Untersuchung von $D^{*\pm}$-Ereignissen am H1–Experiment unter Berücksichtigung des Level-2-Triggers

Dissertation an der Fakultät für Physik
der
Ludwig–Maximilians–Universität München

von
Andreas Gruber
aus Bad Säckingen

München
10. Juni 1997

1. Gutachter: Priv. Doz. Dr. C. Kiesling
2. Gutachter:

Tag der mündlichen Prüfung:

A physicist, an engineer and a computer scientist were discussing the nature of God. “Surely, he is a physicist” said the physicist, “because early in the creation God made light and you know Maxwell’s equations, the dual nature of electromagnetic waves, the relativistic consequences ...”.
“He is an engineer !”, said the engineer, “because before making light, God split the chaos into land and water. It takes a hell of an engineer to handle that big amount of mud and orderly separate the solids from the liquids ...”.
The computer scientist just mumbled, “.... and the chaos ? Where do you think the chaos is coming from, hmm ?”.

— Anonymous [1] —

Inhaltsverzeichnis

Motivation

Unserem heutigen Verständnis vom Aufbau der Materie liegt das sogenannte *Standardmodell der elektroschwachen und starken Wechselwirkung* [2–4] zugrunde. Es beschreibt die Zusammensetzung der Materie aus Fermionen (Quarks und Leptonen) und die Wechselwirkungen der Fermionen durch den Austausch von Eichbosonen (Photonen, W– und Z–Teilchen, Gluonen).

Eines der wichtigsten Hilfsmittel auf dem Weg zum Standardmodell waren, abgesehen von Papier und Bleistift der theoretischen Physiker, eine Vielzahl von Teilchenbeschleunigern und Detektoren. Mit steigender Energie der Teilchen wurde es möglich, immer feinere Strukturen der Materie aufzulösen und damit die Vorhersagen der Theorie zu evaluieren. Bislang ist das Standardmodell in allen zugänglichen Bereichen bestätigt worden, wobei jedoch berücksichtigt werden muß, daß es eine große Anzahl von freien Parametern besitzt, wie z.B. die relativen Stärken der Kopplungskonstanten, die Quarkmassen und die Elemente der Cabibbo-Kobayashi-Maskawa–Matrix [5, 6]. Bis zur Entwicklung einer übergeordneten Theorie müssen diese Größen experimentell bestimmt werden.

Streuexperimente mit Elektronen und Protonen gaben Hinweise darauf, daß Nukleonen keine punktförmigen Objekte sind. Gell-Mann [7] und Zweig [8] schlugen aufgrund von Ergebnissen in der Baryonenspektroskopie vor, die Nukleonen als zusammengesetzte Gebilde zu betrachten, die aus Quarks bestehen und durch Gluonen zusammengehalten werden.

Bei der Berechnung des Wirkungsquerschnitts von Lepton-Nukleon Streuexperimenten fließen die Strukturfunktionen ein, die eine Beschreibung des Quarkgehalts des Nukleons darstellen. Im Bereich hoher Impuls- und Energieüberträge sind die Strukturfunktionen nur noch von einer kinematischen Variablen abhängig. Dieses von Bjorken vorhergesagte Phänomen wird als Skalenhypothese [9] bezeichnet. Es bildet die Basis des Quark–Parton–Modells, in dem der Streuprozeß als elastische Streuung des Leptons an einem Quark beschrieben wird.

Eine exakte Berechnung innerhalb der Quantenchromodynamik berücksichtigt auch die starke Wechselwirkung und Effekte höherer Ordnung, in die Gluonen involviert sind. Diese Rechnungen führen zu Abweichungen von der Skalenhypothese und machen die Strukturfunktion von zwei kinematischen Variablen abhängig. Sind jedoch die Strukturfunktionen bei einer bestimmten Kinematik vollständig bestimmt, erlaubt die Quantenchromodynamik mit Hilfe der Altarelli-Parisi Evolutionsgleichungen [10] im Prinzip die Berechnung der Strukturfunktionen bei jeder beliebigen Kinematik. Die vollständige Vermessung der Strukturfunktionen ist deshalb eine Vorbedingung für die exakte Bestimmung der freien Parameter des Standardmodells, da die Ergebnisse von Nukleon-Streuungen nur bei bekannten Strukturfunktionen mit hoher Präzision ausgewertet werden können.

Die Speicherringanlage HERA bietet ein neuartiges Experimentierfeld für die Untersuchung von Lepton–Nukleon Kollisionen. Hauptinteressengebiete sind die Bestimmung der Protonstrukturfunktionen in bisher unzugänglichen kinematischen Bereichen [11–13], die Ermittlung elektroschwacher Parameter [14], und die Messung der Dichteverteilung der Gluonen im Nukleon [15]. Die hohe Luminosität, die der Beschleuniger zu Verfügung stellt und der erwartete Wirkungsquerschnitt für die Produktion von Charm–Quarks machen HERA zu einer reichen Quelle für Charm–Ereignisse. Die Zahl von Hadronen mit Charm pro 10 pb^{-1} integrierter Luminosität liegt in der Größenordnung von $\mathcal{O}(10^7)$. Sie erlaubt quantitative Tests der Quantenchromodynamik im Bereich der Schwer-Quark-Produktion. Da der dominierende Prozeß in der Charm–Erzeugung mit Gluonaustausch abläuft, bietet die Untersuchung von Ereignissen mit Charm eine ideale Möglichkeit, Informationen über den Gluonanteil innerhalb des Protons zu erhalten. Die Messung der Gluondichteverteilung bei verschiedenen Impulsüberträgen ermöglicht über die Evolutionsgleichungen Tests der QCD. Die wichtigsten Prozesse sind die Produktion von ungebundenen hadronischen Systemen mit Charm und die Erzeugung von J/Ψ–Mesonen. Daneben ermöglicht die Suche nach seltenen Zerfällen des D^0–Mesons die Festlegung von oberen Schranken für die rein leptonischen Moden.

Als Grundvoraussetzung für die obengenannten Untersuchungen müssen bei den HERA-Experimenten große Datenmengen bewältigt werden. Die hohe Rate von Untergrundereignissen in der Größenordnung von $\mathcal{O}$(100 kHz) erweist sich als extrem kritisch. Diese wird hauptsächlich durch Wechselwirkungen des Protonstrahls mit Strahlführungskomponenten oder Restgasatomen im Vakuum des Strahlrohrs erzeugt. Das Datenaquisitionssystem des H1–Experiments kann eine maximale Aufzeichnungsrate in der Größenordnung von $\mathcal{O}$(10 Hz) erreichen. Die Spitzenwerte der spezifischen Luminosität lagen bis 1995 bei 5.0×10^{30}cm^{-2}s^{-1} und damit um den Faktor drei hinter dem projektierten Wert zurück. Bei einer Verdoppelung der momentanen Luminosität überschreitet die Produktionsrate von Charm–Ereignissen, die die Gesamtrate der Physikprozesse dominiert, den Bereich der maximalen Aufzeichnungsrate. Es gilt, die hohen Untergrundraten wirksam zu senken unter der Randbedingung, daß die interessierenden Physikereignisse effektiv herausgefiltert werden.

Das H1–Experiment besitzt hierfür ein vierstufiges Triggersystem, in welchem jeder sogenannte *Level* eine Ratenreduktion um einen Faktor zehn erreichen kann. Der Level-1-Trigger ist ein herkömmlicher Hardware-Schwellentrigger, der binäre Signale aus verschiedenen Detektorkomponenten logisch zu *Subtriggern* verknüpft. Der Level 4 ist ein Software-Trigger, dessen Entscheidungen auf der vollständig ausgelesenen Detektorinformation basieren. Er kann Eingangsraten bis zu $\mathcal{O}$(50 Hz) verarbeiten. Bis Ende 1995 waren nur der Level-1-Trigger und der Level-4-Trigger realisiert. Da dem Level-1-Trigger keine detaillierteren Informationen vorliegen, ist er nicht im Stande, die ähnlichen Ereignistopologien von Untergrund– und Schwer-Quark–Reaktionen effizient zu trennen. Dies hat zur Folge, daß Subtrigger, die für Charm–Ereignisse sensitiv sind, mit einer hohen Ausgangsrate behaftet sind, da sie hauptsächlich Untergrundereignisse triggern. Bei steigender Luminosität blieben zur Einhaltung der maximalen Rate nur die Möglichkeiten des statistischen Verwerfens von Ereignissen oder der Anhebung der Schwellen. Der daraus resultierende Verlust an Ereignissen kann jedoch durch die Verwendung eines Level-2-Triggers vermieden werden, dem detailliertere Information und mehr Zeit zur Entscheidungsfindung zur Verfügung steht. Für die Auswertung der Informationen auf der Ebene des Level-2-Triggers wurden neuronale Netzwerke untersucht. Die Haupteigenschaften von neuronalen Netzwerken wie zum Beispiel die Parallelverarbeitung, die Flexibilität bei der Wahl der Eingabegrößen und die Algorithmen zum Erlernen von Musterunterschieden zwischen Physik und Untergrund, machen sie zu einem idealen Werkzeug für den Einsatz im Triggersektor.

Diese Arbeit beschäftigt sich mit der Untersuchung der $D^{*\pm}$–Produktion und der Steigerung der Triggereffizienz für Charm–Ereignisse mit Hilfe des neuronalen Level-2-Triggers. Im ersten Kapitel werden die wichtigsten physikalischen Grundlagen behandelt. Das Kapitel 2 beschreibt den Beschleuniger und den H1 Detektor mit besonderer Betonung des Triggersystems. Daran schließt sich eine ausführliche Beschreibung des Level-2-Triggers an, in der sowohl die Designprinzipien, die realisierten Algorithmen als auch die implementierte Hardware dargestellt werden. Die Selektion der Daten für die Produktion der $D^{*\pm}$–Mesonen im Jahr 1994 bildet zusammen mit einer vergleichenden Untersuchung von Monte-Carlo–Ereignissen den Kern des vierten Kapitels. Dabei wird das $D^{*\pm}$–Meson durch den Zerfall $D^{*\pm} \longrightarrow D^0\pi^\pm \longrightarrow (K\pi)\pi^\pm$ nachgewiesen. Im Kapitel 5 wird beleuchtet, wie aus diesen Daten und ihren Eigenschaften ein Level-2-Trigger für $D^{*\pm}$–Ereignisse konstruiert werden kann. Die Auswertung der Daten mit einer Bestimmung des Wirkungsquerschnitts für die $D^{*\pm}$–Erzeugung bei HERA bilden den Abschluß der Arbeit.

Kapitel 1

Die Physik der schweren Quarks

Vor etwas mehr als 20 Jahren begann die experimentelle Erforschung der schweren Quarks mit der Entdeckung des J/Ψ–Teilchens [16–18], das den gebundenen 1S Zustand aus Charm– und Anti–Charm–Quark darstellt. Das Charm–Quark ist zusammen mit dem Strange–Quark in die zweite Quark–Familie eingeordnet und besitzt als erstes Quark eine deutlich nicht–vernachlässigbare Masse im Bereich um 1.5 GeV.

Die beiden Quarks der dritten Familie (Bottom und Top) werden ebenfalls zu den schweren Quarks gezählt. Das Top–Quark wurde erst 1995 von den Experimenten CDF und D0 am Fermilab nachgewiesen und besitzt eine Masse von $176 \pm 8 \pm 10$ GeV [19], wobei der erste Fehler die statistische und der zweite die systematische Unsicherheit beschreibt. Das Up– und das Down–Quark der ersten Quarkfamilie komplettieren das Bild des Standardmodells der Teilchenphysik, das aus den drei Quarkfamilien **Top–Quark**

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} \tag{1.1}$$

und den drei Leptonfamilien

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{1.2}$$

mit Elektron, Myon, Tau und ihren Neutrinos besteht.

Mit der Untersuchung von Charm–Physik können die HERA–Experimente wichtige Ergebnisse zum Verständnis der Produktions– und Zerfallsprozesse schwerer Quarks beitragen. Die hohe Masse der Quarks erlaubt es, Vorhersagen aus dem perturbativen Bereich der Quantenchromodynamik (QCD) zu überprüfen. Bei steigender Luminosität in den kommenden Jahren wird auch die Physik der Bottom–Quarks eine wichtige Rolle spielen. Die hohe Schwerpunktsenergie erlaubt es, neue kinematische Regionen zu erkunden und als eine der wichtigsten Größen die Verteilung der Gluondichte im Proton zu messen. **Bottom–Quark**

Die folgenden Abschnitte sollen zunächst auf die grundlegenden Produktionsprozesse eingehen und die kinematischen Variablen beschreiben, die bei H1 Verwendung finden. Danach soll der wichtigste Produktionsmechanismus für Charm–Teilchen herausgegriffen und weitere Details zu dessen kinematischer Beschreibung erläutert werden. Ein Überblick über die Zerfallsmechanismen und –moden schließen dieses Kapitel ab.

1.1 Kinematik

Für die Beschreibung der Kinematik der Lepton–Proton–Stöße wird im allgemeinen das sogenannte Quark–Parton–Modell verwendet [9, 20]. In diesem Modell werden die Protonen als Teilchen betrachtet, die aus Partonen aufgebaut sind. Diese Partonen werden mit Quarks und Gluonen identifiziert. Jedes dieser Partonen trägt einen Teil des Protonimpulses und man kann die inelastische Reaktion des Leptons mit dem Proton als elastische Streuung an einem Konstituenten des Protons auffassen.

tiefinelastische Streuung
W–Boson

In Abbildung 1.1 ist ein Feynmangraph für die tiefinelastische Streuung dargestellt. In diesem Graphen der niedrigsten Ordnung kann man zwischen Reaktionen mit Austausch eines geladenen W–Bosons (*charged current* oder CC–Reaktion) und Austausch eines der neutralen Eichbosonen γ oder Z^0 (*neutral current* oder NC–Reaktion) unterscheiden.

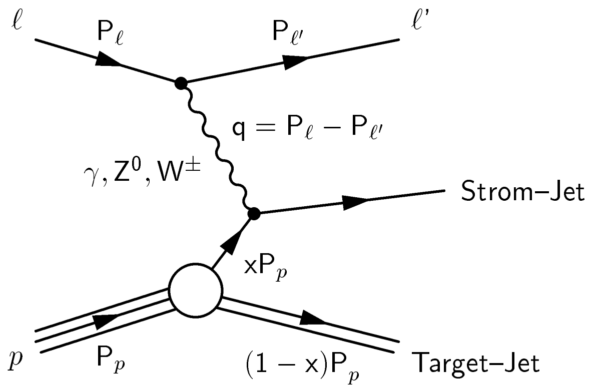

Abbildung 1.1: *Feynmangraph der tiefinelastischen Lepton–Proton–Streuung niedrigster Ordnung.*

Target–Jet

Das einlaufende Lepton (bei HERA bis Mitte des Jahres 1994 ein Elektron, danach ein Positron) wird mit ℓ bezeichnet. Das auslaufende Lepton ℓ' ist im Falle eines neutralen Stromereignisses vom selben Typ wie das einlaufende und im Falle einer Reaktion mit geladenem Strom ein Neutrino. Das Proton p und ist zur Veranschaulichung als Gruppe von drei Linien (je eine pro Valenzquark) dargestellt. Nach der Reaktion bilden die restlichen Fragmente des Protons einen sogenannten Target–Jet. Das an der Wechselwirkung beteiligte Quark bildet in den nachfolgenden Prozessen den Strom–Jet aus. **Strom–Jet**

Die zu den Teilchen gehörenden Viererimpulse sind mit P bezeichnet und tragen als Indizes die Teilchennamen. Das an der quasi-elastischen Streuung beteiligte Quark trägt den Impuls xP_p. Der Viererimpulsübertrag des Austauschteilchens (γ, Z^0 oder $W^{\pm}$) wird mit q beschrieben.

Die Kinematik der ep–Reaktion ist bereits mit zwei gemessenen unabhängigen Variablen eindeutig bestimmt, wenn man die Schwerpunktsenergie als bekannt voraussetzt. Über die bekannten Energien des Leptonstrahls (E_ℓ) und des Protonstrahls (E_p) läßt sich das Quadrat der Schwerpunktsenergie $\sqrt{s}$ bestimmen zu

$$s \equiv (P_\ell + P_p)^2 \approx 4\, E_\ell\, E_p, \tag{1.3}$$

wobei die letzte Näherung von einer Vernachlässigung der Massen der einlaufenden Teilchen ℓ und p ausgeht. Einige Lorentz–invariante Größen werden zur Beschreibung der Kinematik benutzt, so z.B. das negative Quadrat des Viererimpulsübertrags vom Lepton auf das Quark

$$Q^2 \equiv -q^2 = -(P_\ell - P_{\ell'})^2, \tag{1.4}$$

oder die Variable W , die die invariante Masse des hadronischen Endzustands mißt

$$W^2 \equiv (q + P_p)^2. \tag{1.5}$$

Die dimensionslosen Bjorken'sche Skalenvariablen x und y werden als

$$x \equiv \frac{Q^2}{2(P_p \cdot q)} \tag{1.6}$$

$$y \equiv \frac{(P_p \cdot q)}{(P_p \cdot P_\ell)} \tag{1.7}$$

definiert. Im Quark–Parton–Modell hat die Größe x die anschauliche Bedeutung des Impulsanteils des wechselwirkenden Partons am Gesamtimpuls des Protons, y beschreibt die Inelastizität des Prozesses, was im Ruhesystem des Protons dem relativen Energieübertrag des Leptons auf das Quark entspricht. Für die obigen Größen besteht in der Näherung vernachlässigter Teilchenmassen folgende Relation **Quark–Parton–Modell**

$$Q^2 = s\, x\, y. \tag{1.8}$$

Experimentell sind die definierten Größen im Fall von NC–Reaktionen über die Energie und den Winkel des gestreuten Elektrons oder des hadronischen Systems zugänglich.

Wenn Θ den Polarwinkel und $E_{\ell'}$ die Energie des gestreuten Elektrons bezeichnen, dann ergibt sich :

$$Q^2 = 4\, E_\ell\, E_{\ell'}\, \cos^2\frac{\Theta}{2} \quad \text{und} \tag{1.9}$$

$$y = 1 - \frac{E_{\ell'}}{E_\ell}\, \sin^2\frac{\Theta}{2}. \tag{1.10}$$

Die Variable x erhält man über Formel 1.8. Die Kinematik läßt sich auch bestimmen, wenn nur das hadronische System gemessen wird. Über die Methode von Jacquet–Blondel [21] lassen sich die folgenden Variablen berechnen :

$$y = \frac{1}{2E_\ell} \sum_{Hadr.} \left(\mathsf{E}_h - p_h{}^{\|} \right) \tag{1.11}$$

$$Q^2 = \frac{1}{1-y} \sum_{Hadr.} \left(p_h{}^{\perp} \right)^2 \tag{1.12}$$

Jacquet–Blondel

Die Summe erstreckt sich hierbei über alle Hadronen des Endzustands, wobei für die Berechnung von Q^2 nur die transversalen Impulskomponenten $p_h{}^{\perp}$ und für die Berechnung von y nur die gerichteten longitudinalen Komponenten $p_h{}^{\|}$ herangezogen werden. Diese Methode ist relativ unempfindlich gegenüber dem Verlust von Teilchen, die im Strahlrohr bleiben und nicht im Detektor nachgewiesen werden können. Für diese Teilchen gilt $\mathsf{E}_h - p_h{}^{\|} \approx 0$ und damit tragen sie nicht oder wenig zur Summe bei. Im Fall von CC–Ereignissen entfällt die Information über das gestreute Neutrino natürlicherweise.

1.2 Produktionsmechanismen für Charm–Ereignisse

Im Standardmodell der elektroschwachen Theorie wird die Wechselwirkung von Quarks und Leptonen durch den Austausch der Feldquanten γ, Z^0, W^+ und W^- vermittelt. Die Quark-Flavour-Eigenzustände sind jedoch nicht auch die Eigenzustände der elektroschwachen Wechselwirkung. Dadurch ist es erst möglich, Übergänge zwischen Quarks verschiedener Familien zu beobachten, ohne dabei neue Kopplungen einführen zu müssen.

Im Modell von Glashow, Iliopoulus und Maiani [22] wurde diese sogenannte Quarkmischung zwischen den damals nur bekannten ersten beiden Familien beschrieben. Die Matrix, welche den Basiswechsel zwischen den Flavour und elektroschwachen Eigenzuständen vermittelt, kann im Fall der vier leichtesten Quarks (u, d, s, c) durch nur einen Winkel parametrisiert werden. Dies liegt daran, daß die Transformation unitär sein muß und damit vier freie Parameter besitzt, wovon jedoch drei durch Phasentransformationen der Quarkfelder absorbiert werden. Es wurde die Konvention gewählt, daß die Quarks mit Ladung $Q = +\frac{2}{3}e$ nicht mischen. Die Eigenzustände zur schwachen Wechselwirkung sind nun

$$\begin{pmatrix} u \\ d' \end{pmatrix} \begin{pmatrix} c \\ s' \end{pmatrix}, \tag{1.13}$$

die mit den ursprünglichen Eigenzuständen über

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\Theta_c & \sin\Theta_c \\ -\sin\Theta_c & \cos\Theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} \tag{1.14}$$

verknüpft sind. Der Winkel Θ_c wurde von Cabibbo eingeführt und heißt deshalb Cabibbo-Winkel [23].

Cabibbo-Winkel

Dieser Mechanismus läßt sich auf drei Quarkfamilien erweitern. Die Transformation erhält dann die Gestalt

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.15}$$

Die Matrix V wurde von Cabibbo, Kobayashi und Maskawa eingeführt und trägt daher auch den Namen CKM–Matrix. Die Elemente V_{ij} sind proportional zur Kopplungskonstante der Quarks i und j an die W-Bosonen.

CKM–Matrix

Der im vorigen Abschnitt 1.1 beschriebene Prozeß niedrigster Ordnung dominiert nicht bei der Charm–Erzeugung. Dies liegt daran, daß der intrinsische Charmgehalt des Protons vernachlässigbar klein ist, und sich bei Reaktionen mit neutralem Strom der Quarkflavour nicht ändert. CC–Ereignisse ermöglichen durch ihre flavouändernde Wirkung die Produktion von Charm–Quarks aus Down– oder Strange–Quarks.

$$\sigma(e^+p \longrightarrow \bar{\nu}_e\, c\, X) = |V_{cd}|^2 \cdot \sigma_d + |V_{cs}|^2 \cdot \sigma_s \tag{1.16}$$

Die entsprechenden Übergangselemente der CKM–Matrix sind mit V_{cq} bezeichnet, σ_q steht für die Wirkungsquerschnitte an leichten Quarks q. Aufgrund der hohen Masse des W-Bosons ist der Wirkungsquerschnitt für die CC-Reaktionen aber ebenfalls sehr klein bei den bei HERA erreichbaren Schwerpunktsenergien.

Der dominierende Prozeß für die Charm–Erzeugung ist die Boson–Gluon–Fusion, eine QCD–Reaktion der Ordnung $\mathcal{O}(\alpha\alpha_s)$. α ist die Kopplungskonstante der Quantenelektrodynamik (QED) und α_s die Kopplungskonstante der starken Wechselwirkung. In dieser Reaktion koppelt das vom Lepton kommende Eichboson an ein Gluon, das vom Proton stammt (siehe Abb. 1.2). Im Fall eines neutralen Eichbosons wird ein Paar aus Charm und Anti-Charm gebildet, bei W–Boson–Austausch entstehen zwei Quarks mit verschiedenen Flavours. Der Beitrag zum Wirkungsquerschnitt durch Austausch der schweren Eichbosonen W$^{\pm}$ und Z^0 ist aufgrund ihrer hohen Massen um $\mathcal{O}(10^5)$ gegenüber dem Photonaustausch unterdrückt und damit bei den HERA–Experimenten vernachlässigbar klein. Der Photonaustausch mit quasi-reellen Photonen bei $Q^2 \approx 0$ liefert den größten Beitrag zum Wirkungsquerschnitt. Man spricht deshalb auch von Photon–Gluon–Fusion.

Boson–Gluon–Fusion

QED

Photon–Gluon–Fusion

In höherer Ordnung $\mathcal{O}(\alpha\alpha_s^2)$ tragen weitere Subprozesse zum Wirkungsquerschnitt bei, wie sie zum Beispiel in Abbildung 1.3 dargestellt sind.

Im zuvor beschriebenen Prozeß niedrigster Ordnung hat ein Gluon des Protons an der Wechselwirkung teilgenommen und damit die Möglichkeit eröffnet, durch eine Messung Informationen über die Eigenschaften der Gluonen im Proton zu gewinnen. Es gibt jedoch auch den sogenannten Resolved–Photon–Prozeß, bei dem das vom Lepton kommende Photon in ein Quark-Antiquark Paar fluktuiert, sich also gewissermaßen in ein hadronisches Objekt *auflöst*. Damit ist es möglich, $c\bar{c}$–Paare

Resolved–Photon

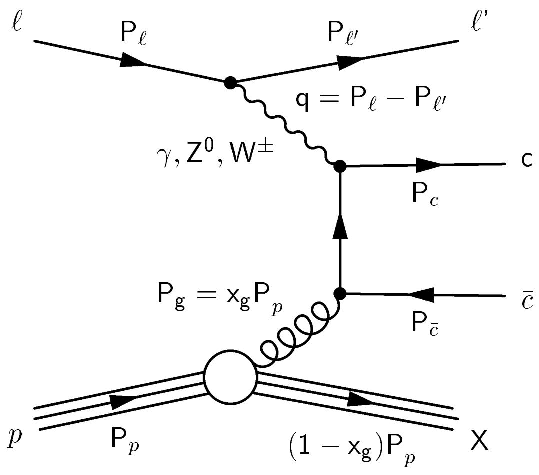

Abbildung 1.2: *Feynmangraph der Boson–Gluonfusion bei der Lepton–Proton–Streuung.*

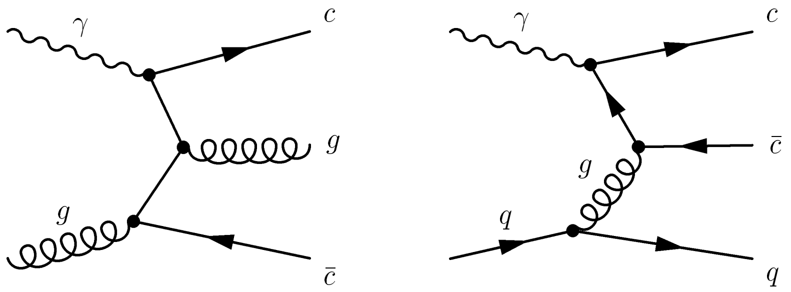

Abbildung 1.3: *Feynmangraphen der Photon–Gluon–Fusion der Ordnung* $\mathcal{O}(\alpha\alpha_s^2)$. *Im Vergleich mit Abbildung 1.2 tritt ein weiteres Parton im Endzustand auf.*

durch $q\bar{q}$–Annihilation zwischen Quarks aus dem Proton und aus dem fluktuierenden Photon zu erzeugen. Dieser Vorgang ist in Abb. 1.4 dargestellt. Abschätzungen [24] ergeben, daß er zum Wirkungsquerschnitt der Schwer–Quark–Erzeugung etwa zu 30% beiträgt und damit nicht vernachlässigbar klein ist.

Die genaue Vermessung der Photon–Gluon–Fusionsereignisse erfordert zusätzlich zu den zwei in der tiefinelastischen Streuung verwendeten kinematischen Variablen noch drei weitere. Üblicherweise [25] verwendet man den Impulsanteil x_g des Gluons am Gesamtimpuls des Protons, der definiert ist durch

$$\mathsf{P}_g \equiv x_g \cdot \mathsf{P}_p \tag{1.17}$$

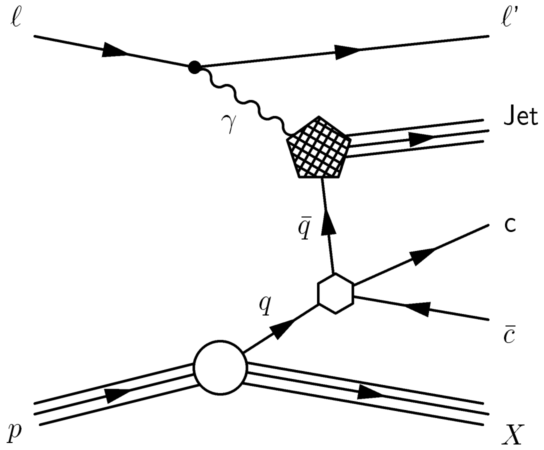

Abbildung 1.4: *Feynmangraph des Resolved–Photon–Prozesses.*

$$x_g \equiv \frac{Q^2 + \hat{s}}{sy} = x + \frac{\hat{s}}{sy} \tag{1.18}$$

wobei $\hat{s}$ das Quadrat der invarianten Masse des Charm–Anticharm–Systems ist

$$\hat{s} \equiv (\mathsf{P}_c + \mathsf{P}_{\bar{c}})^2 = (x_g \mathsf{P}_p + q)^2. \tag{1.19}$$

Wirkungsquerschnitt

Mit Hilfe dieser Variablen kann nun der totale Wirkungsquerschnitt für die Charm–Produktion durch Photon–Gluon–Fusion allgemein angegeben werden

$$\sigma(ep \to c\bar{c}X) = \int dy \int dQ^2 \frac{\alpha}{2\pi} \frac{1}{Q^2} \Big(\frac{1 + (1-y)^2}{y} \cdot \sigma_\perp(\gamma p \to c\bar{c}X) + 2\,(1-y) \cdot \sigma_\parallel(\gamma p \to c\bar{c}X) \Big). \tag{1.20}$$

Er ist als Summe von Wirkungsquerschnitten für transversal ($\sigma_\perp$) und longitudinal polarisierte Photonen ($\sigma_\parallel$) geschrieben. Bei hohen Energien und kleinen Impulsüberträgen ($Q^2 \leq 4$ GeV^2) kann die Weizsäcker–Williams–Näherung [26] verwendet werden, bei der das Photon als reell betrachtet und die longitudinale Komponente bei der Berechnung vernachlässigt wird (zweite Zeile in Formel 1.20). In diesem Fall kann dann der Elektronvertex als unabhängig vom schwerquarkbildenden Subprozeß betrachtet werden. Führt man die Integration über Q^2 mit den Integrationsgrenzen Q^2_{min} und Q^2_{max} für den Faktor vor $\sigma_\perp$ durch, so erhält man die integrierte Spektralfunktion

$$G_{\gamma/e}(y) = \frac{\alpha}{2\pi} \frac{1 + (1-y)^2}{y} \ln \frac{Q^2_{min}}{Q^2_{max}}, \tag{1.21}$$

die die Wahrscheinlichkeit angibt, daß ein Photon mit Energieanteil y vom einlaufenden Lepton abgestrahlt wird. Der Subprozeß zur Bildung der Charm–Quarks

Gluondichteverteilung

kann durch eine Faltung über die Gluondichteverteilung $g(x_g, \mu^2)$ und den Photon–Gluon–Fusionsquerschnitt berechnet werden

$$\sigma_\perp(\gamma p \to c\bar{c}X) = \int dx_g \, g(x_g, \mu^2) \, \sigma(\gamma g \to c\bar{c}). \tag{1.22}$$

Die Gluondichteverteilungsfunktion ist ein Maß für die Wahrscheinlichkeit, ein Gluon mit Impulsanteil x_g im Proton vorzufinden. Der Parameter μ steht für die Massenskala bei der die QCD–Prozesse störungstheoretisch gerechnet werden.

Durch die Kombination der Formeln 1.20, 1.21 und 1.22 erhält man die folgende Darstellung des totalen Wirkungsquerschnitts für die Photoproduktion von Charm–Quarks:

$$\sigma(ep \to c\bar{c}X) = \int dy \, G_{\gamma/e}(y) \int dx_g \, g(x_g, \mu^2) \, \sigma(\gamma g \to c\bar{c})(\hat{s}, \mu^2) \tag{1.23}$$

Unter Berücksichtigung der Beiträge bis zur Ordnung $\mathcal{O}(\alpha\alpha_s^2)$ haben Ellis und Nason [27] den Wirkungsquerschnitt zu

$$\sigma(ep \to c\bar{c}X) = 0.68 \, ^{+0.26}_{-0.18} \, \mu b \tag{1.24}$$

berechnet. In diese Rechnungen gingen die HERA Schwerpunktsenergie, eine Charm–Masse von $m_c = 1.5$ GeV und ein Skalenparameter von $\mu^2 = 10$ GeV^2 ein.

1.3 Zerfallsprozesse

Nachdem im vorigen Abschnitt die wichtigsten Produktionsmechanismen für Charm–Quarks vorgestellt wurden, bleibt festzuhalten, daß aufgrund des Quark–Confinements [28] einzelne Quarks nicht beobachtet werden können, denn Quarks sind keine farbneutralen Objekte. Nur farbneutrale hadronische Teilchen wie Mesonen und Baryonen sind beobachtbar.

Fragmentation

Den Weg von einzelnen Quarks zu beobachtbaren Teilchen läuft über die Prozesse Fragmentation und Hadronisation. Man kann diese Prozesse in drei Phasen unterteilen, die durch die verschiedenen theoretischen Zugänge charakterisiert werden. In der ersten Phase strahlt das Quark Gluonen ab, die wiederum Quark–Antiquark–Paare bilden. Dies geschieht mit hohen Impulsüberträgen und kann aus diesem Grund störungstheoretisch behandelt werden, indem man entweder die exakten Matrixelemente [29] berechnet oder das Parton–Schauer–Modell [30] anwendet. In der zweiten Phase haben die Quarks bereits einen großen Teil ihres Impulses verloren, wechselwirken deshalb nur noch mit geringen Impulsüberträgen und bilden farbneutrale Systeme. Dies ist nicht mehr perturbativ zugänglich und man ist auf Modelle angewiesen, die diesen Prozeß beschreiben. Die zur Zeit populärsten Modelle sind die String–Fragmentation [32] und die Cluster–Fragmentation [33]. In der letzten Phase zerfallen die Teilchen über die schwache Wechselwirkung in stabilere, die dann im Detektor nachgewiesen werden können.

Zerfallsmodus	Verzweigungsverhältnis
$D^{*+} \rightarrow D^0\pi^+$	55 ± 4 %
$D^{*+} \rightarrow D^+\pi^0$	27.2 ± 2.5 %
$D^{*+} \rightarrow D^+\gamma$	18 ± 4 %

Tabelle 1.1: *Zerfälle der $D^{*\pm}$–Vektormesonen [31] (ladungskonjugierte Zustände sind nicht aufgeführt).*

Die gebildeten Hadronen mit Charm lassen sich in Mesonen ($c\bar{q}$), Charm–Baryonen (cqq) und Charmonium–Zustände ($c\bar{c}$) klassifizieren. In dieser Arbeit soll nur auf Mesonen eingegangen werden. Die Mesonen werden als pseudoskalare Mesonen oder Vektormesonen gebildet; wegen der Spinabhängigkeit des Phasenraums ist das Entstehungsverhältnis 1:3 .

Vektormeson

Die Vektormesonen zerfallen über die elektromagnetische Wechselwirkung (letzte Zeile in Tab. 1.1) oder die starke Wechselwirkung (die ersten beiden Zeilen in Tab. 1.1) in pseudoskalare Mesonen. Beim Zerfall des $D^{*\pm}$in das ungeladene D^0 und das geladene Pion resultiert die geringe Massendifferenz zwischen Anfangs- und Endzustand in einem sehr eingeschränkten Phasenraum für das Pion. Dieser Umstand kann zur Identifikation des Zerfalls verwendet werden, da in einem Massendifferenzschaubild zwischen rekonstruiertem $D^{*\pm}$und D^0 in der Nähe der Pionmasse ein Signal sichtbar wird.

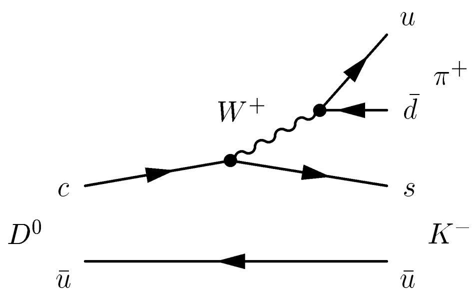

Abbildung 1.5: *Feynmangraph des D^0 –Zerfalls.*

Standard–Modell

Der Zerfall der Charm–Quarks in die leichten Quarks s und d wird im Standard–Modell durch den schwachen geladenen Strom beschrieben. Dabei entsteht aus dem Charm ein $W^{\pm}$–Boson und ein leichtes Quark. Die Zerfallswahrscheinlichkeiten sind proportional zu den CKM–Matrixelementen

$$c \longrightarrow s\, W^+ \sim |V_{cs}|^2 \sim \cos^2\Theta_c \qquad \text{und} \tag{1.25}$$

$$c \longrightarrow d\, W^+ \sim |V_{cd}|^2 \sim \sin^2\Theta_c. \tag{1.26}$$

Zuschauer–Diagramm

Das Eichboson zerfällt dann entweder hadronisch in ein $q\bar{q}$–System oder leptonisch in ein $\ell\,\bar{\nu}_\ell$–System. Da das leichte Quark des ursprünglichen Vektormesons nur eine Zuschauerrolle spielt, nennt man dieses Szenario auch Zuschauer–Diagramm. Ein Beispiel für den Zerfall des D^0 in ein π^+ und ein K^- wird in Abbildung 1.5 gezeigt.

Die Wahrscheinlichkeiten für den W^+–Zerfall sind ebenfalls proportional zu den CKM–Elementen

$$W^+ \longrightarrow u\bar{d} \sim |V_{ud}|^2 \sim \cos^2\Theta_c \quad (1.27)$$

$$W^+ \longrightarrow u\bar{s} \sim |V_{us}|^2 \sim \sin^2\Theta_c. \quad (1.28)$$

Unter Berücksichtigung des Cabibbo-Winkels $\sin^2\Theta_c = 0.23$ ergibt sich anhand des Standardmodells, welche Zerfälle bevorzugt und welche unterdrückt werden. Die Übergänge von $c \longrightarrow d$ oder $W^+ \longrightarrow u\bar{s}$ heißen *einfach Cabibbo-unterdrückt*. Zerfälle, die beide Übergänge enthalten heißen *doppelt Cabibbo-unterdrückt*.

Kapitel 2

Das Experiment

Experimente in der Hochenergiephysik bestehen im allgemeinen aus zwei Komponenten. Zum einen wird ein Teilchenbeschleuniger oder ein Speicherring benötigt, der die zu untersuchenden Physikreaktionen ermöglicht. Zum anderen ist ein Detektor erforderlich, mit dem die Reaktionen erkannt und untersucht werden können. In diesem Kapitel wird auf die Komponenten eingegangen, die am Hamburger Speicherring HERA in Betrieb sind. Es werden die wichtigsten technischen Details des Beschleunigers und des Detektors H1 erläutert.

2.1 Der Beschleuniger

Am Deutschen Elektronen–Synchrotron DESY in Hamburg wird der weltweit erste Elektron–Hadron–Speicherring betrieben. Diese Hadron–Elektron–Ring–Anlage HERA bietet die Möglichkeit, die Kollision von Elektronen oder Positronen (im folgenden nur noch als *Elektronen* bezeichnet) mit Hadronen zu beobachten. Elektronen können in dem etwa 6.3 km langen Ring auf eine Maximalenergie von 30 GeV gebracht werden. Die Protonen erreichen im darunterliegenden Ringsystem eine Energie von bis zu 820 GeV. In vier verschiedenen nach den Himmelsrichtungen benannten Zonen des Beschleunigers werden der Elektron– und der Protonring zusammengeführt und die Strahlen zur Kollision gebracht (siehe Abb. 2.1 rechts). In diesen Zonen befindet sich je ein Experiment (H1, HERMES, ZEUS und HERA–B), wobei nur H1 und ZEUS genuine Elektron–Proton Stöße untersuchen.

DESY
HERA
H1
ZEUS

Die Elektronen und Hadronen durchlaufen zur Erreichung der Maximalenergie ein System aus Vorbeschleunigern (siehe Abb. 2.1 links). Die Elektronen werden zunächst in einem Linearbeschleuniger (LINAC) auf eine Anfangsenergie von 500 MeV gebracht, welche das Synchrotron DESY II auf 7 GeV erhöht. Die für HERA benötigte Einschußenergie von gegenwärtig etwa 11 GeV wird im ehemaligen e^+e^-–Speicherring PETRA II erreicht, der auch eine Bündelung der Teilchen zu Paketen mit einem Abstand von 96 ns vornimmt.

PETRA

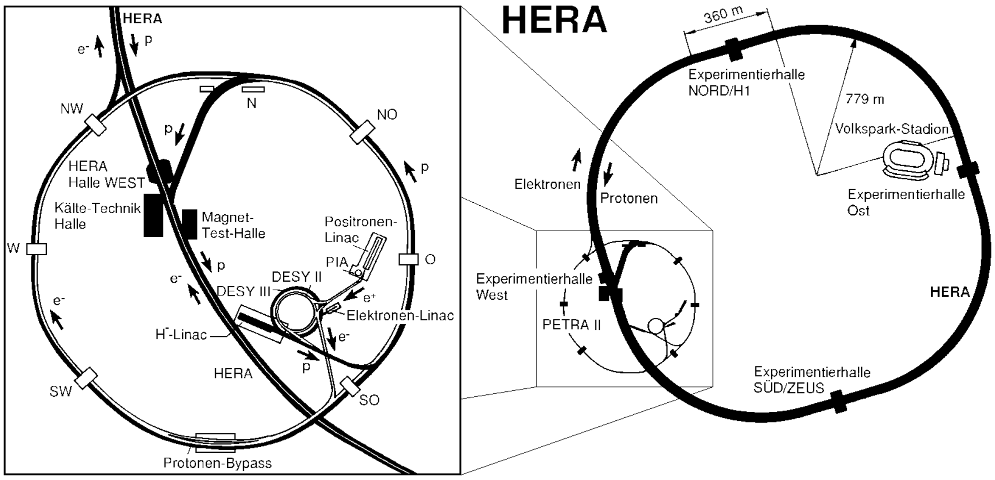

Abbildung 2.1: *Der Speicherring HERA (rechts) mit dem Vorbeschleunigersystem in der Ausschnittvergrößerung (links).*

Die Protonen beginnen ihren Beschleunigungszyklus als H^-–Ionen ebenfalls in einem Linearbeschleuniger, deren Elektronen aber beim Einschuß in das Synchrotron DESY III abgestreift werden. Mit einer Energie von 7.5 GeV werden sie in den Speicherring PETRA II injiziert, auf 40 GeV beschleunigt und dann nach HERA transportiert.

Supraleitung

HERA verwendet normal– und supraleitende Hohlraumresonatoren, um beide Teilchensorten auf ihre Endenergie zu beschleunigen. Im Elektronring werden normalleitende, im Hadronring supraleitende Magnete eingesetzt, um die Teilchen auf ihrer Sollbahn zu halten. Eine Füllung von HERA beginnt mit der Beschleunigung der Protonen, die zuvor in drei Füllungen von PETRA II geliefert werden müssen. Dort können nur jeweils 60 Pakete akkumuliert werden. Danach folgt die Bereitstellung der Elektronen im selben Modus.

Seit August 1994 werden in HERA nicht mehr Elektronen, sondern Positronen beschleunigt, da damit eine höhere Lebensdauer des Leptonstrahls erreicht werden konnte. Im weiteren soll aber trotzdem nur von Elektronen gesprochen werden. Die Füllzeiten betrugen 1994 etwa 60 Minuten für Protonen und 30 Minuten für

Elektronen

Elektronen. Weitere Details zum Speicherring können aus Tabelle 2.1 entnommen werden.

Parameter		Design e–Ring	Design p–Ring	1994 e–Ring	1994 p–Ring
Strahlenergie	[GeV]	30	820	27.5	820
Schwerpunktsenergie	[GeV]	314		300	
Strahlstrom	[mA]	58.0	163.0	17.0	41.0
Teilchenpakete		210	210	153	153
Luminosität	[$cm^{-2}s^{-1}$]	1.5×10^{31}		0.43×10^{31}	
Spez. Lumi.	[$cm^{-2}s^{-1}mA^{-2}$]	4.0×10^{29}		5.2×10^{29}	
Int. Luminosität	[$nb^{-1}y^{-1}$]	10^5		4977	

Tabelle 2.1: *Parameter des Speicherrings HERA.*

2.2 Der Detektor

Zur Erkennung und Aufzeichnung von Elektron–Hadron Stößen bei HERA wurde als Nachweisgerät der H1–Detektor gebaut [34]. Er umschließt nahezu hermetisch die Wechselwirkungszone, in der die Teilchenbündel einander durchdringen. Der Detektor besteht aus einer großen Anzahl unabhängiger Komponenten, deren Signale zusammengefaßt ein möglichst getreues Bild der stattgefundenen Reaktion ergeben.

H1–Detektor

Die beiden stark unterschiedlichen Strahlenergien bewirken, daß das Schwerpunktsystem der Reaktion nicht mit dem Laborsystem identisch ist, und ein Großteil der Reaktionsprodukte in Vorwärtsrichtung[1] fliegt. Dieser sogenannte *Lorentz–Boost* ist durch ein entsprechend asymmetrisches Design des Detektors berücksichtigt worden.

Subsysteme

Die wichtigsten Subsysteme des H1–Detektors sind schalenförmig um den Strahlkreuzungspunkt aufgebaut (siehe Abb. 2.2). Das Strahlrohr $\boxed{1}$ wird umschlossen von den zentralen Spurnachweisgeräten $\boxed{2}$, wie z.B. den Proportionalkammern, den Driftkammern und den Vorwärts–Spurdetektoren $\boxed{3}$. Noch im Inneren des supraleitenden Magneten $\boxed{6}$ folgen dann das elektromagnetische $\boxed{4}$ und das hadronische Kalorimeter $\boxed{5}$, die eine zentrale Rolle bei der Energiemessung spielen. Die Myonkammern $\boxed{9}$ und das mit Streamer–Röhren instrumentierte Rückflußjoch $\boxed{10}$ aus Eisen bilden die äußerste Schale des Detektors und dienen dem Nachweis minimalionisierender Teilchen. Im vorderen Bereich wird dieser Nachweis durch einen toroidalen Magneten $\boxed{11}$ mit Driftkammern unterstützt.

Der H1–Detektor wurde für 1995 in wesentlichen Punkten umgerüstet. So wurde das rückwärtige rein elektromagnetische Kalorimeter $\boxed{12}$ durch ein neuartiges Kalorimeter mit hadronischem Teil und hoher Granularität ersetzt [36]. Nahe des Wechselwirkungspunktes wurden Silizium–Mikrovertexdetektoren installiert, die durch Silizium–Spurdetektoren im rückwärtigen Bereich ergänzt werden [37].

Das H1–Experiment besitzt ein vierstufiges Triggersystem, bei dem die ersten beiden

[1] Es gilt die Konvention, daß die Flugrichtung des Protonstrahls als Vorwärtsrichtung bezeichnet wird.

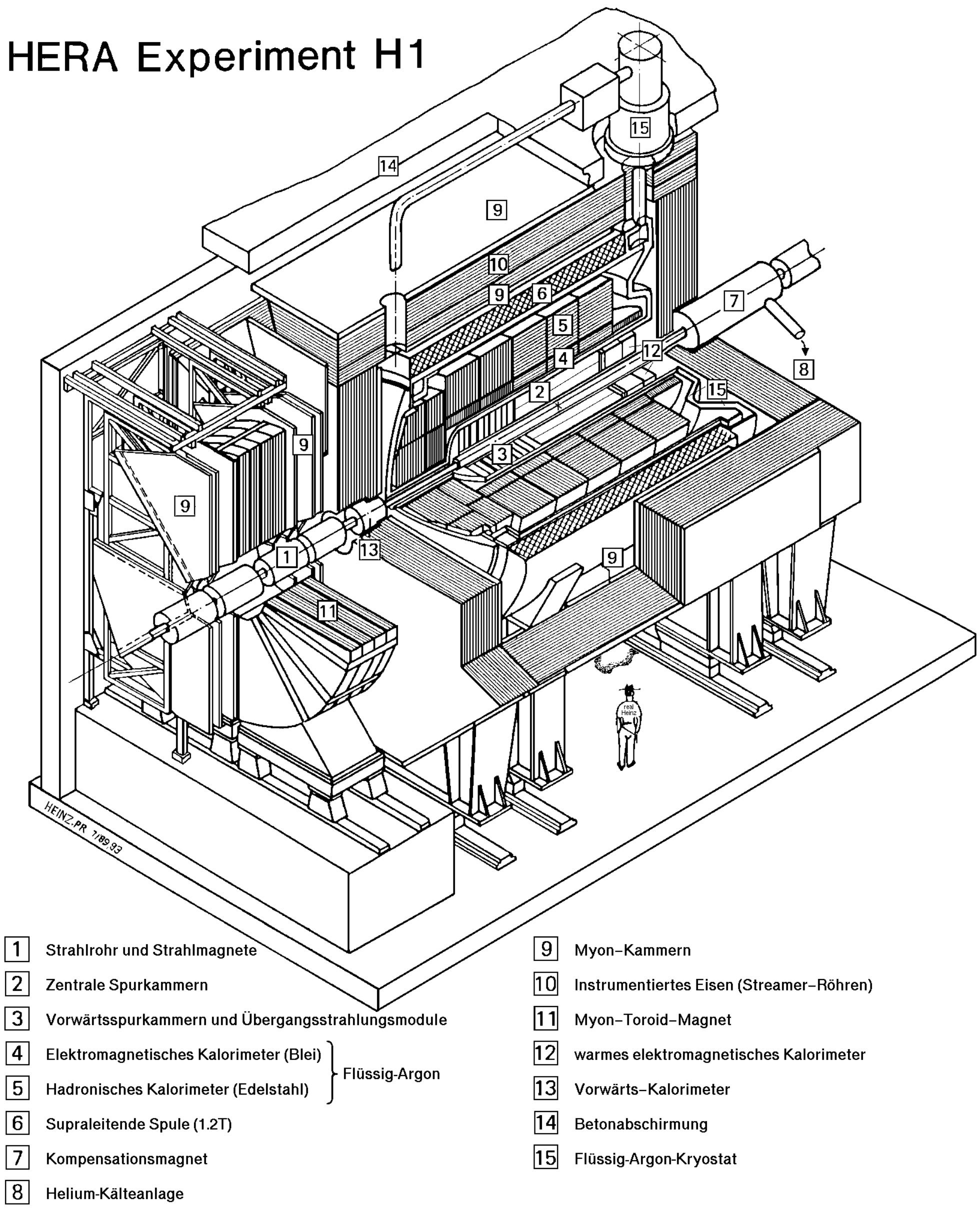

Abbildung 2.2: *Der H1–Detektor im dreidimensionalen Schnittbild [35].*

Stufen – Level-1 und Level-2 – hardwarebasiert sind. Die Stufen 3 und 4 sind softwaretechnisch realisiert. Im weiteren wird auf die einzelnen Subsysteme des Detektors im Detail eingegangen, wobei die Triggerinformation der Subdetektoren einen Schwerpunkt der Darstellung bilden.

2.2.1 Die Subdetektoren und ihre Triggerinformation

Das Flüssig–Argon–Kalorimeter

Die Hauptkomponente des Kalorimetersystems besteht aus einem hochgranularen Flüssig–Argon–Kalorimeter, das sich innerhalb des Solenoid–Magneten in einem Kryostaten befindet. Es handelt sich um ein Sampling–Kalorimeter, das in einen elektromagnetischen und einen hadronischen Bereich unterteilt ist. Die Absorberplatten aus Blei im elektromagnetischen Teil weisen eine Strahlungslänge von 20–30 X_0 auf, die Stahlplatten des Hadronischen haben eine Gesamtdicke von 5–8 Wechselwirkungslängen Λ_{abs}.

Die Triggerinformation des Kalorimeters wird aus den sogenannten Triggerzellen gebildet, die bereits Zusammenfassungen der etwa 45000 Auslesezellen des Kalorimeters darstellen. Separat für elektromagnetische und hadronische Kanäle werden dann durch weitere Summation Triggertower mit projektiver Geometrie geformt. Da auch diese Informationsmenge zu detailliert ist, werden daraus sogenannte Bigtower gebildet. Die 2×252 Bigtower weisen eine 17-fache Unterteilung im Azimutwinkel Θ und eine vom Azimutwinkel abhängige 8– bis 32–fache Unterteilung im Polarwinkel Φ auf. Die Bigtower–Energien werden dann in digitale Form gebracht und daraus topologische Größen wie die Gesamtenergie, die transversale Energie, die fehlende transversale Energie oder eine geometrisch gewichtete Energie berechnet und dem Triggersystem zur Verfügung gestellt.

Bigtower

Das PLUG–Kalorimeter

Im extremen Vorwärtsbereich (12.5 mrad $\leq \Theta \leq$ 60 mrad) zwischen Strahlrohr und Flüssig–Argon–Kalorimeter ist das PLUG–Kalorimeter installiert. Es ist ebenfalls vom Samplingtyp und besitzt Kupfer als Absorbermaterial, das mit Siliziumdetektoren instrumentiert ist. Es dient zur Messung des hadronischen Energieflusses und ist geeignet, Protonfragmente zu untersuchen oder Untergrund aus Strahl–Gas und Strahl–Wand Reaktionen zu identifizieren. Die Informationen über die Energien, die im PLUG gemessen werden, gehen in Form von Bigtowern in die Kalorimeter Triggerinformation ein.

Das BEMC–Kalorimeter

Bis zum Jahr 1994 war in H1 der rückwärtige Teil ($152^\circ \leq \Theta \leq 176^\circ$) des Detektors mit einem rein elektromagnetischen Kalorimeter (BEMC) ausgerüstet. Es basierte auf einer Blei–Szintillator–Sandwich–Struktur in Form von 88 Zellen, die eine Strahlungslänge von 22.5 X_0 aufwiesen. Der Nachweis von unter kleinem Winkel gestreuten Elektronen war die wichtigste Aufgabe dieses Kalorimeters. Jede Zelle lieferte eine über Wellenlängenschieber ausgelesene Energie, die im Triggersystem mit drei verschiedenen Schwellen verglichen wurde. Außerdem wurde noch die Gesamtenergie des Kalorimeters für das Level-1-Triggersignal verwendet. Für die höheren Triggerlevel standen die detaillierten Zelleninformationen zur Verfügung. Diese so gewonnenen Informationen über die Energien wurden ebenso wie beim PLUG in die Bigtower mit aufgenommen.

Das SpaCal–Kalorimeter

Seit 1995 ist das neue Kalorimeter vom SpaCal–Typ im hinteren Bereich des H1–Detektors integriert. Es verwendet Szintillationsfasern, die dichtgepackt in Blei eingebettet sind, und deren Licht über Photovervielfacher ausgelesen wird. Longitudinal ist das SpaCal unterteilt in einen elektromagnetischen Bereich mit einem Faser zu Blei Verhältnis von 1:2 und einen hadronischen Bereich mit einem Verhältnis von 1:4. Der elektromagnetische Teil ist in 1192 Zellen, der hadronische in 136 Zel-

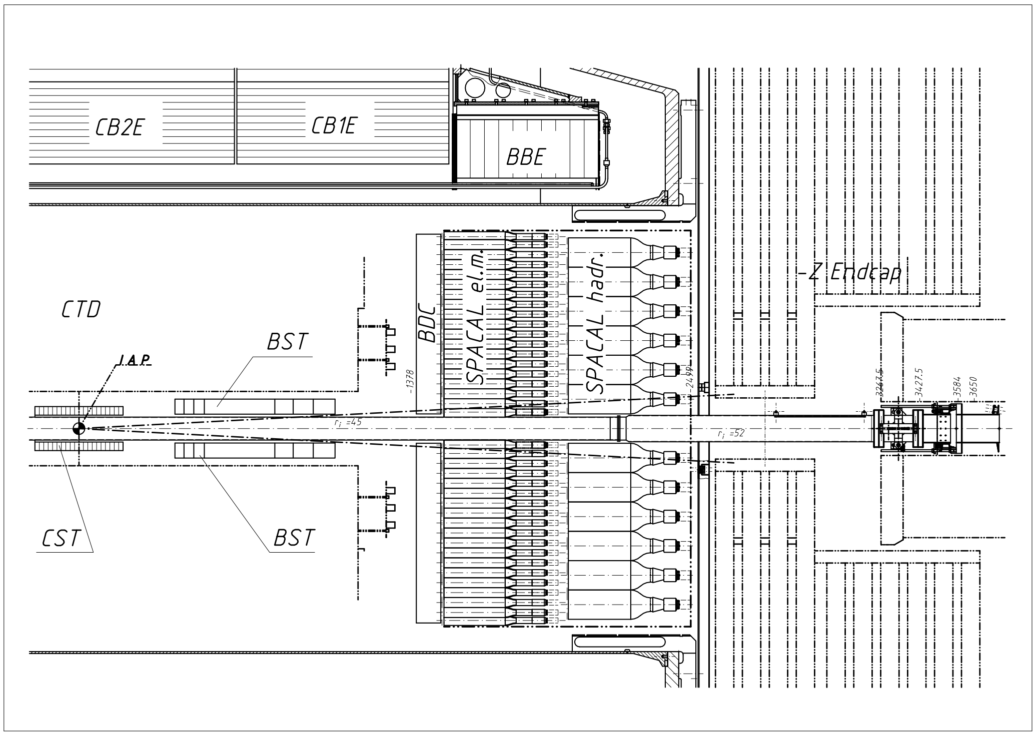

Abbildung 2.3: *Seitenansicht des SpaCal–Kalorimeters im rückwärtigen Teil des H1-Detektors.*

len aufgeteilt [38]. Es ermöglicht die Messung von gestreuten Elektronen bis zu einem Winkel von $\Theta = 177.5^\circ$ mit einer Auflösung, die bei einer Energie von 30 GeV besser als 2% ist. Die genaue Position des SpaCal–Kalorimeters innerhalb des H1–Detektors kann aus Abb. 2.3 ersehen werden.

Das Triggersystem des SpaCal's hat zwei Hauptaufgaben. Zum einen muß es die Funktion eines Flugzeitmeßsystems übernehmen, um Untergrundereignisse zu identifizieren, zum anderen muß es mit verschiedenen Schwellen sensitiv auf deponierte Energien reagieren können. Die Energien von je 4×4 Zellen des SpaCal werden zu Triggertowern zusammengefaßt. Durch Verschiebung dieses Summierfensters um je zwei Zellen in beiden Dimensionen wird eine Effizienz für die Erfassung von Schauern von nahezu 100% gewährleistet. Als Information ist für den Level-1-Trigger ein globales logisches 'ODER' aller Triggertower für drei verschiedenen Energieschwellen erhältlich. Der Level-2-Trigger erhält als detailliertere Informationen die 400 Bits der Triggertower oberhalb der zweiten Schwelle und für Gruppen von 4×4 Triggertowern je ein Bit für jede der drei Schwellen [39].

Das Myon–System

Das Eisenjoch zur Rückführung des magnetischen Flusses ist mit 16 Lagen Streamerkammern instrumentiert, die der Detektion von minimalionisierenden Teilchen dienen. Auch hadronische Teilchen, die im Flüssig–Argon–Kalorimeter nicht komplett absorbiert wurden, können im Myondetektor nachgewiesen werden.

Das gesamte System ist in vier Bereiche unterteilt, einen vorderen und hinteren Zentralteil mit jeweils dazugehöriger Endkappe. Der Zentralbereich ist im Polarwinkel Φ in 16 Module gegliedert, die Endkappen bestehen ebenfalls aus 16 Modulen von ca. 90 cm Höhe, die senkrecht übereinander angeordnet sind (siehe Abb. 2.4).

Die Triggersignale des Myonsystems stammen aus den Streamerlagen 3, 4, 5, 8 und 12 und werden für jedes der insgesamt 64 Module separat ausgelesen. Für den Level-1-Trigger stehen nur die Gesamtzahl der getroffenen Module zur Verfügung. Ein Modul gilt im allgemeinen als 'getroffen', wenn mindestens drei der fünf Lagen ein Signal geliefert haben. Auf den höheren Triggerebenen ist die detailreichere Information erhältlich, welches der Module einen Treffer erhalten hat.

Das Myon–Spektrometer im Vorwärtsbereich

Das Vorwärts–Myon–Spektrometer deckt den Bereich des Polarwinkels Θ zwischen 80 und 300 mrad ab und dient zum Nachweis hochenergetischer Myonen, die sich in Vorwärtsrichtung bewegen. Es besteht aus je drei Paaren von Driftkammern vor und hinter dem Toroid–Magneten. Das jeweils mittlere Driftkammerpaar ist in der Φ–Richtung segmentiert, die anderen Paare sind in der Θ–Richtung orientiert.

Nur die Θ–Driftkammern werden für das Triggern verwendet. Sie sind azimutal in Oktanden unterteilt und radial in 8 Zonen. Dem Level-2-Trigger steht für die prä– und die post–toroidalen Driftkammern die Information zur Verfügung, in welchem

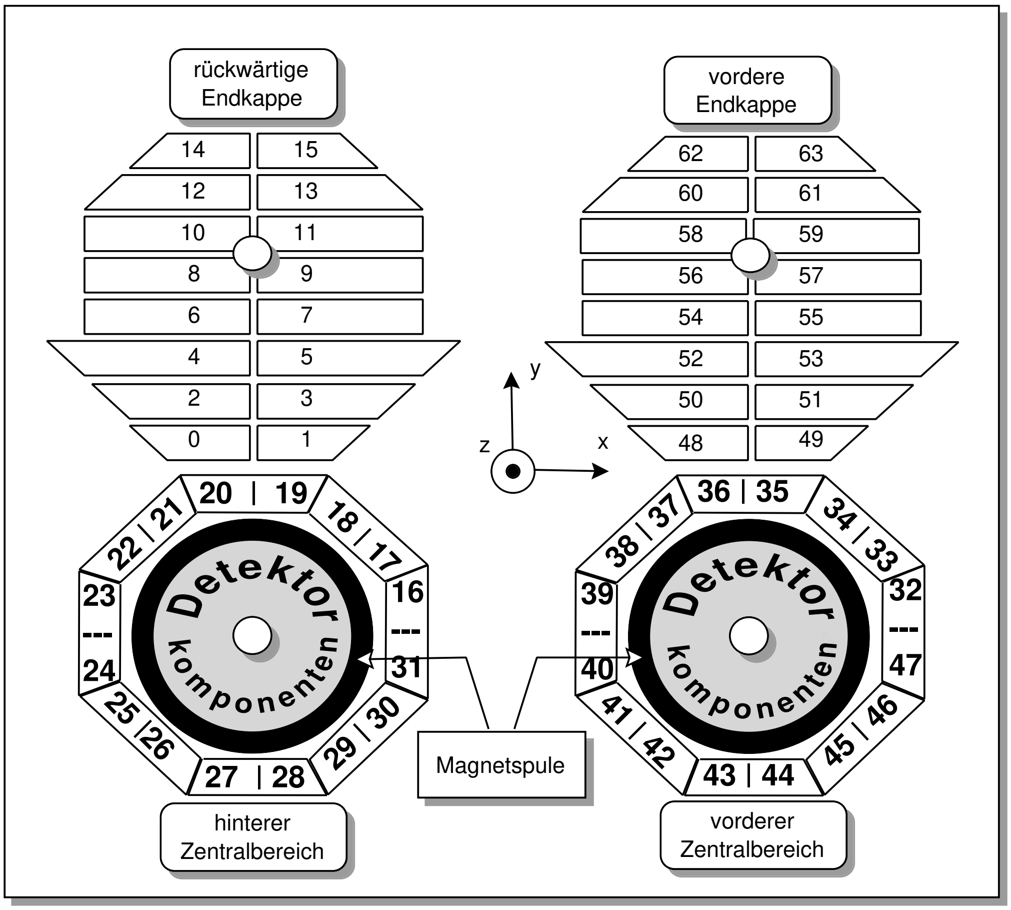

Abbildung 2.4: *Anordnung der Module des Myonsystems.*

der Oktanden und in welcher der radialen Zonen ein Treffer registriert wurde [40]. Ein 'Treffer' wird im allgemeinen dadurch definiert, daß in mindestens drei der vier Driftkammern ein Signal registriert wurde.

Die Proportionalkammern

Die schnelle Erkennung von Teilchenspuren ist essentiell für ein gut funktionierendes Triggersystem. Im H1–Detektor werden dazu sowohl im Vorwärts– als auch im Zentralbereich Vieldrahtproportionalkammern (*Multi Wire Proportional Chambers, MWPC*) eingesetzt.

Im zentralen Spurdetektor befinden sich die innere Proportionalkammer CIP (Central Inner Proportional Chamber) und die äußere Proportionalkammer COP (Central Outer Proportional Chamber) (siehe Abb. 2.5). Die Anodendrähte der zentralen Proportionalkammern sind konzentrisch um die Strahlröhre gespannt. Die Triggersignale werden von den sogenannten Pads der Kathodenebenen abgegriffen. Dabei ist die CIP in 8 $\Phi \times 60$ z Segmente und die COP in 32 $\Phi \times 60$ z Segmente aufgeteilt.

z-Vertex–Histogramm

Diese grobe Ortsbestimmung erlaubt durch Koinzidenzbildung zwischen CIP und COP Informationen über Richtung und Ursprung der Teilchen zu gewinnen. Diese Information kristallisiert sich in Form eines z-Vertex–Histogramms. Es wird gebildet, indem man die aus den Koinzidenzen ermittelten Spuren mit der z-Achse schneidet und die Schnittpunkte histogrammiert. Das Histogramm überdeckt einen Bereich von ± 40 cm um den nominellen Wechselwirkungspunkt und ist in 16 Bins eingeteilt. Dem Level-1-Trigger stehen nur reduzierte Informationen aus dem Histogramm zur Verfügung, wie z.B. die Existenz eines Eintrags oder die Signifikanz eines Maximums. Die Trigger der höheren Levels können auf die Informationen des gesamten Histogramms zugreifen.

Ray–Trigger

Bigray

Das vordere Spurkammersystem beherbergt neben Driftkammern und Übergangsstrahlungsdetektoren auch drei Lagen mit Vieldrahtproportionalkammern. Die Information aus diesen Kammern fließt ebenfalls in das z-Vertex–Histogramm ein. Die Gesamtheit der Proportionalkammern bildet die Basis des Ray–Triggers, der die Spuren auf die Bigtowergeometrie des Kalorimeters abbildet. Aus der Koinzidenz der Spuren mit Energien im Kalorimeter erhält man die sogenannten Bigrays.

Die Jetkammern

Das Jetkammersystem CJC (Central Jet Chamber) des H1–Detektors besteht aus den zwei Driftkammern CJC1 und CJC2 (siehe Abb. 2.5). Es dient über den relativen Energieverlust pro Weglänge $\frac{dE}{dx}$ der Identifizierung und der exakten Vermessung von Teilchenimpulsen.

Die Signal– und Feldformungsdrähte sind parallel zur Strahlachse gespannt und erlauben eine sehr genaue Bestimmung der transversalen Ortskoordinaten.

Das Triggersystem der Jetkammern ist aufgrund der hohen Informationsdichte und der langen Driftzeiten nicht in der Lage, eine vollständige Spurrekonstruktion durchzuführen. Statt dessen wird das von den getroffenen Drähten gebildete Muster mit vorgegebenen Masken verglichen. Damit lassen sich Spurkandidaten finden, die in der transversalen Projektion vom Wechselwirkungspunkt kommen, und es können Spuren mit verschiedenem Ladungsvorzeichen bzw. hohem oder niedrigem Transversalimpuls voneinander unterschieden werden.

Das Luminositätssystem

Die Luminosität L ist definiert über die Beziehung

$$N = \sigma L, \tag{2.1}$$

wobei N die Anzahl der gemessenen Wechselwirkungen und σ den Wirkungsquerschnitt einer Reaktion darstellt. Sie ist ein Maß für die Leistungsfähigkeit eines Beschleunigers.

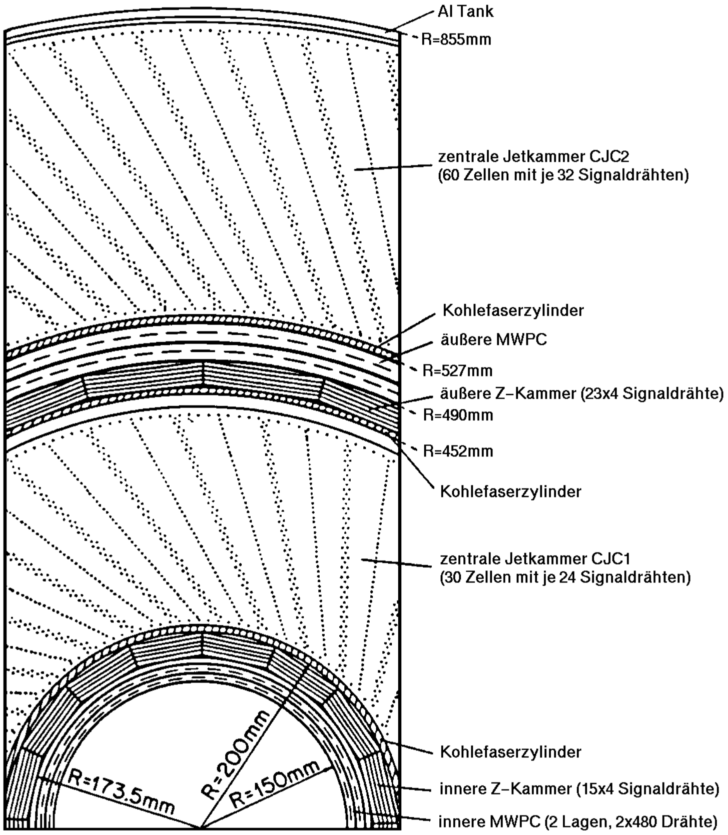

Abbildung 2.5: *Radialer Ausschnitt des Spurkammersystems.*

Elektrondetektor
Photondetektor

Das Luminositätsmeßsystem befindet sich außerhalb des eigentlichen Detektors. Es besteht aus einem Elektrondetektor (*Electron–Tagger, ET*), der 33.4 m vom Wechselwirkungspunkt entfernt ist und einem Photondetektor (*PD*), der in 102.9 m Entfernung installiert ist. Beide Detektoren sind total absorbierende Cerenkov-Kalorimeter aus TlCl/TlBr-Kristallen mit einer Dicke von 21.5 Strahlungslängen X_0. Vor dem Photondetektor ist eine Abschirmung aus Blei (2 X_0) und ein Wasser-Cerenkov-Zähler (*VC*) mit einer Dicke von einer Strahlungslänge angebracht. Beide Detektoren sind als Matrizen von Kristallen realisiert, der Photondetektor als 5×5–, der Elektrondetektor als 7×7–Matrix.

In Abbildung 2.7 ist das gesamte System dargestellt und der grundlegende Prozeß zur Messung der Luminosität erläutert. Der sogenannte Bethe-Heitler-Prozeß $ep \longrightarrow ep\gamma$ dient zur Bestimmung der Luminosität. Die Feynmangraphen zum Bethe-Heitler-Prozeß sind in Abbildung 2.6 dargestellt. Sie sind innerhalb der QED vollständig berechenbar. Deshalb dient die Rate der in Koinzidenz gemessenen Elektronen und Photonen als Meßgröße für die Luminosität. Neben der Hauptaufgabe,

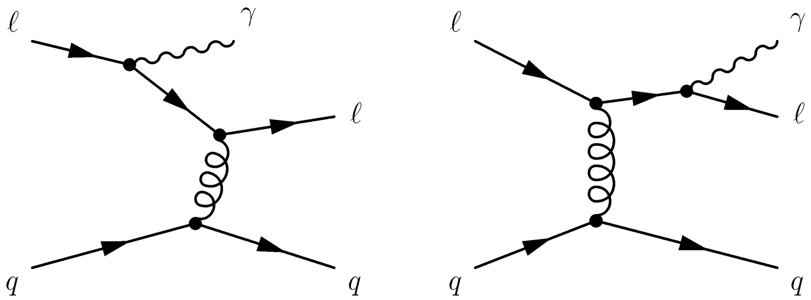

Abbildung 2.6: *Feynmangraphen des Bethe-Heitler Prozesses.*

die Luminosität mit einer Genauigkeit von weniger als 5% zu messen, wird das System auch zum Nachweis von Elektronen verwendet, die in Photoproduktionsereignissen unter flachem Winkel abgestrahlt werden.

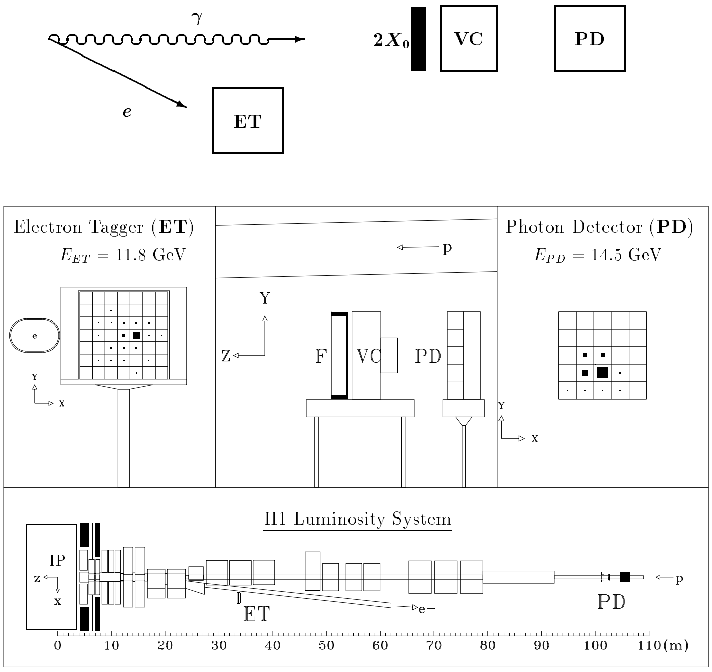

Abbildung 2.7: *Das Luminositätsmeßsystem. Im oberen Teil des Bildes ist der Bremsstrahlungsprozeß dargestellt, im unteren Teil der Aufbau der Detektoren.*

Der Elektrondetektor überdeckt für den Nachweis von Elektronen aus der Photoproduktion einen kinematischen Bereich von Q^2 <0.01 GeV^2/c^2 und 0.2 $< y <$ 0.8 .

2.2.2 Das Trigger– und Datennahmesystem

Die Experimente bei HERA sind durch die technischen Parameter des Beschleunigers vor neue Aufgaben gestellt. Die Zeit zwischen zwei Strahlkreuzungen beträgt 96 ns, wobei im Mittel nur bei jeder zehnten Strahlkreuzung ein Signal im Detektor entsteht. Trotzdem ist es nötig, die Rate der aufgezeichneten Ereignisse stark zu senken, da die Datentransferrate zwischen Detektor und permanentem Speichermedium auf etwa 1 MByte/s beschränkt ist. Bei einer durchschnittlichen Größe von 100 KByte pro Ereignis entspricht dies einer Aufzeichnungsrate von ca. 10 Hz.

Untergrundereignisse

Bei der projektierten Luminosität erwartet man eine Physikrate von $\mathcal{O}$(10 Hz), die durch die Produktion von Schwerquarkereignissen dominiert wird. Daraus folgt die Notwendigkeit, Untergrundereignisse zu unterdrücken und Physikereignisse effizient zu selektieren. Die Untergrundreaktionen sind dominiert durch :

- **Strahl–Gas–Reaktionen**: Dies sind Wechselwirkungen der Protonen mit Restgasatomen im Strahlrohr in der Nähe des Wechselwirkungspunkts. Der hohe Wirkungsquerschnitt für Proton–Nukleon–Streuung erzeugt eine Rate von ungefähr 1 kHz. Die Ereignistopologie dieser Reaktion ähnelt mit ihren vielen niederenergetischen Spuren, die vorzugsweise in Vorwärtsrichtung zeigen, der Topologie von Schwerquarkereignissen.

- **Strahl–Wand–Reaktionen**: Diese stammen von Protonen, die sich nicht auf der Sollbahn bewegen und deshalb mit Strahlführungskomponenten kollidieren. Strahl–Wand–Reaktionen tragen mit einer Rate von etwa 50 kHz zum Untergrund bei. Da die Reaktionen mit Strahlführungskomponenten im allgemeinen weit vor dem Detektor stattfinden, unterscheidet sich der Vertex der entstehenden Teilchenspuren stark von dem der ep–Kollisionen. Aufgrund dieser Topologie sind sie leichter von Physikereignissen separierbar.

- **Synchrotronstrahlung**: Die Elektronen senden bei ihrer Beschleunigung in der Kreisbahn Photonen aus, die aber durch Absorbermasken größtenteils vor dem Detektor eliminiert werden können. Nur in etwa jedem zehnten Ereignis entstehen in den Spurkammern Signale durch Synchrotron–Photonen. Die Rate ist damit sehr gering und stellt kein ernstes Triggerproblem dar.

Selbst wenn höhere Schreibgeschwindigkeiten erreicht werden könnten, wäre es sinnvoll, die Ereignisse so früh wie möglich zu selektieren, da das Datenvolumen, das bereits mehrere Terabyte pro Jahr beträgt, nicht mehr handhabbar wäre.

Das H1–Experiment hat ein vierstufiges Triggersystem entwickelt, um die Eingangsrate von $\mathcal{O}$(100) kHz auf eine Aufzeichnungsrate von 10 Hz zu senken und die Totzeit des Detektors so klein wie möglich zu halten (siehe Abb. 2.8).

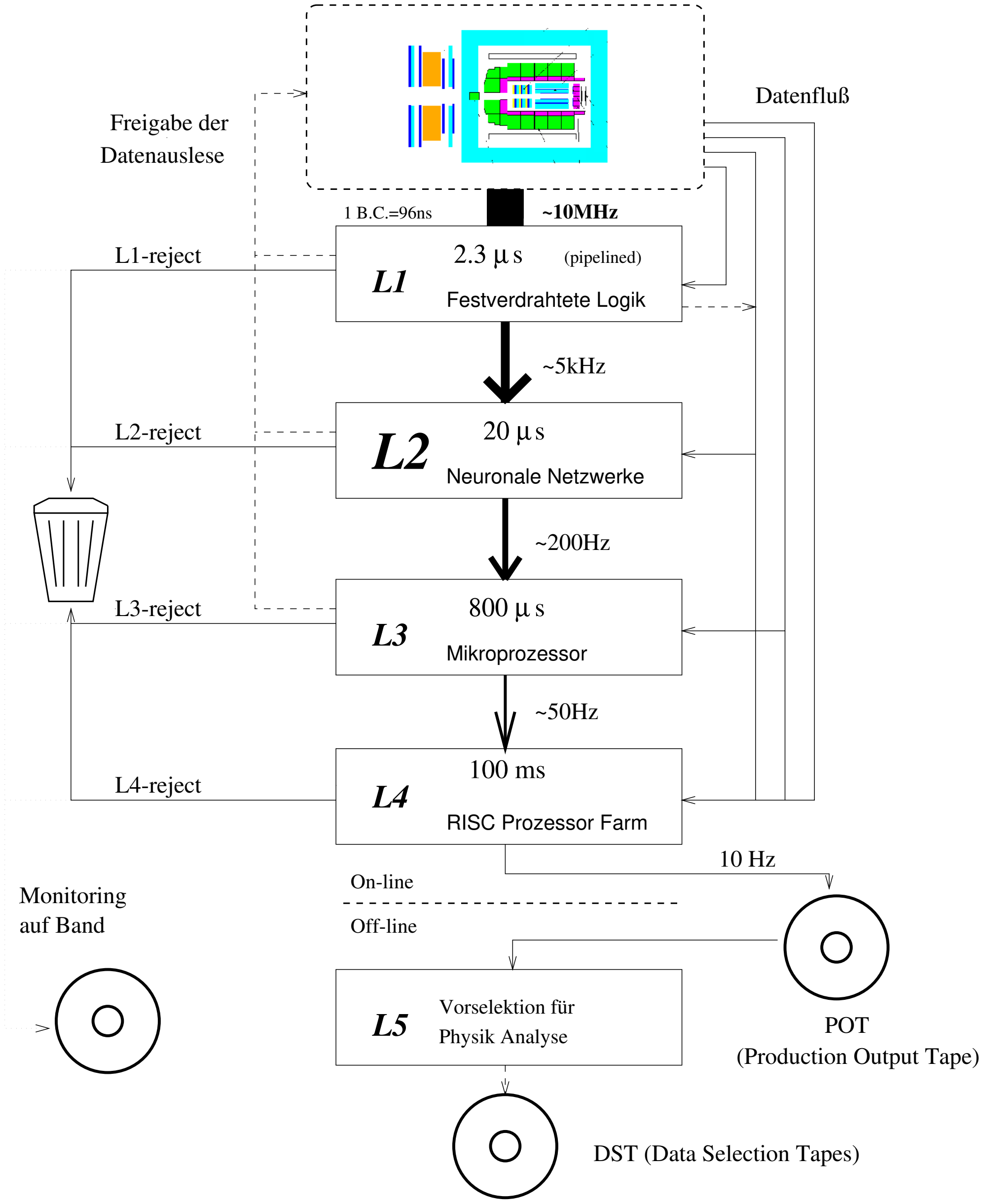

Abbildung 2.8: *Schema des H1 Triggersystems.*

Die ersten beiden Level arbeiten hardwarebasiert, die Level 3 und 4 verwenden Software–Filteralgorithmen.

- **Level 1 (L1)**
  Triggerelemente Die Subdetektoren des H1–Experiments liefern schnelle boole'sche Triggersignale, sogenannte Triggerelemente, an die zentrale Trigger-Logik des Level-1-Triggers CTL1[2]. Die CTL1 ist in der Lage, bis zu 192 Triggerelemente logisch zu 128 Subtriggerelementen zu verknüpfen. Die komplexe Verarbeitung der Triggersignale in den Subdetektoren bedingt, daß der Level-1-Trigger erst 23 Strahlkreuzungen nach der nominellen Kollision seine Entscheidung fällt. Durch Verwendung von Pipelines, die die Subdetektorsignale für die benötigte Zeit zwischenspeichern, wird jedoch keine Totzeit induziert. **Totzeit** Die erste Stufe senkt die Rate auf Werte der Ordnung $\mathcal{O}(1\ \mathrm{kHz})$.

- **Level 2 (L2)**
  Auf der zweiten Stufe des Triggers stehen die Level-1-Triggerinformationen der verschiedenen Subdetektoren zur Verfügung. Es sind zwei Systeme von Level-2-Triggern implementiert worden, die beide eine Entscheidung innerhalb von 20 μs liefern müssen. Der topologische Trigger [41] verwendet Korrelationen zwischen Bigtowern im Kalorimeter und Spuren in den Proportionalkammern. **Neuronaler Trigger** Der neuronale Trigger basiert auf der Ausnutzung von Korrelationen zwischen beliebigen Triggergrößen mit Hilfe von neuronalen Netzwerken. Darauf wird in Kapitel 3 näher eingegangen. Die Rate wird vom Level-2-Trigger auf etwa 100 Hz gesenkt. Nachdem der L2-Trigger das Ereignis akzeptiert hat, beginnt die vollständige Auslese der Detektorsignale.

- **Level 3 (L3)**
  Die dritte Triggerstufe sieht im wesentlichen dieselben Daten wie die zweite. Sie besitzt jedoch einen Mikroprozessor, mit dem beliebige Algorithmen zur Filterung der Ereignisse durchgeführt werden können. Innerhalb von 800 μs muß die Antwort des Triggers vorliegen, wobei eine Ausgangsrate von 50 Hz angestrebt wird.

- **Level 4 (L4)**
  Der Level-4-Trigger besitzt erstmalig die vollständig ausgelesene Detektorinformation und kann beliebige Software–Filter darauf anwenden, solange er unterhalb des Zeitlimits von 100 ms bleibt. Die Ereignisse werden asynchron auf eine Farm von unabhängigen RISC–Prozessorsystemen verteilt, die alle dieselben Algorithmen abarbeiten. Bei Bedarf werden die Daten bereits teilrekonstruiert, z.B. wird eine vereinfachte Version von Spurfindungsalgorithmen eingesetzt. Die vom Level-4 akzeptierten Ereignisse werden permanent in Form von Rohdatenbändern gespeichert und stehen der späteren Analyse zur Verfügung.

[2] Central Trigger Logic Level 1

Die Rohdaten werden innerhalb weniger Stunden quasi on-line rekonstruiert. Dies schließt eine Ereignisklassifikation und nochmalige Untergrundunterdrückung, basierend auf rekonstruierten physikalischen Größen, ein. Als Physikkandidaten klassifizierte Ereignisse werden in komprimiertem Format auf Festplatten verfügbar gemacht.

Kapitel 3

Der Level-2-Trigger

Im mehrstufigen Schema des H1–Triggersystems ist der Level-2 als Hardware–Trigger integriert (siehe auch Abb. 2.8). Seine Aufgabe besteht darin, die Eingangsrate in der Größenordnung von $\mathcal{O}(1\ \text{kHz})$ auf eine Ausgangsrate von etwa $\mathcal{O}(100\ \text{Hz})$ zu senken. Dabei stehen ihm 20 μs als Entscheidungszeit zur Verfügung, um Physikkandidaten effizient von Untergrundereignissen zu trennen. Diese Zeit wurde gewählt, um bei einer maximalen Rate von 5 kHz die vom L2-Trigger induzierte Totzeit unter 10 % zu halten.

Bereits dem Level-1-Trigger steht eine große Informationsmenge der Subdetektoren zur Verfügung. Aufgrund des sehr restriktiven Zeitlimits ist dieser nicht in der Lage, die Information umfassend zu analysieren. Diese Aufgabe fällt dem Level-2-Trigger zu, der dieselben Eingangsdaten besser verwerten muß.

Topologischer Trigger
Neuronaler Trigger

Der Level-2-Trigger basiert auf zwei unterschiedlichen Systemen, dem Topologischen Trigger [41] und dem Neuronalen Trigger [42]. Die Systeme analysieren identische Eingangsinformationen, sie unterscheiden sich aber in der Auswertung der Daten. Der Topologische Trigger ist mit einer spezialisierten Hardware realisiert und verwendet hauptsächlich die Korrelation zwischen verschiedenen Subdetektorinformationen in der durch den Polar– und den Azimutwinkel aufgespannten Ebene.

Parallelrechner

Der neuronale Trigger wertet beliebige Korrelationen im abstrakten Raum der Triggergrößen aus. Er arbeitet im Kern mit einem Parallelrechner, der frei programmierbar ist. Die Entscheidungen beider Systeme müssen innerhalb der vorgegebenen Zeit bei der zentralen Level-2 Triggerlogik (CTL2 [1]) angekommen sein und werden dort zu einer eindeutigen Level-2-Entscheidung kombiniert (siehe Abschnitt 3.5).

Im Rahmen dieses Kapitels wird im besonderen auf den Aufbau und die Arbeitsweise des neuronalen Triggers eingegangen, der sowohl Basis als auch Ziel der $D^{*\pm}$–Analyse bildet.

[1] Central Trigger Logic Level 2

3.1 Überblick über den neuronalen Trigger

Dem Level-2-Trigger stehen erstmals alle Level-1-Informationen an einer zentralen Stelle zur Verfügung, die er damit besser verwerten kann als die lokalen 'Intelligenzen' der einzelnen Subdetektoren. Diese Verwertung ist gleichbedeutend mit einer Analyse der Korrelationen zwischen Triggergrößen, insbesondere wie sich diese bei Untergrund und Physikkandidaten unterscheiden. Je allgemeiner diese Korrelation ist, d.h. nicht unbedingt nur Korrelationen im geometrisch anschaulichen zwei– oder dreidimensionalen Raum, desto größer ist die Wahrscheinlichkeit, die Klasse der Physikereignisse von der Klasse der Untergrundereignisse trennen zu können. Diese Überlegungen implizieren folgende Anforderungen an ein Lösungsschema:

- Es muß Flexibilität bei der Wahl der Eingangsgrößen aufweisen.
- Es muß einen automatischen (algorithmischen) Weg zur Auffindung der hochdimensionalen Korrelationen geben.
- Es muß eine hohe Rechenleistung erreichbar sein, um trotz der großen Komplexität die erforderliche Geschwindigkeit zu gewährleisten.

Künstliche neuronale Netze [43] sind ideal geeignet, um diesen Anforderungen zu genügen. Das Gebiet der neuronalen Netze stellt verschiedene Lösungen zur schnellen Verarbeitung von Informationen bereit; die Art und Anzahl der verwendeten Eingangsdaten ist dabei beliebig.

Die Arbeit des Level-2-Triggers kann man als Mustererkennung im abstrakten Raum der Triggervariablen verstehen. Durch Lernalgorithmen lassen sich Korrelationen in bekannten Datensätzen auffinden, die sich dann auf unbekannte Daten verallgemeinern lassen. Aufgrund der Parallelität der neuronalen Architektur ist eine sehr hohe Rechenleistung erreichbar, wenn man geeignete parallelrechnerbasierte Hardware einsetzt.

Im Prinzip wäre es möglich, nur ein einziges neuronales Netzwerk zu verwenden, dem sämtliche Triggerinformationen zur Verfügung stehen, auf deren Basis dann Ereignisse akzeptiert oder verworfen würden. Es gibt jedoch einige praktische Argumente, die einen anderen Weg nahelegen. Die Erfahrung beim Betrieb des H1-Detektors bei HERA zeigt, daß nur einige wenige Level-1-Subtriggerelemente die hohe L1–Gesamtrate generieren. Diese Subtrigger sind im allgemeinen auf ganz bestimmte Physikkanäle optimiert, die jedoch dem Untergrund sehr ähnlich sind (vgl. Kap. 2.2.2). Ohne eine intelligentere Triggerstufe bliebe nur die statistische Verwerfung der getriggerten Ereignisse. Dieses auch Skalierung genannte Vorgehen führt aber zu einem drastischen Verlust an interessanten Physikereignissen. Ein Satz von Netzwerken, die genau auf diese Physikereignisse spezialisiert sind und die jeweils zur Senkung der Rate der problematischen L1–Subtriggern verwendet werden, bringt eine Reihe von Vorteilen gegenüber einem einzigen Netzwerk:

Skalierung

- Ereignisse, die von Subtriggern niedriger Rate getriggert werden, laufen nicht Gefahr, fälschlicherweise verworfen zu werden, da sie nicht von einem Netzwerk behandelt werden und damit ungehindert den Level-2-Trigger passieren können.
- Eine geringere Anzahl von Triggergrößen reicht aus, um den Untergrund von der jeweiligen Physikreaktion zu trennen.
- Die Komplexität der Netze in Bezug auf die Anzahl der freien Parameter nimmt ab.
- Die Hinzunahme eines Netzes für eine neue Physikklasse kann ohne Störung der anderen Netzwerke erfolgen.
- Die Triggereffizienz für Physikklassen kann separat für jeden Subtrigger in Kombination mit *seinem* neuronalen Netz ermittelt werden.

Untersuchungen haben gezeigt, daß die Kombination der Entscheidungen vieler kleiner dedizierter Netzwerke effizienter arbeitet als ein einzelnes großes Netzwerk, das alle Physikklassen erkennen soll [44]. Diese Erkenntnisse und Überlegungen führten zu einem modularen Design des Gesamtsystems. Das Kernstück des Systems bilden CNAPS[2] Parallelrechner, die als CNAPS/VME–Karten [45] zum Einsatz kommen. Jeder CNAPS–Karte ist eine Vorverarbeitungskarte zugeordnet, die die Triggerinformationen der einzelnen Subdetektoren aufbereitet und zum Parallelrechner überträgt. Diese beiden Karten bilden zusammen ein Modul (vgl. Abschnitt 3.3), das die Aufgabe übernimmt, ein Netzwerk zu berechnen. **CNAPS**

Die Wahl eines frei programmierbaren Parallelrechners hat außerdem den Vorteil, daß nicht nur neuronale Netzwerke implementiert werden können, sondern nahezu jeder Algorithmus, solange die Ausführungszeit innerhalb die Grenze von 20 μs liegt. In den Abschnitten 3.2.2 und 3.2.3 wird gezeigt, wie von dieser Möglichkeit Gebrauch gemacht werden kann.

Grundvoraussetzung für neuronale Netzwerke und auch die meisten anderen Algorithmen sind Datensätze von Physik– und Untergrundreaktionen, anhand derer man die unterschiedlichen Korrelationen finden kann und mit denen die Algorithmen optimiert werden können. Diese Datensätze können sowohl aus echten Experimentdaten, als auch aus Monte-Carlo–Simulationen gewonnen werden. Bei Physikreaktionen mit niedrigem Wirkungsquerschnitt und entsprechend geringer Statistik oder für die Erkundung neuer kinematischer Bereiche bekannter Physikprozesse werden zumeist Daten aus der Simulation verwendet. Diese Monte-Carlo–Daten können auch dazu herangezogen werden, um die Effizienz der Level-2-Trigger zu bestimmen und Vergleiche mit realen Daten durchzuführen. **Monte-Carlo**

Im allgemeinen ist durch die Aufspaltung der Datensätze in zwei unabhängige Mengen auch erreichbar, die Verallgemeinerungsfähigkeit der optimierten Algorithmen

[2]Connected Network of Adaptive ProcessorS

zu untersuchen. Über die softwaretechnischen Aspekte der Aufbereitung der Datensätze, die Selektion von Physik- und Untergrundereignissen, die Auswahl von Triggergrößen und die Optimierung der Algorithmen wird in Abschnitt 3.4.2 berichtet.

3.2 Algorithmen

Im vorangehenden Abschnitt wurde beschrieben, welche Überlegungen dazu führten, neuronale Netzwerke als Hauptalgorithmen für den Level-2-Trigger einzusetzen. Im folgenden Unterabschnitt werden die Grundlagen und die Arbeitsmechanismen neuronaler Netze erläutert und die speziell für den Level-2 gewählte Architektur beschrieben. Ein Haupteinwand gegen die Verwendung neuronaler Netze war, daß sie nur in der Lage seien, Physikereignisse zu erkennen, die ihnen zuvor repräsentativ in Datensätzen vorgelegt wurden. Ereignisse mit neuer Signatur würden jedoch nicht erkannt werden. Diese Problematik ist in Abb. 3.1 dargestellt.

neue Signatur

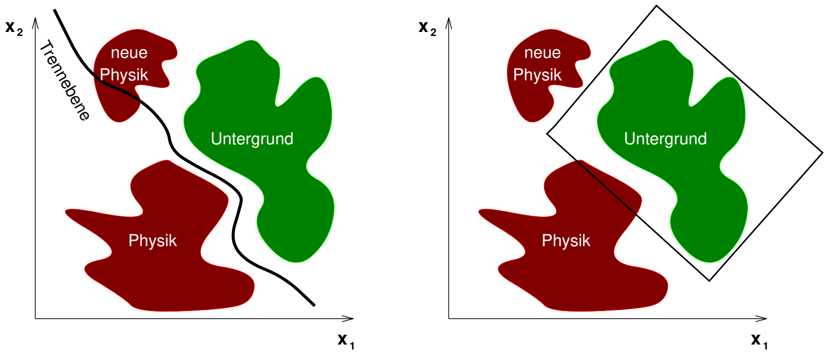

Abbildung 3.1: *Trennung zwischen Untergrund und Physik. Links die Lösung, die ein neuronales Netzwerk realisiert. Rechts die Arbeitsweise des Background Encapsulators.*

Background Encapsulator

Um diesem Einwand entgegenzutreten, wurde ein neuer Algorithmus — der *Background Encapsulator* — entwickelt, der sich nur auf Informationen über die Untergrundereignisse stützt und a priori nichts über die Struktur von Physikereignissen weiß [46]. Details über den Background Encapsulator werden in Abschnitt 3.2.2 erklärt. Es mag der Eindruck entstehen, daß nur noch dieser Algorithmus nötig wäre und daß auf Netzwerke verzichtet werden könnte. Dieser Schein trügt, da der Background Encapsulator nicht so scharfe Trennungen zwischen Untergrund und Physik erreicht wie ein neuronales Netz. Er ist eher als Sicherheits-Trigger gedacht, der auch unvorhergesehene Ereignisse nicht verliert.

Die Kombination von neuronalen Netzwerken und Background Encapsulator ist aber in der Lage, Physikereignisse sehr selektiv auszuwählen, ohne dabei unempfindlich gegenüber neuen Klassen von Physik zu sein. Die dabei verwendete Logik akzeptiert ein Ereignis genau dann, wenn mindestens ein Netzwerk das Ereignis als Physikkandidat erkannt hat, oder wenn alle Netze das Ereignis verwerfen, aber der Background Encapsulator es nicht als Untergrund eingestuft hat. Dieses Ereignis wäre ein Kandidat für potentiell neue Physik. **neue Physik**

Der Optimierungsprozeß bei neuronalen Netzwerken mit dem sogenannten Backpropagation–Algorithmus (siehe Abschnitt 3.2.1) ist sehr rechenaufwendig. Deshalb wurde ein alternativer Algorithmus entwickelt, die *Kegelschnitt–Methode*. Sie wird als Beispiel für die Flexibilität und die universelle Einsetzbarkeit der parallelrechnenden Hardware in Unterkapitel 3.2.3 vorgestellt.

In dieser Arbeit werden neuronale Netzwerke untersucht, die mit dem Backpropagation–Algorithmus trainiert wurden, da die Rechenzeiten bislang noch kein Problem darstellten und mit diesem Algorithmus die besten Effizienzen erreicht werden können.

3.2.1 Neuronale Netzwerke

Das mathematische Modell–Neuron

Künstliche neuronale Netzwerke haben ihren Ursprung in der Biologie. Dort bilden Nervenzellen (Neuronen) die Grundbausteine aller komplexer Strukturen, die an der Sinneswahrnehmung, Informationsverarbeitung und Handlungssteuerung beteiligt sind. Durch den Zusammenschluß von Neuronen entstehen unterschiedliche Topologien, die verschiedene Aufgaben bewältigen können. **Neuronen**

Diese netzartigen Strukturen sind dafür verantwortlich, daß man Neuronen auch als Knoten bezeichnet. Eine einzelne Nervenzelle und ihre mathematische Abstraktion sind in Abb. 3.2 dargestellt. Das mathematische Modell beabsichtigt nicht, die biologischen und chemischen Vorgänge in der Zelle zu beschreiben, sondern lediglich die Funktionalität richtig wiederzugeben. In der Literatur gibt es ausführliche Darstellungen der Zellbiologie von Neuronen [47–49]. **Knoten**

Im mathematischen Modell werden Informationen als Eingangswerte x_i zum Neuron j geschickt. Jeder der Werte wird mit einem Gewicht ω_{ij} multipliziert und im Knoten aufsummiert. Der informationsbewertende Schritt besteht darin, daß von der Summe ein Schwellenwert t_j abgezogen wird.

Dies entspricht im biologischen Fall der Tatsache, daß Nervenzellen nur dann ein Signal weiterleiten, wenn eine minimale Erregung überschritten ist. Das Signal o_j wird im abstrakten Modell als nichtlineare Transformation der Summe berechnet:

$$o_j = g\left(\sum_{i=1}^{n} \omega_{ij} x_i - t_j\right). \tag{3.1}$$

Als sogenannte Aktivierungsfunktion $g(x)$ werden im Modell zumeist nur beschränkte Funktionen zugelassen. Am weitesten verbreitet ist eine sigmoidale Form **Aktivierungsfunktion** **Sigmoid-Funktion**

$$g(x) = \frac{1}{1 + e^{-\frac{x}{T}}} \qquad \text{mit} \qquad T \neq 0, \tag{3.2}$$

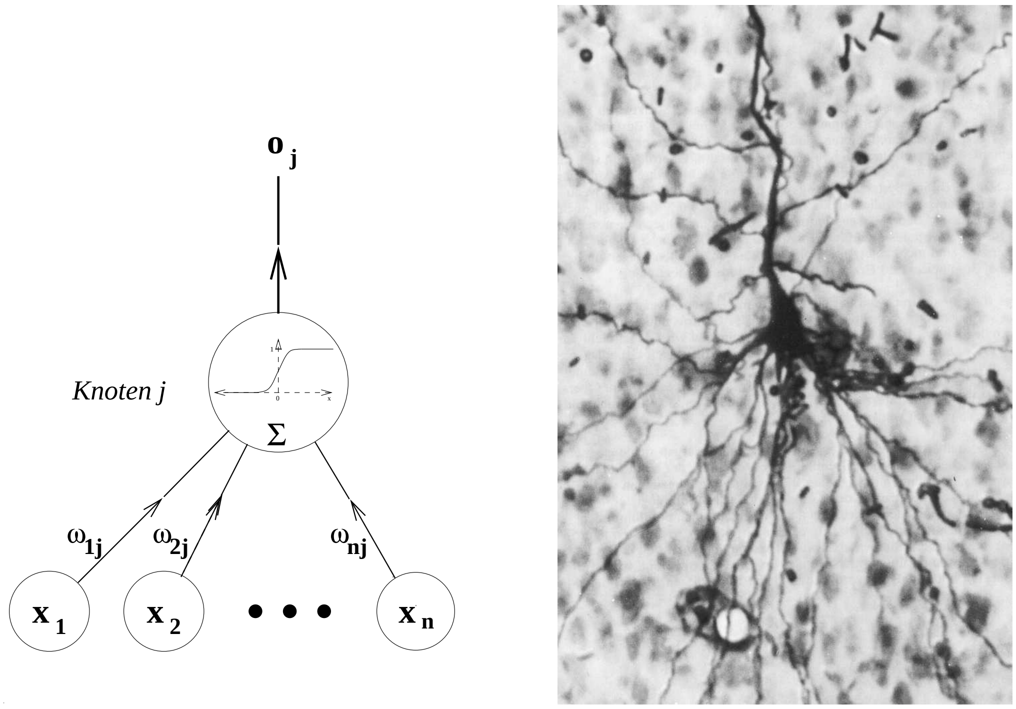

Abbildung 3.2: *Das Neuron. Links: Darstellung des mathematischen Modell–Neurons. Rechts: Schnittbild einer Nervenzelle aus der Sehrinde einer Katze.*

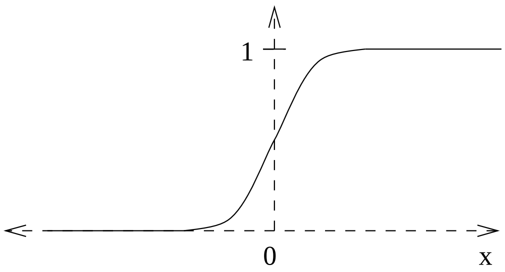

Abbildung 3.3: *Sigmoidale Funktion.*

in der die Wahl von T die Steilheit der Kurve am Nullpunkt bestimmt und als eine Art Temperatur interpretiert werden kann (siehe Abb. 3.3).

Das Feedforward–Netzwerk

Wie bereits angesprochen, bestimmt unter anderem die Anordnung von Neuronen als Knoten in einem Netzwerk dessen Funktionalität. Ebenso entscheidend ist die Verknüpfungsstruktur und die Richtung, in der Signale weitergeleitet werden. In dieser

Arbeit wird nur auf mehrlagige vorwärtsgekoppelte Netzwerke eingegangen, da sie für die Lösung des Klassifikationsproblems am geeignetesten sind. Vorwärtskopplung (feedforward) bedeutet, daß ein Neuron weder direkt noch indirekt von anderen Neuronen Signale (Eingabewerte) erhält, an das es selbst Signale (Ausgabewerte) übertragen hat. In anderen Gebieten der Hochenergiephysik kommen verschiedene andere Netzwerktypen zum Einsatz [50, 51]. In Abbildung 3.4 ist ein dreilagiges Feedforward–Netzwerk dargestellt.

Bei diesem Netztyp sind die Neuronen in Ebenen (Schichten) angeordnet und es bestehen nur zwischen Neuronen aus benachbarten Ebenen Verknüpfungen. Die Information wird von der Eingabeschicht (unten) zur Ausgabeschicht (oben) weitergeleitet. Die dazwischenliegenden Schichten heißen verborgene oder versteckte Schichten . **Ebenen, Schichten**

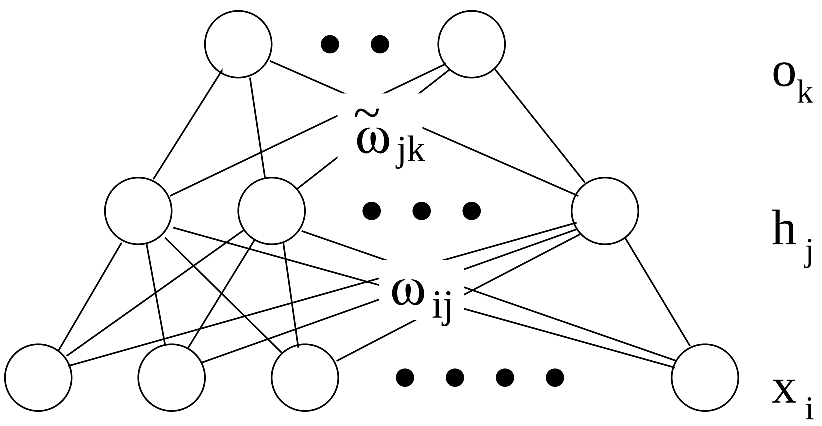

Abbildung 3.4: *Das dreilagige Feedforward–Netzwerk.*

Die Beschränkung auf ein dreilagiges Netz stellt keine Einschränkung dar, da von Kolmogorov gezeigt worden ist [52], daß unter bestimmten Voraussetzungen jede beliebige Funktion mit einem dreilagigen Netz darstellbar ist. Noch wichtiger ist die gleichwertige Aussage, daß man mit dem beim Level-2-Trigger verwendeten Backpropagation–Algorithmus (siehe nächster Abschnitt) jede Funktion beliebig genau approximieren kann [53].

Die für einen Trigger wichtige Funktion ist die Klassifizierungsfunktion. In der Eingabeschicht werden die Meßgrößen (Triggerinformationen) eingespeist. Die Ausgabeneuronen können eine beliebige Repräsentation von Klassen (Physikkandidaten, Untergrund) darstellen. Um nur zwischen zwei Klassen zu unterscheiden, würde ein einziges Ausgabeneuron genügen, das z.B. für Physikkandidaten den Wert '1' annimmt und für Untergrund den Wert '0'. Die Klassifizierungsfunktion F bildet einen n–dimensionalen Vektor $\vec{x}$ auf einen m–dimensionalen Vektor $\vec{o}$ ab, wobei n die **Klassifizierungsfunktion**

Anzahl der Eingabegrößen und m die Dimension des Ausgabevektors beschreiben.

$$\vec{o} = F(\vec{x}) \qquad \text{mit} \qquad \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad \text{und} \qquad \vec{o} = \begin{pmatrix} o_1 \\ \vdots \\ o_m \end{pmatrix} \tag{3.3}$$

Die Funktion F hängt bei gegebener Aktivierungsfunktion nur von den Gewichten ω_{ij} und den Schwellenwerten t_j ab. Durch Einführung eines zusätzlichen virtuellen Eingabeknotens mit einer konstanten Aktivität von $x_0 = 1$, kann man die Sonderbehandlung der Schwellen aufgeben und sie mit in die Gewichtsmatrix integrieren mit $t_j =: x_0\omega_{0j}$. Damit ergibt sich für die Aktivität h_j des j–ten von ℓ Knoten in der verborgenen Schicht

$$h_j = g\Big(\sum_{i=0}^{n} \omega_{ji} x_i\Big) = g(\vec{\omega}_j \cdot \vec{x}) \tag{3.4}$$

und die Aktivität o_k des k–ten Neurons in der Ausgabeschicht

$$o_k = g\Big(\sum_{j=0}^{\ell} \tilde{\omega}_{jk} h_j\Big) = g(\vec{\tilde{\omega}}_k \cdot \vec{h}). \tag{3.5}$$

Der Backpropagation–Algorithmus

Das Auffinden der Gewichte ω_{jk}, die das Klassifizierungsproblem optimal lösen, wird auf dem Gebiet der neuronalen Netze durch sogenanntes Lernen bewerkstelligt. Dies geschieht, indem man dem Netzwerk wiederholt Eingabevektoren einspeist, zu denen der Ausgabevektor bekannt ist. Die Parameter des Netzes werden dann iterativ von einem Algorithmus solange geändert, bis der berechnete Ausgabevektor den gewünschten mit genügender Genauigkeit approximiert. Dieser Lernprozeß findet beim neuronalen L2–Trigger off-line statt, d.h. beim Einsatz im H1–Experiment sind die Gewichte bereits festgelegt und werden nicht mehr verändert.

Lernen

McClelland und *Rumelhart* [54] verhalfen 1986 durch die Wiederentdeckung des von *Werbos* [55] bereits 1974 vorgestellten Backpropagation–Algorithmus dem Gebiet der neuronalen Netze zu einer neuen Blüte. Mit diesem Algorithmus war es zum erstenmal möglich, die Gewichte in der verborgenen Schicht zu adjustieren, indem die Differenz zwischen gewünschtem und berechnetem Ausgabevektor an die darunterliegenden Schichten zurückgereicht wurden (Backpropagation). Der Algorithmus stellt ein Gradientenverfahren zur Lösung nichtlinearer Gleichungssysteme dar, bei dem die freien Parameter des Systems proportional zum Gradienten der Fehlerhyperfläche verändert werden. Als Fehlerfunktion wird dabei im allgemeinen die mittlere quadratische Abweichung vom gewünschten Ausgabevektor verwendet

Backpropagation

$$E(\omega) = \frac{1}{2} \sum_{p=1}^{M} \sum_{k=1}^{m} |o_k^{(p)}(\vec{x}, \omega) - y_k^{(p)}(\vec{x})|^2. \tag{3.6}$$

Dabei werden die M Paare aus dem Trainingsdatensatz durch $(\vec{x}^{(p)}, \vec{y}^{(p)})$ beschrieben, mit $\vec{y}$ als Zielausgabevektor. Die Gesamtheit aller Gewichte des Netzes wird mit ω bezeichnet. Der Startpunkt des Algorithmus wird durch Initialisieren der Gewichte mit Zufallszahlen aus einem Intervall $[-a, +a]$ festgelegt, um etwaige Symmetrien in den Gewichtsvektoren zu vermeiden. Es gibt eine für die Praxis wichtige Modifikation der Fehlerfunktion 3.6, die sich positiv auf die Geschwindigkeit der Konvergenz auswirkt [56]. Dabei wird in der Fehlerfunktion nicht mehr über sämtliche Muster summiert, sondern nach jedem Muster werden die Gewichte verändert (*On-line–Updating*). Bei jedem Lernschritt werden ein Muster aus der Trainingsmenge ausgewählt, die Eingabewerte in das Netz gefüttert und gemäß der Gleichungen 3.4 und 3.5 der Wert $\vec{o}(\vec{x}, \omega)$ berechnet. Nach Bildung der Differenz $E(\omega) = \sum_{k=1}^{m} |o_k(\vec{x}, \omega) - y_k(\vec{x})|^2$ werden die Gewichte des Netzwerks verändert, zuerst die Kopplungen $\tilde{\omega}_{jk}$ zwischen Ausgabe– und versteckter Ebene:

On-line–Updating

$$\Delta\tilde{\omega}_{jk} = -\eta \frac{\partial E}{\partial \tilde{\omega}_{jk}} + \alpha \Delta\tilde{\omega}_{jk}^{alt} = -\eta \delta_k h_j + \alpha \Delta\tilde{\omega}_{jk}^{alt} \tag{3.7}$$

$$\text{mit} \quad \delta_k = (o_k - y_k) \cdot g'(\vec{\tilde{\omega}}_k \cdot \vec{h}). \tag{3.8}$$

Die Kettenregel der Differentialrechnung führt zu den Ausdrücken für die Änderung der Gewichte zwischen verborgener Schicht und Eingabeschicht ('Backpropagation').

$$\Delta\omega_{ij} = -\eta \frac{\partial E}{\partial \omega_{ij}} + \alpha \Delta\omega_{ij}^{alt} = -\eta \delta_j x_i + \alpha \Delta\omega_{ij}^{alt} \tag{3.9}$$

$$\text{mit} \quad \delta_j = \sum_{k=1}^{m} \tilde{\omega}_{jk} \delta_k \cdot g'(\vec{\omega}_j \cdot \vec{x}). \tag{3.10}$$

Die erste Ableitung der Aktivierungsfunktion ist hier als g' geschrieben. Die Proportionalitätskonstante η wird als Lernparameter bezeichnet und muß empirisch dem Problem angepaßt werden. Sie ist ein Maß für die Schrittweite, mit der im Parameterraum nach der optimalen Lösung gesucht wird. Die Größen $\Delta\omega_{ij}^{alt}$ und $\Delta\tilde{\omega}_{jk}^{alt}$ beschreiben die Gewichtsänderungen aus dem vorangegangenen Iterationsschritt. Zusammen mit dem Dämpfungsterm α (Momentumfaktor) dienen sie dazu, die Lernrate in Parameterregionen mit kleinem Gradienten zu vergrößern oder in Bereichen mit großem Gradienten mögliche Oszillationen zu dämpfen. Eine Erweiterung der Formeln für die Gewichtsänderung auf mehr als drei Lagen geschieht nach demselben Schema.

Lernparameter

Momentum

3.2.2 Background Encapsulator

Vorausgesetzt, daß die Trainingsdatensätze eine ausreichende Statistik aufweisen, erreichen neuronale Netze eine nahezu optimale Erkennung der Ereignisklassen. In manchen Gebieten des abstrakten Raumes, der von den Triggergrößen aufgespannt wird, ist die Dichte der Datenpunkte jedoch sehr gering oder verschwindet sogar. Dort ist der Ausgabewert eines Netzwerkes nicht definiert und es kann passieren, daß gerade dort interessante, weil seltene oder unbekannte Physik, falsch klassifiziert wird (siehe z.B. Abb. 3.1).

Neuronale Netzwerke berechnen den Abstand eines Punktes im Hyperraum der Triggergrößen von einer Ebene, die durch die Gewichtsvektoren definiert ist (vgl. Gleichungen 3.1 und 3.2). Auf gleiche Weise kann dieses Abstandsmaß verwendet werden, um einen Untergrund–Datensatz in Gruppen oder sogenannte Cluster einzuteilen, die dann von Hyperwürfeln eingeschlossen und damit von unbekannten Klassen abgegrenzt werden können. Dies ist die dem Background Encapsulator zugrundeliegende Idee.

Cluster

Der erste Schritt des Encapsulator–Algorithmus besteht darin, in den Daten natürliche Gruppierungen zu finden, was mit Hilfe eines Cluster–Algorithmus bewerkstelligt wird. Da das Bilden der Cluster invariant gegenüber linearen Transformationen einzelner Triggergrößen zu sein hat (z.B. Skala der Triggerwerte), wird folgende Funktion gewählt

$$F = \sum_{i=1}^{C} \sum_{\mu_i=1}^{N_i} (\vec{x}^{\mu_i} - \langle \vec{x}_i \rangle)^T \, S^{-1} \, (\vec{x}^{\mu_i} - \langle \vec{x}_i \rangle), \tag{3.11}$$

die es zu minimieren gilt. Dabei bedeutet C die Anzahl der Cluster, N_i die Zahl der Muster im i–ten Cluster und S die Verteilungsmatrix, die definiert ist als

Verteilungsmatrix

$$S = \sum_{\mu=1}^{N} (\vec{x}^{\mu} - \langle \vec{x} \rangle)^T (\vec{x}^{\mu} - \langle \vec{x} \rangle). \tag{3.12}$$

Der mittlere Vektor $\langle \vec{x}_i \rangle$ aller Muster in einem Cluster beschreibt das Cluster–Zentrum und ist definiert durch

$$\langle \vec{x}_i \rangle = \frac{1}{N_i} \sum_{\mu_i=1}^{N_i} \vec{x}^{\mu_i}. \tag{3.13}$$

Es gilt dabei für die Gesamtzahl N aller Muster die Identität $N = \sum_{i=1}^{C} N_i$. Die Funktion F mißt den mittleren quadratischen Fehler, wenn man die Verteilung der N Muster durch die gewichteten C Cluster–Zentren approximieren will. Die Anzahl der möglichen Gruppierungen beträgt $\approx C^N/C!$ und verbietet deshalb einen Lösungsansatz durch Ausprobieren. Der Algorithmus verwendet darum eine zufällige Initialisierung der Cluster–Mittelpunkte mit je einem Muster aus dem Datensatz und weist dann Muster sukzessive einem der Cluster so zu, daß die Fehlerfunktion minimal ist. In diesem Prozeß können auch Muster aus einem Cluster in einen anderen wandern.

Nachdem die Gruppierungen abgeschlossen sind, geht es darum, die Gruppen in ein möglichst kleines Volumen einzuschließen. Dazu wird jeder Cluster zuerst einer Hauptachsentransformation unterworfen, um die Richtungen der stärksten Änderung im Phasenraum der Triggervariablen zu bestimmen. Diese Richtungen werden durch die Eigenvektoren der Verteilungsmatrix S_i jedes Clusters gegeben:

$$S_i = \sum_{\mu_i=1}^{N_i} (\vec{x}^{\mu_i} - \langle\vec{x_i}\rangle)^T (\vec{x}^{\mu_i} - \langle\vec{x_i}\rangle). \tag{3.14}$$

Die zugehörigen Eigenwerte messen die Varianz σ^2 der Punktdichte entlang der Eigenvektorrichtung. Wählt man nun Hyperflächen, deren Normalenvektoren durch die Eigenvektoren gebildet werden, so kann man mit diesen einen Hyperwürfel bilden, der den Cluster einschließt. Für die Entfernung der Hyperflächen vom Clustermittelpunkt werden 3σ gewählt.

Der Algorithmus, der bestimmt, ob ein Punkt in einem Cluster liegt, kann sehr einfach auf ein neuronales Netz abgebildet werden. Für jeden Eingabewert wird ein Neuron pro Cluster benötigt, wenn man eine Transferfunktion der Gestalt

$$g(x) = \Theta(x+a) \cdot (1 - \Theta(x-a)) \qquad \text{mit } \Theta(x) = \begin{cases} 1 & : \ x > 0 \\ 0 & : \ x \leq 0 \end{cases} \tag{3.15}$$

verwendet. Der Wert $2a$ ist der Abstand der beiden Hyperflächen voneinander. Die Gewichte zur Ausgabeschicht sind auf eins gesetzt und die Werte für die Schwellen verschwinden. Damit ergibt sich ein Wert von '1', wenn ein Punkt im Hyperwürfel liegt, ansonsten ist der Ausgabewert '0'.

3.2.3 Kegelschnitte

In diesem Abschnitt soll ein alternativer Algorithmus vorgestellt werden, der das Problem der langen Lernzeiten mit dem Backpropagation-Algorithmus vermeidet und als Beispiel für die Flexibilität der verwendeten Hard– und Software dient. Die Grundidee der Kegelschnitt–Methode besteht darin, die Aktivierungsfunktion $g(x)$ durch eine Reihenentwicklung zu ersetzen. In einem ersten Schritt kann man sie linearisieren ($g(x) = x$) und erhält dann aus den Gleichungen 3.4 und 3.5 für einen Ausgabeknoten

$$o(x_i) = \sum_{i=0}^{n} \hat{\omega}_i x_i \qquad \text{mit } \hat{\omega}_i = \sum_{j=0} \tilde{\omega}_j \omega_{ji}. \tag{3.16}$$

Die Suche nach dem Minimum der Fehlerfunktion 3.6 führt zu einem linearen Gleichungssystem, das sich durch Matrixinversion lösen läßt. Mit dieser Methode werden **Matrixinversion** die Klassen jedoch nur mit *einer* Hyperebene getrennt. Das Verfahren kann durch die Hinzunahme von Termen zweiter Ordnung in den Eingabegrößen verbessert werden, was zur Gleichung

$$o_k(x_i) = \sum_{i=0}^{n} \hat{\omega}_i x_i + \sum_{i \leq j} \hat{\omega}_{ij} x_i x_j \tag{3.17}$$

führt. Wie zuvor kann das Minimum der Fehlerfunktion analytisch bestimmt werden. Die Koeffizienten des sich ergebenden linearen Gleichungssystems sind lineare

Funktionen der Korrelationsmatrix erster bis vierter Ordnung. Die Trennebenen, die dieses 'Netzwerk' implementiert, können als Kegelschnitte in einem mehrdimensionalen Raum interpretiert werden. Neuronale Netzwerke mit derselben Anzahl von Eingabegrößen haben im allgemeinen eine etwas bessere Trennfähigkeit. Durch die Hinzunahme weiterer Größen ist die Leistung der Kegelschnitt–Methode jedoch leicht steigerbar, ohne daß der Rechenzeitaufwand stark wächst. Dies ist einer der Hauptvorteile dieser Methode.

3.3 Hardware

Triggerbox

Die Hardware des neuronalen Level-2-Triggers ist im Überblick in Abbildung 3.5 dargestellt. Die Hauptkomponenten werden von den sogenannten Triggerboxen gebildet. Sie bestehen aus einem *Data Distribution Board* (DDB) und einer CNAPS/VME–Karte. Bei den DDBs handelt es sich um Eigenentwicklungen des Max-Planck-Instituts. Die CNAPS/VME–Karten sind ein kommerziell verfügbares Produkt der Firma *Adaptive Solutions Inc.*. Das voll bestückte System besteht aus zehn Triggerboxen, ist vom Design her jedoch nicht auf diese Zahl beschränkt. Jede dieser Einheiten stellt die Hardware für die Implementierung eines Selektionsalgorithmusses zur Verfügung. Die DDB hat hauptsächlich die Aufgabe, die Daten der Subdetektoren vorzuprozessieren und als Eingabewerte an die CNAPS–Karte weiterzureichen. Der CNAPS–Parallelrechner berechnet dann das Ergebnis mittels des zugehörigen Programms und aufgrund der Eingangsdaten und der zuvor geladenen Parameter (Gewichte). Die Zuordnung einer eigenen DDB zu jedem CNAPS ermöglicht es, für jeden Algorithmus die optimale Untermenge der Daten des Detektors auszuwählen.

Receiver

Bevor die Vorverarbeitung in den DDBs beginnen kann, müssen die Daten des Detektors bzw. der Subdetektoren vom L2–System empfangen werden. Dazu dienen die Receiver–Karten, die die Daten einlesen und auf dem L2–Bus ablegen.

Der neuronale Trigger wird von einer SUN SPARCstation 20 gesteuert, kontrolliert und überwacht. Die Workstation ist über ein SBUS–VME-Interface sowohl mit den CNAPS/VME–Karten als auch mit den DDB–Karten verbunden. Sie ist verantwortlich für das Laden der Programme und der Parameter der CNAPS–Rechner, die Verifizierung des Ladevorgangs, das Starten des Triggers und die Überwachung mittels einer Monitoring–Karte, die *Spy–Card* genannt wird. Diese Karte ist im selben Gehäuse untergebracht wie die DDB–Karten und liest den L2–Bus, die Ergebnisse der Vorverarbeitung und die Ergebnisse der CNAPS–Berechnungen mit.

CTL2

Die boole'schen Entscheidungen der CNAPS Berechnungen werden über Kabel direkt an die zentrale Level-2-Triggerlogik geleitet.

3.3.1 Das Bus–System und die Receiver

Die Subdetektoren speichern ihre Level-2-Triggerinformationen für mehrere Strahlkreuzungszeitpunkte in sogenannten *Storecards* zwischen. Wenn der Level-1-Trigger

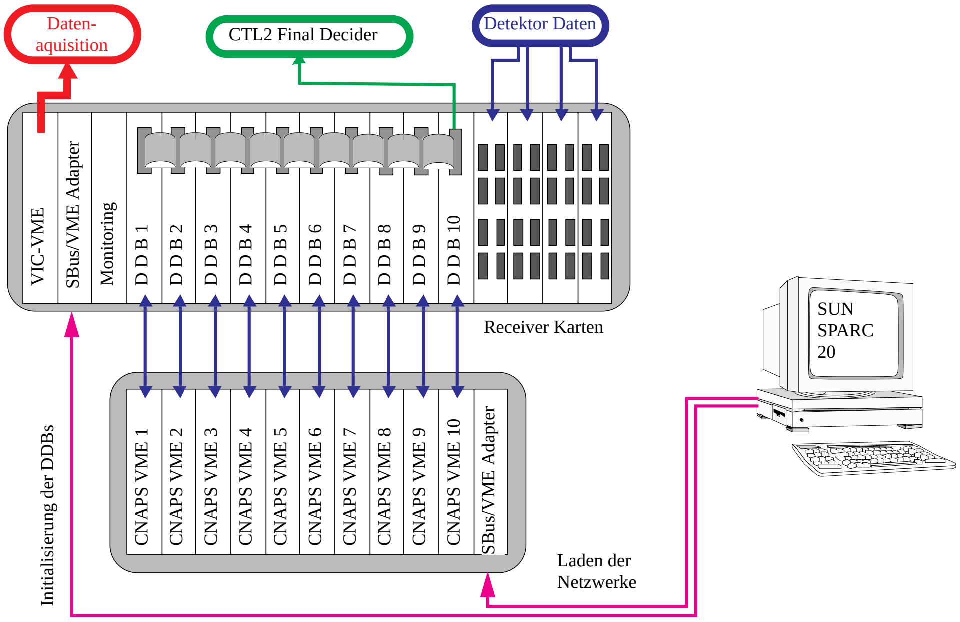

Abbildung 3.5: *Überblick über die Hardware des neuronalen L2-Triggers.*

ein Ereignis akzeptiert, wird die Level-2-Information zum zugehörigen Zeitpunkt aus den Storecards ausgelesen und an die Level-2-Triggersysteme gesendet. Dazu werden zwei verschiedene Bus–Systeme verwendet, für Kalorimeterdaten der Quick–Bus, für alle anderen der PQZP–Bus [57].

Auf Seiten der Level-2-Trigger stehen Receiver–Karten zum Empfang der Daten. Die Receiver sind in der Lage, die zu unterschiedlichen Zeitpunkten auf verschiedenen Leitungen ankommenden Informationen (16-Bit–Worte) zwischenzuspeichern und synchronisiert auf den L2–Bus zu verteilen [58]. Der L2–Bus ist ein paralleler Datenbus mit 8 Sub–Bussen zu je 16 Bits, der mit einer Frequenz von 10 MHz im sogenannten Interleaved-Modus arbeitet und damit eine effektive Übertragungsfrequenz von 20 MHz besitzt. Er bildet die Datenquelle für die Data Distribution Boards.

L2–Bus

3.3.2 Die Vorprozessierungskarte (DDB)

Die Vorprozessierungskarten DDB sind in einem eigenen Gehäuse untergebracht, das an den L2–Bus angeschlossen ist. Die L2–Bus-Daten werden in jeder DDB je nach ihrem Datentyp selektiert und durchlaufen dann verschiedene Vorverarbeitungsprozeduren. Der erste Schritt besteht z.B. darin, daß 16-Bit–Worte in 8-Bit–Bytes aufgeteilt werden oder Wörter in Bits zerlegt werden können. Dies geschieht durch Verwendung einer schnellen Look-Up-Table (LUT), d.h. einer Tabelle, in der die Aus-

DDB

Look-Up-Table

gabewerte für bestimmte Eingabewerte fest vorgegeben sind. Details zum genauen Aufbau und der Funktionalität der Vorprozessierungskarte können der Literatur entnommen werden [59].

Danach geschehen die eigentlichen Vorprozessierungen, in denen z.B. die Anzahl der gesetzten Bits über mehrere Wörter hinweg bestimmt wird, die Summe aus einzelnen Bytes oder Wörtern gebildet wird oder Bits selektiert werden. Die Daten können natürlich auch unverändert weitergeleitet werden. Die Hauptarbeit des Vorprozessierens wird von einem frei programmierbaren XILINX 40xx Chip übernommen, so daß man auf Änderungen der Datenformate flexibel reagieren kann. Die Ergebnisse der Data Distribution Boards stehen im sogenannten Selection RAM, einem Speicherbereich auf der DDB–Karte.

Selection RAM

Teile dieses Speichers können mit Hilfe von Selektionsmasken ausgewählt werden und über einen direkten Parallelbus zu einem Empfänger (Mezzanine–Karte) auf der CNAPS/VME–Karte gesendet werden. Diese Speicherinhalte, die den Eingabewerten der Netzwerke entsprechen, gelangen dann unmittelbar in den lokalen Bus der CNAPS/VME–Karte. Die maximale Übertragungsrate beträgt 26.67 MByte/s.

3.3.3 Die CNAPS/VME–Karte

In Abbildung 3.6 ist die CNAPS/VME–Karte mit den wichtigsten Elementen dargestellt. Die CNAPS–Karte ist ein Parallelrechner, der nach dem *Single Instruction Multiple Data*–Prinzip (SIMD) konzipiert ist. Auf der linken Seite sind die vier Steckplätze für die CNAPS–Chips zu sehen. Jeder dieser Chips (auch Array genannt) besitzt 64 Prozessoren (PN, Processor Nodes).

Der innere Aufbau eines CNAPS–1064 Chips ist in Abbildung 3.7 zu sehen. Die einzelnen Prozessoren sind logisch in einer Reihe nebeneinander angeordnet und mit dem Eingabe–, Ausgabe–, Befehls– und Inter-PN–Bus verbunden. Die 8-Bit breiten Ein– und Ausgabe–Busse werden für den Datentransfer vom und zum CNAPS Chip verwendet. Der Befehls–Bus hat eine Breite von 32 Bit und dient zur Steuerung aller Prozessoren (SIMD). Der Inter-PN–Bus ermöglicht die direkte Kommunikation zwischen benachbarten Prozessoren, die Breite ist jedoch auf zweimal 2 Bit beschränkt.

Der innere Aufbau eines Prozessors ist schematisch in Abbildung 3.8 dargestellt. Die Hauptbestandteile sind die arithmetischen Einheiten wie der Addierer (32 Bit) und der Multiplizierer (24 Bit), die logische Einheit, die Registereinheit, der 4 KByte Speicher und eine Puffereinheit. Sämtliche Berechnungen werden in Festkommaarithmetik durchgeführt. Der Feedforward–Schritt eines neuronalen Netzes mit 64 Eingabegrößen, 64 Knoten in der versteckten Schicht und einem Ausgabeknoten kann mit dem CNAPS in 166 Taktzyklen berechnet werden, was bei einer Taktfrequenz von 20.8 MHz einer Zeit von 8 μs entspricht. Die geringe Anzahl von Taktzyklen ist dadurch bedingt, daß der CNAPS Chip einen Multiplizier– und Addierschritt einer Vektor–Matrixmultiplikation in einer Takteinheit durchführen kann und dabei gleichzeitig einen neuen Eingabewert einlesen kann. Die Aktivierungsfunktion ist

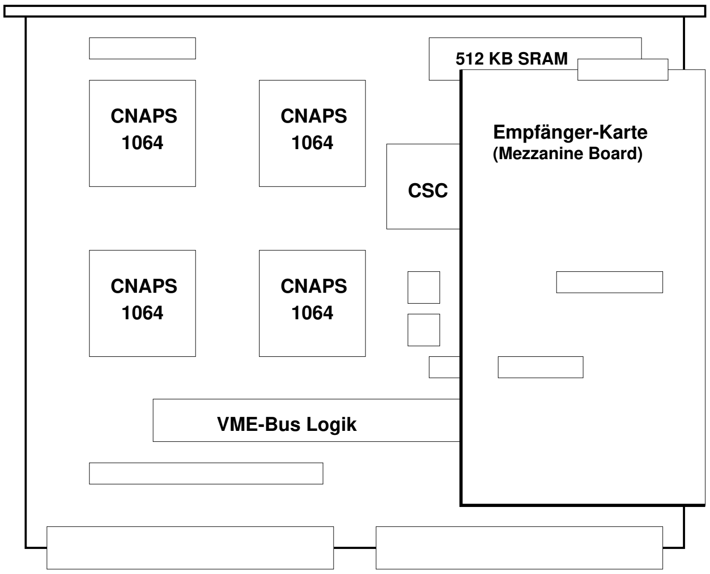

Abbildung 3.6: *Die CNAPS/VME–Karte.*

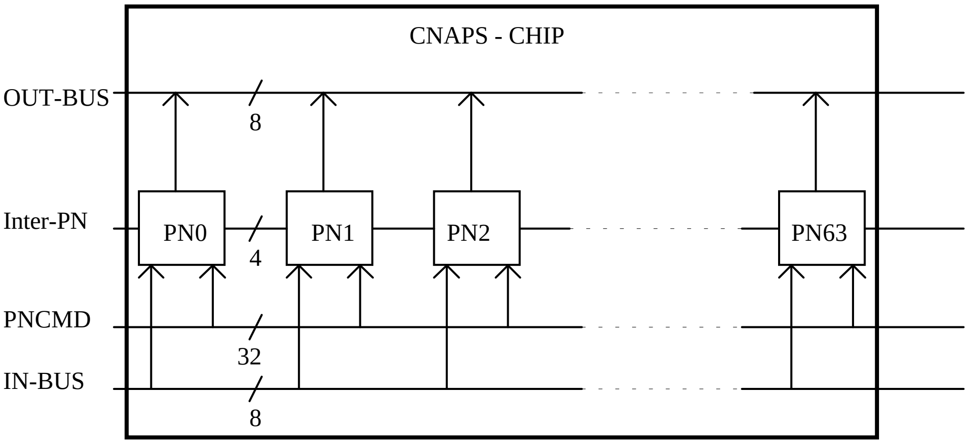

Abbildung 3.7: *Der innere Aufbau eines CNAPS–1064 Chips.*

als 10–Bit Look-Up-Table implementiert und ist deshalb in den Zeitanforderungen vernachlässigbar.

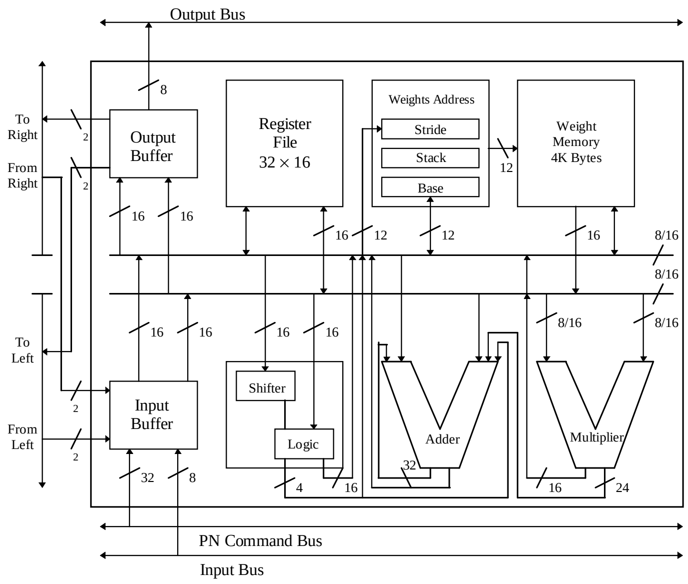

Abbildung 3.8: *Der interne Aufbau eines Prozessors.*

Berücksichtigt man die ungefähre Vorverarbeitungszeit der DDB, die etwa 4–8 μs benötigt, so ist die Triggerentscheidung leicht innerhalb der geforderten 20μs erreichbar.

3.4 Software

On-line Software

Off-line Software

Die Realisierung des neuronalen Level-2-Triggerprojekts war nur durch den Einsatz moderner Softwarewerkzeuge möglich. Die Entwicklung der Software wurde in zwei Bereiche unterteilt. Die Steuerung und Kontrolle des Triggers wurde unter dem Begriff On-line Software zusammengefaßt. Die Bereitstellung der Daten, die Untersuchung von Algorithmen und die Optimierung der Netzwerk– bzw. Encapsulator–Parameter fielen dem Gebiet der Off-line Software zu.

Die programmiertechnische Nähe zur Hardware gebot es, im On-line Sektor als Programmiersprache 'C' zu verwenden. Auch im Off-line Bereich wurde durch die Verwendung von Workstations mit UNIX Betriebssystemen und die Vielzahl von frei erhältlichen Softwaretools, die auf die Programmiersprache 'C' abgestimmt waren, diese Wahl nahegelegt. Gleichzeitig wurde sicher gestellt, daß auf die bestehenden FORTRAN–Programmbibliotheken von CERN [60] und H1 [61] zugegriffen werden konnte [62]. Der Austausch der Software, die gemeinsame Entwicklung von Programmen auf verschiedenen Plattformen und an verschiedenen Orten und die Versionsverwaltung wurden durch das Concurrent Versioning System (CVS) [63] möglich gemacht.

CVS

In den folgenden Abschnitten wird ein Überblick über die wichtigsten Komponenten der On-line und Off-line Software gegeben.

3.4.1 On-line Software

Die Aufgabe der On-line Software besteht darin, den neuronalen L2-Trigger zu steuern und zu überwachen. Dazu müssen folgende Einzelaufgaben gelöst werden :

- Programmierung der Empfänger–Karte
- Programmierung der DDBs
- Handhabung der Algorithmen und Parameter in einer Datenbank
- Herunterladen von Algorithmen und Gewichten zu den CNAPS–Karten
- Programmierung der Spy–Card
- Aufzeichnung der aktuellen Konfiguration in einer Datenbank
- Starten und Beenden des neuronalen Triggers
- lokale Aufzeichnung von Überwachungsdaten und Histogrammen.

Die gesamte On-line Software ist als Client–Server Applikation entworfen worden. Dadurch ist es möglich, von beliebigen Orten aus die Hardware zu überwachen. Auch ein Eingreifen ist möglich, jedoch dadurch beschränkt, daß Hardware–verändernde Aktionen nur von **einem** Benutzer durchgeführt werden können, der zudem über das Experten–Paßwort verfügen muß. Der Zugang zu diesen Funktionen wird mit einer grafischen Benutzerschnittstelle, dem NnCockpit, ermöglicht. Das Kernstück der On-line Software wird von einem Server-Prozeß, dem sogenannten NnServer gebildet, der permanent auf der SUN SPARCstation läuft. Dieser Prozeß stellt Sockets zur Verfügung, über die sich die Clients anbinden können. Der Server startet zwei weitere Prozesse, den NnLoader–Prozeß und den NnMonitor–Prozeß. Mit Hilfe des NnLoader werden die einzelnen Hardwarekomponenten geladen bzw. programmiert und eine Kommunikation mit der zentralen Datenaquisition aufgebaut. Der NnMonitor–Prozeß liest Daten zur Überwachung aus der Empfängerkarte, aus den DDBs und per Spy–Card aus dem L2–Bus aus. Die drei Prozesse koordinieren sich gegenseitig via Inter Process Communication (IPC) und tauschen ihre Informationen über gemeinsamen Speicher (Shared Memory) aus. Die lokale Datenbank bildet für alle drei Komponenten die Basis für Informationen. Das Schema ist grafisch in Abb. 3.9 dargestellt.

NnCockpit
NnServer
NnLoader
NnMonitor
IPC

Der NnMonitor füllt die ausgelesenen Daten in Histogramme und berechnet aus den Entscheidungen der einzelnen Triggerboxen die Rate des Triggers. Da dem Level-2-Trigger auch die Ergebnisse des Level 1 bekannt sind, kann man bereits während des Monitorings die Entscheidungen des Level-2-Triggers überwachen.

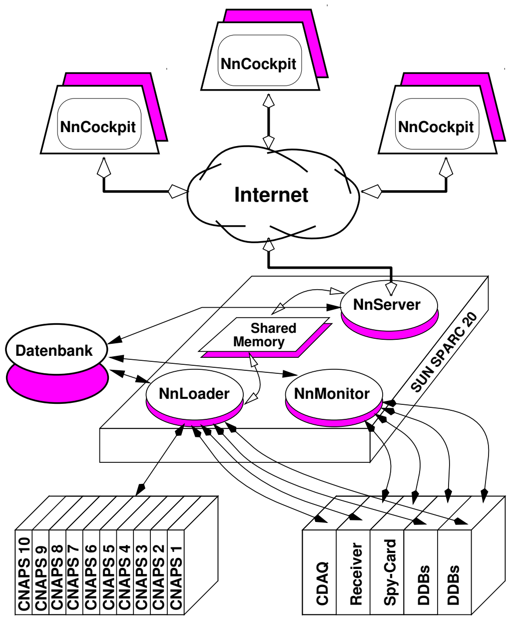

Abbildung 3.9: *Überblick über die On-line Software.*

3.4.2 Off-line Software

Die Vorbereitung von neuronalen Netzen und Background Encapsulatoren für den realistischen Einsatz im H1–Experiment bedarf eines gründlichen Verständnisses der anvisierten Physik und der Triggerdaten. Dazu sind eine Reihe von Programmen erstellt worden, deren Zusammenspiel die folgenden Aufgabenstellungen bearbeiten hilft:

- Selektion der Ereignisse
- Extrahierung von Triggerinformation aus den Ereignisdaten
- Erkennung und Auswahl von wichtigen Triggergrößen
- Erstellung von Datensätzen für Optimierung und Test
- Verglichende Untersuchung der Leistungsfähigkeit verschiedener Algorithmen
- Optimierung der Lernparameter im Fall von neuronalen Netzwerken
- Durchführung der Parameteroptimierung (Training von neuronalen Netzen)
- Überprüfung der Ergebnisse
- Simulation der Hardware.

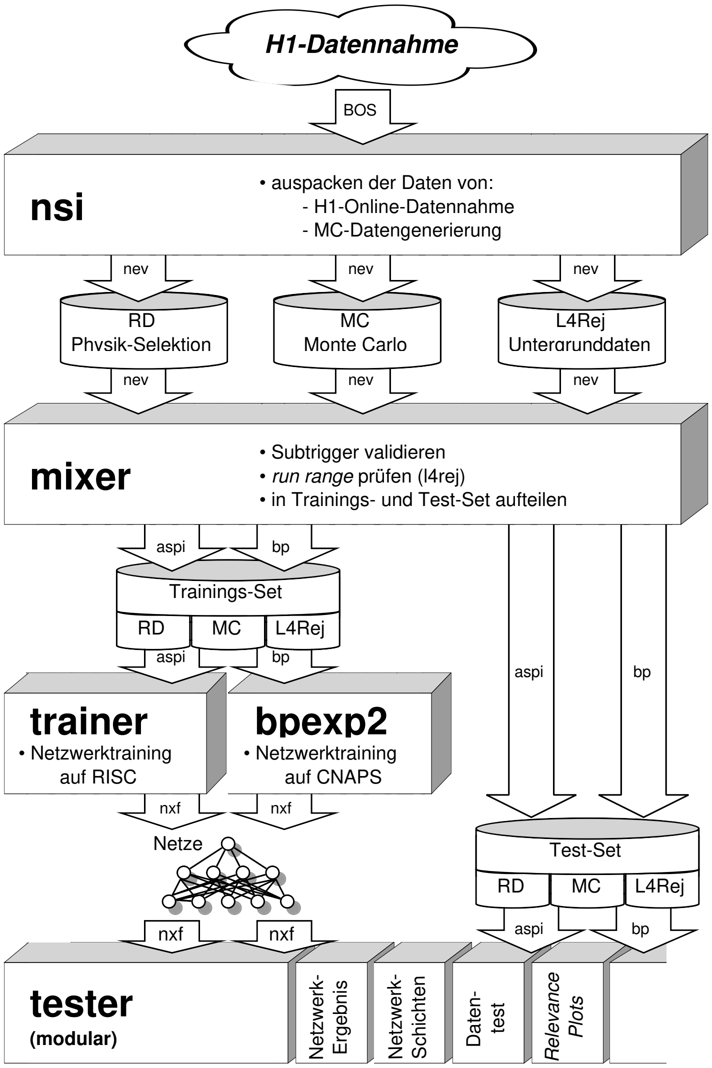

Abbildung 3.10: *Überblick über die Off-line Software.*

In Abbildung 3.10 ist die Kette von einzelnen Modulen und deren Verknüpfung untereinander dargestellt.

Den Anfang und das Ende der Kette bildet der Netzwerk–Simulator *nsi* [64] (siehe auch Anhang A). Er hat zwei unterschiedliche Funktionen. Zum einen dient er der Extraktion der Triggergrößen aus unterschiedlichsten Datenblöcken des Datenstroms. Diese Größen werden im sogenannten nev–Format[3] abgespeichert. Zum anderen ist er in der Lage, den neuronalen Level-2-Trigger in seiner Funktionalität vollständig zu simulieren.

nsi **nev**

[3]New Event Vector

mixer Das nev–Datenformat kann vom Programm *mixer* [65] gelesen werden. Der mixer ermöglicht es, die nev–Daten verschiedener Physik– und Untergrundklassen in einen Trainingsdatensatz zusammenzumischen und gleichzeitig Testdatensätze zu generieren. Es können beim 'mixen' gleichzeitig Selektionen durchgeführt werden, z.B. auf bestimmte Subtriggerelemente oder Zeitabschnitte der Datennahme. Die Daten können, abhängig vom verwendeten Programm für das Training, in verschiedenen Formaten geschrieben werden.

Zum Trainieren von Netzwerken stehen zwei verschiedene Programme zur Verfügung. **trainer** Die *trainer*–Software [66, 67] ist auf allen gängigen Workstations lauffähig und verlangt bei den Eingabedaten nach dem sogenannten aspi–Datenformat. Es benutzt als Kern der Simulation den Backpropagation-Schritt der Aspirin/MIGRAINES Software [68]. **bpexp2** Das *bpexp2* Programm [69] ist speziell für die Verwendung des CNAPS–Parallelrechners für Off-line Zwecke entwickelt worden. Mit ihm sind sehr umfangreiche Studien zur Variation der Lernparameter möglich, die auf einem seriellen Rechner zuviel Rechenzeit in Anspruch nähmen.

cluclu Zur Generierung eines Background Encapsulators ist das Paket *cluclu* [70] entwickelt worden. Alle drei Programme schreiben die Parametersätze im sogenannten nxf–Format[4] , das von nachfolgenden Programmen im On-line und Off-line Bereich eingelesen werden kann.

Die Untersuchung der trainierten Netzwerke und optimierten Background Encapsu- **tester** lator erfolgt mit dem *tester* [71]. Er berechnet die Ergebnisse der verschiedenen Algorithmen zum selben Ereignis, das er aus beliebigen Testdatensätzen auswählen kann. Das modulare Design des testers erlaubt die Erweiterung um neue Funktionen, wie z.B. die Darstellung der mittleren Aktivität in einer Netzwerkschicht oder das Prüfen von Fehlertoleranz gegenüber Störung der Eingangsgrößen.

3.5 Die zentrale Level-2 Triggerlogik

Die Ergebnisse des neuronalen und des topologischen Level-2-Triggers werden in Form von L2–Triggerelementen an die zentrale Level-2 Triggerlogik gesendet. Diese Triggerelemente haben boole'sche Werte. Je nach Algorithmus haben die Bits eine andere Bedeutung, so ist z.B. der Wert '1' beim Background Encapsulator der Indikator für ein Untergrundereignis, wohingegen derselbe Wert bei einem neuronalen Netz für ein Physikereignis steht. Die angemessene Interpretation der Bits ist eine der Aufgaben der CTL2.

In Kapitel 3.1 wurde beschrieben, daß einzelne Netzwerke nur die Rate ganz bestimmter Level-1 Subtrigger senken sollen. Die Bildung von Koinzidenzen zwischen Subtriggern und Level-2 Triggerelementen ist ein anderer Teil der CTL2-Funktionalität.

Falls die Rate einzelner Level-2 Triggerelemente auch nach der Koinzidenzbildung

[4] Network eXchange Format