Einführung in die Telekommunikation

Matthias Neugschwandtner, Matthias Auchmann

12. Februar 2007

Inhaltsverzeichnis

1 Einführung

Dieser Abschnitt wurde nur eingefügt, damit die Kapitelnummerierung synchron mit der des Buches ist.

2 Periodische und transiente Signale

Deterministische Signale können periodisch oder transient sein. Periodische Signale bleiben gleich, wenn man sie um ein ganzzahliges Vielfaches ihrer Periode verschiebt und haben ein diskretes Linienspektrum. Transiente Signale sind aperiodisch und haben ein kontinuierliches Spektrum.

2.1 Periodische Signale

2.1.1 Fourierreihe

Alle periodischen Signale können als gewichtete Summe harmonischer Sinusschwingungen dargestellt werden. Dies erfolgt bei Periode T z.B. als trigonometrische Fourierreihe:

$$v(t) = C_0 + \sum_{n=1}^{\infty} C_n \cos(\omega_n t + \phi_n)$$

$$\omega_n := n\omega_0 \quad \text{mit } w_0 := \frac{2\pi}{T}$$

Das ist die Cosinus-Form der trigonometrischen Fourierreihe, sie kann auch ohne Phasenverschiebung als Cosinus-Sinus-Form angeschrieben werden:

$$v(t) = C_0 + \sum_{n=1}^{\infty} (A_n \cos \omega_n t - B_n \sin \omega_n t)$$

Wobei die A_n die Inphase- und die B_n die Quadratur[1]-Amplituden sind. Diese lassen sich mittels der „Filter-Integrale“ ermitteln:

$$A_n = \frac{2}{T} \int_0^T v(t) \cos \omega_n t dt$$

$$B_n = -\frac{2}{T} \int_0^T v(t) \sin \omega_n t dt$$

Mit dem Satz des Pythagoras sowie der trigonometrischen Beziehung $tan = \frac{sin}{cos}$ bekommt man nun die Darstellung als reine Cosinus-Reihe mithilfe der Formeln:

$$C_n = \sqrt{A_n^2 + B_n^2}$$

$$\phi_n = arctan(\frac{B_n}{A_n})$$

[1] Um $\pi/2$ versetzt

Über die Euler'sche Identität $e^{jx} = cos(x) + jsin(x)$ kommt man zur Exponentialform der Fourierreihe. Dabei wird immer ein Paar aus gegenläufigen, konjugierten Zeigern benutzt um eine reelle Cosinusschwingung darzustellen:

$$v(t) = \sum_{n=-\infty}^{\infty} C'_n e^{-j\omega_n t}$$

Dabei gilt: $C'_n = A_n - jB_n$ für die exponentiellen Fourierkoeffizienten. Trägt man den Betrag der Amplituden auf, so erhält man für reelle $v(t)$ ein zweiseitiges Amplitudenspektrum mit gerader und ein Phasenspektrum mit ungerader Symmetrie. Das einseitige Spektrum für die reellen Sinusschwingungen ergibt sich durch Faltung und Addition der Amplituden der positiven und negativen Frequenzen. Daher sind die komplexen Fourierkoeffizienten auch nur halb so groß wie die trigonometrischen. Das zweiseitige Phasenspektrum bei der Fourierreihendarstellung mit der komplexen Exponentialfunktion rührt daher, dass man für jede Amplitude in der Cosinusdarstellung der Fourierreihe jetzt zwei Zeiger verwendet (darum werden die Amplituden auf beiden Seiten halbiert) die gegengleich drehen (daher kommen die negativen Frequenzen). Zum gegengleich drehenden Zeiger gehört natürlich genau die negative Phase (darum die ungerade Symmetrie im Phasenspektrum).

2.1.2 Leistung

Wenn man die trigonometrischen Koeffizienten durch $\sqrt{2}$ dividiert und auf die Frequenzen aufträgt, erhält man ein einseitiges RMS[2]-Amplitudenspektrum. Das normalisierte Leistungsspektrum erhält man durch Quadrieren des RMS-Amplitudenspektrums. Die Gesamtleistung eines Linienspektrums ist die Summe der Leistungen jeder einzelnen Frequenz. Dies gilt für jedes periodische Signal da Sinus- und Kosinusfunktionen mit ganzzahligen Vielfachen einer festen Grundfrequenz ein Orthogonalsystem bezüglich der quadratischen Integralnorm (normiert auf die Periode) bilden. Diesen Sachverhalt nennt man auch Parseval'sches Theorem, man kann also die normierte Leistung (normalized power P) eines Signals aus der (unendlichen) Reihe seiner Fourierkoeffizienten berechnen:

$$P = \frac{1}{T} \int_0^T |v(t)|^2 \, dt = \sum_{n=-\infty}^{\infty} \left|C'_n\right|^2$$

Man beachte, dass ein periodisches Signal (abgesehen von der Nullfunktion) unendliche Energie hat, da das uneigentliche Integral die notwendige Konvergenzbedingung $v(t) \to 0$ für $t \to \infty$ nicht erfüllt:

$$E = \int_{-\infty}^{\infty} |v(t)|^2 \, dt = \infty$$

Man kann auch ein Leistungsspektrum angeben, man nimmt dazu das Amplitudenspektrum und quadriert jede Frequenzkomponente, man bekommt so das Leistungsspektrum über der Frequenz aufgetragen.

2.1.3 Existenz der Fourierreihe und Gibbssches Phänomen

Es gibt verschiedene kompliziertere Kriterien für die Existenz der Fourierreihe (Dirichlet, Jordan), für unsere Zwecke reicht es jedoch zu bemerken, dass stetig differenzierbare Funktionen stets eine

[2]RMS = root mean square

Fourierreihe haben. Hat man jedoch eine Funktion mit einer Unstetigkeit, so muss man beachten, dass die Fourierreihe das quadratische Integral minimiert, an der Unstetigkeitsstelle haben wir keine punktweise Konvergenz mehr, hier konvergiert die Funktion punktweise gegen das arithmetische Mittel aus links- und rechtsseitigen Grenzwert. Man nennt dieses Phänomen Gibbssches Phänomen, man bekommt an der Unstetigkeitsstelle Über- und Unterschwinger in der Größenordnung von jeweils 9 Prozent der Sprunghöhe.

Abbildung 1: Gibbsches Phänomen an einer Sprungstelle

2.1.4 Bandbreite & Sampling

Die Bandbreite ist der Teil des Frequenzbereichs, in dem ein Signal nennenswerte Leistung enthält. Bandbreite bezeichnet genauer die Differenz zwischen zwei Frequenzen f_{min} und f_{max} unter- respektive oberhalb derer die Spektralkomponenten auf $\frac{1}{\sqrt{2}}$ der Spitzenkomponente abfallen. Signale mit steilen Flanken haben eine hohe Bandbreite. Man beachte dass f_{min} immer positiv ist (hier nicht durch das zweiseitige Spektrum verwirren lassen!).

Hat ein Signal keine Spektralkomponenten oberhalb einer Frequenz f_H, so kann es nach einer Abtastung in der Theorie dann rekonstruiert werden, wenn die Abtastfrequenz f_S mindestens doppelt so hoch wie f_H ist. Diese Tatsache ist auch als Nyquist-Theorem oder Shannon'sches Abtasttheorem bekannt. Die Frequenz $2f_H$ heißt Nyquist-Frequenz, im Fall $f_S > 2f_H$ spricht man von oversampling, im Fall $f_S < 2f_H$ von undersampling. Ist die Abtastfrequenz zu niedrig, so werden Frequenzen die größer als $\frac{f_S}{2}$ sind falsch interpretiert, man nennt sie Alias-Frequenezen und dieses Phänomen Aliasing.

2.2 Transiente Signale

2.2.1 Fouriertransformation

Bekanntlich liefert die Fourierreihe ein diskretes Linienspektrum. Bei transienten Signalen hat man aber ein kontinuierliches Spektrum. Den Übergang kann man sich wie folgt vorstellen: die Periode T strebt gegen unendlich und der Abstand der Spektrallinien $\frac{1}{T}$ gegen Null. Die diskreten f_n des Spektrums werden zu f und aus der Summe $\sum_{n=-\infty}^{\infty}$ wird das Integral $\int_{-\infty}^{\infty}$. Die Fourierkoeffizienten $C'_n(f_n)$ werden zu $V(f)df$ (Einheit Volt). Man bekommt so eine kontinuierliche Darstellung der Funktion $v(t)$ durch die inverse Fouriertransformation:

$$v(t) = \int_{-\infty}^{\infty} V(f)e^{j2\pi ft}df$$

Die Fouriertransformierte $V(f)$ berechnet man wie folgt:

$$V(f) = \int_{-\infty}^{\infty} v(t) e^{-j2\pi ft} dt$$

Auch die (inverse) Fouriertransformation kann in Sinus und Cosinus aufgesplittet werden. Das Aufsplitten ist besonders dann sinnvoll, wenn die zu transformierende Funktion gerade oder ungerade ist, da man dann nur den cos- bzw. den sin-Teil berechnen muss. Es gibt ähnliche Bedingungen für die Existenz der Fouriertransformierten wie im Falle der Fourierreihe.
Jedem Signal im Zeitbereich entspricht genau ein, durch die Fouriertransformation gegebenes Amplituden- und Phasenspektrum im Frequenzbereich.

2.2.2 Korrespondenz der Fouriertransformation eines Rechteckpulses

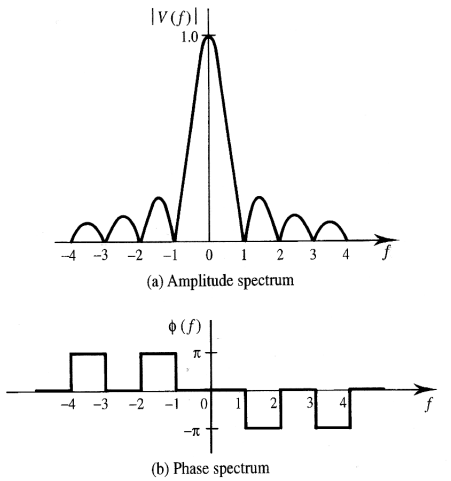
Ein Rechteckpuls ist gegeben durch:

$$\prod(t) := \begin{cases} 1.0, & |t| < \frac{1}{2} \\ 0.5, & |t| = \frac{1}{2} \\ 0, & |t| > \frac{1}{2} \end{cases}$$

Das Spannungsspektrum ergibt sich durch die Fouriertransformation:

$$V(f) = \int_{-\infty}^{\infty} \prod(t) e^{-j2\pi ft} dt = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-j2\pi ft} dt = \left[\frac{e^{-j2\pi ft}}{-j2\pi f} \right]_{-\frac{1}{2}}^{\frac{1}{2}} = \frac{1}{j2\pi f} [e^{j\pi f} - e^{-j\pi f}] =$$

$$= \frac{j2\sin(\pi f)}{j2\pi f} = \frac{\sin(\pi f)}{\pi f} = \text{sinc}(f)$$



Voltage spectrum of unit rectangular pulse shown in Figure 2.27(a).

2.2.3 Theoreme der Fouriertransformation

- Linearität: $av(t) + bw(t) \Leftrightarrow aV(f) + bW(f)$
- Zeitverschiebung: $v(t - T) \Leftrightarrow V(f)e^{-j\omega T}$
- Zeitumkehr: $v(-t) \Leftrightarrow V(-f)$
- Multiplikation: $v(t)w(t) \Leftrightarrow V(f) * W(f)$
- Faltung: $v(t) * w(t) \Leftrightarrow V(f)W(f)$
- Frequenzkonjugation: $v^*(-t) \Leftrightarrow V^*(f)$

Durch die Zeitverschiebungseigenschaft bekommen wir nun auch die Fouriertransformierte für einen verspäteten Rechteckpuls der Breite τ:

$$\prod(\frac{t-T}{\tau}) \Leftrightarrow \tau sinc(\tau f)e^{-j\omega T}$$

Mit dieser neuen Erkenntnis berechnen wir nun ein weiteres Korrespondenzpaar:
Die Impulsfunktion (Dirac-Delta) ist definiert als

$$\delta(t) := \begin{cases} \infty, & t = T \\ 0, & t \neq T \end{cases}$$

Die Fouriertransformierte der Impulsfunktion erhält man durch Grenzwertbildung:

$$FT\{\delta(t-T)\} = FT\{\lim_{\tau\to 0}\frac{1}{\tau}\prod(\frac{t-T}{\tau})\} = \lim_{\tau\to 0} FT\{\frac{1}{\tau}\prod(\frac{t-T}{\tau})\} =$$

$$\lim_{\tau\to 0}\frac{1}{\tau}\tau sinc(\tau f)e^{-j\omega T} = lim_{\tau\to 0} sinc(\tau f)e^{-j\omega T} = e^{-j\omega T}$$

Die Impulsfunktion hat also ein konstantes Amplitudenspektrum (obige e-Funktion rotiert im Uhrzeigersinn am Einheitskreis in der komplexen Ebene). Solche Spektren werden oft weiße Spektren genannt.

Die Faltung $z(t) = f(t) * g(t)$ ist wie folgt definiert (das hier auftretende Integral nennt man auch Superpositionsintegral): $z(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$.
Bei der Faltung wird die Funktion g also zuerst an der y-Achse gespiegelt, danach um t nach rechts verschoben. Das Produkt der so erhaltenen Funktionen wird nun integriert und ist der Wert der Faltung an der Stelle t. Die Faltung ist kommutativ, assoziativ und distributiv gegenüber der Addition.
Das Faltungstheorem ist äußerst praktisch, da eine Faltung im Zeitbereich durch eine Multiplikation im Frequenzbereich gelöst wird.

Will man z.B. ein periodisches Rechtecksignal erzeugen, so kann man dazu einen transienten Rechteck-Impuls mit einem periodischen Dirac-Impulssignal falten.

Zwei weitere wichtige Fouriertransformations-Paare sind:

- Sampling: $\sum_{k=-\infty}^{\infty} \delta(t - kT_s) \Leftrightarrow f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s)$
- Sprungfunktion (heaviside step) $u(t) \Leftrightarrow \frac{1}{2}\delta(f) + \frac{1}{j2\pi f}$

Dabei ist die Sprungfunktion wie folgt definiert:

$$u(t) := \begin{cases} 0, & t < 0 \\ \frac{1}{2}, & t = 0 \\ 1, & t > 0 \end{cases}$$

Weitere, für die Prüfung benötigte Korrespondenzen (f_c bezeichnet die carrier frequency, also die Trägerfrequenz $f_c = \frac{\omega}{2\pi}$):

- Dreieck der Breite 2τ: $\bigwedge(\frac{t}{\tau}) \Leftrightarrow \tau sinc^2(\tau f)$
- Exponentialschwingung: $e^{j(\omega t+\phi)} \Leftrightarrow e^{j\phi}\delta(f - f_c)$
- Sinusschwingung: $sin(\omega t + \phi) \Leftrightarrow \frac{1}{2j}(e^{j\phi}\delta(f - f_c) - e^{-j\phi}\delta(f + f_c))$
  Man beachte dass wir hier ein diskretes Linienspektrum haben, die beiden Summanden in der Klammer haben gleichen Betrag. Der Sinus ist ungerade und die gegenläufig rotierenden Zeiger werden an der komplexen Ebene gespiegelt, darum bekommt man neben dem Minus im Exponenten auch noch ein $1/j$ vor der Klammer (Drehung um 90°) sowie das Minus (man setze Phase $\phi = 0$). Das $1/2$ vor der Klammer sorgt dafür, dass die Amplitude zwischen positiver und negativer Frequenz aufgeteilt wird.
- Vorzeichenfunktion: $sgn(t) \Leftrightarrow \frac{1}{j\pi f}$

2.2.4 Leistungs- und Energiedichtespektren

Das Leistungsspektrum eines periodischen Signals bekommt man durch

$$G_1(f) := \frac{|V(f)|^2}{\sqrt{2}}$$

Das Energiedichtespektrum eines transienten Signals ist

$$E_2(f) := |V(f)|^2$$

Auch hier sind zweiseitige Spektren (man verwendet auf beiden Seiten jeweils den halben Wert) üblich, auch wenn zweiseitige Spektren hier nicht so naheliegend sind.

2.3 Kurvenformen als Vektor

Laut dem Nyquist-Theorem lässt sich ein periodisches Signal mit einer höchsten Frequenzkomponente von f_H durch N Abtastungen mit einer Frequenz von $2f_H$ exakt rekonstruieren. Betrachtet man nun jeden Abtast-Wert als Länge eines Vektors einer Orthogonalbasis, so kann man ein durch N Abtastungen spezifiziertes Signal als $N = 2f_HT$-dimensionalen Vektor darstellen. Nun kann man

mit den Signalen das gleiche anstellen wie mit Vektoren: addieren, skalieren, usw. Man macht diese Betrachtungen um dann eine fundierte mathematische Basis zu haben um Korrelationen zwischen zwei Funktionen definieren zu können. Man definiert zunächst ein Skalarprodukt für zwei periodische Funktionen/Signale:

$$[f(t), g(t)] := \frac{1}{T'} \int_0^{T'} f^*(t)g(t)dt$$

mit der Einheit V^2, man spricht von Kreuz-Leistung. T' ist dabei die Periode der multiplizierten Funktion, $*$ steht für die Konjugation. Man kann die Abhängigkeit von dieser Periode T' eliminieren indem man zu folgender äquivalenten Definition übergeht:

$$[f(t), g(t)] := \lim_{T' \to \infty} \frac{1}{T'} \int_{-\frac{T'}{2}}^{\frac{T'}{2}} f^*(t)g(t)dt$$

Für transiente Signale haben wir folgende Definition, man spricht von Kreuz-Energie:

$$[f(t), g(t)] := \int_{-\infty}^{\infty} f^*(t)g(t)dt$$

Man beachte, dass sich nun alle aus der Fourieranalysis bekannten Resultate auf Signale übertragen, insbesondere die Tatsache, dass man auch andere Orthogonalsysteme als das trigonometrische Orthogonalsystem betrachten kann. Man bekommt hier eine Bestapproximation im Sinne der Integralnorm mit den Fourierkoeffizienten, für ein vollständiges Orthonormalsystem geht der Fehler bei der Approximation bezüglich der Integralnorm definitionsgemäß gegen 0 für $n \to \infty$. D.h. dass die Leistung der Differenz von Fourierreihe und periodischer Funktion gegen 0 geht für n gegen ∞.

2.4 Korrelationsfunktion

Mithilfe der Erkenntnisse aus der Fourieranalysis definiert man nun die (normalisierte) Kreuzkorrelationsfunktion durch Normierung des Skalarprodukts und hat somit eine Funktion, die die Korrelation zweier Funktionen für alle möglichen Zeitverschiebungen angibt. Für transiente Signale lautet sie:

$$\rho_{vw}(\tau) = \frac{\int_{-\infty}^{\infty} v(t)w(t-\tau)dt}{\sqrt{\int_{-\infty}^{\infty} |v(t)|^2 dt}\sqrt{\int_{-\infty}^{\infty} |w(t)|^2 dt}}$$

Für periodische Signale:

$$\rho_{pq}(\tau) = \frac{\lim_{T' \to \infty} \int_{-T'/2}^{T'/2} p(t)q(t-\tau)dt}{\sqrt{\frac{1}{T}\int_0^T |p(t)|^2 dt}\sqrt{\frac{1}{T}\int_0^T |q(t)|^2 dt}}$$

Eigenschaften der normalisierten Kreuzkorrelationsfunktion:

- $-1 \leq \rho(\tau) \leq 1$
- $\rho(\tau) = -1 \Leftrightarrow v(t) = -kw(t-\tau)$
- $\rho(\tau) = 1 \Leftrightarrow v(t) = kw(t-\tau)$
- Wenn $\rho(\tau) = 0$, so sind die Signale orthogonal und haben keine Ähnlichkeit.

2.5 Autokorrelation

Betrachtet man nicht 2 verschiedene Signale, sondern das Skalarprodukt eines Signals mit sich selbst, so bekommt man die Autokorrelationsfunktion (bzw. wenn man normiert dann die normierte Autokorrelationsfunktion bei Dividieren durch Leistung bzw. Energie für periodische bzw. transiente Signale). Hier der Einfachheit halber nochmal die Definition der (nicht normierten) Autokorrelationsfunktion:

$R_p(\tau) := \frac{1}{T} \int_0^T p(t)p(t-\tau)dt$ für periodische Signale (beachte: die Periode ist bekannt!)

$R_v(\tau) := \int_{-\infty}^{\infty} v(t)v(t-\tau)dt$ für transiente Signale

Die Autokorrelationsfunktion gibt die Ähnlichkeit eines Signals mit einer zeitversetzten Kopie seiner selbst an.

Einige Eigenschaften der Autokorrelationsfunktion sind:

- Die Autokorrelationsfunktion ist gerade (substituiere $z = t - \tau$ in der DN, dann steht dort das Integral für $R(-\tau)$)
- Die Maximumstelle der Autokorrelationsfunktion ist 0
- Es bestehen folgende Zusammenhänge über die Fouriertransformation mit den Leistungsdichtespektrum bzw. den Energiedichtespektrum:

$$R_p(\tau) \Leftrightarrow G_p(f)$$

$$R_v(\tau) \Leftrightarrow E_v(f)$$

3 Zufallssignale und Rauschen

Nur Signale, deren Zukunft nicht genau bestimmt ist, können Information tragen. Ebenso sind Störsignale zufälliger Natur. Um Nutzsignale von Störsignalen bzw. Rauschen unterscheiden zu können, müssen sie wahrscheinlichkeitstheoretisch beschrieben werden.

3.1 Wahrscheinlichkeitstheorie

3.1.1 Wahrscheinlichkeit und Bayes

Wird ein Zufallsexperiment N-mal wiederholt, so ist die Wahrscheinlichkeit eines Ereignisses A, das L_N-mal auftritt:

$$P(A) = \lim_{N \to \infty} \frac{L_N}{N}$$

L_N soll hierbei andeuten, dass L von N abhängt (sonst wäre die WK ja 0). Der Ausgang eines Zufallsexperiments, z.B. Münzwurf (diskret) oder Wettlauf (kontinuierlich) ist eine Zufallsvariable. Der Wert einer Zufallsvariablen (stochastische Größe) an einem bestimmten Zeitpunkt in der Zukunft kann

nicht genau vorhergesagt werden, jedoch ist nach einem Wahrscheinlichkeitsmodell bekannt, welche Werte die Variable annehmen könnte.

Die Wahrscheinlichkeit eines Ereignisses A bedingt durch das Auftreten eines Ereignisses B wird mit $P(A|B)$ bezeichnet. Für das gemeinsame Ereignis $P(A, B)$ (also dem Durchschnitt der Ereignisse A und B) gilt:

$$P(A, B) = P(A)P(B|A) = P(B)P(A|B)$$

und nach Umformen die Bayes'sche Regel:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Die Ereignisse A und B sind statistisch unabhängig, wenn das Auftreten des einen nicht die Wahrscheinlichkeit des anderen beeinflusst, also gilt:

$$P(A|B) = P(A) \qquad P(B|A) = P(B)$$

Daraus folgt:

$$P(A, B) = P(A) \cdot P(B)$$

3.1.2 Fehlerwahrscheinlichkeit in einem Datenblock

Gefragt ist die Wahrscheinlichkeit, dass mehr als eine bestimmte Anzahl an Fehlern in einem Codewort auftreten - ein diskretes Wahrscheinlichkeitsproblem. Die gegebene Wahrscheinlichkeit eines Bitfehlers sei P_e, R' ist die Anzahl der Fehler und n ist die Länge des Codeworts. Da die Summe der Wahrscheinlichkeiten 1 ergeben muss, gilt (Gegenwahrscheinlichkeit):

$$P(\text{Fehleranzahl} > R') = 1 - P(\text{Fehleranzahl} \leq R')$$

Die Wahrscheinlichkeit von genau j Fehlern ist gegeben durch:

$$P(j \text{ Fehler}) = \binom{n}{j} (P_e)^j (1 - P_e)^{n-j}$$

Die Wahrscheinlichkeit von mehr als R' Fehlern ergibt sich dann als:

$$P(\text{Fehleranzahl} > R') = 1 - \sum_{j=0}^{R'} P(j)$$

3.1.3 Wahrscheinlichkeitsverteilung und -dichte

Eine Verteilungsfunktion gibt die Wahrscheinlichkeit $P_x(x)$ an, dass der Wert der Zufallsvariablen X kleiner gleich einem bestimmten Wert x ist: $P_x(x) = P(X \leq x)$. Der Wertebereich der Verteilungsfunktion liegt zwischen 0 und 1, der Grenzwert für $x \to -\infty$ ist 0 und der Grenzwert für $x \to \infty$ ist 1. Verteilungsfunktionen sind monoton steigend und es gilt: $P_x(x_2) - P_x(x_1) = P(x_1 < X \leq x_2)$

Eine Dichtefunktion gibt die Wahrscheinlichkeit an, dass der Wert einer Zufallsvariablen zwischen x und $x + dx$ liegt: $p_x(x) = \frac{dP_x(x)}{dx}$. Die Fläche unter der Dichtefunktion ist 1 und es gilt:

$\int_{x_1}^{x_2} p_x(x)dx = P(x_1 < X \leq x_2)$.
Ein Moment ist eine Statistik einer Zufallsvariable und gibt Information über die Form von Dichtefunktionen. Die wichtigsten sind das erste Moment - der Erwartungswert - und das zweite Zentralmoment - die Varianz, deren Wurzel die Standardabweichung ist:

$$E[X] = \overline{X} = \int_{-\infty}^{\infty} xp(x)dx \qquad \overline{(X - \overline{X})^2} = \int_{-\infty}^{\infty} (x - \overline{X})^2 p(x)dx = \sigma^2$$

Bei zentralen Momenten zieht man also stets in der Klammer den Erwartungswert ab. Das dritte und vierte Moment zeigen Schiefe und Kurtosis, sie sind z.B. für die Analyse von Sprache interessant. Hat eine Dichtefunktion nur eine Maximumstelle so spricht man von einer unimodalen Verteilung, sonst von einer multimodalen Verteilung.

Sowohl Verteilungs- als auch Dichtefunktion können diskrete, kontinuierliche und gemischte Zufallsvariablen darstellen.

Es liegt nahe, den ersten Moment (Mittelwert) als DC-Spannung zu interpretieren, und den 2. Zentralmoment (die Varianz) als AC-Anteil. Den zweiten Moment interpretiert man natürlich als Gesamtleistung.

3.1.4 Gemeinsame und Randverteilung

Sind X und Y zwei Zufallsvariablen, so ist eine gemeinsame Dichtefunktion $p_{x,y}(x, y)$ so gegeben, dass $p_{x,y}(x, y)\ dx\ dy$ die Wahrscheinlichkeit ist, dass X im Bereich x und $x + dx$ sowie Y im Bereich y und $y + dy$ liegt. Die Fläche unter der gemeinsamen Dichte ist wiederum 1.

Ist die gemeinsame Dichte bekannt, so wird die Wahrscheinlichkeit, dass X, unabhängig von Y zwischen x_1 und x_2 liegt, Randwahrscheinlichkeit von X genannt und man erhält die Randdichte indem man über ganz Y integriert:

$$p(x) = \int_{-\infty}^{\infty} p_{x,y}(x, y)dy$$

3.1.5 Gemeinsame Momente, Korrelation und Kovarianz

Die Definition der gemeinsamen (zentralen) Momente erfolgt analog zur bisherigen Definition:

$$\overline{X^n Y^m} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^n y^m p_{x,y}(x, y)dxdy$$

$$\overline{(X - \overline{X})^n (Y - \overline{Y})^m} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \overline{X})^n (y - \overline{Y})^m p_{x,y}(x, y)dxdy$$

Das erste gemeinsame Moment ist die Korrelation. Ein hoher positiver Korrelationswert spricht für hohe Ähnlichkeit (wenn x high ist, so ist wahrscheinlich auch y high), ein Wert nahe Null bedeutet, dass sich aufgrund des Zustands für x keine nennenswerten Aussagen über den Zustand von y treffen lassen.
Man sagt X und Y sind unkorreliert, wenn $\overline{XY} = \overline{X} \cdot \overline{Y}$ gilt. Sind X und Y statistisch unabhängig, so sind sie unkorreliert, der Umkehrschluss gilt allerdings nicht (so sind z.B. cos und sinus aufgrund

ihrer Orthogonalität unkorreliert, aber natürlich statistisch abhängig). Für gaußsche zweidimensionale Variablen (das sind solche, die durch geeignete Translationen und Drehungen in Zufallsvariablen übergeführt werden können, die 2-dimensional standardnormalverteilt sind) gilt allerdings die Umkehr.

Das erste zentrale gemeinsame Moment ist die Kovarianz. Bei der Bildung der Kovarianz wurde der Erwartungswert bereits abgezogen; die Kovarianz bezieht sich daher auf die Korrelation der sich unterscheidenden Teile von X und Y. Bei Erwartungswert Null sind Kovarianz und Korrelation ident.

3.1.6 Der zentrale Grenzwertsatz

Addiert man zwei Zufallsvariablen zu einer neuen (d.h. $X + Y = Z$), so stellt sich die Frage wie die aus dieser Addition hervorgehende Dichte $p_z(z)$ aussieht. Aus der trivialen Umformung

$$X + Y = Z \Rightarrow Y = Z - X$$

ergibt sich:

$$p_z(z) = \int_{-\infty}^{\infty} p_{x,y}(x, z - x)dx$$

und bei stochastischer Unabhängigkeit das bekannte Superpositions- bzw. Faltungsintegral:

$$p_z(z) = \int_{-\infty}^{\infty} p_x(x)p_y(z - x)dx$$

Addiert man N stochastisch unabhängige Zufallsvariablen, so hat die Summe eine Dichtefunktion, die sich mit $N \to \infty$ der Gauß'schen Glockenkurve annähert. Die Verteilung der Summe ist also für $N \to \infty$ normalverteilt. Das macht es besonders interessant den Einfluss von gaußschen Störungen zu studieren, da sich die Überlagerung von Zufallsstörungen (Addition) der selben Art nach dem zentralen Grenzverteilungssatz als gaußsche Störung darstellt. Für die Summe von unabhängigen Zufallsvariablen gilt außerdem, dass sich die Mittelwerte und die Varianzen addieren. Da bei identisch verteilten Zufallsvariablen mit Mittelwert $\neq 0$ die Grenzverteilung unendlichen Mittelwert hat, normiert man die Zufallsvariablen zumeist vor der Addition und dividiert jede Summe von n Summanden durch $\sqrt{n}$, so bekommt man eine $N(0, 1)$ verteilte Zufallsvariable für $n \to \infty$.

3.1.7 Beweis des zentralen Grenzwertsatzes

Wie der Beweis in der VO gemacht wurde weiß ich nicht, im Buch wird jedenfalls nichts bewiesen sondern bestenfalls besprochen. Der Standardbeweis ist nicht schwer, baut aber auf Kenntnissen der Komplexen Analysis und der Maßtheorie auf, er sei hier dennoch angegeben:

Zunächst eine Definition: Sei X eine Zufallsvariable. Die Funktion

$$\varphi_X(t) := \mathbb{E}(e^{itX}) = \int_{\Omega} e^{itX} dP_X$$

nennt man charakteristische Funktion von X, Ω bezeichnet den zugrundeliegenden WK-Raum, dP_X die Lebesgue-Integration nach der Verteilung von X. Da e^{itX} eine auf ganz $\mathbb{C}$ holomorphe Funktion

ist darf man Integration und Differentiation vertauschen, man bekommt daher für die Taylorentwicklung von $\varphi_X(t)$:

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \mathbb{E}(X^k)$$

Nun reicht es, den Beweis für normierte Zufallsvariablen X_i mit Mittelwert 0 und Varianz 1 zu führen, man bekommt dann (ein WK-Raum hat Maß 1, Mittelwert ist 0 nach VS, $\mathbb{E}(X^2) = 1$ nach VS und dem Steinerschen Verschiebungssatz):

$$\mathbb{E}(X^0) = 1 \quad \mathbb{E}(X) = 0 \quad \mathbb{E}(X^2) = 1$$

Damit haben wir folgende Taylorentwicklung

$$\varphi_X(t) = 1 - \frac{t^2}{2} + o(t^2)$$

Dividiert man nun die X_i wie oben beschrieben durch $\sqrt{n}$, und betrachtet die transformierten Zufallsvariablen $Y_i = \frac{X_i}{\sqrt{n}}$ so haben wir also, da für unabhängige Zufallsvariablen X und Y gilt: $\varphi_{X+Y} = \varphi_X \varphi_Y$, dass jede endliche Summe die folgende charakteristische Funktion hat (man beachte die Linearität der charakteristischen Funktion, die direkt aus der Integraldefinition folgt):

$$(\varphi_Y(\frac{t}{\sqrt{n}}))^n = (1 - \frac{t^2}{2n} + o(\frac{t^2}{n}))^n \to e^{-t^2/2} \quad \text{für } n \to \infty$$

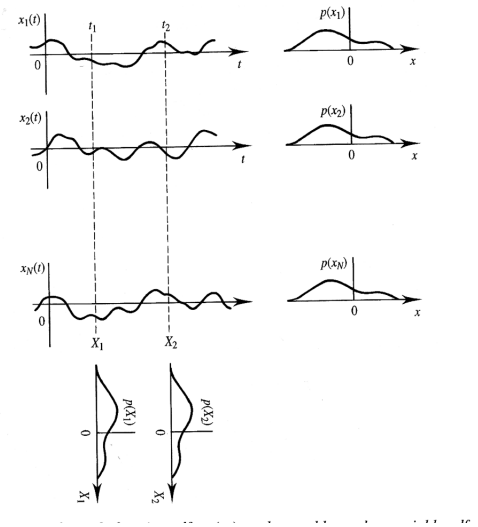
Die charakteristische Funktion der Summe konvergiert also gegen die charakteristische Funktion der $N(0,1)$-Verteilung. Nach dem Stetigkeitstheorem von Lévy folgt aus der punktweisen Konvergenz der charakteristischen Funktion die Verteilungskonvergenz (also die Konvergenz an allen Stetigkeitspunkten der Verteilung), damit wäre alles gezeigt.

3.2 Zufallsprozesse

Ein Zufallsprozess ist eine Zufallsvariable, die sich in Abhängigkeit von Zeit (oder Raum) verändert. Zufallsprozesse werden durch eine Anzahl (i) von sample functions $x_i(t)$ beschrieben. X_j oder $X(t_j)$ bezeichnet ein Ensemble von Werten zur Zeit t_j und bildet eine Zufallsvariable. Man beachte dass ein Zufallsprozess, der durch N sample functions beschrieben wird, KEINE N-dimensionale Zufallsvariable beschreibt. Der Prozess ist eine 1-dimensionale Zufallsvariable, der nur durch N Beobachtungen beschrieben wird. Als Beispiel stelle man sich N Gefäße in einem Chemielabor vor, in denen überall der selbe chemische Prozess abläuft. Der Prozess selbst ist der Zufallsprozess, der hier durch N sampling functions beschrieben wird. $X(t_0) = X_0$ beschreibt dann die Statistik der N Gefäße zum Zeitpunkt t_0.
Zufallsprozesse können kontinuierlich und diskret bezüglich Zeit oder Position, analog oder digital (wiederum kontinuierlich oder diskret) bezüglich ihrer Wahrscheinlichkeitsdichte, (nicht) deterministisch (Bsp. für einen deterministischen Zufallsprozess: sinus mit zufälliger Phase), (nicht) ergodisch und (nicht) stationär sein.
Stationär bezieht sich auf die Zeitabhängigkeit des Zufallsprozess. Ein Zufallsprozess ist

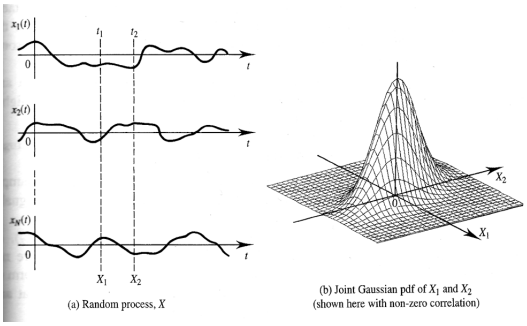
- streng stationär, wenn alle seine Dichten (gemeinsame, bedingte, Randdichte) von $X(t)$ für jeden Zeitpunkt t gleich sind, sich also im Zeitverlauf nicht ändern.
- locker bzw. leicht stationär, wenn sein Erwartungswert zeitunabhängig ist und seine Korrelation $\overline{X(t_1)X(t_2)}$ nur vom Zeitunterschied $t_2 - t_1$, nicht jedoch von t_1 und t_2 selbst abhängt.

Ein Zufallsprozess ist ergodisch wenn jede Zeitfunktion (sample function) des Ensembles die gleiche Verteilung, also das gleiche statistische Verhalten hat wie zu allen anderen Zeitpunkten und diese Verteilung außerdem mit den Verteilungen der einzelnen Funktionen $x_i(t)$ übereinstimmt (siehe Bild). Ergodisch impliziert stationär, aber nicht umgekehrt.

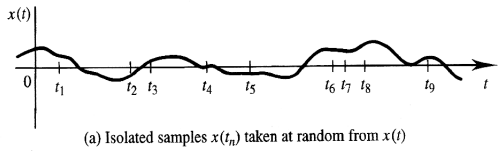


3.2.1 Gauß-Prozess

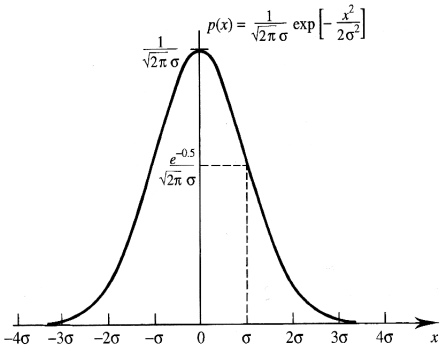
Man sagt eine sample function $x_i(t)$ gehört auf streng-Gaußsche Art zu einem Zufallsprozess, wenn die Zufallsvariablen $X_1 \dots X_N$ ($X_j := X(t_j)$), die die Samples zu N Zeitpunkten enthalten, und dabei auch die Samples von $x_i(t)$ mit in Betracht ziehen, N-dimensional Gauß-verteilt sind. Für ergodische Prozesse gehört also eine sample function genau dann im strengen Sinne zu einem Gaußschen Prozess, wenn das für jede sample function gilt. Ein Ensemble von sample functions bezeichnet man als streng Gauß'schen Prozess, wenn $X_1 \dots X_N$ N-dimensional Gauß-verteilt sind.



Eine sample function gehört auf lockere Gauß'sche Art zu einem Zufallsprozess, wenn einzelne Werte aus $x_i(t)$ Gauß-verteilt sind. Ein Ensemble von sample functions bezeichnet man als lockeren Gauß'schen Prozess, wenn für alle $x_i(t)$ gilt, dass einzelne Werte aus $x_i(t)$ Gauß-verteilt sind.



(a) Isolated samples $x(t_n)$ taken at random from $x(t)$



(b) Gaussian distribution of samples from (a)

Definition of a loose sense Gaussian process.

Ein streng Gauß'scher Prozess ist der strukturloseste, zufälligste, unverhersehbarste Zufallsprozess überhaupt (aber praxisrelevant, Bsp.: thermisches Rauschen). Strenge Gauß-Prozesse sind lockere Gauß-Prozesse. Gauß-Prozesse sind durch erstes und zweites Moment voll spezifiziert.

3.2.2 Autokorrelation und Leistungsdichtespektrum von Zufallsprozessen

Eine Dichtefunktion kann ein Zufallssignal (z.B. sample function) nicht genau beschreiben, da sie kein Wissen über die Änderungsrate des Signals besitzt. Diese Information steckt in den gemeinsamen Dichten von Zufallsvariablen $X(t_1)$ und $X(t_2)$, die $\tau = t_2 - t_1$ auseinanderliegen. Diese Dichte ist meistens nicht bekannt, ein Teil der interessierenden Information ist jedoch in der Korrelation $\overline{X(t_2)X(t_2-\tau)}$ enthalten. Aus diesem Grunde interessiert auch hier die Autokorrelationsfunktion, denn bei ergodischen Prozessen kann man wegen der Zeitunabhängigkeit und der Unabhängigkeit von der sample function die Korrelation mithilfe der Autokorrelationsfunktion berechnen:

$$R_X(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t-\tau)dt \quad [V^2]$$

Die Autokorrelationsfunktion und das doppelseitige Leistungsdichtespektrum von $x(t)$ bilden ein Fouriertransformations-Paar:

$$R_x(\tau) \Leftrightarrow^{FT} G_x(f)$$

auch als Wiener-Kintchine Theorem bekannt. Eine normalisierte Autokorrelationsfunktion erhält man durch Autokorrelation der Funktion nach Abziehen der Gleichstromkomponente (Erwartungswert)

von $x(t)$ und Dividieren durch den RMS-Wert. Für die normalisierte Autokorrelationsfunktion ρ_X gilt, dass sie bei 0 eine Maximumstelle hat mit $\rho_X(0) = 1$ und für $|x| \to \infty$ gilt: $\rho_X \to 0$. Für die Stelle τ_0, für die ρ_X einen festen Wert (meist $\frac{1}{\sqrt{2}}$) nicht mehr überschreitet (ρ_X ist symmetrisch um 0) gilt nach dem Wiener-Kintchine-Theorem die Beziehung:

$$B \propto \frac{1}{\tau_0}$$

Man nennt τ_0 auch Dekorrelationszeit, je größer τ_0, desto größer das Gedächtnis des Zufallssignal, desto kleiner also die Bandbreite die für die Übertragung benötigt wird.

3.2.3 Weißes Rauschen und Gauß'sches Rauschen

Weißes Rauschen ist ein Zufallssignal mit extremen spektralen und Autokorrelations-Eigenschaften. Normalerweise korrelieren Abtastwerte eines Signals umso stärker je kürzer das Abtastintervall ist. Weißes Rauschen hat jedoch kein „Gedächtnis" - es hat keine Ähnlichkeit mit zeitverschobenen Versionen von sich selbst. Die Autokorrelationsfunktion besteht aus einem einzelnen Impuls bei 0 (keine Verzögerung) und das Leistungsdichtespektrum ist flach (Wiener-Kintchine Theorem). Abtastwerte von weißem Rauschen sind immer unkorreliert, egal wie knapp hintereinander sie abgetastet wurden. Da diese Eigenschaft nicht natürlich ist, sind solche Signale nicht realisierbar.

Als Gauß'sches Rauschen bezeichnet man Rauschen, dessen Rauschamplitude Gauß-verteilt ist. Weißes Gauß'sches Rauschen ist Rauschen, dessen Rauschamplitude Gauß-verteilt ist und dessen Leistungsdichtespektrum flach ist. Weißes Gauß'sches Rauschen ist also definitionsgemäß auch Gauß'sches Rauschen, die Umkehrung gilt jedoch nicht, denn bei der Definition von Gauß'schem Rauschen wurde nur dessen Rauschamplitudenverteilung beschreiben und keine Aussage über die Korrelationseigenschaften, also das Leistungsdichtespektrum (Wiener-Kintchine Theorem!) gemacht.

3.2.4 Kreuzkorrelation von Zufallsprozessen

Für zwei reelle, ergodische Prozesse:

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t-\tau)dt$$

Die Fourier-Transformierte von $R_{xy}(\tau)$ wird Kreuz-Leistungsdichtespektrum mit Einheit V^2/Hz genannt. Die Kreuzkorrelation gibt natürlich wieder die Ähnlichkeit zweier um τ verschobener Zufallsprozesse an.

4 Lineare Systeme

Ein System ist eine Anordnung von miteinander verbundenen Komponenten zur Realisierung einer technischen Aufgabenstellung. Ein System kann als Operator aufgefasst werden, der Eingangsgrößen auf Ausgangsgrößen abbildet.

Ein System ist linear, wenn seine Antwort auf die Summe zweier Inputs gleich der Summe der Outputs auf jeden einzelnen Input ist (Überlagerungssatz, Superpositionsprinzip). Antworten auf einzelne Inputs werden am Ausgang einfach überlagert. Proportionalität folgt direkt aus der Linearität.

Lineare Systeme sind außerdem gedächtnislos, d.h. die momentane Ausgabe hängt nur von der momentanen Eingabe ab. Außerdem sind Systeme oft zeitinvariant, es gilt also $y(t) = y_1(t - T)$ falls $x(t) = x_1(t - T)$. Lineare, zeitinvariante Systeme (TILS, time invariant linear systems) sind in der Praxis sehr häufig und können durch folgende Bedingung beschrieben werden:

$$y_1(t) = S\{x_1(t)\} \wedge y_2(t) = S\{x_2(t)\} \Rightarrow S\{ax_1(t - T) + bx_2(t - T)\} = ay_1(t - T) + by_2(t - T)$$

Lineare Systeme verändern die Kurvenform von Sinusschwingungen am Eingang nicht.
Man beachte, dass es in der Praxis kaum Systeme gibt, die immer linear sind, denn die meisten Systeme sind nur in ihrem Betriebsbereich linear und werden, wenn sie in der Sättigung betrieben werden, nichtlinear. Um so interessanter ist, dass man viele Systeme in einer hinreichend kleinen Umgebung als linear betrachten kann (Taylorentwicklung, erster Term).

Beispiele für lineare Systeme sind:

- Elektrische Schaltkreise aus ohmschen Widerständen, Kondensatoren und Spulen
- Elektronische Schaltkreise wie Verstärker und Filter (innerhalb des Versorgungsspannungsbereichs)
- Mechanische Systeme aus Masse-Feder-Dämpfung
- Differentiation und Integration
- Ausbreitung von elektromagnetischen Wellen und Schallwellen in isotropen Medien
- Alle Systeme die durch lineare Differential- oder Differenzengleichungen beschrieben werden können

Beispiele für nichtlineare Systeme sind:

- Leistung eines elektrischen Widerstands $P(t) = R \cdot I^2(t)$
- Nichtlineare elektronische Schaltungen wie Spitzendetektoren, Quadrierer, Frequenzverdoppler, Schwellwertschalter, Komparatoren
- Nichtlineare Effekte in elektronischen Schaltkreisen wie Begrenzen, nichtlineares Verstärken (slew rate), Sättigungseffekte
- Ausbreitung von elektromagnetischen Wellen und Schallwellen in anisotropen Medien
- Alle Systeme deren Beschreibung auf nichlineare Differential- oder Differenzengleichungen führt.

4.1 Lineare Systeme im Zeitbereich

Sei folgende Differentialgleichung gegeben:

$$a_0 y + a_1 \frac{dy}{dt} + a_2 \frac{d^2 y}{dt^2} + \ldots + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} = b_0 x + b_1 \frac{dx}{dt} + b_2 \frac{d^2 x}{dt^2} + \ldots + b_{m-1} \frac{d^{m-1} x}{dt^{m-1}}$$

Jedes System, das sich durch solche lineare Differentialgleichungen beschreiben lässt, folgt dem Prinzip der Überlagerung (bzw. Superposition) und ist daher linear. Man kann die Antwort eines solchen Systems zusammensetzen (überlagern) aus der freien Antwort (Lösung des homogenen Systems, also $x(t) \equiv 0$, mit entsprechenden Anfangsbedingungen, also den Werten aller y-Ableitungen zum Zeitpunkt t_0) und der erzwungenen Antwort ($x(t)$ in die Gleichung eingesetzt, Anfangsbedingungen werden als 0 angenommen).

4.1.1 Ausgangssignal im Zeitbereich

Stellt man sich ein lineares System mit diskreten Inputs $x_1 \ldots x_n$ und diskreten Outputs $y_1 \ldots y_m$ vor, so ist jeder Output y_i durch eine gewichtete Summe aller Inputs gegeben[3]:

$$y_i = \sum_{j=1}^{n} G_{ij} x_j$$

Man beachte dass man wegen dieser Darstellung die Operation eines linearen Systems als Matrixmultiplikation interpretieren kann.
Ersetzt man die diskreten Inputs und Outputs durch kontinuierliche Äquivalente $y(t)$ und $x(\tau)$ (man verwendet hier τ als Variable für x da wir später auf eine Faltungsdarstellung kommen, wo wir getrennte Variablen benötigen... außerdem lässt diese Notation auch eine Beschreibung von Systemen zu, die nicht zeitinvariant sind), so wird aus der diskreten Summe das Integral

$$y(t) = \int_0^{N\Delta\tau} G(t,\tau) x(\tau) d\tau$$

wobei N Abtastwerte $\Delta\tau$ Sekunden auseinanderliegen. Für $\Delta\tau \to 0$ bekommt man allgemein

$$y(t) = \int_{-\infty}^{\infty} G(t,\tau) x(\tau) d\tau$$

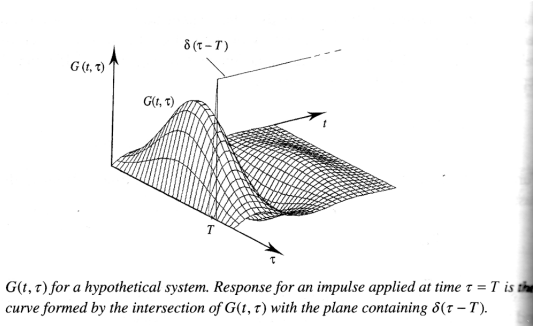
Für kausale Systeme, deren Ausgabewerte nicht von der Zukunft sondern nur von der Vergangenheit bis zum aktuellen Zeitpunkt abhängen, reicht als obere Schranke für das Integral t. Da man meist zu einem definierten Zeitpunkt t_0, meist $t_0 = 0$ mit der Betrachtung beginnt und in der Praxis nur kausale Systeme vorkommen hat man es meist mit der Darstellung

$$y(t) = \int_0^{t} G(t,\tau) x(\tau) d\tau$$

[3] Achtung: Die x_i bzw. y_j sind hier nicht n bzw. m Input/Output-Pins, sondern eine zeitliche, gesampelte Abfolge von *einem* Eingangswertes bzw. Ausgangswertes

zu tun. Für Vektoren als Input und Output ist $G(t,\tau)$ eine Matrix mit Funktionen in jedem Eintrag, die Integration ist dann komponentenweise zu verstehen.

Wird als Input die Impulsfunktion verwendet, so erhält man die Impulsantwort $h(t)$ des Systems. Da $\delta(\tau - T)$ eine Samplingfunktion ist kann $G(t,T)$ als Antwort auf einen Impuls, der zum Zeitpunkt $\tau = T$ angewendet wird, interpretiert werden, d.h. $h(t-T) = G(t,T)$. Der 3D-Plot von $G(t,\tau)$ kann daher als Impulsantwort zu allen möglichen Zeiten interpretiert werden (siehe Bild).



G(t, τ) for a hypothetical system. Response for an impulse applied at time τ = T is the curve formed by the intersection of G(t, τ) with the plane containing δ(τ − T).

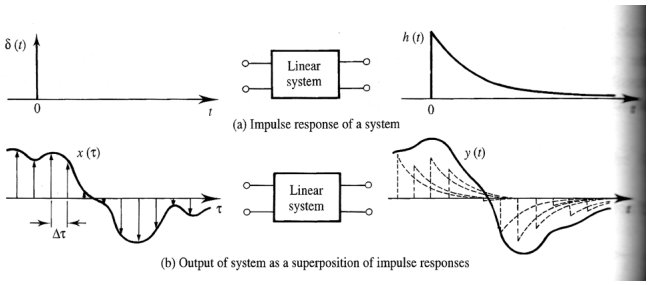
Setzt man nun dieses Äquivalent in das obige Integral ein, so ergibt sich:

$$y(t) = \int_{-\infty}^{t} h(t-\tau)x(\tau)d\tau$$

Dieses Integral ist unschwer als Faltungsintegral zu erkennen. Der Output eines zeitinvarianten linearen Systems ist demnach durch die Faltung von Input und Impulsantwort gegeben:

$$y(t) = h(t) * x(t)$$

In der Praxis kann man sich ein lineares System vorstellen als System, das ein kontinuierliches Eingangssignal abtastet, sein Ausgangssignal ist dann die (durch die Abtastwerte des Eingangssignals) gewichtete Summe von Impulsantworten. Im Grenzfall $\tau \to 0$ bekommt man genau obige Faltung.



4.1.2 Sprungantwort

Die Sprungfunktion ist wie folgt definiert:

$$u(t) = \begin{cases} 1.0, & t > 0 \\ 0.5, & t = 0 \\ 0, & t < 0 \end{cases}$$

Setzt man die Sprungfunktion in das Faltungsintegral ein, so können die Integrationsgrenzen auf 0 und t eingeschränkt werden, da $u(t-\tau) = 0$ bei $\tau > t$ ist. Weiters ist aber $u(t)$ im Bereich 0 bis t gleich 1 und somit ergibt sich für die Sprungantwort $q(t)$:

$$q(t) = \int_0^t h(\tau)d\tau \qquad \text{bzw.} \qquad h(t) = \frac{d}{dt}q(t)$$

Die Sprungantwort ist besonders deswegen sinnvoll, weil sie in der Praxis leichter zu bestimmen ist als die Impulsantwort.

4.2 Lineare Systeme im Frequenzbereich

Im Frequenzbereich ist der Output eines linearen Systems durch Anwendung der Fouriertransformation und des Faltungstheorems auf beide Seiten der Ausgangsgleichung im Zeitbereich gegeben:

$$Y(f) = H(f)X(f)$$

Hierbei ist $Y(f)$ das Ausgansspannungsspektrum, $X(f)$ das Eingangsspannungsspektrum und $H(f)$ die Frequenzantwort des Systems. Bei Frequenz f_0 ist die Frequenzantwort eine komplexe Zahl, die die Spannungsverstärkung (oder -dämpfung) sowie die Phasenverschiebung einer Sinusschwingung mit Frequenz f_0 angibt.

4.3 Zufallssignale und lineare Systeme

Zufällige Signale können nicht als deterministische Funktionen wie $x(t)$ spezifiziert werden. Ihre Eigenschaften werden durch ihre Dichtefunktion und das Leistungsdichtespektrum festgelegt.

4.3.1 Leistungsdichtespektren linearer Systeme

Quadriert man die Beschreibung eines linearen Systems im Frequenzbereich und betrachtet nur die Amplitude, so erhält man $|Y(f)|^2 = |H(f)X(f)|^2$. Da $|Y(f)|^2$ und $|X(f)|^2$ Leistungsdichtespektren sind, ergibt sich

$$G_y(f) = |H(f)|^2 G_x(f)$$

Über die Fouriertransformation und das Wiener-Kintchine-Theorem ergibt sich schließlich für den Zeitbereich:

$$R_{yy}(\tau) = R_{hh}(\tau) * R_{xx}(\tau)$$

Die R-Funktionen sind Autokorrelationsfunktionen. So kann man zwar nichts über das genaue zeitliche Verhalten des Ausgangssignals eines linearen Systems, das mit einem Zufallssignal angeregt wird, aussagen, aber zumindest etwas über die Autokorrelationsfunktion des Ausgangssignals.

4.3.2 Rauschbandbreite

Die Rauschbandbreite eines Filters ist die Breite, die eine rechteckige Frequenzantwort bräuchte, um die gleiche Rauschleistung wie der Filter zu haben. Ist f_p die Frequenz der höchsten Amplitude der

Antwort, so ist die Rauschbandbreite B_N gegeben durch:

$$B_N = \int_0^\infty \frac{|H(f)|^2}{|H(f_p)|^2} df$$

Die Division durch $|H(f_p)|^2$ bewirkt hier eine Normierung der Frequenzantwort auf eine Funktion mit Maximum 1. Die Rauschbandbreite ist nicht gleich der 3dB-Bandbreite (also der $1/\sqrt{2}$-Definition). Bei einem Tiefpass-Filter ist die Rauschbandbreite um den Faktor $\pi/2$ größer als die 3dB-Bandbreite.

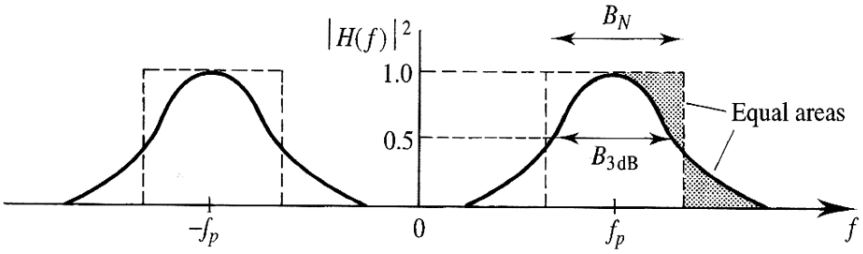


Illustration of noise bandwidth, B_N.

Im Bild sieht man schön die Normierung des Frequenzganges (Maximum 1) und dass $B_N \neq B_{3dB}$.

4.3.3 Wahrscheinlichkeitsdichte von gefiltertem Rauschen

Bei nicht-gedächtnislosen Systemen gibt es keine generelle Methode um von der Eingabe- auf die Ausgabedichte zu schließen, eine Ausnahme ist die Dichte von gefiltertem Gauß'schem Rauschen. Das Ausgangsrauschen ist durch die Faltung von Eingangsrauschen und Frequenzantwort gegeben. Man kann also, wie man an der folgenden Faltungsdarstellung des Outputs sehen kann, den Output auch als Summe gewichteter Inputimpulse sehen.

$$n_{0,i}(t) = \int_{-\infty}^{t} h(t-\tau) n_i(\tau) d\tau$$

Sind die Inputimpulse nun weißes Gauß'sches Rauschen, so sind die einzelnen benachbarten Impulse unabhängige (wegen weißem Rauschen) Gauß'sche Zufallsvariablen. Das Ausgangsrauschen zu jedem Zeitpunkt ist daher eine lineare Summe Gauß'scher Zufallsvariablen und daher selbst eine Gauß'sche Zufallsvariable. Die Ausgangsamplitude nach dem Filtern ist also Gauß-verteilt. Da ein Filter in der Praxis aber endliche Bandbreite hat, und das flache Eingangsspektrum beim Filtern daher begrenzt wird, ist das Ausgangsspektrum nicht mehr weiß.

$$\Rightarrow \text{Gefiltertes weißes Gauß-Rauschen ist Gaußsch}$$

Durch ein Gedankenexperiment (siehe Bild), in dem man einen einzigen Filter durch 2 Teilsysteme ersetzt kann man daraus einen weiteren Schluss ziehen: Da im ersten Teilsystem aus Gaußschem, weißen Rauschen Gaußsches Rauschen wird (obige Schlussfolgerung) und der Output des Gesamtsystems Gaußsches Rauschen ist (ebenfalls obige Schlussfolgerung), muss gelten:

$$\Rightarrow \text{Gefiltertes Gaußsches Rauschen ist Gaußsch}$$

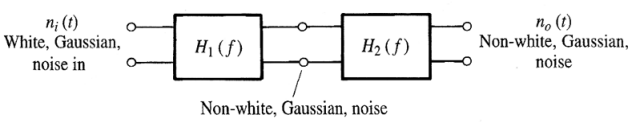


Figure 4.18 *Reinterpretation of $H(f)$ in Figure 4.17 as two cascaded sections.*

4.4 Nichtlineare Systeme und Transformation von Zufallsvariablen

Nichtlineare Systeme sind schwer zu analysieren, jedoch kann die Dichte des Outputs von gedächtnislosen nicht-linearen Systemen durch die Transformation von Zufallsvariablen leicht ermittelt werden. Zum Beispiel ist ein System mit $y(t) = x^2(t)$ nichtlinear, dennoch kann die Dichte der transformierten Zufallsvariablen leicht bestimmt werden. Die Wahrscheinlichkeit der transformierten Variable ist durch die Wahrscheinlichkeit des Urbilds gegeben, man bezeichnet das auch als Gesetz der Erhaltung der Wahrscheinlichkeit.

Wenn Gauß'sches Rauschen am Eingang eines Hüllkurvendetektors[4] anliegt, so ist die Ausgangsdichte Rayleigh-verteilt. Eine Rayleigh-Verteilung mit Parameter σ hat folgende Dichtefunktion:

$$f_\sigma(x) = \frac{xe^{-\frac{x^2}{2\sigma^2}}}{\sigma^2}$$

Das sieht man unmittelbar, da bei einem Hüllkurvendetektor natürlich nur die Amplitude des Signals interessiert und nicht die Phase; nach Substitution in Polarkoordinaten folgt die Behauptung also unmittelbar aus der (aus der WK-Theorie bekannten) Aussage[5]:

$$\text{Sind } X \sim N(0,\sigma^2),\ Y \sim N(0,\sigma^2), \text{ so ist } R = \sqrt{X^2+Y^2} \text{ Rayleigh}(\sigma) \text{ verteilt}$$

Wenn Gauß'sches Rauschen am Eingang eines Quadrierers[6] anliegt, so ist die Ausgangsdichte Chi-Quadrat-verteilt mit einem Freiheitsgrad.
Allgemein ist eine Chi-Quadrat-Verteilung mit n Freiheitsgraden die Summe von n quadrierten, unabhängigen, normalverteilten Zufallsvariablen, die alle gleiche Varianz σ^2 haben:

$$Y = X_1^2 + \ldots X_n^2$$

Es gilt:

$$\overline{Y} = N\sigma^2, \qquad \sigma_Y^2 = 2N\sigma^4$$

5 Abtasten, Multiplexen und PCM

Man spricht von einem Basisband-Signal, wenn $B \geq f_L$ gilt und von einem Bandpass-Signal wenn $B < f_L$ gilt.

[4] Wird zur Demodulation von AM-Signalen verwendet
[5] siehe z.B.: http://de.wikipedia.org/wiki/Rayleighverteilung
[6] engl. square law device

5.1 Pulsmodulation

Pulsmodulation beschreibt einen Prozess, bei dem die Amplitude, Breite oder Position einzelner Pulse einer periodischen Pulsreihe entsprechend der Amplitude des (analogen) Basisband-Informationssignals variiert werden. Pulsamplitudenmodulation (PAM) benötigt einen höheren Signal-Rauschabstand[7] als Pulspositionsmodulation (PPM) und Pulsweitenmodulation (PWM), da sich Störungen stark auf die Amplitude auswirken. Ist die Pulsrate hoch genug (Nyquist-Kriterium), so kann das Informationssignal aus dem pulsmodulierten Signal rekonstruiert werden; besonders einfach ist das für PAM (Tiefpass-Filter) bzw. PWM (Integrator, also Kondensatorschaltung). Man beachte jedoch, dass die Bandbreite die für Pulsmodulation benötigt wird aufgrund der steilen Flanken höher ist als jene des ursprünglichen Signals.

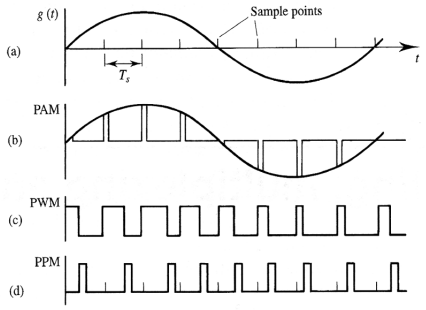


Illustration of pulse amplitude, width and position modulation: (a) input signal.

5.2 Abtasten

Unter Abtasten (Sampling) versteht man das Aufzeichnen der Ordinaten-Werte (Amplitude) einer kontinuierlichen Funktion, normalerweise zu äquidistanten Werten der Abszisse (Zeit).

Werden bei einer Abtastrate f_s ein konstantes Gleichstromsignal und ein periodisches Inputsignal mit einer Frequenz, die ein ganzzahliges Vielfaches von f_s beträgt, abgetastet so ergeben beide die gleichen Ausgabewerte. Daraus folgt, dass sich im Frequenzbereich das abgetastete Basisbandspektrum bei f_s und ganzzahligen Vielfachen von f_s wiederholt.

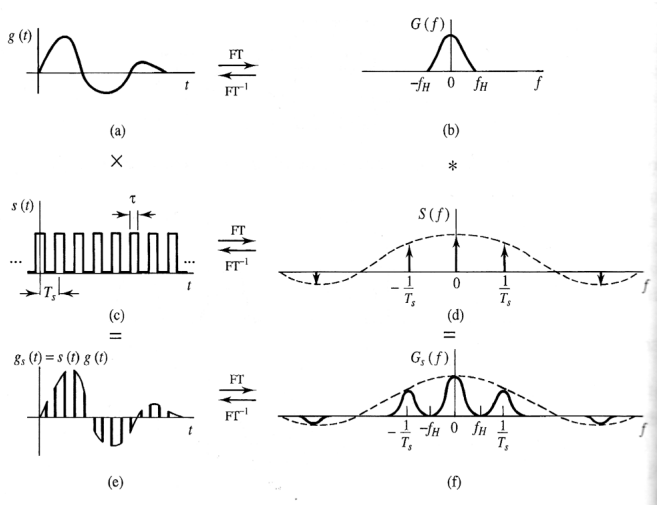
Es gibt zwei Arten von Sampling:

Natural Sampling Ein natürlich abgetastetes Signal ergibt sich, wenn man das Basisband-Informationssignal mit der periodischen Pulsreihe multipliziert. Dabei entspricht die Form der Pulsspitzen der Form des abgetasteten Signals. Im Frequenzbereich entspricht der Multiplikation die Faltung und so ergibt sich ein Signalspektrum, bei dem sich das Spektrum des Informationssignals, gewichtet durch die Amplituden des Pulsreihen-Spektrums, um die ganzzahligen Vielfachen von f_s wiederholt (diese Eigenschaft folgt hier aus der Faltung, und ist eine nochmalige Bestätigung der Überlegungen von vorher). Dass das Spektrum der periodischen Pulsreihe tatsächlich $\tau sinc(\tau f) \cdot \sum_{n=-\infty}^{\infty} \delta(f - nf_s)$ ist folgt aus dem Faltungstheorem, wenn man die periodische Pulsreihe im Zeitbereich als Faltung von *einem* Rechteck der Breite τ und der aus Impulsfunktionen bestehenden Samplingfunktion anschreibt. Nun ist auch klar, warum ein Tiefpassfilter

[7] SNR, Signal to Noise Ratio

vonnöten ist, um das (Spektrum des) Informationssignals rückzugewinnen. Einen solchen Filter nennt man auch einen Rekonstruktionsfilter, da er das Signal rekonstruiert.

Flat Top Sampling Beim natural Sampling entsprechen die Pulsspitzen des Resultats der Form des abgetasteten Signals. Beim flat top Samping werden sie künstlich abgeflacht. Dieses Signal erhält man, wenn das Informationssignal mit einer Impulsreihe abgetastet (multipliziert) wird und die resultierenden, gewichteten Impulse mit einem Rechteckimpuls gefaltet werden. Ebenso wird das Frequenzspektrum des Resultats mit dem Spektrum des Rechteckimpulses (sinc-Funktion) multipliziert. Um hier das (Spektrum des) Informationssignals rückzugewinnen muss nach dem Tiefpass noch eine Entzerrung (Equalisation), nämlich eine Multiplikation mit der Inversen des Rechteckpulsspektrums ($\frac{1}{sinc(x)}$), durchgeführt werden.



ıre 5.4 *Time and frequency domain illustrations of natural sampling: (a) signal g(t); (b) signal spectrum; (c) sampling function; (d) spectrum of sampling function; (e) sampled signal; (f) spectrum of sampled signal.*

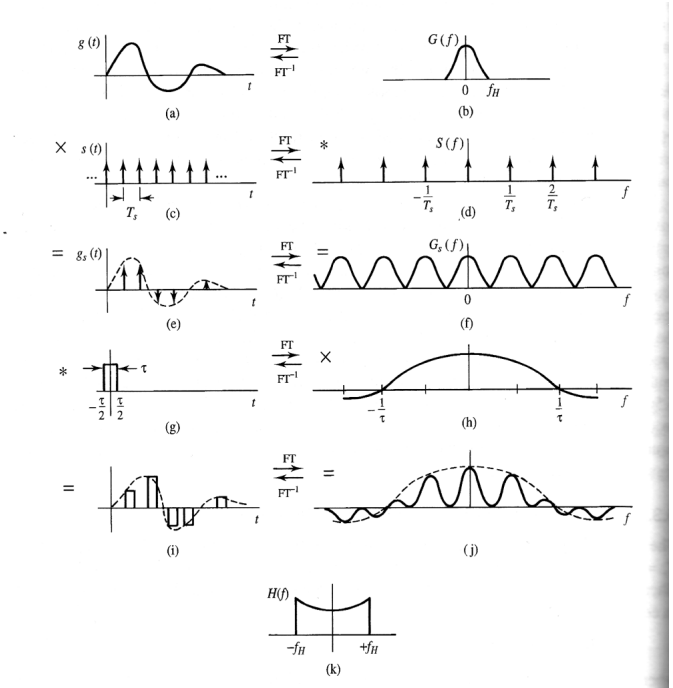


Figure 5.5 *Time and frequency domain illustrations of PAM or flat topped sampling: (a) signal; (b) signal spectrum; (c) sampling function; (d) spectrum of (c); (e) sampled signal; (f) spectrum of (e); (g) finite width sample; (h) spectrum of (g); (i) sampled signal; (j) spectrum of (i); (k) receiver equalising filter to recover g(t).*

5.3 Aliasing

Wird ein Basisband-Signal mit einer zu niedrigen Frequenz abgetastet, so überlappen sich die wiederholenden Frequenzspektren. Das gefilterte Signal enthält dann auch die „abgeschnittenen“ Frequenzen in „zurückgeklappter“ Form. Um das zu verhindern kann man entweder vor dem Samplen einen Antialiasing-Filter verwenden, oder man stellt sicher dass die Abtastrate dem Nyquisttheorem entspricht.
In der Praxis muss man außerdem beachten, dass ein Signal endlicher Dauer unendliche Bandbreite, also unbeschränktes Spektrum hat. Die höchste, vorkommende Frequenz f_H muss daher in der Praxis als die höchste relevante Frequenz gewählt werden. Außerdem gibt es keine idealen Filter mit rechteckiger Übertragungsfunktion, folglich muss $f_S \geq 2f_H$ in der Praxis oft durch $f_S \geq 2.2f_H$ ersetzt werden.
Als quantitatives Maß für die Güte einer Rekonstruktion verwendet man den Signal-Störungs-Abstand (SDR, signal to distortion ratio):

$$SDR := \frac{\int_0^{\frac{f_S}{2}} G(f)df}{\int_{\frac{f_S}{2}}^{\infty} G(f)df}$$

bzw. in der Praxis, um die Eigenschaften des Rekonstruktionsfilters einfließen zu lassen:

$$SDR := \frac{\int_0^{\infty} G(f)|H(f)|^2df}{\int_0^{\infty} G(f-f_S)|H(f)|^2df}$$

wobei in beiden Fällen $G(f)$ das Spektrum des abzutastenden Signals ist. Man beachte, dass hier nur Amplituden, jedoch keine Phasenverzerrungen betrachtet werden.

5.4 Abtasten von Bandpassignalen

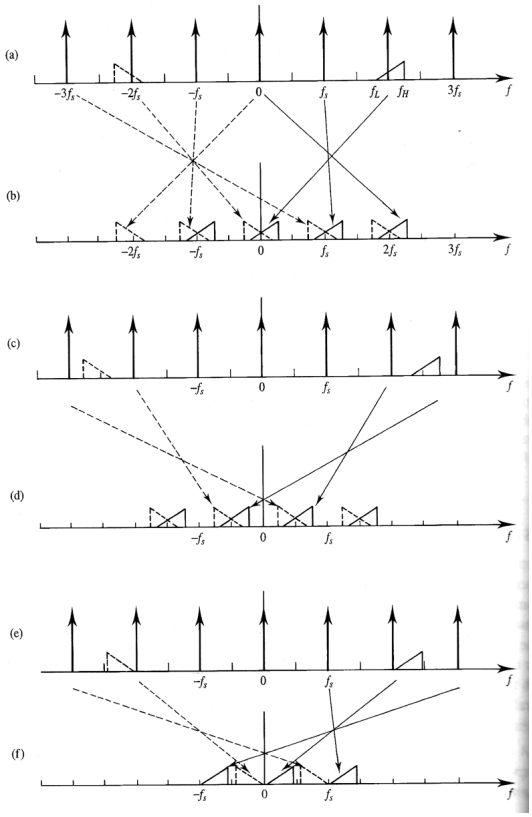
Ist die zentrale Frequenz weit höher als die Signalbandbreite, so ist es möglich, auch mit einer Abtastrate unter der Nyquist-Rate das Signal so zu erfassen, dass es nachher wieder vollständig rekonstruiert werden kann. Die Abtastrate f_s muss dann in folgendem Frequenzband liegen:

$$2B \left\{ \frac{Q}{n} \right\} \leq f_s \leq 2B \left\{ \frac{Q-1}{n-1} \right\}$$

wobei $Q = f_H/B$ und $n := \lfloor Q \rfloor$ das Frequenzband festlegt. Ist Q eine Ganzzahl so ist mit $n = Q$ die Abtastrate genau auf die Nyquist-Rate festgelegt. Ist $Q < 2$ so gilt $n = 1$ und man erhält (da $BQ = f_H$) das Nyquist-Kriterium.

Die Richtigkeit des Abtastkriteriums ergibt sich bei Betrachtung folgender Fälle der Faltung des Frequenzspektrums des Bandpassignals mit der Impulsreihe:

- Das Bandpassspektrum kreuzt nf_s - bei der Faltung ergeben sich Interferenzen zwischen positiven und negativen spektralen Frequenzrepliken, mit Mittelpunkt nf_s.
- Das Bandpassspektrum kreuzt $(n + \frac{1}{2})f_s$ - es ergeben sich ähnliche Interferenzen, mit Mittelpunkt $(n + \frac{1}{2})f_s$.
- Das Bandpassspektrum kreuzt keine der obigen Frequenzen und es ergibt sich für korrektes Abtasten: $2f_H \leq nf_s$ und $2f_L \geq (n-1)f_s$, was umgeformt (mit $Q := f_H/2$) obiges Abtastkriterium ergibt.



Bei der Rekonstruktion muss man natürlich beachten, dass man nun ein Bandpasssignal rekonstruieren will. Wurde das Basisbandsignal nur aus übertragungstechnischen Gründen in ein Bandpasssignal gewandelt und will man das Basisbandsignal rekonstruieren, dann kann man wie bisher mit Tiefpass arbeiten.

5.5 Multiplexen analoger Impulse

Oft müssen mehrere Signale über das gleiche Medium (Kabel, Funk, ...) übertragen werden. Die Signale müssen daher genügend getrennt werden - die Güte ihrer Getrenntheit wird mit Orthogonalität bezeichnet. Orthogonale Signale können unabhängig voneinander empfangen werden. Orthogonalität kann auf mehrere Arten erreicht werden:

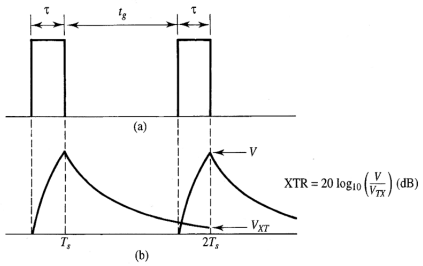
Frequency Division Multiplexing Bei FDM werden die Signale auf unterschiedliche Trägerfrequenzen aufmoduliert (d.h. es wird das Signal mit dem Träger multipliziert, $g(t) \cdot cos(\omega t)$) und belegen so unterschiedliche Frequenzbänder des Kanals. Empfängerseitig können sie dann mittels eines Bandpassfilters rückgewonnen werden. Bei Glasfaser-Kanälen wird das gleiche Prinzip mit Wellenlängen verwendet und Wavelenght Division Multiplexing genannt.

Time Division Multixplexing Bei TDM werden die PAM-Signale auf unterschiedliche Zeitschlitze aufgeteilt. Dabei steigt allerdings die Bandbreite, da die Abstände zwischen Pulsen entsprechend verkürzt werden. Die nötige Bandbreite kann jedoch stark reduziert werden, denn an den

Abtastzeitpunkten muss die Amplitude korrekt sein, dazwischen jedoch kann man die Funktion glätten und so die Bandbreite reduzieren. Das jedoch setzt voraus, dass Sender und Empfänger den selben Abtastzeitpunkt verwenden... kein Vorteil ohne Nachteil.
Die Kanäle können auch Übersprechen, wenn das Medium die Bandbreite limitiert und die Flanken daher exponentiell mit einer großen Zeitkonstante RC abklingen. Als Maß für das Verhältnis von Signal und übersprechendem Signal gilt XTR (cross talk ratio):

$$XTR = 20\ log_{10} e^{\frac{t_g+\tau}{RC}} \quad (DB)$$

Dabei bezeichnet τ die Breite des Impulses und t_g den Abstand zwischen zwei Pulsen, siehe Bild. Man beachte, dass man XTR nur verbessern kann, indem man die Anzahl der übertragenen Pulse pro Zeiteinheit verkleinert (also $t_g + \tau$ vergrößert) oder ein anderes Medium verwendet (RC verkleinert), durch kürzere Pulse alleine (τ verkleinern) gewinnt man nichts. Das mag kontraintuitiv sein, man beachte jedoch, dass der Abtastzeitpunkt nicht in die Rechnung eingeht, ein größeres τ vergrößert daher auch den Beitrag des beabsichtigten Signals.



Crosstalk between tributary channels of a TDM signal: (a) signal at channel input; (b) signal at channel output.

5.6 Quantisierte Pulsamplitudenmodulation

Ein PAM-Signal ist zwar diskret im Zeitbereich, aber kontinuierlich im Wertbereich, da alle Amplitudenwerte erlaubt sind - die Dichte der Pulsamplituden ist kontinuierlich. Wird das PAM-Signal quantisiert (ADC), d.h. die Amplituden können nur mehr einen Wert aus einer endlichen Menge an Werten annehmen, wobei natürlich immer der der tatsächlichen Amplitude am nächsten gelegene ausgewählt wird, so ist das entstehende Signal digital und besitzt eine diskrete Dichte. Das digitale Signal kann durch eine endliche Menge an Symbolen repräsentiert werden, wobei ein Symbol einem Quantisierungslevel entspricht.

Quantisieren ist mit Qualitätsverlust verbunden, da das quantisierte Signal nicht mehr dem Originalsignal entspricht. Der Unterschied zwischen den letzteren ergibt ein neues Zufallssignal, das Quantisierungsrauschen.
Als Maß für das Quantisierungsrauschen verwendet man den sogenannten SN_qR (signal to noise ratio), also das Verhältnis von Signal zu Quantisierungsrauschen. Bei äquidistanten Quantisierungsstufen mit Abstand q zwischen den Quantisierungsstufen bekommt man:

$$SN_qR := \frac{\overline{v^2}}{\overline{\epsilon_q^2}} = \frac{\int_{v_{min}}^{v_{max}} v^2 p(v) dv}{\int_{-q/2}^{q/2} \epsilon_q^2 p(\epsilon_q) d\epsilon_q}$$

Dabei bezeichnet v das Signal, ϵ_q das Quantisierungsrauschen, q den Abstand zwischen Quantisierungsstufen und $p(v)$ die Wahrscheinlichkeitsdichte des Signals. Hier wird also nach $dP(v)$ bzw. $dP(\epsilon_q)$ integriert, man betrachtet also das Verhältnis der beiden Erwartungswerte (Mittelwerte). Außerdem betrachtet man noch den Wert $(SN_qR)_{peak}$, das maximal auftretende Verhältnis von Signal zu Quantisierungsrauschen.
Setzt man lineare Quantisierung und ein Signal ohne Gleichanteil mit Gleichverteilung voraus, so berechnet man für die PAM leicht (M bezeichnet die Anzahl der Quantisierungsstufen):

$$SN_qR = M^2 - 1 = 20\, log_{10}(M) \quad dB$$

$$(SN_qR)_{peak} = 3M^2 = 4.8 + SN_qR \quad dB$$

(Erinnerung: Umrechnung in dB mit $10\, log_{10}$)
Das SN_qR wird also für steigendes M mit dem Quadrat der Anzahl an Quantisierungsstufen besser.

5.7 Pulscodemodulation

Nachdem ein PAM-Signal quantisiert wurde, kann man anstatt der Pulse selbst lediglich Nummern, die die Höhe der Pulse angeben, übertragen. Acht Amplitudenlevel könnten durch eine dreistellige Binärzahl[8] repräsentiert werden, wobei die Einsen und Nullen durch unterschiedliche Spannungen dargestellt werden. Generell ist bei M Amplitudenwerten das Codewort $n = \lceil \log_2 M \rceil$ Bit lang. SN_qR hängt nicht von der Kodierung ab und bleibt daher gleich, als Funktion von n haben wir hier also $SN_qR = (2^n)^2 - 1$ und $(SN_qR)_{peak} = 3(2^n)$. Als Schätzwert für SN_qR verwendet man oft $SN_qR = 6(n-1) \quad dB$.

PCM benötigt mehr Bandbreite als das native PAM-Signal, da mehr Pulse statt einem übertragen werden. Jedoch können die Spannungslevel bei PCM weitaus leichter unterschieden werden als bei PAM, PCM ist also weniger anfällig auf Rauschen. Bei der Berechnung des SNR muss man jedoch einige Feinheiten beachten, z.B. ist der Fehler davon abhängig, welches der Bits gestört ist (LSB ist weniger wichtig als MSB). Man rechnet daher meist mit nur einem Bitfehler pro Codewort und setzt voraus, dass die Wahrscheinlichkeit, dass ein Bit gestört wird, gleichverteilt ist über alle Bits mit Wahrscheinlichkeit P_e. Da Quantisierungsrauschen und das Rauschen auf der Leitung statistisch unabhängig sind addieren sich die Erwartungswerte (der quadrierten, transformierten Zufallsvariablen), wir haben:

$$SNR = \frac{\overline{v^2}}{\overline{\epsilon_q^2} + \overline{\epsilon_{de}^2}}$$

wobei $\overline{\epsilon_{de}^2}$ den Erwartungswert des Dekodierungsfehlers angibt. Man rechnet nach:

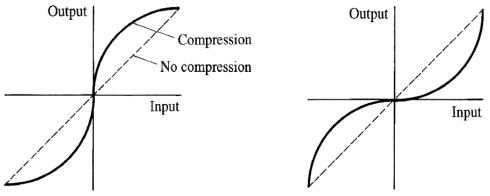
$$SNR = \frac{SN_qR}{1 + 4P_eSN_qR}$$

5.7.1 Companded PCM

Bei „gewöhnlichen“ Signalen kann nicht davon ausgegangen werden, dass alle Quantisierungslevel gleichmäßig genutzt werden, d.h. die Dichte des Informationssignals gleichmäßig verteilt ist. Um das

[8] ein Codewort → PCM

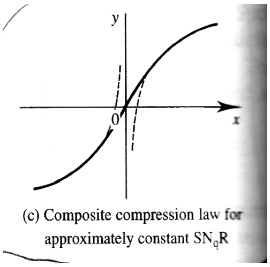
Quantisierungsrauschen zu minimieren ist es also sinnvoll, die Quantisierungslevel an den häufig genutzten Stellen zu verdichten und an den weniger genutzten Stellen zu erweitern. Diese nicht-lineare Quantisierung nennt man Companding[9], da das Signal mit einer nicht-linearen Amplitudencharakteristik komprimiert und dann das empfangene Signal mit der inversen Charakteristik dekomprimiert wird. Nach der Komprimierung hat die transformierte Zufallsvariable idealerweise eine Gleichverteilung.



(a) Compression characteristic (b) Expansion characteristic

Typical compression and expansion (compander) characteristics.

Leider ist die Dichte des Informationssignals oft nicht bekannt, z.B. bei der menschlichen Stimme. In diesem Fall ist die übliche Companding-Strategie, einen konstanten SN_qR für alle Signallevel zu erzielen. Da die Quantisierungsrauschleistung proportional zu q^2 ist, wobei q den Abstand zwischen zwei Quantisierungslevel bezeichnet, muss q proportional zum Signallevel, d.h. $\frac{v}{q}$ konstant sein. Dazu muss der Fehler bei kleinen Amplitudenwerten des Signals klein sein, bei großen Werten ist der absolute Fehler nicht mehr so relevant. Daher scheint der Logarithmus eine geeignete Funktion zu sein, er hat die Eigenschaft, die multiplikative Einteilung der Amplitudenwerte in eine Skala in einen additiven Fehler beim Quantisierungsrauschen zu verwandeln. Der konstante Quantisierungsrauschabstand wird daher durch eine logarithmische Funktion realisiert, die noch für negative reelle Zahlen fortgesetzt wird (Amplituden können auch negativ sein) und von -1 bis 1 durch eine Gerade approximiert wird (siehe Bild). Beispiele für solche Kurven sind das A-law (verwendet für Telefonie in Europa) und das μ-law (Telefonie USA).



(c) Composite compression law for approximately constant SN_qR

5.7.2 PCM Multiplexing

PCM-Multiplexing kann auf zwei verschiedene Arten erfolgen:

- Ein bereits TDM-gemultiplextes (PAM-)Signal wird PCM-kodiert

[9] Companding = Compression & Expanding

- Die zu multiplexenden Signale werden zuerst PCM-kodiert und dann per TDM gemultiplext

Die zweite Methode hat den Vorteil, dass die A/D-Wandlung näher an der Signalquelle stattfindet (man will möglichst früh digitalisieren) und dass nicht nur ganze Codewörter, sondern auch die einzelnen Bits der Codewörter ineinander geschachtelt werden können.

5.8 Möglichkeiten zur Bandbreitenreduktion

Da Bandbreite ein begrenztes und wertvolles Gut ist und es außerdem teuer ist z.B. zusätzliche Leitungen zu legen muss die vorhandene Bandbreite effizient genutzt werden. Dies geschieht durch spektral effiziente Signalgebungstechniken für (Sprach-)Signale.

5.8.1 Delta PCM

Bei Delta PCM wird statt der Abtastwerte selbst nur die Differenz zum vorangegangenen Abtastwert übertragen. Da die Differenz üblicherweise kleiner als die tatsächlichen Abtastwerte ist, da nahe bei einanderliegende Abtastwerte üblicherweise korrelieren, sind kürzere Codewörter vonnöten. Delta PCM kann sich schnell ändernde Signale allerdings nicht so gut behandeln wie konventionelles PCM.

5.8.2 Differentielles PCM

Differentielles PCM (DPCM) macht sich ebenfalls die Korrelation benachbarter Abtastwerte (Bsp.: Bilder, ist ein Pixel schwarz, so hat man eine große Chance dass ein benachbarter Pixel auch ein Schwarzton ist) zunutze und benutzt einen Algorithmus um zukünftige Abtastwerte vorauszusagen. Übertragen wird dann nur ein Wert zur Korrektur des vorhergesagten Signals - dieses Korrektursignal beschreibt also den unvorhersehbaren Teil des Informationssignals. Die Vorhersage-Einheit (sowohl Transmitter als auch Receiver haben eine solche) besteht oft aus einer linear gewichteten Summe vorheriger Abtastwerte und wird durch ein Schieberegister realisiert.

5.8.3 Adaptives DPCM

Bei adaptivem DPCM (ADPCM) sind die Koeffizienten bzw. Gewichte der Vorhersage-Einheit nicht fix festgelegt, sondern werden entsprechend der sich ändernden Signalstatistik (Bsp.: zuerst wird ein helles, dann ein dunkles Bild übertragen $\Rightarrow$ Erwartungswert wird dann „dünkler“) des Informationssignals angepasst. Die neuen Koeffizienten der Vorhersageeinheit (auch hier oft gewichtete lineare Summe) werden auch übertragen.

5.8.4 Deltamodulation

Wird bei einem DPCM-System die Auflösung des Quantisierers auf 1 bit reduziert, so erhält man das Schema der Deltamodulation... es wird also in jedem Schritt das Signal nur um $\pm\Delta$ verändert. Der Vorhersage-Algorithmus wird darauf beschränkt, stets anzunehmen, dass der neue Abtastwert gleich dem vorigen ist, es wird also einfach ein „One-Sample-Delay“ eingeführt.

Aus den Anforderungen für geringes Quantisierungsrauschen und slope overload noise ergibt sich ein Konflikt: um das Quantisierungsrauschen klein zu halten sollte die Korrekturschrittgröße

Δ möglichst klein sein um das Signal fein aufzulösen; um slope overload noise zu verkleinern, also schnell auf große Änderungen reagieren zu können, sollte Δ relativ groß sein (siehe Bild). Um das Rauschen beider Arten gering zu halten, kann man die Δ klein halten und dafür die Abtastfrequenz erhöhen (dann hat man mehr Schritte und kann schneller auf starke Steigungen reagieren). Muss man mit der Abtastfrequenz allerdings zu stark hinaufgehen, so sind die Bandbreiten-Vorteile der Deltamodulation schnell weg.

Bleibt das Informationssignal konstant für eine nennenswerte Zeitdauer, so wird das resultierende rechteckige Quantisierungsrauschen Ruherauschen[10] genannt. Durch einen Glättungsfilter wird dieses Rauschen, das üblicherweise die halbe Sampling-Frequenz hat, weggefiltert. Das geht deswegen, da die Samplingfrequenz bei Deltamodulation meist weit höher ist als $2f_H$, Ruherauschen ist also *hochfrequent*.

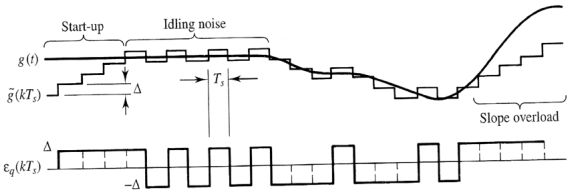


Figure 5.33 *DM signal waveforms illustrating slope overload and quantisation noise.*

Ist der Signalrauschabstand nicht ausreichend hoch, so interpretiert der Empfänger gelegentlich ein empfangenes Symbol falsch und es kommt zu einem Fehler in Höhe von $+\Delta - (-\Delta) = 2\Delta$. Man beachte, dass dieser Fehler, im Vergleich zu normalem PCM, im nächsten übertragenen Codewort nicht notwendig korrigiert wird. Erst mit dem nächsten Fehler, der den ersten Fehler ausgleicht, findet eine Korrektur statt, in der Zwischenzeit hat man einen konstanten Fehler (siehe Bild). Allerdings sind, wenn die Bitfehlerwahrscheinlichkeit P_e klein ist, diese Fehler niederfrequent, und man kann sie daher meist filtern, vor allem wenn f_L entsprechend groß ist. Das drückt sich auch in nachfolgender Formel aus, die die durchschnittliche Leistung dieses Fehlersignals angibt (f_L im Nenner):

$$N_e = \frac{\Delta^2}{\pi} \frac{f_S}{f_L} P_e$$

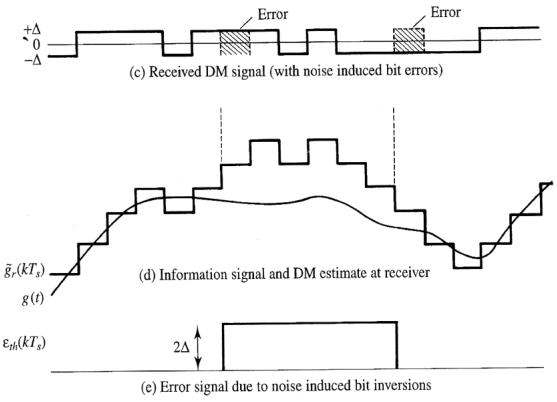


Figure 5.35 *Stepped error signal in DM receiver due to thermal noise.*

[10] idle noise

5.8.5 Adaptive Deltamodulation

Bei konventioneller Deltamodulation wird, um sowohl Quantisierungs- als auch slope overload noise gering zu halten, Δ klein und die Abtastfrequenz hoch gewählt. Bei adaptiver Deltamodulation ist Δ variabel. Die Korrekturschrittgröße wird dabei gemäß der Vergangenheit des Quantisierungsfehlers ϵ_q angepasst - ist ϵ_q immer gleich $+\Delta$, so steigt das Informationssignal schneller, als das vorhergesagte Signal folgen kann und Δ wird erhöht. Ist ϵ_q abwechselnd $-\Delta$ und $+\Delta$, so ändert sich das Informationssignal nur langsam und Δ wird verkleinert.

6 Basisbandübertragung und Basisbandmodulation

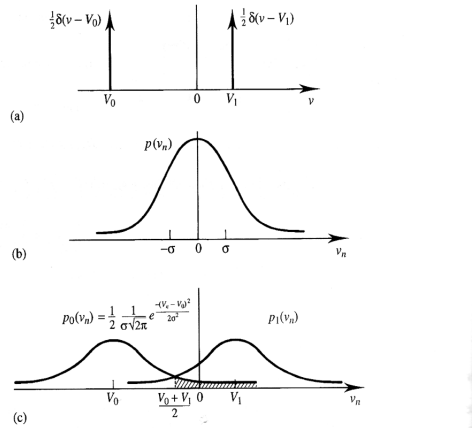
6.1 Basisband Centre Point Detection

Die Erkennung digitaler Signale beinhaltet zwei Prozesse: erstens das Reduzieren jedes empfangenen Spannungspulses auf einen numerischen Wert und zweitens das Vergleichen dieses Wertes mit einer Referenzspannung um festzustellen, welches Symbol übertragen wurde. Im Falle binärer Symbole liegt die Referenzspannung meist genau in der Mitte zwischen den Spannungswerten für 0 und 1.
Zusätzlich stellt sich noch die Frage, wann während einer Periode der empfangene Wert gesamplet wird. Bei der centre point detection wird das Signal in der Mitte der Symbolperiode abgetastet.

6.2 Bitfehlerwahrscheinlichkeit einer binären Basisbandübertragung

Nimmt ein binäres Signal zwei Spannungslevel V_0 und V_1 an und ist das Signal mit Gauß-Rauschen (Mittelwert 0, Standardabweichung σ) belegt, so erhält man die Dichte des Signals durch Falten der Dichte des Gauß-Rauschens (Glockenkurve) mit den Impulsen in der Dichte des Binärsignals (die Dichte der Summe zweier unabhängiger Zufallsvariable erhält man durch Faltung der Dichten). Die gesamte Fläche der neuen Dichte ist wiederum 1, bei den Impulsen der Bitwahrscheinlichkeiten repliziert sich jeweils die Glockenkurve des Gaußschen Rauschens (siehe Bild). Wir wollen nun die Wahrscheinlichkeit berechnen, dass durch die Gaußsche Störung ein Bit missinterpretiert wird. Es handelt sich hier um bedingte Wahrscheinlichkeiten der Art „WK dass 1 empfangen wird wenn 0 gesendet wird“. Gehen wir davon aus, dass die beiden Symbole gleichwahrscheinlich sind, dann hat jede der beiden Glockenkurve Fläche $\frac{1}{2}$ (die Gesamtfläche beider Kurven ist natürlich 1), wir müssen also bei bedingten Fehlerwahrscheinlichkeiten durch $\frac{1}{2}$ dividieren. Für den Schwellwert ergibt sich bei gleichwahrscheinlichen Symbolen $\frac{V_0+V_1}{2}$, also der Schnittpunkt der beiden Glockenkurven (man kann zeigen dass das der ideale Schwellwert ist, was in unserem Fall intuitiv klar ist). Die Wahrscheinlichkeit P_{e1} dass 1 empfangen wird obwohl 0 gesendet wurde ist also 2 Mal die schattierte Fläche (siehe Bild), also (das $\frac{1}{2}$ von der Höhe der Verteilung und das $\cdot 2$ wegen der bedingten WK hebt sich auf):

$$P_{e1} = \int_{(V_0+V_1)/2}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(v-V_0)^2}{2\sigma^2}} dv$$



Probability density function of: (a) binary symbol; (b) noise; (c) signal plus noise.

Durch Verwendung der Gegenwahrscheinlichkeit, Substitution und der gut tabellierten Fehlerfunktion $erf(z) := \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2} du$ erhält man schließlich:

$$P_{e1} = \frac{1}{2}\left[1 - \text{erf}\left(\frac{V_1 - V_0}{2\sigma\sqrt{2}}\right)\right]$$

Aus Symmetriegründen gilt natürlich $P_{e0} = P_{e1}$. Da die Fehlerwahrscheinlichkeit eigentlich nur vom Spannungsunterschied $\Delta V = V_1 - V_0$ abhängt, schreibt man:

$$P_e = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\Delta V}{2\sigma\sqrt{2}}\right)\right]$$

Man beachte dass ΔV der Spannungsunterschied zum Abtastzeitpunkt ist, obige Formel gilt für alle Spannungswerte und Pulsformen.

Für die Berechnung der Bitfehler eines Kanals mit AWGN (Additive White Gaussian Noise) und NRZ Signale lässt sich der Bruch $\frac{\Delta V}{\sigma}$ durch SNR$= \frac{S}{N}$ ausdrücken:
Für unipolares NRZ ist $S_{peak} = \Delta V^2$ und die mittlere Signalstärke $S = \frac{\Delta V^2}{2}$. Die normalisierte Gaußsche Rauschstärke beträgt $N = \sigma^2$. Daraus folgt:

$$\frac{\Delta V}{\sigma} = \left(\frac{S}{N}\right)_{peak}^{1/2} = \sqrt{2}\left(\frac{S}{N}\right)^{1/2}$$

und daher

$$P_e = \frac{1}{2}\left[1 - \text{erf}(\frac{1}{2\sqrt{2}}\left(\frac{S}{N}\right)_{peak}^{1/2})\right] = \frac{1}{2}\left[1 - \text{erf}(\frac{1}{2}\left(\frac{S}{N}\right)^{1/2})\right]$$

Für polares NRZ sind die Spitzensignalstärke und die mittlere Signalstärke gleich: $S_{peak} = S = \left(\frac{\Delta V}{2}\right)^2$. Daraus folgt:

$$\frac{\Delta V}{\sigma} = 2\left(\frac{S}{N}\right)^{1/2} \qquad \text{und daher} \qquad P_e = \frac{1}{2}\left[1 - \text{erf}(\frac{1}{\sqrt{2}}\left(\frac{S}{N}\right)^{1/2})\right]$$

Hat man $M > 2$ Symbole, wieder alle mit gleicher Wahrscheinlickeit, deren Spannungslevel alle ΔV Volt auseinanderliegen, dann hat man M Glockenkurven nach der Faltung, wobei die „inneren“ Glockenkurven jeweils 2 Nachbarn, also die doppelte Schnittfläche haben. Folglich bekommt man hier:

$$P_{eM} = \frac{M-2}{M} 2P_e + \frac{2}{M} P_e = \frac{2(M-2)}{M} P_e$$

Last but not least: Die Symbolfehlerrate (SER, Symbol Error Rate) und die Bitfehlerrate (BER, Bit Error Rate) berechnet man als:

$$BER = P_b R_b \qquad SER = P_e R_s$$

wobei P_b und P_e die Fehlerwahrscheinlichkeiten für binäre bzw. n-äre Übertragung beschreiben und R_b und R_s die Bitrate bzw. die Symbolrate sind.

6.3 Fehlersummierung über mehrere Hops

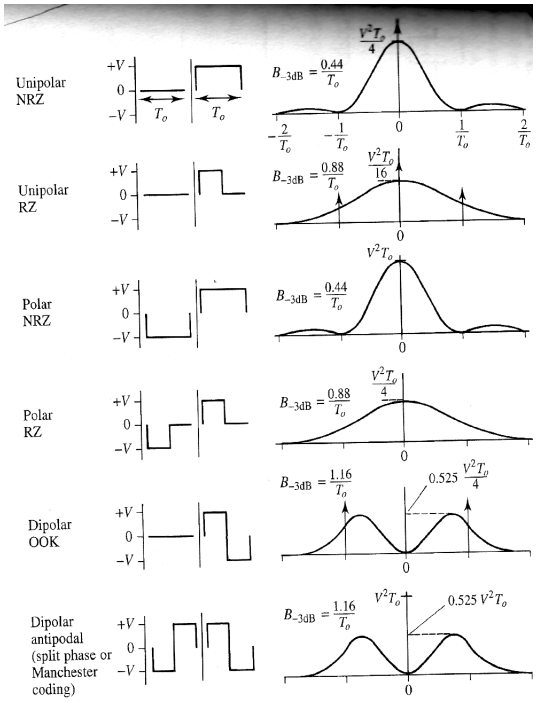
Übertragungsmedien dämpfen Signale. Bei der Übertragung über längere Strecken werden daher Repeater eingesetzt, die die Signale wieder verstärken, wobei zwischen verstärkenden und regenerativen Repeatern unterschieden wird.

Verstärkende Repeater heben das Signal beim Empfang lediglich auf die ursprüngliche Signalstärke an und verstärken dabei das Rauschen ebenso, sodass nach n Hops das Signal die gleiche Stärke hat wie zu Beginn, die Rauschstärke allerdings um den Faktor n zugenommen hat.

Regenerative Verstärker entscheiden ob eine Eins oder Null an ihrem Eingang anliegt und erzeugen dann aufgrund ihrer Entscheidung ein neues Signal, wodurch sich das Rauschen nicht pro Hop aufsummiert.

6.4 Signalisierung (line coding)

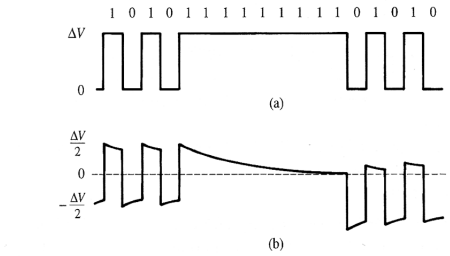
Binäre Daten können mittels unterschiedlichster Pulstypen übertragen werden. Diese unterscheiden sich unter anderem durch Aspekte wie Gleichstromanteil, Leistungsdichtespektrum (besonders im Hinblick auf den Wert bei 0 Hz), Bandbreite, Rauschimmunität und Taktrückgewinnung.



6.4.1 Unipolare Signalisierung

Unipolare Signalisierung, auch On-Off Keying (OOK) genannt, repräsentiert ein Binärsymbol durch die Abwesenheit eines Pulses und das andere durch das Auftreten eines Pulses. Dabei wird zwischen non-return to zero (NRZ) und return to zero (RZ) unterschieden - bei NRZ ist die Dauer des Pulses gleich der Länge des Symbolzeitschlitzes, bei RZ ist die Pulsdauer kürzer als der Symbolzeitschlitz. RZ-Signale, die üblicherweise ein Tastverhältnis (=Pulsdauer/Symbolzeitschlitz) von 50% haben, benötigen so die doppelte Bandbreite von NRZ-Signalen. Sie haben allerdings auch den Vorteil einer Spektrallinie bei der Symbolrate $f_0 = 1/T_0$ Hz und ermöglichen damit eine einfachere Taktrückgewinnung. Das Spektrum von NRZ-Signalen hat eine $sinc$-Form.

Sowohl RZ als auch NRZ haben in ihrem Spektrum bei 0 Hz eine Linie, also einen Gleichstromanteil. Bei der Übertragung über Wechselstrom-gekoppelte Repeater, die meist über einen Kondensator angebunden sind, wird dieser Gleichstromanteil entfernt und das Signal in ein polares umgeformt. Da sich die AC-gekoppelten Medien wie Hochpassfilter verhalten, kommt es zu einem exponentiellen Abflachen der Signalspitzen nach jedem Übergang („signal droop“).



Distortion due to AC coupling of unipolar NRZ signal: (a) input; (b) output.

6.4.2 Polare Signalisierung

Bei polarer Signalisierung wird eine Eins durch einen Puls $g_1(t)$ und eine Null durch einen Puls $g_0(t) = -g_1(t)$ repräsentiert. Es gilt im wesentlichen das selbe wie für unipolare Signalisierung bezüglich Spektrum, Bandbreite, AC-coupling. Ein Vorteil jedoch ist dass der DC-Anteil über längere Zeit gesehen 0 ist. Außerdem ist ΔV doppelt so groß, was höhere Leistung voraussetzt, aber die Fehlerrate verkleinert. Ein weiterer Vorteil polarer Signalisierung ist dass der Entscheidungsschwellwert 0 ist, das ist aus schaltungstechnischer Sicht meist einfacher umzusetzen. Bei polarer Signalisierung sind die 0- und 1-Signale genau um 180° phasenverschoben, die Spektrallinie bei $\frac{1}{T_0}$ fehlt daher, was sich negativ auf die Taktrückgewinnung auswirkt.

6.4.3 Dipolare Signalisierung

Dipolare Signalisierung hat keinen Gleichstromanteil und eignet sich daher gut für AC-gekoppelte Medien. Das Symbolintervall T_0 wird in zwei gleich große Pulse, einer negativ und einer positiv, aufgeteilt. Somit ist die Fläche unter jedem Puls gleich Null. Beim „Manchester-Coding“ wird eine Eins durch einen Eins-Null-Übergang und eine Null gegengleich dargestellt. Man beachte (siehe Grafik), dass beim Manchestercode die Spektrallinie bei $\frac{1}{T_0}$ im Gegensatz zum dipolaren OOK fehlt, da die Signale für 0 und 1 genau 180° phasenverschoben sind.

6.4.4 Sonstige Signalisierungen

HDBn Kodierung stellt durch spezielle Kodierung sicher, dass Taktrückgewinnung auch noch bei langen 0 oder 1 Folgen möglich ist. nBmT Kodierung kodiert Blöcke von n Bits durch Blöcke von m ternären Symbolen.

Man beachte, dass die Spektrallinie bei Unipolar RZ und Dipolar OOK für die Taktrückgewinnung zwar vorteilhaft ist, aber nutzlos wird, wenn lange 0-Folgen übertragen werden. Polar RZ und Manchestercode haben zwar keine Spektrallinie bei $\frac{1}{T_0}$, aber ein Spektrum $\neq 0$ bei $\frac{1}{T_0}$ unabhängig von der übertragenen Bitfolge.

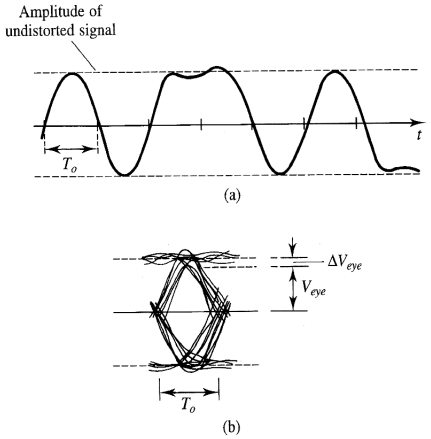
6.5 Signalrückgewinnung

6.5.1 Entzerrung von Impulsen

Verzerrung von Amplitude und Phase eines Signals durch das Übertragungsmedium stellt ein signifikantes Problem dar. Man muss daher einen langen Link immer durch Repeater unterteilen, üblicherweise macht man die Strecken zwischen Repeatern so kurz, dass sie ein SNR $>$ 18 dB haben.
Bei einer Datenrate von 2Mbit/s hat ein bipolarer RZ-Puls eine Breite von $\frac{1}{4}\mu sec$ und daher eine Bandbreite von ungefähr 2MHz. Bei der Übertragung über ein Metallkabel wird der Puls stark verzerrt, gedämpft und in der Folge zeitlich gestreckt. Es kommt weiters zu einer Symbolinterferenz der einzelnen empfangenen Symbole, wenn die Signale so langsam abklingen (RC Zeitkonstante des Mediums) dass sie das nachfolgende Symbol beeinflussen. Diesem Phänomen kann man durch einen Entzerrfilter, der die inverse Übertragungscharakteristik des Übertragungsmediums hat, entgegenwirken - wo das Medium dämpft, verstärkt der Filter.

6.5.2 Augendiagramm

Das Augendiagramm entsteht, wenn man auf einem Oszilloskop den Trigger auf T_s einstellt und den Schirm so träge macht, dass mehrere Symbole übereinander dargestellt werden (siehe Bild). Das Augendiagramm ist sehr hilfreich bei der Untersuchung von Übertragungsmedien. Die vertikale Augenöffnung V_{eye} zeigt, wie empfindlich das Signal auf weitere Störungen ist, wie leicht also ein Bit „umfallen“ kann. An dem Zeitpunkt, wo die vertikale Öffnung des Auges am größten ist, ist der optimale Abtastzeitpunkt. Die horizontale Öffnung gibt den Zeitbereich an, in dem eine Abtastung möglich ist. Der vertikale Rand ΔV_{eye} gibt an, wie stark die Störungen sind, der horizontale Rand gibt den Jitter bei der Taktrückgewinnung an.



(a) Noiseless but distorted signal; and (b) corresponding eye diagram.

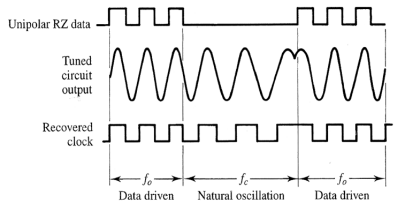
6.5.3 Übersprechen

Man spricht bei der kapazitiven Kopplung zwischen den Pulsen eines ausgehenden, noch starken Signals und eines schwachen Eingangssignals von near end crosstalk (NEXT). Far end crosstalk (FEXT)

bezeichnet die Kopplung zwischen den Pulsen zweier ausgehender Signale. Die für Crosstalk verantwortliche kapazitive Kopplung kann als Highpass-Filter modelliert werden, um Crosstalk-Effekte zu kompensieren kann der Entzerrer also höhere Frequenzen etwas weniger verstärken als wenn man nur die Leitungscharakteristik in Betracht ziehen würde.

6.6 Taktrückgewinnung

Um die Abtastrate richtig einzustellen muss aus dem bereits entzerrten Signal der Takt rückgewonnen werden. Das geht durch Bandpass-Filter wenn der Code bei f_0 ein Spektrum hat, das nicht 0 ist. Für Codes die bei f_0 Spektrum 0 haben (z.B.: (Uni)Polare NRZ Codes) kann durch Quadrieren oder Gleichrichten die Signalbandbreite verdoppelt werden, so wird die Nullstelle bei f_0 Hz im Spektrum entfernt und der neue f_0-Anteil kann durch eine Resonanzschaltung extrahiert werden. Während nur Nullen übertragen werden und kein Takt aus dem Signal entnommen werden kann, springt die Resonanzschaltung (PLL, Phase Locked Loop), die mit ihrer natürlichen Frequenz schwingt, ein. Wenn wieder Takt rückgewonnen werden kann entsteht ein Timing-Jitter zwischen der PLL und dem rückgewonnenen Takt, der dazu führen kann, dass Symbole zu ungültigen Zeiten abgetastet werden, was (siehe Augendiagramm) die Wahrscheinlichkeit, dass ein Bit umfällt, erhöht.
Je mehr Repeater zwischengeschalten sind, desto mehr Oszillatoren sind involviert, und das Problem wird größer. In der Praxis löst man es durch entsprechende Codierung (z.B. HDBn), die sicherstellt, dass es keine langen 0- oder 1-Bitströme gibt.



Clock recovery oscillator operation when a string of no symbols (spaces) is received.

7 Entscheidungstheorie

Bei der Entscheidungstheorie geht es darum, optimale Empfängertrennspannungen zu finden, um sich zu entscheiden, welches Symbol dem Signalwert zugeordnet wird. Bei weichen (soft) Entscheidungen wird das Signal zum Entscheidungszeitpunkt auf mehrere Level (üblicherweise acht) quantisiert um die Vertrauenswürdigkeit der Entscheidung mit angeben zu können. Letztere ist von Interesse sofern man eine Sequenz mehrerer Entscheidungen in Hinblick auf Fehlererkennung betrachtet. Bei harten Entscheidungen gibt es nur zwei Level.

7.1 A priori, bedingte und a posteriori Wahrscheinlichkeiten

Es gibt unterschiedliche Wahrscheinlichkeiten und Dichten bei der Symbolübertragung:

- $P(x)$ - a priori Wahrscheinlichkeit, dass Symbol x übertragen wird

- $p(v|x)$ - bedingte Dichte der Wahrscheinlichkeit Spannung v zu messen, wenn Symbol x gesendet wurde
- $P(x|v)$ - a posteriori Wahrscheinlichkeit, dass Symbol x gesendet wurde, wenn Spannung v gemessen wurde

Die a priori Wahrscheinlichkeiten sind normalerweise bereits im Vorhinein bekannt. Die a posteriori Wahrscheinlichkeiten können erst nach vielen Übertragungen ermittelt werden. Bedingte Wahrscheinlichkeiten werden mit kleinem p geschrieben. Übergangswahrscheinlichkeiten der Form $P(1|0)$ (WK dass 1 empfangen wird wenn 0 gesendet wird) werden oft in einer Übergangsmatrix zusammengefasst (2x2 Matrix). Gilt $P(1|0) = P(0|1)$, so spricht man von einem symmetrischen Binärkanal.

7.2 Das Entscheidungskriterium von Bayes

Das Ziel des Bayes'schen Kriteriums ist es, die Kosten (Fehler oder verlorene Information) zu minimieren: die Kosten wenn Symbol x empfangen und als y missinterpretiert wird, werden mit C_y bezeichnet. Korrekter Empfang ist kostenlos. Die erwarteten bedingten Kosten $C(0|v)$, die anfallen wenn Spannung v als Null interpretiert wird, belaufen sich auf: $C(0|v) = C_0 P(1|v)$ ($P(1|v)$ ist eine a posteriori WK), und vice versa für $C(1|v)$.

Eine rationale Entscheidungsregel entscheidet sich nun anhand der bedingten Kosten für ein Symbol: gilt $C(1|v) < C(0|v)$, sind also die Kosten für das Interpretieren einer Eins geringer, so wird für eine Eins entschieden; Umgekehrtes gilt für eine Null... hier werden also die bedingten Kosten minimiert. Setzt man $C(0|v) = C_0 P(1|v)$ bzw. $C(1|v) = C_1 P(0|v)$ in die Ungleichung $C(1|v) < C(0|v)$ ein, so bekommt man das Entscheidungskriterium:

$$\frac{P(0|v)}{P(1|v)} \underset{>0}{\overset{<1}{}} \frac{C_0}{C_1}$$

Da die verwendeten a posteriori Wahscheinlichkeiten normalerweise nicht bekannt sind, lässt sich über das Bayes'sche Theorem

$$P(x|v) = \frac{p(v|x)P(x)}{p(v)}$$

und Einsetzen/Dividieren das Kriterium folgendermaßen angeben:

$$\frac{p(v|0)}{p(v|1)} \underset{>0}{\overset{<1}{}} \frac{C_0 P(1)}{C_1 P(0)} =: L_{th}$$

Die linke Seite ist eine Funktion der Spannung und gibt das Wahrscheinlichkeitsverhältnis an und die rechte Seite einen Wahrscheinlichkeitsschwellwert L_{th}, der bei gleicher Wahrscheinlichkeit und gleichen Kosten den Wert 1 annimmt. Ersetzt man das Ungleichheitszeichen durch ein Gleichheitszeichen, so ergibt sich für gegebenes L_{th} eine Gleichung in der Variablen v für die optimale Schwellspannung. Man beachte, dass bei multimodalen Wahrscheinlichkeitsverteilungen auch mehrere Schnittpunkte von Wahrscheinlichkeitsdichten auftreten können, obige Gleichung für die optimale Schwellspannung v_{th} muss also nicht eindeutig lösbar sein. Es könnte also auch eine Entscheidungsregel

folgender Form herauskommen:

$$S(v) := \begin{cases} 0, & v < 0 \\ 1, & 0 \leq v < 1 \\ 0, & 1 \leq v < 2 \\ 1, & v \geq 2 \end{cases}$$

Das Bayes'sche Entscheidungskriterium minimiert auch die mittleren Entscheidungskosten, also den Erwartungswert der Kosten. Das sieht man, indem man die Schwellspannung so zu bestimmen versucht, dass der Erwartungswert minimiert wird:
Der Erwartungswert berechnet sich als:

$$\overline{C} = C_1 P(0) P(1_{RX}|0_{TX}) + C_0 P(1) P(0_{RX}|1_{TX})$$

wobei $P(X)P(Y_{RX}|X_{TX})$ die WK angibt, dass ein gesendetes X versehentlich als Y interpretiert wird.
Man versucht also v_{th} so zu bestimmen, dass $\overline{C}$ minimiert wird, man muss also die Gleichung

$$\overline{C} = C_1 P(0) \int_{v_{th}}^{\infty} p(v|0)dv + C_0 P(1) \int_{-\infty}^{v_{th}} p(v|1)dv$$

differenzieren und 0 setzen. Die bestimmten Integrale lassen sich leicht differenzieren (beim ersten Integral muss man vorher die Grenzen umdrehen und bekommt daher ein Minus), man setzt die differenzierte Gleichung 0 und erhält:

$$-C_1 P(0) p(v_{th}|0) + C_0 P(1) p(v_{th}|1) = 0$$

und nach Umformen genau wieder

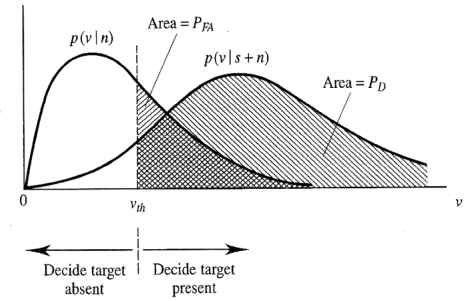
$$\frac{p(v_{th}|0)}{p(v_{th}|1)} = \frac{C_0 P(1)}{C_1 P(0)}$$

7.3 Das Neyman-Pearson Entscheidungskriterium

Das Neyman-Pearson Kriterium basiert nur auf a posteriori Wahrscheinlichkeiten und funktioniert besonders gut, wenn $C_0 \gg C_1$. Es wird daher vor allem beim Radar eingesetzt, wo die Kosten von C_0 (ein Flugzeug fehlerhaft nicht erkennen) weitaus größer sind als C_1 (ein Flugzeug fehlerhaft zu erkennen). Für eine bestimmte Schwellspannung sind die Wahrscheinlichkeiten ein Ziel zu erkennen P_D und eines Fehlalarms P_{FA} gegeben durch:

$$P_D = \int_{v_{th}}^{\infty} p(v|s+n)dv \qquad bzw. \qquad P_{FA} = \int_{v_{th}}^{\infty} p(v|n)dv$$

wobei s Signal und n Rauschen bezeichnet. Der Schwellwert wird so gewählt, dass die Falschalarme unter einer akzeptablen Wahrscheinlichkeit liegen. Die Qualität der Entscheidungen hängt sowohl von der Wahl von P_{FA} als auch dem Verhältnis zwischen empfangener Pulsenergie und Rauschleistungsdichte ab: E/N_0.



Conditional pdfs and threshold voltage for Neyman–Pearson radar signal detector.

8 Optimale Filterung für die Übertragung und Detektion

Es gibt zwei Filtertechniken, die bei digitaler Kommunikation von essentieller Bedeutung sind: das Filtern bei der Übertragung um die Signalbandbreite zu minimieren und das Filtern beim Empfang um den Signal-Rauschabstand zum Abtastzeitpunkt zu maximieren.

8.1 Pulsformung für optimales Senden

8.1.1 Spektrale Effizienz

Die spekatrale Effizienz η_s ist ein Maß für die übertragene Informationsmenge pro Bandbreite und definiert als

$$\eta_s := \frac{R_s H}{B} \quad (bits/s/Hz)$$

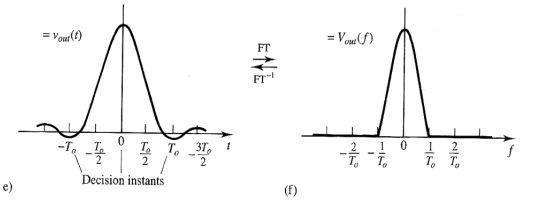
wobei R_s die Symbolrate und H die Entropie (mittlerer Informationsgehalt pro Bit, $H = log_2(M)$ für M statistisch unabhängige, gleichwahrscheinliche Symbole) ist. Ziel ist es, η_s zu maximieren.

8.1.2 Intersymbolinterferenz (ISI)

Signalisierung mit Rechteckimpulsen hat theoretisch unendliche Bandbreite, praktisch können Rechteckpulse auch über Kanäle mit begrenzter Bandbreite übertragen werden, wenn ein gewisses Maß an Verzerrung in Kauf genommen wird. Ist die Verzerrung allerdings zu stark, so überlappen sich die Pulse im Zeitbereich und die Spannung zum Abtastzeitpunkt kann nicht nur vom gewünschten, sondern zusätzlich von einem oder mehreren vorigen Pulsen herrühren. Das Verschmieren eines Pulses in einen anderen wird als Intersymbolinterferenz bezeichnet.

Die Performance digitaler Kommunikation wird dabei allerdings nur von der ISI zum Abtast- bzw. Entscheidungszeitpunkt beeinträchtigt. Außerhalb des Entscheidungszeitpunktes ist die ISI irrelevant. Dass man sich Gedanken machen muss, wie man ein rechteckiges Signal vor der Übertragung filtert, zeigt folgendes Beispiel:
Filtert man ein rechteckiges NRZ-Signal mit Breite T_0, welches eine Bandbreite ($\sqrt{2}$ Regel!) von $B = \frac{1}{T_0}$ Hz hat, mit einem (idealen) Low-Pass-Filter mit Bandbreite B, so berechnet sich das Spektrum des gefilterten Signals, indem man das Spektrum des Rechtecks ($sinc$-Form) mit dem Rechteckspektrum des Filters multipliziert (siehe Plot von $V_{out}(f)$ im Bild). Rücktransformation ergibt, dass

zum Abtastzeitpunkt T_0 des nächsten Zeichens der Beitrag dieses Zeichens nicht 0 ist, das nächste Zeichen also noch beeinflusst wird (Plot von $v_{out}(t)$ im Bild). Wir haben also hier ISI.



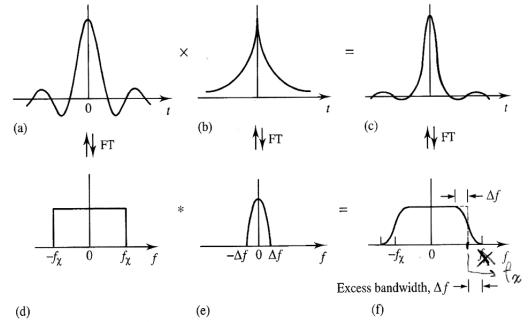
.2 NRZ rectangular pulse distortion due to rectangular frequency response filtering.

Wir definieren also ein ISI-freies Signal als ein solches, das zu allen Abtastzeitpunkten bis auf einen Zeitpunkt t_n einen Nulldurchgang hat, dann kann das n-te Zeichen zum Zeitpunkt t_n abgetastet werden und stört an den anderen Abtastzeitpunkten $t_j, j \neq n$ nicht. Mathematisch kann man das so formulieren:

$$v(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_0) = v(0)\delta(t)$$

wobei hier 0 als Abtastzeitpunkt gewählt wurde. Man beachte, dass wir hier eine Funktionalgleichung haben, $v(t)$ muss also wirklich zu allen Zeitpunkten nT_0 mit $n \neq 0$ Null sein, die Summanden können sich nicht gegenseitig aufheben. Diese Bedingung stellt sicher, dass das Signal $v(t)$ an allen Abtastzeitpunkten bis auf den Zeitpunkt $t = 0$ Null ist.

Ein gutes Beispiel für eine Funktion die diese Bedingung erfüllt ist der sinc-Puls, der allerdings praktisch nicht realisierbar ist[11], und außerdem zu langsam abklingt. Um ein schnelleres Abklingen zu erreichen, müsste man den sinc-Puls mit einer schnell abklingenden, monotonen Funktion multiplizieren. Führt man das anhand eines Beispiels durch und verwendet eine gerade Funktion für die monotone Funktion, so erkennt man, dass das Spektrum des multiplizierten Signals ungerade symmetrisch um $f_\chi = \frac{1}{2T_0}$ ist (siehe Bild). Ungerade Symmetrie ist hier im grafischen Sinne zu interpretieren und nicht im streng mathematischen Sinne.



3.5 Suppression of sinc pulse sidelobes and its effect on pulse spectrum.

[11] da rechteckige Tiefpass-Filter nicht realisierbar sind

Dass diese Bedingung notwendig ist zeigt die folgende Rechnung: Wir transformieren die vorher formulierte notwendige Bedingung

$$v(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_0) = v(0)\delta(t)$$

in den Frequenzbereich, mit den Korrespondenzen

$$\delta(t - T) \Leftrightarrow e^{-j2\pi fT}$$

$$\sum_{k=-\infty}^{\infty} \delta(t - kT_s) \Leftrightarrow f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s)$$

sowie dem Faltungstheorem und der Linearität der Fouriertransformation ($v(0)$ ist eine Konstante, $\delta(t) \Leftrightarrow e^0 = 1$) bekommen wir so

$$V(f) * \frac{1}{T_0} \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T_0}) = v(0)$$

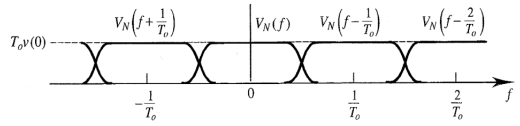
Da die Deltafunktion bei der Faltungsopertion eine replizierende Eigenschaft hat, bekommen wir

$$\frac{1}{T_0} \sum_{n=-\infty}^{\infty} V(f - \frac{n}{T_0}) = v(0)$$

Das im Abstand T_0 replizierte Spektrum hat also konstante Summe. Daraus folgt, dass V um $\frac{1}{2}\frac{1}{T_0} = \frac{1}{2T_0}$ ungerade symmetrisch ist.

Daraus ergibt sich: Ein ISI-freies Signal hat ein um $\frac{1}{2T_0}$ Hz ungerade symmetrisches Spannungsspektrum.

Da zwischen der Fouriertransformierten und der Zeitfunktion eine 1:1 Korrespondenz besteht, ist diese Bedingung auch hinreichend.



Constant sum property of replicated ISI-free signal spectra.

8.1.3 Das Nyquist'sche Symmetrietheorem

Das Nyquist'sche Symmetrietheorem definiert eine Symmetriebedingung die $H(f)$ erfüllen muss um eine ISI-freie Impulsantwort zu geben:

Wird die Amplitudenantwort eines rechteckigen Tiefpassfilters mit linearer Phase und Bandbreite f_χ durch Addition mit einer reellwertigen Funktion mit ungerader Symmetrie um die Trennfrequenz des Filters modifiziert, dann besitzt die sich daraus ergebende Impulsantwort zumindest die in der originalen $\text{sinc}(2f_\chi t)$-Impulsantwort enthaltenen Nulldurchgänge und ist damit ISI-frei.

Das Theorem muss nicht weiter begründet werden, da es direkt aus den Spektraleigenschaften eines ISI-freien Signals folgt. Ein Filter, der aus dem Nyquist-Symmetrietheorem hervorgeht, heißt Nyquist-Filter.

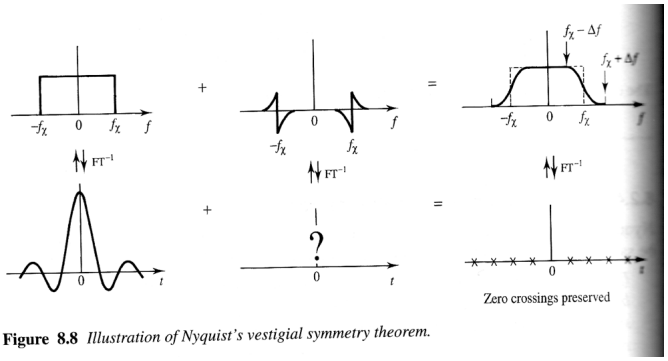


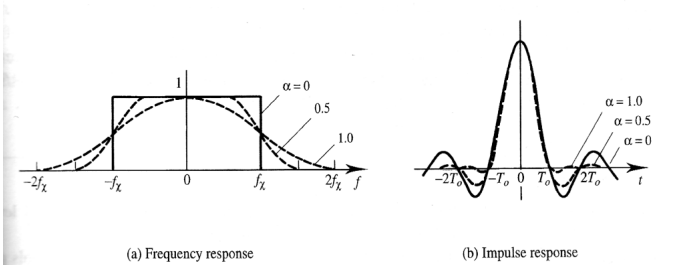
Figure 8.8 *Illustration of Nyquist's vestigial symmetry theorem.*

8.1.4 Raised-Cosine Filter

Die raised-cosine Filter bilden eine Untermenge der Nyquist-Filter. Die ungerade Symmetrie wird durch eine halbe Cosinusschwingung - den abschwingenden Teil von der Spitze bis zum Nulldurchgang - erreicht. Das Amplitudenspektrum hat folgende Form (siehe Abbildung):

$$|H(f)| = \begin{cases} 1.0, & |f| \leq (f_\chi - \Delta f) \\ \frac{1}{2}\left\{1 + \sin\left[\frac{\pi}{2}\left(1 - \frac{|f|}{f_\chi}\right)\frac{f_\chi}{\Delta f}\right]\right\}, & (f_\chi - \Delta f) < |f| < (f_\chi + \Delta f) \\ 0, & |f| \geq (f_\chi + \Delta f) \end{cases}$$

Dabei bezeichnet f_X die Trennfrequenz des zugrundeliegenden rechteckigen Tiefpassfilters, f_χ verhält sich zur Symbolperiode T_0 wie $f_\chi = 1/(2T_0)$. Δf bezeichnet die absolute bzw. Exzess-Bandbreite des Filters über f_χ hinaus. Die normalisierte Exzess-Bandbreite ist durch $\alpha = \frac{\Delta f}{f_\chi}$ gegeben und wird roll-off Faktor genannt, er kann zwischen 0 und 1 liegen. Der Einfluss des roll-off Faktors ist ebenfalls aus dem Bild ersichtlich, je größer er ist desto schneller klingt die Impulsantwort im Zeitbereich ab, und desto größer ist auch die Bandbreite des Filters.



ure 8.10 *Responses of raised cosine filters with three different roll-off factors.*

Raised-Cosine Filter haben lineare Phase. Wie man ebenfalls in der Abbildung sehen kann hat die Impulsantwort eine $sinc$ Form, diese geht jedoch nicht wie $1/t$ gegen 0 für $t \to \infty$ wie bei der normalen $sinc$ Funktion, sondern wie $1/t^3$.
Die absolute Bandbreite (also 0, nicht $\sqrt{2}$) des Filters berechnet sich folgendermaßen:

$$B = \frac{1}{2T_0}(1+\alpha)$$

Ein Filter mit $\alpha = 1$ heißt Full Raised Cosine Filter. Für ihn erhält man die Übertragungsfunktion:

$$|H(f)| = \begin{cases} cos^2(\frac{\pi f}{4f_\chi}), & |f| \leq 2f_\chi \\ 0, & |f| > 2f_\chi \end{cases}$$

8.1.5 Duobinäre Zeichengebung

Duobinäre Zeichengebung verwendet statt des schwierig zu konstruierenden rechteckigen Tiefpasses einen Cosinus-Tiefpass, der ebenfalls eine Trennfrequenz bei $1/2T_0$ hat. Der abgestuftere roll-off des Cosinusfilters verglichen mit dem rechteckigen Tiefpass hat eine längere Impulsantwort und damit höhere ISI zur Folge. Die ISI tritt allerdings nur zwischen benachbarten Symbolen auf und ist von vorhersagbarem Ausmaß.

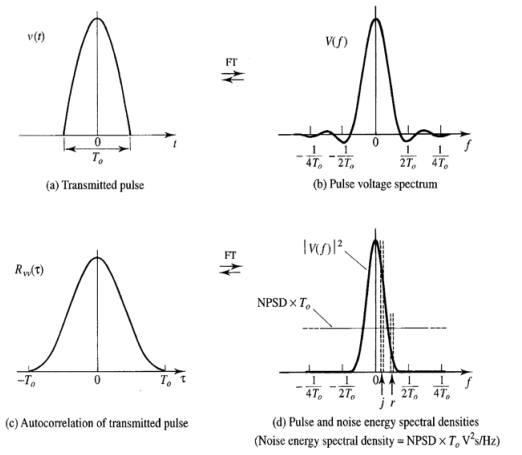
8.2 Pulsfilterung für optimales Empfangen

Bei der centre point detection Methode wird aufgrund eines Abtastwertes eine Entscheidung für ein Symbol getroffen. Betrachtet man mehrere Abtastwerte innerhalb der Symboldauer, so kann man zuverlässigere Entscheidungen treffen, als mit nur einem Abtastwert wie bei der Centre Point Detection. Man könnte die n Abtastwerte getrennt betrachten und eine Mehrheitsentscheidung durchführen (majority voting), oder aber es werden die n Abtastwerte summiert und mit dem n-fachen des Schwellwertes verglichen. Bildet man dabei den Grenzwert $n \to \infty$, so wird aus der Summe der Abtastwerte (inklusive Rauschen) eine Integration des Signals (inklusive Rauschen) über die Symboldauer. Dieses Verfahren wird integrate and dump (I+D) detection genannt. Integrier-Schaltungen mit nachfolgendem Vergleich sind leicht in Hardware zu implementieren (Kondensator und Komparator). I+D ist ein Sonderfall eines generellen Entscheidungsprozesses, der auf jede Signalform angewandt werden kann, nämlich Matched Filtering.

8.2.1 Matched Filtering

Ein Filter, der dem Entscheidungsschaltkreis eines Empfängers direkt vorgeschalten ist, wird „matched“ (also ansprechend auf, oder passend zu) einem bestimmten Symbolpuls genannt, wenn er den Signalrauschabstand zum Abtastzeitpunkt maximiert, sofern der Puls dieses bestimmten Symbols, auf das er „matched“, am Filtereingang anliegt.

Betrachtet man das Energiedichtespektrum eines Pulses verglichen mit dem Leistungsdichtespektrum weißen Rauschens (eine konstante Linie), so sieht man, dass manche Frequenzbänder des Pulses einen hohen und manche einen niedrigen Signalrauschabstand haben. Die Frequenzbänder mit hohem SNR sollten einen höheren Einfluss im Entscheidungsprozess haben, als diese mit niedrigem SNR.



8.28 *Relationship between energy spectral densities of signal pulse and white noise to illustrate matched filtering amplitude criterion.*

Dieser Umstand führt auf das Bilden einer gewichteten Summe der einzelnen Frequenzband Rausch- und Signalenergien, wobei die Gewichte direkt proportional zum SNR des jeweiligen Frequenzbandes sind. Da das Leistungsdichtespektrum des Rauschens konstant ist, ist SNR proportional zu $|V(f)|^2$. Da weiters die Leistung bzw. Energie, die von einem Filter weitergegeben wird proportional zu $|H(f)|^2$ ist, muss das Quadrat der Amplitudenantwort eines Matched Filter formgleich zum Energiedichtespektrum des Pulses sein, auf den er abgestimmt ist, also:

$$|H(f)|^2 =)k^2|V(f)|^2$$

für eine Konstante k.

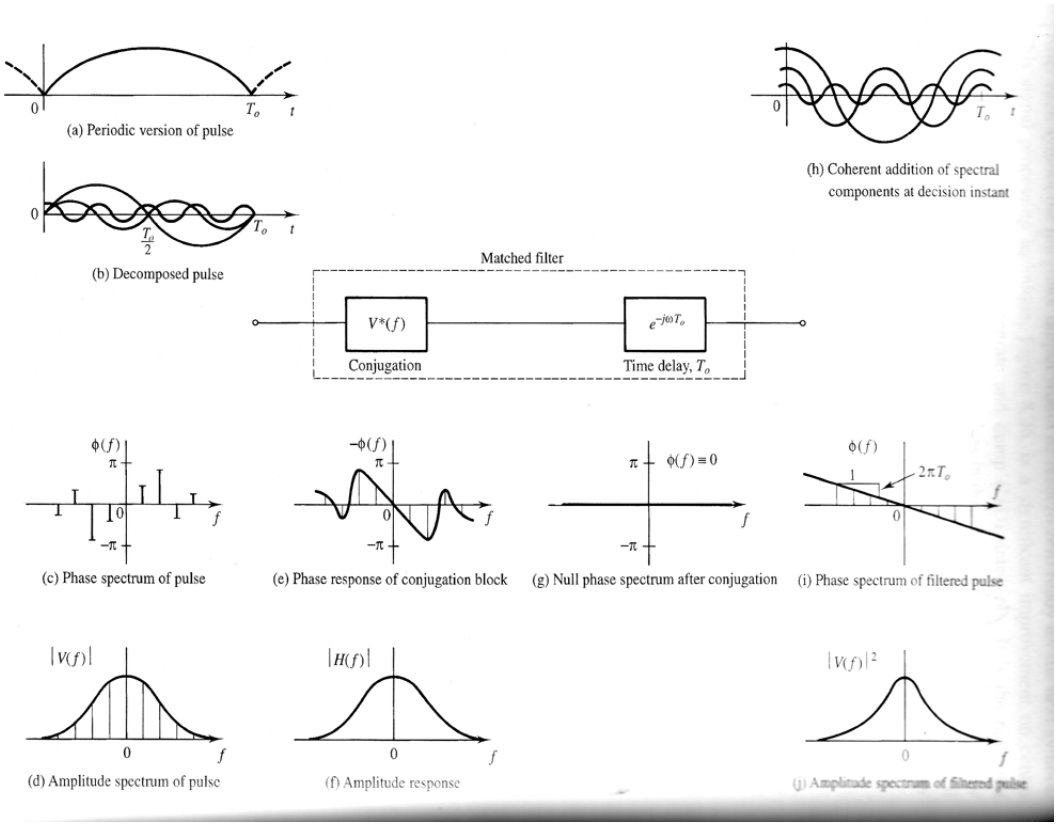
Bleibt noch die Frage nach der Phasenantwort des Matched Filter. Zuerst zerlegt man eine periodische Version der Pulsfolge in seine einzelnen Cosinusschwingungen. Damit die Signalstärke zum Entscheidungszeitpunkt maximal ist müssen alle einzelnen Schwingungen zum Entscheidungszeitpunkt ihren Spitzenwert erreichen, Phasenverschiebung wäre hier nicht wünschenswert da sie den Zeitpunkt frequenzabhängig verschiebt. Dies wird dadurch erreicht, dass die Phasenantwort des Matched Filter genau dem invertierten Phasenspektrum des Pulses entspricht. Der gefilterte Puls hat dann ein Nullphasenspektrum. In der Praxis muss noch eine lineare Phasenverschiebung um $e^{-j\omega T_0}$ (also eine Gerade im Phasenspektrum), die einer Zeitverschiebung um T_0 entspricht, hinzugefügt werden, damit die einzelnen Cosinuskomponenten genau zum Zeitpunkt T_0 ihre konstruktive Interferenz bilden. Gesamt bekommen wir also für die Phasenantwort des Filters:

$$\phi(f) = -\phi_v(f) - 2\pi T_0 f \quad (rad)$$

Schließlich ergibt sich für die Übertragungsfunktion des Matched Filter:

$$H(f) = kV^*(f)e^{-j\omega T_0}$$

wobei die Konjugation $*$ für das Nullphasenspektrum sorgt.



Um die Impulsantwort zu erhalten, muss lediglich die Frequenzantwort invers Fourier transformiert werden. Dabei ergibt sich:

$$h(t) = k\, v^*(T_0 - t)$$

D.h. die Impulsantwort eines Matched Filter ist die zeitumgekehrte Version des erwarteten Pulses (also des Pulses, für den er designed wurde), die zusätzlich um die Pulsdauer verzögert ist (das muss sie auch, sonst wäre der Filter nicht kausal wenn er das Ende des Pulses schon ausgeben kann bevor er am Eingang einen Impuls bekommen hat).

Das Ausgangssignals des Filters ergibt sich durch Faltung der Impulsantwort mit dem Eingangssignal. Die Faltung besteht bekannterweise aus dem Umkehren einer Funktion und der Zeitverschiebung des Resultats multipliziert mit der anderen Funktion, und die Faltung ist kommutativ. Da die Impulsantwort bereits eine umgekehrte Kopie des erwarteten Eingangssignals ist, und die Faltung noch eine Umkehrung erfordert ist das Ausgangssignal durch das Integral über das Produkt des Eingangssignals mit der umgekehrten Impulsantwort gegeben. In dem Fall, dass das Eingangssignal der erwartete Puls ist, multipliziert man unter dem Integral also das Eingangssignal (oder in diesem Fall gleichbedeutend die Impulsantwort) mit sich selbst. Das Ausgangssignal ist für das erwartete Signal also die Autokorrelation des Eingangssignals (oder gleichbedeutend die Autokorrelation der Impulsantwort):

$$v_{out}(t) = v_{out}(T_0 - \tau) = k\, R_{v_{in}v_{in}}(\tau) = k\, R_{hh}(\tau) \qquad mit \tau := T_0 - t$$

Das Ausgangssignals eines Matched Filters ist also, wenn er mit dem erwarteten Puls, für den er designed wurde, angeregt wird, bis auf eine multiplikative Konstante und eine Zeitverzögerung T_0 die

Autokorrelation des (erwarteten) Eingangssignals.

Man beachte auch, dass in obiger Formel $v_{out}(T_0) = kR_{v_{in}v_{in}}(T_0 - T_0) = kR_{v_{in}v_{in}}(0)$ ist, und die Autokorrelation hat bekanntlich an 0 seine Maximumstelle.

Verwendet man kein OOK oder mehrere Signallevel oder Pulsformen, so kann man für jede vorkommende Pulsform einen eigenen Matched Filter verwenden und sich dann für das Symbol entscheiden, dessen Matched Filter den höchsten Wert lieferte.

8.2.2 SNR zum Entscheidungszeitpunkt

Das SNR zum Entscheidungszeitpunkt wird hier für Korrelatoren hergeleitet. Man beachte bei voriger Herleitung dass ein Matched Filter (mit Konstante $k = 1$) zum Zeitpunkt T_0 (bzw. allgemeiner zum Zeitpunkt kT_0) die Kreuzkorrelation zwischen Eingangssignal v_{in} und erwartetem Puls v_{exp} berechnet, also $R_{v_{in}v_{exp}}$, ein Matched Filter ist also, zumindest für die Zeitpunkte kT_0, ein Multiplikator mit $v_{exp}(t)$ und danach ein Integrator, also:

$$f(kT_0) = \int_{(k-1)T_0}^{T_0} v_{in}(t)v_{exp}(t)dt$$

Sei von nun an $v(t) := v_{exp}(t)$.
Für $v_{in}(t) = v_{exp}(t)$ gibt der Filter also $\int_0^{T_0} v_{in}^2 dt$ zurück, also die Symbolenergie E_S, das Betragsquadrat liefert die Symbolleistung:

$$f(kT_0) = E_S \Rightarrow |f(kT_0)|^2 = E_S^2$$

Wir berechnen nun (das brauchen wir später) die Autokorrelationsfunktion von Rauschen $n(t)$, also

$$R_{nn}(\tau) = \langle n(t), n(t+\tau)\rangle$$

Weißes Gaußsches Rauschen hat ein konstantes zweiseitiges Spektrum der Höhe $N_0/2$, nach dem Wiener-Kintchine Theorem können wir daher $R_{nn}(\tau)$ durch inverse Fouriertransformation von $N_0/2$ berechnen, also:

$$R_{nn}(\tau) = \frac{N_0}{2}\delta(\tau)$$

Nun wird das Rauschen aber noch durch $v(t)$ multiplikativ verstärkt, um daher seine Leistung berechnen zu können wollen wir die Autokorrelationsfunktion von $R_{xx}(\tau)$ mit $x(t) := v(t)n(t)$ berechnen, um danach wieder das Wiener-Kintchine Theorem anwenden zu können, dann in die andere Richtung. Wir berechnen:

$$R_{xx}(\tau) = \langle n(t)v(t), n(t+\tau)v(t+\tau)\rangle = \langle n(t), n(t+\tau)\rangle\langle v(t), v(t+\tau)\rangle = \frac{N_0}{2}\delta(\tau)R_{vv}(\tau)$$

Dabei gilt das mittlere Gleichheitszeichen deshalb, da für unabhängige Zufallsvariablen gilt $E(XY) = E(X)E(Y)$, das lässt sich (durch einen konstanten Faktor $\frac{1}{b-a}$) unmittelbar auf $\int_a^b x(t)y(t)dt =$

$\int_a^b x(t)dt \cdot \int_a^b y(t)dt$ verallgemeinern, wenn $x(t)$ und $y(t)$ unabhängig sind. In unserem Fall sind $v(t)$ und $n(t)$ natürlich statistisch unabhängig. Da $\delta(\tau) = 0$ für $t \neq 0$ gilt weiters:

$$R_{xx}(\tau) = \frac{N_0}{2}\delta(\tau)R_{vv}(0)$$

$R_{vv}(0)$ ist der mean square Wert von $v(t)$, also:

$$R_{xx}(\tau) = \frac{N_0}{2}\delta(\tau)\frac{1}{T_0}\int_0^{T_0} v^2(t)dt = \frac{N_0}{2}\delta(\tau)\frac{E_S}{T_0} \quad [\mathrm{V}^2]$$

Über das Wiener-Kintchie Theorem erhält man das Leistungsdichtespektrum von $x(t)$:

$$G_x(f) = \frac{N_0}{2}\frac{E_S}{T_0} \quad [\mathrm{V}^2/\mathrm{Hz}]$$

Wir haben nun also das Leistungsdichtespektrum des Störsignals unter dem Integral berechnet. Um das Leichtungsdichtespektrum des Störsignals nach dem Integrieren zu bestimmen, berechnen wir zunächst die Übertragungsfunktion des Integrals. Die Impulsantwort jedenfalls ist ein Rechteck der Breite T_0 und Höhe 1 mit Mittelpunkt 0, also $\prod \frac{(t-T_0/2)}{T_0}$. Die Fouriertransformierte gibt uns die Übertragungsfunktion

$$H(f) = T_0 \mathrm{sinc}(T_0 f)e^{-j\omega T_0/2}$$

Jetzt setzen wir zusammen was wir haben und bekommen das Leichtungsdichtespektrum $N(f)$ des Rauschens mit

$$N(f) = G_x(f)|H(f)|^2 = \frac{N_0}{2}E_S T_0 \mathrm{sinc}^2(T_0 f)$$

Die gesamte Rauschleistung bekommen wir daher mit

$$N = \frac{N_0}{2}E_S T_0 \int_{-\infty}^{\infty} \mathrm{sinc}^2(T_0 f)df = \frac{N_0}{2}E_S \quad [\mathrm{V}^2]$$

(Das Integral über $sinc^2$ ist $1/T_0$, das folgt aus dem sogenannten „value at origin theorem“ der Fouriertransformation und der Fourierkorrespondenz zwischen $sinc^2$ und einem Dreieck). Nun berechnet man leicht:

$$\frac{S}{N} = \frac{E_S^2}{\frac{N_0}{2}E_S} = \frac{2E_S}{N_0}$$

Der Signalrauschabstand hängt also nur von der Pulsenergie und dem Rauschleistungsdichtespektrum, nicht jedoch der Pulsform ab, wir haben hier keine Abhängigkeit von $v(t) = v_{exp}(t)$ mehr.

8.2.3 Matched Filter Detektion vs. Korrelationsdetektion

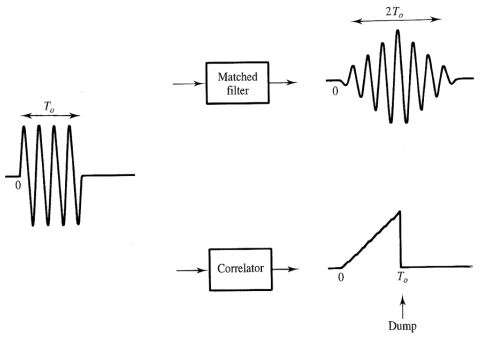
Korrelations- und Matched Filter Detektoren geben zwar bei gleichem Eingangssignal identische Rausch- und Signalspannungen zum Abtastzeitpunkt ab, und haben zum Abtastzeitpunkt auch die selben Werte, erzeugen jedoch nicht notwendigerweise die gleiche Ausgangspulsform. Die beiden Ausgangssignale sind:

$$f(t) = \int_0^t v^2(t')dt'$$

für den Korrelator und

$$f(t) = \int_0^t v(t')v(t' + T_0 - t)dt'$$

für den Matched Filter, bei den die erwartete Funktion über das Eingangssignal gleitet. Nur für $t = T_0$ haben wir notwendig Gleichheit in obigen Formeln. Nachfolgendes Bild zeigt die unterschiedlichen Ausgangssignale für Korrelator und Matched Filter, man sieht unmittelbar, dass der Matched Filter hier wesentlich anfälliger auf Timerjitter, also leichte Fehler beim Abtastzeitpunkt T_0, ist.



8.3 Root Raised Cosine Filterung

Bisher haben wir senderseitig einen Nyquistfilter als ideale Lösung gefunden (ISI-frei) und empfängerseitig den Matched Filter. Man beachte jedoch, dass die Nyquist-Frequenzantwort nicht nur senderseitig, sondern für den Puls bis zum Empfänger vor der Abtastung spezifiziert wurde, man kann also nicht den Nyquistfilter einfach nur senderseitig und den Matched Filter empfängerseitig implementieren. Für die Übertragungsfunktion des Nyquistfilters gilt, wenn man den Einfluss des Kanals als vernachlässigbar voraussetzt, $H_N(f) = H_T(f)H_R(f)$ (T für Transmitter, R für Receiver). Die Frequenzantwort kann also beliebig zwischen Sender und Emfpänger aufgeteilt werden. Kann man die Aufteilung so vornehmen, dass $H_T^*(f) = H_R(f)$, so wäre auch die Bedingung für den Matched Filter erfüllt[12], und man könnte sowohl das Nyquist als auch das Matched Filter Kriterium gleichzeitig erfüllen. Diesem Umstand wird der Root Raised Cosine Filter gerecht, man teilt also den Full Raised Cosine Filter in 2 Teile auf (darum die Wurzel):

$$H_T(f) = H_R(f) = \begin{cases} \sqrt{\cos^2(\pi f/(4f_\chi))}, & f \leq 2f_\chi \\ 0, & f > 2f_\chi \end{cases}$$

wobei $f_\chi = 1/(2T_0)$. Die Bandbreite des Root Raised Cosine Filters ist $B = 1/T_0 \quad (Hz)$.

9 Informationstheorie, Quellkodierung und Verschlüsselung

Die Informationstheorie beschäftigt sich mit der Repräsentation von Information durch Symbole und ermöglicht die Messung der Effizienz von Kommunikationssystemen. Folgende Termini werden benutzt:

[12] Hier geht man von Impulsen als Pulse aus, nur so ist das Spektrum der am Empfänger ankommenden Pulsform gleich dem des Sendefilters

- Symbol: ein Zeichen des Quellalphabets
- Baud: Symbolübertragungsrate (Symbole/sec)
- Bit: Informationsmenge eines Symbols mit Wahrscheinlichkeit $P = 0.5$

9.1 Information und Entropie

Der Informationsgehalt steht im Bezug zur Wahrscheinlichkeit: desto wahrscheinlicher es ist, dass eine gewisse Nachricht übertragen wird, desto niedriger ist deren Informationsgehalt. Der Informationsgehalt sollte also mit steigender Nachrichtenwahrscheinlichkeit $P(m)$ abnehmen und für eine Wahrscheinlichkeit von $P(m) = 1$ Null betragen. Weiters sollte der Informationsgehalt zweier Nachrichten die Summe des Informationsgehaltes der einzelnen Nachrichten sein. Da die Wahrscheinlichkeit der gesamten Nachricht gleich dem Produkt der Wahrscheinlichkeiten der einzelnen Nachrichten ist, muss also der Informationsgehalt summiert werden, wenn die Wahrscheinlichkeiten multipliziert werden. Diese Anforderung erfüllt der Logarithmus und der Informationsgehalt I_m einer Nachricht m ist folgendermaßen definiert:

$$I_m = \log \frac{1}{P(m)} = -\log P(m)$$

9.2 Entropie einer binären Quelle

Entropie ist die mittlere Information, die ein Symbolalphabet beinhaltet, also der Erwartungswert des Informationsgehaltes eines Alphabetes. Für ein Alphabet der Größe 2 und unter der Annahme, dass alle Symbole statistisch unabhängig sind, gilt:

$$H = \sum_{m=1}^{2} P(m) \log_2 \frac{1}{P(m)} \quad [\text{Bit/Symbol}]$$

Die Entropie erreicht ihr Maximum wenn die Symbole gleichwahrscheinlich sind (siehe Abb. 2). Hat irgendeine Nachricht eine Wahrscheinlichkeit von Eins, so ist die Entropie Null - die entsprechende Nachricht wird fortwährend gesendet.

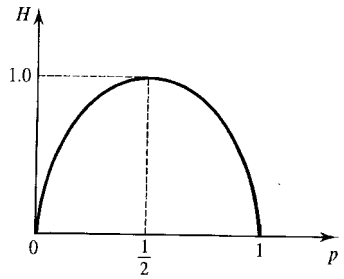


Abbildung 2: Entropie einer binären Quelle

9.3 Informationsverlust durch Rauschen

Bei einem verrauschten Übertragungskanal besteht empfängerseitig immer eine gewissen Ungewissheit, welches Symbol gesendet wurde. Diese Ungewissheit lässt sich durch die bedingte Wahrscheinlichkeit von $P(i_{TX}|j_{RX})$ ausdrücken. Die tatsächliche Entropie H_{eff} beim Empfänger ist um diese Ungewissheit (E)[13] kleiner, es gilt:

$$H_{eff} = H - E$$

mit

$$E = \sum_i \sum_j P(i_{TX}|j_{RX}) \log_2(\frac{1}{P(i_{TX}|j_{RX})}) \quad (bit/symbol)$$

E beschreibt hier sozusagen die mittlere Information, also die Entropie, aller „Verwechslungen". Man beachte, dass E für einen Fehlerfreien Kanal 0 ist, da dann $P(i_{TX}|j_{RX}) = 0$ für $i \neq j$ und $P(i_{TX}|j_{RX}) = 1$ für $i = j$ ist, also entweder der erste Faktor oder der Logarithmus ist stets 0. Je größer $P(i_{TX}|j_{RX})$ für $i = j$, desto kleiner wird der Ausdruck.
Die Ungewissheit E kann auch als unechte oder negative Information angesehen werden.

9.4 Quellkodierung

In der Nachrichtentechnik wird das Kodieren von Nachrichten aus einer Quelle durch einen Sender als Quellkodierung bezeichnet. Quellkodierung beeinflusst die Quellentropie nicht; die Entropie ist also eine fundamentale Eigenschaft der Quelle selbst. Quellkodierung beeinflusst allerdings die Symbolentropie und kann auch Fluktuationen in der Informationsrate reduzieren.

9.4.1 Codeeffizienz

Wird ein Symbolsatz durch Kodierung auf einen anderen Symbolsatz übersetzt, so ändert sich die Entropie, die Entropie des kodierten Symbolsatzes sei mit H bezeichnet. Die Codeeffizienz wird durch den Quotient von tatsächlicher und maximaler Entropie definiert:

$$\eta_{code} = \frac{H}{H_{max}} \times 100\%$$

Die größtmögliche Entropie einer Quelle H_{max} wird erreicht, wenn alle Symbole gleichwahrscheinlich sind. Sind alle M Symbole gleichwahrscheinlich, so kann man H_{max} berechnen mit:

$$H_{max} = log_2(M)$$

Werden Quellsymbole in ein *binäres* Alphabet übersetzt, so haben die M einzelnen Codewörter eine Länge l_m und eine mittlere Codewortlänge L ist folgendermaßen definiert:

$$L = \sum_{m=1}^{M} P(m) l_m$$

Bei der Codeeffizienz kann dann statt H_{max} die mittlere Codewortlänge L eingesetzt werden. Das mag zunächst verwirrend sein, man beachte jedoch, dass sich mit L normalerweise auch H ändert, anderes Quellalphabet ⇒ anderes Zielalphabet.

[13] E für equivocation (=Mehrdeutigkeit)

9.4.2 Dekodieren von Codewörtern variabler Länge

Ein effizienter Code repräsentiert Information mit möglichst wenigen Ziffern. Dies hat eine variable Codewortlänge zur Folge. Zwei Eigenschaften sind von essentieller Bedeutung, sollen solche Codewörter variabler Länge dekodiert werden:

- Eindeutiges Dekodieren: Ein Code, bei dem ein A durch 0 und ein B durch 00 repräsentiert wird, lässt sich die Folge 0000 nicht eindeutig dekodieren - es könnte sich um ABA, AAAA, BB oder ähnliches handeln.
- Sofortiges Dekodieren: Um sofortiges Dekodieren zu ermöglichen, darf kein vollständiges Codewort Präfix eines größeren Codewortes sein. Bsp.: A $\leftrightarrow$ 1, B $\leftrightarrow$ 10 kann nicht sofort dekodiert werden, da nach Empfang von 1 noch gewartet werden muss, ob nachher eine 0 kommt... das kann problematisch sein... wie lange wartet man?

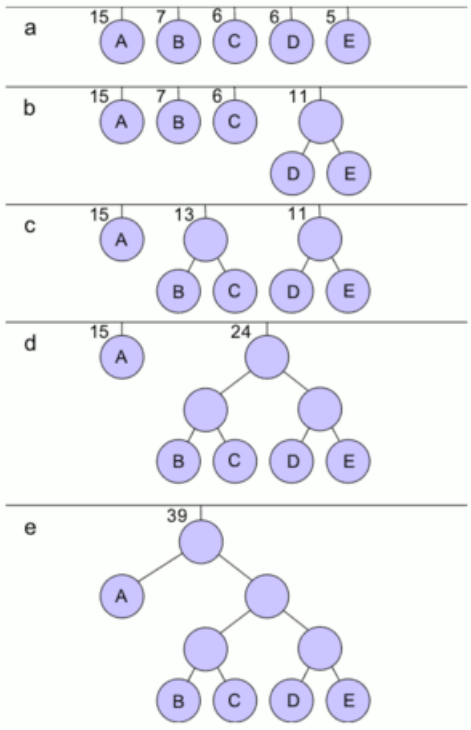
9.4.3 Huffman-Kodierung

Der Huffman-Code ist ein optimaler Code (hinsichtlich der Codeeffizienz), er ist präfix-frei und daher sofort dekodierbar, und natürlich eindeutig.

Er läuft in 2 Schritten ab:

- Reduktion: In jedem Schritt wird zuerst nach WK sortiert und danach jeweils die beiden unwahrscheinlichsten Knoten zu einem zusammengefasst... so lange bis es nur noch 2 Knoten gibt.
- Splitting: Jetzt wird in der umgekehrten Reihenfolge ein binärer Baum erstellt, alle linken Kanten bekommen dann z.B. eine 0, alle rechten eine 1. Will man den Code eines Zeichens wissen, so muss man die binären Symbole auf allen Kanten am Weg von der Wurzel bis zum Symbol zusammensetzen.

Üblicherweise führt man beide Schritte in einem aus und fügt zwischendurch Pseudoknoten ein (siehe Bild). Da der Code präfixfrei ist können nur die Blätter des Baumes Symbole sein.



9.5 Beispiele für Quellkodierung

Morsecode Der am öftesten auftretende Buchstabe, das E (.), hat das kürzeste Codewort, das Y ein 4-bittiges (– . – –).

Fax Beim Group 3 Fax Standard wir das Dokument jeweils zeilenweise abgetastet. Es werden die Längen von schwarzen bzw. weißen Linien kodiert (Run Length), also sowas wie 10 Pixel weiß, 3 Pixel Schwarz, 14 Pixel weiß usw. Man fertigte Statistiken von Dokumenten, die üblicherweise gefaxt werden, an und kodierte die Run Lengths mit dem Huffman-Code um so häufigen Run-Lengthes kleine Codelängen zuordnen zu können.

Vocoder Vocoder benutzen Modelle des physischen Mechanismus der Spracherzeugung, um verständliche Sprachsignale kleiner Bitrate zu erzeugen. Grundlegend ist die Beobachtung, dass es 4 formierende Signalformen (engl. Formants) gibt, aus denen Sprache aufgebaut wird. Die Sprache wird nun (mit Korrelatoren) auf diese Formants untersucht, und danach wird nur noch Dauer und Amplitude der Formants übertragen. Üblicherweise wird das Eingangssignal in Frames von 20 ms aufgeteilt. Beim Empfänger werden Pulsgeneratoren verwendet, um das ursprüngliche Signal wiederherzustellen.

Subbandkodierung Bei der Subbandkodierung wird das übertragene Signal in zwei oder mehrere Subfrequenzbänder unterteilt. Da die unterschiedlichen Frequenzbänder Signale unterschiedlicher Energie enthalten kann die Genauigkeit (Anzahl der Bits) der einzelnen Quantisierer im Encoder für niedrigenergetische Bänder reduziert werden. Diese Technik kann auch für Videosignale verwendet werden.

Audiokodierung Effiziente Audiokodierung nutzt die Frequenzcharakteristik des menschlichen Ohrs aus: Signale mit hohem Level maskieren (im Frequenzbereich) benachbarte Signale mit niedrigem Level. Höhere Frequenzen sind vom Ohr sowieso nur bei viel höheren Signalstärken hörbar, das Ohr hat also eine frequenzabhängige Empfindlichkeit, der sogenannte sound pressure level treshold ist frequenzabhängig. Auch das macht man sich natürlich zu Nutze, hohe Töne, die zu leise sind, müssen nicht übertragen werden.

Stringkodierung Stringkodierung ist besonders effizient um zeichenbasierte Daten zu kodieren: oft vorkommende Zeichenketten werden durch kürzere Codewörter repräsentiert. Manchmal macht man sich die Redundanz auch so zu Nutze, dass man, wenn ein String wiederholt auftritt, bei Wiederholungen nur noch die Position des ersten Auftretens übertragt und nicht mehr den String selbst. Der zeitliche Aufwand ist jedoch hoch, da die oft auftretenden Zeichenketten erst herausgefunden werden müssen.

10 Kanalkodierung (engl. error coding oder channel coding)

Fundamentale Ressourcen bei digitaler Kommunikation sind Signalstärke, -zeit und -bandbreite, die gegeneinander eingetauscht werden können. Ziel ist es üblicherweise, bei einer maximalen Datentransferrate und einer minimalen Bandbreite noch eine akzeptable Qualität der Übertragung zu gewährleisten. Error control coding (Kanalkodierung) dient dazu, Fehler in der Übertragung der Symbole zu entdecken und eventuell auch zu korrigieren.

Es gibt zwei Messgrößen für die Fehlerperformance: die Bitfehlerrate (BER, bit error rate) und die bekannte Wahrscheinlichkeit eines Bitfehlers P_b. Die Bitfehlerrate ist die mittlere Rate zu der Fehler auftreten und durch das Produkt von $P_b R_b$ gegeben, wobei R_b die Bitübertragungsrate des Kanals ist.

10.1 Kanalkodierungs-Konzepte

Ist die Fehlerrate eines Systems zu hoch, so gibt es mehrere Möglichkeiten, sie zu verbessern:

Senderstärke Das Erhöhen der Senderstärke ist ein einfaches Konzept, aber ungünstig im Fall begrenzter Energieressourcen (z.B. Batteriebetrieb).

Diversität Es gibt drei Arten von Diversität: Raumdiversität, Frequenzdiversität und Zeitdiversität. Bei allen drei Konzepten wird Redundanz umgesetzt: Raumdiversität benutzt zwei Antennen; Frequenzdiversität sendet ein Signal auf zwei unterschiedlichen Frequenzen; bei Zeitdiversität wird eine Nachricht öfters hintereinander gesendet.

Full Duplex Sobald ein Sender Daten an den Empfänger überträgt, sendet letzterer sie sofort auf einem seperaten Kanal zurück. Sieht der Sender, dass eine Nachricht fehlerhaft war, so sendet er sie erneut. Diese Technik benötigt die doppelte Bandbreite einer Half Duplex Übertragung.

ARQ Wird bei automatic repeat request ein Fehler in der Übertragung erkannt, so fordert der Empfänger über einen Feedback-Kanal den fehlerhaften Datenblock erneut an. Man beachte dass bei langsamen Links wie Satellitenübertragung ARQ oft keine Option ist, hier benötigt man FECC:

FECC Bei forward error correction coding wird mittels eines data check Bits unter den Informationsbits Redundanz implementiert. Ein Nachteil ist die benötigte Zeit um die Nachrichten zusammenzusetzen und empfängerseitig zu prüfen. Die Hardwarekomplexität ist heutzutage dank VLSI-Technologien kein nennenswertes Problem mehr.

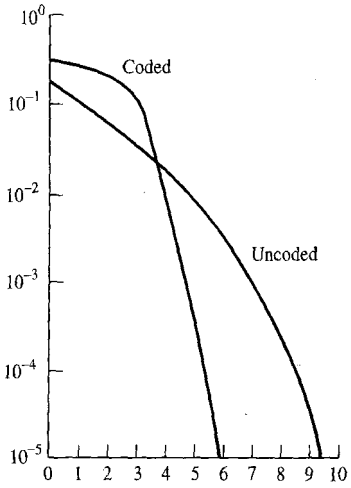
10.1.1 ARQ-Techniken

Es gibt prinzipiell zwei ARQ-Techniken, wobei beiden gemein ist, dass, sofern der Sender keine (weder positive noch negative) Rückmeldung erhalten hat, ein Datenblock spätestens nach dem Ablaufen eines Time-Outs erneut übertragen wird.

- Bei *stop and wait* ARQ wird jeder Datenblock vom Empfänger als positiv oder negativ bestätigt, bevor ein neuer Datenblock übertragen wird
- Bei *continuous* ARQ werden, ohne die empfängerseitige Bestätigung abzuwarten, Datenblöcke gesendet. Dabei wird wiederum zwischen zwei Varianten unterschieden:
  - Bei der *go-back-N* Version hat jeder Datenblock eine Sequenznummer N. Jedes Quittungssignal enthält die Sequenznummer eines Datenblocks und bestätigt damit den Empfang aller Datenblöcke bis $N-1$. Bei Time-Out oder einer negativen Quittierung werden alle Datenblöcke ab dem fehlerhaften erneut übertragen.
  - Bei der *selective repeat* Version werden nur die Datenblöcke mit negativer Quittierung bzw. Time-Out erneut übertragen. Hier müssen die ankommenden Pakete eventuell umgeordnet werden, das ist bei go-back-N nicht nötig.

10.1.2 Der Schwelleneffekt in der Bitfehlerwahrscheinlichkeit

Betrachtet man nachfolgende Abbildung, wo SNR (x-Achse) gegen P_b, also Bitfehlerrate (y-Achse) aufgetragen ist, so fällt auf, dass das FECC-kodierte Signal bei ca. 6dB quasi keine Fehler mehr macht. Allerdings ist die FECC-kodierte Version für schlechtes SNR, also viel Noise, sogar schlechter als die unkodierte Version. Grund dafür ist, dass in einem Bereich geringen Signal-Rauschabstandes der Dekodierer durch den Versuch Fehler zu korrigieren, die Fehleranzahl cirka verdoppelt. Unter 4dB SNR ist es also sogar besser, auf FECC zu verzichten.



10.2 Güteparameter für Codes

Es gibt zwei Güteparameter für Codes:

- Die *Hammingdistanz* gibt an, um wie viele Bits sich zwei Codewörter unterscheiden, d.h. wie leicht es ist ein Codewort falsch als ein anderes zu interpretieren.
- Das *Gewicht* eines Codewortes gibt an, wie viele Einsen das Codewort enthält.

10.3 Block Codes

Ein Blockkodierer nimmt ein k-stelliges Informationscodewort und macht daraus ein n-stelliges kodiertes Codewort. Das n-stellige Codewort besteht daher aus k Informationsbits und $(n-k)$ redundanter Paritäts-Bits. Das Verhältnis von k zu n ist die *Rate* oder *Effizienz* des Codes $R = \frac{k}{n}$ und bewegt sich normalerweise zwischen 0,5 und 1.

Es gibt zwei Arten von Blockcodes:

- Bei *systematischen* Codes werden, wie soeben beschrieben, die Informationsbits explizit zusammen mit den Paritätsbits übertragen. Gemäß der strengen Definition müssen die Informationsbits und die Paritätsbits jeweils als konsistenter Block übertragen werden, bei der lockeren Definition müssen die Informationsbits lediglich im Codewort enthalten sein.
- Bei nicht systematischen Codes darf das n-stellige Codewort keine der Informationsbits explizit enthalten.

Ein Beispiel ist der single parity check code. Dem Codewort wird ein Paritätsbit P angehängt: enthält das Codewort eine ungerade Anzahl an Einsen, so ist $P = 0$, bei einer geraden Anzahl an Einsen ist $P = 1$. Die Effizienz beträgt $\frac{7}{8}$ und zeigt so von geringer Redundanz. Einfache Paritätsüberprüfung kann nur eine ungerade Anzahl an Fehlern erkennen. Einfache Paritätsüberprüfung wird oft auch bei Matrizen für jede Zeile und jede Spalte verwendet. Tritt nur ein Fehler auf, so stimmt genau eine Zeilen- und eine Spaltenprüfsumme nicht, der Fehler kann daher erkannt und korrigiert werden. Für 2 Fehler können die Fehler zwar lokalisiert, aber nicht korrigiert werden.

Paritätsbits können auf Basis einer Paritätsüberprüfungsmatrix H berechnet werden. Eine solche Matrix sieht z.B. folgendermaßen für einen $(7, 4)$ Blockcode aus:

$$H = \begin{bmatrix} 1011 : 100 \\ 1101 : 010 \\ 1110 : 001 \end{bmatrix}$$

Dabei gibt die linke Seite die Koeffizienten für die Nachrichtenbits an und die rechte Seite die Information darüber, welches Paritätsbit betroffen ist. So gibt die erste Reihe die Information für das Paritätsbit P_1 an: $P_1 = 1I_1 \oplus 0I_2 \oplus 1I_3 \oplus 1I_4$.

Eine weitere beliebte Technik für Prüfsummen ist einfach die Summation modulo 2^{8n}, man bekommt dann eine n Byte lange Prüfsumme. Für $n = 1$ wird meist das Komplement der Summe übertragen, sodass der Empfänger nur noch alle Bytes aufsummieren muss und überprüfen muss, ob die Summe 0 ist.

10.4 Fehlerwahrscheinlichkeit eines Codewortes

Die Wahrscheinlichkeit von genau j Fehlern in n Zeichen mit einer Fehlerwahrscheinlichkeit von P_e pro Zeichen in einem Codewort beträgt:

$$P(j \text{ Fehler}) = \binom{n}{j} P_e^j (1 - P_e)^{n-j}$$

Die Wahrscheinlichkeit von mehr als R' Fehlern beträgt somit:

$$P(\text{Fehleranzahl} > R') = 1 - \sum_{j=0}^{R'} P(j)$$

Die Anzahl an Fehlern in einem großen Datenblock ist nahe bei $P_e n$. Die Anzahl der Blöcke mit von $P_e n$ stark unterschiedlichen Fehlerraten sinkt wenn die Blocklänge n steigt (die beiden letzten Aussagen folgen daraus, dass die Fehler benachbarter Bits als unabhängig vorausgesetzt werden und man dann die Summe der Fehler betrachtet. Diese nähert sich nach dem Gesetz der großen Zahlen dem Erwartungswert nP_e, die Grenzverteilung ist eine Normalverteilung.) Ein Code, der $P_e n$ Fehler in einem Block korrigieren kann, ist bei hohem n also ein Garant dafür, dass das Kodiersystem nur in wenigen Fällen fehlschlägt. Blockcodes, im Speziellen lineare Gruppencodes, lassen sich sehr gut hinsichtlich ihrer Performance analysieren.

10.5 Lineare Gruppencodes

Gruppencodes beinhalten das Null-Codewort und sind abgeschlossen, sie verhalten sich also wie eine mathematische Gruppe. Nimmt man zwei Codewörter C_i und C_j so gilt $C_i \oplus C_j = C_k$ (bitweise Addition modulo 2), wobei C_k ebenfalls ein Codewort ist. Gruppencodes werden normalerweise polynomiell generiert und lassen sich in zwei Hauptgruppen unterteilen: die binären Bose-Chaudhuri-Hocquenghem (BCH)-Codes und die in CD-Playern und Computerspeichern eingesetzten symbolweise organsierten Reed-Solomon Codes.

10.5.1 Performance von Gruppencodes

Die wichtigste Messgröße für die Vorhersage von Code-Performance ist die minimale Hammingdistanz zwischen allen Codewortpaaren. Normalerweise müssen dazu alle Codewortpaare untersucht werden, die Anzahl dieser Paare steigt kombinatorisch. Bei Gruppencodes reicht es jedoch jedes der Codewörter mit dem Null-Codewort zu vergleichen: Die Wahrscheinlichkeit, dass C_i als C_j missinterpretiert wird, hängt von dem Abstand zwischen C_i und C_j ab. Der Abstand ist aber das Gewicht des Codewortes C_k, wobei $C_i \oplus C_j = C_k$ gilt.

10.6 Fehlerkorrektur von Codes

Die maximale Anzahl an korrigierbaren Fehlern t ist, wenn D_{min} die minimale Hammingdistanz und e die Anzahl an erkennbaren Fehlern ist, gegeben durch:

$$t = int(\frac{D_{min} - 1}{2}) \qquad wobei\ D_{min} - 1 = e + t$$

Man muss also einen Mittelweg zwischen erkennbaren und korrigierbaren Fehlern finden.

10.7 Hamming-Bound

Die Fehlerkorrigierungsstärke eines Codes beträgt t, wenn er alle Codewörter mit t oder weniger Bitfehlern korrigieren kann. Es kann sein dass ein Code zwar mehr Fehler erkennen kann, jedoch nicht korrigieren, da sich ein empfangenes, als inkorrekt erkanntes Codewort von 2 korrekten Codewörtern durch die selbe Hamming-Distanz unterscheidet. Dann ist eine Entscheidung auf Basis eines Nearest Neighbour Algoritmus' nicht mehr möglich.
Der Hamming Bound gibt eine obere Schranke für die Performance von Blockcodes an. Für einen Code mit Codewortlänge n und k Informationsbits gilt:

$$2^k \leq \frac{2^n}{1 + n + \binom{n}{2} + \binom{n}{3} + \ldots + \binom{n}{t}}$$

Es gibt insgesamt 2^k dekodierte Codewörter. Für jedes dieser dekodierten Codewörter gibt es 1 Codewort ohne Fehler, n kodierte Codewörter mit einem Fehler, $\binom{n}{2}$ Codewörter mit zwei Fehlern bis $\binom{n}{t}$ Codewörter mit t Fehlern, die als dieses dekodierte Codewort missinterpretiert werden können. Dividiert man die Anzahl aller möglichen n-stelligen Codewörter, also 2^n durch die soeben aufgestellte Summe (der Divisor in obiger Ungleichung), so erhält man ein Maß für die Performance eines Codes (je kleiner desto besser). Dieser Performance ist aber durch obige Ungleichung eine Schranke nach unten gesetzt, will man also einen Code mit Codewortlänge n und k Informationsbits haben, der t Fehler erkennen soll, so kann man sich vorher mit dem Hamming Bound vergewissern, ob das überhaupt möglich ist. Für einen perfekten Code gilt in obiger Ungleichung das Gleichheitszeichen.

10.8 Syndrom

10.8.1 Kodierung

Blockcodes können mittels einer Generator-Matrix, einer Matrix aus Basisvektoren, konstruiert werden. Wir betrachten hier zur Illustration eine G-Matrix für den weiter oben verwendeten $(7, 4)$-

Blockcode mit folgender H Matrix:

$$H = \begin{bmatrix} 1011:100 \\ 1101:010 \\ 1110:001 \end{bmatrix}$$

Dann konstruiert man die G-Matrix durch die kanonischen Basisvektoren des $\mathbb{R}^4$, die linke Seite von G ist die Einheitsmatrix. Die rechte Seite von G ist der transponierte linke Teil der H-Matrix:

$$G = \begin{bmatrix} 1000:111 \\ 0100:011 \\ 0010:101 \\ 0001:110 \end{bmatrix}$$

In unserem Beispiel haben wir einen Code der einen Bitfehler korrigieren kann.
Die Berechnung des Codeworts c erfolgt durch Multiplikation der zu kodierenden Informationen d mit der Matrix G von links:

$$c = d \cdot G$$

In unserem Beispiel für $d = [1001]$:

$$[1001001] = [1001] \cdot \begin{bmatrix} 1000111 \\ 0100011 \\ 0010101 \\ 0001110 \end{bmatrix}$$

10.8.2 Dekodierung

Das Problem an der Nearest Neighbour Dekodierung[14] ist die Größe der entstehenden Tabelle. Die Syndrom-Dekodierung speichert aber nicht für jedes Codewort alle möglichen gestörten Codewörter, sondern nur sogenannte Syndrome. Die Syndrom-Dekodiertabelle ist unabhängig vom übertragenen Codewort und damit um den Faktor 2^k kleiner als eine Nachbarschaftstabelle. Die Fehlererkennung funktioniert wie folgt:

Ist d ein k-stelliger Nachrichtenvektor und c das n-stellige Codewort (noch ohne Störung), so gilt natürlich: $H\ c = 0$. Ist $r = c \oplus e$ eine empfangene Datensequenz, die aus Überlagerung von c mit einem Fehler e entsteht, so berechnet man das Syndrom s wie folgt:

$$s := Hr = H(c \oplus e) = Hc \oplus He = 0 \oplus He$$

Sind keine Fehler aufgetreten, so ist s gleich dem Nullvektor. Man wählt H so, dass H bijektiv ist, also zwischen $s = He$ und e eine eindeutige Zuordnung existiert, natürlich immer vorausgesetzt, dass nicht mehr Fehler aufgetreten sind als die t Fehler wie in der Konstruktion des Codes angenommen. Man speichert nun also nur noch die wesentlich kleinere Syndromtabelle und die zugehörigen Fehlervektoren e, man bekommt dann die wahre Nachricht durch $c = r \ominus e$.
Treten mehr als t Fehler auf, dann kann es natürlich zu Fehlern bei der Dekodierung kommen, auch wenn der Dekodierer auch in diesem Fall immer noch nach dem Nearest Neighbour Prinzip vorgeht.

[14] Steigt die Wahrscheinlichkeit einer gewissen Anzahl t von Bitfehlern schnell, gilt also $P(t) \ll P(t+1)$, so ist Nearest Neighbour Dekodierung äquivalent zur Maximum Likelihood Dekodierung

10.9 Zyklische Codes

Zyklische Codes sind eine Unterklasse der Gruppencodes, die kein Null-Codewort besitzen (z.B. Hamming-Code). Ihre Code-Struktur kann in Hardware einfach mittels Schieberegistern und XOR-Gattern implementiert werden. Ihr Name kommt daher dass man durch ein Codewort und seine zyklischen Verschiebungen (Bitshifts) alle Codewörter des zugehörigen Codes bekommt. RS und BCH sind zyklische Codes.

10.9.1 Kodierung & Dekodierung

Alle Operationen werden modulo 2 durchgeführt, das ist auch der Grund für die einfache Implementierung in Hardware (Addition und Subtraktion können modulo 2 mit XOR-Gattern implementiert werden). Ein CRC (cyclic redundancy check) ist der Rest einer binären Division modulo 2. CRCs werden vielfach zur Fehlererkennung eingesetzt. Wird eine Nachricht der Länge k mit den Bits $m_{k-1} \dots m_0$ übertragen, so wird sie als Polynom der Ordnung $k-1$ betrachtet:

$$M(x) = m_{k-1}x^{k-1} + \ldots + m_1 x + m_0$$

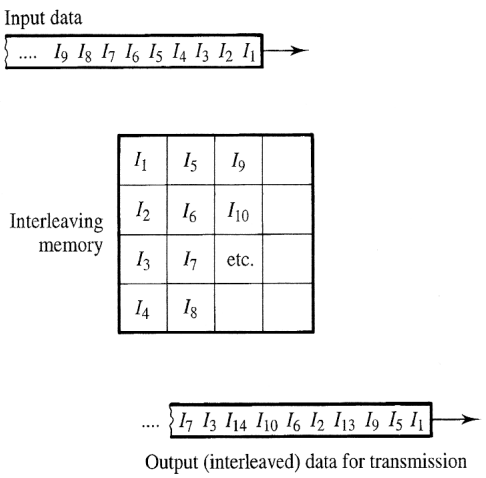
Die Nachricht $M(x)$ wird durch das Generatorpolynom $P(x)$ modifiziert um daraus die kanalkodierte Version von $M(x)$ zu erzeugen. Dazu wird zuerst $M(x)$ um i Stellen bitweise nach links geschoben, wobei i die Ordnung von $P(x)$ bezeichnet... das entspricht also der Multiplikation von $M(x)$ mit $P(x)$. Danach wird die nun erweiterte Version von $M(x)$ durch $P(x)$ durchdividiert. Der Rest, der bei dieser Division überbleibt, überschreibt dann die im ersten Schritt (beim bitweisen Shift) zu $M(x)$ hinzugefügten Nullen. Hardwaremäßig kann man mit XOR-Gattern und Schieberegistern während der Übertragung der zu sendenden Information die CRC berechnen und diese gleich im Anschluss übertragen, es entsteht hier also keine Verzögerung.

Beim Empfänger wird die empfangene Nachricht durch $P(x)$ durchdividiert - ist der Rest Null, so wurde die Nachricht korrekt übertragen. Ist der Rest nicht Null, so können über eine Syndromtabelle die Fehlerstellen ermittelt und die Fehler durch Addition mit dem entsprechenden Fehlermuster korrigiert werden.

Das Generator-Polynom wird sorgfältig ausgewählt, um möglichst viele Fehler korrigieren zu können. Ein Generatorpolynom der Ordnung k erlaubt die Fehlererkennung von bis zu k Bits die in Folge gestört wurden (sogenannte Burst Error).

10.9.2 Interleaving

Will man auch Burst-Error korrigieren, die relativ lang sind, also eventuell sogar einen ganzen Datenblock zerstören, so verwendet man Interleaving, wo man, grob gesagt, die einzelnen Bits oder Teilblöcke mischt, damit ein Burst-Fehler nur Teile des später wieder zusammengesetzten Blocks betrifft. Dabei verwendet man oft ein Memory-Array, in das man Daten spaltenweise schreibt und dann reihenweise ausliest und übertragt (siehe Bild). Anwendungsbeispiele sind Audio-CDs.

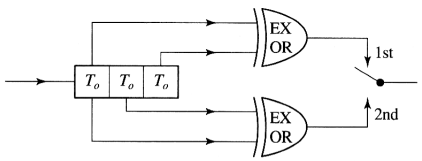


10.10 Faltungscodes

Der Name Faltungscode kommt daher, dass die Ausgabe aus der Faltung von Eingabedaten und der Impulsantwort des Kodierers besteht.

10.10.1 Kodierung

Faltungscodes werden generiert, indem die Eingabebits in ein Schieberegister geschoben werden. Pro Eingabebit werden dann mehrere Ausgabebits generiert, für jedes Ausgabebit gibt es ein eigenes Generatorpolynom. Das jeweilige Ausgabebit wird dann, dem Generatorpolynom entsprechend, durch XOR-Gatter aus den Bits im Schieberegister berechnet.
Am einfachsten ist das anhand eines Beispiels zu erklären: Wir betrachten hier einen Kodierer der Rate $1/2$, für 1 Eingabebit werden also 2 Ausgabebits generiert. Für jedes Ausgabebit gibt es ein Generatorpolynom, hier $P_1(x) = 1 + x^2$ und $P_2(x) = 1 + x$. Der höchste Grad dieser Polynome ist 2, man braucht also ein Schieberegister der Länge 3. An folgendem Bild sieht man wie die beiden Ausgabebits durch die XOR-Gatter den Polynomen entsprechend berechnet werden, der Switch reiht dann die Ausgabebits aneinander:

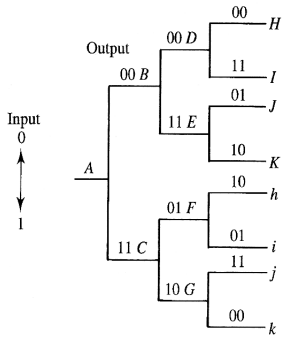


A simple example of a rate ½ convolutional encoder.

Dieser Kodierer ist nicht systematisch, da die Eingabebits nicht explizit in der Ausgabe vorkommen. Im Resetzustand wird das Schieberegister des Kodierers mit lauter Nullen gefüllt.

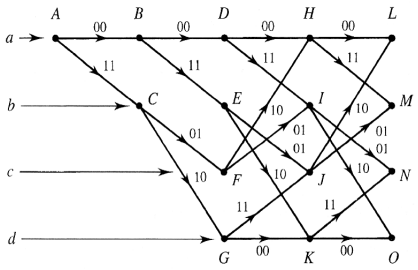
Man kann einen Kodierer durch verschiedene Zustandsdiagramme beschreiben, die alle mehr oder weniger übersichtlich sind.

- Baumdiagramm: Die Zustände repräsentieren immer die letzten $n-1$ Input-Bits, beschreiben also das Gedächtnis des Kodierers, und das nächste Eingabebit bestimmt die Übergänge. Auf die Kanten tragt man die Ausgabe auf.



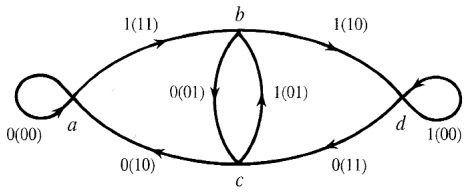
Tree diagram representation of the coder in Figure 10.18.

- Trellis-Diagramm: Hier versucht man redundante Zustände zu vermeiden, darunter leidet allerdings die Übersichtlichkeit. An den Kanten stehen wieder die Ausgabebits, für eine 0 als nächstes Eingabebit verwendet man immer die obere, für eine 1 als nächstes Eingabebit die untere Kante zum nächsten Zustand.



Trellis diagram representation of the coder in Figure 10.18.

- State Transition Diagramm: Aus dem Trellis-Diagramm gewinnt man das State Transition Diagramm. Über den Kanten stehen in der Form E(AA) das nächste Eingabebit E und das zugehörige Ausgabebit AA.



State transition diagram representation of the coder in Figure 10.18.

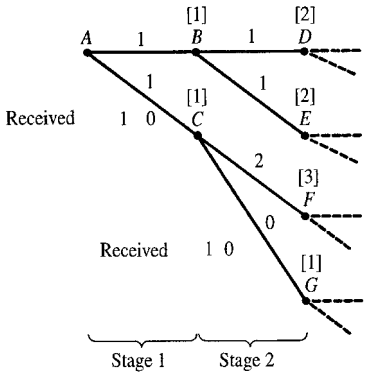
10.10.2 Dekodierung

Die Dekodierung von Faltungscodes erfolgt in der Praxis fast immer mit dem Viterbi-Algorithmus, der eine Nearest Neighbour Dekodierung implementiert. Mit Faltungscodes kodierte Nachrichten haben eine enorme Länge. Da der Speicherbedarf des Dekodierers mit der Länge der Nachricht wächst, werden kleinere Datenblöcke verarbeitet. Diese Datenblöcke enden immer mit Nullen um den Kodierer wieder in den Anfangszustand zu bringen... das verringert natürlich die Effizienz des Algorithmus'. Der Viterbi-Algorithmus wird bei jedem Dekodierschritt benutzt um den wahrscheinlichsten Pfad durch das Dekodiergitter zu finden.
Die Dekodierung verläuft wie folgt:

1. Das Trellisdiagramm beginnt im Startzustand
2. Ein Codewort wird eingelesen. Die Hammingdistanz zwischen dem eingelesenen Codewort und allen möglichen Codewörtern wird berechnet und als Kantengewicht eingetragen. D.h. dass der Dekodierer „denkt“ wie ein Kodierer und sich anhand der bisherigen Eingabe (also den bisherigen Verlauf durch das Zustandsdiagramm) überlegt, was der Kodierer für eine 0 bzw. eine 1 als nächste Eingabe ausgegeben hätte. Über die beiden Kanten (nächste Eingabe war 0, obere Kante, und nächste Eingabe war 1, untere Kante), wird die Hammingdistanz zwischen dem errechneten Codewort für die jeweilige Eingabe und dem tatsächlich empfangenen Codewort aufgetragen.
3. Die in Schritt 2 berechneten Kantengewichte werden vom Startzustand aus aufsummiert und in die Zustände eingetragen.
4. Nun werden die Zustände mit den minimalen Pfadkosten behalten, alle anderen verworfen (es kann mehrere Zustände geben, die alle die selben, minimalen Kosten haben)
5. Sprung zu Schritt 2 bis alle empfangenen Codewörter abgearbeitet sind.

Ein Beispiel für ein teilweises Trellisdekodierdiagramm ist in nachfolgender Abbildung ersichtlich. Im Startzustand A hätte ein Kodierer für die Eingabe 0 das Codewort 00 ausgegeben; empfangen wurde aber 10, also wird die Hammingdistanz 1 über der Kante aufgetragen. Hätte der Kodierer im Startzustand eine 1 bekommen, so wäre die Ausgabe 11 gewesen, wieder haben wir eine Hammingdistanz von 1 zum tatsächlich empfangenen Codewort. In diesem Schritt kann kein Pfad eliminiert werden, da beide Zustände Gesamtkosten von 1 haben. Betrachten wir nun noch Zustand C. Das letzte Eingabebit war eine 1, und wir nehmen an wir bekämen jetzt wieder eine 1 als Eingabebit ($C \rightarrow G$).

Der Kodierer würde in diesem Fall 10 ausgeben, und da wir auch 10 empfangen haben tragen wir hier also 0 als Hammingdistanz über die Kante $C \rightarrow G$. Die anderen 3 Übergänge macht man analog. Nach diesem Schritt kann man die Zustände D, E und F eliminieren. So sichert man, dass der Baum nur linear wächst und nicht exponentiell.



Da normalerweise große Datenblöcke verarbeitet werden und es nicht möglich ist, den ganzen Baum zu speichern, muss der Baum reduziert werden. Das passiert in Schritt 4 obiger Beschreibung. Dabei kann es allerdings passieren, dass man nicht mehr die optimale Entscheidung trifft: Es könnte in obigem Beispiel theoretisch passieren, dass nach Schritt C für die verbleibenden, sagen wir 10 Bits immer Hammingdistanz 1 berechnet wird, von Zustand F aus aber immer Hammingdistanz 0, und man also besser F hätte behalten sollen. In der Praxis werden oft die letzten N Übergänge gespeichert, die Größe N heißt dann Decision Window. So kann man zwar Speicher sparen, dennoch steigt der Speicher hier exponentiell. Die Fehlererkennung des Dekodierers wird aber besser als wenn man immer nur die Knoten mit minimalen Kosten behalten würde. Das Decision Window muss jedenfalls groß genug sein, um Burst-Fehler erkennen zu können. Eine vernünftige Größe des Decision Window wird meist in Computersimulationen bestimmt.

11 Bandpassmodulation eines Trägersignals

Modulation bezieht sich auf die Modifizierung der Signalcharakteristiken eines Trägersignals aufgrund eines Informationssignals. In der Folge wird Intermediate Frequency (IF) Modulation behandelt, wobei der Träger eine Sinusschwingung ist. Die Eigenschaften des Trägers, die verändert werden, sind: Amplitude, Frequenz oder Phasenlage.
Man betrachtet diese Art von Modulationen, da die Transformation von Basisbandsignalen in Bandpasssignale mehrere Vorteile hat:

1. Signale können den Charakteristika des Übertragungskanals besser angepasst werden
2. Es kann Frequenzmultiplexing verwendet werden um mehrere Signale gleichzeit zu übertragen
3. Effiziente Antennen, die vernünftig klein sind, können verwendet werden

11.1 Spektrale Effizienz und Leistungseffizienz

Als spektrale Effizienz bezeichnet man das Verhältnis von Informationsrate zur Bandbreite, je höher desto besser.
Eine ähnlich einfache Möglichkeit Leistungseffizienz zu definieren gibt es nicht, da die benötigte Leistung immer von der vorhandenen Rauschleistung abhängt. Um die Leistungseffizienz verschiedener Modulationen vergleichen zu können verwendet man daher das CNR (Carrier to Noise Ratio), das die Leistung des Trägers C mit der Leistung des Rauschens N vergleicht: $CNR := \frac{C}{N}$. Je kleiner das CNR, desto effizienter die Modulation, wenn sonst alle Eigenschaften (z.B. Bandbreite, BER) gleich sind.

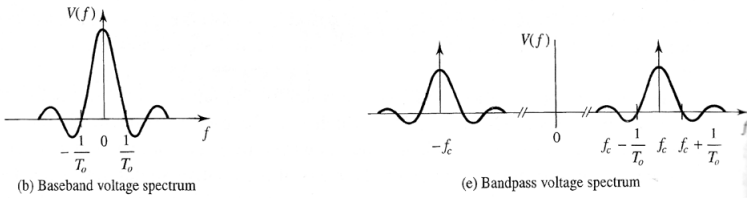
11.2 Binäre IF Modulation

11.2.1 Amplitudenmodulation

Bei binary amplitude shift keying (BASK) werden Null und Eins durch Pulse der Träger-Sinusschwingung mit Carrier-Frequenz f_c mit unterschiedlichen Amplituden A_0 und A_1 repräsentiert. Eine der Amplituden, meistens A_0 wird in der Praxis meistens auf Nullpegel festgelegt. Dies ergibt sogenanntes on-off keying (OOK):

$$f(t) = \begin{cases} A \prod(t/T_0) \cos 2\pi f_c t, & \text{digitale Eins} \\ 0, & \text{digitale Null} \end{cases}$$

Ein OOK-Modulator kann auf mehrere Arten realisiert werden: als ein simpler Schalter, der das Trägersignal ein- und ausschaltet, oder als Multiplikator, der den Träger mit einem unipolaren Basisband-OOK-Signal multpliziert. Das Spektrum des Basisbandsignals wird durch die IF Modulation rund um die Trägerfrequenz f_c reproduziert.



(b) Baseband voltage spectrum (e) Bandpass voltage spectrum

Die Erkennung von OOK-Signalen kann auf zwei Arten erfolgen:

- Beim *kohärenten* Emfpang ist die Phasenlage des empfangenen Signals von Bedeutung. Dazu werden vor dem Sampling Matched Filter oder Korrelatoren eingesetzt, die ja bekanntlich von der Phase abhängen ($sin(f_c t)$ und $cos(f_c t)$ sind unkorreliert, obwohl sie das selbe Signal um $\pi/2$ verschoben darstellen). Verglichen wird jeweils mit dem Trägersignal, also der Sinusschwingung.
- Beim *inkohärenten* Empfang ist die Phasenlage des empfangenen Signals nicht von Bedeutung. Hier wird ein Hüllkurvendetektor gefolgt von centre point Abtastung oder I+D Entscheidung eingesetzt. Alternativ kann ein inkohärenter Detektor auch mittels zweier Korrelationskanäle, einer für Inphase und einer für Quadraturkomponenten realisiert werden (also einmal wird mit dem Sinus, einmal mit dem Cosinus korreliert, beide schwingen natürlich mit

der Trägerfrequenz f_c). Die Ergebnisse der beiden Korrelatoren werden summiert, und man wird so unabhängig von der exakten Phasenlage des Eingangssignals. Die beim inkohärenten Empfang verlorene Information (die Phasenlage) entspricht ca. einem Anstieg des Träger-Rauschabstandes um 1dB... das ist in der Praxis wenig, darum sind auch die inkohärenten Empfänger die weitaus häufiger eingesetzten.

Wir leiten nun die Bitfehlerwahrscheinlickeit für OOK-Modulation her, falls ein kohärenter Empfänger verwendet wird, also Matched Filter bzw. Korrelator. Zum Entscheidungszeitpunkt haben wir bekanntlich den Wert kE_1 Volt für eine 1 bzw. 0 Volt für eine 0 am Ausgang des Korrelators/Matched Filters, wobei E_1 die normalisierte Symbolenergie ist. Auch die Rauschleistung am Ausgang des Korrelators/Matched Filters haben wir bereits im Kapitel über Matched Filter hergeleitet, sie beträgt $\sigma^2 = k^2 E_1 N_0/2$ (im Kapitel über Matched Filter wurde das k mit $k = 1$ angenommen, darum kam k dort nicht vor). Da der Entscheidungsprozess nach der Filterung gleich dem eines binären Basisbandentscheidungsprozesses ist, erhält man die Bitfehlerwahrscheinlichkeit durch Einsetzen in die entsprechende Formel aus dem Kapitel über Bitfehlerwahrscheinlichkeiten, nämlich:

$$P_e = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\Delta V}{2\sigma\sqrt{2}}\right)\right]$$

Dabei ist $\Delta V = k(E_1 - 0)$. Setzt man ΔV und σ ein so bekommt man:

$$P_e = \frac{1}{2}\left[1 - \text{erf}\frac{1}{2}\left(\frac{E_1}{N_0}\right)^{1/2}\right]$$

Drückt man E_1 durch das zeitliche Mittel $\langle E\rangle = (E_1 + E_0)/2$ aus, so erhält man:

$$P_e = \frac{1}{2}\left[1 - \text{erf}\frac{1}{\sqrt{2}}\left(\frac{\langle E\rangle}{N_0}\right)^{1/2}\right]$$

Die Bitfehlerwahrscheinlichkeit kann weiters durch den Träger-Rauschabstand (C/N) ausgedrückt werden, wobei $C = \langle E\rangle/T_0$ und $N = N_0 B$ (B ist die betrachtete Bandbreite unserer Bandpassmodulation), also:

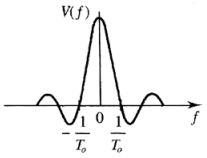
$$\langle E\rangle/N_0 = T_0 BC/N \Rightarrow P_e = \frac{1}{2}\left[1 - \text{erf}\frac{\sqrt{T_0 B}}{\sqrt{2}}\left(\frac{C}{N}\right)^{1/2}\right]$$

11.2.2 Phasenlagenmodulation

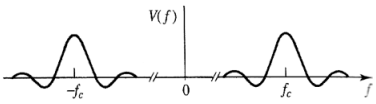
Bei binary phase shift keying (BPSK) wird die Basisbandinformation durch Änderung der Phasenlage des Trägers auf den Träger moduliert:

$$f(t) = \begin{cases} A\prod(t/T_0)\cos(2\pi f_c t), & \text{digitale Eins} \\ A\prod(t/T_0)\cos(2\pi f_c t + \phi), & \text{digitale Null} \end{cases}$$

Sind die Zeiger von Null und Eins antipodal, d.h. um 180° verschieden, so spricht man von phase reverersal keying (PRK), das Spektrum eines PRK Signals ist im nachfolgenden Bild angegeben:



(b) Baseband voltage spectrum of



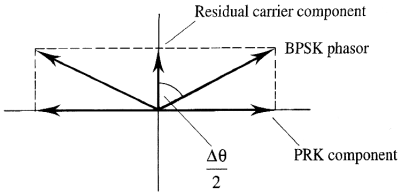
(e) PRK voltage spectrum

Zum Empfang sind natürlich nur kohärente Empfänger geeignet, wobei bei PRK nur ein Matched Filter/Korrelator von Nöten ist, denn bei 180° Phasenverschiebung liefert der Korrelator/Filter kE für eine 1 und $-kE$ für eine 0 (hier ist $\langle E \rangle = E = E_0 = E_1$).

Für die Bitfehlerwahrscheinlichkeit von PRK setzt man wieder in die Formel für Bitfehler ein, diesmal ist $\Delta V = 2kE$ und $\sigma^2 = k^2 E N_0/2$:

$$P_e = \frac{1}{2}\left[1 - \text{erf}\left(\frac{E}{N_0}\right)^{1/2}\right] \qquad \text{bzw.} \qquad P_e = \frac{1}{2}\left[1 - \text{erf}\sqrt{T_0 B}\left(\frac{C}{N}\right)^{1/2}\right]$$

Zur Berechnung der Bitfehlerwahrscheinlichkeit für einen Winkel $\Delta\theta \neq \pi$ zerlegt man die beiden Zeiger jeweils in ihre x-Komponente (PRK-Komponente) und ihre y-Komponente (Residual Carrier Komponente), und betrachtet diesen Fall als speziellen PRK-Fall.



Resolution of BPSK signal into PRK signal plus residual carrier.

Die y-Komponente trägt dann keine Information, da $cos(f_c)$ und $sin(f_c)$ um $\pi/2$ phasenverschoben und daher unkorreliert sind. Je mehr die beiden Zeiger auseinanderliegen, desto näher kommt die Modulationsart PRK und desto höher ist der Anteil am übertragenen Signal, der Energie enthält... diesen Anteil nennt man Modulationsindex m:

$$m = \sin(\Delta\theta/2)$$

Für $\Delta\theta = 0$ haben wir keine Information, für $\Delta\theta = \pi$ haben wir eine PRK und eine Maximumstelle von m als Funktion von $\Delta\theta$. Die Symbolenergie bekommen wir nun durch $m^2 E$, und damit für die Bitfehlerwahrscheinlichkeit:

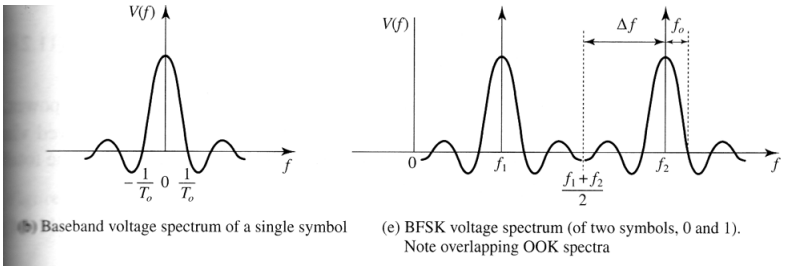
$$P_e = \frac{1}{2}\left[1 - \text{erf}\, m\, \left(\frac{E}{N_0}\right)^{1/2}\right]$$

11.2.3 Frequenzmodulation

Bei binary frequency shift keying (BFSK) werden Nullen und Einsen durch unterschiedliche Pulsfrequenzen des Trägersignals repräsentiert:

$$f(t) = \begin{cases} A\prod(t/T_0)\cos(2\pi f_1 t), & \text{digitale Eins} \\ A\prod(t/T_0)\cos(2\pi f_2 t), & \text{digitale Null} \end{cases}$$

In der Praxis wird BFSK meist mittels eines numerisch gesteuerten Oszillators implementiert, zum Empfang kann sowohl ein kohärenter als auch ein inkohärenter Emfpänger benutzt werden... man braucht für jede der beiden Frequenzen einen eigenen Empfänger (also z.B. zwei Korrelatoren). Die Bandbreite eines BFSK-Signals ergibt sich als $B = f_2 - f_1 + 2f_0$ mit $f_0 = \frac{1}{T_0}$ (siehe Bild).



(b) Baseband voltage spectrum of a single symbol

(e) BFSK voltage spectrum (of two symbols, 0 and 1). Note overlapping OOK spectra

Damit ist die Bandbreite von BFSK um $f_2 - f_1$ größer als die Bandbreite $B = 2f_0$ von BASK und BPSK. Für orthogonale Frequenzen f_1 und f_2, also für Frequenzen für welche

$$\int_0^{T_0} cos(2\pi f_1 t) cos(2\pi f_2 t) dt = 0$$

gilt berechnet man die Bitfehlerwahrscheinlichkeit bei kohärentem Empfang mit 2 Korrelatoren, die voneinander subtrahiert werden (dann hat man wieder $\Delta V = 2kE$), als

$$P_e = \frac{1}{2}\left[1 - \text{erf}\frac{1}{\sqrt{2}}\left(\frac{\langle E\rangle}{N_0}\right)^{1/2}\right]$$

Man beachte dass die Bitfehlerwahrscheinlichkeiten für BASK und BFSK gleich sind, und jene für BPSK besser ist ($\frac{1}{\sqrt{2}} < 1$).

11.3 Trägerrückgewinnung

Bisher haben wir bei der Berechnung von Bitfehlerwahrscheinlichkeiten immer vorausgesetzt, dass es sich um einen kohärenten Empfänger handelt; das ist auch nötig, um eine möglichst kleine Fehlerrate zu erreichen. Dazu wird ein Referenzsignal benötigt, das gegenüber dem Trägersignal nicht phasenverschoben ist. Sieht man sich nochmal die Bilder der Spektren von BASK und BFSK an, so sieht man dass diese Modulationen eine diskrete Linie bei der Trägerfrequenz f_c haben, bei diesen Modulationen kann man das Trägersignal also filtern und als kohärente Referenz verwenden. Wie man am Spektrum des PRK-Signals sieht, fehlt diese Linie hier. Für BPSK mit $\Delta\theta \neq \pi$ existiert diese Linie

wieder, die Trägerfrequenz kann hier also wieder herausgefilter werden, man filtert hier im Prinzip die Residual Carrier Komponente des Signals. Je näher der Modulationsindex bei 1 ist, desto kleiner ist die Residual Carrier Komponente, und desto schwerer wird es das Trägersignal zu filtern.

Für PRK gibt es zwei verschiedene Möglichkeiten der Trägerrückgewinnung:

- Quadrieren des empfangenen Signals. Man erhält dadurch ein Signal mit der doppelten Trägerfrequenz, aus der das kohärente Referenzsignal generiert wird. Durch das Quadrieren ist das Referenzsignal allerdings um 90° phasenverschoben.
- Die Costas Schleife besteht aus zwei PLLs, die um 90° phasenverschoben arbeiten, jeweils eine der beiden PLLs korrigiert die andere und das unterdrückte Trägersignal wird so zurückgewonnen.

Beide Möglichkeiten haben einen Nachteil: eine 180° Phasenzweideutigkeit, d.h. das Referenzsignal kann entweder phasengleich mit dem Träger oder um 180° verschoben sein, was zu einer Invertierung der Symbole führt. Wiederum gibt es zwei Möglichkeiten diese Zweideutigkeit zu beheben:

- Mittels einer bekannten Datensequenz, die den Nutzdaten vorausgesendet wird (Präambel), kann eine Inversion des Signals erkannt werden.
- Differentielle Kodierung vor der Modulation. Dies hat zur Folge, dass z.B. eine Eins durch eine Phasenänderung und eine Null durch das Fehlen einer Phasenänderung repräsentiert wird. So wird die Zweideutigkeit irrelevant. Systeme, die ein Symbol als kohärente Referenz für das nächste Symbol verwenden, werden differentielles PSK genannt.

11.4 Weitere Varianten von PSK

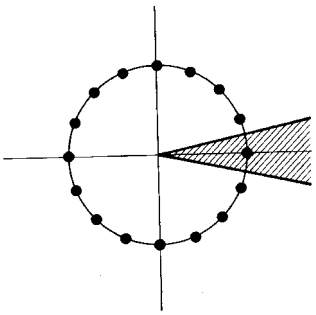
Die spektrale Effizienz wurde im Kapitel über optimale Filterung als

$$\eta_s := \frac{R_s log_2(M)}{B} \quad (bits/s/Hz)$$

definiert, wobei M die Anzahl der Symbole ist, die stochastisch unabhängig sind. Man sieht hier, dass bei fester Bandbreite und Symbolrate $R_s = 1/T_0$ die Anzahl der Symbole die einzige Möglichkeit ist, um die spektrale Effizienz zu erhöhen. Bisher hatten wir immer 2 Symbole, jetzt betrachten wir Modulationen die mehrere Symbole zulassen.

11.4.1 MPSK

MPSK steht für M-symbol PSK, das heißt die Erweiterung von PSK auf $M = 2^n$ Symbole. Das Zeigerdiagramm für 16 Symbole ist in folgender Abbildung ersichtlich, wobei der schattierte Bereich angibt, wie sehr die Phase maximal gestört werden darf, damit das Symbol mit Phase 0 noch richtig erkannt wird. 4-PSK kann als die Superposition zweier PRK-Signale, die um 90° phasenverschobene Träger (also sin und cos als Trägerfrequenz) benutzen, betrachtet werden. Diese Variante wird als QPSK bezeichnet.

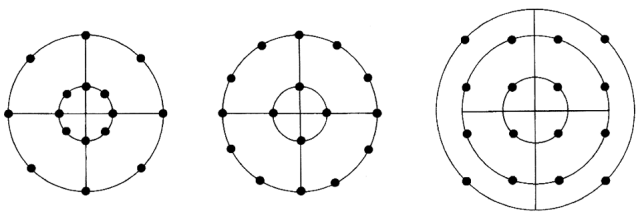


Multisymbolsignalisierung kann als Kodierungsprozess angesehen werden, wobei mehrere binäre Symbole in ein M-äres Symbol kodiert werden. Ein Erkennungsfehler in einem Symbol kann daher in mehrere Fehler in der dekodierten Bitsequenz ausarten. Der wahrscheinlichste Fehler ist eine Missinterpretation eines Zeigerzustands mit seinem Nachbarn. Wird ein Gray Code[15] benutzt um binäre Symbole den Zeigerzuständen zuzuweisen, so hat solch ein Fehler nur einen einfachen dekodierten Bitfehler zur Folge. In jedem Fall muss man aber die Art der Kodierung bei der Berechnung von Fehlerwahrscheinlichkeiten in Betracht ziehen.

Bei differentiellem MPSK (DMPSK) werden binäre Symbole den Phasenunterschieden zwischen benachbarten Symbolen zugewiesen und jedes Symbol wird empfängerseitig durch das vorherige Symbol als kohärente Referenz dekodiert. Der Empfänger muss also nicht mehr die Phase des Trägers kennen, sondern berechnet nur die Phasenänderung zum letzten Symbol. DMPSK hat eine höhere Fehlerwahrscheinlichkeit als MPSK.

11.4.2 APK

Beim Amplitude Phase Keying (APK) wir (M)ASK und MPSK kombiniert. Die Signalkonstellationen sind hier anhand von 3 Beispielen in folgender Abbildung ersichtlich.



(a) 16 APK (8,8) constellation

(b) 16 APK (4,12) constellation

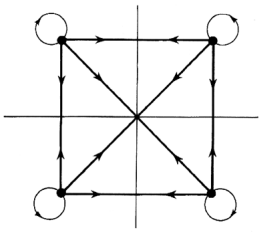
(c) 16 APK (4,8,4) constellation (or 16 QAM)

11.4.3 (O)QPSK

Durch Addition zweier PRK-Signale mit Amplitude $A/\sqrt{2}$ mit Bitrate $R_b/2$, die um 90° phasenverschobene Träger (I und Q Kanal, also Inphase Kanal und Quadraturphase Kanal) benutzen erhält man quadrature phase shift keying (QPSK) mit Amplitude A mit Bitrate R_b. Es handelt sich hierbei um

[15] Code, bei dem sich zwei aufeinanderfolgende Werte nur durch ein Bit unterscheiden

ein Alphabet mit 4 Zuständen, die den Phasenzeigern mit den Phasenverschiebungen $k\pi/2 + \pi/4$ mit $k = 0, 1, 2, 3$ entsprechen.



Quadrature phase shift keying (QPSK) signal constellation showing allowed state transitions.

Dass man dieses Verfahren wirklich als Addition zweier um 90° phasenverschobener PRK-Signale implementieren kann basiert auf der trigonometrischen Identität:

$$cos(u) + cos(v) = 2cos(\frac{u+v}{2})cos(\frac{u-v}{2})$$

aus der man hier erhält:

$$\frac{1}{\sqrt{2}}(cos(\omega + k\pi) + cos(\omega + j\pi + \pi/2)) = \sqrt{2}cos(\frac{2\omega + (j+k)\pi + \pi/2}{2})cos(\frac{(k-j)\pi - \pi/2}{2})$$

Der letzte Faktor ist immer $\pm 1/\sqrt{2}$ und wir bekommen daher ein Signal, das Frequenz ω und Phasenverschiebung $k\pi/2 + \pi/4$ mit $k = 0, 1, 2, 3$ hat.
Hardwaremäßig implementiert man das mit einem Seriell-Parallel Wandler (2 Bit breit), jedes der Bits speist einen von zwei PRK Sendern, die um 90° phasenverschoben mit der halben Bitrate des gesamten QPSK Systems operieren.

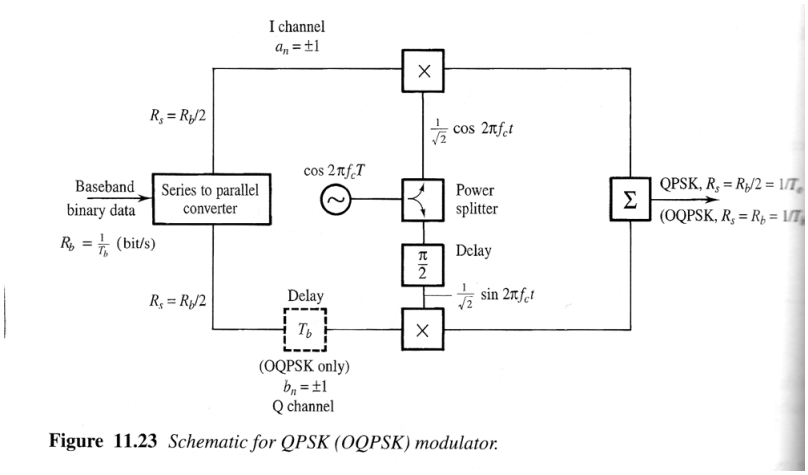


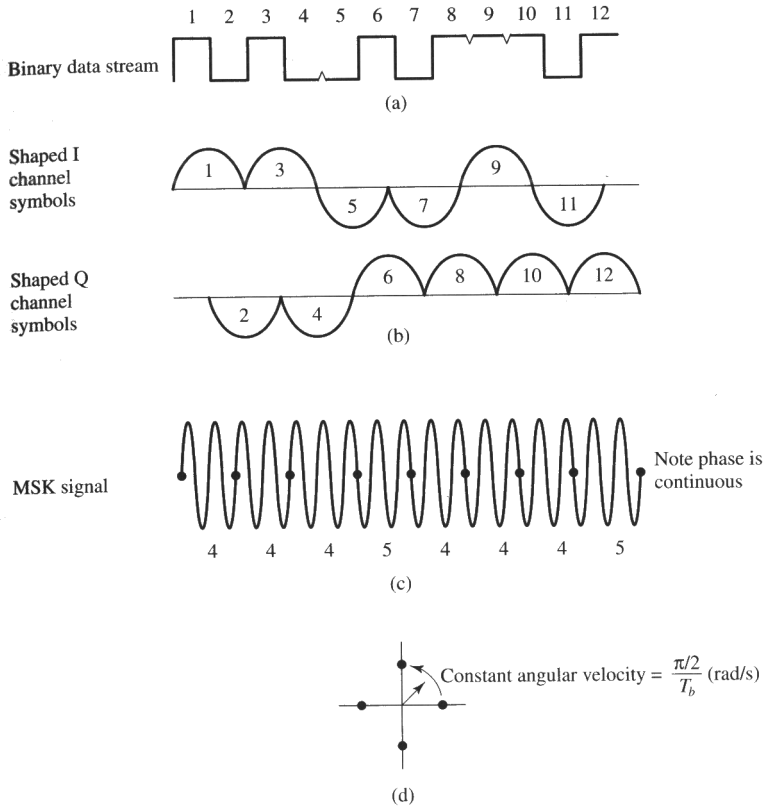
Figure 11.23 *Schematic for QPSK (OQPSK) modulator.*

Die spektrale Effizienz von QPSK ist doppelt so hoch wie bei BPSK, da die Symbole in jedem Quadraturkanal den gleichen Spektralraum und die halbe Spektralbreite (halbe Übertragungsrate!) eines BPSK Signals gleicher Datenrate benötigen. Die prinzipielle Fehlerwahrscheinlichkeit von QPSK ist höher als die von BPSK (die Phasen-Deltas sind kleiner), die Bitfehlerwahrscheinlichkeit ist jedoch gleich groß. Das rührt daher, dass die beiden Kanäle (I und Q) orthogonal sind, man sich die Übertragung also wie 2 getrennte Kanäle vorstellen kann. Da ein QPSK-Symbol nur halb so lang ist und nur die halbe Leistung eines äquivalenten BPSK-Symbol hat ($(1/\sqrt{2})^2 = 1/2$), ist die gesamte Nachrichtenenergie gleich groß.

Offset QPSK (OQPSK) ist identisch zu QPSK bis auf den Umstand, dass vor der Aufmodulierung auf den Träger der Datenstrom des Q-Kanals um eine QPSK-Bitdauer, also eine halbe QPSK-Symboldauer bzw. eine volle PRK-Symboldauer, versetzt wird. Da so nie auf beiden Kanälen gleichzeitig ein Null-Eins-Übergang stattfindet werden Phasensprünge von 180° unterbunden. Ein 180° Sprung wird also in zwei 90° Sprünge aufgeteilt die im Abstand von dieser QPSK-Bitdauer erfolgen. Übergänge treten so häufiger auf; ihre Auswirkungen sind allerdings geringer, die spektrale Effizienz von OQPSK und QPSK sind also gleich. Der Vorteil von OQPSK ist jedoch, dass nichtlineare Effekte (wie sie Filter und Übertragungskanäle usw. haben) durch die kleineren Phasensprünge von 90° statt der 180° bei QPSK die Übertragungsqualität weniger beeinflussen.

11.4.4 (G)MSK

Minimum shift keying (MSK) ist eine modifizierte Form von OQPSK. Bei MSK werden die Eingangspulsströme der I- und Q-Kanäle mit einem Filter umgeformt auf eine Sinusschwingungs-Form (siehe Abbildung).



Die Summe der so geformten Signale folgt nun der Funktion:

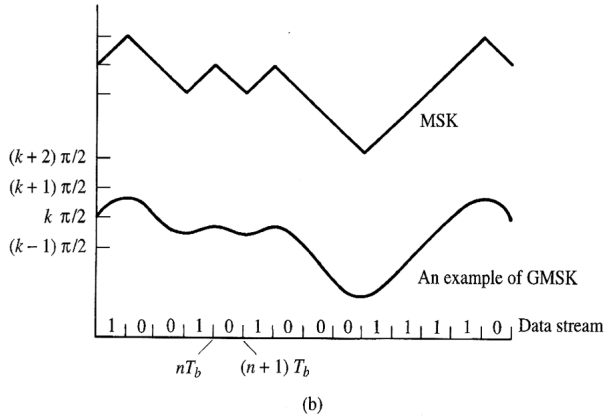
$$f(t) = a_n sin(\frac{2\pi t}{4T_b})cos(2\pi f_c t) + b_n cos(\frac{2\pi t}{4T_b})sin(2\pi f_c t)$$

Beim b_n hat man deswegen einen Cosinus, da MSK auf OQPSK basiert, die Phasenverschiebung gegenüber dem a_n-Puls schlägt sich hier nieder. Untersucht man die Funktion $f(t)$ näher, so bemerkt man, dass für $a_n \neq b_n$ ein Signal mit Frequenz $f_c + 1/(4T_b)$ ausgegeben wird und für $a_n = b_n$ ein Signal mit Frequenz $f_c - 1/(4T_b)$[16]. MSK kann also entweder als BFSK betrachtet werden, oder man veranschaulicht MSK wie folgt: Ein MSK-Symbol wird durch eine Zeigerbewegung repräsentiert; der Zeiger rotiert mit konstanter Winkelgeschwindigkeit vom „Anfangspunkt“ eines Symbols bis zum „Endpunkt“, und zwar im Uhrzeigersinn für $a_n = b_n$ und gegen den Uhrzeigersinn für $a_n \neq b_n$. Dabei muss man sich die Achsen, also das ganze Diagramm, als gegen den UZS drehend vorstellen. Wie auch in OQPSK ist die MSK Symbolperiode gleich der Bitperiode des unmodulierten Datenstroms.

Bei MSK ist zwar die Phase des Signals stetig, nicht aber differenzierbar wenn sich a_n und b_n entsprechend ändern... das ist auch deshalb klar, weil die Ableitung der Phase die Frequenz ist, und diese ja springen kann. Das führt zu einer höheren Bandbreite, die Nebenkeulen klingen aber sehr

[16]Man beachte, dass das MSK Signal in der Abbildung nicht ganz stimmt. Die Anzahl der Nulldurchgänge stimmt zwar, die Kurve stimmt aber beim 3. Vierer und beim 1. Fünfer nicht

schnell ab was vor allem für FDM-Systeme vorteilhaft ist.
Wird der Pulsstrom für den MSK-Modulator nicht in Sinusform gebracht, sondern in Form einer Gaußschen Glockenkurve, dann spricht man von GMSK. GMSK hat stetig differenzierbare Phase, die Frequenz springt also nicht mehr, sondern ändert sich stetig (wenn sie sich denn überhaupt ändert). GMSK hat ein schmaleres Spektrum als MSK und wird z.B. bei GSM angewendet. Durch die Gauß-Formung kommt es zwar zu nicht zu vernachlässigbarer ISI, die Vorteile der spektralen Effizienz des Verfahrens überwiegen jedoch oft.



Es sei noch bemerkt, dass bei (G)MSK die 1:1 Entsprechung von Phase und Symbol verloren geht, dass man diese Entsprechung durch vorherige differentielle Kodierung des binären Datenstroms jedoch wiederherstellen kann.

12 Systemrauschen und Linkbudget

12.1 Thermisches Rauschen

Thermisches Rauschen wird durch freie Ladungsträger, meist Elektronen, und ihrer zufälligen Bewegung erzeugt. Durch die zufälligen Bewegungen treten zufällige Spannungen und Ströme auf, und zwar bei allen Temperaturen über 0K. Man rufe sich in Erinnerung, dass die Temperatur nur ein Maß für die kinetische Energie (Bewegungsenergie) eines Teilchens ist. Thermisches Rauschen kann ein limitierender Faktor für Kommunikationssysteme sein.

Nach dem sogenannten Äquipartitionstheorem ist die kinetische Energie eines Moleküls gleichmäßig auf seine 3 räumlichen Freiheitsgrade (x-, y- und z-Richtung) verteilt, die durchschnittliche thermische Energie pro Freiheitsgrad beträgt $\frac{1}{2}kT$ Joule, wobei k die Boltzmann Konstante und T die Temperatur ist. Man berechnet so die Leistung des thermischen Rauschens als

$$N = kTB \quad (W)$$

pro Bandbreite B, diese Formel nennt man Nyquistformel. Sie ist gültig aus Sicht der klassischen Mechanik, nicht jedoch wenn man quantenmechanische Effekte in Betracht zieht. Das ist auch gut

so, denn würde man auf Basis der Nyquistformel das Leistungsdichtespektrum $G_n(f)$ des thermischen Rauschens berechnen, so käme man auf $G_n(f) = kT$, das Spektrum wäre also nicht frequenzabhängig und das thermische Rauschen hätte unendliche Leistung. Dieses Paradoxon nennt man *Ultraviolettkatastrophe*. Zieht man quantenmechanische Effekte mit in Betracht, so bekommt man das Leistungsdichtepektrum thermischen Rauschens als

$$G_n(f) = \frac{hf}{e^{\frac{hf}{kT}} - 1} \quad (W/Hz)$$

wobei h die Planck-Konstante ist, und berechnet dann die Leistung thermischen Rauschens bei einer Bandbreite B als

$$N = \int_0^B G_n(f) df \quad (W)$$

12.2 Nichtthermisches Rauschen

Schrotrauschen ist eine Form des Rauschens, das immer dann auftritt, wenn ein elektrischer Strom eine Potentialbarriere überwinden muss. Das Schrotrauschen rührt daher, dass sich der Gesamtstromfluss aus der Bewegung einzelner Ladungsträger (Elektronen oder Löcher) zusammensetzt, und jeder Ladungsträger für sich diese Barriere überquert. Dies geschieht nicht gleichmäßig, sondern ist ein statistischer Prozess. In der Summe sind gewisse Schwankungen des Stromflusses zu beobachten.
Die Eigenschaften von Schrotrauschen werden meist für eine Vakuum-Diode hergeleitet, die gewonnenen Erkenntnisse gelten jedoch ebenfalls für die Potentialbarrieren von PN-Dioden und Bipolartransistoren. Bei einer Vakuumdiode werden Elektronen emittiert, sobald sie die Potentialbarriere durchbrechen können. Die Zeiten, wann Elektronen freigesetzt werden, folgen einer Poisson-Verteilung. Jedes emittierte Elektron erzeugt einen rechteckigen Strompuls der Dauer τ auf seinem τ sec dauernden Weg von Kathode zu Anode, sein Spektrum hat daher $sinc^2$ Form. Für Frequenzen $f \ll 1/\tau$ bekommt man im Wesentlichen ein weißes Spektrum mit einer Höhe von q_e^2 wobei q_e die Ladung des Elektrons ist, erst für höhere Frequenzen kann man Schrotrauschen vernachlässigen. Den gesamten durch Schrotrauschen verursachten Strom berechnet man als

$$I_n = \sqrt{2 I_{DC} q_e B} \quad (A)$$

er ist also proportional zum DC-Strom und der betrachteten Bandbreite. Diese Formel gilt auch für eine PN-Diode und für Basis und Kollektor eines Bipolartransistors (mit $I_{DC} = I_B$ bzw. $I_{DC} = I_C$, also Basis- bzw. Kollektorstrom statt DC-Strom).

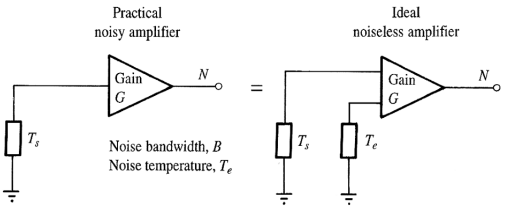
Flickerrauschen, auch $1/f$-Rauschen genannt, tritt auch bei den meisten Komponenten auf, ist aber sehr komponentenabhängig. Das Leistungsdichtespektrum von Flickerrauschen hat ca. $1/f$ Charakteristik für Frequenzen unterhalb einiger Kilohertz, und ein flaches, vernachlässigbar niedriges Spektrum darüber. Flickerrauschen konzentriert sich bei niedrigen Frequenzen und wird oft Rosa Rauschen (pink noise) genannt. Flickerrauschen tritt vor allem bei Feldeffektransistoren (FETs), ganz besonders bei MOSFETs auf.

12.3 Rauschtemperatur

Ein angenehmer Weg um Rauschleistung zu spezifizieren ist die sogenannte Rauschtemperatur. Die Rauschtemperatur einer Komponente ist jene Temperatur, die ein idealer Widerstand haben müsste, um die selbe Rauschleistung wie die Komponente zu haben. Die Rauschtemperatur hat also nichts mit der Komponententemperatur oder der Umgebungstemperatur zu tun. Die Rauschtemperatur T_e berechnet man wie folgt:

$$T_e = \frac{N}{kB} \quad (K)$$

wobei N die thermische Rauschleistung der betrachteten Komponente und B die betrachtete Bandbreite ist. Das Konzept der Rauschtemperatur kann man sich am Beispiel eines Verstärkers veranschaulichen: Anstatt eines rauschenden Widerstands und eines rauschenden Verstärkers betrachtet man zwei rauschende Widerstände und einen idealen Verstärker. Man beachte, dass der Widerstand, der zur Modellierung der Rauschtemperatur hinzugefügt wird, nach wie vor temperaturabhängige Rauschleistung hat, das Konzept der Rauschtemperatur maskiert die Temperaturabhängigkeit des Rauschens also nicht. Meist verwendet man die Rauschtemperatur jedoch um die Rauscheigenschaften verschiedener Komponenten zu vergleichen.



Equivalent noise temperature (T_e) of an amplifier.

Die Gesamtrauschleistung am Ausgang des Transistors berechnet man dann, wie man unmittelbar am Bild sieht, als:

$$N = (kT_sB + kT_eB)G = (N_s + N_e)G$$

wobei das Subskript s für Eigenschaften des rauschenden Widerstands und das Subskript e für Eigenschaften des rauschenden Verstärkers verwendet wird, G ist der Verstärkungsfaktor. Diese Gleichung gilt allgemein für ein System mit Verstärkung G, man beachte jedoch dass G auch < 1 sein kann, so hat eine Leitung zum Beispiel meist eine Dämpfung und keine Verstärkung. Rauscharme Verstärker erreichen Rauschtemperaturen von 40K (=40-273=-233°C) und weniger.

12.3.1 Rauschtemperatur kaskadierter Systeme

Verwendet man die Gleichung aus dem letzten Kapitel für z.B. 2 Systeme mit den Rauschtemperaturen T_{ei} und Verstärkung G_i für $i = 1, 2$, so bekommt man nach Division durch kB

$$T_{casc} = ((T_s + T_{e1})G_1 + T_{e2})G_2$$

Diese Gleichung kann man rekursiv fortsetzen für n Teilsysteme. Manchmal dividiert man alle Verstärkungsfaktoren aus der Gleichung und schreibt die Rauschtemperatur dann als Produkt der Verstärkungen

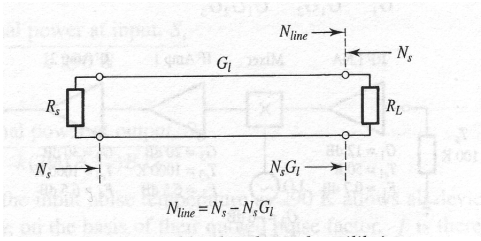
und der „Eingangsrauschtemperatur“:

$$T_{out} = T_{in} \prod_{i=1}^{n} G_i \quad \text{mit} \quad T_{in} = \frac{T_{casc}}{\prod_{i=1}^{n} G_i}$$

Man kann von T_{in} auch noch T_s abziehen, dann hat man mit $T_e = T_{in} - T_s$ eine Rauschtemperatur für das kaskadierte System, kann also Berechnungen in der Form $T_{out} = (T_s + T_e) \prod_{i=1}^{n} G_i$ durchführen.

12.3.2 Rauschtemperatur verlustbehafteter Systeme

Um die Rauschtemperatur verlustbehafteter Systeme zu berechnen betrachtet man eine Übertragungsleitung mit Dämpfung $G_l < 1$ und zwei gleich große Widerstände an der Quelle und am Empfänger ($R_S = R_L$, S = Source, L = Load).



Die thermische Rauschleistung von R_S ist gleich jener von R_L (wir gehen davon aus dass beide Widerstände die selbe physikalische Temperatur T_{ph} haben):

$$N_S = N_L = kT_{ph}B$$

Wenn die Übertragungsleitung eine Dämpfung G_l hat bekommen wir am Widerstand R_L eine Leistungdifferenz, diese muss die Rauschleistung der Übertragungsleitung sein, da in einem geschlossenen System der Energieerhaltungssatz gilt, wir haben also:

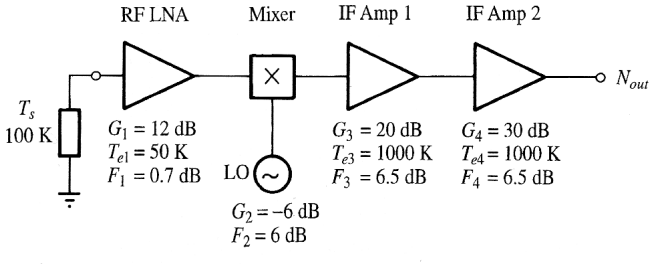
$$N_{line} = N_L - N_s G_l = kT_{ph}B - kT_{ph}BG_l \Rightarrow T_{line} = T_{ph}(1 - G_l)$$

Je näher G_l der Null kommt, desto häher wird die Rauschtemperatur.

12.3.3 Superheterodynempfänger

Der Überlagerungsempfänger (Superheterodynempfänger) ist eine elektrische Schaltung zum Empfang und zur Verarbeitung von hochfrequenten elektromagnetischen Signalen (HF-Signalen). Im Superheterodynempfänger wird das zu empfangende HF-Signal mit dem Signal eines sich im Empfänger befindenden sogenannten Lokaloszillators (LO-Signal) gemischt (multipliziert). Die Frequenz des Lokaloszillators wird je nach gewünschter Empfangsfrequenz eingestellt. Das Ausgangssignal des Mischers (also das Produkt von Eingangssignal und Lokaloszillatorsignal) weist zwei Frequenzbänder auf, nämlich $f_{LO} \pm f_{HF}$, meist wird das niedrigfrequente Signal weiterverwendet, also das hochfrequente gefiltert. Es wird als Zwischenfrequenzsignal (ZF-Signal) (engl. IF = Intermediate Freqeuency)

bezeichnet, da dieses Signal noch weiterverarbeitet werden muss und lediglich einen Zwischenschritt in der Signalverarbeitung darstellt. Folgende Abbildung zeigt eine schematische Darstellung eines Superheterodynempfängers.



LNA steht für Low Noise Amplifier, der LNA verstärkt das RF-Signal. Verstärkung und Rauschtemperatur der einzelnen Komponenten sind in folgender Tabelle angegeben:

Komponente	Verstärkung	T_e
LNA	12 dB	50 K
Mixer	-6 dB	
IF Amp 1	20 dB	1000 K
IF Amp 2	30 dB	1000 K

12.3.4 Rauschfaktor und Rauschzahl

Der Rauschfaktor f einer Komponente ist definiert als das Verhältnis von SNR am Eingang und SNR am Ausgang wenn das Eingangsrauschen einer Temperatur von 290 K entspricht:

$$f = \frac{(S/N)_i}{(S/N)_o} \quad \text{bei } N_i = 290kB$$

Ist SNR_o klein im Vergleich zu SNR_i, so wird f groß, daher auch der Name Rauschfaktor. Diese Spezifikation erlaubt es, verschiedene Komponenten fair auf Basis des Rauschfaktors zu vergleichen. Man kann bei bekannter Rauschtemperatur T_e den Rauschfaktor f berechnen als

$$f = 1 + \frac{T_e}{290}$$

f kann nur dann direkt als Verhältnis von Eingangs- zu Ausgangs-SNR interpretiert werden, wenn die Umgebungstemperatur 290 K ist und die Rauschtemperatur am Eingang der Komponente ebenfalls 290 K ist. Man kann aber wegen

$$T_e = (f-1)290 \quad (K)$$

dennoch genaue Berechnungen durchführen, wenn man nur den Rauschfaktor gegeben hat. Den Rauschfaktor kaskadierter Systeme leitet man über die letzte Formel und die bereits bekannten Formeln für

die Rauschtemperatur kaskadierter Systeme her, und bekommt die sogenannte *Formel von Friis*, hier für 3 Systeme:

$$f = f_1 + \frac{f_2 - 1}{G_1} + \frac{f_3 - 1}{G_1 G_2}$$

Allgemein lautet die Formel von Friis also:

$$f = f_1 + \sum_{i=2}^{n} \frac{f_i - 1}{\prod_{j=1}^{i-1} G_j}$$

Die Rauschzahl F berechnet man aus dem Rauschfaktor durch Umrechnung in eine logarithmische (dB) Skala:

$$F = 10\ log_{10} f \quad (dB)$$

Für verlustbehaftete Komponenten erhält man

$$f = 1/G_l \quad \text{und} \quad F = -G_l \quad (dB)$$

wobei G_l im ersten Fall ein Faktor sein muss und im zweiten Fall bereits in dB ausgedrückt sein muss.

12.4 Linkbudget

Für Übertragungskanäle werden oft, ähnlich wie in der Finanzwelt, eine Art Kosten berechnet. Für Kabelübertragung gehen in diese Rechnung Übertragungsleistung, Kabeldämpfung, Empfängerverstärkung und Rauschzahl ein. Für Funkübertragungen ist die Situation etwas komplizierter, da Signalenergie nicht nur von der Dämpfung, die von der Luft bewirkt wird (die Dämpfung schlägt sich in Erwärmung der Luft nieder), abhängt, sondern man auch beachten muss, dass sich das Signal in alle Richtungen und nicht nur am direkten Weg zur Antenne ausbreitet. Für analoge Repeater bei Funkübertragungen betrachtet man Verstärkung und Rauschzahl, für digitale Repeater die BER. Bevor man das Linkbudget für Funkübertragungen genauer behandeln kann muss man einige Vorbetrachtungen anstellen.

12.4.1 Antennen

Eine isotrope Antenne ist eine Antenne, die in alle Richtungen gleich stark sendet. Sendet sie mit einer Leistung von P_{rad}, so nimmt die Sendeleistungsdichte W im Fernfeld der Antenne ($R \geq 2D^2/\lambda$ wobei D die größte Dimension der Antenne ist, also das Maximum von Länge und Durchmesser, und λ die Wellenlänge) mit R^2 ab:

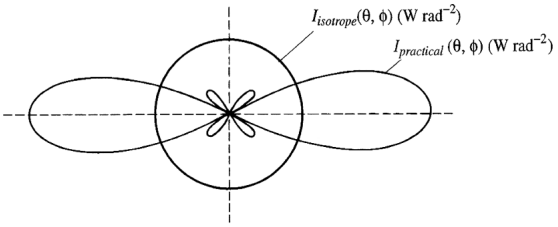
$$W_{isotrop}(R) = \frac{P_{rad}}{4\pi R^2} \quad (W/m^2)$$

Die Strahlungsintensität I einer isotropen Antenne ist in alle Richtungen gleich groß:

$$I_{isotrop} = \frac{P_{rad}}{4\pi}$$

In der Praxis strahlen Antennen jedoch in bevorzugte Richtungen stärker, und die Strahlungsintensität ist eine Funktion der Richtung, in sphärischen Koordiaten gibt man die Strahlungsintensität einer

realen Antenne als $I(\theta, \phi)$ an. Für eine isotrope Antenne ist das Bild der Funktion $I_{isotrop}(\theta, \phi)$ die Oberfläche einer Kugel im $\mathbb{R}^3$. Einen Vergleich zwischen der Strahlungsintensität einer isotropen und einer realen Antenne zeigt nachfolgendes Bild.



Two-dimensional polar plots of antenna radiation intensity for isotropic and practical antenna.

Als Antennengewinn $G(\theta, \phi)$ bezeichnet man das Verhältnis der Strahlungsintensität einer realen Antenne im Vergleich zu einer isotropen Antenne mit der selben Sendeleistung:

$$G(\theta, \phi) = \frac{I(\theta, \phi)}{I_{isotrop}}$$

In der bevorzugten Richtung der Antenne ist der Antennengewinn > 1, sonst nicht. Der Antennengewinn wird oft in dB angegeben.
Beim Senden wird ein Teil der Energie in Verlustwärme umgewandelt, da eine reale Antenne einen ohmschen Widerstand hat. Der Wirkungsgrad einer Antenne wird mit η_Ω bezeichnet und berechnet sich als:

$$\eta_\Omega = \frac{\text{Nutzleistung}}{\text{Nutzleistung + Verlustleistung}}$$

Man definiert nun naheliegend die Sendeleistung P_T einer Antenne als:

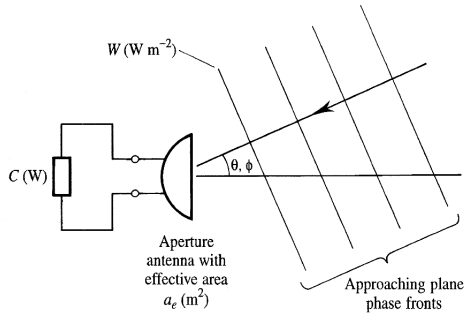
$$P_T = \eta_\Omega P_{rad}$$

Mithilfe des Wirkungsgrades einer Antenne und dem Antennengewinn kann man nun die Leistungsdichte einer realen Antenne im Fernfeld angeben als:

$$W(\theta, \phi, R) = \frac{P_T}{4\pi R^2} G(\theta, \phi)$$

Eine Empfangsantenne entnimmt aus einer ebenen Wellenfront Energie. Kommt die Welle aus der Richtung (θ, ϕ) so ist die Absorptionsfläche (Wirkfläche) a_e einer Antenne definiert als das Verhältnis der entnommenen Leistung $C(\theta, \phi)$ zur Leistungsdichte W der Welle:

$$a_e(\theta, \phi) = \frac{C(\theta, \phi)}{W}$$



Aperture antenna receiving plane wave from θ, ϕ direction.

Man kann nun den Antennengewinn mithilfe der Absorptionsfläche berechnen:

$$G(\theta, \phi) = \frac{4\pi a_e(\theta, \phi)}{\lambda^2}$$

Diese Berechnung beruht auf einer Eigenschaft, die Reziprozität genannt wird, Sende- und Empfangseigenschaften einer Antenne sind demnach gleich, Antennen sind stets reziprok. Werden θ und ϕ nicht angegeben, so ist immer die Richtung mit der größten Leistungsintensität gemeint, man nennt diese Richtung die Peilrichtung.
Die physikalische Fläche A_{ph} einer Parabolantenne und ihre Absorptionsfläche stehen zueinander im Verhältnis. Leider ist die Absorptionsfläche in der Praxis kleiner als die physikalische Fläche, der Faktor η_{ap} in der Gleichung

$$a_e = \eta_\Omega \eta_{ap} A_{ph}$$

heißt Strahlungswiderstand der Antenne.

12.4.2 Empfangene Trägerleistung

Will man die Trägerleistung am Empfänger berechnen (wir gehen hier immer von der Peilrichtung aus), so bekommt man zunächst für die Leistungsdichte der von der Sendeantenne ausgestrahlten Welle:

$$W = \frac{P_T}{4\pi R^2} G_T$$

Der Teil der Welle, der an der Empfängerantenne empfangen wird hängt von der Absorptionsfläche der Empfangsantenne ab, man bekommt daher für den empfangenen Träger

$$C = W\ a_e = \frac{P_T}{4\pi R^2} G_T\ a_e$$

Im vorigen Abschnitt haben wir uns G mithilfe von a_e ausgedrückt, für die Peilrichtung bekommen wir $G_R = \frac{4\pi a_e}{\lambda^2}$ und damit

$$C = \frac{P_T}{4\pi R^2} G_T G_R \frac{\lambda^2}{4\pi} = P_T G_T (\frac{\lambda}{4\pi R})^2 G_R$$

Die Größe $P_T G_T$ wird Effective Isotropic Radiated Power EIRP genannt:

$$EIRP := P_T G_T$$

Die Größe $(\frac{\lambda}{4\pi R})^2$ wird Free Space Path Loss FSPL genannt:

$$FSPL := (\frac{\lambda}{4\pi R})^2$$

FSPL ist auch von der Wellenlänge abhängig. Man kann nun

$$C = EIRP \cdot FSPL \cdot G_R$$

schreiben, bzw in dBW

$$C = EIRP - FSPL + G_R \quad (dBW)$$

mit

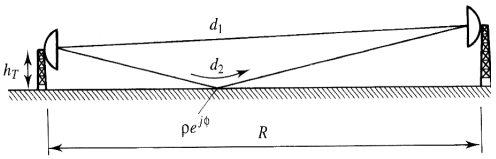
$$EIRP = 10\ log_{10}(P_T) + 10\ log_{10}(G_T) \quad (dBW)$$

und

$$FSPL = 20\ log_{10}(\frac{4\pi R}{\lambda}) \quad (dB)$$

12.4.3 Mehrwegeausbreitung

In der Praxis breitet sich eine Welle nicht nur am direkten Weg von Sender zu Empfänger aus, sondern wird auch am Boden, der Ionosphäre sowie Gebäuden usw. reflektiert. Der Empfänger sieht dann die Summe all dieser Signale, das Signal kann dann stärker sein als wenn man nur das direkte Signal betrachten würde (konstruktive Interferenz) oder schwächer (destruktive Interferenz).
Beim sogenannten Zwei-Pfade-Modell berechnet man die Trägerleistung am Empfänger unter Berücksichtigung des direkten Weges und einer Reflektion am Boden.



Geht man davon aus, dass das Signal ungedämpft reflektiert wird, so bekommt man einen Verstärkungsfaktor $|F|$, der von der Höhe des Senders h_T, der Höhe des Empfängers h_R, der Wellenlänge λ und der Distanz der beiden Antennen R abhängt:

$$|F| = |2sin(\frac{2\pi h_T h_R}{\lambda R})|$$

Um die Leistung des Signals an der Empfangsantenne berechnen zu können muss man diesen Faktor noch quadrieren, man bekommt dann:

$$C = EIRP \cdot FSPL \cdot G_R \cdot |F|^2$$

Man sieht an der letzten Formel, dass das Trägersignal beim Zwei-Pfade-Modell im Vergleich zur direkten Übertragung auf das Vierfache verstärkt sein kann für $\frac{2\pi h_T h_R}{\lambda R} = \frac{\pi}{2}$ (konstruktive Interferenz). Es ist allerdings auch möglich, dass die Trägerleistung fast 0 ist (destruktive Interferenz). In der Praxis berechnet man mit obigen Formeln für festes R und λ meist die optimalen Höhen der Antennen.

12.4.4 Antennentemperatur

Man kann natürlich auch für eine Antenne die Rauschtemperatur bestimmen, die sogenannte Antennentemperatur. Für die Antennentemperatur sind verschiedene Quellen verantwortlich:

- *Atmosphärisches Rauschen*: Für Frequenzen unter 30 MHz sind meistens Blitze in Gewittern hauptverantwortlich für das Rauschen. Die Entladungen werden von der Ionosphäre eingefangen und das Rauschen kann sich so weltweit ausbreiten.
- *Galaktisches Rauschen*: Für Frequenzen zwischen 30 MHz und 1 GHz stammt das Rauschen von sich bewegenden Elektronen in der Galaxis. Es hängt stark von der Ausrichtung der Antenne ab, wie stark der Einfluss dieser Art von Rauschen ist. Das Spektrum galaktischen Rauschens fällt schnell mit steigender Frequenz und ist oberhalb von 1 GHz vernachlässigbar.
- *Atmosphärisches thermisches Rauschen*: Dieses Rauschen ist von 1 GHz bis 10 GHz dominant.

Dreht man eine Antenne richtung Himmel, so nimmt der Einfluss galaktischen Rauschens zu, und der Einfluss atmosphärischen Rauschens ab (in der „Sichtlinie“ der Antenne befindet sich weniger Atmosphäre, bildlich gesprochen).
Die gesamte Antennentemperatur setzt sich nun aus der Rauschtemperatur T_A, bedingt durch die drei Rauschquellen, die eben beschrieben wurden, abgeschwächt durch η_Ω, sowie der Rauschtemperatur der Antenne selbst zusammen. Da die Antenne eine verlustbehaftete Komponente ist bekommen wir für die Antennentemperatur:

$$T_{ant} = \eta_\Omega T_A + T_{ph}(1 - \eta_\Omega)$$

13 Simulation von Kommunikationssystemen

Dieser Abschnitt wurde nur eingefügt, damit die Kapitelnummerierung synchron mit der des Buches ist.

14 Fixpunkt Mikrowellen Kommunikation

Dieser Abschnitt wurde nur eingefügt, damit die Kapitelnummerierung synchron mit der des Buches ist.

15 Mobile Kommunikation

Private mobile radio (PMR) ist ein Überbegriff für Funkübertragungen für Feuerwehr, Rettung, Taxis, und auch Handytelefonie. PMR benützt Teile der VHF und der UHF-Frequenzbänder. Jeder Kanal hat eine Bandbreite von 12,5kHz. Die Probleme bei PMR sind einerseits die eingeschränkte Anzahl an Kanälen - 1000 - und andererseits die Tatsache, dass die mobilen Geräte nur dann gut funktionieren, wenn sie sich nahe der Basisstation befinden.

15.1 Kanaleigenschaften

Der Mobilfunkkanal leidet unter mehreren Problemen bei Verbindungen mit Sichtbehinderung:

- Dopplereffekt durch relative Bewegung zwischen den Terminals
- Langsamer räumlicher Schwund durch topographische Verdeckung auf der Übertragungsstrecke (Baum steht im Weg)
- Schneller räumlicher Schwund durch konstruktive und destruktive Interferenz zwischen gleichen Signalen, die unterschiedliche Übertragungsstrecken benutzt haben
- Zeitlicher Schwund durch die Bewegung des Terminals durch das sich räumlich verändernde Feld
- Frequenzselektiver Schwund bei Breitbandsignalen
- Timingprobleme durch unterschiedliche Übertragungsstrecken
- Zeitliche Veränderung der Kanaleigenschaften durch Bewegung des Terminals

Einige dieser Probleme zeigen unterschiedliche Auswirkungen des gleichen physikalischen Prozesses auf.

15.1.1 Mittlere Empfangsleistung

Durch die unterschiedlichen Arten von Schwund lassen sich nur statistische Aussagen über die empfangene Leistung treffen. Die mittlere Empfangsleistung hängt außer von Größen wie der Senderleistung P_T, Senderantennengewinn G_T und Empfängerantennengewinn G_R auch vom Landnutzungsfaktor L und dem Urbanisierungsfaktor U ab. L leitet sich aus dem Anteil verbauten Gebietes und U aus dem Anteil von mit 4- oder mehr-stöckigen Gebäuden verbauten Gebietes ab. Weiters wird auch noch der Höhenunterschied H zwischen Sender und Empfänger mit einbezogen, sowie die Frequenz f_{MHZ} mit der gesendet wird und der Plain Earth Path Loss PEPL. In die Berechnung von PEPL wiederum geht zusätzlich der Abstand zwischen Sender und Empfänger ein.
Die genaue Formel ist kompliziert, sie lautet:

$$C = P_T + G_T - PEPL + G_R - \beta$$

mit

$$\beta = 20 + \frac{f_{MHZ}}{40} + 0.18L - 0.34H + K$$

K ist hier $(0.094U - 5.9)$ für Städte und 0 sonst. PEPL berechnet man als

$$PEPL = 20\, log_{10}(\frac{R^2}{h_T h_R})$$

wobei h_T die Höhe des Senders und h_R die Höhe des Empfängers ist und R die Distanz von Sender und Empfänger. PEPL ist ein Schätzwert für FSPL und hat daher die selbe Bedeutung wie FSPL.

15.1.2 Langsamer und schneller räumlicher Schwund

Langsamer Schwund, der aufgrund von topographischer Beugung entlang der Übertragungsstrecke auftritt folgt log-normal Statistiken und tritt in ländlichem Gebiet auf. Schneller Schwund, der durch den zeitlich versetzten Empfang mehrerer Signalrepliken, die unterschiedlichen Übertragungsstrecken gefolgt sind, auftritt, folgt Rayleigh Statistiken. Schneller Schwund ist in verbautem Gebiet dominierend. Schneller Schwund hat jeder Autofahrer schon einmal erlebt: Man bleibt an der Ampel stehen und der Radioampfang ist schlecht. Man fährt nur einen halben Meter weiter vor, und plötzlich ist das Radiosignal wieder ganz klar.

15.1.3 Zeitdispersion, frequenzselektiver Schwund, Kohärenzbandbreite und Dopplereffekt

Breitet sich ein Signal über mehrere Übertragungsstrecken aus, so erhält der Empfänger mehrere zeitversetzte Signalkopien. Bei digitaler Signalisierung führt dies zu ISI. Der Grad an ISI hängt von dem Verhältnis von Zeitdispersion (wie stark sind die Signalrepliken zeitlich versetzt) zu Symboldauer ab.

Die Streuung der Übertragungsstrecken-Verzögerung kann im Frequenzbereich durch die *Kohärenzbandbreite* ausgedrückt werden. Die Kohärenzbandbreite ist das maximale Frequenzintervall, für das der Schwund zweier Frequenzen noch korreliert ist. Wenn die Verzögerungen durch Mehrwegeausbreitung bis zu D Sekunden beträgt, dann ist die Kohärenzbandbreite W_c gegeben als

$$W_c \approx \frac{1}{2\pi D}$$

Die Kohärenzbandbreite gibt an, über welchen Bereich ein Kanal ein „flaches“ Spektrum besitzt, d.h. in welchem Bereich zwei Frequenzen eines Signals annährend gleichem Schwund ausgesetzt sind. Liegen zwei Signalkomponenten weiter als die Kohärenzbandbreite auseinander, so schwinden sie unabhängig voneinander und das Gesamtsignal ist *frequenzselektivem Schwund* ausgesetzt.

Der Dopplereffekt spielt ebenfalls eine Rolle bei mobiler Kommunikation. Bekanntlich ergibt sich, wenn ein Empfänger sich mit der Geschwindigkeit v zur (fixen) Basisstation bewegt, beim Empfänger eine durch den Dopplereffekt bedingte Frequenzverschiebung um f_d:

$$f_d = \frac{v}{c} f$$

wobei c die Lichtgeschwindigkeit bezeichnet. Bei 1800 MHz und 120 km/h bekommt man schon eine Verschiebung der Frequenz um 200 Hz.

15.2 Zellulare Kommunikation

Die Kapazität des PMR-Spektrums ist verglichen mit den Anforderungen klein. Darum werden Basisstationen mit bescheidener Sendeleistung eingesetzt, die alle Teilnehmer in einem beschränkten Bereich, der Zelle, versorgen. Benachbarte Zellen setzen unterschiedliche Betriebsfrequenzen ein, so kann die selbe Frequenz in anderen Zellen, die hinreichend weit entfernt sind, wieder eingesetzt werden.

Der protection ratio ist das Verhältnis von Sendersignalstärke der gewünschten Basisstation in der Zelle, in der man sich befindet, zur Sendersignalstärke einer entfernten Zelle, die die selbe Frequenz benutzt.

15.2.1 Zellgrößen

Die Zellgröße hängt von den Anruf-Anforderungen ab. Über die Anzahl der Teilnehmer, die Wahrscheinlichkeit dass diese einen Anruf tätigen sowie die mittlere Dauer ihrer Gespräche kann die Verkehrsintensität berechnet werden (die Wahrscheinlichkeitsverteilung für die Intervall-Längen zwischen Telefonanrufen ist die Erlangverteilung). Der Wiederverwendungsabstand d gibt den Abstand von zwei Zellzentren an, deren Zellen mit gleicher Frequenz arbeiten. Man ist in der Praxis daran interessiert, die Anzahl an Zellen n pro Cluster zu berechnen, um dabei ein gewisses Carrier-to-Interference-Verhältnis C/I zwischen Zellen mit der selben Frequenz nicht zu unterschreiten. In einem Cluster mit n Zellen vom Radius r mit Wiederherstellungsabstand d kann man C/I berechnen als

$$\frac{C}{I} = \frac{1}{n-1}(\frac{d}{r})^4$$

$(n-1)$ im Nenner deswegen, da man C/I am Zellenrand, wo es am kleinsten ist, betrachtet.
Wegen der Beziehung

$$\frac{d}{r} \approx \sqrt{3n}$$

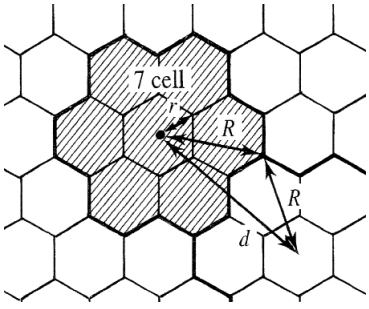
kann die Anzahl n an Zellen pro Cluster also durch Lösen der quadratischen Gleichung

$$\frac{C}{I} = \frac{9n^2}{n-1}$$

berechnet werden. Die Beziehung $\frac{d}{r} \approx \sqrt{3n}$ leitet man wie folgt her:

$$n = \frac{\text{Clusterfläche}}{\text{Zellfläche}} \Rightarrow \frac{R}{r} \approx \frac{\sqrt{\text{Clusterfläche}}}{\sqrt{\text{Zellfläche}}} = \sqrt{n} \Rightarrow R = \sqrt{n}r$$

Nun betrachte man im folgenden Bild das rechtwinklige Dreieck mit Hypothenuse R, wobei eine der Katheten durch $d/2$ gegeben ist:

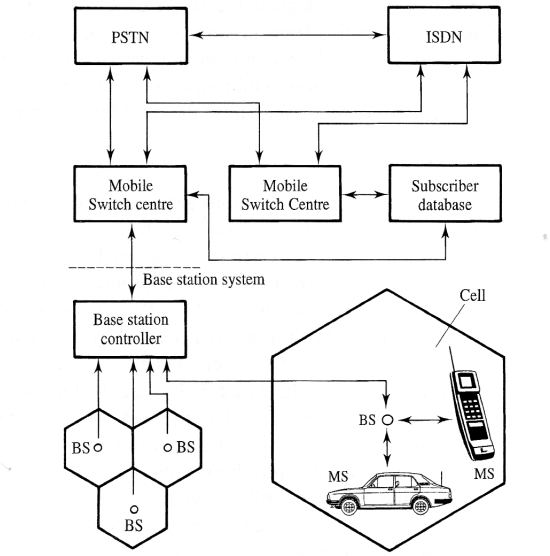


Der Winkel zwischen der Hypothenuse und dieser Kathete ist 30°, wir bekommen daher

$$\frac{d}{2} = R\,cos(30°) \Rightarrow d = \sqrt{3}R \Rightarrow \frac{d}{r} = \sqrt{3n}$$

15.3 Architektur von Mobilfunksystemen

In jeder Zelle eines Mobilfunksystems befindet sich eine *Basisstation* (BS). Mehrere Basisstationen sind über einen *Base Station Controller* an sogenannte *Mobile Switch Centre* angebunden. Die *Mobile Switch Centre* sind mit dem *PSTN* (Public Switched Telephone Network) verbunden, um Festnetzanrufe gleich ins Festnetz einzuspeisen. In einer zentralen *Subscriber Database* wird die Zelle, in der sich jeder mobile Benutzer befindet, gespeichert, um Handover (Wechsel von Zellen) zu ermöglichen.



15.4 Terrestrische Mobilfunksysteme

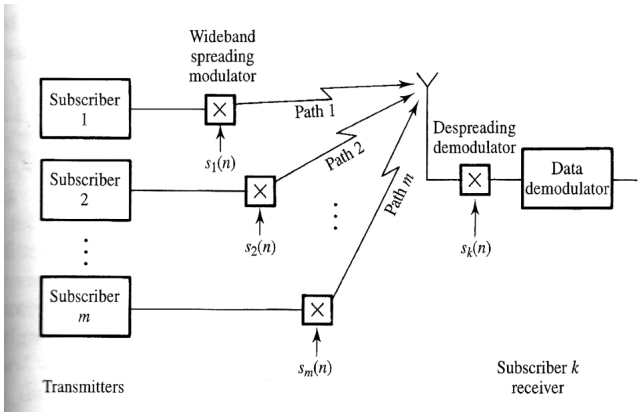
Mobilfunksysteme basieren in der Regel auf TDMA. Bei GSM 900 und DCS 1899 kommunizieren acht TDMA Benutzer über einen Träger und teilen sich eine Basisstation. Das GSM TDMA Format besteht aus acht Zeitschlitzen mit einer Dauer von rund einer halben Millisekunde, die in einen TDMA-Rahmen gepackt werden. Uplink und Downlink haben unterschiedliche Frequenzen. GSM ist relativ komplex und hat hohen Overhead für die Kanalkodierung. Die resultierende Redundanz wirkt sich jedoch positiv auf die Signalqualität aus.

GSM benutzt GMSK und hat eine Datenrate von rund 270kbit/s bei einer Bandbreite von 200kHz. Das Sprachsignal wird mittels eines erweiterten LPC-Vocoders kodiert. Aufgrund der niedrigen Bitrate des Kodierers, der die Sprachqualität stark unter Bitfehlern leiden lässt, wird ein 3bit CRC eingesetzt.

15.5 CDMA

In spread-spectrum Systemen wird der langsame Bitdatenstrom jedes Teilnehmers mit einem Spreizcode $s_k(n)$ hoher Datenrate multipliziert. Ein Datenbit wird in eine dem Sender zugewiesene Bitfolge, den Spreizcode übersetzt; es wird also, was eher untypisch ist, ein Bit in viele Bits übersetzt.

Zur Übertragung des Bitwerts 0 wird der Spreizcode selbst, für den Bitwert 1 der inverse Spreizcode übertragen. Somit füllt das schmalbandige Datensignal nun eine breite Kanalbandbreite aus. Bei *Code Division Multiple Access* wird jedem Teilnehmer ein spezifischer Spreizcode zugewiesen. Die Bits des Spreizcodes nennt man Chips. Die Signale, die sich im Übertragungskanal aufsummieren, haben ein flaches, rauschähnliches Spektrum. Das hat als Vorteil gegenüber FDMA, dass bei Störungen auf einem schmalen Frequenzband zwar alle Kanäle gestört werden, aber alle nur in geringem Ausmaß, während bei FDMA eventuell ein Teilnehmer gar nicht mehr kommunizieren kann.



Der Empfänger filtert das gewünschte Signal mittels Korrelation mit einem lokalen Referenzcode, der dem senderseitigen Spreizcode entspricht. Da die Spreizcodes alle orthogonal sind, gibt der Korrelator 1 aus, wenn das übertragene Bit 1 war, -1 wenn das übertragene Bit 0 war und 0 wenn gar nichts übertragen wurde.

CDMA wird bei UMTS eingesetzt. Hier können jedem Nutzer, je nach Bedarf, verschieden lange Spreizcodes zugewiesen werden. Braucht der Benutzer gerade eine hohe Bandbreite, so bekommt er einen kurzen Code, ansonsten einen langen... der Bandbreitenbedarf kann so also flexibel angepasst werden und ist nicht fix geregelt wie bei TDMA oder FDMA.

16 Übertragung und Speicherung von Videosignalen

Ein Videosignal muss, um ein Fernsehbild darstellen zu können, zwei Grundinformationen übertragen: eine Beschreibung des repräsentierten Bildteils (z.B. Hellligkeit) und den Ort (und Zeit) dieses Bildteils. Das Fernsehbild ist aus einer Vielzahl dieser Bildteile, auch Pixel oder Bildpunkte genannt, aufgebaut.

16.1 Farbdarstellung

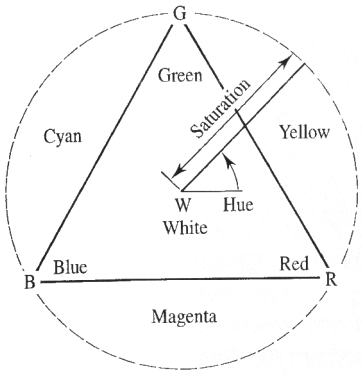
Ein Bildpunkt eines Farbbildes wird meist auf eine der folgenden zwei Arten repräsentiert:

- Ein unabhängiges Intensitätssignal (oder Bildhelligkeits-, sprich Luminanzsignal) und zwei Farbsignale (Chrominanzsignal), nämlich Farbton (im Winkelmaß) und Sättigung.

- Drei Farbsignale, üblicherweise die Intensitätswerte von Rot, Grün und Blau.

Das Farbdreieck (siehe Abbildung) zeigt, wie die unterschiedlichen Farben repräsentiert werden. Der Vorteil der Luma/Chroma-Variante liegt darin, dass die Luminanzkomponente ausreicht, um ein Schwarzweißbild zu repräsentieren, Rückwärtskompatibilität zu Schwarz-Weiß-Fernsehern wird so gewahrt. Die Luminanzkomponente Y berechnet sich als

$$Y = 0.3R + 0.59G + 0.11B$$



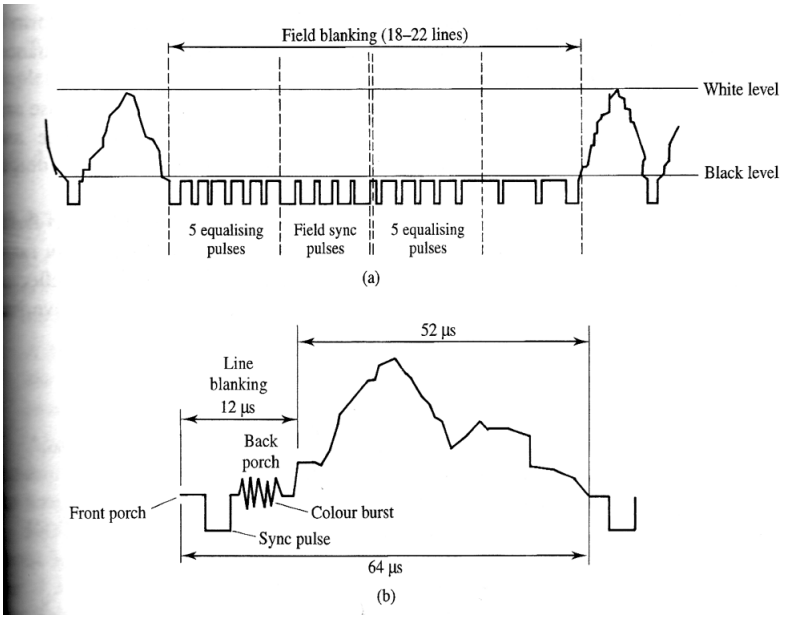
In der Theorie kann man alle Farben durch die Luminanz und 2 der Grundfarben wiederherstellen. Es werden daher meist Y, U und V (YUV) übertragen, wobei U und V differenzielle Farbsignale sind:

$$U = 0.88(R - Y) \qquad V = 0.49(B - Y)$$

16.2 TV Übertragungssysteme

16.2.1 PAL

Der PAL-Standard überträgt Farbsignale. Ein Einzelbild wird dazu in eine Sequenz von 625 Linien aufgeteilt; es werden 25 Einzelbilder pro Sekunde übertragen. Spezielle Pulse geben Anfang einer neuen Zeile bzw. eines neuen Bildes an. Bei PAL wird ein Einzelbild in zwei Halbbilder (sogenannte Fields) unterteilt, wobei ein Halbbild alle geraden und eines alle ungeraden Zeilen umfasst. Die Synchronisation zwischen Halbbildern sieht man am oberen Teil nachfolgender Abbildung. Die Pixelinformation in einer Zeile wird abgetastet und als analoges Signal übertragen (unterer Teil nachfolgender Abbildung). Die Chrominanz-Information wird in einem Teil des Hochfrequenzabschnitts des Luminanz-Signals eingefügt. Dies ist nicht weiter tragisch, da das menschliche Auge Farbinformation schwächer als Helligkeitsinformation auflöst. Durch diese Technik ist es außerdem möglich, auch weiter Schwarz-Weiß-Fernseher zu verwenden (bzw. war es historisch umgekehrt?!). Für die pixelweise Abtastung der Zeilen müssen Takt von Sender und Empfänger synchronisiert werden, um die Chrominanz-Komponente richtig dekodieren zu können müssen außerdem die Phase von Sender und Empfänger synchron sein. Dazu dient das „Colour Burst“ Signal am Anfang einer Zeile.



Von den 625 übertragenen Linien enthalten nur 575 tatsächlich Bildinformation; die restlichen Linien enthalten Halbbild-Synchronisationspulse und Information für Teletext.

16.2.2 andere Standards

NTSC ist PAL ähnlich, hat aber eine unterschiedliche Übertragungsrate: es werden 30 Bilder pro Sekunde übertragen; die dafür aber nur mit 525 Linien. Außerdem hat NTSC keine Phasenverzerrungskorrektur wie PAL und so können bei schwachem NTSC-Empfang starke Farbfehler auftreten.

SECAM ist PAL noch ähnlicher, überträgt aber pro Linie nur eine Chrominanzkomponente.

PAL ist eine Weiterentwicklung von NTSC und auch SECAM überlegen.

16.2.3 HDTV

HDTV ist definiert als ein System, das es erlaubt Bilder in dreifacher Höhe anzuzeigen, um eine Qualität wie am Originalschauplatz zu erzielen. Um dies zu ermöglichen sind zumindest doppelt so hohe Auflösungen wie bei konventionellem TV vonnöten. Weiters haben die Bilder ein Seitenverhältnis von 16:9 anstatt 4:3 (Stichwort Breitbild), das dem menschlichen Blickfeld besser angepasst ist. HDTV kann analog oder digital übertragen werden, letzteres ist meistens der Fall. HDTV benötigt eine Bandbreite von ca. 12 MHz (PAL: 6 MHz). HDTV braucht, je nach Auflösung, zumindest 1000 Zeilen. Meistens wird ein Vielfaches der bereits vorhandenen Zeilenzahlen (z.B. $2 \cdot 625$ bei PAL) verwendet.
Man beachte, dass HDTV zum Zeitpunkt des Buchdrucks noch eine sehr neue Entwicklung war.

16.3 JPEG

JPEG ist ein internationaler Standard zur Kompriminierung und Dekomprimierung von Bildern. Es besteht aus mehreren Techniken, das Kernsystem verwendet Diskrete Cosinus-Transformation (DCT) um Blöcke von 8x8 Pixeln abzutasten, es werden 64 Koeffizienten pro Block erzeugt. Diese Koeffizienten werden dann quantisiert anhand einer vom User vorgegebenen Quantisierungsmatrix, die eine Quantisierungsgenauigkeit für Koeffizienten angibt. Diese Matrix wird im Header des JPEGs gespeichert. Da das Auge auf grobe Strukturen sensibler reagiert als auf Feinheiten, werden meistens kleine örtliche Frequenzen feiner aufgelöst. Die Koeffizienten werden differentiell zum letzten übertragenen Block kodiert, da sie sich von Block zu Block meistens nur wenig ändern. Als letzter Schritt werden die quantisierten Koeffizienten nach ihrer Auftrittswahrscheinlichkeit mit dem Huffman Code kodiert (Entropiekodierung).

16.4 MPEG-1 und MPEG-2

MPEG ist ein Standard zur Komprimierung und Dekomprimierung von Videos. Die Grundblöcke von MPEG-Kodierern sind

- Bewegungskompensation
- DCT
- Variable Length Coding

MPEG ist für Videos, also für Bildersequenzen, gedacht, und kann sich daher zunutze machen, dass sich von einem Frame zum nächsten oft nur wenig ändert. MPEG verwendet 3 Arten von Frames, die hintereinander in 3 Arten angeordnet werden können (für Details siehe www):

I-Frame Sogenannte Intra-Frames werden unabhängig von allen anderen kodiert und erlauben so ein gewisses Maß an Random-Access, man kann also im Video herumspringen und muss nicht immer alles von Anfang bis Ende sehen.

P-Frame Predictive Frames sagen Bewegungen voraus (der Bewegungsvektor wird geschätzt), und kodieren nur die Art der Bewegung, nicht jedoch das Bild selbst. Das Bild selbst muss in einem vorherigen Frame gespeichert sein.

B-Frame Bidirectional Predictive Frames haben die stärkste Komprimierung, brauchen aber 2 Referenzframes, ein Frame in der Vergangenheit und eines in der Zukunft.

16.5 MPEG-4

MPEG-4 hat noch stärkere Komprimierung als MPEG-1 und MPEG-2, und ist außerdem fehlertoleranter bei Netzwerkübertragungsfehlern.

Bei MPEG-4 untersucht man das Video auf Objekte, sogenannte Video Objects, und verwendet für jedes Objekt (z.B. Person im Vordergrund, Auto im Vordergrund, Hintergrund) einen eigenen Layer zur Kodierung. Auch hier wird wieder die Bewegung der Objekte vorhergesagt. Die Aufteilung in verschiedene Layer ermöglicht es, bei einem Tennismatch z.B. den mehr oder weniger statischen Hintergrund nur einmal zu übertragen anstatt immer wieder. Der Kodierungsprozess selbst ist ähnlich dem von MPEG-1 bzw. MPEG-2, es wird ebenfalls die DCT verwendet.

16.6 Digital Audio Broadcast

Das Digital Audio Broadcasting (DAB) ist ein digitaler Übertragungsstandard für terrestrischen Empfang von Hörfunkprogrammen. Es ist für den Frequenzbereich von 30 MHz bis 3 GHz geeignet und schließt somit auch eine Verbreitung über Kabel und Satellit ein. Die Audiodaten der Programme werden bei DAB zunächst mittels MUSICAM (MP2) mit Datenraten von 32 bis 256 kbit/s codiert. Die Bitrate liegt zwar deutlich unter der einer Audio-CD, durch spezielle Codec-Verfahren wird aber eine der Audio-CD vergleichbare Qualität erreicht, solange die Bitrate nicht zu gering gewählt wird (vgl. mp3). Für die Übertragung werden mehrere Audiodatenströme zusammen mit ebenfalls möglichen reinen Datendiensten zu einem sogenannten Ensemble mit hoher Datenrate zusammengeführt. Der so entstandene Multiplex wird mittels Coded Orthogonal Frequency Division Multiplex (COFDM) moduliert. Dieses Verfahren ist im Vergleich zur analogen Ausstrahlung deutlich robuster gegenüber Störungen. Zudem ist es möglich, große geographische Bereiche mit nur einer Frequenz abzudecken.

16.6.1 Orthogonal Frequency Division Multiplex

Jeder einzelne Träger ist phasen- und/oder amplitudenmoduliert und kann daher die Information von mehreren Bits (typischerweise 2 bis 6 Bit) pro Symbol tragen. Dieses Modulationsverfahren nutzt alle drei freien Parameter Frequenz, Amplitude und Phase für die Übertragung der Information. Die Symboldauer ist bei OFDM gegenüber Einträgerverfahren sehr viel länger, da die Daten gleichzeitig statt nacheinander übertragen werden. Die längere Symboldauer bringt auch Vorteile insbesondere beim Mehrwegempfang (d.h. bei Echos). Bis zu einer bestimmten, durch das Schutzintervall festgelegten Zeitspanne der Laufzeitdifferenzen der verschiedenen Signalwege verschlechtern Echos den Empfang nicht. Die Bedeutung des Schutzintervalles besteht darin, dass der Funkkanal im Empfänger erst dann ausgewertet wird, wenn alle Einschwingvorgänge abgeklungen sind. Je größer das Schutzintervall, desto länger darf das Echo sein. OFDM-Signale werden mit komplex rechnenden inversen diskreten Fouriertransformationen (IDFT) erzeugt. Die IDFT setzt voraus, dass alle Subträgerfrequenzen orthogonal zueinander stehen, daher auch der Name Orthogonal Frequency Division Multiplex (OFDM).

16.6.2 Coded Orthogonal Frequency Division Multiplex

Coded Orthogonal Frequency Division Multiplex (COFDM) ist ein digitales Modulationsverfahren, welches das Modulationsverfahren OFDM um eine Vorwärts-Fehlerkorrektur und ein Guard Intervall ergänzt. COFDM bietet eine hohe Stabilität gegen Mehrwegempfang, Burst-Fehler und frequenzselektive Auslöschungen (Fading) und eignet sich auch für den mobilen Empfang damit übertragener Signale.