

Entwicklung von Algorithmen zur automatischen Datenqualitätsanalyse des KASCADE-Grande FADC-Systems

BACHELORARBEIT
zur Erlangung des akademischen Grades
Bachelor of Science (B.Sc.)

dem Fachbereich Physik der
Universität Siegen

vorgelegt von
Felipe Gerhard

Juli 2007

Inhaltsverzeichnis

1 Die kosmische Strahlung und ihr Nachweis	2
1.1 Das Energiespektrum der kosmischen Strahlung	2
1.2 Die Beschleunigungsmechanismen	4
1.3 Ausgedehnte Luftschauer	4
1.4 Detektoren	6
2 Das KASCADE-Grande-Experiment	8
2.1 Das KASCADE-Experiment	8
2.2 Das KASCADE-Grande-Experiment	10
3 Das KASCADE-Grande FADC-DAQ-System	12
3.1 Motivation und Design	12
3.2 Komponenten	12
3.3 Das Datenausleseformat	15
4 Algorithmen zur Datenqualitätsanalyse	16
4.1 Übersicht über die verwendeten Daten	16
4.2 Grundlegende Plausibilitätsbetrachtungen	16
4.3 Die Bitfehler	17
4.4 Die Ereignisrate	18
4.5 Die Totzeit	20
4.6 Das Problem der angehängten Pulse	21
4.7 Die Schwankungen der FADC-Grundlinien	22
4.8 Das Filtern von Rausch-Ereignissen	23
5 Zusammenfassung	29

1 Die kosmische Strahlung und ihr Nachweis

1.1 Das Energiespektrum der kosmischen Strahlung

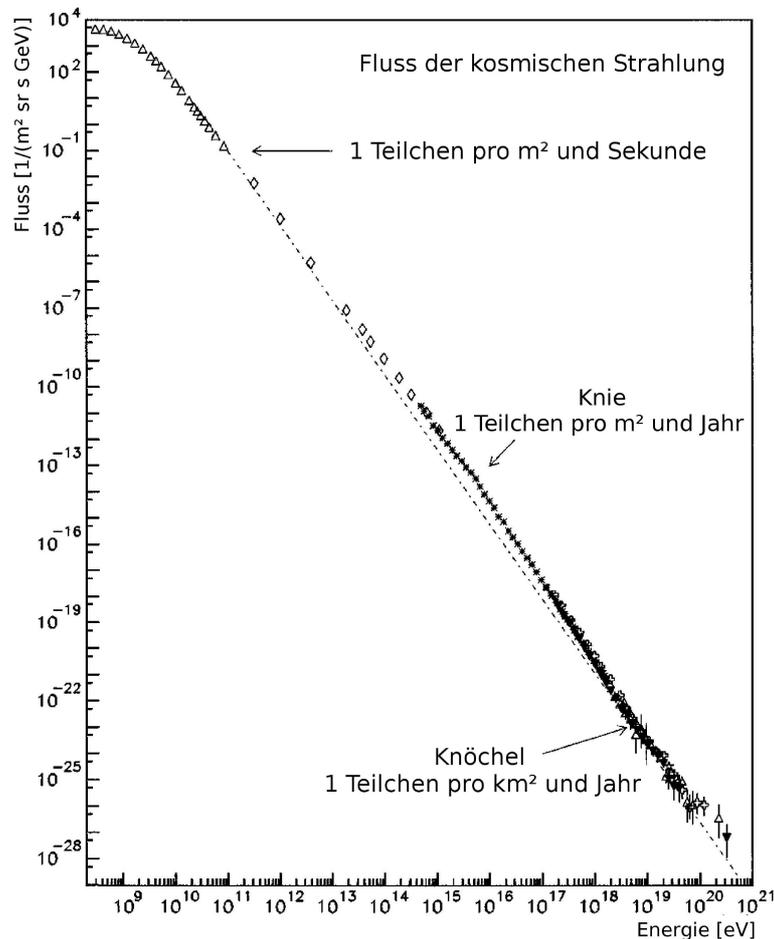


Abbildung 1: Das differentielle Energiespektrum der kosmischen Strahlung. Erkennbar sind zwei Strukturen: Das „Knie“ bei 10^{15} eV und ein Abflachen bei 10^{18} eV, der sogenannte „Knöchel“ (nach [1]).

Die Atmosphäre unserer Erde wird ständig von zum Teil hochenergetischen Teilchen aus dem Weltall getroffen. Während man zu Beginn des 20. Jahrhunderts noch davon ausging, dass es sich bei dieser Höhenstrahlung um hochenergetische γ -Strahlung handelte, wissen wir heute, dass die kosmische Strahlung hauptsächlich aus Protonen, Heliumkernen, schweren vollständig ionisierten Kernen sowie zu einem geringen Teil aus Elektronen besteht. Die Messung der Anzahl der Teilchen als Funktion ihrer atomaren Massenzahl und Energie kann Auskunft über ihre Quellen geben. Die Teilchen der kosmischen Strahlung stellen für Physiker die einzige Materie dar, die ihren Ursprung außerhalb des Sonnensystems hat und sich direkt analysieren lässt. Ihre Untersuchung ist deshalb von besonderem astrophysikalischen Interesse.

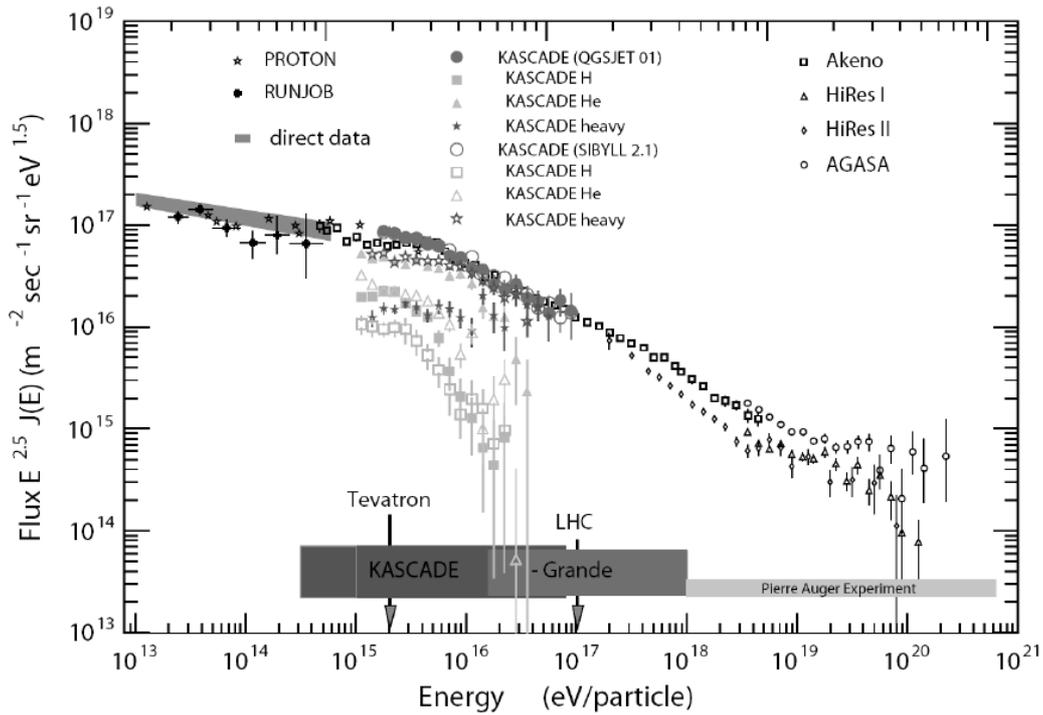


Abbildung 2: Das differentielle Energiespektrum der kosmischen Strahlung im Bereich des „Knies“. Zur übersichtlicheren Darstellung wurden die Ordinatenwerte mit $E_0^{2,5}$ multipliziert [2].

Die Intensität der kosmischen Strahlung als Funktion ihrer Energie ist im Energiespektrum dargestellt (Abb. 2). Die Strahlung überdeckt einen Energiebereich von 14 Dekaden; ihre Intensität fällt um etwa 30 Dekaden ab. Der Abfall gehorcht oberhalb von 10^9 eV einem Potenzgesetz: $I(E) \propto E^{-\gamma}$ mit $\gamma \approx 2,7$. Dabei sind Teilchen unterhalb von 10 GeV zum großen Teil solaren Ursprungs. Bei Energien über 10 GeV geht man von einem galaktischen Ursprung aus. Bei einem markanten Knick des Spektrums bei etwa $4 \cdot 10^{15}$ eV, dem sogenannten „Knie“, ändert sich der Exponent zu $\gamma \approx 3,1$. Das Magnetfeld unserer Galaxie kann nur Teilchen bis zu einem bestimmten Impuls festhalten, dem eine Energie von ca. 10^{15} eV entspricht. Ein zweiter Grund für das „Knie“ könnte sein, dass bei 10^{15} eV das Energiemaximum erreicht ist, das durch Supernova-Explosionen geliefert werden kann. Ein weiterer Knick im Energiespektrum, der „Knöchel“, scheint sich bei einer Energie von $3 \cdot 10^{18}$ eV zu befinden, bei dem γ wieder auf einen Wert von 2,7 wechselt. Dieses Abflachen wird allgemein als Beitrag einer extragalaktischen Komponente verstanden. Protonen oberhalb einer Schwellwertenergie von $6 \cdot 10^{19}$ eV wechselwirken mit den Photonen der kosmischen Hintergrundstrahlung und verlieren dadurch Energie (Greisen-Zatsepin-Kuzmin Cut-off). Die Herkunft der beobachteten hadronischen Teilchen mit Energien oberhalb von 10^{20} eV muss noch verstanden werden. Die einfachste Erklärung sehr naher Quellen ist dabei durch die Vielzahl verfügbarer Daten eher unwahrscheinlich.

1.2 Die Beschleunigungsmechanismen

Verschiedene Modelle versuchen zu beschreiben, wie Teilchen mit solch hohen Energien entstehen oder beschleunigt werden können. Man kann davon ausgehen, dass die Teilchen der kosmischen Strahlung schon bei ihrer Entstehung z.B. bei Supernova-Explosionen, in Kernen aktiver Galaxien, Doppelsternsystemen und Pulsaren beschleunigt werden. Es ist aber auch möglich, dass die Teilchen im interstellaren Raum durch ausgedehnte Magnetfelder eine zusätzliche Beschleunigung erfahren.

Fermi stellte 1949 die Theorie auf, dass geladene Teilchen an den Schockfronten sich ausdehnender magnetischer Plasmawolken reflektiert werden, wie sie beispielsweise durch Supernovae-Explosionen entstehen [3, 4]. Eine relativistische Rechnung mit plausiblen Annahmen kann eine Maximalenergie von etwa 100 TeV erklären. Durch die Wechselwirkung mit ausgedehnten, sich bewegenden Magnetwolken können die Teilchen zusätzlich beschleunigt werden. Es zeigt sich, dass dieser Beschleunigungsmechanismus quadratisch in der Wolkengeschwindigkeit ist. Weil die Teilchen einen Teil ihrer Energie zwischen zwei Wechselwirkungen wieder verlieren können, ist für diese Beschleunigung (Fermi-Mechanismus 2. Art) eine minimale Injektionsenergie notwendig. Für Teilchen, die die nötige Injektionsenergie von 200 MeV aufweisen, erhält man so eine maximale Energie von bis zu 10^{16} eV. Der Fermi-Mechanismus sagt das inverse Potenzgesetz des Energiespektrums voraus und erklärt, warum es keine hochenergetischen Elektronen in der primären kosmischen Strahlung gibt. Die maximal erreichbare Energie ist dabei proportional zur Ladung des beschleunigten Teilchens.

Rotierende Neutronensterne, die nach dem Gravitationskollaps von Sternen entstehen, haben wegen ihrer geringen Größe sehr hohe Rotationsfrequenzen. Nimmt man an, dass der magnetische Fluss während des Kollapses konstant bleibt, ergibt sich ein magnetisches Feld des Pulsars in der Größenordnung von mehreren 100 Millionen Tesla. In solch hohen Magnetfeldern können Teilchen sehr stark beschleunigt werden.

Bei Doppelsternsystemen mit einem beteiligten Pulsar oder Neutronenstern wird ständig Materie vom zweiten Stern abgesaugt und wirbelt in einer Akkretionsscheibe um den Begleitstern. Bei diesen Plasmabewegungen werden starke elektromagnetische Felder erzeugt, in denen geladene Teilchen ebenfalls beschleunigt werden können. Mit plausiblen Annahmen erhält man eine maximal erreichbare Energie von ca. $3 \cdot 10^{16}$ eV [5].

1.3 Ausgedehnte Luftschauer

Beim Eintritt der Teilchen in die Erdatmosphäre wechselwirken diese mit den Atomkernen der Luft, sodass von der primären kosmischen Strahlung praktisch nichts auf der Erdoberfläche ankommt. Stattdessen entsteht in einem kaskadenartigen Prozess, den man als ausgedehnten Luftschauer bezeichnet, eine Vielzahl von Sekundärteilchen. Dadurch wird bei entsprechend hoher Energie des einfallenden

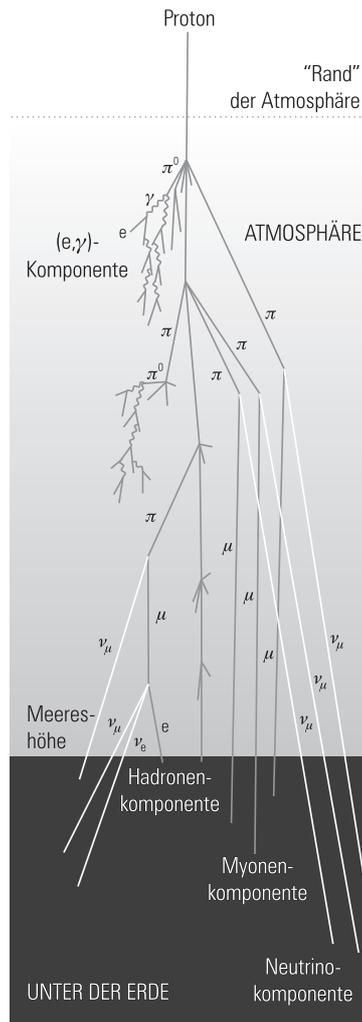


Abbildung 3: Schematische Darstellung eines ausgedehnten Luftschauers nach [5]. Das primäre Teilchen tritt mit der Atmosphäre in Wechselwirkung und erzeugt eine Kaskade von Millionen niederenergetischer Teilchen.

Teilchens diese auf Millionen niederenergetischer Teilchen verteilt.

Protonen, die mit ca. 85% den größten Anteil an primären Teilchen stellen, starten eine Hadronenkaskade. Bei den frühen Wechselwirkungen entstehen dabei Pionen. Die neutralen Pionen initiieren über ihren Zerfall eine elektromagnetische Kaskade (durch abwechselnde Paarbildung und Bremsstrahlung), während die geladenen Pionen und Kaonen in Myonen und Neutrinos zerfallen:

$$\begin{aligned}\pi^0 &\rightarrow \gamma + \gamma, \quad \pi^+ \rightarrow \mu^+ + \nu_\mu, \quad \pi^- \rightarrow \mu^- + \bar{\nu}_\mu, \\ K^+ &\rightarrow \mu^+ + \nu_\mu, \quad K^- \rightarrow \mu^- + \bar{\nu}_\mu.\end{aligned}$$

Beim Zerfall der Myonen entstehen wiederum Elektronen, Positronen und weitere Neutrinos:

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu, \quad \mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu.$$

In der Atmosphäre entstandene Myonen können trotz ihrer kurzen Lebensdauer aufgrund ihrer hohen relativistischen Energie und der daraus folgenden Zeitdilatation auf der Erde detektiert werden. Myonen stellen mit 80% aller geladenen Teilchen den größten Anteil der sekundären Teilchen auf Meereshöhe. Die restlichen 20% bilden Nukleonen (Protonen und Neutronen) sowie Elektronen, Positronen und Gammaquanten aus der elektromagnetischen Kaskade. Auf Meereshöhe findet man nur wenige Pionen und Kaonen, die in lokalen Wechselwirkungen erzeugt wurden.

Ein Schauer besitzt somit mehrere Komponenten: eine elektromagnetische (Elektronen und γ -Teilchen), eine myonische, hadronische sowie eine Neutrino-Komponente, die einzeln nachgewiesen werden können. Der größte Teil der primären Energie geht in die elektromagnetische Komponente. Die Teilchenzahl wächst zunächst mit der Schauertiefe, bis Ionisationsprozesse für geladene Teilchen bzw. Compton-Streuung und Photoeffekt den Schauer aussterben lassen. Um einen Schauer auf der Erdoberfläche detektieren zu können, muss das primäre Teilchen mindestens eine Energie von 100 TeV besitzen.

Der Zweck der Messung von ausgedehnten Luftschauern liegt zum einen in der Analyse der chemischen Zusammensetzung, zum anderen in der Suche nach den Quellen und Beschleunigern der primären kosmischen Strahlung.

1.4 Detektoren

Um möglichst viele Ereignisse mit hohen Energien zu registrieren, ist eine große Detektorfläche erforderlich. Bei den höchsten Energien trifft nur noch ein Teilchen pro Quadratkilometer und Jahrhundert auf die Erde. Eine direkte Messung der Primärstrahlung z.B. mit Ballon- oder Satellitenexperimenten wird damit für hohe Energien unpraktikabel.

In den letzten Jahrzehnten wurden erdgebundene Luftschauerexperimente mit verschiedenen experimentellen Ansätzen durchgeführt, wie z.B. AGASA (1991), ASCE (1989), GREX (1986), SUGAR (1968-1979) und Auger (geplante Fertigstellung 2007). Aufgrund der Vielzahl der in einem Luftschauer befindlichen Teilchen genügt es, ihre Verteilung stichprobenartig am Boden zu messen. Dies ist mit den gleichen Detektoren möglich, die auch in Beschleunigerexperimenten benutzt werden, also Szintillations- oder Wasser-Cherenkov-Zähler. Die Energiebestimmung des primären Teilchens ist dadurch begrenzt, dass nur ein kleiner Anteil der gesamten Sekundärteilchen detektiert werden kann. Die Zahl der erzeugten Myonen ist ein Maß für die Energie des Primärteilchens. Aufwendige Monte-Carlo-Simulationen zeigen, dass die Masse des Primärteilchens aus dem Verhältnis der Hadronenzahl zur Myonenzahl bestimmt werden kann. Die Bestimmung der Lateralverteilungen der jeweiligen Teilchensorten, die Höhe des Schauermaximums und die Produktionshöhen der Myonen tragen ebenfalls zur Rekonstruktion bei. Die Einfallrichtung lässt sich aus den Ankunftszeiten der Sekundärteilchen rekonstruieren (siehe Abb. 4).

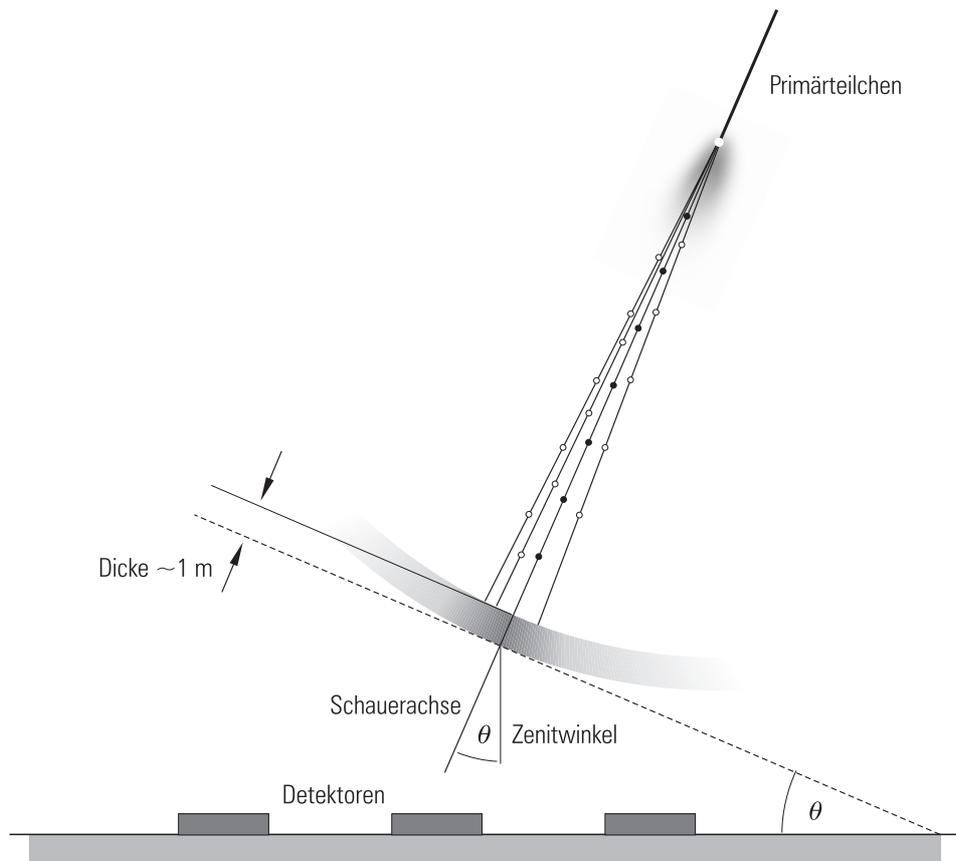


Abbildung 4: Prinzip der Schauerrekonstruktion mithilfe eines erdgebundenen Experiments nach [5]. Der ausgedehnte Teilchenschauer trifft unter dem Zenitwinkel auf die Detektoren. Aufgrund der unterschiedlichen Ankunftszeiten kann die Richtung der Schauerachse rekonstruiert werden.

Daneben gibt es optische Nachweismethoden, die Cherenkov-Strahlung sowie Fluoreszenzlicht detektieren. Cherenkovstrahlung ist stark in Bewegungsrichtung der Teilchen konzentriert und entsteht, wenn sich geladene Teilchen schneller als mit der Lichtgeschwindigkeit im Medium bewegen. Mit dieser Methode ist es u.a. möglich, auch Primärteilchen mit Energien unterhalb von 100 TeV zu detektieren, deren ausgedehnter Luftschauer die Meereshöhe nicht erreicht.

Fluoreszenzlicht entsteht durch Wechselwirkung der geladenen Teilchen mit Stickstoffmolekülen der Luft. Diese kann (in klaren Nächten) mit Spiegelteleskopen bis über 40 km Entfernung beobachtet werden. Oberhalb von 10^{17} eV lässt sich die Fluoreszenz vom diffusen Hintergrund des Sternenlichtes unterscheiden. Der Vorteil eines solchen Detektionsmechanismus liegt in der Rekonstruktion des longitudinalen Schauerprofils. Aus der gemessenen Lichtmenge kann die Schauerenergie bestimmt werden.

Das mit $3 \cdot 10^{20}$ eV bisher energiereichste Teilchen wurde vom Fly's Eye Detektor im Jahre 1991 nachgewiesen [6]. Insgesamt wurden bisher sechs Schauer mit rekonstruierten Primärenergien von über 10^{20} eV gemessen [5].

2 Das KASCADE-Grande-Experiment

Ein Experiment zur Messung von ausgedehnten Luftschauern mit einer Energie der Primärteilchen im Bereich des „Knies“ ist das KASCADE-Experiment [7] (Karlsruhe Shower Core and Array Detector) auf dem Gelände des Forschungszentrums Karlsruhe. KASCADE-Grande [8] ist eine Erweiterung des KASCADE-Experimentes, welches seit 1996 in Betrieb ist. Neben der Bestimmung des Energiespektrums soll die Massenzusammensetzung der primären Strahlung untersucht werden. Mithilfe des Experiments können außerdem verschiedene hadronische Wechselwirkungsmodelle für hohe Energien getestet werden. Weiterhin soll das Experiment dazu beitragen, den Ursprung und die Transportmechanismen der kosmischen Strahlung zu klären. KASCADE-Grande verfügt über einen Detektoraufbau mit einer effektiven Fläche von einem halben Quadratkilometer, der es erlaubt, so viele Observablen wie möglich zu messen um eine präzise Rekonstruktion des Luftschauers zu ermöglichen.

2.1 Das KASCADE-Experiment

Das KASCADE-Experiment am Forschungszentrum Karlsruhe misst das Energiespektrum der kosmischen Strahlung im PeV-Energiebereich. Die elektromagnetische Komponente eines ausgedehnten Luftschauers wird mit einem Array von Szintillationszählern, die myonische Komponente durch Szintillatoren und Spurkammern gemessen. Die hadronische Schauerkomponente wird mit einem Sampling-Kalorimeter detektiert.

Das KASCADE-Array besteht aus 252 Stationen, die mit einem durchschnittlichen Abstand von 13 m in einem Gitter angeordnet sind. 15 bzw. 16 Stationen werden für die elektronische Auslese zu einem Cluster zusammengefasst. In einer Station befinden sich je nach Lage im Array zwei oder vier e/γ -Detektoren, in diesem Fall PDP-Flüssigszintillatoren mit einer Dicke von 5 cm. Unterhalb einer Blei-/Eisenabschirmung befinden sich nur in den äußeren Stationen jeweils vier segmentierte Plastikszintillatoren mit einer Dicke von 3 cm um Myonen zu detektieren.

Die Datennahme und der Trigger agieren auf Cluster-Ebene, sodass jeder Cluster als unabhängige Komponente des Experiments wirkt. Die Front-End-Elektronik besteht pro Cluster aus acht KAFE-Modulen (KASCADE Array Frontend Electronics) um die Signale aus den Stationen zu verarbeiten. Um die Daten verschiedener Cluster zusammenzuführen, werden gemeinsame Zeitsignale benutzt, die von einem zentralen GPS-Empfänger, einem 5 MHz- sowie einem 1 Hz-Frequenzgenerator geliefert werden.

Alle Detektorsignale über einer festgelegten Diskriminatorschwelle werden zusammen mit den Zeitinformationen zwischengespeichert. Die Triggerelektronik löst ein Signal aus, sobald eine festgelegte Zahl von Stationen im Cluster Teilchen detektiert haben. Das Triggersignal wird an alle Cluster weitergegeben und Signale mit übereinstimmenden Zeiten innerhalb eines Fensters von einigen μs werden aus

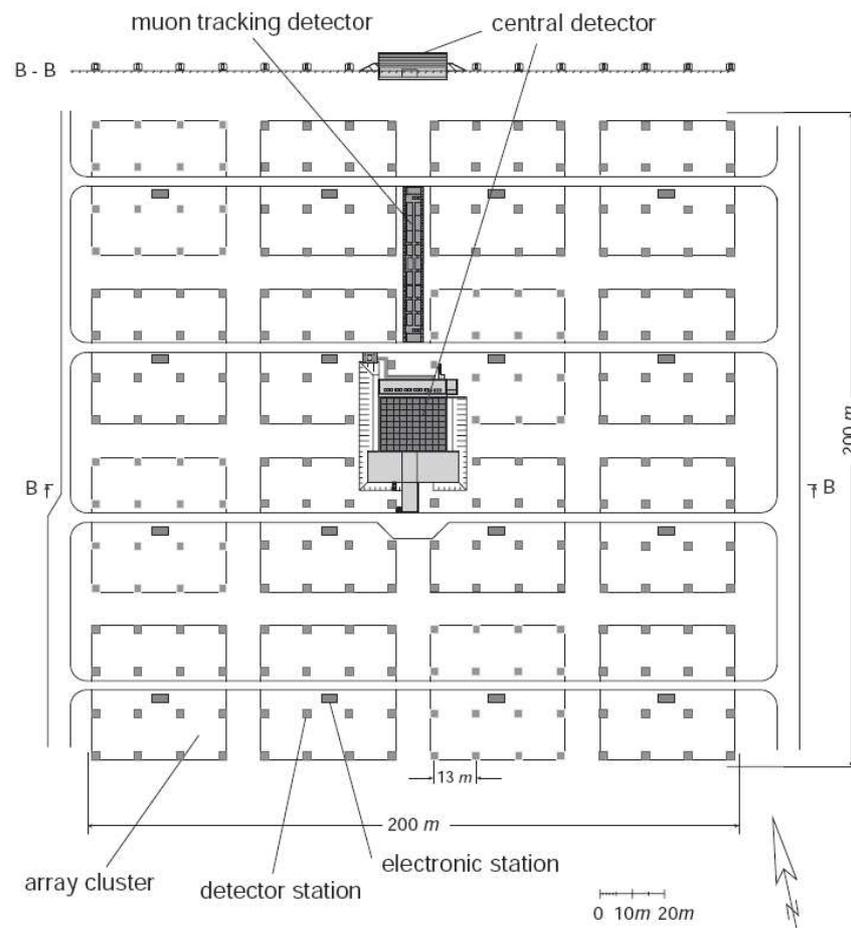


Abbildung 5: Schematische Darstellung des KASCADE-Experiments mit den Hauptkomponenten: Detektorarray, Zentraldetektor und Myonenspurdetektor auf einer Fläche von $200 \cdot 200 \text{ m}^2$ [7].

den Zwischenspeichern ausgelesen und an die zentrale Datennahme geschickt, wo sie zu einem Ereignis zusammengefasst und endgültig abgespeichert werden.

Im Gegensatz zu den Myon-Detektoren in dem KASCADE-Array, die Myonen durch ihre Energieabgabe im Szintillationsmaterial detektieren, kann die Myonspurkammer die Teilchenspur rekonstruieren, aus der schließlich die Ankunftsrichtung bestimmt werden kann. In einem 48 m langen Tunnel befinden sich drei horizontale Lagen von sogenannten Streamertubes. Dies sind gasgefüllte, unter Hochspannung stehende Detektoren, in denen Teilchen Entladungen auslösen, die mit einer geeigneten Elektronik nachgewiesen werden. Jede Lage deckt eine effektive Detektionsfläche von 128 m^2 ab. Zwei zusätzliche vertikale Lagen an den Seiten erhöhen die Detektoreffizienz für schräg einfallende Myonen.

Die Lagen liefern drei Koordinaten für jedes Teilchen, sodass es möglich ist, den Einfallswinkel zu rekonstruieren. Störende Untergrundteilchen werden dabei durch Eisenplatten abgeschirmt. Als Myonen werden solche Teilchen identifiziert, die eine gerade Spur hinterlassen und aus einer ähnlichen Richtung wie der Schauer

kommen. Die Produktionshöhe der Myonen lässt sich aus der Extrapolation der gemessenen Spur und Triangulation bestimmen. Sie ist ein Maß für die Masse des primären Teilchens.

Um den hadronischen Kern der Schauer zu untersuchen, kommt ein Zentraldetektor zum Einsatz. Hauptkomponente dieses Detektors ist ein Hadronkalorimeter. Es besteht aus ca. 11000 Kammern mit 44000 Auslesekanälen. Der Detektor besteht neben den Absorberschichten aus neun Lagen von Ionisationskammern und einer Lage aus Plastikszintillatoren als Triggerquelle und zur Messung von Ankunftszeiten. Die Signale der einzelnen Kanäle werden vor Ort zu Ereignissen zusammengefasst und können vom zentralen Datennahmesystem angefordert werden.

Unterhalb des Kalorimeters befinden sich 32 Vieldrahtproportionalkammern in zwei Schichten sowie 90 Streamertubes zum Nachweis hochenergetischer Myonen. Zur Messung von kleinen Schauern und der elektromagnetischen Schauerkomponente befinden sich oberhalb der Abschirmung des Hadronkalorimeters Szintillatoren und eine Lage von Flüssigionisationskammern.

Das KASCADE-Array, die Myonspurkammer und die Komponenten des Zentraldetektors (u.a. Hadronkalorimeter, Triggerebene, Vieldrahtproportionalkammern) sind voneinander unabhängige Komponenten und können getrennt gestartet, ausgelesen und gestoppt werden. Alle Komponenten werden mit den gleichen Zeitsignalen versorgt. Im Falle eines Triggers werden die lokal in den einzelnen Detektoren zwischengespeicherten Daten zusammen mit den dazugehörigen Zeitinformationen von dem zentralen Datennahmesystem angefordert. Die Hauptaufgabe dieses Systems ist es, die Daten zu bündeln, um sie schließlich für eine spätere Analyse abzuspeichern. Ein Trigger wird ausgelöst, wenn eine bestimmte Zahl von inneren und äußeren Arraystationen Teilchen detektiert hat. Typischerweise führt dies zu einer Triggerrate von 1-3 Hz. Pro Tag werden ca. zwei Ereignisse mit einer rekonstruierten Primärenergie von über 10^{17} eV aufgenommen. Andere Triggerbedingungen können eingestellt werden, um zum Beispiel die in den Monte-Carlo-Simulationen verwendeten Wechselwirkungsmodelle der Teilchen zu testen.

2.2 Das KASCADE-Grande-Experiment

KASCADE-Grande baut auf dem zuvor beschriebenen KASCADE-Experiment auf und vergrößert die effektive Detektorfläche auf einen halben Quadratkilometer durch zwei neue Detektoren: das Grande- und das Piccolo-Array.

Das Grande-Array ist aus 37 ehemals zum EAS-TOP-Experiment [9] gehörenden Stationen aufgebaut, die in einem hexagonalen Gitter mit einem durchschnittlichen Abstand von 137 m in der Nähe des KASCADE-Experiments angeordnet sind. Eine Station beinhaltet 16 Szintillationsdetektoren (mit einer Fläche von 10 m^2), die von Photomultipliern ausgelesen werden. Zur Berücksichtigung der variablen Teilchendichte stehen zwei Verstärkungsstufen zur Verfügung. Die vier zentralen Detektoren werden von zusätzlichen Photomultipliern ausgelesen, die mit

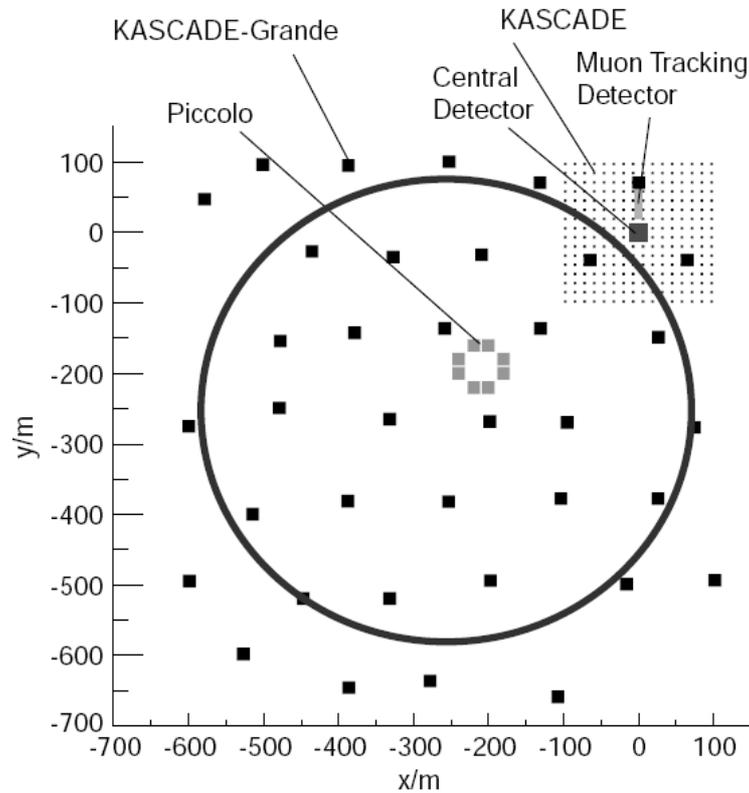


Abbildung 6: Schematische Darstellung des KASCADE-Grande-Experiments mit den 37 Grande-Stationen und acht Stationen des Piccolo-Arrays [8].

einer niedrigen Verstärkung arbeiten. Damit können selbst hohe Teilchendichten von $20 - 6000 \text{ m}^{-2}$ gemessen werden, die einer rekonstruierten Primärenergie bis zu 10^{18} eV entsprechen.

Alle 37 Stationen sind mit 700 m langen Glasfaserkabeln mit der zentralen Datennahme verbunden. In dieser werden die Stationen elektronisch zu 18 hexagonalen Zellen zusammengefasst, die jeweils aus einer zentralen und ihren sechs benachbarten Stationen bestehen.

In der Nähe des Zentrum des Grande-Arrays befindet sich ein zusätzlicher Cluster (Piccolo), der aus acht jeweils 10 m^2 großen Stationen mit je zwölf Plastikszintillatoren besteht. Der Cluster erlaubt ein schnelles Triggern aller Detektoren mit einer hundertprozentigen Effizienz für Primärteilchen oberhalb einer Energie von 10 PeV. Die Triggerrate beträgt ca. 0,5 Hz [10].

Mit der aktuellen Datennahmeelektronik beginnt nach dem Auslösen eines Hardware-Triggers die Datenauslese aller Stationen. Die analogen Signale werden verlängert, verstärkt und zur zentralen Datennahme geschickt. Das Signal wird schließlich mittels amplitudensensitiven ADCs digitalisiert und ausgewertet. Die so ermittelten Ankunftszeiten und Teilchendichten erlauben es, die Position des Schauerkerns, die Gesamtzahl geladener Teilchen, die Ankunftsrichtung und die Form der Schauerfront zu rekonstruieren.

3 Das KASCADE-Grande FADC-DAQ-System

3.1 Motivation und Design

Mit dem bisherigen Grande-Datennahmesystem tritt nach jedem Schauerereignis eine Totzeit von 0,5 ms auf, die dadurch entsteht, dass während des Auslesens der Grande-Stationen nach einem Trigger keine neuen Ereignisse aufgezeichnet werden können. Die Daten enthalten zudem keine Informationen über den zeitlichen Verlauf der Energieabgabe im Szintillator.

Zusätzlich zum bereits laufenden System wurde deshalb ein auf Flash-ADCs basiertes Datennahmesystem für das Grande-Array entwickelt und installiert. Mithilfe des FADC-Systems zur totzeitfreien Datenaufnahme und -auslese kann die gesamte Pulsform, die von den Photomultipliern erzeugt wird, aufgezeichnet werden. Dies erlaubt neben einer effizienten Rauschunterdrückung zusätzliche Analysemöglichkeiten. Die Stationen nehmen selbstgetriggerte Daten mittels Überwachung der Signalhöhe auf. Die Digitalisierung erfolgt durch kaskadierte Flash-ADCs. Anschließend werden die digitalen Daten optisch über 700 m lange Glasfaserkabel übertragen. Eine Samplingfrequenz von 250 MHz erlaubt eine Zeitauflösung in der Größenordnung von Nanosekunden [11]. Die zusätzlichen Zeitinformationen unterstützen die Eingrenzung der Primärteilchenmasse. Durch das flexible Triggersystem ist die Suche nach korrelierten Signalen in beliebigen Zeitfenstern möglich.

3.2 Komponenten

Das FADC-System setzt sich modular aus verschiedenen Komponenten zusammen.

Jede der 37 Stationen wird mit einer Digitalisierungskarte (KASCADE-Grande electromagnetic detector digitizer board, KGEMD) ausgestattet. Sie digitalisiert kontinuierlich mit jeweils vier FADCs die analogen Signale der zwei Verstärkungsstufen der Photomultiplier. Die FADCs arbeiten mit einer 12-Bit-Auflösung und durch die Kaskadierung im sog. Interleaved-Modus mit einer effektiven Samplingfrequenz von 250 MHz. Aufgrund der endlichen Fertigungstoleranzen der einzelnen Komponenten kann es Verschiebungen der Amplituden sowie unterschiedliche Verstärkungsfaktoren der vier FADCs geben, die eine Interkalibration zwischen den vier Kanälen notwendig machen.

Ein 12-Bit-FADC beherbergt 4095 Komparatoren, die das Eingangssignal gleichzeitig mit einer gleichverteilten Referenzspannung vergleichen. Die Ausgabe der Komparatoren wird in einen digitalen zwölfstelligen Binärkode übersetzt, der sogenannten Kanalnummer.

Trotz der kontinuierlichen Digitalisierung des Signals werden erst beim Überschreiten einer programmierbaren Komparator-Schwelle die Daten für $1 \mu\text{s}$ in den Zwischenspeicher geschrieben. Sollte im letzten Teil der Digitalisierungsperiode die Schwelle erneut überschritten werden, werden ohne Unterbrechung für eine weitere Mikrosekunde Daten genommen.

Aufgrund der FADC-internen Datenpuffer enthält die Pulsform bereits Daten,

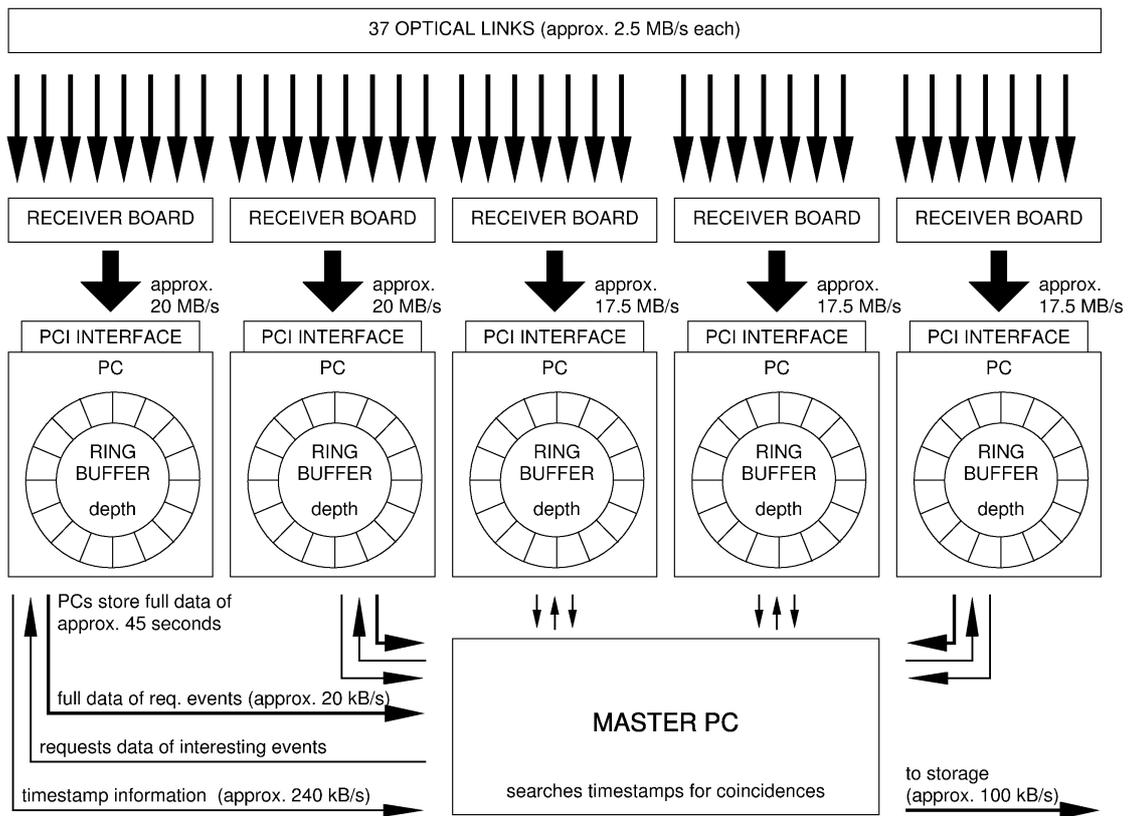


Abbildung 7: Schematischer Überblick über die FADC-System-Komponenten[12]: Fünf Empfängerkarten verarbeiten die Signale von 37 Digitalisierungskarten und geben diese über eine PCI-Schnittstelle an fünf First-Level-PCs und einen Master-PC weiter, der die Signale auf Koinzidenzen überprüft und gegebenenfalls abspeichert.

die 112 ns vor dem Überschreiten der Schwelle aufgenommen wurden. Dies erlaubt eine für jedes Paket individuelle Grundlinien-Korrektur. Die Daten zusammen mit den Zeitmarken und weiteren Meta-Informationen bilden ein Datenpaket, das via optischer Links (1300 nm Laserdiode, Reichweite bis zu 800 m, Datentransferrate von 1,25 GBaud [13]) zur zentralen DAQ-Station gesendet wird, in der es von einer Empfängerkarte weiterverarbeitet wird.

Die Zeitinformationen, die mit jedem Paket gesendet werden, enthalten den Wert eines 1 Hz-, 5 MHz- und 62,5 MHz-Zählers. Die ersten beiden werden mit jedem Zyklus des gemeinsamen KASCADE-Grande-Signals erhöht, mit dem das Digitalisierungsmodul über ein optisches Kabel versorgt wird. Der 62,5 MHz-Zähler wird von einer internen Clock auf dem Board gesteuert. Das zentrale 1 Hz-Signal setzt sowohl den 5 MHz- als auch den 62,5 MHz-Zähler zurück.

Um ein totzeitfreies System zu erhalten, müssen die Digitalisierung und Übertragung der Daten unabhängig voneinander laufen. Die Daten werden daher zunächst in einem FIFO zwischengespeichert, der ausgelesen wird, sobald die Übertragung eines neuen Pakets beginnen kann. Der FADC-Daten-FIFO kann bis zu

32 Ereignisse speichern, während ein separater FIFO für die Zeit-Informationen maximal Informationen von fünf Paketen speichern kann [12].

Das Gegenstück zu den Digitalisierungskarten in den Stationen bilden fünf Empfängerkarten (KASCADE-Grande electromagnetic detector storage board, KGEMS) in der DAQ-Station, die Daten von jeweils bis zu acht Stationen empfangen und verarbeiten. Die Aufgabe der Boards ist es, die ankommenden optischen Signale wieder in elektronische Signale umzuwandeln und die Datenpakete unverändert an das PCI-Interface weiterzugeben, das die Verbindung des FADC-Systems zu den Computern bildet.

Die PCI-Schnittstelle schreibt die Daten via direct memory access (DMA) direkt in den 896 MB großen PC-Speicher. Die erreichbare Datenrate liegt mit 85 MB/s deutlich über der erwarteten Datenrate von etwa 20 MB/s [12]. Damit kann der Ringbuffer Daten für eine Dauer von bis zu 45 Sekunden speichern.

Ein First-Level-PC verwaltet die Daten von bis zu acht Stationen. Da unkorrelierte Teilchen den größten Teil der digitalisierten Pulse bilden, ist die Hauptaufgabe der Datenverarbeitung, die Ereignisse der Stationen zu filtern, die in Koinzidenz aufgetreten sind. Da keiner der First-Level-PCs einen Überblick über die gesamten Daten hat, wird ein weiterer PC benötigt.

Dieser Master-PC empfängt nur die Zeitmarken der Pakete und untersucht sie auf Koinzidenzen. Mithilfe eines Triggerempfangsmoduls (KASCADE-Grande electromagnetic detector trigger receiver board, KGEMT) ist er zudem in der Lage, externe Triggersignale anderer Komponenten bzw. Subdetektoren zu verarbeiten. Eine Koinzidenz ist durch das gleichzeitige Auftreten von Ereignissen in den Stationen innerhalb eines Zeitintervalls definiert. Im Fall einer solchen Koinzidenz der Zeitwerte werden alle Pakete in einem frei programmierbaren Zeitfenster angefordert, die anschließend abgespeichert werden und somit zur weiteren Offlineanalyse zur Verfügung stehen.

3.3 Das Datenausleseformat

Das ursprüngliche Format für die Datenpakete ist dahingehend geändert worden, dass es nun mehrere Prüfsummen und ein abschließendes „end-of-block“-Wort enthält [12]. Die Paketstruktur sieht wie folgt aus (die Zahl in Klammern gibt die Größe des jeweiligen Eintrags in Einheiten von 16-Bit-Wörtern an):

- „magic word“ („KASG“) (2),
- Stationsnummer (1),
- Daten des „high gain“-Kanals (252),
- Daten des „low gain“-Kanals (252),
- Wert des 1 Hz-Zählers (1),
- Wert des 5 MHz-Zählers (2),
- Wert des 62,5 MHz-Zählers (2),
- Paketzähler (1),
- CRC-Prüfsumme über die ersten 512 Wörter (verwendetes Polynom $0x4405$, Startwert 0) (1),
- Paritätsbit-Prüfsumme (1),
- zusätzliche CRC-Prüfsumme über die ersten 514 Wörter (verwendetes Polynom $0x4505$, Startwert 0) (1),
- „end-of-block“-Wort (2).

Ein Paket hat damit eine Größe von 518 16-Bit-Wörtern. Solange nicht anders angegeben, sind die Algorithmen in dieser Arbeit auf der Grundlage des obigen Datenformats entwickelt worden.

4 Algorithmen zur Datenqualitätsanalyse

4.1 Übersicht über die verwendeten Daten

Es gibt drei Typen von Daten, die man zur Analyse heranziehen kann:

- intern-getriggerte Rohdaten: Durch das Setzen eines Jumpers auf dem Digitalisierungsmodul werden fortlaufend alle $400 \mu\text{s}$ Pulsformen digitalisiert. Dies führt zu einer Paketrage von exakt 2500 Hz. Im späteren Betrieb wird dieser Modus nicht aktiviert sein; er kann jedoch während der Installationsphase u.a. für eine Rauschmessung verwendet werden.
- signal-getriggerte Rohdaten: Im Standardmodus werden Daten genommen, wenn das Signal die programmierbare Schwelle überschreitet. Die Ereignisrate lässt sich so auf 2000-2500 Hz einstellen. In diesen Datensets finden sich alle relevanten Ereignisse (Schauer- und Ein-Teilchen-Ereignisse zur Energiekalibration).
- Schauerdaten: Hier werden nur Pakete berücksichtigt, die von der Triggerlogik mit einer speziellen Bedingung (z.B. Ereignisse in 7 Stationen innerhalb von 3000 ns) ausgesucht wurden.

Da es wichtig ist, schon in einem frühen Stadium Aussagen über die Datenqualität zu treffen und fehlerhafte Pakete rechtzeitig aussortieren zu können, basieren die folgenden Algorithmen hauptsächlich auf signal-getriggerten Ereignissen. An diesen Daten können alle notwendigen Analysen vollzogen werden.

Die Bezeichnungen der Datensätze beziehen sich auf die folgende Zusammenstellung:

BEZEICHNUNG	DATENTYP	DATUM	EREIGNISSE	STATIONEN
raw2102	signal-getriggert	21/02/2005	100000	1
raw1808	signal-getriggert	18/08/2005	210227	1
event1001	Schauerdaten	10/01/2006	74655	21

Tabelle 1: Übersicht über die in dieser Arbeit ausgewählten Datensätze.

4.2 Grundlegende Plausibilitätsbetrachtungen

Die einzelnen Elemente eines Datenpakets (siehe Abschnitt 3.3) bieten zahlreiche Möglichkeiten, die Integrität der Daten zu gewährleisten.

Durch Prüfen des magic-Worts am Anfang und des end-of-block-Worts am Ende eines Paketes wird sichergestellt, dass die Daten vollständig übertragen und in den Speicher geladen wurden. Daten zwischen einem end-of-block- und dem nachfolgenden magic-Wort werden ignoriert.

Zwischen der Boardnummer, die von der Digitalisierungskarte mit jedem Paket versandt wird, und den 37 Stationsnummern gibt es eine eindeutige Zuordnung, so dass die Boardnummer nur einen von maximal 37 Werten annehmen kann. Ein Paket mit ungültiger Stationsnummer wird verworfen.

Die Zeitmarken müssen innerhalb des erlaubten Bereichs liegen sowie in aufsteigender Reihenfolge im Datenstrom auftauchen. Die Werte des 5 MHz- und 62,5 MHz-Zählers sollten stark korreliert sein und bei einer Erhöhung des 1 Hz-Zählers zurückgesetzt werden. Bei der Fehleranalyse wird unterschieden, ob die ungültigen Zeitmarken einzeln oder in Kombination aufgetreten sind.

Der Paketzähler nummeriert fortlaufend die Ereignisse einer Station. Damit kann direkt herausgefunden werden, ob Pakete nach der Kodierung auf dem Digitalisierungsmodul verlorengegangen sind.

Schließlich können die Puls-Daten auf Sättigung und Overflow-Wörter untersucht werden. Bei einer Sättigung des High-Gain-Kanals sollte in der späteren Analyse dynamisch auf den Low-Gain-Kanal zurückgegriffen werden. Die FADC-Werte in beiden Kanälen sollten natürlich korrelieren.

4.3 Die Bitfehler

Die KGEMDs erstellen für die Daten zwei CRC-Checksummen, die auf verschiedenen Polynomen basieren. Zudem wird eine Paritätsbit-Prüfsumme erstellt. Durch die optische Übertragung können mit diesen Informationen die verursachten Bitfehler festgestellt werden. Mit einer systematischen Simulation verschiedener Bitfehler konnte herausgefunden werden, dass durch die Betrachtung aller drei Prüfsummen ein einzelner Bitfehler pro Paket immer erkannt wird.

Durch einen Reparaturalgorithmus ist es möglich, das fehlerhafte Bit in allen Fällen zu identifizieren [14]. Auch alle Kombinationen von zwei Bitfehlern pro Paket werden erkannt. Umfangreiche Simulationen mit mehr als zwei Bitfehlern pro Paket bestätigen die Zuverlässigkeit des Prüfverfahrens. Wenn Bitfehler erkannt, aber nicht repariert werden konnten, wird das Paket verworfen.

Die Bestimmung der Anzahl der Pakete mit fehlerhafter Checksumme erlaubt eine Abschätzung der Bitfehlerrate λ . Unter der Annahme, dass die Bitfehler poisson-verteilt mit λ auftreten und durch die Tatsache, dass ein Datenpaket 518 16-Bit-Wörter, also 8288 Bits enthält, ergibt sich für die Wahrscheinlichkeit, dass genau k Bitfehler in einem Paket auftreten:

$$P_k = \frac{(8288\lambda)^k \cdot e^{-8288\lambda}}{k!} .$$

Die Wahrscheinlichkeit für ein fehlerhaftes Paket ohne Reparaturversuch beträgt damit: $P_{\geq 1} = 1 - P_0$ und für ein fehlerhaftes Paket trotz Reparaturversuchs: $P_{\geq 2} = 1 - P_0 - P_1$.

Tabelle 2 zeigt allerdings, wie selten ein Paket aufgrund eines oder mehrerer Bitfehler wirklich verworfen werden muss. Die Berechnung der Bitfehlerrate dient

BITFEHLERRATE	FEHLERFREIE PAKETE (P_0)	FEHLERHAFTE PAKETE VOR REPARATUR ($P_{\geq 1}$)	FEHLERHAFTE PAKETE NACH REPARATUR ($P_{\geq 2}$)
$5,57 \cdot 10^{-8}$	0,9995	$4,615 \cdot 10^{-4}$	$1,065 \cdot 10^{-7}$
$3,52 \cdot 10^{-12}$	$> 0,9999$	$2,917 \cdot 10^{-8}$	$4,255 \cdot 10^{-16}$
$1,00 \cdot 10^{-12}$	$> 0,9999$	$8,288 \cdot 10^{-9}$	$3,434 \cdot 10^{-17}$

Tabelle 2: Die Tabelle zeigt, wie Bitfehlerrate und die Rate von fehlerhaften Paketen zusammenhängen. Die erste Zeile fasst die Resultate der Datennahmeperiode „raw1808“ mit relativ geringer Messdauer zusammen. Die Angaben der zweiten Zeile stammen von einer Datennahme mit wesentlich längerer Laufzeit und modifiziertem Hardware-Design [14]. Die dritte Zeile zeigt die Rate von fehlerhaften Paketen, die man erwarten würde, wenn die in den Hardware-Spezifikationen angegebene Bitfehlerrate erreicht wird. Zum Vergleich: Bei einer Ereignisrate von 2000 Hz pro Station ergeben sich für eine kontinuierliche Laufzeit ca. $2 \cdot 10^{12}$ Pakete pro Jahr.

der Überwachung der Übertragungsqualität. Beobachtet man eine signifikante Steigerung der BER („bit error rate“) um mehrere Größenordnungen, kann es nötig sein, die Empfangsdiode neu zu justieren.

Mithilfe der Prüfsummen können nur Bitfehler entdeckt werden, die während der Übertragung entstanden. Eine andere Ursache von falsch gesetzten Bits sind Hardwarefehler, z.B. Lötfehler auf den Boards. Wenn man die Samplewerte eines einzelnen FADC-Kanals histogrammiert und dabei äquidistante Lücken gleicher Breite findet (siehe Abbildung 8), so zeigt dies an, dass entweder ein Bit immer gesetzt ist oder zwei benachbarte Bits miteinander verbunden sind. Durch die genaue Diagnose, welcher Kanal und welche Bits betroffen sind, ist eine zielgenaue Reparatur möglich.

4.4 Die Ereignisrate

Für die Bestimmung der momentanen Ereignisrate r einer Station gibt es mehrere Möglichkeiten: Zum einen kann die Anzahl der Ereignisse gezählt werden, die jeweils die gleiche 1-Hz-Zeitmarke haben. Zum anderen kann die Zeitdifferenz zwischen zwei aufeinanderfolgenden Ereignissen einer Station histogrammiert werden. Da die Ereignisse unkorreliert auftreten, sollte die Zeitdifferenz exponentialverteilt sein:

$$p(\Delta t) \sim e^{-r \cdot \Delta t} .$$

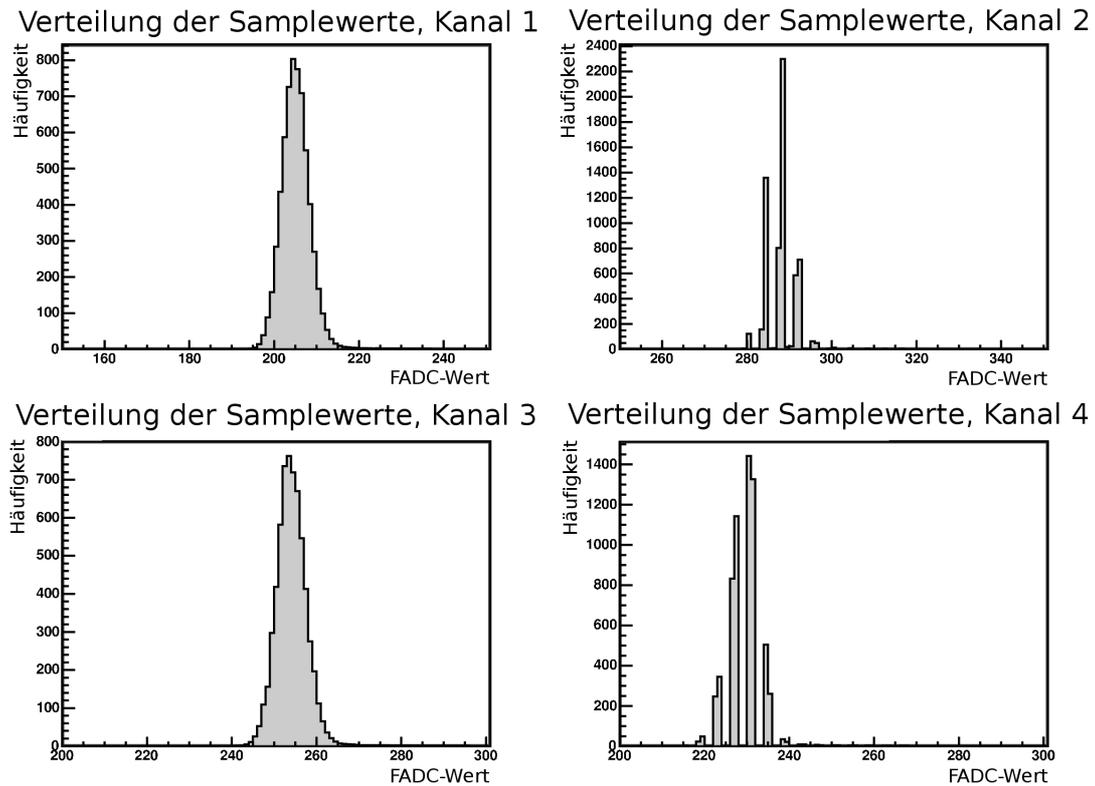


Abbildung 8: Die Histogramme zeigen die Werteverteilungen der vier FADC-Kanäle einer Station (Datensatz „raw2102“). Kanal 1 und 3 zeigen die erwartete Form, während bei den anderen beiden Kanälen Bits dauerhaft gesetzt bzw. zwei benachbarte Bits durch eine fehlerhafte Lötstelle verbunden waren.

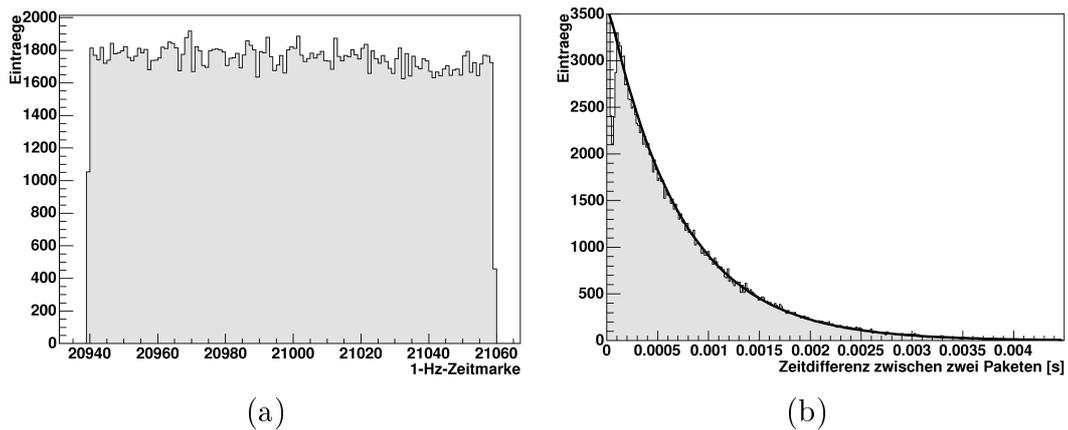


Abbildung 9: Bestimmung der Ereignisrate. (a) Die Häufigkeiten der 1 Hz-Zeitmarken wurden histogrammiert (Datensatz „raw1808“). Die durchschnittliche Rate ergibt sich hier zu 1755 Hz. (b) Die Zeitdifferenzen zwischen zwei aufeinanderfolgenden Paketen wurden aufgetragen. Die exponentielle Verteilung ist gut zu erkennen. Die mittlere Zeitdifferenz ist $5,69 \cdot 10^{-4}$ s; sie entspricht dem Kehrwert der mittleren Rate.

Der Mittelwert des Histogramms entspricht dann dem Kehrwert der Ereignisrate:

$$\overline{p(\Delta t)} = \frac{1}{r} .$$

Die Rate bekommt man auch direkt als Parameter eines exponentiellen Fits. Die drei so ermittelten mittleren Ereignisraten sollten innerhalb ihrer Unsicherheiten übereinstimmen (siehe Abbildung 9). Eine Abweichung von der exponentiellen Form müsste in jedem Fall näher untersucht werden.

Die mittlere Ereignisrate ergibt sich einfach durch:

$$\bar{r} = \frac{\text{Zahl der Ereignisse}}{\text{Länge der Datennahme}} .$$

Die mittlere Ereignisrate für Ein-Teilchen-Ereignisse beträgt ca. 2500 Hz, wie man aus Messungen mit dem bestehenden System weiß. Die Bestimmung der Ereignisrate ist wichtig um zu kontrollieren, ob die Komparatorschwelle vernünftig eingestellt ist.

4.5 Die Totzeit

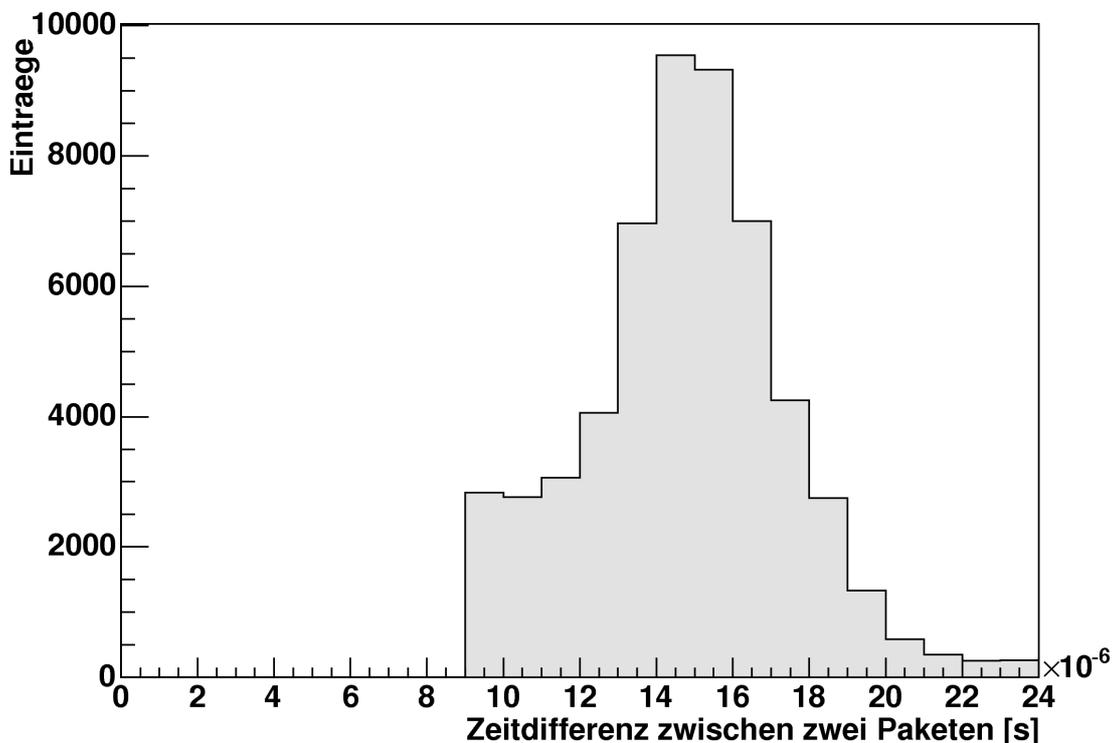


Abbildung 10: Die effektive Totzeit beträgt $9 \mu\text{s}$ (Datensatz „raw1808“): Es gibt keinen Eintrag mit einer kleineren Zeitdifferenz zwischen zwei Paketen.

Das Flash-ADC-System wurde als praktisch totzeitfreies System konzipiert. Um dies zu überprüfen, kann man den kleinsten gefundenen Zeitabstand zwischen zwei Events überwachen. Die Länge der aufgezeichneten Pulsform bestimmt den minimalen Zeitabstand; er sollte daher bereits nach kurzer Laufzeit einer Station bei ca. $1\ \mu\text{s}$ liegen.

In den anfangs analysierten Daten lag dieser minimale Zeitabstand bei etwa $9\ \mu\text{s}$, was einer ebenso großen effektiven Totzeit entspricht (siehe Abb. 10). Es handelte sich hier jedoch offenbar nicht um ein grundsätzliches Design-Problem, da bei einem Test mit vorgegebenen Pulsen keine Totzeit festgestellt werden konnte. Dabei wurde ein Pulsgenerator an den Eingang eines Digitalisierungsmoduls angeschlossen und auf kontinuierliche Pulse mit einem zeitlichen Abstand von wenigen Mikrosekunden eingestellt.

Untersucht wurden folgende Konfigurationen [15]:

- Drei Pulse in einem Abstand von $1,0\ \mu\text{s}$; Abstand zu den nächsten drei Pulsen: $1,2\ \text{ms}$,
- zwei Pulse in einem Abstand von $4,0\ \mu\text{s}$; Abstand zu den nächsten zwei Pulsen: $0,9\ \mu\text{s}$,
- zwei Pulse in einem Abstand von $980\ \text{ns}$; Abstand zu den nächsten zwei Pulsen: $0,8\ \mu\text{s}$.

Die Pulse waren groß genug, um die Komparatorschwelle zu überschreiten. Die Pakete wurden fehlerfrei übertragen. Der durch die Zeitmarken bestimmte Abstand stimmte mit dem am Pulsgenerator eingestellten überein. Unter diesen Laborbedingungen konnte also das Phänomen der Totzeit nicht reproduziert werden.

Bei neueren Daten (z.B. Datensatz „event1001“) zeigt das System keine Totzeit mehr, nachdem mehrere Veränderungen am Design der programmierbaren Logik vorgenommen wurden. Die Bestimmung des minimalen Zeitabstands zwischen zwei Paketen stellt aber weiterhin einen schnellen Test dar, ob das System praktisch totzeitfrei (nach den Spezifikationen) läuft.

4.6 Das Problem der angehängten Pulse

Sollte während der Digitalisierung eines Pulses dieser kurz vor Ende der Digitalisierungsperiode erneut über die Komparatorschwelle steigen, so wird die Aufnahme eines weiteren Pulses mit einer Länge von einer Mikrosekunde ausgelöst. Dadurch wird gewährleistet, dass auch besonders ausgedehnte Schauer komplett aufgenommen werden. Solche „verketteten“ Pakete sind im Datenstrom dadurch zu identifizieren, dass sich ihre Zeitmarken um $\lesssim 1$ Mikrosekunde unterscheiden. Eine weitere Möglichkeit zur Bestimmung der Rate, mit der solche Verkettungen auftreten, besteht in der Analyse des hinteren Teils der Pulsform selbst. Die dadurch bestimmte, erwartete Rate lässt sich nun mit der tatsächlich beobachteten vergleichen. In frühen Datennahmepetoden (Datensatz „raw1808“) gab es eine erhebliche

Diskrepanz: Erwartet wurden angehängte Pulse bei ca. 12% der Pakete, beobachtet wurden praktisch keine. In neueren Runs stimmen erwartete und ermittelte Rate in etwa überein.

Zwei Parameter entscheiden über die Anzahl der angehängten Pulse: Zum einen die Komparatorschwelle selbst, die jedoch bereits für die Festlegung der Eventrate eingestellt ist. Der zweite Parameter ist das Zeitfenster, in dem nach einer erneuten Überschreitung des Signals über die Komparatorschwelle gesucht wird. Für den Fall, dass verkettete Pulse zu sehr durch Rauschen getriggert werden, kann man dieses Fenster verkleinern.

4.7 Die Schwankungen der FADC-Grundlinien

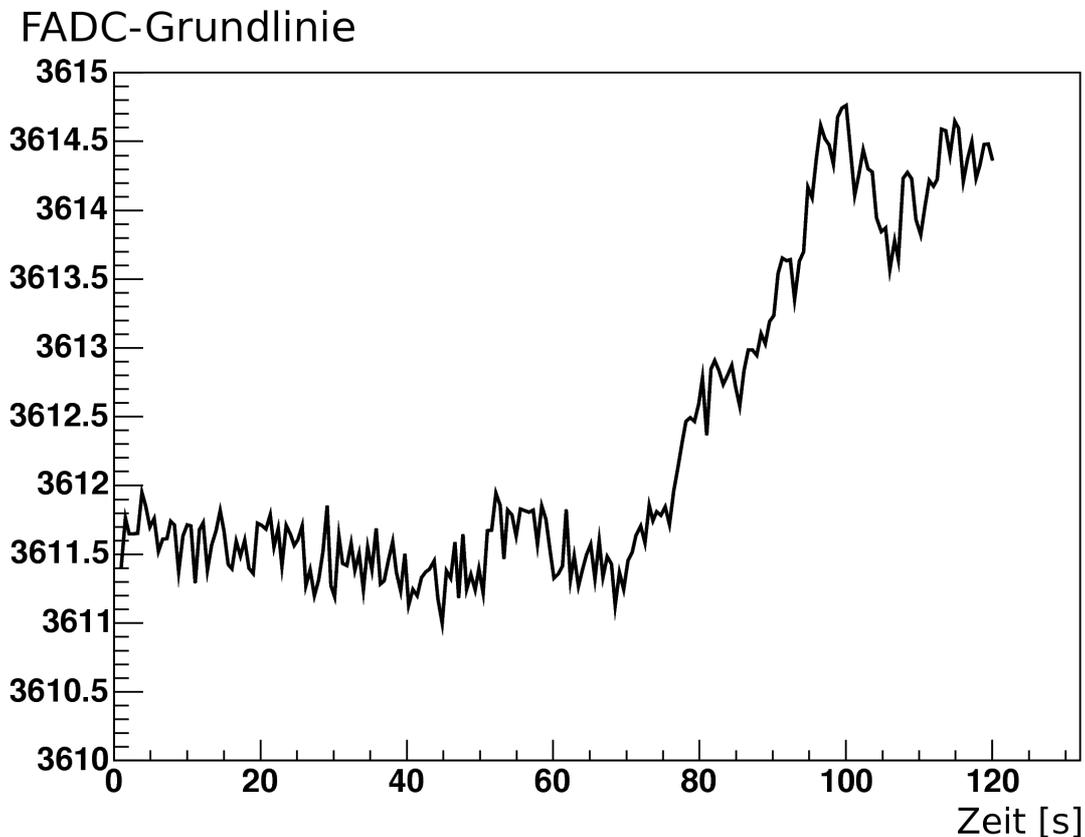


Abbildung 11: Die zeitliche Entwicklung der Grundlinie eines ausgewählten FADC-Kanals über den Zeitraum von ca. zwei Minuten (Datensatz „raw1808“). Die Grundlinie steigt in kurzer Zeit um über drei FADC-Werte an.

Jedes digitalisierte Pulssignal sitzt auf einer Grundlinie („baseline“ oder „pedestal“), die nach Invertierung des Signals typischerweise bei ca. 50 FADC-Werten liegt. Diese Grundlinie ist für jeden der vier FADC-Kanäle pro Station sowie von Station zu Station unterschiedlich. Sie kann sich durch einen Neustart ei-

nes einzelnen Boards ändern und ist auch nicht notwendigerweise während einer Datennahmeperiode konstant. Für die späteren Analysen der Pulse (z.B. für die Energiespektren) ist es notwendig, die Grundlinie für jeden aufgenommenen Puls möglichst genau zu bestimmen. Die digitalisierte Pulsform enthält bereits einige Samples, bevor die Komparatorschwelle überschritten wurde. Auf den ersten Blick scheint dieses eine praktikable Möglichkeit, die Grundlinie zu bestimmen. Da hier jedoch pro FADC-Kanal nur über maximal vier Samples gemittelt werden kann, erweist sich die Unsicherheit bei der Bestimmung als zu groß.

Bei signalgetriggerten Daten beschränkt sich der Puls in fast allen Fällen auf das erste Drittel der Digitalisierungsperiode. Die letzten zwei Drittel der Samples können daher für die Grundlinienbestimmung herangezogen werden. Nun gibt es die Möglichkeit, die Grundlinie für jedes Paket einzeln zu bestimmen oder über mehrere, zeitlich nah beieinanderliegende Pakete zu mitteln. Letztere Möglichkeit bietet den Vorteil, dass mehr Samples in die Mittelung einbezogen werden können, was zu einer geringeren statistischen Unsicherheit führt. Dieses Vorgehen ist dann gerechtfertigt, wenn die Grundlinie für diesen Zeitraum auf einem etwa konstanten Niveau verbleibt. Dass dies nicht unbedingt der Fall sein muss, zeigt Abb. 11. Man erkennt Schwankungen der Grundlinie auf einer Zeitskala von wenigen Sekunden sowie größere Verschiebungen um mehr als drei FADC-Einheiten, die sich in der Größenordnung von Minuten vollziehen. Diese kurzzeitigen Schwankungen machen eine individuelle Grundlinien-Bestimmung für jedes einzelne Paket notwendig und sinnvoll. Der statistische Fehler bei der Bestimmung entsteht dadurch, dass man über 42 Samples der Grundlinie mittelt. Dieser Fehler liegt typischerweise bei 0,7 FADC-Werten (siehe Abb. 12) und ist signifikant kleiner als derjenige, den man dadurch erhalten würde, wenn man über mehrere Pakete mittelt (da dort die Schwankungen der Grundlinien von Paket zu Paket überwiegen). Demgegenüber beträgt die Standardabweichung des Rauschens, das über der gesamten Pulsform liegt, etwa fünf FADC-Einheiten.

Es kann ebenfalls vorkommen, dass die Grundlinie abrupt um mehrere FADC-Werte springt, manchmal sogar nur bei einem einzelnen Paket (siehe Abb. 13). Diese Tatsache untermauert die Notwendigkeit einer individuellen Grundlinienbestimmung und -korrektur.

4.8 Das Filtern von Rausch-Ereignissen

Um auch Ein-Teilchen-Pulse aufnehmen zu können, ist eine niedrige Triggerschwelle unvermeidbar. Dadurch werden jedoch auch Pakete aufgenommen, deren Pulse lediglich Rauschen mit mehr oder minder stark ausgeprägten Oszillationen zeigen (siehe Abb. 14). Diese Rauschereignisse können bis zu 25% der gesamten Pakete ausmachen. Als Instrument zur Überwachung der Datenqualität kann eine gemittelte Pulsform über viele Pakete erstellt werden. Ein solch durchschnittlicher Puls ist in Abb. 15 dargestellt.

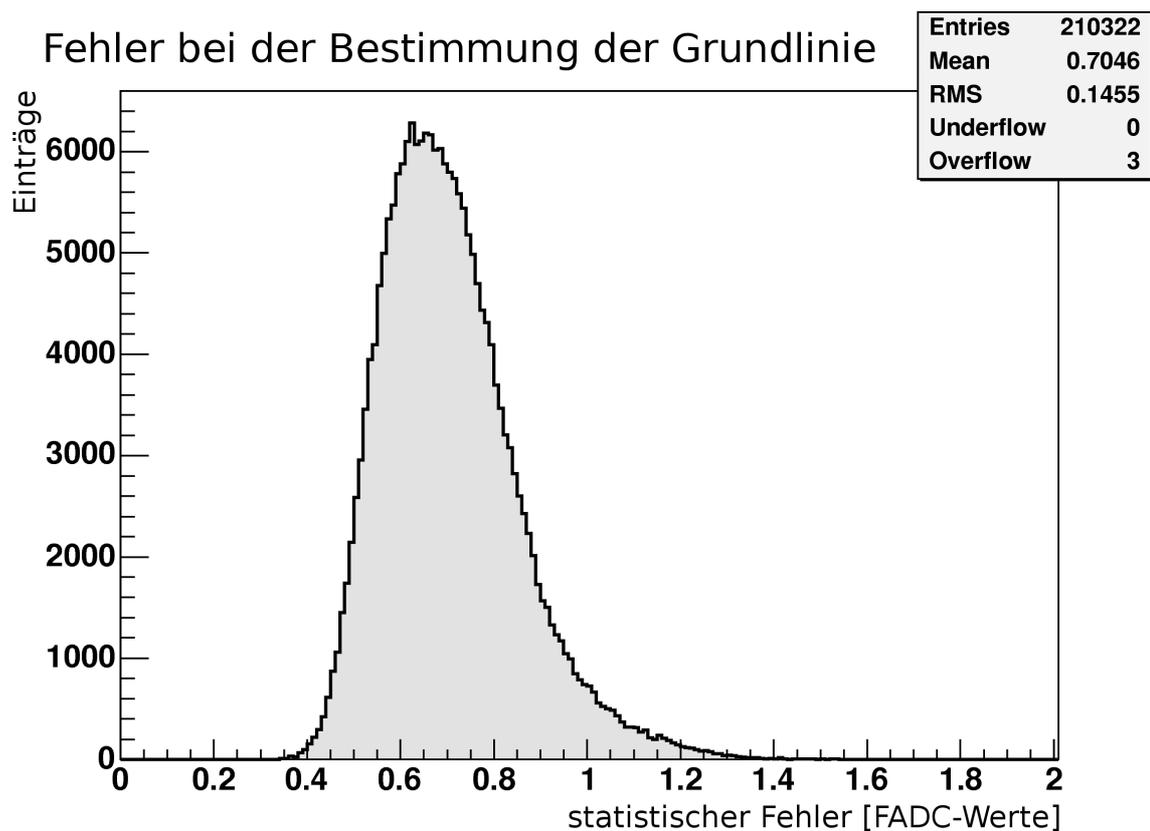


Abbildung 12: Der Fehler bei der individuellen Bestimmung der Grundlinie für jedes Paket beträgt in diesem Fall (Datensatz „raw1808“) durchschnittlich 0,7 Einheiten. Bei den meisten Paketen liegt der Fehler unter einer Einheit.

Mittelt man nur über die Rauschereignisse (die man zunächst per Hand in einer kleinen Stichprobe von einigen hundert Paketen herausgesucht hat), so heben sich die Schwingungen nicht gegenseitig auf. Das deutet darauf hin, dass sie phasengleich mit dem Start der Digitalisierungsperiode sind (siehe Abb. 15). Die Ausprägung dieser Oszillationen ist unterschiedlich für verschiedene Stationen.

Es wurde ein Algorithmus entwickelt, um diese Rauschereignisse automatisch identifizieren zu können. Dazu werden die Pulse einer diskreten Fourier-Transformation (DFT) unterzogen. Die Fourier-Spektren von normalen Schauer-Ereignissen unterscheiden sich in den Frequenzen bis ca. 45 MHz signifikant von denen einer durch Rauschen getriggerten Pulsform (siehe Abb. 16). Während bei Ein- bzw. Mehr-Teilchen-Pulsen die Frequenzanteile bis 45 MHz den dominierenden Teil des Spektrums bilden, sind die Frequenzanteile bei Rauschereignissen sehr gering. Die elektrischen Oszillationen und die Kaskadierung der FADC-Kanäle erzeugen Frequenzbeiträge um 50 MHz. Niedrigere Frequenzen sind kaum vorhanden. Integriert man das Spektrum über diesen Bereich und vergleicht es mit einem festen Schwellwert, kann der Algorithmus eine binäre Entscheidung über die Qualität des Pulses treffen.

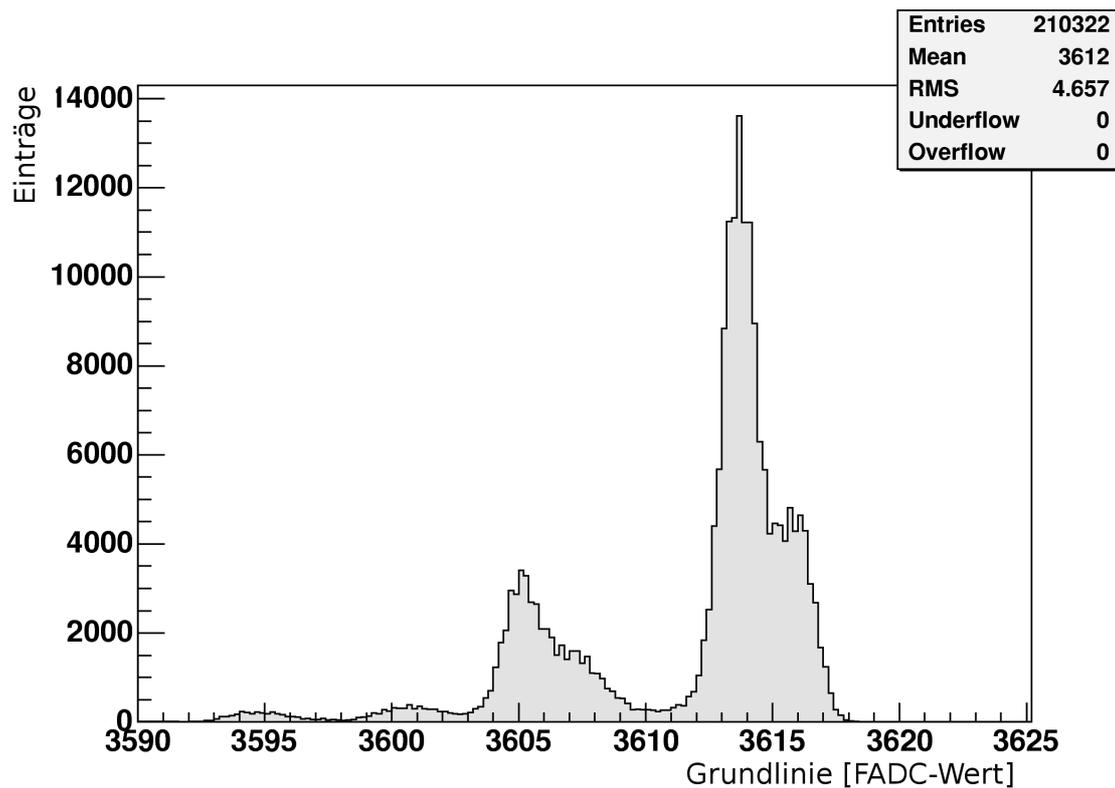


Abbildung 13: Dargestellt ist ein Histogramm über die ermittelten Grundlinien (Datensatz „raw1808“). Die Häufung bei unterschiedlichen Werten lässt darauf schließen, dass die Grundlinie abrupt um mehrere Einheiten „springt“.

Um die Effizienz dieses binären Tests zu ermitteln, wurden 100 Datenpakete manuell klassifiziert und die Ergebnisse mit denen des Algorithmus verglichen (siehe Tab. 3). Folgende Kenngrößen spielen bei der Beurteilung eines binären Tests eine Rolle:

- **Sensitivität:** Die Sensitivität gibt den Anteil der richtig positiven Ergebnisse an der Gesamtheit der positiven Ergebnissen an. Der Wert für die Sensitivität ergibt sich in diesem Beispiel zu 100%, da alle 70 „guten“ Ereignisse auch als solche erkannt wurden.
- **Spezifität:** Die Spezifität gibt die Wahrscheinlichkeit an, ein negatives Ergebnis auch als solches zu erkennen. Der Wert für die Spezifität beträgt in diesem Fall 90%, da 27 von 30 Rauschereignissen als solche identifiziert wurden.
- **Reinheit:** Die Reinheit des Datensets gibt an, wie hoch die Rate an guten Ereignissen ist. Vor dem Filtern durch den Algorithmus beträgt sie in diesem Beispiel 70%. Nach dem Filtern konnte die Reinheit auf 96% gesteigert werden (70 von 73 Ereignissen).

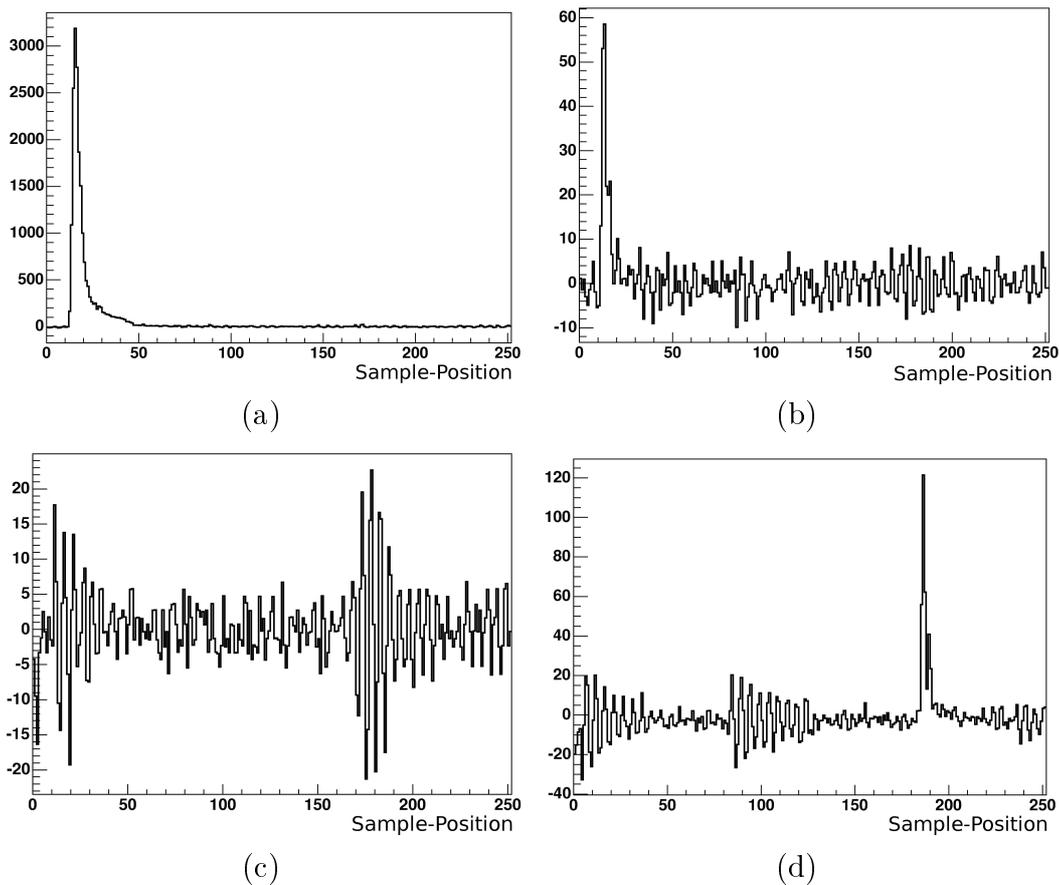


Abbildung 14: Einzelne ausgewählte Pulse (Datensatz „raw1808“). (a) Puls aus einem größeren Schauer-Ereignis. Der Puls ist kurz vor der Sättigungsgrenze. Das Rauschen der Grundlinie fällt kaum ins Gewicht. (b) Ein typischer kleinerer Puls, der sich jedoch noch klar vom Rauschen abgrenzt. (c) Die Digitalisierung wurde durch elektrische Oszillationen getriggert. Die Peakhöhe liegt in derselben Größenordnung wie die eines Einzelteilchen-Pulses. (d) Auch dieses Paket wurde durch die Oszillationen getriggert. Zufällig wurde später ein „echter“ Puls mit aufgenommen.

Bei der Festlegung des Schwellenwertes ist ein Kompromiss zwischen Sensitivität und Spezifität notwendig. Dabei ist der Wert für die Sensitivität bedeutender: Eine Sensitivität signifikant geringer als 1 bedeutet, dass physikalisch wertvolle Ereignisse als Rauschen identifiziert werden und damit eventuell für eine spätere Analyse verworfen werden.

Tests mit großen Datensets haben ergeben, dass die Rechenzeit, die für den obigen Algorithmus benötigt wird, in der Größenordnung einer einfachen Trapez-Integration liegt. Der Algorithmus ist daher schnell genug, um ihn auf alle Pulse auf dem First-Level-PC anzuwenden.

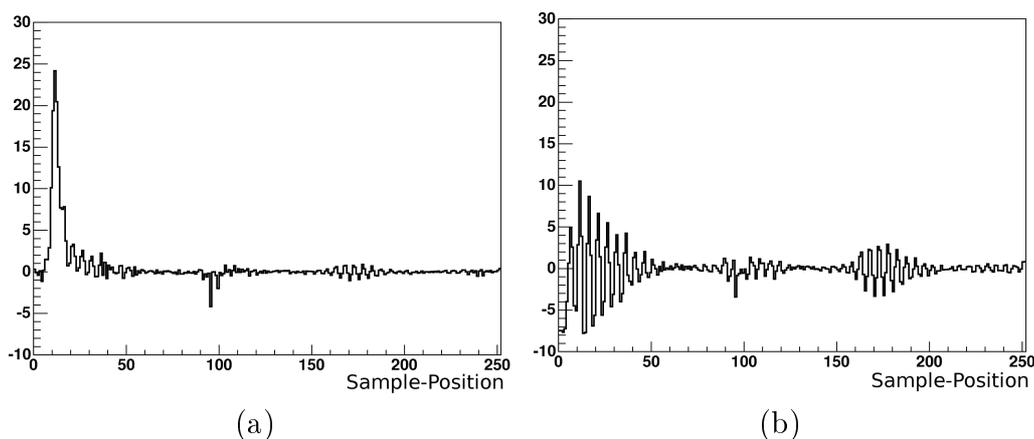


Abbildung 15: (a) Dargestellt ist eine gemittelte Pulsform (Datensatz „raw1808“). Gut sichtbar ist der Ein-Teilchen-Peak am Anfang. Danach fällt der Puls auf die Grundlinie zurück. Zu sehen ist auch der Einfluss der durch Rauschen getriggerten Pakete, die für die Schwankungen im vorderen und hinteren Drittel der Pulsform verantwortlich sind. (b) Durchschnittliche Pulsform von Rauschereignissen (Datensatz „raw1808“). Die Schwingungen heben sich bei der Mittelung über viele Pakete nicht auf. Das bedeutet, dass sie nicht zufällig verteilt auftreten, sondern stets phasengleich mit dem Start der Pulsdigitalisierung beginnen.

Pakete, die nur durch Rauschen getriggert wurden, treten unkorreliert im Datenstrom auf und tauchen deshalb nur selten in aufgezeichneten Schauerereignissen auf. Ihr Einfluss auf die Analyse von ausgedehnten Luftschauern kann daher vernachlässigt werden. Da sie aber einen nicht unerheblichen Teil der signalgetriggerten Daten ausmachen, werden Rauschereignisse bei der Erstellung von Energiespektren einbezogen und beeinflussen die Energiekalibrierung: Sie können das Spektrum verzerren beziehungsweise verbreitern (wie in Abb. 17 gezeigt). Eine Filterung kann daher die Qualität der Energiekalibration signifikant steigern, ohne dabei physikalische Ereignisse zu verwerfen.

		manuell	
		„gutes“ Ereignis	Rauschereignis
Algorithmus	„gutes“ Ereignis	70	3
	Rauschereignis	0	27

Tabelle 3: Die Entscheidungen des Algorithmus gegenüber einer manuellen Klassifizierung der Pulse nach „guten“ und verrauschten Ereignissen. Die Datengrundlage bilden 100 Pulse aus dem Datensatz „raw1808“.

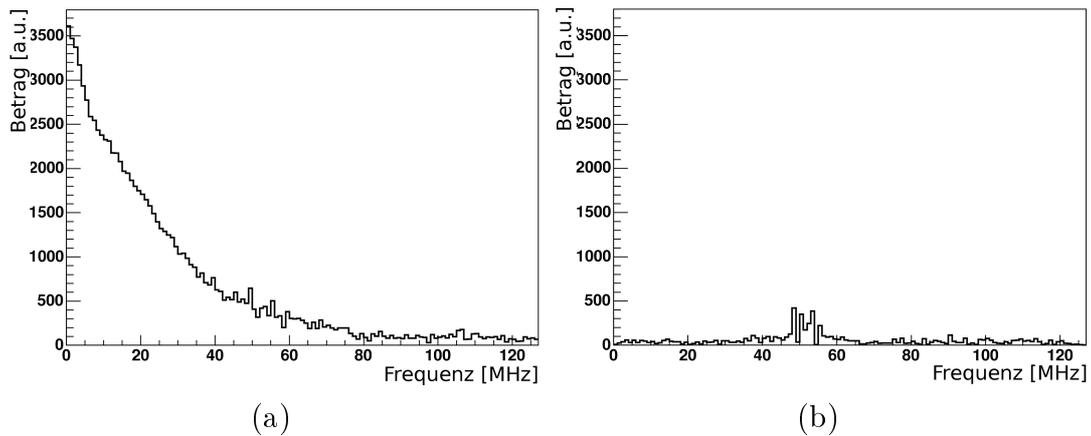


Abbildung 16: Fourier-Spektren ausgewählter Pulse (Datensatz „raw1808“). (a) Das Fourier-Spektrum eines typischen Schauerereignisses. (b) Das Fourier-Spektrum eines durch Rauschen getriggerten Pakets.

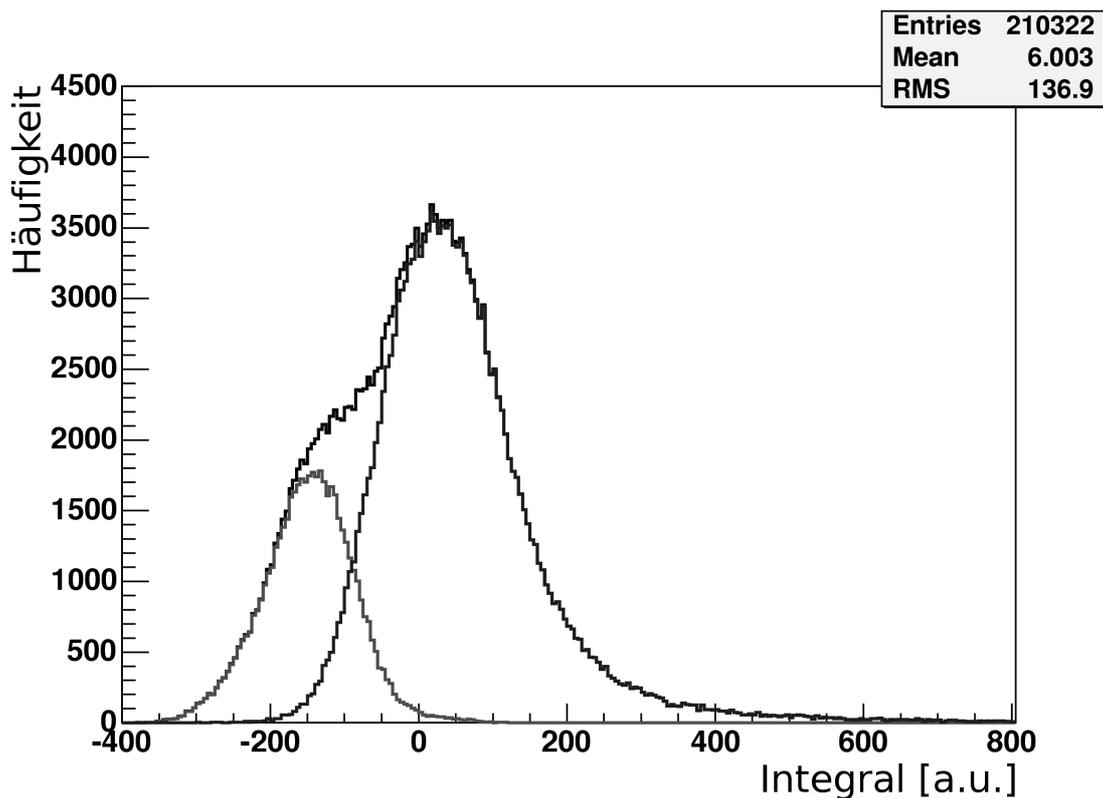


Abbildung 17: Das Integralspektrum des Datensatzes „raw1808“. Die Einhüllende zeigt das Spektrum aller Ereignisse. Der linke Peak zeigt das Spektrum der Ereignisse, die vom Algorithmus als Rauschen klassifiziert wurden. Ein entsprechendes Filtern dieser Pakete würde das rechts liegende Spektrum ergeben. Der Peak ist deutlich schmaler: Eine genauere Energiekalibration ist möglich.

5 Zusammenfassung

Das FADC-System zur Auslese der Daten des KASCADE-Grande-Experiments bietet durch die Digitalisierung der Pulsform und des flexiblen Triggerprinzips eine Vielzahl von zusätzlichen Analysemöglichkeiten. Auf verschiedenen Ebenen der Datennahme ist es möglich, die Datenqualität zu überwachen: Rudimentäre Überprüfungen des Datenpakets und die eingebauten Prüfsummen erlauben die Integrität der Daten sicherzustellen und Übertragungsfehler zu erkennen. Die individuelle Bestimmung der Grundlinie für jeden einzelnen Puls scheint notwendig und ist essentiell für die späteren Analysen (beispielsweise der Energiebestimmung). Wegen der unvermeidlich niedrigen Trigger-Schwelle werden auch Rauschereignisse aufgezeichnet, die das Energiespektrum verzerren. Es wurde ein Algorithmus zur Rauscherkennung entwickelt, der auf einer Diskreten-Fourier-Analyse basiert. Es wurde gezeigt, dass er effizient und sicher Rauschereignisse filtern kann, schnell genug ist, um für jeden Puls auf dem First-Level-PC angewandt werden zu können und die Datenqualität signifikant erhöhen kann, ohne dabei physikalische Daten irrtümlich zu verwerfen.

Die in dieser Arbeit besprochenen Tests beziehen sich meist auf die erste Ebene der Datenprozessierung, wo die signal-getriggerten Pakete individuell untersucht werden. Im nächsten Schritt, wenn die Ereignisse zu Schauern zusammengefasst werden, ergeben sich eine Vielzahl weiterer Möglichkeiten, die Datenqualität zu überprüfen, die im Rahmen dieser Arbeit nur angedeutet werden können: Dazu gehören u.a. die Wahl der optimalen Triggerbedingung, die Überprüfung von Koinzidenzen und die gleichmäßige Beteiligung der Stationen an den getriggerten Ereignissen.

Einige Fragen lassen sich nur klären, wenn Daten über längere Zeit aufgenommen werden können. Interessant ist die Veränderung der Grundlinie auf einer zeitlichen Skala von Tagen und Wochen. Auch für die Einschätzung der Bitfehlerrate sind größere Datenmengen nötig.

Diese Arbeit entstand in der Phase der Inbetriebnahme des FADC-Systems, in der es immer wieder zu Änderungen im Hardware-Design und in der Software-Umgebung kam, sodass die obigen Fragen nicht abschließend geklärt werden konnten. Endgültige quantitative Aussagen über die Datenqualität des FADC-System können daher in dieser Arbeit nicht gegeben werden.

Literatur

- [1] A. Angelis. The mysteries of cosmic rays. *Cern Courier*, 39(1), 1999.
- [2] M. Brüggemann et al. Cosmic ray studies with Kascade-GRANDE. In *ECRS*, 2006.
- [3] E. Fermi. On the origin of the cosmic radiation. *Phys. Rev.*, 75(8):1169–1174, 1949.
- [4] R. Blandford and D. Eichler. Particle acceleration at astrophysical shocks: A theory of cosmic ray origin. *Phys. Rep.* 154, 1987.
- [5] C. Grupen. *Astroparticle Physics*. Springer, 2005.
- [6] D. J. Bird et al. Evidence for correlated changes in the spectrum and composition of cosmic rays at extremely high energies. *Phys. Rev. Lett.*, 71(21):3401–3404, 1993.
- [7] T. Antoni et al. The cosmic-ray experiment KASCADE. *Nucl. Instrum. Meth. A*, vol. 513, 3, 490-510, 2003.
- [8] G. Navarra et al. KASCADE-Grande: A large acceptance, high-resolution cosmic-ray detector up to 10^{18} eV. *Nucl. Instrum. Meth.*, vol. A518, 207-209, 2004.
- [9] M. Aglietta et al. UHE cosmic ray event reconstruction by the electromagnetic detector of EAS-TOP. *Nucl. Instrum. Meth. A*, vol. 336, pp. 310-321, 1993.
- [10] F. Di Pierro et al. Status of the KASCADE-Grande Experiment. In *CRIS 2006: Ultra-High Energy Cosmic Rays: Status and Perspectives*, Catania, Italien, 2006.
- [11] V. Andrei. Reconstruction algorithms for the FADC DAQ system of the KASCADE-Grande experiment. Masterarbeit, Universität Siegen, 2004.
- [12] S. Over. Development and commissioning of data acquisition systems for the KASCADE-Grande experiment. Diplomarbeit, Universität Siegen, 2004.
- [13] W. Walkowiak et al. A FADC-based Data Acquisition System for the KASCADE-Grande experiment. *IEEE Nuclear Science Symposium*, 2004.
- [14] S. Over. Persönliche Mitteilung, August 2005.
- [15] Y. Kolotaev. Persönliche Mitteilung, August 2005.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate und Ergebnisse Anderer kenntlich gemacht habe.

Felipe Gerhard