algorithmus (vgl. Abschnitt 4.4.2) angewendet werden.

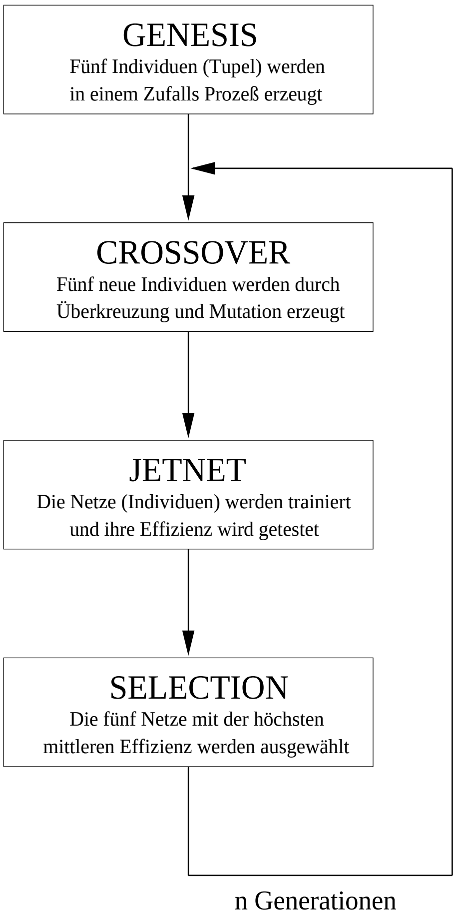


Abbildung 6.2: *Darstellung der einzelnen im Opimierungsprozeß ablaufenden Programmschritte*

In der Abbildung 6.2 ist der Programmablauf der Optimierung eines neuronalen Netze zur Trennung zweier Ereignissklassen dargestellt:

- Im ersten Programmteil *GENESIS* werden fünf Tupel $\mathbf{g}_i$ erzeugt. Für jedes Bit im Tupel wird durch ein Zufallsexperiment entschieden, ob es gesetzt werden soll oder nicht.

- Aus diesen fünf zufällig erzeugten Tupeln werden im folgenden Programmteil *CROSSOVER* durch die im Abschnitt 4.4.2 definierten Operatoren, Überkreuzung und Mutation, fünf zusätzliche Individuen erzeugt. Die Rate der Mutation wird durch den einzigen freien Parameter im Algorithmus, die sogenannte *Mutationswahrscheinlichkeit* p_m bestimmt.
- Für jedes der zehn Tupel wird im Programmabschnitt *JETNET* ein zugehöriges Netz mit den in $\mathbf{g}_i$ enthaltenen Parametern initialisiert, trainiert und dessen mittlere Effizienz bestimmt.
- Anschließend werden die fünf Netze mit den höchsten mittleren Effizienzen innerhalb des Programmteils *SELECTION* selektiert und die zugehörigen Tupel als "Eltern" einer neuen Generation ausgewählt. Die Fitnessfunktion im Algorithmus ist die mittlere Effizienz $\bar{\varepsilon}$:

$$\bar{\varepsilon} = \bar{\varepsilon}(\mathbf{g}) = \bar{\varepsilon}(Kn_1, Kn_2, Eingabevar., T, w_{init}). \tag{6.2}$$

Der Optimierungsprozeß wird beendet, falls innerhalb von 20 Generationen keine Steigerung der mittleren Effizienz erzielt wurde.

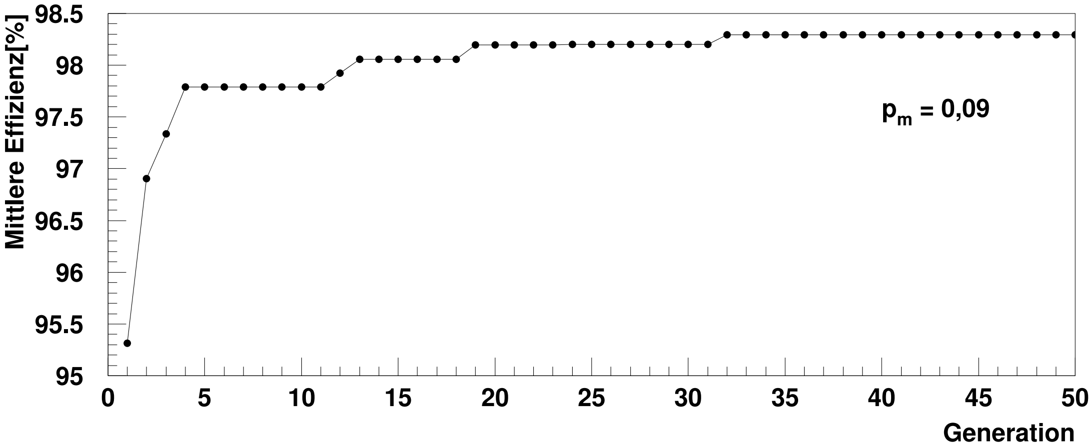


Abbildung 6.3: *Anwachsen der mittleren Effizienz mit der Anzahl der Generationen am Beispiel eines Netzes zur Trennung von Halo- und CC-Ereignissen.*

Am Beispiel der Optimierung eines Netzes zur Trennung von Halo- und CC-Ereignissen zeigt die Abbildung 6.3 das Anwachsen der mittleren Effizienz mit der Zahl von Generationen.
Die bereits in der ersten Generation erreichte hohe Effizienz, erklärt sich durch die erläuterte Beschränkung der in den Tupeln enthaltenen Parameter auf Werte, die große resultierende Effizienzen erwarten lassen. Außerdem wurde eine Vorauswahl der Menge der möglichen Eingabevariablen getroffen (mehr dazu im folgender Abschnitt).

Nachdem der Optimierungsalgorithmus innerhalb der ersten 15 Generation die mittlere Effizienz um 3% verbessert hat, findet keine wesentliche Steigerung mehr statt. Innerhalb der ersten Generationen findet der Algorithmus möglicherweise mehrere lokale Maxima der mittleren Effizienz $\bar{\varepsilon}(\mathbf{g})$. Nach dem das größte Maximum lokalisiert wurde, nähert sich in den folgenden Schritten der Algorithmus dessen Scheitelpunkt, wobei nur kleine weitere Steigerungen der Effizienz erzielt werden. Ein möglicher Ablauf der Suche nach dem globalen Maximum der mittleren Effizienz $\bar{\varepsilon}$, ist in der Abbildung 6.4 dargestellt.

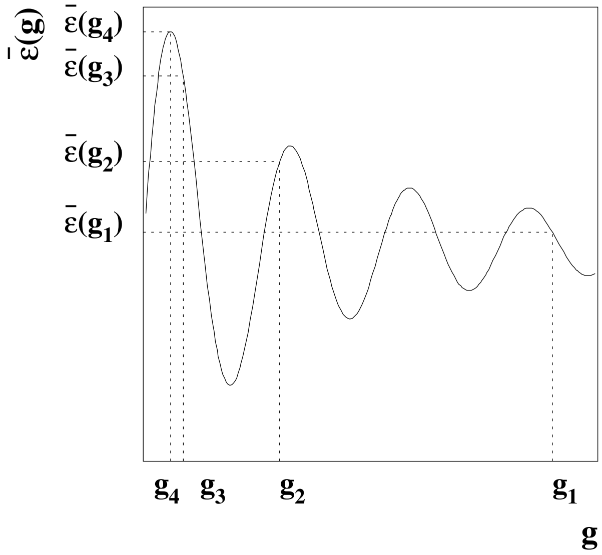


Abbildung 6.4: *Dargestellt ist eine hypothetische mittlere Effizienz* $\bar{\varepsilon}(\mathbf{g})$ *mit mehreren lokalen Maxima. Die Tupel* $\mathbf{g}$ *werden in dieser Abbildung durch Werte* g *auf der Abszissenachse symbolisiert. Der Wert* g_i *repräsentiert das Tupel mit der höchsten Effizienz in der i-ten Generation.*

Nachdem im letzten Abschnitt die Funktionsweisen der Programme zum Training und zur Optimierung neuronaler Netze dargestellt wurden, wird im folgenden Teil eine Beschreibung der als Eingabevariablen ausgewählten Größen gegeben.

6.3 Auswahl der möglichen Eingabevariablen

Mit dem im vorigen Abschnitt dargestellten genetischen Algorithmus ist es prinzipiell möglich, die geeigneten Eingabevariablen eines neuronalen Netzes aus allen zur Verfügung stehenden auszuwählen. Um die vom genetischen Algorithmus benötigte Rechenzeit zu beschränken, wird in dieser Analyse eine Vorauswahl der Eingabevariablen vorgenommen. Es wurden im Allgemeinen solche Variablen ausgewählt, mit denen in früheren Analysen eine teilweise Separierung der Ereignisklassen erzielt werden konnte. Eine Auswahl der verwendeten Variablen beruht auf dem Vergleich der, mit der Spurkammer bzw. mit dem LAr-Kalorimeter des H1-Detektors bestimmten, Kinematik eines Ereignisses. Voruntersuchungen im Rahmen dieser Arbeit haben gezeigt, daß derartige Variablen zur teilweisen Trennung der Ereignisklassen geeignet sind.

Zur Rekonstruktion der Ereigniskinematik können die Informationen des Spurkammersystems, des Kalorimeters oder beider in Kombination verwendet werden. Aus der Position $\vec{r}_i = (x_i, y_i, z_i)$ der Kalorimeterzelle i in der die Energie E_i über der Rauschschwelle deponiert wurde und der Position des Wechselwirkungspunktes $\vec{r}_{Vertex}$ lassen sich unter Vernachlässigung der zugehörigen Masse die einzelnen Komponenten des Impulses berechnen:

$$p_{x,i} = \frac{E_i \cdot (x_i - x_{Vertex})}{d_i} \quad p_{y,i} = \frac{E_i \cdot (y_i - y_{Vertex})}{d_i} \quad p_{z,i} = \frac{E_i \cdot (z_i - z_{Vertex})}{d_i} \tag{6.3}$$

$d_i = \sqrt{(x_i - x_{Vertex})^2 + (y_i - y_{Vertex})^2 + (z_i - z_{Vertex})^2}$ ist der Abstand der Zelle i vom Wechselwirkungspunkt. Der zur Zelle i gehörige Vierervektor ist somit $(E_i, p_{x,i}, p_{y,i}, p_{z,i})$.

In der *Clustermethode* werden benachbarte Zellen im Kalorimeter, in denen die Energiedeposition die Schwelle überschritten hat, zu Gruppen, sogenannte *Cluster* zusammengefaßt. Die Vierervektoren der einzelnen Zellen des Clusters werden addiert und der resultierende Vierervektor als der des Clusters betrachtet. Einzelne Cluster bzw. Gruppen von ihnen können mit der Energiedeposition einzelner Teilchen identifiziert werden. So erzeugt der kompakte elektromagnetische Schauer, den ein gestreutes Positron auslöst, in der Regel einen Cluster. Hadronische Schauer können aus mehreren Clustern bestehen, da sie eine größere Ausdehnung im Kalorimeter besitzen.

Durch die mit den zentralen- und Vorwärts Spurkammern beobachteten Spuren der geladenen Teilchen im Endzustand eines Ereignisses, kann der Anteil der geladenen Teilchen an der Kinematik analysiert werden. Im ersten Schritt wird mit den Informationen der Proportionalkammern CIP, COP und FPC die Position des Wechselwirkungspunktes bestimmt. Im folgenden werden nur solche Spuren betrachtet die mit dem Vertex verbunden werden können. Im magnetischen Feld der supraleitenden Spule bewegen sich geladene Teilchen auf einer Helix, die durch einen Satz von Parametern beschrieben werden kann. Durch einen Anpassungsprozeß werden die Parameter der Teilchenspuren bestimmt. Aus der so gewonnen Parametrisierung einer Spur, wird der zugehörige Impulsvektor des Teilchens berechnet [Ste97]. Der Nachteil dieser Methode besteht darin, daß ungeladene Teilchen nicht mit den Spurkammern nachgewiesen werden (so ist z.B. $p_{\perp,Spur} \lesssim \frac{2}{3} p_{\perp,Cluster}$).

Im folgenden werden die verwendeten Eingabevariablen vorgestellt. Es sind dies:

- Die mit der Cluster- bzw. Spurmethode bestimmten transversalen Impulse $p_{\perp,Cluster}$ und $p_{\perp,Spur}$ (Gleichung 2.8).
- Das Verhältnis der beiden vorgenannten Größen $p_{\perp,Cluster}/p_{\perp,Spur}$.
- Der aus der Energiedeposition im instrumentierten Eisen bestimmte transversale Impuls $p_{\perp,Eisen}$.
- Die mit der Clustermethode bestimmte gesamte bzw. transversale Energie des hadronischen Endzustandes E_{had} und $E_{\perp,had}$ (Gleichung 2.9).
- Der nach der Jaquet Blondel Methode bestimmte Energieübertrag des Positrons auf das hadronische System y_{JB} (Gleichung 2.5).
- Die mit der Cluster- bzw. Spurmethode bestimmten Polar- und Azimutwinkel θ_h und φ_h des gestreuten Partons:

$$\theta_{h,Cluster} = \arccos\left(\frac{p^2_{\perp,Cluster} - (E_{had} - p_{z,Cluster})^2}{p^2_{\perp,Cluster} + (E_{had} - p_{z,Cluster})^2}\right) \tag{6.4}$$

$$\varphi_{h,Cluster} = \arctan\left(\frac{p_{y,Cluster}}{p_{x,Cluster}}\right) \tag{6.5}$$

$$\theta_{h,Spur} = \arccos\left(\frac{p^2_{\perp,Spur} - (E_{had} - p_{z,Spur})^2}{p^2_{\perp,Spur} + (E_{had} - p_{z,Spur})^2}\right) \tag{6.6}$$

$$\varphi_{h,Spur} = \arctan\left(\frac{p_{y,Spur}}{p_{x,Spur}}\right). \tag{6.7}$$

- Der mit den Energiedepositionen im instrumentierten Eisen berechnete Polarwinkel $\theta_{h,Eisen}$ des gestreuten Partons:

$$\theta_{h,Eisen} = \arccos\left(\frac{p^2_{\perp,Eisen} - (E_{had} - p_{z,Eisen})^2}{p^2_{\perp,Eisen} + (E_{had} - p_{z,Eisen})^2}\right). \tag{6.8}$$

- Die Differenzen der mit den einzelnen Methoden berechneten Polarwinkel des gestreuten Partons:

$$\Delta\theta_h = \theta_{h,Cluster} - \theta_{h,Spur} \tag{6.9}$$

$$\Delta\theta_{h,Eisen-Spur} = \theta_{h,Eisen} - \theta_{h,Spur} \tag{6.10}$$

$$\Delta\theta_{h,Eisen-Cluster} = \theta_{h,Eisen} - \theta_{h,Cluster}. \tag{6.11}$$

- Die Differenz der mit der Cluster- bzw. Spurmethode berechneten Azimutwinkel des gestreuten Partons:

$$\Delta\varphi_{h,Cluster-Spur} = \varphi_{h,Cluster} - \varphi_{h,Spur} \tag{6.12}$$

- Die mit der Cluster- bzw. Spurmethode vorhergesagten Polarwinkel des gestreuten Leptons und ihre Differenz:

$$\theta_{l,Cluster} = 2\arctan\sqrt{\frac{E_e(1-y_{Cluster})}{E_p x_{Cluster}\, y_{Cluster}}} \tag{6.13}$$

$$\theta_{l,Spur} = 2\arctan\sqrt{\frac{E_e(1-y_{Spur})}{E_p x_{Spur}\, y_{Spur}}} \tag{6.14}$$

$$\Delta\theta_l = \theta_{l,Cluster} - \theta_{l,Spur}. \tag{6.15}$$

- Die Anzahl der Treffer und der rekonstruierten Spuren in der zentralen Spurkammer N_{CJC} und $N_{Spur,CJC}$.

- Die Symmetrie der Energieverteilung in der radialen Ebene V_{ap}/V_p (Gleichungen 5.4 und 5.3).

- Die Anzahl der Teilchenjets N_{Jets} im Ereignis. Dabei wurde zur Jetsuche ein Kegelalgorithmus im Laborsystem mit einem Radius von 0,7 und einer minimalen transversalen Energie $E_{\perp,min} = 5{,}0\,GeV$ verwendet.

- Die Energie E_{Jet1}, der transversale Impuls $p_{\perp,Jet1}$ und die Rapidität η_{Jet1} des Teilchenjets mit der größten Energie E_{Jet1}.
  Die mit dem Polarwinkel korrelierte Rapidität ist definiert als:

$$\eta_{Jet1} = \frac{1}{2}\ln\left(\frac{E_{Jet1}+p_{z,Jet1}}{E_{Jet1}-p_{z,Jet1}}\right). \tag{6.16}$$

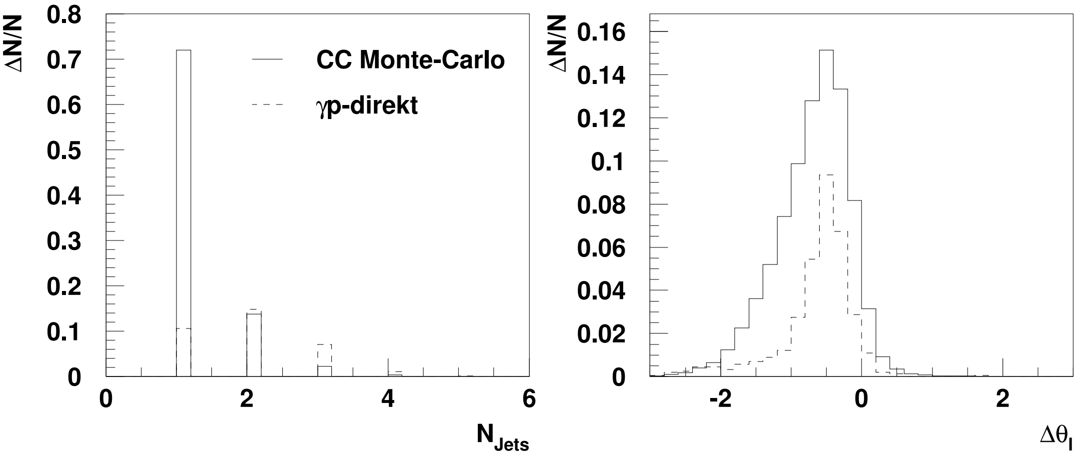


Abbildung 6.5: *Die Monte Carlo Verteilungen der möglicher Eingabevariablen N_{Jets} und $\Delta\theta_l$, für das Netz zur Trennung von CC- und γp-Ereignissen.*

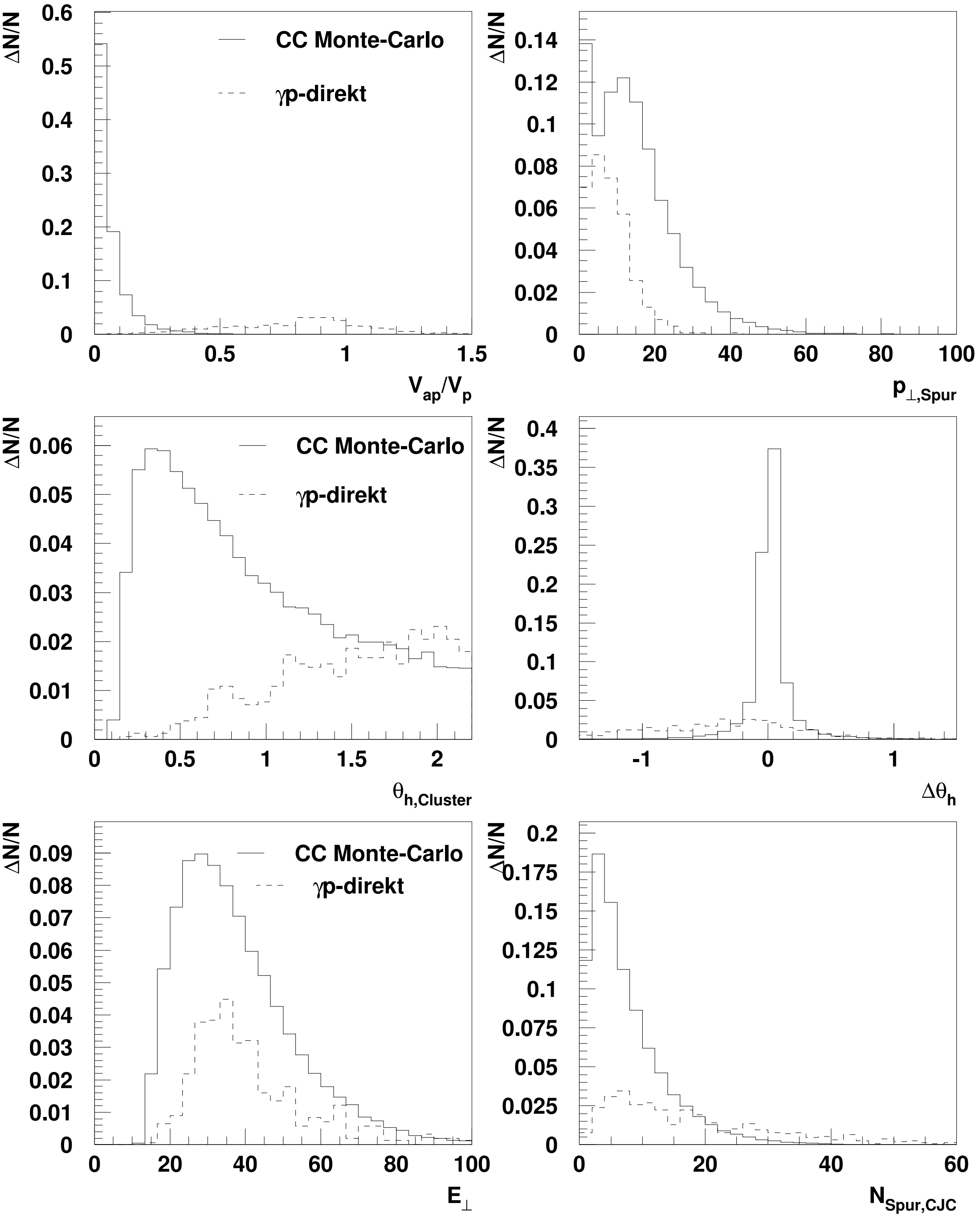


Abbildung 6.6: *Monte Carlo Vergleich der Verteilungen der Variablen* V_{ap}/V_p *und* $p_{\perp,Spur}$ *(oben),* $\theta_{h,Cluster}$ *und* $\Delta\theta_h$ *(Mitte) sowie* $E_\perp$ *und* $N_{Spur,CJC}$ *(unten), die als mögliche Eingabegrößen eines neuronalen Netzes zur Trennung von Ereignissen des geladenen Stroms und der Photoproduktion verwendet werden. Aufgetragen ist die relative Anzahl von Ereignissen gegen die Eingabevariable.*

In den Abbildungen 6.5 und 6.6 sind für die Verteilungen der Eingabevariablen mehrere Beispiele dargestellt. Es wird jeweils die Verteilung der Ereignisse des geladenen Stroms mit der der Photoproduktion verglichen. Dabei wurden die in Tabelle 6.1 und 6.2 beschriebenen Datensätze CCMC und γpdir3 verwendet. Die größten Unterschiede der Verteilungen der beiden Ereignisklassen sind in den Variablen V_{ap}/V_p, $p_{\perp,Spur}$, $\theta_{h,Cluster}$ und $\Delta\theta_h$ zu erkennen. Die Variablen $E_\perp$ und $\Delta\theta_l$ scheinen als Eingabegrößen weniger geeignet zu sein. Man erkennt außerdem, daß aufgrund der Überlappung der Verteilungen in allen Beispielen, durch eine Beschränkung einzelner Variablen auf bestimmte Bereiche, keine vollständige Trennung der Ereignisklassen erreichbar ist.

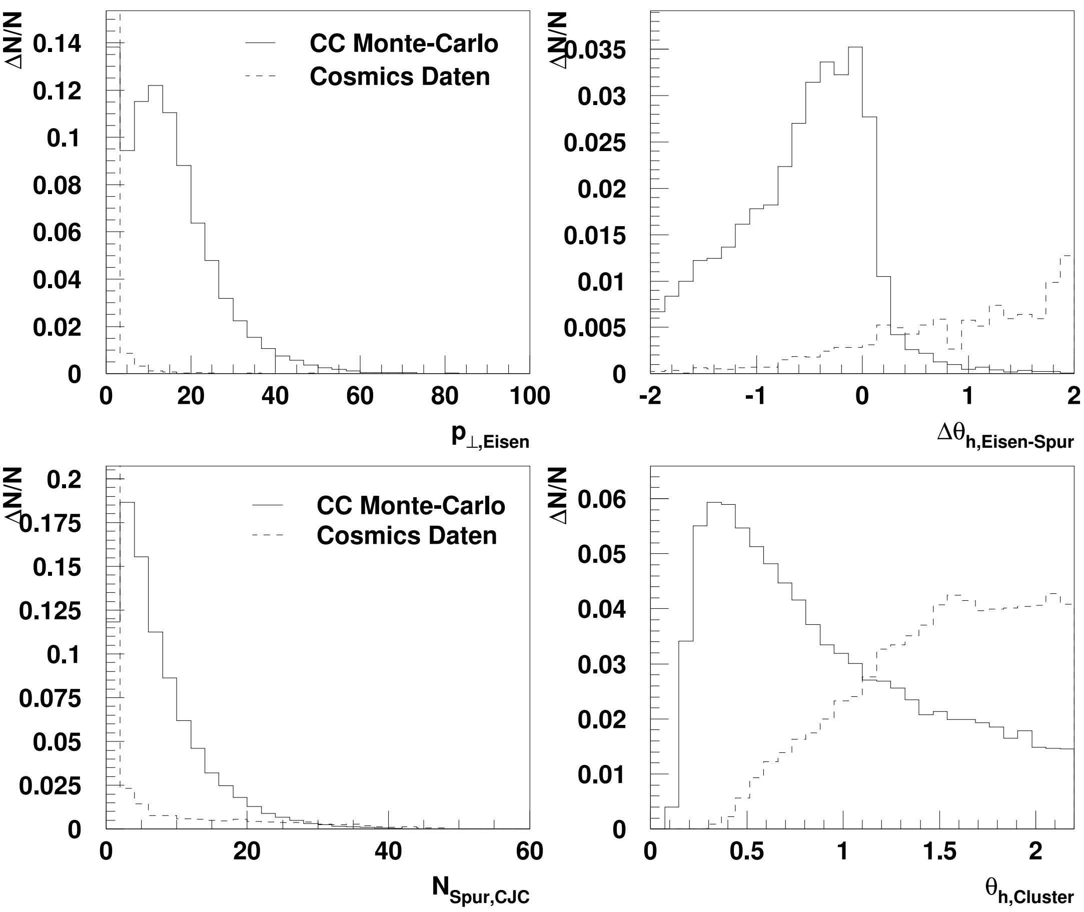


Abbildung 6.7: *Die Verteilungen der möglicher Eingabevariablen* $p_{\perp,Eisen}$, $\Delta\theta_{h,Eisen-Spur}$, $N_{Spur,CJC}$ *und* $\theta_{h,Cluster}$ *eines Netzes zur Trennung von CC-Ereignissen und Cosmics.*

In der Abbildung 6.7 sind Beispiele der Verteilungen möglicher Eingabevariablen eines neuronalen Netzes zur Unterscheidung von Ereignissen des geladenen Stroms und den durch kosmische Myonen induzierten Ereignisse dargestellt. Dabei wurden die in Tabelle 6.1 und 6.3 beschriebenen Datensätze CCMC und COSM96 verwendet. Zu erkennen sind deutliche Differenzen zwischen den Verteilungen der beiden Ereignisklassen in allen Variablen. Besonders die mit den im instrumentierten Eisen deponierten Energien berechneten Variablen $p_{\perp,Eisen}$ und $\Delta\theta_{h,Eisen-Spur}$ zeigen deutliche Unterschiede in ihren Verteilungen.

Beispiele möglicher Eingabevariablen eines Netzes zur Erkennung von Halo-Myon Ereignissen werden in der Abbildung 6.8 gezeigt. Wiederum zeigen sich in allen Variablen deutliche Unterschiede zwischen den Verteilungen der Ereignissklasen. Im Vergleich wurden die Datensätze CCMC und HALO96 aus den Tabellen 6.1 und 6.3 verwendet.

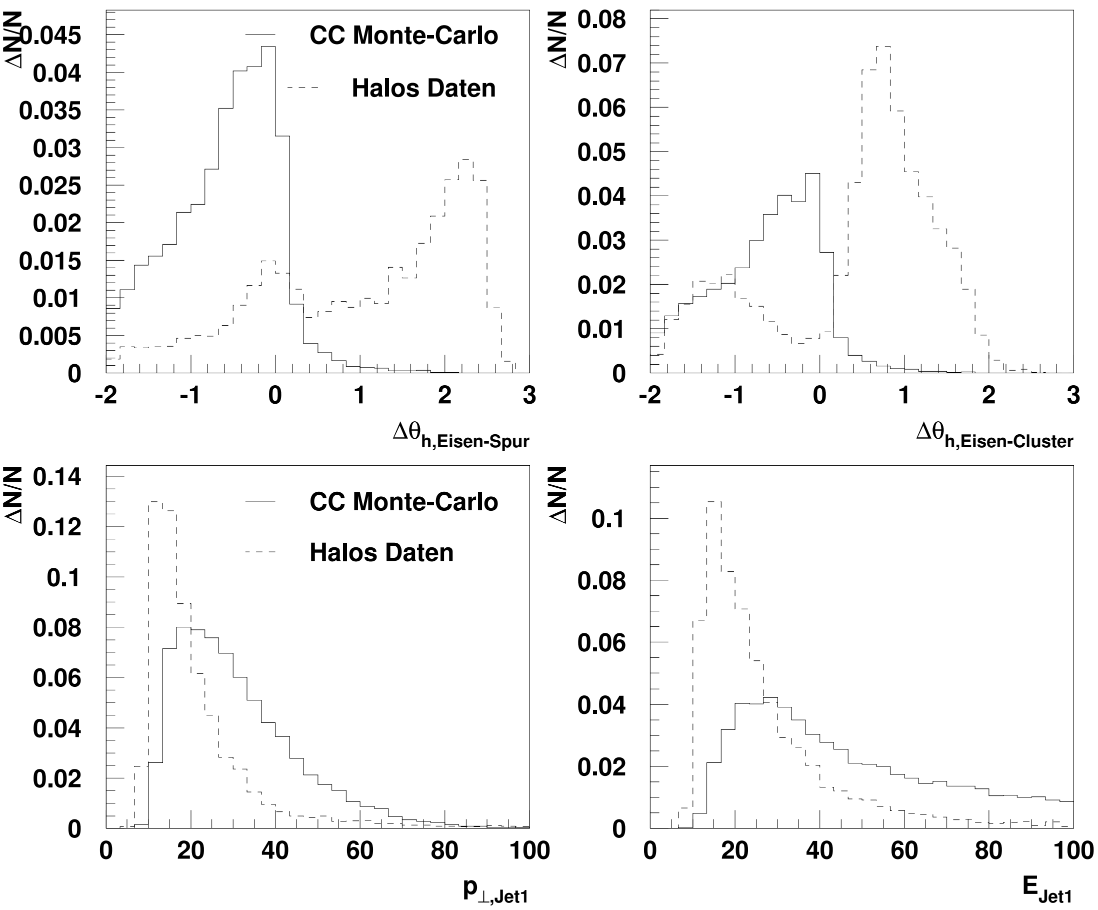


Abbildung 6.8: *Verteilungen der möglicher Eingabevariablen $\Delta\theta_{Eisen-Spur}$ und $\Delta\theta_{Eisen-Cluster}$ (oben) sowie $p_{\perp,Jet1}$ und E_{Jet1} (unten) eines neuronalen Netzes zur Erkennung von Halo-Ereignissen.*

Die Abbildung 6.9 zeigt die Verteilung der möglichen Eingabevariablen $E_\perp$ und y_{JB} für ein Netz zur Trennung von Ereignissen des geladenen- und neutralen Stroms. Verwendet wurden die Datensätze CCMC und NCMC.

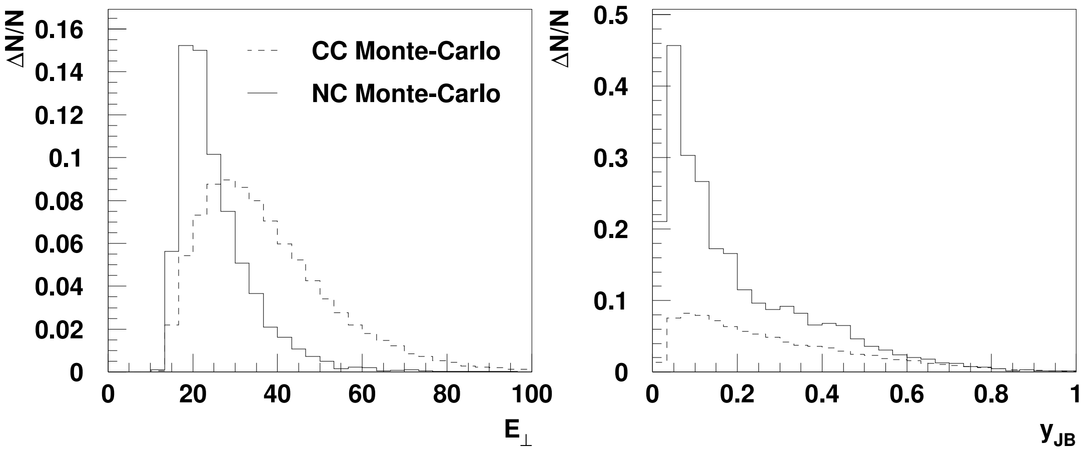


Abbildung 6.9: *Die Verteilungen der möglichen Eingabevariablen $E_\perp$ und y_{JB} für ein Netz zur Erkennung von Ereignissen des neutralen Stroms.*

6.4 Training und Test der Netze

Nachdem in den vorigen Abschnitten die Auswahl der möglichen Eingabevariablen getroffen und die Algorithmen zum Training und zur Optimierung der Netze vorgestellt wurden, werden in diesem Abschnitt die Resultate des Trainings dargestellt.

6.4.1 Netzwerke zur Erkennung von CC- und γp-Ereignissen

Ein neuronales Netz zur Unterscheidung der Ereignisklasse CC von der Untergrundklasse U benötigt in seiner Ausgabeebene nur einen Knoten. Die Ausgabefunktion dieses Knotens läßt Netzantworten y zwischen Null und Eins zu. Beim Training des Netzes verlangt man bei Eingabe eines Ereignisses der Klasse CC am Ausgabeknoten den Wert Eins und bei einem Ereignis der Klasse U den Wert Null.

In der Abbildung 6.10 wird die Verteilung der Netzantwort y eines Netzes zur Unterscheidung von CC- und γp-Ereignissen veranschaulicht. Das Netz besitzt in der Eingabeebene 11 Knoten, in der ersten verborgenen Schicht 4 Knoten und in der zweiten verborgenen Schicht 25 Knoten. Um die Auswirkung der Anzahl der verwendeten Trainingsmuster auf die Effizienz des Netzwerkes zu analysieren, wurde eine Testreihe mit vier verschiedenen Trainingsmusterzahlen N_{Train} durchgeführt. Die maximale mittlere Effizienz $\bar{\varepsilon}(N_{Train})$ des mit N_{Train} Mustern trainierten Netzwerkes wurde jeweils durch einen Test mit 2000 unabhängigen Testmustern bestimmt:

N_{Train}	$\bar{\varepsilon}(N_{Train})$
10	86,7%
100	95,6%
1000	97,5%
2000	97,7%

Es besteht eine Abhängigkeit der mittleren Effizienz $\bar{\varepsilon}$ von der Anzahl N_{Train} der verwendeten Trainingsmuster. Mit steigender Anzahl N_{Train} wächst die erreichte mittlere Effizienz des Netzwerkes. Ab einer Anzahl von $N_{Train} = 1000$ Trainingsmustern erreicht die Effizienz einen Sättigungswert. Man erkennt, daß bereits 10 Trainingsmuster ausreichen um eine hohe Effizienz des Netzwerkes zu garantieren. Dieses Beispiel zeigt die Fähigkeit des neuronalen Netzwerks zur Generalisation: Ein neuronales Netz ist durch das Training mit Mustern einer Klasse in der Lage weitere Muster dieser Klasse zu erkennen, die ihm zuvor nicht präsentiert wurden. Im folgenden wurde das Netzwerk mit je 1000 γp- und CC-Ereignissen trainiert. Dabei wurde für Ereignisse des geladenen Stroms der Ausgabewert $y = 1$ und für solche der Photoproduktion $y = 0$ trainiert.

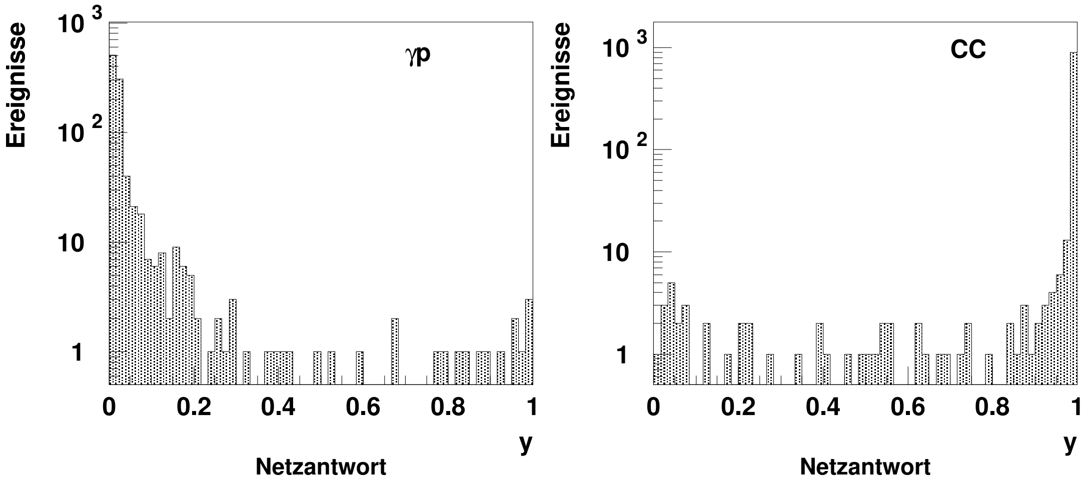


Abbildung 6.10: *Die Verteilungen der Netzantwort y beim Test eines mit dem Programm Crossover erzeugten Netzes zur Trennung von CC- und γp-Ereignissen im logarithmischen Maßstab.*

Man erkennt in der Abbildung, daß der überwiegende Teil der Testmuster korrekt erkannt wurde: Die Maxima der beiden Verteilungen befinden sich bei dem im Training für die Ereignissklassen geforderten Ausgabewerten ($y_{cc} = 1$, $y_{\gamma p} = 0$). Allerdings wurden einige Ereignisse als solche der jeweils anderen Klasse fehlerkannt. Durch den Verlauf der Ausgabefunktion (vgl. Abbildung 4.6) bei einer Temperatur von $T = 0{,}625$ im Netz, erklären sich die Verteilung der Ausgebewerte y zwischen 0 und 1. Um die beiden Klassen zu trennen, muß definiert werden welcher Bereich in y der jeweiligen Klasse zugeordnet werden kann. Die Formen der Verteilungen des Ausgabewertes y lassen die

Einordnung der Ereignisse mit einem $y \geq 0{,}5$ in die Klasse des geladenen Stroms und der Ereignisse mit $y < 0{,}5$ in eine Untergrundklasse als sinnvoll erscheinen.

Die Abbildung 6.11 verdeutlicht die Entwicklung der mittleren Effizienz $\bar{\varepsilon}$ beim Training eines Netzes zur Unterscheidung von Ereignissen des geladenen Stroms und der Photoproduktion. Aufgetragen ist zum einen die durch den Trainingsmustersatz bestimmte mittlere Effizienz und im Vergleich dazu die mit dem unabhängigen Testmustersatz ermittelte. Im Gegensatz zur Effizienz während des Trainings, die kontinuierlich mit zunehmender Zahl von Epochen ansteigt, erreicht die Effizienz beim Test ein Maximum und fällt danach auf einen konstanten Wert ab. Dieses Phänomen wird als "Auswendiglernen" des Trainingsmustersatzes gedeutet. Der Punkt an dem die mit dem Testmustersatz bestimmte mittlere Effizienz ihren maximalen Wert erreicht, wird als maximal mögliche *Generalisation* bezeichnet.

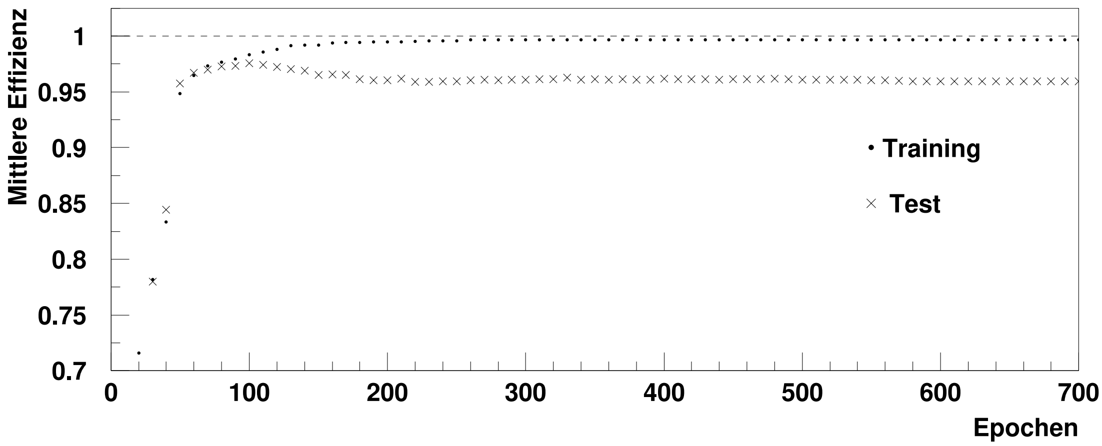


Abbildung 6.11: *Die mit den Trainings- bzw. Testmustersatz bestimmte mittlere Effizienz* $\bar{\varepsilon}$ *des Netzes in Abhängigkeit von der Zahl der Trainings-Epochen.*

Um das Auswendiglernen des Trainingsmustersatzes zu verhindern, dürfen die Netze nur bis zum erreichen der maximalen, mit einem unabhängigen Testmustersatz bestimmten, mittleren Effizienz trainiert werden. Innerhalb des genetischen Algorithmus werden die Netze jeweils 120 Epochen trainiert. Gleichzeitig wird nach jeweils 10 Epochen die mittlere Effizienz mit einem unabhängigen Testmustersatz bestimmt und die dabei ermittelte maximale Effizienz als Fitnesswert des Netzes betrachtet. Dies ist gerechtfertigt, da sich in dieser Analyse gezeigt hat, daß der RPROP-Trainingsalgorithmus den Punkt der maximalen Generalisation innerhalb von 110 Epochen erreicht. Das abschließenden Training, des mit dem genetischen Algorithmus gefundenen neuronalen Netzes mit der höchsten mittleren Effizienz, wird nach dem Erreichen von $\bar{\varepsilon}_{max}$ abgebrochen.

Im Verlauf der Analyse wurden mit dem genetischen Algorithmus Netzwerke entwickelt, die jeweils eine Untergrundklasse von der des geladenen Stroms unterscheiden sollen. Für jede Untergrundklasse wurden mehrere Netze entwickelt, wobei die Mutationsrate im Bereich $6\% \leq p_m \leq 10\%$ variiert wurde. Größere Mutationsraten verhindern die effektive Weitergabe der Erbinformation in der Generation, da eine zu große Anzahl von Genen bei der Mutation invertiert wird. Der in Abbildung 6.12 dargestellte Vergleich der mit unterschiedlichen Mutationsraten erzeugten Netze läßt erkennen, daß deren Effizienzen, im Rahmen der statistischen Fehler, vergleichbare Werte erreichen.

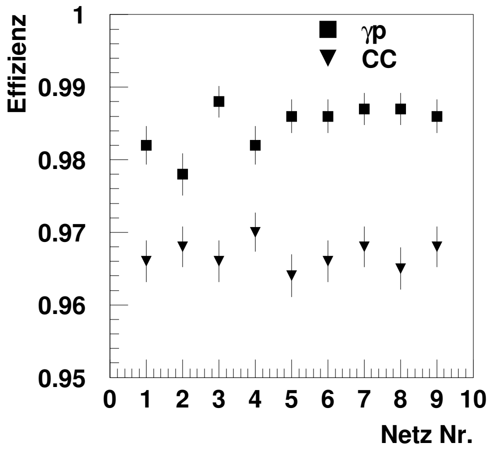


Abbildung 6.12: *Die Effizienzen der Netze zur Unterscheidung von CC- und γp-Ereignissen im Vergleich. Die Effizienzen wurden mit den während des Trainings verwendeten Trainings- und Testmustersätzen bestimmt. Angegeben ist der statistische Fehler der Effizienz.*

Das in Abbildung 6.12 mit Nr. 9 bezeichnete Netz CCγp soll im Folgenden näher analysiert werden. In der Tabelle 6.4 sind die vom genetischen Algorithmus ausgewählten Parameter und Eingabevariablen aufgeführt.

1.Ebene Eingabe	2.Ebene	3.Ebene	$1/T$	w_{init}	Eingabegrößen
11 Knoten	4 Knoten	25 Knoten	1,6	0,02	$\Delta\theta_h$; $p_{\perp,Spur}$; $\theta_{h,Spur}$; $\theta_{l,Spur}$; $E_\perp$ $N_{Spur,CJC}$; y_{JB}; V_{ap}/V_p; $p_{\perp,Eisen}$ N_{Jets}; η_{Jet1}

Tabelle 6.4: *Die Parameter des mit dem genetischen Algorithmus nach 50 Generationen bei einer Mutationsrate von $p_m = 0{,}09$ erzeugten Netzes CCγp. Das Training erfolgte 110 Epochen lang mit dem RPROP-Lernalgorithmus.*

Die Verteilungen der Netzantwort y beim Test des Netzes CCγp mit den in Tabelle 6.1 beschriebenen Testsätzen CCMC und γpdir3 ist in der Abbildung 6.13 dargestellt. Die überwiegende Zahl von Testmustern wurde der richtigen Klasse zugeordnet. Das Netz CCγp ist somit in der Lage die Ereignisklassen mit hoher Effizienz zu unterscheiden. Die

Abhängigkeit der Effizienz der CC-Erkennung vom Transversalimpuls $p_{\perp,Cluster}$ erkennt man in der Abbildung 6.14. Es werden die Effizienzen beim Test des Netzes mit drei unabhängigen Testmustersätzen verglichen.

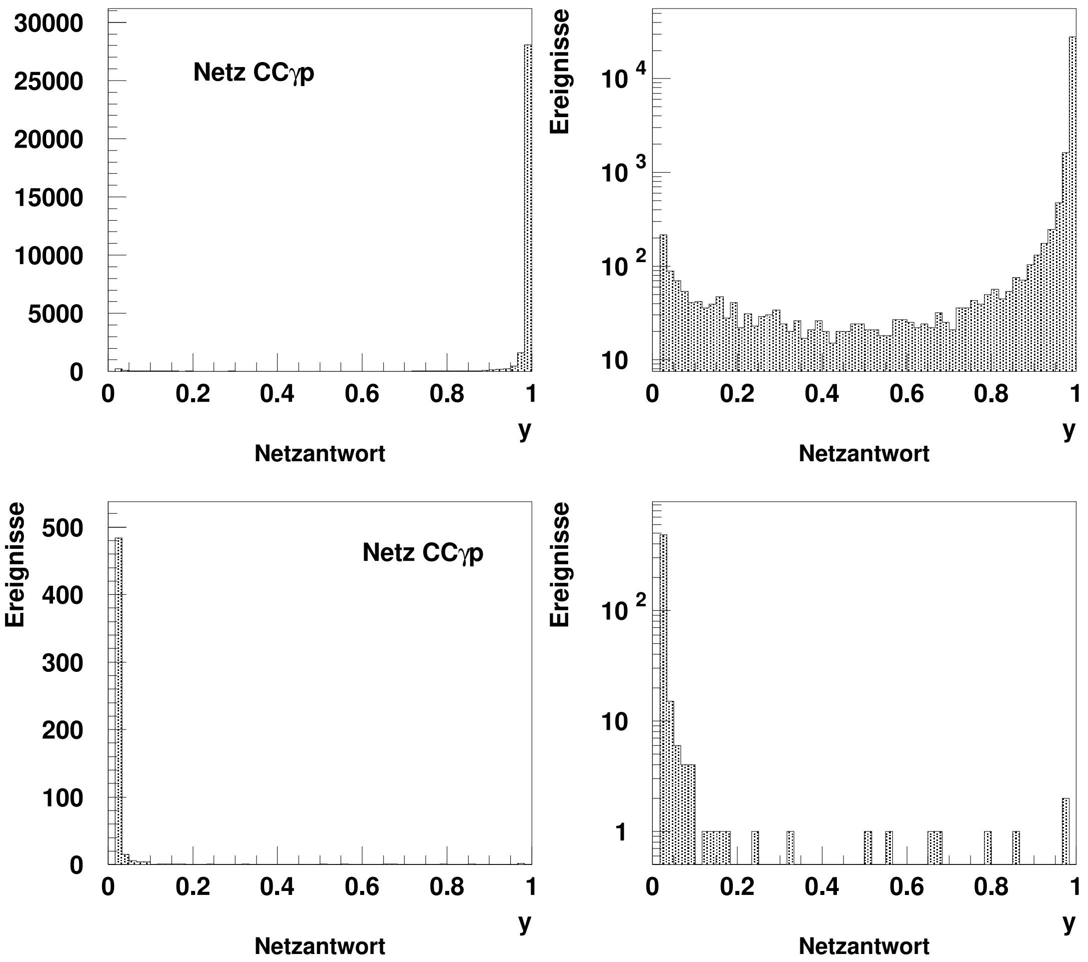


Abbildung 6.13: *Die Verteilungen der Netzantwort y beim Test von Netz CCγp mit den Mustersätzen CCMC (obere Bildhälfte) und γpdir3.*

Da sich die kinematischen Verteilungen von CC- und NC-Ereignissen unterscheiden, müssen Pseudo-CC-Ereignisse entsprechend dem Verhältnis der Wirkungsquerschnitte, bei der Berechnung der Effizienz umgewichtet werden. Der Gewichtsfaktor

$$g(\widetilde{x},\widetilde{y}) = \frac{\frac{d^2\sigma_{CC}(\widetilde{x},\widetilde{y})}{(dxdy)}}{\frac{d^2\sigma_{NC}(\widetilde{x},\widetilde{y})}{(dxdy)}} \tag{6.17}$$

wird am Punkt $(\widetilde{x},\widetilde{y})$ in der kinematischen Ebene durch die Variablen der Elektronmethode definiert [Mar97].

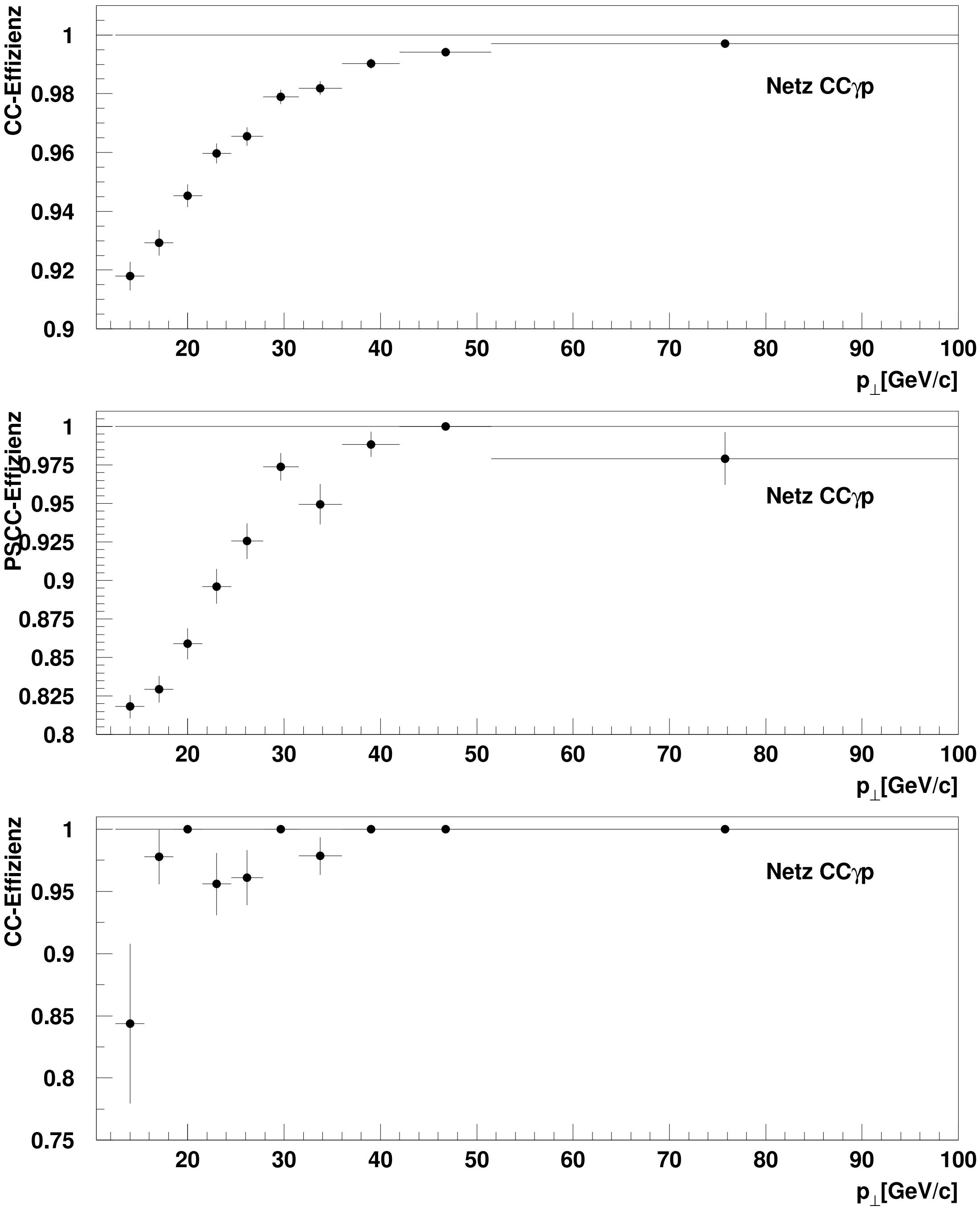


Abbildung 6.14: *Die Effizienz der CC-Erkennung von Netz CC*γp *in Abhängigkeit vom Transversalimpuls* $p_\perp$. *In der oberen Abbildung beim Test mit dem Monte-Carlo Datensatz CCMC in der Mitte beim Test mit den Pseudo-CC Ereignissen des Satzes PSCC und in der unteren Abbildung mit dem CC-Datensatz CCDATA. Man beachte die verschiedenen Effizienzskalen.*

Der Abfall der Effizienzen bei kleinem Transversalimpuls in allen drei Verteilungen, erklärt sich aus dem starken Anstieg des γp-Wirkungsquerschnitts bei sinkendem $p_\perp$ (vgl. Abbildung 5.5). In diesem $p_\perp$-Bereich können die beiden Ereignissklassen nicht an ihren Transversalimpulsen unterschieden werden.

Auffällig ist die im Vergleich zu den übrigen Verteilungen schlechtere Effizienz des Netzes beim Test mit Pseudo-CC-Ereignissen. In der Pseudo-CC-Ereignismenge befindet sich, wie in der Menge der Ereignisse des neutralen Stroms, ein nicht zu vernachlässigender Anteil von Photoproduktionsereignissen. Dieser Untergrundanteil beträgt im Bereich 12,5GeV $< p_\perp <$ 17,5GeV *ca.* 6% [Mar97]. Dadurch erklärt sich die Verschlechterung der CC-Effizienz des Netzes beim Test mit Pseudo-CC-Ereignissen.

Ein Teil der vom Netz als γp-Ereignisse erkannten Pseudo-CC's wurde einer visuellen Inspektion unterzogen, wobei ein großer Anteil als Ereignisse der Photoproduktion identifiziert werden konnte.

Bei der Bewertung der beim Test mit dem CC-Datensatz CCDATA erzielten Effizienzen, muß die Beschränkung der Variablen V_{ap}/V_p bei der Datenselektion auf Werte $V_{ap}/V_p < 0{,}2$ berücksichtigt werden. Wie man in Abbildung 6.6 sieht, entfernt eine solche Einschränkung den größten Teil des γp-Untergrundes aus der CC-Ereignismenge. Das heißt, der für das Netz problematische Bereich $V_{ap}/V_p > 0{,}2$ ist in dieser Ereignismenge nicht enthalten.

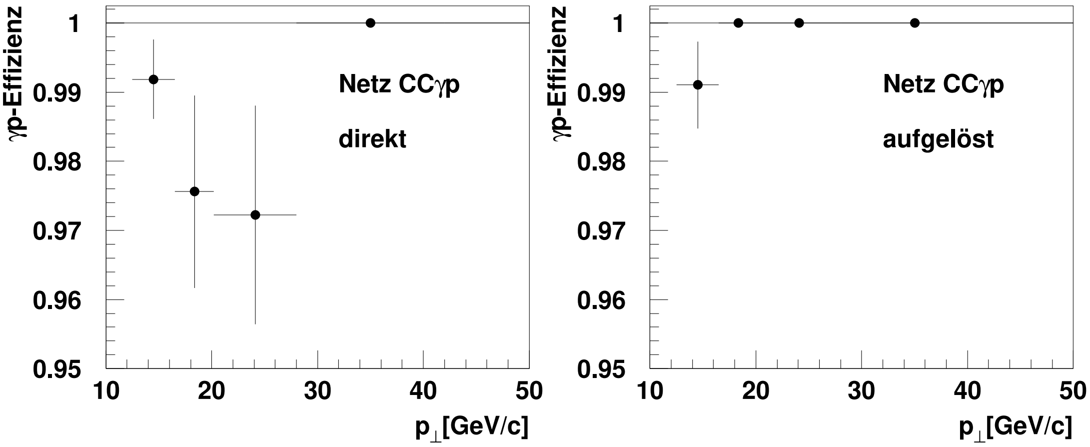


Abbildung 6.15: *Die $p_\perp$-Abhängigkeit der Effizienz der γp-Erkennung des Netzes CCγp. Im linken Bild beim Test mit dem Monte Carlo Datensatz γpdir3 und rechts beim Test mit γpres3.*

Um die Effizienz der Erkennung von Ereignissen der Photoproduktion durch das Netz CCγp zu analysieren, wurden die Testmustersätze γpdir3 und γpres3 verwendet. Abbildung 6.15 verdeutlicht, daß Ereignisse der direkten und aufgelösten Photoproduktion mit unterschiedlichen Effizienzen erkannt werden. Zu erklären ist diese Beobachtung anhand der in Abbildung 6.16 dargestellten Verteilung der γp-Ereignisse, in der von den Variablen V_{ap}/V_p und N_{Jets} aufgespannten Ebene. Im Vergleich der Verteilungen erkennt man, daß der Datensatz der aufgelösten Photoproduktion, im Gegensatz zu dem der direkten Photoproduktion keine Ereignisse mit einem $V_{ap}/V_p < 0{,}2$ enthält. Gerade in diesem Bereich befindet sich jedoch das Maximum der Verteilung der CC-Ereignisse (vgl. Abbildung 6.6). Das Netz wird bevorzugt derartige Ereignisse als solche des geladenen Stromes identifizieren.

Die Größe der Wirkungsquerschnitte des geladenen Stroms und der Photoproduktion unterscheidet sich erheblich. Es ist daher nicht möglich die Effizienzen der CC- und γp-Erkennung des Netzes, in Hinblick auf den zu erwartenden γp-Untergrund direkt zu vergleichen. Eine quantitative Bestimmung des Untergrundes ist mit der sogenannten Reinheit möglich. Die Reinheit P ist definiert als:

$$P := \frac{N_{cc}}{N_{cc} + (N_{\gamma p} \cdot LF)} \quad \text{mit} \quad LF = \frac{\mathcal{L}_{cc}}{\mathcal{L}_{\gamma p}}. \tag{6.18}$$

Dabei ist N_{cc} die Anzahl der als CC's erkannten Ereignisse des CC-Datensatzes, $N_{\gamma p}$ die Anzahl der als CC's fehlerkannten Ereignisse des γp-Datensatzes, $\mathcal{L}_{cc}$ und $\mathcal{L}_{\gamma p}$ die integrierte Luminosität des CC- bzw. γp-Datensatzes.

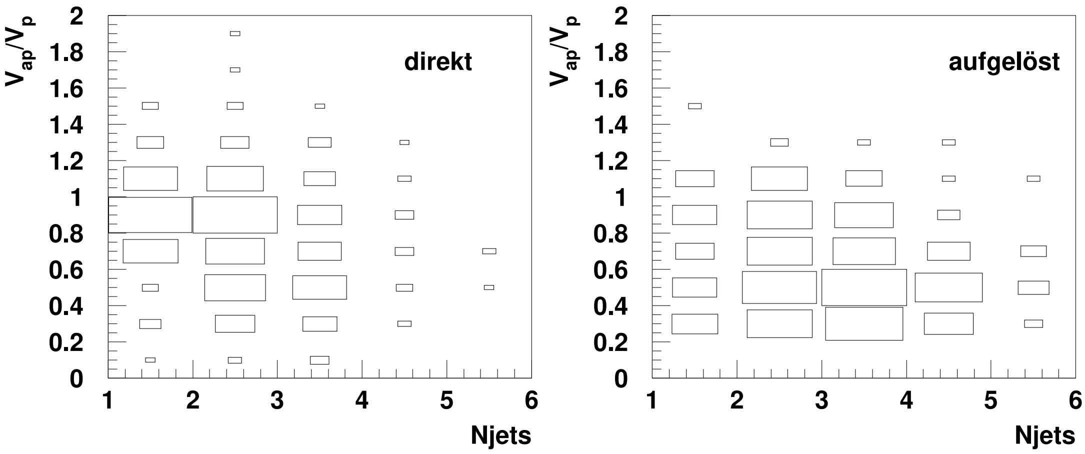


Abbildung 6.16: *Verteilung der γp-Ereignisse in der von den Variablen V_{ap}/V_p und N_{Jets} aufgespannten Ebene. Die Fläche der einzelnen Rechtecke gibt die relative Anzahl von Ereignissen im jeweiligen Bin an. Die Ebene wurde in 50 Bins unterteilt.*

Die Effizienzen und Reinheiten des neuronalen Netzwerks CCγp findet man in der Tabelle 6.5.

CC-Datensatz	CC-Effizienz	γp-Datensatz	γp-Effizienz	Reinheit
CCMC	$(96{,}4 \pm 0{,}1)\%$	γpdir3	$(98{,}5 \pm 0{,}5)\%$	$(99{,}0 \pm 0{,}3)\%$
PSCC	$(93{,}1 \pm 0{,}3)\%$	γpres3	$(99{,}4 \pm 0{,}4)\%$	$(99{,}4 \pm 0{,}3)\%$
CCDATA	$(97{,}7 \pm 0{,}6)\%$	γpres1	$(97{,}2 \pm 0{,}9)\%$	$(96{,}0 \pm 1{,}2)\%$

Tabelle 6.5: *Zusammenstellung der Effizienzen des Netzes CCγp. Angegeben sind die statistischen Fehler. Die Beschreibung der einzelnen Datensätze findet man in den Tabellen 6.1 und 6.2.*

6.4.2 Netzwerke zur Erkennung von CC- und Myon-Ereignissen

Analog der im vorhergehenden Abschnitt beschriebenen Vorgehensweise, wurden mit dem genetischen Algorithmus verschiedene Netzwerke zur Erkennung von CC- und Halo- bzw. Cosmic-Ereignissen erzeugt. Zum Training und Test der Netze wurden die in Tabelle 6.3 beschriebenen, mit den Halo- und Cosmic-Filtern selektierten Datensätze verwendet. Die Datensätze wurden in je einen Trainings- und einen Testmustersatz gleicher Größe aufgeteilt. Bei der abschließenden Bestimmung der Effizienzen wurden nur die Testmustersätze verwendet. Die Parameter und Eingabevariablen der Netze CCHalo und CCCosmic zur Trennung von CC- und Halo- bzw. Cosmic-Ereignissen, sind in der Tabelle 6.6 zusammengefaßt. In beiden Netzwerken werden mehrere auf den Energiedepositionen im instrumentierten Eisen basierende Variablen verwendet. Diese Variablen sind insbesondere geeignet die charakteristische Signatur der Halo- und Cosmic-Ereignisse im instrumentierten Eisen zu erkennen.

Eingabe Knoten	2.Ebene Knoten	3.Ebene Knoten	$1/T$	w_{init}	Eingabegrößen
Netz CCHalo					
14	28	9	0,6	0,18	$\Delta\varphi_h; p_{\perp,Spur}; \theta_{h,Spur}; \theta_{l,Spur}; E_\perp$ $\varphi_{h,Spur}; N_{Spur}; y_{JB}; V_{ap}/V_p; p_{\perp,Eisen}$ $\Delta\theta_{h,E-C}; \Delta\theta_{h,E-S}; N_{Jets}; p_{\perp,Jet1}$
Netz CCCosmic					
12	21	2	2,2	0,1	$\Delta\theta_h; p_{\perp,Cluster}; p_{\perp,Spur}; p_{\perp,C}/p_{\perp,S}$ $\theta_{h,Spur}; \varphi_{h,Cluster}; y_{JB}; V_{ap}/V_p$ $p_{\perp,Eisen}; \Delta\theta_{h,E-S}; N_{Jets}$

Tabelle 6.6: *Parameter und Eingabevariablen der Netze CCHalo und CCCosmic*

In der Abbildung 6.17 sind die Effizienzen der CC-Erkennung des Netzes CCHalo in Abhängigkeit vom Transversalimpuls $p_\perp$ dargestellt. Dazu wurde das Netzwerk mit den drei in Tabelle 6.1 beschriebenen CC-Datensätzen getestet. In jeder der drei Abbildungen ist ein Anstieg der CC-Effizienz mit wachsendem Transversalimpuls zu beobachten.

Um die Effizienz der Halo-Erkennung des Netzwerkes CCHalo zu ermitteln, wurde es

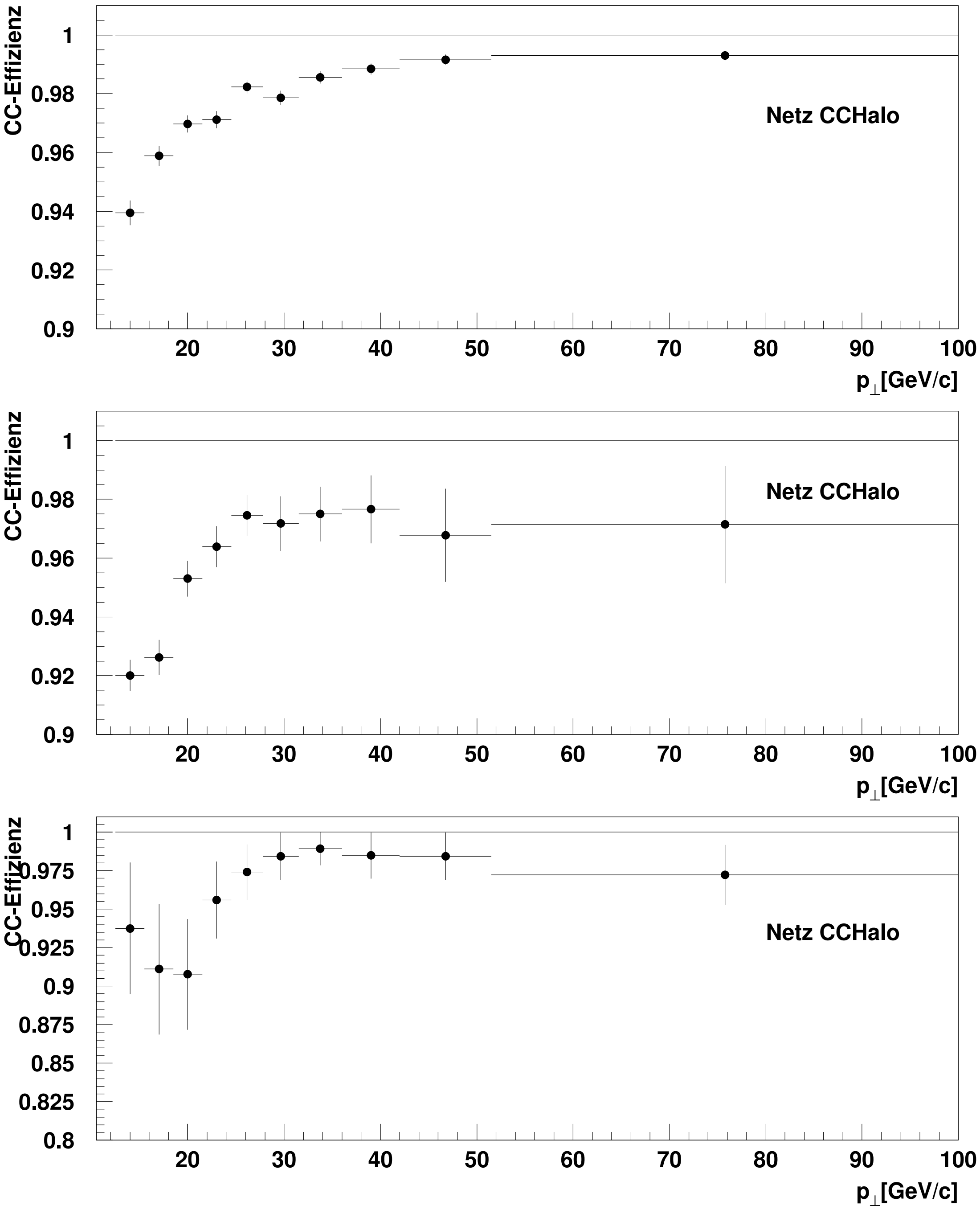


Abbildung 6.17: *Die Effizienz der CC-Erkennung des Netzes CCHalo in Abhängigkeit vom Transversalimpuls $p_\perp$. In der oberen Abbildung beim Test mit dem Monte-Carlo Datensatz CCMC in der Mitte beim Test mit den Pseudo-CC Ereignissen des Satzes PSCC und in der unteren Abbildung mit dem CC-Datensatz CCDATA. Man beachte die verschiedenen Effizienzskalen.*

mit den Testmustern des Datensatzes HALO96 der Tabelle 6.3 getestet. Die Effizienz der Halo-Erkennung ist in der Abbildung 6.18 als Funktion des Transversalimpulses $p_\perp$ abgebildet. Bei steigendem Transversalimpuls fällt die Effizienz auf 93% im Bereich 20GeV/c $\leq p_\perp \leq$ 27,5GeV/c ab, um danach wieder anzusteigen. Um die Erkennung von Halo-Ereignissen, die nicht von den topologischen Filtern als solche erkannt wurden zu untersuchen, wurde das Netzwerk CCHalo mit 29 durch visuelle Inspektion gefundenen Halo-Ereignissen getestet. Das Netzwerk erkannte 24 dieser Ereignisse als Halos. Dies entspricht einer Effizienz von (82,8 ± 7,0)%.

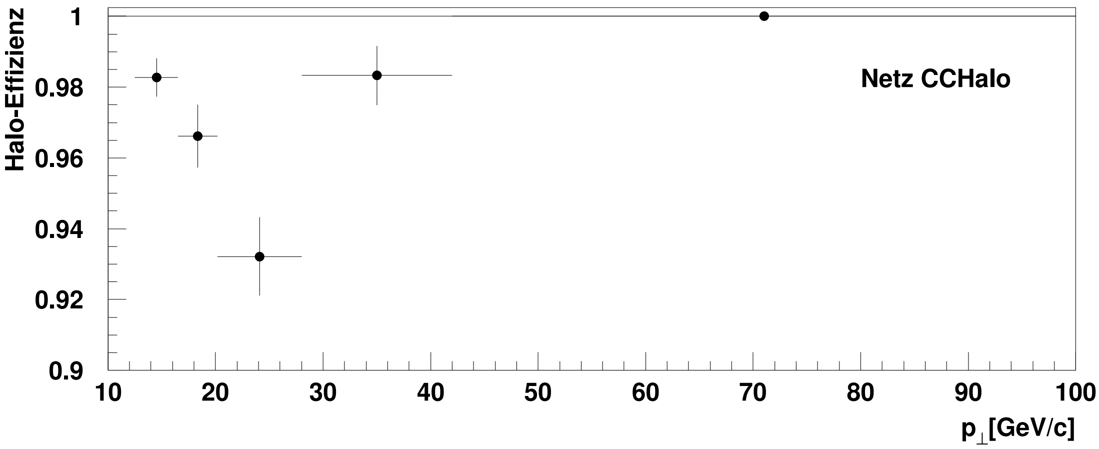


Abbildung 6.18: *Die Effizienz der Halo-Erkennung des Netzes CCHalo in Abhängigkeit vom Transversalimpuls* $p_\perp$.

Die Tabelle 6.8 gibt einen Überblick über die vom Netz CCHalo erreichten Effizienzen. Die mit den Datensätzen CCMC, CCDATA und PSCC ermittelten Effizienzen sind vergleichbar.

Die Parameter und Eingabevariablen des Netzwerkes CCCosmic zur Unterscheidung der Ereignisen des geladenen Stroms von denen durch kosmische Myonen induzierten Ereignissen, findet man in der Tabelle 6.6.

Die Abbildung 6.19 zeigt die $p_\perp$-Abhängigkeit der CC-Effizienz beim Test des Netzwerkes CCCosmic mit den CC-Datensätzen der Tabelle 6.1. Man erkennt eine große Differenz zwischen der mit den Monte Carlo-CC's und der mit den Pseudo- bzw. Daten-CC's ermittelten CC-Effizienz. Eine visuelle Inspektion der als Cosmics fehlerkannten Pseudo-CC-Ereignisse zeigte, daß insbesondere Ereignisse mit einer hohen Energiedeposition in der vorderen Endkappe des instrumentierten Eisens nicht korrekt identifiziert wurden. Der Monte Carlo Generator scheint Ereignisse dieses Typs nicht korrekt zu beschreiben.

Die Effizienz der Cosmic-Erkennung des Netzwerkes CCCosmic ist in der Abbildung 6.20 in Abhängigkeit vom Transversalimpuls dargestellt. Man erkennt, daß Cosmic-Ereignisse in allen $p_\perp$-Bereichen mit einer Effizienz größer 96% erkannt werden.

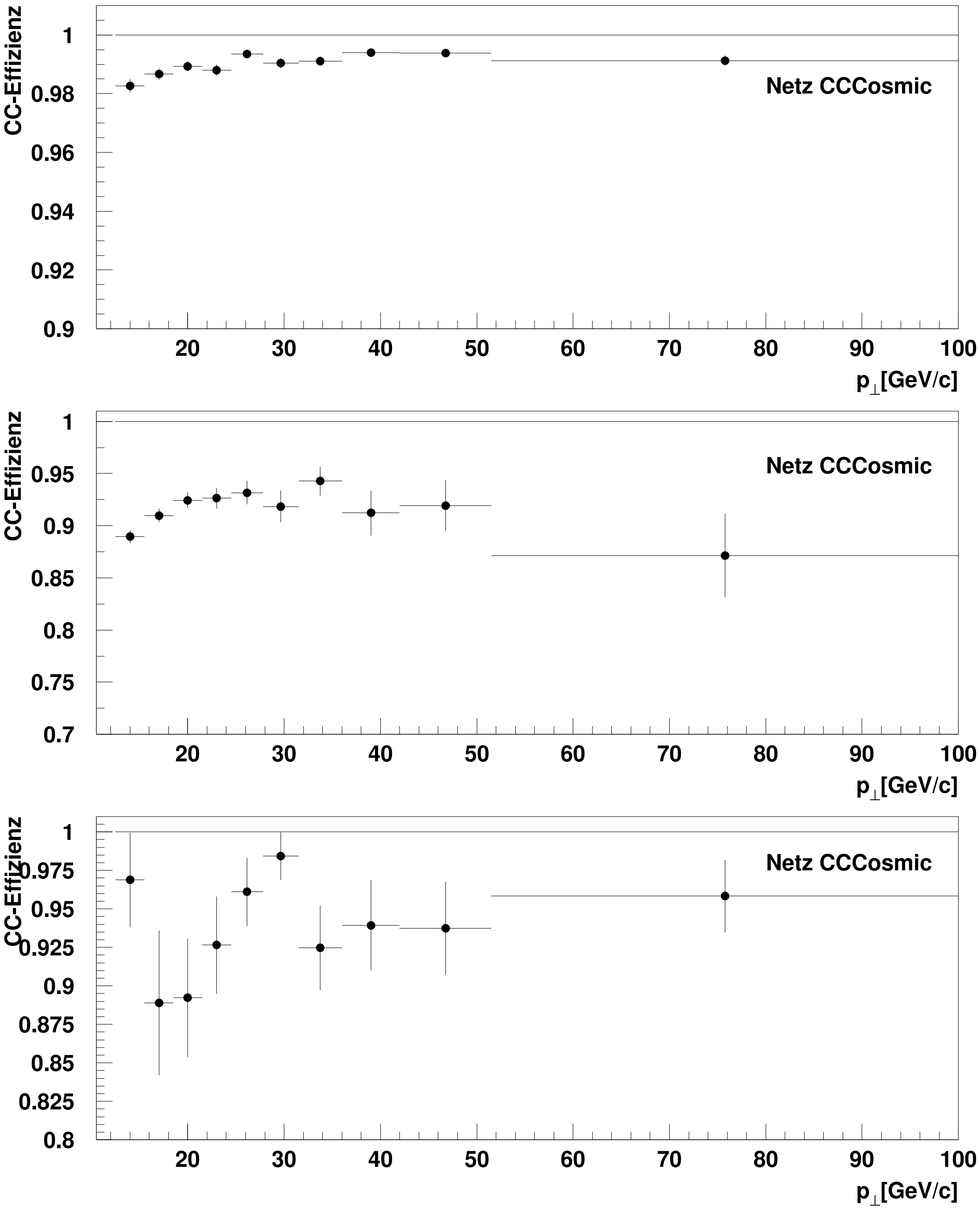


Abbildung 6.19: *Die Effizienz der CC-Erkennung des Netzes CCCosmic in Abhängigkeit vom Transversalimpuls* $p_\perp$*. In der oberen Abbildung beim Test mit dem Monte-Carlo Datensatz CCMC in der Mitte beim Test mit den Pseudo-CC Ereignissen des Satzes PSCC und in der unteren Abbildung beim Test mit dem CC-Datensatz CCDATA. Man beachte die verschiedenen Effizienzskalen.*

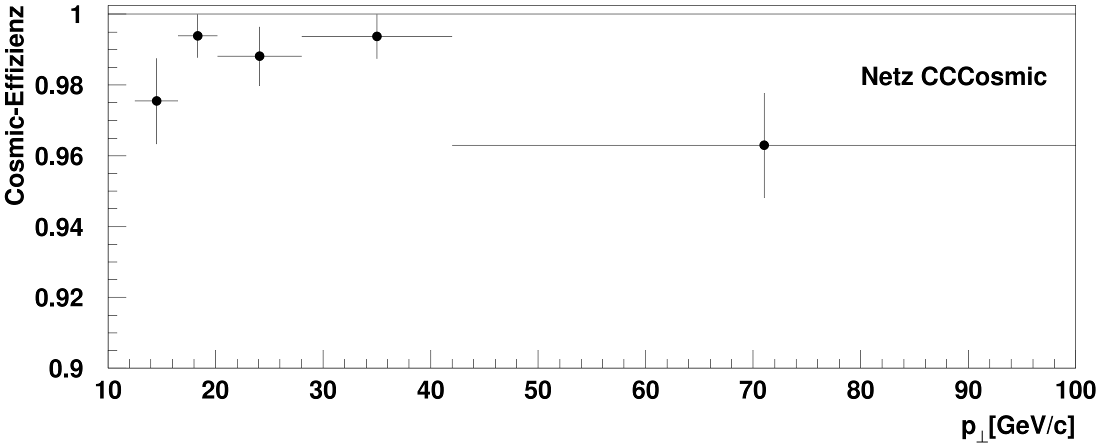


Abbildung 6.20: *Die Effizienz der Cosmic-Erkennung des Netzes CCCosmic in Abhängigkeit vom Transversalimpuls* $p_\perp$.

Um die ungenügend modellierte Energiedeposition im Instrumentierten Eisen der Monte Carlo CC's zu untersuchen, wurde ein weiteres Netzwerk zur Unterscheidung von CC's und Cosmics mit dem evolutionären Algorithmus erzeugt. Dieses Netz wurde nicht mit Monte Carlo CC's, sondern mit Pseudo-CC-Ereignissen trainiert. Im Folgenden wird dieses Netz als CCCosmic2 bezeichnet. Die Parameter und Eingabevariablen dieses Netzwerks sind in der Tabelle 6.7 zusammengefaßt.

Eingabe Knoten	2.Ebene Knoten	3.Ebene Knoten	$1/T$	w_{init}	Eingabegrößen
13	25	31	0,4	0,24	$\Delta\theta_h; p_{\perp,Spur}; \theta_{l,Cluster}; \Delta\theta_l : \varphi_{h;Cluster}$ $N_{Spur}; V_{ap}/V_p; p_{\perp,Eisen}; \Delta\theta_{h,Eisen-Cluster}$ $\Delta\theta_{Eisen-Spur}; N_{Jets}; E_{Jet1}; p_{\perp,Jet1}$

Tabelle 6.7: *Parameter und Eingabevariablen des mit Pseudo-CC Ereignissen trainierten Netzwerks CCCosmic2.*

Der Vergleich der in Tabelle 6.8 aufgeführten Effizienzen der Netze CCCosmic und CCCosmic2 zeigt, daß durch ein Training mit Pseudo CC's die Effizienz beim Test mit Daten CC's um 3%, beim Test mit Pseudo CC's um 3% und beim Test mit Cosmics um 1% verbessert werden konnte. Allerdings zeigt sich, daß die CC-Effizienz bei großem Transversalimpuls $p_\perp > 20$ GeV unter der des Netzwerks CCCosmic liegt. Der im zum Training verwendete Datensatz PSCC enthaltene Anteil von Ereignissen mit einem $p_\perp > 20$ GeV beträgt ca. 34,1%. Im Vergleich dazu beträgt dieser Anteil im CC-Datensatz ca. 71,8%. Die Abnahme der Effizienz bei großem $p_\perp$ des Netzes CCCosmic2 erklärt sich aus dieser Diskrepanz der Anteile von Ereignissen mit großen Transversalimpulsen in den Trainingsdaten.

Netz CCHalo		Netz CCCosmic		Netz CCCosmic2	
Datensatz	Effizienz	Datensatz	Effizienz	Datensatz	Effizienz
CCMC	$(97{,}6 \pm 0{,}1)\%$	CCMC	$(98{,}9 \pm 0{,}1)\%$	CCMC	$(95{,}6 \pm 0{,}1)\%$
PSCC	$(95{,}7 \pm 0{,}3)\%$	PSCC	$(93{,}1 \pm 0{,}3)\%$	PSCC	$(96{,}4 \pm 0{,}2)\%$
CCDATA	$(96{,}4 \pm 0{,}7)\%$	CCDATA	$(93{,}7 \pm 1{,}0)\%$	CCDATA	$(96{,}8 \pm 0{,}7)\%$
HALO96	$(96{,}8 \pm 0{,}4)\%$	COSM96	$(98{,}3 \pm 0{,}4)\%$	COSM96	$(99{,}0 \pm 0{,}3)\%$

Tabelle 6.8: *Tabelle der Effizienzen der Netze CCHalo, CCCosmic und CCCosmic2.*

6.4.3 Netzwerke zur Erkennung von CC- und NC-Ereignissen

Nachdem die Entwicklung von Netzwerken zur Erkennung der Ereignisse der Untergrundklassen γp, Halo und Cosmic innerhalb der Menge der CC-Ereignisse beschrieben wurde, wird in diesem Abschnitt ein Netzwerk zur Trennung der Ereignisse des geladenen- und neutralen Stromes vorgestellt.

Analog zu den in den vorhergehenden Abschnitten beschriebenen Methoden, wurde das Netzwerk zur Trennung von CC- und NC-Ereignissen mit dem genetischen Algorithmus erzeugt. Zum Training wurden Monte Carlo NC und CC Daten verwendet. In der Tabelle 6.9 sind die Parameter und Eingabevariablen des Netzwerkes CCNC aufgelistet. Die Effizienzen des Netzwerkes CCNC findet man in der Tabelle 6.10. Bei allen vier zum Test verwendeten Datensätzen erreicht das Netzwerk eine Effizienz größer 99%. Mit dem Netzwerk CCNC ist im Rahmen der statistischen Fehler eine Trennung der Ereignisklassen CC und NC erreicht worden.

Eingabe Knoten	2.Ebene Knoten	3.Ebene Knoten	$1/T$	w_{init}	Eingabegrößen
10	26	31	1,9	0,1	$\Delta\theta_h$; $\Delta\varphi_h$; $p_{\perp,Spur}$; $p_{\perp,Cluster}$; $p_{\perp,C}/p_{\perp,S}$ $\Delta\theta_l$; $E_\perp$; $\varphi_{h,Spur}$; y_{JB}; V_{ap}/V_p

Tabelle 6.9: *Parameter und Eingabevariablen des Netzwerks CCNC.*

Datensatz	Effizienz
CCMC	$(99{,}5 \pm 0{,}1)\%$
PSCC	$(99{,}2 \pm 0{,}2)\%$
CCDATA	$(99{,}4 \pm 0{,}3)\%$
NCMC	$(99{,}7 \pm 0{,}1)\%$

Tabelle 6.10: *Zusammenstellung der Effizienzen des Netzwerks CCNC.*

6.4.4 Relevanz der Eingabevariablen

Um die Bedeutung der einzelnen Eingabevariablen für die Netzentscheidung zu untersuchen, wird die Netzantwort y für jedes Eingabemuster eines Datensatzes mit der auf ihren Mittelwert gesetzten i-ten Eingabevariable $\bar{x}_i$ berechnet. Von diesem Ergebnis wird die Differenz zum wahren, mit dem Wert x_i der Eingabevariable berechneten Ausgabewert, gebildet. Die normierte quadratische Summe aller N_p im Datensatz enthaltenen Eingabemuster ergibt die Relevanz der Eingabevariablen i:

$$R_i = \frac{1}{N_p} \sum_{j=1}^{N_p} (y(\vec{x}_j) - y(\vec{x}_j|_{x_i=\bar{x}_i}))^2 \tag{6.19}$$

Dabei ist $\vec{x}_j$ der Vektor der Eingabevariablen des Eingabemusters j. Die so ermittelte Relevanz der Eingabevariable hängt vom verwendeten Mustersatz und von der Architektur des Netzwerkes ab. Nur die Relevanzen der einzelnen Eingabevariablen innerhalb eines Netzwerkes können mit einander verglichen werden. Ein Vergleich der Relevanzen der Eingabevariablen verschiedener Netzwerke ist nicht sinnvoll. In der Abbildung 6.21 sind die Relevanzen der Eingabevariablen der Netzwerke CCγp, CCHalo und CCCosmic dargestellt.

Im Netzwerk CCγp übersteigen die Relevanzen der Eingabevariablen $p_{\perp,Spur}$ und V_{ap}/V_p die der übrigen Variablen um eine Größenordnung. Das Training eines Netzwerkes, in dem nur $p_{\perp,Spur}$ und V_{ap}/V_p als Eingabevariablen verwendet werden, hat gezeigt, daß die Effizienzen dieses Netzes ca. 3% unter denen des Netzwerkes CCγp liegen. Die Kombination der Variablen $p_{\perp,Spur}$ und V_{ap}/V_p ist somit ausreichend, den Großteil der CC- und γp-Ereignisse voneinander zu unterscheiden.

Im Netzwerk CCHalo sind die Relevanzen von fünf Eingabevariablen miteinander vergleichbar. Zwei von diesen, $\Delta\theta_{h,Eisen-Cluster}$ und $\Delta\theta_{h,Eisen-Spur}$, verwenden die Energiedepositionen im instrumentierten Eisen. Im Netzwerk CCCosmic liegen die Relevanzen der Variablen $p_{\perp,Spur}$ und $p_{\perp,Eisen}$ um eine Größenordnung über denen der restlichen Eingabevariablen. In dem mit Pseudo-CC's trainierten Netz CCCosmic2 (vgl. Abbildung 6.22) hat die Variable $p_{\perp,Spur}$ die größte Relevanz gefolgt von drei Variablen mit vergleichbaren Relevanzen. Die Variable $p_{\perp,Eisen}$ hat im Vergleich an Bedeutung verloren, da das Netzwerk CCCosmic die in den MC CC-Daten mangelhafte Modellierung der Energiedepositionen im instrumentierten Eisen als Unterscheidungsmerkmal von Cosmic-Ereignissen verwendete.

In der Abbildung 6.23 sind die Relevanzen der im Netzwerk CCNC verwendeten Eingabevariablen aufgetragen. Die Eingabegrößen $p_{\perp,Cluster}$ und V_p/V_{ap} besitzen die größten Relevanzen im Netzwerk.

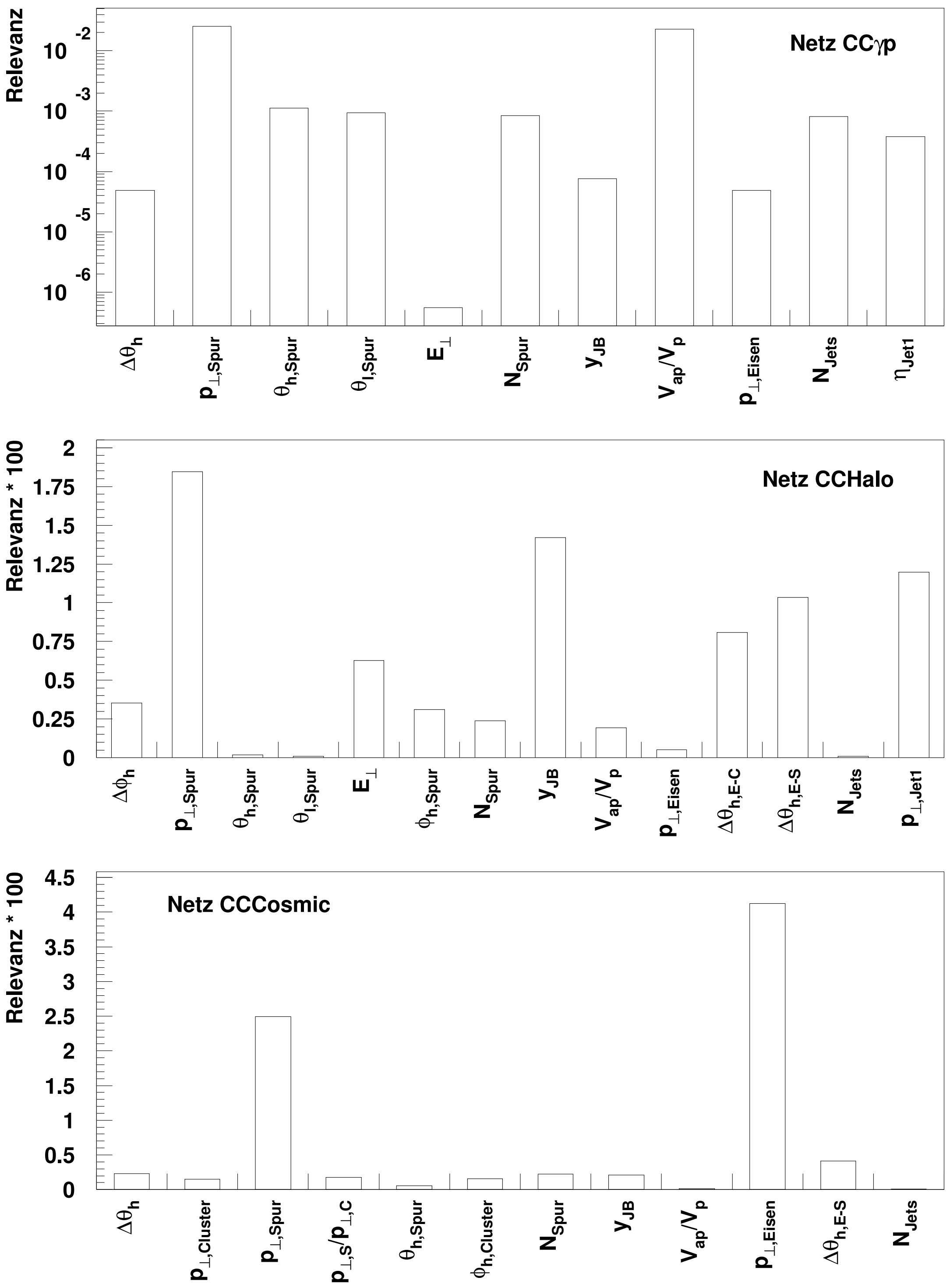


Abbildung 6.21: *Die Relevanzen der in den Netzwerken CCγp (oben), CCHalo (Mitte) und CCCosmic (unten) verwendeten Eingabevariablen. Die Relevanz ist in der oberen Abbildung im logarithmischen Maßstab aufgetragen.*

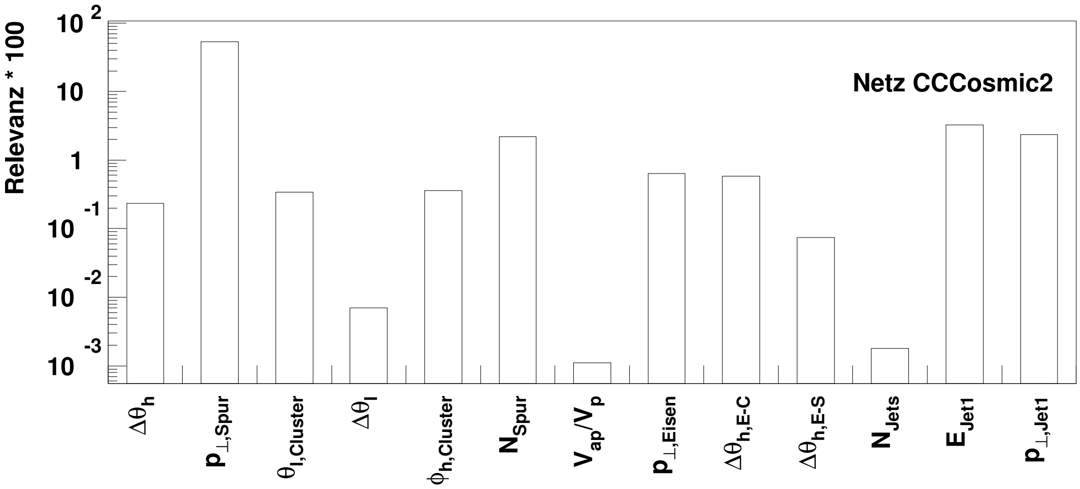


Abbildung 6.22: *Die Relevanzen der im Netzwerk CCCosmic verwendeten Eingabevariablen. Die Relevanz ist im logarithmischen Maßstab aufgetragen.*

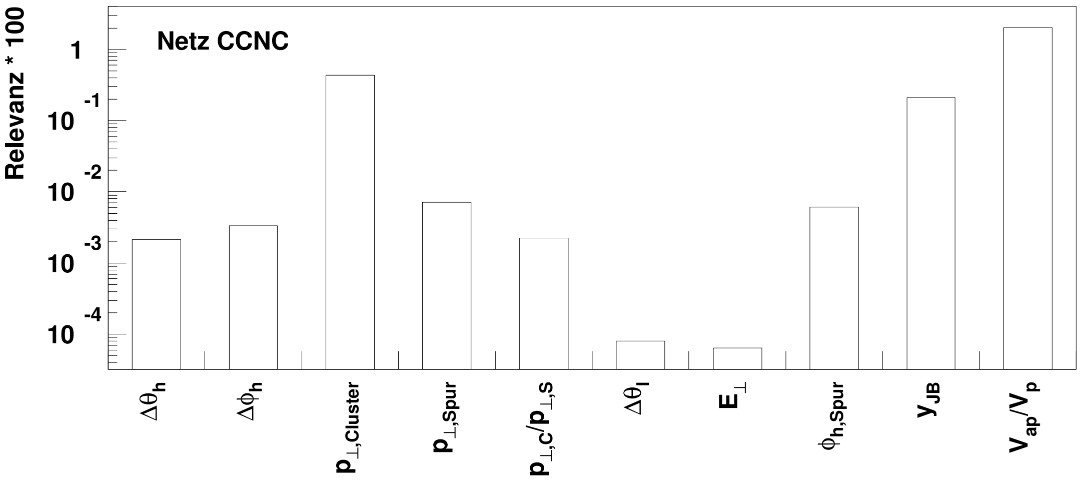


Abbildung 6.23: *Die Relevanzen der im Netzwerk CCNC verwendeten Eingabevariablen.*

6.5 Vergleiche mit bisherigen Methoden

In diesem Abschnitt werden die von den Netzwerken erreichten Effizienzen mit denen bisheriger Methoden zur Untergrundreduktion in der Ereignismenge des Geladenen Stroms verglichen. Den Schwerpunkt bildet der Vergleich der Effizienz des Netzwerkes CCγp mit der durch die Beschränkung der Variablen V_{ap}/V_p erreichten Effizienz.

In der Tabelle 6.11 wird die CC-Effizienz des Netzwerkes CCγp mit den durch die Einschränkung der Variablen V_{ap}/V_p auf Werte kleiner 0,13 bzw. kleiner 0,2 erreichten Effizienzen verglichen. Dazu werden die CC-Datensätze CCMC, PSCC und CCDATA verwendet.

	Effizienzen der Methoden		
CC-Datensätze	Netz CCγp	$V_{ap}/V_p < 0{,}13$	$V_{ap}/V_p < 0{,}2$
CCMC	$(96{,}4 \pm 0{,}1)\%$	$(87{,}9 \pm 0{,}2)\%$	$(94{,}8 \pm 0{,}1)\%$
PSCC	$(93{,}1 \pm 0{,}3)\%$	$(85{,}1 \pm 0{,}5)\%$	$(93{,}4 \pm 0{,}3)\%$
CCDATA	$(97{,}7 \pm 0{,}6)\%$	$(90{,}9 \pm 1{,}1)\%$	100%

Tabelle 6.11: *Vergleich der CC-Effizienzen der verschiedenen Methoden.*

Der Vergleich zeigt, daß die vom neuronalen Netzwerk erreichten Effizienzen, für die drei CC-Datensätze um ca. 8% über denen der $V_{ap}/V_p < 0{,}13$-Methode liegen. Die Effizienzen die aus der Beschränkung der Variablen V_{ap}/V_p auf Werte kleiner 0,2 resultieren, sind mit denen des Netzwerkes vergleichbar. Die von der $V_{ap}/V_p < 0{,}2$-Methode erreichte Effizienz von 100% beim Test mit dem Datensatz CCDATA, beruht auf der Selektion der Ereignisse mit der Bedingung $V_{ap}/V_p < 0{,}2$.

	Reinheiten der Methoden		
Datensätze	Netz CCγp	$V_{ap}/V_p < 0{,}13$	$V_{ap}/V_p < 0{,}2$
γpres3	$(99{,}4 \pm 0{,}3)\%$	100%	100%
γdir3	$(99{,}0 \pm 0{,}3)\%$	$(99{,}8 \pm 0{,}2)\%$	$(99{,}1 \pm 0{,}3)\%$
γpres1	$(96{,}0 \pm 1{,}2)\%$	$(99{,}0 \pm 0{,}6)\%$	$(97{,}2 \pm 1{,}1)\%$

Tabelle 6.12: *Vergleich der mit den verschiedenen Methoden erreichten Reinheiten eines CC-Datensatzes.*

In der Tabelle 6.12 werden die Reinheiten der drei Methoden CCγp, $V_{ap}/V_p < 0{,}13$ und $V_{ap}/V_p < 0{,}2$ miteinader verglichen. Dazu wurde der Datensatz CCMC und die γp-Datesätzen γpres3, γpdir3 und γpres1 verwendet. Die beim Test mit den Datensätzen γpres3 und γpdir3 erzielten Reinheiten betragen über 99% und sind für alle drei Methoden im Rahmen der statistischen Fehler vergleichbar. Beim Test mit dem Datensatz γpres1 werden von den drei Methoden weniger hohe Reinheiten ereicht. Im Datensatz γpres3 besitzen ca. 10% der Ereignisse nur einen Teilchenjet. Im Gegensatz dazu, sind im Datensatz γpres1 ca. 20% der Ereignisse von diesem Typ. An der in Abbildung 6.24 dargestellten Abhängigkeit der γp-Effizienz des Netzes von der Anzahl der Jets N_{Jets}, erkennt man, daß gerade die Ereignisse mit einer geringen Effizienz erkannt werden, die nur einen Jet besitzen.

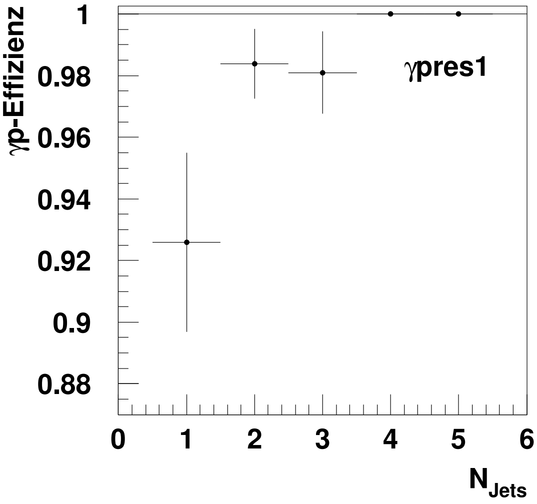
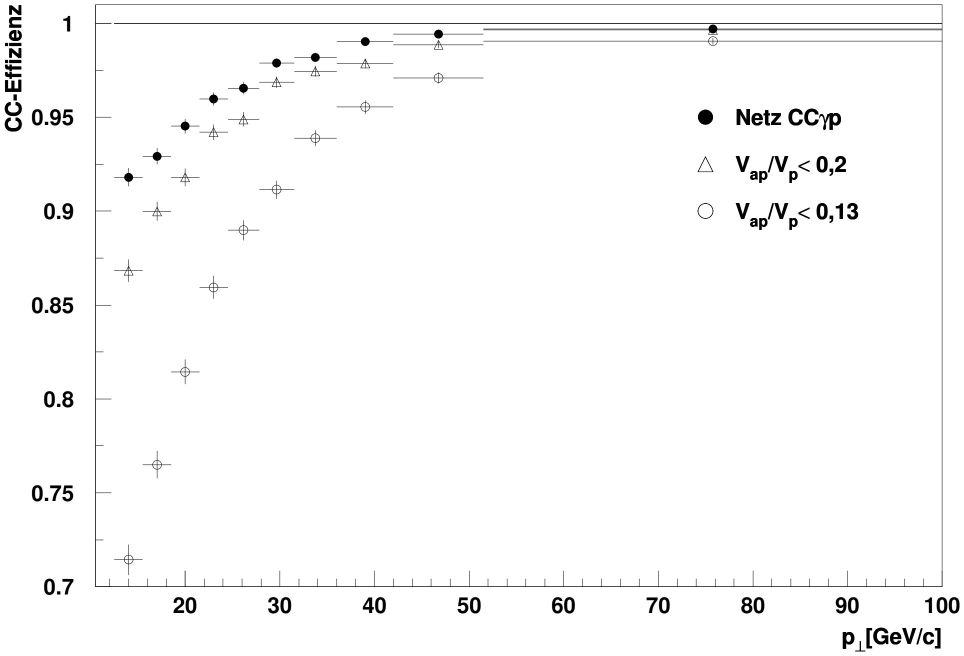


Abbildung 6.24: *Abhängigkeit der γp-Effizienz des Netzwerkes CCγp von der Anzahl der Teilchenjets N_{Jets} beim Test mit dem MC-Datensatz γpres1.*

In den Abbildungen 6.25 (oben) und 6.25 (unten) werden die Abhängigkeiten der von den drei Methoden ereichten CC-Effizienzen vom transversalen Impuls $p_\perp$ miteinander verglichen. Wird beim Test der Effizienz der MC-Datensatz CCMC verwendet (Abbildung 6.25 oben), liegt die CC-Effizienz des Netzes CCγp in allen Bins über denen der V_{ap}/V_p-Methoden. Insbesondere im Bereich $p_\perp < 20$ GeV übersteigt die Effizienz des Netzes die der $V_{ap}/V_p < 0{,}2$-Methode um bis zu 5%. Die CC-Effizienz der $V_{ap}/V_p < 0{,}2$-Methode liegt in allen $p_\perp$-Bins über der Effizienz der $V_{ap}/V_p < 0{,}13$-Methode.

Wird beim Test der CC-Effizienz der Pseudo-CC-Datensatz PSCC verwendet (Abbildung 6.25 unten), ereicht das neuronale Netzwerk im Bereich $p_\perp < 25$ GeV eine höhere Effizienz als die V_{ap}/V_p-Methoden. In den Bins des Bereiches $p_\perp > 25$ GeV übersteigt die Effizienz der $V_{ap}/V_p < 0{,}2$-Beschränkung die des Netzes oder ist mit ihr vergleichbar.

Das neuronale Netzwerk CCγp ist in der Lage insbesondere Ereignisse mit einem transversalen Impuls $p_\perp < 25$ GeV mit höherer Effizienz zu erkennen als bisherige Methoden. Die vom Netzwerk erreichten Reinheiten sind mit denen der diskutierten bisherigen Methoden vergleichbar.



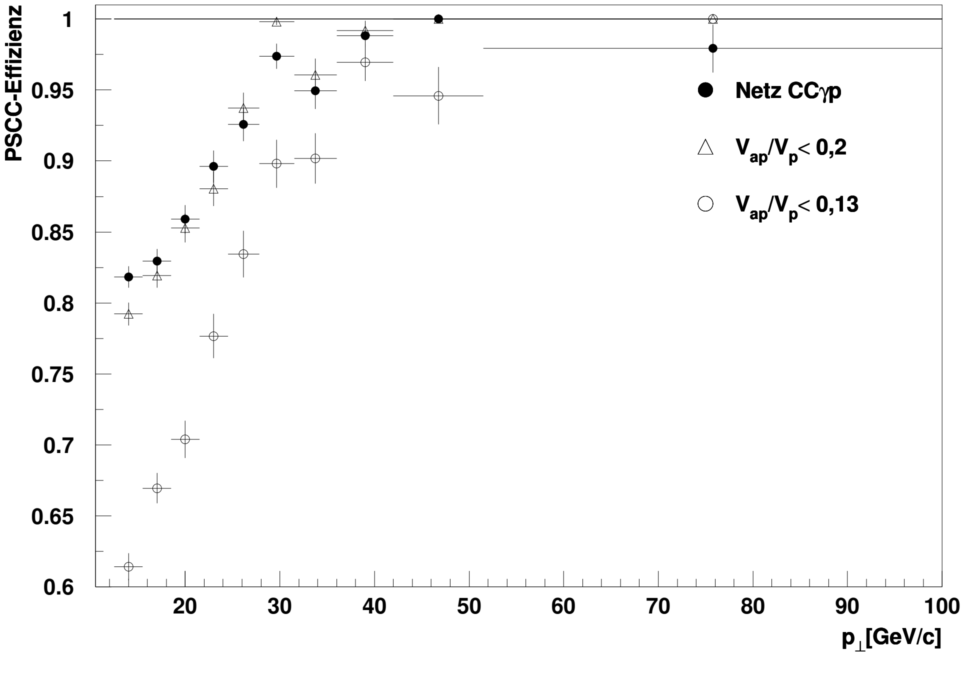


Abbildung 6.25: *Die Abhängigkeit der CC-Effizienz vom transversalen Impuls* $p_\perp$ *beim Test mit den MC-Datensätzen CCMC (oben) und PSCC (unten). Es werden die Methoden CC*γp *(Punkte),* $V_{ap}/V_p < 0{,}13$ *(Kreise) und* $V_{ap}/V_p < 0{,}2$ (Dreiecke) verglichen.

Um das Verhalten des neuronalen Netzes in Abhängigkeit von den kinematischen Variablen x_{JB}, y_{JB} und Q^2_{JB} zu untersuchen, wird im folgenden die CC-Effizienz des Netzes als Funktion dieser Größen mit den Effizienzen der V_{ap}/V_p-Methoden verglichen.

In der Abbildung 6.26 sind die von den einzelnen Methoden erreichten CC-Effizienzen in Abhängigkeit von den kinematischen Variablen x_{JB}, y_{JB} und Q^2_{JB} beim Test mit dem MC-Datensatz CCMC dargestellt. Die betrachteten Bereiche der kinematischen Variablen wurden in jeweils vier Bins unterteilt. Die Abbildungen zeigen, daß die CC-Effizienz beim Test des Netzwerkes mit dem MC-Datensatz in sämtlichen Bins über denen der V_{ap}/V_p-Methoden liegt. Besonders in den Bereichen ($0{,}01 < x_{JB} \leq 0{,}023$) und ($300\mathrm{GeV}^2/\mathrm{c}^2 < Q^2_{JB} \leq 700\mathrm{GeV}^2/\mathrm{c}^2$) erzielt das Netzwerk um ca. 5% bzw. ca. 3% höhere Effizienzen. Die CC-Effizienz des Netzes steigt mit wachsenden Q^2_{JB} bzw. x_{JB} an und fällt mit wachsendem y_{JB} ab. Dieses Verhalten des Netzwerkes stimmt mit dem der V_{ap}/V_p-Methoden überein.

Die Abbildung 6.27 zeigt den Vergleich zwischen den Effizienzen der V_{ap}/V_p-Methoden und des Netzes unter Verwendung des Pseudo-CC-Datensatzes PSCC. Im Rahmen der statistischen Fehler sind die CC-Effizienzen der $V_{ap}/V_p < 0{,}2$-Beschränkung mit denen des Netzwerks vergleichbar. Das Netz besitzt in allen Bins Effizienzen, die um bis zu 10% über denen mit dem $V_{ap}/V_p < 0{,}13$-Schnitt ereichten liegen. Der beim Test mit MC-Daten beobachtete Anstieg der CC-Effizienz des Netzes mit wachsendem Q^2_{JB} bzw. x_{JB} und der Rückgang derselben mit wachsendem y_{JB}, wird im Test mit Pseudo-CC-Ereignissen bestätigt.

Im folgenden sollen die Effizienzen der Netze zur Erkennung von NC-, Halo- und Cosmic-Ereignissen mit denen bisheriger Methoden verglichen werden.

Das Netzwerk CCNC zur Erkennung von NC-Ereignissen in der CC-Ereignismenge erreicht beim Test mit Pseudo-CC-Daten eine CC-Effizienz von $\varepsilon_{cc} = (99{,}2 \pm 0{,}2)\%$. Die CC-Effizienz der in Abschnitt 5.2.3 beschriebenen Spurmethode beträgt im Vergleich dazu $\varepsilon_{cc} = (98{,}3 \pm 0{,}3)\%$.

Die Effizienzen der Netzwerke CCHalo und CCCosmic2 zur Erkennung der durch Myonen induzierten Untergrundereibnisse, werden in der Tabelle 6.13 mit denen der topologischen Filter verglichen [Neg96]. Der Vergleich zeigt, daß die Netzwerke die CC-Effizienzen der topologischen Filter nicht erreicht haben. Dies beruht vermutlich auf der zu geringen Anzahl von Ereignissen mit großem $p_\perp$ im Pseudo-CC-Datensatz der zum

	Datensatz			
Methode	PSCC	CCMC	HALOS	COSMICS
Halo-Netz	(95,7±0,3)%	(97,6±0,1)%	(96,0±0,6)%	/
Cosmic-Netz	(96,4±0,2)%	(95,6±0,1)%	/	(85,7±0,3)%
Topo.Filter	99,3%	99,7%	94,4%	86,6%

Tabelle 6.13: *Vergleich der Effizienzen der Netze CCHalo und CCCosmic2 mit denen der topologischen Filter*

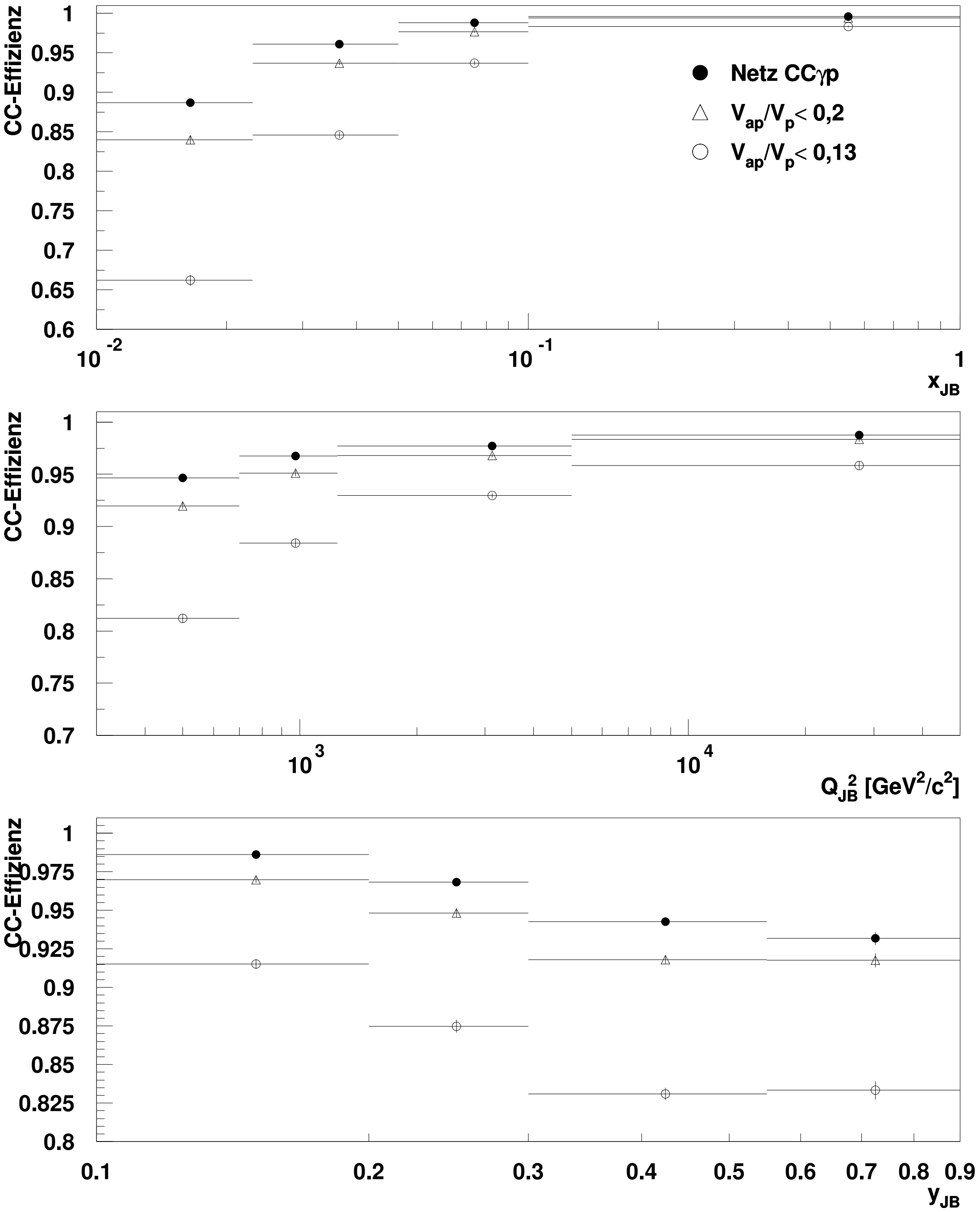


Abbildung 6.26: *Die CC-Effizienz des Netzwerkes CCγp in Abhängigkeit von den kinematischen Variablen x_{JB} (oben), Q^2_{JB} (Mitte) und y_{JB} (unten) beim Test mit dem Datensatz CCMC.*

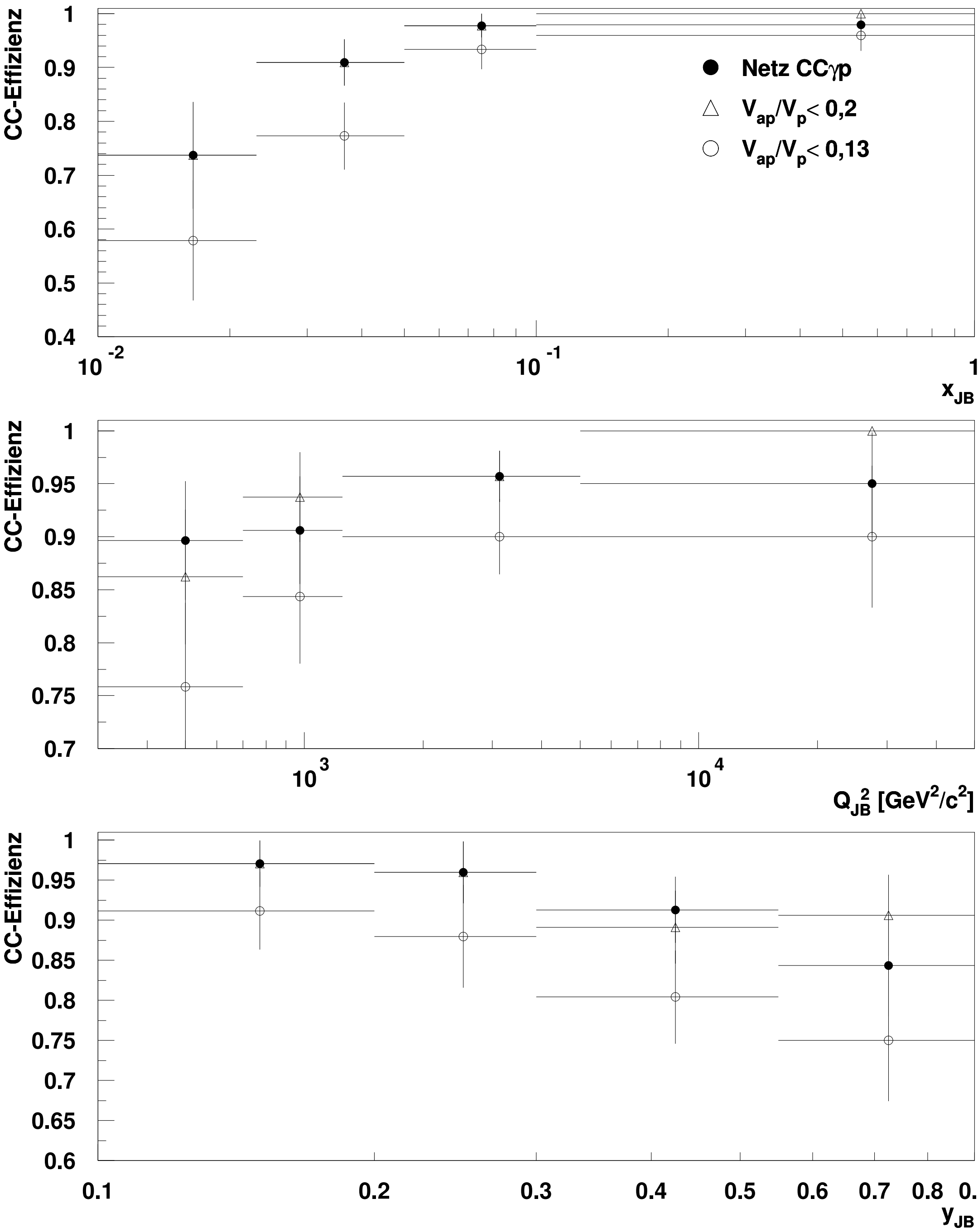


Abbildung 6.27: *Die CC-Effizienz des Netzwerkes CCγp in Abhängigkeit von den kinematischen Variablen x_{JB} (oben), Q^2_{JB} (Mitte) und y_{JB} (unten) beim Test mit dem Datensatz PSCC.*

Training verwendet wurde. Die Halo-Effizienz des Netzes CCHalo wurde mit dem Datensatz HALO96 und mit 29 nur durch visuelle Inspektion gefundenen Halo-Ereignissen bestimmt. Sie liegt ca. 2% über der Halo-Effizienz der topologischen Filter. Die Cosmic-Effizienz des Netzes CCCosmic2 wurde mit dem Datensatz COSMIC96, der nur von den Filtern identifizierte Cosmic-Ereignisse enthält, getestet und ist daher kleiner als die Cosmic-Effizienz der Filter.

Im folgenden Abschnitt soll versucht werden, durch die Einführung spezieller topologischer Eingabevariablen, die auch in den topologischen Filtern verwendet werden, Netzwerke mit höheren Effizienzen zu entwickeln.

6.6 Netze mit topologischen Eingabevariablen

In diesem Abschnitt werden zwei Netzwerke vorgestellt, die spezielle topologische Variablen der Halo- und Cosmic-Filter als Eingabegrößen verwenden [Cha98]. Die Variablen sind geeignet die typischen Signaturen der Halo- und Cosmic-Ereignisse im H1-Detektor zu beschreiben. Sie beinhalten nur Informationen über die Topologie und beruhen nicht auf den kinematischen Größen der Ereignisse. Die Netze wurden, analog dem in den vorigen Kapitel beschriebenen Methoden, mit dem genetischen Algorithmus erzeugt.

Zum trainieren und testen der Netzwerke standen folgende Datensätze zur Verfügung:

- 38519 Monte Carlo CC-Ereignisse mit einem transversalen Impuls $p_\perp > 12{,}0$ GeV.
- 854 Daten CC-Ereignisse mit einem transversalen Impuls $p_\perp > 12{,}0$ GeV.
- 2319 Halo-Ereignisse aus den Daten des Jahres 1996 mit einem transversalen Impuls $p_\perp > 25{,}0$ GeV. 69 dieser Ereignisse wurden nicht von den Filtern identifiziert.
- 1575 Cosmic-Ereignisse aus den Daten des Jahres 1996 mit einem transversalen Impuls $p_\perp > 25{,}0$ GeV. 204 dieser Ereignisse wurden nicht von den Filtern identifiziert.

Es standen insgesamt 20 topologische Variablen zur Verfügung. Hier sollen nur die in den Netzwerken als Eingabegrößen verwendeten vorgestellt werden. Es sind dies:

- Die maximale Anzahl *n1* von Sektoren im LAr-Kalorimeter mit Energiedepositionen über der Rauschschwelle entlang z bei konstantem Azimutwinkel φ . Das Kalorimeter wird dabei in 36 φ- und 26 z-Sektoren unterteilt.
- Die nächst kleinere Anzahl *n11* von Sektoren im LAr-Kalorimeter mit Energiedepositionen über der Rauschschwelle entlang z bei konstantem Azimutwinkel φ.
- Die Energiesumme $e3 = E(\varphi_3)$ entlang z, wobei φ_3 der Azimutwinkel ist bei dem eine maximale Anzahl von z-Sektoren mit Energien $E(\varphi_3, z)$ oder $E(\phi_3 \pm 10°, z)$ auftritt.

- Die nächst kleinere Energiesumme *e33* $= E(\varphi)$ entlang z, wobei φ nicht gleich $\varphi_3 \pm 10$ oder φ_3 ist.
- Der minimale transversale Abstand *dmubw* des Schwerpunktes der Energiedepositionen bei φ_3 vom Treffer eines Myons in der hinteren Endkappe des Myon-Systems.
- Das Verhältniss *eoutf* der Energie ausserhalb eines Zylinders mit Radius $r = 25$ cm um den Cluster mit dem höchsten transversalen Impuls im IF-Bereich (vgl. Abbildung 3.4) des LAr-Kalorimeters und der Gesamtenergie im LAr-Kalorimeter.
- Der minimale transversale Abstand *difsp* zwischen dem Cluster mit dem höchsten transversalen Impuls im IF-Bereich und einem Cluster im SPACAL.
- Die minimalen Polarwinkel *thmufw* bzw. *thmubw* einer Myonspur in der vorderen bzw. hinteren Endkappe des Myon-Systems.
- Das Maximum *efmu* des Verhältnisses der Energie innerhalb eines Zylinders mit Radius $r = 90$ cm um die Spur eines Myons im FB-CB-Bereich des LAr-Kalorimeters oder der hinteren Endkappe und der gesamt Energie im LAr-Kalorimeter.
- Das Maximum *eftc* des Verhältnisses der Energie eines Zylinders mit dem Radius $r = 90$ cm um die Verbindungslinie zweier diametral angeordneter Cluster im instrumentierten Eisen und der gesamt Energie im LAr-Kalorimeter.
- Das Minimum *thdtnv* des Betrags $|\theta_1 - \theta_2 - 180°|$ der Polarwinkel zweier Spuren in der zentralen Spurkammer im Winkelbereich zwischen 10° und 170°.
- Den Anteil *efemc* der elektromagnetischen Energie innerhalb eines Zylinders mit dem Radius $r = 25$ cm, der die maximale Energie bei der Zentrierung auf die Cluster im LAr-Kalorimeter enthält.
- Das Verhältniss *efout* der Energie im vorgenannten Zylinder und der Gesamtenergie im LAr-Kalorimeter.

Zum Training der Netzwerke wurden 750 Cosmics und 750 Monte Carlo CC's bzw. 1100 Halos und 1100 Monte Carlo CC's verwendet. Die Parameter und die vewendeten Eingabevariablen der Netze mit den höchsten Effizienzen sind in Tabelle 6.14 zu finden.

Der Test der Netzwerke und der Vergleich mit den Effizienzen der topologischen Filter wurde mit den kompletten zur Verfügung stehenden Datensätzen durchgeführt. Die Ergebnisse sind in den Tabellen 6.15 und 6.6 zusammengefaßt.

Das Halo-Netzwerk erreicht Effizienzen, die mit denen der Halo-Filter vergleichbar sind. Zusätzlich erkennt das Netzwerk ca. 35% der Halo-Ereignisse, die nicht von den Filtern identifiziert wurden.

Eingabe Knoten	2.Ebene Knoten	3.Ebene Knoten	$1/T$	w_{init}	Eingabegrößen
Halo-Netzwerk					
8	16	26	0,4	0,06	n1 n11 e33 dmubw eoutf difsp thmufw thmubw
Cosmic-Netzwerk					
10	19	13	1,9	0,02	n11 e3 e33 dmubw thmubw efmu eftc thdtnv efemc efout

Tabelle 6.14: *Parameter und Eingabevariablen des Netzwerks zur Trennung von CC- und Halo- bzw. Cosmic-Ereignissen.*

	Effizienzen	
Datensätze	Halo-Filter	Netzwerk
MC CC	(98,7 ± 0,1)%	(99,0 ± 0,1)%
CC Daten	(98,7 ± 0,4)%	(98,0 ± 0,5)%
Halos	(97,0 ± 0,3)%	(97,1 ± 0,3)%
Halos(Filt.)	100%	(99,1 ± 0,2)%
Halos(Scan.)	0%	(34,8 ± 5,6)%

Tabelle 6.15: *Vergleich der Effizienzen des Halo-Netzwerks mit denen der Halo-Filter*

Die CC-Effizienzen des Cosmic-Netzes sind mit denen der Cosmic-Filter vergleichbar. Die Effizienz der Cosmic-Erkennung des Netzes liegt um 9% über der Effizienz der Cosmic-Filter. Insbesondere können mit dem Netzwerk 73% der Ereignisse identifiziert werden, die nicht mit den Filtern gefunden wurden.

	Effizienzen	
Datensätze	Cosmic-Filter	Netzwerk
MC CC	(98,7 ± 0,1)%	(98,1 ± 0,1)%
CC Daten	(98,7 ± 0,4)%	(98,0 ± 0,5)%
Cosmics	(87,1 ± 0,9)%	(96,4 ± 0,5)%
Cosm.(Filt.)	100%	(99,8 ± 0,1)%
Cosm.(Scan.)	0%	(73,4 ± 3,1)%

Tabelle 6.16:
Vergleich der Effizienzen des Cosmic-Netzwerks mit denen der Cosmic-Filter

Die Verwendung der topologischen Variablen als Eingabegrößen neuronaler Netzwerke ermöglicht es, Netze zu entwickeln, mit denen die Ereignisklassen des geladenen Stroms und der Halos bzw. Cosmics mit hoher Effizienz getrennt werden können. Dabei sind diese Variablen noch nicht optimal an die Verwendung in neuronalen Netzwerken angepasst. Die Berechnungen einiger der Variablen liefert nur für bestimmte Ereignistopologien (z.B. zwei gegenüberliegende Cluster im Kalorimeter eftc) Werte, für die übrigen Ereignisse wird die Variable auf einen bestimmten Wert gesetzt. Dadurch entstehen binäre Eingabeinformationen, welche die in den Verteilungen der Variablen enthaltenen Informationen überdecken können.

Kapitel 7

Zusammenfassung und Ausblick

In dieser Arbeit wird untersucht, ob eine Erkennung von Ereignissen des geladenen Stroms und deren Separierung von Untergrundereignissen, mit neuronalen Netzwerken möglicht ist. Es werden eine Reihe von Netzwerken zur Unterscheidung der CC-Ereignisse von den Ereignissen einer Untergrundklasse trainiert und getestet. Die Parameter und die Auswahl der Eingabevariablen des Netzwerkes werden mit einem genetischen Algorithmus optimiert. Innerhalb des genetischen Algorithmus werden die Eingabegrößen aus einer Menge von möglichen Eingabevariablen so ausgewählt, daß eine maximale Effizienz erreicht wird. Die entwickelten Netzwerke erreichen bei der CC-Erkennung und der Separierung der Untergrundereignisse eine hohe Effizienz, die in allen Netzen über 90% liegt. Die Effizienzen der Netzwerke werden mit denen bisheriger Methoden zur Untergrundreduktion verglichen.

Das Netzwerk zur Separierung der CC-Ereignisse von den Untergrundereignissen der Photoproduktion ist, insbesonderen bei kleinem transversalen Impuls $p_\perp$ des hadronischen Endzustandes, den bisherigen Methoden überlegen. Die Analyse der Abhängigkeit der CC-Effizienz von den kinematischen Variablen x_{JB}, y_{JB} und Q^2_{JB0} zeigt, daß durch die Anwendung der Netzwerke im Vergleich zu den bisherigen Methoden keine physikalischen Informationen verloren gehen.

Das Netzwerk zur Trennung von Ereignissen des geladenen und neutralen Stroms erzielt CC- und NC-Effizienzen von annähernd 100%.

Die Netzwerke zur Separierung der durch Myonen induzierten Untergrundereignisse erreichen bei der Erkennung von Halo- bzw. Cosmic-Ereignissen Effizienzen, die größer bzw. vergleichbar mit denen der topologischen Filter sind. Die CC-Effizienz der Netzwerke liegt unter denen der Filter. Durch die Verwendung von speziellen topologischen Eingabevariablen, die auf den Filteralgorithmen beruhen, kann sowohl die CC-Effizienz als auch die Effizienz der Untergrundreduktion der Netzwerke verbessert werden. Die CC-Effizienzen dieser Netzwerke entsprechen denen der topologischen Filter, die Effizienzen der Untergrundreduktion übersteigen die der Filter.

Eine weitere Effizienzsteigerung der Netzwerke ist durch folgende Möglichkeiten zu erzielen:

- Die Konstruktion von weiteren Eingabevariablen und die Anwendung des genetischen Algorithmus auf diese vergrößerte Auswahl.
- Die Erhöhung der Anzahl von Trainingsmustern in Hinblick auf Ereignistypen die von den Netzwerken mit geringer Effizienz erkannt werden.
- Das Training von mehreren Netzen für bestimmte Bereiche der Eingabevariablen (z.B. für Transversalimpulse $p_\perp < 20$ GeV und $p_\perp \geq 20$ GeV im Fall der Erkennung von Untergrundereignissen der Photoproduktion).

Neuronale Netzwerke sind zur Selektion von Ereignissen des geladenen Stroms geeignet und erreichen hohe Effizienzen die zum Teil über denen der herkömmlichen Methoden liegen. Die Selektion von CC-Ereignissen mit neuronalen Netzwerken führt im Vergleich mit bisherigen Methoden zu keinem Verlust von physikalischen Informationen.

Abbildungsverzeichnis

Tabellenverzeichnis

Literaturverzeichnis

[Abt96] I. Abt et al.: *The H1 Detector at HERA* Internal Report, DESY H1-96-01 (1996)

[Alb95] H. Albrecht et al.: *A guide to PHAN an H1 physics analysis package* H1PHAN Manual (1995)

[Bec94] K. H. Becks: *Genetische Algorithmen in der Elementarteilchenphysik* Physikalische

[Blo92] V. Blobel et. al.: *Central Tracker Software documentation* H1 Software Note 30 (1992)

[Cha98] Dr. Claude Vallee u. Eric Chabert : Persönliche Mitteilung (1998)

[H194] H1 Collaboration: *First Measurement of the Charged Current Cross Section at HERA* Desy Report 94-012, Physical Lettrs B 324 (1994)

[H195] H1 Collaboration: *Measurement of the* e^+ *and* e^- *induced Charged Current cross section at HERA* Desy Report 95-102 (1995)

[H195b] H1 Collaboration: *Jets and Energy Flow in Photon-Proton Collisions at HERA* DESY 95-219 (1995)

[Hei96] Beate Heinemann : *Messung des transversalen Energieflusses bei hohen Impulsüberträgen mit dem H1-Detektor bei HERA* Univ. Hamburg, Diplomarbeit (1996)

[Hin92] G. E. Hinton: *Wie neuronale Netze aus Erfahrungen lernen* Spektrum der Wissenschaft (11/1992)

[Hop97] M. Höppner: *Anwendung neuronaler Netze in der Analyse hadronischer Schauer in feinsegmentierten Flüssigargon-Kalorimetern* Dissertation Uni. Dortmund (1997)

[Jac79] A. Blondel, F. Jacquet: *Proceedings of the study of an ep facility for Europe* DESY 79-48 (1979)

[Koe97] J. H. Köhne et al.: *Realization of a second level neural network trigger for the H1 experiment at HERA* internal Note, DESY H1-01/97-509

[Ku94] M. Kunze: *Application of artificial neural networks in the analysis of multi-particle data* Proceedings of the CORINNEII conference, Nantes World Scientific (1994)

[Leb96] A. I. Lebedev: *Study of Photoproduction Processes at HERA* DESY H1-08/97-527 (1997)

[Lev96] A. Levy: *Low-x physics at HERA* TAUP 2398-96 (1996)

[Lis92] P. G. J. Lisboa: *Neural networks current applications* Chapman and Hall London (1992)

[Lon94] L. Lönnblad: *Jetnet 3.0 - A versatil artificial neural network package* CERN-TH.7135/94

[Lue97] St. Lueders: *Erkennung semimyonischer Zerfälle schwerer Quarks durch Neuronale Netze am H1-Detektor bei HERA* Diplomarbeit Uni. Hamburg (1997)

[Mar97] Günter Martin: *Untersuchung von Prozessen des geladenen und neutralen Stroms mit dem H1-Detektor bei HERA* Univ. Hamburg, Dissertation (1997)

[Mey97] P. O. Meyer: *Detector Simulation* H1 Workshop Seminar Cross Talk (1997)

[MRS93] A. D. Martin, R. G. Roberts, W. J. Stirling: *Proc. on the Workshop on Quantum Field Theory, Theor. Aspects of QED, eds. B. Geyer and M. Igenfritz* (1993)

[Neg96] I. Negri et al.: *A Minimal Comprehensive Set of Muon Background Topological Finders For High $p_\perp$ Physics Analysis* H1-Note H1-10/96-498 (1996)

[Nez95] P. Di Nezza: *Application of neural networks to the analysis of HERA physics with the H1 apparatus* H1-Note H1-12/95-468 (1995)

[Per90] D. H. Perkins: *Introduction to High Energy Physics* Addison-Wesley Publishing London (1990)

[Pre96] A. Prell: *Entwicklung des Steuer- und Auslesemoduls für die Siliziumstreifendetektoren des H1-Detektors* Univ. Hamburg, Dissertation (1996)

[Ried93] M. Riedmiller, H. Braun: *A direct adaptiv methode for faster backpropagation learning:The RPROP algorithm* Proceedings of the IEEE Intern.Confer.on NN. San Francisco (1993)

[Rie98] Dr. Stefan Riess : Persönliche Mitteilung (1998)

[Rud94] G. Rudolph: *Evolutionäre Algorithmen ein robustes Optimierungskonzept* Physikalische Blätter (3/1994)

[Sch95] Peter Schmüser: *Feynman-Graphen und Eichtheorien für Experimentalphysiker* Springer (1995)

[She93] V. Shekelyan: H1 note H1-04/93-288 (1993)

[Sjo92] T. Sjöstrand: *PYTHIA at HERA* Proceedings of the Workshop, Physics at HERA (1992)

[Spi97] H. Spiesberger: *Physics with DJANGO6 and HERACLES* Interner Vortrag, DESY (1997)

[Ste97] J. Steinhart: *Central Drift Chambers* H1 Workshop Seminar Cross Talk (1997)

[Swa97] M. Swart: *SPACAL Calorimeter* H1 Workshop Seminar Cross Talk (1997)

Danksagung

An dieser Stelle möchte ich mich bei allen Menschen bedanken, die zum Gelingen dieser Arbeit beigetragen haben. Mein besonderer Dank gilt

- Herrn Prof. Dr. Büßer für die Aufgabenstellung und das stete Interesse sowie die freundliche Betreuung,
- Herrn Dr. Stefan Rieß für die vielen Anregungen und den fundierten fachlichen Rat,
- Herrn Arndt Boris Fahr und Herrn Robert Casties für die Unterstützung bei Problemen zur Rechnernutzung,
- Herrn Arndt Boris Fahr und Herrn Dirk Laube für das Korrekturlesen dieser Arbeit,
- Herrn Dr. Claude Vallee und Herrn Eric Chabert für die Bereitstellung der topologischen Eingabevariablen,
- Herrn Dr. Friedrich Niebergall für das stete Intresse am Gelingen dieser Arbeit,
- Herrn Prof. Dr. Heinzelmann für die Bereitschaft als Gutachter zur Verfügung zu stehen,
- allen Mitgliedern der EW-Gruppe der H1-Kollaboration,
- und meinem Bruder Bernd Rauschenberger für die moralische und finanzielle Unterstützung.

Ich versichere, daß ich die vorliegende Arbeit unter Angabe aller verwendeter Referenzen selbstständig angefertigt habe.