Untersuchungen zur Erkennung von Sekundärvertices in hochenergetischen Teilchenreaktionen mit Hilfe Neuronaler Netze

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

von

Martin Schied

angefertigt am

Max-Planck-Institut für Physik

München
Juli 1998

Erstgutachter: Prof. Dr. Christian Kiesling

Zweitgutachterin: Prof. Dr. Dorothee Schaile

Inhaltsverzeichnis

Kapitel 1

Einleitung

Die Elementarteilchenphysik, die auch als Hochenergiephysik bezeichnet wird, beschäftigt sich mit dem Aufbau der subatomaren Materie.

Vor nicht einmal 90 Jahren machten Rutherford, Geiger und Marsden ein für die Physik äußerst entscheidendes Experiment [1]. Sie untersuchten die Ablenkung von α-Teilchen an einer sehr dünnen Goldfolie (wenige μm). Die meisten α-Teilchen durchdrangen diese unabgelenkt. Allerdings erfuhren einige wenige Ablenkungen unter großen Streuwinkeln. Man bekam dadurch eine völlig neue Vorstellung vom Aufbau des Atoms. Danach ist fast die gesamte Masse eines Atoms in seinem Atomkern konzentriert. Dieser Atomkern hat einen mehr als 10000 mal kleineren Radius als das Atom selbst. Bereits 1920 vermutete Rutherford, daß zum Aufbau des Atomkerns die damals bekannten Teilchen Proton und Elektron nicht genügen. Es wurde also ein weiteres Teilchen, nämlich das Neutron postuliert. Das Neutron wurde dann auch 1932 von Chadwick experimentell nachgewiesen. In den folgenden Jahren setzte sich die Suche nach weiteren Teilchen fort. Es wurden das Positron, der positive Partner des Elektrons (Anderson 1932, 1933), das Myon (Anderson 1937, 1938, Street 1937), das π-Meson (Lattes 1947) und das K-Meson (Rochester 1947) gefunden [2].

Um einen immer besseren Aufschluß über die Struktur der Materie zu bekommen, benötigt man schnelle Teilchen als Geschoße auf die zu untersuchende Materie. Aufgrund der Wellennatur der Materie, die sich speziell durch die deBroglie-Gleichung $\lambda = h/p$ ausdrückt, ist klar, daß man zur Auflösung immer kleinerer Strukturen immer höhere Impulse p der Geschoßteilchen braucht. Schließlich ist die Neuerzeugung eines Teilchens der Masse m frühestens bei einer Geschoßenergie $W = mc^2$ möglich. Oft ist eine sehr viel höhere Energie nötig. Neben der hohen Energie kommt es jedoch auch auf eine hohe Intensität der Teilchen an. Mit wachsender Intensität nimmt nämlich die Wahrscheinlichkeit für die Produktion neuer Teilchen zu. Apparate, die diesen Anforderungen entsprechen, sind die Teilchenbeschleuniger.

In ersten Versuchen beschleunigten J.D. Cockroft und E. Walton Protonen auf 300 keV [3]. Weitere Entwicklungen von Beschleunigungsanlagen waren der Van-de-Graaf-Generator (1933) mit Energien bis zu 10 MeV, das Zyklotron (E.O. Lawrence 1932, um 10 MeV) und das Synchrotron (V.I. Vehsler, E.M. McMillan) mit einigen 100 MeV. Die Experimente zu dieser Zeit waren sogenannte Festtarget-Experimente, da der beschleunigte Teilchenstrahl auf ein ruhendes Ziel gelenkt wurde. Der für die Reaktion verfügbare Anteil der Teilchenenergie wächst jedoch im Schwerpunktssystem nur mit der Wurzel der Teilchenenergie. Eine Lösung, den Energieverlust durch die Bewegung des Schwerpunkts zu vermeiden, war es zwei gegenläufige Teilchenstrahlen zur Kollision zu bringen. So wird die gesamte kinetische Energie beider Teilchen für die Reaktion verfügbar.

Mit Hilfe solcher Beschleunigungsanlagen und entsprechender Detektoren konnten bis zum heutigen Zeitpunkt mehr als hundert Teilchen nachgewiesen werden [4]. Diese Teilchen werden in drei Gruppen eingeteilt: Leptonen, Hadronen und Eichbosonen. Die Hadronen setzen sich ihrerseits aus punktförmigen Teilchen, den sogenannten Quarks zusammen. Als fundamentale Teilchen existieren damit Leptonen, Quarks und Eichbosonen.

Diese Elementarteilchen mit ihren Wechselwirkungen untereinander finden ihre Erklärung im Standardmodell der Teilchenphysik (SM). Das SM wird kurz in Kapitel 2 dieser Arbeit beschrieben. Ein Teilchen im SM wurde jedoch bis jetzt noch nicht entdeckt: das Higgs-Boson. Um dieses Higgs-Boson, falls es existiert, zu erzeugen, ist ein Beschleuniger mit einer Schwerpunktsenergie im TeV-Bereich von Nöten. Am CERN soll daher der Proton-Proton-Speicherring LHC (**L**arge **H**adron **C**ollider) mit einer Schwerpunktsenergie von $\sqrt{s} = 14$ TeV errichtet werden [5]. Zum Nachweis der entstehenden Teilchen ist unter anderem der ATLAS-Detektor (**A** **T**oroidal **L**HC **A**pparatu**S**) geplant [6]. Dieser Beschleuniger und dieser Detektor werden in Kapitel 3 genauer beschrieben.

Auf die Physik, die am LHC möglich sein wird, geht in einem kurzen Überblick Kapitel 2 ein.

Interessante physikalische Teilchenzerfälle, wie beispielsweise der des Higgs-Bosons zeichnen sich durch schwere Quarks im Endzustand aus. Bei solchen Zerfällen treten praktisch immer Teilchen mit endlicher Lebensdauer ($\tau > 0$ s) auf. Solche Teilchen zerfallen noch innerhalb des Detektors weiter. Der Ort, an dem dieser Zerfall stattfindet, wird als sekundärer Vertex bezeichnet. Sekundäre Vertices bilden klare Signaturen bei Higgs-Zerfällen. Man könnte somit beispielsweise solche Higgs-Zerfälle durch eine Selektion von Ereignissen, die sich durch sekundäre Vertices auszeichnen, anreichern. Wegen der hohen Kollisionsrate, die es am LHC geben wird, ist es unmöglich zuerst alle Ereignisse aufzuzeichnen und diese nachträglich (offline) in einer Rekonstruktion zu bearbeiten. Daher sollten die Ereignisse, die sekundäre Vertices enthalten, bereits online ausgewählt werden. Man braucht also einen selektiven Trigger. Bisher ist jedoch eine solche Selektion erst nach einer aufwendigen Spurfindung möglich. Dies bedeutet lange Rechenzeiten, die eine Vertexrekonstruktion nur offline zulassen. Herkömmliche Triggermethoden versuchen schwere Quarks durch ihre semileptonischen Zerfälle zu markieren. Der Nachteil dieser Methode ist, daß nur etwa 10% der Zerfälle semileptonischen Charakter haben, während sich jedoch alle diese Ereignisse durch einen sekundären Vertex auszeichnen.

In dieser Arbeit wird untersucht werden, ob man alleine aus den schnell verfügbaren Detektortrefferpunkten von geladenen Teilchen mit Hilfe von neuronalen Netzen auf das Auftreten solcher sekundären Vertices in einem Ereignis schließen kann. Erläuterungen zum Thema neuronale Netze werden daher in Kapitel 4 gemacht. Die

Idee der Verwendung neuronaler Netze begründet sich in ihrer hochgradigen Parallelität. Sie könnten daher auf Triggerebene eingesetzt werden. In Kapitel 5 wird der Zerfall $D^{\star +} \rightarrow D^0\pi^+$ mit $D^0 \rightarrow K^-\pi^+$ betrachtet. Die Teilchen K^- und π^+ stellen hadronische Endzustände dar und entstehen an einem sekundären Vertex. Des weiteren wird hier die selbstprogrammierte Simulation des Barrelbereichs des inneren ATLAS-Detektors beschrieben und dargestellt wie Detektortrefferpunkte für die Eingabe in ein neuronales Netz vorbereitet werden müssen. Kapitel 6 gibt Einblicke wie neuronale Netze unter verschiedenen vereinfachten Bedingungen Zerfallslängen schätzen können. In Kapitel 7 werden anhand von einem $t\bar{t}$-Monte-Carlo Teilchenspurmultiplizitäten in bestimmten Detektorbereichen untersucht. Diese Ergebnisse werden dann benutzt, um realistischere Eingaben für ein neuronales Netz zu testen. Schließlich wird in Kapitel 8 eine Zusammenfassung der Ergebnisse und ein Ausblick für zukünftige Verfahren gemacht.

Kapitel 2

Physik bei ATLAS

Die Physik, die es zu erforschen und verstehen gilt, nimmt am LHC neue Dimensionen an. Dieses Kapitel beschäftigt sich mit dem Top-Quark und dem Higgs-Boson, dem wichtigsten Grund für den Bau des LHC mit seinen Detektoren. Bei ATLAS werden jedoch neben dem Higgs-Boson und dem Top-Quark auch andere interessante Bereiche untersucht werden, auf die in dieser Arbeit nicht näher eingegangen wird. Zu nennen sind detaillierte Studien zur B-Physik (CP-Verletzung, Messungen zu B_s^0-Mischzuständen, B-Baryonen-Zerfallsdynamik, Spektroskopie seltener B-Hadronen) und die Suche nach supersymmetrischen Teilchen [7, 8].

2.1 Das Standardmodell der Teilchenphysik

Die experimentelle und theoretische Arbeit der letzten 50 Jahre führte zur Entwicklung des Standardmodells (SM). Das Standardmodell beschreibt die Elementarteilchen und ihre Wechselwirkungen miteinander. Nachdem das SM in vielen Büchern [2, 9–11] detailliert beschrieben wird, soll hier nur ein kurzer Überblick gegeben werden. In der Theorie stellt das SM eine Eichfeldtheorie, die auf der Gruppe $SU(3) \otimes SU(2) \otimes U(1)$ basiert, dar. Die SU(3)-Gruppe beschreibt die starke, die $SU(2) \otimes U(1)$-Gruppe die Vereinheitlichung aus elektromagnetischer und schwacher Wechselwirkung.

Es gibt zwei grundlegende Typen von Teilchen. Das eine sind Spin 1/2-Teilchen, also Fermionen, das andere sind Eichbosonen mit Spin 1. Die Wechelwirkung der Fermionen miteinander erfolgt über Eichbosonen (siehe Tabelle 2.1). Man unterscheidet zwei elemantare Gruppen von Fermionen: Leptonen und Quarks. Die Gesamtheit der Fermionen teilt sich in drei „Familien“ oder „Generationen“ auf (siehe Tabelle 2.2). Die erste Klasse enthält das Up- und Down-Quark, das Elektron und das Elektronneutrino. „Normale“ Materie setzt sich aus diesen Teilchen zusammen. Das Charme- und das Strange-Quark bilden mit dem Myon und dem Myonneutrino die zweite Familie, das Bottom- und das Top-Quark bilden mit dem Tau und dem Tauneutrino die dritte Generation. Zu all diesen Fermionen gibt es auch Antiteilchen.

Wechselwirkung	Stärke	Boson	Ladung (e)	Masse (GeV/c^2)
elektromagnetisch	1/137	γ	0	0
schwach	10^{-5}	$W^{\pm}$	± 1	80.2
		Z^0	0	91.2
stark	$\simeq 1$	8 Gluonen	0	0

Tabelle 2.1: *Wechselwirkungen zwischen den Teilchen mit ihrer Stärke und den Eichbosonen. Die Stärke der jeweiligen Wechselwirkung ist über die Kopplungskonstante α gegeben.*

Familie	Leptonen			Quarks		
	Flavour	Ladung (e)	Masse (GeV/c^2)	Flavour	Ladung (e)	Masse (GeV/c^2)
erste	ν_e	0	0	u	2/3	0.002 - 0.008
	e	-1	0.0005	d	-1/3	0.005 - 0.015
zweite	ν_μ	0	0	c	2/3	1.0 - 1.6
	μ	-1	0.106	s	-1/3	0.1 - 0.3
dritte	ν_τ	0	0	t	2/3	180 $\pm$ 12
	τ	-1	1.8	b	-1/3	4.1 - 4.5

Tabelle 2.2: *Fermionen (Spin 1/2-Teilchen) im Standardmodell mit ihren Ladungen und Massen. Nicht dargestellt sind die jeweiligen Antiteilchen.*

Die einzelnen Kräfte werden mathematisch durch Eichtheorien beschrieben. Die elektromagnetische Kraft ist verbunden mit einer lokalen U(1) Eichinvarianz. Sie findet ihre Beschreibung in der Quantenelektrodynamik (QED). Alle geladenen Teilchen unterliegen der QED. Das Eichboson ist das masselose Photon γ.
Die schwache Wechselwirkung wird durch $W^{\pm}$- und Z^0-Bosonen vermittelt. Geladene Ströme resultieren vom Austausch eines $W^{\pm}$-Bosons. Diese treten beispielsweise auch beim Neutron β-Zerfall auf. Das Z^0 ist verantwortlich für den neutralen Strom. Da das Z^0 nur an Fermionen innerhalb derselben Familie koppelt, gibt es keine neutralen Ströme, die den Flavour ändern. Das $W^{\pm}$ koppelt im leptonischen Sektor auch nur[1] an Leptonen innerhalb derselben Familie. Das gilt jedoch nicht für den

[1] Am Super-Kamiokande wurden Neutrino-Oszillationen beobachtet. Damit hätten Neutrinos Masse und das $W^{\pm}$ würde nicht mehr nur an Leptonen innerhalb derselben Familie koppeln. Es wird dann ähnliches wie für den Quark-Sektor erwartet. [12]

Quark-Sektor. Hier koppelt das $W^{\pm}$ nicht direkt an die Masseneigenzustände d, s und b, sondern an Linearkombinationen daraus. Das wird mathematisch durch die Cabibbo-Kobayashi-Maskawa (CKM) Matrix ausgedrückt. Die Matrix läßt sich durch drei Eulerwinkel und eine nicht triviale Phase parametrisieren.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.1}$$

Die Werte der einzelnen Matrixelemente lauten:

$$V = \begin{pmatrix} 0.9745 - 0.9757 & 0.219 - 0.224 & 0.002 - 0.005 \\ 0.218 - 0.224 & 0.9736 - 0.9750 & 0.036 - 0.046 \\ 0.004 - 0.014 & 0.034 - 0.046 & 0.9989 - 0.9993 \end{pmatrix} \tag{2.2}$$

Es gibt damit beispielsweise folgende Möglichkeiten von Übergängen in das u-Quark:

$$\mathrm{d} \to \mathrm{u} + \mathrm{W}^- \qquad (\propto \mathrm{V_{ud}}) \tag{2.3}$$
$$\mathrm{s} \to \mathrm{u} + \mathrm{W}^- \qquad (\propto \mathrm{V_{us}}) \tag{2.4}$$
$$\mathrm{b} \to \mathrm{u} + \mathrm{W}^- \qquad (\propto \mathrm{V_{ub}}) \tag{2.5}$$

Die Übergangsamplituden sind proportional den entsprechenden Elementen der CKM-Matrix.
Die schwache Kraft verletzt die Parität P. An der schwachen Wechselwirkung nehmen nur linkshändige Neutrinos und rechtshändige Antineutrinos teil. Dies spiegelt sich wider in der V-A-Theorie (Vektor-Axialvektor-Theorie). Auch das Produkt aus Ladungskonjugation C und Parität P ist verletzt (CP-Verletzung). Die CP-Verletzung wurde im neutralen Kaonensystem entdeckt.
Die starke Kraft ist dafür verantwortlich, daß die Quarks innerhalb der Hadronen zusammengehalten werden. Die zugehörige Eichfeldtheorie ist die Quantenchromodynamik (QCD), eine Theorie mit lokaler SU(3)-Symmetrie. Die Teilchen tragen die Quantenzahl Farbe und die Wechselwirkung zwischen ihnen vermitteln acht Eichbosonen (Gluonen). Quarks haben drei, Gluonen acht verschiedene Farbzustände. Da die Gluonen selbst Farbladung tragen, können sie auch untereinander wechselwirken. Die starke Kraft vergrößert sich bei zunehmenden Abständen.
In den Jahren 1967-1968 entwickelten Weinberg und Salam eine Vereinheitlichung der schwachen und elektromagnetischen Wechselwirkung, die sogenannte elektroschwache Wechselwirkung. Sie beruht auf der SU(2)-Gruppe des „schwachen Isospins“ I und der U(1)-Gruppe der „schwachen Hyperladung“ Y (einer Phasentransformation). Die Theorie enthält einen Parameter, den sie nicht vorhersagen kann, den sogenannten Weinberg-Winkel Θ_W. Über diesen schwachen Mischwinkel werden aus einem

hypothetischen masselosen vierkomponentigen Feld die beobachtbaren Eichbosonen erzeugt. Das Photon ist masselos, jedoch haben das Elektron und die Bosonen $W^{\pm}$ und Z^0 Masse. Wenn man für das Elektron und für die Bosonen der Lagrange-Dichte explizite Massenterme hinzufügt, verletzt dies die Eichinvarianz und die „Theorie“ wird sogar nicht renormierbar.
Eine Lösung dieses Problems liefert der Higgs-Mechanismus. Die W-Bosonen und das Z-Boson erhalten über spontane Symmetriebrechung eines skalaren Feldes, des sogenannten Higgs-Feldes, Masse. Dieser Mechanismus fordert ein weiteres Spin 0-Teilchen - das Higgs-Boson.

2.2 Das Top-Quark

Das CDF Experiment lieferte 1994 den Beweis für die Existenz des Top-Quarks [13]. Dieses Ergebnis wurde von der D0 Kollaboration bestätigt [14]. Die Topmasse wurde bestimmt zu 180 ± 12 GeV/c^2 [15]. Am LHC werden schon bei geringer Anfangsluminosität von $\mathcal{L} = 10^{32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ etwa 6000 $t\bar{t}$ Paare pro Tag produziert. Das macht den LHC zu einer wahren „Top-Fabrik“ und erlaubt es, die Produktion und den Zerfall des Top-Quarks detailliert zu studieren.
Die häufigsten Produktionsmechanismen für Top-Quarks sind in Abbildung 2.1 dargestellt [16].

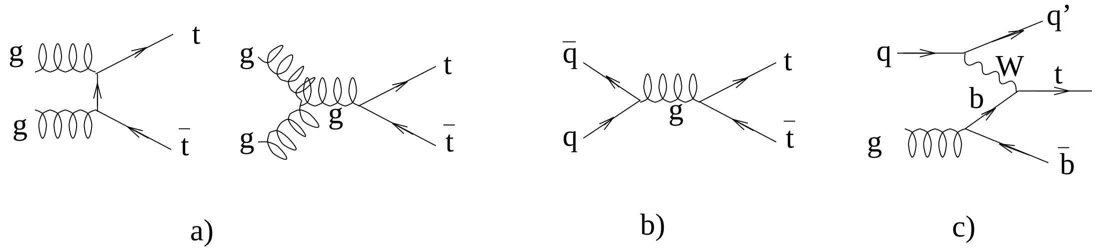


Abbildung 2.1: *Die einzelnen Produktionsmechanismen für Top-Quarks. a, Gluon-Gluon Fusion. b, Quark-Antiquark Fusion. c, W-Gluon Fusion*

Die Produktion eines einzelnen Top-Quarks über W-Gluon Fusion ist viel seltener als die $t\bar{t}$ Produktion über Gluon-Gluon Fusion oder Quark-Antiquark Fusion. Das liegt daran, daß die W-Gluon Fusion über die schwache Wechselwirkung abläuft, während bei den anderen beiden Prozessen nur starke Kopplungen auftreten.
Innerhalb des Standardmodells zerfällt das Top-Quark fast ausschließlich in ein b-Quark und ein W-Boson (siehe Abbildung 2.2).
Die Begründung liegt in der CKM-Matrix. Der Wert des Matrixelements V_{tb} ist nahe bei Eins (siehe Gl. 2.2). Die Elemente V_{ts} und V_{td} sind fast null (siehe Gl. 2.2) und damit ist der direkte Übergang $t \to s$ und $t \to d$ sehr stark unterdrückt. Das W-Boson zerfällt entweder in ein Lepton und ein Neutrino oder in ein Quark und

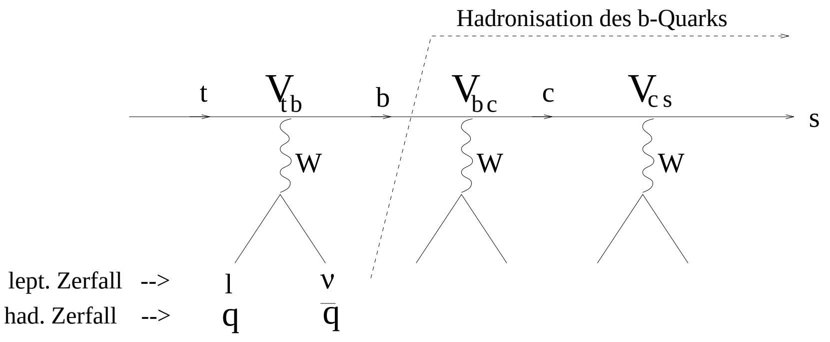


Abbildung 2.2: *Der Zerfall des Top-Quarks.*

ein entsprechendes Antiquark. Dabei kann das W-Boson jedoch aus energetischen Gründen nicht mehr in ein t-Quark oder ein b-Quark übergehen.
Die in Tabelle 2.3 und in Abbildung 2.2 mit q und $\overline{\mathrm{q}}$ bezeichneten Quarks stehen also für die vier Quarks u, d, c, s.
Die möglichen Zerfälle des W^+-Boson sind:

$$\mathrm{W}^+ \to \mathrm{e}^+ + \nu_\mathrm{e} \tag{2.6}$$

$$\mathrm{W}^+ \to \mu^+ + \nu_\mu \tag{2.7}$$

$$\mathrm{W}^+ \to \tau + \nu_\tau \tag{2.8}$$

$$\mathrm{W}^+ \to \mathrm{u} + \overline{\mathrm{d}}_\mathrm{c} \tag{2.9}$$

$$\mathrm{W}^+ \to \mathrm{c} + \overline{\mathrm{s}}_\mathrm{c} \tag{2.10}$$

Die in den Gleichungen 2.9 und 2.10 eingeführten Bezeichnugen $\overline{\mathrm{d}}_\mathrm{c}$ und $\overline{\mathrm{s}}_\mathrm{c}$ stellen Mischzustände aus Quarks dar:

$$\overline{\mathrm{d}}_\mathrm{c} = \cos\Theta_\mathrm{c}\ \overline{\mathrm{d}} + \sin\Theta_\mathrm{c}\ \overline{\mathrm{s}} \tag{2.11}$$

$$\overline{\mathrm{s}}_\mathrm{c} = \cos\Theta_\mathrm{c}\ \overline{\mathrm{s}} - \sin\Theta_\mathrm{c}\ \overline{\mathrm{d}} \tag{2.12}$$

Θ_c bezeichnet man als Cabibbo-Winkel. Es gilt $\Theta_\mathrm{c} \approx 13^0$. Damit ergibt sich die Kopplungsstärke der Quarks an die W-Bosonen:

$$\begin{array}{ccc}
W^+ \to u\overline{d} & \propto |V_{ud}|^2 \approx \cos^2\Theta_c \approx (0.98)^2 & (Cabibbo - erlaubt) \\
W^+ \to u\overline{s} & \propto |V_{us}|^2 \approx \sin^2\Theta_c \approx (0.23)^2 & (Cabibbo - unterdrückt) \\
W^+ \to c\overline{s} & \propto |V_{cs}|^2 \approx \cos^2\Theta_c \approx (0.98)^2 & (Cabibbo - erlaubt) \\
W^+ \to c\overline{d} & \propto |V_{cd}|^2 \approx \sin^2\Theta_c \approx (0.23)^2 & (Cabibbo - unterdrückt)
\end{array}$$

Zerfallskanal	Verzweigungsverhältnis
$t\bar{t} \rightarrow (q\bar{q}b)(q\bar{q}\bar{b})$	36/81
$t\bar{t} \rightarrow (q\bar{q}b)\ (e\nu\bar{b})$	12/81
$t\bar{t} \rightarrow (q\bar{q}b)\ (\mu\nu\bar{b})$	12/81
$t\bar{t} \rightarrow (q\bar{q}b)\ (\tau\nu\bar{b})$	12/81
$t\bar{t} \rightarrow (e\nu b)\ (\mu\nu\bar{b})$	2/81
$t\bar{t} \rightarrow (e\nu b)\ (\tau\nu\bar{b})$	2/81
$t\bar{t} \rightarrow (\mu\nu b)\ (\tau\nu\bar{b})$	2/81
$t\bar{t} \rightarrow (e\nu b)\ (e\nu\bar{b})$	2/81
$t\bar{t} \rightarrow (e\nu b)\ (e\nu\bar{b})$	1/81
$t\bar{t} \rightarrow (\mu\nu b)\ (\mu\nu\bar{b})$	1/81
$t\bar{t} \rightarrow (\tau\nu b)\ (\tau\nu\bar{b})$	1/81

Tabelle 2.3: *Zerfallskanäle eines* $t\bar{t}$*-Paares mit den in niedrigster Ordnung genäherten Verzweigungsverhältnissen, aus [13]*

Die Zerfälle des W^- aus dem $\bar{t}$-Quark verlaufen analog. Mit diesen Überlegungen und Ansätzen aus der Theorie lassen sich die in Tabelle 2.3 angegebenen Verzweigungsverhältnisse berechnen.
Das aus dem Top-Quark entstandene b-Quark zerfällt selbst wieder weiter. Da das CKM-Matrixelement $V_{bc} \approx 1$ ist, geht es hauptsächlich in ein c-Quark über. Dieser Prozeß wird in Abbildung 2.2 als Hadronisation des b-Quarks bezeichnet.
Insgesamt dominieren die hadronischen Zerfallskanäle gegenüber den leptonischen (siehe Tabelle 2.3). Eine wichtige Möglichkeit Top-Signale zu extrahieren, ist daher die Möglichkeit b-Quarks zu identifizieren. Am ATLAS-Experiment werden Siliziumpixellagen in der Nähe der Strahlachse verwendet, um b-Quarks genau über sekundäre Vertices zu erkennen [17]. Die vorliegende Arbeit beschäftigt sich mit Möglichkeiten sekundäre Vertices, die von Top-Quark-Zerfällen ausgehen zu erkennen. Nähere Erläuterungen dazu werden ab Kapitel 5 gemacht.

2.3 Das Higgs-Boson

Der Large-Electron-Positron Beschleuniger (LEP) liefert als Untergrenze für die Higgsmasse $m_H > 63$ GeV/c^2 [18] und LEP2 wird diese Grenze auf $m_H > 80$ GeV/c^2 aufstocken [19], falls das Higgsboson nicht schon dort entdeckt wird. Die Experimente am LHC werden das Higgsboson in einem Massenbereich von 80 GeV/c^2 $< m_H <$ 1 TeV/c^2 suchen.

2.3.1 Produktion und Zerfall des Higgsbosons

Nach theoretischen Berechnungen mit der Lagrangedichte des Standardmodells koppelt das Higgs-Boson an die W- und Z-Bosonen, an alle Fermionen und an sich selbst. Als notwendige Folge der Massenerzeugung über spontane Symmetriebrechung ist die Kopplungsstärke des Higgs-Bosons proportional den Massen beziehungsweise den Massenquadraten. Daher koppelt es nur mit sehr geringer Wahrscheinlichkeit an leichte Teilchen. Es müssen folglich schwere reelle oder virtuelle Zustände angeregt werden, um ein Higgsboson zu erzeugen. In Abbildung 2.3 sind die Produktionsmechanismen des Higgsbosons dargestellt. Die dominierenden Mechanismen sind die Gluon-Gluon-Fusion und die WW- beziehungsweise ZZ-Fusion. Die Prozesse $t\bar{t}$-Fusion und W- beziehungsweise Z-Bremsstrahlung sind von untergeordneter Bedeutung.

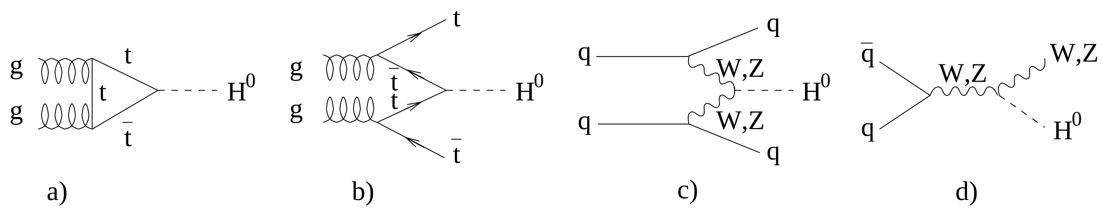


Abbildung 2.3: *Die einzelnen Produktionsmechanismen für ein SM Higgsboson. a, Gluon-Gluon-Fusion. b,* $t\bar{t}$*-Fusion. c, WW- und ZZ-Fusion. d, W- und Z-Bremsstrahlung.*

Bei voller Luminosität erwartet man Produktionsraten von 10^4 - 10^6 Ereignissen pro Jahr [20].
Falls sie energetisch erlaubt sind, werden für das Higgsboson folgende Zerfälle erwartet:

$$\mathrm{H} \rightarrow \mathrm{W}^+\mathrm{W}^- \qquad (\sim \mathrm{m}_\mathrm{W}^2) \tag{2.13}$$

$$\mathrm{H} \rightarrow \mathrm{Z}^0\mathrm{Z}^0 \qquad (\sim \mathrm{m}_\mathrm{Z}^2) \tag{2.14}$$

$$\mathrm{H} \rightarrow \mathrm{f}\bar{\mathrm{f}} \qquad (\sim \mathrm{m}_\mathrm{f}) \tag{2.15}$$

Mit f werden die fundamentalen Fermionen bezeichnet. Das Higgs-Boson sollte aufgrund der Proportionalität der Kopplungsstärke zur Masse vorwiegend in das jeweils schwerste Teilchen zerfallen, in das es aus energetischen Gründen zerfallen kann. Abbildung 2.4 zeigt die Verzweigungsverhältnisse eines Higgsboson-Zerfalls im Massenbereich $60\ \mathrm{GeV/c^2} < m_H < 600\ \mathrm{GeV/c^2}$ [4, 21].
Für eine Masse des Higgs-Bosons unterhalb der Grenze für Zerfälle in ein Paar von Vektorbosonen ($\mathrm{m_H} < 2\mathrm{m_W}, 2\mathrm{m_Z}$) beträgt das Verzweigungsverhältnis für $\mathrm{H} \rightarrow \mathrm{b}\bar{\mathrm{b}}$ im wesentlichen 100 %, da das b-Quark das schwerste erreichbare Teilchen darstellt. Ein weiterer Zerfall in diesem Massenbereich ist $\mathrm{H} \rightarrow \gamma\gamma$. Auch in diesem Kanal erhofft man sich Higgs-Signale zu sehen.
Ist $\mathrm{m_H} > 2\mathrm{m_W}$ beziehungsweise $\mathrm{m_H} > 2\mathrm{m_Z}$, so dominieren Zerfälle des Higgs-Bosons

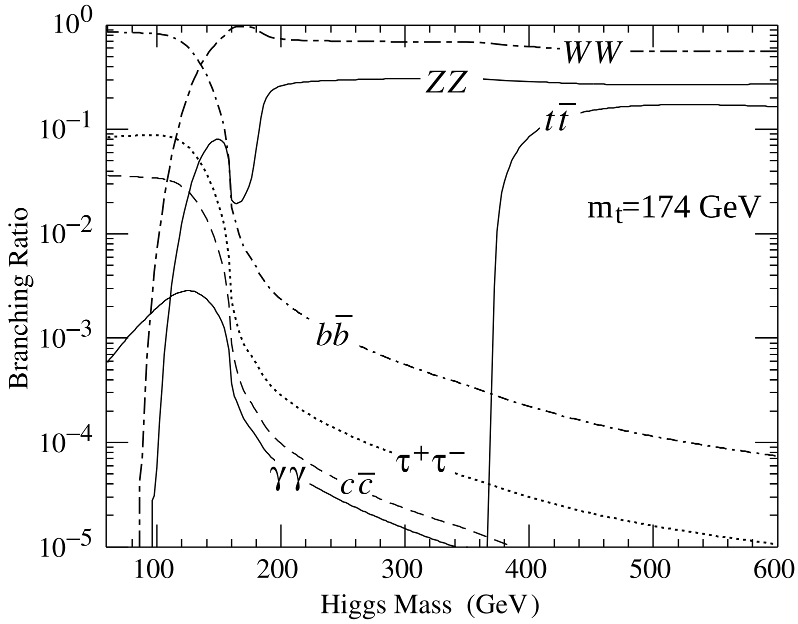


Abbildung 2.4: *Die Verzweigungsverhältnisse für ein Higgsboson im Massenbereich 60* GeV/c^2 $< m_H <$ *600* GeV/c^2

in Z- und W-Bosonen. Es ist sogar möglich, daß das Higgs-Boson schon im Bereich 120 GeV/c^2 $< m_H <$ 180 GeV/c^2 in zwei Z-Bosonen übergeht. Dann ist jedoch zu beachten, daß eines der beiden Z-Bosonen virtuell ist. In Z- und W-Zerfallskanälen untersucht man besonders den leptonischen Sektor, also den jeweiligen Zerfall der Eichbosonen in vier Leptonen und den vier dazugehörigen Neutrinos.
Übersteigt die Higgsmasse sogar die doppelte Top-Quark-Masse ($m_H >$ 360 GeV/c^2), so kann ein $t\bar{t}$-Paar entstehen. Der Zerfall des $t\bar{t}$ wurde bereits im Abschnitt 2.2 behandelt. Es entsteht ein b-Jet und sekundäre Vertices treten auf.
Auch beim $b\bar{b}$-Zerfall wird man nach sekundären Vertices Ausschau halten. Wie beim kaskadenartigen Zerfall des Top-Quarks wird hier das b-Quark in ein c-Quark und ein W-Boson zerfallen.

2.3.2 Der Higgssektor im mimimal supersymmetrischen Modell (MSSM)

Unter mehreren möglichen Erweiterungen des Standardmodells hat sich insbesonde re das MSSM hervorgetan [22]. In diesem Modell wird die Existenz zweier Higgs-Doubletts gefordert, die zu fünf, zum Teil geladenen Higgsbosonen (H^+, H^-, H, h, A) führen.
Im MSSM und in der Bornnäherung sind die Higgsbosonmassen und die Kopplungen vollständig mit zwei unabhängigen Parametern bestimmt. Diese Parameter sind die Masse m_A und $\tan\beta$ (das Verhältnis der Vakuum-Erwartungswerte der zwei

Doubletts). Wichtige Zerfallskanäle sind A/H$\rightarrow \tau\tau$, h/H$\rightarrow \gamma\gamma$, $\mathrm{H} \rightarrow \mathrm{ZZ} \rightarrow 4\ell$ und t$\rightarrow$bH^+.

Kapitel 3

Das ATLAS Experiment

3.1 Der Proton-Proton Speicherring

Im Dezember 1994 wurde vom CERN-Rat der Bau des Proton-Proton Speicherrings Large Hadron Collider (siehe Abbildung 3.1) beschloßen. Der LHC soll mit einer Schwerpunktsenergie von $\sqrt{s} = 14$ TeV im Jahre 2005 in Betrieb gehen [23]. Er wird den im Moment laufenden Beschleuniger Large Electron-Positron (LEP) ablösen und im bereits vorhandenen 27 Kilometer langen Tunnel installiert werden [5].

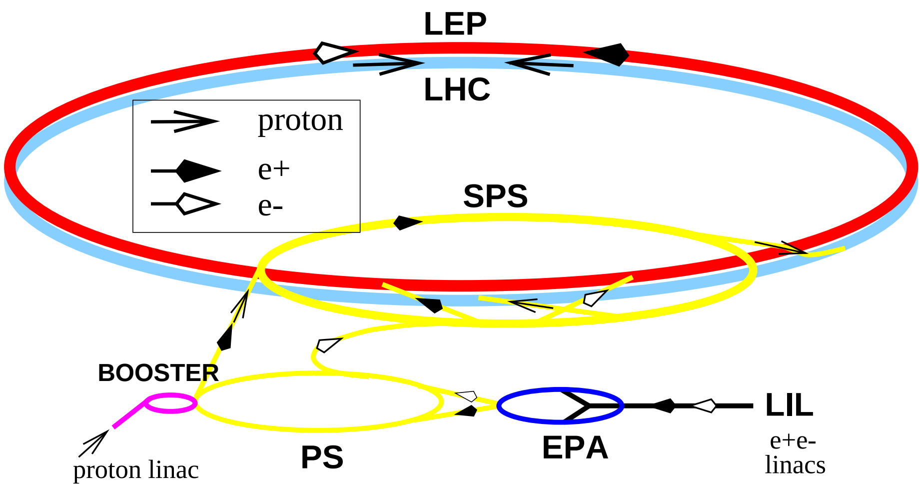


Abbildung 3.1: *Der LHC Speicherring und seine Vorbeschleuniger.*

Mit dem LHC wird es möglich sein, einen noch tieferen Einblick in die Struktur der Materie zu bekommen. Er wird Bedingungen herstellen, wie sie 10^{-12} Sekunden nach dem Urknall waren.

Es werden zwei Universalexperimente, ATLAS [6][8] und CMS [24][25] am LHC stattfinden. Das Experiment LHC-B [26], welches speziell der B-Physik gewidmet werden soll, ist in Plannung. Die LHC Maschine wird aber auch als Schwerionenbeschleuniger (Pb-Pb) verwendet. Die Kollisionen finden mit 5.4 TeV pro Nukleonpaar bei einer Luminosität von bis zu $\mathcal{L} = 10^{27}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ statt. Das zugehörige Experiment heißt ALICE [27]. In Abbildung 3.2 werden die Standorte der einzelnen Experimente am LHC dargestellt.

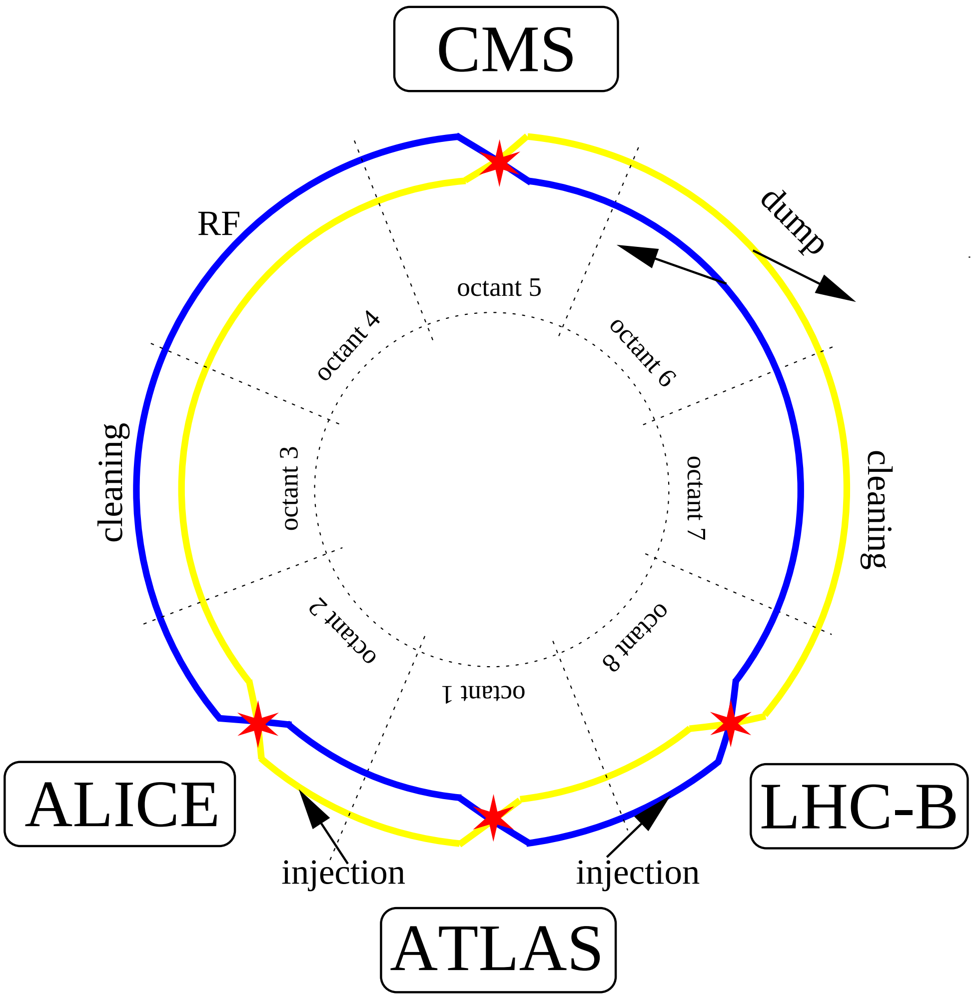


Abbildung 3.2: *Die Experimente am LHC Speicherring.*

Als Proton-Proton Speicherring wird der LHC in der Anfangsphase mit einer „geringen“ Luminosität von $\mathcal{L} = 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ in Betrieb sein, um dann für mindestens 10 Jahre mit einer höheren Luminosität von $\mathcal{L} = 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ zu laufen. Gebiete mit hoher Luminosität werden dort sein, wo sich das ATLAS- und das CMS-Experiment befinden. Kollisionen zwischen den Protonen wird es nur an den vier Geradenstücken, an denen sich Experimente befinden, geben. Die Zeit zwischen zwei

Bündelkollisionen (Bunch crossing) in der Wechselwirkungszone wird 25 ns betragen. Die bisherigen Beschleuniger (Linac, Booster, PS, SPS) werden mit einigen Änderungen zum Teilcheneinschuß für den LHC benutzt. Diese Vorbeschleuniger sind in Abbildung 3.1 zu sehen. Vom SPS und den anderen Vorbeschleunigern bekommen die Protonen eine Energie von 450 GeV und werden dann mit dieser Energie in den LHC eingespeist. Damit die beiden Protonenstrahlen jeweils auf eine Energie von 7 TeV gebracht werden können, müssen die höchstentwickelten Magnet- und Beschleunigungstechnologien benutzt werden. Dies stellt natürlich eine enorme technische Herausforderung dar. Insgesamt werden über 8000 Magnete gebraucht, von denen 1232 supraleitende Dipolmagnete mit einer jeweiligen Länge von 14.2 m sind. Das so entstehende Dipolfeld hat 8.3 T. Das magnetische Feld muß in einem superfluidem Helium unter 2 K arbeiten.
In Tabelle 3.1 sind wichtige Parameter des LHC Speicherrings zusammengefaßt dargestellt.

LHC-Parameter		
Umfang des Rings	(m)	26658.883
Protonenenergie	(TeV)	7.0
Schwerpunktsenergie	(TeV)	14.0
Dipolfeld	(T)	8.3
Luminosität	($cm^{-2}s^{-1}$)	10^{34}
Einschußenergie	(GeV)	450
Bunch crossing	(ns)	25
Teilchen pro Bündel		10^{11}
Intensität pro Strahl	(mA)	530

Tabelle 3.1: *Wichtige Parameter des LHC Speicherring*

3.2 Der ATLAS Detektor

3.2.1 Allgemeines Detektorkonzept

Der ATLAS Detektor ist ein universelles Proton-Proton Experiment, das entworfen wurde um das volle Entdeckungspotential des LHC auszuschöpfen. Die einzelnen Detektorkomponenten umschließen den nominellen Wechselwirkungspunkt zwiebelschalenförmig. Als Detektorkomponenten sind zu nennen: elektromagnetisches und hadronisches Kalorimeter, innerer Detektor und Myonenkammern. Eine Ansicht des

ATLAS Detektors ist in Abbildung 3.3 zu sehen. Das Layout des Detektors wird vorwiegend von der Konfiguration der Magneten bestimmt. Eine supraleitende Ringspule umschließt den Innenraum des Detektors, der für die Spurerkennung zuständig ist. Außerhalb dieses Solenoiden befindet sich das Kalorimeter. Dieses wird seinerseits umgeben von 8 supraleitenden Spulen ohne Eisenkern (Air Core). Ein Magnet mit Eisenkern hätte den Nachteil, daß an ihm Vielfachstreuung auftreten könnte, was zu einem Auflösungsverlust bei der Myonspurrekonstruktion führen würde. Diese Gestaltung der Magneten ermöglicht also eine hochauflösende Myonkammer.

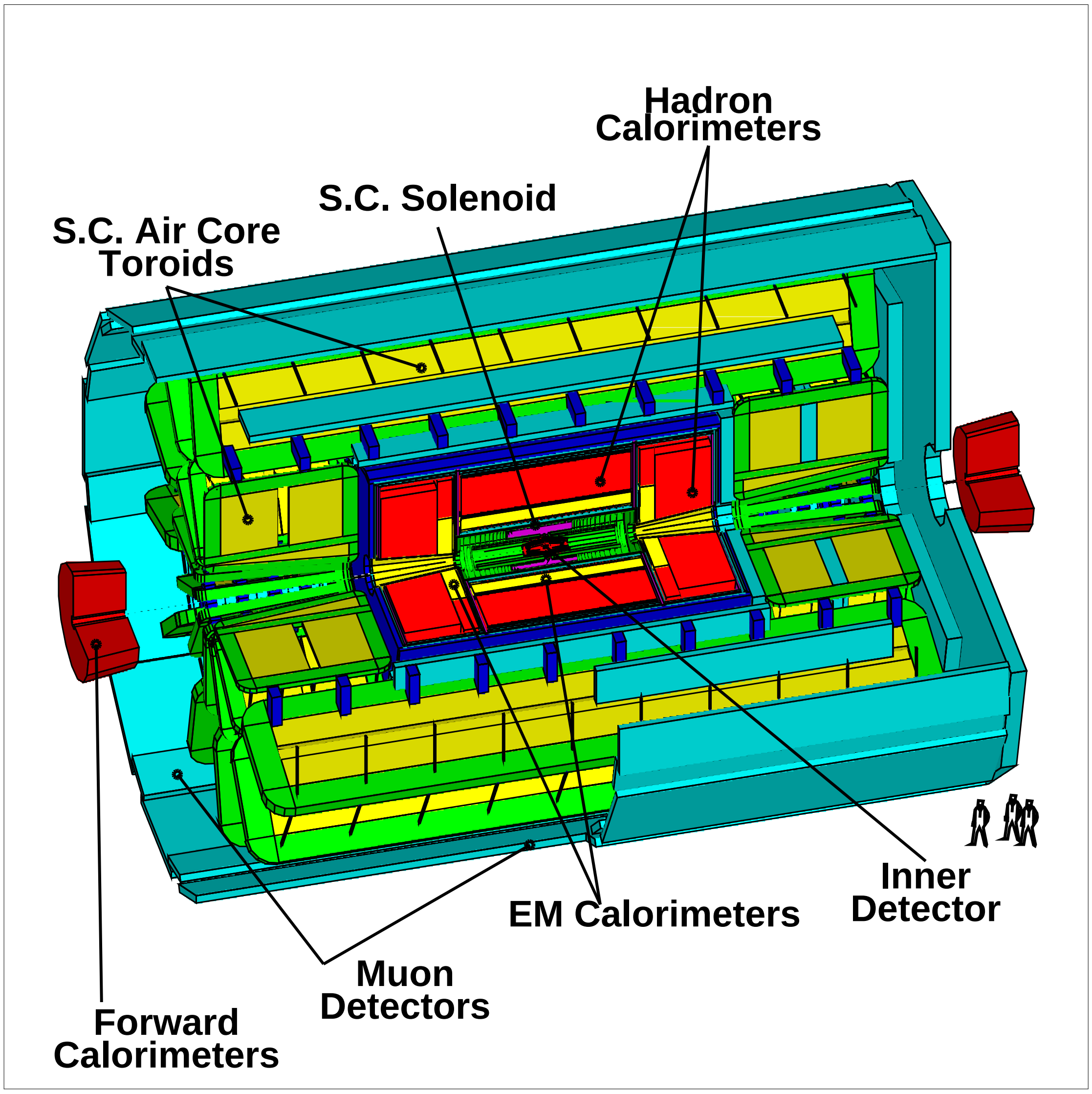


Abbildung 3.3: *Der Atlas-Detektor am LHC.*

Um ein besseres Verständnis des Detektorsystems zu bekommen, muß man vertraut sein mit dem Koordinatensystem, das verschiedene Positionen des Atlas-Detektors beschreibt. Da ATLAS eine zylindrische Geometrie aufweist, erscheinen die Zylinderkoordinaten r, ϕ und z geeignet. Anstelle von z wird die Pseudorapidität η (siehe Abbildung 3.4) eingeführt:

$$\eta = -ln(tan\frac{\theta}{2}). \tag{3.1}$$

Wenn θ von der Strahlachse aus gemessen wird, ist $\eta = 0$ bei $\theta = 90^0$ und geht für $\theta = 0^0$ gegen unendlich. Die Pseudorapidität wird speziell in der Teilchenphysik verwendet. Sie ist allerdings nur eine Näherung der Rapidität y:

$$y = \frac{1}{2}ln(\frac{E + p_z}{E - p_z}). \tag{3.2}$$

Die Größen E und p_z in Gleichung 3.2 bezeichnen die Energie eines Teilchens und dessen Impuls in z-Richtung. Der Grund für die Einführung von y liegt darin, daß Δy eine lorentzinvariante Größe ist. Dies ist wichtig, da die im Detektor auftretenden Teilchen relativistische Energien haben. Verwendet wird jedoch praktisch immer die Pseudorapidität η, da sie immer gemessen werden kann, wohingegen für die Rapidität y Energie und Impuls der Teilchen bekannt sein müssen.

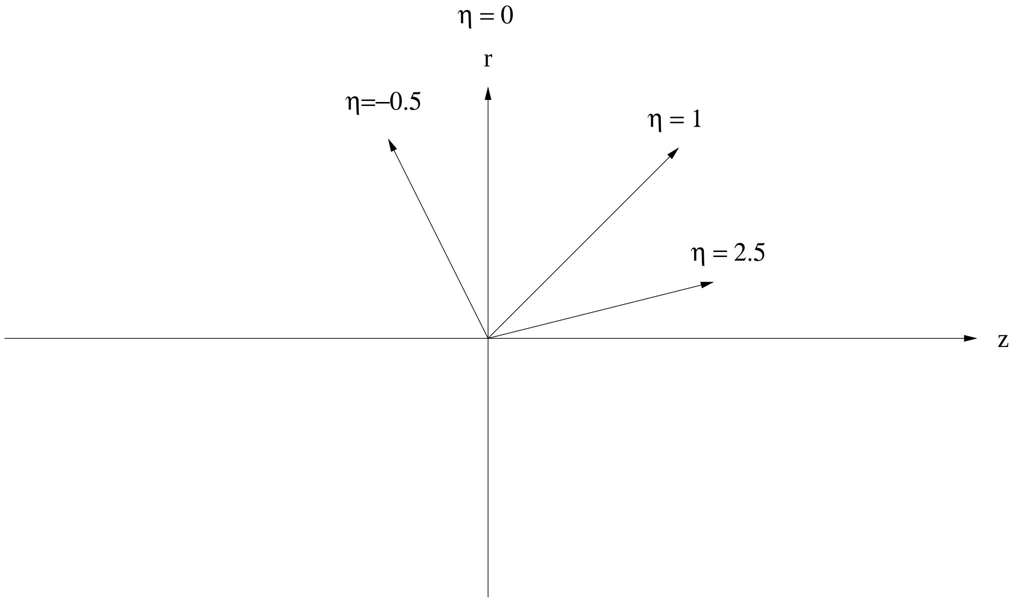


Abbildung 3.4: *Veranschaulichung der Pseudorapidität.*

Im folgenden wird eine knappe Beschreibung des Kalorimeters und der Myonenkammern gemacht. Dem inneren Detektor wird ein eigenes Kapitel gewidmet, da er für diese Arbeit von größerer Bedeutung ist.

Aufgabe des Kalorimeters ist es, die Energie und die Position von Teilchen, die das Kalorimeter durchqueren, zu bestimmen. Die Technologie für den elektromagnetischen Teil stellt ein Kalorimeter aus flüssigem Argon (LAr) mit Blei-Absorbern

dar. Das Kalorimeter zeichnet sich durch hohe Granularität aus und überstreicht den Pseudorapiditätsbereich $|\eta| < 3.2$. Für die Endkappen des hadronischen Teils wird ebenfalls die LAr-Technologie verwendet. Der Großteil des hadronischen Kalorimeters benutzt neue Szintillator-Platten, welche versetzt innerhalb eines Eisen-Absorbers plaziert sind. Das spezielle Kalorimeter im Vorwärtsbereich überstreicht einen Bereich von $3.2 < |\eta| < 4.9$. Das elektromagnetische Kalorimeter hat zylindrische Struktur mit einem Radius von 2.25 m Radius und ist in das ebenfalls zylindrische hadronische Kalorimeter mit einem Radius von 4.25 m eingebettet. Die Länge des Kalorimeters beträgt 13 m und es hat ein Gewicht von 4000 Tonnen.
Die Myonenkammern umgeben dieses Kalorimeter. Das Air-Core Toroidsystem besteht aus einem langen Barrel und zwei Endkappenmagneten. Da in diesem System, im Gegensatz zu herkömmlichen Eisenmagneten, Vielfachstreunung deutlich reduziert wird, erlangt man eine deutlich verbesserte Impulsmessung. Die Bestimmung des Myonimpulses erfolgt mit der Sagitta-Messung: Drei Positionen des Myons werden mit hochauflösenden Spurkammern bestimmt. Vervollständigt wird das Myonsystem durch schnelle Triggerkammern.
Die Ausmaße des gesamten ATLAS-Detektors werden von den Myonkammern bestimmt. Der Radius des Barrels der äußeren Kammern beträgt 11 m. Die Myonenkammern haben eine Länge von 42 m. Insgesamt hat der ATLAS-Detektor ein Gewicht von 7000 Tonnen.

3.2.2 Der innere Detektor

Der innere Detektor liegt innerhalb eines Zylinders mit einer Länge von 6.90 m und einem Radius von 1.15 m. Durch eine Ringspule wird ein magnetisches Feld von 2 T hervorgerufen. Die einzelnen Komponenten, aus denen sich der innere Detektor zusammensetzt, sind in Abbildung 3.5 dargestellt und werden weiter unten genauer erläutert.

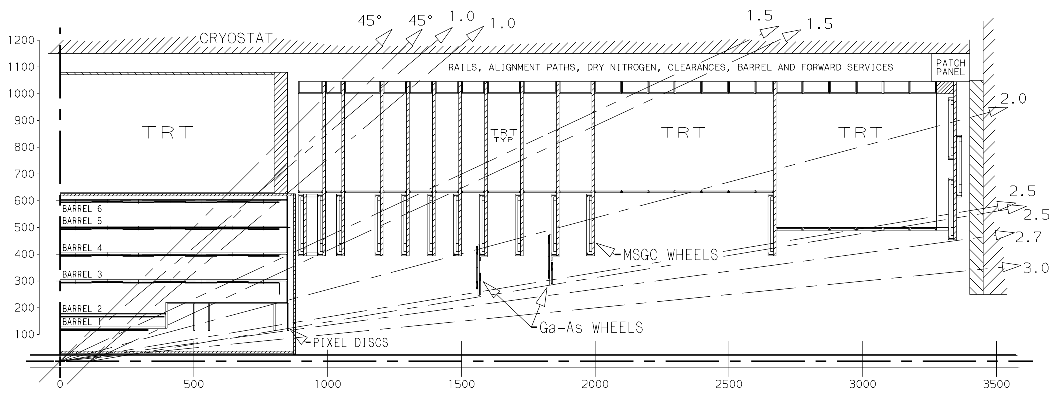


Abbildung 3.5: *Schematische Seitenansicht des inneren Detektors.*

Die Aufgaben des inneren Detektors sind Mustererkennung, Impuls- und Vertexmessung und verbesserte Elektronenidentifikation. Aufgrund der Ziele der Physik müssen eine ganze Reihe an Anforderungen an den inneren Detektor gestellt werden. Im folgenden werden die Hauptanforderungen benannt:

- gute Möglichkeiten der Spurerkennung über einen Pseudorapiditätsbereich $|\eta| < 2.5$.
- transversale Impulsauflösung von mehr als 30% bei $p_{\mathrm{T}} = 500$ GeV.
- hohe Effizienz beim Finden isolierter und assoziierter Spuren
- gute Möglichkeiten der Elektronen- und Photonenidentifikation
- hohe Effizienz beim Erkennen von b-Jets
- präzises Erkennen von sekundären Vertices
- Einsatz als Spurtrigger auf Triggerstufe 2

Die hohe Spurdichte, die am LHC erwartet wird, stellt enorme Anforderungen an die Impuls- ,Winkel- und Vertexauflösung. Man braucht also Detektorlagen mit einer sehr feinen Granularität. Dies ist natürlich mit sehr hohen Material- und Energiekosten verbunden. Auf der anderen Seite benötigt man aber auch gute Mustererkennung, die viele Messungen pro Spur fordert, um Unklarheiten möglichst zu vermeiden. Die Lösung ist eine Kombination von hochauflösenden „diskreten“ Detektoren auf inneren Radien und „kontinuierlichen“ Spurelementen auf äußeren Radien. Der innere Teil besteht aus Halbleiterdetektoren (SemiConductor Tracker, SCT). Der äußere Teil wird vom Übergangsstrahlung-Detektor (Transition Radiation Tracker, TRT) gebildet, der auch eine gute Elektronenidentifikation ermöglicht.
Der SCT hat einen Radius von 60 cm und untergliedert sich in einen Barrel-Teil, der sich entlang der Strahlachse über 160 cm ausdehnt und zwei vorwärtsgerichteten Teilen auf beiden Seiten. Je nach Stärke der Strahlung, die abhängig ist vom Abstand der Strahlachse, werden unterschiedliche Detektortechnologien verwendet. Im Barrel-Teil befinden sind auf kleineren Radien Silizium-Pixel-Detektoren und auf größeren Radien Silizium-Streifen-Detektoren. Dabei sind die Pixel-Detektoren um eine Größeneinheit widerstandsfähiger gegen Strahlungszerstörung als Streifendetektoren. Während der Anfangsphase, in der der LHC mit niedrigerer Luminosität betrieben wird, ist eine zusätzliche Vertexlage vorgesehen. Diese soll ziemlich nahe an der Strahlachse liegen. Man diskutiert entweder über einen Pixeldetektor mit einem Radius von ≈ 4 cm oder eine doppeltbeschichtete Siliziumstreifenlage mit einem Radius von ≈ 6 cm. In der Vorwärtsregion verwendet man auf kleineren Radien GaAs-Streifendetektoren, auf größeren Radien werden Mikrostreifen-Gas-Zähler (MicroStrip Gas Counters MSGCs) eingesetzt. Die MSGCs haben einen Kostenvorteil und zeichnen sich durch Einfachheit aus, die vom Gebrauch von großflächigen Glasplattensubstraten herrührt.
Der TRT überdeckt den radialen Bereich ab etwa 50 cm bis 110 cm. Er besteht aus Driftröhren mit 4 mm Durchmesser (Straw Tubes). Diese sind eingebettet in

System	Position	Auflösung σ (μm)	η-Bereich
Pixel	(1 Barrellage bei kl. Lumi.) 2 Barrellagen 4 vorwärtsorientierte Scheiben auf jeder Seite	$(R\phi = 14, z = 87)$ $R\phi = 14, z = 87$ $R = 87$	(± 2.5) ± 2.5
Silizium-streifen	(1 Barrellage bei kl. Lumi.) 4 Barrellagen	$(R\phi = 10, z = 20)$ $R\phi = 15, z = 770$	(± 2.1) ± 1.4
GaAs	2 vorwärtsorientierte Scheiben auf jeder Seite	$R\phi = 10, R = 1200$	2.0–2.5
MSGC	12 vorwärtsorientierte Scheiben und 1 Rad auf jeder Seite	$R\phi = 35, R = 1800$	1.4–2.5
TRT	Straws im Barrelbereich Straws im Vorwärtsbereich	170 (pro Straw)	± 2.5

Tabelle 3.2: *Parameter des inneren Detektors. Die beiden Möglichkeiten für eine Vertexlage sind in Klammern gesetzt.*

Polypropylenschaum und Polypropylenfolie, die Röntgenstrahlung produzieren. Die Straw Tubes sind axial orientiert, 160 cm lang und in 64 konzentrischen Radien mit einem Abstand von 6.8 mm angeordnet. In den Endkappen sind die Straw Tubes radial orientiert und haben eine Länge von etwa 40 cm bis 50 cm. Mit dieser Anordnung wird erwartet, daß Spuren mit $p_T > 1$ GeV/c mindestens 36 Straws kreuzen werden. Ein Überblick über die einzelnen Komponenten des inneren Detektors wird in Tabelle 3.2 gegeben.

Abbildung 3.6 zeigt ein gutes Beispiel für Mustererkennung im Barrelbereich des inneren Detektors. Dargestellt ist das Ereignis $\mathrm{B_d^0} \to \mathrm{J}/\psi \mathrm{K_S^0}$ bei einer Luminosität von $\mathcal{L} = 5{\cdot}10^{33}\mathrm{cm^{-2}s^{-1}}$.

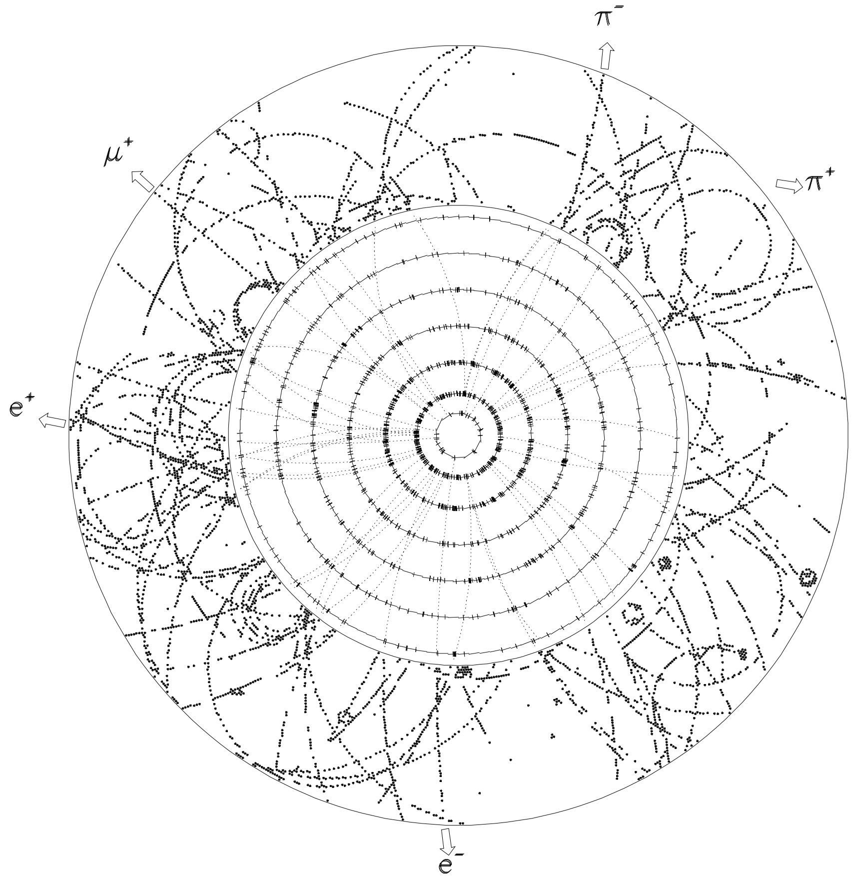


Abbildung 3.6: $B_d^0 \to J/\psi K_S^0$ *Ereignis im Barrelbereich des innerern Detektors.*

3.3 Das Trigger- und Datennahmesystem

Am LHC werden sich bei voller Luminosität alle 25 ns 18 Proton-Proton Kollisionen ereignen. Die Aufgabe des ATLAS-Trigger-Systems ist es, aus dieser Vielzahl an Kollisionen physikalisch interessante Ereignisse zu erkennen. Das Triggersystem muß die Strahldurchdringungsrate von 40 MHz (alle 25 ns) auf etwa 10 Hz reduzieren, um Aufzeichnungen auf Speichermedien überhaupt zu gewährleisten. Damit eine so enorme Datenreduktion erfolgreich ermöglicht werden kann, wird bei ATLAS ein 3-Stufen-Triggersystem (LVL1, LVL2, LVL3) eingesetzt (siehe Abbildung 3.7).

Level 1 Trigger (LVL1)

Der LVL1 nimmt Daten mit reduzierter Granularität vom Myon- und Kalorimeter-Trigger. Der Kalorimeter-Trigger versucht Cluster zu finden, die von Elektronen, Photonen oder Jets kommen. Der Myon-Trigger sucht nach Myon-Spuren mit hohem p_T. Wegen der Komplexität der Ereignisse wird auf LVL1 keine Information des Spurkammersystems benutzt. Die Daten kommen am LVL1 mit der vollen

Strahldurchdringungsrate von 40 MHz an. Während der Wartezeit (Zeit, die gebraucht wird um die Daten zu sammeln und zu verarbeiten und schließlich eine Triggerentscheidung zu fällen) von $\approx 2\mu s$ werden die Daten der Subdetektoren in Pipelines gehalten. Die maximale Ausgaberate an den LVL2 ist begrenzt auf 100 kHz. Diese Begrenzung ist gegeben durch die Möglichkeiten des Auslesesystems der Subdetektoren und durch den LVL2 selbst. Die Hauptanforderungen an den LVL1 sind also eine eindeutige Identifikation der Bündeldurchdringungen, die interessante Ereignisse enthalten und eine Minimierung der Totzeit.

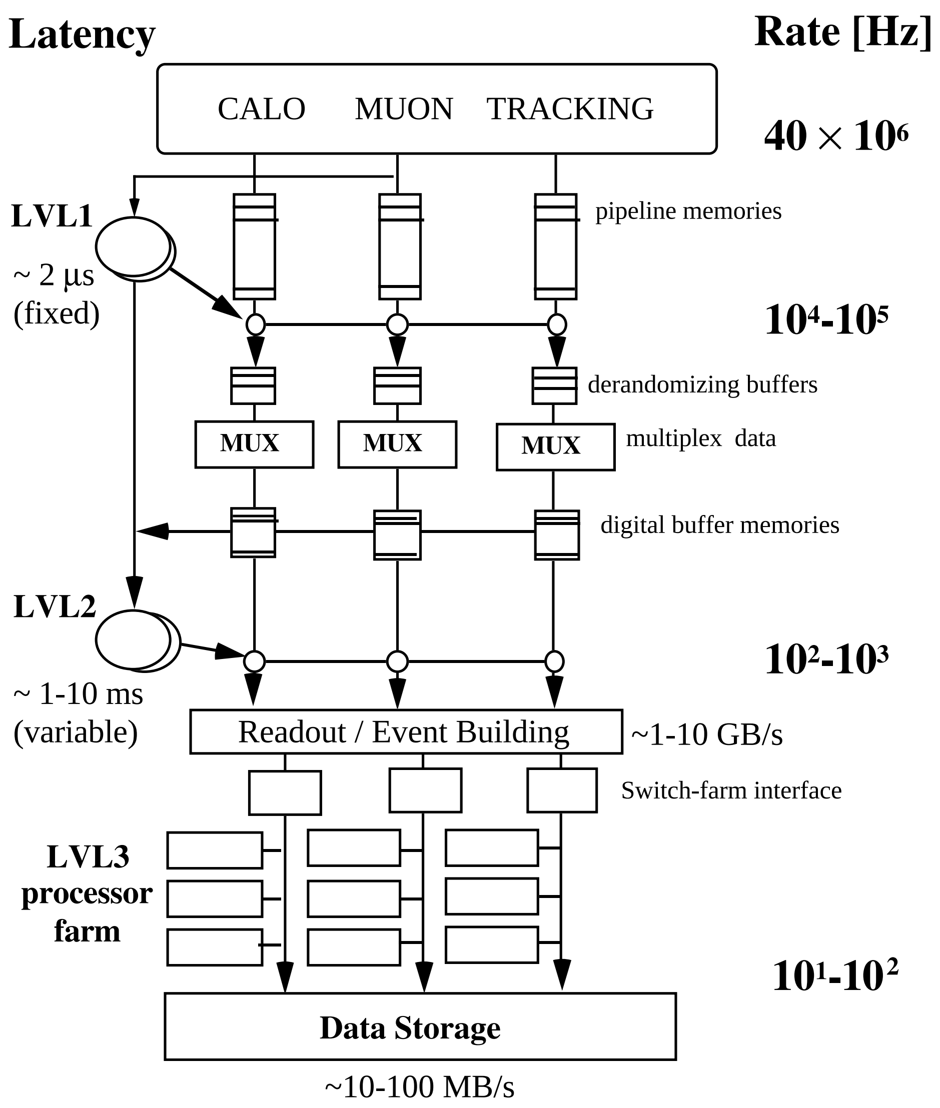


Abbildung 3.7: *Das 3-stufige Triggersytem bei ATLAS .*

Level 2 Trigger (LVL2)

Der LVL2 soll die Triggerrate von 100 kHz nach LVL1 auf $\approx$ 1 kHz reduzieren. Das angestrebte Konzept auf LVL2 ist die Benutzung von Detektorbereichen, die von LVL1 als interessant deklariert wurden, sogenannten „Regions of Interest" (RoIs). Genauer gesagt sind die RoIs Gebiete im (η, ϕ)-Raum, die vom LVL1 Kalorimeter- und Myontrigger gefunden werden und möglicherweise interessante Physik enthalten. Diese Vorgehensweise hat den entscheidenden Vorteil, daß nur ein kleiner Bruchteil der gesamten Detektordaten verarbeitet werden muß, was zu einer erheblich verminderten Bandbreite der zu transferierenden Daten und zu kleineren Anforderungen an die Rechenkapazität führt. Die Datenverarbeitung auf LVL2 besteht aus zwei Teilen. Der erste Teil ist die Extraktion wichtiger Information aus den einzelnen RoIs. Im zweiten Teil werden diese Informationen aus den unterschiedlichen RoIs und Subdetektoren kombiniert und miteinander verglichen. Danach wird eine Triggerentscheidung erfolgen. Die Wartezeiten auf LVL2 sind variabel von etwa 1 ms bis zu 10 ms. Im Gegensatz zu LVL1 steht auf der zweiten Triggerstufe die volle Datengranularität vom inneren Detektor, vom Kalorimeter, wie auch von den Myonenkammern zur Verfügung.

Level 3 Trigger (LVL3)

Nach einer positiven LVL2-Entscheidung werden die Daten des gesamten Ereignisses in den verschiedenen Auslesespeichern (Readout Buffers) gehalten. Ein Hochgeschwindigkeitsnetzwerk (Event Building Network) sorgt dafür, daß das ganze Ereignis übersetzt wird und die Daten aller Subdetektoren koordiniert werden. Die so aufbereiteten Daten werden zu einer riesigen Prozessorfarm geschickt. Dort wird die komplette Ereignisrekonstruktion mit einer Entscheidungszeit von $\approx$ 1 s durchgeführt. Die Größenordnung der Rechenleistung für derartige Anforderungen liegt bei $\approx 10^6$ MIPS. Die von LVL3 selektierten Ereignisse werden mit einer Rate von 10 - 100 Ereignisse pro Sekunde abgespeichert (maximale Ereignisrate $\approx$ 100 Hz). Daher werden voraussichtlich in einem Jahr ca. 10^6 GB Daten aufgezeichnet werden.

Kapitel 4

Neuronale Netze

4.1 Biologischer Hintergrund

Das menschliche Gehirn besteht aus 10^{10} - 10^{11} Nervenzellen, die auch als Neuronen bezeichnet werden. Ein einzelnes Neuron besteht aus einem Zellkörper (Soma) mit einer Vielzahl an kurzen faserartigen Fortsätzen (Dendriten) und einer langen Nervenfaser (Axon). Dies ist in Abbildung 4.1 dargestellt.

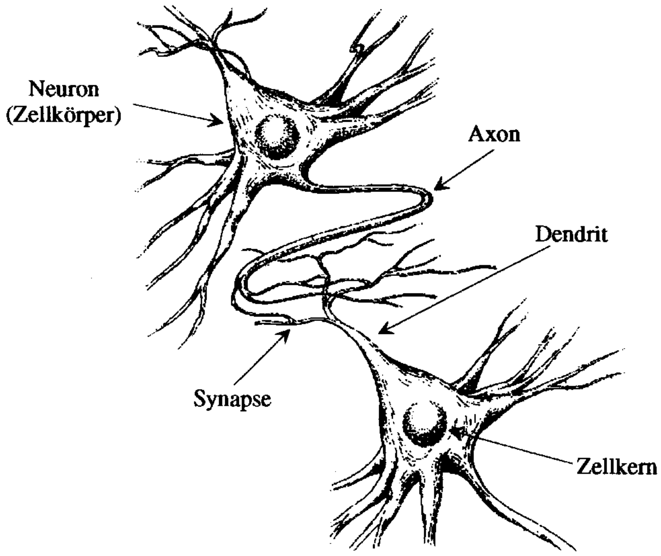


Abbildung 4.1: *Aufbau und Verbindung zweier biologischer Neuronen*

Die Neuronen tauschen untereinander Informationen mit Hilfe von elektrischen Signalen aus. Über die Dendriten nimmt das Neuron von anderen Neuronen Signale auf. Nur über das Axon wird Information weitergeleitet. Das Axon spaltet sich am Ende in über 1000 Ästchen auf. Diese Ästchen sind an ihren Enden über schmale Spalten (Synapsen) mit den Dendriten eines Empfängerneurons verbunden. Die Übertragung einer vom Axon kommenden Erregungswelle auf die Dendriten eines Empfängerneurons erfolgt nun auf chemischem Weg. In den synaptischen Spalt werden chemische Botenstoffe, sogenannte Neurotransmitter, ausgeschüttet. Diese Neurotransmitter diffundieren zu den Dendriten hinüber und können sich dort an spezifische Rezeptoren heften. Ob das kontaktierte Neuron erregt oder gehemmt wird, hängt von der Art des Rezeptors ab. Wenn nun das an diesem Neuron entstehende Signal, welches sich aus raum-zeitlich kumulierten Signalen aller Dendriten zusammensetzt, einen bestimmten zellabhängigen Schwellenwert überschreitet, so wird ein sogenanntes Aktionspotential über das Axon ausgesandt. Man spricht vom „Feuern“ des Neurons. Für eine genauere Beschreibung der Neurologie des Gehirns sei auf [28][29][30] verwiesen.

4.2 Das künstliche Neuron

Die im vorangegangenen Abschnitt eingeführte Nervenzelle wird nun abstrahiert und als mathematisches Neuron (siehe Abbildung 4.2) für neuronale Netze verwendet:

- Das Neuron stellt ein binäres Schaltelement dar, das entweder aktiv oder inaktiv ist und einen festen Schwellenwert Θ_j besitzt.
- Die Eingabeleitungen sind mit verschieden großen Gewichten ω_{ij} versehen, was in der Biologie der Art und Stärke der synaptischen Verbindung entspricht.
- Die Ausgabeleitung o_j ist analog dem biologischen Axon.

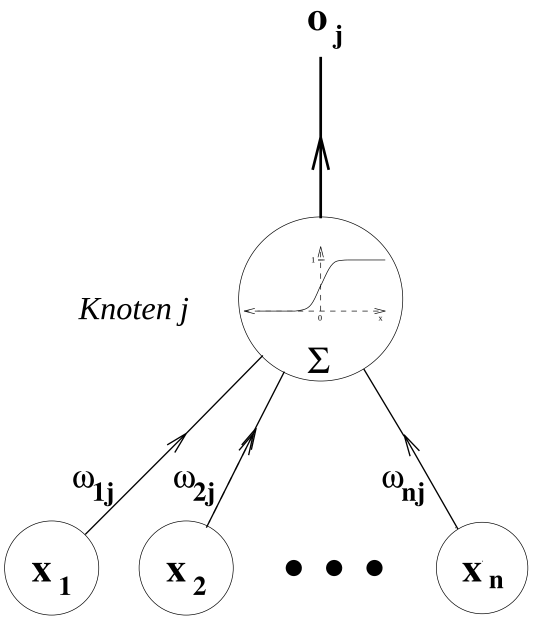


Abbildung 4.2: *Mathematisches Modell eines Neurons*

Die wesentliche Grundidee dieses mathematischen Modellneurons geht auf W. Pitts und W.S. McCulloch [31] zurück. Es wird im weiteren als Knoten bezeichnet. Ein

Knoten j nimmt n Signale x_i auf, die jeweils mit verschiedenen Gewichten ω_{ij} versehen sind. Diese werden nun aufaddiert und mit dem Schwellenwert Θ_j des Knotens verglichen. Mathematisch läßt sich das wie folgt beschreiben:

$$o_j = g\left(\sum_{i=1}^{N} \omega_{ij} x_i - \Theta_j\right). \tag{4.1}$$

Die in Gleichung 4.1 eingeführte Funktion g wird Aktivierungsfunktion genannt. Damit wird die Ausgabe o_j des Knotens berechnet. Die von Pitts und McCulloch verwendete Aktivierungsfunktion war eine Heavyside'sche Sprungfunktion. Mit ihr wurde man der Binarität gerecht.

$$g(a) := \begin{cases} 0 & \text{für} \quad a \leq 0 \\ 1 & \text{für} \quad a > 0. \end{cases} \tag{4.2}$$

Von der Biologie her ist es jedoch sinnvoller eine stetig differenzierbare und sigmoidale Funktion als Aktivierungsfunktion zu benutzen. Das ist beispielsweise die Fermi-Funktion:

$$g(a) = \frac{1}{1 + \exp(-a/T)} \tag{4.3}$$

Im Grenzfall $T \to 0$ geht die Fermi-Funktion aus Gleichung 4.3 in die Heavyside-Stufenfunktion aus Gleichung 4.2 über. Die Größe T kann als Temperaturkoeffizient interpretiert werden. In Abbildung 4.3 ist die Fermi-Funktion für zwei verschiedene Werte von T dargestellt.

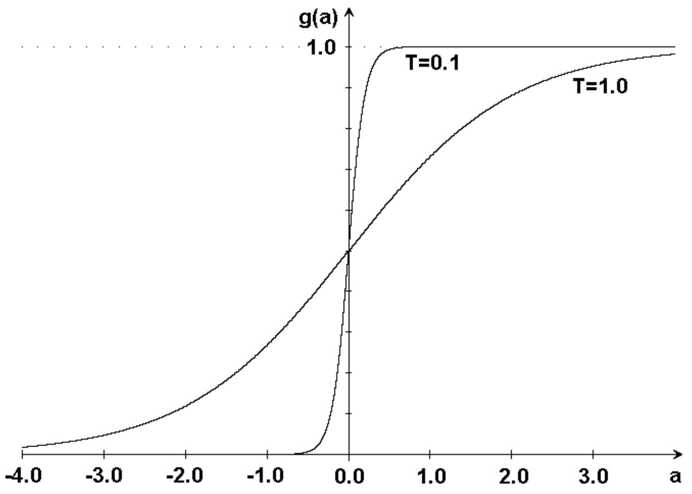


Abbildung 4.3: *Die Fermi-Funktion als sigmoidale Aktivierungsfunktion.*

Bisher wurde nur von einem einzelnen Neuron oder Knoten gesprochen. Die große Leistungsfähigkeit des menschlichen Gehirns liegt aber darin, daß es sehr viele biologische Neuronen und noch mehr Verbindungen zwischen diesen besitzt. Wenn man nun Verbindungen zwischen vielen einzelnen Knoten schafft und so praktisch ein „künstliches neuronales Netz“ (KNN) herstellt, so sollte dieses in seiner Leistungsfähigkeit die eines einzelnen Neurons übertreffen. Jedoch wird es niemals an die Leistung des Gehirns anknüpfen können. Als positive Eigenschaften und Leistungsmerkmale neuronaler Netze sind beispielweise zu nennen: Mustererkennung und Klassifikation, Lernfähigkeit, Parallelität, hohe Fehlertoleranz und Robustheit gegen Störungen und verrauschte Daten [32].

4.3 Das Feedforward-Netz

Künstliche neuronale Netze lassen sich hinsichtlich ihrer Netztopologie unterscheiden. Die Netztopologie beschreibt die Art und Weise wie die einzelnen Knoten miteinander verknüpft sind und sie hat entscheidenden Einfluß auf Funktionalität und Leistungsfähigkeit eines KNN. In dieser Arbeit werden ausschließlich „Feedforward-Netze“, die manchmal auch als Mehrschicht-Perzeptron bezeichnet werden, verwendet. In diesen Netzen sind die Knoten in einzelnen Schichten angeordnet. Jeder Knoten gibt Signale an andere Knoten weiter, von denen er selbst jedoch keine Signale empfängt. Ein solches Feedforward-Netz besteht aus einer Eingabeschicht (Input-layer), einer Ausgabeschicht (Output-Layer) und aus einer oder mehreren dazwischenliegenden Schichten (Hidden-layers). Abbildung 4.4 zeigt ein dreilagiges Feedforward-Netz.

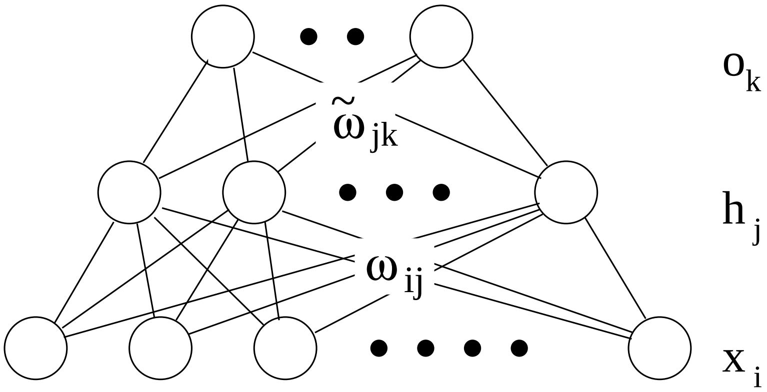


Abbildung 4.4: *Ein dreilagiges Feedforward-Netzwerk. $\vec{x}$ ist der Eingabevektor, $\vec{o}$ der Ausgabevektor und $\vec{h}$ der Ausgabevektor der versteckten Schicht.*

Der Name Feedforward-Netz rührt daher, daß nachgeschaltete Schichten ihre Information von darunterliegenden Schichten bekommen und ihre eigene Information nur an die Nachfolgeschicht weitergeben. Informationen werden also ohne Rückkopplung von einer zur nächsten Schicht in Vorwärtsrichtung durchgereicht.

4.3.1 Die versteckte Schicht

Mit einem einzigen Knoten, also einem Perzeptron [33] können nur linear seperable Probleme gelöst werden. Dies wurde von M. Minsky und S. Papert 1969 gezeigt.

Das XOR-Problem [34] ist beispielsweise mit einem Perzeptron nicht lösbar. Im folgenden wird gezeigt, daß das XOR-Problem jedoch mit einen Feedforward-Netz mit einer versteckten Lage gelöst werden kann.
Zwei Eingabegrößen x_1 und x_2 können die Werte Null und Eins annehmen. Mit Hilfe eines neuronalen Netzes soll nun entschieden werden, ob die beiden Eingaben den gleichen Wert hatten oder ob die Eingabewerte verschieden waren. Im ersten Fall soll als Ausgabewert eine Null, im zweiten Fall eine Eins produziert werden. Man erkennt in Abbildung 4.5, daß mit einer einzigen Geraden der zweidimensionale Raum nicht nach den beiden Klassen getrennt werden kann. Man benötigt also eine zweite Gerade. Diese zweite Gerade hat zur Folge, daß man eine versteckte Schicht mit zwei Knoten benötigt. Der zweidimensionale Raum wird dadurch in drei Gebiete aufgegliedert. Die beiden Knoten in der versteckten Schicht liefern also als Ausgabe die Möglichkeiten 10, 00 oder 01. Die Eingaben (0,0) und (1,1) liegen somit in der gleichen Region, nämlich der Region 00. Der Ausgabeknoten hat nur noch die Aufgabe die Bezeichnung der Regionen zu dekodieren.
Man sieht, daß mit Hilfe einer versteckten Schicht das XOR-Problem lösbar wurde.

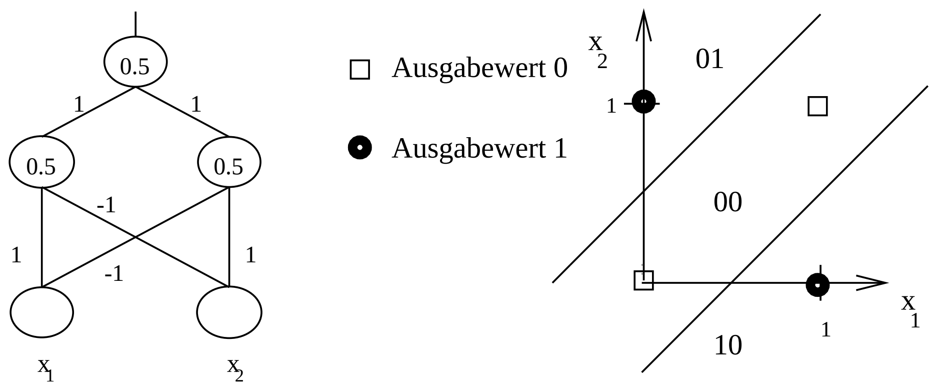


Abbildung 4.5: *Links: Ein dreilagiges Netz für das XOR-Problem. In den Knoten sind die Schwellenwerte und an den Verbindungslinien die Gewichte eingetragen. Rechts: Geometrische Darstellung des XOR-Problems im zweidimensionalen Raum. Der Raum wird durch die versteckte Schicht in drei Bereiche aufgeteilt. Das Ausgabeneuron dekodiert die Bezeichnung der Bereiche.*

Aus der vorangegangenen Diskussion wurde klar, daß die Verwendung einer versteckten Lage wirkliche Vorteile mit sich bringt. Nun stellt sich die Frage, wieviele versteckte Lagen sinnvoll sind. Mit Hilfe eines theoretisches Theorems von Kolmogorov [35] kann gezeigt werden, daß unter nur geringen Einschränkungen jede beliebige stetige Funktion $R^n \rightarrow R^m$ mit nur einer versteckten Schicht approximiert werden kann [36, 37]. Das Netz hat dabei n Eingabeknoten, $2n+1$ Knoten in der versteckten Schicht und m Ausgabeknoten. Da aber beispielsweise Einschränkungen an Aktivierungsfunktionen bestehen und noch kein konstruktives Verfahren zur Bestimmung der Aktivierungsfunktionen existiert, ist der praktische Wert dieses Theorems von äußerst geringer Bedeutung.

4.4 Der Backpropagation-Algorithmus

Dieses Kapitel beschäftigt sich mit einem in der Praxis häufig verwendeten Algorithmus, dem Backpropagation-Lernalgorithmus (BPG). Der Algorithmus wurde 1974 von Werbos [38] entwickelt. Er geriet dann in Vergessenheit, wurde jedoch mehrfach „wiederentdeckt“, zuletzt 1986 von McClelland und Rumelhart [39].
Mit einem Backpropagation-Netz möchte man bestimmte Eingabewerte auf bestimmte Ausgabewerte abbilden. Demzufolge können Backpropagation-Netze als Funktionen angesehen werden. Das Lernverhalten des Backpropagation-Algorithmus stellt einen Optimierungsprozeß dar, nämlich diejenige Funktion zu finden, welche die Eingabedaten bestmöglich auf die Ausgabedaten abbildet. Die vom Netz berechnete Funktion ist abhängig von den Netzgewichten. Ziel ist es daher, die optimalen Netzgewichte zu finden. Die Eingabedaten $\vec{x_p}$ bilden mit den jeweils dazugehörigen Zielwerten (Targetvektoren) $\vec{t_p}$ ein Lernmuster. Diese Lernmuster, auch Trainingsdaten genannt, werden dem neuronalen Netz sequentiell angeboten, was zum Begriff der Lernepochen führt. Das Netz wird aus den Eingabevektoren $\vec{x_p}$ Ausgabevektoren $\vec{o_p}$ berechnen. Der quadratische Fehler des Netzes läßt sich dann über eine Fehlerfunktion E berechnen. Sie ist gegeben durch

$$E(W) = \frac{1}{2} \sum_{p=1}^{N} |\vec{t}(\vec{x}_p) - \vec{o}(\vec{x}_p, W)|^2. \tag{4.4}$$

In Gleichung 4.4 wird die Gesamtheit der Gewichte mit W bezeichnet und N steht für die Anzahl der Musterpaare. Der Backpropagation-Algorithmus stellt ein Gradientenabstiegsverfahren dar. Er versucht die Fehlerfunktion E zu minimieren. Ziel ist es das globale Minimum von E zu finden. Zu Beginn des Netzwerktrainings werden die Gewichte initialisiert. Danach erfolgt die Berechnung der Ausgabewerte der versteckten Schicht und der Eingabeschicht mit folgenden Gleichungen:

$$h_j = g\left(\sum_{i=0}^{n} \omega_{ij} x_i\right). \tag{4.5}$$

$$o_k = g\left(\sum_{j=0}^{m} \tilde{\omega}_{jk} h_j\right). \tag{4.6}$$

Die Größe g in den Gleichungen 4.5 und 4.6 bezeichnet die bereits im Abschnitt 4.2 eingeführte Fermifunktion. Die Variable m stellt die Anzahl der Knoten der versteckten Schicht und die Variable n die Anzahl der Eingabeknoten dar. Die Schwellenwerte der Knoten sind schon indirekt in diesen Formeln über x_0 = -1 und h_0 = -1 verankert.
Der eigentliche Algorithmus beginnt damit, daß die Gewichtsvektoren vor der Ausgabeschicht und dann die Gewichtsvektoren vor der versteckten Schicht korrigiert

werden. Die Änderung der Gewichte erfolgt dabei in Richtung des steilsten Abstiegs auf der Hyperebene der Fehlerfunktion.

$$\Delta\tilde{\omega}_{jk} = -\eta \frac{\partial E}{\partial \tilde{\omega}_{jk}} = -\eta \delta_k h_j \tag{4.7}$$

$$\text{mit dem Fehlersignal } \delta_k = (o_k - t_k) \cdot g'(\widetilde{\vec{\omega}_k} \cdot \vec{h}). \tag{4.8}$$

Das Fehlersignal wird nun Schicht für Schicht bis zur Eingabelage zurückvermittelt, wodurch der Algorithmus seinen Namen bekam. Analog zu Gleichung 4.7 ergibt sich für die Gewichtsänderungen zwischen der Eingabeschicht und der versteckten Schicht:

$$\Delta\omega_{ij} = -\eta \frac{\partial E}{\partial \omega_{ij}} = -\eta \delta_j x_i \tag{4.9}$$

$$\text{mit dem Fehlersignal } \delta_j = \sum_{k=1}^{m} \tilde{\omega}_{jk} \delta_k g'(\vec{\omega}_j \cdot \vec{x}). \tag{4.10}$$

Der in den Gleichungen 4.7 und 4.9 eingeführte Proportionalitätsfaktor η gibt die Größe der Gewichtsänderungen an. Er heißt Lernparameter und kann zwischen $0 \leq \eta \leq 1$ variiert werden.

4.4.1 Probleme und Modifikationen des Backpropagation-Algorithmus

Im folgenden werden einige Probleme, die der Backpropagation-Algorithmus mit sich bringt kurz erörtert [32].

- Lokale Minima der Fehlerfläche
  Das generelle Problem bei Gradientenabstiegsverfahren ist, daß sie in einem lokalen Minimum steckenbleiben können. Je größer die Netzdimensionen sind, desto zerklüfteter wird die Fehlerfläche und somit erhöht sich die Wahrscheinlichkeit, nur ein lokales Minimum statt das globale Minimum zu finden. Allerdings kann man bei ausreichend kleinem Lernparameter oft ein gutes lokales Minimum erreichen, das für Anwendungen akzeptabel ist.

- Flache Plateaus
  Die Stärke der Gewichtsänderungen hängt vom Betrag des Gradienten ab. Wenn man sich auf einem flachen Plateau befindet, stagniert daher der Algorithmus. Die Gewichte verändern sich nicht mehr. Außerdem ist es schwierig zu erkennen, ob man sich auf einem flachen Plateau der Fehlerfläche befindet oder ob man ein lokales oder sogar das globale Minimum gefunden hat.

- Oszillationen in steilen Schluchten
  Hat die Fehlerfunktion steile Schluchten, so kann, wenn der Gradient am Rande einer Schlucht sehr groß ist, ein Sprung auf die andere Seite der Schlucht

stattfinden. Falls dort der Gradient auch sehr groß ist, wird ein Sprung zurück auf die ursprüngliche Seite erfolgen. Man sagt das Lernverhalten oszilliert.

- Verlassen guter Minima
Bei engen Tälern der Fehlerfunktion kann der Backpropagation-Algorithmus aus einem bereits guten Minimum herausspringen und dadurch vielleicht zu einem nur suboptimalen Minimum gelangen.

Um solche Probleme des Backpropagation-Algorithmus ein wenig besser in den Griff zu bekommen, kann man den Gewichtsänderungen aus den Gleichungen 4.7 und 4.9 einen Momentum-Term hinzufügen:

$$\Delta\tilde{\omega}_{jk} = -\eta\delta_k h_j + \alpha\Delta\tilde{\omega}_{jk}^{alt} \tag{4.11}$$

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

α wird als Momentumfaktor bezeichnet und liegt im Bereich $0 \leq \alpha \leq 1$. In den Momentum-Term gehen die Gewichtsänderungen des vorausgegangen Schrittes mit ein. Somit sollten speziell die Probleme der flachen Plateaus und der Oszillationen der Fehlerfunktion verringert werden.

Eine weitere Ergänzung des BPG stellt der Hidden Learning Scale β dar. Er bewirkt, daß der Lernparameter in der versteckten Schicht gegenüber dem Lernparameter in der Ausgabeschicht erhöht wird.

$$\eta_{Ausgabeschicht} = \eta \tag{4.13}$$

$$\eta_{versteckteSchicht} = \beta \cdot \eta \tag{4.14}$$

Der Grund für β liegt darin, daß Probleme im Lernverhalten auftreten können, falls das Netz nur wenige Ausgabeknoten besitzt. Wie man an Gleichung 4.10 erkennen kann, ist es möglich, daß im Falle weniger Ausgabeknoten der zur versteckten Schicht zurückpropagierte Fehler sehr gering sein kann. Das liegt daran, daß sich der Fehler an einem versteckten Knoten als gewichtete Summe der Ausgabeknoten darstellt. Im Extremfall kann es passieren, daß die versteckte Schicht kein Lernverhalten zeigt [40].

Eine weitere mögliche Modifikation des in Abschnitt 4.4 eingeführten BPG stellt das Lernintervall ℓ dar. Dort wurde als Lernintervall die Anzahl der Trainingsmuster gewählt. Die Gewichtsänderungen werden, nachdem alle Muster präsentiert worden sind, durchgeführt (batch learning). Oft verwendet wird als Lernintervall $\ell = 1.0$. Die Gewichtsänderungen erfolgen für diesen Fall nach jedem einzelnen Muster (online learning) [41].

Es gibt noch eine ganze Reihe an Abwandlungen des Backpropagation-Algorithmus, wie beispielsweise den Quickprop oder Weight Decay [32]. Es bedarf aber immer einer genauen Analyse des Problems, um eine Entscheidung für den einen oder anderen Algorithmus zu treffen.

Kapitel 5

Erzeugung von Trainings- und Testdaten für den CNAPS-Neurocomputer

5.1 Physikalische Motivation

Die in den nächsten Kapiteln folgenden Untersuchungen beschäftigen sich mit Mustererkennungsaufgaben. In Kapitel 3.2.2 wurde in Abbildung 3.6 bereits ein Beispiel für ein sehr komplexes Muster gezeigt: Die Spurpunkte von Teilchen, die den Barrelbereich des inneren ATLAS-Detektors durchqueren.
Mustererkennungsaufgaben sind typische Anwendungsbereiche neuronaler Netze. Aufgrund der hochgradigen Parallelität von neuronalen Netzen erfüllen sie zudem Zeitkriterien, die an die zweite Triggerstufe gestellt werden. Wie in Kapitel 2 erwähnt wurde, hat die Produktion von Top-Quarks oder von möglichen Higgs-Bosonen schwere Quarks in den Endzuständen zur Folge. Bei deren Zerfällen treten sekundäre Vertices auf. Durch eine Untersuchung mit PYTHIA 5.7 und JETSET 7.4 [42] zeigte sich, daß bei ca. 62 % der $t\bar{t}$-Zerfälle mindestens ein $D^{\star\pm}$-Meson entsteht. Es können pro Zerfall sogar bis zu sechs $D^{\star\pm}$-Mesonen produziert werden, wie in Abbildung 5.1 zu erkennen ist.
Diese $D^{\star\pm}$-Mesonen zerfallen dann weiter, wobei als hadronische Endzustände überwiegend Pionen und Kaonen auftreten. Ein spezieller Zerfall des $D^{\star}$-Mesons ist $D^{\star+} \to D^0\pi^+$ mit $D^0 \to K^-\pi^+$ ($D^{\star-}$-Zerfall verläuft analog, im folgenden wird die Ladungssignatur bei allen Teilchen weggelassen). Dieser Zerfall, der im weiteren untersucht werden soll, zeichnet sich durch einen sekundären Vertex aus.
Werden bereits auf Triggerebene sekundäre Vertices erkannt, so können Kandidaten für Higgs- oder Top-Zerfälle angereichert werden. Da am LHC pro Bündeldurchdringung ca. 18 Ereignisse stattfinden, ist es schwierig zu entscheiden, ob die Teilchen vom primären Vertex eines Ereignisses oder von einem sekundären Vertex eines anderen Ereignisses kommen. Somit ist es günstiger die Projektion des Detektors in die Ebene senkrecht zur Strahlachse (z-Achse) zu betrachten. Wichtig ist damit das Erkennen des transversalen Anteils der Zerfallslänge l_T. In Abbildung 5.2 sind kinematische Daten des oben beschriebenen $D^{\star}$-Zerfall aus einem $t\bar{t}$-Zerfall zu sehen.

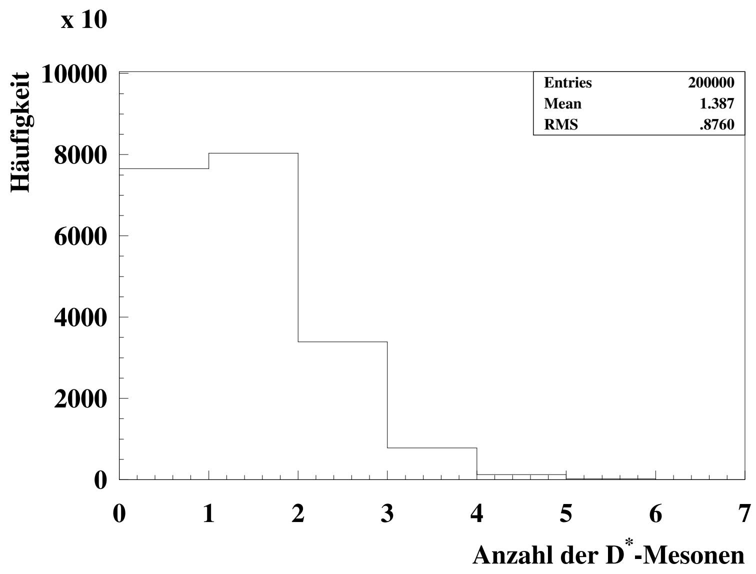


Abbildung 5.1: D*-*Ereignisse in* $\mathrm{t\bar{t}}$-*Zerfällen.*

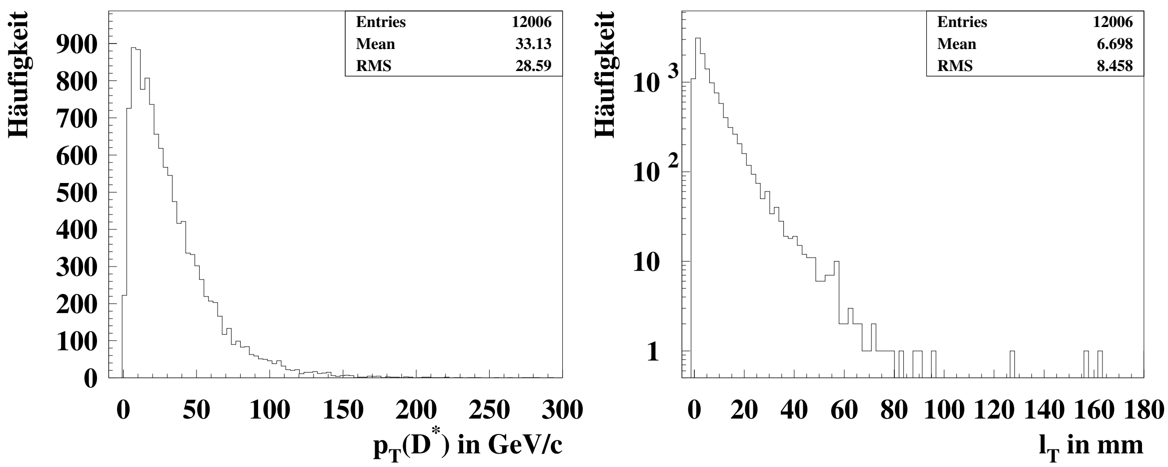


Abbildung 5.2: *Im linken Bild ist die Verteilung des transversalen Impulses der* D*-*Mesonen aus dem* $\mathrm{t\bar{t}}$-*Monte-Carlo abgebildet. Im rechten Bild ist aus selbigem Monte-Carlo der transversale Ort, an dem das* D^0-*Meson zerfällt zu sehen.*

Die Aufgabe eines neuronalen Netzes sollte also zunächst eine optimale Schätzung der transversalen Zerfallslänge aus den Spurpunkten im Detektor sein. Für einfache Studien, ob ein neuronales Netz einer solchen Anforderung genügen kann, war es zweckmäßig die Kinematik des oben aufgezeigten $D^\star$-Zerfalls und eine einfache Simulation des Barrelbereichs des inneren ATLAS-Detektors zu programmieren.

5.2 Kinematik eines $D^\star$-Zerfalls

Für erste Untersuchungen zu einem neuen Thema ist es sinnvoll, Wertebereiche und Verteilungen einiger Variablen selbst zu definieren.
Ein dazu erstelltes Programm liefert bei Eingabe des Betrags des $D^\star$-Impulses und der Abflugrichtung im Laborsystem die 4er-Impulse der Zerfallsteilchen K, π, π im Laborsystem. Wesentliche physikalische Elemente, die bei der Programmierung berücksichtigt werden mußten, waren der Energie- und Impulserhaltungssatz, die betragsmäßig gleich großen, aber entgegengesetzt gerichteten Impulse der Zerfallsteilchen im jeweiligen Schwerpunktssystem ($D^\star$- und D^0-Ruhesystem) und die Lorentztransformation der 4er-Impulse ins Laborsystem. Bei den Zerfallswinkeln der Teilchen in den Ruhesystemen wurde jeweils für $\cos(\theta)$ und ϕ eine Gleichverteilung angenommen. Eine Rückrechnung aus den 4er-Impulsen der Zerfallsteilchen K, π, π auf die invariante Masse des $D^\star$ bestätigte die Richtigkeit der Kinematik.
Für die durchgeführten Studien wurde $\phi(D^\star)$ zwischen 0^0 und 360^0 und $\theta(D^\star)$ zwischen 36^0 und 143^0 gleichverteilt gewählt. Die Grenzen für $\theta(D^\star)$ sind durch die Abmessungen des betrachteten Detektorbereichs vorgegeben. Zunächst wurde der Transversalimpuls des $D^\star$ zwischen $0\ \text{GeV/c} < p_\text{T}(D^\star) < 20\ \text{GeV/c}$ und die transversale Zerfallslänge zwischen $0\ \text{mm} < l_\text{T} < 40\ \text{mm}$ gleichverteilt erzeugt. Diese Annahmen wurden getätigt, da neuronale Netze ein besseres Lernverhalten auf einer Gleichverteilung zeigen [43]. Die beiden Verteilungen sind in Abbildung 5.3 zu sehen.

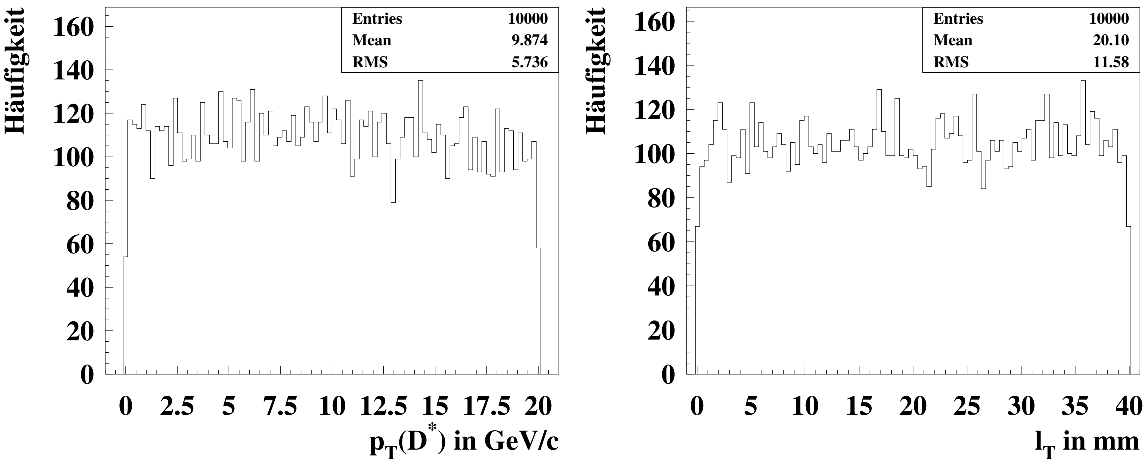


Abbildung 5.3: *Im linken Bild ist die Verteilung des transversalen Impulses des* $D^\star$ *dargestellt. Das rechte Bild zeigt die Verteilung der transversalen Zerfallslänge.*

Für alle Zerfallsteilchen, einschließlich des neutralen und damit im Detektor nicht sichtbaren D^0, ergeben sich dann die Impulsverteilungen im Laborsystem nach Abbildung 5.4. Aufgrund des hohen Massenunterschieds zwischen D^0 und π, die direkte Zerfallsprodukte des $D^\star$ sind, hat das Pion nur einen kleinen Phasenraum zur Verfügung, was sich in einem sehr niedrigen Impuls im Ruhesystem des $D^\star$ äußert. Es wird daher als *slow* Pion π_s bezeichnet.

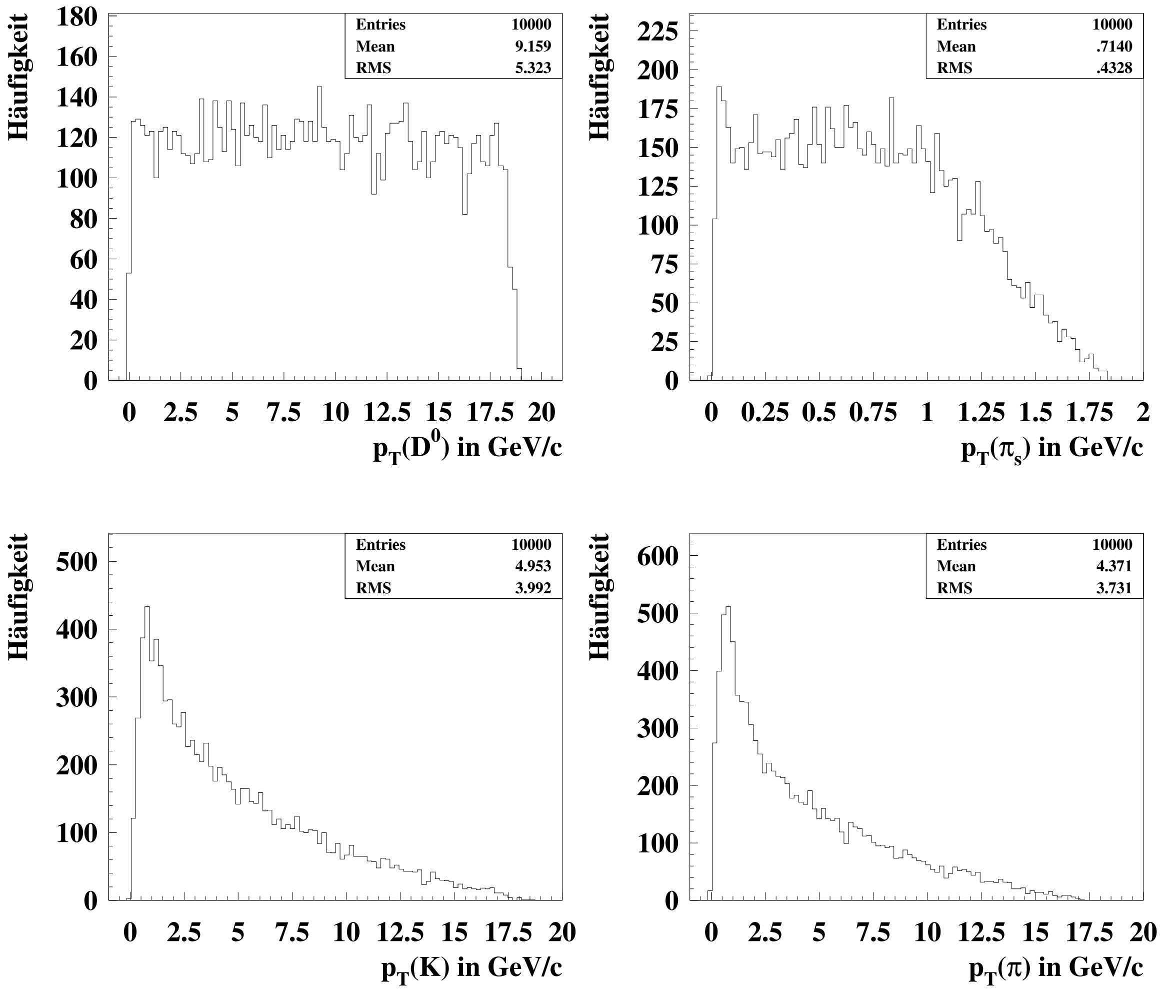


Abbildung 5.4: *Im linken oberen Bild ist die Verteilung des transversalen Impulses des* D^0, *daneben die des* π_s *dargestellt. Das linke untere Bild zeigt die Verteilung des transversalen Impulses des Kaons. Daneben ist die Verteilung für das Pion als Zerfallsteilchen des* D^0 *zu sehen.*

5.3 Die Detektorsimulation

Nachdem man die Impulse der $D^\star$-Zerfallsteilchen gewonnen hat, muß man wissen, wo diese Teilchen im Detektor Spurpunkte hinterlassen. Daher wurde eine einfache Simulation des Barrelbereichs des inneren ATLAS-Detektors programmiert. Zu berücksichtigende Größen waren dabei das Magnetfeld von 2.0 T entlang der Strahlachse (z-Achse), die Ausdehnung des Detektors entlang der z-Achse von 1.6 m und die sieben kreisförmigen Detektorlagen senkrecht zur z-Achse (transversale (r-ϕ) Ebene). Zur Übersicht sind in Tabelle 5.1 die für diese Simulation benutzten Detektorlagen mit ihrer jeweiligen Position und Auflösung $\sigma_{\mathrm{R}\phi}$ angegeben.

System	Siliziumstreifen (Vertexlage)	Pixel		Siliziumstreifen			
Position in cm	6.0	11.5	16.5	30.0	40.0	50.0	60.0
Auflösung $\sigma_{R\phi}$ in μm	10	14	14	15	15	15	15

Tabelle 5.1: *Wichtige Parameter der Detektorsimulation. Sie entsprechen den vergleichbaren Parametern des ATLAS-Detektors.*

Für Untersuchungen, inwieweit solche Detektorgrößen Einfluß auf das Lernverhalten eines neuronalen Netzes haben, können diese Größen beliebig verändert werden. Die Simulation benutzt als Eingabe die Impulsvektoren, die Massen, die Ladungen und den Entstehungsort der Zerfallsteilchen im Laborsystem.
Im folgendem wird immer der Zerfall eines positiv geladenen $D^{\star +}$ betrachtet. Damit haben die beiden Pionen positive, und das Kaon negative Ladung. Die Ladungssignatur wird jedoch im Text weggelassen.
Aus dem transversalen Impuls der einzelnen Teilchen wird jeweils der Teilchenradius, der aufgrund des Magnetfeldes hervorgerufen wird, bestimmt. Der Mittelpunkt des Teilchenradius ist abhängig von der Teilchenladung, dem Entstehungsort des jeweiligen Teilchens in der transversalen Ebene und dem Transversalimpuls. Er berechnet sich so, daß der Teilchenkreis den Nullpunkt des Koordinatensystems schneidet, falls das Teilchen am primären Vertex erzeugt wurde, beziehungsweise den Ort des sekundären Vertex in transversaler Ebene schneidet, wenn es von diesem stammt. Der Teilchenkreis wird mit dem Detektorkreis zum Schnitt gebracht und so der exakte Spurpunkt in der r-ϕ-Ebene bestimmt. Die exakte z-Koordinate ergibt sich aus der z-Komponente des Teilchenimpulses und aus der Flugzeit des Teilchens zwischen den einzelnen Detektorkreisen in der r-ϕ-Ebene. Die in Tabelle 5.1 angegebene Auflösung wird berücksichtigt, indem die einzelnen Detektorkreise in genau soviele Zellen untergliedert werden, wie es der jeweiligen Detektorauflösung entspricht. Auf dem innersten Detektorradius hat jede Zelle eine Bogenlänge von 10 μm, auf dem äußersten eine Bogenlänge von 15 μm. Diese Annahme ist gerechtfertigt, da die erste Triggerstufe bereits eine derart genaue Ortsbestimmung zulassen sollte. Die Siliziumstreifenlagen und Pixellagen werden dabei identisch behandelt.
Abbildung 5.5 zeigt die Zerfallsteilchen des $D^{\star}$-Mesons im simulierten Detektor. Die am äußersten Detektorradius angegebenen Zahlen entsprechen der Nummer einer Detektorzelle für diesen Detektorradius. Die Nummer der Detektorzelle, die ein Teilchen durchfliegt, wird aus dem Winkel des exakten Schnittpunktes des Teilchenradius mit dem Detektorradius berechnet.

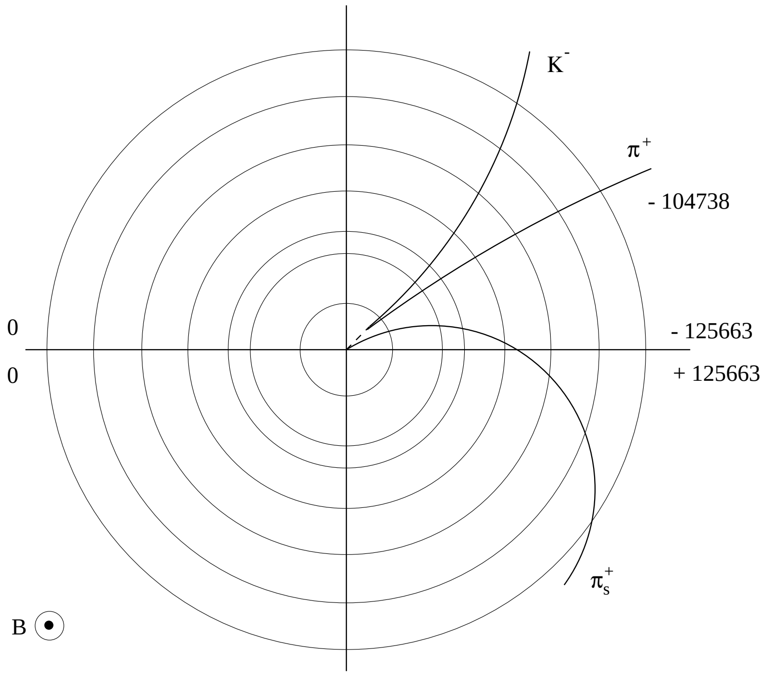


Abbildung 5.5: *Teilchenspuren im simulierten Detektor (Projektion in die r-ϕ-Ebene). Gestrichelt ist die transversale Zerfallslänge angedeutet.*

Die z-Koordinate mußte nicht mit der entsprechenden Detektorauflösung quantisiert werden, da sie nicht als Eingabegröße für das Training des neuronalen Netzes verwendet wurde. Sie diente lediglich dazu festzustellen, ob sich die betrachteten Teilchen noch innerhalb des gesamten oder zumindest eines gewissen Bereiches des inneren Detektors befinden oder nicht. Der Grund, warum die z-Koordinate nicht als Eingabegröße benutzt wurde, liegt darin, daß die Aufgabe des neuronalen Netzes die richtige Erkennung der transversalen Zerfallslänge ist. Zur Ermittlung der transversalen Zerfallslänge sind aber bereits r- und ϕ-Koordinaten der Detektordurchstoßpunkte ausreichend. Die Diskussionen in Kapitel 5 und Kapitel 6 beziehen sich also immer auf die $r - \phi$-Ebene.
Wie die genaue Netzeingabe aussieht, wird in Abschnitt 5.5 beschrieben.

5.4 Der CNAPS-Neurocomputer

Alle Netzwerktrainings wurden auf einem CNAPS-Parallelrechner durchgeführt [32, 44]. Dieser ist speziell zum Berechnen neuronaler Netze entworfen und arbeitet in Festkomma-Arithmetik. Die neuronalen Netze sind auf CNAPS/VME-Karten[1] implementiert. Der auf dem CNAPS verwendete Algorithmus war der Backpropagation-Algorithmus.
Ein CNAPS-1064 Chip (in SIMD[2]-Architektur) besteht aus 64 parallelen Prozessoren

[1] CNAPS steht für Connected Network of Adaptive ProcessorS
[2] single-instruction-multiple-data

(PNs; processor nodes). Auf den CNAPS/VME-Karten befinden sich bis zu acht dieser Chips, das heißt die Karten beinhalten 64 bis 512 PNs. Die Taktfrequenz dieser PNs beträgt 20 MHz. Die CNAPS-1064 Chips werden über den CNAPS-Sequencer Chip (CSC) gesteuert. Dieser erhält sein Programm aus dem Programmspeicher (4 MByte) und kann über einen internen Bus auf 16 MB Datenspeicher zugreifen. Die Verbindung zur Außenwelt geschieht über ein VME-Bus-Interface. Für das Laden und Schreiben der Trainings- und Testdateien ist ein Host-Rechner (eine Sun) zuständig.
Zu der Rechengeschwindigkeit des CNAPS-Rechners ist anzumerken, daß ein Multiplizier-Addier-Schritt einer Vektor-Matrixmultiplikation in einem Taktzyklus (50 ns) erfolgt. Ein Feedforward-Netz mit 14 Eingabeknoten, 32 Knoten in der ersten versteckten Schicht und 16 Knoten in der zweiten versteckten Schicht, benötigt mindestens 63 Taktzyklen. Dazu addiert sich noch ein unvermeidlicher Überhang von ca. 45 Taktzyklen. Das bedeutet eine Rechenzeit von etwa 5.4 μs.
Das Programm bpexp2 [45] wurde zur Steuerung der Trainings auf dem CNAPS-Parallelrechner verwendet. bpexp2 läuft auf einer Workstation. Es liest aus einer Steuerdatei verschiedene Netzwerkparameter, dadurch wird ein variables Training ermöglicht:

- Einstellung und Variation des „Lernparameters“, des „Hidden Learning Scale“, des „Momentumfaktors“ und des „Lernintervalls“.
- Bestimmung der Netzwerkgröße:
  Dabei kann man zwischen einer oder zwei versteckten Schichten wählen und die Anzahl der Knoten in den versteckten Schichten bestimmen. Außerdem müssen die Anzahl der Eingabe- und der Ausgabeknoten gewählt werden.
- Festlegung der Lernepochen für das Training.

Als weitere wichtige Option kann der CNAPS in zwei Softwaremodi betrieben werden:

- 8/16-Bit Modus (low precision)
  In diesem Modus werden 8 Bit Daten und 16 Bit Gewichte verwendet. Die Bestimmung der Gewichtsänderungen während des Trainings erfolgt mit einer 16 Bit Genauigkeit.
- 16/32-Bit Modus (high precision)
  Dieser Modus benutzt 16 Bit Daten und 16 Bit Gewichte. Gewichtsänderungen werden mit einer Genauigkeit von 32 Bit berechnet. Dies erweist sich für das Training als günstiger, da kleinere Gewichtsänderungen vorgenommen werden können.

5.5 Datengenerierung für das neuronale Netz

In den Netzwerktrainings wurde ausschließlich der 16/32-Bit Modus verwendet, da er immerhin 65535 verschiedene Eingabedaten zuläßt. Damit kann als RoI ein

ϕ-Bereich von maximal 93.87^0 bei einer Auflösung von $\sigma_{R\phi} = 15\ \mu$m und einem äußeren Detektorradius von 0.60 m gewählt werden. Im 8/16-Bit Modus wäre nur ein maximaler ϕ-Bereich von 0.37^0 möglich. Andererseits wäre es sowieso nicht sinnvoll einen ϕ-Bereich von 360^0 zu benutzen, da das neuronales Netz als Trigger auf LVL2 laufen soll und somit LVL1 bereits eine Datenreduktion durchgeführt hat. Das LAr-Kalorimeter kann mit seiner hohen Granularität sehr genau den Ort einer Energiedeposition feststellen. Dieser Bereich wird dann mit einem bestimmten Bereich (RoI) im inneren Detektor in Verbindung gebracht.
Als Mitte einer sehr vereinfacht angenommenen RoI wurde in den Studien die $p_{\mathrm{T}}(\mathrm{D}^\star)$-Richtung gewählt. Sie bestimmt auf jedem Detektorradius die Zellennummer Null. Die Eingaben für das neuronale Netz sind die Nummern der getroffenen Detektorzellen relativ zu dieser Null-Zelle. Die so konstruierte RoI ist in Abbildung 5.6 zu sehen. Vorerst wurden nur die beiden Spuren, welche vom sekundären Vertex ausgehen, betrachtet.

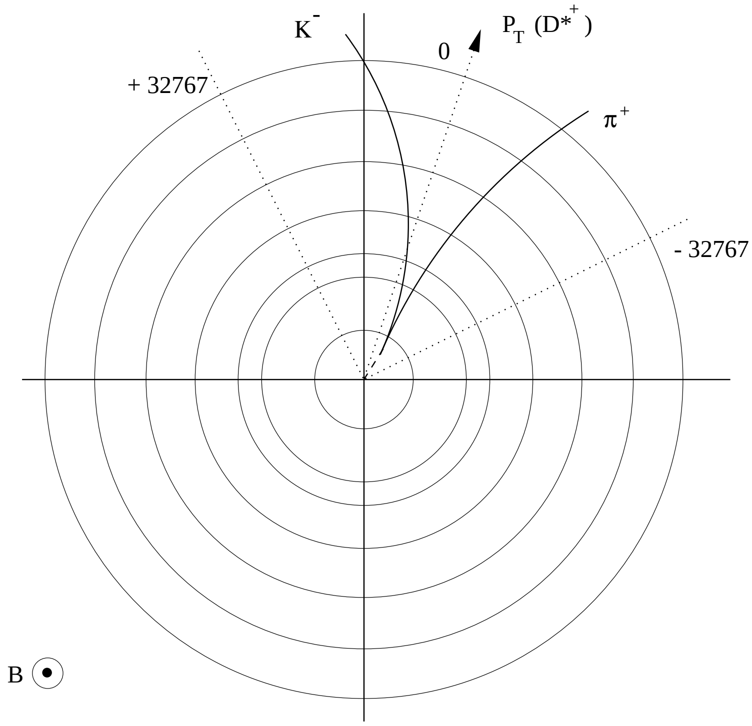


Abbildung 5.6: *Schematische Darstellung der maximal möglichen RoI bei Verwendung des CNAPS-Neurocomputers*

Da der Detektor sieben Radien besitzt, erhält man Zellennummern von maximal 14 getroffenen Detektorzellen als Netzwerkeingabegrößen. Stellt man die Forderung nach einer vollständigen Trefferzahl in einer RoI von $\phi = \pm\ 45^0$ um die $p_{\mathrm{T}}(\mathrm{D}^\star)$-Richtung, so ergibt sich eine Abweichung von der in Abbildung 5.3 gezeigten Gleichverteilung für $p_{\mathrm{T}}(\mathrm{D}^\star)$. In Abbildung 5.7 ist zu erkennen, daß kleine transversale Impulse des $\mathrm{D}^\star$ weniger häufig vorkommen als größere.
Die Erklärung liegt darin, daß kleine transversale Impulse des $\mathrm{D}^\star$-Mesons kleine

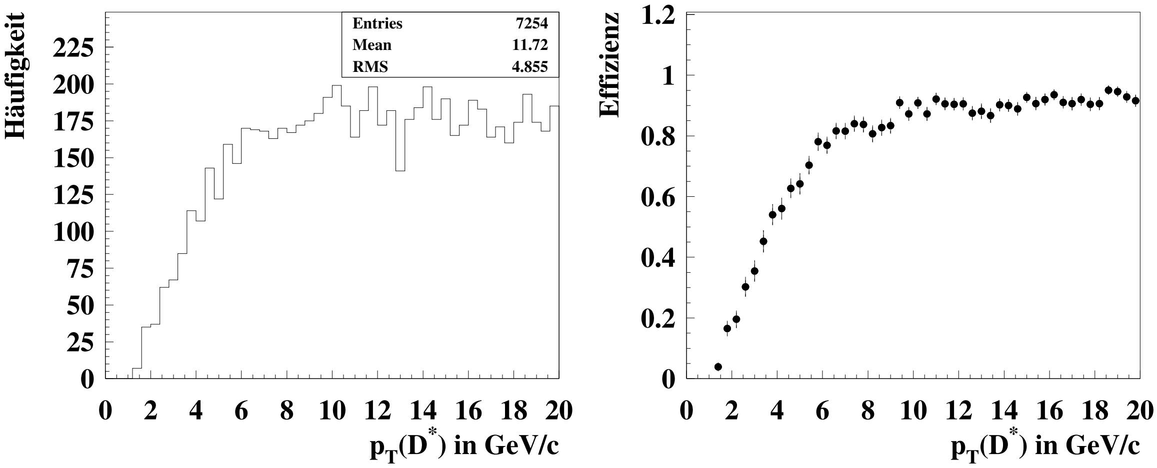


Abbildung 5.7: *Links sieht man die $p_{\mathrm{T}}(\mathrm{D}^{\star})$-Verteilung für die Netzeingabe, rechts die Effizienz mit der diese Impulse für das Netztraining ausgewählt wurden.*

transversale Impulse der Zerfallsteilchen zur Folge haben und diese Zerfallsteilchen aufgrund dessen die RoI mit unvollständiger Trefferzahl verlassen. Solche Ereignisse wurden nicht als Eingabegrößen herangezogen.
Die Art und Weise wie diese Detektorzellen ausgelesen wurden und die Information dem neuronalen Netz angeboten wurde, ist in Abbildung 5.8 dargestellt. Zu beachten ist, daß hier nur vier anstatt der sieben Detektorradien gezeigt werden.

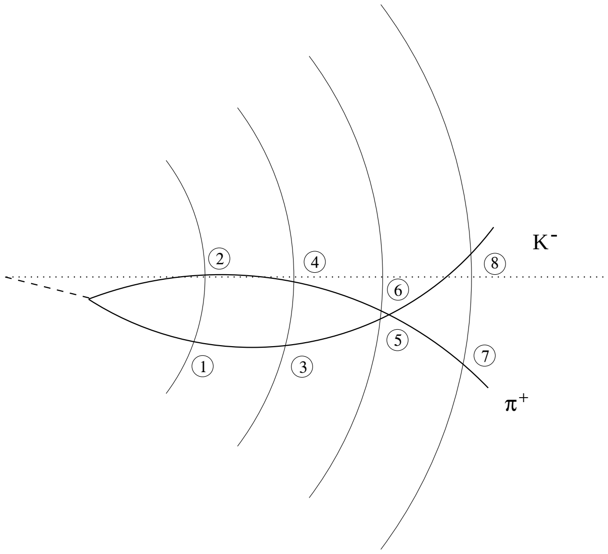


Abbildung 5.8: *Reihenfolge der Eingabe in das neuronale Netz*

Es wurde die realitätsnahe Methode der radienweisen Auslese benutzt. Der erste Knoten des neuronalen Netzes bekommt als Eingabe die kleinste von einem Teilchen getroffene Zellennummer des innersten Detektorradius. Der zweite Knoten bekommt

die nächsthöhere getroffene Zellennummer des innersten Detektorradius. Schließlich erhält der letzte Netzeingabeknoten die höchste getroffene Zellennummer des größten Detektorradius.
Eine Häufigkeitsverteilung der Detektorzellennummern, die von einem K durchdrungen werden zeigt Abbildung 5.9. Dieselbe Darstellung für das π ist in Abbildung 5.10 zu sehen. Die Verteilungen beziehen sich jeweils auf den ersten, dritten , fünften und siebten Detektorradius.

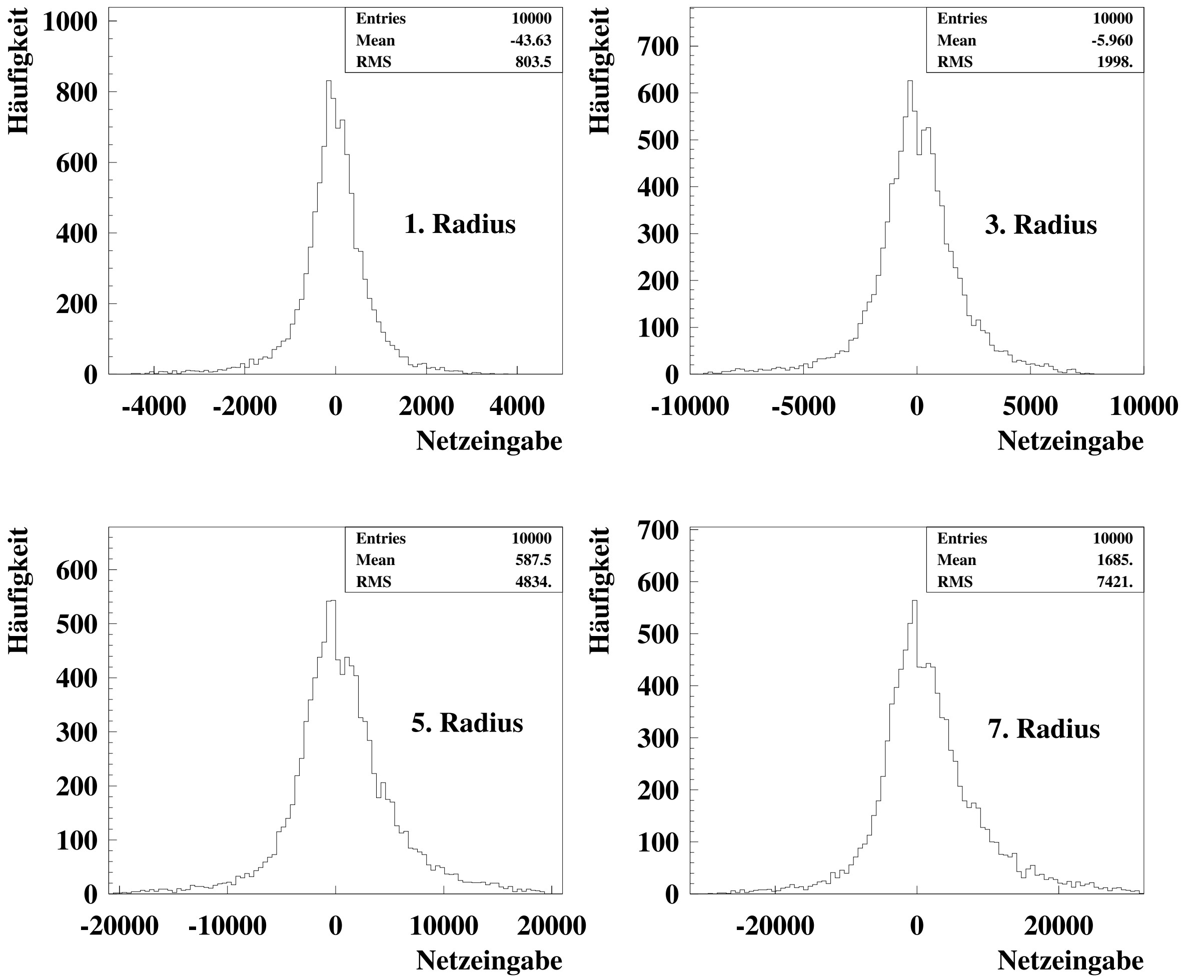


Abbildung 5.9: *Darstellung der Zellennummernverteilung der Eingabe des Kaons auf unterschiedlichen Detektorradien. Zu beachten ist der unterschiedliche Wertebereich bei den einzelnen Bildern.*

Die geforderte Netzausgabe (transversale Zerfallslänge l_T) mußte zum Netztraining auf den CNAPS-Wertebereich [-32767, 32767] transformiert werden. Sie wird dem CNAPS-BPG als Zielwert mitgegeben. In den folgenden Studien wurden mit dieser Methode Trainings- und Testdaten für das neuronale Netz erzeugt.

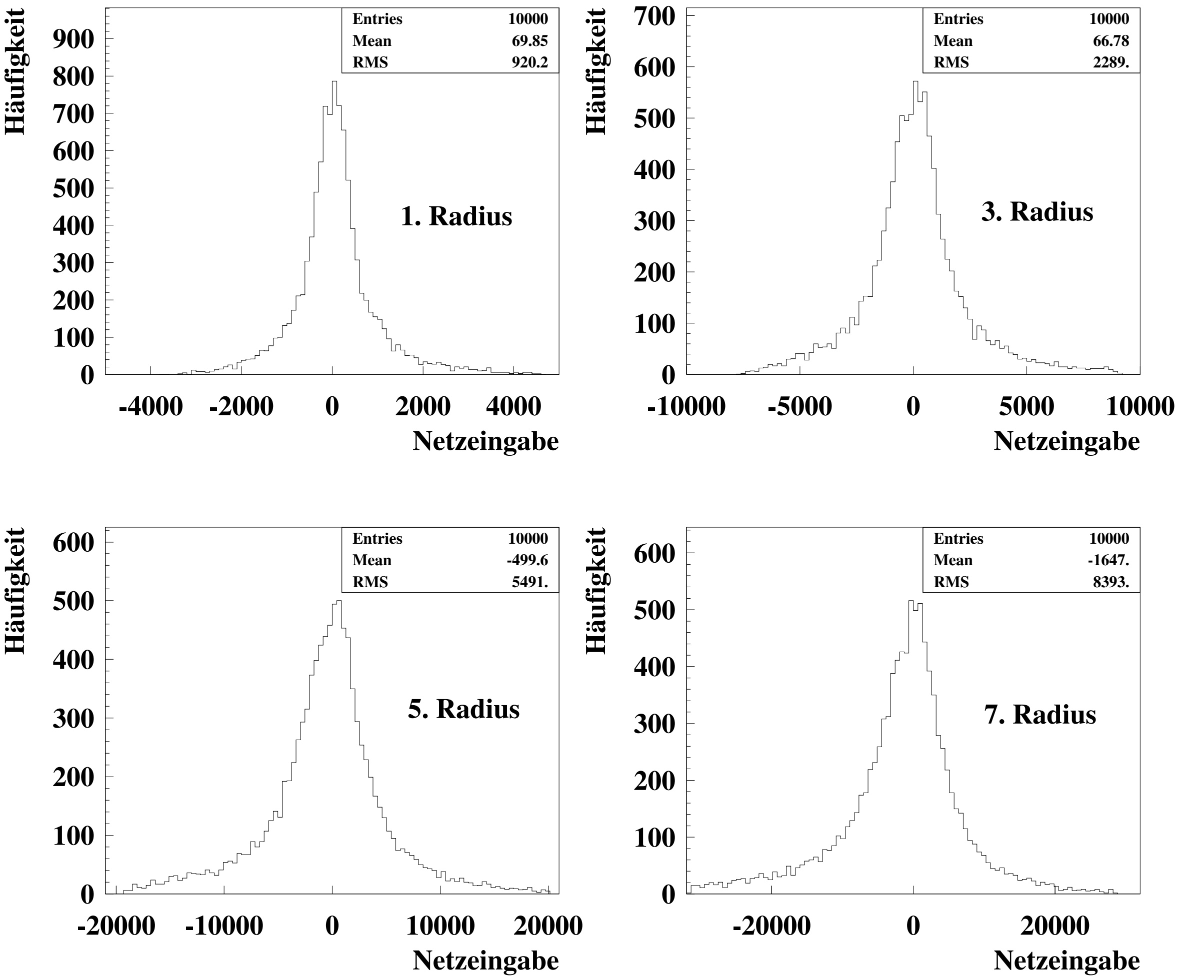


Abbildung 5.10: *Darstellung der Zellennummernverteilung der Eingabe des Pions auf unterschiedlichen Detektorradien. Zu beachten ist der unterschiedliche Wertebereich bei den einzelnen Bildern.*

Kapitel 6

Allgemeine Studien zum Erkennen transversaler Zerfallslängen

6.1 Parameterstudien

6.1.1 Lernkurven und Bewertungskriterien für Netze

Um die Lernfähigkeit eines neuronalen Netzes beurteilen zu können, wird als Fehlermaß der mittlere quadratische Fehler (mean squared error)

$$MSE = \frac{1}{N} \sum_{p=1}^{N} |t(\vec{x}_p) - o(\vec{x}_p, W)|^2 \tag{6.1}$$

gewählt.
Der Vergleichbarkeit verschiedener Netze wird mit der Normierung durch die Zahl der Trainingsmuster N Rechnung getragen. Die Zielwerte t und die Netzausgabewerte o wurden für die nachfolgenden MSE-Fehlerangaben auf das Intervall [0, 1] abgebildet. Lernkurven charakterisieren das Lernverhalten eines Netzes mit zunehmender Epochenzahl. In den Abbildungen 6.1 sind Lernkurven aus im Rahmen dieser Arbeit getätigten Untersuchungen ausgewählt. Aufgetragen ist der mittlere quadratische Fehler MSE gegen die Epochenzahl. Entstanden sind die Kurven durch Netzwerktrainings auf die transversale Zerfallslänge l_{T} mit 10000 Trainingsdaten und 10000 davon unabhängigen Testdaten über 4000 Epochen. Da der Berechnungsmodus des MSE für Trainingsdaten und Testdaten nicht direkt vergleichbar ist, wurden hier nur die Fehler auf den Testdaten untersucht.
Das linke Bild in Abbildung 6.1 zeigt ein optimales Verhalten des Netzes. Der MSE zeigt ein fallendes Verhalten mit Anwachsen der Epochenzahl. Bei hohen Epochenzahlen fällt der Fehler nur noch minimal bis nicht mehr. Im Gegensatz dazu ist im rechten Bild ein erneutes Ansteigen des MSE ab etwa Epoche 2200 zu beobachten. Dort wo der mittlere quadratische Fehler am niedrigsten ist, hat das neuronale Netz am besten gelernt zu generalisieren. Das Verhalten des Wiederanstiegs des Fehlers ab einer gewissen Epochenzahl wird „Overtraining“ genannt. Das neuronale Netz hat begonnen die Trainingsdaten auswendig zu lernen. Dies geschieht, wenn die Anzahl

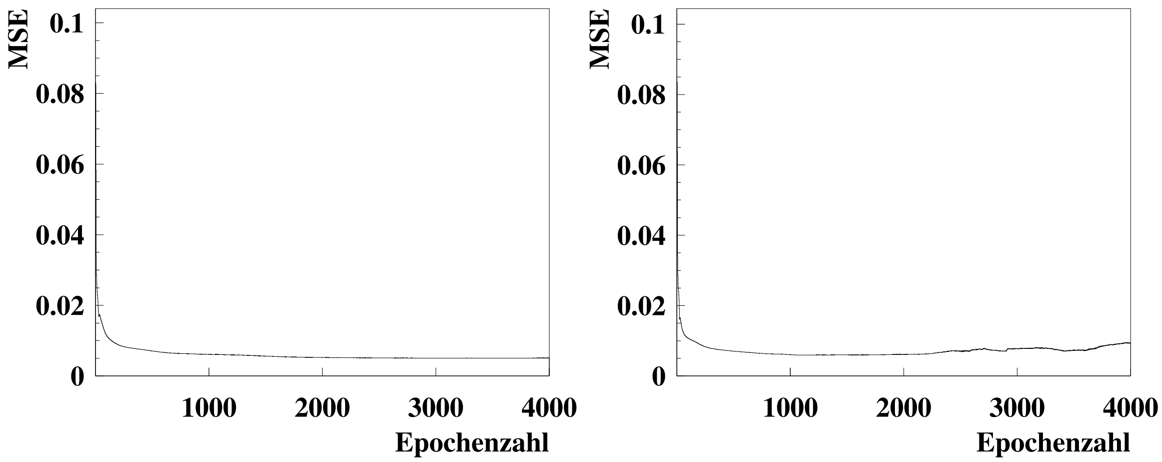


Abbildung 6.1: *Links ist der optimale Verlauf einer Lernkurve dargestellt, rechts sieht man ein leichtes Overtraining.*

der Trainingsdaten im Vergleich zur Größe des Netzwerks (Anzahl der Gewichte) zu klein gewählt wurde. In den hier gemachten Studien war ein Overtraining nahezu ausgeschlossen. Trotzdem wurde vorsichtshalber in den Parameterstudien immer der MSE zur besten Epoche benutzt.

6.1.2 Netzgröße und Lernparameter

Zunächst wurde für das vorliegende Problem der Schätzung der transversalen Zerfallslänge aus Detektortrefferpunkten eine Untersuchung zur Netzwerkarchitektur unternommen. Die folgenden Parameter (siehe Kapitel 4.4.1) wurden anfangs nicht variiert und auf folgende Werte gesetzt: $\beta = 4.0$, $\alpha = 0.0$, $\ell = 1.0$. Der Wert $\ell = 1.0$ liefert nach [46] die besten Ergebnisse und wird im weiteren nicht mehr verändert.
Da auf dem CNAPS-Neurocomputer nur die Wahl zwischen einer versteckten Schicht oder zwei versteckten Schichten existiert, wurden diese beiden Fälle analysiert. Für eine Lernrate $\eta \in \{0.00075, 0.003, 0.012, 0.048, 0.192\}$ wurden Netze mit 14 Eingabegrößen und variabler Anzahl von versteckten Knoten trainiert. Bei dreilagigen Netzen (eine versteckte Schicht) waren die Knotenzahlen in dieser versteckten Schicht 8, 16, 32, 64 und 128.
Es ist deutlich zu erkennen, daß sich im allgemeinen der MSE mit zunehmender Anzahl versteckter Knoten verringert (siehe Abbildung 6.2). Das Netzwerktraining wird also erfolgreicher. Der kleinste verwendete Lernparameter scheint relativ unabhängig von der Anzahl dieser Knoten zu sein. Äußerst wichtig ist die Wahl des „richtigen" Lernparameters. Der Bereich des Lernparameters zwischen $0.003 \leq \eta \leq 0.012$ zeigt die geringsten Fehler. Das gute Ergebnis, das diese Lernparameter bei nur 16 versteckten Knoten liefern, wird weder von dem kleinsten, noch von dem größten Lernparameter mit 128 versteckten Knoten erreicht. Sehr hohe Lernparameter sind für dieses Problem nicht einsatzfähig. Der MSE ist gegenüber den günstigen Lerparametern um mehr als einen Faktor zwei erhöht.

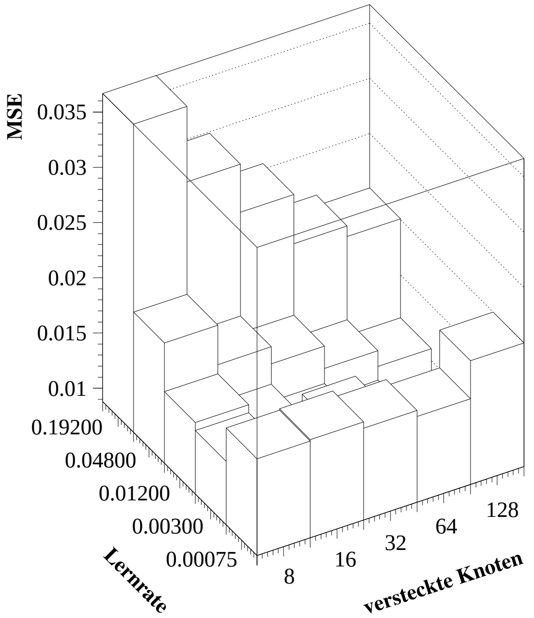


Abbildung 6.2: *Reaktion dreilagiger Netze auf Variation der Knotenzahl und des Lernparameters.*

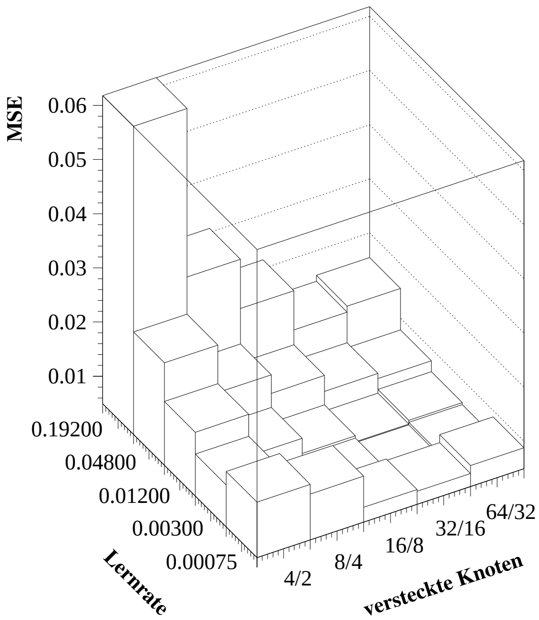


Abbildung 6.3: *Reaktion vierlagiger Netze auf Variation der Knotenzahl und des Lernparameters.*

Die besten erzielten mittleren quadratischen Fehler lagen bei ≈ 0.009.
Die Tendenzen, die sich schon bei Netzen mit einer versteckten Schicht gezeigt haben, treten bei Untersuchungen mit vierlagigen Netzen (zwei versteckte Lagen) in ähnlicher Form wieder auf (siehe Abbildung 6.3). Die Knoten in den beiden versteckten Lagen wurden im Verhältnis 2:1 gewählt. In der ersten versteckten Lage befinden sich also doppelt so viele Knoten wie in der zweiten. Mit einer großen Anzahl an versteckten Knoten wird schon wie bei dreilagigen Netzen ein deutlich geringerer Fehler erzielt. Mit 64/32 versteckten Knoten tritt jedoch gegenüber 32/16 Knoten praktisch keine Verbesserung mehr auf. Der größte getestete Lernparameter $\eta = 0.192$ zeigt auch bei vierlagigen Netzen das schlechteste Lernverhalten. Für die kleinste Netzgröße und den größten Lernparameter ist der MSE bei vierlagigen Netzen gegenüber dreilagigen Netzen fast verdoppelt. Dies legt die Vermutung nahe, daß bei Netzen mit mehreren versteckten Lagen und nur wenigen Knoten in diesen Lagen, eine hohe Lernrate nur nachteilige Auswirkungen für das Lernverhalten mit sich bringt.
Die Lernraten, die den geringsten MSE aufweisen, liegen im Bereich $0.003 \leq \eta \leq 0.012$. Der MSE liegt für diese Lernraten bei ≈ 0.005. Dieser Wert wurde von einem dreilagigen Netz mit 128 versteckten Knoten bei einer Lernrate von $\eta = 0.012$ auch nach einem Training über 20000 Epochen nicht unterboten. Der MSE war mit ≈ 0.008 sogar noch deutlich erhöht.

Bei dem Problem des Erkennens von transversalen Zerfallslängen sollte man daher vierlagige Netze mit genügend großer Knotenzahl verwenden. In den nachfolgenden Untersuchungen wurde deshalb mit einem 32/16-Netz operiert.

6.1.3 Initialisierung der Gewichte

Wie in Kapitel 4.4 erwähnt, müssen zu Beginn eines Trainings die Gewichte initialisiert werden. Sie geben den Startpunkt auf der zu minimierenden Fehlerfunktion an. Werden alle Anfangsgewichte auf identische Werte gesetzt, so ist es dem BPG nicht möglich, die Gewichte richtig auszurichten. Auf dem CNAPS werden die Anfangsgewichte zufällig, aber dennoch reproduzierbar gemäß

$$\omega_{ij} \in [-\kappa, \kappa] \quad \text{mit} \quad \kappa = \frac{3}{\sqrt{n}} \tag{6.2}$$

berechnet. n steht dabei für die Anzahl der Eingabeknoten. Die Formel 6.2 betrifft die Gewichtsinitialisierungen zwischen der Eingabeschicht und der ersten versteckten Schicht. Die Bestimmung der Anfangsgewichte zwischen der ersten und der zweiten versteckten Schicht, beziehungsweise zwischen der zweiten versteckten Schicht und der Ausgabeschicht erfolgt analog.
Bei dieser Studie wurden für $\eta \in \{0.00075, 0.003, 0.012, 0.048, 0.192\}$ fünf verschiedene Anfangsgewichte gewählt. In Abbildung 6.4 erkennt man sehr schön, daß der Fehler relativ unabhängig von der Initialisierung der Gewichte ist. Lediglich bei dem hier wegen Darstellungsproblemen (zu großer Fehler verdeckt Rest der Grafik) nicht

aufgetragenem Lernparameter $\eta = 0.00075$ ergaben sich etwas stärkere Schwankungen. Aufgrund des kleinen Lernparameters gelangt man hier wahrscheinlich leicht in ein Miminum in der Umgebung der Anfangsgewichte, welches nicht mehr verlassen werden kann. Diese Minima sind in der Nähe unterschiedlicher Startgewichte auch unterschiedlich tief. Bei den in Abbildung 6.4 dargestellten größeren Lernraten deutet alles darauf hin, daß für den jeweiligen Lernparameter immer das gleiche Minimum der Fehlerfunktion erreicht wird.

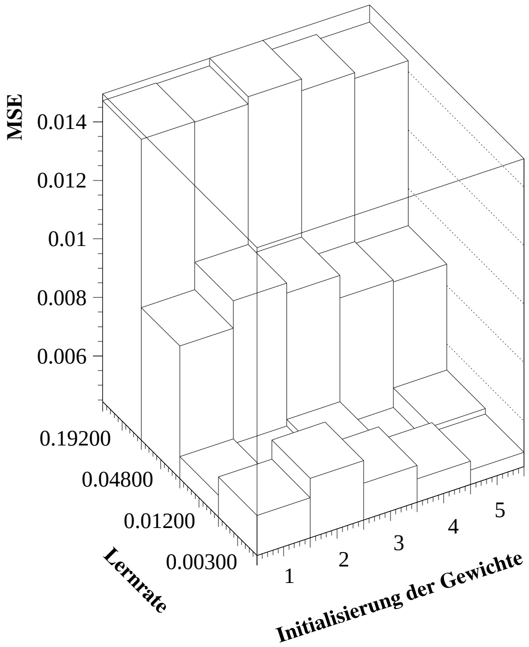


Abbildung 6.4: *Unterschiedliche Initialisierung der Gewichte bei variierender Lernrate.*

6.1.4 Hidden Learning Scale

Diese Untersuchung beschäftigt sich mit dem in Kapitel 4.4.1 eingeführten Hidden Learning Scale HLS. Nachdem in den vorangegangenen Untersuchungen deutlich wurde, daß gute Werte der Lernrate zwischen $0.003 \leq \eta \leq 0.012$ liegen, wird in diesem Abschnitt $\eta \in \{0.003, 0.006, 0.009, 0.012\}$ gewählt. Dazu wird der HLS $\beta \in \{1.0, 4.0, 7.0, 13.0\}$ variiert. Nach den Gleichungen 4.13 und 4.14 ist zu erwarten, daß günstige Werte für β mit steigender Lernrate abnehmen. In Abbildung 6.5 ist zu erkennen, daß für die größte hier untersuchte Lernrate $\eta = 0.012$ die Werte $\beta = 1.0$ und $\beta = 4.0$, also die kleinen Werte des HLS auch einen kleinen Fehler liefern. Dies wurde auch in [46] gezeigt. Der beim CNAPS-Parallelcomputer als Standardwert benutzte Wert $\beta = 4.0$ erweist sich über den gesamten getesteten η-Bereich als günstiger HLS-Parameter. Damit wird dieser Wert in den weiteren Untersuchungen verwendet.

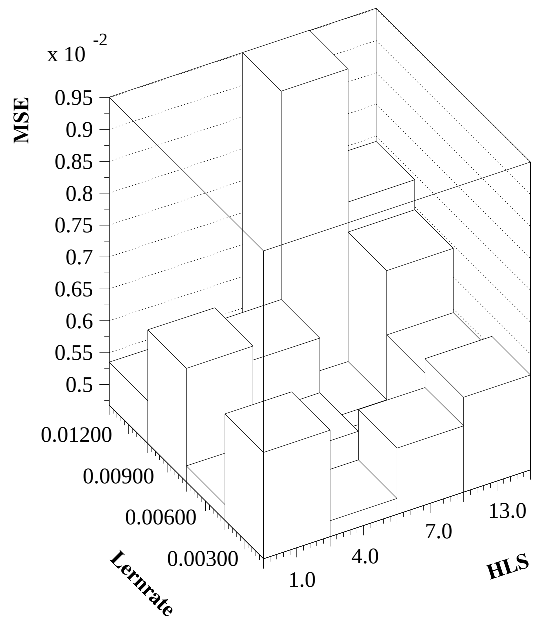


Abbildung 6.5: *Einfluß von Hidden Learning Scale und Lernrate auf das Lernverhalten.*

6.1.5 Momentumfaktor

Eine weitere Variation des BPG ist der Momentumfaktor α. Dieser wurde wie der HLS bereits in Kapitel 4.4.1 beschrieben. Die Schrittweite der Gewichtsänderungen wird durch die Kombination aus Lernrate η und Momentumfaktor α beeinflußt. Deshalb wurden für $\eta \in \{0.003, 0.006, 0.009, 0.012\}$ und für $\alpha \in \{0.0, 0.5, 0.8, 0.9\}$ die Veränderungen des mittleren quadratischen Fehlers beobachtet. In Abbildung 6.6 ist zu sehen, daß mit zunehmendem Momentumfaktor α für jede Lernrate η der MSE zunimmt. Da sich zudem die Rechenzeiten auf dem CNAPS-Parallelcomputer bei einem von Null verschiedenen Momentumfaktor etwa verdoppeln, wird im folgenden immer $\alpha = 0.0$ benutzt.
Die Lernrate η wurde in sämtlichen nachfolgenden Untersuchungen auf $\eta = 0.012$ festgelegt.

6.2 Einführung von Skalierungsfaktoren

Bei den bisher vorgenommenen Netzwerkuntersuchungen waren die Eingabegrößen für das neuronale Netz in einfacher Weise ausgewählt worden: Als Netzeingabe diente eine von der ersten Triggerstufe vorbereitete Detektorzelle. In Abbildung 5.9 ist diese Zellennummerverteilung der Eingabe für verschiedene Detektorradien für das Kaon zu sehen. Nur auf dem äußersten Detektorradius wird fast der gesamte Wertebereich von 65535 Zellennummern ausgenutzt. Hier tritt wegen des größten Detektorradius die größte Bogenlänge bei einer RoI von $\phi = \pm\ 45^0$ auf. Der in-

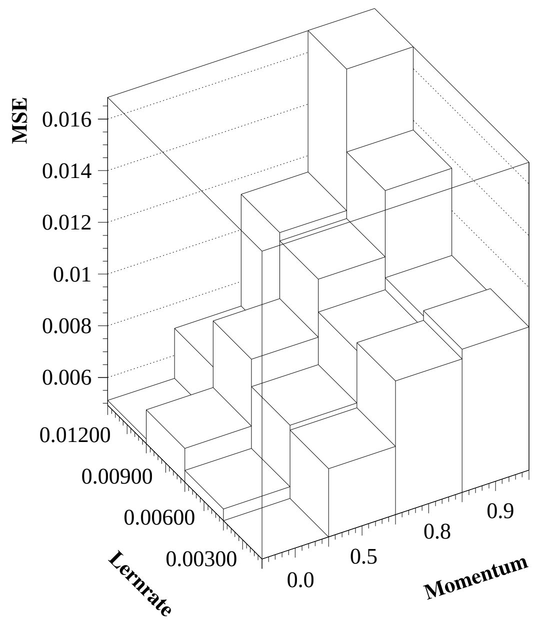


Abbildung 6.6: *Einfluß von Momentumfaktor und Lernrate auf das Lernverhalten.*

nerste Detektorradius dagegen hat nur 9425 verschiedene Detektorzellen und damit nur Werte von [-4712, +4712]. Man kann nun die Netzeingabe mit Hilfe eines für jeden Detektorradius verschiedenen Skalierungsfaktors S_j (j $\in \{1, 2, ..., 7\}$) auf den vollständigen CNAPS-Wertebereich [-32767, +32767] transformieren. Damit ergibt sich beispielsweise für den innersten Detektorradius ein Skalierungsfaktor S_1 von 6.95. Die Anzahl von 9425 Detektorzellen bleibt gleich. Es wird nur der CNAPS-Wertebereich [-32767, +32767] quantisiert, indem nur bestimmte Werte daraus als Eingabe zugelassen werden.
Im linken Bild der Abbildung 6.7 ist für die herkömmliche Methode die von [-32767, +32767] auf [0 mm, 40 mm] rücktransformierte Netzausgabe $l_{\mathrm{T}}^{\mathrm{Netz}}$ gegen den geforderten Wert der transversalen Zerfallslänge $l_{\mathrm{T}}^{\mathrm{Ziel}}$ aufgetragen. Daneben ist logarithmisch die Differenz aus beiden Werten dargestellt. Abbildung 6.8 veranschaulicht dies für eine mit Skalierungsfaktoren transformierte Netzeingabe.
Eine in den folgenden Untersuchungen entscheidende Größe ist die Standardabweichung, hier mit RMS (root mean squared error) bezeichnet:

$$RMS = \sqrt{\frac{1}{N-1}\sum_{p=1}^{N}[(o(\vec{x}_p, W) - t(\vec{x}_p)) - \overline{(o-t)}]^2} \qquad (6.3)$$

Zur Berechnung des RMS wird in Gleichung 6.3 für o $l_{\mathrm{T}}^{\mathrm{Netz}}$ und für t $l_{\mathrm{T}}^{\mathrm{Ziel}}$ eingesetzt. Dieser RMS beträgt bei Netzeingabe mit Skalierungsfaktoren 1.6 mm gegenüber der herkömmlichen Methode, die eine Standardabweichung von 2.8 mm hat. Das ist eine Verbesserung von mehr als 40 %.

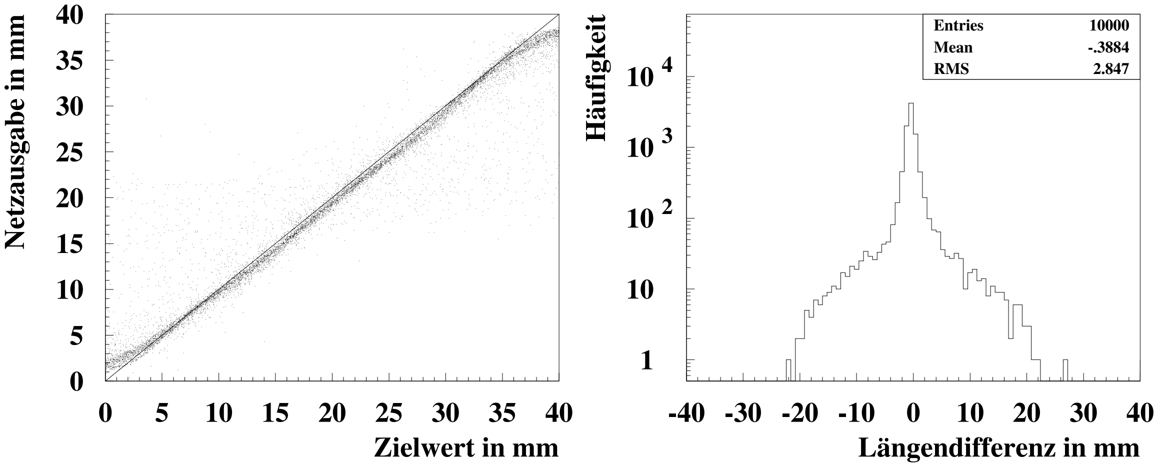


Abbildung 6.7: *Abbildungen für ein Netztraining ohne Verwendung von Skalierungsfaktoren. Links ist die Netzausgabe gegen den geforderten Wert aufgetragen, rechts die Differenz aus beiden Werten.*

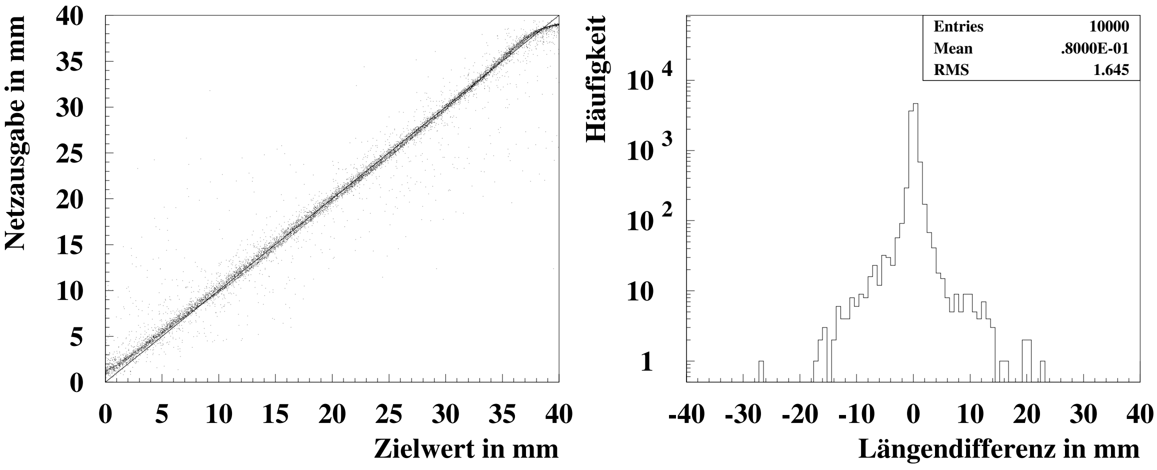


Abbildung 6.8: *Abbildungen für ein Netztraining mit Verwendung von Skalierungsfaktoren. Links ist die Netzausgabe gegen den geforderten Wert aufgetragen, rechts die Differenz aus beiden Werten.*

Vergleicht man die beiden linken Bilder der Abbildungen 6.7 und 6.8, so entdeckt man auch eine Gemeinsamkeit: Ein Abknicken der Kurve für kleine und große Werte von $l_{\mathrm{T}}^{\mathrm{Ziel}}$. Es spiegelt sich die in Abbildung 4.3 dargestellte Aktivierungsfunktion wider. Der Grund dafür ist, daß beim CNAPS-BPG die Netzausgabe der Aktivierungsfunktion unterliegt. Damit ist klar, daß das neuronale Netz kleine Werte tendentiell zu groß und große Werte tendentiell zu klein schätzt.
Aber es zeigt sich natürlich auch ein Unterschied hinsichtlich der beiden Eingabemethoden. Die großen Streuungen und Abweichungen von der in den Abbildungen eingezeichneten Geraden, die bei Eingabe ohne Skalierungsfaktoren auftreten, sind bei Eingabe mit Skalierungsfaktoren deutlich reduziert. Es ist also von erheblichem Vorteil, wenn der numerische Wertebereich des CNAPS-Neurocomputers vollständig ausgenutzt wird, da der CNAPS-Rechner mit Festkomma-Arithmetik arbeitet. Man erhält so eine viel bessere Abbildung der Netzausgabe $l_{\mathrm{T}}^{\mathrm{Netz}}$ auf den Zielwert $l_{\mathrm{T}}^{\mathrm{Ziel}}$.
Im folgenden wird daher die Eingabe immer mit Skalierungsfaktoren transformiert.

6.3 Analyse der Netzeingabe

Trotz der Einführung von Skalierungsfaktoren S_j sind immer noch, wenn auch entscheidend weniger, größere Abweichungen von Netzausgabe $l_{\mathrm{T}}^{\mathrm{Netz}}$ und Zielwert $l_{\mathrm{T}}^{\mathrm{Ziel}}$ zu beobachten. Es stellt sich die Frage wie Eingabegrößen beschaffen sind, die zu solchen Abweichungen führen und welche Eingaben sehr gut gelernt werden können.

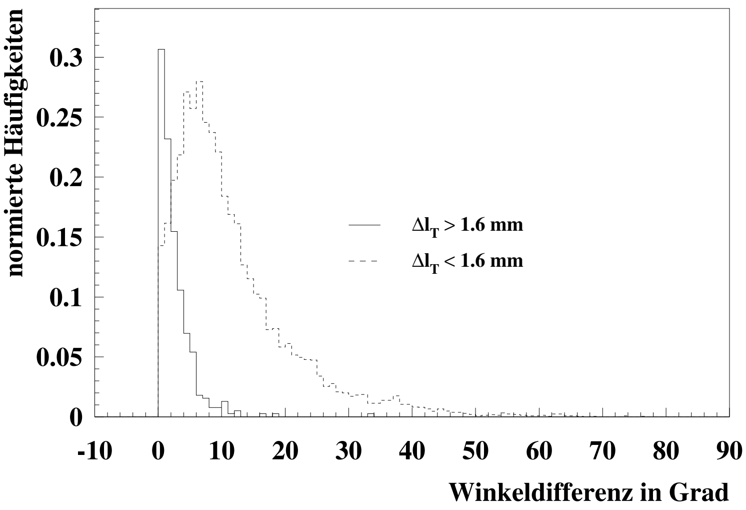


Abbildung 6.9: *Vergleich von gut gelernten Ereignissen (gestrichelte Kurve) zu schlecht gelernten Ereignissen (durchgezogene Linie). Da die Anzahl der Ereignisse mit* $\Delta l_{\mathrm{T}} < 1.6$ mm *mehr als 96 % der gesamten Ereignisse darstellt, sind die beiden Kurven zum besseren Vergleich unterschiedlich normiert.*

Als äußerst aussagekräftige Größe erweist sich diesbezüglich die Winkeldifferenz von Detektortrefferpunkten auf dem innersten Detektorradius. Der Winkel wird von der

Strahlachse (= Mittelpunkt der Detektorradien) aus berechnet. In Abschnitt 6.2 zeigte sich ein RMS von 1.6 mm. Daher wurde als Unterscheidungskriterium für gut, beziehungsweise schlecht gelernte Werte $\Delta l_{\mathrm{T}} = 1.6$ mm gewählt. Abbildung 6.9 zeigt, daß Ereignisse mit kleinen Winkeldifferenzen sehr viel schwieriger zu erlernen sind als solche mit großen. Dies mag darauf zurückzuführen sein, daß selbst die transformierten Zellennummern für die beiden Eingaben auf dem innersten Detektorradius noch zu ähnlich sind .
Eine weitere Untersuchung betrifft die Lage der Teilchenspuren zueinander. Grundsätzlich gibt es hier die zwei in Abbildung 6.10 dargestellten Möglichkeiten. Entweder entfernen sich die Spuren voneinander, oder sie bewegen sich aufeinander zu.

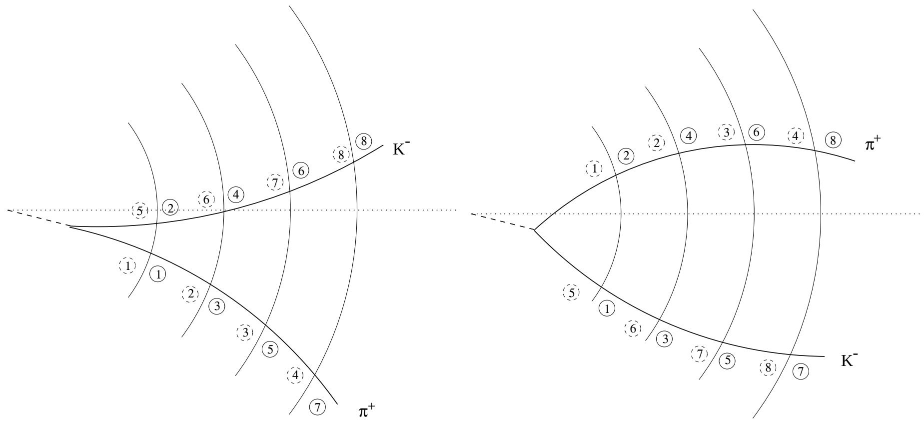


Abbildung 6.10: *Links sind Spuren dargestellt, die sich voneinander entfernen, rechts solche, die sich einander annähern und gleichzeitig außerhalb des Detektors schneiden. Die Zahlenfolge in vollen Kreisen gibt die radiusweise Auslese an, die in gestrichelten Kreisen, diejenige mit Spur- und Teilchenerkennung.*

Bei dem letzteren Verhalten können sich die Spuren innerhalb oder außerhalb des Detektors schneiden. Zunächst werden nur die Fälle des Auseinanderbewegens der Spuren und des Zusammenbewegens mit Schnittpunkt außerhalb des Detektors betrachtet. Im ersten Fall erhält der erste Netzknoten als Eingabe den Detektortreffpunkt des π, der zweite Knoten erhält den des K auf dem ersten Detektorradius u.s.w. Im zweiten Fall tritt das Umgekehrte auf (siehe Abbildung 6.10). Es bekommen jedoch in beiden Fällen alle ungeraden Eingabeknoten die Detektortreffpunkte der einen Teilchenspur und alle geraden Eingabeknoten diejenigen der anderen Teilchenspur. Für das neuronale Netz bedeutet das im Prinzip eine vorweggenommene Spurerkennung. Bild 6.11 zeigt die Qualität der Netzantwort. Bei den auseinanderlaufenden Spuren treten einige größere Abweichungen der Netzausgabe gegenüber dem Zielwert auf, wohingegen dies bei auseinanderlaufenden Spuren überhaupt nicht erkennbar ist. Dies ist vereinbar mit der Aussage, daß es schwierig ist, ein gutes Ergebnis zu schätzen, wenn die Winkeldifferenz auf dem ersten Detektorradius klein ist. Bei Spuren, die sich erst außerhalb des Detektors schneiden, ist diese Winkeldifferenz immer ausreichend, um eine gute Schätzung zu erhalten. Bei auseinanderlaufenden Spuren gibt es hin und wieder solche, die sich durch kleine

Winkeldifferenzen auszeichnen. Diese bereiten Probleme bei einer guten Schätzung.

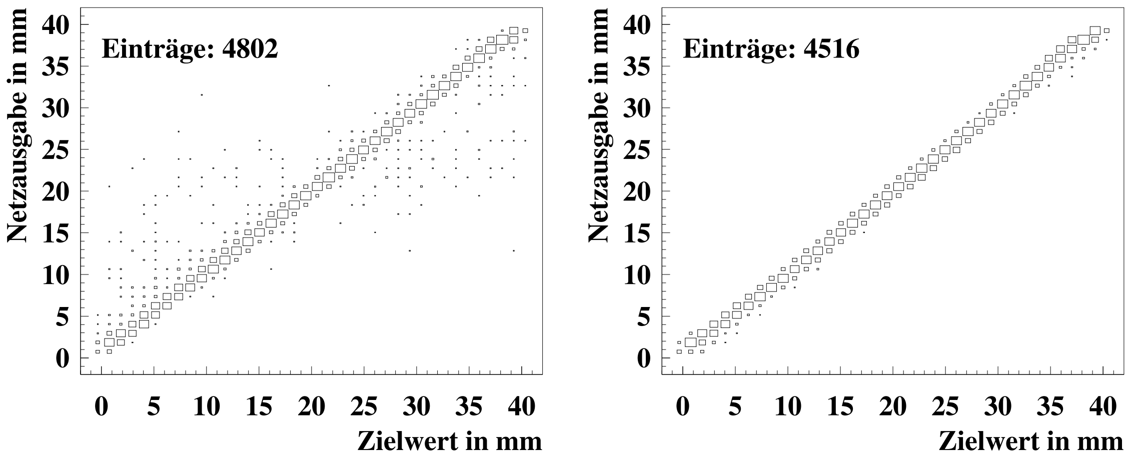


Abbildung 6.11: *Darstellung der Netzausgabe gegen den Zielwert. Links für sich entfernende Spuren, rechts für sich annäherde Spuren mit Schnittpunkt außerhalb des Detektors.*

Wenn man die Spuren betrachtet, die sich innerhalb des Detektors schneiden, wie es in Abbildung 6.12 zu sehen ist, sollte man zwei verschiedene Eingabearten vergleichen. Zum einen wurde die radienweise Auslese, zum anderen die Auslese mit vorheriger Spur- und Teilchenerkennung (Paricle Identification, ID) benutzt. Das bedeutet, daß die ersten sieben Eingabeknoten die Detektortrefferpunkte des π und die Eingabeknoten 8 bis 14 die des K in aufsteigender Radiusfolge bekommen.

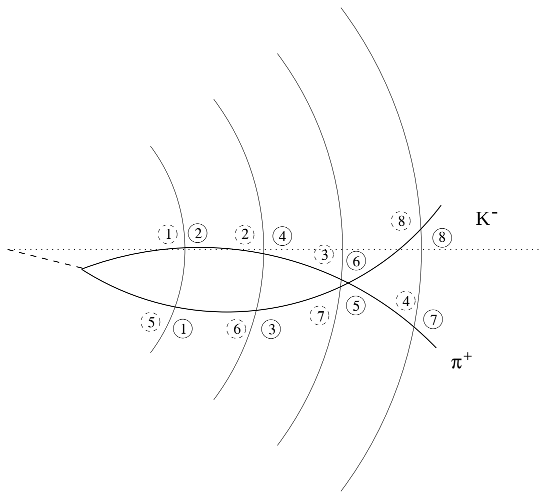


Abbildung 6.12: *Darstellung von Spuren, die sich im Detektor schneiden. Die Zahlenfolge in vollen Kreisen gibt die radiusweise Auslese an, die in gestrichelten Kreisen, diejenige mit Spur- und Teilchenerkennung.*

Bei der radienweisen Auslese entsteht für Eingaben nach dem Schnittpunkt der Spuren ein Bruch. Die geraden Eingabeknoten bekommen für diesen Fall nicht mehr nur Detektortrefferpunkte, die zu einer Spur gehören. Dasselbe gilt natürlich auch für die ungeraden Eingabeknoten. Das erschwert dem Netz die richtige Schätzung der transversalen Zerfallslänge (siehe Abbildung 6.13).

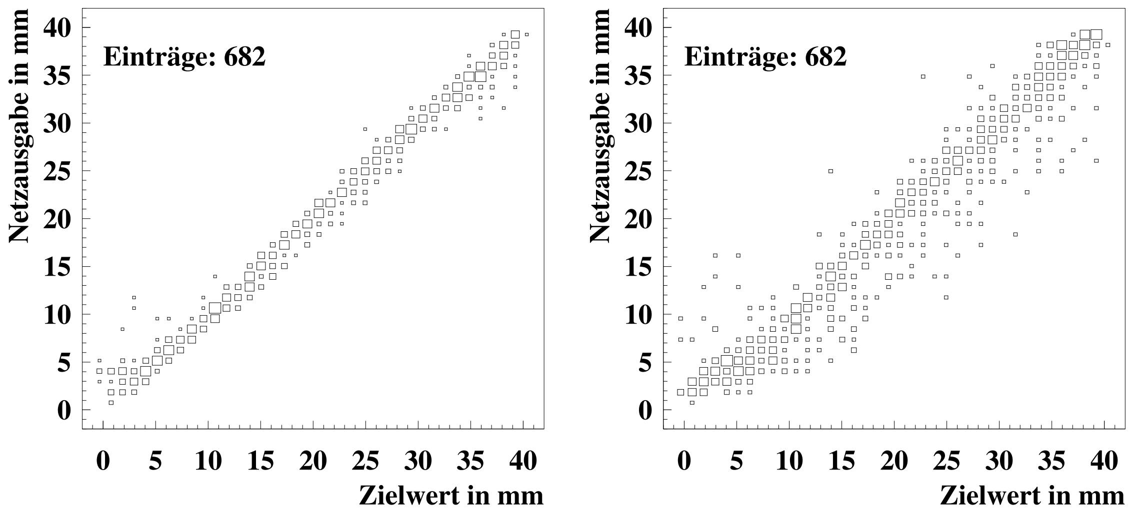


Abbildung 6.13: *Darstellung der Netzausgabe gegen den Zielwert für sich im Detektor schneidende Spuren. Links wird bei der Netzeingabe Spur- und Teilchenerkennung vorausgesetzt, rechts wird die bekannte radienweise Eingabe gemacht.*

Generell bilden Detektortrefferpunkte von sich innerhalb des Detektors schneidenden Spuren eine schwierigere Eingabe als solche von Spuren, die sich außerhalb des Detektors schneiden. Das Argument ist wiederum die kleine Winkeldifferenz.
In Tabelle 6.1 finden sich die mittleren quadratischen Abweichungen von Netzausgabe $l_{\mathrm{T}}^{\mathrm{Netz}}$ und Zielwert $l_{\mathrm{T}}^{\mathrm{Ziel}}$ für die beiden vorgestellten Möglichkeiten der Netzeingabe.

Eingabeart	RMS_{all} mm	$\mathrm{RMS}_{incross}$ mm	$\mathrm{RMS}_{outcross}$ mm	$\mathrm{RMS}_{nocross}$ mm
ID	1.25	1.33	0.35	1.69
Radienweise	1.64	3.22	0.49	1.93

Tabelle 6.1: *Darstellung der Netzleistung für unterschiedliche Detektorauslesearten. Der RMS ist bestimmt für alle Spuren (RMS_{all}), für sich innerhalb des Detektors schneidende Spuren ($RMS_{incross}$), für sich außerhalb des Detektors scheidende Spuren ($RMS_{outcross}$), und für auseinanderlaufende Spuren ($RMS_{nocross}$)*

Bei der radienweisen Auslese sind auch die auseinanderlaufenden und sich außerhalb des Detektors schneidenden Spuren etwas schwieriger zu erlernen als bei der Methode

mit Spur- und Teilchenerkennung. Das hat seine Ursache vermutlich darin, daß das auf diese Weise trainierte Netz auch schwierige Information (sich im Detektor schneidende Spuren) erlernen mußte.
Eine Detektorauslese mit vorheriger Spurerkennung („Tracking“), aber keiner Teilchenerkennung, lieferte einen RMS_{all} von 1.25 mm. Es zeigte sich also kein Unterschied, ob man nur Tracking oder Tracking mit Teilchenerkennung (ID) voraussetzte.

6.4 Vergleich mit anderen Algorithmen

In diesem Abschnitt soll die Leistung des neuronalen Netzes mit der anderer, einfacher Algorithmen verglichen werden. Getestet wurden ein Geraden-Algorithmus und ein Kreis-Algorithmus.

Der Geraden-Algorithmus

Bei diesem Algorithmus werden nur die Detektortrefferpunkte der zwei innersten Detektorradien als Eingabe benutzt. Die untersuchte Ausleseform ist die radienweise Auslese. Aus der ersten und dritten und aus der zweiten und vierten Auslesezelle werden jeweils eine Gerade berechnet. Diese beiden Geraden bringt man dann zum Schnitt. Der Abstand des erhaltenen Schnittpunktes von der Strahlache ist für diesen Algorithmus die transversale Zerfallslänge. Da die beiden Teilchenspuren gekrümmt sind, wird man niemals eine sehr gute Übereinstimmung mit dem wirklichen Wert erhalten. Ist die transversale Zerfallslänge relativ groß, so sollten die Abweichungen jedoch geringer werden, da die Krümmung der Spuren erst über eine kürzere Strecke erfolgt ist als bei Spuren, die von einem sekundären Vertex kommen, der nahe beim primären liegt. Diese Beobachtung wurde auch gemacht (siehe Abbildung 6.14).

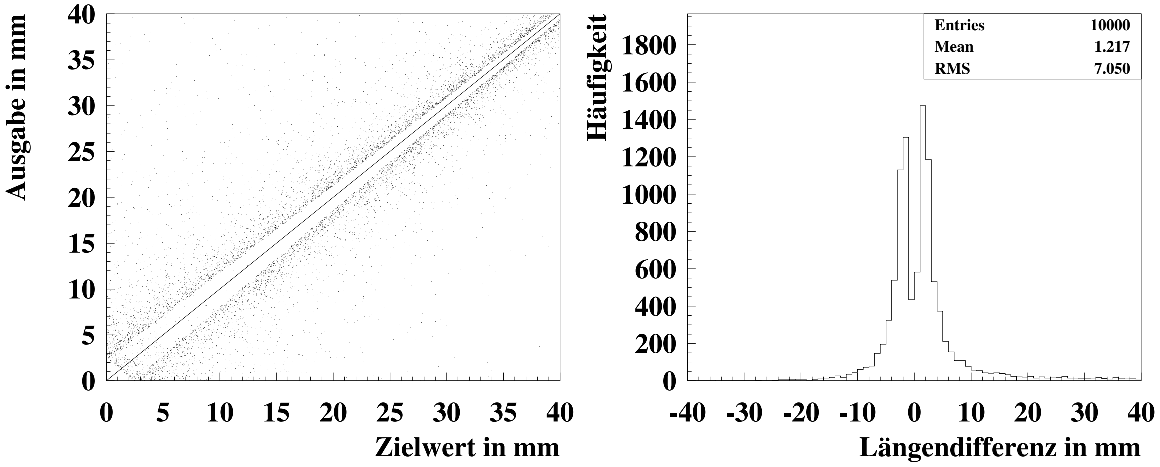


Abbildung 6.14: *Links ist die Ausgabe des Geradenalgorithmus gegen die geforderte Ausgabe abgebildet. Rechts ist die Differenz aus dieser Algorithmusausgabe und dem Zielwert dargestellt.*

Um einen fairen Vergleich mit dem Ergebnis eines neuronalen Netzes zu erhalten, wird der Abstand des Geradenschnittpunktes, falls dieser 40 mm überschreitet auf genau 40 mm gesetzt. Damit wird auch hier der Wertebereich der transversalen Zerfallslänge auf den Bereich [0, 40 mm] festgelegt. Wie schon in den vorangegangenen Abschnitten erwähnt, erreichte das Netz einen RMS von 1.6 mm. Dieser Wert wird von dem einfachen Geraden-Algorithmus weit verfehlt. Er liefert einen RMS von 7.1 mm.

Der Kreis-Algorithmus

Dieser Algorithmus benutzt als Eingabe die Detektortrefferpunkte des innersten, des mittleren und des äußersten Detektorradius (0.06 m, 0.3 m, 0.6 m). Um dem Netz bei radienweiser Auslese gerecht zu werden, wird der erste Kreis mit Hilfe des ersten, siebten und dreizehnten Auslesepunktes, der zweite Kreis mit Hilfe des zweiten, achten und vierzehnten Auslesepunktes bestimmt. Die beiden auf diese Art erhaltenen Kreise werden geschnitten. Der Abstand des Schnittpunktes von der Strahlachse beschreibt die transversale Zerfallslänge, welche der Kreis-Algorithmus berechnet.
In Abbildung 6.15 ist die Differenz aus Algorithmusausgabe und gefordertem Wert und die Differenz aus Netzausgabe und Zielwert für eine Abweichung kleiner als 1.0 mm dargestellt. Wie erwartet liefert der Kreis-Algorithmus für diesen Abweichungsbereich eine bessere Lösung als das neuronale Netz. Allerdings kann es beim

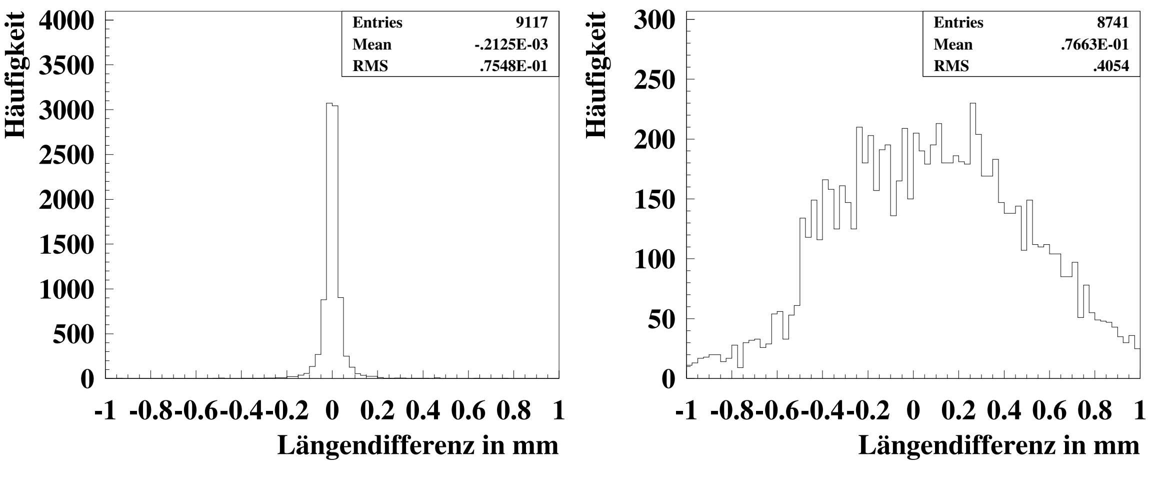


Abbildung 6.15: *Links ist für den Kreisalgorithmus die Differenz aus dessen Ergebnis und dem geforderten Wert im Bereich [-1.0 mm, 1.0 mm] abgebildet. Rechts sieht man dieselbe Abbildung für das neuronale Netz.*

Kreis-Algorithmus passieren, daß sich Kreise überhaupt nicht schneiden. Dies liegt an der reduzierten Detektorauflösung. Errechnet der Algorithmus keinen Schnittpunkt wird er automatisch auf Null gesetzt. Schnittpunkte, die aufgrund radienweiser Auslese der Detektortrefferpunkte von falsch rekonstruierten Teilchenradien

stammen, verschlechtern das Ergebnis ebenfalls signifikant. Andere erhebliche Abweichungen können durch schwierige geometrische Bedingungen hervorgehen. Dies alles führt dazu, daß bei Betrachtung des gesamten Wertebereich der Differenzen von [-40 mm, 40 mm], das Netzergebnis eine kleinere Standardabweichung liefert als das Ergebnis des Kreis-Algorithmus.
Fordert man für den Kreis-Algorithmus eine Auslese der Detektortrefferpunkte mit vorheriger Spur- und Teilchenerkennung (ID), so wird das Ergebnis zwar besser, kann aber das Ergebnis des neuronalen Netzes nicht unterbieten. Dies ist in Tabelle 6.2 dargestellt. Es sollte allerdings bemerkt werden, daß das Netz 14 Eingabedaten erhält, der Kreis-Algorithmus nur sechs und daß der Kreisalgorithmus noch optimiert werden könnte.

Eingabeart	RMS(Kreis) mm	RMS(Kreis$_{mod}$) mm	RMS(Netz) mm
ID	1.60	1.25	1.25
Radienweise	6.02	4.41	1.64

Tabelle 6.2: *Zusammenfassung der Ergebnisse eines neuronalen Netzes und eines einfachen Kreisalgorithmus. Kreis$_{mod}$ bedeutet, daß Kreise ohne gemeinsamen Schnittpunkt in dieser Berechnung keine Brücksichtigung fanden.*

6.5 Lernverhalten bei Variation der Detektorgeometrie

In diesem Abschnitt soll die Abhängigkeit der Detektorgeometrie auf das Erlernen der transversalen Zerfallslänge überprüft werden. Untersucht wurden die Auflösung des Detektors, die Anzahl der Detektorlagen und die Größe des äußersten Detektorradius (Länge des Hebelarms). Nach der Theorie haben alle drei Variationsmöglichkeiten Einfluß auf die Genauigkeit der Krümmungsbestimmung einer Teilchenkurve. Die Krümmung einer Teilchenkurve ist gegeben durch $\kappa = 1/R$, wobei R für den Teilchenradius steht. Der Fehler dieser Krümmung kann durch folgende Formel [47, 48] angegeben werden:

$$\sigma_\kappa = \frac{\sigma_{R\phi}}{L^2}\sqrt{\frac{720}{N+4}} \tag{6.4}$$

In dieser Formel bedeutet:

N = *Anzahl der Detektortrefferpunkte einer Teilchenspur*
L = *Länge der Teilchenspur in der transversalen Ebene*
$\sigma_{R\phi}$ = *räumlicher Meßfehler des Spurdetektors in der r-ϕ-Ebene senkrecht zum Magnetfeld*

Die Genauigkeit dieser Krümmungsbestimmung hat letztendlich Einfluß darauf, wie exakt man den Ort des sekundären Vertex in der Rekonstruktion bestimmen kann.
Bei den nachfolgenden Untersuchungen wurden jeweils für eine variierende Größe drei Netztrainings durchgeführt. Eine größere Anzahl an Trainings war aufgrund der Trainingszeit eines einzelnen Netzes von mehreren Stunden nicht möglich. Das läßt natürlich nur eine unzureichende Fehlerabschätzung zu. Trotzdem sollten sich zumindest Tendenzen erkennen lassen.
Die Zahl der Trainigsdaten betrug weiterhin 10000, da sich in stichprobeartigen Untersuchungen keine Verbesserungen der Netzergebnisse bei der für den CNAPS-Parallelcomputer maximalen Zahl von 65535 Trainigsdaten abzeichnete. Das Training würde sich jedoch um etwa einen Faktor 6.5 verlängern.

Detektorauflösung

Die Detektorauflösung geht in die Formel 6.4 über den Parameter $\sigma_{R\phi}$ ein. Bei einer Rekonstruktion würde man also erwarten, daß mit verbesserter Auflösung eine genauere Bestimmung der Krümmung κ und damit eine exaktere Bestimmung der transversalen Zerfallslänge möglich ist. Diese Untersuchung soll nun zeigen wie ein neuronales Netz auf Veränderungen der Detektorauflösung reagiert.
Die Detektorauflösung wurde ausgehend von der in Tabelle 5.1 gewählten Auflösung verschlechtert. Die Verschlechterung geschah um einen Faktor zwei, fünf, zehn und hundert. Die äußeren Siliziumstreifenlagen hatten also eine Auflösung von 30 μm, 75 μm, 150 μm und 1500 μm, wie dies in Abbildung 6.16 beschriftet ist. Als Qualität der Schätzung wurde jeweils der Mittelwert aus den RMS (Gl. 6.3) genommen.

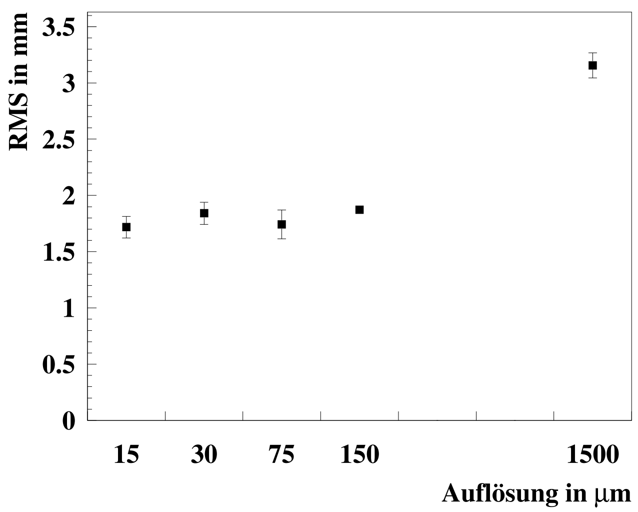


Abbildung 6.16: *Aufgetragen ist der RMS gegen die Detektorauflösung. Ein kleinerer RMS stellt eine geringere Abweichung der Netzanwort gegenüber dem geforderten Wert dar.*

Das Netz ist relativ robust bei Verschlechterung der Detektorauflösung bis zu einem Faktor zehn oder eventuell noch höher. Erst bei einer Verschlechterung um einen Faktor 100 verdoppelt sich der RMS fast. Netze benötigen folglich nicht die volle Detektorgranularität.

Anzahl der Detektorlagen

In der Rekonstruktion sollte nach Gleichung 6.4 eine Erhöhung der Zahl der Spurpunkte eine Verbesserung der Krümmungsbestimmung zulassen. Dieser Effekt wurde durch eine Vergrößerung der Detektorlagen für das neuronales Netz untersucht. Es wurde die Benutzung von vier, sieben und zwölf Detektorlagen getestet. Die Position der verwendeten Lagen ist der Tabelle 6.3 zu entnehmen. Der äußerste Detektorradius bleibt unabhängig von der Lagenzahl bei 60 cm (Länge des Hebelarms). Die Auflösung wurde auf allen Detektorlagen einheitlich $\sigma_{\mathrm{R}\phi} = 15\ \mu$m gewählt.

Anzahl der Detektorlagen	Position cm											
4	6			20				40				60
7	6	10		20		30		40		50		60
12	6	10	15	20	25	30	35	40	45	50	55	60

Tabelle 6.3: *Position der Detektorradien bei Variation der Lagenzahl.*

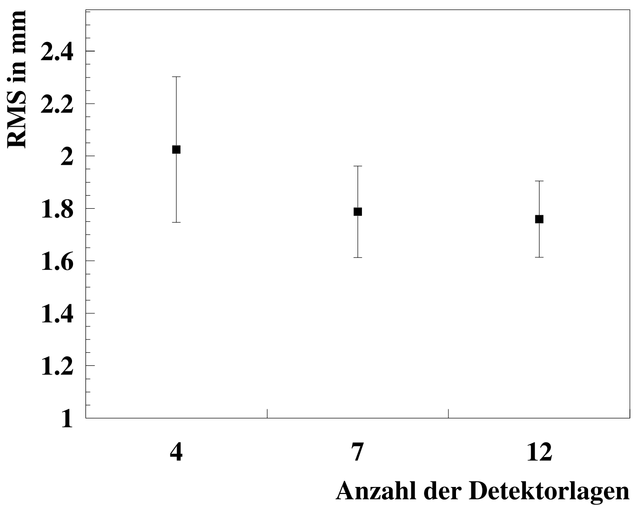


Abbildung 6.17: *Aufgetragen ist der RMS gegen die Anzahl der Detektorlagen.*

Abbildung 6.17 zeigt, daß im Rahmen der Fehler kein Unterschied bei Veränderung der Detektorlagen auftritt.
Es ist eine Verbesserung der Schätzung der transversalen Zerfallslänge durch ein neuronales Netz erkennbar, wenn man nur die Mittelwerte der RMS bei vier und sieben Detektorlagen vergleicht. Zwischen sieben und zwölf Detektorlagen besteht diesbezüglich kaum ein Unterschied. Um die vermutete Abhängigkeit der Netzleistung von der Anzahl der Detektorlagen tatsächlich zu bestätigen, sind jedoch pro Datenpunkt erheblich mehr als nur die drei durchgeführten Trainings notwendig. Wie bereits oben erwähnt, war der begrenzende Faktor dieser Untersuchung die extrem lange Rechenzeit.

Länge des Hebelarms

Nun wird überprüft, wie ein neuronales Netz reagiert, wenn man bei gleichbleibender Detektorlagenzahl (sieben) und fester Auflösung ($\sigma_{\mathrm{R}\phi} = 15\ \mu$m) auf diesen Lagen die Position des äußersten Detektorradius verschiebt. Die Rekonstruktion sollte nach Gleichung 6.4 eine Verbesserung der Krümmungsbestimmung mit wachsendem Hebelarm liefern. Der äußerste Detektorradius wurde für diese Untersuchung auf 30 cm, 60 cm, und 90 cm gesetzt. Die Positionen der einzelnen Detektorlagen finden sich in Tabelle 6.4.

Länge des Hebelarms in cm	Position cm						
30	6	10	14	18	22	26	30
60	6	15	24	33	42	51	60
90	6	20	34	48	62	76	90

Tabelle 6.4: *Position der Detektorradien bei Variation des Hebelarms.*

Um eine Auflösung von $\sigma_{\mathrm{R}\phi} = 15\ \mu$m auch bei einem Hebelarm von 90 cm zu gewährleisten, mußte wegen des CNAPS-Wertebereichs von [-32767, 32767] der $\phi-$Bereich auf $\phi = \pm\ 30^0$ erniedrigt werden. Diese RoI von 60^0 wurde für alle drei Längen des Hebelarms beibehalten.
Wie schon für die Anzahl der Detektorlagen, zeigt das Netzergebnis im Rahmen der Fehler auch für die Variation der Hebelarmlänge keine Auswirkungen. Die Qualität der Netzantwort ist in üblicher Darstellung in Abbildung 6.18 zu sehen. Ohne Berücksichtigung der Fehler könnte man eine Verbesserung bei einem Hebelarm von 60 cm gegenüber den beiden anderen vermuten. Diese Vermutung kann dadurch gestützt werden, daß einerseits bei Verlängerung des Hebelarms eine Verbesserung der Netzantwort erwartet wird, andererseits jedoch bei einem größeren Hebelarm mehr sich innerhalb des Detektors schneidende Spuren auftreten, was eine Verschlechterung der Netzantwort nach sich zieht. Diese beiden gegenläufigen Trends

könnten die beste Netzschätzung bei einer Länge des Hebelarms von 60 cm ergeben. Zur Bestätigung dieser Vermutung müßten jedoch, wie schon bei der Analyse der Anzahl der Detektorlagen, deutlich mehr Trainings durchgeführt werden.

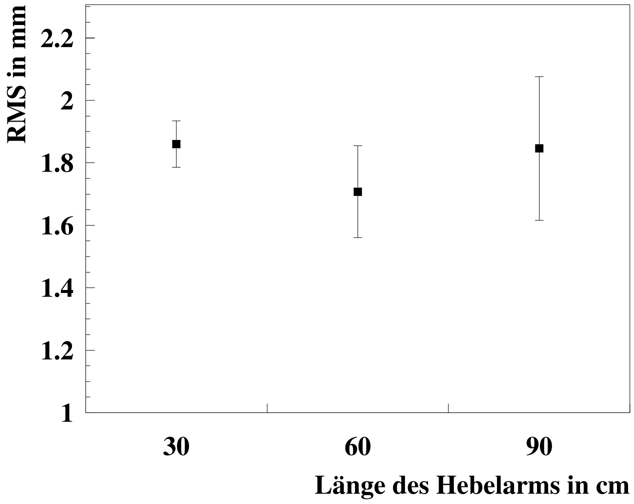


Abbildung 6.18: *Aufgetragen ist der RMS gegen die Länge des Hebelarms.*

6.6 Lernverhalten bei Variation kinematischer Variablen

Zur Untersuchung, wie das neuronale Netz auf Variation des Wertebereichs der transversalen Zerfallslänge $l_{\mathrm{T}}^{\mathrm{Ziel}}$ und auf Variation des Transversalimpulses $p_{\mathrm{T}}(\mathrm{D}^{\star})$ reagiert, wird die Detektorgeometrie wie in Kapitel 5.3 beschrieben, beibehalten. Zum Trainieren und Testen wurden, wie schon im vorangegangenen Kapitel erörtert, drei Versuche unternommen, um zumindest eine kleine Statistik zu erhalten.

Bereich der transversalen Zerfallslänge

In den Trainings- und Testdaten wurde der Wertebereich von $l_{\mathrm{T}}^{\mathrm{Ziel}}$ in der Weise beschränkt, daß als maximales $l_{\mathrm{T}}^{\mathrm{Ziel}}$ 10 mm, 20 mm, 30 mm und 40 mm vorgegeben wurden. Die Verteilungen von $l_{\mathrm{T}}^{\mathrm{Ziel}}$ waren Gleichverteilungen zwischen 0 mm und dem jeweils maximalen $l_{\mathrm{T}}^{\mathrm{Ziel}}$.
Interessant ist bei dieser Untersuchung, ob das neuronale Netz kleine Zerfallslängen noch genügend gut auflösen kann und inwieweit die Größe des Wertebereichs eine Auswirkung auf das Lernvermögen hat. Eine Vergrößerung des Wertebereichs der geforderten Netzantwort erschwert dem Netz die richtige Schätzung der transversalen Zerfallslänge (siehe Abbildung 6.19).

Dies liegt vermutlich einerseits am größeren Bereich des Eingaberaums und andererseits daran, daß sich bei einem größeren Wertebereich von $l_{\mathrm{T}}^{\mathrm{Ziel}}$ auch stärkere Abweichungen von Netzausgabe $l_{\mathrm{T}}^{\mathrm{Netz}}$ und Zielwert $l_{\mathrm{T}}^{\mathrm{Ziel}}$ ergeben können, was sich in einem größeren RMS äußert. Der zuletzt genannte Grund dürfte dabei der entscheidende sein. Als sehr angenehm erweist sich die Tatsache, daß es keine Auflösungsprobleme von kleinen Zerfallslängen gibt. Es war nicht von Anfang an klar, daß das Netztraining für kleine Zerfallslängen überhaupt funktioniert.

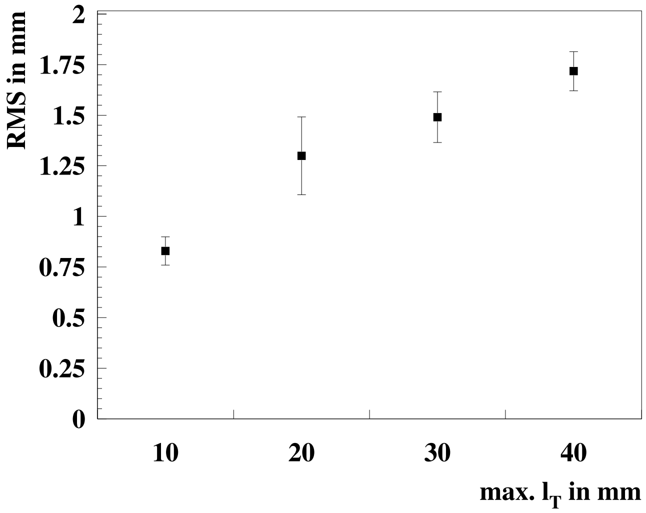


Abbildung 6.19: *Qualität der Netzantwort in Abhängigkeit vom Wertebereich der transversalen Zerfallslänge* $l_{\mathrm{T}}^{\mathrm{Ziel}}$.

Bereich des Transversalimpulses des $\mathrm{D}^\star$

Die zweite durchgeführte Variation einer kinematischen Variablen war die des Transversalimpulses des $\mathrm{D}^\star$. Dabei wurden drei disjunkte Bereiche für $p_{\mathrm{T}}(\mathrm{D}^\star)$ gewählt (0-20 GeV/c, 20-60 GeV/c, 60-100 GeV/c). Die geforderte transversale Zerfallslänge war gleichverteilt zwischen 0 mm $< l_{\mathrm{T}}^{\mathrm{Ziel}} <$ 40 mm.
Große $p_{\mathrm{T}}(\mathrm{D}^\star)$ ziehen große Transversalimpulse der Zerfallsteilchen nach sich. Dies hat wiederum zur Folge, daß die Krümmungen der Spuren im Detektor abnehmen. Ein weiterer Effekt ist die nähere Ausrichtung der Spuren zur $p_{\mathrm{T}}(\mathrm{D}^\star)$-Richtung. Damit wird auch speziell der Winkel zwischen den beiden Detektortrefferpunkten vom Kaon und vom Pion auf dem kleinsten Detektorradius geringer. Man erwartet, diesen Argumenten folgend, eine Verschlechterung der Netzausgabe mit steigendem $p_{\mathrm{T}}(\mathrm{D}^\star)$. Das Verhalten tritt auch deutlich auf, wie in Abbildung 6.20 zu erkennen ist.
Diese Studie wurde bei der üblichen Beschränkung der RoI auf $\phi = \pm\ 45^0$ durchgeführt. Trotz der Verwendung von Skalierungsfaktoren für die Eingabegrößen, sind diese Eingabegrößen bei hohem $p_{\mathrm{T}}(\mathrm{D}^\star)$ sehr stark um die Null-Zelle ($p_{\mathrm{T}}(\mathrm{D}^\star)$-Richtung) konzentriert. Der Wertebereich des CNAPS-Neurocomputers

wird also nicht gut ausgenutzt. Eine Verbesserung bringt eine Verkleinerung der RoI auf $\phi = \pm\, 30^0$, da damit die Skalierungsfaktoren vergrößert werden. Der numerische Bereich des CNAPS wird so günstiger ausgeschöpft. Für 20 GeV/c $< p_{\mathrm{T}}(\mathrm{D}^\star) <$ 60 GeV/c betrug dann der RMS $= 2.00 \pm 0.11$ mm, bei einer RoI von $\phi = \pm\, 45^0$ war der RMS $= 2.48 \pm 0.17$ mm.

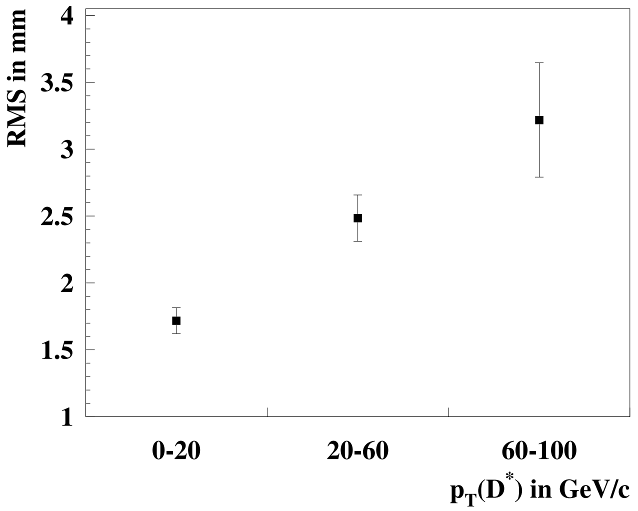


Abbildung 6.20: *Verschlechterung der Netzleistung bei Erhöhung von $p_{\mathrm{T}}(\mathrm{D}^\star)$.*

Kapitel 7

Untersuchungen zur Spurmultiplizität

7.1 Spurmultiplizität in RoIs bei $t\bar{t}$-Zerfällen

Im letzten Kapitel wurde gezeigt, daß es prinzipiell möglich ist, alleine aus Detektortrefferpunkten transversale Zerfallslängen zu schätzen. Es wurden jedoch im Gegensatz zum realistischen Fall einige Vereinfachungen angenommen. Die entscheidende Vereinfachung war, daß man nur Spurpunkte der Teilchen als Netzeingabe benutzte, die an einem sekundären Vertex produziert wurden.
In Wirklichkeit ist der gesamte Detektor „voll“ von Teilchenspuren. Das bedeutet, daß um die Impulsrichtung des $D^\star$ mehr Spuren als nur die des Kaons und des Pions verteilt sind. Man muß sich folglich darüber klar werden, wieviele Teilchen in einer bestimmten Region um das $D^\star$-Meson auftreten. Das Maß für die Größe dieser Region ist der Konus Δr:

$$\Delta r = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{7.1}$$

$\Delta\phi$ $(\phi \in [0, 2\pi])$ und $\Delta\eta$ $(\eta \in]-\infty, +\infty[)$ werden dabei immer im Vergleich zu einer fest vorgegebenen Richtung gerechnet. Δr stellt gewissermaßen eine RoI dar, die von der ersten Triggerstufe als interessant deklariert wurde.
Die nachfolgenden Untersuchungen zur Bestimmung der Teilchendichte um das $D^\star$-Meson basieren auf einem $t\bar{t}$-Monte-Carlo, aus dem speziell der in dieser Arbeit betrachtete $D^\star$-Zerfall extrahiert wurde (siehe Kapitel 5.1).
Liegt die Impulsrichtung eines Teilchens in einem Konus bestimmter Größe um das $D^\star$-Meson, so wird die Annahme gemacht, daß dieses Teilchen eine Spur in diesem Konus hinterläßt. Dabei wird die Krümmung dieser Spuren durch das Magnetfeld nicht berücksichtigt. Wenn daher von Spurzahl in einem Konus die Rede ist, ist die Anzahl der Teilchenimpulse innerhalb dieses Konus gemeint.
Die Teilchendichte zeigt keine Unterschiede bei Betrachtung der Vorwärtsregion und des Zentralbereiches. Als Vorwärtsregion wurde $1.01 < \eta(D^\star) < 1.73$, als Zentralregion $-0.17 < \eta(D^\star) < 0.17$ gewählt. Die Verteilung der Spuren innerhalb eines Konus $\Delta r < 0.3$ und $\Delta r < 1.0$ zeigt Abbildung 7.1. Es war auch zu erwarten, daß die Vertei-

lungen annähernd gleiches Ausehen zeigen, da dies eine Folgerung der Verwendung der Pseudorapidität η in der Teilchenphysik ist. Somit kann zur Untersuchung von Spurmultiplizität in einem Konus bei der Simulation der Kinematik die Flugrichtung des $D^{\star}$ auf $\eta = 0$ gesetzt werden.

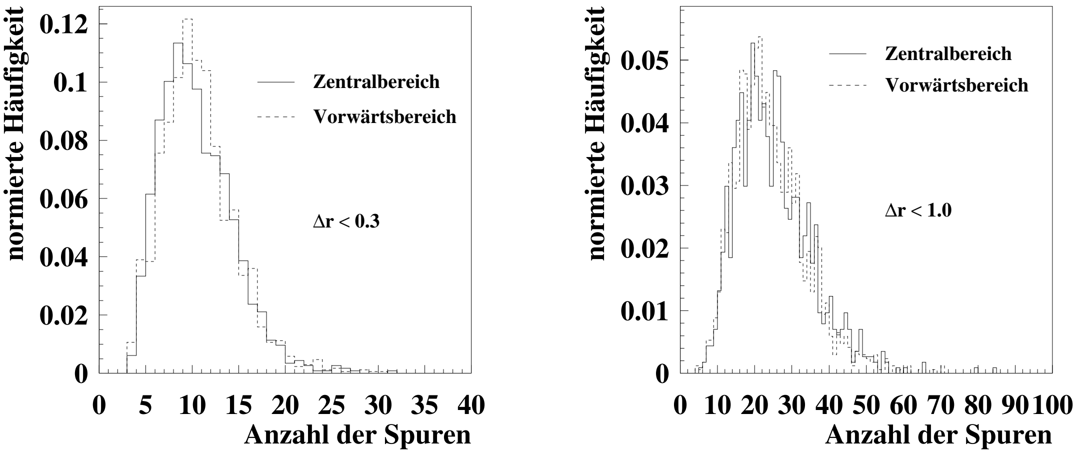


Abbildung 7.1: *Dargestellt ist die Spurdichte um das* $D^{\star}$*-Meson im Zentralbereich und im Vorwärtsbereich für zwei verschiedene Konusgrößen* Δr.

Zur Bestimmung der transversalen Zerfallslänge ist es notwendig, daß die Spuren der Teilchen, die an einem sekundären Vertex entstehen, innerhalb des Konus zu finden sind. Bei einem kleinen Konus um die Impulsrichtung des $D^{\star}$-Mesons sollte die Wahrscheinlichkeit, daß diese Teilchenspuren innerhalb des Konus liegen geringer sein als bei Betrachtung eines großen Konus. Im vorliegenden Fall sind diese Teilchenspuren, wie schon in den vorangegangenen Kapiteln erwähnt, die des Kaons und des Pions, die aus dem D^0 entstehen. Die Effizienz diese beiden Teilchenspuren in einem Konus $\Delta r < 0.3$ zu finden, liegt bei etwa 85 %. Die Abhängigkeit dieser Effizienz von der Konusgröße ist in Abbildung 7.2 dargestellt.
Auf der anderen Seite ist es interssant, wieviele Teilchenspuren in einem Konus bestimmter Größe zu finden sind. Eine Verteilung der Spuranzahl bei unterschiedlicher maximaler Konusgröße zeigt Abbildung 7.3.
Die Mittelwerte dieser Verteilungen wurden benutzt, um daraus die Kurve zu ermitteln, welche die Zunahme der mittleren Spurzahl mit der Konusgröße zeigt (Abbildung 7.4). Der Konus $\Delta r < 1.0$ beinhaltet im Mittel 24 Spuren. Das bedeutet, daß das neuronale Netz bei einer Auslese eines Konus von dieser Größe Detektortreffpunkte von mehr als 20 Spuren bekommen würde. Wesentlich zum Erlernen der Position des sekundären Vertex sind jedoch nur zwei Spuren. Der Rest stellt keine Information dar. Im Gegenteil, er bedeutet eigentlich „Rauschen“ für das neuronale Netz. Ziel ist es als RoI eine Konusgröße zu wählen, die eine möglichst hohe Effizienz für zwei Spuren vom sekundären Vertex liefert und zugleich eine möglichst geringe Spuranzahl beinhaltet. Aus diesen Gründen wurde für die nachfolgende Simulation $\Delta r < 0.3$ gewählt.

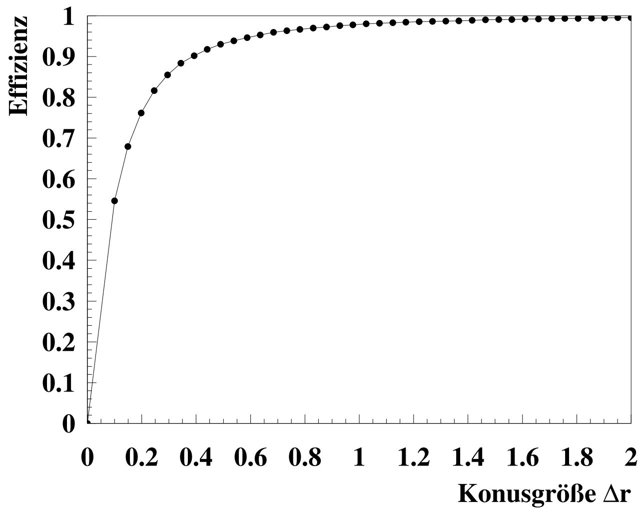
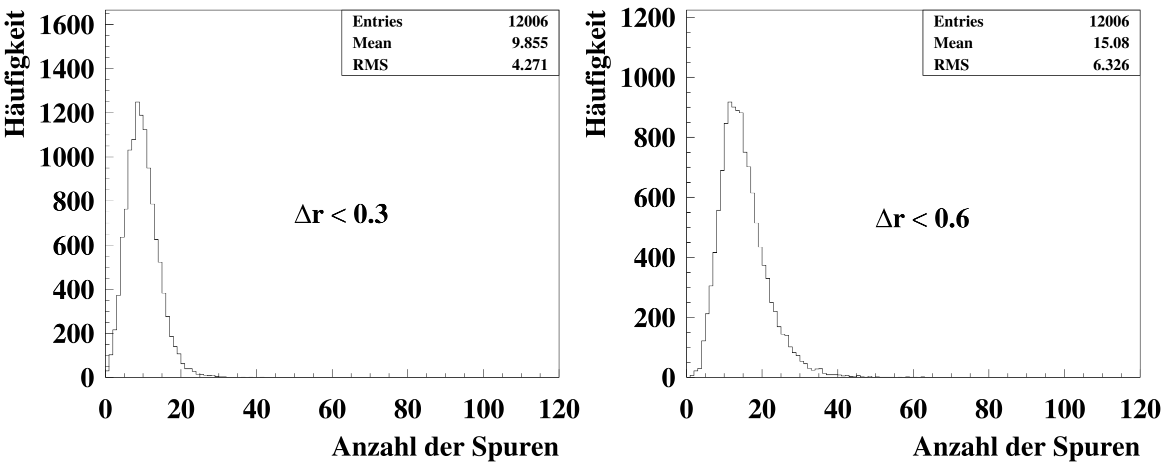


Abbildung 7.2: *Dargestellt ist die Effizienz, daß die beiden Spuren vom sekundären Vertex (Kaon und Pion) in einer variierenden Konusgröße* Δr *liegen.*



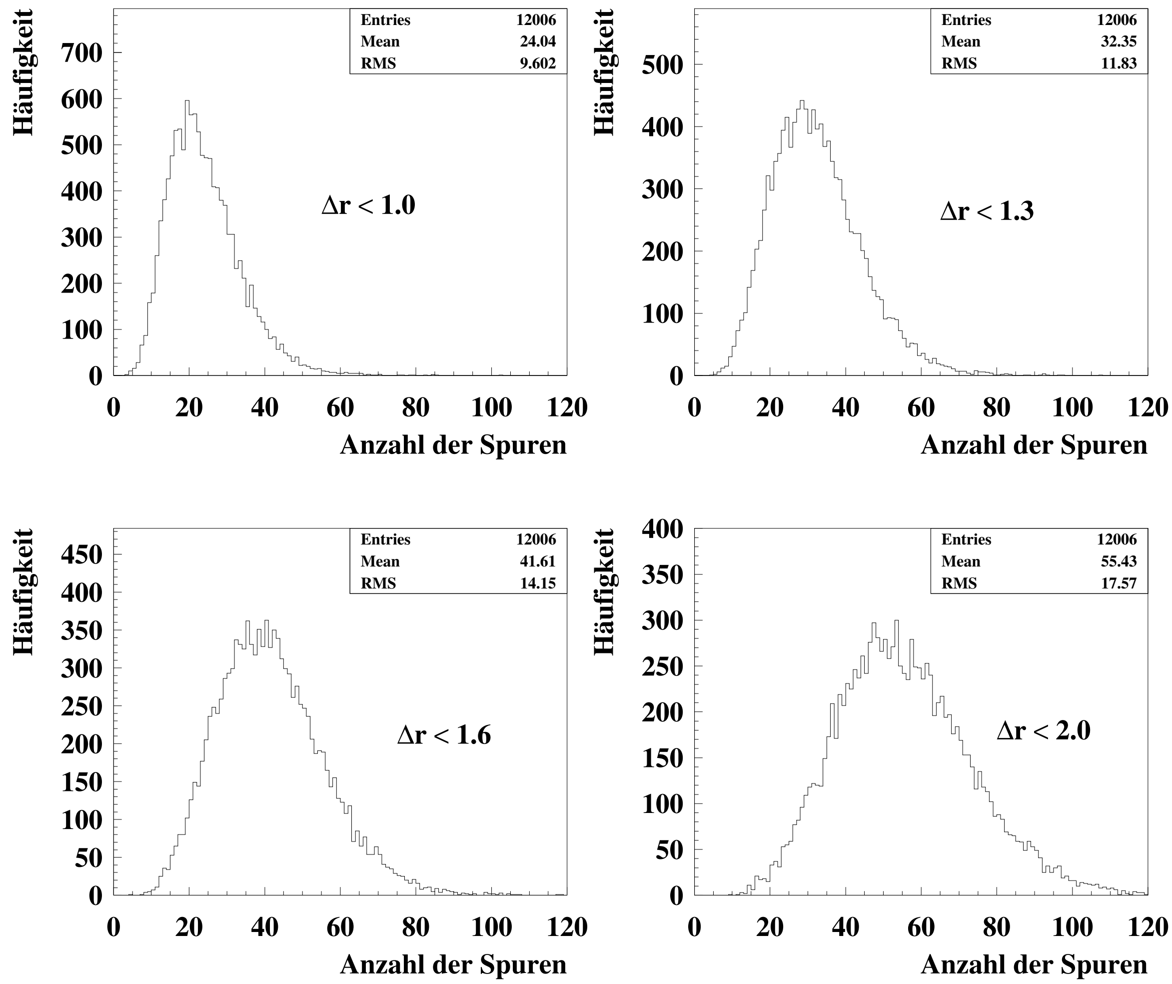


Abbildung 7.3: *Spurverteilungen bei unterschiedlicher Konusgröße* Δr.

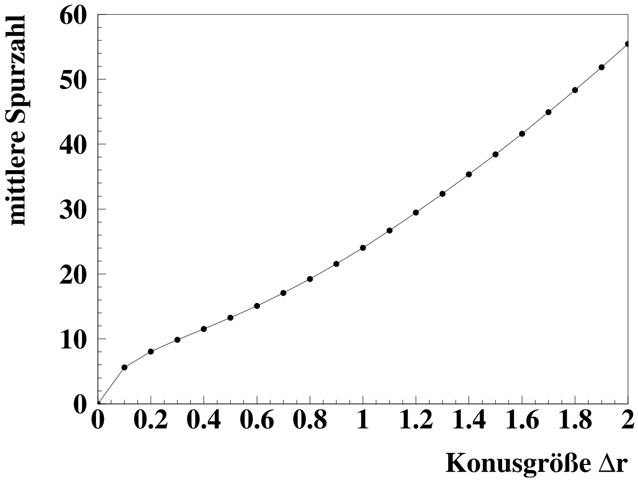


Abbildung 7.4: *Mittlere Spuranzahl bei steigender Konusgröße* Δr.

7.2 Simulation von Störspuren

Die Kinematik, die bisher verwendet wurde, ging von einer Gleichverteilung von $p_\mathrm{T}(\mathrm{D}^\star)$ aus (siehe Kapitel 5.2). Diese Gleichverteilung wird nun zugunsten einer realistischeren aufgehoben.
Die Verteilung aus dem $\mathrm{t\bar{t}}$-Monte-Carlo (durchgezogene Linie) und die vom eigenen Monte-Carlo simulierte Verteilung (gestrichelte Linie) zeigt Abbildung 7.5.

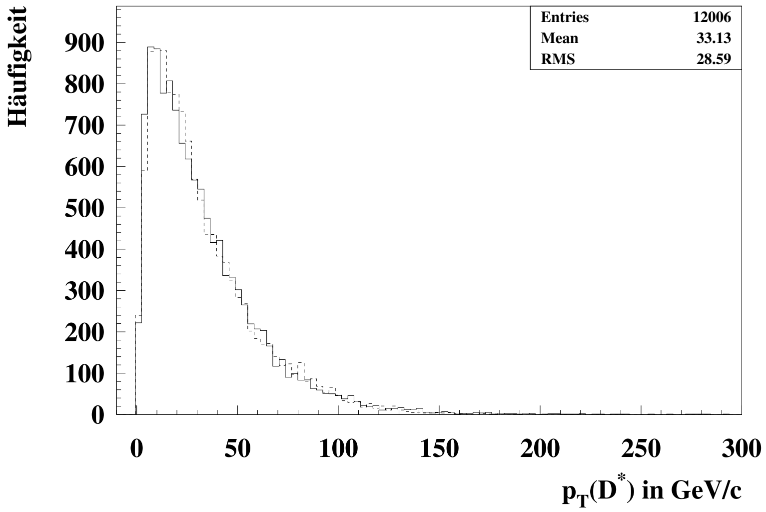


Abbildung 7.5: *Verleich des transversalen* $\mathrm{D}^\star$*-Impulses* $p_\mathrm{T}(\mathrm{D}^\star)$ *aus* $\mathrm{t\bar{t}}$*-Monte-Carlo (durchgezogene Linie) und eigenem Monte-Carlo (gestrichelte Linie).*

Die $\mathrm{D}^\star$-Mesonen werden nur in $\eta = 0$, also in der transversalen Ebene produziert (Begründung siehe Abschnitt 7.1) . Die selbstprogrammierte Kinematik liefert daraus die Impulse des Kaons, des Pions, und des *slow* Pions. Das *slow* Pion stellt im Prinzip schon ein Störteilchen dar, da es nicht am sekundären Vertex entsteht. Um zusätzliche Störspuren zu generieren, wurde die Verteilung der Transversalimpulse der Teilchen aus dem $\mathrm{t\bar{t}}$-Monte-Carlo nachgebildet, die in einem Konus der Größe $\Delta r < 0.3$ liegen und nicht vom $\mathrm{D}^\star$-Zerfall herrühren (siehe Abbildung 7.6).
Das *slow* Pion und ein zusätzlich simuliertes Teilchen haben ihren Entstehungsort an der Strahlachse. Dies stellt im Gegensatz zur Wirklichkeit eine gewisse Vereinfachung dar, da diese Teilchen sehr wohl an einem anderen Vertex produziert werden können. Allerdings entspricht dieser Vertex nicht dem des Kaons und des Pions. Diese vereinfachte Annahme spielt daher keine große Rolle. Die transversale Zerfallslänge (Abstand des Zerfallsortes des D^0 von der Strahlachse) wurde wie schon in Kapitel 5.2 gleichverteilt zwischen $0\ \mathrm{mm} < l_\mathrm{T}^\mathrm{Ziel} < 40\ \mathrm{mm}$ gewählt.

Es werden nun im wesentlichen zwei verschiedene Netzwerktrainings durchgeführt:

- Sowohl für das Training als auch für den Test des Netzwerks wurde die Forderung nach vollständiger Trefferzahl in einer RoI von $\Delta r < 0.3$ gestellt.

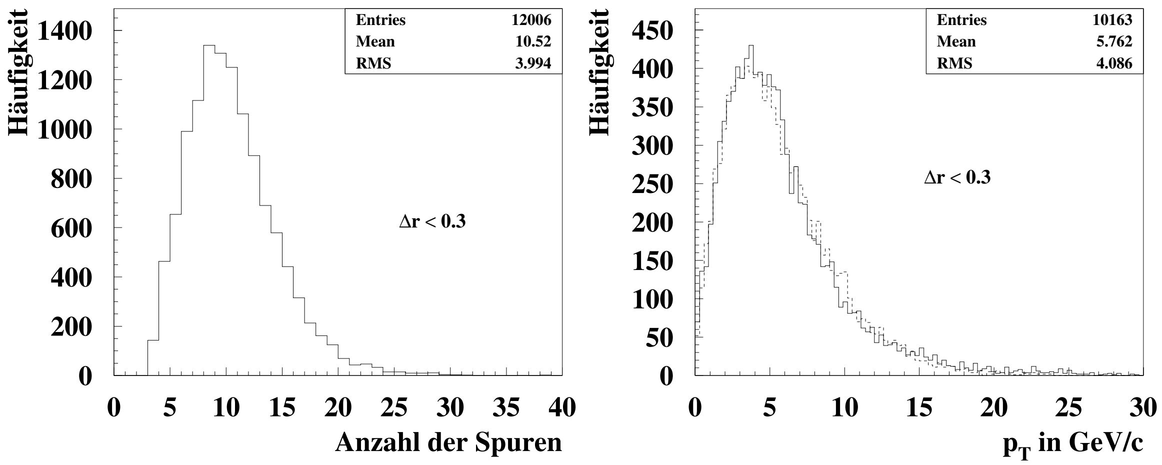


Abbildung 7.6: *Links: Anzahl der zusätzlichen Spuren in einem Konus* $\Delta \mathrm{r} < 0.3$. *Rechts:* p_{T}*-Verteilung dieser zusätzlichen Spuren. Die durchgezogene Linie bedeutet, daß sich die Darstellung auf das* $\mathrm{t\bar{t}}$*-Monte-Carlo bezieht, die gestrichelte Linie bedeutet die Darstellung mit dem eigenen Monte-Carlo.*

Dies wird erreicht, indem für jedes Teilchen und jeden Detektorradius überprüft wird, ob $\Delta\phi(\text{Teilchen}) < 0.21$ und $\Delta\eta(\text{Teilchen}) < 0.21$ ist. Somit kann annähernd der Forderung $\Delta r < 0.3$ nachgekommen und zugleich die Spurkrümmung durch das Magnetfeld berücksichtigt werden.

- Die Bedingung einer vollständigen Trefferzahl wurde nicht gestellt. Für die Trainingsdaten wurden lediglich mindestens drei Detektortrefferpunkte des Kaons und drei Detektortrefferpunkte des Pions gefordert. Diese Minimalforderung ist wichtig, da sie die Information zur Schätzung der transversalen Zerfallslänge enthält. Die Netzeingabe wurde dann auf die Zahl Null gesetzt, wenn die Detektortrefferpunkte außerhalb einer RoI von $\Delta r < 0.3$ liegen. Für die Testdaten wurde nur vorausgesetzt, daß es summiert über alle Teilchenspuren mindestens sechs Detektortrefferpunkte in dieser RoI gibt. Dies soll einer realistischen Annahme der Datenauslese entsprechen, bei der nicht klar ist, ob sich Teilchenspuren von einen sekundären Vertex in der RoI befinden. Bei den hier als Testdaten deklarierten Daten werden allerdings immer Trefferpunkte des Kaons und des Pions als Eingabe vorhanden sein, da die simulierte Kinematik interessante Ereignisse mit Sekundärvertices produziert.

Zu diesen zwei grundsätzlich unterschiedlichen Netzwerktrainings wurde nun jeweils die Anzahl der Teilchen in der RoI variiert:

- zwei Teilchen: das Kaon und das Pion, die am sekundären Vertex entstehen $\Rightarrow$ 14 Eingabegrößen.

- drei Teilchen: zusätzlich wurde das *slow* Pion als Eingabe verwendet $\Rightarrow$ 21 Eingabegrößen.
- vier Teilchen: zusätzlich wurde ein fiktives Pion mit der in Abbildung 7.6 dargestellten Impulsverteilung simuliert $\Rightarrow$ 28 Eingabegrößen.

Die Detektorauslese erfolgte wie im Kapitel zuvor radienweise.

7.3 Ergebnisse

Um eine grobe Fehlerabschätzung zu erhalten, wurden zu jedem der sechs verschiedenen Eingabearten jeweils drei Trainings und Tests durchgeführt. Erwartungsgemäß liegt der mittlere quadratische Fehler RMS, der sich bei Betrachtung der Abweichungen von Netzausgabe und Zielwert ergibt, für ein Netztraining mit vollständiger Trefferzahl deutlich unter dem für unvollständige Trefferzahl. Des weiteren werden die Netzergebnisse jeweils mit steigender Zahl an Störspuren (*slow* Pion, fiktives Pion) schlechter (siehe Abbildung 7.7).

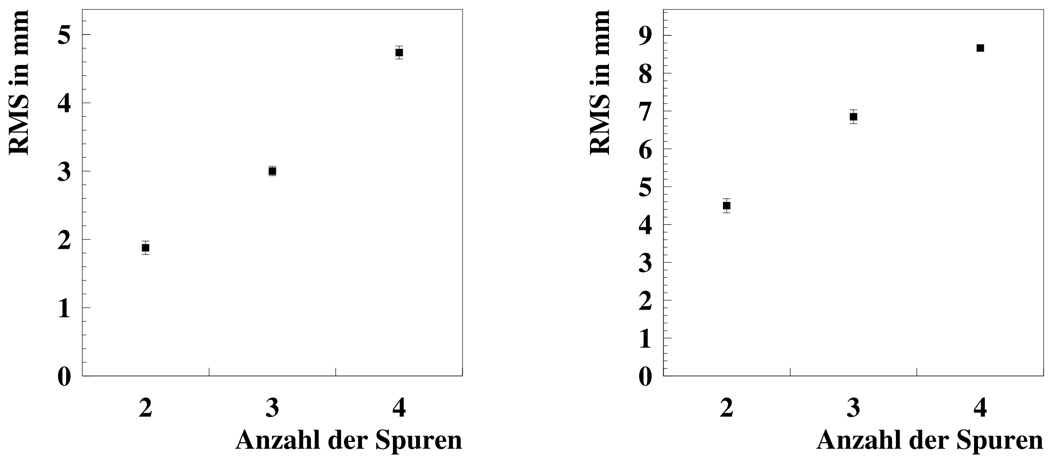


Abbildung 7.7: *Anzahl der Spuren bedeutet die Gesamtanzahl der Spuren in der RoI. Vier Spuren bedeutet zwei Störspuren. Links: Netzqualität ausgedrückt durch den RMS für variierende Teilchenzahl und der Vorraussetzung von vollständiger Trefferzahl in einer RoI. Rechts: Gleiche Darstellung für unvollständige Trefferzahl.*

Es ist bemerkenswert, daß ein neuronales Netz bei Eingabe von zwei Störspuren, was einem Rauschen von 50% entspricht, die transversale Zerfallslänge immer noch mit einer Genauigkeit von RMS = 4.74 ± 0.09 mm auflösen kann. Läßt man die Forderung nach vollständiger Trefferzahl außer acht, so bedeutet dies weit mehr als 50 % Rauschanteil. Das neuronale Netz zeigt auch hier ein Lernverhalten ! Allerdings hat es keinen Sinn noch weitere Störspuren zu simulieren und Netztrainings zu machen, da mit einer weiteren Verschlechterung zu rechnen ist.

Vielmehr interessiert, wie effizient ein neuronales Netz Physikereignisse erkennen und Untergrundereignisse verwerfen kann. Als Klasse der Physikereignisse werden Ereignisse mit einer transversalen Zerfallslänge von 0 mm $< l_{\mathrm{T}}^{\mathrm{Ziel}} <$ 40 mm bezeichnet. Die Klasse Untergrundereignisse sind Daten, die mit einer Zerfallslänge von $l_{\mathrm{T}}^{\mathrm{Ziel}} = 0$ mm generiert werden.
Die Effizienzuntersuchungen werden für einen „einfachen“ und einen „schwierigen“ Fall durchgeführt. „Einfacher“ Fall bedeutet Betrachtung von zwei Spuren mit vollständiger Trefferzahl, „schwieriger“ Fall heißt Betrachtung von vier Spuren mit unvollständiger Trefferzahl.
Zunächst zum einfachen Fall. In Abbildung 7.8 sieht man wie das neuronale Netz auf die Klasse Physikereignisse reagiert. Im rechten Bild ist die Differenz von Netzausgabe und gefordertem Wert dargestellt. Die Erklärung der Physikeffizienz wird mit Hilfe des linken Bildes durchgeführt.

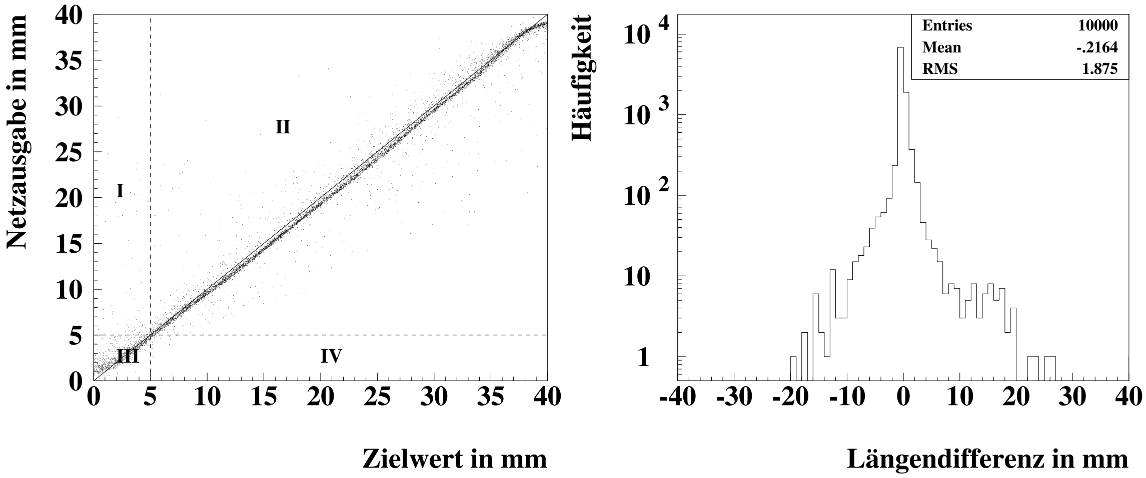


Abbildung 7.8: *Links: Netzausgabe gegen Zielwert. Rechts: Differenz aus beiden Werten.*

Sobald das neuronale Netz als Ausgabe einen Wert $l_{\mathrm{T}}^{\mathrm{Netz}} > 5$ mm liefert, soll dieses Ereignis vom Triggersystem als weiterverarbeitungswürdig behandelt werden. Damit würden alle Ereignisse in den Bereichen I und II getriggert. Wünschenswert wären jedoch die Bereiche II und IV, da hier der Zielwert $l_{\mathrm{T}}^{\mathrm{Ziel}}$ größer als 5 mm ist. Der Bereich IV geht folglich mit einem solchen Trigger verloren. Damit ist die Physikeffizienz bestimmt als das Verhältnis der Ereignisse aus dem Bereich II zu den Bereichen II + IV. Diese Effizienzbestimmung kann für variierende Schnitte („Cuts“) auf der transversalen Zerfallslänge durchgeführt werden. Im linken Bild der Abbildung 7.9 sieht man die Verteilung des Zielwertes und der Netzausgabe für die Physikklasse.
Nun zum Begriff der Untergrundreduktion. Der Trigger soll möglichst den gesamten Untergrund verwerfen. Dem neuronalen Netz werden als Eingaben nur Untergrundereignisse mit einer Zerfallslänge von 0 mm präsentiert. Optimal wäre eine Netzausgabe, die ebenfalls immer 0 mm als Ausgabe produziert. Die erhaltene Netzantwort ist rechts in Abbildung 7.9 dargestellt. Die Untergrundreduktion errechnet sich aus dem Verhältnis der Ereignisse, die unterhalb einem gewissen Cut auf der

Netzausgabe liegen, zu den gesamten Ereignissen.

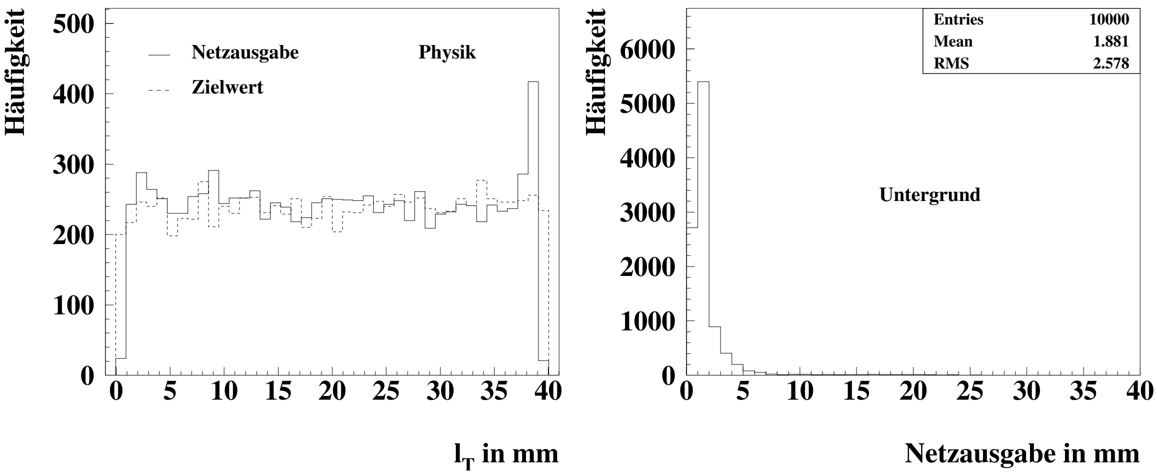


Abbildung 7.9: *In der linken Abbildung sieht man die Verteilung der Netzausgabe (durchgezogene Linie) und die Verteilung des geforderten Wertes (gestrichelte Linie) für die Physikklasse. Die rechte Abbildung zeigt die Netzantwort $l_{\mathrm{T}}^{\mathrm{Netz}}$ auf den Untergrund (transversale Zerfallslänge $l_{\mathrm{T}}^{\mathrm{Ziel}} = 0$ mm).*

Aus diesen Überlegungen kann man die Physikeffizienz über der Untergrundunterdrückung bei variierendem Cut auf der transversalen Zerfallslänge darstellen (siehe Abbildung 7.10).
Macht man einen Cut bei einer Netzausgabe von 5 mm, so können etwa 96% des Untergrunds bei einer gleichzeitigen Physikeffizienz von etwa 98% unterdrückt werden. Auch bei einem Cut von 3 mm erhält man noch eine Untergrundreduktion von etwa 90%.
Die Diagramme 7.11 zeigen die Physikeffizienz bei zwei verschiedenen Cuts auf der transversalen Zerfallslänge. Die Effizienz für Ereignisse unterhalb eines Cuts, bedeutet eine Wahrscheinlichkeit, daß solche Ereignisse mit $l_{\mathrm{T}}^{\mathrm{Ziel}} <$ Cut mitgetriggert werden. Diese Ereignisse stellen im Prinzip Physik dar, da Untergrund über $l_{\mathrm{T}}^{\mathrm{Ziel}} = 0$ mm definiert ist. Die Breite der Stufe spiegelt gewissermaßen den RMS aus Abbildung 7.8 wider.
Auf die gleiche Art und Weise wurden für den schwierigen Fall Physikeffizienz und Untergrundunterdrückung bestimmt. Die Qualität des Netzes beim Erkennen von Physikereignissen und Untergrundereignissen ist in den Abbildungen 7.12 und 7.13 dargestellt.
Um eine Untergrundunterdrückung von 80% zu erhalten, mußte man hier einen Cut bei einer Netzausgabe von 17 mm machen. Dadurch erhielt man etwas mehr als 80% Physikeffizienz (siehe Abbildung 7.14). Die mit dem einfachen Fall vergleichbare Untergrundreduktion von 96% würde einen Cut auf der Netzausgabe bei 30 mm fordern. Die Physikeffizienz läge dann bei etwas mehr als 65%. Ein Cut auf einer Netzausgabe von 5 mm lieferte nur noch eine Untergrundreduktion von 46% (siehe Abbildung 7.15). Die Physikeffizienz lag dann immer noch bei etwas mehr als 90%.

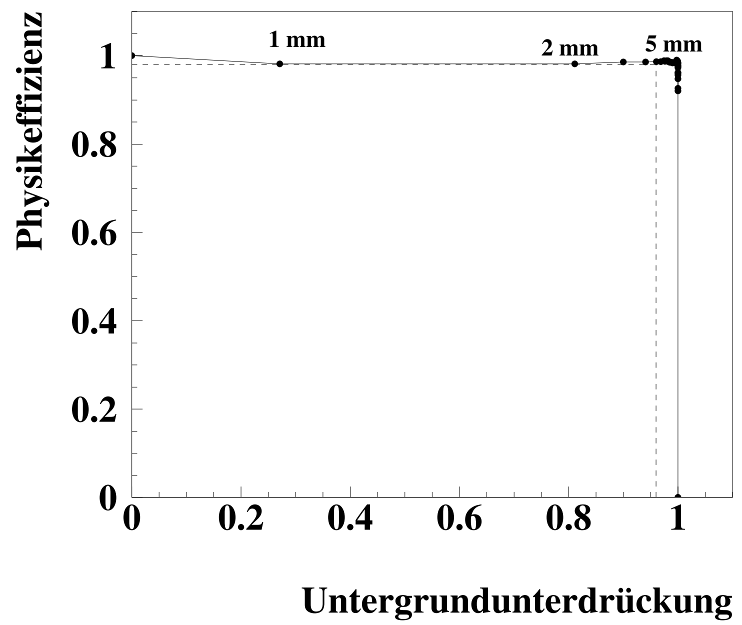


Abbildung 7.10: *Darstellung von Physikeffizienz zu Untergrundunterdrückung bei variierendem Cut auf* $l_{\mathrm{T}}^{\mathrm{Netz}}$.

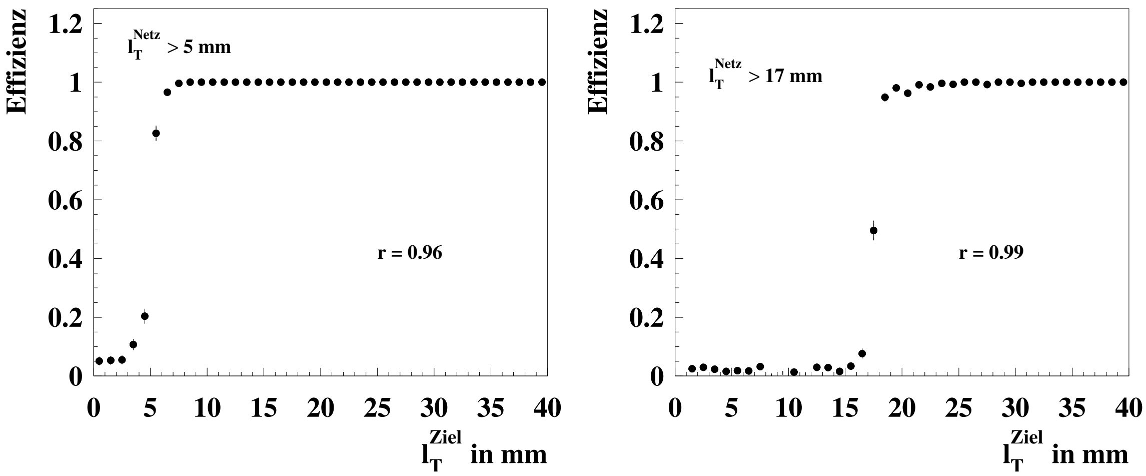


Abbildung 7.11: *Dargestellt wird die Physikeffizienz bei zwei verschiedenen Cuts (links:* $l_{\mathrm{T}}^{\mathrm{Netz}} > 5$ *mm, rechts:* $l_{\mathrm{T}}^{\mathrm{Netz}} > 17$ *mm). Mit* r *wird die zugehörige Untergrundreduktion bezeichnet.*

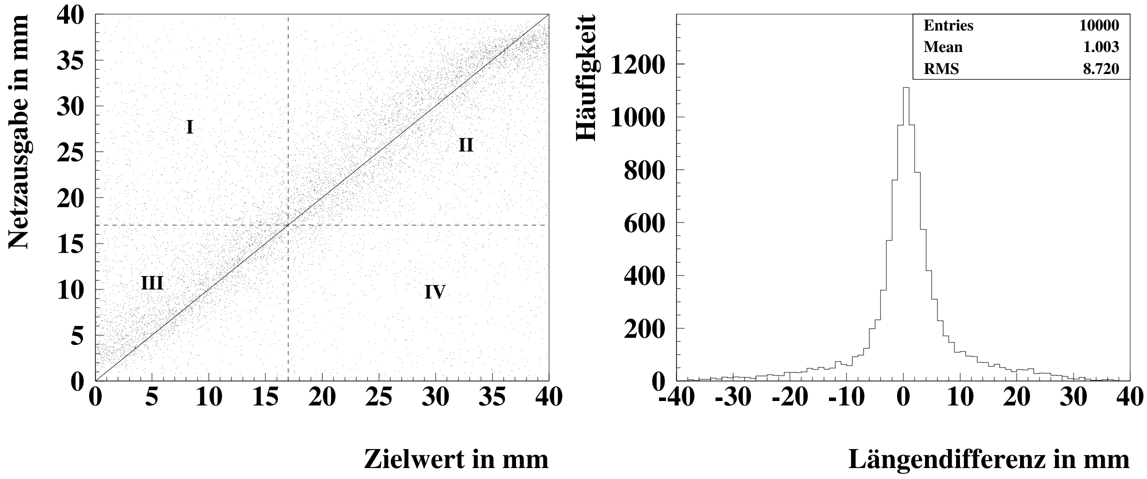


Abbildung 7.12: *Schwieriger Fall: Links: Netzausgabe gegen Zielwert. Rechts: Differenz aus beiden Werten.*

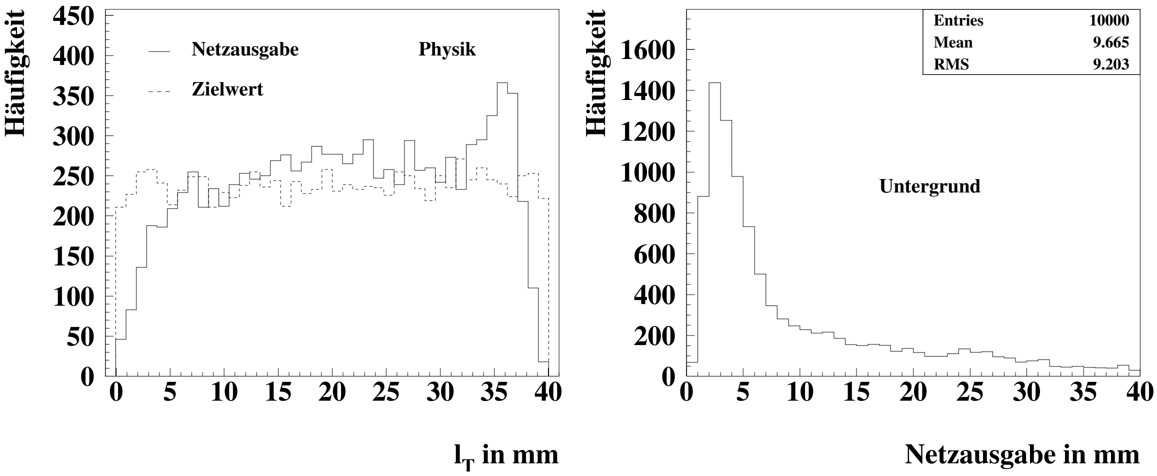


Abbildung 7.13: *Schwieriger Fall: In der linken Abbildung sieht man die Verteilung der Netzausgabe (durchgezogene Linie) und die Verteilung des geforderten Wertes (gestrichelte Linie) für die Physikklasse. Die rechte Abbildung zeigt die Netzantwort $l_{\mathrm{T}}^{\mathrm{Netz}}$ auf den Untergrund (transversale Zerfallslänge $l_{\mathrm{T}}^{\mathrm{Ziel}}$ = 0 mm).*

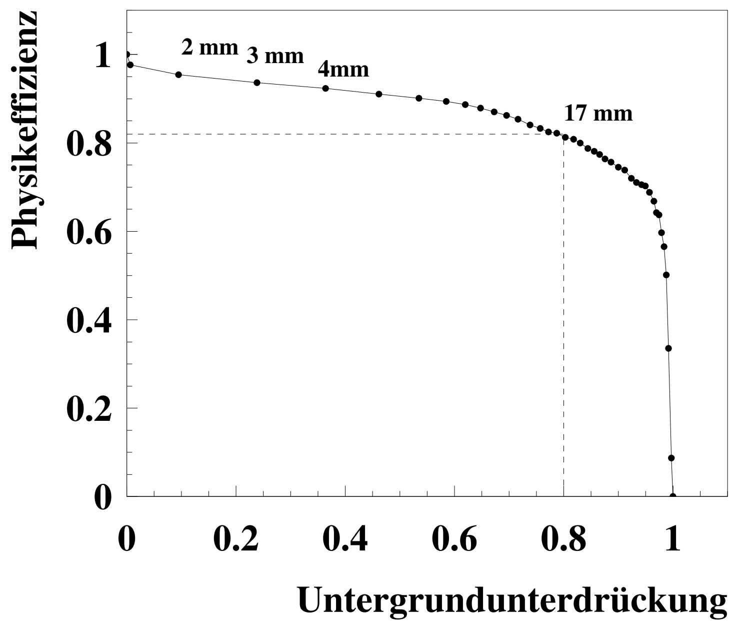


Abbildung 7.14: *Schwieriger Fall: Darstellung von Physikeffizienz zu Untergrundunterdrückung bei variierendem Cut auf* $l_{\mathrm{T}}^{\mathrm{Netz}}$.

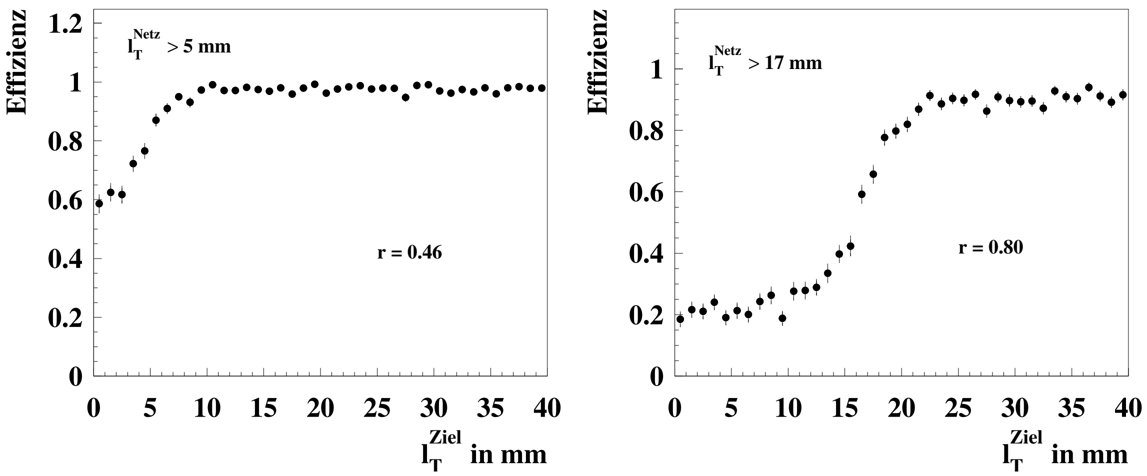


Abbildung 7.15: *Schwieriger Fall: Dargestellt wird die Physikeffizienz bei zwei verschiedenen Cuts (links:* $l_{\mathrm{T}}^{\mathrm{Netz}} > 5$ *mm, rechts:* $l_{\mathrm{T}}^{\mathrm{Netz}} > 17$ *mm). Mit* r *wird die zugehörige Untergrundreduktion bezeichnet.*

Die durchgeführten Studien zeigten, daß mit zunehmender Anzahl von Störspuren die Lernerfolge des neuronalen Netzes kontinuierlich abnehmen. Bei zwei Störspuren und realitätsnaher Auslesemethode beträgt der RMS bereits mehr als 1/5 des gesamten angebotenen Wertebereichs 0 mm $< l_{\mathrm{T}}^{\mathrm{Ziel}} <$ 40 mm. Eine weitere Erhöhung der Anzahl von Störspuren ist bei derzeit realisierbaren Netzwerkarchitekturen auf dem CNAPS-Parallelcomputer nicht mehr sinnvoll.

Kapitel 8

Zusammenfassung und Ausblick

In dieser Arbeit wurde diskutiert, ob man aus Detektortrefferpunkten von geladenen Teilchen auf das Auftreten sekundärer Vertices in einem Ereignis schließen kann.
Dabei soll die Erkennung von sekundären Vertices bereits auf Triggerebene geschehen. Für diese Aufgabe wurden neuronale Netze, wegen ihrer hochgradigen Parallelität, verwendet. Die dafür benötigten Trainings- und Testdaten wurden mit Hilfe zweier selbsterstellter Programme ($D^{\star}$-Monte-Carlo und einfache Simulation des Barrelbereichs des inneren ATLAS-Detektors) erzeugt.
Mit den durchgeführten Untersuchungen konnte gezeigt werden, daß neuronale Netze prinzipiell in der Lage sind, aus Detektortrefferpunkten transversale Zerfallslängen zu schätzen.
Das dabei verwendete neuronale Netz war ein Feedforward-Netz, der Trainings-Algorithmus der Backpropagation-Algorithmus (BPG). Gearbeitet wurde mit dem Neurocomputer CNAPS, einem Parallelrechner, der sich gegenüber den sequentiell arbeitenden Rechnern, wie etwa einem Pentium, durch eine erheblich größere Rechengeschwindigkeit auszeichnet.
Für erste Untersuchungen wurden als Netzeingabe nur Ereignisse, die eine vollständige Anzahl der Detektortrefferpunkte in einer „Region of Interest“ (RoI) haben, verwendet. Bei diesen Ereignissen wurden nur die beiden Spuren, die an einem sekundären Vertex entstehen, zum Training benutzt.
Es konnte gezeigt werden, daß die Lernqualität des neuronalen Netzes neben der Lernrate η auch empfindlich von der benutzten Netzarchitektur abhängt. Vierlagige Netze erreichten im Vergleich zu dreilagigen Netzen einen um nahezu 50% kleineren mittleren quadratischen Fehler.
Eine weitere wichtige Erkenntnis war, daß es, um gute Ergebnisse zu erhalten, nicht genügt, lediglich die Zellennummern der von geladenen Teilchen durchquerten Detektorzellen als Netzeingabe zu benutzen. Vielmehr sollte die Eingabe so transformiert werden, daß der numerische Wertebereich des CNAPS-Neurocomputers vollständig ausgenutzt wird. Die Standardabweichung, die sich bei Betrachtung der Differenz aus Netzausgabe und gefordertem Wert ergab, konnte somit nochmals um fast 50% verbessert werden. Bei einer geforderten gleichverteilten transversalen Zerfallslänge von 0 mm $< l_{\mathrm{T}} <$ 40 mm, betrug die Standardabweichung weniger als 2 mm.
Einen Vergleich mit einem einfachen Algorithmus, der aus Detektortrefferpunkten gewonnene Geraden zum Schnitt bringt, um so die transversale Zerfallslänge zu er-

kennen, gewann das neuronale Netz deutlich. Ein einfacher Kreisschneidealgorithmus konnte die Netzleistung ebenfalls nicht unterbieten.
Auf Veränderungen der Detektorgeometrie (Auflösung, Anzahl der verwendeten Detektorlagen, Hebelarm) reagierte das neuronale Netz äußerst robust. Nur bei einer Verschlechterung der Detektorauflösung um einen Faktor 100 fiel die Netzleistung etwas ab.
Eine realistischere Studie zeigte, daß es dem neuronalen Netz auch noch möglich war, die Position eines Sekundärvertex zu schätzen, wenn man zusätzlich zu den zwei Spuren, die am Sekundärvertex entstehen, zwei weitere Spuren als Eingabe verwendete. Da diese Spuren nicht am sekundären Vertex entstehen, bedeuten sie im Prinzip 50% „Rauschen“ für das neuronale Netz. Die Standardabweichung lag für diesen Fall bei knapp 5 mm. Wurde außerdem eine unvollständige Trefferzahl erlaubt, so lag die Standardabweichung bei fast 9 mm. Ein Lernverhalten des neuronalen Netzes war aber auch hier erkennbar.
Bezeichnet man Ereignisse mit gleichverteilten transversalen Zerfallslängen zwischen $0 \text{ mm} < l_T < 40$ mm als Klasse der Physikereignisse und Ereignisse mit einer transversalen Zerfallslänge von 0 mm als Klasse der Untergrundereignisse, so konnte man damit eine Bestimmung der Physikeffizienz und der Untergrundunterdrückung durchführen. Erlaubte man einen Cut auf einer Netzausgabe von 5mm, so erreichte man für einen „einfachen“ Fall (zwei Spuren vom sekundären Vertex mit gleichzeitiger Forderung einer vollständigen Trefferzahl in der RoI) eine Physikeffizienz von etwa 98% bei einer Untergrundunterdrückung von ca. 96%. Für eine kompliziertere Situation (zwei zusätzliche Spuren und Aufhebung der Forderung nach vollständiger Trefferzahl in der RoI) errechnete sich eine Physikeffizienz von etwas mehr als 90% bei einer Untergrundunterdrückung von ca. 46%.
Damit ein neuronales Netz effektiv als selektiver Trigger von Sekundärvertices benutzt werden kann, sind jedoch weiterführende Untersuchungen nötig.
Die im Rahmen dieser Arbeit erzielten Ergebnisse basieren alle auf dem CNAPS-Parallelrechner. Dieser bringt aber einige Limitierungen mit sich. Zum einen arbeitet er mit Festkomma-Arithmetik, günstiger wäre jedoch die Verwendung von Fließkomma-Arithmetik. Zum anderen können mit dem CNAPS höchstens vierlagige Netze untersucht werden. Da sich schon eine Verbesserung des Ergebnisses von vierlagigen Netzen gegenüber dreilagigen Netzen abzeichnete, liegt die Vermutung nahe, daß durch die Verwendung von noch mehr versteckten Lagen weitere Steigerungen der Netzqualität möglich sein sollten.
Des weiteren wäre eine umfassende Datenvorverarbeitung sinnvoll. Es ist eventuell an eine Spurfindung („Tracking“) zu denken. Hier gibt es bereits Ansätze auf der Basis neuronaler Netze: Spurfindung mit dem „Elastic Arms“-Algorithmus [49] und mit einem „Growing Neural Gas“-Algorithmus [50]. Um in der Online-Analyse Anwendung zu finden, müssen diese Algorithmen jedoch parallelisiert werden. Dies ist bei Betrachtung der Leistungssteigerung von Prozessoren in den vergangenen Jahren kein unmöglicher Schritt, zumal bis zur Inbetriebnahme des LHC noch fast sieben Jahre Entwicklungszeit zur Verfügung stehen.
Zusammenfassend kann gesagt werden, daß neuronale Netze ein vielversprechender Weg in Richtung Sekundärvertexerkennung auf Triggerebene zu sein scheinen.

Literaturverzeichnis

[1] H. Vogel. Gerthsen Physik. Springer Verlag, Berlin, Heidelberg, 18. Aufl., 1995.

[2] O. Nachtmann. Elementarteilchenphysik, Phänomene und Konzepte. Vieweg Verlag, Braunschweig, 1991.

[3] G. Musiol et al. Kern- und Elementarteilchenphysik. VCH Verlagsgesellschaft, Weinheim, 1988.

[4] Particle Data Group. Particles and Fields. *Physical Review D*, **54**, 1. July 1996.

[5] P. Lefevre et al. The Large Hadron Collider, Conceptual Design. *CERN/AC/95-05(LHC)*, 1995.

[6] ATLAS Collaboration. Letter of Intent by the ATLAS Collaboration for a General-Purpose pp Experiment at the LHC. *CERN/LHCC/92-4LHCC/I2*, 1. Oktober 1992.

[7] G. Jarlskog D. Rein. Proceedings of the Large Hadron Collider Workshop, Aachen. *CERN 90-10*, **Vol. I, II, III**, 4-9 Oktober 1990.

[8] ATLAS Collaboration. Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN. *CERN/LHCC/94-43LHCC/P2*, 15. Dezember 1994.

[9] W. S. C. Williams. Nuclear and Particle Physics. Oxford University Press, 1994.

[10] I. J. R. Aitchison und A. J. G. Hey. Gauge Theories in Particle Physics. *Graduate student series in physics*. Adam Hilger, Bristol and Philadelphia, second. Aufl., 1989.

[11] D. H. Perkins. Hochenergiephysik. Addison Wesley Publishing Company, Bonn, 1990.

[12] The Super-Kamiokande Collaboration. Evidence for Oszillation of atmospheric neutrinos. 1998. Vorgestellt auf: The 18th international conference on neutrino physics and astrophysics.

[13] F. Abe et al., CDF Collaboration. Evidence for Top Quark Production in $\bar{p}p$ Collisions at $\sqrt{s} = 1.8$ TeV. *FERMILAB-PUB-94/097-E, CDF/PUB/TOP/PUBLIC/2561*, 1994.

[14] D0 Collaboration F. Abachi et al. ,. Observation of the Top Quark. *Phys. Rev. Lett.*, **74**, pp. 2632–2637, 1995.

[15] M. Mangano T. Trippe. The Top Quark. Veröffentlicht in [4].

[16] E. Reya et al. Top Physics. Veröffentlicht in [7].

[17] M. Cobal. Top Physics. Atlas Internal Note PHYS-No-93, 1996.

[18] A. Blondel. Precision Electroweak Physics at LEP. *CERN-PPE/94-133*, August 1994.

[19] A. Sopczak. Status of the Higgs Hunting at the Z Resonance and its Prospects at LEP2. *CERN-PPE/94-73*, 1994.

[20] D. Denegri. Standard Model Physics at the LHC (pp Collisions). Veröffentlicht in [7].

[21] I. Hinchchliffe LBNL. The Higgs Boson. Veröffentlicht in [4].

[22] F. Zwirner Z. Kunszt. Testing the Higgs Sector of the Minimal Supersymmetric Standard Model at LHC. Veröffentlicht in [7].

[23] J.-P. Gourber. Status of the LHC. Technischer Report LHC Project Report 167, CERN, 1997.

[24] CMS Collaboration. Letter of Intent by the CMS Collaboration for a General Purpose Detector at LHC. *CERN/LHCC/92-3 LHCC/I1*, 1. Oktober 1992.

[25] CMS Collaboration. The Compact Muon Solenoid Technical Proposal. *CERN/LHCC/94-38 LHCC/P1*, 15. Dezember 1994.

[26] LHC-B Collaboration. Letter of Intent: A Dedicated LHC Collider Beauty Experiment for Precision Measurements of CP-Violation. *CERN/LHCC95-5 LHCC/I8*, 25. August 1995.

[27] ALICE Collaboration. Technical Proposal for A Large Ion Collider Experiment at the CERN LHC. *CERN/LHCC/95-71 LHCC/P3*, 15. Dezember 1995.

[28] J. C. Eccles. The Understanding of the Brain. McGraw-Hill, New York, 1977.

[29] J. W. Cohen. Funktionelle Anatomie des Nervensystems. Schattauer, Stuttgart/New York, 4. Aufl., 1985.

[30] Spektrum der Wissenschaft. Gehirn und Nervensysteme. Spektrum der Wissenschaft Verlag, Heidelberg, 10. Aufl., 1992.

[31] W. S. McCulloch und W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Math. Bio.*, **5**, pp. 115–133, 1943.

[32] Andreas Zell. Simulation Neuronaler Netze. Addison–Wesley, Reading, Massachusetts, 1994.

[33] F. Rosenblatt. The Perceptron: A Probabilistic Model for Inforamtion Storage and Organisation in the Brain. *Psychological Review*, **Vol.65**, 1958.

[34] S. Papert M. Minsky. Perceptrons: An Introduction to Computional Geometry. MIT Press, Cambridge, 1969.

[35] A. N. Kolmogorov. On the Representation of Continous Functions of several Variab les by Superposition of One Variable and Addition. *Dokl.Akad.Nauk. USSR*, **115**, pp. 953–956, 1957.

[36] R. Hecht-Nielsen. In *Proc. of the Int. Conf. on Neural Networks*, volume II, pp. 19–32. IEEE press, 1987.

[37] Hecht-Nielsen. Neurocomputing. Addison-Wesley, 1991.

[38] P. J. Werbos. Beyond Regression. Appl. math. , Harvard University, November 1974.

[39] D. E. Rumelhart und J. L. McClelland. Parallel Distributed Processing : Foundations, volume 1. MIT Press Cambridge, 1986. Im wesentlichen p318-362.

[40] Adaptive Solutions Inc., Beaverton, OR 97006. CNAPS Data Book, 2.0 Version, 1995.

[41] Adaptive Solutions Inc., Beaverton, OR 97006. CNAPS Back-Propagation Guide, 1.0 Version, 1992.

[42] Torbjörn Sjöstrand. PYTHIA 5.7 and JETSET 7.4: Physics and Manual. 1995. hep-ph/9508391.

[43] Markus Enzenberger. Persönliche Mitteilung, 1997.

[44] Adaptive Solutions Inc., Beaverton, OR 97006. CNAPS/VME Board Reference Manual, 1.0 Version, 1993.

[45] S. Udluft. Training und Einsatz neuronaler Netze auf dem CNAPS-Parallelcomputer. http://wwwh1.mppmu.mpg.de/projects/neuro/doc/bpexp2, 1995.

[46] S. Udluft. Untersuchungen zu Neuronalen Netzen als Vertextrigger im H1-Experiment bei HERA. Diplomarbeit, LMU München, 1996.

[47] R. L. Glückstern. Uncertainties in track momentum and direction, due to multiple scattering and measurement errors. *Nucl. Instrum. Meth.*, **24**, pp. 381–389, 1963.

[48] K. Kleinknecht. Detektoren für Teilchenstrahlung. Teubner Taschenbücher Physik, Stuttgart, 1984.

[49] C. Peterson. Neural Networks in High Energy Physics - Algorithms and Results. Veröffentlicht in [51].

[50] M. Kunze J. Steffens. Growing Cell Structure and Neural Gas. Veröffentlicht auf: Proceedings of the 4th AIHEP Workshop, Pisa, World Scientific, 1995.

[51] D. Perret-Gallix. New Computing Techniques in Physics Research II. World Scientific Publishing Co. Pte. Ltd., Singapur, 1992.

Danksagung

Diese Arbeit wurde am Max-Planck-Institut für Physik in München angefertigt. Ich möchte all denjenigen danken, die durch ihre Hilfe zum Entstehen dieser Diplomarbeit beigetragen haben.

An erster Stelle namentlich zu nennen ist Prof. Dr. Christian Kiesling. Er vertraute mir eine reizvolle und auch herausfordernde Aufgabe an. In zahlreichen Fachgesprächen, nicht nur physikalischer Natur, erwies er sich als wertvoller Ansprechpartner.

Mein herzlicher Dank gilt auch Philip Scheideler für die gute und erfolgreiche Zusammenarbeit.

Eine große Hilfe bei physikalischen Fragestellungen und der Datenauswertung waren Dr. Armin Wegner, Burkard Reisert, Steffen Udluft und Markus Enzenberger.

Der gesamten H1-Gruppe sei Dank für die freundliche und harmonische Arbeitsatmosphäre.

Der Max-Planck-Gesellschaft, insbesondere Prof. Dr. Gerd Buschhorn, verdanke ich die Bereitstellung der erforderlichen Arbeitsmittel und die Finanzierung meiner Dienstreisen zu DESY Hamburg.

Erklärung

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

München, Juli 1998

(Martin Schied)