

II. Charakterisierung von Meßfehlern

Nachdem im vorangegangenen Abschnitt die Charakterisierung von Meßgrößen und Meßsignalen im Vordergrund stand, soll im folgenden der Einfluß von Meßfehlern und ihre Beschreibung behandelt werden. Dabei muß man prinzipiell zwischen systematischen (deterministischen) Meßfehlern und unsystematischen (statistischen) Meßfehlern unterscheiden. Systematische Meßfehler entstehen aufgrund des zugrundeliegenden Meßprinzips, das unter Umständen eine direkte Einwirkung auf die zu messende Größe hat und daher eine deterministische Meßverfälschung bewirkt. Derartige Fehler lassen sich durch geeignete Konstruktion der Apparatur vermeiden, was jedoch nur in Grenzen möglich ist und z. B. bei allen quantenmechanischen Systemen an prinzipielle Grenzen stößt. Außerdem können verschiedene Meßprinzipien zur Bestimmung derselben Meßgröße eingesetzt werden, so daß durch Vergleich der mit diesen voneinander unabhängigen Meßmethoden erzielten Meßwerte der systematische Fehler abgeschätzt werden kann.

Der unsystematische (statistische) Meßfehler entsteht dagegen durch nicht im Detail vorhersehbare, statistische Schwankungen, deren Ausmaß zwar abschätzbar, aber die jeweilige Auswirkung auf jeden einzelnen Meßvorgang nicht genau und nur im Mittel abgeschätzt werden kann. Diese Art von Meßfehlern lassen sich zwar ebenfalls durch die Konstruktion der Meßapparatur verringern, dabei treten jedoch prinzipielle Grenzen auf. Beispielsweise läßt sich der statistische Fehler durch Mittelung des Meßwertes über verschiedene Meßvorgänge (bzw. Meßzeitpunkte) mit derselben Apparatur verringern und durch Mittelung der Meßwerte von verschiedenen Apparaturen, die identisch aufgebaut sind. Dieser statistische Meßfehler ist damit eine Zufallsgröße, die eng mit den Begriffen „stochastischer Prozeß“ und „Rauschen“ verbunden ist und im folgenden näher behandelt werden soll. Dazu werden zunächst einige grundlegende Definitionen und Eigenschaften von Begriffen erläutert, die in der Statistik eine große Bedeutung haben.

II. 1 Zufallsvariable und Verteilungen

Von prinzipieller Bedeutung ist der Begriff der **Zufallsvariable X**. Anschaulich versteht man darunter eine Funktion, die jedem Ereignis, das mit einer Wahrscheinlichkeit P auftreten kann, einen reellen Wert x zuweist. Dieses kann z. B. im Fall eines Würfels der Wert $x =$ (Anzahl der Augen) sein, wobei x die Werte 1 bis 6 annehmen kann und eine diskrete Zufallsvariable ist. x kann aber auch eine kontinuierliche Zufallsvariable sein, z. B. die Körpergröße eines beliebig ausgewählten Studenten der Universität Oldenburg. Die exakte mathematische Definition dieser Zu-

fallsvariablen ist begrifflich nicht sehr einfach, hier sei auf die entsprechende Literatur (z. B. Papoulis, A.: Probability, random variables and stochastic processes, McGraw Hill, 1965) hingewiesen. Bei der formalen mathematischen Definition bezeichnet die Zufallsvariable X eine Funktion vom Ereignisraum Ω in die reellen Zahlen \mathbb{R} , also: $X: \Omega \rightarrow \mathbb{R}$.

Zusätzlich zum Ereignisraum Ω muß eine σ -Algebra (oder im Fall einer endlichen, diskreten Zufallsvariable eine Potenzmenge) $F(\Omega)$ existieren, die jede mögliche Kombination von Ereignissen aus dem Ereignisraum Ω enthält. Insbesondere muß das Urbild jeder Zahl und jedes Intervalls aus \mathbb{R} in F enthalten sein. Auf dieser σ -Algebra ist schließlich das Wahrscheinlichkeitsmaß $P: F(\Omega) \rightarrow [0, 1]$ definiert, so daß für jede Realisation der Zufallsvariable X bzw. für jeden Bereich von Realisationen von X eine Wahrscheinlichkeit P angegeben werden kann, die zwischen dem Wert 0 und dem Wert 1 variiert. Für den Fall des Würfels besteht $F(\Omega)$ aus der Menge sämtlicher Kombinationen von den 6 Elementar-Ereignissen (einschließlich der möglichen Dreier-, Vierer-, Fünfer- und Sechserkombinationen). Jeder dieser möglichen Kombinationen wird eine Wahrscheinlichkeit zugeordnet, wobei diese Wahrscheinlichkeit leicht unter der Annahme konstruiert werden kann, daß jedes Elementarereignis mit der Wahrscheinlichkeit $1/6$ auftritt. Für das Beispiel der Körpergröße eines männlichen Studenten ist die Angabe von Ω und $F(\Omega)$ nicht so leicht möglich, allerdings wird durch die Definition festgelegt, daß für jede mögliche Körpergröße, d. h. jede Realisierung der Zufallsvariable X ein Wahrscheinlichkeitswert angegeben werden kann, daß diese Körpergröße auftritt.

Für eine Zufallsvariable läßt sich dann eine **Verteilungsfunktion**

$F: \mathbb{R} \rightarrow [0, 1]$ angeben, für die gilt:

$$F(x) = P(x \geq X) \quad (\text{II.1})$$

d. h. F bezeichnet die Wahrscheinlichkeit, daß die Zufallsvariable X den Wert x nicht übersteigt. Als Beispiel sei die Verteilungsfunktion für den „idealen“ Würfel angegeben, der eine Treppenfunktion darstellt, sowie die Verteilungsfunktion für die Körpergröße eines männlichen Studenten, bei der der Einfachheit halber angenommen wird, daß Studenten der Universität Oldenburg im Mittel eine Körpergröße von 1,8 m haben.

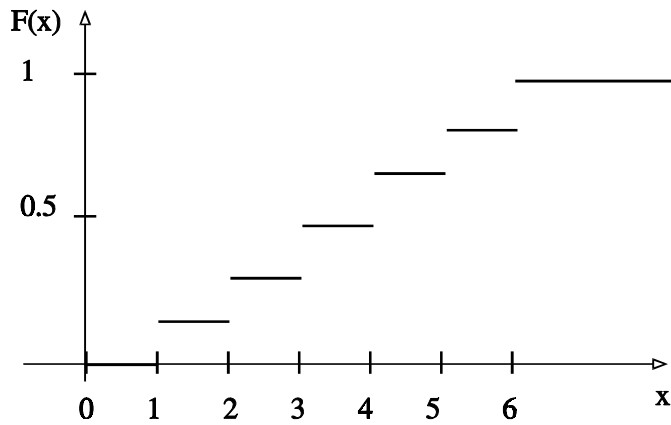


Abb. 2.1: Verteilungsfunktion für den Würfel

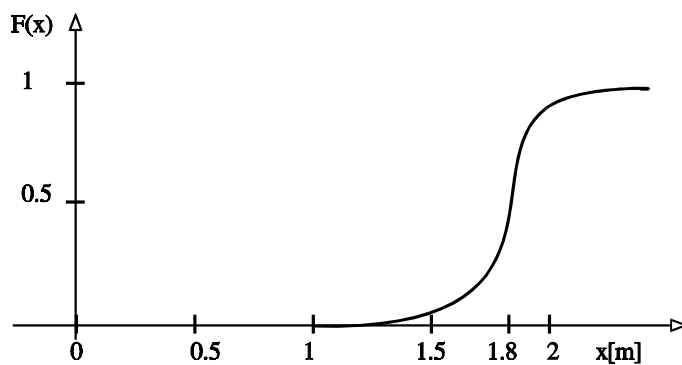


Abb. 2.2: Verteilungsfunktion für die Körpergröße

Ausgehend von der Verteilungsfunktion kann nun die **Verteilungsdichte-Funktion** $f: \mathbb{R} \rightarrow \mathbb{R}$ angegeben werden, für die gilt:

$$F(x) = \int_{-\infty}^x f(x') dx', \quad (II.2)$$

$$\int_{-\infty}^{\infty} f(x') dx' = 1$$

D. h. die Wahrscheinlichkeit, daß X in dem Intervall $[x, x+dx]$ liegt, beträgt $f(x)dx$. Die Verteilungsdichte-Funktion ist somit die Ableitung der Verteilungsfunktion und nimmt für das Beispiel des Würfels die folgende Form an:

$$f(x) = \sum_{i=1}^6 \frac{1}{6} \cdot \delta(x-i) \quad (II.3)$$

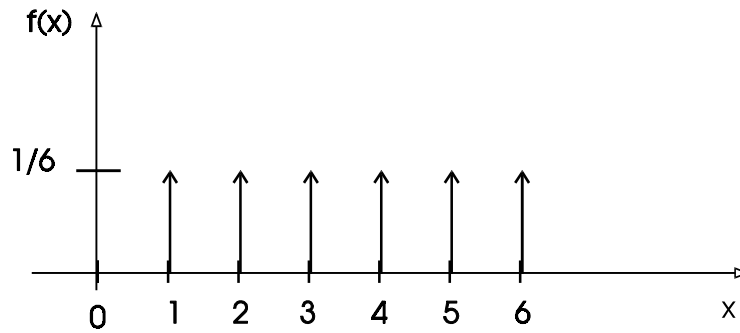


Abb. 2.3: Verteilungsdichtefunktion für den Würfel

Für die Körpergröße des männlichen Studenten beträgt die Verteilungsdichte-Funktion:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(Normalverteilung) (II.4)

z. B. $\mu \approx 1,8\text{m}$

$\sigma \approx 0,1\text{m}$

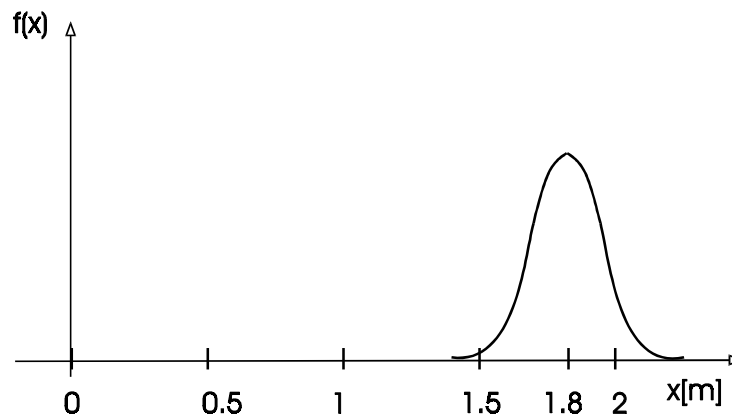


Abb. 2.4: Verteilungsdichtefunktion für die Körpergröße

Die Verteilungsfunktion bzw. Verteilungsdichte-Funktion kann durch verschiedene Parameter festgelegt sein, die z. B. die Lage der Funktion und ihre Breite beschreiben. Der wichtigste Parameter, der die Lage der Verteilungsfunktionen beschreibt, ist der **Erwartungswert**

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \bar{x} \quad (\text{II.5})$$

Der Erwartungswert ist somit eine Eigenschaft der Verteilungsfunktion

und ist nicht mit dem später noch zu behandelnden Mittelwert zu verwechseln, d. h. dem Mittel aus mehreren Realisationen derselben Zufallsvariable. Für das Beispiel des Würfels beträgt der Erwartungswert $E(X) = 3,5$ und für das Beispiel der Körpergröße beträgt der Erwartungswert $E(X) = \mu$ (z. B. $\mu = 1,8$ m wie oben angenommen).

Die „effektive“ Breite der Verteilungsdichte-Funktion wird durch die Standardabweichung σ bzw. die Varianz σ^2 ausgedrückt, für die gilt:

$$\sigma^2 = E([X - E(X)]^2) \quad (\text{II.6})$$

Die Varianz ist damit ebenfalls ein Parameter, der die zugrundeliegende Verteilungsfunktion bzw. Verteilungsdichte-Funktion charakterisiert und als mittlere quadratische Abweichung vom Erwartungswert \bar{x} aufgefaßt werden kann. σ^2 läßt sich auch ausrechnen als:

$$\sigma^2 = E([X - E(X)]^2) = \int_{-\infty}^{\infty} (x - \bar{x})^2 \cdot f(x) dx = E(X^2) - E^2(X) \quad (\text{II.7})$$

Für den Würfel errechnet sich $\sigma^2 = \frac{35}{12}$, $\sigma = 1,708$ und für die im Beispiel angegebene Körpergröße ergibt sich

$$E(X^2) - E^2(X) = \sigma^2, \text{ z. B. } \sigma = 0,1\text{m} \quad (\text{II.8})$$

Aus dem letzten Beispiel ist abzulesen, daß in der Formel für die Normalverteilung der Parameter σ bereits als die anhand der hier gegebenen Definition zu erwartenden Varianz eingesetzt wurde.

Neben der oben bereits eingeführten Normalverteilung gibt es eine Reihe weiterer wichtiger Verteilungen, die im folgenden vorgestellt werden sollen. Die **Binomialverteilung** ist die diskrete Verteilung für das Ergebnis bei einer n-fachen Wiederholung eines **Bernoulli-Experiments**. Ein derartiges Experiment hat nur zwei mögliche Ereignisse mit der Wahrscheinlichkeit p bzw. $1 - p$. Beispielsweise ist das Werfen einer Münze mit dem Ereignis A (Münze zeigt „Zahl“, Wahrscheinlichkeit $p = 0,5$) und dem Ereignis \bar{A} (Münze zeigt „Kopf“, Wahrscheinlichkeit $1 - p = 0,5$) ein derartiges Bernoulli-Experiment. Die Binomialverteilung gibt nun die Wahrscheinlichkeit an, daß k -mal das Ereignis A auftritt, wobei die genaue Rei-

henfolge des k -maligen Auftretens von A und des $(n-k)$ -maligen Auftretens des Ereignisses \bar{A} egal ist:

$$b_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (\text{II.9})$$

Der Zahlenfaktor $\binom{n}{k}$ gibt die Anzahl der möglichen Anordnungen von k Ereignissen A und $n-k$ Ereignissen \bar{A} als Sequenz der Länge n an. Die zugehörige Zufallsvariable X beschreibt dabei die Anzahl der „Zahl“-Ereignisse bei n Versuchen (d. h. im vorliegenden Fall Anzahl der Ereignisse A). Die Verteilungsdichte-Funktion dieser Zufallsvariable beträgt damit:

$$\begin{aligned} f(x) &= \sum_{k=0}^n b_{n,p}(k) \cdot \delta(x-k) \\ E(X) &= n \cdot p \\ \sigma^2 &= n \cdot p(1-p) \end{aligned} \quad (\text{II.10})$$

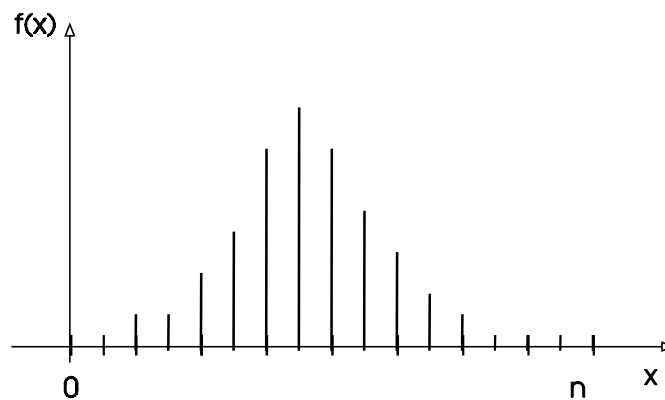


Abb. 2.5: Verteilungsdichtefunktion für Münzwurf

Als Erwartungswert ergibt sich damit bei n Ereignissen genau das „erwartete“ Ergebnis, nämlich daß bei der Auftretenswahrscheinlichkeit p das Ereignis $n \cdot p$ -mal zu erwarten ist. Interessant ist für die Varianz σ^2 , daß sie maximal wird falls $p = \frac{1}{2}$ ist. Für $n \rightarrow \infty$ (d. h. für eine beliebig häufige Wiederholung des Bernoulli-Experiments) geht die Verteilung in eine Standard-Normalverteilung über, wenn anstelle der Zufallsvariable X eine „normierte“ Zufallsvariable

$$Y = \frac{X - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}} \quad (\text{II.11})$$

gebildet wird. Diese Zufallsvariable Y konvergiert für $n \rightarrow \infty$ gegen die Standard-Normalverteilung

$$f(Y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Y^2}{2}} \quad (\text{II.12})$$

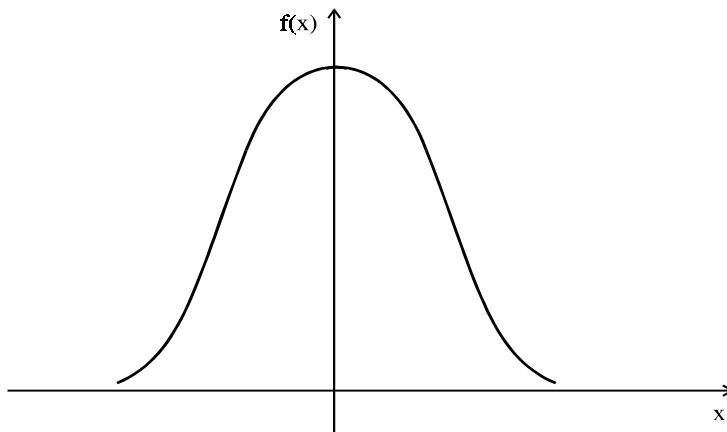


Abb. 2.6: Standard-Normalverteilung

Eine weitere wichtige Verteilung, die als Grenzfall der Binomialverteilung bzw. der Normalverteilung eine Rolle spielt, ist die **Poissonverteilung**: Sie tritt beispielsweise beim radioaktiven Zerfall auf, bei dem die Wahrscheinlichkeit eines Zerfalls im Zeitintervall der Länge t λt beträgt. Die Wahrscheinlichkeit der k unabhängigen Zerfällen im Zeitintervall der Länge t ist daher proportional zu $(\lambda t)^k$. Allerdings sind in diesem Produkt sämtliche möglichen Anordnungen der k Zerfälle im gesamten Zeitintervall als getrennte Möglichkeiten enthalten. Um diese Zahl von $k!$ Anordnungsmöglichkeiten von k Zerfallszeitpunkte innerhalb des Intervalls vorgegebener Länge zu berücksichtigen, muß die Wahrscheinlichkeit korrigiert werden, so daß sich insgesamt eine Wahrscheinlichkeit proportional zu folgendem Ausdruck ergibt:

$$p(k) \approx \frac{(\lambda t)^k}{k!} \quad (\text{II.13})$$

Weiterhin muß die Gesamt-Wahrscheinlichkeit, daß $k = 0$ bis $k = \infty$ Zerfälle im Zeitintervall der Länge t auftreten, den Wert 1 ergeben, so daß sich die folgenden Normierungsbedingung ergibt:

$$\sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} = e^{\lambda t} \Rightarrow p(x = k) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!} \quad (\text{II.14})$$

Die zu dieser Wahrscheinlichkeit gehörenden Zufallsvariable X ist die Anzahl der Zerfälle im Intervall der Länge t . Ihre Verteilungsfunktion beträgt demnach:

$$f(X) = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \cdot \delta(X - k)$$

$$E(X) = \lambda \cdot t$$

$$\sigma^2 = \lambda t = E(X) \quad (\text{II.15})$$

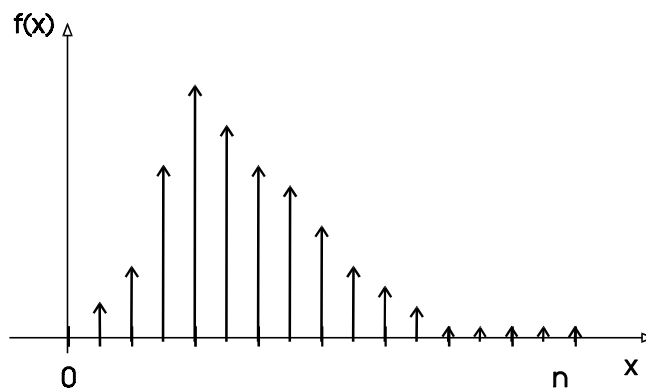


Abb. 2.7: Poisson-Verteilung

Die Poissonverteilung ist der Grenzwert der Binomialverteilung für $n \rightarrow \infty$ und $n \cdot p \rightarrow \lambda \cdot t = \text{const}$, d.h. für verschwindend kleine Bernoulli-Wahrscheinlichkeit p . Da die Binomialverteilung für $n \rightarrow \infty$ gegen die Normalverteilung gilt, geht die Poisson-Verteilung für große Werte von $\lambda \cdot t$ ebenfalls gegen die Normalverteilung.

Der Vollständigkeit halber sei die **Gleichverteilung** erwähnt, für die die Verteilungsdichte-Funktion $f(x)$ den konstanten Wert $1/a$ annimmt:

$$f(x) = \begin{cases} \frac{1}{a} & \text{für } 0 \leq x \leq a \\ 0 & \text{sonst} \end{cases}$$

$$E(X) = \frac{1}{2a}$$

$$\sigma^2 = \frac{a^2}{12} \quad (\text{II.16})$$

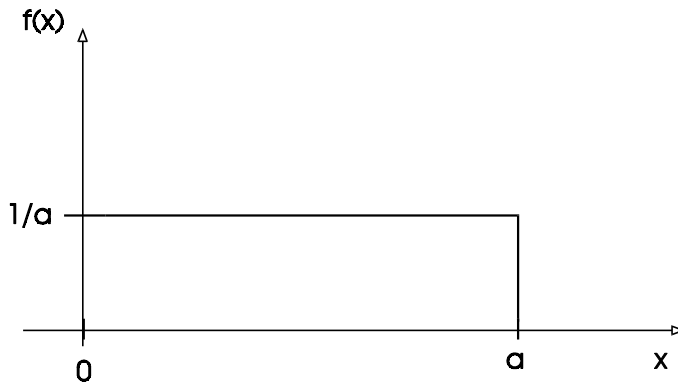


Abb. 2.8: Gleichverteilung

II.2 Summen von Zufallsvariablen, Zentraler Grenzwertsatz

In der Realität hat man es häufig nicht nur mit einer Zufallsvariablen zu tun, sondern mit der Funktion von einer oder mehreren Zufallsvariablen, so daß man sich die Verkettung von mehreren Zufallsvariablen ansehen muß. Der einfachste Fall ist die **Summe zweier Zufallsvariablen**: sei X eine Zufallsvariable mit der Verteilungsdichte-Funktion $f(x)$ und Y eine Zufallsvariable mit der Verteilungsfunktion $g(y)$, dann ist $Z = X+Y$ die Summe dieser beiden Zufallsvariablen, d. h. die Realisation z von Z wird jeweils aus der Summe der Realisation x von X und y von Y gebildet:

$$Z = X+Y \quad (\text{II.17})$$

Die Verteilungsdichtefunktion $h(z)$ der Summe ergibt sich nun als **Faltung** der ursprünglichen Verteilungsdichtefunktion:

$$h(z) = f(x)*g(y) = \int_{x=-\infty}^{\infty} f(z-x) \cdot g(x) dx \quad (\text{II.18})$$

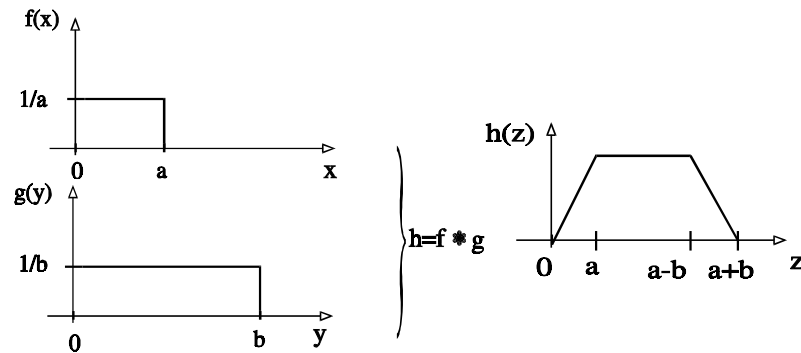


Abb. 2.9: Summe zweier Zufallsvariablen

Als Beispiel ist die Faltung der Verteilungsdichte-Funktion von zwei gleichverteilten Zufallsvariablen angegeben, die eine trapezförmige Verteilungsdichte-Funktion ergibt. Diese Trapez-Form kann man sich anschaulich dadurch erklären, daß es sehr unwahrscheinlich ist, daß die Summe $x+y=0$ ist (in diesem Fall müßte sowohl x als auch y Null sein). Mit zunehmender Größe der Summe wird jedoch das Auftreten wahrscheinlicher bis es ein Plateau erreicht, das von dem Verhältnis a und b der zugrundeliegenden Zufallsvariablen abhängt. Für sehr große Werte der Summe tritt ein ähnlicher Fall wie für sehr kleine Werte auf, d. h. das Auftreten des Wertes $a+b$ ist ebenso unwahrscheinlich wie das Auftreten des Wertes 0 .

In der Physik wird oft nicht nur die Summe von *zwei* Zufallsvariablen, sondern (z. B. im thermodynamischen Grenzwert) die Summe von unendlich vielen Zufallsvariablen betrachtet. Dies ist z. B. der Fall bei thermischem Rauschen, dessen Spannung sich aus (fast) unendlich vielen Einzelspannungen zusammensetzt, die jeweils von einem Elektron erzeugt werden. Für die Verteilung eines derartigen Grenzwertes von Summen von Zufallsvariablen besagt der **zentrale Grenzwertsatz**, daß die Verteilungsfunktion die Form einer Normalverteilung aufweist:

Sei X_i , $i = 1, \dots \rightarrow \infty$ eine Folge von Zufallsvariablen mit identischen Verteilungsdichtefunktionen $f_i(x)$.

Ohne Beschränkung der Allgemeinheit sei $E(X_i) = 0$ und $E(X_i^2) = \sigma^2 < \infty$, d. h. die Form der Verteilungsfunktion ist irrelevant, solange ihre Varianz endlich ist. Dann gilt für die Folge des normierten Mittelwertes über n dieser Zufallsvariablen:

$$Z_n = \sum_{i=1}^n \frac{X_i}{\sqrt{n} \cdot \sigma} \tag{II.19}$$

$$f(Z_n) \rightarrow n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

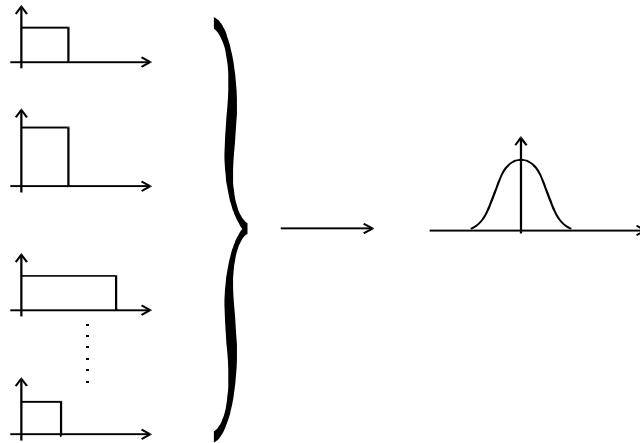


Abb. 2.10: Illustration des zentralen Grenzwertsatzes

Das heißt mit zunehmender Zahl der Zufallsvariablen, die aufaddiert werden, nähert sich die Verteilung der (normierten) Summe immer mehr einer Standard-Normalverteilung an. Im Grenzwert $n \rightarrow \infty$ ist die Summe beliebig verteilter Zufallsvariablen damit normalverteilt. Aus diesem Satz resultiert u. a. die Sonderstellung der Normalverteilung, daß sie für eine Vielzahl von physikalischen Phänomenen die „richtige“ statische Beschreibung ist: Wann immer sich eine Beobachtungsgröße als Summe von beliebigen unabhängigen Teil-Zufallsgrößen darstellen läßt, ist die resultierende Verteilungsfunktion eine Normalverteilung. Mathematisch sind für die Anwendbarkeit des zentralen Grenzwertsatzes eine Reihe einschränkender Annahmen notwendig, die in der Praxis jedoch (fast) immer erfüllt sind. Der Grund, wieso ausgerechnet die Normalverteilung als Grenzwert im zentralen Grenzwertsatz auftaucht, ist ihre Invarianz bei Faltung mit sich selbst bis auf den Faktor $\sqrt{2}$ im Argument: d. h. die Summe zweier normal verteilter Zufallsvariablen ist auch wieder normalverteilt:

$$e^{-x^2} * e^{-x^2} = e^{-\frac{x^2}{2}} \tag{II.20}$$

$$e^{-\left(\frac{f}{x}\right)^2} \cdot e^{-\left(\frac{f}{x}\right)^2} = e^{-2\left(\frac{f}{x}\right)^2}$$

Die Fourier-Transformierte der Verteilungsdichtefunktion wird auch als charakteristische Funktion bezeichnet. Sie wird oft als Hilfsmittel in der Statistik verwendet und dient in diesem Fall zum einfachen Nachweis, daß

die Summe zweier normalverteilter Zufallsvariablen wiederum normalverteilt ist. Wenn es also überhaupt eine Verteilungsfunktion gibt, die als Grenzwert der Summen-Verteilungsfunktion für $n \rightarrow \infty$ auftritt, so müßte diese Summen-Verteilungsfunktion auch für die zwei Teilsummen über $n/2$ mit $n/2 \rightarrow \infty$ ebenso gelten wie für die Gesamtsumme $n \rightarrow \infty$. D. h. eine notwendige Bedingung für die Verteilungsfunktion des Grenzwertes ist gerade, daß sie invariant ist gegenüber der Summenbildung von zwei Zufallsvariablen, die gemäß dieser Funktion verteilt sind.

II.3 Statistik

Bisher haben wir Modellannahmen über Zufallsvariablen und ihre Verteilungsfunktionen gemacht, bei denen die Art der Verteilungsfunktionen bekannt waren und insbesondere ihre Momente, d. h. Erwartungswert und Varianz als bekannt vorausgesetzt wurde. In der Praxis sind diese Verteilungsfunktionen und die sie bestimmenden Parameter unbekannt. Es können nur die Realisierungen der Zufallsvariablen beobachtet werden und daraus Schätzungen für die Parameter der zugrundeliegenden Verteilungsfunktionen abgeleitet werden. Mit dieser Schätzung von Parametern anhand von Beobachtungswerten beschäftigt sich die Statistik und die statistische Test-Theorie (vgl. auch Sachse, 1990).

Als Beispiel sei die **Schätzung des Erwartungswertes** μ vorgestellt, in dem über N Beobachtungen x_i gemittelt wird. Dabei wird vorausgesetzt, daß die x_i voneinander unabhängige Realisierungen der Zufallsvariable X sind, die jeweils den Erwartungswert μ und die Standardabweichung σ aufweisen:

$$\text{Mittelwert } M = \frac{1}{N} \cdot \sum_{i=1}^N x_i \quad (\text{II.21})$$

Der empirische Mittelwert M ist selber eine Zufallsvariable, für die ein Erwartungswert und eine Standardabweichung angegeben werden kann:

$$E(M) = \frac{1}{N} \cdot \sum_i E(x_i) = \frac{N}{N} \cdot \mu = \mu$$

$$E([M - \mu]^2) = E\left(\frac{1}{N^2} \left[\sum_{i,j} (x_i - \mu)(x_j - \mu) \right]\right) = \frac{1}{N^2} \cdot N \cdot \sigma^2 = \frac{\sigma^2}{N} \quad (\text{II.22})$$

Der Erwartungswert des Mittelwerts ist damit gleich dem Erwartungswert der zugrundeliegenden Zufallsvariable, so daß der empirische Mittelwert M ein erwartungstreuer Schätzer für μ ist. Mit zunehmender Stichprobengröße N geht die relative Schwankung von M gegen 0 mit $1/N$, d. h. die Standardabweichung des Mittelwerts aus N unabhängigen Meßwerten desselben Vorgangs nimmt mit $\frac{1}{\sqrt{N}}$ ab. Die zugrundeliegende normierte

Verteilungsfunktion des Mittelwertes wird also immer schärfer um den Erwartungswert konzentriert. Dieser schärfer werdenden Verteilung liegt das **Gesetz der großen Zahlen** zugrunde, das für $N \rightarrow \infty$ besagt, daß die Verteilung des Mittelwertes nur noch beim Erwartungswert Wahrscheinlichkeiten ungleich 0 aufweist. Dieses Gesetz der großen Zahlen spielt in der Thermodynamik beim thermodynamischen Limes eine große Rolle. Dort sind die makroskopischen Systeme aus einer gegen ∞ gehenden Zahl von mikroskopischen Systemen zusammengesetzt, die eine relativ breite Verteilungsfunktion für ihre jeweiligen Zustände aufweisen. Durch Mittelwertbildung über die große Anzahl wird die Zustandsdichte um den Erwartungswert so scharf, daß im makroskopischen System praktisch nur noch ein beobachtbarer Zustand auftreten kann, der dem Erwartungswert der mikroskopischen Systeme entspricht.

Ein weiteres Beispiel ist die **Schätzung der Standardabweichung** σ durch Berechnung der **empirischen Standardabweichung** S :

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - M)^2 \quad (\text{II.23})$$

Die empirische Standardabweichung besitzt demnach eine $N-1$ -Gewichtung bei der Berechnung des mittleren quadratischen Fehlers. Diese Gewichtung ist notwendig, um die Erwartungstreue von S^2 für die zu schätzende Standardabweichung σ^2 zu gewährleisten: Ohne Beschränkung der Allgemeinheit sei $\mu = 0$ (andernfalls würde x_i durch $x_i - \mu$ und M durch $M - \mu$ ersetzt werden). Der Erwartungswert von S^2 ist dann gleich:

$$\begin{aligned}
E(S^2) &= \frac{1}{N-1} \sum_{i,j=1}^N E(x_i \cdot x_j) - \frac{2}{N-1} \sum_i E(x_i \cdot M) + \frac{N}{N-1} E(M^2), \\
&\text{und da } M^2 = \frac{1}{N} \sum_j x_j \cdot \frac{1}{N} \sum_i x_i \\
&= \left(\frac{1}{N-1} - \frac{2}{N(N-1)} + \frac{1}{N(N-1)} \right) \cdot \underbrace{\sum_{i,j} E(x_i \cdot x_j)}_{\sum_i E(x_i^2), \text{ da } x_i \text{ und } x_j \text{ unabhängig}} \\
&= \frac{N-1}{N(N-1)} \cdot (N \cdot E(x_i^2)) = \sigma^2 \tag{II.24}
\end{aligned}$$

D. h. S^2 ist erwartungstreu. Falls in der Formel für die empirische Standardabweichung anstelle von M der „wahre“ Erwartungswert μ eingesetzt würde (der z. B. durch die Konstruktion der Messung schon bekannt sein könnte), muß anstelle der $N-1$ -Gewichtung die N -Gewichtung benutzt werden, da die Zahl der Freiheitsgrade nicht mehr $N - 1$, sondern N ist.

Es stellt sich nun die Frage, welche Verteilungsfunktionen für die Schätzwerte M bzw. S^2 auftreten. Dieses läßt sich am einfachsten dann angeben, wenn die x_i normalverteilt sind, weil die resultierenden Verteilungen dann bestens bekannt und tabelliert sind: Für das Quadrat einer normalverteilten Zufallsvariable x_i^2 tritt die χ^2 -Verteilung auf, die für eine unterschiedliche Zahl von Freiheitsgraden tabelliert ist. Das Verhältnis aus zwei χ^2 -verteilten Variablen, d. h. die Schätzgröße der Form x_i^2/x_j^2 ist F -verteilt. Die um die empirische Standardabweichung normierte Testgröße

$\frac{x_i}{\sqrt{\sum_j \frac{1}{N-1} x_j^2}}$ ist t -verteilt. Weitere Einzelheiten über diese Verteilungs-

funktionen sowie Tabellen der Funktionen für bestimmte Perzentilen (d. h. Argumente der Verteilungsfunktion, die zu einem bestimmten Wahrscheinlichkeitswert z. B. von 95 % gehören) sind z. B. bei Sachse, 1986 nachzulesen.

Diese o. a. Verteilungsfunktionen werden bei einem **statistischen Test** verwendet, um die Signifikanz eines Schätzwertes nachzuweisen. Bei einem derartigen Test wird eine Nullhypothese H_0 aufgestellt, die besagt, daß die Prüfvariable X einer bestimmten Verteilungsfunktion $P_0(x)$ entstammt. Beispielsweise soll nachgewiesen werden, daß die männlichen Studenten der Universität A signifikant größer als die Studenten der Universität B sind. Die Prüfgröße wäre dann die Größendifferenz zwischen jeweils einem Mitglied der Stichprobe männlicher Studenten aus der Universität A und aus der Universität B, die im Falle der Nullhypothese

H_0 („Die Studenten an beiden Universitäten sind gleich groß“) den Mittelwert 0 und eine Normalverteilung aufweisen würde.

Die Alternativhypothese H_1 ist so zu konstruieren, daß sie automatisch Gültigkeit besitzt, falls H_0 verworfen werden muß, d. h. sie besagt, daß die Prüfvariable X einer anderen Verteilungsfunktion entstammt als der Verteilungsfunktion $P_0(X)$. Das Ziel des Testes ist es nun, die Nullhypothese abzulehnen und damit bei einem bestimmten Signifikanzniveau die Alternativhypothese annehmen zu müssen. Dabei kann der **Fehler erster Art** gemacht werden. Er bezeichnet die Wahrscheinlichkeit, daß H_0 fälschlicherweise abgelehnt wird. Dieser Fehler hängt mit dem „Signifikanzniveau“ zusammen. Es kann aber auch der **Fehler zweiter Art** auftreten, mit dem die Wahrscheinlichkeit bezeichnet wird, H_1 fälschlicherweise abzulehnen. Als Beispiel seien N unabhängige Meßwerte einer physikalischen Größe aufgeführt (z. B. e/M). Um die Frage zu beantworten, ob diese Größe vom Literaturwert x_0 signifikant abweicht (unter der Annahme einer Normalverteilung der N Meßwerte um ihren Erwartungswert) bildet man also die Prüfgröße

$$X = \frac{\sum_i (x_i - x_0) / N}{\sqrt{\sum_i \frac{1}{N-1} (x_i - \bar{x})^2}} \quad (\text{II.25})$$

Diese Größe X ist unter H_0 t-verteilt mit dem Freiheitsgrad $N - 1$

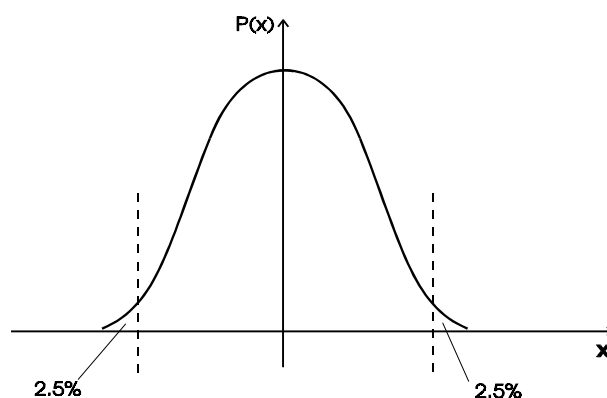


Abb. 2.11: t-Verteilung mit Grenzen bei -2,5 % und +2,5%

In der Tabelle für die t-Verteilung mit passender Anzahl von Freiheitsgraden wird nun derjenige t-Wert t_0 gesucht, der mit 5% (oder einem anderen Signifikanzniveau) über- bzw. unterschritten wird. Falls $|X| \geq t_0$ ist, wird die Nullhypothese abgelehnt, d. h. eine signifikante Abweichung (mit Irrtumswahrscheinlichkeit z. B. 5 %) zwischen dem Literaturwert und den

beobachteten Meßgrößen tritt auf. Ist dagegen $|X| \leq t_0$, ist keine Ablehnung der Nullhypothese möglich und es muß davon ausgegangen werden, daß die vorliegenden Meßwerte keine signifikante Abweichung zu dem Literaturwert aufweisen. Diese Situation mag zwar unbefriedigend sein, wenn z. B. Hinweise auf eine Diskrepanz zwischen den Meßwerten und den Literaturdaten auftreten. Diese Diskrepanz kann aber mit statistischen Mitteln nur bei einer hinreichend großen Stichprobenzahl, bei kleiner Varianz innerhalb der Stichprobe und bei entsprechend großem Effekt der Abweichung zwischen Meßwert und erwarteten Literaturwert verbessert werden.

II.4 Stochastischer Prozeß

Rauschen in physikalischen Systemen läßt sich als stochastischer Prozeß beschreiben, mit dem eine „Familie“ oder ein Ensemble von Zeitfunktionen $X(t)$ bezeichnet wird. $X(t)$ ist dabei für jedes feste $t \in \mathbb{R}$ eine Zufallsvariable, für jede Realisierung der Zufallsvariable eine bestimmte Zeitfunktion, und für jedes feste t und eine feste Realisierung eine bestimmte Zahl aus \mathbb{R} .

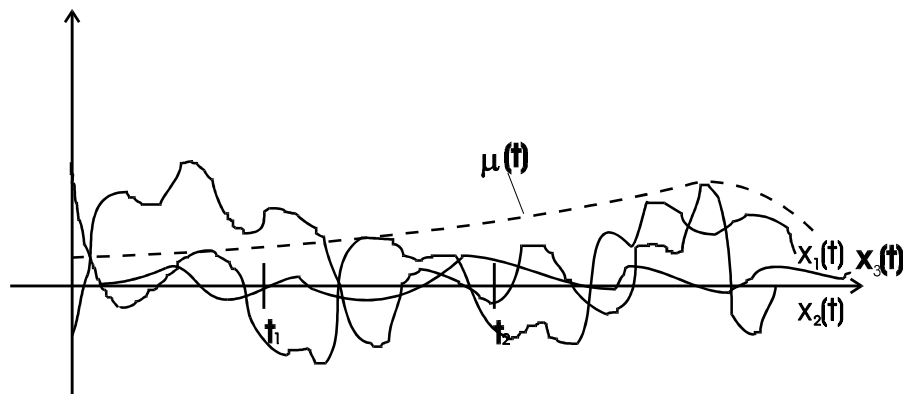


Abb. 2.12: Unterschiedliche Realisierungen desselben stochastischen Prozesses

Für einen festen Zeitpunkt t wird die Zufallsvariable $X(t)$ durch die Verteilungsdichtefunktion $f(x,t)$ beschrieben. Mit ihr läßt sich der Erwartungswert $\mu(t)$ ausdrücken als

$$\mu(t) = E[X(t)] = \int_{-\infty}^{\infty} x \cdot f(x,t) dx \quad (\text{II.26})$$

Dabei ist der Erwartungswert $\mu(t)$ selbst eine Funktion der Zeit. In ähnlicher Weise lassen sich auch die höheren Momente (z. B. die Varianz als Maß für die Leistung des Prozesses) ausdrücken. Um auf die spektrale

Verteilung schließen zu können, wird die **Autokorrelationsfunktion** $\varphi(\mathbf{t}_1, \mathbf{t}_2)$ über die Verbundverteilungsdichtefunktion $f(x_1, x_2; t_1, t_2)$ definiert. Diese Verbundverteilungsdichtefunktion bezeichnet die Wahrscheinlichkeitsdichte, daß zum Zeitpunkt t_1 der Wert x_1 und zum Zeitpunkt t_2 der Wert x_2 auftritt:

$$\varphi(t_1, t_2) = E[x(t_1) \cdot x(t_2)] = \int_{-\infty}^{\infty} x_1 \cdot x_2 \cdot f(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (\text{II.27})$$

Für die meisten der in der Praxis vorkommenden stochastischen Prozesse können die folgenden vereinfachenden Annahmen gemacht werden:

$X(t)$ ist **stationär** falls $X(t)$ und $X(t + \varepsilon)$ dieselbe Statistik für alle ε aufweisen. Diese Stationarität wird auch als „**streng stationär**“ bezeichnet, weil sämtliche statistischen Eigenschaften identisch für alle Zeitpunkte sein müssen. $X(t)$ ist dagegen „**schwach stationär**“, falls der Mittelwert $\mu(t) = \mu = \text{konstant}$ ist, also nicht von der Zeit abhängt, und außerdem die Autokorrelationsfunktion $\varphi(t_1, t_2) = \varphi(t_1 - t_2)$ ist. Damit hängt die Autokorrelationsfunktion nur von dem Zeitunterschied zwischen den zwei ausgewählten Zeitpunkten ab, nicht jedoch von der wirklichen Lage dieser Zeitpunkte (d. h. Zeitinvarianz der Autokorrelationsfunktion).

Eine weitere wichtige Eigenschaft von $X(t)$ ist die Ergodizität: $X(t)$ heißt **ergodisch** wenn Ensemble-Mittelwerte durch Zeitmittelwerte (jeweils eines Repräsentanten von $X(t)$) ersetzt werden können. Im Gegensatz zu der oben (Glg. II.26) definierten Berechnung des Erwartungswertes (Glg. II. 27) und der Autokorrelationsfunktion können diese dann geschrieben werden als

$$\mu(t) = E[x(t)] = \int_{-\infty}^{\infty} x \cdot f(x, t) dx = \int_{-\infty}^{\infty} x(t) dt \quad (\text{II.28})$$

$$\varphi(\tau = t_1 - t_2) = E[x(t)x(t + \tau)] = \int_{-\infty}^{\infty} x(t)x(t + \tau) dt \quad (\text{II.29})$$

Bei einem ergodischen Prozeß ist die gesamte „Statistik“ bereits in einem Zeitsignal vorhanden, so daß die zeitliche Mittelung an einer einzigen Realisierung dieses Prozesses (z. B. an einer einzigen Meßapparatur), die Mittelung über mehrere Realisierungen dieses Prozesses (z. B. über mehrere gleichartige Apparaturen) ersetzen kann. Mathematisch ist diese Ei-

genschaft der Ergodizität relativ schwer handhabbar, in der Praxis ist jedoch fast immer der beobachtete stochastische Prozeß ergodisch.

Es stellt sich nun die Frage, wie das Leistungsspektrum eines stationären, stochastischen Prozesses berechnet, bzw. geschätzt werden kann. Gemäß des Wiener-Kchintchine-Theorems kann das Leistungsspektrum aus der Autokorrelationsfunktion durch Fouriertransformation gebildet werden:

$$S^2(f) = \mathfrak{F}[\varphi(t)] = \mathfrak{F}[E(x(t') \cdot x(t'+t))] \quad (II.30)$$

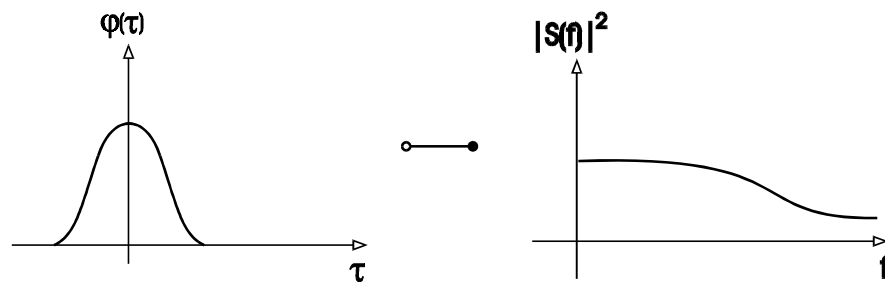


Abb. 2.13: Leistungsspektrum des stationären stochastischen Prozesses

Damit ist $S^2(f)$ die Fourier-Transformierte eines Erwartungswertes. Das Leistungsspektrum des stochastischen Prozesses ist daher entweder von vornherein bekannt oder kann aufgrund einer Realisierung dieses Prozesses nur geschätzt werden. Es kann aber höchstens im Grenzwert für unendlich aufwendige Mittelung exakt bestimmt werden. In der Praxis ist es daher ein Problem, daß nur eine (oder wenige) Realisierungen von $X(t)$ vorliegen, so daß dieser Schätzwert des Leistungsspektrums eine Zufallsvariable darstellt. Speziell für den ergodischen, stationären Prozeß gilt

$$x_T(t) = x(t) \cdot \text{rect}_T(t) \quad \circ - \bullet \quad X_T(f) \neq S^2(f) = \lim_{T \rightarrow \infty} [E(X_T^2(f))] \quad (II.31)$$

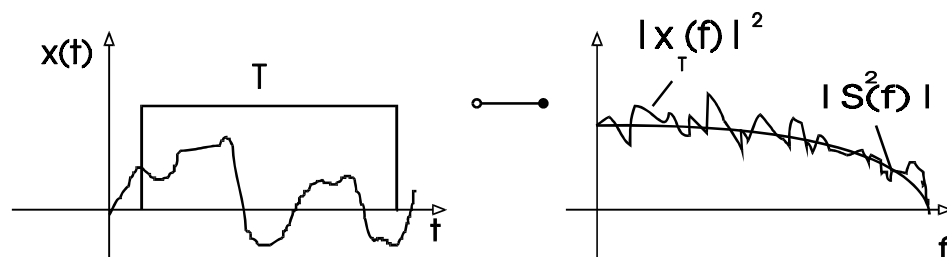


Abb. 2.14: Ergodischer stationärer Prozeß

Wenn aus dem zeitbegrenzten (d. h. mit einem Rechteckfenster gefensterten) Prozeß das Leistungsspektrum $X_T(f)^2$ errechnet wird, stellt

dieses nicht das gesuchte Leistungsspektrum des gesamten Prozesses $S^2(f)$ dar, was nur als Grenzwert für $T \rightarrow \infty$ des Erwartungswertes auftritt. Es müßten also sowohl der Grenzübergang $T \rightarrow \infty$ als auch die Mittelung über das gesamte Ensemble durchgeführt werden. Einen „glatteren“ Schätzwert des Leistungsspektrums erreicht man durch Mittelung von $X_T^2(f)$ über mehrere Zeitfenster, d. h. man ermittelt das „mittlere“ Leistungsspektrum:

$$\bar{X}_T^2(f) = \sum_{i=1}^n \frac{1}{n} X_T^2(f) \rightarrow |S(f) * \text{sinc}(fT)|^2 \quad (\text{II.32})$$

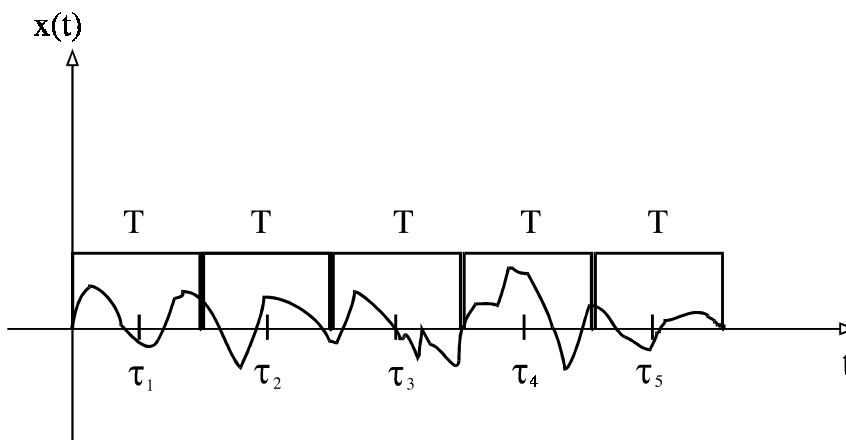


Abb. 2.15: Mittleres Leistungsspektrum

Das geschätzte Leistungsspektrum kann daher geglättet werden, in dem der Schätzwert aus mehreren Zeitfenstern der Länge T gemittelt wird. Durch die zeitliche Begrenzung auf die Analyse-Fensterlänge T wird zwar die spektrale Auflösung geringer, dafür steigt die Genauigkeit bei der Schätzung des Mittelwerts. Würde man anstelle über n Intervalle der Länge T eine Analyse über ein einziges Zeitfenster der Länge nT durchführen, würde zwar die spektrale Auflösung höher werden, das resultierende (geschätzte) Leistungsspektrum würde jedoch eine sehr hohe Schwankung der einzelnen Spektralwerte aufweisen, da jeder Wert dieses Leistungsspektrums wieder eine Zufallsvariable darstellt. Daher ist es in der Praxis wichtig, das Spektrum eines stochastischen Prozesses durch Mittelung über mehrere Zeitfenster mit einem möglichst „glattem“ Schätzwert zu schätzen.

II.5 Rauschen in physikalischen Systemen

Nachdem wir nun die Eigenschaften und die prinzipielle Beschreibungsweise eines stochastischen Prozesses kennengelernt haben, wenden wir uns den für die physikalische Meßtechnik relevanten Fällen von Rauschen

in physikalischen Systemen zu. Dabei handelt es sich um stationäre, ergodische Prozesse, aus deren Autokorrelationsfunktion auf das Leistungsspektrum geschlossen werden kann. Eine detaillierte Darstellung findet sich bei Müller, R.: Rauschen (Springer-Verlag, 1990).

Beim **thermischen Rauschen** oder Widerstandsrauschen ist die Ursache die thermische Bewegung von Ladungsträgern in einem Leiter bzw. Halbleiter, den wir uns mit idealisierten Zuleitungen wie folgt vorstellen können:

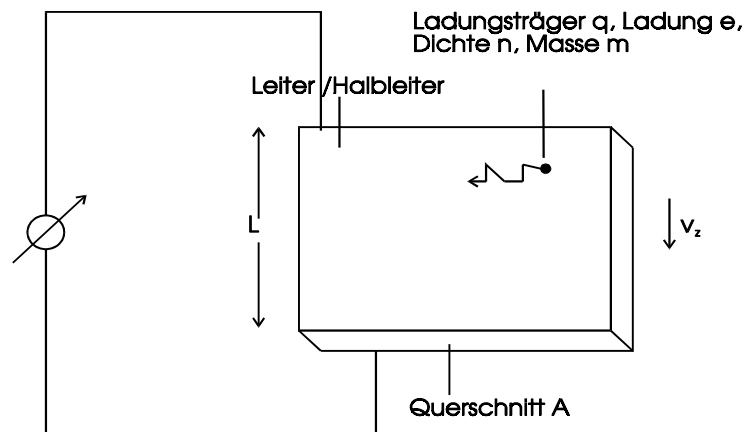


Abb. 2.16: Leiter der Länge L und Querschnitt A

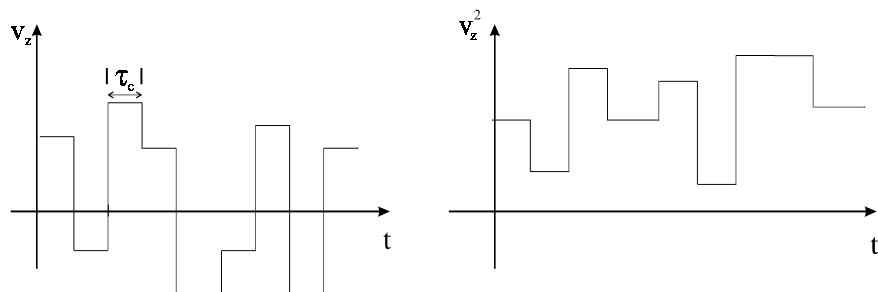


Abb. 2.17: z-Komponente der Bewegung der Ladungsträger

In dem Leiter, den man sich mikroskopisch als Metall-Kristall bzw. Halbleiter-Kristall mit einem „Gas“ von Ladungsträgern vorstellen kann, bewegen sich diese Ladungsträger der Ladung e und der Dichte n (Anzahl geteilt durch Volumen) in einer statistischen Weise. Wenn hier nur die z-Komponente der Geschwindigkeit des Ladungsträgers q betrachtet wird, so besteht die Modellvorstellung darin, daß mit dem mittleren zeitlichen Abstand τ_c jeder Ladungsträger Stöße mit dem Gitter ausführt und seine jeweilige Geschwindigkeit ändert. Insgesamt findet damit eine Art Diffusionsprozeß statt, der im Fall einer äußeren Spannung von 0 den Mittelwert $\overline{v_z} = 0$ aufweist. Der von der Bewegung dieses Ladungsträgers in den

Zuleitungen verursachte Strom beträgt damit

$$i_q = \frac{e}{L} \cdot v_{zq} \quad (\text{II.33})$$

Dabei bezeichnet L die Länge des Leiters/Halbleiters und v_{zq} die jeweilige z -Komponente der Geschwindigkeit des Teilchens q . Der vom Teilchen q verursachte Strom integriert sich daher im zeitlichen Mittel zu 0 auf. Die von diesem Strom verursachte Leistung ist jedoch proportional zum mittleren Quadrat des Stroms (d. h. Mittelwert einer positiven Größe), so daß sie sich errechnet zu:

$$\bar{i}_q = 0, \quad \bar{i}_q^2 = \left(\frac{e}{L}\right)^2 \cdot v_z^2 \neq 0 \quad (\text{II.34})$$

Diese nicht verschwindende Leistung des Stroms, der durch die innere thermische Bewegung der Ladungsträger im Leiter in den Zuführungen verursacht wird, stellt somit die Quelle des Rauschens dar, das an einem Widerstand gemessen werden kann. Für den insgesamt am Widerstand gemessenen Strom i ergibt sich damit die Gesamtleistung:

$$\begin{aligned} \bar{i}^2 &= \sum_q \bar{i}_q^2 = A \cdot L \cdot n \cdot \bar{i}_q^2 = e^2 \cdot \frac{An}{L} v_z^2 \\ &= e^2 \cdot \frac{A \cdot n}{L} \frac{kT}{m}, \quad \text{da } \frac{1}{2} m v_z^2 = \frac{k \cdot T}{2} \end{aligned} \quad (\text{II.35})$$

Hierbei bezeichnet A den Querschnitt des Leiters, so daß $A \cdot L \cdot n$ die Gesamtzahl der im Leiter vorhandenen Ladungsträger darstellt. Die in der letzten Zeile gemachte Substitution von $\frac{1}{2} m v_z^2 = \frac{kT}{2}$ resultiert aus dem Gleich-

verteilungssatz der statistischen Mechanik, in dem die kinetische Energie für jeden Freiheitsgrad die Größe $\frac{1}{2} \cdot k \cdot T$ annimmt. Hierbei ist k die Boltzmannkonstante und T die absolute Temperatur, sowie m die Masse des Ladungsträgers. Mit dieser Gleichung ist daher die Leistung des Gesamt-Rauschens direkt mit der mittleren thermischen Energie verknüpft, die wiederum proportional zur (absoluten) Temperatur ist.

Um über die spektrale Leistungsdichte des thermischen Rauschens nun eine Aussage treffen zu können, müssen wir zunächst die **Autokorrelationsfunktion** $\phi(\tau)$ berechnen: Sie läßt sich aus der Idealisierung herleiten, daß die Geschwindigkeit jedes Ladungsträgers für die mittlere freie Zeit

zwischen zwei Stößen τ_c konstant bzw. determiniert ist, während die Geschwindigkeit nach einem Stoß statistisch unabhängig von der Geschwindigkeit vor einem Stoß ist. Daher ist bei der Autokorrelationsfunktion zu erwarten, daß für Verzögerungszeiten $> \tau_c$ die Autokorrelationsfunktion verschwinden muß. Sie hat dagegen ein Maximum für $\tau = 0$, das gerade der Gesamtleistung $\overline{i^2}$ entsprechen muß. Unter der Annahme, daß v_z bzw. i_q jeweils während eines Zeitintervalls τ_c konstant ist, ergibt sich die Autokorrelationsfunktion dann als die Faltung zweier Rechteckfunktionen der Breite τ_c :

$$\varphi(\tau) = \frac{\overline{i^2}}{\tau_c} \cdot \text{rect}\left(\frac{t_c}{\tau_c}\right) * \text{rect}\left(\frac{t}{\tau_c}\right) = \begin{cases} \overline{i^2} \cdot \left|1 - \frac{\tau}{\tau_c}\right| & \text{für } |\tau| \leq \tau_c \\ 0 & \text{für } |\tau| > \tau_c \end{cases} \quad (\text{II.36})$$

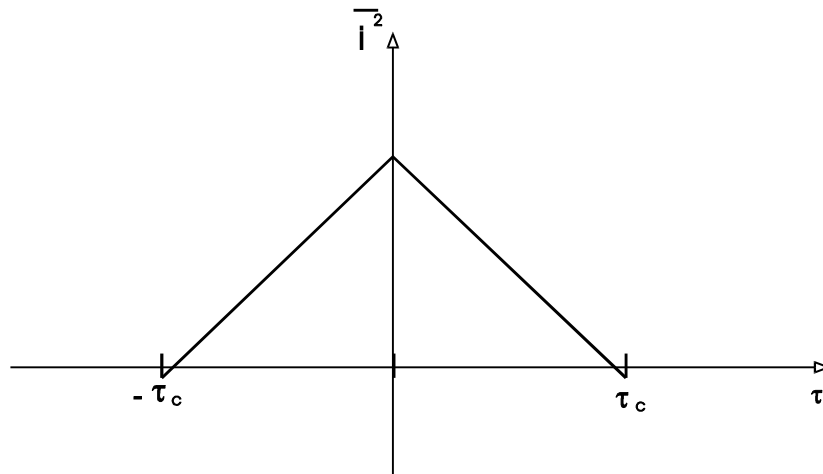
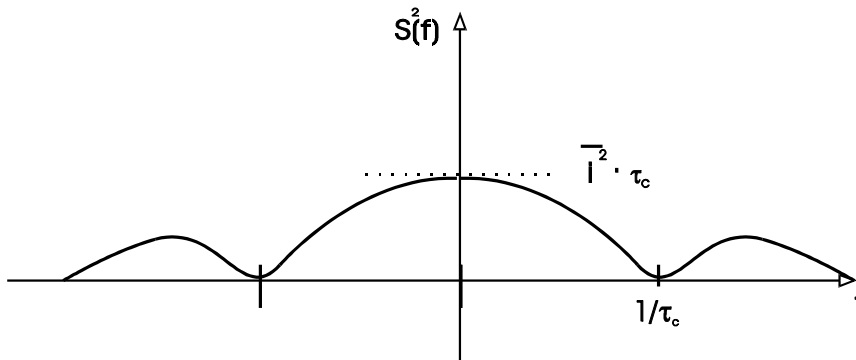


Abb. 2.18: Autokorrelationsfunktion von i

Selbst wenn die Geschwindigkeit zwischen aufeinanderfolgenden Stößen nicht konstant ist oder die Zeit zwischen diesen Stößen von dem Erwartungswert τ_c aufgrund von Fluktuationen abweicht, bleibt die generelle Form der Autokorrelationsfunktion weitgehend erhalten, so daß die nachfolgenden Betrachtungen weiter ihre Gültigkeit behalten.

Das Leistungsspektrum $S^2(f)$ berechnet sich als Fouriertransformierte der Autokorrelationsfunktion zu :

$$\Rightarrow S^2(f) = \bar{i}^2 \cdot \tau_c \cdot \text{sinc}^2(f\tau_c) \approx \begin{cases} f^0 & \text{für } f \ll \frac{1}{\tau_c} \\ \frac{1}{f^2} & \text{für } f \gg \frac{1}{\tau_c} \end{cases} \quad (\text{II.37})$$

Abb. 2.19: Leistungsspektrum $S^2(f)$

Für Frequenzen weit unterhalb der Grenzfrequenz $1/\tau_c$ ist die spektrale Leistungsdichte daher konstant, so daß sich ein weißes Rauschen ergibt. Für Frequenzen weit oberhalb der Grenzfrequenz fällt die spektrale Leistungsdichte dagegen mit $1/f^2$ ab. Damit ist eine weitere wichtige Eigenschaft des thermischen Rauschen, nämlich die Eigenschaft eines weißen Spektrums für Frequenzen unterhalb einer gewissen Grenzfrequenz gezeigt. Diese Eigenschaft gilt auch, wenn die Autokorrelationsfunktion nicht exakt die o. a. Form aufweist, wobei allerdings eine andere Form des Leistungsspektrums für höhere Frequenzen auftritt.

Es bleibt noch zu klären, wie die errechnete Rauschleistung von dem Ohmschen Widerstand R des betrachteten Leiters bzw. Halbleiters abhängt. Dazu wird die mittlere Geschwindigkeitsverteilung in z -Richtung und der daraus resultierende statische Strom \bar{i}_z für den Fall betrachtet, daß ein elektrisches Feld der Feldstärke E an die äußeren Anschlüsse des Leiters bzw. Halbleiters angelegt wird. Dabei tritt eine Beschleunigung der Ladungsträger in z -Richtung auf, für die aufgrund der Gleichheit der Kräfte gelten muß:

$$\frac{dv_z}{dt} = -\frac{e \cdot E}{m} \quad (\text{II.38})$$

Während jedes „freien“ Zeitintervalls der Länge τ_c führt der Ladungsträger daher eine gleichmäßig beschleunigte Bewegung in z -Richtung auf, d. h.

die zusätzliche Geschwindigkeit in z-Richtung steigt von jeweils 0 bis $v_{z,\max} = -\frac{e \cdot E}{m} \cdot \tau_c$. Die mittlere Geschwindigkeit ergibt sich demnach zu:

$$\bar{v}_z = -\frac{e \cdot E}{m} \cdot \frac{\tau_c}{2} \quad (\text{II.39})$$

Der mittlere Gleichstrom in z-Richtung I_z errechnet sich dann zu:

$$I_z = A \cdot n \cdot e \cdot \bar{v}_z \quad (\text{II.40})$$

Zusammen mit der anliegenden Spannung $U_z = E \cdot L$ folgt daraus für den ohmschen Widerstand (Leitwert) des Leiters:

$$\frac{1}{R} = \frac{I_z}{U_z} = \frac{e^2}{m} \cdot \frac{A \cdot n}{L} \cdot \frac{\tau_c}{2} \quad (\text{II.41})$$

Setzt man nun in der Formel (II.35) für die Gesamt-Leistung des thermischen Wechselstroms \bar{i}^2 die Größe von R ein, so ergibt sich:

$$\bar{i}^2 = \frac{1}{R} \cdot \frac{2m}{\tau_c} \cdot v_z^2 = \frac{2kT}{R \cdot \tau_c} \quad (\text{II.42})$$

Als spektrale Leistungsdichte ergibt sich dann für tiefe Frequenzen:

$$P = \frac{U^2}{R} = I^2 \cdot R \Rightarrow S_i^2(f) \approx \frac{2k \cdot T}{R} \quad (\text{II.43})$$

$$S_u^2(f) \approx 2k \cdot T \cdot R$$

$S_i^2(f)$ bezeichnet damit die spektrale Leistungsdichte des Stromrauschens, während $S_u^2(f)$ das Leistungsspektrum des Spannungsrauschens bezeichnet, das bei Messen der Rauschspannung am Widerstand auftreten würde. Insgesamt tritt also die Rauschleistungsdichte P auftritt:

$$P = \frac{U^2}{R} = I^2 \cdot R, \quad P = 2k \cdot T \quad (\text{II.44})$$

Das Widerstandsrauschen ist damit ein weißes Rauschen mit einer konstanten Leistungsdichte, die proportional zu $k \cdot T$ ist. Diese Näherung gilt für sämtliche Frequenzen des technisch nutzbaren Bereichs der elektromagnetischen Wellen, da die Grenzfrequenz selbst für eine Temperatur $T = 4\text{k}$ im Infrarot liegen würde. Es ist jedoch zu beachten, daß für hohe Spannungen („heiße Elektronen“) die Annahme der thermischen Gleichverteilung der Energie nicht gilt, so daß man zu einer anderen Berechnung dabei auftretenden Rauschphänomene zurückgreifen muß.

Ein anderer Entstehungsmechanismus, der aber zu einer ähnlichen spektralen Verteilung der Energie wie für das thermische Rauschen führt, liegt dem **Schrotrauschen** zugrunde. Hervorgerufen wird das Schrotrauschen (im Englischen: shot noise) durch einen Poissonprozess, d. h. eine statistische Verteilung von Stromstößen kurzer Dauer, die als „schroteln“ oder „shots“ bezeichnet werden.

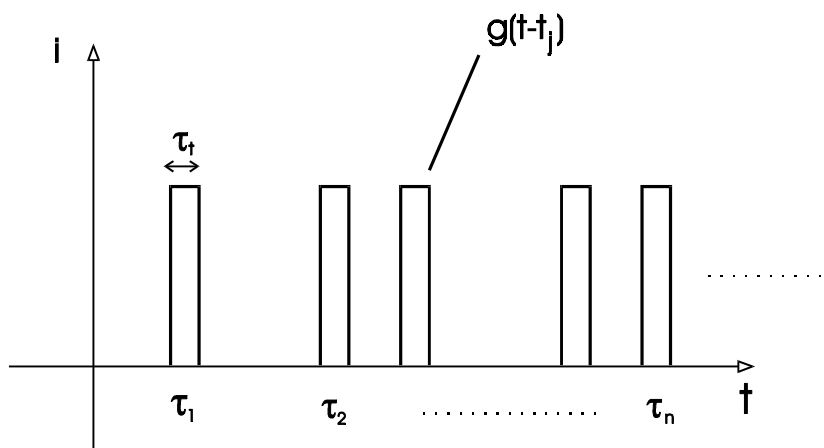


Abb. 2.20: Schrotrauschen

Ein Beispiel für diese Art des Rauschens stellt der Stromfluß in einer Hochvakuumdiode dar, bei der jedes aus der Glühkatode austretende Elektron durch die zwischen Kathode und Anode anliegende Beschleunigungsspannung praktisch sofort zur Anode „gesaugt“ wird.

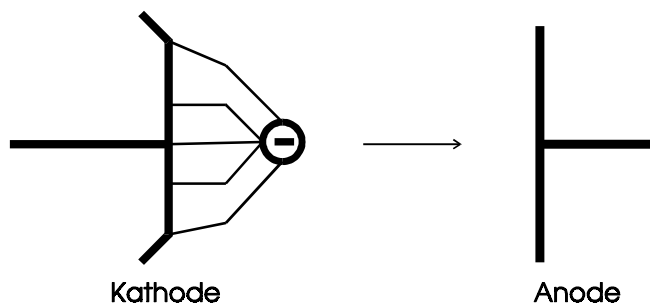


Abb. 2.21: Kathode und Anode

Jedes einzelne Elektron der Ladung e erzeugt damit während der Flugzeit τ_i einen Stromstoß $g(t)$, der als rechteckförmig mit der Dauer τ_i und dem zeitlichen Integral e angenommen wird:

$$g(t) = \frac{e}{\tau_i} \cdot \text{rect}\left(\frac{t}{\tau_i}\right) \quad (\text{II.45})$$

Diese Stromstöße treten nun zu statistisch verteilten Zeitpunkten t_j auf, die durch einen Poisson-Prozeß mit der Auftretenshäufigkeit λ charakterisiert werden. Der Strom $i(t)$ kann daher als Summe dieser statistisch verteilten Stromstöße aufgefaßt werden:

$$i(t) = \sum_j g(t - t_j) \quad (\text{II.46})$$

Der mittlere Strom (Gleichstrom-Komponente) i_0 errechnet sich dann als zeitlicher Mittelwert (Erwartungswert des ergodischen Prozesses) über diesen Strom $i(t)$:

$$I_0 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T i(t) dt = \sum_j \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T g(t - t_j) dt = \lambda \int_{-\infty}^{\infty} g(t) dt = \lambda \cdot e \quad (\text{II.47})$$

Dabei bezeichnet λ die Häufigkeit der Impulse und

$$\bar{g} = \int_{-\infty}^{\infty} g(t) dt = e$$

den zeitlichen Mittelwert des Stromstoßes.

Um das Leistungsspektrum ausrechnen zu können, muß wiederum die Autokorrelationsfunktion des Schrotrauschens bestimmt werden:

$$\begin{aligned}
\varphi_{ii}^p(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T i(t)i(t+\tau)dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \sum_j g(t-t_j) \cdot \sum_k g(t-t_k + \tau)dt \\
&= \lim_{T \rightarrow \infty} \frac{1}{2T} \left(\sum_j \int_{-T}^T g(t-t_j)g(t-t_j + \tau)dt \right) + \\
&\quad + \lim_{T \rightarrow \infty} \frac{1}{2T} \underbrace{\left(\sum_{j \neq k} \int_{-T}^T g(t-t_j)g(t-t_k + \tau)d\tau \right)}_{\lim_{T \rightarrow \infty} \frac{1}{2T} \left(\sum_j \left(\int_{-T}^T g(t-t_j)dt \right)^2 \right)} = \lambda \cdot \varphi_{gg}(\tau) + (\lambda \cdot \bar{g})^2 \quad (II.48)
\end{aligned}$$

Dabei wurde ausgenutzt, daß die Doppelsumme in einen Anteil mit $j = k$ und eine Summe $j \neq k$ aufgespalten werden kann. Während die erste Summe die Autokorrelationsfunktion eines einzelnen Stromstoßes $\varphi_{gg}(\tau)$ angibt, beschreibt der Summand der zweiten Doppelsumme das Produkt zweier statistisch unabhängiger Prozesse, die auch als Produkt der Erwartungswerte beschrieben werden können.

Dieser letzte Ausdruck ist eine Konstante und ist gleich dem Gleichstromquadrat, während die Autokorrelationsfunktion des einzelnen Strompulses genau wie im Fall des thermischen Rauschens eine Dreiecksfunktion ergibt:

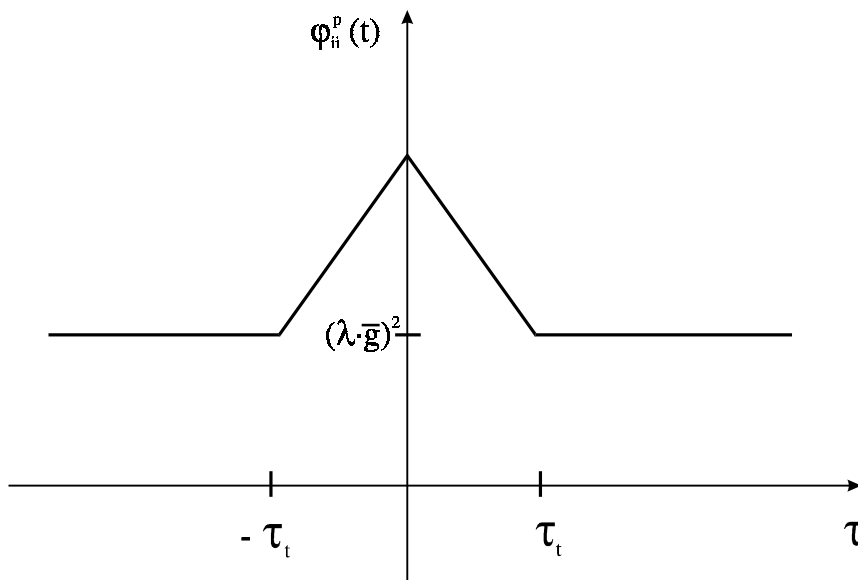


Abb. 2.22: Autokorrelationsfunktion des Schrotrauschens

$$\varphi_{gg}(\tau) = \begin{cases} \frac{e^2}{\tau_t} \left(1 - \frac{|\tau|}{\tau_t} \right) & \text{für } |\tau| \leq \tau_t \\ 0 & \text{für } |\tau| > \tau_t \end{cases}$$

Die Gesamtleistung des Schrotrauschens ergibt sich als:

$$\overline{I^2} = \varphi_{ii}^p(0) = \lambda \cdot \frac{e^2}{\tau_t} + I_0^2 = I_0 \cdot \frac{e}{\tau_t} + I_0^2 \quad (\text{II.49})$$

Der erste Term beschreibt die Leistung des Rauschens, also des Wechselspannungsanteils, der mit kleiner werdendem τ_t ansteigt. Der zweite Term beschreibt die oben bereits berechnete Leistung des Gleichstromes. Für das Leistungsspektrum des Wechselanteils (Rauschanteils) ergibt sich damit:

$$S^2(f) = \mathfrak{F}(\varphi_{gg}(\tau) \cdot \lambda) = e \cdot I_0 \cdot \sin^2(f \cdot \tau_t) \quad (\text{II.50})$$

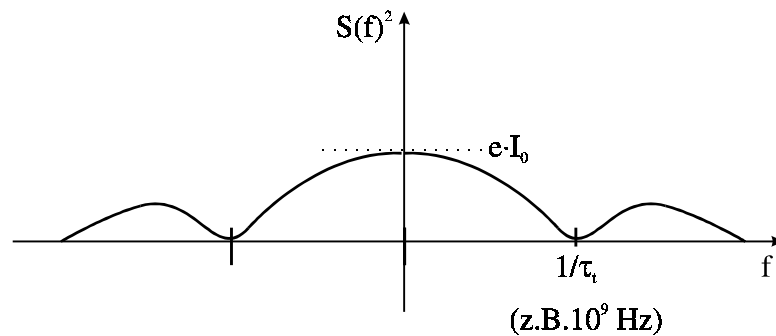


Abb. 2.23: Leistungsspektrum des Rauschanteils

Genau wie für das thermische Rauschen ergibt sich eine konstante Leistungsdichte für $f \ll \frac{1}{\tau_t}$ und eine mit $\frac{1}{f^2}$ abfallende Leistungsdichte für

$f \gg \frac{1}{\tau_t}$. Aufgrund der makroskopischen Abmessungen der „freien“ Weg-

strecke des Elektrons (Abstand Katode-Anode) ist die Zeitkonstante τ_t wesentlich größer als für das thermische Rauschen, bei dem die freie Laufzeit durch die mikroskopischen Dimensionen der mittleren freien Weglänge der Ladungsträger gegeben war. Daher hat das Schrotrauschen in der Regel eine wesentlich niedrigere Grenzfrequenz, die z. B. in der Gegend von 10^9 Hz liegt. Für eine andere Form des Strompulses $g(t)$ (z. B. Sägezahn-Form) ändert sich nur die Hochfrequenz-Flanke des Leistungsspektrums, ohne daß sich die Näherung einer konstanten Leistungsdichte für kleine Frequenzen und die generelle Form des Spektrums ändert. Damit kann man die wesentlichen Eigenschaften des Schrotrauschens auch mit dem hier vereinfacht angenommenen Rechteck-

Strompuls bereits sehr gut beschreiben.

In physikalischen Systemen ist das Auftreten von weißem Rauschen eher selten, da es bei entsprechend hoher Grenzfrequenz eine beliebig hohe Gesamt-Energie aufweisen müßte. Sehr viel häufiger ist in der Praxis das **rosa Rauschen** (1/f-Rauschen) anzutreffen. Bisher haben wir entweder weißes Rauschen betrachtet, d. h.

$$S^2(f) \approx f^0 \quad \text{für } f \ll \frac{1}{t} \quad (\text{II.51})$$

oder braunes Rauschen, d.h.

$$S^2(f) \approx f^{-2} \quad \text{für } f \ll \frac{1}{t} \quad (\text{II.52})$$

Im Übergangsbereich zwischen weißem und braunem Rauschen ergibt sich jedoch eine Abhängigkeit

$$S^2(f) \approx f^{-1} \quad \text{für } f \approx \frac{1}{t} \quad (\text{II.53})$$

In der Praxis wird die Abhängigkeit von f^{-1} (d. h. rosa Rauschen) auch für sehr niedrige Frequenzen beobachtet und nicht erst für Frequenzen in der Gegend einer bestimmten Grenzfrequenz. Eine plausible Begründung liegt darin, daß bei den meisten in der Natur vorkommenden Rauschprozessen nicht nur eine charakteristische Zeitkonstante τ vorliegt, sondern mehrere Zeitkonstanten τ_i auftreten, die jeweils auf verschiedene Zeitskalen verteilt sind. Dabei resultiert als Summenprozeß ein stochastischer Prozeß mit einer spektralen Leistungsdichte, die proportional $1/f$ ist.

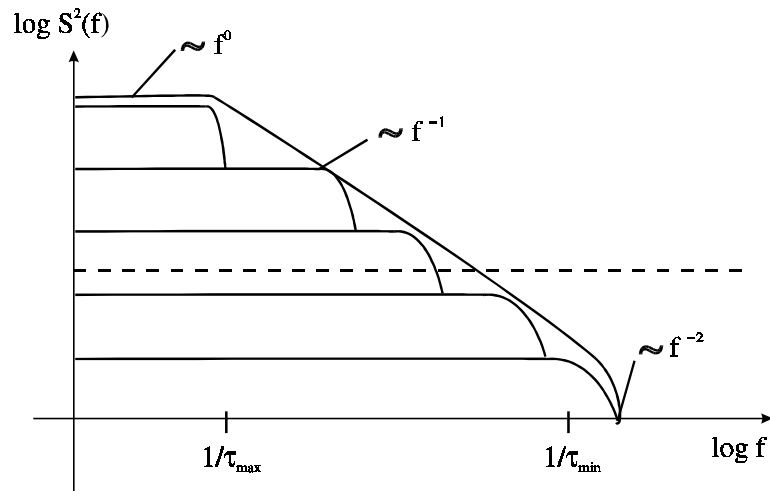


Abb. 2.24: 1/f- Rauschen

Zu den ausgezeichneten Eigenschaften eines derartigen 1/f-Rauschens zählt die Selbstähnlichkeit des Prozesses, d. h. bei Streckung oder Stauchung der Zeitachse bleibt die Form des Spektrums erhalten. Der Prozeß ist damit auf jeder Zeitskala im statistischen Sinne gleich. Diese Selbstähnlichkeit ist ein in der Natur sehr häufig zu beobachtendes Prinzip (vgl. Schroeder, 1991)

Eine weitere wichtige Eigenschaft ist die konstante Leistungsdichte des rosa Rauschens auf einer logarithmischen Frequenzachse. Derartige Frequenzachsen sind beispielsweise in der Akustik und der Musik von großer Bedeutung, da die Erhöhung der Frequenz um eine Oktave gerade einer Frequenzverdopplung entspricht und im Ohr (ab Frequenzen oberhalb von ca. 500 Hz) ein konstantes Frequenzintervall in einen konstanten Längenbereich der Basilarmembran abgebildet wird. Damit wird nicht eine konstante **absolute** Bandbreite, sondern eine konstante **relative** Bandbreite zugrundegelegt, die der sogenannten Bark-Skala entspricht. Rosa Rauschen hat in jedem Frequenzintervall und in jedem Bark somit die gleiche Leistung.

Eine weitere Eigenschaft des 1/f-Rauschens sind die geringen Konvergenz-Probleme für den Übergang zu sehr hohen Frequenzen, bei denen weißes Rauschen zu einer unendlich hohen Gesamt-Leistung führen würde. Gleichzeitig stellt sich beim Übergang zu sehr kleinen Frequenzen ein geringeres Problem als bei braunem Rauschen heraus, bei dem die Gesamt-Leistung $\rightarrow \infty$ gehen würde. Rosa Rauschen kann daher relativ problemlos sowohl sehr kleine als auch sehr große Frequenzen gleichzeitig überdecken. Es hat eine praktikable Gesamtenergie selbst bei Überdeckung eines beliebig großen Frequenzbereichs. Der größte physikalisch sinnvolle Frequenzbereich, der überdeckt werden kann, läßt sich aus

dem Verhältnis zwischen atomaren Zeiteinheiten und dem Alter des Weltalls abschätzen, das einem Frequenzverhältnis von etwa 10^{40} entspricht. Da $1/f$ -Rauschen in jeder Dekade der Frequenz dieselbe Leistung enthält, enthält die Leistung eines über sämtliche möglichen Frequenzen sich erstreckenden $1/f$ -Rauschens etwa nur die 40-fache Leistung eines $1/f$ -Rauschens, das nur eine einzige Dekade überstreicht.

Die „Wahrscheinlichkeit“ für das Vorliegen von $1/f$ -Rauschen ist in der Natur daher sehr hoch. Es wurde bereits in einer Fülle von Phänomenen entdeckt (z. B. Verteilung der Intervalle in Bachs brandenburgischen Konzerten, Intermittenz beim Übergang zum Chaos, vgl. Schroeder, 1991).

Eine Art von Rauschen, die nicht eine Eigenschaft eines physikalischen Systems kennzeichnet, sondern sich als Testsignal zum Ausmessen von Eigenschaften physikalischer Systeme hervorragend eignet, ist das **Schieberegister-Rauschen** (Pseudo-Zufalls-Signale). Es hat eine Reihe von Eigenschaften, die typisch für Rauschsignale sind, obwohl es deterministischer Natur ist und eignet sich für eine Vielzahl von Anwendungen, die eine oder mehrere der im folgenden aufgeführten Eigenschaften erfordern:

- a) Ein Testsignal sollte einen geringen „Peak-Faktor“ aufweisen, d. h. das Verhältnis zwischen Maximalaussteuerung und der mittleren Leistung des Signals sollte möglichst gering sein. Diese Eigenschaft ist notwendig, damit bei einer limitierten Aussteuerung des Übertragungsmediums (z. B. beim Ausmessen eines Verstärkers oder eines linearen Filters) möglichst viel Energie „transportiert“ wird, um z. B. das Signal-Rauschverhältnis zu optimieren.
- b) Das Testsignal sollte ein flaches Leistungsspektrum aufweisen. Diese gleichmäßige Anregung sämtlicher Frequenzen ist z. B. zur Ermittlung einer Übertragungsfunktion von linearen Systemen notwendig, da das Ausgangsspektrum durch das Eingangsspektrum dividiert werden muß, so daß das Eingangsspektrum möglichst keine Nullstellen und einen möglichst konstanten Wert annehmen sollte. Selbst ein Ausschnitt eines „weißen“ Rauschsignals ist hierzu nicht sehr geeignet, weil im Spektrum zufällig auch sehr kleine Werte auftreten können, die zu numerischen Problemen führen (s. o.). Die Verwendung eines deterministischen Signals mit einem bekannten, konstanten Spektrum vermeidet diese Probleme.

- c) Das Testsignal sollte reproduzierbar sein und eine endliche Länge aufweisen, so daß eine exakte Rekonstruktion eines Meßvorgangs möglich ist. Dies ist bei der Verwendung von stochastischen Signalen nicht der Fall.

Die in der Praxis verwendeten Schieberegister-Rauschsignale werden durch die logische Verschaltung eines Schieberegisters mit n Bits erzeugt, bei dem der Zustand des vorhergehenden Bits in jedem Taktzyklus der Länge T auf das nachfolgende Verzögerungsglied übertragen wird. Der Ausgang des Schieberegisters ist zusammen mit einem oder mehreren Abgriffen des Schieberegisters über ein XOR-Glied (Exklusiv- Oder - Verknüpfung) an den Eingang gekoppelt. Bei einer geeigneten Ankopplung durchläuft das Schieberegister sukzessive sämtliche 2^{n-1} Zahlenkombinationen, die sich mit n Bits darstellen lassen. Auf diese Weise werden sämtliche mögliche Zustände des Schieberegisterauschens zyklisch durchlaufen und es entsteht eine Ausgangsfolge aus „0“ und „1“, die eine Periode der Länge $(2^n - 1) \cdot T$ aufweist.

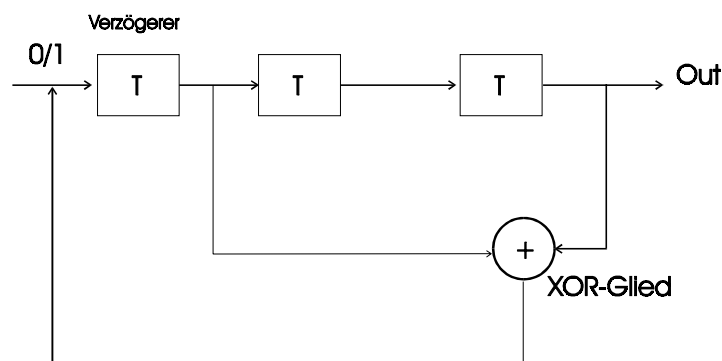


Abb. 2.25: Schieberegisterkonstruktion



Abb. 2.26: Beispiel für Ausgangsfolge aus „0“ und „1“

Der 0-Zustand (Zustand 0 in allen n Bits) wird dabei als einzige mögliche Zahl nicht durchlaufen, weil sie nur wieder den gleichen Zustand „0“ produzieren würde. Ebenfalls würde bei einer ungeeigneten Verkopplung der Schieberegisterzustände auf den Eingang nur eine Teilmenge sämtlicher 2^{n-1} „erlaubter“ Zustände durchlaufen werden, so daß sich das Problem der geeigneten „Anzapfung“ des Schieberegisters auf das Problem der Faktorisierung von sogenannten irreduziblen Polynomen des Grades n zurückführen läßt, d. h. Polynomen n -ten Grades, die sich nicht weiter in Faktoren von Polynomen niedrigeren Grades zerlegen lassen. (Zur zahlentheo-

retischen Grundlage von Schieberegisterrauschen vgl. Schroeder, 1986).

Das Leistungsspektrum eines derartigen Signals mit der Periode $(2^n-1) \cdot T$ errechnet sich aus der Autokorrelationsfunktion, die jeweils Maxima für Verzögerungswerte von $\tau = k \cdot (2^n-1) \cdot T$ aufweist ($k \in \mathbb{N}$). Für alle anderen Verzögerungszeiten ist die Autokorrelationsfunktion konstant gleich n . Für sehr kleine Verzögerungszeiten um dieses Maximum herum ist die Autokorrelationsfunktion dreiecksförmig. Das Leistungsspektrum als Fouriertransformierte hat daher die Grob-Strukturen einer sinc^2 -Funktion mit der ersten Nullstelle bei $1/T$ (T : Taktfrequenz des Schieberegisters). Für Frequenzen, die kleiner sind als die Taktfrequenz ist, das Leistungsspektrum daher konstant (bis auf einen etwaigen Gleichspannungsanteil). Aufgrund der Periodizität des Signals mit der Länge $(2^n-1) \cdot T$ besitzt das Spektrum als Feinstruktur eine Linienstruktur mit dem Linienabstand $1/(2^n-1) \cdot T$. Damit erfüllt das Schieberegisterrauschen sämtliche Eigenschaften, die oben gefordert wurden.

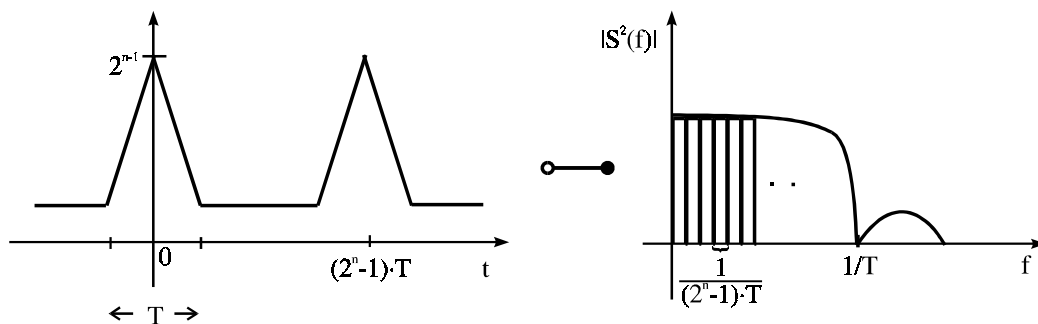


Abb. 2.27: Autokorellationsfunktion

Ein weiterer Vorteil liegt in der Auswerte-Möglichkeit durch die schnelle Hadamard-Transformation, die etwa vergleichbar ist mit der FFT. Bei dieser Transformation werden sämtliche Multiplikationen durch einfache Binär-Operationen wie Additionen ersetzt, so daß sie sich sehr schnell und effizient ausrechnen läßt. Aufgrund dieser Eigenschaften reicht die Anwendung der Maximalfolgen von der exakten Entfernungsbestimmung zwischen Erde und Mond mittels Radar-Messungen bis hin zu Ausmessungen von Konzertsaal-Akustik-Eigenschaften bei vollbesetzten Auditorien. Eine weitere Beschreibung findet sich bei Alrutz (1993) und bei Golomb (1967).

Eine Auswahl von Schieberegister-Längen n und dafür erforderlichen Abgriffen gibt die folgende Tabelle.

n	Schieberegister-Abgriffe	Periode in Taktzyklen ($2^n - 1$)
3	1, 3	7
5	1, 3, 4, 5	127
11	1, 3, 4, 5, 6, 11	2.047
15	1, 3, 6, 8, 10, 15	32.767
19	1, 3, 6, 9, 12, 19	524.287
21	1, 3, 4, 5, 7, 9, 10, 21	2.097.151