

Dataacquisition und Trigger

Vorlesungsnotizen im Rahmen der Veranstaltung
“Experimentelle Methoden der Teilchenphysik”

U. Straumann, Physik - Institut Universität Zürich

29. Dezember 2006

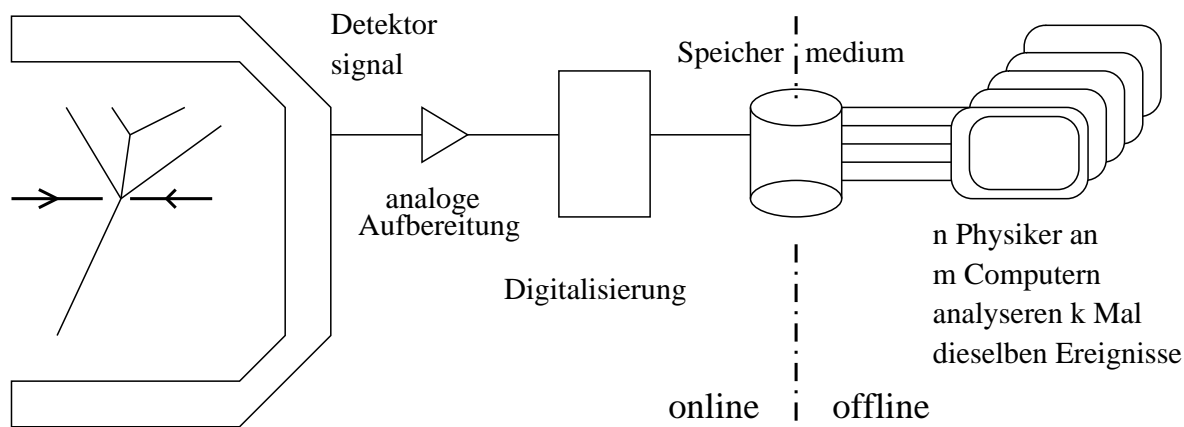
Inhaltsverzeichnis

1	Uebersicht über die Datennahme	2
2	Analoge Aufbereitung der Daten	3
3	Digitalisierung der Messwerte	7
4	Einfache Triggerkonzepte	8
5	Pipelines	9
6	Occupancy, Pileup, Spill over, Cross talk. Random coincidences	10
7	Totzeit	11
8	Höhere Triggerkonzepte und Ereignisselektion	12
9	Ergänzung Geant	13

1 Uebersicht über die Datennahme

Experimente in der Kern- und Elementarteilchenphysik studieren grundsätzlich immer **Ereignisse**. Seien es Reaktionen zwischen in einem Beschleuniger kontrolliert zur Kollision gebrachten Teilchen, spontane Zerfälle von nuklearer Materie oder Reaktionen von Teilchen aus dem Kosmos mit Detektoren auf der Erde, immer geht es darum diese Ereignisse zu zählen und die Reaktionsprodukte zu vermessen. Im Anfang der Teilchenphysik steht das Zählen. (Die begehbaren Räume in denen die Datennahme-Elektronik untergebracht ist, werden immer noch als “counting house” bezeichnet, ob- schon natürlich längst viel komplexere Aufgaben wahrgenommen werden). Entsprechend der quan- tenmechanischen Natur der Prozesse, werden die so gewonnenen **Zählraten** (Anzahl Reaktionen pro Zeiteinheit) statistisch analysiert.

Wie Sie in dieser Veranstaltung gelernt haben, gibt es eine Menge verschiedener Detektoren, die Er- eignisse zählen, und verschiedene Eigenschaften des Endzustandes der Reaktion messen können. Die Aufgabe der Datennahme besteht grundsätzlich darin, die Information der Dektoren für jedes inter- essante Ereignis in digitaler Form zu speichern, sodass später detaillierte Analysen vorgenommen werden können.



Diese Aufgabe scheint auf den ersten Blick rein technischer Natur zu sein; man sollte sie am besten an Elektronikeneure und Informatiker delegieren.

Die Zahl der Messkanäle können aber durchaus 10^7 und mehr betragen; die Reaktionsraten liegen im MHz Bereich (HERA 10 MHz, PSI 50 MHz, LHC 40 MHz, Tevatron 2.5 MHz). Wollen wir nur eine einziges Byte pro Messkanal aufschreiben, müssten wir pro Messtag die unvorstellbare Datenmenge von 3×10^{19} Bytes = 3 Mio TB speichern. Wollte ein Physiker jede Sekunde eines dieser Bytes anschauen, bräuchte er dafür 100 mal solange, wie das Universum alt ist...

Die **Datenmenge** muss also **online reduziert** werden. Um die richtige Information auszuwählen, muss eine grosse Menge von Wissen eingebracht werden: Relevante Eigenschaften von Detektoren, erforderliche und erzielbare Messgenauigkeiten, physikalische Eigenschaften der Teilchen und der Prozesse. Schliesslich müssen forschungsstrategische Entscheide fallen, welche Prozesse und damit welche physikalischen Fragestellungen studiert werden sollen, und welche ignoriert werden können. Planung und Betrieb der Datennahme besteht also aus physikalischen, technischen und politischen Vorgängen.

Die Datennahme eines Experimentes besteht im wesentlichen aus den folgenden Schritten, an die sich gleichzeitig die Struktur dieser Vorlesung anlehnt:

1. **Analoge Aufbereitung der Detektorsignale.** Die dazugehörige Elektronik muss optimal an den Detektor angepasst werden, um Messfehler aufgrund von Rauschen etc. kleinzuhalten. Die Elektronik wird ausserdem verschieden aussehen, je nachdem ob zum Beispiel eine Ladungsmenge, der Zeitpunkt des Signales oder doch nur eine ja/nein Antwort verlangt wird.
2. **Uebertragung der Signale.** Hier spielen Informationsverlust wegen elektrischer Kabeleigenschaften eine Rolle, aber auch das durch die Uebertragungskabel eingebrachte "tote Material" muss minimiert werden.
3. **Digitalisierung der Messwerte.** Hier wird durch Quantisierung in Amplitude und Zeit ein wesentlicher Informationsverlust erzeugt, der entsprechend unter Kontrolle gehalten werden muss.
4. **Selektion von Detektorbereichen,** die untersucht werden sollen. (zum Beispiel Clusters oder Spur).
5. **Selektion von Ereignissen mit bestimmten Signaturen,** die aufgrund physikalischer Quantitäten der Ereignisses gespeichert werden sollen (zum Beispiel gibt es ein Muon mit grossem transversalem Impuls? Gibt es einen zweiten Zerfallsvertex?, ist die Gesamtenergie grösser als ein bestimmter Schwellwert?)
6. **Selektion von Ereignissen, bestimmter physikalischer Prozesse**

Bei jedem dieser Schritte können verschiedenste Arten von Fehler passieren, Messwerte werden verfälscht oder die besonders interessanten Ereignisse werden gar nicht selektioniert.

2 Analoge Aufbereitung der Daten

Detektoren erzeugen elektrische Signale. Die Signale sind in der Regel sehr klein, sodass sie elektronisch verstärkt werden müssen. Die Optimierung dieser Verstärkung muss verschieden ausfallen, je nachdem welche Grössen im Zentrum des Interesses stehen. In allen Fällen wird das Signal durch elektronisches Rauschen (noise) verfälscht, was zu einem Messfehler führt, den man genau kennen und möglichst klein halten muss.

Die folgenden drei Arten von elektrischem Rauschen sind von praktischer Relevanz.

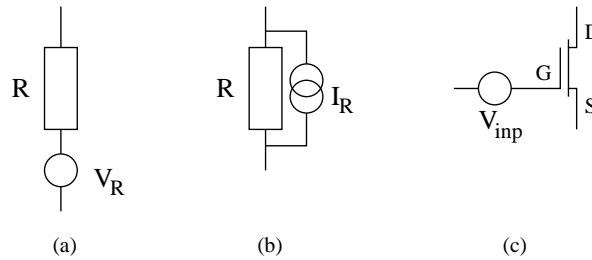
1. **Thermisches Rauschen (Johnson, 1928):** Entsteht in allen Elementen, in denen eine bewegte Ladung eine Spannung erzeugt, also insbesondere in Widerständen und Ladungskanälen von Feldeffekttransistoren (aber nicht in Kondensatoren). Die dadurch entstehenden Spannungs- und Stromschwankungen führen zu einer Leistung P , deren Erwartungswert $\langle P \rangle$ pro Frequenzintervall

$$\frac{d \langle P \rangle}{df} = 4 \cdot k T \quad (1)$$

unabhängig von der Frequenz selbst ist (f =Frequenz, k =Boltzmannkonstante, T =Temperatur). Es handelt sich also um weisses Rauschen. mit Hilfe von $P = RI^2 = V^2/R$ kann auch die Varianz der Spannung und des Stromes an einem Widerstand angegeben werden:

$$\frac{d \langle V_R^2 \rangle}{df} = 4 k T R \quad \text{oder} \quad \frac{d \langle I_R^2 \rangle}{df} = 4 k T \frac{1}{R} \quad (2)$$

Das Rauschen muss entsprechend (a) oder (b) im Schaltschema berücksichtigt werden:



Am meisten trägt immer die erste Verstärkerstufe zum thermischen Rauschen bei, da hier das Signal noch sehr klein ist. Das Rauschen des Kanalwiderstandes erzeugt eine effektive Rauschspannung gemäss Skizze (c) mit der spektralen Rauschdichte von

$$\frac{d \langle V_{\text{inp}}^2 \rangle}{df} = 4 k T \frac{2}{3 g_m} \quad (3)$$

Dabei ist $g_m = \partial I_{\text{drain}} / \partial V_{\text{gate}}$ der Verstärkungsfaktor des FET, die sogenannte “forward transconductance”. (siehe auch meine Elektronikvorlesung).

Diese Rauschspannung wirkt sich umso mehr aus, je grösser die elektrische Kapazität des Detektors ist, da die Rauschspannung dann mehr Strom erzeugen kann, der dem Signalstrom des Detektors überlagert wird. (Beachte aber, dass Kapazitäten selbst kein thermisches Rauschen erzeugen).

Im weiteren wird die Rauschspannung am Ausgang des Verstärkers proportional zur dessen Bandbreite. Schnelle Verstärker sehen deshalb mehr Rauschen, als solche mit grosser Integrationszeit.

2. Das **Schrotrauschen (Schottky, 1918)** entsteht durch die Quantisierung der elektrischen Ladung und führt zur Fluktuation eines Stromes. Die Variationen sind in sehr guter Näherung gaussförmig und der Erwartungswert der Varianz beträgt

$$\frac{d \langle \Delta I^2 \rangle}{df} = 2 \cdot e \cdot I \quad (4)$$

(e =Elektronladung, I =mittlerer Strom). Das Schrotrauschen ist also ebenfalls weiss. Das Schrotrauschen ist relevant bei Detektoren, durch die ein erheblicher Strom fliesst. Prominentestes Beispiel sind stark bestrahlte Halbleiterdetektoren, bei denen aufgrund der Strahlenschäden ein grosser Sperrstrom fliesst.

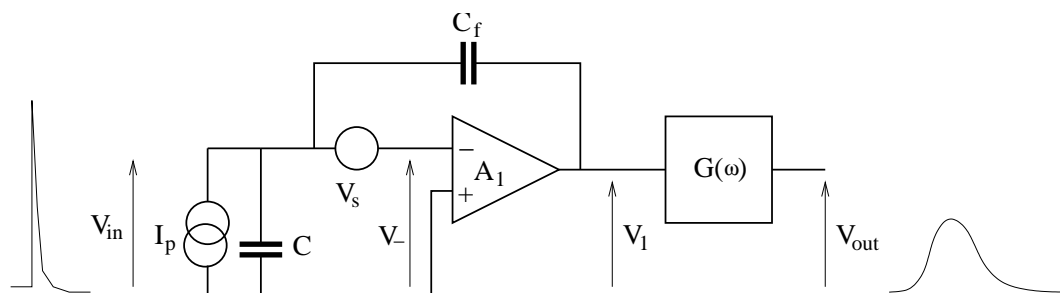
3. **Pickupnoise** entsteht durch externe Einstrahlung durch elektromagnetische Felder in den Detektor oder direkt in den Eingang des Verstärkers. Auch Spannungs-Schwankungen der Massenreferenz können Rauschen auslösen. Diese Art von Rauschen kann durch professionellen Layout im Prinzip fast vollständig vermieden werden. Im Gegensatz zu den beiden anderen Rauscharten ist seine Grösse aber sehr schwierig vorzuberechnen.

Für kalorimetrische oder Energieverlustmessungen (dE/dx) möchte man die gesamte erzeugte Ladung kennen. In diesem Fall sollte ein Ladungsintegrator mit anschliessendem Pulsformer ein einzelnes Signal erzeugen, dessen Amplitude proportional zur Ladung ist. Das Signalmaximum wird anschliessend gemessen.

Die gleichen Anforderungen gelten für sogenannte “binäre” Systeme, bei denen nur eine ja/nein Antwort des Detektors erwartet wird.

Für Driftkammern oder Time-of-Flight Messungen steht die genaue Signalkunftszeit im Vordergrund. Man muss also einen sehr schnellen Verstärker haben.

In praktisch allen Fällen wird die Eingangsverstärker aus einem Integrator und einer Pulsformerstufe wie in der folgenden Skizze bestehen:



Die Signalformen am Eingang und am Ausgang sind schematisch skizziert. I_p steht für alle parallelen Rauschquellen (z.B. Schrotrauschen), V_s für das thermische Rauschen aufgrund des ersten EingangsfET des Operationsverstärkers (“serielles” Rauschen). A_1 ist die maximale (“open loop”) Verstärkung des Integrators und $G(\omega)$ bedeutet die frequenzabhängige Verstärkungsfunktion des anschliessenden Pulsformers.

Falls man am Eingang instantan eine Ladung Q plziert, erzeugt das eine Deltafunktion als Eingangssignal. Die Spannung V_1 wird dadurch um Q/C_f erhöht. Das Ausgangssignal V_{out} wird durch die Laplacetransformierte der Uebertragungsfunktion $G(\omega)/\omega$ bestimmt.

Falls zusätzliche Widerstände am Eingang vorkommen, muss deren Rauschen ebenfalls berücksichtigt werden (zum Beispiel Vorwiderstand der Detektorhochspannung, Schutzwiderstände gegen Hochspannungsdurchschläge am Eingang des Verstärkers usw.)

In unserem Modellschema kann das Rauschen am Ausgang leicht berechnet werden (siehe zum Beispiel <http://www.physik.unizh.ch/~strauman/simnote.pdf>). Die beiden Rauschquellen sind statistisch voneinander unabhängig, ihre Amplituden werden deshalb quadratisch addiert. Mit den üblichen Regeln des Integrators und für ein lineares System erhält man für die spektrale Rauschdichte am Ausgang des Verstärkers:

$$\frac{d \langle V_{out}^2 \rangle}{df} = \left(\frac{8 k T}{3 g_m} \cdot (C + C_f)^2 + 2 e I_{bias} \cdot \frac{1}{\omega^2} \right) \cdot \left(\frac{1}{C_f + \frac{C+C_f}{A_1}} \right)^2 \cdot |G(i\omega)|^2 \quad (5)$$

wobei für den parallelen Rauschstrom die Rauschstromdichte nach (4) und für die serielle Spannungsrauschen das FET Rauschen nach (3) schon eingesetzt wurde. Um die am Ausgang gemessen Rauschspannung zu erhalten, muss noch über die Frequenz integriert werden. Man gibt dieses Rauschsignal meistens als **Equivalent noise charge (ENC)** an. Die ENC wird in Einheiten der Elementarladung gemessen und gibt an, aus wieviel Elektronen ein normales Detektorsignal bestehen müsste, damit das entsprechende Ausgangssignal am Verstärker gerade gleich der Rauschamplitude ist. Es wird

$$ENC^2 = \frac{\int_0^{\infty} df \cdot \frac{d \langle V_{out}^2 \rangle}{df}}{\frac{V_{max}^2}{Q^2}} \quad (6)$$

wobei der V_{max} die Ausgangssignalhöhe für ein Detektorsignal der Ladung Q bedeutet.

Gleichung (5) besitzt einige interessante Eigenschaften, die für den Detektorbauer wichtig sind. Vor-erst sieht man, dass das serielle Rauschen in der Tat mit der Detektorkapazität C ansteigt (in der Regel hat man $C_f \ll C$, um eine grosse Signalverstärkung zu bekommen, aber gleichzeitig A_1 sehr gross, um gute Linearität zu haben). Weiter wird das Rauschen kleiner, wenn man den Verstärker erheblich kühlen kann. g_m soll möglichst gross sein, was man mit einem möglichst “dicken” FET mit grossem Drain-Strom erreichen kann.

Der zweite Summand beschreibt den Effekt des parallelen Rauschens (zum Beispiel Schrotrauschen). Hier nimmt wegen dem Term $1/\omega^2$ bei höheren Frequenzen das Rauschen ab, da es durch die Kapazität C immer mehr “kurzgeschlossen” wird.

Je grösser die Bandbreite, je schneller der Verstärker, umso grösser wird das thermische Rauschen. Diesen Zusammenhang lässt sich an einem besonders einfachen Pulsformer zeigen, der den Frequenzgang einer Kombination von einem RC Tiefpass und einem RC Hochpass mit der gleichen Zeitkonstante $\tau = RC$ besitze. Die Verstärkungsfunktion lautet dann

$$G_1(\omega) = \frac{\omega/\tau}{(\omega + 1/\tau)^2} \quad (7)$$

Mit Hilfe der Laplacetransformation erhält man den Deltaresponse der Gesamtschaltung:

$$V_{out}(t) = \frac{Q}{C_f + \frac{C+C_f}{A_1}} \cdot \frac{t}{\tau} e^{-t/\tau} \quad (8)$$

Man spricht von einem RCCR Filter erster Ordnung, oder gelegentlich von einem Semigaussfilter, da der Anstieg des Ausgangspulses eine entfernte Aehnlichkeit mit einem halben gaussförmigen Signal besitzt. Die charakteristische Zeit τ heisst die “shaping time”. Die Breite des Signals bei halber Höhe beträgt etwa 2.5τ . Der Filter ist alles andere als ideal, sein Signal besitzt eine lange Abfallzeit, die zu hohem pileup führt; und es wird in der Praxis nicht verwendet. Hingegen ist es didaktisch interessant, da sich damit die Gleichung (5) analytisch integrieren lässt. Mit (6) wird die äquivalente Rauschladung

$$ENC_1^2 = \frac{e^2}{8 \cdot \tau} \cdot \frac{8 k T}{3 g_m} \cdot C^2 + \frac{e^2 \cdot \tau}{8} \cdot 2 q_e I \quad (9)$$

wobei e nun die Eulersche Zahl und q_e die Elementarladung ist. Der Faktor vor dem Strom wird numerisch $e^2 q_e/4 = 11.56 \text{ electron}^2/\text{nA ns}$. Beachte, dass die ENC so definiert wurde, dass sie unabhängig von der Rückkopplungskapazität C_f und dem Verstärkungsfaktor wird.

Das RCCR Filter rauscht aufgrund des seriellen Rauschens umso mehr, je schneller es ist (kleines τ). Das parallele Rauschen (zum Beispiel Schrotrauschen) wird im Gegensatz dazu für schnelle Filter immer kleiner. Hat man ein System mit mehreren parallelen und seriellen Rauschquellen, gibt es offenbar eine optimale Wahl der Zeitkonstante des Verstärkers, bei dem das Rauschen minimal wird.

Für praktische Filter muss das Integral von (5) numerisch berechnet werden, typischerweise ergeben sich Abweichungen von Faktoren 2 im Vergleich mit dem RCCR - Filter.

Ein wichtiges Kriterium für die Wahl von $G(\omega)$ stellt die Rate dar, mit der dieser Detektor kanal Signale sieht. Die Breite des Ausgangssignals soll kurz sein, im Vergleich zum mittleren Intervall zwischen zwei Teilchendurchgängen, da sonst das Signal vom alten Ereignis verfälscht wird ("pileup").

Der Integrator muss natürlich auch wieder entladen werden, da er sonst irgendwann im Anschlag der Stromversorgung landet. Es gibt geschaltete Entladungen (die meist zusätzliches Rauschen erzeugen), oder man kann einen Widerstand parallel zu C_f schalten. Dessen Dimensionierung hängt davon ab, wie gross der mittlere Eingangsstrom ist. Dieser hängt wiederum von der Teilchenrate ab, und von allfälligen Ruheströmen des Detektors.

3 Digitalisierung der Messwerte

Die Digitalisierung der analogen Detektorsignale erfolgt mit Hilfe eines **Analog-Digital-Converters (ADC)**. Es gibt verschiedene Techniken, die alle in irgendeiner Form Referenzspannungen mit dem Eingangssignal vergleichen. Ist das Eingangssignal höher, wird der entsprechende digitale Wert angenommen. Im einfachsten (und schnellsten) Fall besteht ein n-Bit ADC aus 2^n solchen Komparatoren (Flash-ADC).

Digitalisierung führt definitionsgemäss zu einer Quantisierung sowohl im Amplituden- also auch im Zeitraum. In der Amplitude wird ein Messfehler von der Grössenordnung einem bit des digitalen Wertes erzeugt. Die dadurch erzeugte Verfälschung des Signales bezeichnet man als **Quantisierungsrauschen**. Etwas genauer dargestellt, schwankt der digitale Wert eines (idealen) n-bit ADCs um maximal $\pm 0.5 \times V_0/2^n$, wo V_0 die maximal digitalisierbare Amplitude bedeutet. Die Fehlerverteilung ist also kastenförmig, die Standardabweichung des Fehlers und damit der Erwartungswert des Quantisierungsrauschens wird demnach $1 \text{ bit} / \sqrt{12}$. Diese Überlegungen gelten natürlich nur für einen idealen ADC, der keine "lokale Nichtlinearität" besitzt, das heisst die analogen Amplitudendifferenzen zwischen zwei um eins verschiedenen Digitalwerten müssen exakt gleich sein.

Um eine einzelne Digitalisierung auszuführen, braucht ein ADC ein **Konversionssignal**, das den genauen Zeitpunkt bestimmt, zu welchem das analoge Signal gemessen wird. Je nach Anwendung kommen verschiedene Methoden zum Einsatz:

Wird beim Vorliegen eines analogen Signales dessen Amplitude genau einmal gemessen werden soll (zum Beispiel zum Zeitpunkt des Maximums) spricht man von **sample and hold** und der Messzeitpunkt heisst "sampling time". Diese Technik wird angewendet, wenn Amplitudeninformation gewonnen werden soll, zum Beispiel bei kalorimetrischen oder dE/dx Messungen. Das Justieren der "sampling time" kann dabei ziemlich anspruchsvoll sein. Dafür misst man "delay curves", das heisst man verändert den sampling Zeitpunkt systematisch und sucht das Maximum der gemessenen Pulshöhe. Dafür sind reproduzierbare Testpulse notwendig.

Wird ein Signal mit einem regelmässigen Takt ("Abtastfrequenz") abgetastet, in Zeitintervallen, die klein sind gegen die Zeitstruktur des Signals, spricht man von **waveform digitizing**. Digitale Oszilloskope stellen die prominentesten Beispiele solcher Systeme dar. Bei Teilchendetektoren wird diese Technik wegen den grossen anfallenden Datenmengen nur in speziellen Fällen angewendet (zum Beispiel aktives Target des Experimentes $\pi \rightarrow e\nu$ am PSI). Hier muss auf das sogenannte "aliasing" aufgepasst werden: Enthält nämlich das analoge Signal eine Komponente mit einer Frequenz, die über der halben Abtastfrequenz liegt, wird es auf der digitalen Seite mit kleinerer Frequenz dargestellt (um

die halbe Abtastfrequenz gespiegelt, Nyquist - Theorem). Antialiasfilter werden vor der Digitalisierung eingesetzt, um höhere Frequenzen zu unterdrücken.

Binäre Systeme stellen einen besonders einfachen Fall dar. Sie bestehen aus einem Komparator oder "Ein-Bit-ADC", dessen Antwort 1 ist, falls das Signal einen bestimmten Schwellwert überschreitet und 0 sonst. Die Abtastung eines binären Systems ist zeitlich weniger kritisch. Sie werden zum Beispiel bei MWPCs und bei Siliziumstreifendetektoren eingesetzt. Binäre Auslesesysteme sollten aber nur bei völlig unkritischen Systemen verwendet werden, da man im Betrieb nur sehr wenig Informationen über die tatsächliche Form des Detektorsignales erhält. Deshalb ist praktisch keine Kontrolle über das Detektorverhalten möglich (zum Beispiel Strahlenschäden, Effekt aufgrund von Temperaturschwankungen oder Detektorgaszusammensetzungen usw.).

Gelegentlich werden auch **nichtlineare ADCs** eingesetzt. Zum Beispiel möchte man bei sampling Kalorimetern bei kleinen Signalen eine höhere Genauigkeit haben, da dessen absolute Messfehler ja mit \sqrt{E} geht. Es werden zum Beispiel annähernd logarithmische ADCs gebaut. Allerdings ist in der Praxis die Ungenauigkeit der Uebertragungsfunktion meist so gross, dass diese Lösung für Präzisionsexperimente nicht praktikabel ist

In Driftkammern oder bei Time-of-Flight Messungen muss die Zeit zwischen einem bestimmten Startpunkt (zum Beispiel der Stop eines Teilchens im Target oder der Kollisionszeitpunkt im Kolliderexperiment) und dem Detektorsignal gemessen werden. In einem **Time to Digital Converter (TDC)** wird durch ein Startsignal eine Spannungsrampe gestartet, mit einer bestimmten Steilheit (angegeben in Volt/ns). In dem Moment, wo das Detektorsignal einen bestimmten Schwellwert überschreitet wird die Rampe gestoppt und die erreichte Spannung mit einem ADC digitalisiert. Der digitale Wert ist dann proportional zur Zeitdifferenz zwischen Start und Detektorsignal. Für eine hohe Genauigkeit muss ein schneller analoger Verstärker (mit kleinem τ) verwendet werden. Mit Hilfe dieser Methode können Zeitdifferenzen von ps bis einige μs sinnvoll gemessen werden. Für längere Zeiten verwendet man besser die klassische Zeitmessmethode, bei der die Zahl der Schwingungen eines genauen Oszillators gezählt werden.

Die Digitalisierung der Messsignale soll so nahe wie möglich am Detektor und so schnell wie möglich nach dem Ereignis stattfinden, da digitale Informationen sich fehler- und rauschfrei und mit weniger Materiel und Aufwand transportieren und zwischenspeichern lassen.

Die Digitalisierung erfordert aber schnelle Kontrollsignale, Takt- (Clock-) signale usw., die störend auf den Detektor einkoppeln können, und sich so als zusätzliches Rauschen bemerkbar machen. Je näher dran man digitalisiert, umso sorgfältiger ist der Layout zu planen.

4 Einfache Triggerkonzepte

Den einfachsten Fall eines Triggers haben Sie wohl alle schon in Praktikas kennengelernt: Produziert ein Sensor ein Signal, dessen Zeitpunkt nicht vorhersehbar ist (zum Beispiel kosmische Strahlung, radioaktive Zerfälle), wird das Signal verstärkt, und dann durch eine Verzögerungsleitung (**delay line**) einem ADC zugeführt. Das Konversionssignal für den ADC wird mit Hilfe eines **Diskriminators** gebildet, dessen Eingang vor der delay line am Verstärker angeschlossen ist.

Mit dem Diskriminator werden also Signale ausgewählt, die mindestens einer bestimmten Signalthöhe entsprechen. Es wird also eine "Selektion eines Ereignisses mit einer bestimmten Signatur" (die Pulshöhe) vorgenommen.

Das Konversionssignal kann weiter dazu verwendet werden, einem Computer oder einem Zwischenspeicher mitzuteilen, dass ein Datensatz abholbereit ist.

Hat man mehrere Detektorsignale können im Trigger auch logische Funktionen verlangt werden. Verlangt man zwei oder mehr aktive Triggersignale innerhalb eines gewissen Zeitfensters Δt , spricht man von **Koinzidenzschaltung**. Möchte man zum Beispiel ein π^0 detektieren, dann muss man gleichzeitig die beiden Photonen in zwei Detektoren sehen. Da die Laufzeitdifferenz der beiden Photonen bei einem symmetrischen Aufbau sehr klein ist, kann Δt in diesem Fall so klein wie die Zeitauflösung der Detektoren gewählt werden. - Dies ist die einfachste Version einer "Selektion von Ereignissen, bestimmter physikalischer Prozesse".

Nun gibt es zwei Wege zu höherer Komplexität:

Erstens könnten wir ein Experiment mit einer grösseren Kanalzahl wählen. Dann können mit den Triggersignalen kompliziertere Verknüpfungen erstellt werden. Dazu verwendet man am besten FPGA - Bausteine (field programmable gate arrays). Solche gibt es mit mehreren hundert Eingängen, und es können - programmierbar - praktisch beliebig komplexe logische Verknüpfungen erstellt werden. Dafür muss aber die Delay Line länger werden, da wir jetzt für die Entscheidung mehr Zeit brauchen.

Zweitens könnten wir auch nach dem ADC mit den digitalen Werten noch etwas rechnen. Im π^0 Beispiel könnten wir verlangen, dass die Summe der Energieen der beiden Photonen etwa der Masse des π^0 's entspricht. Dafür brauchen wir eine Addierschaltung und etwas digitale Logik, deren Ausgang schliesslich entscheidet, ob das Ereignis dem Speicher übergeben werden soll. Wir haben also eine zweite Entscheidungsstufe verwendet, den einfachsten Fall eines Multilevel-Triggers. Solche Multilevel-Trigger können ebenfalls mit FPGA's oder auch mit gewöhnlichen Prozessoren aufgebaut werden, falls letztere schnell genug sind.

In diesen beiden Fällen braucht es eine lange delay line oder einen Speicher, der Daten während der Zeit aufbewahrt, in dem die Triggerentscheidung getroffen wird. Das wird im nächsten Abschnitt diskutiert.

5 Pipelines

Das Verlängern der Delay Lines, um genügend Zeit für die Triggerentscheidung zu bekommen, kann schnell an seine Grenzen stossen. 100 ns delay erfordert etwa 20 m Kabel. Alternativ kann man **pipelines** verwenden. Dabei verwendet man mit Schaltern verbundene Kondensatoren (analog pipeline) oder digitale Register. Eine solche Speicherstelle heisst "pipe line cell". Ein externer Takt puls schiebt das Ereignis jeweils eine Zelle weiter. Konkret wird in der Regel nicht die Information verschoben, sondern die Zelle kriegt vom Takt nur einfach eine neue Nummer. Ein "write pointer" zeigt zu der Zelle, in die als nächstes geschrieben werden soll, und ein "read pointer" zeigt auf die Zelle, die als nächstes ausgegeben wird. Der Takt schiebt die den write pointer jeweils eine Zelle weiter. Befindet sich einer der pointer am Ende der pipeline springt er zur Position eins. Man nennt eine solche Anordnung auch **Ringspeicher oder circular buffer**.

Die Länge der pipeline, also die Differenz-Zeit zwischen read und write pointer, heisst die **latency**. Die Latency kann höchstens so gross eingestellt werden, wie der Ringspeicher Plätze hat.

Bei der Wahl der Taktfrequenz muss man die erforderliche Zeitauflösung des Detektorsignales berücksichtigen. Ausserdem werden selbstverständlich vor allem in in den analog pipelines Messfehler in der Ampli-

tude erzeugt, die man verstehen muss. In dieser Hinsicht ist natürlich der einfache Kabeldelay allen anderen Methoden überlegen.

Bei Kolliderexperimenten mit hoher Taktfrequenz dauert die Triggerentscheidung meistens länger als das Intervall zwischen zwei bunch crossings (BC oder BX). In diesem Fall muss man auch die Entscheidungslogik als pipeline aufbauen: Eine erste Operation wird ausgeführt, dann das Resultat der nächsten Stufe übergeben und gleichzeitig das nächste BX eingelesen. Beispiel Energiesumme des π^0 : Im ersten Schritt erzeugt der ADC den digitalen Wert, im zweiten werden die beiden Werte addiert, und im dritten wird die Energie gleichzeitig (“parallel”) mit zwei verschiedenen Werten verglichen. Befindet sich die Energie dazwischen, dann wird im vierten Schritt das Abspeichern ausgelöst. Für diesen Trigger beträgt die latency 4 BX, es befinden sich immer 3 verschiedene Ereignisse gleichzeitig in der Triggerlogik. Das Beispiel illustriert auch den Unterschied zwischen **parallel processing** und **pipelining**.

Triggersysteme mit einer festen latency heissen **synchron**. Asynchrone Systeme bekommt man, wenn man Prozessoren einsetzt, die zum Beispiel bei einem Ereignis alle gefundenen Spuren untersuchen sollen, und von allen gewisse Berechnungen machen. Dann dauert die Rechnung verschieden lang, je nach dem wieviel Spuren detektiert worden sind. Man bekommt eine variable latency. Es werden viele Prozessoren eingesetzt, der jeder jeweils ein Ereignis bearbeitet. Nun ist es möglich, dass sich die Ereignisse zeitlich “überholen”, weshalb bei einem solchen System eine eindeutige Nummerierung der Ereignisse so früh nach dem Ereigniszeitpunkt als möglich erforderlich ist.

6 Occupancy, Pileup, Spill over, Cross talk. Random coincidences

Unter **Occupancy** versteht man die Wahrscheinlichkeit, dass bei einem Ereignis ein Messkanal ein Signal sieht. Die Occupancy sollte immer sehr klein gegen eins sein. Denn die Occupancy gibt gleichzeitig die Wahrscheinlichkeit an, dass ein Signal einem falschen physikalischen Objekt zugeordnet wird (zum Beispiel Zuordnung von Signalen zu falschen Spuren).

Wenn ein Signal von früheren Ereignissen her gestört wird, spricht man von **pileup**. Es gibt einen elektronischen pileup, wenn ein Signal auf dem auslaufenden Signal des vorherigen Ereignisses liegt und damit das Signal zum Beispiel eine zu grosse Amplitude bekommt. Die Wahrscheinlichkeit dafür hängt von der Zeitkonstanten und Form des Signales ab, sowie von der occupancy.

Es gibt auch “physikalischen” pileup, wenn zum Beispiel ein von einem früheren Ereignis herührendes positives Müon im Target oder Detektor stoppt, und es nach $2 \mu\text{sec}$ gleichzeitig mit einem neuen Ereignis zerfällt. Dann gibt es in diesem neuen Ereignis ein scheinbar unerklärbares Positron. Durch Bestrahlung oder durch natürliche Radioaktivität entstehende Ionisation stellt ebenfalls eine Art pileup dar, man spricht hier aber meistens eher von “Aktivierungsuntergrund”.

Wenn bei einem Collider nach einem Ereignis noch Signale übrigbleiben, die sich im nächsten BX wie gewöhnliche Signale darstellen, spricht man von **spill over**.

Bei getriggerten Experimenten sind pileup und spill over oft schwierig zu erkennen. Weil der grösste Teil der Ereignisse ja weggeworfen und nicht gespeichert wird, sind bei der Analyse eines Ereignisses in der Regel die Daten der vorhergehenden bunch crossings nicht vorhanden. Man kann deshalb nicht nachsehen, ob ein spill over erwartet werden muss. Die Wahrscheinlichkeit, dass das vorkommt, kann man aber natürlich leicht berechnen, bzw. simulieren.

Schliesslich versteht man unter **Cross talk** das elektrische Uebersprechen von einem Messkanal auf einen anderen. Der einfachste Fall besteht darin, dass durch die Kapazität zwischen zwei Dektoelementen oder Anschlussleitungen, ein Signal auch von einem falschen Kanal gesehen wird.

Random coincidences kann man am besten anhand des Beispiels des π^0 -Zerfalls verstehen. Es ist ja namlich moglich, dass die beiden Photonen gar nicht vom gleichen π^0 stammen, sondern von zwei verschiedenen, die zufallig etwa zur gleichen Zeit zerfallen sind. Oder es gibt in den entsprechenden Prozessen weitere Moglichkeiten, Photonen zu produzieren. Wie gross ist die Zahlrate P_r , fur zufallige Koinzidenzsignale, in welchen die Photonen rein zufallig beide in das gleiche Intervall Δt fallen? Sie ist sicher proportional zu Δt , und zu den Einzelzahlraten Z_i der beiden Detektoren, es gilt

$$P_r = Z_1 \cdot Z_2 \cdot \Delta T$$

7 Totzeit

Produziert ein Detektorkanal ein Signal, das digitalisiert und vielleicht in einer pipeline verzogert wird, entsteht fur das System eine Totzeit, das heisst es existiert ein Zeitintervall um das Ereignis herum, in dem keine weiteren Ereignisse detektiert werden konnen. Die Totzeit entsteht oft auch durch die endliche Rechenzeit eines Prozessors, der die Daten auslesen muss.

Damit die gespeicherten Daten nicht verfalscht werden, wird in der Regel ein ‘‘Computergate’’ oder ‘‘run gate’’ verwendet. Es handelt sich um ein Register, das entweder den Wert ‘‘on’’ oder ‘‘off’’ annehmen kann. Ein Ereignis wird nur akzeptiert, wahrend der ‘‘on’’ Zeit. Ein akzeptiertes Ereignis setzt das gate auf ‘‘off’’. Erst nachdem der Prozessor mit der Verarbeitung des Ereignisses fertig ist, setzt er das gate wieder auf ‘‘on’’. Das gate wird in Form eines digitalen Flip-Flops implementiert (RS-Flip-Flop).

Die **gesamte Totzeit** d des experimentellen Aufbaus wird in Prozenten der Messzeit angegeben. Die Verarbeitungszeit, oder die Totzeit pro Ereignis τ wird in ms angegeben. Ein Ereignis, das wahrend der Totzeit auftritt, geht verloren. Der prozentuale Anteil der akzeptierbaren Ereignisse ϵ hangt direkt von d ab.

Bei **zeitlich rein zufallig verteilten** Ereignissen, wie es bei radioaktiven Zerfallen oder Experimenten mit kosmischer Strahlung auftritt, erzeugt jeder Datenaufbau grundsatzlich eine Totzeit. Es gilt $\epsilon = 1 - d$. Sei R_{true} die wahre Ereignisrate (Ereignisse pro Zeiteinheit), und R_{acc} die akzeptierte Ereignisrate. Offensichtlich gilt:

$$R_{acc} = \epsilon \cdot R_{true} = (1 - d) \cdot R_{true}$$

Die akzeptierten Ereignisse erzeugen eine Totzeit:

$$d = R_{acc} \cdot \tau = (1 - d) \cdot R_{true} \cdot \tau$$

Damit wird die prozentuale Totzeit:

$$d = \frac{R_{true} \cdot \tau}{1 + R_{true} \cdot \tau} \quad (10)$$

und

$$\epsilon = \frac{1}{1 + R_{true} \cdot \tau} \quad (11)$$

Dies gibt die Abhangigkeit der prozentualen Totzeit von der wahren Ereignisrate an. Es gibt immer eine Totzeit, aber bei grossen Raten nimmt der Anteil der akzeptierten Experimente proportional zur Rate ab.

Die obigen Formeln lösen auch die praktische Aufgabe, aus der gemessenen Ereignisrate R_{acc} und der ebenfalls bekannten Totzeit τ die wahre Ereignisrate R_{true} zu berechnen. Bei zufällig verteilten Ereignissen, gibt es immer Datenverlust wegen Totzeit, und es besteht keine technische Möglichkeit, das vollständig zu verhindern.

Auf der anderen Seite führt eine kleine Totzeit dazu, dass die in der Regel nicht gerade billige Prozessorhardware während der meisten Zeit herumsteht und nichts zu tun hat, als auf das nächste Ereignis zu warten.

Treten die Ereignisse **in einem festen Taktintervall** t_{BX} auf (z.B. bei einem Colliderexperiment), verändert sich die Situation drastisch. Jetzt müssen wir zwei Fälle unterscheiden: Ist $\tau < t_{BX}$ gibt es überhaupt keine Totzeit. Dies war zum Beispiel bei den LEP - Experimenten der Fall, da dort $t_{BX} = 22\mu s$ und damit viel Zeit für die Verarbeitung war. Bereits bei HERA und bei allen neueren Kollidern ist jedoch t_{BX} sehr klein. Hier muss mit einem rigorosen Multilevel-Triggerkonzept die akzeptierte Ereignisrate klein gehalten werden.

Zusätzlich kann mit einer Kombination von pipelines und sogenannter **“Derandomizer buffers”** in Form von FIFOs ein Ereignisverlust fast vollständig vermieden werden. Während der ersten Triggerstufe werden die Daten in einer pipeline gespeichert, die in Form eines Ringspeichers organisiert ist. Getriggerte Ereignisse werden anschliessend in einem FIFO (first in - first out register) zwischengespeichert. Die zweite (und ebenso alle höheren) Triggerstufen arbeiten mit maximaler Ausnutzung der Rechenleistung diese FIFOs ab.

Dieser Aufbau erlaubt die vollständige Ausnutzung der vorhandenen Prozessorleistung, ohne dass Totzeit erzeugt wird. Die Tiefe der eingesetzten FIFOs (die maximale Anzahl zwischenspeicherbarer Ereignisse) ist sorgfältig zu optimieren.

Bei komplexen Datennahmesystemen muss mit Hilfe von **Simulationen** bestimmt werden, wie es sich bei hohen Ereignisraten verhält. Dabei soll einerseits die eingesetzte Prozessorleistung möglichst gut genutzt werden, andererseits die Ereignisverlustrate sehr klein gehalten werden, da es keinen Sinn hat mit grossem Aufwand eine hohe Rate zu erzeugen, die dann wegen Totzeit des Datennahmesystems nicht verarbeitet werden kann.

8 Höhere Triggerkonzepte und Ereignisselektion

Wie im vorhergehenden Abschnitt beschrieben, stehen die Ereignisse nun vollständig zur Verfügung, sodass im Prinzip eine komplette physikalische Analyse damit möglich ist. Diese besteht aus den Schritten: Signalaufbereitung, Rekonstruktion der einzelnen produzierten Teilchen und Bestimmung der charakteristischen Grössen des gesamten Ereignisses.

Normalerweise müssen auch auf dieser Stufe noch ein grosser Anteil der Ereignisse verworfen werden, da Prozessorleistung und Speichervolumen beschränkt sind. Im Gegensatz zu “offline” Analysen muss hier mit festen Bedingungen gearbeitet werden: Ereignisse werden genau dann akzeptiert, wenn sie eine wohldefinierte Liste von Bedingungen an einzelne physikalische Grössen (oder an eine Kombination von solchen) erfüllen, sogenannte **cuts**. Beispiele: Gesamtenergie des Ereignisses, höchster vorkommender transversaler Impuls im Ereignis, mindestens eine feste Zahl von Müonen.

Im Gegensatz zu einer offline Analyse, müssen die cuts vor dem Start der Messung festgelegt werden. Die nicht akzeptierten Ereignisse sind unwiderruflich für immer verloren. Logischerweise können deshalb multivariate Optimierungen von Analysen nur offline durchgeführt werden.

Um die Prozessorleistung effizient zu nutzen, werden die Ereignisse nicht zuerst vollständig rekonstruiert, und erst dann auf die Bedingungen der cuts geprüft. Vielmehr werden die cuts möglichst früh im algorithmischen Ablauf der Rekonstruktion angewendet. Hat zum Beispiel ein Ereignis keine Müonen, obschon solche verlangt sind, dann muss man die Gesamtenergie und die Impulse gar nicht mehr ausrechnen, das Ereignis kann sofort weggeworfen und mit der Verarbeitung des nächsten Ereignisses begonnen werden.

Normalerweise werden für diese Arbeiten eine grosse Anzahl von Prozessoren (z.B. kommerzielle PC's) verwendet (sogenannte **processor-farms**). Ein Netzwerk verteilt jedem Rechner ein neues Ereignis, sobald dieser mit dem letzten fertig ist. Dabei wird ausgenützt, dass die Ereignisse voneinander vollständig unabhängig sind, sodass ausser dem Einlesen des Ereignisses und der Ausgabe des Resultates keine Kommunikation zwischen den Rechnern notwendig ist.

Wichtig ist es natürlich, dass man genau versteht, was die cuts tun, damit man nicht die wichtigste Physik wegen falscher cuts verliert. Es ist aber klar, dass man im Trigger immer nur Ereignisarten selektionieren kann, die man auch erwartet. Unerwartete Physik kann so grundsätzlich nicht detektiert werden, es sei denn sie äussere sich mit sehr spektakulären Ereignissen, mit zum Beispiel einer sehr hohen Gesamtenergie.

Um sicher zu sein, dass alles richtig läuft, werden sogenannte **Monitortrigger** verwendet. An solche werden nur ganz schwache cuts angewendet, die von fast allen Ereignissen erfüllt werden (auch **minimum bias trigger** genannt). Um nicht von solchen Ereignissen überschwemmt zu werden, werden davon nur ein fest eingestellter Bruchteil (zum Beispiel jedes hundertste Ereignis) ausgelesen und gespeichert: **prescaling**.

9 Ergänzung Geant

Geant, in der aktuellen Version Geant 4, ist ein Programmpaket, das Teilchen durch Materie und Elektro-magnetische Felder transportiert, und dabei alle elektromagnetischen und hadronischen Wechselwirkungen berechnet. Es werden zum Beispiel Energieverluste durch Ionisation berücksichtigt und Sekundärteilchen generiert (z.B. Photonen und Elektronen bei elektromagnetischen Schauern in den elektrischen Feldern der Atomkerne). Die langlebigen Teilchen (z.B. Müonen, geladene Pionen und Kaonen) zerfallen auch korrekt aufgrund ihrer Lebensdauer.

Solche Monte-Carlo Simulationsprogramme werden benötigt, um das Aussehen und die statistischen Verteilungen der Produkte der elementaren Prozesse in den Experimenten vorherzusagen. Der Messprozess besteht dann darin, die tatsächlich im Experiment gemessenen Verteilungen mit denen der Simulation zu vergleichen.

Es wird empfohlen im Selbststudium die webseite <http://cern.ch/geant4/> und speziell darin den link "more about Geant, Physics Manual" zu studieren. Letzteres stellt auch eine gute Zusammenfassung aller Wechselwirkungen von Teilchen mit Materie dar.