# Inhaltsverzeichnis

# Einleitung

Das Satelliten-Experiment EGRET (**E**nergetic **G**amma **R**ay **E**xperiment **T**elescope) auf dem CGRO-Satelliten (**C**ompton **G**amma **R**ay **O**bservatory) konnte bisher 161 astronomische Objekte mit $\gamma$-Strahlung oberhalb von 100 MeV nachweisen. Insgesamt 57 $\gamma$-Quellen [1] wurden oberhalb von 1 GeV gefunden, teilweise mit (27), teilweise ohne (30) identifiziertes Gegenstück im unteren Energiebereich.

Von dieser Vielzahl von $\gamma$-Quellen konnten bis heute, mit Hilfe von abbildenden Čerenkov-Teleskopen, nur vier Quellen im unteren VHE Bereich (1 TeV bis 20 TeV) nachgewiesen werden: Der Krebs-Nebel, der Supernova-Überrest SNR1006 sowie die aktiven galaktischen Kerne (AGN) Mkn 421 und Mkn 501. Bei höheren Energien, also ab 20 TeV und darüber, konnten bis heute noch keine $\gamma$-Quellen statistisch gesichert etabliert werden[1].

Der Nachweis weiterer $\gamma$-Quellen, insbesondere im Energiebereich oberhalb von 20 TeV hätte weitreichende Konsequenzen für die Astrophysik:

> Der Nachweis von $\gamma$-Strahlung von Supernova-Überresten ist von Bedeutung für die Überprüfung von Aussagen über die Orte und die Beschleunigungsmechanismen der geladenen kosmischen Strahlung.
>
> Durch die Beobachtung extragalaktischer AGN können weitere Informationen über die zwischen der Quelle und dem Beobachtungspunkt liegenden intergalaktischen Strahlungsfelder gesammelt werden.

Falls man den differentiellen Energiespektren der Quellen Potenzgesetze (wobei typischerweise Exponenten zwischen 2 und 3 angenommen werden) zu Grunde legt und die von Ballon- und Satellitenexperimenten oder Čerenkov Teleskopen gelieferten Daten extrapoliert, scheint das Ausbleiben statistisch gesicherter Nachweise oberhalb von 20 TeV auf der nicht ausreichenden Fluß-Sensitivität heutiger Instrumente zum Nachweis kosmischer $\gamma$-Strahlung in diesem Bereich zu beruhen.

Die experimentelle Technik in diesem Energiebereich ist der auch bei HEGRA (**H**igh **E**nergy **G**amma **R**ay **A**stronomy) mit den Detektormatrizen durchgeführte Nachweis *ausgedehnter atmosphärischer Luftschauer* durch eine auf dem Erdboden aufgebaute ausgedehnte Anordnung von Einzeldetektoren (*shower front sampling*).

Für die Suche nach den von sehr hochenergetischen Photonen induzierten Luftschauern stellt der Untergrund der geladenen kosmischen Strahlung, dessen Fluß um 2 bis 3 Größenordnungen höher ist als die erwarteten Photonen-Flüsse, das größte Problem dar.

[1] Das Experiment CANGAROO hat allerdings positive Evidenz für $\gamma$-Strahlung bis 50 TeV vom Krebs-Nebel veröffentlicht [2]. Dieses Signal konnte bisher von keinem anderen Experiment bestätigt werden.

Die Erhöhung der Sensitivität der Nachweisinstrumente zur $\gamma$-Punktquellensuche im TeV-Energiebereich ist also mit einer möglichst effektiven Reduzierung des Untergrunds möglich. Dies kann entweder durch die Verbesserung der Winkelauflösung für $\gamma$-induzierte Luftschauer oder durch eine Unterscheidungsmöglichkeit zwischen $\gamma$-induzierten und Hadron-induzierten Schauern, die sogenannte $\gamma/H$-Separation, erreicht werden.
Ziel dieser Diplomarbeit ist die systematische Untersuchung von Möglichkeiten zur $\gamma/H$-Separation mit Methoden der Theorie künstlicher Neuronaler Netze.
Oder anders ausgedrückt: Die Zuhilfenahme einer an die mikroskopischen Vorgänge im Gehirn angelehnten Theorie bei der Klärung wichtiger astrophysikalischer Fragestellungen.

Diese Arbeit gliedert sich in folgende Teile :

- Einführung und Motivierung der astrophysikalische Fragestellung
- Beschreibung des HEGRA-Experiments
- Überblick über die Methode Neuronaler Netze
- Anwendung von Neuronalen Netzen zur $\gamma/H$-Separation mit dem HEGRA Detektormatrizen
- Vergleich zwischen der mit dem HEGRA-Detektor erreichten $\gamma/H$-Separation und der mit einem 100% flächenbedeckenden idealen Detektor zu erreichenden $\gamma/H$-Separation
- Anwendung eines zur $\gamma/H$-Separation trainierten Neuronalen Netzes auf reale Detektor-Daten und die Suche nach TeV $\gamma$-Strahlung von 12 extragalaktischen Objekten
- Zusammenfassung

# Kapitel 1

# $\gamma-$Astronomie im TeV Energiebereich

Die experimentellen Techniken der $\gamma-$Astronomie werden durch die Wechselwirkungswahrscheinlichkeiten in der Atmosphäre und die niedrigen Photonenflüsse bestimmt. Für den Radio-, große Teile des Infrarot- und den optischen Bereich ist die Atmosphäre transparent und somit der direkte erdstationäre Nachweis möglich. Oberhalb des UV-Bereichs wird die Atmosphäre für Photonen undurchlässig und die Beobachtungen müssen mit Satelliten oder Ballons durchgeführt werden. Auch diese Technik hat, wegen der mit der Energie stark abnehmenden Photonenflüsse und der somit erforderlichen großen Nachweisflächen, ihre Grenzen bei einigen GeV. Im Bereich der VHE (**V**ery **H**igh **E**nergy 0.1 TeV bis 100 TeV) und der UHE (**U**ltra **H**igh **E**nergy 100 TeV bis 100 PeV) $\gamma-$Astronomie kommen deshalb Luftschauerexperimente zum Einsatz, bei denen der Nachweis hochenergetischer Photonen indirekt durch den induzierten Luftschauer geführt wird. Diese indirekte Nachweismethode mit der Atmosphäre als Detektor-Volumen stellt im Vergleich zum direkten Nachweis mit Satelliten oder Ballons erhöhte Anforderungen an die Unterdrückung des durch die geladene Kosmische Strahlung verursachten Untergrunds.

Die Beobachtung hochenergetischer $\gamma$-Quanten setzt kosmische Beschleuniger, die Energien im VHE/UHE-Bereich erzeugen können, voraus. Einige der möglichen astronomischen Objekte und die Prozesse, die zur Erzeugung hochenergetischer $\gamma$-Strahlung führen können, werden in Abschnitt 1.1 vorgestellt. Abschnitt 1.2 beschäftigt sich mit den Abschwächungsmechanismen, denen Photonen im intergalaktischen Raum unterliegen. Die Eigenschaften der geladenen kosmischen Strahlung werden in Abschnitt 1.3 zusammengefaßt. Abschnitt 1.4 stellt die physikalischen Prozesse vor, die zur Beobachtbarkeit ausgedehnter atmosphärischer Luftschauer führen, und schließlich werden in 1.5 verschiedene experimentelle Methoden der $\gamma/H$-Separation (der Trennmöglichkeit zwischen $\gamma$-induzierten und Hadron-induzierten Schauern) angesprochen.

Am Ende diese Kapitels wird die Zielsetzung dieser Arbeit beschrieben.

## 1.1 Quellen der sehr hochenergetischen $\gamma$-Strahlung

Hochenergetische $\gamma$-Quanten sind nicht wie Photonen des sichtbaren Lichts thermalen Ursprungs, sondern sie entstehen bei Wechselwirkungen hochenergetischer geladener Teilchen in den folgende Prozessen:

- hochenergetische $e^+, e^-$:
  - Bremsstrahlung
  - Synchrotronstrahlung
  - inverse Compton-Streuung
- hochenergetische Protonen und Kerne:
  - $\pi^0$-Produktion bei Kern-Kern-Stößen , anschließender Zerfall $\pi^0 \rightarrow 2\gamma$
  - $\pi^0$-Produktion bei Kern-Photon-Stößen , anschließender Zerfall $\pi^0 \rightarrow 2\gamma$

Die Erzeugung hochenergetischer $\gamma$-Strahlung ($\gtrsim$0.5 MeV) setzt also kosmische Beschleuniger voraus, die geladene Teilchen sehr hoher Energie hervorbringen, und kann im Umkehrschluß zu deren Identifikation und Erforschung benutzt werden. Im folgenden werden 3 galaktische und anschließend die zur Zeit einzigen bekannten extragalaktischen Objekte kurz vorgestellt, für die sehr hochenergetische $\gamma$-Strahlung beobachtet oder erwartet wird.

**Pulsare**
Als Pulsare werden sehr dichte (ca. $10^{13}\,\mathrm{g/cm^3}$) und schnell rotierende Neutronensterne mit einem starken Magnetfeld bezeichnet. Hochenergetische geladene Teilchen können nur an den magnetischen Polkappen dieses Magnetfeld verlassen. Dies führt zusammen mit den oben aufgeführten Prozesse zu einer lokalisierten Strahlungs-Emission, die durch die Rotation, ähnlich wie bei einem Leuchtturm, periodisch moduliert wird. Weil Neutronensterne das Ergebnis von Supernova-Explosionen massereicher Sterne sind, befindet sich der Pulsar häufig inmitten eines ausgedehnten Materie-Nebels, der aus der abgestoßenen Materiehülle des Vorgänger Sterns entstanden ist. Falls die Materie des Nebels durch die Strahlung des Pulsares zum Leuchten angeregt wird, man spricht man von einem Plerion. Eines der prominentesten Beispiele für einen Plerion ist der M1 im Zentrum des Krebs-Nebels. Dem *Synchrotron-Self-Compton* Modell (de Jaeger & Harding 92) nach emittieren relativistische Elektronen in dem starken Magnetfeld des Neutronensternes Synchrotronstrahlung. Durch inverse Compton-Streuung, wiederum an dengleichen relativistischen Elektronen, werden die Photonen der Synchrotronstrahlung bis zu Energien im TeV-Bereich beschleunigt. Im Gegensatz zu der niederenergetischen Komponente ist diese Komponente der $\gamma$-Abstrahlung isotrop und wird somit durch die Rotation des Pulsars nicht mehr moduliert. Der Krebs-Nebel wurde vom Radio- bis in den TeV-Bereich [3] mit hoher Signifikanz nachgewiesen und gilt als die "Standard Kerze" der $\gamma-$Astronomie.

**Supernova-Überreste**
Als Quellen der geladenen kosmischen Strahlung gelten Supernova-Überreste in der Sedovphase - hundert bis mehrere tausend Jahre nach der Explosion. Bis heute konnte erst für einen Supernova-Überrest, SNR1006 [4], hochenergetische $\gamma$-Strahlung nachgewiesen werden. Für die Beschleunigungsmechanismen und die $\gamma$-Erzeugung existieren detaillierte Modelle [5]. Durch die

Ausbreitung der Materiehülle in die interstellare Materie entsteht eine Schockfront, in der durch diffuse Schockbeschleunigung (Fermi I) geladene Teilchen auf Energien über $10^{15}$ eV beschleunigt werden können. Bei der beobachteten $\gamma$-Strahlung des Supernova-Überrests SNR1006 handelt es sich dabei wahrscheinlich um Elektronen, die invers compton-gestreut werden an den Photonen der 2.7K Hintergrundstrahlung. Der Nachweis, daß in Supernova-Überresten auch Protonen beschleunigt werden, könnte sie als die Haupt-Quellen der geladenen kosmischen Strahlung oberhalb einiger GeV etablieren, weil nur Supernova-Überreste aufgrund ihrer Häufigkeit die geschätzte Energiedichte der kosmischen Strahlung von $\approx 1\,\mathrm{eV/cm^3}$ hervorbringen könnten.

**Binärsysteme**
Binärsysteme bestehen aus einem nicht-degenerierten Stern und einem schweren, kompakten Begleiter, einem weißen Zwerg, Pulsar oder schwarzem Loch. Die von dem kompakten Begleiter dem Stern durch Gravitation entzogene Materie sammelt sich in einer Akkretionsscheibe und wird von dort entweder durch Schockbeschleunigung oder starke elektrische Felder[1] auf Energie über $10^{15}$ eV beschleunigt. Obwohl für Binärsysteme detaillierte Modelle bestehen, die hochenergetische $\gamma$-Strahlung vorhersagen, ist bis jetzt deren Nachweis noch nicht gelungen.

**Aktive Galaxienkerne**
Die Kerne aktiver Galaxien (engl. **A**ctive **G**alactic **N**uclei) sind die energiereichsten und leuchtstärksten bekannten Objekten im Universum. Vermutlich wird die Energie bei der Akkretion von Materie durch ein schwarzen Loch im Zentrum der Galaxie erzeugt. Senkrecht zur rotierenden Akkretionsscheibe entstehen zwei Jets, die zuerst im Radiobereich beobachtet werden konnten. Die Mechanismen in diesen Jets, die zur Erzeugung hochrelativistischer Teilchen und der beobachteten Abstrahlung hochenergetische Photonen [6][7] führen, werden noch diskutiert. In dem zur Zeit favorisierten *Elektronen-Induzierten* Modell wird davon ausgegangen, daß die $\gamma$-Strahlung durch inverse Compton-Streuung von optischen, UV- und Röntgen-Photonen an hochrelativischen Elektronen der Jets entsteht, beziehungweise die von den Elektronen in den starken Magnetfeldern der Jets erzeugt Synchrotronstrahlung, durch den *Synchrotron-Selbst-Compton-Effekt* zu höheren Energien transformiert wird. AGN zeichnen sich durch Emission in einem großen spektralen Bereich aus, z.B. umfaßt die nachgewiesene Strahlung der AGN Mrk421 18 Größenordnungen in der Energie. Im Rahmen des vereinheitlichten AGN-Modells werden AGNs nach Entfernung und Beobachtungswinkel, der aufgrund unterschiedlicher Abstrahlungscharakteristiken der für die verschiedenen Energiebereiche relevanten Prozesse unterschiedliche beobachtete spektrale Eigenschaften zufolge hat, in verschiedenen Kategorien klassifiziert: Blazare, Seyfert Galaxie oder Radio Galaxie.

## 1.2 Ausbreitung von VHE/UHE Photonen im intergalaktischen Medium

Neben der Absorption von $\gamma$-Strahlung an dem interstellaren Medium oder an Dunkelwolken spielt im VHE-Bereich insbesondere Paarerzeugung mit niederenergetischen Photonen eine wichtige Rolle als Abschwächungsprozeß.

$$\gamma\gamma_{bg} \rightarrow e^- e^+ \qquad (1.1)$$

[1] Rotieren der kompakte Begleiter und die Akkretionsscheibe mit unterschiedlicher Frequenz bildet sich eine Art Dynamo aus.

Die Schwellenenergie dieser Reaktion ist:

$$E \geq E_{tresh} = \frac{2m_e^2}{\epsilon(1 - cos(\theta))} \approx 0.5 \left[\frac{1TeV}{\epsilon}\right] eV \tag{1.2}$$

Wobei E die Energie des einfallenden, $\epsilon$ die Energie des Hintergrund-Photons und $\theta$ der Winkel zwischen den Photonen ist.
Für den sehr gut bekannten 2.7K Mikrowellenhintergrund ($\epsilon \approx 10^{-3}$ eV) folgt aus der Formel 1.2, daß für Photonen mit einer Energie größer als 100 TeV das Universum undurchlässig wird. Unterhalb von 100 TeV ist die Paarerzeugung an Photonen des optischen- und des Infrarot Bereichs möglich. Durch die mangelnde Kenntnis der diffusen intergalaktischen Photonendichte in diesem Energieintervall sind Abschätzungen der Wechselwirkungs-Längen unsicher. Die Beobachtung weiterer extragalaktischer Objekte über einen möglichst großen Energiebereich wird zu schärferen oberen Grenzen für den intergalaktischen IR Hintergrund führen.

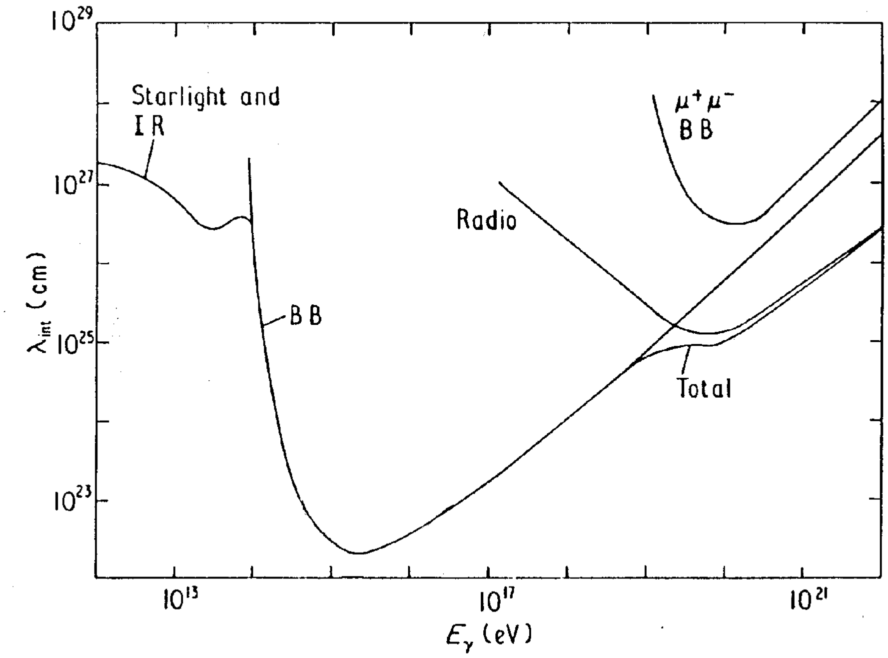



*Abbildung 1.1: Mittlere freie Weglänge für hochenergetische Photonen als Funktion der Energie $E_\gamma$. Unterhalb von $E_\gamma = 10^{15} eV$ tragen Photonen des Mikrowellenhintergrunds (BB) sowie des Sternen- und Infrarotlichts (IR) zur Absorption bei (aus [8]).*

Eine genauere Kenntnis des IR Hintergrunds hätte interessante und weitreichende kosmologische Implikationen. MacMinn & Primack [9] konnten durch Berechnungen zeigen, daß Modelle des IR Hintergrunds $n(\epsilon, z)$ ($n$ ist die Photonendichte, $z$ die Rotverschiebung) entscheidend beeinflußt werden durch den Zeitpunkt der Galaxien-Entstehung. Modelle mit einer angenommenen frühen Galaxien-Bildung sagen einen hohen IR Hintergrund voraus, während Modelle mit einer späten Galaxien-Entstehung niedrigere Werte vorhersagen. Der Zeitpunkt der kosmischen Strukturbildung wiederum gibt Hinweise auf die Art der dunklen Materie, denn Modelle mit kalter dunkler Materie (CDM) sagen eine frühe Galaxien-Entstehung vorher, während Modelle mit kalter und heißer dunkler Materie (CHDM) eine späte Galaxien-Entstehung vorhersagen.

## 1.3 Geladene Kosmische Strahlung

Abbildung 1.3 zeigt das durch Ballon-, Satelliten- und Luftschauerexperiment bestimmte Spektrum der kosmischen Strahlung integral für alle Teilchensorten. Der Verlauf läßt sich über weite Bereiche gut durch ein Potenzgesetz beschreiben:

$$\frac{d\Phi}{dE} = \Phi_0 \cdot E^{-(\gamma+1)} \tag{1.3}$$

Man unterscheidet bezüglich des Exponenten drei Bereiche:

Von $10^3$ GeV bis 3 PeV beträgt $\gamma \sim 1.7$ , der Bereich bei 3 PeV wird als "Knie" bezeichnet.

Von 3 PeV bis 1 EeV beträgt $\gamma \sim 2.1$, der Bereich bei 1 EeV wird als "Knöchel" bezeichnet.

Oberhalb des "Knöchels" scheint das Spektrum wieder abzuflachen, dies gilt wegen der großen statistischen Fehler der Messungen jedoch nicht als gesichert.

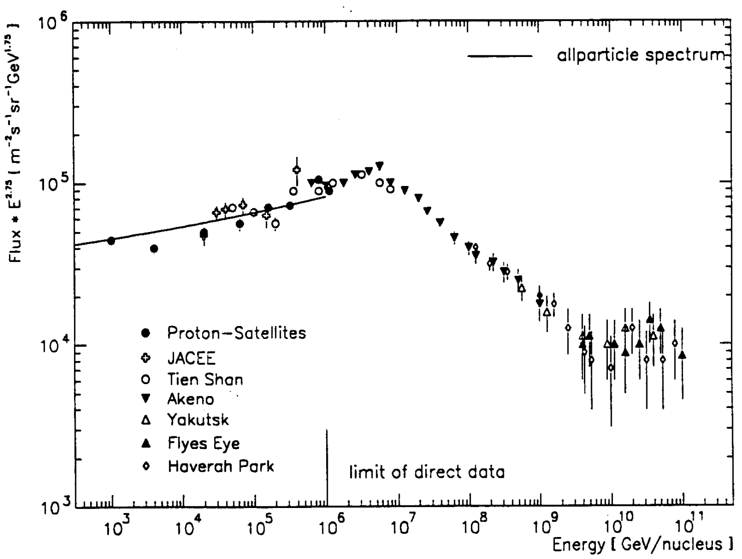



*Abbildung 1.2: Differentielles Energiespektrum der geladenen kosmischen Strahlung (Die Werte wurden für die Darstellung mit $E^{2.75}$ skaliert).*

Die Form des Spektrums schließt einen thermische Ursprung der Strahlung aus. Von besonderem Interesse für die Astrophysik ist der Energiebereich des "Knies". Allgemein anerkannt sind zwei unterschiedliche Erklärungsweisen für das bei 3 PeV plötzlich steiler werdende Spektrum.

**A** Es wird durch die energieabhängige Änderung des Ausbreitungsverhaltens verursacht: Teilchen können nur bis zu einer bestimmten Energie (abhängig von ihrem Masse zu Ladung Verhältnis) durch galaktische Magnetfelder gebunden werden.

**B** Es handelt sich um eine den Beschleunigungsmechanismen der kosmischen Strahlung gemeinsame Eigenschaft.

Da die Modelle unterschiedliche Aussagen über die Zusammensetzung der kosmischen Strahlung vor und nach dem Knie machen (**A**: Komposition verschiebt sich zu schwereren Elementen, **B**: Komposition bleibt gleich oder verschiebt sich nur leicht zu schwereren Elementen.) kommt der Bestimmung der Komposition eine entscheidene Rolle zu. Unterhalb von 100 TeV gilt die Komposition als sehr gut vermessen, so daß sie in dem für diese Arbeit interessierenden Energiebereich als bekannt angenommen werden kann. Bei 10 TeV setzt sie sich aus 50% Protonen, 25% Helium und 25% schwereren Kernen zusammen([10]). Diese Werte wurden auch für die in Abschnitt 2.5 vorgestellten $M^{te}$ Carlo-Ereignisse verwandt. Neben dem Spektrum und der Komposition wurden von der HEGRA Kollaboration weitere Untersuchungen zur kosmischen Strahlung veröffentlicht :

- Im TeV-Bereich ist die Abweichung des Flusses der kosmischen Strahlung von der Isotropie kleiner als 0.1 % [11].
- Das Verhältnis von diffuser $\gamma$-Strahlung zur geladenen Komponente der kosmischen Strahlung ist oberhalb 65 TeV kleiner als 1% [12].

## 1.4 Ausgedehnte Luftschauer

### 1.4.1 Longitudinale Entwicklung von Luftschauern

Nach dem Eintritt eines hochenergetischen Teilchens der kosmischen Strahlung in die Atmosphäre kommt es zur Wechselwirkung mit den dort vorhandenden Atomkernen. In einem multiplikativen Prozeß bildet sich schnell eine Kaskade von Sekundärteilchen der Wechselwirkungsprozesse aus. Je nach Typ des einfallenden Primärteilchens spricht man von $\gamma$-induzierten oder Hadron-induzierten ausgedehnten Luftschauern. In $\gamma$-induzierten Schauern sind Prozesse der elektromagnetischen Wechselwirkung entscheidend[2], dies ist die Paarerzeugung $\gamma \rightarrow e^+e^-$ im Kernfeld eines Targets sowie Bremsstrahlung der Elektronen[3]. Abbildung 1.3 zeigt das Zusammenspiel der beiden Prozesse innerhalb der Kaskaden, die relevanten Längenskalen sind die Strahlungslänge in Luft $X_0 = 37g/cm^2$ und die Konversionslänge von Photonen $9/7X_0$. Mit fortschreitendem Alter der elektromagnetischen Kaskade nimmt die mittlere Energie der beteiligten Elektronen ab und erreicht schließlich die kritische Energie ($E_c \approx$80 MeV für Luft). Unterhalb dieser Energie dominieren Ionisationsverluste den weiteren Verlauf der Kaskade und Bremsstrahlung spielt nur noch eine untergeordnete Rolle dies bewirkt eine Abnahme der Schauerteilchen. Die Höhe, in der die Anzahl der Schauerteilchen ihr Maximum erreicht hat, wird als *Schauermaximum* bezeichnet.

Die Vorgänge in Hadron-induzierten Schauern sind vielfältiger. Durch starke Wechselwirkungen enstehen in den Kollisionen neben Kernfragmenten auch geladene und neutrale Pionen sowie Kaonen. Die geladenen Pionen und die Kaonen zerfallen schnell in Myonen, die neutralen Pionen in zwei Photonen, die ihrerseits wie oben beschrieben eine elektromagnetische Subkaskade auslösen. Der Hadron-induzierte Schauer stirbt aus, wenn die Energie der Sekundärteilchen nicht mehr ausreicht, um in den Kollisionen Pionen zu erzeugen. Die hadronische Wechselwirkungslänge (z.B 1 TeV Protonen) in Luft beträgt 83 $g/cm^2$, ist also mehr als doppelt so groß

[2] Man spricht deshalb auch von elektromagnetischen Schauern.
[3] Im weiteren wird Elektron synonym zu Elektron und Positron verwand.

wie die Strahlungslänge $X_0$. Hadron-induzierte Schauer entwickeln sich daher im Mittel langsamer als elektromagnetische Schauer der gleichen Energie und sterben auch langsamer aus, ihre elektromagnetische Komponente wird ständig über den Pionen Zerfall der in hadronischen Wechselwirkungen entstandenden neutralen Pionen nachgespeist. Mit Hilfe des in Abbildung

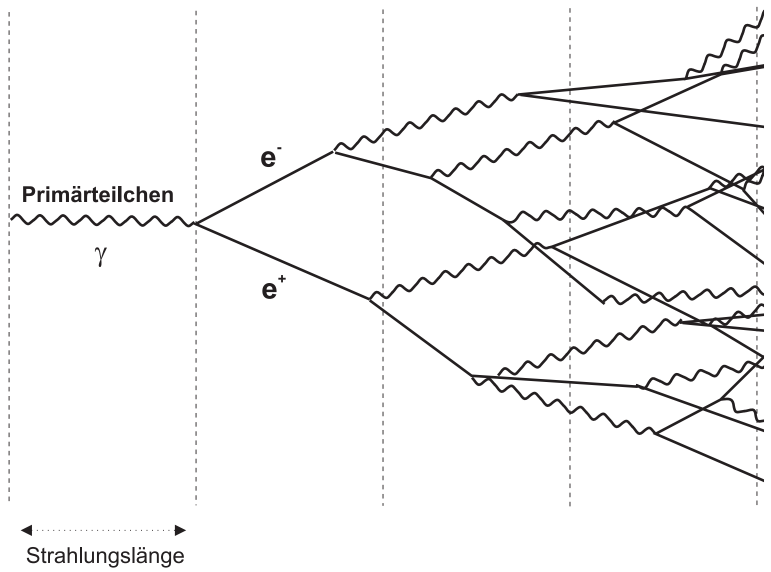



*Abbildung 1.3: Skizze der Kaskadenentwicklung eines elektromagnetischen Schauers*

1.3 skizzierten Modell kann für elektromagnetische Schauer eine einfache Relation zwischen der Primärenergie und der Höhe des Schauermaximums hergeleitet werden. Der Erwartungswert der mittleren Energie ist nach k Strahlungslängen:

$$\langle E(k)\rangle = E_0\ 2^{-k} \tag{1.4}$$

Weil das Schauermaximum bei $E(k) = E_c$ erreicht ist, folgt daraus:

$$k = \frac{1}{\log(2)} \log\left(\frac{E_o}{E_c}\right) \propto \log(E_0) \tag{1.5}$$

D.h. die Höhe des Schauermaximums nimmt logarithmisch mit der Energie des Primärteilchens ab. Mit M$^{te}$ Carlo-Simulationen kann gezeigt werden, daß diese Relation auch für Hadron-induzierte Schauer gilt, und daß die Maxima $\gamma$-und Proton-induzierter Schauer bei gleicher Energie auch auf gleicher Höhe liegen. Ebenfalls aus M$^{te}$ Carlo-Simulationen folgt, daß für schwerere Hadronen in Formel 1.5 $E_0$ durch $E_0/A$ substituiert werden muß[4]. Die durch Kerne mit A>1 erzeugten Schauer erreichen ihr Maximum also deutlich früher als Proton-induzierte Schauer, so daß die elektromagnetische Komponente auf Detektorniveau sich bereits in einer späteren Phase der Entwicklung befindet. Dadurch fällt das Verhältnis von Čerenkovlichtzur Anzahl der elektromagnetischen Partikel auf Detektorniveau entsprechend kleiner aus und schwerere Elemente werden im Kontext der $\gamma/H$-Separation als "hadronischer" bezeichnet (siehe Abschnitt 4.1.1).

[4]Wie dies auch aus dem Modell der einfachen Superposition der Kerne durch A Nukleonen folgen würde.

### 1.4.2 Laterale Entwicklung von Luftschauern

Bei elektromagnetischen Schauern wird die laterale Dichteverteilung der Sekundärteilchen bestimmt durch Coulomb-Vielfachstreuung an den Atomen der Atmosphäre. Da dieser Prozeß im Schwerpunktssystem isotrop ist, kommt es zusammen mit der stark vorwärts gerichteten Bewegung der Schauerteilchen zu einer resultierenden Dichteverteilung, die rotationssymetrisch zur Schauerachse[5] ist. Die Lateralverteilung (also senkrecht zur Schauerachse) der Sekundärteilchen kann durch die aus der Kaskadentheorie stammenden NKG Funktion (J. Nishimura, K. Kamata und K. Greisen) [13] beschrieben werden. Sie beschreibt die mittlere Flächendichte der $e^-$ und $e^+$ im Abstand r zur Schauerachse für elektromagnetische Schauer:

$$\rho = \frac{\Gamma(4.5 - s)}{2\pi \cdot \Gamma(s) \cdot \Gamma(4.5 - 2s)} \cdot N_e \cdot \frac{1}{r_M^2} \cdot \left(\frac{r}{r_M}\right)^{s-2} \cdot \left(\frac{r}{r_M} + 1\right)^{s-4,5} \tag{1.6}$$

Der Molière-Radius $r_M$ und das sogenannte Schaueralter $s$ sind von der atmosphärischen Tiefe des Beobachtungspunktes abhängig. Der Moliére-Radius $r_M$ ist eine Größe der für die laterale Ausdehnung des Schauers verantwortlichen Coulomb-Vielfachstreuung. Der Normierungsfaktor $N_e$ gibt die Gesamtzahl der im Schauer enthaltenen $e^-$ und $e^+$ an. Das Schaueralter s parametrisiert die Stadien der longitudinalen Schauerentwicklung, s steigt streng monoton mit der Eindringtiefe an und ist im Schauermaximum gleich 1. Die Herleitung der NKG Formel gilt nicht für Hadron-induzierte Schauer oder für den Fall, daß die Schauerentwicklung abschnittsweise in unterschiedlichen Medien (charakterisiert durch $E_C$ und $r_M$) stattfindet, wie dies bei HEGRA aufgrund einer Bleischicht über den Szintillator-Detektoren der Fall ist. In diesem Fall kann s jedoch weiterhin als Charakterisierung der Steilheit der Lateralverteilung verwendet werden. Ein kleines s deutet somit auf eine steile und ein großes s auf eine flache Verteilung hin. Wegen der oben genannten Gründe wird im weiteren Verlauf statt s und $N_e$ die aus der Anpassung der NKG Formel bestimmten Parameter als $Age$ und $Ne$ bezeichnet.

### 1.4.3 Atmosphärisches Čerenkovlicht

Geladene Teilchen, die sich in einem dielektrischen Medium mit Überlichtgeschwindigkeit $v/c_0 = \beta > 1/n(\omega)$ fortbewegen, erzeugen Čerenkov-Photonen, die unter dem Winkel

$$cos(\theta_c) = \frac{1}{\beta n(\omega)} \tag{1.7}$$

emittiert werden. Die Intensität des abgestrahlten Čerenkovlichts ist proportional zu $\frac{1}{\lambda^2}$, der Schwerpunkt des Spektrums wird jedoch durch Raleigh-Streuung, Mie-Streung und Absorption an Ozon Molekülen zu blauen Wellenlängen verschoben, so daß das am Erdboden ankommende Čerenkovlicht sein Maximum bei ca. 330 nm hat. Für Elektronen liegt die Schwellenenergie der Čerenkov-Emission im MeV Bereich, der Anteil der übrigen Schauerteilchen zur gesamten Čerenkov-Photonen-Produktion ist wegen ihrer hohen Ruhemasse und der daher höheren Energieschwelle praktisch zu vernachlässigen. Dies führt dazu, daß die in der elektromagnetischen Kaskade deponierte Energie eng korreliert ist mit dem insgesamt erzeugten Čerenkovlicht. Für die Form der Lateralverteilung sind im wesentlichen zwei Effekte verantwortlich:
Mit steigender Eindringtiefe wird der Brechungsindex $n$ der Atmosphäre größer und daher auch der Čerenkov-Winkel sowie die Energieschwelle der Čerenkov-Emission niedriger.

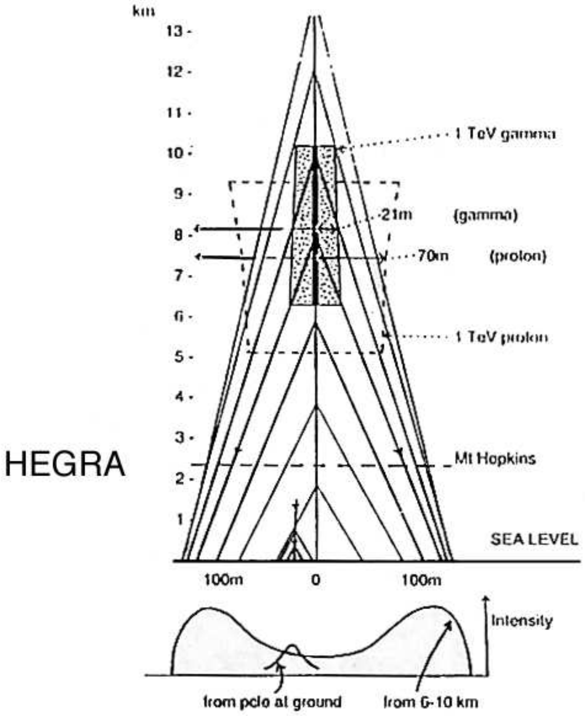



*Abbildung 1.4: Erzeugungshöhe und Auftreffpunkt des Čerenkovlichts (aus [14]).*

Insgesamt ergibt die Superposition der emittierten Čerenkov-Photonen eine Lateralverteilung die im Gegensatz zur Lateralverteilung der elektromagnetischen Komponente die gesamte longitudinale Entwicklung widerspiegelt: Im Bereich nahe des Kernortes (bis 20 m) [6] wird sie bestimmt durch die Čerenkov-Photonen, die in den letzten 1 bis 2 Strahlungslängen über der Beobachtungsebene abgestrahlt wurden. Der Bereich von 20 m bis 120 m wird durch die im Schauermaximum emittierten Čerenkov-Photonen dominiert und den Bereich außerhalb von 120 m erreicht nur das von vielfachgestreuten Teilchen abgestrahlte Čerenkovlicht.
Die Lateralverteilung im Bereich von 20 m bis 100 m Abstand zum Kernort kann in sehr guter Näherung durch eine Exponentialfunktion beschrieben werden.

$$\rho_{Cl}(r) = \rho_o \cdot \exp\left(-\frac{r}{r_l}\right) \tag{1.8}$$

Die Konstante $r_l$ (Lichtradius) beschreibt den exponentiellen Abfall der Intensitätsverteilung in diesem Bereich und ist korreliert mit der Höhe des Schauermaximums: Je tiefer das Schauermaximum, desto mehr Čerenkov-Photonen erreichen das Detektorniveau in der Nähe des Schauerzentrums - die Lateralverteilung des Čerenkovlichts $\rho_{Cl}(r)$ erscheint steiler und $r_l$ ist entsprechend klein.

[5] Entspricht der Trajektorie des Primärteilchens.

[6] Alle Angaben in Metern beziehen sich auf die Beobachtungshöhe des HEGRA Detektors.

## 1.5 $\gamma/H$-Separation bei erdstationären Detektoren

Wie bereits erwähnt, ist für die $\gamma-$Astronomie mit Luftschauer-Experimenten die Unterdrückung des dominierenden Hadron-induzierten Untergrundes, der weit mehr als 99 % [15] aller nachweisbaren Schauer ausmacht, von entscheidener Bedeutung. In Hadron-induzierten Luftschauern sind durch den Zerfall geladener Pionen eine große Anzahl von Myonen enthalten (siehe 1.4.1). Im Gegensatz dazu können in $\gamma$-induzierten Schauern Myonen nur durch Photoproduktion erzeugt werden. Der kleine Wechselwirkungsquerschnitt dieses Prozesses führt dazu, daß in $\gamma$-induzierten Schauern der Myonen-Anteil um den Faktor $\sim$ 30-100 im Vergleich zu Hadron-induzierten Schauern unterdrückt ist. Der Nachweis von Myonen als Indikator Hadron-induzierter Schauer wurde bei HEGRA mit den inzwischen abgeschalteten Geiger-Turm-Detektoren durchgeführt [16]. Die Schwierigkeiten, einer auf dem Myonen-Anteil basierenden Separations-Methode liegen in der geringen Anzahl und der flachen Lateralverteilung der Schauer-Myonen

$$\rho_\mu \propto {N_e}^{0.75} \cdot r^{-0.75} \left(1 + \frac{r}{300m}\right)^{-2.5} N_\mu \propto 10 \cdot \left(\frac{E_0}{m_\pi c^2}\right)^{0.82} \tag{1.9}$$

und der daraus resultierenden notwendigen hohen Überdeckung des Detektors mit aktiver Fläche begründet. Hinzu kommen experimentelle Schwierigkeiten Myonen von dem sehr viel größeren Anteil hochenergetischer Elektronen (*punch through-*) zu unterscheiden.

Eine andere Eigenschaft Hadron-induzierter Schauer, nämlich die kontinuierliche Produktion von Sub-Schauern und die damit verbundene "Selbstähnlichkeit", wird für fraktale Analysen der elektromagnetischen Lateralverteilung am Erdboden benutzt. Dieses Verfahren wurde unter anderem auf die Photoplatten des "Kieler Luftschauerexperiments" zur Trennung von Proton- und durch schwerere Elemente induzierter Schauer angewand [17]. Mit Ausnahme experimenteller Schwierigkeiten steht der Übertragung dieser Methode auf das Problem der $\gamma/H$-Separation nichts entgegen.

Für ACT (**A**bbildende **Č**erenkov **T**eleskope) existieren leistungsfähige $\gamma/H$-Separations-Methoden, die auf der Bildanalyse des in den Teleskop-Kameras aufgezeichneten Čerenkovlichts basieren. In Abschnitt 4.6 wird eine Vergleich gezogen zwischen der ACT-Methode und der in dieser Arbeit behandelten Methode der $\gamma/H$-Separation mit simultaner Messung des atmospärischen Čerenkovlichts und der elektromagnetischen Teilchen-Komponente.

## 1.6 Zielsetzung dieser Arbeit

Zielsetzung dieser Arbeit ist die Untersuchung der Separationsmöglichkeit zwischen $\gamma$-und Hadron-induzierten ausgedehnten Luftschauern mit Hilfe von Neuronalen Netzen.
Als Eingangswerte des Neuronalen Netzes werden die in den HEGRA AIROBICC- und der Szintillator-Matrix gemessenen Größen verwendet. In diesem Rahmen wird der Vergleich gezogen zwischen Neuronalen Netzen, die mit direkten Meßgrößen der Detektoren trainiert werden, und Neuronalen Netzen, die auf bereits sehr weit vorverarbeiteten Kenngrößen des Schauers basieren. Bei der Quantifizierung der Ergebnisse wird im Hinblick auf die Suche nach $\gamma$-Quellen eine realistische Modellierung der unteren Energieschwelle sowie der Rekonstruktionsgenauigkeiten berücksichtigt.
Als Anwendungsbeispiel wird eines der Neuronalen Netze für die $\gamma/H$-Separation bei der Suche nach TeV $\gamma$-Strahlung von 12 extragalaktischen Objekten verwendet werden.
Es wird außerdem die Frage untersucht werden, bis zu welcher Qualität die $\gamma/H$-Separation mit den verwendeten physikalischen Größen möglich ist, falls man eine nahezu ideale Bestimmung der Meßgrößen zugrunde legt.
Das hierfür entwickelte Programmpaket kann wegen der Allgemeinheit des Ansatzes, auf die Schätzung weiterer nicht direkt beobachtbarer Größen angewendet werden. Dies wird im Anhang exemplarisch am Beispiel eines Regressionsnetzes für die Höhe des Schauermaximums dargestellt.

# Kapitel 2

# Das HEGRA Experiment

Das Experiment HEGRA (**H**igh **E**nergy **G**amma **R**ay **A**stronomy) besteht aus insgesamt vier verschiedenen Detektor Systemen: Szintillator Matrix, AIROBICC Matrix, 17 Geigertürme und einem System aus 6 abbildenden Čerenkov Teleskopen. Da die Messungen der Detektoren auf unterschiedlichen Komponenten und/oder Meßprinzipien beruhen, die Energie-Intervalle sich jedoch teilweise überdecken, bietet sich die einmalige Chance mit Hilfe von koinzident aufgezeichneten Ereignissen systematische Untersuchungen der einzelnen Detektor-Komponenten durchzuführen.
Der Standort des Experiments liegt auf der kanarischen Insel La Palma ($28.8^o$ nördl. Breite, $17.9°$ westl. Länge) in 2200 Meter Höhe über dem Meeresspiegel auf dem Roque de los Muchachos. Er zeichnet sich durch verschiedene Vorteile aus :
Die geographische Position ermöglicht die Beobachtung des gesamten Nordhimmels und eines Teils des Südhimmels. Die hohe Lage ermöglicht eine niedrige Energieschwelle zum Nachweis der von Luftschauern induzierten Teilchenkaskaden. Der hohe Prozentsatz an klaren und wolkenlosen Nächten ermöglichen die Beobachtung des von Luftschauern erzeugten Čerenkovlichts.

Die Themengebiete/Schwerpunkte des HEGRA Experimentes sind :

- Suche nach VHE/UHE $\gamma$-Quellen
- Nachweis einer VHE Komponente der von Satellitenexperimenten gemeldeten $\gamma$-Ray Bursts
- Suche nach isotropen Photonen-Hintergrund.
- Nachweis von Anisotropien in der kosmischen Strahlung
- Bestimmung des Spektrums und der Komposition der geladenen kosmischen Strahlung

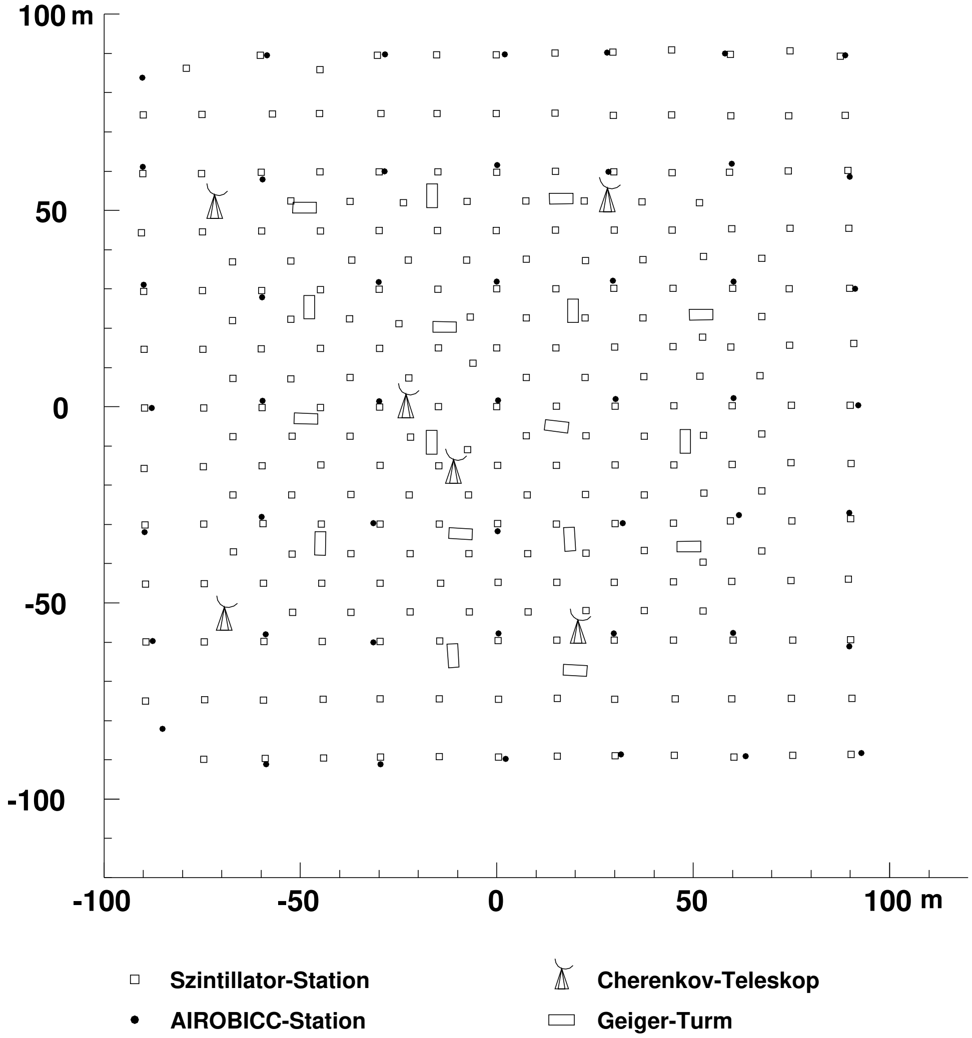



*Abbildung 2.1: Lageplan des HEGRA Experiments*

## 2.1 Gesamtaufbau

Im Zentrum der HEGRA Anlage befindet sich ein Container mit der Elektronik zur Steuerung und Datennahme der einzelnen Detektor Komponenten. Die Matrizen aus Szintillator, AIROBICC (**AIR**shower **OB**servation with angle **I**ntegrating **C**herenkov **C**ounters) und Geiger-Türmen, die über Kabel mit der zentralen Elektronik verbunden sind, besitzen eine gemeinsame Logik zum Starten der Datenauslese. Die Logik ist so geschaltet, daß sowohl die Szintillator- als auch die AIROBICC-Matrix bei Überschreiten der minimalen Anzahl von Einzeldetektoren mit Signaleintrag den Ausleseprozeß starten kann. Die Datennahme des Systems von Čerenkov Teleskopen ist völlig eigenständig und wird von einer eigenen Elektronik ausgelöst. Nur speziell ausgewählte Ereignisse lösen die Datennahme der anderen Komponenten aus und können für wechselseitige Überprüfungen genutzt werden. In dieser Arbeit werden nur Daten der AIROBICC- sowie der Szintillator-Matrix verwendet.

## 2.2 Szintillator-Matrix

Die Szintillationszählermatrix wurde 1988 mit 37 Zählern in Betrieb genommen. In der heutigen Ausbaustufe besteht sie aus einer $13 \times 13$ Matrix von Detektoren mit einem Zählerabstand von 15m m, die im Zentrum durch eine zweite überlagerte Matrix mit weiteren 52 Zählern auf einen minimalen Abstand von 10 m verdichtet wird. Grund für die Verdichtung ist die Tatsache, daß die Energieschwelle in etwa umgekehrt proportional zur Zählerdichte ist, und somit durch die Verdichtung die Energieschwelle der Szintillator-Matrix von $\approx 50$ TeV auf $\approx 30$ TeV gesenkt werden konnte.
Die einzelnen Detektoren sind in lichtdichten Blechgehäusen untergebracht und bestehen aus horizontalen Schichten mit Blei ($\approx 0.9$ Strahlungslängen), Holz und Szintillationsmaterial (von Oben nach Unten), sowie einem darunter angebrachten Lichttrichter an dessen Verjüngung ein SEV (**S**ekundär **E**lektronen **V**ervielfacher) angebracht ist. Die Bleiabdeckung wurde nachträglich hinzugefügt, um niederenergetische Teilchen ($\lesssim 6$ MeV) zu absorbieren (diese verringern die Zeitauflösung, da sie aufgrund der Coulomb-Vielfachstreuung verspätet den Detektor erreichen) sowie die Sensitivitiät gegenüber hochenergetischen Photonen ($\gtrsim 10$ MeV) zu erhöhen [18] (diese konvertieren im Blei zu $e^-$ und $e^+$). Die Anoden- und Dynoden-Signale der SEV werden über 150 m lange Kabel zur zentralen Datennahme geführt. Dort werden aus den Anoden-Pulsen durch Schwellwertbildung in einem "Constant Fraction Discriminator", der einen Zeit-Digital Wandler ansteuert, die Ankunftszeiten ermitteln, sowie die Detektor Signale aus den Spitzenwerten der Dynoden- und Anoden-Pulse mit Analog-Digital Wandlern bestimmt. Aus den Zeitinformationen wird in der "Trigger Logik" das Datennahme-Signal gebildet. Die Standard Bedingung hierfür ist zur Zeit, daß 14 Einzeldetektoren-Signale in einem 150 ns Zeitfenster liegen [1].
Im statistischen Mittel sind die Signalamplituden der SEV-Pulse proportional zur Dichte (Fläche bezogen auf die Senkrechte zur Schauerachse) der geladenen hochenergetischen Teilchen, die das Szintillationsmaterial durchdringen. Hierbei gehen zwei Effekte ein :
Teilchen mit einer Primärenergie $\gtrsim 8\,MeV$ (**M**inimal **I**onisating **P**articles) deponieren im Szintillationsmaterial Energie, die nur schwach abhängig von ihrer Gesamt-Energie ist, aber proportional zur durchdrungenen Weglänge (Bethe-Bloch Formel). Im gleichen Maße, wie die Weglänge bei nicht-senkrechtem Einfall größer wird, schrumpft die effektive Fläche des Szintillators. Als Nettoeffekt bleibt eine Abhängigkeit von der Detektorfläche senkrecht zur Schauerachse übrig.

## 2.3 AIROBICC-Matrix

Der AIROBICC-Detektor wird seit Januar 1992 betrieben. Die 49 Zähler dieses Typs sind in einem quadratischen Raster aufgestellt, wobei der Abstand benachbarter Detektoren 30 m beträgt und somit ein Areal von insgesamt 33000 m$^2$ abgedeckt wird. Mitte 97 wurde eine Erweiterung auf 97 AIROBICC-Detektor in Betrieb genommen, die allerdings kurze Zeit später durch ein Feuer stark beschädigt wurden. In dieser Arbeit wird von einem Feld mit 49 AIROBICC-Detektoren ausgegegangen. Die Detektoreinheiten sind in zinkverkleideten Blechhütten installiert, deren Deckel ferngesteuert mit Motoren geöffnet werden können. Jede Detektoreinheit besteht aus einem nach oben hin geöffneten Winston-Cone-Spiegel, an dessen unterem Ende ein hochempfindlicher SEV angebracht ist. Durch eine blaue Folie vor Streu- und Sternlicht mit einer Wellenlänge von größer als 470 $nm$ geschützt, registriert der SEV direkt das einfallende Licht

[1] Diese Bedingung wird im weiteren mit S14 bezeichnet.

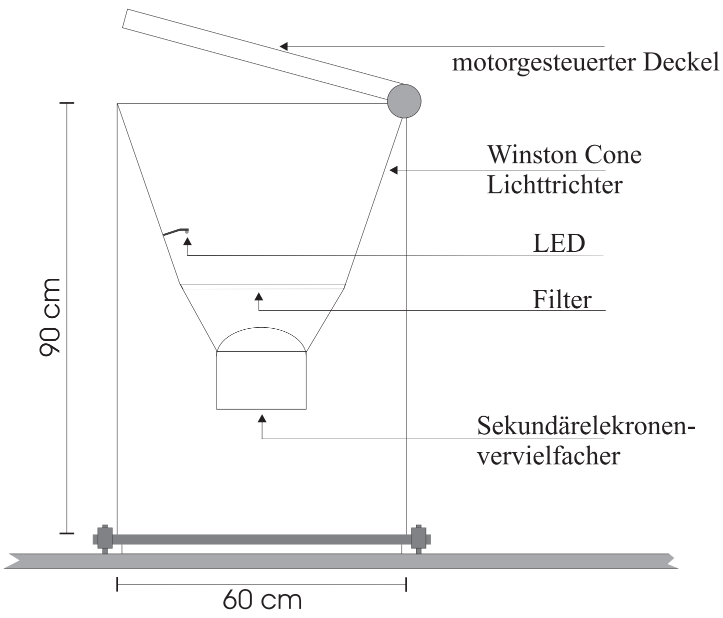



*Abbildung 2.2: Skizze eines AIROBICC-Detektors*

aus einem Raumwinkelbereich von $1\,sr$. Direkt im Lichttrichter ist für Kabellängenmessungen eine LED als Impulsgeber angebracht.
In möglichst wolkenfreien und klaren Nächten werden in der Zeit zwischen Mondauf- und untergang die Deckel der Hütten geöffnet und die Messung durchgeführt. Zusätzlich hierzu können auf Grund der Geometrie des Winston-Cone-Spiegels und der deshalb für Zenitwinkel $> \ 30^o$ stark abfallenden Detektor Akzeptanz [19] auch bei flach über dem Horizont stehendem Mond Messungen mit verminderter Hochspannung der SEV durchgeführt werden. Die Zeitauflösung der Zähler wird durch die Auflösung der Elektronik bestimmt und beträgt $\approx 0,3\,ns$.

Die SEV-Amplituden sind proportional zu den von den Zählern registrierten Lichtmengen. Um die Pulsamplituden mit den absoluten Lichtmengen in Verbindung zu bringen, wird auf einen Vergleich der mit AIROBICC vermessenen Amplitudenverteilung mit der aus der Szintillatormatrix gewonnenen Teilchen-Dichteverteilung zurückgegriffen. Mit der aus M$^{te}$ Carlo-Ereignissen gewonnenen Kenntnis der Relation zwischen Čerenkovlicht-Menge und Teilchendichte läßt sich eine absolute Kalibration der Pulsamplituden durchführen.
Dazu wird der Čerenkov-Amplituden zu Photon Eichfaktor variiert, so daß die Abweichungen der Korrelation $L_{90}$ zu $Ne$(Čerenkov-Lichtdichte bei 90 m Abstand vom Kernort zu angepaßter Anzahl der den Detektor erreichenden Teilchen) zwischen M$^{te}$ Carlo-Ereignissen und den Messungen minimal wird. Wenn dieses Verfahren auf zeitlich getrennte Abschnitte einer Meßreihe angewendet wird, können auf diese Weise so auch Schwankungen in den atmosphärischen Absorptionsbedingungen und den Detektor-Akzeptanzen kompensiert werden (siehe Kapitel 5.2).
Die Standardbedingung für die von AIROBICC ausgelösten Ereignisse sind mindestens 6

AIROBICC-Einzeldetektoren mit Signaleintrag[2]. Die mit der Standardbedingung erreichte Energieschwelle ist $\approx$ 15 TeV, also halb so hoch wie die Energieschwelle der Szintillator-Matrix, und liegt für $\gamma$-induzierte Ereignisse noch leicht unter diesem Wert.

## 2.4 Rekonstruktion

Die Rekonstruktion eines Ereignisses erfolgt für die beiden Detektor-Matrizen getrennt, wobei für den erste Schritt, die **Bestimmung des Schauer-Kernorts**[3], ein kombiniertes Verfahren angewendet wird. Es besteht aus einem modifizierten Schwerpunktsverfahren, das die inhomogene Verteilung der Szintillator-Detektoren und die endliche Ausdehnung der Detektor-Matrizen berücksichtigt, und zusätzlich aus einem Anpassungsverfahren an die laterale Verteilung der mit AIROBICC gemessenen Čerenkov-Photonendichte. Alle Teilergebnisse gehen anschließend, gewichtet durch verschiedenen Qualitätsmerkmale (z.B Anzahl der Detektoren mit Signaleintrag etc.) in das Gesamtergebnis ein [20].
Die **Bestimmung der Richtung** der Schauer-Front basiert auf zwei Maximum Likelihood Anpassungen einer Kegelform (Polynom 3. Ordnung im Abstand r) an die mittleren Ankunftszeiten der mit AIROBICC gemessenen Čerenkov-Photonen sowie der in den Szintillator-Detektoren gemessen minimal ionisierenden Partikel. Für beide Anpassungen wird die Kegelspitze an dem bereits bestimmten Kernort fixiert. Da aber im Gegensatz zur Čerenkov-Photonen-Front, deren Breite hauptsächlich durch Detektoreffekte bestimmt wird, die Schauer-Front der elektromagnetischen Komponente mit steigender Entfernung zum Kernort auf Grund von Coloumb-Vielfachstreuung verbreitert wird, muß bei der Anpassung an die Szintillator Daten die Ausdehnung der Partikel-Front berücksichtigt werden. Dies geschieht durch Vernachlässigung von sehr weit vom erwarteten Mittelwerte abweichenden Ankunftszeiten und durch Verwendung einer Gewichtsfunktion, die die Poisson-Statistik der Fluktuation in den Ankunftszeiten berücksichtigt [21].
Als letzter Schritt der Rekonstruktion erfolgt die **Anpassung der NKG Funktion** (Formel 1.6) an die laterale Verteilung der elektromagnetische Partikeldichte und die **Anpassung der Exponentialfunktion** (Formel 1.8) an die laterale Verteilung der Čerenkov-Photonen. Dazu werden die lateralen Verteilungen mit Hilfe des zuvor ermittelten Kernorts und der Schauerrichtung in ein Koordinatensystem mit Ursprung im Kernort und senkrecht zur Schauerrichtung transformiert.
Abbildung 2.3 zeigt den Prozeß der Datennahme und der Rekonstruktion im Überblick.
Deutlich werden die drei verschiedenen Möglichkeiten Ereignis-Informationen einem Neuronalen Netz zuzuführen :

- ① Kalibrierte Amplituden der Einzelhütten
- ② Aus den kalibrierte Amplituden konstruierte arithmetrische Größen.
- ③ Parameter der NKG- oder der Exponential-Anpassung

In Kapitel 4 wird auf diese Möglichkeiten näher eingegangen.

[2] Diese Bedingung wird im weiteren mit A6 bezeichnet.
[3] Als Kernort wird der Schnittpunkt der Trajektore des Primärteilchens mit der Detektor Ebene bezeichnet.

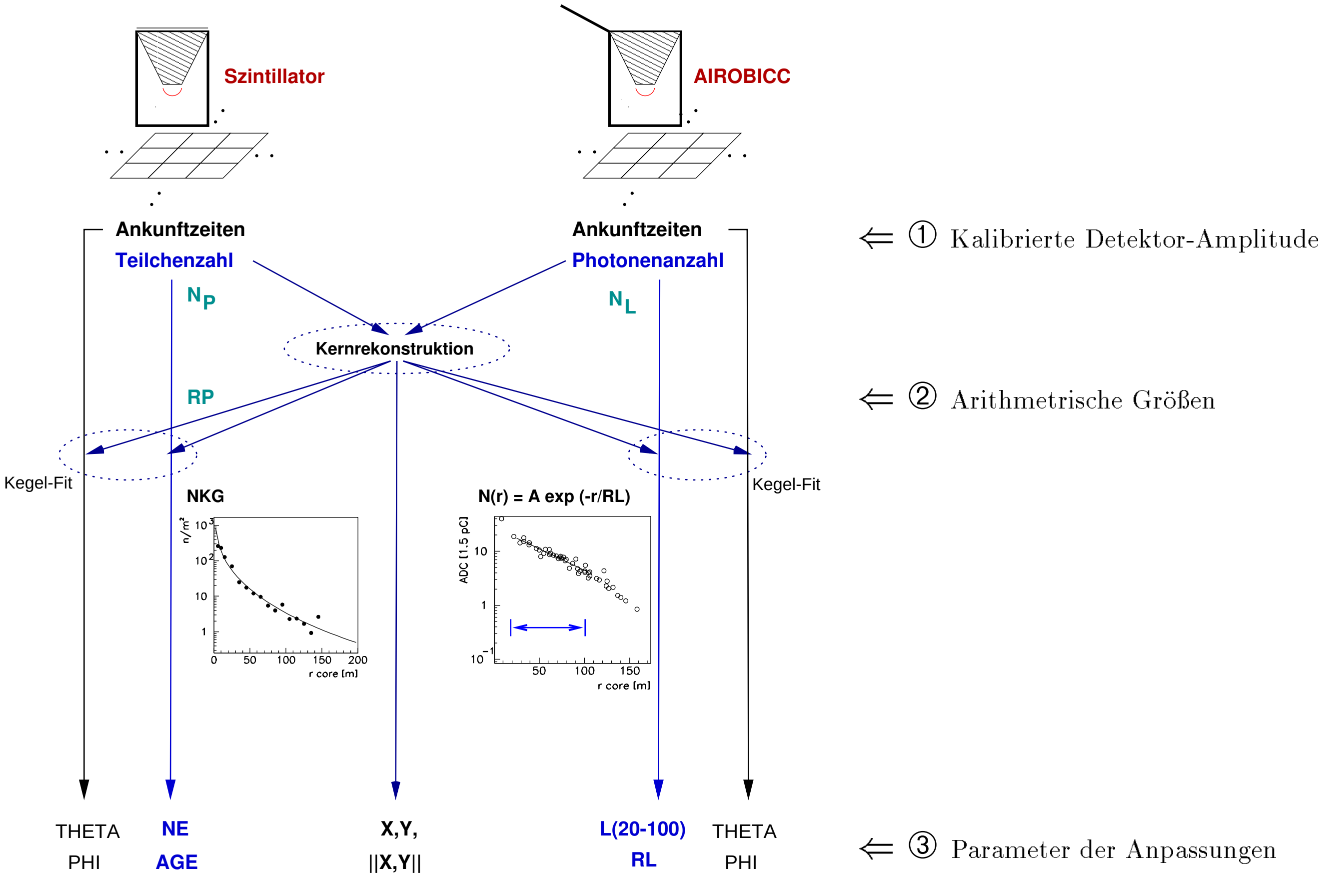



*Abbildung 2.3: Überblick der Rekonstruktionsschritte eines HEGRA-Ereignisses*

Tabelle 2.1 führt die wichtigsten Ergebnisse der verschieden Rekonstruktionsstufen auf.

| Name | Beschreibung |
|---|---|
| $a_i$ | Anzahl der im AIRIBICC-Detektor $i$ registrierten Čerenkov-Photonen |
| $s_i$ | Anzahl der im Szintillator-Detektor $i$ registrierten Schauerteilchen |
| $r$ | Vektor des rekonstruierten Kernorts |
| $N_l$ | Summe der registrierten Čerenkov-Photonenen $\sum_i a_i$ |
| $N_p$ | Summe der registrierten Schauerteilchen $\sum_i s_i$ |
| $r_p$ | Mittlerer Abstand der Schauerteilchen vom Kernort $\frac{\sum_i \lVert r_i - r \rVert s_i}{\sum_i s_i}$ ($r_i$ Position Sz-Detektor $i$) |
| $\theta$ , $\phi$ | Rekonstruierte Richtung des Schauers |
| $Ne$,$Age$ | Parameter der NKG-Anpassung |
| $L_{(20-100)}$ | Integral über die Čerenkov-Photonendichte von 20 m bis 100 m Kernort-Abstand |
| $r_l$ | Parameter der AIROBICC Lateral-Anpassung, Lichtradius |

*Tabelle 2.1: Die wichtigsten Ergebnisse der Ereignis-Rekonstruktion*

## 2.5 $\mathbf{M}^{te}$ Carlo-Simulation

Die Simulation der Luftschauer Ereignisse wurde mit einer erweiterten Version des Programms CORSIKA 4.01 [22] durchgeführt. Für die Simulation der hadronischen Wechselwirkung wurden

- bei Schwerpunktsenergien $< 10$ GeV das Isobarenmodell verwendet
- bei Schwerpunktsenergien $> 10$ GeV das Dual Parton Modell verwendet

mit den jeweils dafür aus Beschleunigerexperiment bestimmten bzw. extrapolierten Wechselwirkungsquerschnitten.
Kern-Kern Reaktionen werden unter der Annahme des Superpositionsmodells als n aufeinanderfolgende Kern-Nukleon Reaktionen gerechnet. Der Berechnung der elektromagnetischen Komponente liegt der EGS4 Programmcode [23] zugrunde. Es werden oberhalb einer Teilchenenergie von 3 MeV Annihilation, Compton-Streuung, Bremsstrahlung sowie Bhabha- und Möllerstreuung bzw. für Photonen Paarerzeugung, inverse Comptonstruung und photoelektrische Prozesse berücksichtigt.
Die Abstrahlung von Čerenkov-Photonen der auf diese Weise simulierten (geladenen) Sekundärteilchen des Luftschauer wird bei jedem Transportschritt ermittelt und aus Speichergründen zu Bündeln von 30 bis 200 Photonen zusammengefaßt. Die Zusammenfassung der Čerenkov-Photonen zu Bündeln zeigt ein Punkt auf, der bei der Verwendung von Neuronalen Netzen beachtet werden muß. Die unterschiedlichen Longitudinal Entwicklungen von $\gamma$-und Hadron-induzierten Schauern werden sich in systematischer Weise in die Größe und Dichte der Bündel übersetzen und somit auf dem Niveau der Einzelhütten-Signale zu physikalisch nicht motivierten Systematiken führen, die bei entsprechender Anwendung eines Neuronalen Netzes die physikalischen Eigenschaften überlagern oder sogar unterdrücken können. Grundsätzlich gilt, je elementarer die Eingangsgrößen des Netzes sind, desto detaillierter muß auch die Modellierung der physikalischen Prozesse sein. Die weiteren Schritte der $\mathrm{M}^{te}$ Carlo-Generierung sind :

- Simulation der Blei- und der Holzlage der Szintillator-Detektoren, sowie Reduktion der Datenmenge durch Abtastung der Schauerteilcheninformationen auf einem 7.5m $\times$ 7.5m Gitter und entsprechend der Čerenkov-Photonen auf einem 15m $\times$ 15m Gitter [4] mit dem Programm ELBA [24].
- Nochmalige Reduktion der Abtastung entsprechend der Geometrie des HEGRA Detektors und Anwendung der Triggerbedingung mit dem Programm TEXEL [24]. Vervielfachung der Ereignisse durch mehrfache Anwendung des vorigen Schritts auf ein von ELBA geliefertes Ereignis, für unterschiedliche Kernorte und jeweils um $90^0$ gedrehte Einfallsrichtungen des Schauers.

Insgesamt wurden auf diese Weise mit dem Programm CORSIKA 10170 Schauer erzeugt von denen 240 $\gamma$- und 460 Hadron-induzierte Schauer mindestens ein simuliertes HEGRA-Ereignis auslösen konnten. Durch die mehrfache Ausnutzung der Schauerinformationen mit dem Programm TEXEL standen insgesamt 6576 Schauer zur Verfügung, die den Bereich von 5 bis 300 TeV abdecken. Für die Simulation der Schauer wurde die Komposition bei 10 TeV aus Abschnitt 1.3 verwendet. Sowohl den $\gamma$ als auch den hadronischen Simulationsereignissen wurde ein Energiespektrum mit einem differenziellen Index von -2.7 zugrunde gelegt.

[4] Dies entspricht für beide Matrizen einer gegenüber dem realen Detektor verdoppelten Abtastfrequenz.

| | | $\gamma$ | P | He,O,Fe | $\sum$ |
|---|---|---|---|---|---|
| Nach CORSIKA | 2-12 TeV | 1000 | 5700 | 2400 | 9100 |
| | 12-80 TeV | 360 | 600 | 30 | 990 |
| | >80 TeV | 20 | 30 | 30 | 80 |
| | $\sum$ | 1380 | 6330 | 2460 | 10170 |
| Nach ELBA und TEXEL | insgesamt | 2313 | 3068 | 1195 | 6576 |

*Tabelle 2.2: Anzahl der generierten Schauer der verwendeten M$^{te}$ Carlo-Menge*

Da die Akzeptanz des Detektors mit abnehmender Energie stark nachläßt, wurde die Simulation für verschiedene Energie-Intervalle mit unterschiedlichen Gewichtsfaktoren durchgeführt, um auch in der Nähe der Akzeptanz-Schwelle genügend Ereignisse analysieren zu können. Im weiteren Verlauf der Arbeit wird, falls nicht anders gesagt, immer die oben genannte M$^{te}$ Carlo-Menge mit Gewichtsfaktoren verwendet. Dem Abschnitt 4.6 (idealer Detektor) liegen die von CORSIKA gelieferten Informationen zugrunde.

# Kapitel 3

# Neuronale Netze

Dieses Kapitel bietet eine kurze Übersicht des Themengebiets Neuronaler Netze (NN). Es wird speziell auf die sogenannten Feedforward-Netze eingegangen. Hierbei werden verschiedene Bewertungs- und Interpretationsmöglichkeiten der Netz-Resultate dargestellt, auf das Problem der Überanpassung eingegangen und der Bezug zu anderen Strategien der statistischen Klassifikationsfindung hergestellt. Am Ende dieses Kapitels wird der Ablauf der in dieser Arbeit durchgeführten NN-Analyse dargestellt.

## 3.1 Überblick

Die Entwicklung der Theorie Neuronaler Netze begann 1943 mit den Arbeiten von McCulloch und Pitts. Angeregt durch physiologische Untersuchungen des Hirns entwickelten sie die mathematische Formulierung einer elementaren Verarbeitungseinheit als Summationseinheit mit nachgeschaltetem Diskriminator [25]. McCulloch-Pitts Perceptrons können einfache logische Operationen ausführen, besitzen jedoch eine Reihe von elementaren Einschränkungen, die auf der Linearität des Ansatzes beruhen. Die Blütezeit der Theorie begann erst als durch die Formulierung des Backpropagation Algorithmus [26] Anfang der 80er Jahre eine leistungsfähige Lernmethode für die allgemeinere Klasse der Feedforward-Netze gefunden wurde. Das gesamte Spektrum Neuronaler Netze, die heute Anwendung finden, zerfällt in die Kategorien :

- Lineare und nichtlineare Netze
- Rekursive und nicht rekursive Netze
- Stochastische und nicht stochastische Netze
- Überwachtes und nicht überwachtes Training

Allen Modellen ist gemeinsam: Die elementaren Recheneinheiten des Neuronalen Netzes, die Knoten (Neuronen), erhalten von anderen Knoten Signale, die mit Skalaren, den Werten der synaptischen Matrix, gewichtet und aufsummiert[1] werden. Das Training besteht aus der Veränderung der Gewichte in geeigneter Weise. Die oben genannten Unterteilungen der Modelle unterscheiden sich dabei sowohl in der Art, wie die Knoten aus den Eingangswerten die Ausgabewerte

[1] Teilweise werden die Eingänge auch multipliziert → Sigma-Pi Neuronen.

(die Aktivität) ermitteln, als auch in den Trainingsabläufen und diversen Beschränkungen der erlaubten Verknüpfungen (Netztopologie). Neuronale Netze haben sich in den letzten Jahren für viele Aufgaben der Informationsverarbeitung bewährt. Regressions-Aufgaben, Funktionsapproximation, Zeitserien- Vorhersagen/Analysen, nichtlineare Dimensionsreduktion, Clustering und Klassifikations-Aufgaben sind Beispiele für Anwendungsgebiete. Abgesehen von der neurobiologischen Motivation erscheinen Neuronale Netze so attraktiv, weil bestimmte Modelle, wie z.B. das FeedforwardModell, als universelle Funktionsapproximatoren aufgefaßt werden können, die nach Wahl angemesser Trainingsparameter als semi-parametrische Verfahren angewendet werden und somit nur wenige A priori Annahmen über die den Daten zugrunde liegenden Modelle gemacht werden müssen.

## 3.2 Feedforward-Modell

Feedforward-Netzesind zyklenfreie Neuronale Netze, in denen die Informationsverarbeitung in einer vorbestimmten Richtung stattfindet. Die Knoten werden in Ebenen eingeteilt, die nach ihren Verbindungen in Eingabeschicht, versteckte Schicht und Ausgabeschicht klassifiziert werden. Die Anzahl der Schichten und der Knoten wird nach dem Schema (# Eingabeknoten)-(# Knoten versteckte Schicht)... -(# Ausgabeknoten) angegeben, also z.B. 4-2-2-1. Die Eingabesignale werden an die Eingabeschicht angelegt, deren Ausgänge sind mit den Knoten der folgenden Ebene verbunden, ebenso wie diese wieder mit den Knoten der darauf folgenden Ebene verbunden sind. Auf diese Weise werden die Eingabesignale, von der Eingabeschicht durch die versteckten Schichten bis zur Ausgabeschicht propagiert.
Die Aktivitäten $V$ der Knoten ergeben sich zu:

Eingabeschicht:

$$V_j^1 = g(\sum_k \omega_{jk}^0 x_k^\mu)$$

Verborgene Schicht $n$ :

$$V_j^n = g(\sum_k \omega_{jk}^n V_k^{n-1})$$

Ausgabeschicht:

$$O_j^\mu = g(\sum_k \omega_{jk}^\mu V_k^{\mu-1})$$

$x$ Eingangsaktivität an Knoten $k$ für Muster $\mu$

$g$ Transferfunktion

$\omega_{jk}^n$ Gewicht der Verbindung zwischen Knoten $k$ in der Schicht $n$ zu Knoten $j$ in der Schicht $n-1$

In Abbildung 3.1 sind die Bezeichnungen für ein Netz mit einer versteckten Schicht eingezeichnet. Die Aktivitäten der Ausgabe-Knoten werden als Ausgabevektor, oder im Fall nur eines Knotens als Ausgabewert bezeichnet.
Die Bestimmung der Gewichte $w$ auf Basis der Daten wird als Training (Lernen) bezeichnet und die hierfür verwendeten Daten-Beispiele entsprechend Trainingsmenge. Das Training von Feedforward-Netzen basiert üblicherweise auf der Minimierung einer Kostenfunktion (Energiefunktion, Fehlerfunktion). Unterschiedliche Aufgaben[2] korrespondieren mit unterschiedlichen Kostenfunktionen (siehe Abschnitt 3.3).

[2] und Annahmen über die den Daten zugrunde liegen statistischen Verteilungen

Der bekannteste Trainings-Algorithmus für mehrschichtige Feedforward-Netze ist der Backpropagation Algorithmus, der eine sehr effiziente Formulierung eines Gradienten-Abstiegsverfahrens darstellt. Im Laufe der letzten Jahre wurden eine Vielzahl von ähnlichen aber leistungsfähigeren Algorithmen entwickelt (zur Übersicht siehe z.B. [27]). Allen Algorithmen ist die dynamische Beschreibung durch eine Aktualisierungsregel für die Netz-Gewichte $w^{(n+1)} = w^{(n)} + \delta w^{(n)}$ ($n$ beschreibt die zeitliche Entwicklung und wird als Trainingsschritt bezeichnet) gemeinsam. Unterschiede existieren nur in der Bestimmung der Gewichts-Änderung $\delta w^{(n)}$. Findet die Aktualisierung nach jedem Trainingsmuster statt, spricht man von *normal updating*, findet sie erst nach der Auswertung der Kostenfunktion mit allen Trainingsmustern statt (Epoche), spricht man von *batch updating*.
Vom theoretischen Standpunkt kommt dem Trainings-Algorithmus eine untergeordnete Rolle zu. Entscheidend sind die Wahl der Kostenfunktion, der Transferfunktion, die der Anzahl der Knoten und Schichten sowie eventueller Regularisierungs Parameter. In der Praxis zeigen sich jedoch deutliche Unterschiede der verschiedenen Trainings-Algorithmen bzgl. Konvergenz und Sensitivität auf lokale Minima der Kostenfunktion.

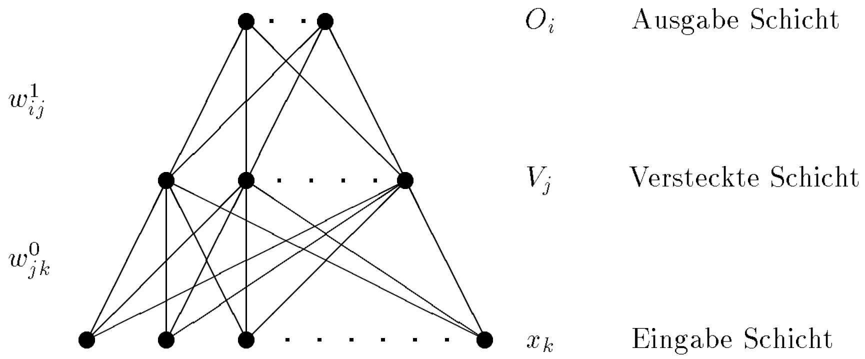



*Abbildung 3.1: Feed-Forward Netz mit einer verborgenen Ebene*

Bei Feedforward-Netzen werden, mit Ausnahme der Ausgabeschicht, faßt ausschließlich auf dem Bereich $[-1, +1]$ oder $[0, +1]$ monotone und beschränkte Funktionen als Transferfunktion verwendet. Eine ganze Klasse dieser Funktionen sind die Sigmoidfunktion, die zusätzlich noch unendlich oft differenzierbar sind und somit ein robustes Verhalten des Netzes bezüglich kleinen Änderungen der Gewichte gewährleisten, z.B.

$$g(x) = \frac{1}{1 + \exp(-x)} \qquad \textit{oder} \qquad g(x) = \tanh(x)$$

Nur in der Ausgabeschicht wird für Regressions-Aufgaben die lineare Funktion $g(x) = x$ verwendete. Die Verwendung *nichtlinearer* Transferfunktionen ist von entscheidener Bedeutung. Ein Neuronales Netz mit linearen Transferfunktionen stellt eine Reihe von linearen Transformation dar, die insgesamt linear sind und durch ein Netz mit nur einer linearen Schicht dargestellt werden können. Im Gegensatz dazu wird bei der Verwendung nichtlinearer Funktionen beim Übergang von keiner zu einer versteckten Schicht eine erhebliche Leistungssteigerung erreicht.

Der Grund hierfür ist, daß die gewichteten Ausgaben der Eingangsschicht in der versteckten Schicht in den Termen des Arguments der Transferfunktionen auftreten, dadurch kann das System der beiden Schichten als Repräsentation einer *adaptiven Funktion* aufgefaßt werden.

Abbildung 3.2 zeigt dies für ein einfaches Klassifikations-Problem (weiße Punkte -1, schwarze Punkte +1) anhand zweier 2-2-1 Netze. Die linke Seite zeigt die aus Geraden bestehende Lösung eines linearen Netzes, die rechte Seite die eines mit der $tan(x)$ Funktion trainierten Netzes. Die Entscheidungsgrenzen (*Diskriminante*) wurden in beiden Fällen aus der Bedingung Netzausgabe gleich Null bestimmt. Die Verwendung einer versteckten Schicht zusammen mit der $tan(x)$ Funktion bewirkt im rechten Bild eine glatte Diskriminante - die Diskriminanten-Funktion ist in der Lage sich "anzuschmiegen".
Daß bereits eine versteckte Schicht aureichend ist, besagt ein ein Satz von Cybenko [28], der zeigt daß für nichtlineare Feedforward-Netze jedes Netz mit mehreren versteckten Schichten zu einem Netz mit nur einer versteckten Schicht äquivalent ist. In der Praxis werden jedoch häufig Netze mit mehr als einer versteckten Schicht angewandt, weil diese Netze während des Trainings ein robusteres Konvergenz-Verhalten zeigen.

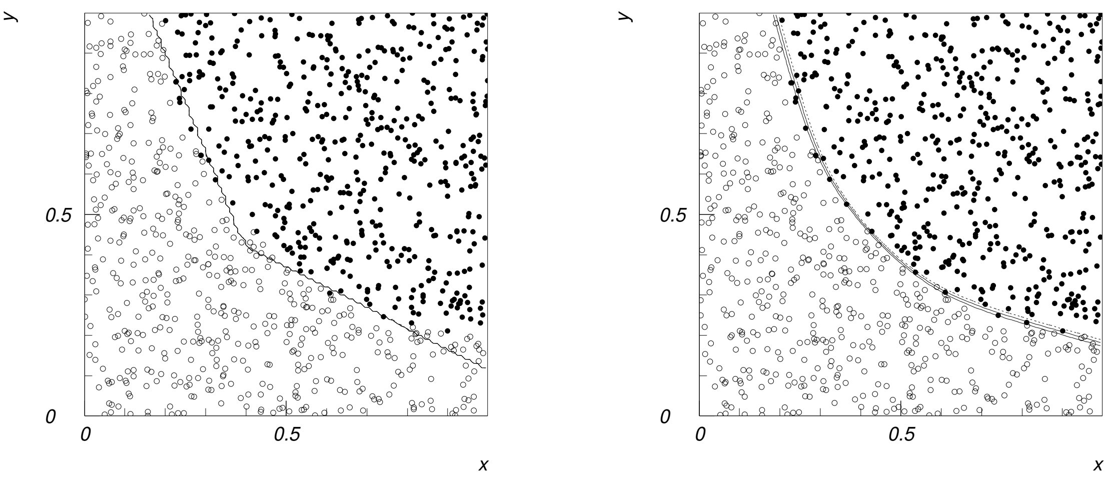



*Abbildung 3.2: Links, das Ergebnis der Klassifikation mit linearen, rechts mit nichtlinearen (logistischer) Aktivierungsfunktionen.*

## 3.3 Klassifikation mit Feedforward-Netzen

Ein übliches Verfahren, Klassifikations-Aufgaben zu lösen, ist die Überlagerung mehrerer unabhängiger linearer Schnitte. Ein linearer Schnitt in den Variablen $x_k$ kann sehr effizient z.B. durch die lineare Fischer-Diskriminante F ausgedrückt werden:

$$\begin{aligned} F &= \Theta(\sum_k a_k x_k - \theta) \\ \Theta(x) &\quad \text{Heavyside Funktion} \\ a_k &\quad \text{Koordinaten-Projektionen der Diskriminante} \end{aligned} \tag{3.1}$$

Wenn man der letzten Formel die Darstellung der Ausgabe-Funktion eines Feedforward-Netzes mit einer versteckten Schicht gegenüberstellt

$$F = \Theta[\sum_j \Theta(\sum_k a_k x_k - \theta)] \tag{3.2}$$

wird deutlich, daß sie als Superposition von Entscheidungen der Form 3.1 aufgefaßt werden kann. Darüber hinaus werden durch den Übergang von der Heavyside- zu Sigmoid-Funktionen glatte Entscheidungsgrenzen modelliert. Zusammengenommen mit der im Kontext der nichtlinearen Transferfunktion erwähnten Fähigkeit der versteckten Schicht zur Approximation adaptiver Basis-Funktionen, stellen sich Feedforward Neuronale Netze also als Verallgemeinerung des Ansatzes der linearer Schnitte dar. Übrig bleibt zu zeigen, daß auch der Minimierung der Netz-Energiefunktion eine sinnvolle Interpretation im Rahmen der mathematischen Klassifikationstheorie hat. Der Rest dieses Abschnitts richtet sich nach den Darstellungen in [29].
Ausgehend von dem Prinzip des Maximalen Likelihood, das besagt daß diejenige Wahrscheinlichkeitsdichte $p(t, x)$ die Daten am besten beschreibt, die die Likelihood Funktion $L$ maximiert, definiert man die Energiefunktion E des Netzes als negativen Logarithmus des Likelihood.

$$E := -ln\,(L) = -\sum_n ln\, p(t^n \mid x^n) \tag{3.3}$$

Wobei $x^n$ Eingabevektoren und $t^n$ die Zielvektoren darstellen. Unter der Annahme, daß die einzelnen Komponenten $(t_k)$ der Trainingvektoren voneinander unabhängig sind, ergibt sich:

$$p(t \mid x) = \prod_{k=1}^{c} p(t_k \mid x) \tag{3.4}$$

Und somit für die Energiefunktion des Neuronalen Netzes:

$$E = -ln(L) = -\sum_{n=1}^{N} \sum_{k=1}^{c} p(t_k^n \mid x^n) \tag{3.5}$$

An dieser Stelle der Herleitung muß eine Annahme über die Form der Verteilung $p(t \mid x)$ gemacht werden. Falls man den Zielvariablen $t$ ein Modell bestehend aus einer analytischen Funktion h und einem additiven gaussverteiltem Rauschen $\epsilon$, mit einem Mittelwert von 0 und einer von $x$ und $k$ unabhängigen Standardabweichung $\sigma$, zu Grunde legt

$$t_k = h_k(x) + \epsilon_k \tag{3.6}$$

und dies in 3.5 einsetzt, ergibt sich :

$$p(t_k \mid x) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp(-\frac{(y_k(x;w) - t_k)^2}{2\sigma^2}) \tag{3.7}$$

In Formel 3.7 wurde $h_k(x)$ durch $y_k(x;w)$, die Netzwerkausgabe des k-ten Knoten für den Eingabevektor $x$ bei festen Netzgewichten $w$, ersetzt. Diese Formel eingesetzt in 3.5 ergibt nun direkt die übliche Form der summierten und quadrierten Energiefunktion (*summed squared error function*).

$$E \propto \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{c} (y_k(x^n;w) - t_k^n)^2 = \frac{1}{2} \sum_{n=1}^{N} \|y(x^n;w) - t^n\|^2 \tag{3.8}$$

Die von $w$ unabhängige Terme und Faktoren wurden in Formel 3.8 weggelassen, da sie für den Lernvorgang unwesentlich sind. Entsprechende Beweise existieren auch für die anderen Formen des Modells 3.6, z.B. folgt aus der Annahme einer den Trainingsvektoren zugrunde liegenden Binomialverteilung die bedingte Wahrscheinlichkeit $p(t \mid x, w) = y^t(1-y)^{1-t}$ und entsprechend wie oben die *Cross Entropy* Funktion als Kostenfunktion des Neuronalen Netzes (siehe 4.4.3). Wie die Ableitung der *summed squared Error* Kostenfunktion zeigt, gilt sie streng nur für Regressionsprobleme, bei denen die $t_k$ kontinuierliche Größen sind. Da sie jedoch in praktisch allen Arbeiten mit vergleichbarer physikalische Fragestellung verwendet wird und sich für die Endresultate aus Kapitel 4 keine Unterschiede zeigten, wird mit Ausnahme von Kapitel 4.4.3 im weiteren immer die Form 3.8 der Kostenfunktion verwendet.

## 3.4 Generalisierung und Überanpassung

Als Generalisierungsleistung wird die Fähigkeit eines Neuronalen Netzes bezeichnet, die während der Trainingsphase anhand der Trainingsmenge erlernte Aufgabe auf eine disjunkte Testmenge zu übertragen. Als Bewertung wird dieselbe Kostenfunktion verwendet wie während der Trainingsphase, nur daß die Summation über die Vektoren der Testmenge ausgeführt wird.
In dem Fall, daß eine ausreichend große Menge von M$^{te}$ Carlo-Ereignissen vorliegt, wird diese in eine Trainingsmenge und eine Testmenge zerlegt; diese Methode wird als Testmengen-Validation bezeichnet. [3] In der Literatur werden für das Verhältnis von Trainings- zu Testmenge üblicherweise 1:1 bis 3:1 als Richtwerte angegeben. Insbesondere für den Fall, daß nur eine begrenzte Anzahl von Ereignissen zu Verfügung steht, kann von diesen Werten abgewichen werden. Dabei sind zwei konkurrierende Effekte zu berücksichtigen:
**(A)** Ein kleine Trainingsmenge bedingt eine große Modellvarianz, d.h. insbesondere bei einem Netz mit einer großen Anzahl von Knoten (Freiheitsgraden des Modells) wird für die Trainingsmenge eine Lösung gefunden, die für die Menge optimal ist, aber für die Testmenge und somit dem gestellten Problem suboptimal ist.
**(B)** Andererseits kann mit einer zu kleinen Testmenge der Generalisierungsfehler aufgrund der kleinen Stichprobenanzahl nur ungenau bestimmt werden.

[3] Gelegentlich wird die Gesamtmenge auch in Trainings-,Test- und Validierungsmenge unterteilt,wobei die Testmenge zur Überprüfung der Generalisierung während und die Validationsmenge zur Überprüfung nach dem Training verwendet wird. Diese Verfahren ist insbesondere bei der Regularisierung durch *early stopping* weniger problematisch.

Der Fall **(A)** wird als Überanpassung (*overfitting*) des Problems bezeichnet. Es gibt eine Reihe von Lösungsansätzen für dieses Problem; die bekanntesten sind:

- Regularisierung (im Kontext von Neuronalen Netzen häufig als *pruning* bezeichnet) durch einen Zusatzterm in der Kostenfunktion. Die einfachste Form ist der *weight decay* Term.

  $$E \to E + \frac{\lambda}{2} \sum_{i,j} w_{ij}^2 \quad ; \quad i, j \text{ indizieren die Knoten des Netzes}$$

  Große Gewichte werden durch diesen Term bestraft, insgesamt reduziert sich die Anzahl der effektive Parameter des Netzes [30]. Nachteilig bei diesem Verfahren ist, daß die lokalen Minima der modifizierten Kostenfunktion nicht mit denen der eigentlichen übereinstimmen. In der Praxis stellt dies bei nur leichter Regularisierung allerdings kein Problem dar.

- Regularisierung durch *early stopping*, d.h. das Training wird abgebrochen, wenn die Generalisierungsleistung über mehrere Trainingsschritte abnimmt[4].

- Regularisierung durch Anpassung der Anzahl der Eingabevariablen und Knoten an das Problem (unter Berücksichtigung der Anzahl von Trainingsmustern).

Abbildung 3.3 zeigt den Effekt der Überanpassung für das Beispiel aus Abschnitt 3.2. Die Daten wurden zusätzlich so verrauscht, daß die Klassenzugehörigkeit jedes zehnten Datenpunktes vertauscht wurde. Oben links ist die Diskriminante eines 2-12-12-1 Netzes trainiert mit 3000 Trainingsmustern zu sehen (der Deutlichkeit wegen wurde nur jeder zehnte Punkt eingezeichnet). Rechts davon ist das Ergebnis desselben Netzes trainiert mit nur 90 Mustern zu sehen. Es ist offensichtlich, daß das mit 90 Mustern trainierte Netz die den Daten zugrunde liegende Wahrscheinlichkeitsverteilung unter dem Einfluß der Störungen nicht mehr vollständig extrahieren kann. Durch Regularisierung mit einem *weight decay* Term von $\lambda = 0.12$ (unten links) nähert sich die Lösung auch bei 90 Mustern der Lösung mit 3000 Mustern an. Die Reduktion der effektiven Parameter des Netzes durch den *weight decay* Term bewirkt eine glattere Entscheidungsgrenze. Als letztes Beispiel ist die Regularisierung des Netzes durch Anpassung der Knotenanzahl gezeigt. Die Knotenanzahlen der versteckten Schichten wurden hierbei durch die in Abschnitt 3.5.3 vorgestellten PCA und CDA Methoden festgelegt. Diese Form der Regularisierung wird in Kapitel 4 verwandt werden.

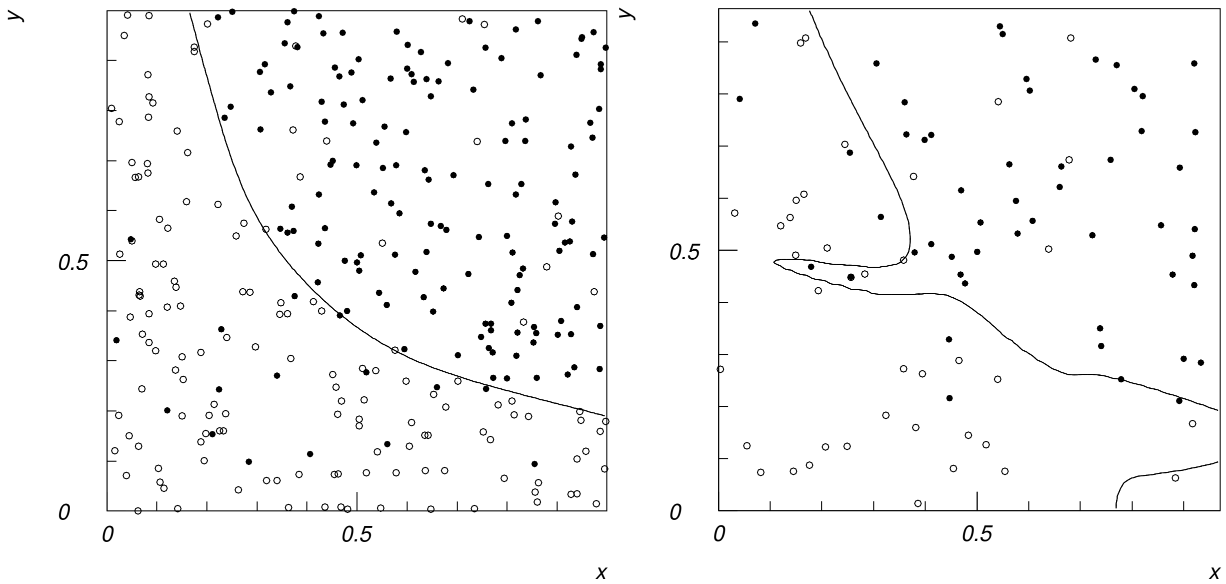
## 3.5 Bewertungs- und Interpretationsmöglichkeiten

### 3.5.1 NCV-Methode

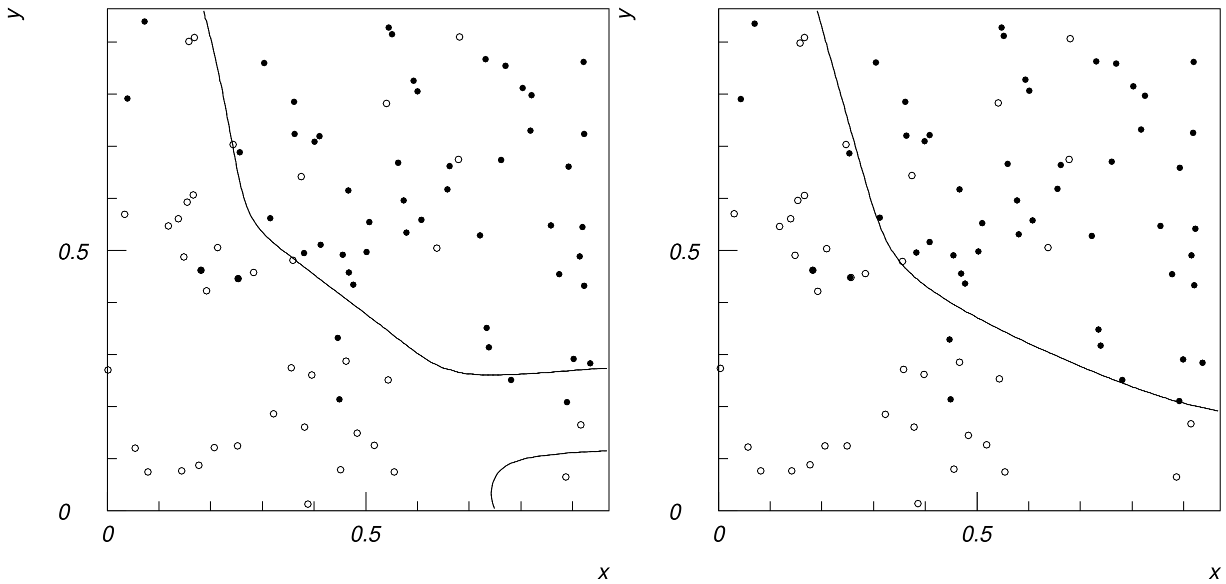
Im vorigen Abschnitt wurde gezeigt, daß die Gefahr der Überanpassung drastisch mit der Reduktion der Trainingsmenge ansteigt. Dieses Problem umgeht die Methode der Kreuz Validation durch Verwendung der gesamten Daten-Menge, sowohl für die Trainingsmenge als auch für die Testmenge. In ihrer einfachsten Form, der *leave one out cross validation*, nimmt sie folgende Gestalt an:

$$CV(\lambda) = \frac{1}{N} \sum_{j=1}^{N} t_j - \mu_{\lambda(j)(x_j)} \tag{3.9}$$

[4] Dies ist gleichbedeutend mit dem Ansteigen des Trainingsfehlers.




*Das 2-12-12-1 Netz trainiert mit 3000 Datenpunkten (links, nur jeder zehnte Punkt ist gezeichnet), und mit 90 Datenpunkten (rechts).*




*Das Netz mit 90 Datenpunkten trainiert, aber durch einen weight decay Term* $\lambda = 0.12$ *regularisiert (links). Ein mit der aus der PCA-Analyse gewonnenen Information auf 2-2-2-1 reduziertes Netz (rechts).*

*Abbildung 3.3: Beispiel zur Überanpassung, Erläuterung im Text.*

Das j-te Muster der Gesamtmenge $\lambda$ wird als Testmuster für das mit der Menge $\lambda(j) = \{\lambda \text{ ohne } x_j\}$ trainierte Netz $\mu$ verwendet. Der Generalisierungsfehler $CV(\lambda)$ ergibt sich als Ensemble-Mittelwert dieser Netze. Der mit Formel 3.9 verbundene Rechenprozeß ist extrem zeitaufwendig, so daß in einer Abwandlung statt dessen nicht nur ein Trainingsmuster ausge-

lassen wird, sondern eine größere Untermenge (*$\nu$-fold cross validation*). Nichtlineare Verfahren besitzen im allgemeinen viele lokale Minima. Jedes lokale Minimum repräsentiert eine erlaubte Lösung des Modells und entspricht einem Satz von Netz-Gewichten.
Dies macht eine Modifikation an dem vorgestellten Verfahren notwendig. Um die Mittelwertbildung über möglichst ähnliche Lösungen des Modells durchzuführen, wird die Validierung nicht mit neu initialisierten Anfangszuständen, sondern mit einem bereits austrainierten Zustand begonnen. Abbildung 3.4 veranschaulicht die Methode der nichtlinearen Kreuz-Validation (**N***onlinear* **C***ross* **V***alidation, NCV*):

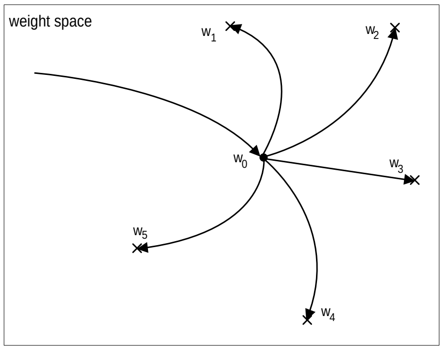



*Abbildung 3.4: Skizze zum Prinzip der nichtlinearen Kreuz-Validation*

Das Neuronale Netz wird bis zum Zustand $w_0$ trainiert. Von diesem Zustand aus wird das Training mit den Mengen $T_n$ fortgeführt, die resultierenden Zustände sind mit $w_n$ bezeichnet. Der Generalisierungsfehler ergibt sich als Mittelwert der Validation mit den Mengen $V_n$ auf den Zuständen $w_n$.
Die Mengen $T_n$ und $V_n$ ergeben sich wie bei der *$\nu$-fold cross validation* aus den Bedingungen $T_n \bigcup V_n = \lambda$, $\bigcup_n V_n = \lambda$, $V_n \bigcap V_m = \emptyset$ für alle n$\neq$m.
Abbildung 3.5 zeigt anhand des vorigen Beispiels (Abbildung 3.3, unten rechts) den Vorteil der Methode bei einer begrenzten zur Verfügung stehende Anzahl von M$^{te}$ Carlo-Mustern. NN1 wurde mit 90 Mustern trainiert und mit 10 Mustern getestet. NN2 wurde wie NN1 trainiert aber mit 1000 Mustern getestet.
Der Generalisierungsfehler für NN1 ist klar zu groß abgeschätzt, während er für NN2 dem durch das Rauschen verursachte erwarteten Fehler von 0.1 entspricht. Der für NN1 ermittelte NCV-Fehler ist am rechten Rand eingezeichnet, er beträgt 0.07 und ist somit eine deutlich bessere Schätzung der wahren Generalisierungleistung. Wie in diesem Beispiele wird der Generalisierungsfehler im allgemeinen mit der NCV-Methode leicht zu klein abgeschätzt, da das Neuronales Netz die zu $V_n$ gehörenden Muster erst "vergessen" muß. Dieser Vorgang ist in dem Kasten oben rechts zu sehen. Die Fluktuation der Fehler zwischen den verschiedenen NCV-Mengen wird durch die kleinen Testmengen verursacht.

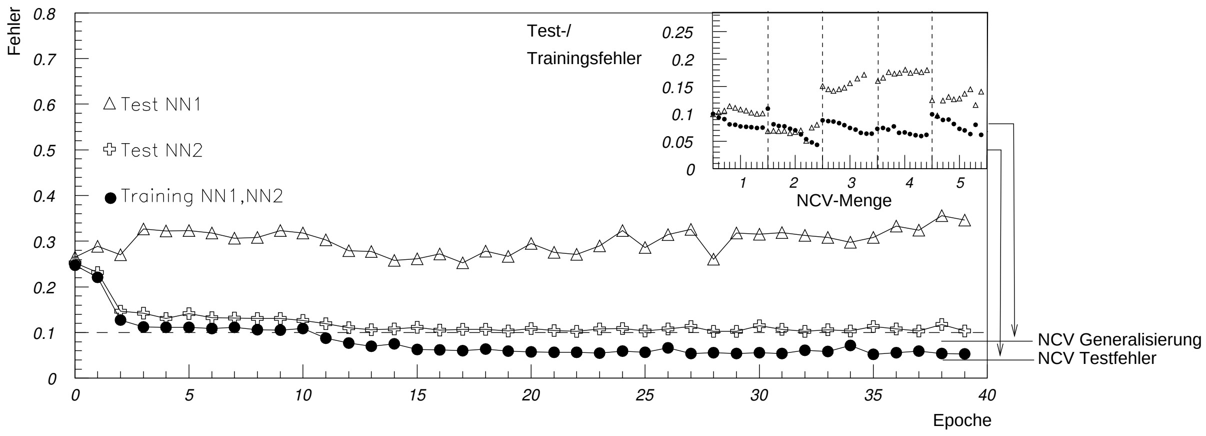



*Abbildung 3.5: Verlauf des Trainings-/Testfehlers für NN1 und NN2.*
*Oben rechts ist der Verlauf für die 5 NCV-Menge (NN2) eingezeichnet, Erläuterung im Text.*

### 3.5.2 Relevanzen und skalierte Gewichte

**Relevanzen der Eingangsgrößen**

Einer der Vorteile Neuronaler Netze liegt in der einfachen Anwendbarkeit der Methode auf mehrdimensionale Eingängsräume mit vielen Variablen. Die in Abschnitt 3.4 dargestellte Gefahr von Überanpassung zeigt jedoch, daß eine möglichst geringe Anzahl von Eingabegrößen wünschenswert ist. Abgesehen von der Selektion der Eingangsgrößen durch physikalisch motivierte Vorüberlegungen formuliert sich daraus die Forderung nach der Bestimmung einer Teilmenge der für die Netzantwort wesentlichen Variablen. Die aufwendigste Art dies durchzuführen ist das sukzessive Löschen des der Eingabegröße zugeordneten Knotens und das erneute Training des Netzes. Da hierfür bei N Eingangsgrößen $2^N - 1$ Netze überprüft und bewertet werden müssen, ist diese Verfahren extrem rechenzeitaufwendig[5]. Ein häufig verwendetes Verfahren, das ohne erneutes Training auskommt, ist die Bewertung der Eingangsgrößen durch folgenden Ausdruck, der im weiteren Relevanz der Eingangsgröße genannt wird:

$$R_i \quad := \quad \frac{1}{N}\sqrt{\sum_{j=1}^{N}(o(\mathbf{x}_j) - o(\mathbf{x}_j \mid_{x_{j,i}=\mathbf{x}_i}))^2} \tag{3.10}$$

$N$ Anzahl der Eingabemuster

$o(\mathbf{x}_j)$ Netzausgabe für $\mathbf{x}_j$

$o(\mathbf{x}_j \mid_{x_{j,i}=\mathbf{x}_i})$ mit $\mathbf{x}_i = \frac{1}{N}\sum_{j=1}^{N} x_{j,i}$

Daß diese Methode nur einen ersten Hinweis auf die Bedeutung der Eingabegrößen liefert und bei einer genaueren Analyse überprüft werden muß, kann anhand eines einfachen Beispiels verdeutlicht werden :

[5] In Kapitel 4 wird eine hierzu ähnliches Verfahren verwendet werden.

Gegeben sei ein NN mit den drei Eingangssignalen a,b und c. Die Größen a und b seien miteinander korreliert und ermöglichen jede zusammen mit c eine vollständige Separation der Trainingsmenge in zwei Klassen. Falls a oder b durch den entsprechenden Mittelwert ersetzt werden, wird der Einfluß der jeweils anderen Größe übrig bleiben und $R_{a/b}$ entsprechend klein ausfallen. Falls c auf seinen Mittelwert gehalten wird, wird $R_c$ entsprechend groß ausfallen. Das Ergebnis der Analyse wird also sein, daß a *und* b im Vergleich zu c weniger wichtig sind, obwohl mit c ohne eine der beiden anderen Größen keine Lösung der Aufgabe gefunden werden kann[6].

**Skalierte Gewichte**
Um die Funktionsweise zu verstehen und die Verläßlichkeit des Netzwerks abzuschätzen, ist es wichtig, nicht nur den Einfluß bestimmter Eingaben zu quantifizieren, sondern auch zu untersuchen, auf welche Weise die Eingaben verwandt werden. Da die Funktionsweise durch die Gewichte der Verbindungen bestimmt wird, also in den Elementen der synaptischen Matrix kodiert ist, kann deren Visualisierung für eine Interpretation benutzt werden. Wegen der unterschiedlichen Aktivitäten der Knoten enthälten die Gewichte der versteckten Knoten einen unbestimmten Normalisationsfaktor. Um die Darstellung davon zu befreien, werden skalierte Gewichte [31] definiert :

$$w'_{ij} = \frac{1}{< x_i >} * w_{ij} \tag{3.11}$$

wobei $< x_i >$ den Mittelwert des i-ten Eingangs des betrachteten Knotens über alle Trainingseingaben darstellt. Insbesondere bei Netzen mit mehreren versteckten Schichten ist dies Verfahren problematisch, da zu derselben Kostenfunktion unterschiedliche Aktivitätsmuster innerhalb und zwischen den Schichten korrespondieren können[7]. Für Eingangssignale mit räumlicher Anordnung ist diese Form der Visualisierung oft sehr hilfreich (vgl. Abschnitt 4.5.2).

### 3.5.3 PCA- und CDA-Analyse

Eines der schwierigsten Probleme bei der Anwendung von Neuronalen Netzen ist die Bestimmung einer geeigneten Netzwerkarchitektur. Im Falle der Feedforward-Netze reduziert sich dies Problem auf die Frage nach der Anzahl der versteckten Schichten und der Knoten pro Schicht. Die **P**rincipal **C**omponent **A**nalysis (PCA) ist eine Statistik-Methode, um aus einer Menge M von teilweise korrelierten Größen durch Linearkombination eine Menge M' von unkorrelierten (orthogonalen) Variablen zu konstruieren. Dazu ermittelt man zunächst den Richtungsvektor der stärksten Datenänderung (Varianz groß) als neuen Basisvektor. Dann wird sukzessiv zu den bereits gefundenen Basisvektoren der Basisvektor in Richtung der größten Varianz orthogonal dazu bestimmt, bis der Raum der Variablen aufgespannt ist. In der Physik entspricht dieses Verfahren der bekannten Hauptachsentransformation und, wie weiter unten zu erkennen ist, findet die Masseverteilung bei der Anwendung auf NN ihre Entsprechung in der Verteilung der Aktivitäten.
Diese PCA-Methode, angewendet auf die Aktivitäten (bzw. Ausgabewerte) $O_i^\mu$ der Knoten einer versteckten Schicht, liefert ein Maß für Komplexität der von dieser Schicht repräsentierten

[6] Für weitere Beispiele: [27]

[7] Z.B. ist die Energiefunktion invariant gegenüber Permutationen der Knoten und der entsprechenden Gewichte innerhalb einer versteckten Schicht.

Information [32]. Indem man die Dimension $d_h$ der Information im Raum der Aktivitäten durch abzählen der wesentlich von Null verschiedenen Eigenwerte bestimmt, erhält man die minimale Anzahl der Knoten dieser Schicht, die ausreichend sind, um die Informationen dieser Schicht zu repräsentieren.
Im Einzelnen ergibt sich folgendes Verfahren :

- Bestimmung der Kovarianz-Matrix $C_{ij} = \sum_\mu (O_i^\mu - O_i^{mittel}) \cdot (O_j^\mu - O_j^{mittel})$
  $O_i^\mu$ = Aktivität des Knotens i in der betrachteten Layer bei Test mit Eingabe $\mu$
  $O_i^{mittel}$ = Mittelwert der Aktivitäten des Knotens i für die Testmenge

- Bestimmung der Eigenwerte $\lambda_k$ von $C_{ij}$, die wegen der Symmetrie der Matrix existieren und alle positiv sind

- Diese Eigenwerte stellen die Kovarianzen und somit die Breite der Aktivitätsverteilung in einem Orthogonalsystem dar.

- Die Dimension wird als Anzahl der wesentlich von Null verschiedenen Eigenwerte definiert, da die Aktivitätsverteilung nur in Richtung der dazu konjugierten Eigenvektoren eine wesentliche Ausdehnung hat.

Im Gegensatz zur PCA-Analyse nutzt die **C**anonical **D**iscriminant **A**nalysis (CDA) die Zuordnung eines Eingabemusters zu einer Klasse aus. Die Information wird verwendet, um Richtungen im Eingaberaum zu suchen, in denen Muster innerhalb der Klassen möglichst nahe angeordnet sind, während gleichzeitig die Separation zwischen den Klassen maximal wird.

- Bestimmung der Matrix WSS (*within summed square*):
  $\mathrm{WSS}_{ij} = \sum_{\mu,g} (O_i^\mu - m_i^g) \cdot (O_j^\mu - m_i^g)$
  $m^g$ Mittelwert der Aktivitäten der Klasse g;
  die Summe $\mu$ läuft über die Elemente der Klassen.

- Bestimmung der Matrix BSS (*between summed square*):
  $\mathrm{BSS}_{ij} = \sum_g (m_i^g - M_i) \cdot (m_j^g - M_j)$
  mit $M$ Mittelwert über alle Aktivitäten

- Berechnung der Eigenvektoren der Matrix $\mathrm{WSS}^{-1}\mathrm{BSS}$

Die gesuchten Richtungen entsprechen auch hier wieder den Eigenvektoren der Matrix. Im Unterschied zur PCA-Analyse ist die resultierende Matrix nicht notwendigerweise symmetrisch und somit sind die Eigenvektoren nicht orthogonal. Dies führt bei der Darstellung der Eingabemuster in dem neuen Koordinatensystem der Eigenvektoren der CDA-Matrix zu einer verzerrten Darstellung. Da auch hier wieder nur die Absolutwerte der Eigenwerte als Maß der repräsentierten Information verwendet werden sollen, stellt dies jedoch kein Problem dar.
Zusammen zeichnen PCA- und CDA-Analyse ein Bild der in den versteckten Schichten gewählten Repräsentation. In dem Fall, daß die gewählte Klasseneinteilung der Analysemenge unangemessen ist oder eine weitere Klasseneinteilung sinnvoll erscheinen könnte, zeigt sich dies in einer Diskrepanz der beiden Methoden, da nur in der CDA-Analyse die vorgenommene Klassenunterteilung explizit enthalten ist.

### 3.5.4 Bemerkungen zur Durchführung der NN-Analyse

Als Basis der Implementation wird die Funktionsbibliothek JETNET 3.0 [33] verwendet. JETNET stellt eine Reihe von Routinen zur Simulation von Feedforward-Netzen sowie von selbstorganisierenden Mappen (SOM) zur Verfügung. Die Wahl der Aktivierungsfunktionen, der Netztopologie, des Lern-Algorithmus und der Lernparameter wird durch eine Reihe von Parametern gesteuert, die von dem externen Programm CNET gesetzt werden. Falls nicht anders angegeben, gelten für die in Kapitel 4 benutzen Netze folgende Einstellungen:

- Für $\gamma$-induzierte Ereignisse wird eine -1, für Hadron-induzierte Ereignisse eine +1 als Trainingswert vorgegeben.
- Das Training wird im *batch modus* mit dem RPROP Algorithmus [34] durchgeführt und einer Lernrate[8], die je nach Gesamtanzahl der Knoten zwischen 0.01 und 0.1 liegt. RPROP ist ein lokal adaptiver Lernalgorithmus, der wesentlich schneller konvergiert und bezüglich der Start-Parameter robuster ist, als der gewöhnliche Backpropagation Algorithmus
- Es wird kein Regularisierungsparameter verwendet, da statt dessen die Anzahl der Knoten reduziert oder das Training bei Ansteigen des Generalisierungsfehlers abgebrochen wird.

Die Eingangsvariablen werden, wie in Bild 3.6 gezeigt, vor der Verwendung für das Neuronale Netze so skaliert, daß sie einen vergleichbaren dynamischen Bereich haben. Die Skalierung wirkt

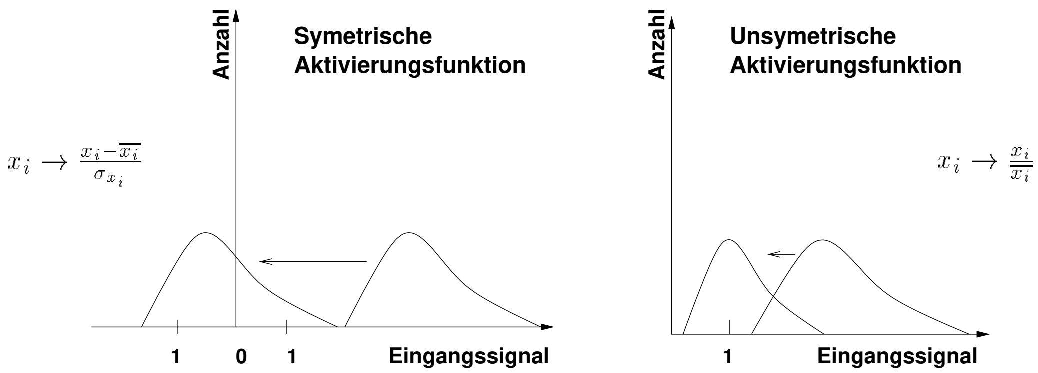



*Abbildung 3.6: Methode zur Skalierung der Eingangsgrößen*

sich nicht auf die gefundene Lösung aus, kann jedoch unter Umständen das Konvergenzverhalten des Neuronalen Netzes verbessern [27].
Aus den in Abschnitt 2.5 gegebenen Gründen sind die Ereignisse mit einem Gewichtsfaktor versehen, dieser Faktor wird während des Trainings durch eine Gewichtung der zugehörigen Netz-Ausgabe berücksichtigt. Dies entspricht einer Gewichtung der einzelnen Terme in der Kostenfunktion 3.8. Die Auswirkung dieser Korrektur ist verhältnismäßig gering, da die Gewichtsfaktoren nicht stark variieren und wirkt sich unterhalb von 5% auf das Endresultat[9] aus.

[8] Als Lernrate wird der Parameter des Training-Algorithmus bezeichnet, der die Geschwindigkeit des Gradienten-Abstiegs bestimmt.

[9] D.h. auf den für die $\gamma/H$-Separation entscheidenden Q-Faktor, siehe nächste Seite.

Simultan zum Training wird nach jeder Lernepoche die PCA-Analyse und nach jeder fünften Lernepoche die zeitaufwendigere CDA-Analyse[10] durchgeführt. Am Ende des Trainings wird die NCV-Methode durchgeführt, die 5 disjunkten Testmengen der Analyse werden zusammengestellt und zur abschließenden Bestimmung des Q-Faktors, der für die $\gamma$-Quellensuche ein Maß der erreichten Qualität der $\gamma/H$-Separation angibt (siehe Abschnitt 4.2), verwendet.
Die in Kapitel 4 gegebenen Zahlenwerte sind also als Ensemble-Werte über 5 Netze mit gleichen Eigenschaften, aber voneinander abweichenden Trainingsmengen, aufzufassen.
Dieses Verfahren stellt eine Übertragung der Idee der Kreuzvalidation auf die Bestimmung abgeleiteter Bewertungsmerkmale dar.
Um aus der kontinuierlichen Netz-Ausgabe ein binäres Entscheidungs-Merkmal zu extrahieren, wird die Berechnung des Q-Faktors[11] für alle möglichen Schnittwerte $N_c$ der Netzausgabe durchgeführt und der optimale Q-Wert aus dem Maximum des erzeugten Histogramms bestimmt. Für die Bestimmung des statistischen Fehlers wird die oben beschriebene Berechnung für 11 disjunkte Teilmengen durchgeführt und der Fehler auf den Mittelwert dieser Verteilung errechnet. Der auf diese Weise für den Q-Faktor ermittelte Fehler ist in guter Übereinstimmung mit der Annahme, daß die Anzahl der Ereignisse mit einer Netz-Ausgaben kleiner bzw. größer $N_c$ einer Binomialverteilung entstammen. Insgesamt stellt sich der Ablauf wie in Abbildung 3.7 dar.

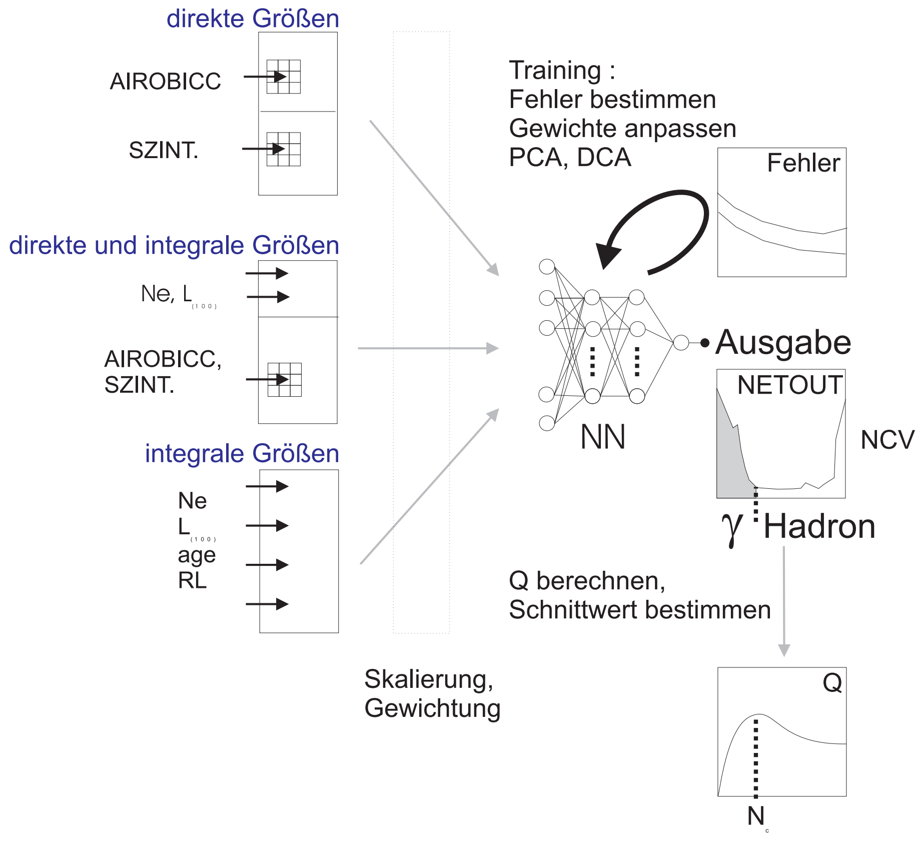



*Abbildung 3.7: Ablauf des Trainings und der anschließenden Bewertung*

[10] Die CDA-Routinen basieren auf Quellcode, der dem Programmpaket ASPIRIN entnommen wurde.

[11] $Q := \frac{N_\gamma(<N_c)\sqrt{N_{bg}}}{N_\gamma\sqrt{N_{bg}(>N_c)}}$ $N_\gamma$ gewichtete Anzahl der in der Testmenge enthaltenen $\gamma$-Ereignisse, $N_\gamma(<N_c)$ $\gamma$-Ereignisse mit Netz-Ausgabe kleiner $N_c$. Entsprechend $N_{bg}$ und $N_{bg}(>N_c)$ für Hadron-induzierte Ereignisse.

# Kapitel 4

# Anwendung von Neuronalen Netzen zur $\gamma/H$-Separation mit atmosphärischem Čerenkovlicht

Für die $\gamma/H$-Separation mit den HEGRA Szintillator- und AIROBICC-Matrizen wurden im Rahmen der HEGRA-Kollaboration die Methoden LES [19] und PEARL [35] entwickelt.
In diesem Kapitel werden die physikalischen Grundlagen dieser Verfahren anhand von M$^{te}$ Carlo-Studien erläutert. Es wird gezeigt, daß Feedforward-Netze bei gleichen Eingabegrößen zu äquivalenten Aussagen und Ergebnissen führen. Daran anschließend werden Netze diskutiert, die ebenfalls auf diesen Größen basieren, aber bei denen von der Möglichkeit Neuronaler Netze Gebrauch gemacht wurde, Korrelationen in höherdimensionalen Merkmalsräumen aufzufinden und in den Separationsprozeß einzubeziehen.
Um zu klären, ob die für den obigen Ansatz verwandten Größen, die bereits stark vorverarbeitet und von ihrer Konstruktion her integrale Größen sind, den kalibrierten Detektor-Meßgrößen (im folgenden direkte Größen genannt) vorzuziehen sind, werden verschiedene Möglichkeiten vorgestellt, die direkte Größen einbeziehen.

## 4.1 Physikalische Grundlagen

### 4.1.1 Longitudinale Entwicklung

Die Abbildung 4.1, links zeigt die Entwicklung der elektromagnetischen Komponente am Beispiel zweier exemplarischer Schauer (20 TeV $\gamma$, 25 TeV Proton), aufgetragen ist die Anzahl der elektromagnetischen Schauerteilchen mit Energien höher als 3 MeV für verschiedene atmosphärische Tiefen. Im Energiebereich von einigen TeV erzeugen primäre $\gamma$ im Mittel 30% mehr Schauerteilchen als Protonen gleicher Primärenergie. $\gamma$- und Proton-induzierte Schauer erreichen ihr Maximum ungefähr bei der gleichen atmosphärischen Tiefe. Zusammen mit der Tatsache, daß Proton-induzierte Schauer langsamer nach dem Maximum abklingen (siehe 1.4.1), folgt daraus, daß sich auf der atmosphärischen Höhe des HEGRA-Detektors die beiden Longitudinalentwicklungen für einen großen Primärenergiebereich schneiden.

Der AIROBICC-Detektor mißt das insgesamt abgestrahlte Čerenkovlicht eines Luftschauers, daß in erster Näherung proportional zum Integral über die Wegstrecke der Čerenkovlicht-emittierenden Teilchen ist (dies entspricht der Fläche unterhalb der Kurve in Abbildung 4.1, links). Daher besitzt der AIROBICC-Detektor eine niedrigere Energieschwelle für $\gamma$-als für Hadron-induzierte Schauer. Im Gegensatz dazu hat der Szintillator-Detektor die gleiche Energieschwelle für beide Primärteilchensorten. Die $\gamma/H$-Separation mit Hilfe der Kombination aus AIROBICC- und Szintillator-Detektor basiert auf dem folgenden Prinzip:
Der AIROBICC-Detektor selektiert aufgrund der Ereignisauslösung an der Nachweisschwelle Schauer mit gleichem Čerenkovlichtgehalt. Mit der Messung der Anzahl der elektromagnetischen Partikel $Ne$ lassen sich dann $\gamma$- induzierte Schauer anreichern (Abbildung 4.1, rechts). Oberhalb der Energieschwelle läßt sich dasselbe Prinzip anwenden, wenn man auf das Verhältnis von Čerenkovlichtmenge zur Anzahl der elektromagnetischen Partikel auf Detektorniveau[1] schneidet, dies ist in Abbildung 4.2, links zu sehen[2]. Mit der Kombination aus $Ne$ und dem Čerenkovlicht $L_{(20-100)}$ kann also die unterschiedliche longitudinale Entwicklung von $\gamma$-und Hadron-induzierten Schauern zur $\gamma/H$-Separation ausgenutzt werden. Obwohl sich die Profile der Longitudinalent-

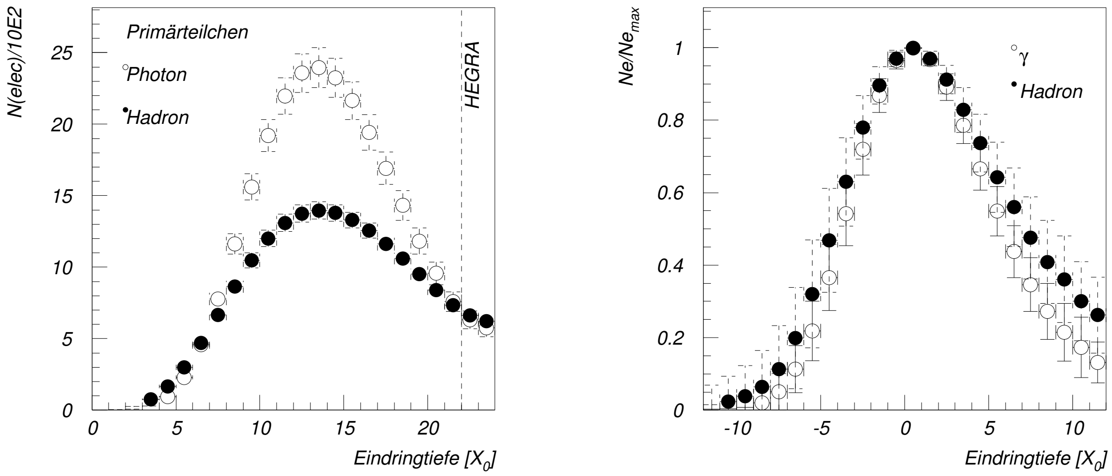



*Abbildung 4.1: Vergleich der longitudinale Entwicklung von $\gamma$-und Hadron-induzierten Luftschauern. Aufgetragen ist die Anzahl der $e^+$ und $e^-$ oberhalb von 3 MeV für verschiedene atmosphärische Tiefen. Rechts wurden die Maxima der longitudinalen Profile in horizontaler Richtung in Deckung gebracht und auf eins normiert. Erläuterung im Text.*

wicklung im Mittel deutlich voneinander unterscheiden, wird die $\gamma/H$-Separation durch die individuelle Fluktuation der Schauer in dem für die $\gamma$-Quellensuche besonders interessanten Energiebereich knapp oberhalb der Detektorschwelle erschwert.

[1] Dieser Quotient wird im folgenden Licht- zu Partikel-Verhältnis genannt.

[2] Die in Abbildung 4.2 dargestellten Meßgrößen basieren auf dem in Abschnitt 4.6 beschriebenen idealen Detektor. Beide Effekte zeigen sich ebenso für die Observablen der HEGRA-Rekonstruktion, sind aber weniger deutlich.

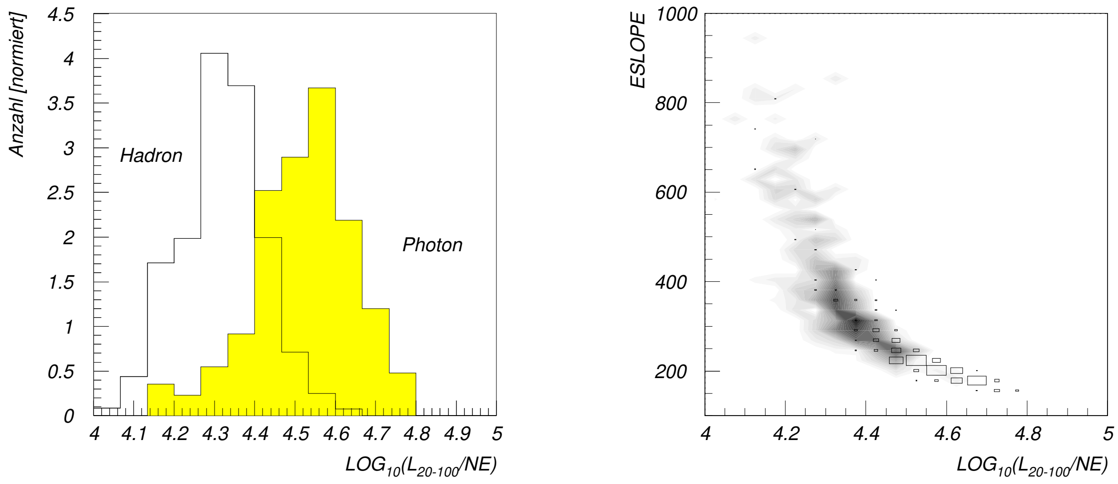



*Abbildung 4.2: Links: Das unterschiedliche logarithmische Verhältnis von Čerenkovlicht- zu Partikel-Komponente auf Detektorniveau für $\gamma$- und Hadron-induzierte Schauer.*
*Rechts: Korrelation des Licht- zu Partikel-Verhältnisses mit ESLOPE ($\gamma$: Kästchen, Hadronen: schattiert).*

Die Zahl der elektromagnetischen Schauerpartikel auf Detektorniveau $Ne$ ist sehr viel stärker von der Höhe des Maximums und des Profils der longitudinalen Entwicklung abhängig als die integrale Čerenkovlichtmenge und somit ist die damit erreichte Separations-Fähigkeit insgesamt abhängig von der individuellen longitudinalen Entwicklung der einzelnen Schauer.
Eine oder mehrere weitere Observable, die möglichst vollständig das Profil der Longitudinalentwicklung und/oder die Höhe des Schauermaximums beschreiben, ist also notwendig.
Charakterisiert man das Profil der Longitudinalentwicklung durch die Größen $HMAX$ (Schauerhöhe), $Ne_{max}/Ne$ ($Ne_{max}$ Anzahl der Schauerteilchen im Maximum) oder die Länge des Schauermaximums[3] zeigen sich in allen Größen Korrelationen mit dem Licht- zu Partikel-Verhältnis. Am deutlichsten ist die Korrelation jedoch für die Größe $ESLOPE$, die die Abklinglänge des Schauers nach dem Maximum angibt.

$$ESLOPE := \frac{HMAX}{\log_{10} \frac{NE_{max}}{NE}} \qquad (4.1)$$

Die anschauliche Bedeutung von Abbildung 4.2, rechts ist, daß das Licht- zu Partikel-Verhältnis das Abklingen nach dem Maximum mißt. Die $\gamma/H$-Separation mit $Ne/L_{(20-100)}$ beruht also auf der Tatsache, daß $\gamma$- induzierte Schauer nach dem Maximum schneller abklingen.

[3]Beschrieben durch die Strahlungslängen zwischen dem 50%-Wert auf der ansteigenden- und dem 80%-Wert auf der abfallenden Flanke des longitudinalen Profils.

### 4.1.2 Laterale Entwicklung

Nicht nur die Unterschiede in der longitudinalen, sondern auch die Unterschiede in der lateralen Entwicklung zwischen $\gamma$- und Hadron-induzierten Schauern liefern wertvolle Informationen und können zusätzliche für die $\gamma/H$-Separation verwendet werden. Der in hadronischen Wechselwirkungen auftretende Transversalimpuls ist verantwortlich für die breitere Lateralverteilung Hadron-induzierter Schauer. In einem beliebigen Schnitt senkrecht zur Schauerachse erscheint die Verteilung der elektromagnetischen Komponente bei Hadron-induzierten Schauern flacher, dies setzt sich auf Detektorebene in die Größen $Age$ bzw. $r_p$ (Abschnitt 4.3.2) um. Abbildung 4.3, links zeigt diesen Effekt integral für die gesamte M$^{te}$ Carlo-Menge. Der durch die unterschiedlichen Eindringtiefen verursachte mittlere Effekt im Alter $Age$ der Schauer wurde durch eine von der Schauerhöhe abhängige Parametrisierung kompensiert, um den Unterschied deutlicher zu machen[4].

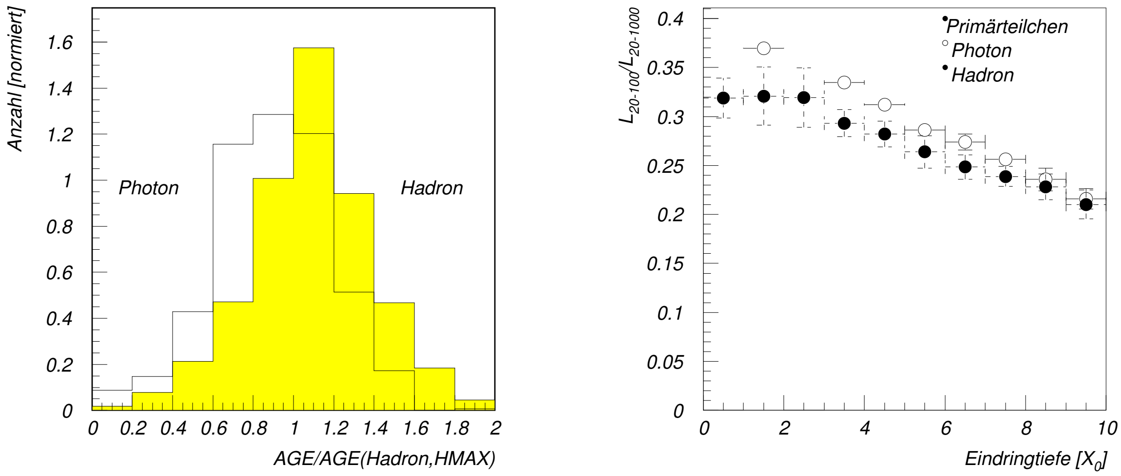



*Abbildung 4.3: Links: Age für $\gamma$- und für Hadron-induzierte Schauer. Rechts: Anteil des Čerenkovlichts bis 100 m Entfernung von der Schauerachse am gesamten Čerenkovlicht (bis 1000 m) in Abhängigkeit von der Eindringtiefe des Schauers.*

Auch in der Lateralverteilung des Čerenkovlichts zeigt sich die kompaktere Form der $\gamma$-Schauer (Abbildung 4.3,rechts). Insbesondere für tief eindringende Schauer ist der Anteil des bis 100 m um den Kernort nachweisbaren Čerenkovlichts am gesamten erzeugten Čerenkovlicht bei $\gamma$-Schauern größer. Somit wird das Licht- zu Partikel-Verhältnis innerhalb des Detektorradiuses (100 m) nochmals günstiger für die Separation.
Wie aus der NKG-Theorie für elektromagnetische Schauer folgt, ist $Age$ ein Maß der longitudinalen Schauerentwicklung (also der Eindringtiefe des Schauers). Das aus der tatsächlich im Detektor gemessenen Lateralverteilung angepaßte $Age$ ist jedoch zusätzlich sensitiv auf den

[4]Auch ohne diese Parametrisierung zeigt sich der Unterschied in der Steilheit der Lateralverteilung, ist aber deutlich kleiner.

hadronischen Transversalimpuls und die individuellen Fluktuationen in den Longitudinalentwicklungen der Schauer. Zur Verdeutlichung des Zusammenhanges zwischen longitudinaler Ent-

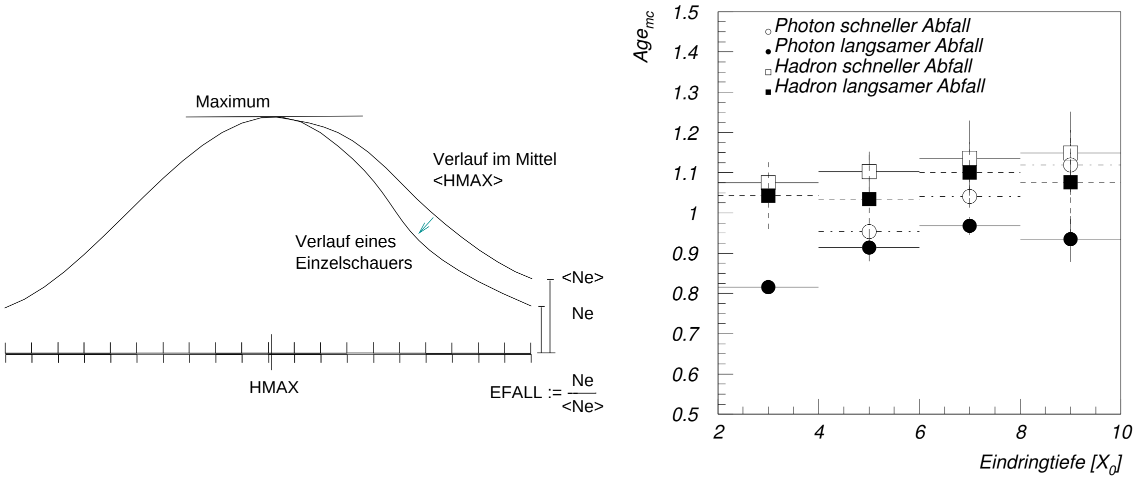



*Abbildung 4.4: Links: Zur Konstruktion der Größe EFALL*
*Rechts: $Age_{mc}$ bei schnellem/langsamem Abfall nach dem Schauermaximum. Deutlich wird, daß sich ein schneller Abfall der elektromagnetischen Komponente des Schauers nach dem Maximum in eine flache Lateralverteilung am Boden übersetzt.*

wicklung und der Steilheit der Lateralverteilung (charakterisiert durch $Age_{mc}$, siehe Abschnitt 4.6) zeigt Abbildung 4.4, rechts $Age_{mc}$ in Abhängigkeit von der Höhe des Schauermaximums $HMAX$ für Schauer, die nach dem Maximum schnell abklingen, und für Schauer, die langsamer abklingen. Die Klassenunterteilung wurde durch den Median in der Größe $EFALL$ definiert, die das Verhältnis aus gemessener zu erwarteter Partikel-Komponente auf Detektorniveau als Funktion von $HMAX$ angibt (siehe Abbildung 4.4, links):

$$EFALL := \frac{Ne}{< Ne_{h,\gamma}(HMAX) >} \tag{4.2}$$

Der Erwartungswert $< Ne_{h,\gamma}(HMAX) >$ wurde getrennt für alle Primärteilchensorten aus den Mittelwerten bei vorgegebenen Schauerhöhen bestimmt. Ein langsamer Abfall der Longitudinalentwicklung nach dem Maximum bewirkt kleinere $Age_{mc}$ Werte, wie in Abbildung 4.4, rechts neben der erwarteten Abhängigkeit des $Age_{mc}$ Parameters von der Entfernung des Maximums zum Detektor zu sehen ist. Eine starke Vorwärtsrichtung von Schauern hat ein langsames Abklingen des Schauers nach dem Maximum und eine lateral kompakte Teilchenkomponente zur Folge.

## 4.2 Q-Faktor und Ereignisselektion

Die Signifikanz einer Messung wird in Vielfachen der Streuung der Untergrund-Verteilung angegeben. Für die Messung einer hypothetische $\gamma$-Quelle nimmt sie die Form an[5]:

$$S = \frac{Signal}{Untergrund - Streuung} = \Phi_\gamma \sqrt{\frac{A\Delta t}{\Phi_{bg}\Delta\Omega}} \tag{4.3}$$

$A$ ist die effektive Fläche des Detektors, $\Delta t$ die Meßdauer und $\Delta\Omega$ der Raumwinkel des Suchbereichs. $\Phi_\gamma$ gibt den Fluß der Quelle und $\Phi_{bg}$ den des Untergrunds aus Schauern,die von geladenen Teilchen erzeugt wurden, pro Raumwinkel an.
Wendet man eine Datenselektion mit Signaleffizienz $\epsilon_\gamma$ und Untergrundeffizienz $\epsilon_{bg}$ an, so verändert sich die Signifikanz um den sogenannten Q-Faktor:

$$Q = \frac{\epsilon_\gamma}{\sqrt{\epsilon_{bg}}} \tag{4.4}$$

Wendet man zuerst eine Datenselektion mit dem Q-Faktor $Q_1$ und dann auf den reduzierten Datensatz eine zweite Datenselektion mit dem $Q$-Faktor $Q_2$ an, so ist der gesamte Q-Faktor $Q_g$ trivialer weise:

$$Q_g = Q_1 \cdot Q_2 \tag{4.5}$$

Gleichung 4.5 separiert also bezüglich der Schnittanwendung. Im Fall parametrischer Verfahren, wie z.B. Neuronaler Netze, bei denen die gewählten Parameter von der zugrunde liegenden Gesamtheit abhängen, gilt die Separabilität nur für den Fall fester Parameter, jedoch nicht für den Fall optimal gewählter Parameter, so daß die Qualität der Methode für alle Vorselektionen neu bestimmt werden muß. Die Kette der verschiedenen Selektions-Schritte ist in Abbildung 4.5 gezeigt. Die maximal zur Verfügung stehende Menge von Ereignissen wird durch die Hardware-Bedingung A6S14 definiert, dies entspricht der Standard-HEGRA-Bedingung von 6 AIROBICC-*oder* 14 Szintillator-Signalen. Von den Ereignissen wird dann entweder ein Szintillator-Auslöser (S14) und/oder gefordert, daß die beiden Lateralanpassungen konvergieren[6]. Erst nach diesen Vorselektionen erfolgt die speziell auf die $\gamma$-Anreicherung optimierte $\gamma/H$-Separationsmethode.
In Tabelle 4.1 sind die Effizienzen der verschiedenen Vorselektionen aufgeführt. Die Werte in der Tabelle sind auf die Bedingung A6S14 normiert. Deutlich zu erkennen ist, daß durch die AIROBICC-Bedingung A6 gegenüber der Szintillator-Bedingung S14 deutlich mehr $\gamma$ Ereignisse selektiert werden (vgl. Abschnitt 4.1.1). Im Umkehrschluß bedeutet dies, daß eine Methode, die hohe Anforderungen an die Szintillator-Daten stellt, um den Faktor $\approx 2$ besser sein muß, um die gleiche Sensitivität für die $\gamma$-Quellensuche zu erreichen, wie eine Methode, für die alle A6S14-Ereignisse verwendet werden können. Eine analoge Argumentation gilt für die Forderung nach konvergierten Lateralanpassungen. In diesem Fall ist der durch den intrinsischen Q-Faktor der Selektion verursachte Verlust jedoch deutlich geringer (ohne Kernort-Schnitt: 0.97). Auf diesen Punkt wird in Abschnitt 4.4.2 nochmals eingegangen.
Im Rahmen dieser Arbeit wird die Qualität der erreichten $\gamma/H$-Separation durchgängig mit Q bewertet, wie dies für die Quellensuche sinnvoll ist. Bei anderen Fragestellungen der $\gamma-$Astronomie,

[5] Diese Formel gilt für sehr kleine Statistiken nicht mehr, sondern muß durch die Formel von Li & Ma [36] ersetzt werden.

[6] Also, daß die NKG-Anpassung an die Szintillator-Daten und die Anpassung der Exponentialfunktion an die AIROBICC-Daten konvergierten.

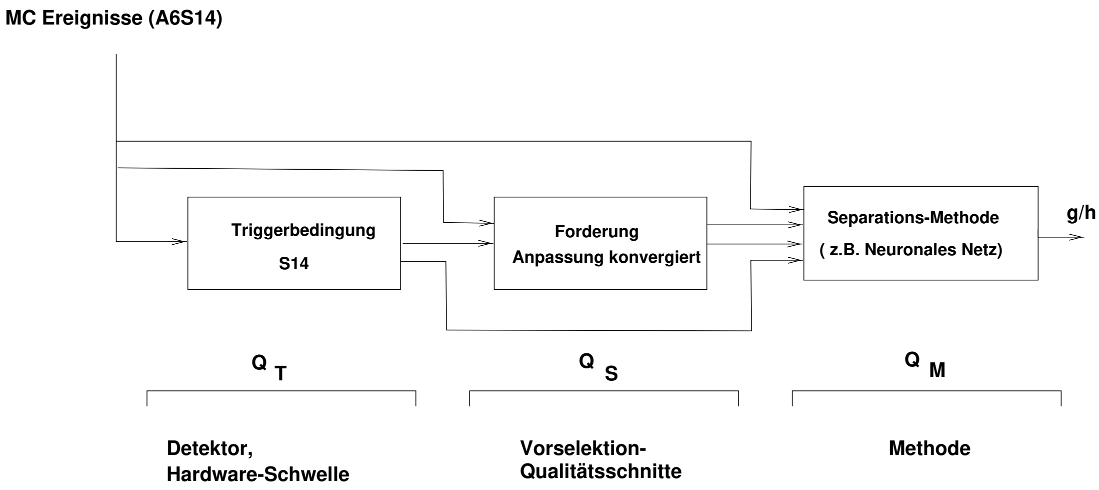



*Abbildung 4.5: Ereignisselektions-Kette*

| Vorselektion | | | | rek. Kernort im Feld | | |
|---|---|---|---|---|---|---|
| | $\epsilon_\gamma$ | $\epsilon_{bg}$ | $Q$ | $\epsilon_\gamma$ | $\epsilon_{bg}$ | $Q$ |
| A6S14 | 1 | 1 | 1 | 0.84 | 0.92 | 0.87 |
| S14 | 0.37 | 0.79 | 0.42 | 0.35 | 0.76 | 0.39 |
| konvergierte Lateralanpassung | 0.69 | 0.51 | 0.97 | 0.58 | 0.44 | 0.87 |
| S14 und konvergierte Lateralanpassung | 0.27 | 0.35 | 0.46 | 0.25 | 0.33 | 0.44 |

*Tabelle 4.1: Resultierende Effizienzen und Q-Faktoren für verschiedene Ereignis-Vorselektionen. Die Vorselektionsbedingungen sind im Text erläutert.*

z.B. nach der oberen Grenze des Verhältnisses von diffuser $\gamma$-Strahlung zur geladenen kosmischen Strahlung $f_\gamma = \frac{I_\gamma}{I_{cr}}$ [12], bei denen keine Möglichkeit zur Untergrundabschätzung und Subtraktion besteht, muß anstelle des Q-Faktors der Quotient $\epsilon_\gamma/\epsilon_{bg}$ (Effizienz des Schnittes) möglichst optimal gewählt werden. Bis zur Bestimmung des Schnittwerts in der Netzausgabe bleibt das in Abschnitt 3 vorgestellte Verfahren für diesen Fall identisch.

## 4.3 Bestehende Verfahren

Keines der in Abschnitt 1.5 genannten Verfahren nutzt die simultane Messung der elektromagnetischen Partikel-Komponente und der des atmosphärischem Čerenkovlichts von Luftschauern, die durch den HEGRA AIROBICC-Detektor ermöglicht wird.
Die beiden bei HEGRA entwickelten Methoden basieren im Wesentlichen auf der Tatsache, daß bei $\gamma$-induzierten Luftschauern im Mittel das Licht- zu Partikel-Verhältnis größer ist als bei Hadron-induzierten Luftschauern (siehe Abschnitt 4.1.1). In der Wahl der verwendeten Observablen unterscheiden sich die beiden Verfahren. Im folgenden werden sie kurz erläutert und ihre

Güte auf Grundlage der verwendeten M$^{te}$ Carlo-Menge[7] quantifiziert.

### 4.3.1 Methode LES

Die Methode LES (**L**ight, **E**lectrons and **S**lope) wurde von F. Arqueros [37] entwickelt und von A. Karle für die Suche nach $\gamma$-Punktquellen [12] sowie für die Abschätzung der isotropen $\gamma$-Strahlung oberhalb von 65 TeV [19] angewandt. Auch hier wird ausgenutzt, daß Hadron-induzierte Schauer bei gleicher Maximumshöhe langsamer abklingen als $\gamma$-induzierte Schauer (Kapitel 4.1.1). Wie in Abschnitt 4.1.1 beschrieben wurde, kann dieser Unterschied zur $\gamma/H$-Separation genutzt werden, da bei gleicher Höhe des Schauermaximums und gleicher Anzahl von Elektronen im Schauermaximum die elektromagnetische Komponente auf Detektorniveau für Hadron-induzierte Schauer größer ist als für $\gamma$-induzierte Schauer. Die Höhe des Schauermaximums wird durch $r_l$, die Anzahl der Elektronen im Schauermaximum $Ne_{max}$ durch das Čerenkovlicht bei 90 m Abstand vom Kernort $L_{90}$ ($L_{90} \propto E_{Prim\ddot{a}r} \propto Ne_{max}$) und die elektromagnetische Komponente auf Detektorniveau durch $Ne$ beschrieben [38].
Als Separationsparameter wird $\log_{10}(Ne\ /L_{90})$ in Abhängigkeit von $r_l$ verwendet. Die Separationslinie wird dabei entweder durch Optimierung dreier linearer Funktionen [12] oder durch die eines Polynoms dritten Grades für verschiedene Winkel- und Energiebereiche bestimmt [38].

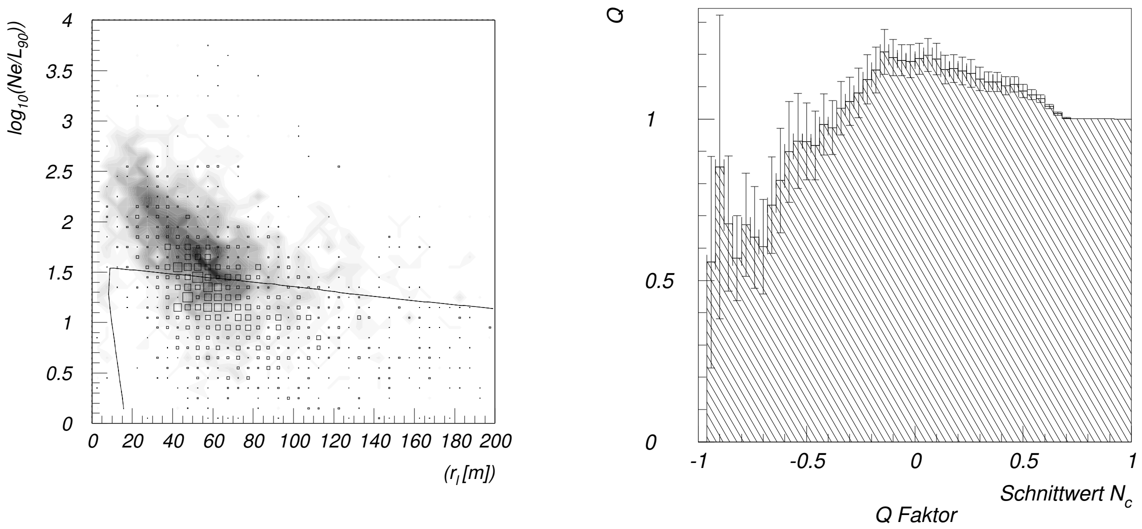



*Abbildung 4.6: Neuronales Netz mit den für LES verwendeten Größen. Es ist die Diskriminante für $N_c = 0.0$ eingezeichnet. Die $\gamma$-induzierten Ereignisse sind als Kästchen, die Hadron-induzierten als Schraffur gezeichnet. Die rechte Abbildung zeigt den ermittelten Q-Wert.*

[7] Außer der Forderung nach konvergierten Lateralanpassungen werden keine weiteren Selektionsschnitte verwendet, dies entspricht Zeile 3, Spalte 1-3 der Tabelle 4.1.

Abbildung 4.6,links zeigt die von einem 2-3-1 Netz für die LES-Größen gefundene Diskriminante. Obwohl zu höheren $r_l$ Werten die Steigung ein wenig flacher wird, handelt es sich bei der Diskriminante in guter Näherung um eine Gerade. Die negative Steigung der Diskriminante bezogen auf $r_l$ spiegelt wider, daß $Ne$ für $\gamma$-induzierte Schauer mit steigender Schauerhöhe schneller abnimmt als für Hadron-induzierte Schauer, da $\gamma$-Schauer im Mittel kürzer sind. Deshalb verschiebt sich das für die Separation günstigste Verhältnis $\log_{10}(Ne/L_{90})$ mit steigendem $r_l$ zu kleineren Werten.
Das Abknicken der Diskriminante im linken Teil der Abbildung ist als Artefakt der begrenzten $M^{te}$ Carlo-Statistik zu verstehen. In diesem Bereich stehen nur wenige Ereignisse zur Verfügung und sind deshalb dafür entscheidend, daß einer der Knoten, aufgrund der Abwesenheit stärkerer Merkmale, sensitiv auf eine sehr kleinen Abnahme der Fehlerfunktion des Neuronalen Netzes wird[8]. Der mit dem LES-Netz erreichte maximale Q-Faktor beträgt Q=1.2±0.1 und ist in Abbildung 4.6, rechts zu sehen. Der optimale Wert wird für einen Schnitt in der Netzausgabe bei $N_c$=0.0 erreicht[9].
Bei den in der Abbildung dargestellten Fehlern handelt es sich um die statistischen Fehler, die mit dem in Abschnitt 3.5.4 beschriebenen Verfahren berechnet wurden. Da es sich bei den Schnitt-Effizienzen, die der Q-Faktor-Bestimmung zugrunde liegen um integrale Verteilungen handelt, sind die einzelnen Werte der Abbildung nicht statistisch unabhängig voneinander.
In Abschnitt 4.4.1 wird gezeigt werden, daß Q=1.2 auch ohne den Parameter $r_l$ zu erreichen ist und vollständig durch die Separationsfähigkeit des Licht- zu Partikel-Verhältnisses erklärt werden kann.

### 4.3.2 Methode PEARL

Bei niedrigen Primärenergien, nahe der Detektor-Nachweisgrenze von 15-20 TeV, registrieren nur wenige Einzeldetektoren-Signale, so daß die Lateralanpassungen nicht konvergieren oder unsichere Ergebnisse liefern. Der Vergleich zwischen den mit der HEGRA Rekonstruktion ermittelten und denen mit der vollständiger $M^{te}$ Carlo-Überdeckung (Abschnitt 4.6) ermittelten Werten bestätigt dies: Für $r_l$ und $L_{(20-100)}$ steigen die relativen Abweichungen unterhalb von 10 AIROBICC-Detektoren schnell auf das 2-3 fache an. Auch die relativen Abweichungen auf die Bestimmung von $Age$ und $Ne$ sind für Ereignisse, die allein durch AIROBICC ausgelöst wurden (A6), doppelt so groß wie der Plateau-Werte oberhalb von 13 Szintillator-Detektoren (S14).
Um diesen für die Quellensuche besonders interessanten Energiebereich abzudecken, benutzt die Methode PEARL nur mit einfachen arithmetischen Operationen konstruierte Größen, die für alle Ereignisse zur Verfügung stehen.

$$
\begin{aligned}
PEARL &= \log_{10}(r_p[m]) + 0.8432 \cdot \log_{10}(\frac{N_p}{N_l}) \qquad (4.6)\\
& r_p = \text{mittlerer Abstand der registrierten Teilchen vom Kernort}\\
& N_p = \text{summierte Anzahl der in den Szintillatoren registrierten Teilchen}\\
& N_l = \text{Summe aller von AIROBICC registrierten Photonen}
\end{aligned}
$$

[8]Diese Eigenschaft Neuronaler Netze wird häufig mit "NN können gut interpolieren, aber schlecht extrapolieren." beschrieben.

[9]Da es sich bei Q um eine abhängige Variable handelt, die keinen direkten Zusammenhang mit der Energiefunktion des Neuronalen Netzes hat, und wegen der in Abschnitt 3.3 gemachten Bemerkungen zum *summed squared Error*, kann im Allgemeinen nicht davon ausgegangen werden, daß der optimale Schnitt bei $N_c = 0$ liegt.

Als optimaler Schnittwert wird PEARL = - 1.67 verwendet [35]. Abbildung 4.7 zeigt die von J. Prahl angegebene Diskriminante und die mit einem 2-3-1 Netz errechnete. Es ist eine gute Übereinstimmung der beiden Verfahren zu sehen.

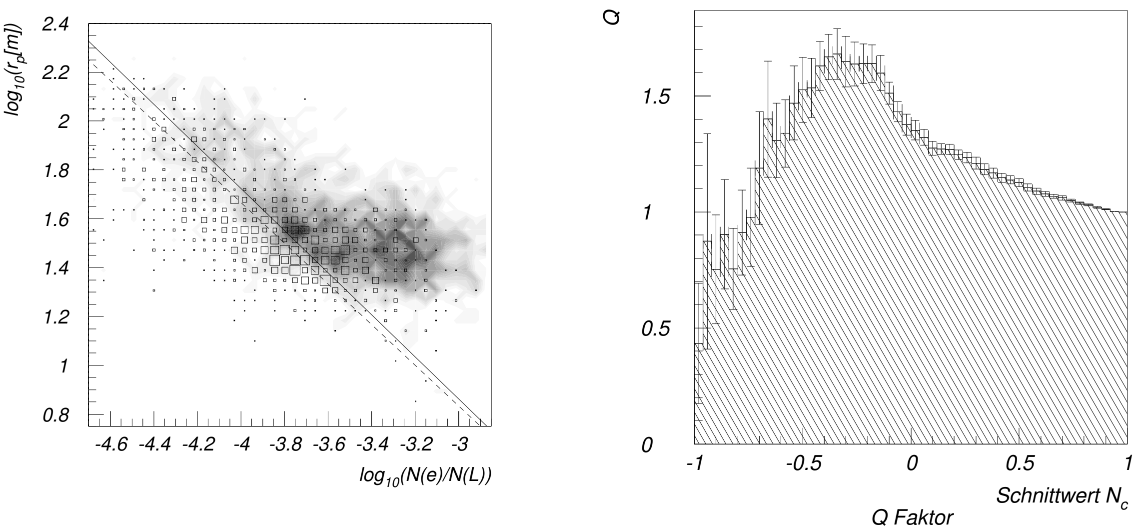



*Abbildung 4.7: Vergleich der Separationslinie der Methode PEARL (durchgezogene Linie) mit der des äquivalenten Neuronalen Netzes (gestrichelte Linie). Die $\gamma$-induzierten Ereignisse sind (wieder) als Kästchen, die Hadron-induzierten als Schraffur dargestellt. Die rechte Abbildung zeigt den ermittelten Q-Wert in Abhängigkeit vom Schnittwert in der Ausgabevariable.*

Sowohl die PCA-Analyse als auch die CDA-Analyse (Abbildung 4.8) zeigen jeweils für die versteckte Schicht nur einen wesentlich von Null verschiedenen Eigenwert und somit, daß von den drei versteckten Knoten nur einer wesentlich zur Entscheidung beiträgt. Eine weitere Schicht ist ebenfalls nicht notwendig, da die Konvergenz auch mit nur einer Schicht hinreichend schnell erreicht wird.

Wie in Abbildung 4.7 zu erkennen ist, beträgt der maximale Q-Faktor für das PEARL-Netz 1.6±0.1. Dies ist in guter Übereinstimmung mit dem in [39] angegebenen Wert von 1.5±0.1, wenn man berücksichtigt, daß dort nicht die Forderung nach konvergierten Lateralanpassungen an die M$^{te}$ Carlo-Ereignisse gestellt wurde. In Abschnitt 4.4.2 wird gezeigt werden, daß durch die unterschiedliche Vorselektion der ohnehin geringe Unterschied der Q-Werte erklärt werden kann.

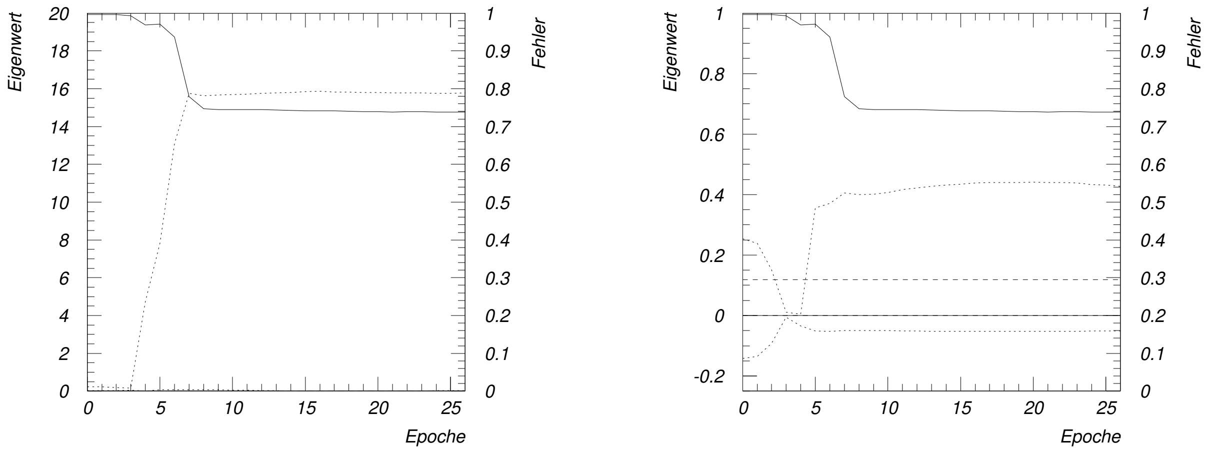



*Abbildung 4.8: Links: PCA-Analyse der versteckten Schicht, die gepunktete Kurve zeigt die Entwicklung des einzigen wesentlich von Null verschieden Eigenwert. Rechts: CDA-Analyse der versteckten Schicht (gepunktet) und der Eingabeschicht (gestrichelt). In beiden Fällen ist der Trainingsfehler als durchgezogen Kurve eingezeichnet.*

## 4.4 Neuronale Netze mit integralen Größen

Die in diesem Abschnitt verwendeten Neuronale Netze basieren auf zwei unterschiedlichen Variablen-Sätzen. Im wesentlichen sind dies die in Abschnitt 2.4 vorgestellten angepaßten Größen und die in Abschnitt 4.3.2 vorgestellten arithmetischen PEARL-Größen. Die Strategie zur Bestimmung der notwendigen Netztopologie und zur Findung der geeigneten Eingabegrößen ist für beide Variablen-Sätze die gleiche. Es wird mit einer möglichst großen Anzahl von Eingabeknoten[10] und versteckten Knoten begonnen, die dann schrittweise reduziert werden. Es wird dieselbe M$^{te}$ Carlo-Menge wie in den beiden vorigen Abschnitten verwendet.

### 4.4.1 Angepaßte Größen

Als Ausgangspunkt der Analyse wird ein 7-x-x-1 Netz verwendet, wobei x die Anzahl der noch zu bestimmenden Knoten pro versteckter Schicht ist. Um eine schnelle Konvergenz zu gewährleisten, wurde ein Netz mit zwei versteckten Schichten gewählt. Tabelle 4.2 führt die für das Neuronale Netz verwendeten Eingangsgrößen auf. Die Motivation für die ersten vier Observablen wurde in den vorigen Abschnitten gegeben. Die Maximumsnorm des Ortsvektors des Kernorts, $\|\binom{x}{y}\|$, stellt ein Beispiel für eine Reihe anderer Größen dar, die indirekt über die Abhängigkeit systematischer Rekonstruktions-Effekte eine Korrelation mit den tatsächlich gemessenen physikalischen Observablen besitzen und somit den Einfluß von Rekonstruktions-Unsicherheiten reduzieren könnten. Die Motivation für die aus den verschiedenen Anpassungen ermittelten $\chi$-Werte wurde in Abschnitt 1.5 gegeben und beruht auf der ungleichmäßigeren Lateralverteilung der Schauerteilchen in Hadron-induzierten Schauern.

| Name | Bedeutung |
|---|---|
| $Ne$ | Parameter der NKG-Anpassung, Anzahl der elektromagnetischen Partikel |
| $Age$ | Parameter der NKG-Anpassung, Steilheit der Lateralverteilung |
| $L_{(20-100)}$ | Aus der AIROBICC Lateral-Anpassung berechnete Anzahl von Čerenkov-Photonen im Kreisring von 20 m bis 100 m um den Kernort |
| $r_l$ | Parameter der AIROBICC Lateral-Anpassung, Lichtradius |
| $\|\binom{x}{y}\|$ | Maximumsnorm des Ortsvektors des Kernorts |
| $\chi_{air,dir}$ | $\chi$ pro Freiheitsgrad der AIROBICC Lateral-Anpassung |
| $\chi_{air,lat}$ | $\chi$ pro Freiheitsgrad aus der Richtungs-Anpassung an die AIROBICC-Amplituden |
| $\chi_{sz,lat}$ | $\chi$ pro Freiheitsgrad der NKG-Anpassung |
| $\chi_{sz,mip}$ | $\chi$ pro Freiheitsgrad der NKG-Anpassung, beschränkt auf den Bereich, in dem der Wert der NKG-Funktion größer als 1 MIP ist. |

*Tabelle 4.2: Die für das Netz mit angepaßten Größen verwendeten Observablen.*

Zuerst wird die notwendige Anzahl von versteckten Knoten bestimmt. Dazu zeigt Abbildung 4.9 den Trainings- und den Generalisierungsfehler für 1,2,4,6 und 12 versteckte Knoten pro Schicht. Schon ab zwei versteckten Knoten verringert sich der mit der NCV-Methode ermittelte Testfehler des trainierten Netzes nicht mehr und ab 4 Knoten steigt er sogar leicht an. Dieses Verhalten des Testfehlers ist typisch, da bei einer zu geringe Anzahl von Knoten die Merkmale der

[10] Nur die im Laufe der Arbeit als aussichtsreich eingestuften Größen werden diskutiert.

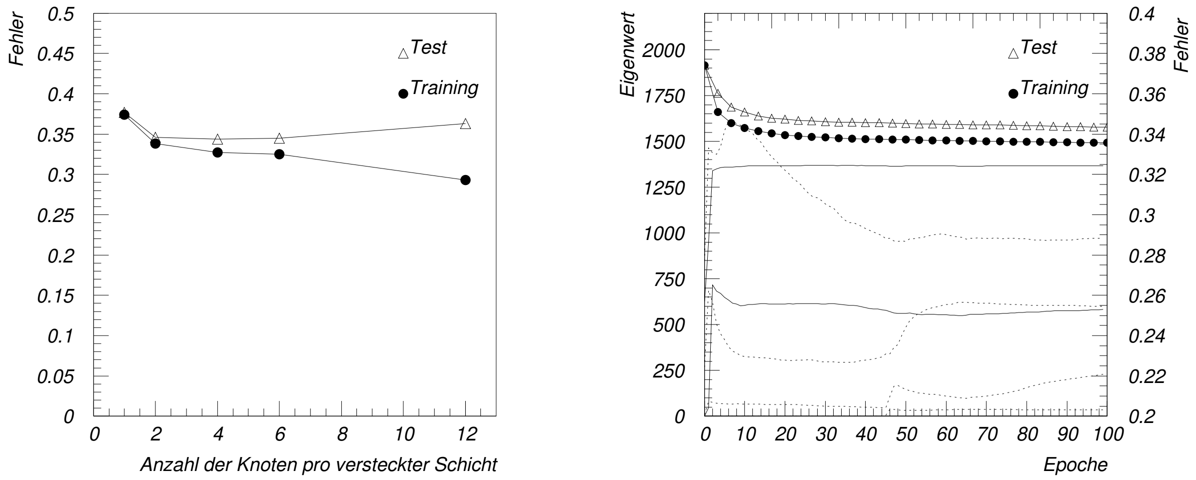



*Abbildung 4.9: Links: Netzfehler in Abhängigkeit der Anzahl der Knoten pro versteckter Schicht für die im Text beschriebenen Neuronalen Netze.*
*Rechts: PCA-Analyse für das Netz mit 6 Knoten. Es sind die Eigenwerte der ersten (gepunktet) und zweiten versteckten Schicht (durchgezogen) eingezeichnet.*

Trainingsmenge nicht ausreichend extrahiert werden können, während bei einer zu großen Anzahl von Knoten die Trainingsmenge überangepaßt wird und die Generalisierungsleistung zurückgeht. In jedem Fall nimmt der Trainingsfehler mit zunehmender Anzahl von Knoten monoton ab (oder bleibt konstant), da die Erhöhung der Modell-Freiheitsgrade (Knoten) eine immer genauere Anpassung der Testmenge ermöglicht.
Daß bereits 3 Knoten in der ersten und 2 Knoten in der zweiten versteckten Schicht ausreichend sind, wird durch Abbildung 4.10, rechts bestätigt, die die PCA-Analyse eines Netzes mit 6 Knoten pro Schicht zeigt. Für die erste versteckte Schicht (gepunktete Linie) sind nur 2-3 Eigenwerte wesentlich von Null verschieden, für die zweite versteckte Schicht sogar nur zwei (durchgezogene Linie).
Nachdem feststeht, daß bereits 3 versteckte Knoten ausreichend sind, wird die Berechnung der Relevanzen der Eingabevariablen mit einem 7-3-3-1 Netz durchgeführt. Neben den erwarteten hohen Relevanzen für $Ne$, $L_{(20-100)}$ und $Age$ werden $\chi_{sz,lat}$ und $\chi_{sz,mip}$ ebenfalls hohe Relevanzen zugeordnet. Aus den in Abschnitt 3.5.2 gegebenen Gründen sollten die Ergebnisse der Relevanzanalyse überprüft werden. Dazu werden sukzessive Eingangsgrößen weggelassen, beginnend mit der zur niedrigsten Relevanz gehörenden Größe, und das Netz bis zum Erreichen des neuen Gleichgewichts weiter trainiert.
Falls das Löschen eines Eingabeknotens zu einem höheren Testfehler führt, wird das Löschen rückgängig gemacht und mit dem nächsten Knoten fortgefahren. Das Verfahren ist beendet, wenn kein weiterer Knoten entfernt werden kann, ohne daß der Testfehler ansteigt[11]. Tabelle 4.3 zeigt das Ergebnis und bestätigt den Verdacht, daß es sich bei den hohen Relevanzen für

[11] Dieses Verfahren wird als *input sensitivity pruning* bezeichnet.

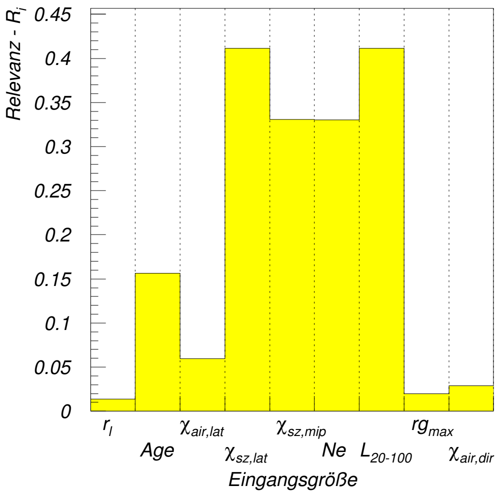



*Abbildung 4.10: Relevanzen der angepaßten Eingangsgrößen für das auf drei versteckte Knoten reduzierte Netz*

$\chi_{sz,lat}$ und $\chi_{sz,mip}$ um einen vorgetäuschten Effekt handelt, da das Löschen der zugeordneten Eingabeknoten zu keiner Erhöhung des Testfehlers führt.
Der Lichtradius $r_l$ ist in Abbildung 4.10 die Eingabegröße mit der geringsten Relevanz, deshalb wird $r_l$ als erste Variable entfernt. Das Löschen wirkt sich minimal in einem Anstieg des Testfehlers aus, ohne allerdings den Q-Wert zu reduzieren. Da der Testfehler beim Entfernen ansteigt, wird $r_l$ bis zur vorletzten Zeile in Tabelle 4.3 weiter verwendet. Obwohl erst beim Löschen des $Age$-Knotens eine signifikante Reduzierung der $\gamma/H$-Separation beobachtet werden kann, ist $r_l$ die einzige der untersuchten Variablen, die neben $Ne$, $L_{(20-100)}$ und $Age$ zumindest einen Einfluß auf den Testfehler hat.
Der geringe Einfluß von $r_l$ auf das Gesamtergebnis ist überraschend, da die Methode LES den Parameter $r_l$ als einzige zusätzliche Information der Lateralentwicklung nutzt. Aufgrund der in Abschnitt 4.3.1 gegebenen Erklärung für LES liegt es nah, $r_l$ durch die direkt aus dem M$^{te}$ Carlo-Simulation gewonnene Schauerhöhe $HMAX$ zu ersetzen. Der von einem Netz mit $Ne$, $L_{(20-100)}$ und HMAX maximal erreichbare Q-Faktor ist 1.4±0.2, und liegt somit über dem in Tabelle 4.3, Zeile 9 für das Licht- zu Partikel-Verhältnis gegebenen Wert von 1.2±0.1, aber unter dem in derselben Tabelle in Zeile 8 gegebenen Wert für $Ne$, $L_{(20-100)}$ und $Age$.

| Eingangsvariablen | Trainingsfehler | Testfehler | Q Wert |
|---|---|---|---|
| Alle | 0.336 | 0.344 | 1.5±0.1 |
| (-)$r_l$ | 0.337 | 0.346 | 1.5±0.1 |
| (-)$\|\binom{x}{y}\|$ (+)$r_l$ | 0.340 | 0.344 | 1.5±0.1 |
| (-)$\chi_{air,dir}$ | 0.340 | 0.344 | 1.5±0.1 |
| (-)$\chi_{air,lat}$ | 0.340 | 0.343 | 1.5±0.1 |
| (-)$\chi_{sz,mip}$ | 0.341 | 0.343 | 1.5±0.1 |
| (-)$\chi_{sz,lat}$ | 0.341 | 0.343 | 1.5±0.1 |
| (-)$r_l$ | 0.342 | 0.345 | 1.5±0.1 |
| (-)$Age$ | 0.392 | 0.394 | 1.2±0.1 |

*Tabelle 4.3: Änderung der Netzfehler bei Löschen von Eingabeknoten. Die Tabelle ist von oben nach unten zu lesen. Das Symbol (-) zeigt die Entfernung einer Eingabegröße an, ein (+) bedeutet, daß die Eingabegröße wieder hinzugefügt wurde.*

Daraus läßt sich schließen, daß selbst dann, wenn $r_l$ die Höhe des Maximums ideal messen würde, die durch $Age$ gewonnene Separationsfähigkeit immer noch überwiegen würde.

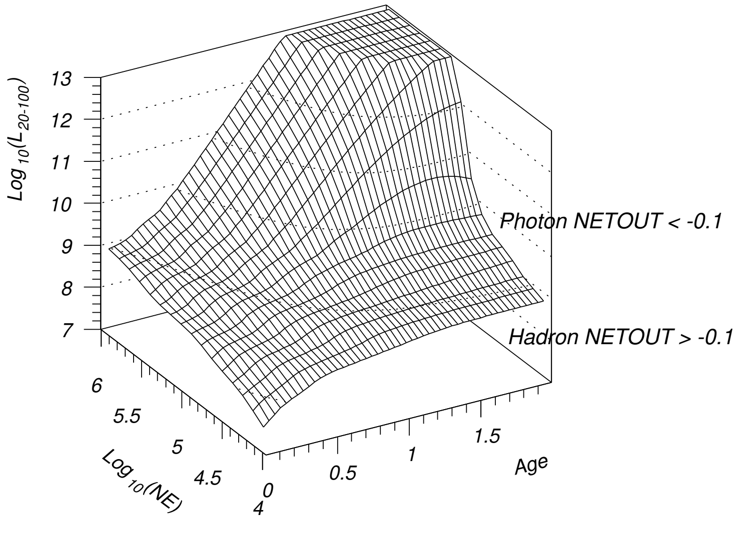



*Abbildung 4.11: Die Entscheidungsgrenze für das auf drei Eingänge reduzierte Netz mit angepaßten Größen.*

Abbildung 4.11 zeigt die dreidimensionale Darstellung der Diskriminante des Netzes aus Tabelle 4.3, Zeile 8. Es handelt sich dabei um einen mit $Age$ monoton ansteigenden Schnitt im Licht- zu Partikel-Verhältnis $L_{(20-100)}/Ne$.

Insgesamt läßt sich aus der Analyse des Neuronalen Netzes schließen, daß von den verwendeten angepaßten Größen im wesentlichen $Ne$, $L_{(20-100)}$ und $Age$ und mit starker Einschränkung $r_l$ für die $\gamma/H$-Separation tauglich sind und einen bestmöglichen Satz an Observablen darstellen.

### 4.4.2 Arithmetische Größen

Der Satz von Eingabegrößen für das Neuronale Netz mit arithmetischen Größen besteht in erster Linie aus den PEARL-Größen $\log_{10}(N_l)$, $\log_{10}(N_p)$, $r_p$ . Hinzugenommen wurden $\|\binom{x}{y}\|$ und $r_l$. Obwohl $r_l$ aus der Lateralanpassung bestimmt wird, wurde es in den Variablen-Satz aufgenommen, da kein entsprechendes arithmetisches Äquivalent existiert. Für den in Abschnitt 4.3.2 erwähnten Ansatz stellt dies eine Einschränkung dar. Da wegen der Vergleichbarkeit mit den Resultaten des vorigen Abschnitts auch hier Ereignisse mit konvergierten Lateralanpassungen gefordert werden, ist dies an dieser Stelle jedoch unproblematisch.

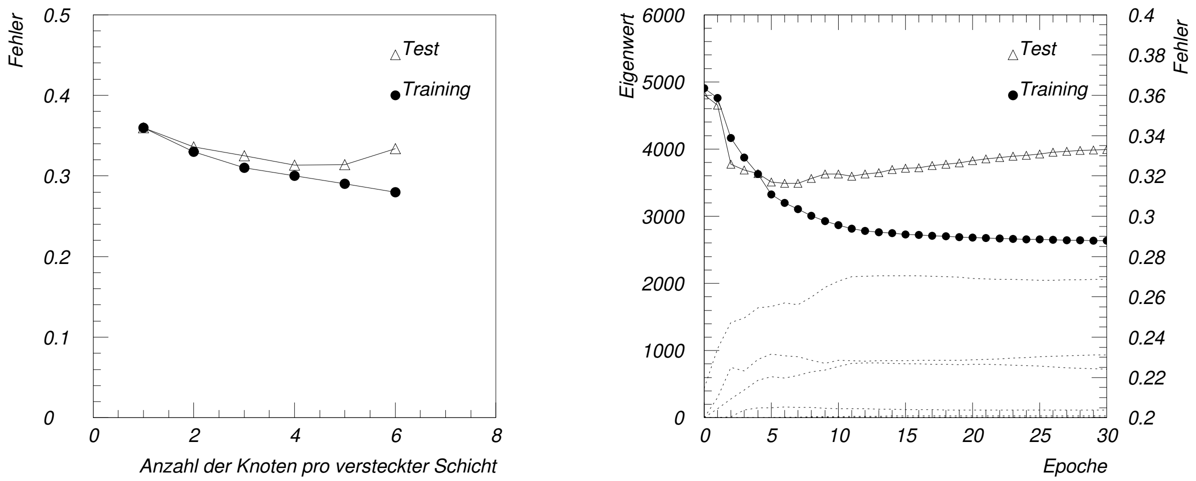



*Abbildung 4.12: Links: Netzfehler in Abhängigkeit der Anzahl der versteckten Knoten. Rechts: PCA-Diagramm der ersten versteckten Schicht für ein 5-6-6-1 Netz.*

Wie bei dem Netz für angepaßte Größen zeigt sich, daß bereits eine geringe Anzahl von versteckten Knoten ausreichend ist. In diesem Fall sind es vier Knoten (Abbildung 4.12). Das PCA-Diagramm für die erste versteckte Schicht bestätigt dieses Ergebnis. Deutlich sind 4 Eigenwerte (bei insgesamt 6 Knoten in der versteckten Schicht) der Aktivitätsverteilung zu sehen. Daß im Vergleich zum PEARL-Netz aus Abschnitt 4.3.2 mehr Knoten erforderlich sind, liegt zum einen daran, daß $\log_{10}(N_l)$ und $\log_{10}(N_p)$ als getrennte Eingabegrößen behandelt werden, und zum anderen daran, daß anstatt der für das PEARL-Netz günstigen Variable $\log_{10}(r_p)$, direkt $r_p$ verwendet wird.
Die für ein 5-4-4-1 Netz berechneten Relevanzen (Abbildung 4.13) zeigen, daß $r_l$ und $\|\binom{x}{y}\|$ für die Separationsfähigkeit des Netzes eine untergeordnete Rolle spielen. Wie bei dem Neuronalen Netz mit angepaßten Größen wirkt sich das Löschen von $r_l$ und $\|\binom{x}{y}\|$ nur auf den Fehler des Neuronalen Netzes aus, nicht auf den erreichten Q-Faktor[12]. Es bleiben die drei für PEARL verwendeten Größen übrig.

[12] Auf die Darstellung der Überprüfung analog zu Abschnitt 4.4.1 wurde verzichtet.

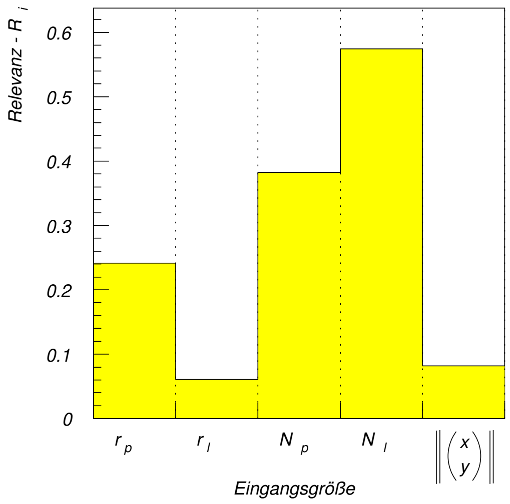



*Abbildung 4.13: Relevanzen für das 5-4-4-1 Netz mit arithmetischen Größen*

Das resultierende Netz besteht somit aus den zu Abschnitt 4.4.1 korrespondierenden arithmetischen Größen. $Ne$ wird durch das summierte $N_p$, $L_{(20-100)}$ durch das summierte $N_l$ und die Steilheit der Lateralverteilung $Age$ durch $r_p$ ersetzt.

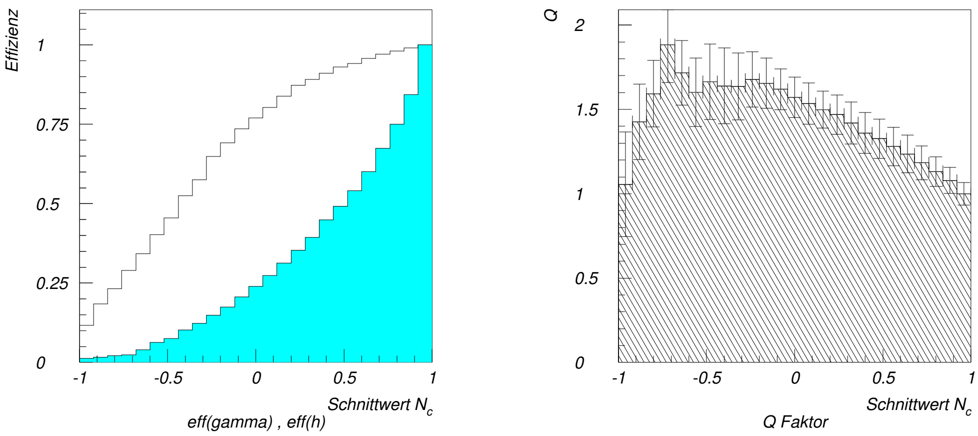



*Abbildung 4.14: Links: Effizienz für verschiedene Schnitte $N_c$. Als Linie die $\gamma$-Effizienz, schattiert die Hadronen-Akzeptanz. Rechts: Q-Faktor für das reduzierte Netz mit arithmetischen Größen.*

Die insgesamt auf der M$^{te}$ Carlo-Menge ermittelte Qualität ist Q=1.7±0.2 (Abbildung 4.14, rechts)[13], und ist im Rahmen der Fehlerschwankungen vergleichbar mit dem Resultat aus Abschnitt 4.4.1 für das einfache PEARL-Netz. Die getrennte Verwendung von $N_l$ und $N_p$ bringt also keinen Vorteil gegenüber der Verwendung des Quotienten dieser Größen, dies deckt sich mit der Argumentation in Abschnitt 4.1.1.

[13] Es wird angenommen das die Erhöhung bei $N_c$=-0.75 nicht statistisch signifikant ist.

Interessant ist nun wie sich dieser Wert ändert, wenn die Forderung nach konvergierten Lateralanpassungen fallengelassen wird. Tabelle 4.1 zeigt, daß in diesem Fall zur Gesamtanzahl der die Daten-Vorselektion passierenden Ereignisse nochmals 2/3 relativ schlecht ausgeprägter Ereignisse neu hinzukommen.
Mit einem neu trainierten Netz und den oben ausgewählten Observablen wurde auf der Gesamtmenge ein Wert von Q=1.5±0.1 ermittelt. Die Reduktion des Q-Faktors der Methode durch Minderung der Qualitätsanforderung an die Ereignisse ist also marginal. Zusammen mit dem ebenfalls in Tabelle 4.1 angegebenen intrinsischen Q-Faktor Q=0.97 der $M^{te}$ Carlo-Menge mit konvergierten Lateralanpassungen ergibt sich also bei der Verwendung arithmetischer Größen kein entschiedener Vorteil für die eine oder andere Ereignis-Vorselektion.

### 4.4.3 Verallgemeinerung auf mehrere Ereignisklassen

In vielen Anwendungsfällen von Neuronalen Netzen in der Physik sollen zwei Ereignisklassen voneinander getrennt werden, Signal- und Untergrundklasse. Das Standardverfahren ist dann, auf die in den vorigen Abschnitten dieses Kapitels dargestellte Weise durch einen Schnitt in einer Ausgabevariable des Netzes ein binäres Merkmal der Klassenzuordnung zu gewinnen.
Bei mehr als zwei Klassen ist die Verwendung einer einzigen Ausgabegröße nicht mehr möglich, da jeder Zuordnung der Klassen zu einem bestimmten Wert in der kontinuierlichen Ausgabevariable eine A priori-Annahme über Ähnlichkeiten der Klassen zugrunde liegt. Die Lösung dieses Problems ist die binäre Kodierung der Trainingsvektoren, d.h. jeder Klasse wird ein Ausgabeknoten zugeordnet und das Training wird beispielsweise bei drei Klassen für die Vektoren $(1,0,0),(0,1,0),(0,0,1)$ durchgeführt. Für diese Kodierung muß die in den vorigen Abschnitten verwendete Kostenfunktion durch die auf mehr als zwei Klassen verallgemeinerte *Cross Entropy* Fehlerfunktion ersetzt werden,

$$E(w) = -\sum_n \sum_i t_{n,i} \ln(y_{n,i}) \tag{4.7}$$

und als Transferfunktion der Ausgabeknoten die *Softmax* Funktion eingesetzt werden.

$$y_i = \frac{e^{z_i}}{\sum_k e^{z_k}} \tag{4.8}$$

Die Normierung in Formel 4.8 bewirkt, daß die $y_i$ direkt als A posteriori Wahrscheinlichkeiten der Klasse $i$ interpretiert werden können.
Es wird kurz der Fall diskutiert, daß drei Klassen von Primärteilchen unterschieden werden sollen, um zu untersuchen, ob die physikalisch motivierte Zuordnung aller Hadron-induzierten Ereignisse zu einer Klasse (He, O, Fe sind "*hadronischer*" als Protonen, vgl. Abschnitt 1.4.1) auch innerhalb des Rahmens der Klassifikation mit Neuronalen Netzen optimal ist. Die drei Klassen werden definiert durch[14]:

- $\gamma$-induzierte Ereignisse $(N_1 = 1, 0, 0)$
- durch Protonen induzierte Ereignisse $(0, N_2 = 1, 0)$
- durch He,O,Fe induzierte Ereignisse $(0, 0, N_3 = 1)$

Für den Test wird ein Netz mit den Größen $N_l$, $N_p$, $r_p$ und $r_l$ verwendet und mit der Netztopologie 4-5-5-3. Weitere Eingangsgrößen und Knotenanzahlen wurden untersucht, zeigten aber keine wesentlich anderen Resultate. Abbildung 4.15 zeigt die anhand der NCV-Validationsmenge ermittelten Netzausgaben der drei Knoten getrennt nach den drei Klassen. Die Anzahl der Einträge in den einzelnen Histogrammen wurde auf eins normiert. In Zeile 3 der Abbildung 4.15 ist deutlich zu sehen, daß der den schwereren Kernen zugeordnete Ausgabeknoten $N_3$ nur eine geringe Klassifikationsfähigkeit erlernt und praktisch alle Ereignisse der Klasse $\{He, O, Fe\}$ durch den Knoten $N_2$ dargestellt werden (Zeile 2). Dies zeigt, daß unter Verwendung der benutzen Eingangsgrößen eine weitere Unterteilung der Hadronen nicht möglich ist oder die dazu

[14] Die Entscheidung für die Unterteilung der Hadronen in die Klassen $\{P\}$ $\overline{\{P\}}$ wurde aufgrund der geringen Statistik der M$^{te}$ Carlo-Menge für schwerere Elemente getroffen, in physikalischer Hinsicht wäre auch die Einteilung $\{P, He\}$ $\overline{\{P, He\}}$ sinnvoll.

notwendigen Informationen in den Eingangsvektoren durch stärkere Merkmale zur Unterscheidung zwischen $\gamma$-und Hadron-induzierten Ereignissen unterdrückt werden (siehe z.B. auch [40] für eine analoge Diskussion).

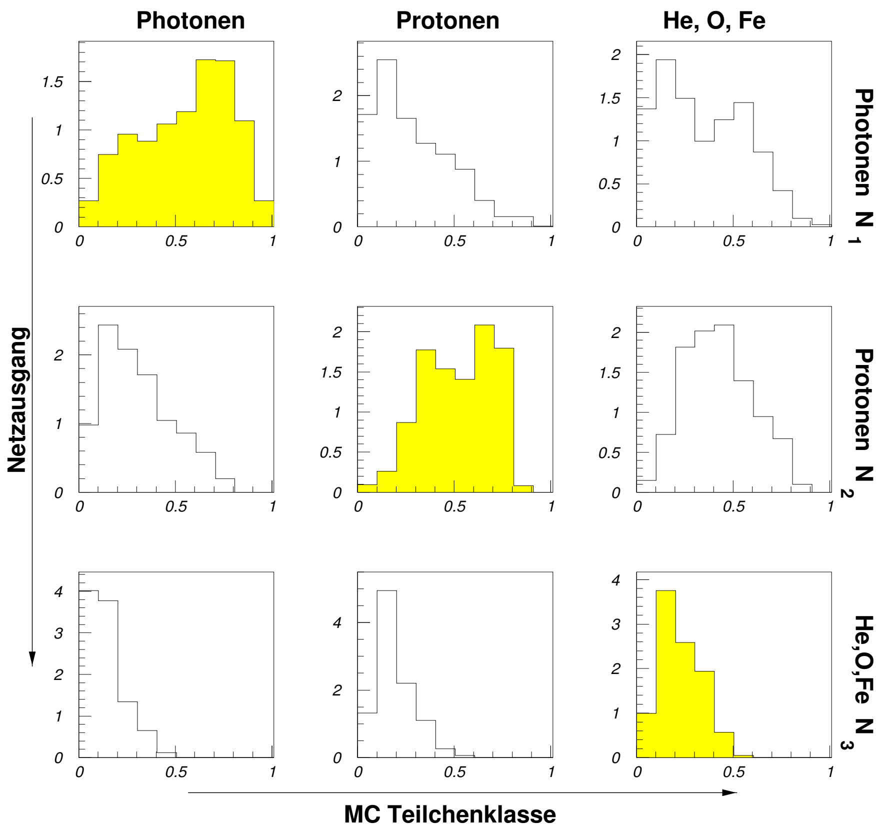



*Abbildung 4.15: Netzantworten der drei Ausgabeknoten nach $M^{te}$ Carlo-Klasse unterteilt.*

In Abbildung 4.16, links ist die Netzausgabe $N_2$ gegen die Netzausgabe $N_3$ aufgetragen. Aus der Normierungsbedingung der Ausgabeschicht (Formel 4.8) ergibt sich für die $\gamma/h$-Diskriminante S1 eine Gerade (dies ist äquivalent zu einem Schnitt in der den Photonen zugeordneten Ausgabe $N_1$). Der Abszissenabschnitt und die Steigung der Geraden wurden durch ein einfaches Gitterverfahren bestimmt und auf den Q-Faktor optimiert.
In Abbildung 4.16, rechts ist die aus der Bedingung

$$Err_{Schnitt} = (1 - \epsilon_P) + (1 - \epsilon_{He,O,Fe}) \stackrel{!}{=} Min \tag{4.9}$$

für die Unterteilung der Hadron-induzierten Ereignisse ermittelte Diskriminante S2 der Klassen $\{P\}$ und $\{He, O, Fe\}$ eingezeichnet.

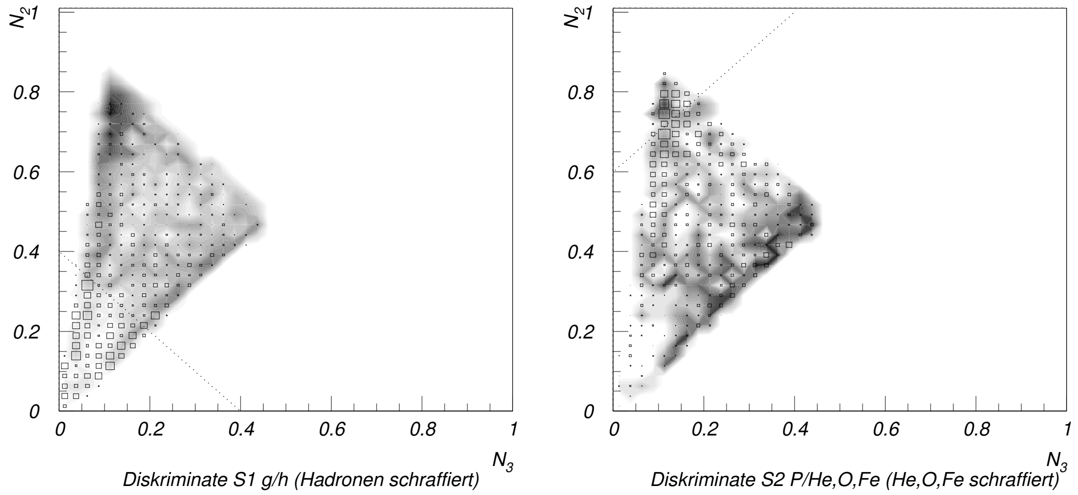



*Abbildung 4.16: Diskriminanten zur $\gamma/H$-Separation und zur Unterscheidung $\{P\}/\{He,O,Fe\}$, dargestellt in der Ebene der Ausgänge $N_2$ und $N_3$.*

| Netz-Klasse | | $\mathrm{M}^{te}$ Carlo-Klasse $\gamma$ | P | He,O,Fe |
|---|---|---|---|---|
| $\gamma$ | $(S_1)$ | 45% | 7% | 9% |
| P | $(\overline{S_1} \cap S_2)$ | 53% | 81% | 85% |
| He,O,Fe | $(\overline{S_1} \cap \overline{S_2})$ | 1% | 12% | 6% |

*Tabelle 4.4: Resultierende Effizienzen für das Netz mit drei Ausgängen.*
*Die Bedingung $S_1$ bedeutet, daß das Ereignis in der $N_2$-$N_3$-Ebene als $\gamma$ klassifiziert wurde, die Bedingung $S_2$ bedeutet das es als Proton klassifiziert wurde.*

Der aus der Projektion der Einträge auf die Senkrechte zur $\gamma/h$ Diskriminante (Abbildung 4.16, links) ermittelte Q-Wert ist 1.6±0.1 und somit kompatibel zu den Resultaten der beiden vorigen Abschnitte. Eine weitere Unterteilung der Hadron-induzierten Ereignisse ist also nicht möglich, wie Tabelle 4.4.3 für die Selektions-Effizienzen der Klassen (ausgedrückt durch die Diskriminanten S1 und S2) zeigt, und für die $\gamma/H$-Separation auch nicht sinnvoll.

## 4.5 Neuronale Netze mit direkten Eingabegrößen

Den beiden folgenden Abschnitten liegt das Motiv zugrunde, daß Integrations- und Summationprozesse immer einen Verlust an Information bedeuten. Ein "klassischer" Ansatz im Bezug auf Neuronale Netze dies zu verhindern ist es, möglichst minimal vorverarbeitete Merkmale zu verwenden. Dadurch wird die Möglichkeit Neuronaler Netze genutzt, in höherdimensionalen Räumen zu interpolieren und Korrelationen aufzufinden.

Es bieten sich eine Reihe von Möglichkeiten für die Abbildung der Detektordaten auf die Eingangsgrößen des Neuronalen Netzes an:

1. Jeder kalibrierten Einzel-Amplitude der AIROBICC- und/oder Szintillator-Matrix wird ein Eingangsknoten zugeordnet.
2. Transformation in ein vom Kernort abhängiges Koordinatensystem und Darstellung wie bei 1.
3. Verwendung der über Kreisringe gemittelten AIROBICC- oder Szintillator-Amplituden.
4. Sortierung der Signale nach Amplitude oder Distanz zum Kernort, mit und ohne zusätzlich räumliche Information der sortierten Einträge.
5. Räumliche Kombination der von den beiden Matrizen gelieferten Werte, z.B. Verhältnis aus AIROBICC-Amplitude und Erwartungswert der NKG-Funktion am Ort des zugeordneten AIROBICC-Detektors.

Alle diese Darstellungsformen für die Eingangsgrößen wurden untersucht, jedoch zeigten sie nirgends eine den Netzen mit integralen Größen gleichwertige oder sogar überlegene Leistungsfähigkeit. Exemplarisch werden in den beiden folgenden Abschnitten Netze beschrieben, die auf den Darstellungsformen 2 und 3 basieren.

### 4.5.1 Ringweise gemittelte Čerenkov-Lichtdichte

In [37] beschreiben die Autoren ein Verfahren zur $\gamma/H$-Separation, daß auf den Čerenkov-Lichtdichten bei 30 m, 80 m und 150 m Kernortabstand basiert. Sie beziehen sich hierbei auf vollständige M$^{te}$ Carlo-Informationen, also auf alle Čerenkovphotonen und elektromagnetischen Partikel auf Detektorniveau. Grundsätzlich sollte ein Neuronales Netz diese Informationen auch bei realistischen Detektorbedingungen in die $\gamma/H$-Separation zumindest teilweise einbeziehen können[15]. Als Repräsentation der Ereignisse wird die Darstellung 3 benutzt. Die Anzahl und Größe der Ringflächen wird durch die maximale Ausdehnung (205 m) und die Detektorabstände (30 m) bestimmt. Nach einigen Versuchen mit anderen Werten wurden 7 Ringe mit jeweils konstanter radialer Ausdehnung von 30 m gewählt. An die Ereignisse wird die Bedingung gestellt, daß die rekonstruierten Kernorte innerhalb der Matrix liegen, so daß zumindest die inneren 4 Ringe immer gefüllt sind und bei Schauern am Matrix-Rand der Bereich bis 200 m beobachtet werden kann. Folgende Netze wurden untersucht:

- Netz **A** : 7 Eingangsknoten für die über Ringflächen gemittelten AIROBICC-Amplituden.
- Netz **B** : wie Netz **A**, aber die 7 Eingangsknoten für die AIROBICC-Amplituden sind normiert und die Normierungskonstante wird an einen zusätzlichen Eingabeknoten angelegt.
- Netz **C** : wie Netz **B**, aber mit einem zusätzlichen Eingabeknoten für $N_p$.

Keines der oben aufgeführten Netze zeigt ein auf der Ringinformation basierendes signifikantes Lernverhalten. Alle Netze, insbesondere Netz **A**, neigen stark zur Überanpassung, dies ist

[15] Unterhalb von 100 TeV Primärenergie ist die Lichtdichte bei 150 m Entfernung zum Kernort nicht mehr meßbar, da sie unterhalb der Nachweisschwelle der AIROBICC-Detektoren liegt.

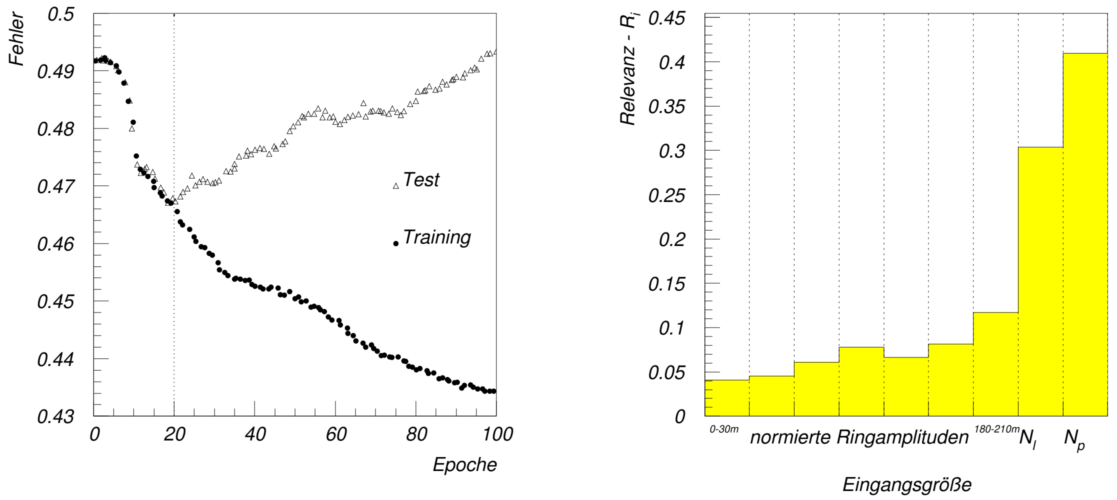



*Abbildung 4.17: Links: Fehlerverlauf für Netz A, das Netz neigt nach Epoche 20 deutlich zur Überanpassung.*
*Rechts: Relevanzen der Eingangsgrößen für Netz C. Die ersten 7 Relevanzen sind den normierten Ringamplituden zugeordnet.*

in Abbildung 4.17, links für ein 7-7-7-1 Netz nach bereits 20 Epochen (mit je 10 Trainingsdurchläufen) zu sehen. Nur Netz **C** zeigt eine im wesentlichen auf $N_p$ und die Normierungskonstante (entspricht $N_l$) basierende Lernleistung. Das Relevanzen-Diagramm für Netz **C** (Abb. 4.17,rechts) weist neben den hohen Werten für $N_p$ und der Normierungskonstante einen interessanten Aspekt auf: Die nach außen ansteigenden Relevanzen der Ring-Knoten scheinen ein Substitut für $r_l$ zu sein, da nach der Normierung der Ringamplituden deren Werte nach außen hin immer sensitiver auf die Lateralverteilung des Čerenkovlichts werden. Dieser Interpretation wird gestützt durch die Tatsache, daß dieser Effekt für ein Netz mit einem zusätzlichen Eingang für $r_l$ verschwindet. Zusammenfassend zeigt sich, daß die Verwendung der kalibrierten AIROBICC-Signale unter Berücksichtigung der Detektor-Ausdehnung und dessen Granularität zu keiner Verbesserung der $\gamma/H$-Separationsleistung gegenüber den in diesem Kapitel bereits gezeigten Ansätz führt. Als problematisch stellt sich dabei die notwendige Erhöhung der Eingabeknoten und der für die Extraktion komplexerer Merkmale erforderliche Anzahl von versteckten Knoten dar. Die verwendeten Neuronalen Netze neigen in diesem Fall dazu, die in den ringweise gemittelten AIROBICC-Signalen enthaltene zusätzliche Information zur Überanpassung der Trainingsmenge zu verwenden.

### 4.5.2 Translationsinvariante Repräsentation der Szintillator Amplituden

Der Ausnutzung der dem jeweiligen Problem zugrunde liegenden Symmetrien kommt bei dem Einsatz von Neuronalen Netzen eine besondere Bedeutung zu. Die dadurch erreichte Einschränkung des Lösungsraums bewirkt eine erhebliche Reduktion der Konnektivität (Gewichte) und der Komplexität (Knoten) und somit eine Milderung der Gefahr von Überanpassung.
Der Translations-Freiheitsgrad kann z.B. durch die Verwendung *rezeptiver Felder* genutzt werden, bei denen die Eingabeknoten (die Eingabevektoren werden als Darstellung einer 2D-Matrix aufgefaßt) in Sub-Felder unterteilt werden, die ihre Gewichte zur nächsten Schicht untereinander teilen. Die robusteste und im Hinblick auf das Training effizienteste Möglichkeit jedoch ist die Vorverarbeitung der Daten auf eine Weise, die eine invarianten Repräsentation der Eingabe sicherstellt. Die in Abschnitt 4.5.1 gezeigt Verwendung ringweiser gemittelter Amplituden leistet diese in natürlicher Weise, sowohl für Rotation als auch Translation. Um jedoch zu untersuchen, ob die in Hadron-induzierten Schauern beobachtete größere Fluktuation in der lateralen Verteilung der elektromagnetischen Komponente zusätzlich eine Verbesserung der Separations-Fähigkeit ermöglicht, ist der Mittelungsprozeß ungünstig.

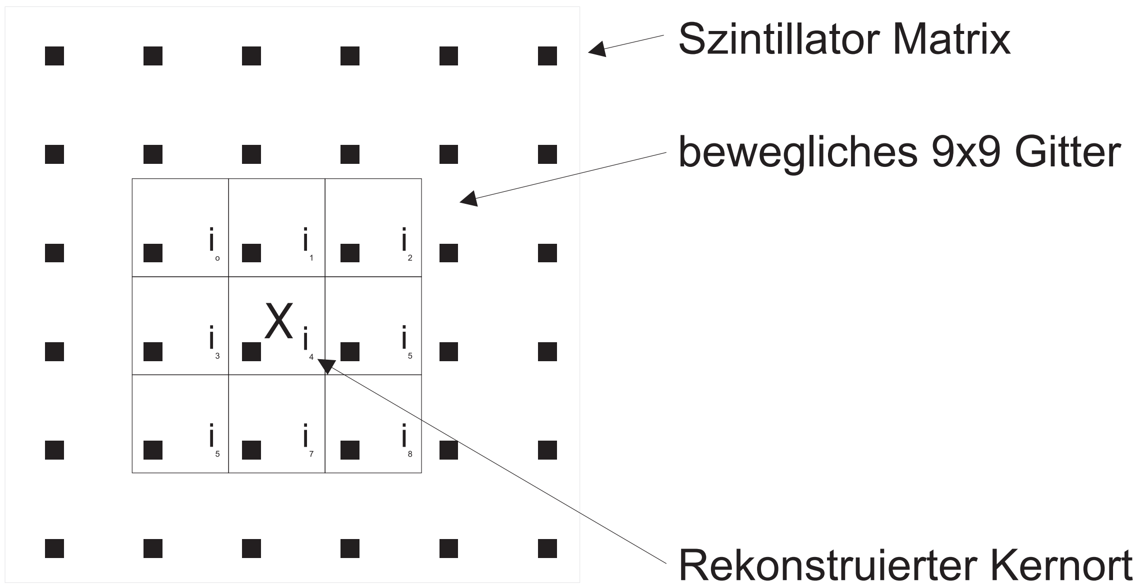
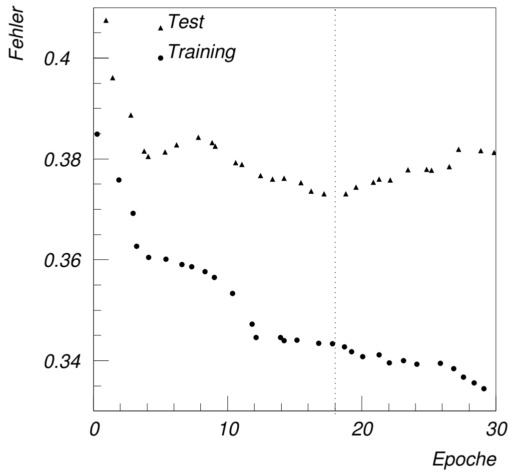



*Abbildung 4.18: Skizze des Hilfsgitters zur translationsinvarianten Darstellung der Szintillator-Signalamplituden.*

Das Verfahren basiert auf einem virtuellen Gitter, dessen Mittelpunkt durch den rekonstruierten Kernort bestimmt wird. Die Einträge der Gitterzellen werden durch die mittleren gewichteten[16] Einträge der enthaltenen Szintillator-Detektoren bestimmt. Anschließend wird die Summe der Einträge normiert und der Merkmalsraum somit um eine weitere Dimension reduziert. Nach mehreren Versuchen hat sich eine Zellgröße von 15 m $\times$ 15 m und 9 Zellen als günstig[17] erwiesen. Um

[16] entsprechend der Schnittmenge zwischen der Zellfläche und der Brillouin-Zone des Szintillator-Detektors [20].
[17] Die untere Grenze wird durch die Granularität des Detektors bestimmt.

auf die Verwendung von virtuellen Einzeldetektoren [20] verzichten zu können (die Verwendung könnte zu unvorhergesehenen Systematiken führen) wurde die verwendete M$^{te}$ Carlo-Menge auf Schauer beschränkt, deren Kernorten mindestens 15 m innerhalb der Matrix liegen. Das Netz besteht aus einer 11-4-4-1 Topologie mit 9 Eingängen für die Zellen und zusätzlich je einem Eingang für $N_p$ und $N_l$. Diese Form der Zuordnung gewährleistet, daß nur den 9 Zell-Knoten Informationen der lateralen Verteilung zugänglich sind. Die Verwendung des Partikel- zu Licht-Verhältnisses wird von den übrigen beiden Knoten übernommen. Der Verlauf des Netzfehlers zeigt, daß auch bei einer moderaten Wahl von 4 Knoten in den versteckten Schichten das Netz stark zur Überanpassung neigt (Abbildung 4.19, links). Das Training wurde deshalb nach der Epoche 18 abgebrochen und der Q-Faktor ermittelt; dies entspricht einer Regularisierung mit *early stopping*.



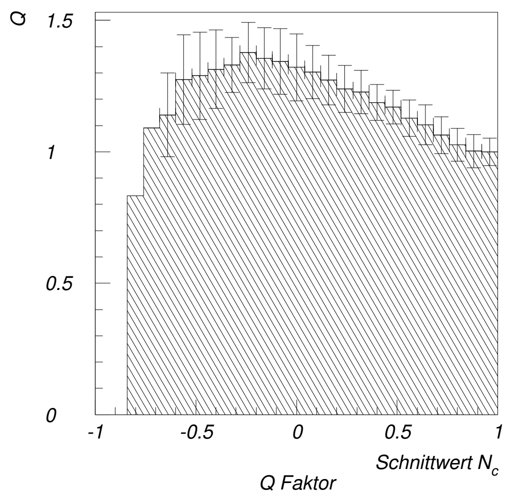




*Abbildung 4.19: Links: Fehlerverlauf des Netzes mit Zell-Eingängen. Rechts: Erreichter Q-Wert*

Im Vergleich zu den in 4.4.1 und 4.4.2 dargestellten Netzen zeigt sich trotz der durch den Kernort-Schnitt härteren Bedingungen an die Ereignisselektion eine schlechtere $\gamma/H$-Separation des Netzes. Dies läßt sich entweder durch einen deutlich höheren Bedarf an M$^{te}$ Carlo-Ereignissen oder durch die Abwesenheit verwendbarer Merkmale auf dem Niveau der Einzeldetektor-Signale erklären. Der errechnete Q-Wert von 1.4$\pm$0.1 liegt leicht oberhalb des mit demselben Kernort-Schnitt und nur $N_l$ und $N_p$ erreichbaren Werts von 1.3$\pm$0.1 .
Die Vermutung, daß aus den Zell-Informationen ein zu $r_p$ äquivalentes Merkmal extrahiert wird, wird anhand der Abbildung 4.20 verstärkt, die die skalierten Gewichte der den Zellen zugeordneten Knoten zeigt. Links sind die skalierten Gewichte für das zur $\gamma/H$-Separation trainierte Neuronale Netz gezeigt. Die Zahlen in der x-y-Ebene geben die Knoten-Nummern an. Von den

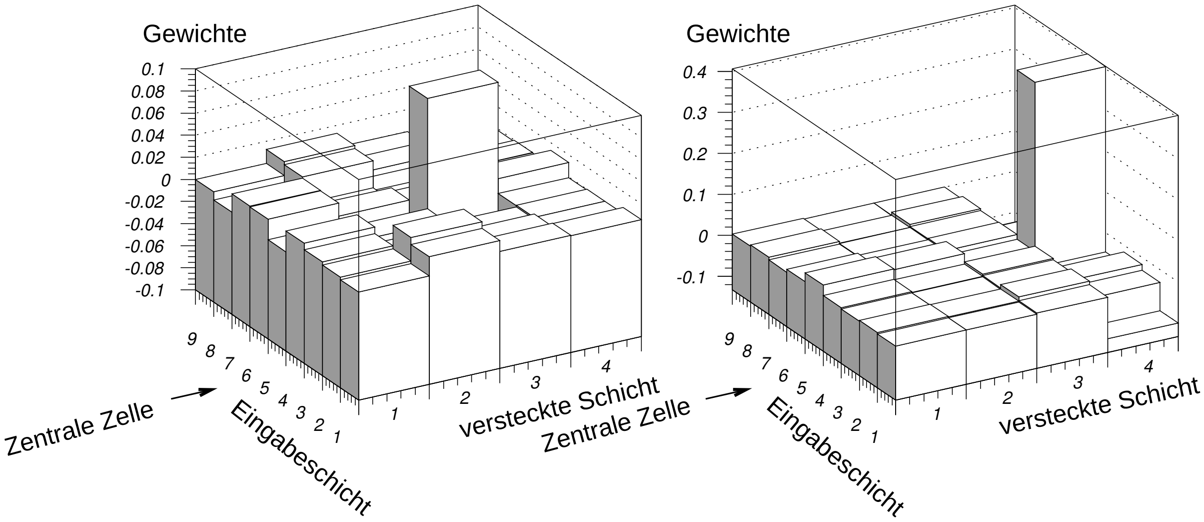



*Abbildung 4.20: Skalierte Gewichte der Zellknoten,links für das daß $\gamma/h$-Netz, rechts für das $r_p$-Netz.*

der zentralen Zelle (Knoten 6 der Eingabeschicht) zugeordneten Gewichten besitzt eines einen deutlich erhöhten Wert. Die Gewichte der Randzellen sind vergleichsweise schwach ausgeprägt und ergeben zusammen einen leicht negativen (inhibitorischen) Überschuß.
Rechts sind die entsprechenden Gewichte für ein mit $r_p$ als Zielgröße trainiertes Netz gezeigt. Ganz äquivalent existiert ein stark positives Gewicht für die zentrale Zelle und leicht negative Gewichte der Randzellen[18]. Der Korrelationskoeffizient zwischen $r_p$ und der Netzausgabe beträgt 0.5. Wenn man berücksichtigt, daß die Anzahl von Zell-Knoten mit 9 Knoten relativ klein gewählt wurde, zeigt dieser Korrelationskoeffizient, daß tatsächlich bezüglich $r_p$ eine Lernleistung vorhanden ist. Aus dem Vergleich der skalierten Gewichte kann also geschlossen werden, daß es sich bei der in Abbildung 4.20, links vorhandenen Signatur tatsächlich um eine Repräsentation von $r_p$ handelt.
Insgesamt zeigt sich, wie bei dem Netz mit ringweise gemittelter Čerenkov-Lichtdichte, daß die analytischen Observablen $r_l$ bzw. $r_p$ den aus den Detektor-Amplituden extrahierten Merkmalen überlegen sind; interessant ist jedoch, daß in beiden Fällen tatsächlich zu den analytischen Größen ähnliche Merkmale extrahiert werden.

[18]Die negativen Gewichte befinden sich alle zwischen den Randzellen und Knoten 4 der versteckten Schicht.

## 4.6 Der quasi-ideale Detektor und $\gamma/H$-Separation mit Čerenkov-Teleskopen

Dieser Abschnitt beschäftigt sich mit der Frage, inwieweit die $\gamma/H$-Separation durch intrinsischen Fluktuationen der Schauer und prinzipielle Ähnlichkeiten der physikalisch beobachtbaren Observablen begrenzt wird, oder ob sie durch den speziellen Aufbau (niedrige Granularität und im Vergleich zum Molière Radius geringe Ausdehnung) des HEGRA-Detektors determiniert wird. Insbesondere die niedrige Energieschwelle des Detektors führt dazu, daß teilweise nur wenig Informationen des Luftschauers abgetastet werden können.
Der im folgenden untersuchte ideale Detektor besitzt eine 100% Bedeckung mit Szintillator- und mit AIROBICC-Detektoren[19]. Es werden nur Ereignisse benutzt, die den realen Detektor ausgelöst hätten. Die Größe $L_{(20-100)}$ wird aus der Verteilung aller Čerenkov-Photonen auf Detektorniveau bestimmt. Die Anpassung der Exponentialfunktion wird dabei von 20 m bis 100 m in Ringintervallen von 1 m durchgeführt. Die Größen $Ne$ und $Age$ werden aus einer NKG-Anpassung an die Verteilung der elektromagnetischen Teilchen im Abstand von 10 m - 200 m von der Schauerachse ermittelt, so wie sie vom M$^{te}$ Carlo-Ereignisgenerator CORSIKA auf Detektorniveau geliefert werden. Hier basiert die Anpassung allein auf der Teilchendichte der $e^+e^-$. Hochenergetische $\gamma$ werden nicht berücksichtigt, da dies die Simulation des Szintillationsmaterials und der Bleischicht voraussetzen würde, die hier nicht zur Verfügung stand. Obwohl die Lateralverteilung der $e^+e^-$ ohne Simulation der Bleischicht deutlich flacher ist als die aus der vollen HEGRA-Simulation, sollte sich hieraus prinzipiell für die betrachtete $\gamma/H$-Separation kein Effekt ergeben, da dies in guter Näherung gleichermaßen für $\gamma$- als auch Hadron-induzierte Schauer gilt. Abbildung 4.21, links zeigt alle Schauerteilchen auf Detektorniveau für einen exemplarischen 80 TeV Proton-Schauer, zusammen mit der in radialen Intervallen bestimmten Partikelanzahl. Es ist zu erkennen, daß ein solcher Schauer eine große Anzahl von Partikeln mit detaillierter Information über deren Lateralverteilung enthält. Abbildung 4.21, Mitte und rechts zeigen die aus der Information des realen und des idealen Detektors berechnete Größe $Age$ für $\gamma$ und Hadronen. Die Korrelation der jeweiligen Größen ist moderat, dies zeigt die großen Rekonstruktionsungenauigkeiten bei den mit dem realen Detektor gemessenen Größen.

Für die Untersuchung der Auswirkungen der Kernort-Bestimmung werden die folgenden Ereignismengen definiert:

- Menge A: keine Einschränkung des Kernortes
- Menge B: Einschränkung des M$^{te}$ Carlo-Kernorts auf Werte innerhalb der Detektor Matrix

[19] Man könnte so einen Detektor realisieren, indem man die AIROBICC- auf die Szintillator-Detektoren stellt.

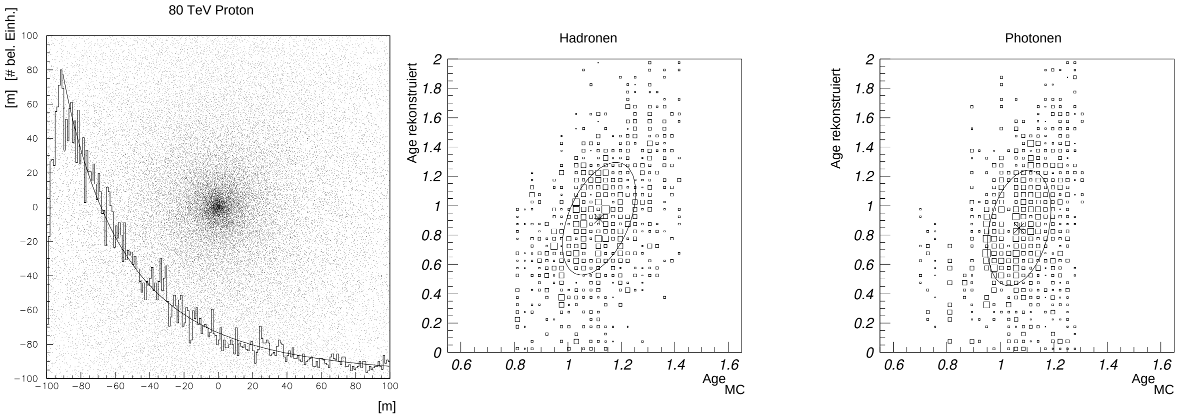



*Abbildung 4.21: Links: Räumliche Verteilung der $e^+e^-$ auf Detektorniveau eines 80 TeV Proton Schauers. Der Kernort des simulierten Schauers liegt im Koordinaten-Ursprung. Die überlagerten Kurven zeigen die Teilchenanzahlen innerhalb von 1m Ringsegmenten um den Ursprung und die NKG Anpassung. Rechts: Korrelation zwischen dem so bestimmten Age und dem mit der HEGRA-Rekonstruktion ermittelten. Die Ellipsen verdeutlichen die 2te Momente Korrelationsergebnisse.*

$Q_{max}$

| | M$^{te}$ Carlo-Kernort | | Rekonstruierter Kernort | |
|---|---|---|---|---|
| | Ereignismenge B | Ereignismenge A | Ereignismenge B | Ereignismenge A |
| nur M$^{te}$ Carlo- | $> 4$ | $> 4$ | – | – |
| $L_{(20-100)}$ real | 3.9 ±0 .4 | 3.6 ±0.4 | – | – |
| $Ne, Age$ real | 3.6 ±0 .4 | 3.4 ±0.4 | – | – |
| alle real | 3.4 ±0 .4 | 2.7 ±0.3 | 2.5 ±0.3 | 1.7 ±0.2 |

*Tabelle 4.5: Effekt der Rekonstruktionsgüte auf die resultierende $\gamma/H$-Separation: In der ersten Zeile werden Oberservablen, die nur im M$^{te}$ Carlo zur Verfügung stehen, als Eingangsgröße des NN verwendet. In der zweiten und dritten Zeile werden diese durch die äquivalenten Größen aus den Anpassungen an die simulierten HEGRA Ereignisse ersetzt. In der letzten Zeile werden nur Größen aus den Anpassungen an die simulierten HEGRA Ereignissen verwendet. Die Ereignismengen A und B werden im Text erläutert.*

In Tabelle 4.6 sind die mit verschiedenen Neuronalen Netzen gewonnen Ergebnisse zusammengefaßt. Werden die vollständigen M$^{te}$ Carlo-Informationen verwendet (Reihe 1), so läßt sich eine sehr gute Trennung mit Q>4 erreichen (siehe Abbildung 4.22). Der Austausch der idealen M$^{te}$ Carlo-Größen $L_{(20-100)}$ bzw. $Ne$ und $Age$ durch die reale Observablen (Reihe 2 und 3) verschlechtert die Trennungsmöglichkeit erheblich und der Q-Faktor sinkt deutlich unter 4.
Dieser Effekt ist wie zu erwarten bei der Ersetzung der Observablen $Ne$ und $Age$ etwas größer, weil zwei Größen durch ihrer realistischen Werte ersetzt werden, und weil die Messung der elektromagnetischen Komponente prinzipiell aufgrund der kleineren Teilchenanzahl und den größeren Fluktuationen ungenauer ist.

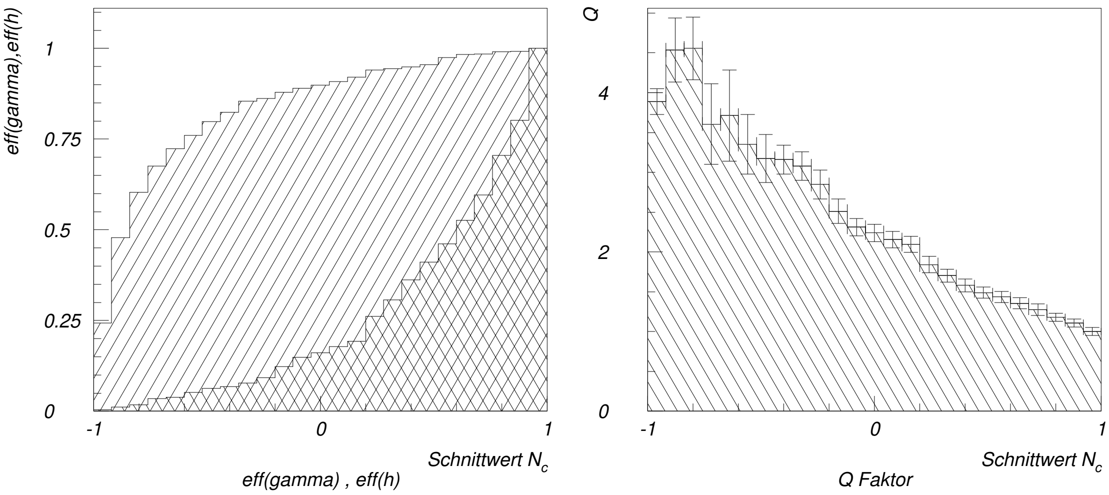



*Abbildung 4.22: Effizienzen ($\gamma$: gestrichelt, Hadronen: gekreuzt ) und Q-Faktor der $\gamma/H$-Separation mit idealen $M^{te}$ Carlo-Größen (Tabelle 4.6, Zeile 1).*

Die Werte in Zeile 4 zeigen die Auswirkung der Ereignisrekonstruktion auf die Separationsfähigkeit: Die prinzipiell schlechtere Rekonstruktionsgüte von Ereignissen mit $M^{te}$ Carlo-Kernort außerhalb der Detektor-Matrix (ca. 30% der $M^{te}$ Carlo-Menge) bewirkt eine drastische Minderung des integralen Q-Faktors, ersichtlich im Übergang von Spalte 1 zu Spalte 2. Der Fehler der Kernortrekonstruktion $\sigma_{63\%}$, der innerhalb des Detektorfelds 6 m - 10 m und außerhalb 20 m - 50 m beträgt [20], bewirkt eine weitere Minderung, bis zu den in diesem Kapitel für die realistische HEGRA Rekonstruktionsgüte angegegebenen Werten.
Für Menge B sinkt Q durch den Übergang vom idealen zum realen Detektor von Q>4 auf Q=2.5 und für Menge A von Q>4 auf Q=1.7 .
Der mit dem idealen Detektor erzielte Q-Faktor ist also im Vergleich zu den in den vorigen Abschnitten aufgeführten Werten erstaunlich hoch.
$M^{te}$ Carlo-Studien für Photonen mit Energien oberhalb von 100 TeV zeigen [41], daß für Luftschauer, die sehr viele AIROBICC- und Szintillator-Detektoren ansprechen, auch mit dem realen HEGRA-Detektormatrizen Q-Faktoren oberhalb von Q=4 erreicht werden können. Diese guten Ergebnisse, die mit der *Shower Front Sampling-Technik* bei einer hohen Abtastung der Schauerinformation erreichbar sind, legen einen Vergleich mit der bei den **A**bbildenden **Č**erenkov **T**eleskope (ACT) erreichten $\gamma/H$-Separation nah.

ACT erreichen im Energiebereich zwischen mehreren hundert GeV und einigen TeV die höchste Sensitivität erdstationärer Detektoren.

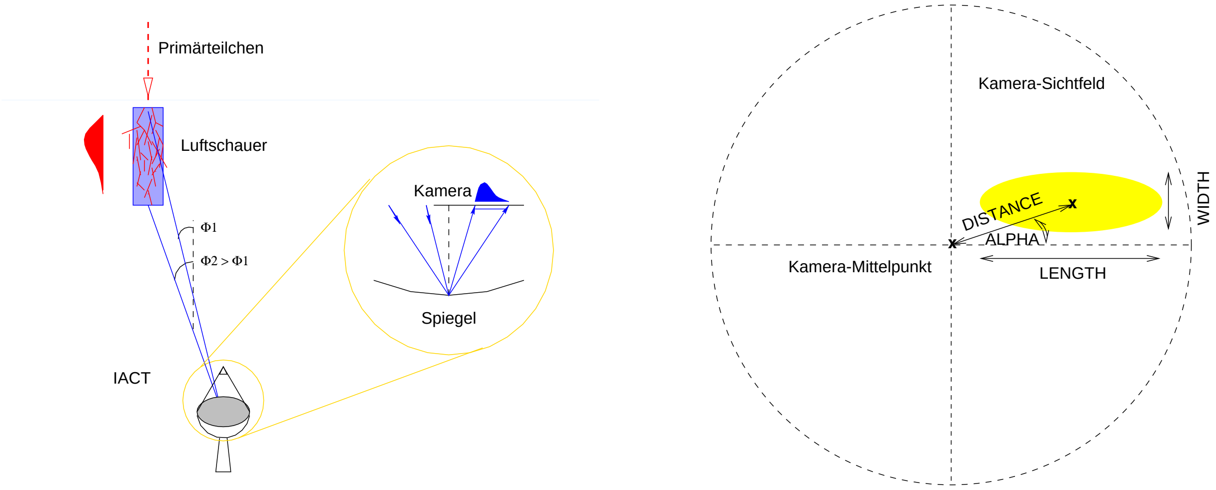



*Abbildung 4.23: Prinzipskizze der Imaging-Technik und Definition der Hillas-Parameter.*

Ein ACT bildet das vom Schauer erzeugte Bild auf eine Matrix von SEVs ab (*Imaging-Technik*). Die emittierenden Partikel erzeugen ein ellipsenartiges Bild (siehe Abbildung 4.6, links).
Die $\gamma/H$-Separation wird dann mit Schnitten in verschiedenen Bild-Parametern (den sog. Hillas-Parametern Abbildung 4.6, rechts) durchgeführt.
Sie basiert auf folgenden Tatsachen:

- Die in der Nukleon-Nukleon Kollision erzeugten Pionen haben im Schwerpunktsystem einen mittleren Transversalimpuls von 200 $\frac{MeV}{c}$ [42]. Dadurch ist die elliptische Form des Bildes für Hadron-induzierte Schauer wesentlich breiter.

- Elektromagnetische Kaskaden sind signifikant kompakter. Insbesondere bei niedrigen Energien führen in Hadron-induzierten Schauern $\pi$ und $\mu$ mit hohen Transversalimpulsen zu einer im Vergleich zu $\gamma$-induzierten Schauern ungleichmäßigeren lateralen (und winkelabhängigen) Verteilung des Čerenkovlichts [43]. Dadurch ist das Abbild Hadron-induzierter Schauer wesentlich ungleichmäßiger.

Die Hillas-Parameter sind in Tabelle 4.6 zusammengefaßt und den Matrix-Observablen gegenübergestellt. Für einzelne ACTs wird der ALPHA-Parameter zur Selektion auf die Quellrichtung verwendet. LENGTH und WIDTH spiegeln die longitudinale und laterale Schauerentwicklung wider. Für Schauer, deren Schauerachse parallel zur Teleskop-Achse verläuft, hängt DIST vom Kernortabstand und von der Höhe des Schauermaximums ab. SIZE ist von der elektromagnetischen Schauerenergie abhängig.
Abbildung 4.24 zeigt, wie sich die unterschiedlichen Merkmale für $\gamma$- und Hadron-induzierte Schauer in den Teleskop-Kameras darstellen. Die geringere Breite des $\gamma$-Schauers ist deutlich zu erkennen.

| Teleskop-Observable | Bildparameter | physikalische Bedeutung | Matrix-Observable |
|---|---|---|---|
| ALPHA | Winkel zwischen großer Hauptachse und Verbindungslinie von Bildschwerpunkt und Kamerazentrum | Orientierung der Schauerachse zur Teleskop-Achse | Richtung $(\theta, \phi)$ |
| DIST | Winkelabstand des Bildschwerpunkts vom Kamerazentrum | Lage des Schauermaximums (Höhe und Kernort) | Kernort $(x, y)$, $r_l$ |
| LENGTH | RMS-Wert der Lichtverteilung entlang der großen Hauptachse des Schauerbilds | Länge des Schauers | $N_l$ / $N_p$ , $L_{(20-100)}$ $/Ne$ |
| WIDTH | RMS-Wert der Lichtverteilung entlang der kleinen Hauptachse des Schauerbilds | Breite des Schauers | $Age$, $r_p$ |
| SIZE | Gesamtlichtmenge des Schauerbilds | Elektromagnetische Energie des Schauers | $\gamma_{100}$ |

*Tabelle 4.6: Vergleich zwischen Teleskop- und Matrix-Observablen*

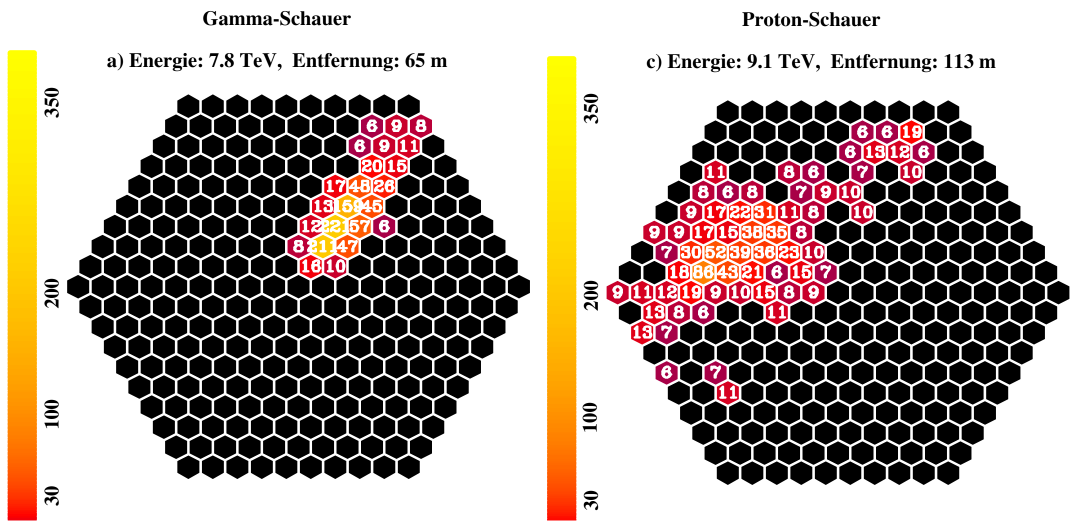



*Abbildung 4.24: Abbilder des von einem $\gamma$ bzw. Protonen Schauer erzeugten Čerenkovlichts in den aus SEV aufgebauten Kameras der ACT.*

Die $\gamma/H$-Separation mit ACTs basiert auf Schnitten in den Hillas Parametern WIDTH und LENGTH. Wie in der Tabelle 4.6 zu sehen ist, basieren diese Observablen auf den gleichen physikalischen Eigenschaften der Luftschauer, wie die für die $\gamma/H$-Separation mit AIROBICC- und Szintillator-Matrix verwandten Observablen.
Mit Einzelteleskopen erreicht man durch Schnitte in WIDTH und LENGTH Q-Faktoren zwischen 2.5 und 4, also durchaus vergleichbar mit denen, die mit einer ausgedehnten, 100% flächenbedeckenden Szintillator/AIROBICC-Kombination erreicht würden.

# Kapitel 5

# Anwendung auf Daten

Der Übergang von der auf $\mathrm{M}^{te}$ Carlo-Ereignissen basierenden Trainingsphase zu der Anwendungsphase mit realen Detektordaten stellt einen wichtigen Test für den verwendeten $\mathrm{M}^{te}$ Carlo-Generator sowie das Neuronale Netz mit seinen Eingangsgrößen und der gewählten Topologie dar. Da keine $\gamma$-Quellen bekannt sind, die mit den Matrix-Detektoren beobachtet werden konnten, bleibt dieser Test auf die Suche nach $\gamma$-Quellen und die Analyse des Hadron-induzierten Untergrunds beschränkt.
Der Untergrund bietet also die einzige Testmöglichkeit des Neuronalen Netzes. Falls Energiespektrum, chemische Zusammensetzung der hadronischen Komponente und die Akzeptanzen des Detektors korrekt modelliert wurden und keine vom $\mathrm{M}^{te}$ Carlo-Charakter der Trainingsmenge verursachten Systematiken erlernt wurden, wird man einen qualitativ und quantitativ ähnlichen Verlauf der Netzausgabe zwischen gemessenem Untergrund und Hadron-induzierten $\mathrm{M}^{te}$ Carlo-Ereignissen erwarten.

## 5.1 Beschreibung des AGN-Quellensatzes

Für die Analyse wurden 12 AGN des Blazar Typs (siehe Abschnitt 1.1) mit einer Rotverschiebung zwischen z=0.017 und z=0.062 ausgewählt [44],[16].
Bei der Auswahl wurde nach folgenden Kriterien vorgegangen :

- Der Blazar liegt im Sichtfeld des HEGRA-Detektors, d.h. die Deklination liegt in dem Bereich zwischen $0^o < \delta < 52^o$.
- Der Fluß des beobachteten Spektrums im Radiobereich fällt flacher als mit $\alpha = 0.5$, $s_\nu \approx \nu^{-\alpha}$ ab.
- Der Blazar ist in geringer Entfernung und somit innerhalb des Bereichs, in dem man trotz der Absorption am angenommenen IR-Hintergrund Photonen oberhalb der HEGRA-Energieschwelle erwarten kann.

Der Roh-Datensatz besteht aus $69 \cdot 10^6$ Ereignissen, die in der Zeit von 1. Juni 1994 bis 27. September 1995 aufgezeichnet wurden. Es wurden nur Ereignisse zugelassen, bei denen die rekonstruierte Richtung innerhalb einer $6^o$ Umgebung um eine der potentiellen Quellen lag und der Zenitwinkel in lokalen Koordinaten kleiner als $30^o$ war.

## 5.2 Analysemethode

Um Fehlfunktionen des Detektors auszuschließen und weil Messungen mit AIROBICC empfindlich den atmosphärischen Bedingungen unterliegen, wurden folgende Bedingungen an die zur Analyse zugelassenen Meß-Intervalle (Runs) des Beobachtungszeitraums gestellt :

- Median $r_l > 60m$
- Median AIROBICC Detektoren mit Signal $> 10$
- Median $\chi^2$ AIROBICC Kegel-Anpassung $\leq 1.6ns$
- Median der summierten AIROBICC Signale $(N_l) \geq 40 \cdot 10^3 phot/m^2$
- $\geq 70\%$ der Ereignisse wurden von AIROBICC ausgelöst

Auf Ereignis-Ebene wurde zusätzlich eine Korrektur der AIROBICC Einzeldetektor-Signale durchgeführt [45]. Die Korrektur kompensiert Akzeptanz-Änderungen der Detektoren in lokalen $\theta$ und $\phi$ Koordinaten, wie sie z.B. durch Änderungen der atmosphärischen Bedingungen oder langfristig durch Alterungsprozesse der SEV verursacht werden können. Dieses Verfahren ist insbesondere für Quellen bei hoher oder niedriger Deklination, die während der gesamten Periode ausschließlich am Rande der AIROBICC Akzeptanz sichtbar sind, sinnvoll.

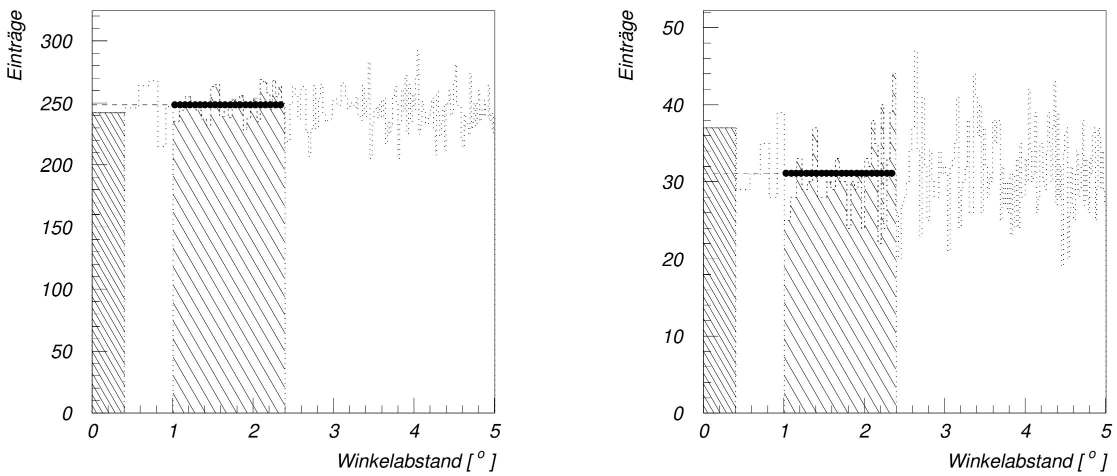



*Abbildung 5.1: Signal- und Untergrundbestimmung für Ereignisse an der Position von Quelle 7, links ohne, rechts mit $\gamma/H$-Separation. Aufgetragen sind die Einträge in den radialen Winkelintervallen; Quell- und Untergrund-Intervall sind schraffiert eingezeichnet. Die horizontale Linie gibt die Untergrund-Abschätzung an.*

Abbildung 5.1 illustriert das verwendete Verfahren der Signal- und Untergrundbestimmung, links ohne $\gamma/H$-Separation rechts mit $\gamma/H$-Separation. In dem gezeigten Beispiel erhöht die $\gamma/H$-Separation also die Signifikanz der Messung. Die Anzahl der Ereignisse wird für Winkelintervalle mit festen aber unterschiedlichen Größen ermittelt und gegen den Winkelabstand zur Quell-Position aufgetragen. Die Größen der Winkelintervalle werden so bestimmt, daß sie gleichen Raumwinkeln entsprechen, und deshalb die Ereignis-Inhalte der Winkelintervalle untereinander vergleichbar sind, insbesondere mit dem der Quell-Position zugeordneten Intervall.
Für das Quell-Intervall wurde der Wert $\alpha = 0.4^o$ verwendet, dieser ergibt sich aus der AIROBICC Winkelauflösung von $\sigma_{63} = 0.32^o$ und dem Mißweisungs-Winkel $\delta \leq 0.2^o$ [46] unter der Annahme einer idealen Punktquelle gemäß folgender Formel [47]:

$$\alpha = 1.12\, \sigma_{63} \sqrt{1 + \frac{\delta^2}{\sigma_{63}^2}} \tag{5.1}$$

Die Region $1.0^o$ - $2.4^o$ wird zur Abschätzung des Untergrunds verwendet. Dieser Bereich sollte zum einen weit genug entfernt von der Quell-Region sein, so daß keine $\gamma$-Ereignisse der Quelle darin liegen, und zum anderen möglichst klein sein, damit die Untergrund-Bestimmung nicht durch andere potentielle $\gamma$-Quellen oder durch Variationen der Winkelakzeptanz gestört werden kann. An die Winkelintervalle dieses Bereichs wird eine Gerade mit verschwindender Steigung angepaßt, deren Abzissen-Abstand direkt (s.o.) die Untergrund-Erwartung angibt.
Alternativ hierzu könnten auch Polynome höheren Grades [1] angepaßt werden. Wegen der höheren systematischen Unsicherheiten dieses Verfahrens wird für diese Analyse eine Gerade zur Untergrundabschätzung verwendet.
Die Signifikanz $S$ der Messung ergibt sich dann aus der Anzahl der Ereignisse im Quell-Intervall $N_{source}$ und der Untergrund-Erwartung $N_{bg}$ gemäß der Formel:

$$S = \frac{N_{source}}{\sqrt{N_{bg}}} \tag{5.2}$$

## 5.3 Vergleich der Netzantwort zwischen M$^{te}$ Carlo-Simulation und Daten

Für die $\gamma/H$-Separation der Ereignisse des AGN-Datensatzes wurde das in Abschnitt 4.3.2 definierte 5-4-4-1 Netz benutzt. Wie dort gezeigt wurde, verbessern $r_l$ und $||\binom{x}{y}||$ die $\gamma/H$-Separation nur unerheblich. Da in diesem Abschnitt aber die prinzipielle Anwendbarkeit von Neuronalen Netzen gezeigt werden soll und um eine möglichst zu PEARL unabhängige $\gamma/H$-Separations-Methode anzuwenden, wurden $r_l$ und $||\binom{x}{y}||$ als Eingabegrößen übernommen. Abbildung 5.2, links zeigt die Netzausgabe für Daten- und M$^{te}$ Carlo-Ereignisse. Insbesondere am rechten Rand der Abbildung weicht die für den gesamten AGN-Datensatz ermittelte Kurve von der mit M$^{te}$ Carlo-Ereignissen ermittelten erheblich ab. Es kommt zu einer Diskrepanz für deutlich "hadronisch aussehende" Ereignisse[2] . Für den in der Analyse verwendeten Schnitt von $N_c$=$-0.5$ stimmen die Akzeptanzen jedoch hinreichend gut überein.

[1] Terme ungerader Potenzen verschwinden aus Symmetrie-Gründen.
[2] D.h. NETOUT nahe bei +1

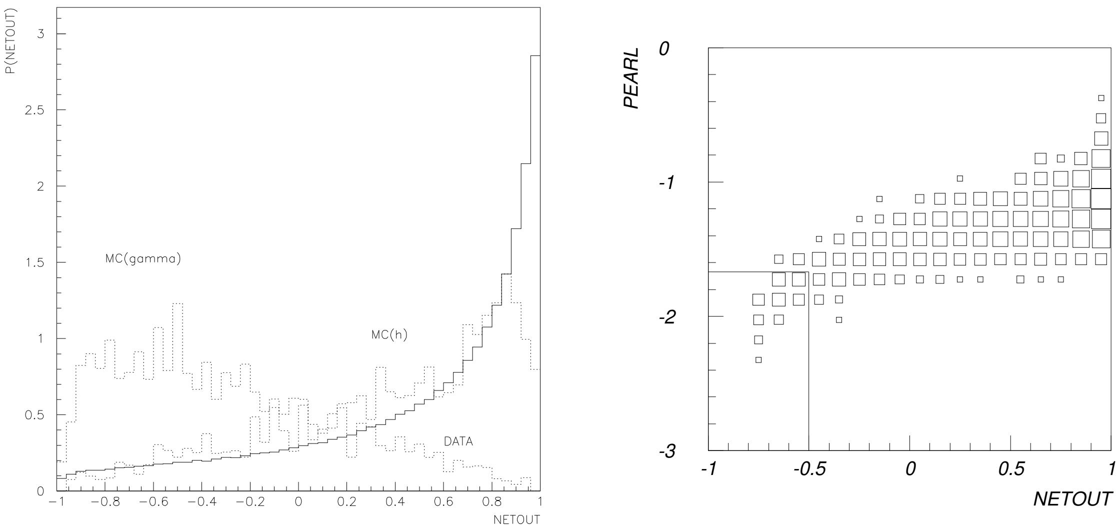



*Abbildung 5.2: Links : Vergleich der Netzausgaben für Daten- und $M^{te}$ Carlo-Ereignisse, Rechts: Vergleich des verwendeten Neuronalen Netzes und der Methode PEARL (logarithmische Auftragung).*

Ein abschließende Erklärung des Unterschieds kann an dieser Stelle nicht gegeben werden, da die Einzelverteilungen der Eingabegrößen zwischen Daten und $M^{te}$ Carlo-Simulation in ihren Mittelwerten und Breiten gut übereinstimmen und es sich deshalb um einen Effekt höherer Ordnung handeln muß. Möglich ist, daß es sich bei dem für Daten beobachteten Überschuß nahe NETOUT=1 um hochenergetische Hadronen handelt (im $M^{te}$ Carlo sind nur Primärenergien bis 300 TeV simuliert) oder um einen zenitwinkelabhängigen Effekt (im $M^{te}$ Carlo sind nur Zenitwinkel bis $15^o$ simuliert). Beide Argumente können die Diskrepanz quantitativ jedoch nicht erklären.

| $NN := (N_C = -0.5)$, $PEARL := (P_C = -1.67)$ | Ereignisse | Effizienz |
|---|---|---|
| $NN \cap PEARL$ | 41206 | 5.71% |
| $\overline{NN} \cap PEARL$ | 25741 | 3.56% |
| $NN \cap \overline{PEARL}$ | 10445 | 1.45% |
| $\overline{NN} \cap \overline{PEARL}$ | 653710 | 90.5% |

*Tabelle 5.1: Anzahl der Ereignisse des AGN-Datensatzes, die vom Neuronalen Netz und von PEARL gleich klassifiziert wurden (Zeilen 1 und 4) und die unterschiedlich klassifiziert wurden (Zeilen 2 und 3).*

In Abbildung 5.2, rechts ist die Korrelation zwischen Netzausgabe und der in Abschnitt 4.3.2 definierten Größe PEARL zu sehen, ebenfalls ermittelt für den gesamten AGN-Datensatz. Ereignisse die in einer der beiden Größen eindeutig "$\gamma$-ähnlich" oder "Hadron-ähnlich" aussehen, werden auch von der jeweils anderen Methode gleich klassifiziert.

In dem Bereich dazwischen ist die Korrelation allerdings sehr breit[3]. Wendet man den für das verwendete Netz optimalen Schnitt $N_c = -0.5$ und den für PEARL gegebenen Schnitt $P_c = -1.67$ an, ergeben sich die in Tabelle 5.1 aufgeführten Werte für die Schnitteffizienzen. Die Bedingungen $NN$ und $PEARL$ bedeuten hier, daß die Ereignisse vom Neuronalen Netz bzw. von PEARL als Photon klassifiziert wurden.
Insgesamt werden also $\approx 95\%$ der Ereignisse von beiden Verfahren gleich klassifiziert, entweder als $NN \cap PEARL$ oder als $\overline{NN} \cap \overline{PEARL}$.

## 5.4 Ergebnisse der Analyse

| Nr. | Quelle | Bezeichnung | z | $\sigma(0.41°)$ PEARL | $\sigma(0.41°)$ Neuronales Netz |
|---|---|---|---|---|---|
| 1 | 0055+300 | NGC 315 | 0.017 | -1.47 | -0.01 |
| 2 | 2201+044 | | 0.028 | -0.60 | -0.60 |
| 3 | 1101+384 | MRK 421 | 0.031 | 0.35 | -0.11 |
| 4 | 0430+052 | 3C 120 | 0.033 | 0.68 | 1.81 |
| 5 | 1652+398 | MRK 501 | 0.034 | -0.71 | 0.20 |
| 6 | 2344+514 | | 0.044 | -0.69 | 1.03 |
| 7 | 1514+004 | | 0.052 | 1.49 | 1.05 |
| 8 | 0402+379 | 4C +37.11 | 0.055 | -0.96 | 0.15 |
| 9 | 1727+502 | 1Zw 187 | 0.055 | 0.78 | -0.86 |
| 10 | 0116+319 | MS 01166+31 | 0.059 | -0.01 | -1.37 |
| 11 | 1214+381 | MS 12143+38 | 0.062 | 0.35 | 0.60 |
| 12 | 2200+420 | Bl Lac | 0.069 | 0.01 | 0.33 |

*Tabelle 5.2: Vergleich der Signifikanzen für die 12 untersuchten AGN.*

In Tabelle 5.2 sind die mit dem Neuronalen Netz ermittelten Signifikanzen den mit PEARL (Abschnitt 4.3.2) für denselben Datensatz ermittelten Signifikanzen, gegenübergestellt.
Die Verteilung der mit dem Neuronalen Netz ermittelten Signifikanzen hat einen Mittelwert von $0.19 \pm 0.25$ und einen RMS-Wert von $0.85 \pm 0.17$, wobei nur der statistische Fehler berücksichtigt wurde. Somit sind die Signifikanzen mit einer um 0 zentrierten Standardverteilung mit Varianz 1 verträglich und zeigen, wie die mit PEARL ermittelten Signifikanzen, keine positive Evidenz für TeV $\gamma$-Strahlung der 12 AGN.
Wie anhand der Tabelle und den Sternen in Abbildung 5.3 zu sehen ist, sind die Ergebnisse der beiden Methoden nur sehr schwach miteinander korreliert.

[3] Wie zu erwarten, ist die Korrelation ohne die Eingabegrößen $r_l$ und $\|\binom{x}{y}\|$ deutlich schärfer.

Daß die Resultate der beiden Methoden trotzdem, unter der Annahme der Nullhypothese (es sind keine Photonen gefunden worden), kompatibel sind, kann mit einem einfachen $M^{te}$ Carlo-Modell gezeigt werden:
Die Zahlenangaben der Tabelle 5.1 können als Schätzung für die Erwartungswerte der verschiedenen Schnitteffizienzen aufgefaßt werde. Die Verteilung der Signifikanzen aufgrund der statistischen Fluktuation der Ereignisse im Quell-Intervall [4] läßt sich dann darstellen als:

$$\begin{aligned} S_{PEARL} &= \frac{(N_1+N_2)-N_{bg,pearl}}{\sqrt{N_{bg,pearl}}} \\ S_{NN} &= \frac{(N_1+N_3)-N_{bg,nn}}{\sqrt{N_{bg,nn}}} \\ N_1 &\quad \text{poissonverteilte Zufallszahl mit Erwartung } N_{on,nn\cap pearl} \\ N_2 &\quad \text{poissonverteilte Zufallszahl mit Erwartung } N_{on,\overline{nn}\cap pearl} \\ N_3 &\quad \text{poissonverteilte Zufallszahl mit Erwartung } N_{on,nn\cap\overline{pearl}} \end{aligned} \tag{5.3}$$

Die Zufallszahl $N_1$ stellt die durch beide Methoden selektierten, $N_2$ und $N_3$ die jeweils nur durch PEARL bzw. durch das Neuronale Netze selektierten Ereignisse dar. Abbildung 5.3 zeigt

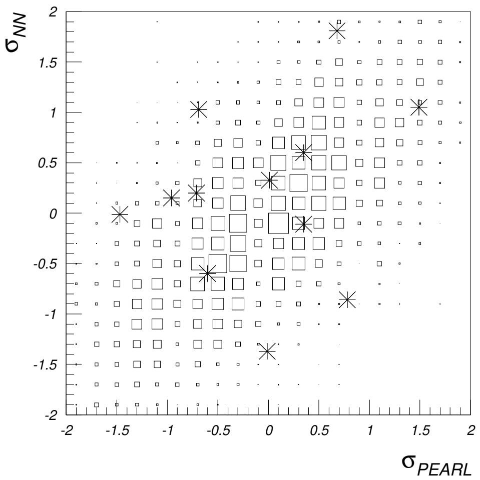



*Abbildung 5.3: Vergleich der mit dem Neuronalen Netz und der mit PEARL ermittelten Signifikanzen. Als Sterne sind die Werte der 12 AGN dargestellt, die Quadrate zeigen die Verteilung der Signifikanz-Paare des im Text beschriebenen $M^{te}$ Carlos.*

die Ergebnisse von 1000 generierten Signifikanz-Paaren. Wie die Breite der Verteilung zeigt, kann für die in Tabelle 5.1 gegebenen Schnitteffizienzen unter der Annahme der Nullhypothese keine ausgeprägte Korrelation für die Signifikanzen der beiden Methoden erwartet werden.

[4] Die Anzahl der Ereignisse im Untergrund-Intervall sei so hoch, daß deren Fluktuation zu vernachlässigen ist.

# Zusammenfassung

Für die $\gamma-$Astronomie mit Luftschauerexperimenten ist die $\gamma/H$-Separation, die Unterscheidung $\gamma$-induzierter und Hadron-induzierter Luftschauer, von entscheidender Bedeutung.
In dieser Arbeit wurde eine Reihe von Anwendungen Neuronaler Netze zur $\gamma/H$-Separation mit den HEGRA-Detektor-Matrizen vorgestellt. Für die verwendeten Feedforward-Netze wurden Bewertungs- und Interpretationsmöglichkeiten der Netzresultate vorgestellt und ein Vergleich der verschiedenen möglichen Eingabegrößen durchgeführt. Es wurde gezeigt, daß die Theorie Neuronaler Netze ein leistungsfähiges Werkzeug darstellt, um existierende analytische Verfahren zu überprüfen und nach neuen relevanten Größen und Korrelationen zu suchen, die die $\gamma/H$-Separation verbessern könnten. Die bisher bei HEGRA zur $\gamma$-Quellensuche verwendete analytische Methode PEARL wurde bestätigt. Ebenso wie mit PEARL lieferte die mit einem Neuronalen Netz zur $\gamma/H$-Separation durchgeführte Suche nach TeV $\gamma$-Strahlung von 12 extragalaktischen Quellen keine positive Evidenz für Emission oberhalb von 20 TeV.

Die wichtigsten Ergebnisse können folgendermaßen zusammengefaßt werden:

- Das Verhältnis der Intensität des atmosphärischem Čerenkovlichts zur Anzahl der geladenen Teilchen auf Detektorniveau ist das wichtigste Unterscheidungsmerkmal zwischen $\gamma$- und Hadron-induzierten Luftschauern.

- Die Trennung wird durch Informationen über die lateralen Verteilung der geladene Teilchen auf Detektorniveau verbessert.

- Informationen über die laterale Verteilung des Čerenkovlichts innerhalb der mit der HEGRA-AIROBICC-Matrix meßbaren Schauerachsen-Abstände tragen nicht zu einer weiteren Verbesserung der $\gamma/H$-Separation bei.

- Unter Berücksichtigung der Energieschwelle der HEGRA-Detektor-Matrizen ist die mit Neuronalen Netzen erreichte Güte der $\gamma/H$-Separation unabhängig davon, ob die verwendeten Eingabegrößen durch einfache arithmetische Operationen oder durch Anpassungen theoretisch motivierter Verteilungsfunktionen an die Daten gewonnen werden.

- Die Verwendung von Eingabegrößen, die das Ergebnis von Integrations- oder Summations-Operationen sind, ist vorteilhaft. Anstelle solcher Größen direkt die Einzeldetektor-Signale als Eingabegrößen zu benutzen führt, bei Berücksichtigung der Detektor-Granularität, zu einer Verschlechterung der erreichbaren $\gamma/H$-Separations-Ergebnisse.

- In dem für die $\gamma$-Quellensuche relevanten Energiebereich nahe der Energieschwelle des Detektors ist die erreichbare Qualität der $\gamma/H$-Separation durch die spärliche Detektorinformation limitiert. Bei einer deutlich erhöhten Überdeckung des Detektors mit aktiver Fläche würde sich die $\gamma/H$-Separation wesentlich verbessern.

# Anhang A

# Beispiel eines Regressionsnetzes zur Bestimmung der Höhe des Schauermaximums

### Bedeutung des Schauermaximums für die Energierekonstruktion

Die Primärenergie eines hadronischen Teilchens, das in der Atmosphäre aufschauert, teilt sich in vier Kanäle auf:
Den elektromagnetischen Kanal, der durch die Reaktion $\pi^0 \to \gamma + \gamma$ gespeist wird, den hadronischen Kanal, der aus Fragmenten der Projektile besteht, und eine Myonen- sowie Neutrino-Komponente, die aus den Zerfällen der $\pi^\pm$ stammen.
Der Anteil der insgesamt in der elektromagnetische Kaskade deponierten Energie ist proportional zur Anzahl der elektromagnetischen Teilchen im Schauermaximum $Ne_{max}$. Diese Größe kann aus der Schauerhöhe $HMAX$ und dem $Ne$ auf Detektorniveau mit Hilfe einer empirisch gewonnenen Korrelation bestimmt werden. Um von der elektromagnetischen Energie auf die Gesamtenergie zu schließen, wird eine weitere Korrektur vorgenommen, die den Energieverlust in den übrigen drei nicht beobachteten Kanälen kompensiert. Diese Korrektur ist von der Energie pro Nukleon $(E/A)$ abhängig [48] und kann deshalb (siehe Kapitel 1.4.1) ebenfalls durch $HMAX$ parametrisiert werden. Da diese Energierekonstruktion weitgehend primärteilchenunabhängig ist, kann in einem weiteren Schritt, wiederum mit der aus $HMAX$ gewonnenen Kenntnis von $(E/A)$, die Art des Primärteilchens bestimmt werden und Aussagen über die chemische Komposition der kosmischen Strahlung gewonnen werden.

### Regressionsnetze zur Bestimmung der Höhe des Schauermaximums

Die Anwendung von Feedforward-Netzen zur Regressionsanalyse kann auf verschiedene Weisen erfolgen. Die Verwendung der kontinuierlichen abhängigen Variable als Trainingsvektor haben alle Verfahren gemeinsam; in der Wahl der Eingabeaktivitäten jedoch unterscheiden sie sich:

- Basisfunktion-Netze: Die Eingabeaktivitäten werden durch Projektion auf ein System von variablen Basisfunktionen gewonnen. Die Koeffizienten und Parameter der Basisfunktionen werden durch den Lernvorgang optimiert.

- Feedforward-Netze, bei denen festen Intervallen des Argumentbereichs ein Eingabeknoten zugeordnet wird.

- Feedforward-Netze, deren Eingabeaktivitäten direkt durch den Argumentvektor bestimmt werden, und die über mindestens eine versteckte Schicht mit einer hohen Knoten-Anzahl verfügen, so daß die Knoten in der versteckten Schicht sensitiv auf bestimmte Bereiche der Funktion werden.

Hier soll der letzte Weg verfolgt werden und mit einem 3-12-12-1 Feedforward-Netz aus drei Parametern der radialen ČerenkovlichtVerteilung [48] die Schauerhöhe bestimmt werden.

$$\rho_{Cl}(r) = \rho_o \cdot \exp(-\frac{r}{r_l}) \rightarrow \rho_{Cl}(r) = \rho_o \cdot \exp(Ar^3 + Br^2 + Cr) \qquad \text{(A.1)}$$

Für jedes der 300 $\mathrm{M}^{te}$ Carlo-Ereignisse (mit Primärenergien zwischen 300 TeV und 3 PeV) werden die Werte der Koeffizienten A, B und C als Komponenten der Eingabevektoren verwand.

Da der Wertebereich sowohl positive als auch negative Koeffizienten umfaßte, wird für die versteckten Knoten die $\tan(x)$ und für den Ausgabeknoten die lineare Aktivierungsfunktion[1] eingesetzt.

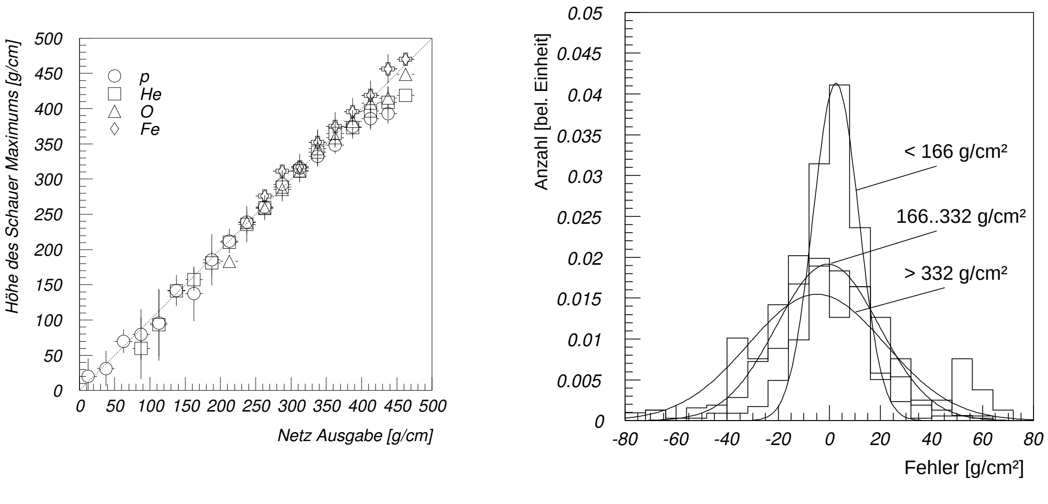



*Abbildung A.1: Links : Trainingwerte und Netzausgabe des im Text beschriebenen Regressionsnetzes für verschiedene Primärteilchen, die exakte Lösung ist gepunktet eingezeichnet.*
*Rechts: Fehler der Bestimmung für verschiedene Schauerhöhen.*

[1] siehe Abschnitt 3.2

Abbildung A.1 zeigt das Resultat für verschiedene Primärteilchen, wobei während des Trainings nicht nach Primärteilchen unterschieden wurde. In der linken Abbildung ist die Netz-Vorhersage verglichen mit den aus der Longitudinalentwicklung bestimmten Werten. Der obere Teil der Korrelation zeigt, daß die Werte für Protonen und Helium bei großer Entfernung des Schauermaximums zum Detektor systematisch überschätzt werden. Da wegen der Abhängigkeit der Eindringtiefe von $\log(E/A)$ in diesem Bereich hauptsächlich Protonen und Helium aus dem unteren Energiebereich der Trainingsmenge liegen, deckt sich dies mit der in [48] gemachten Aussage zur Energieabhängigkeit der lateralen Čerenkovlicht-Verteilung.

Der RMS-Fehler der Bestimmung liegt je nach Höhe des Schauermaximums zwischen 15 $g/cm^2$ und 25 $g/cm^2$ (Abbildung A.1, rechts) und ist in guter Übereinstimmung mit der in [48] durchgeführten Likelihood-Anpassung, die ebenfalls einen wachsenden RMS-Fehler bei steigenden Schauermaxima, bzw. sinkenden Primärenergien, aufweist.

Dieses Beispiel zeigt, daß auch für mehrdimensionale Funktionsanpassungen Neuronale Netze ein leistungsfähiges Werkzeug darstellen.

# Anhang B

# Glossar wichtiger Begriffe der Theorie Neuronaler Netze

**Bayes Theorem:** Relation aus der Wahrscheinlichkeitstheorie
$P(C_k \mid \mathbf{x}) = \frac{p(\mathbf{x}|C_k)P(C_k)}{p(\mathbf{x})}$ $P(C_k \mid \mathbf{x})$ wird A posteriori Wahrscheinlichkeit genannt. $p(\mathbf{x} \mid C_k)$ ist die bedingte Wahrscheinlichkeit, das Muster $x$ zu finden unter der Annahme der Klasse $C_k$.

**Diskriminanten-Funktion:** Eine Funktion der Eingabe-Vektoren, die zur Zuordnung der Vektoren zu einer Klasse verwendet wird.

**Entscheidungsgrenze:** Projektion des Schnittwerts der Diskriminanten Funktion in den Eingabe-Raum

**Feedforward-Netz:** Die bekannteste Topologie für Neuronale Netze. Es handelt sicht um zyklenfreie und gerichtete Netze, für die leistungsfähige Trainingsalgorithmen existieren. Die untereinander nicht verbundenen Knoten des Netzes sind in Schichten organisiert.

**Gewichtsmatrix, synaptische Matrix:** Matrix der Verbindungsgewichte zwischen den Knoten. Der Lernenvorgang von Neuronalen Netzen besteht im Auffinden der optimalen Gewichte für die gestellte Aufgabe.

**Klassifikation:** Lernproblem bei dem das Ziel ist, jeden Eingabe-Vektor einer (oder mehreren) Klasse zuzuordnen.

**Knoten,** *Node, Unit*: Elementare Einheit eines Netzes. Basierend auf den durch die synaptischen Werte gewichteten Eingängen wird mit Hilfe der Transferfunktion ein Ausgabewert ermittelt.

**Kostenfunktion, Energiefunktion:** Funktion der adaptierbaren Parameter eines Netzes, deren optimaler Wert durch das Minimum der Funktion geben wird. Die Funktion kann als Likelihood-Funktion mit zusätzlichen Termen aufgefaßt werden (siehe Regularisierung).

**Mehrschichtiges Perzeptron,** *Multi Layer Perceptron*: Feedforward-Netz mit mindestens einer versteckten Schicht.

**Netztopologie:** Anordnung und Verbindung der Knoten des Neuronalen Netzes untereinander.

**A posteriori Wahrscheinlichkeit:** $P(C_k \mid \mathbf{x})$ Wahrscheinlichkeit, daß ein Ereignis $x$ zur Klasse $C_k$ gehört, nachdem $x$ eingetroffen ist. Falls die rechte Seite des Bayes Theorems (s.o.) bekannt ist, reduziert sich die Klassifikation auf die Zuordnung von $x$ zur Klasse mit der maximimalen a posteriori Wahrscheinlichkeit (bei gleichverteilten Kosten).

**Überanpassung,** *Overfitting*: Der Fall daß das Modell zu viele freie Parameter (Gewichte) hat und die Daten der Trainingsmenge sehr viel besser anpaßt als neue Daten.

**Versteckte Schicht,** *Hidden Layer*: Die Schichten eines Netzes, die keine Verbindung zur Eingabe- oder Ausgabeebene besitzen. Netze mit mindesten einer versteckten Schicht sind in der Lage, nichtlinear-separable Problem zu lösen und jede stetige Funktion zu approximieren.

**Regression:** Ein Lernproblem, bei dem das Ziel ist, jedem Eingabe-Vektor auf einen Ausgabe-Vektor mit nicht diskreten Komponenten abzubilden.

**Regularisierung:** Eine Methode, um die Modell-Komplexität zu kontrollieren und die Generalisierungsleistung zu erhöhen. Der häufigste Fall der Regularisierung ist das *weight pruning*, die Bestrafung großer Gewichte durch einen Zusatzterm in der Kostenfunktion.

**Transferfunktion, Aktivierungs-, Basis- oder synaptische Funktion:** (*Siehe Knoten*); übliche Transferfunktionen sind $g(x) = x$, $g(x) = \Theta(x)$, $g(x) = \tan(x)$ oder die logistische Funktion $g(x) = \frac{1}{1+\exp(-x)}$

# Literaturverzeichnis


[1] D. J. Macomb R. C. Lamb. Point Sources of GeV Gamma-Rays. *Astrophys. Journal*, 488:872, 1997.

[2] T. Kifune et al. Detection of Gamma Rays of Up to 50 TeV From the Crab Nebula. *astro-ph/9710272*, 1997.

[3] T.C. Weeks et al. Observation of TeV Gamma Rays from the Crab Nebula using the Atmospheric Cherenkov Imaging Technic. *Astrophys. Journal*, 342:379–395, 1989.

[4] T. Kifune et al. Discovery of TeV Gamma Rays from SN1006: Further Evidence for the SNR Origin of Cosmic Rays. *astro-ph/9801275*, 1998.

[5] H.J. Völk. Particle Acceleration and Gamma-Ray Production in Shell Remnants. In Towards a Major Atmospheric Cherenkov Detector V. (Kruger Park Workshop), astro-ph/9711204, 1997.

[6] M. Punch et al. Detection of TeV Photons from the Active Galaxy Markarian 421. *Nature*, 352:477–478, 1992.

[7] R.J. Protheroe et al. Very High Energy Gamma Rays from Markarian 501. *Proc. 25th ICRC, Durban*, 1997.

[8] A. Wolfendale R. Murthy. Gamma-ray astronomy. Cambrigde University Press, 1986.

[9] D. MacMinn and J. Primack. Probing the Era of Galaxy Formation via TeV Gamma Ray Absorption by the Near Infrared Extragalactic Background. *Sp. Sc. Rev.*, 75, 1996.

[10] P. Sokolsky. Introduction to ultrahigh energy cosmic ray physics. Addison-Wesley Publishing Company, Inc., 1989.

[11] D. Schmele. HEGRA Himmelsdurchmusterung. Dissertation, Universität Hamburg, 1998.

[12] A. Karle et al. Search for isotropic gamma radiation in the cosmological window between 65 and 200 TeV. *Phys. Lett. B*, 347, 1995.

[13] K. Greisen. *Ann. Rev. Nucl. Sci.*, 10:63, 1960.

[14] A.M. Hillas. Differences between Gamma-Ray and Hadron Showers. *Sp. Sc. Rev.*, 75:17–30, 1996.

[15] J.W. Cronin et al. Search for discrete sources of 100 TeV gamma radiation. *Phys. Rev. D*, V45,N12:4385–4391, 1992.

[16] S. Westerhoff. Search for extragalactic $\gamma$-ray point sources with the HEGRA air shower array. Dissertation, Bergische Universität Gesamthochschule Wuppertal, 1996.

[17] M. Samorski J. Kempa. Results of fractal analysis of the Kiel extensive air Shower data. *J. Phys. G*, 24:1039–1060, 1998.

[18] R. Eckmann. Simulation von elektromagnetischen Luftschauern der Energie 30 TeV und Untersuchungen zur Richtungsbestimmung. Diplom-Arbeit, Universität Hamburg, 1991.

[19] A. Karle. Entwicklung eines neuartigen atmosphärischen Tscherenkovdetektors und Messung an hochenergetischer Kosmischer Strahlung zwischen 15 und 1000 TeV. Dissertation, Ludwig-Maximilians-Universität München, 1994.

[20] D. Dommenget. Verbesserte Rekonstruktion des Zentrums eines Luftschauers für das HEGRA-Experiment. Diplom-Arbeit, Universität Hamburg, 1994.

[21] H. Krawczynski. Verbesserte Richtungsrekonstruktion von Luftschauern für ein Szintillatorfeld und Reanalyse eines Signals von Cygnus X-3. Diplom-Arbeit, Universität Hamburg, 1994.

[22] J.N. Capdevielle. A Monte Carlo Generator for Cosmic-Ray Collisions. *J. Phys. G*, 15:909, 1989.

[23] W. R. Nelson et al. The EGS4 Code System. *SLAC Report*, 265, 1985.

[24] V. Haustein. Messungen der Ankunftszeitverteilung von Luft-Cherenkovphotonen mit einem Flash-ADC-System zur Bestimmung der Energie und der Komposition der primären kosmischen Strahlung im Energiebereich von $10^{14}$ - $10^{16}$ eV. Dissertation, Universität Hamburg, 1996.

[25] W. Pitts W.S. McCulloch. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[26] R.J. Williams D.E. Rummelhart, G.E. Hinton. Learning Represenations by Back-Propagating Errors. *Nature*, 323:533–536, 1986.

[27] Archive name: ai-faq/neural nets. `ftp://ftp.sas.com/pub/neural`. elektronisch erhältlich, 1998.

[28] R.G. Palmer J. Hertz, A. Krogh. Introduction to the Theory of Neural Computation. Addison-Wesley, 1992.

[29] C.M. Bishop. Neural Networks for Pattern Recognition. Oxford Univ. Press, New York, 1995.

[30] J. Moody. Prediction Risk and Architecture Selection for Neural Networks. *NATO ASI Series F, Springer Verlag*, 1994.

[31] A. Gruber et al. A Neural Network Architecture for the Second Level Trigger in the H1 Experiment at HERA. tech. report, 1996.

[32] S. Dennis and S. Phillips. Analysis Tools for Neural Networks. University of Queensland (Department of Computer Science) tech. report 207, 1991.

[33] T. Rögnvaldsson L. Lönnblad, C. Peterson. JETNET 3.0 - A Versatile Artificial Neural Network Package. *Computer Physics Communication*, 1994.

[34] M. Riedmiller. Advanced Supervised Learning in Multi-layer Perceptrons. *Computer Standards and Interfaces*, Special Issue on Neural Networks 5, 1994.

[35] J. Prahl et al. Search for the Crab Nebula and Mrk 421 with an Improved Sensitivity and a Lower Threshold of the HEGRA Arrays. In Proc. 24th ICRC, Rome, volume 2, page 354, 1995.

[36] T.P. Li and Y.Q. Ma. Analysis Methodes for Results in Gamma-Ray Astronomy. *Astrophys. Journal*, 272, 1983.

[37] F. Arqueros et al. Separation of gamma and hadron initiated air showers with energies between 20 and 500 TeV. *Astropart. Phys.*, 4:309–332, 1996.

[38] C. Prosch. Suche nach Supernova-Überresten als Quellen der kosmischen Strahlung oberhalb von 14 TeV. Dissertation, Ludwig-Maximilians-Universität München, 1997.

[39] J. Prahl. Suche nach $\gamma-$Punktquellen oberhalb von 15 TeV. Dissertation, Universität Hamburg, 1998. (in Vorbereitung).

[40] A. L. Coryn et al. Automated Classification of Stellar Spectra II: Two-Dimensional Classifikation with Neural Networks and Principal Components Analysis. *Mon. Not. R. Astron. Soc.*, 1997. (submitted).

[41] A. Lindner. DPG-Vortrag, München 97, 1997.

[42] M.S. Longair. High Energy Astrophysics. Cambridge University Press, 1981.

[43] J.W. Cronin et al. The Search for Discrete Astrophysical Sources of Energetic Gamma Radiation. *Annual Review Nuclear Particle Science*, 43:883–925, 1993.

[44] K. Mannheim et al. Beacons at the gamma ray horizon. *Astron. & Astrophys.*, 315:77–85, 1996.

[45] J. Prahl. A search for TeV $\gamma-$Rays from 11 Nearby Blazers with 94/95 AIROBICC Data. HEGRA Internal Note, 1997.

[46] H. Krawczynski. Search for TeV Counterparts of Gamma-Ray Bursts with the HEGRA Detector. Dissertation, Universität Hamburg, 1997.

[47] J. Prahl. Untersuchungen zur Richtungsrekonstruktion von Luftschauern and Randeffekten in einem Szintillationszählerfeld. Diplom-Arbeit, Universität Hamburg, 1992.

[48] A. Lindner. A New Method to Reconstruct the Energy and Determine the Composition of Cosmic Rays from Measurement of Cherenkov Light and Particle Densities in Extensive Air Showers. *Astropart. Phys.*, 8:235–252, 1998.

# Abbildungsverzeichnis

# Danksagung

Ich möchte Prof. Dr. Götz Heinzelmann für die interessante Aufgabenstellung danken und dafür, daß er es mir ermöglicht hat, eine Diplomarbeit im Rahmen der HEGRA-Kollaboration zu erstellen. Die Arbeitsatmosphäre war die gesamte Zeit über sehr angenehm.

Der gesamten HEGRA-Gruppe, insbesonders den Hamburger Kollegen Mathias Albers, Volker Haustein, Dieter Horns, Henrik Krawczynski, Axel Lindner, Sven Mildahn, Till Mundzeck, Jürgen Prahl, Torsten Prietz, Michael Raabe, André Röhring, Dirk Schmele möchte ich für die Unterstützung und die vielen netten Tage danken.

Meinen Eltern und FreundInnen gilt mein großer Dank für Ihre Unterstützung, Geduld und Verständnis die sie mir entgegengebracht haben.

Ich versichere, daß ich diese Arbeit nur unter Zuhilfenahme der angegeben Quellen und Hilfsmittel selbstständig angefertigt habe.

Hamburg, Dezember 1998 Kay Pätzold