

Überblick zu Kapitel 4 des Buches „Modulationsverfahren“

Das vierte Kapitel behandelt *die digitalen Modulationsverfahren* ASK, PSK und FSK sowie einige davon abgeleitete Modifikationen. Die meisten der in den Kapiteln 2 und 3 genannten Eigenschaften der analogen Modulationsverfahren gelten weiterhin. Unterschiede ergeben sich aus der nun erforderlichen Entscheiderkomponente des Empfängers.

Wir beschränken uns hier im wesentlichen auf die systemtheoretischen und übertragungstechnischen Aspekte. Die Fehlerwahrscheinlichkeit wird zwar für ideale Bedingungen angegeben. Die Herleitungen und die Berücksichtigung nichtidealer Randbedingungen – Phasenversatz bei kohärenter Demodulation, nichtkohärente Demodulation, Einfluss von Impulsinterferenzen und vieles mehr – finden Sie im Buch „Digitalsignalübertragung“. In den nachfolgenden Unterkapiteln werden behandelt:

- die *Pulsmodulation* und deren Komponenten Abtastung – Quantisierung – Codierung,
- die *linearen Modulationsverfahren* ASK, BPSK und DPSK sowie zugehörige Demodulatoren,
- die *Quadraturamplitudenmodulation* sowie kompliziertere Signalraumzuordnungen,
- *FSK – Frequency Shift Keying* als Beispiel einer nichtlinearen digitalen Modulation,
- die FSK mit *kontinuierlicher Phasenanpassung*, insbesondere das (G)MSK–Verfahren.

Die theoretischen Grundlagen werden auf 52 Bildschirmseiten mit 82 Grafiken, vielen Gleichungen und Beispielen verdeutlicht. Das Kapitel 15 beinhaltet auch Aufgaben und zehn Zusatzaufgaben mit insgesamt 124 Teilaufgaben. Dazu gibt es noch zwei Lernvideos (LV) und sechs Interaktionsmodule (IM):

Gesetze der Fouriertransformation (LV zu den Grundlagen, 2–teilig, Dauer 5:57 – 5:55)

Pulsmodulation (LV zu Kapitel 4.1, 3–teilig, Dauer 11:35 – 12:53 – 22:15)

Abtastung periodischer Signale und Signalrekonstruktion (IM zu Kapitel 4.1)

Komplementäre Gaußsche Fehlerfunktionen (IM zu Kapitel 4.2)

Frequenzgang und zugehörige Impulsantwort (IM zu Kapitel 4.3)

Prinzip der Quadratur–Amplitudenmodulation (IM zu Kapitel 4.3: Zusammenfassende Darstellung)

Quaternary Phase Shift Keying (IM zu Kapitel 4.3: QPSK und O–QPSK)

Continuous Phase Modulation (IM zu Kapitel 4.4: FSK mit kontinuierlicher Phase)

Literatur: [BC02] – [Joh92] – [Kam04] – [Klo01] – [KöZ08] – [Lin05] – [Mäu85] – [Pro01] – [Söd01]

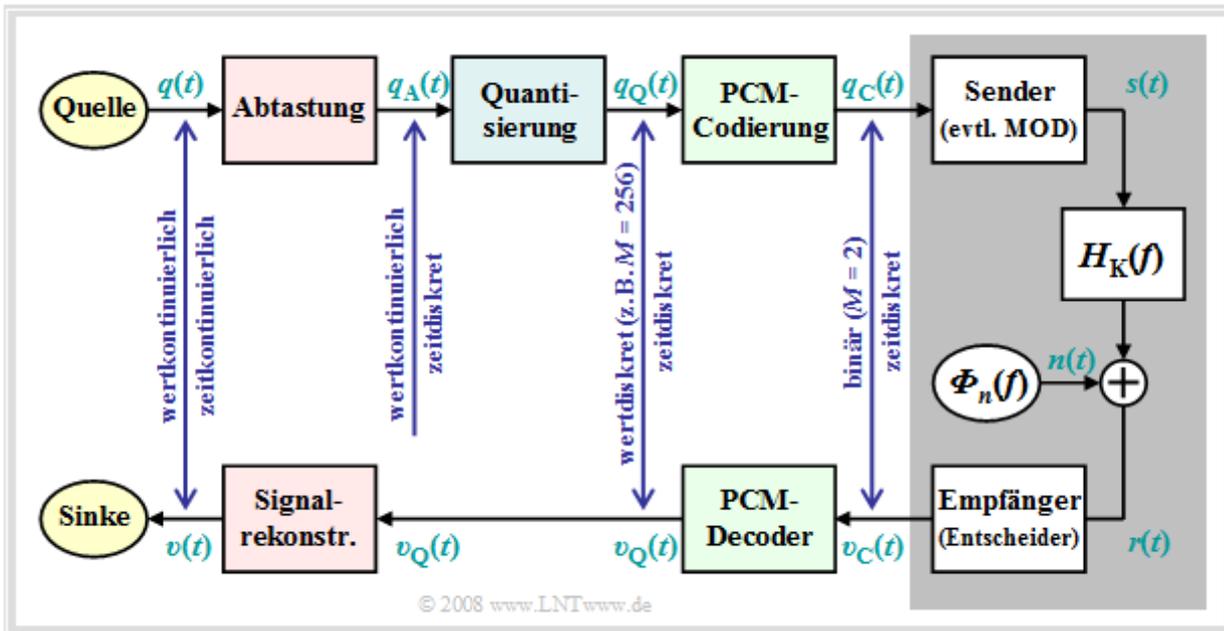
Zur Verdeutlichung der Thematik empfehlen wir die Grafik- und Simulationsprogramme *DMV* und *PSK*. Diese sind Teil des Lehrsoftwarepakets *LNTsim* und werden seit Jahren im Praktikum *Simulation digitaler Übertragungssysteme* am Lehrstuhl für Nachrichtentechnik der TU München eingesetzt:

Download der Programme von *LNTsim* (ZIP–Version)

Download der Anleitung „Digitale Modulationsverfahren“ (PDF–Datei)

Prinzip und Blockschaltbild (1)

Nahezu alle heute eingesetzten Modulationsverfahren arbeiten digital. Deren Vorteile wurden schon im **Kapitel 1.1** aufgeführt. Das erste Konzept zur digitalen Signalübertragung wurde bereits 1938 von **Alec Reeves** entwickelt und wird seit den 1960er Jahren unter dem Namen **Pulscodemodulation** (PCM) auch in der Praxis eingesetzt. Auch wenn sich viele der in den letzten Jahren konzipierten digitalen Modulationsverfahren von der PCM im Detail unterscheiden, so eignet sich diese doch sehr gut, um das Prinzip all dieser Verfahren zu erklären.

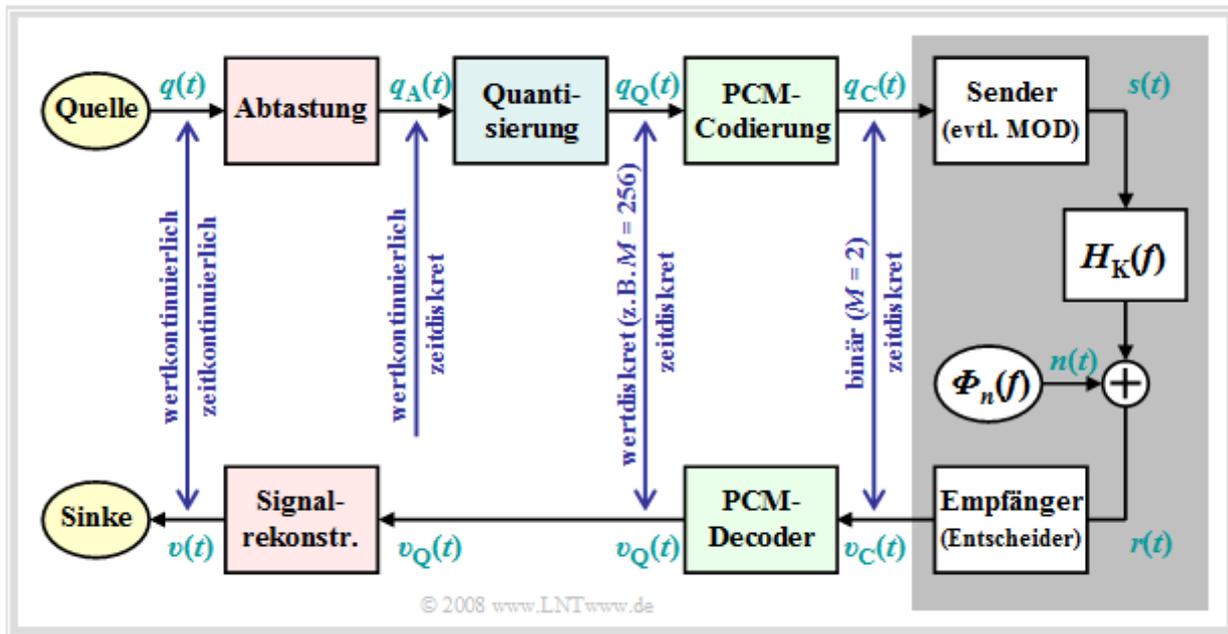


Die Aufgabe des PCM-Systems ist es,

- das analoge Quellsignal $q(t)$ in das Binärsignal $q_C(t)$ umzusetzen – diesen Vorgang bezeichnet man auch als **A/D-Wandlung**,
- dieses Signal über den Kanal zu übertragen, wobei das empfängerseitige Signal $v_C(t)$ wegen des Entscheiders ebenfalls binär ist,
- schließlich aus dem Binärsignal $v_C(t)$ das analoge, wert- und zeitkontinuierliche Sinkensignal $v(t)$ zu rekonstruieren \Rightarrow **D/A-Wandlung**.

Die Bildbeschreibung wird auf der nächsten Seite fortgesetzt.

Prinzip und Blockschaltbild (2)



Weiterhin ist zum obigen PCM-Blockschaltbild anzumerken:

- Der PCM-Sender (bzw. der A/D-Wandler) setzt sich aus den drei Funktionsblöcken **Abtastung** – **Quantisierung** – **PCM-Codierung** zusammen, die auf den nächsten Seiten noch im Detail beschrieben werden.
- Der grau hinterlegte Block zeigt das **digitale Übertragungssystem** mit digitalem Sender und Empfänger (letzterer beinhaltet auch einen Entscheider), sowie dem analogen Übertragungskanal, gekennzeichnet durch den Frequenzgang $H_K(f)$ und die Rauschleistungsdichte $\Phi_n(f)$.
- Dieser Block wird in den ersten drei Kapiteln des Buches „Digitalsignalübertragung“ eingehend behandelt. Im **Kapitel 5** des gleichen Buches finden Sie auch **digitale Kanalmodelle**, die das Übertragungsverhalten anhand der Binärsignale $q_C(t)$ und $v_C(t)$ phänomenologisch beschreiben.
- Weiter erkennt man aus dem Blockschaltbild, dass es für die Quantisierung empfängerseitig keine Entsprechung gibt. Deshalb wird sich auch bei fehlerfreier Übertragung, also für $v_C(t) = q_C(t)$, das analoge Sinkensignal $v(t)$ vom Quellensignal $q(t)$ unterscheiden.
- Als Maß für die Qualität des (digitalen) Übertragungssystems verwenden wir das **Sinken-SNR** als der Quotient der Leistungen von Nutzsinal $q(t)$ und Fehlersignal $\varepsilon(t) = v(t) - q(t)$:

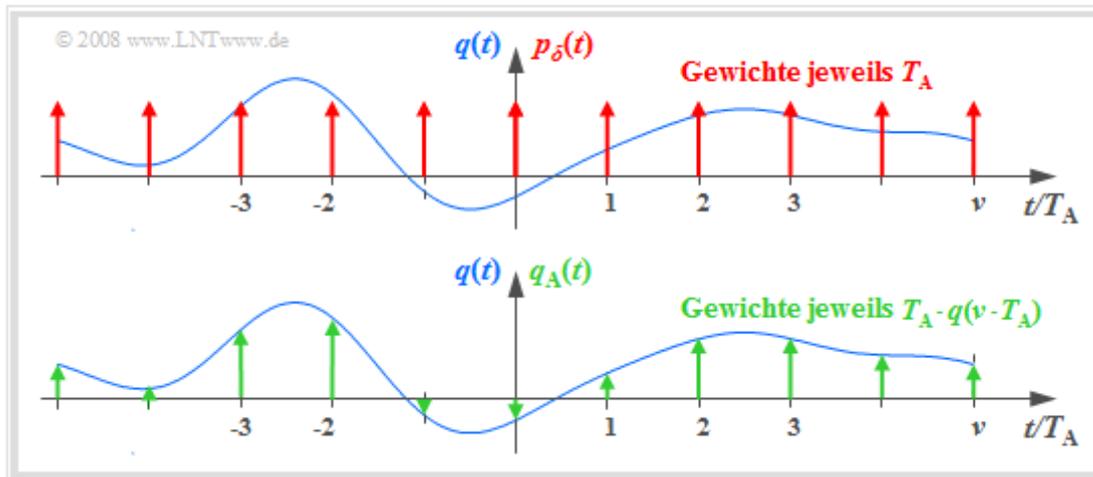
$$\rho_v = \frac{P_q}{P_\varepsilon} \quad \text{mit} \quad P_q = \overline{q(t)^2}, \quad P_\varepsilon = \overline{[v(t) - q(t)]^2}.$$

- Hierbei ist ideale Amplitudenanpassung vorausgesetzt, so dass im Idealfall (das heißt: Abtastung gemäß dem Abtasttheorem, bestmögliche Signalrekonstruktion, unendlich feine Quantisierung) das Sinkensignal $v(t)$ mit dem Quellensignal $q(t)$ exakt übereinstimmen würde.

Wir möchten Sie bereits hier auf das 3-teilige Lernvideo **Pulscodemodulation** (Gesamtdauer 46:45) hinweisen, dass alle Aspekte der PCM beinhaltet. Das Prinzip wird im ersten Teil ausführlich erläutert.

Abtastung und Signalrekonstruktion (1)

Die Abtastung – also die Zeitdiskretisierung des Analogsignals $q(t)$ – wurde im **Kapitel 5.1** des Buches „Signaldarstellung“ ausführlich behandelt. Hier folgt eine Kurzzusammenfassung dieses Abschnitts.



Die Grafik verdeutlicht die Abtastung im Zeitbereich. Das (blaue) Signal $q(t)$ ist zeitkontinuierlich und das im Abstand T_A abgetastete (grüne) Signal $q_A(t)$ zeitdiskret. Dabei gilt:

- Die Abtastung lässt sich durch die Multiplikation des Analogsignals $q(t)$ mit dem **Diracpuls** $p_\delta(t)$ darstellen, der sich auf den Zeitbereich bezieht:

$$q_A(t) = q(t) \cdot p_\delta(t) \quad \text{mit} \quad p_\delta(t) = \sum_{\nu=-\infty}^{\infty} T_A \cdot \delta(t - \nu \cdot T_A).$$

- Das Gewicht der Diracfunktion bei $t = \nu \cdot T_A$ ist gleich $T_A \cdot q(\nu \cdot T_A)$. Da die **Diracfunktion** $\delta(t)$ die Einheit 1/s aufweist, hat somit $q_A(t)$ die gleiche Einheit wie $q(t)$, zum Beispiel „V“.
- Die Fouriertransformierte des Diracpulses ist ebenfalls ein Diracpuls (im Frequenzbereich), wobei der Abstand der einzelnen Diraclinien $f_A = 1/T_A$ beträgt. Alle Impulsgewichte von $P_\delta(f)$ sind 1:

$$p_\delta(t) = \sum_{\nu=-\infty}^{+\infty} T_A \cdot \delta(t - \nu \cdot T_A) \quad \circ \bullet \quad P_\delta(f) = \sum_{\mu=-\infty}^{+\infty} \delta(f - \mu \cdot f_A).$$

- Das Spektrum $Q_A(f)$ des abgetasteten Signals ergibt sich aus dem **Faltungssatz**, wobei $Q(f)$ das kontinuierliche Spektrum des Analogsignals $q(t)$ bezeichnet:

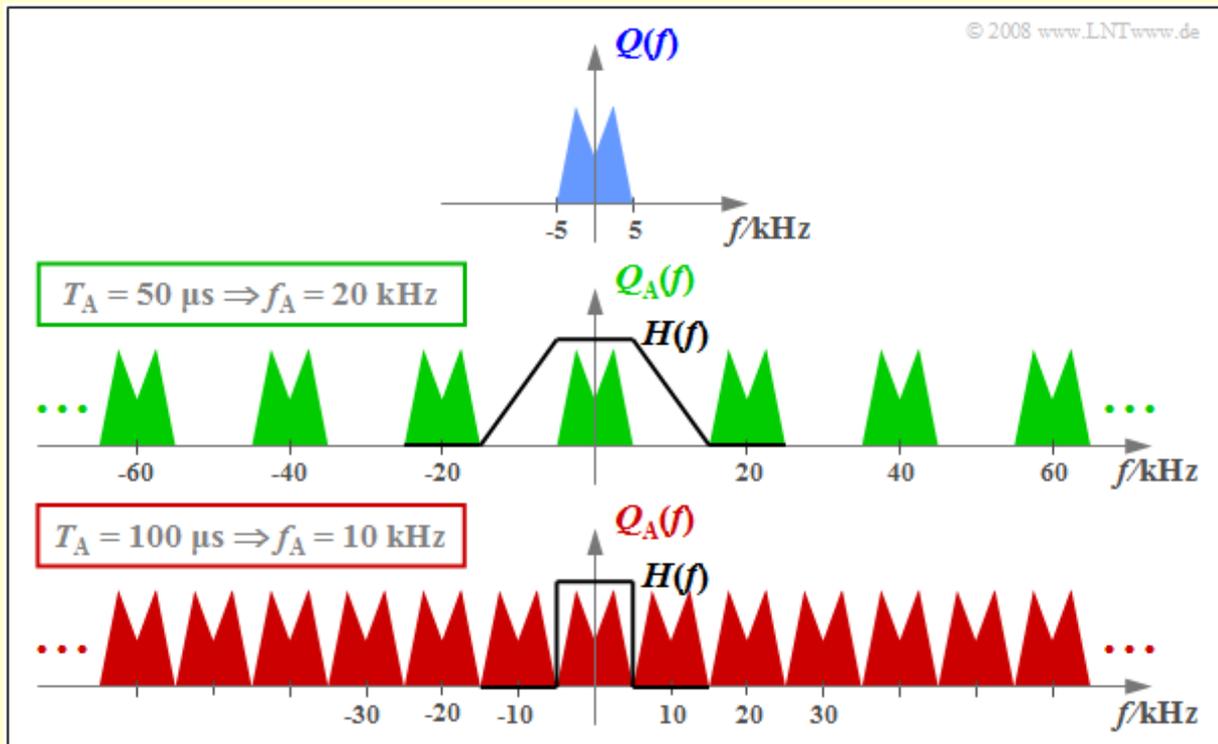
$$Q_A(f) = Q(f) \star P_\delta(f) = \sum_{\mu=-\infty}^{+\infty} Q(f - \mu \cdot f_A).$$

Diese Gleichungen werden auf der nächsten Seite durch ein Beispiel verdeutlicht.

Wir weisen Sie hier auf den zweiten Teil des Lernvideos **Pulscodemodulation** (Dauer 12:53) hin, das die Abtastung und die Signalrekonstruktion systemtheoretisch erklärt.

Abtastung und Signalrekonstruktion (2)

Beispiel: Die obere Grafik zeigt schematisch das Spektrum $Q(f)$ eines analogen Quellensignals $q(t)$, das Frequenzen bis $f_{N, \max} = 5$ kHz beinhaltet. Tastet man das Signal mit der Abtastrate $f_A = 20$ kHz (also im jeweiligen Abstand $T_A = 50 \mu\text{s}$) ab, so erhält man das grün skizzierte periodische Spektrum $Q_A(f)$. Da die Diracfunktionen unendlich schmal sind, beinhaltet $q_A(t)$ auch beliebig hochfrequente Anteile und dementsprechend ist $Q_A(f)$ bis ins Unendliche ausgedehnt (mittlere Grafik). Darunter (rot) gezeichnet ist das Spektrum $Q_A(f)$ für die Abtastparameter $T_A = 100 \mu\text{s} \Rightarrow f_A = 10$ kHz.



Aus diesem Beispiel lassen sich folgende wichtige Erkenntnisse bezüglich der Abtastung gewinnen:

- Beinhaltet $Q(f)$ Frequenzen bis $f_{N, \max}$, so muss die Abtastrate $f_A \geq 2 \cdot f_{N, \max}$ gewählt werden \Rightarrow **Abtasttheorem**. Bei kleinerer Abtastrate f_A (also größerem Abtastabstand T_A) kommt es zu Überlappungen der periodifizierten Spektren und damit zu irreversiblen Verzerrungen.
- Gilt exakt $f_A = 2 \cdot f_{N, \max}$ wie in der unteren Grafik des obigen Beispiels, so kann $Q(f)$ aus $Q_A(f)$ – bzw. im **PCM-System** $V(f)$ aus $V_Q(f)$ – durch einen idealen rechteckförmigen Tiefpass $H(f)$ mit der Grenzfrequenz $f_G = f_A/2$ vollständig rekonstruiert werden.
- Erfolgt dagegen die Abtastung mit $f_A > 2 \cdot f_{N, \max}$ wie in der mittleren Grafik des Beispiels, so kann empfängerseitig zur Signalrekonstruktion auch ein Tiefpass $H(f)$ mit kleinerer Flankensteilheit verwendet werden, solange die folgende Bedingung erfüllt ist:

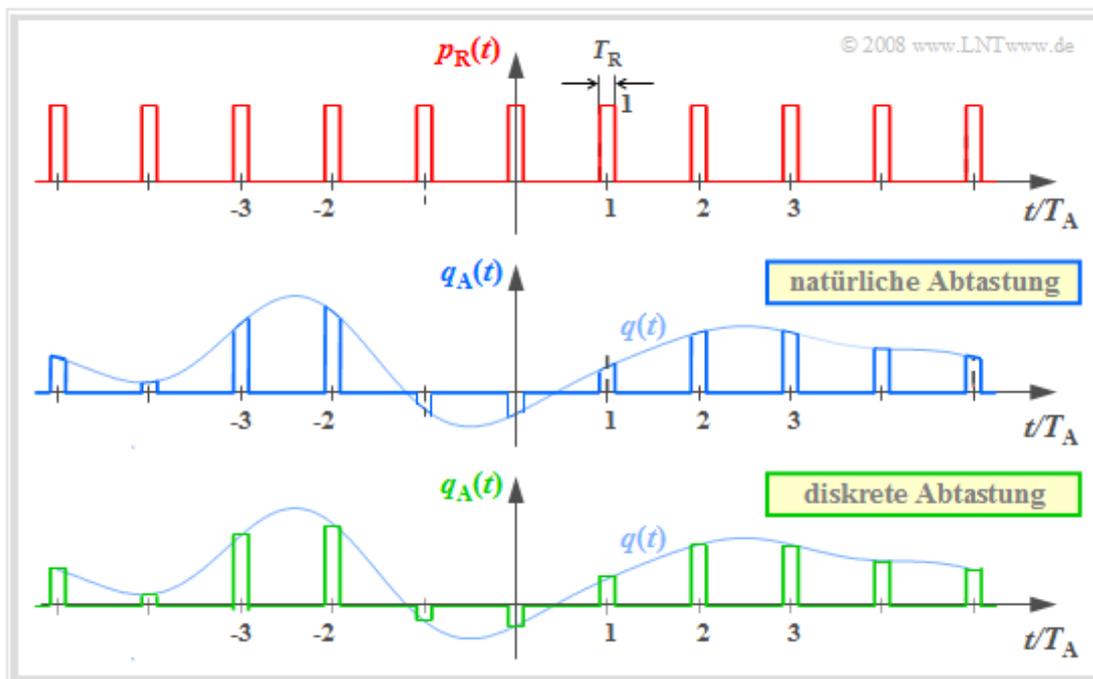
$$H(f) = \begin{cases} 1 & \text{für } |f| \leq f_{N, \max}, \\ 0 & \text{für } |f| \geq f_A - f_{N, \max}. \end{cases}$$

Natürliche und diskrete Abtastung (1)

Die Multiplikation mit dem Diracpuls liefert nur eine idealisierte Beschreibung der Abtastung, da eine Diracfunktion (Dauer $T_R \rightarrow 0$, Höhe $1/T_R \rightarrow \infty$) nicht realisierbar ist. In der Praxis muss der Diracpuls $p_\delta(t)$ zum Beispiel durch einen **Rechteckpuls**

$$p_R(t) = \sum_{\nu=-\infty}^{+\infty} g_R(t - \nu \cdot T_A) \quad \text{mit} \quad g_R(t) = \begin{cases} 1 & \text{für } |t| < T_R/2, \\ 1/2 & \text{für } |t| = T_R/2, \\ 0 & \text{für } |t| > T_R/2 \end{cases}$$

ersetzt werden, wobei die Rechteckimpulsdauer T_R deutlich kleiner als der Abtastabstand T_A sein sollte.



Die Grafik zeigt oben den Rechteckpuls $p_R(t)$. Darunter sind zwei verschiedene Abtastverfahren mit diesem Rechteckpuls dargestellt:

- Bei der **natürlichen Abtastung** ergibt sich das abgetastete Signal $q_A(t)$ durch die Multiplikation von $q(t)$ mit $p_R(t)$. In den Bereichen $p_R(t) = 1$ hat somit $q_A(t)$ den gleichen Verlauf wie $q(t)$.
- Dagegen wird bei der **diskreten Abtastung** das analoge Signal $q(t)$ – zumindest gedanklich – zuerst mit dem Diracpuls $p_\delta(t)$ multipliziert und danach wird jeder Diracimpuls $T_A \cdot \delta(t - \nu \cdot T_A)$ durch einen Rechteckimpuls $g_R(t - \nu \cdot T_A)$ ersetzt.

Hier und bei der auf der nächsten Seite folgenden Frequenzbereichsbetrachtung ist zur Vereinfachung eine akasale Beschreibungsform gewählt. Für eine (kausale) Realisierung müsste $g_R(t) = 1$ im Bereich von 0 bis T_R gelten, und nicht wie hier für $-T_R/2 < t < T_R/2$.

Natürliche und diskrete Abtastung (2)

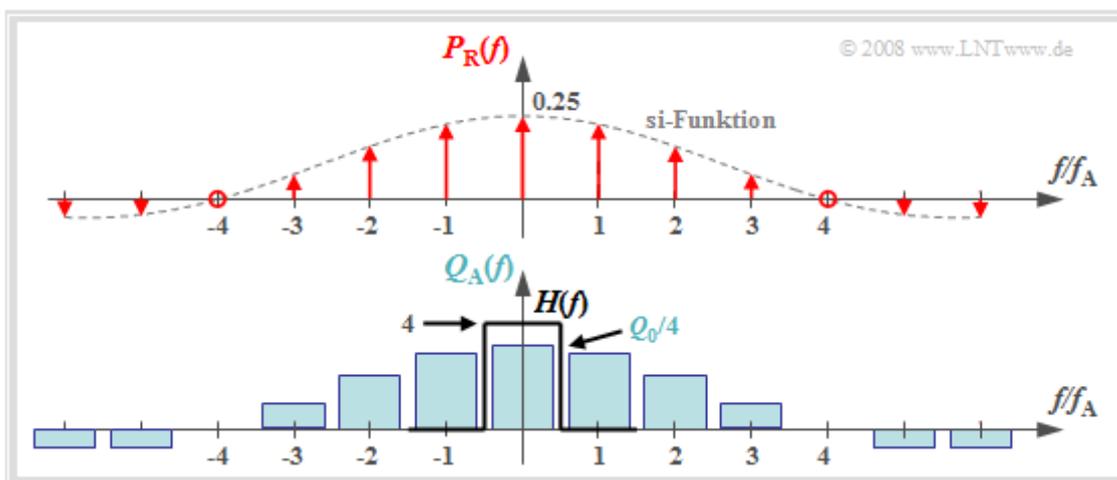
Die natürliche Abtastung lässt sich mit dem **Faltungssatz** im Spektralbereich wie folgt darstellen:

$$q_A(t) = p_R(t) \cdot q(t) = \left[\frac{1}{T_A} \cdot p_\delta(t) \star g_R(t) \right] \cdot q(t)$$

$$\Rightarrow Q_A(f) = \left[P_\delta(f) \cdot \frac{1}{T_A} \cdot G_R(f) \right] \star Q(f) = P_R(f) \star Q(f).$$

Die Grafik zeigt das Ergebnis für

- ein rechteckförmiges Spektrum $Q(f) = Q_0$, das auf den Bereich $|f| \leq 4$ kHz begrenzt ist,
- die Abtastrate $f_A = 10$ kHz $\Rightarrow T_A = 100$ μ s sowie
- die Rechteckimpulsdauer $T_R = 25$ μ s $\Rightarrow T_R/T_A = 0.25$.



Man erkennt aus dieser Darstellung:

- Das Spektrum $P_R(f)$ ist im Gegensatz zu $P_\delta(f)$ kein Diracpuls (alle Gewichte gleich 1), sondern die Impulsgewichte sind hier mit der Funktion $G_R(f)/T_A = T_R/T_A \cdot \text{si}(\pi f T_R)$ bewertet. Auf Grund der Nullstelle der si-Funktion verschwinden die Diraclinien bei $\pm 4f_A$ vollständig.
- Das Spektrum $Q_A(f)$ ergibt sich aus der Faltung mit $Q(f)$. Das Rechteckspektrum um $f = 0$ hat die Höhe $T_R/T_A \cdot Q_0$, die Anteile um $\mu \cdot f_A$ ($\mu \neq 0$) sind weniger hoch.
- Verwendet man zur Signalrekonstruktion einen idealen, rechteckförmigen Tiefpass

$$H(f) = \begin{cases} T_A/T_R = 4 & \text{für } |f| < f_A/2, \\ 0 & \text{für } |f| > f_A/2, \end{cases}$$

so gilt für das Ausgangsspektrum $V(f) = Q(f)$ und dementsprechend ist auch $v(t) = q(t)$.

Dieses Ergebnis kann wie folgt zusammengefasst werden:

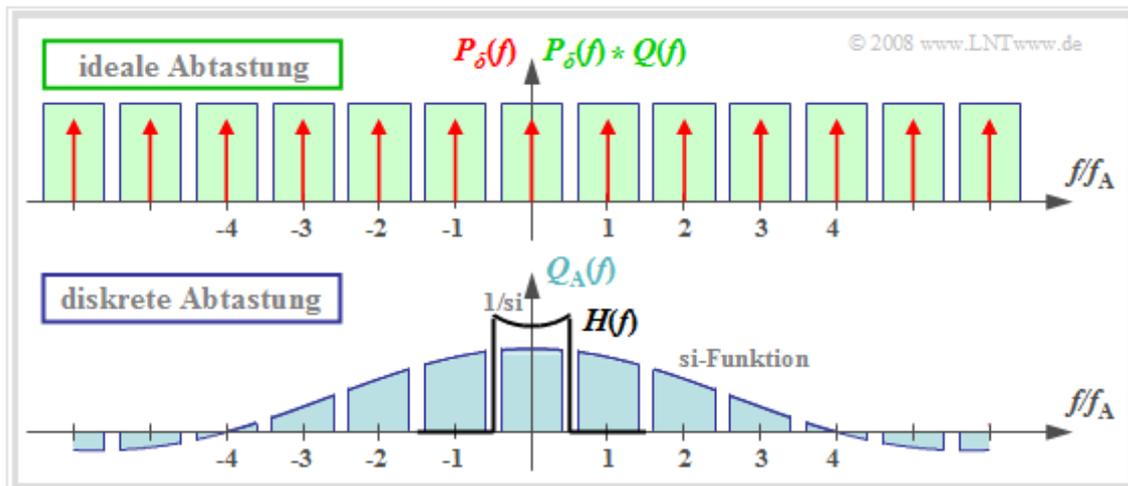
- Bei natürlicher Abtastung kann zur Signalrekonstruktion wie bei der idealen Abtastung (mit einem Diracpuls) ein idealer rechteckförmiger Tiefpass verwendet werden.
- Allerdings muss zur Amplitudenanpassung im Durchlassbereich eine Verstärkung um den Faktor T_A/T_R berücksichtigt werden.

Natürliche und diskrete Abtastung (3)

Bei der diskreten Abtastung erfolgt – zumindest gedanklich – zunächst die Multiplikation des Diracpulses $p_\delta(t)$ mit dem Quellsignal $q(t)$ und erst danach die Faltung mit dem Rechteckimpuls $g_R(t)$:

$$q_A(t) = \left[\frac{1}{T_A} \cdot p_\delta(t) \cdot q(t) \right] \star g_R(t) \Rightarrow Q_A(f) = [P_\delta(f) \star Q(f)] \cdot G_R(f) / T_A.$$

Es ist unerheblich, aber durchaus zweckmäßig, dass hier der Faktor $1/T_A$ zur Bewertungsfunktion $G_R(f)$ hinzugefügt wurde. Damit gilt wieder $G_R(f)/T_A = T_R/T_A \cdot \text{si}(\pi f T_R)$.



Die obere Grafik zeigt die Spektralfunktion $P_\delta(f) \star Q(f)$ nach idealer Abtastung. Bei diskreter Abtastung mit einem Rechteckpuls ergibt sich dagegen das Spektrum $Q_A(f)$ entsprechend dem unteren Diagramm. Man erkennt:

- Jedes der unendlich vielen Teilspektren hat nun eine andere Form. Wichtig ist allerdings nur das Spektrum mit der Mitte bei der Frequenz $f = 0$, da alle anderen Spektralanteile empfängerseitig durch den Tiefpass der Signalrekonstruktion entfernt werden.
- Verwendet man für diesen Tiefpass wieder ein Rechteckfilter mit der Verstärkung um T_A/T_R im Durchlassbereich, so erhält man für das Ausgangsspektrum:

$$V(f) = Q(f) \cdot \text{si}(\pi f T_R).$$

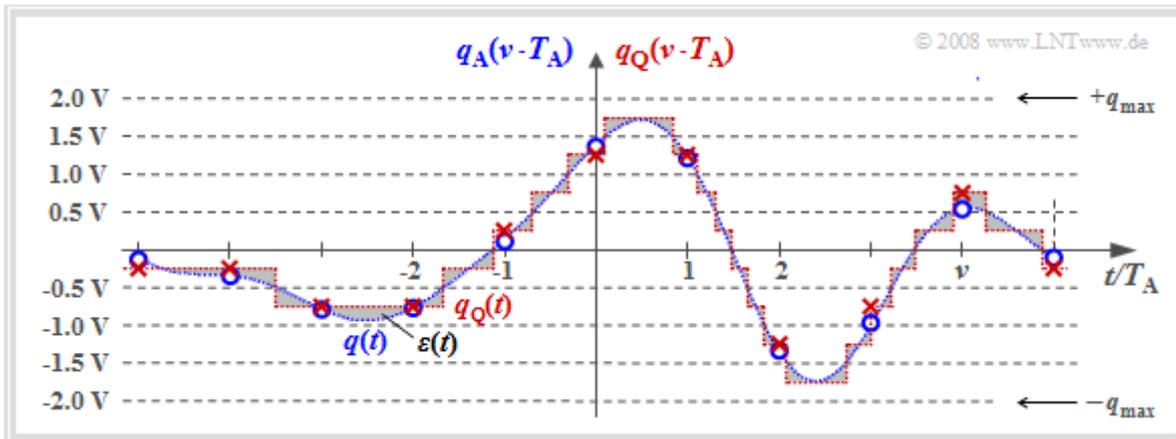
- Das bedeutet: Bei diskreter Abtastung und Rechteckfilterung kommt es zu Dämpfungsverzerrungen entsprechend der Bewertungsfunktion $\text{si}(\pi f T_R)$. Diese sind um so stärker, je größer T_R ist. Nur im Grenzfall $T_R \rightarrow 0$ gilt $\text{si}(\pi f T_R) = 1$.
- Allerdings können durch eine ideale Entzerrung diese linearen Dämpfungsverzerrungen vollständig kompensiert werden. Mit

$$H(f) = \begin{cases} (T_A/T_R) / \text{si}(\pi f T_R) & \text{für } |f| < f_A/2, \\ 0 & \text{für } |f| > f_A/2 \end{cases}$$

erhält man $V(f) = Q(f)$ bzw. $v(t) = q(t)$.

Quantisierung und Quantisierungsrauschen

Die zweite Funktionseinheit **Quantisierung** des PCM-Senders dient der Wertediskretisierung. Hierzu wird der gesamte Wertebereich des analogen Quellensignals (zum Beispiel der Bereich $\pm q_{\max}$) in M Intervalle aufgeteilt und jedem Abtastwert $q_A(v \cdot T_A)$ wird anschließend ein Repräsentant $q_Q(v \cdot T_A)$ des zugehörigen Intervalls (beispielsweise die Intervallmitte) zugewiesen.



Die Grafik verdeutlicht die Quantisierung am Beispiel der Quantisierungsstufenzahl $M = 8$. Tatsächlich wird für M in der Praxis wegen der anschließenden Binärcodierung stets eine Zweierpotenz gewählt. Jeder der durch Kreise markierten Abtastwerte $q_A(v \cdot T_A)$ wird durch den dazugehörigen quantisierten Wert $q_Q(v \cdot T_A)$ ersetzt. Die quantisierten Werte sind als Kreuze eingetragen.

Dieser Vorgang der Wertdiskretisierung ist allerdings mit einer irreversiblen Verfälschung verbunden. Die Verfälschung $\varepsilon_v = q_Q(v \cdot T_A) - q_A(v \cdot T_A)$ hängt dabei von der Quantisierungsstufenzahl M ab. Es gilt:

$$|\varepsilon_v| < 1/2 \cdot 2/M \cdot q_{\max} = q_{\max}/M.$$

Man bezeichnet den quadratischen Mittelwert der Fehlergröße ε_v als **Quantisierungsrauschleistung**:

$$P_Q = \frac{1}{2N + 1} \cdot \sum_{\nu=-N}^{+N} \varepsilon_\nu^2 \approx \frac{1}{N \cdot T_A} \cdot \int_0^{N \cdot T_A} \varepsilon(t)^2 dt \quad \text{mit} \quad \varepsilon(t) = q_Q(t) - q(t).$$

Zur Berechnung von P_Q wird meist die angegebene Näherung der „Spontanquantisierung“ verwendet. Wie oben skizziert lässt man dazu die Abtastung außer Betracht und bildet das Fehlersignal aus den beiden zeitkontinuierlichen Signalen $q_Q(t)$ und $q(t)$.

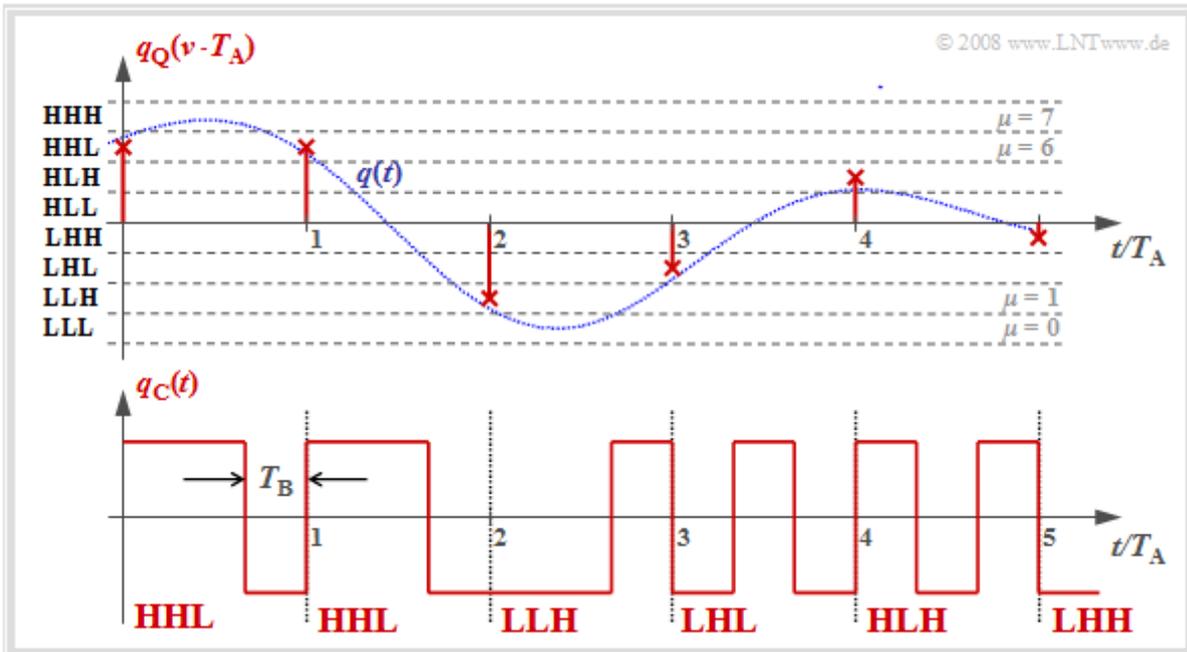
Die Quantisierungsrauschleistung hängt auch vom Quellensignal $q(t)$ ab. Unter der Voraussetzung, dass $q(t)$ alle Werte zwischen $\pm q_{\max}$ mit gleicher Wahrscheinlichkeit annimmt und der Quantisierer genau für diesen Bereich ausgelegt ist, ergibt sich (siehe Aufgabe A4.4):

$$P_Q = \frac{q_{\max}^2}{3 \cdot M^2}.$$

Hinweis: Bei einem Sprach- oder Musiksignal können – wenn auch nur sehr selten – beliebig große Amplitudenwerte auftreten. In diesem Fall wird für q_{\max} meist derjenige Amplitudenwert herangezogen, der nur zu 1% aller Zeiten (betragsmäßig) überschritten wird.

PCM-Codierung und -Decodierung

Der Block **PCM-Codierung** dient der Umsetzung der zeitdiskreten (nach Abtastung) und wertdiskreten (nach Quantisierung mit M Stufen) Signalwerte $q_Q(v \cdot T_A)$ in eine Folge von $N = \text{ld}(M)$ Binärwerte. Hierbei steht „ld“ für den Logarithmus zur Basis 2 $\Rightarrow \log_2 \Rightarrow \text{Logarithmus dualis}$.



In der Grafik ist jeder Binärwert (jedes Bit) durch ein Rechteck der Dauer $T_B = T_A/N$ dargestellt, woraus sich das Signal $q_C(t)$ ergibt. Man erkennt:

- Es wird hier der Dualcode verwendet. Das bedeutet, dass die Quantisierungsintervalle μ von 0 bis $M-1$ durchnummeriert und anschließend in einfacher Binärform geschrieben werden. Mit $M = 8$ gilt beispielsweise $\mu = 6 \Leftrightarrow 110$.
- Die drei Binärsymbole des codierten Signals $q_C(t)$ ergeben sich, wenn man **0** durch **L** („Low“) und **1** durch **H** („High“) ersetzt. Im Beispiel erhält man so: **HHL HHL LLH LHL HLH LHH**.
- Die Bitdauer T_B ist hier um den Faktor $N = \text{ld}(M) = 3$ kürzer als der Abtastabstand $T_A = 1/f_A$, so dass sich die Bitrate zu $R_B = \text{ld}(M) \cdot f_A$ ergibt.
- Verwendet man bei der Decodierung ($v_C \Rightarrow v_Q$) die gleiche Zuordnung wie bei der Codierung ($q_Q \Rightarrow q_C$), so gilt $v_Q(v \cdot T_A) = q_Q(v \cdot T_A)$, falls es zu keinen Übertragungsfehlern kommt.
- Eine Alternative zum Dualcode ist der Graycode, bei dem sich benachbarte Binärwerte nur in einem Bit unterscheiden, zum Beispiel für $N = 3$:

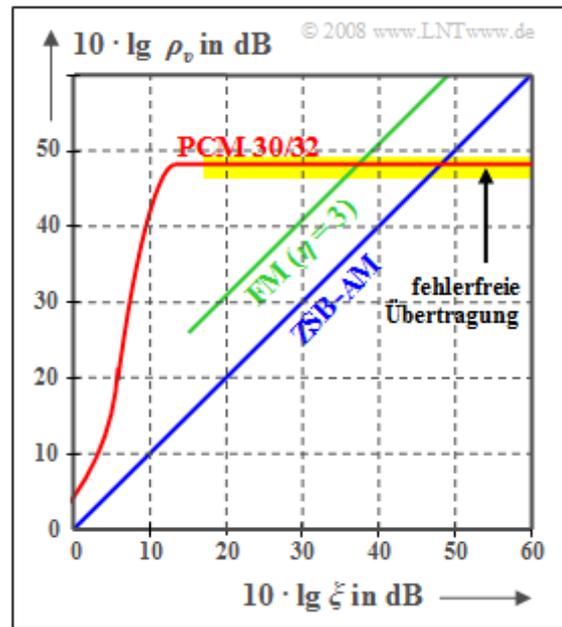
$\mu = 0$: LLL, $\mu = 1$: LLH, $\mu = 2$: LHH, $\mu = 3$: LHL,
 $\mu = 4$: HHL, $\mu = 5$: HHH, $\mu = 6$: HLH, $\mu = 7$: HLL.

Signal–zu–Rausch–Leistungsverhältnis (1)

Das digitale Pulscodemodulation (**PCM**) wird nun den analogen Modulationsverfahren (AM, FM) hinsichtlich des erreichbaren **Sinken–SNR** $\rho_v = P_q/P_\varepsilon$ bei AWGN–Rauschen vergleichend gegenüber gestellt. Wie in **Kapitel 3.3** bezeichnet $\xi = \alpha_K^2 \cdot P_S/(N_0 \cdot B_{NF})$ die Leistungskenngröße. Der Parameter ξ fasst

- den Kanaldämpfungsfaktor α_K (quadratisch),
- die Sendeleistung P_S ,
- die AWGN–Rauschleistungsdichte N_0 sowie
- die Bandbreite B_{NF} des Analogsignals

zusammen. Bei einer harmonischen Schwingung ist die Bandbreite B_{NF} durch die Frequenz f_N zu ersetzen.



Die beiden Vergleichskurven **AM** (Kapitel 2.3) und **FM** (Kapitel 3.3) lassen sich wie folgt beschreiben:

- ZSB–AM ohne Träger: $\rho_v = \xi \Rightarrow 10 \cdot \lg \rho_v = 10 \cdot \lg \xi$,
- FM mit Modulationsindex $\eta = 3$: $\rho_v = 3/2 \eta^2 \cdot \xi = 13.5 \cdot \xi \Rightarrow 10 \cdot \lg \rho_v = 10 \cdot \lg \xi + 11.3 \text{ dB}$.

Die Kurve für das PCM 30/32–System ist wie folgt zu interpretieren:

- Ist die Leistungskenngröße ξ hinreichend groß, so treten keine Übertragungsfehler auf und das Fehlersignal $\varepsilon(t) = v(t) - q(t)$ ist allein auf die Quantisierung zurückzuführen ($P_\varepsilon = P_Q$).

- Mit der Quantisierungsstufenzahl $M = 2^N$ gilt dann näherungsweise:

$$\rho_v = \frac{P_q}{P_\varepsilon} = M^2 = 2^{2N} \Rightarrow 10 \cdot \lg \rho_v = 20 \cdot \lg M = N \cdot 6.02 \text{ dB}$$

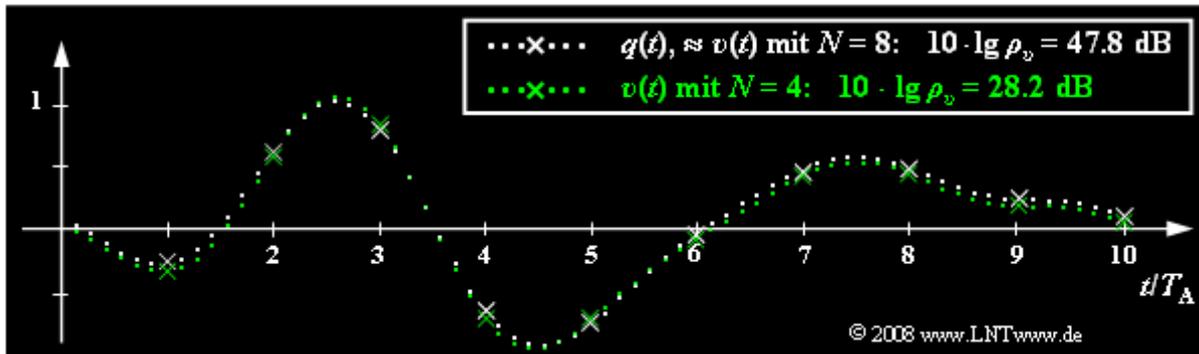
$$\Rightarrow N = 8, M = 256 : 10 \cdot \lg \rho_v = 48.16 \text{ dB}.$$

- Anzumerken ist, dass die angegebene Gleichung nur für ein sägezahnförmiges Quellsignal exakt gültig ist. Bei sinusförmigem Quellsignal ist die Abweichung jedoch nicht sehr groß.
- Mit kleiner werdendem ξ (kleinere Sendeleistung oder größere Rauschleistungsdichte) nehmen die Übertragungsfehler zu. Damit wird $P_\varepsilon > P_Q$ und der Sinken–Störabstand wird kleiner.
- Die PCM (mit $M = 256$) ist den Analogverfahren (AM und FM) nur im unteren und mittleren ξ –Bereich überlegen. Spielen dagegen Übertragungsfehler keine Rolle mehr, so ist durch eine größere Leistungskenngröße keine Verbesserung mehr zu erzielen (horizontaler Kurvenabschnitt).
- Eine Verbesserung bringt dann nur eine Erhöhung von N (Bitanzahl pro Abtastwert) und damit auch die Erhöhung von $M = 2^N$ (Quantisierungsstufenzahl). Beispielsweise erreicht man bei einer Compact Disc mit dem Parameter $N = 16 \Rightarrow M = 65536$ den Wert $10 \cdot \lg \rho_v = 96.32 \text{ dB}$.

Hinweis: Die näherungsweise Berechnung des SNR für kleine ξ –Werte folgt auf den nächsten Seiten.

Signal-zu-Rausch-Leistungsverhältnis (2)

Die folgende Grafik zeigt den begrenzenden Einfluss der Quantisierung. Weiß gepunktet eingezeichnet ist das Quellensignal $q(t)$ und grün gepunktet das Sinkensignal $v(t)$ nach einer PCM mit $N = 4 \Rightarrow M = 16$. Die Abtastzeitpunkte sind durch Kreuze markiert. Übertragungsfehler werden vorerst ausgeschlossen und die Abtastung sowie die Signalrekonstruktion sind bestmöglich an das Quellensignal angepasst.

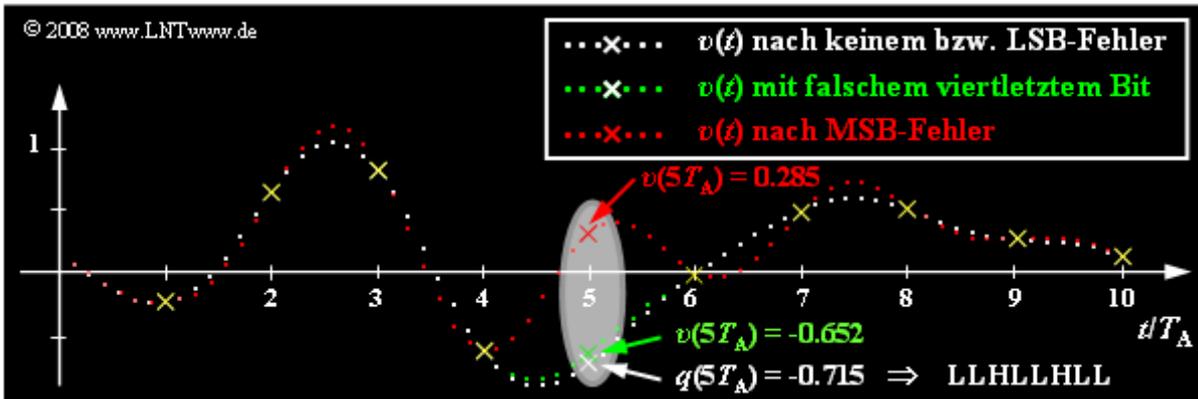


Dieses Bild kann wie folgt interpretiert werden:

- Mit $N = 8 \Rightarrow M = 256$ ist das Sinkensignal $v(t)$ vom Quellensignal $q(t)$ mit dem bloßen Auge nicht zu unterscheiden. Für beide gilt näherungsweise der weiß gepunktete Signalverlauf.
- Am Sinkenstörabstand $10 \cdot \lg \rho_v = 47.8$ dB erkennt man aber, dass das Quantisierungsrauschen (Leistung des Fehlersignals) nur etwa um den Faktor $1.6 \cdot 10^{-5}$ kleiner ist als die Leistung des Quellensignals. Dieses SNR wäre bei einem Sprach- oder Musiksignal schon deutlich hörbar.
- Obwohl das hier betrachtete Quellensignal weder sägezahnförmig noch cosinusförmig verläuft, sondern sich aus mehreren Frequenzanteilen zusammensetzt, weicht die angegebene Näherung $\rho_v \approx M^2 \Rightarrow 10 \cdot \lg \rho_v = 48.16$ dB nur unwesentlich vom tatsächlichen Wert ab.
- Dagegen erkennt man für $N = 4 \Rightarrow M = 16$ bereits Abweichungen zwischen dem Sinkensignal (grün markiert) und dem Quellensignal (weiße Markierung), was auch durch den sehr kleinen Störabstand $10 \cdot \lg \rho_v = 28.2$ dB quantitativ zum Ausdruck kommt.

Einfluss von Übertragungsfehlern (1)

Ausgehend vom gleichen Analogsignal $q(t)$ wie auf der letzten Seite und einer linearen Quantisierung mit $N = 8 \text{ Bit} \Rightarrow M = 256$ werden nun die Auswirkungen von Übertragungsfehlern anhand des jeweiligen Sinkensignals $v(t)$ verdeutlicht.



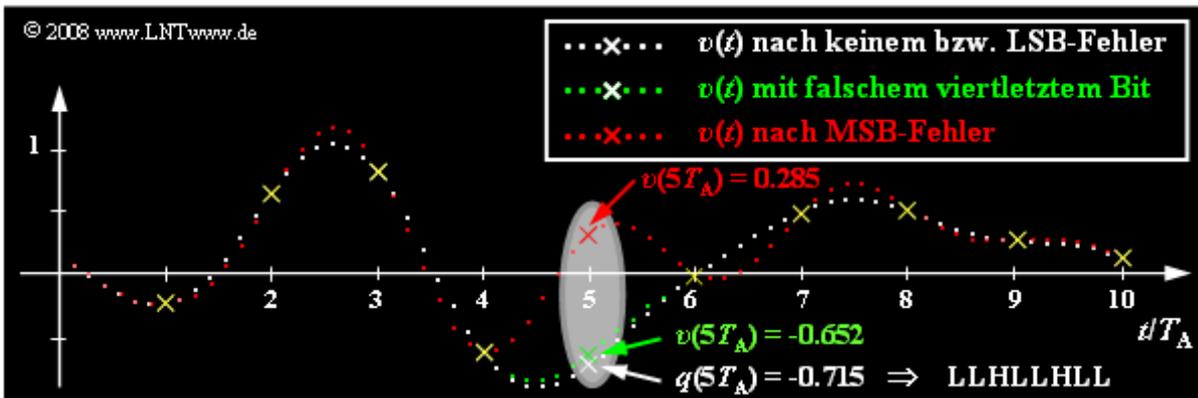
Die weißen Punkte markieren das Quellsignal $q(t)$. Ohne Übertragungsfehler ist das Sinkensignal $v(t)$ bei Vernachlässigung der Quantisierung genau so groß. Nun wird jeweils genau ein Bit des 5. Abtastwertes $q(5 \cdot T_A) = -0.715$ verfälscht, wobei dieser Abtastwert mit **LLHLLHLL** codiert wurde. Dieser Grafik zugrunde liegt der Dualcode, das heißt, dass das unterste Quantisierungsintervall ($\mu = 0$) mit **LLLL LLLL** und das oberste Intervall ($\mu = 255$) mit **HHHH HHHH** dargestellt wird.

© 2008 www.LNTwww.de	$v(5T_A)$	$10 \cdot \lg \rho_v$	in Grafik
kein Übertragungsfehler	-0.715	47.8 dB	weiß
niederwertigstes Bit (LSB) verfälscht	-0.707	44.3 dB	ss weiß
viertniedrigstes Bit verfälscht	-0.652	28.2 dB	grün
höchstwertigstes Bit (MSB) verfälscht	+0.285	4.1 dB	rot

Die Tabelle zeigt die Ergebnisse dieser Untersuchung. Der angegebene Störabstand $10 \cdot \lg \rho_v$ wurde aus dem dargestellten (sehr kurzen) Signalausschnitt der Dauer $10 \cdot T_A$ berechnet. Bei jeweils einem Fehler bei der Übertragung von $10 \cdot 8 = 80$ Bit entspricht dies einer Bitfehlerrate von 1.25%.

Die Bildbeschreibung wird auf der nächsten Seite fortgesetzt.

Einfluss von Übertragungsfehlern (2)



Die in der Grafik und der Tabelle dargestellten Ergebnisse dieser Fehleranalyse (siehe unten) können wie folgt zusammengefasst werden:

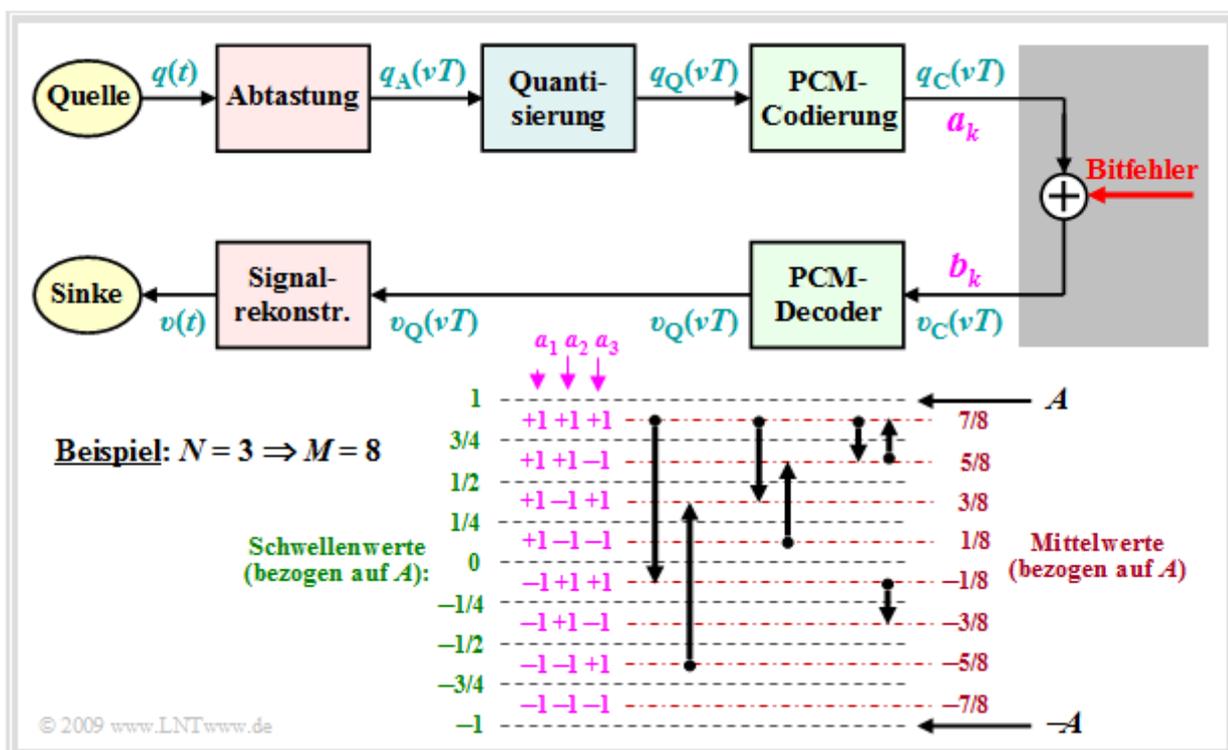
- Wird nur das letzte Bit des Binärwortes verfälscht (LSB: *Least Significant Bit*, LLHLLHLL \Rightarrow LLHLLHLL), so ist mit dem bloßen Auge kein Unterschied zur fehlerfreien Übertragung zu erkennen (weißer Kurvenzug). Trotzdem wird der Störabstand um 3.5 dB vermindert.
- Ein Übertragungsfehler des viertletzten Bits (grüne Kurve, LLHLLHLL \Rightarrow LLHLLHLL) führt bereits zu einer deutlich erkennbaren Verfälschung um 8 Quantisierungsintervalle. Das heißt: $v(5T_A) - q(5T_A) = 8/256 \cdot 2 = 0.0625$ und der Störabstand sinkt auf $10 \cdot \lg \rho_v = 28.2$ dB.
- Die rote Kurve zeigt schließlich den Fall, dass das MSB (*Most Significant Bit*) verfälscht wird: LLHLLHLL \Rightarrow HLHLLHLL. Dies führt zur Verfälschung $v(5T_A) - q(5T_A) = 1$ (was dem halben Aussteuerbereich entspricht). Der Störabstand beträgt nun nur mehr etwa 4 dB.
- Zu allen Abtastzeitpunkten mit Ausnahme von $5T_A$ stimmt $v(t)$ bis auf den Quantisierungsfehler mit $q(t)$ exakt überein. Außerhalb der durch gelbe Kreuze markierten Zeitpunkte führt der eine Fehler bei $5T_A$ in einem ausgedehnten Bereich zu starken Abweichungen, was auf die Interpolation mit der si-förmigen Impulsantwort des Rekonstruktionstiefpasses $H(f)$ zurückzuführen ist.

© 2008 www.LNTwww.de	$v(5T_A)$	$10 \cdot \lg \rho_v$	in Grafik
kein Übertragungsfehler	-0.715	47.8 dB	weiß
niederwertigstes Bit (LSB) verfälscht	-0.707	44.3 dB	weiß
viertniedrigstes Bit verfälscht	-0.652	28.2 dB	grün
höchstwertigstes Bit (MSB) verfälscht	+0.285	4.1 dB	rot

Einfluss von Übertragungsfehlern (3)

Nun soll versucht werden, die SNR-Kurve des PCM-Systems unter Berücksichtigung von Bitfehlern zumindest näherungsweise zu bestimmen. Wir gehen dabei vom folgenden Blockschaltbild aus und setzen weiter voraus:

- Jeder Abtastwert $q_A(vT)$ wird mit M Stufen quantisiert und mit $N = \text{ld}(M)$ Binärzeichen (Bit) dargestellt. Im Beispiel gilt $M = 8 \Rightarrow N = 3$.
- Die Binärdarstellung von $q_Q(vT)$ liefert die Amplitudenkoeffizienten a_k ($k = 1, \dots, N$), die durch Bitfehler in die Koeffizienten b_k verfälscht werden können. Sowohl a_k als auch b_k sind ± 1 .
- Ein Bitfehler ($b_k \neq a_k$) tritt mit der Wahrscheinlichkeit p_B auf. Jedes Bit wird gleichwahrscheinlich verfälscht und in jedem PCM-Wort ist maximal ein Fehler (nur eines der N Bits kann falsch sein).



Aus dem unteren Diagramm ist für $N = 3$ und die natürliche Binärcodierung (Dualcode) zu erkennen:

- Eine Verfälschung des Koeffizienten a_1 verändert den quantisierten Wert $q_Q(vT)$ um $\pm A$.
- Eine Verfälschung des Koeffizienten a_2 verändert den quantisierten Wert $q_Q(vT)$ um $\pm A/2$.
- Eine Verfälschung des Koeffizienten a_3 verändert den quantisierten Wert $q_Q(vT)$ um $\pm A/4$.

Durch Verallgemeinerung erhält man für die Abweichung $\varepsilon_k = v_Q(vT) - q_Q(vT)$ unter der Voraussetzung, dass der Amplitudenkoeffizient a_k falsch übertragen wurde:

$$\varepsilon_k = -a_k \cdot A \cdot 2^{-k+1}.$$

Auf der nächsten Seite wird hieraus die so genannte Fehlerrauschleistung $P_F = E[\varepsilon_k^2]$ berechnet.

Einfluss von Übertragungsfehlern (4)

Für die so genannte **Fehlerrauschleistung** erhält man nach Mittelung über alle Verfälschungswerte ε_k (mit $1 \leq k \leq N$) unter Berücksichtigung der Bitfehlerwahrscheinlichkeit p_B :

$$\begin{aligned}
 P_F = E[\varepsilon_k^2] &= \sum_{k=1}^N p_B \cdot (-a_k \cdot A \cdot 2^{-k+1})^2 = \\
 &= p_B \cdot A^2 \cdot \sum_{k=0}^{N-1} 2^{-2k} = p_B \cdot A^2 \cdot \frac{1 - 2^{-2N}}{1 - 2^{-2}} \approx 4/3 \cdot p_B \cdot A^2.
 \end{aligned}$$

Hierbei ist die Summenformel der geometrischen Reihe sowie die Näherung $1 - 2^{-2N} \approx 1$ verwendet. Für $N = 8 \Rightarrow M = 256$ beträgt der damit verbundene relative Fehler beispielsweise etwa 10^{-5} .

Ohne Berücksichtigung von Übertragungsfehlern hat sich für das Signal-zu-Rausch-Leistungsverhältnis $\rho_v = P_S/P_Q$ ergeben, wobei bei einem gleichverteilten Quellensignal (zum Beispiel sägezahnförmig) die Signalleistung und die Quantisierungsrauschleistung wie folgt zu berechnen ist:

$$P_S = \frac{A^2}{3}, \quad P_Q = \frac{A^2}{3} \cdot 2^{-2N}.$$

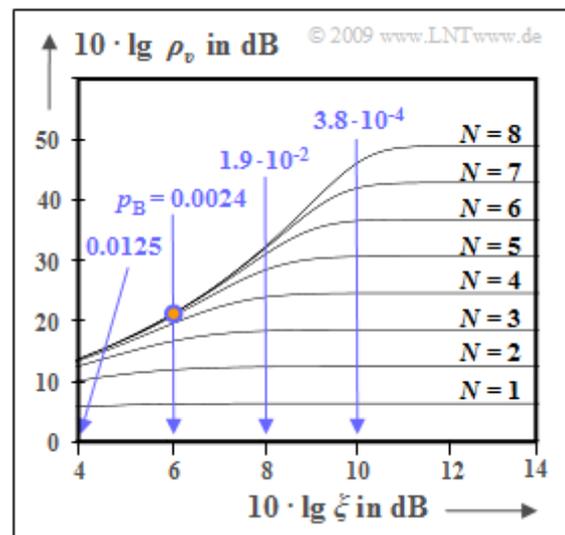
Unter **Berücksichtigung der Übertragungsfehler** erhält man mit obigem Ergebnis:

$$\rho_v = \frac{P_S}{P_Q + P_F} = \frac{A^2/3}{A^2/3 \cdot 2^{-2N} + A^2/3 \cdot 4 \cdot p_B} = \frac{1}{2^{-2N} + 4 \cdot p_B}.$$

Die Grafik rechts zeigt den Rauschabstand $10 \cdot \lg \rho_v$ in Abhängigkeit der (logarithmierten) Leistungskenngröße $\xi = P_S/(N_0 \cdot B_{NF})$, wobei B_{NF} die Signalbandbreite angibt. Die konstante Kanaldämpfung sei $\alpha_K = 1$.

Beim **optimalen Binärsystem** und AWGN-Rauschen gilt aber für die Leistungskenngröße auch $\xi = E_B/N_0$ (Energie pro Bit bezogen auf die Rauschleistungsdichte), und die **Bitfehlerwahrscheinlichkeit** ist dann mit der Gaußschen Fehlerfunktion $Q(x)$ wie folgt gegeben:

$$p_B = Q\left(\sqrt{2E_B/N_0}\right).$$



Für $N = 8 \Rightarrow 2^{-2N} = 1.5 \cdot 10^{-5}$ sowie $10 \cdot \lg \xi = 6$ dB $\Rightarrow p_B = 0.0024$ (siehe Grafik) ergibt sich:

$$\rho_v = \frac{1}{1.5 \cdot 10^{-5} + 4 \cdot 0.0024} \approx 100 \Rightarrow 10 \cdot \lg \rho_v \approx 20 \text{ dB}.$$

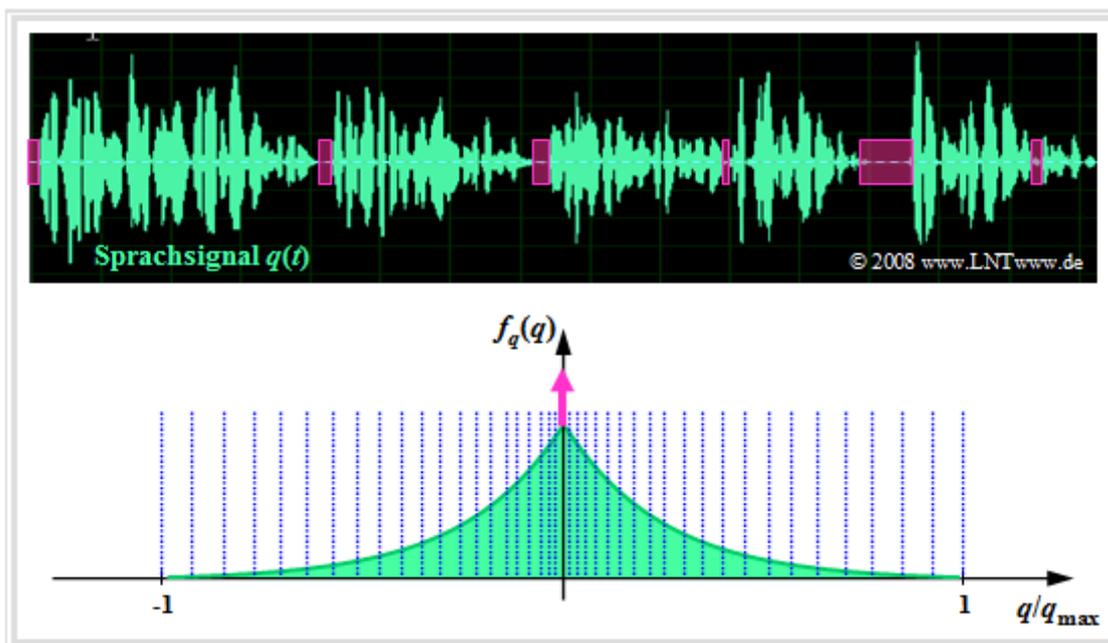
Dieser kleine ρ_v -Wert geht auf den Term $4 \cdot 0.0024$ (Einfluss des Übertragungsfehlers) zurück, während im horizontalen Kurvenzug für jedes N der Term 2^{-2N} – also das Quantisierungsrauschen – dominiert.

Nichtlineare Quantisierung (1)

Häufig werden die Quantisierungsintervalle nicht gleich groß gewählt, sondern man verwendet für den inneren Amplitudenbereich eine feinere Quantisierung als für große Amplituden. Dafür gibt es mehrere Gründe:

- Bei Audiosignalen werden Verfälschungen der leisen Signalanteile (also Werte in der Nähe der Nulllinie) subjektiv als störender empfunden als eine Beeinträchtigung großer Amplitudenwerte.
- Eine solche ungleichmäßige Quantisierung führt bei einem solchen Musik- oder Sprachsignal auch zu einem größeren Sinkeinstörabstand, da hier die Signalamplitude nicht gleichverteilt ist.

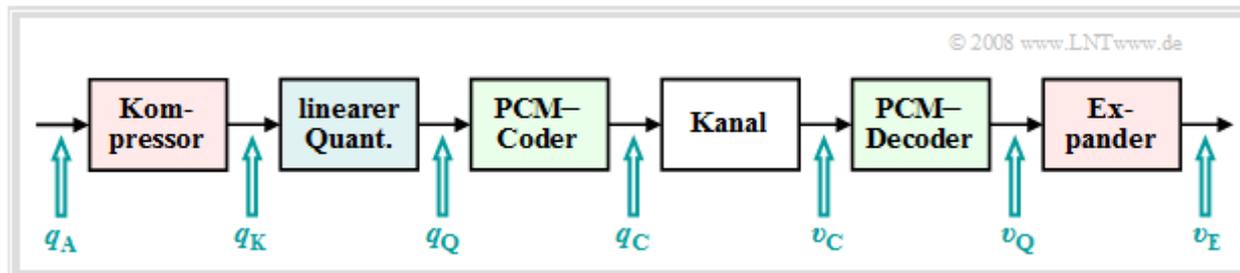
Die Grafik zeigt ein Sprachsignal $q(t)$ und dessen Amplitudenverteilung $f_q(q)$. Es handelt sich um die sogenannte Laplaceverteilung, die man durch eine kontinuierliche, zweiseitige Exponentialverteilung und eine Diracfunktion $\delta(q)$ zur Berücksichtigung der Sprachpausen (magentafarben) annähern kann.



Nur angedeutet ist die nichtlineare Quantisierung, zum Beispiel mittels der 13-Segment-Kennlinie, die in **Aufgabe A4.5** genauer beschrieben ist. Die Quantisierungsintervalle werden hierbei zu den Rändern hin abschnittsweise immer breiter. Die häufigeren kleinen Amplituden werden dagegen sehr fein quantisiert.

Nichtlineare Quantisierung (2)

Eine ungleichmäßige Quantisierung kann zum Beispiel dadurch realisiert werden, in dem die abgetasteten Werte $q_A(v \cdot T_A)$ zunächst durch eine **nichtlineare Kennlinie** $q_K(q_A)$ verformt und die entstehenden Ausgangswerte $q_K(v \cdot T_A)$ gleichmäßig quantisiert werden. Damit ergibt sich folgende Signalkette:



Eine solche ungleichmäßige Quantisierung bedeutet:

- Durch die nichtlineare Kennlinie $q_K(q_A)$ werden kleine Signalwerte verstärkt und große Werte abgeschwächt \Rightarrow **Kompression**.
- Diese bewußte Signalverzerrung muss man beim Empfänger durch die Umkehrfunktion $v_E(v_Q)$ wieder rückgängig machen \Rightarrow **Expandierung**.
- Den Gesamtvorgang von sendeseitiger Kompression und empfängerseitiger Expansion nennt man auch **Kompondierung**.

Für das PCM-System 30/32 wurde von der *Comité Consultatif International des Télégraphique et Téléphonique* (CCITT) die sogenannte A-Kennlinie empfohlen:

$$y(x) = \begin{cases} \frac{1+\ln(A \cdot x)}{1+\ln(A)} & \text{für } 1/A \leq x \leq 1, \\ \frac{A \cdot x}{1+\ln(A)} & \text{für } -1/A \leq x \leq 1/A, \\ -\frac{1+\ln(-A \cdot x)}{1+\ln(A)} & \text{für } -1 \leq x \leq -1/A. \end{cases}$$

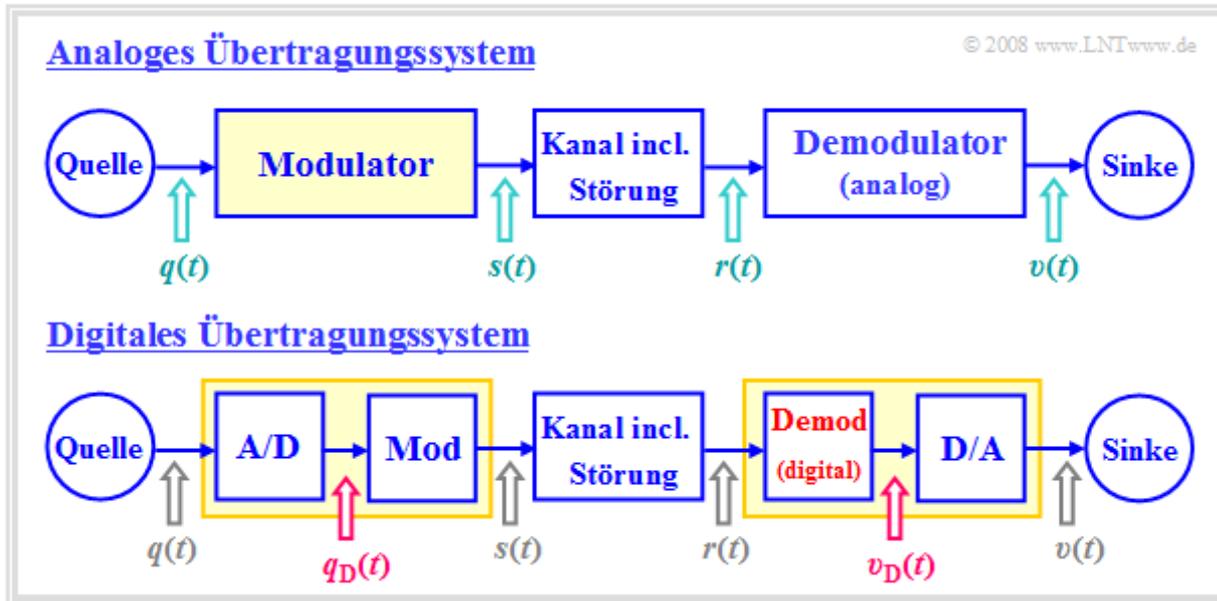
Hierbei ist zur Abkürzung $x = q_A(v \cdot T_A)$ und $y = q_K(v \cdot T_A)$ verwendet. Diese Kennlinie mit dem in der Praxis eingeführten Wert $A = 87.56$ hat eine sich ständig ändernde Steigung. Nähere Angaben zu dieser Art der ungleichmäßigen Quantisierung finden Sie in der **Aufgabe Z4.5**.

Hinweis: Im dritten Teil des Lernvideos **Pulscodemodulation** (Dauer 22:15) werden behandelt:

- die Definition des Signal-zu-Rausch-Leistungsverhältnisses (SNR),
- der Einfluss von Quantisierungsrauschen und Übertragungsfehlern,
- die Unterschiede zwischen linearer und nichtlinearer Quantisierung.

Unterschiede zwischen analogen und digitalen Modulationsverfahren

Die Grafik zeigt oben ein analoges Übertragungssystem und darunter gezeichnet ein Digitalsystem.



Die wesentlichen Unterschiede sind rot hervorgehoben:

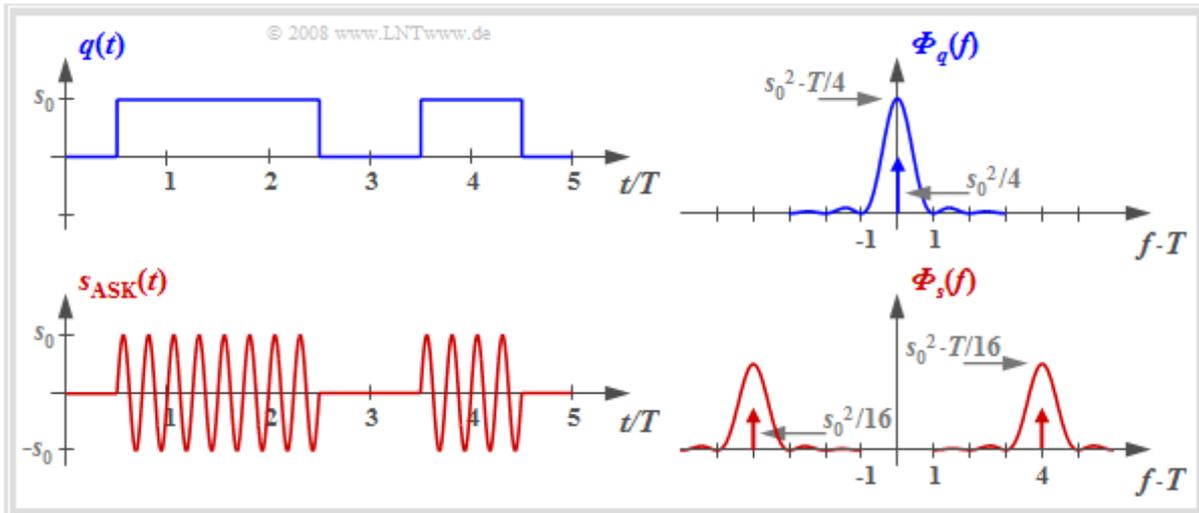
- Während beim oberen System am Modulatoreingang das analoge Quellensignal $q(t)$ anliegt, ist beim unteren Digitalsystem das modulierende Signal $q_D(t)$ ein Digitalsignal, gekennzeichnet durch die Amplitudenkoeffizienten a_ν , den Grundimpuls $g_q(t)$ sowie die Symboldauer T :

$$q_D(t) = \sum_{\nu=-\infty}^{+\infty} a_\nu \cdot g_q(t - \nu \cdot T).$$

- Die A/D-Wandlung kann z. B. mittels **Pulscode modulation** erfolgen und umfasst die Funktionen Abtastung, Quantisierung, Binärcodierung und Signalförmung. Der Grundimpuls $g_q(t)$ wird im Folgenden meist als NRZ-rechteckförmig mit Amplitude s_0 und Dauer T angenommen, so dass für die Spektralfunktion $G_q(f) = s_0 \cdot T \cdot \text{si}(\pi f T)$ mit $\text{si}(x) = \sin(x)/x$ gilt.
- Die Modulatoren können bei beiden Systemen durchaus gleich sein. Sie verändern einen der drei Signalparameter des Trägersignals $z(t)$ entsprechend dem Modulatoreingangssignal. Die digitalen Varianten von AM, PM und FM heißen *Amplitude Shift Keying* (ASK), *Phase Shift Keying* (PSK) und *Frequency Shift Keying* (FSK).
- Dagegen unterscheidet sich der Demodulator des Digitalsystems grundsätzlich von einem analogen Demodulator durch die erforderliche Entscheiderkomponente (in Hardware oder Software). Das Ausgangssignal $v_D(t)$ ist ebenso wie $q_D(t)$ digital. Dieses Signal muss anschließend noch in das analoge Sinkensignal $v(t)$ D/A-gewandelt werden.
- Das entscheidende Gütekriterium ist bei beiden Systemen das **Sinken-SNR** als der Quotient der Leistungen von Quellensignal $q(t)$ und Fehlersignal $\varepsilon(t) = v(t) - q(t)$. Bei einem Digitalsystem begnügt man sich meist mit dem Qualitätsmerkmal **Bitfehlerquote** (engl.: *Bit Error Rate*, BER), das sich auf die beiden Digitalsignale $q_D(t)$ und $v_D(t)$ bezieht. Diese ist in ein SNR umrechenbar.

ASK – Amplitude Shift Keying

Die Grafik zeigt das digitale Quellensignal $q(t)$ – auf den Index „D“ wird ab sofort verzichtet – sowie das ASK–Sendesignal $s_{\text{ASK}}(t) = q(t) \cdot \sin(2\pi \cdot f_T \cdot t)$, wobei hier von unipolaren Amplitudenkoeffizienten $a_v \in \{0, 1\}$ und einem sinusförmigen Träger ausgegangen wird. Dieses Verfahren wird insbesondere bei optischen Übertragungssystemen eingesetzt (da es bekanntlich keine negativen Lichtimpulse gibt) und ist auch unter der Bezeichnung **On–Off–Keying** bekannt.



In der rechten Bildhälfte sind – allerdings nicht maßstäblich – die dazugehörigen Leistungsdichtespektren (abgekürzt: LDS) dargestellt. Bei rechteckförmigem Grundimpuls $g_q(t)$ und gleichwahrscheinlichen unipolaren Amplitudenkoeffizienten gilt:

$$\Phi_q(f) = \frac{s_0^2 \cdot T}{4} \cdot \text{si}^2(\pi f T) + \frac{s_0^2}{4} \cdot \delta(f),$$

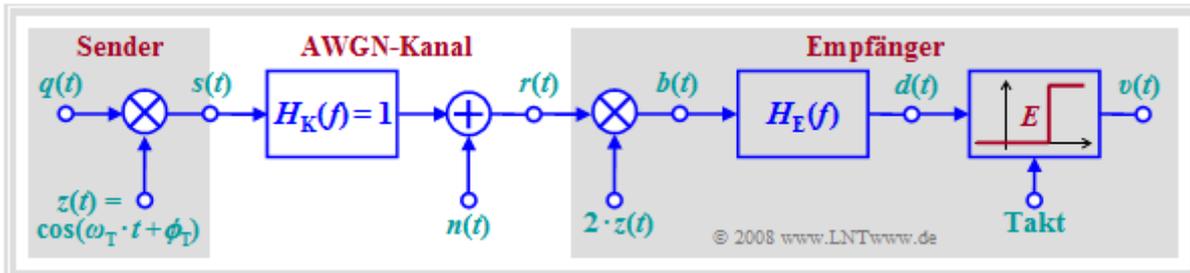
$$\Phi_s(f) = \frac{1}{4} \cdot [\Phi_q(f - f_T) + \Phi_q(f + f_T)].$$

Zu diesen Gleichungen ist zu bemerken:

- Der Gleichanteil $m_q = s_0/2$ des Quellensignals führt im Leistungsdichtespektrum $\Phi_q(f)$ zu einer Diracfunktion bei der Frequenz $f = 0$ mit dem Gewicht $s_0^2/4$.
- Das Leistungsdichtespektrum des ASK–Sendesignals ist gleich $\Phi_s(f) = \Phi_q(f) * \Phi_z(f)$, wobei sich das LDS $\Phi_z(f)$ des Trägersignals $Z(t)$ aus zwei Diracfunktionen bei $\pm f_T$ mit jeweiligem Gewicht $1/4$ zusammensetzt. Die Gleichung gilt auch bei anderer Trägerphase, „*“ beschreibt die Faltung.
- Das Leistungsdichtespektrum $\Phi_s(f)$ ist bis auf die Verschiebung um $\pm f_T$ formgleich mit $\Phi_q(f)$. Deshalb gehört ASK zu den **linearen digitalen Modulationsverfahren**.

Kohärente Demodulation von ASK-Signalen (1)

Die Grafik zeigt das Blockschaltbild eines ASK-Systems inklusive der Empfängerkomponenten. Das Quellensignal $q(t)$ sei NRZ-rechteckförmig und unipolar, das heißt, es gilt $a_v \in \{0, 1\}$. Der Kanal sei zunächst ideal, gekennzeichnet durch $H_K(f) = 1$ und $n(t) = 0 \Rightarrow r(t) = s(t)$.



Die Demodulation erfolgt hier kohärent mittels **Synchrondemodulator**, dessen Funktionsweise bereits bei den analogen Modulationsverfahren AM und PM beschrieben wurde.

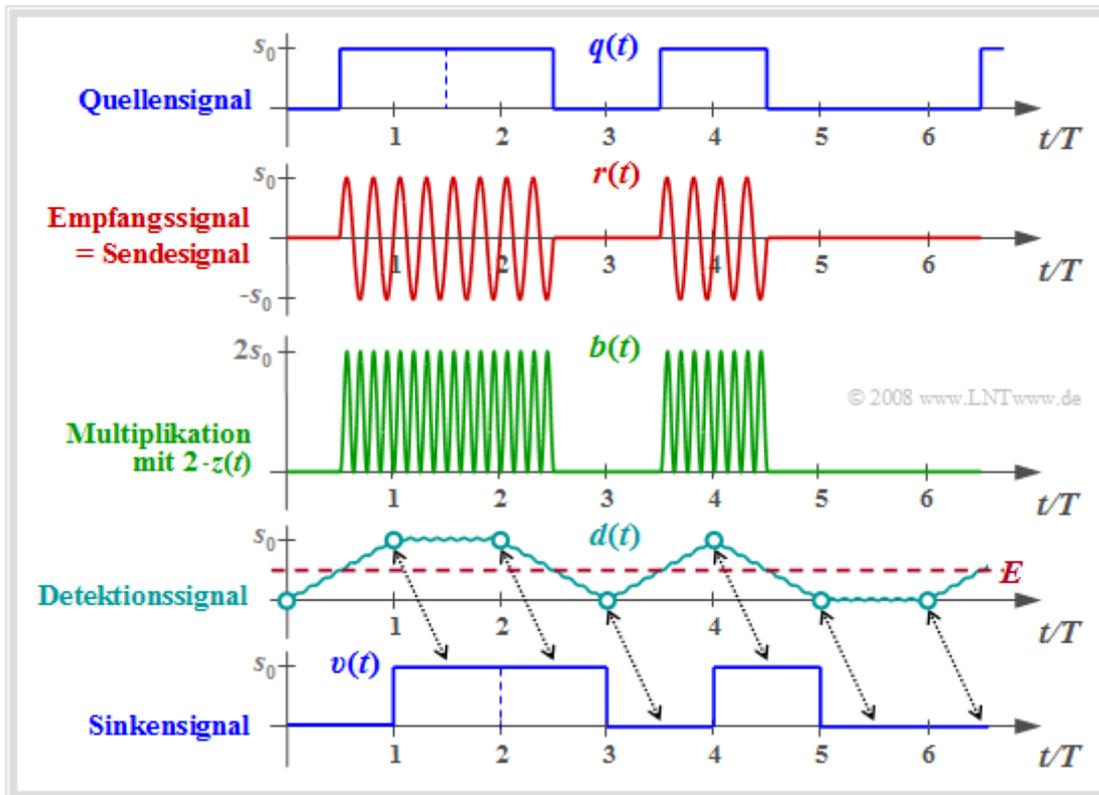
Zusammenfassend lässt sich sagen:

- Beim Empfänger wird das gleiche Trägersignal zugesetzt wie beim Sender, jedoch mit doppelter Amplitude. $z(t)$ bezeichnet den Träger beim Sender und $2 \cdot z(t)$ den Träger beim Empfänger.
- Nach der Multiplikation folgt ein geeignet dimensionierter Tiefpass mit dem Frequenzgang $H_E(f)$, der die höherfrequenten Anteile des Signals $b(t)$ entfernt.
- Schließlich wird das Detektionssignal $d(t)$ zu den Detektionszeitpunkten $v \cdot T$ abgetastet und mit Hilfe eines Schwellenwertentscheiders mit der Entscheidungsschwelle $E = s_0/2$ entschieden.
- Das Sinkensignal $v(t)$ am Ausgang des Entscheiders ist rechteckförmig und im rauschfreien Fall (oder bei nur kleinen Rauschstörungen) bis auf die Laufzeit $T/2$ gleich dem Quellensignal $q(t)$.

Zu beachten ist: Eine **kohärente Demodulation** erfordert, dass dem Empfänger die Trägerfrequenz f_T und die Trägerphase ϕ_T exakt bekannt sind. Der Empfänger muss diese beiden Größen aus dem Empfangssignal $r(t)$ extrahieren, was bei starken Kanalverzerrungen und großen Rauschstörungen durchaus aufwändig sein kann. Solche Realisierungsaspekte werden zum Beispiel in der **Aufgabe A4.8** zu diesem Kapitel behandelt.

Kohärente Demodulation von ASK-Signalen (2)

Die Grafik zeigt die im ASK-Blockschaltbild genannten Signale bei idealem Kanal: $H_K(f) = 1$, $n(t) = 0$.



Die einzelnen Signalverläufe können wie folgt interpretiert werden:

- Das Sendesignal $s(t)$ ist das Produkt aus dem unipolaren Quellensignal $q(t)$ und dem Trägersignal $z(t) = \sin(2\pi f_T t)$, wobei hier $f_T = 4/T$ gilt (nur jeweils vier Schwingungen pro Symboldauer).
- Das Empfangssignal $r(t) = s(t)$ wird zunächst mit dem Träger $z_E(t) = 2 \cdot \sin(2\pi f_T t) \Rightarrow$ doppelte Amplitude gegenüber $z(t)$, kein Frequenz- und Phasenversatz – multipliziert. Damit ergibt sich:

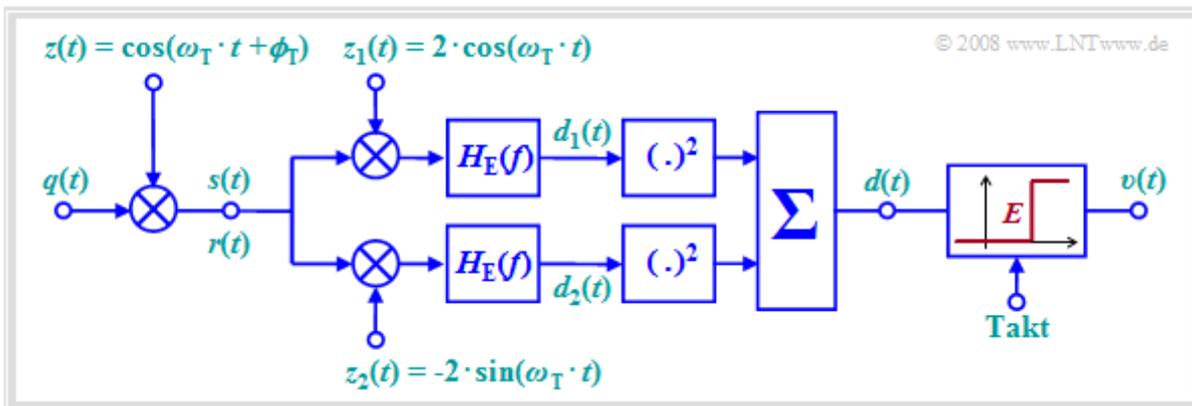
$$b(t) = 2 \cdot z(t) \cdot r(t) = 2 \cdot z^2(t) \cdot q(t) = q(t) \cdot [1 - \cos(4\pi f_T t)] .$$
- Das TP-Filter mit dem Frequenzgang $H_E(f) = \text{si}(\pi f_T T)$ und dementsprechend rechteckförmiger Impulsantwort $h_E(t)$ formt aus dem Signal $b(t)$ das Detektionssignal $d(t) = b(t) * h_E(t)$.
- $h_E(t)$ ist an den rechteckförmigen Grundimpuls $g_q(t)$ angepasst; man spricht vom sog. **Matched-Filter** \Rightarrow bestmöglicher Kompromiss zwischen Entzerrung und Rauschleistungsbegrenzung.
- Ohne Rauschen gilt $d(vT) = q(vT) \in \{0, s_0\}$. Bei (moderaten) Rauschstörungen ist mit großer Wahrscheinlichkeit $d(vT) > s_0/2$, falls $a_v = +1$, und es wird $d(vT) < s_0/2$ für $a_v = 0$ gelten.
- Der Schwellenwertentscheider gewinnt aus dem Vergleich der Detektionsabastwerte $d(vT)$ mit der Entscheiderschwelle $E = s_0/2$ das Sinkensignal $v(t)$, das bei fehlerfreier Entscheidung bis auf die Laufzeit $T/2$ identisch mit $q(t)$ ist.

Inkohärente Demodulation von ASK-Signalen

Wir gehen weiter von ASK-Modulation sowie dem idealen (das heißt: verzerrungs-, dämpfungs- und rauschfreien) Übertragungskanal aus, so dass gilt:

$$r(t) = s(t) = q(t) \cdot \cos(2\pi \cdot f_T \cdot t + \phi_T).$$

Weiter wird für diese Seite vorausgesetzt, dass dem Empfänger zwar die Trägerfrequenz f_T , nicht jedoch die Trägerphase ϕ_T bekannt ist. Üblich ist, auch diesen Demodulator als **inkohärent** zu bezeichnen.



Die Grafik zeigt einen solchen inkohärenten Demodulator, dessen Funktionsweise hier nur stichpunktartig angegeben werden soll. Das Demodulationsergebnisse ist unabhängig von der Trägerphase ϕ_T , die der Empfänger nicht kennt.

- Die Signale $d_1(t)$ und $d_2(t)$ nach den beiden Matched-Filtern mit jeweiligem Frequenzgang $H_E(f)$ sind formgleich mit dem Detektionssignal $d_{\text{koh}}(t)$ entsprechend der **Grafik** auf der letzten Seite, aber gegenüber diesem im Allgemeinen wegen der fehlenden Phasenanpassung gedämpft:

$$d_1(t) = d_{\text{koh}}(t) \cdot \cos(\phi_T), \quad d_2(t) = -d_{\text{koh}}(t) \cdot \sin(\phi_T).$$

- Ist der Amplitudenkoeffizient $a_\nu = 0$, so sind im rauschfreien Fall die beiden Signalwerte $d_1(\nu \cdot T)$ und $d_2(\nu \cdot T)$ jeweils 0. Andernfalls ($a_\nu = 1$) gilt für den Zeitpunkt $\nu \cdot T$:

$$d_1(\nu \cdot T) = s_0 \cdot \cos(\phi_T), \quad d_2(\nu \cdot T) = -s_0 \cdot \sin(\phi_T).$$

- Nach Quadrierung der zwei Teilsignale erhält man für das Summensignal:

$$d(\nu \cdot T) = \begin{cases} 0 & \text{falls } a_\nu = 0, \\ s_0^2 & \text{falls } a_\nu = 1. \end{cases}$$

- Durch die Schwellenwertentscheidung – sinnvollerweise mit der Entscheiderschwelle $E = s_0^2/4$ – können die Amplitudenkoeffizienten a_ν entschieden werden. Allerdings ergibt sich eine etwas größere **Bitfehlerwahrscheinlichkeit** als bei kohärenter Demodulation.

BPSK – Binary Phase Shift Keying

Bei analoger Phasenmodulation lautet das Sendesignal entsprechend **Kapitel 3.1** allgemein:

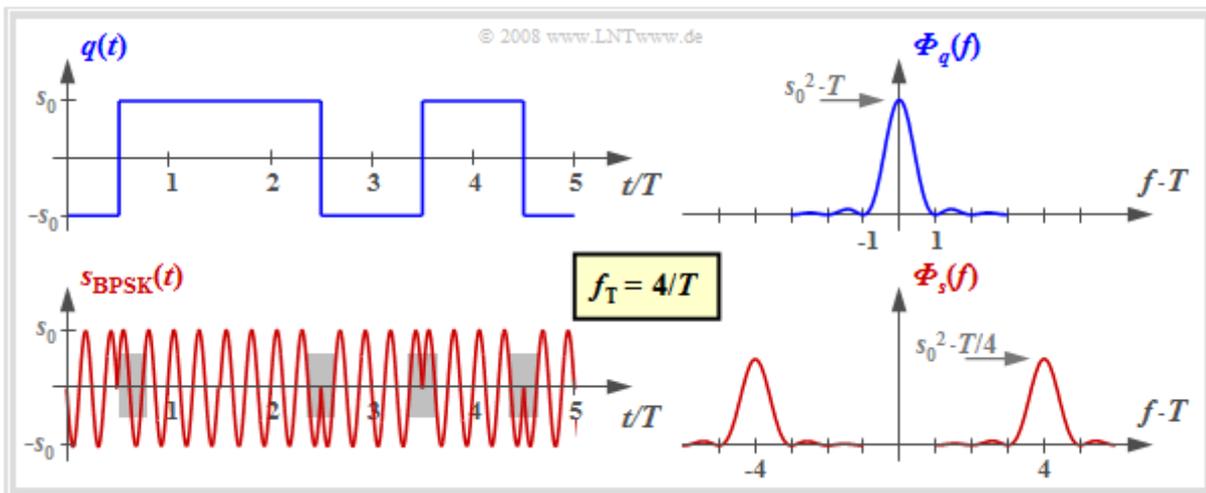
$$s_{\text{PM}}(t) = s_0 \cdot \cos(2\pi f_T t + \phi_T + K_{\text{PM}} \cdot q(t)).$$

Bei bipolarem Quellensignal $\Rightarrow a_\nu \in \{-1, +1\}$, der angenommenen Trägerphase $\phi_T = \pi$ (180°) und mit der geeignet dimensionierten Modulatorkonstanten $K_{\text{PM}} = \pi/(2s_0)$ ergibt sich im ν -ten Zeitintervall:

$$s_{\text{BPSK}}(t) = \begin{cases} s_0 \cdot \cos(2\pi f_T t + \pi + \pi/2) & \text{falls } a_\nu = +1, \\ s_0 \cdot \cos(2\pi f_T t + \pi - \pi/2) & \text{falls } a_\nu = -1. \end{cases}$$

Diese Gleichung für die binäre Phasenmodulation (BPSK) lässt sich wie folgt umformen:

$$s_{\text{BPSK}}(t) = a_\nu \cdot s_0 \cdot \sin(2\pi f_T t) = \begin{cases} s_0 \cdot \sin(2\pi f_T t) & \text{falls } a_\nu = +1, \\ -s_0 \cdot \sin(2\pi f_T t) & \text{falls } a_\nu = -1. \end{cases}$$

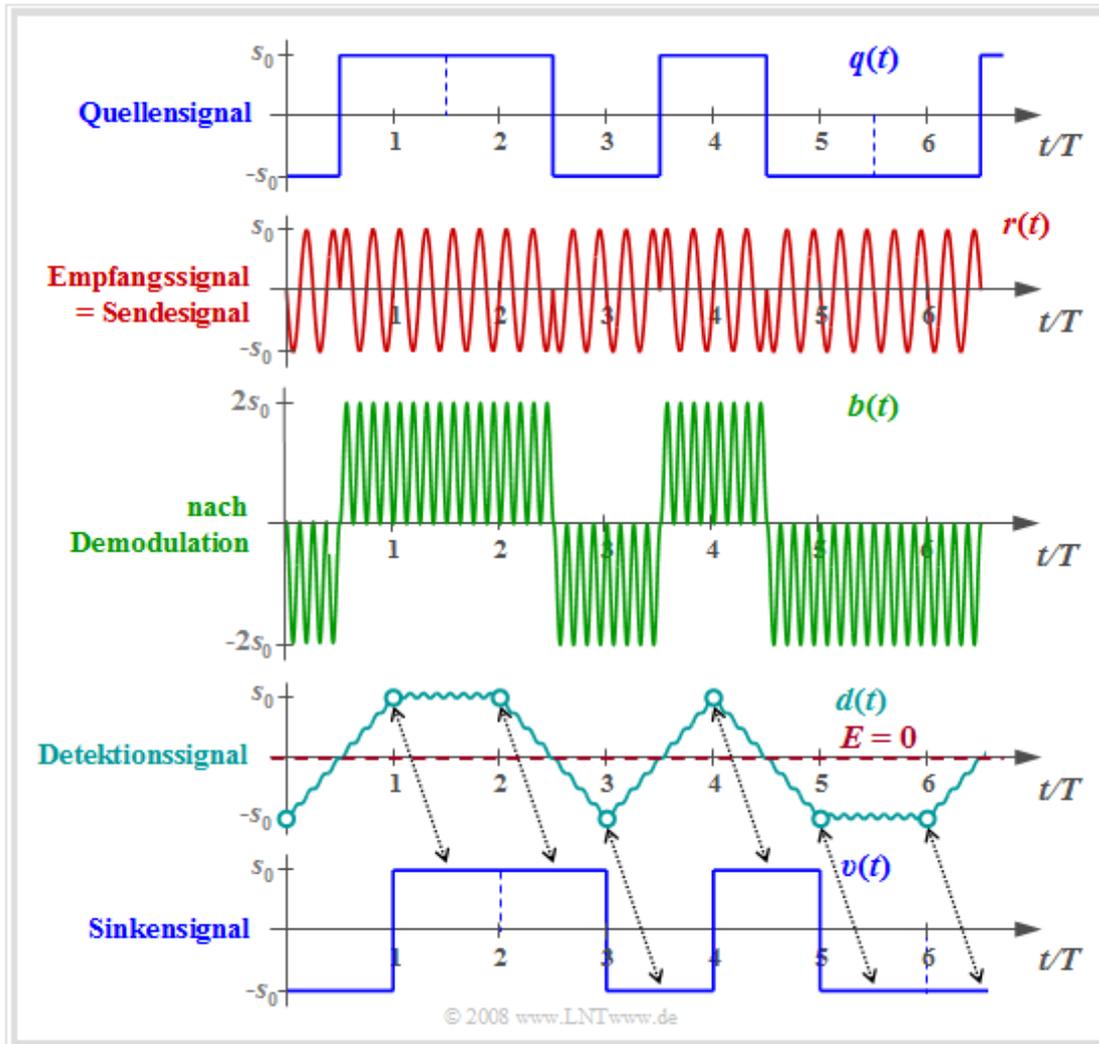


In der Grafik sind die Signale und die dazugehörigen Leistungsdichtespektren skizziert. Man erkennt:

- Das BPSK-Signal lässt sich wie das ASK-Signal als Produkt von Quellensignal $q(t)$ und Trägersignal $z(t)$ darstellen. Der einzige Unterschied liegt in den bipolaren Amplitudenkoeffizienten $a_\nu \in \{-1, +1\}$ gegenüber den unipolaren Koeffizienten (0 oder 1) bei ASK.
- Im Gegensatz zur ASK ist bei der BPSK – wie bei jeder Form von Phasenmodulation – die Hüllkurve konstant. Die Information wird hier durch die Phasensprünge innerhalb des Sendesignals $s(t)$ übermittelt (graue Hinterlegungen in der Grafik).
- Die Leistungsdichtespektren bei BPSK unterscheiden sich von denen bei ASK lediglich durch die fehlenden Diracfunktionen (da nun $q(t)$ keinen Gleichanteil beinhaltet) sowie durch den Faktor 4 bezüglich der kontinuierlichen LDS-Anteile.
- Daraus folgt weiter, dass die binäre Phasenmodulation zu den **linearen Modulationsverfahren** gezählt werden kann. Im Allgemeinen ist nämlich die (analoge) Phasenmodulation bis auf wenige Ausnahmen hinsichtlich des Quellensignals nichtlinear.
- Für die Grafiken wurden bei ASK (Sinus) und BPSK (Minus-Cosinus) aus Darstellungsgründen verschiedene Trägerphasen gewählt. Diese willkürliche Festlegung ist jedoch keine Einschränkung. Beide Verfahren funktionieren bei anderen Trägerphasen in gleicher Weise.

Demodulation und Detektion von BPSK-Signalen

Aufgrund der konstanten Hüllkurve des BPSK-Signals muss hier die Demodulation kohärent erfolgen. Es kann dabei vom gleichen **Blockschaltbild** wie bei der kohärenten ASK-Demodulation ausgegangen werden. Die Grafik zeigt die Signale $q(t)$, $r(t) = s(t)$, $b(t)$, $d(t)$ und $v(t)$.



Ein Vergleich mit den entsprechenden ASK-Signalen zeigt:

- Die Rechtecksignale $q(t)$ und $v(t)$ sind nun bipolar und für das Detektionssignal bei BPSK gilt:

$$d_{\text{BPSK}}(t) = 2 \cdot d_{\text{ASK}}(t) - s_0.$$
- Im betrachteten dämpfungs-, verzerrungs- und rauschfreien Fall sind alle Detektionsabstastwerte $d(v \cdot T) = \pm s_0$. Deshalb muss hier die Entscheidungsschwelle $E = 0$ verwendet werden.
- Man erkennt den doppelten Abstand der BPSK-Detektionsabstastwerte (Kreismarkierungen) von der Schwelle, was sich bezüglich der Fehlerwahrscheinlichkeit entscheidend auswirkt.

DPSK – Differential Phase Shift Keying

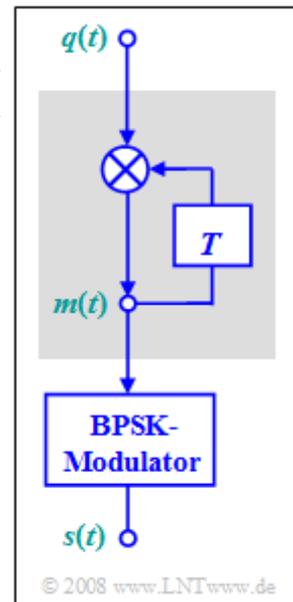
Die nebenstehende Grafik zeigt das Blockschaltbild des DPSK-Modulators (*Differential Phase Shift Keying*). Das (bipolare) Quellsignal $q(t)$ mit den Amplitudenkoeffizienten $q_v \in \{-1, +1\}$ wird entsprechend dieses Mappings in das Signal $m(t)$ mit den Amplitudenkoeffizienten

$$m_v = m_{v-1} \cdot q_v \in \{-1, +1\}$$

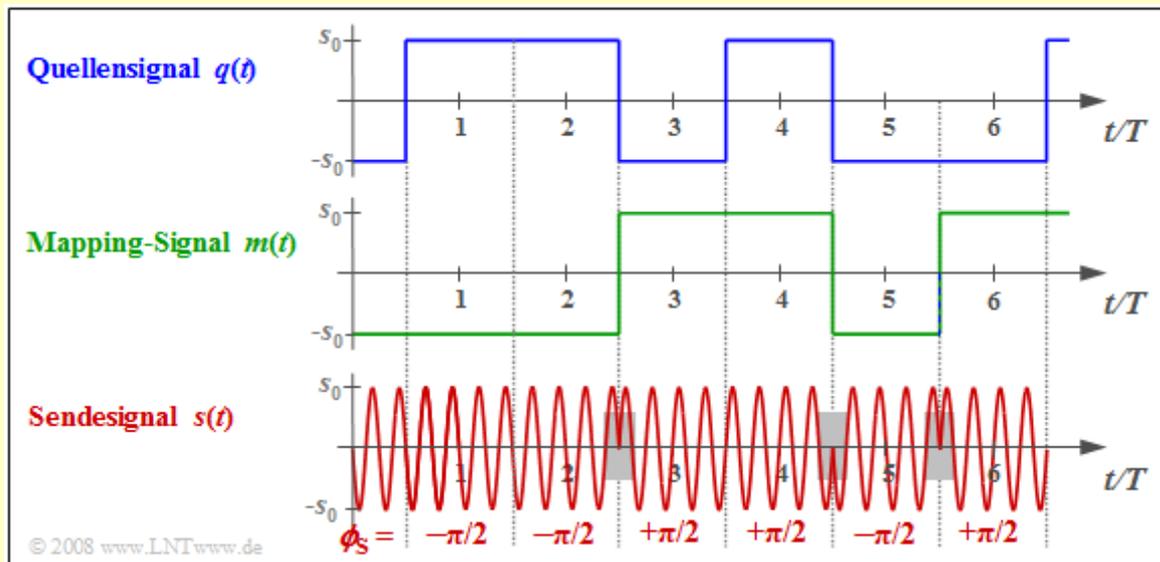
abgebildet, bevor es dem BPSK-Modulator zugeführt wird.

Ist $q_v = m_{v-1}$, so ergibt sich der gemappte Amplitudenkoeffizient $m_v = +1$. Dagegen weist $m_v = -1$ darauf hin, dass sich die Amplitudenkoeffizienten q_v und m_{v-1} unterscheiden.

Wesentlicher Vorteil der differentiellen binären Phasenmodulation ist, dass das so entstehende Signal $s(t)$ auch ohne Kenntnis der Trägerphase ϕ_T demoduliert werden kann, siehe **nächsten Seite**. Obwohl dem Empfänger die genaue Trägerfrequenz f_T bekannt sein muss, spricht man trotzdem von einem inkohärenten, manchmal auch vom **differentiell-kohärenten PSK-Demodulator**.



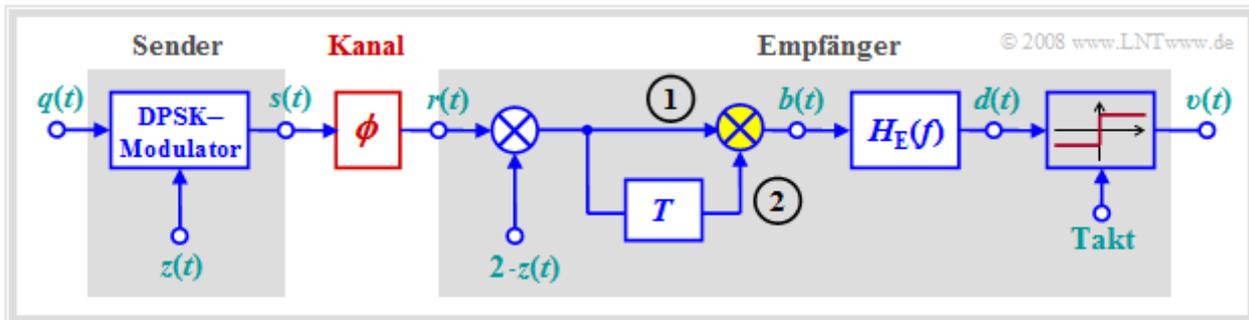
Beispiel: Die nachfolgende Grafik zeigt die Signale $q(t)$ und $m(t)$ sowie das DPSK-Sendesignal $s(t)$. Immer dann, wenn der Empfänger einen Phasensprung erkennt, entscheidet er sich für sich $v_v = -1$ (Zeitpunkte $v = 3, 5$ und 6), andernfalls für $+1$. v_v bezeichnet die Amplitudenkoeffizienten nach der Entscheidung, die möglichst mit den sendeseitigen Amplitudenkoeffizienten q_v übereinstimmen sollten.



Unterhalb des Sendesignals sind für die ersten sechs Symbole die Phasenwerte ϕ_s angegeben. Durch eine zusätzliche Phasendrehung auf dem Kanal, zum Beispiel um 70.3π , ändern sich zwar die absoluten Phasenwerte auf $69.8\pi, 69.8\pi, 70.8\pi, 69.8\pi, 70.8\pi$, und 69.8π , die Phasendifferenz benachbarter Symbole bleibt jedoch erhalten, so dass die differentiell-kohärente Demodulation trotzdem funktioniert. Ein entsprechender Demodulator wird auf der nächsten Seite vorgestellt.

Differenziell-kohärente Demodulation des DPSK-Signals

Die nachfolgende Grafik zeigt das Blockschaltbild eines Übertragungssystems mit DPSK-Modulation (*Differential Phase Shift Keying*) und differenziell-kohärenter Demodulation.



Stichpunktartig lässt sich die Funktionsweise wie folgt beschreiben:

- Ohne Berücksichtigung der Modulation mit den Trägersignalen $z(t)$ bzw. $2 \cdot z(t)$ liegt im Intervall v am Eingang (1) des gelb hervorgehobenen Multiplizierers das Symbol $m_v = m_{v-1} \cdot q_v$ an und am Eingang (2) das Symbol m_{v-1} . Die Multiplikation von (1) und (2) ergibt das gewünschte Ergebnis, nämlich $v_v = m_{v-1} \cdot q_v \cdot m_{v-1} = q_v$, da $m_{v-1} \in \{+1, -1\}$ gilt.
- Das Matched-Filter mit dem Frequenzgang $H_E(f)$ eliminiert die unerwünschten Anteile um die doppelte Trägerfrequenz, die durch die zweifache Multiplikation mit $z(t)$ bzw. $2 \cdot z(t)$ entstehen. Bei rechteckförmigem Grundimpuls $g_q(t)$ lässt sich der Frequenzgang $H_E(f)$ auch sehr einfach durch einen Integrator realisieren.
- Wir nehmen an, dass der Kanal eine Phasendrehung um ϕ bewirkt, die der Empfänger nicht kennt (rote Markierung). Geht man beispielsweise vom sendeseitigen Trägersignal $z(t) = \cos(2\pi \cdot f_T \cdot t)$ aus, so beinhaltet das Empfangssignal $r(t)$ einen multiplikativen Anteil mit $\cos(2\pi \cdot f_T \cdot t + \phi)$. Die Zusetzung des empfangsseitigen Trägers $2 \cdot z(t)$ erfolgt also nicht phasensynchron.
- Aber auch das um eine Symboldauer T verzögerte Signal $r(t - T)$ weist die gleiche Phase ϕ auf. Durch die Korrelation zwischen $2 \cdot r(t) \cdot z(t)$ und $2 \cdot r(t - T) \cdot z(t - T)$ wird erreicht, dass das Entscheidungsergebnis unabhängig von der zufälligen Phase ϕ ist. Man bezeichnet diese Art der Demodulation als **differenziell-kohärent**.

Fehlerwahrscheinlichkeiten - ein kurzer Überblick

Die Fehlerwahrscheinlichkeiten der behandelten digitalen Modulationsverfahren (ASK, BPSK, DPSK) werden in **Kapitel 1.5** des Buches „Digitalsignalübertragung“ unter verschiedenen Randbedingungen berechnet. Hier werden nur einige Ergebnisse ohne Beweis vorweg genommen, gültig für

- ein Sendesignal mit der mittleren Energie E_B pro Bit,
- AWGN-Rauschen mit der Rauschleistungsdichte N_0 ,
- bestmögliche Empfängerrealisierung nach dem Matched-Filter-Prinzip.

Betrachten wir zunächst die Bitfehlerwahrscheinlichkeit der **BPSK** (*Binary Phase Shift Keying*) unter der Voraussetzung eines kohärenten Empfängers:

$$p_B = Q\left(\sqrt{\frac{2 \cdot E_B}{N_0}}\right) = \frac{1}{2} \cdot \operatorname{erfc}\left(\sqrt{\frac{E_B}{N_0}}\right).$$

Dagegen gilt für die **ASK** (*Amplitude Shift Keying*) bei kohärenter Demodulation:

$$p_B = Q\left(\sqrt{\frac{E_B}{N_0}}\right) = \frac{1}{2} \cdot \operatorname{erfc}\left(\sqrt{\frac{E_B}{2 \cdot N_0}}\right).$$

In den Formeln wurden zwei Varianten der **komplementären Gaußschen Fehlerfunktion** verwendet:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-u^2/2} du, \quad \operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-u^2} du.$$

Trägt man die Bitfehlerwahrscheinlichkeit p_B über den Quotienten E_B/N_0 in doppelt-logarithmischem Maßstab auf, so liegt die ASK-Kurve stets um 3 dB rechts von der BPSK-Kurve. Diese Degradation ist auch ein Grund dafür, dass ASK in der Praxis nur selten eingesetzt wird.

Der entscheidende Vorteil der **DPSK** (*Differential Phase Shift Keying*) ist es, dass diese auch ohne Kenntnis der Trägerphase demoduliert werden kann. Diese einfache Realisierung erkaufte man sich durch eine gegenüber der kohärenten BPSK erhöhte Fehlerwahrscheinlichkeit:

$$p_B = \frac{1}{2} \cdot e^{-E_B/N_0}.$$

Die inkohärente Demodulation eines BPSK-Signals ist dagegen nicht möglich. Für die ASK erhält man bei inkohärenter Demodulation:

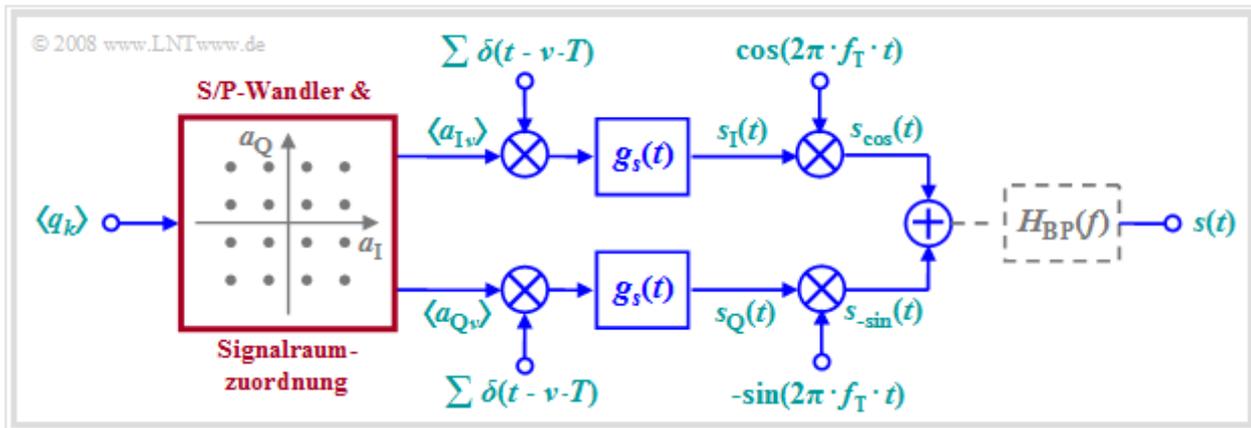
$$p_B = \frac{1}{2} \cdot e^{-E_B/(2N_0)}.$$

Beispielsweise benötigt man bei der BPSK das logarithmische Verhältnis $10 \cdot \lg E_B/N_0 \approx 8.4$ dB, um die Fehlerwahrscheinlichkeit $p_B = 10^{-4}$ zu erreichen. Bei der (differentiell-kohärenten) DPSK sind hierfür 9.3 dB, also fast ein dB mehr, erforderlich und bei der ASK 11.4 dB (kohärent) bzw. 12.3 dB (inkohärent).

Die hier angegebenen Gleichungen sollen in der **Aufgabe A4.7** ausgewertet werden.

Allgemeine Beschreibung und Signalraumzuordnung (1)

Aufgrund der Orthogonalität von Cosinus und (Minus-)Sinus kann man über einen Übertragungskanal zwei Datenströme unabhängig voneinander übertragen. Die Grafik zeigt das allgemeine Blockschaltbild.

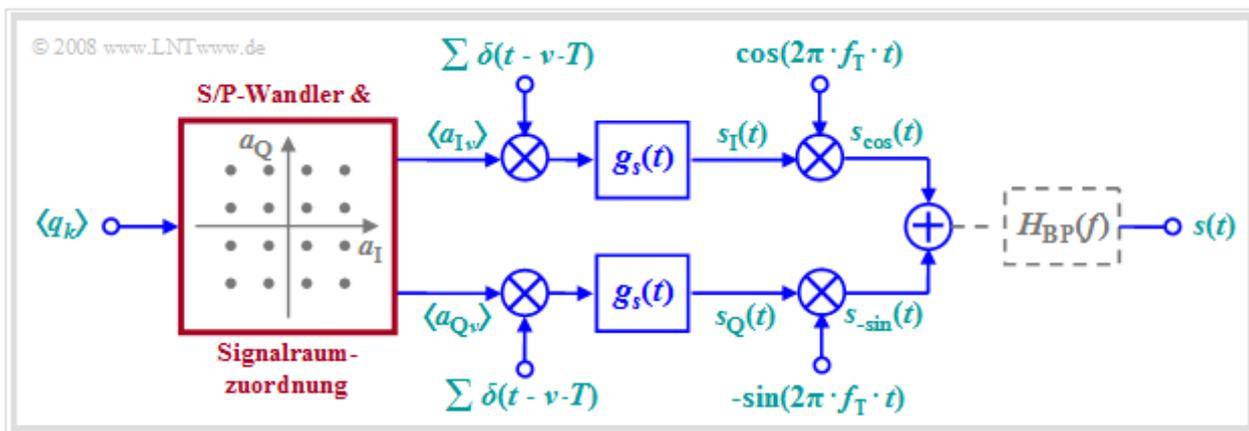


Dieses sehr allgemeine Modell lässt sich wie folgt beschreiben:

- Am Eingang liegt die binäre Quellensymbolfolge $\langle q_k \rangle$ mit der Bitrate R_B an. Der zeitliche Abstand zweier Symbole ist damit $T_B = 1/R_B$.
- Aus jeweils b binären Eingangssymbolen q_k werden zwei mehrstufige Amplitudenkoeffizienten $a_{I,v}$ und $a_{Q,v}$ abgeleitet, wobei „I“ für Inphase und „Q“ für Quadraturkomponente steht.
- Ist b geradzahlig und die Signalraumzuordnung quadratisch, so können die Koeffizienten $a_{I,v}$ und $a_{Q,v}$ jeweils einen von $M = 2^{b/2}$ Amplitudenwerten mit gleicher Wahrscheinlichkeit annehmen. Man spricht dann von **Quadratur-Amplitudenmodulation (QAM)**.
- Das in der Grafik betrachtete Beispiel gilt für die 16-QAM mit $b = M = 4$ und dementsprechend 16 Signalraumpunkten. Bei einer 256-QAM würde $b = 8$ und $M = 16$ gelten ($2^b = M^2 = 256$).

Die Bildbeschreibung wird auf der nächsten Seite fortgesetzt.

Allgemeine Beschreibung und Signalraumzuordnung (2)



Fortsetzung der Bildbeschreibung zur obigen Grafik:

- Anschließend werden die Koeffizienten $a_{I\nu}$ und $a_{Q\nu}$ jeweils einem Diracpuls als Impulsgewichte eingepreßt. Nach der Impulsformung mit dem Sendegrundimpuls $g_s(t)$ gilt somit in den beiden Zweigen des Blockschaltbildes:

$$s_I(t) = \sum_{\nu=-\infty}^{+\infty} a_{I\nu} \cdot g_s(t - \nu \cdot T),$$

$$s_Q(t) = \sum_{\nu=-\infty}^{+\infty} a_{Q\nu} \cdot g_s(t - \nu \cdot T).$$

- Anzumerken ist, dass wegen der redundanzfreien Umsetzung die **Symboldauer** T dieser Signale um den Faktor b größer ist als die Bitdauer T_B des binären Quellensignals. Im gezeichneten Beispiel (16-QAM) gilt $T = 4 \cdot T_B$.
- Das **QAM-Sendesignal** $s(t)$ ist dann die Summe der beiden mit Cosinus bzw. Minus-Sinus multiplizierten Teilsignale:

$$s_{\cos}(t) = s_I(t) \cdot \cos(2\pi f_T t),$$

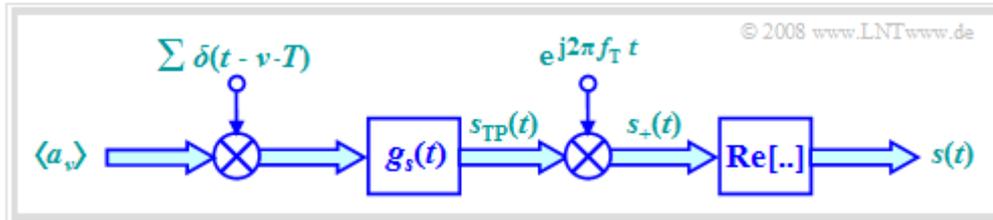
$$s_{-\sin}(t) = -s_Q(t) \cdot \sin(2\pi f_T t)$$

$$\Rightarrow s(t) = s_{\cos}(t) + s_{-\sin}(t) = s_I(t) \cdot \cos(2\pi f_T t) - s_Q(t) \cdot \sin(2\pi f_T t).$$

- Die beiden Übertragungszweige (I, Q) können als zwei völlig getrennte M -stufige ASK-Systeme aufgefasst werden, die sich gegenseitig nicht stören, solange alle Komponenten optimal ausgelegt sind. Die Quadratur-Amplitudenmodulation ermöglicht somit (im Idealfall) eine Verdoppelung der Datenrate bei gleichbleibender Qualität.

Systembeschreibung durch das äquivalente TP-Signal

Da die Multiplikation von $s_I(t)$ und $s_Q(t)$ mit einer Cosinus- bzw. Minus-Sinus-Schwingung nur eine Verschiebung im Frequenzbereich bewirkt und eine solche Verschiebung eine lineare Operation darstellt, lässt sich die Systembeschreibung mit Hilfe der **äquivalenten TP-Signale** wesentlich vereinfachen.



Die Grafik zeigt das vereinfachte Modell im Basisband. Dieses ist äquivalent zum bisher betrachteten **Blockschaltbild**. Beachten Sie bitte die folgenden Hinweise:

- Die im Blockschaltbild rot gezeichnete Seriell-Parallel-Wandlung und die Signalraumzuordnung bleibt erhalten, obwohl dieser Block hier nicht mehr eingezeichnet ist. Lassen wir zunächst auch den oft aus schaltungstechnischen Gründen eingebrachten Bandpass $H_{BP}(f)$ außer Betracht.
- Alle Doppelpfeile in dem obigen Basisbandmodell kennzeichnen komplexe Größen. Die damit verbundenen Operationen sind ebenfalls komplex zu verstehen. Beispielsweise fasst der komplexe Amplitudenkoeffizient a_ν je einen Inphase- und einen Quadraturkoeffizienten zusammen:

$$a_\nu = a_{I\nu} + j \cdot a_{Q\nu}.$$

- Die äquivalente Tiefpass-Repräsentation des tatsächlichen, physikalischen und damit per se reellen Sendesignals $s(t)$ ist bei QAM stets komplex und es gilt mit den Teilsignalen $s_I(t)$ und $s_Q(t)$:

$$s_{TP}(t) = s_I(t) + j \cdot s_Q(t) = \sum_{\nu=-\infty}^{+\infty} a_\nu \cdot g_s(t - \nu \cdot T).$$

- Zum **analytischen Signal** $s_+(t)$ kommt man von $s_{TP}(t)$ durch Multiplikation mit der komplexen Exponentialfunktion. Das physikalische Sendesignal $s(t)$ ergibt sich dann als der Realteil von $s_+(t)$.
- Damit die Vorzeichen im Blockschaltbild der letzten Seite und im skizzierten Basisbandmodell übereinstimmen, ist im Quadraturzweig die Multiplikation mit der negativen Sinus-Schwingung erforderlich, wie die nachfolgende Rechnung zeigt:

$$\begin{aligned} s(t) &= \text{Re}[s_+(t)] = \text{Re}[s_{TP}(t) \cdot e^{j2\pi f_T t}] = \\ &= \text{Re} \left[\left(\sum (a_{I\nu} + j \cdot a_{Q\nu}) \cdot g_s(t - \nu \cdot T) \right) (\cos(2\pi f_T t) + j \cdot \sin(2\pi f_T t)) \right] = \\ &= s_I(t) \cdot \cos(2\pi f_T t) - s_Q(t) \cdot \sin(2\pi f_T t). \end{aligned}$$

- Der Einfluss des Bandpasses $H_{BP}(f)$, der in der Praxis oft am Ausgang des QAM-Modulators zu berücksichtigen ist, kann dem Impulsformfilter $g_s(t)$ beaufschlagt werden. Ist der Durchlassbereich des BP-Filters symmetrisch um f_T , so ist sein Tiefpass-Äquivalent (im Zeitbereich) $h_{BP \rightarrow TP}(t)$ rein reell und man kann im Modell $g_s(t)$ durch $g_s(t) * h_{BP \rightarrow TP}(t)$ ersetzen.

Leistung und Energie von QAM-Signalen

Wie im **Kapitel 4.3** von „Signaldarstellung“ gezeigt wird, kann die **Leistung** des QAM-Sendesignals $s(t)$ auch aus dem äquivalenten TP-Signal $s_{TP}(t)$ berechnet werden, das stets komplex ist:

$$P = \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \cdot \int_{-T_M/2}^{+T_M/2} s(t)^2 dt = \frac{1}{2} \cdot \lim_{T_M \rightarrow \infty} \frac{1}{T_M} \cdot \int_{-T_M/2}^{+T_M/2} |s_{TP}(t)|^2 dt.$$

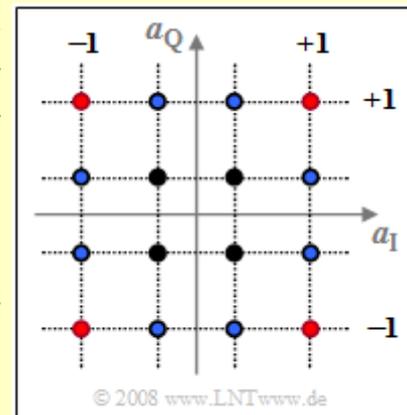
Dagegen ist die Energie der Signale $s(t)$ und $s_{TP}(t)$ unendlich groß. Beschränkt man sich jedoch auf eine Symboldauer T , so erhält man die **Energie pro Symbol**

$$E_S = \frac{E[|a_\nu|^2]}{2} \cdot \int_{-\infty}^{+\infty} |g_s(t)|^2 dt = \frac{E[|a_\nu|^2]}{2} \cdot \int_{-\infty}^{+\infty} |G_s(f)|^2 df.$$

Dagegen gibt $E_B = E_S/b$ die **Energie pro Bit** an, wenn gemäß der gegebenen Signalraumzuordnung jeweils b Binärsymbole zu einem komplexen Amplitudenkoeffizienten zusammengefasst werden.

Beispiel: Die obere Grafik zeigt die Signalraumzuordnung bei **16-QAM**, wobei sowohl der Real- als auch der Imaginärteil der komplexen Amplitudenkoeffizienten a_ν jeweils einen von vier Werten (± 1 sowie $\pm 1/3$) annehmen kann. Durch Mittelung über die 16 Abstandsquadrate zum Ursprung erhält man:

$$E[|a_\nu|^2] = \frac{4}{16} \cdot (1^2 + 1^2) + \frac{4}{16} \cdot [(1/3)^2 + (1/3)^2] + \frac{8}{16} \cdot [1^2 + (1/3)^2] = \dots = \frac{10}{9} \approx 1.11.$$



Die Summanden gehören in dieser Reihenfolge zu den vier roten, den vier schwarzen und den acht blauen Punkten.

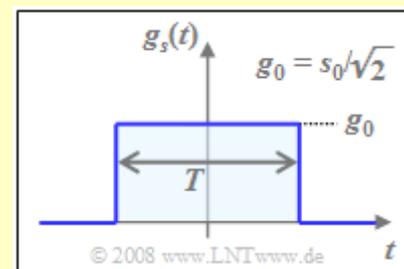
Bei NRZ-rechteckförmigem Sendegrundimpuls $g_s(t)$ mit der Amplitude g_0 und der Symboldauer T ist das Spektrum $G_s(f)$ si-förmig. In diesem Fall gilt für

- die mittlere Energie pro Symbol:

$$E_S = \frac{E[|a_\nu|^2]}{2} \cdot g_0^2 \cdot T = \frac{5}{9} \cdot g_0^2 \cdot T,$$

- die mittlere Energie pro Bit:

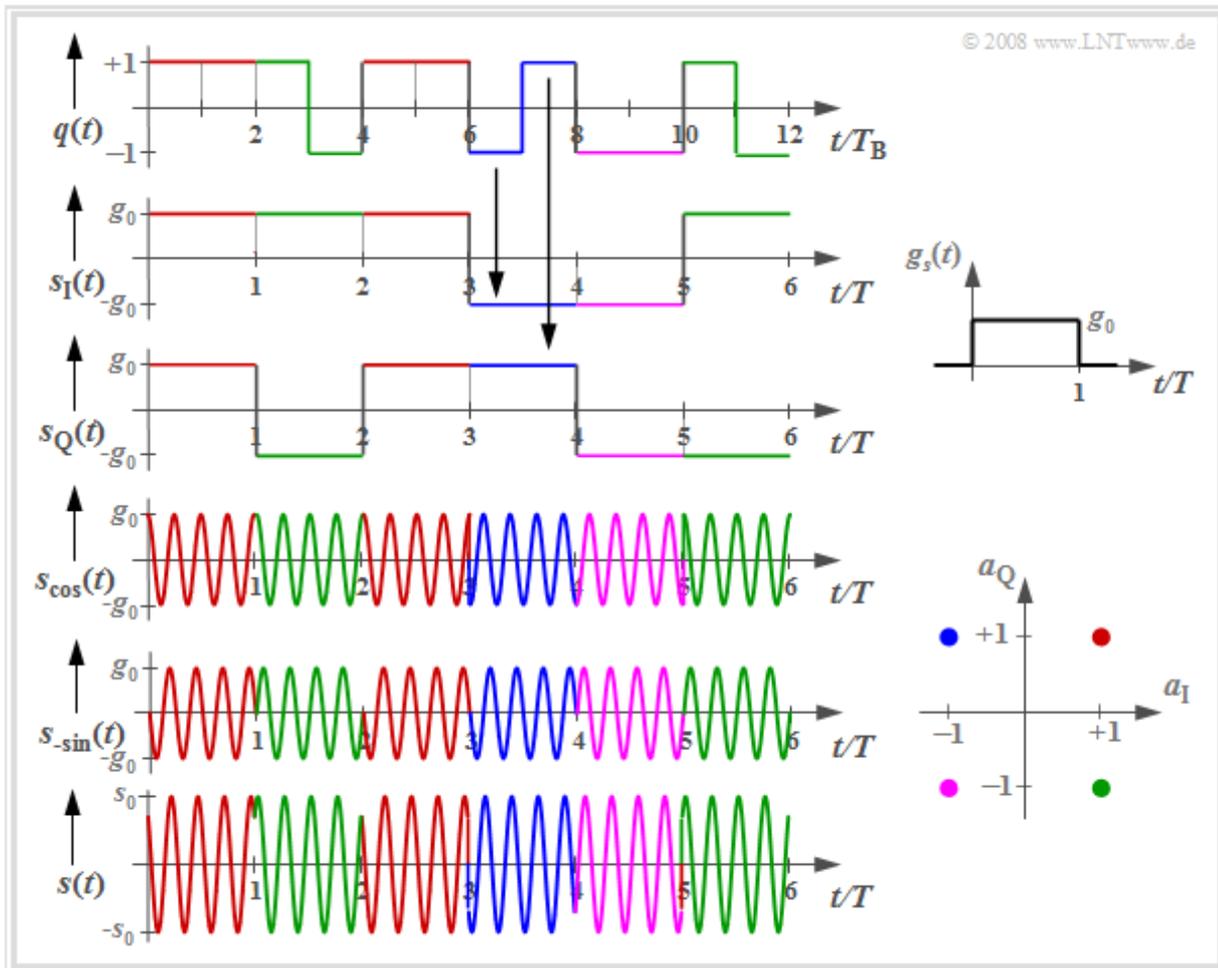
$$E_B = \frac{E_S}{4} \approx 0.139 \cdot g_0^2 \cdot T.$$



Die maximale Hüllkurve s_0 ist um den Faktor „Wurzel aus 2“ größer als die Amplitude g_0 des Rechteckimpulses (siehe untere Skizze) und tritt bei einem der roten Amplitudenkoeffizienten auf, also immer dann, wenn $|a_{I\nu}| = |a_{Q\nu}| = 1$ ist.

Signalverläufe der 4-QAM

Die folgende Grafik zeigt die Signalverläufe der 4-QAM, wobei die Farbgebung mit der eingezeichneten Signalraumzuordnung übereinstimmt.



Man erkennt aus diesen Darstellungen:

- die Seriell-Parallel-Wandlung des Quellensignals $q(t)$ in die beiden Komponentensignale $s_I(t)$ und $s_Q(t)$, jeweils mit der Symboldauer $T = 2T_B$ und den Signalwerten $\pm g_0$ (T_B ist die Bitdauer);
- die beiden trägerfrequenzmodulierten Signale $s_{\cos}(t)$ und $s_{\sin}(t)$ mit Phasensprüngen um $\pm\pi$:

$$s_{\cos}(t) = s_I(t) \cdot \cos(2\pi f_T t), \quad s_{\sin}(t) = -s_Q(t) \cdot \sin(2\pi f_T t),$$

- das Sendesignal $s(t) = s_{\cos}(t) - s_{\sin}(t)$ mit Phasensprüngen um Vielfache von $\pm\pi/2$; deren Hüllkurve ist gegenüber den beiden Komponentensignalen um den Faktor „Wurzel aus 2“ größer:

$$s_0 = \sqrt{2} \cdot g_0.$$

Anzumerken ist, dass hier der Sendegrundimpuls $g_s(t)$ zur Vereinfachung der Darstellung im Bereich von 0 bis T als rechteckförmig (also unsymmetrisch bezüglich $t = 0$) angenommen wurde. Die zugehörige Spektralfunktion $G_s(f)$ dieses kausalen Impulses ist komplex, was jedoch in diesem Zusammenhang keine Auswirkungen hat.

Fehlerwahrscheinlichkeit der 4-QAM

Im **Kapitel 4.2** wurde die Fehlerwahrscheinlichkeit der BPSK angegeben. Nun werden die Ergebnisse auf die 4-QAM übertragen, wobei weiterhin folgende Voraussetzungen gelten:

- ein Sendesignal mit der mittleren Energie E_B pro Bit,
- AWGN-Rauschen mit der Rauschleistungsdichte N_0 ,
- bestmögliche Empfängerrealisierung nach dem Matched-Filter-Prinzip.

Die obere Abbildung zeigt das BPSK-Phasendiagramm des Detektionssignals $d(t)$, also inklusive dem Matched-Filter. Der Abstand des Nutzsignals von der Schwelle (d_Q -Achse) beträgt zu den Detektionszeitpunkten jeweils s_0 . Mit den weiteren Gleichungen

$$p_B = Q\left(\frac{s_0}{\sigma_d}\right), \quad E_B = \frac{1}{2} \cdot s_0^2 \cdot T_B, \quad \sigma_d^2 = \frac{N_0}{T_B}$$

erhält man für die **BPSK-Fehlerwahrscheinlichkeit**:

$$p_{B, \text{BPSK}} = Q\left(\sqrt{2 \cdot E_B/N_0}\right) = \frac{1}{2} \cdot \text{erfc}\left(\sqrt{E_B/N_0}\right).$$

Bei der 4-QAM entsprechend der unteren Abbildung

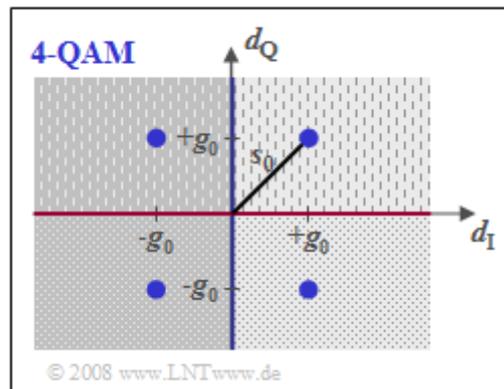
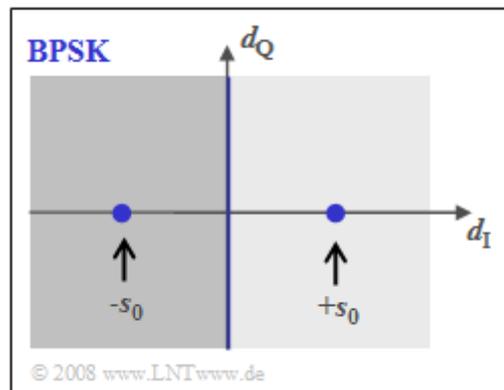
- gibt es nun zwei Schwellen zwischen den Bereichen mit hellerem/dunklerem Hintergrund (blaue Linie) sowie zwischen den gepunkteten/gestrichelten Flächen (rot),
- ist der Abstand von den Schwellen jeweils nur noch g_0 anstelle von s_0 ,
- ist aber die Rauschleistung wegen der halb so großen Symbolrate in jedem Teilzweig gegenüber der BPSK auch nur halb so groß.

Mit den Gleichungen

$$p_B = Q\left(\frac{g_0}{\sigma_d}\right), \quad g_0 = \frac{s_0}{\sqrt{2}}, \quad E_B = \frac{1}{2} \cdot s_0^2 \cdot T_B, \quad \sigma_d^2 = \frac{N_0}{2 \cdot T_B}$$

erhält man für die **4-QAM-Fehlerwahrscheinlichkeit** genau das gleiche Ergebnis wie für die BPSK:

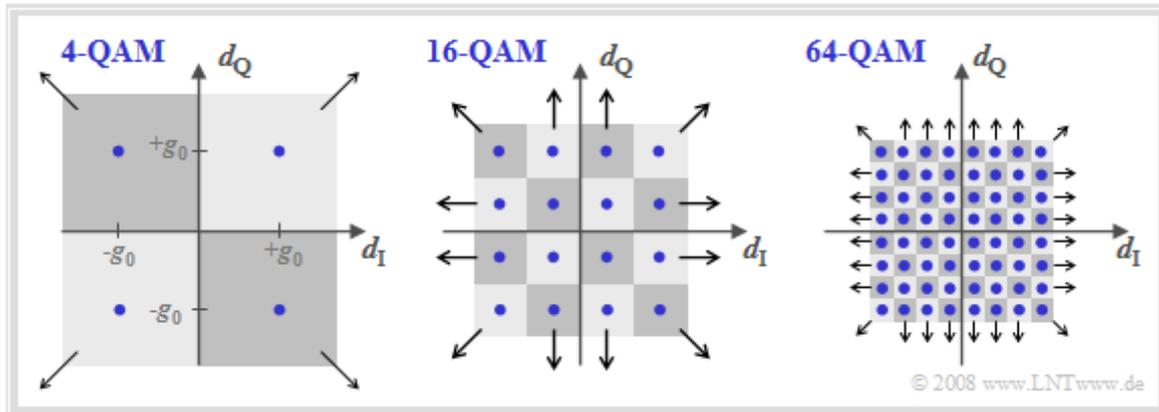
$$p_{B, 4\text{-QAM}} = Q\left(\sqrt{2 \cdot E_B/N_0}\right) = \frac{1}{2} \cdot \text{erfc}\left(\sqrt{E_B/N_0}\right).$$



Fazit: Die 4-QAM weist bei idealen Bedingungen die gleiche Fehlerwahrscheinlichkeit wie die BPSK auf, obwohl die doppelte Informationsmenge übertragen werden kann. Sind allerdings die Bedingungen nicht mehr ideal – zum Beispiel bei einem ungewollten Phasenversatz zwischen Sender und Empfänger – so gibt es bei der 4-QAM eine deutlich stärkere Degradation als bei der BPSK. Dieser Fall wird im **Kapitel 1.5** des Buches „Digitalsignalübertragung“ noch genauer betrachtet.

QAM-Signalraumkonstellationen

Die nachfolgende Grafik zeigt die Signalraumkonstellationen von 4-QAM, 16-QAM und 64-QAM. Mit den hier gewählten Achsenbeschriftungen beschreiben die Bilder auch das Detektionsnutzsignal (zu den Detektionszeitpunkten) im äquivalenten Tiefpassbereich. Ebenfalls eingezeichnet sind die verschiedenen Entscheidungsgebiete, die dem verrauschten Detektionssignal zugeordnet werden. Die Pfeile geben an, wenn Entscheidungsgebiete bis ins Unendliche ausgedehnt sind.

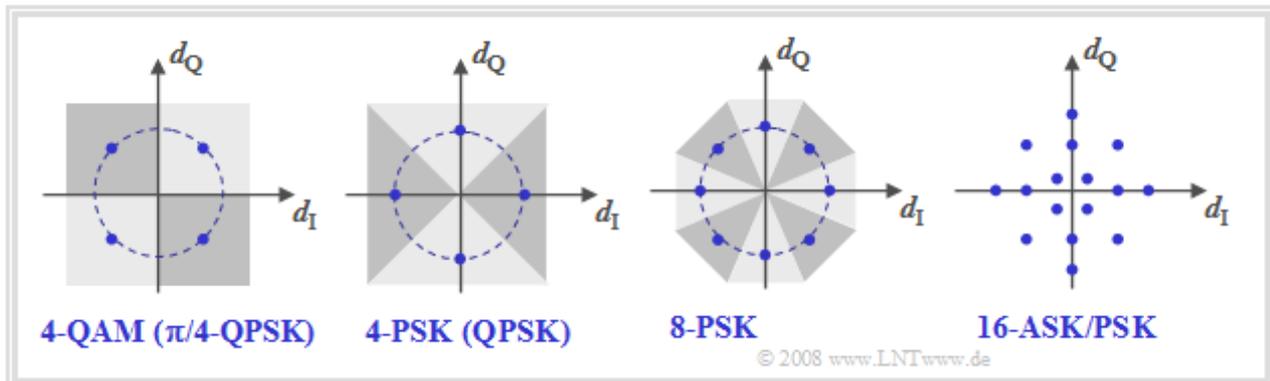


Zu diesen Darstellungen ist anzumerken:

- Die Bilder beziehen sich nur auf die Detektionszeitpunkte und gelten für alle Nyquistsysteme wie die **Rechteck-Rechteck-Konfiguration** oder ein **Wurzel-Cosinus-Rolloff-Nyquistsystem**.
- Die hier nicht dargestellten Übergänge außerhalb der Detektionszeitpunkte zwischen den einzelnen Punkten hängen dagegen sehr wohl vom gewählten Nyquistsystem ab.
- Bei echter QAM-Struktur – das heißt: die Signalraumkonstellation ist quadratisch oder zumindest rechteckförmig – lässt sich die 2D-Detektion durch zwei „eindimensionale“ Detektionsvorgänge vereinfacht lösen.
- Die 16-QAM ist somit nichts anderes als die parallele Übertragung zweier Digitalsignale mit jeweils $M = 4$ Amplitudenstufen. Bei der 64-QAM würde entsprechend $M = 8$ gelten und bei der 256-QAM ist die „eindimensionale“ Stufenzahl $M = 16$.
- Alle im **Kapitel 2.2** von Buch „Digitalsignalübertragung“ genannten Eigenschaften für mehrstufige Signale gelten auch hier, wobei allerdings die Zusammensetzung der orthogonalen Trägerfrequenzsignale noch geeignet zu berücksichtigen ist.

Weitere Signalraumkonstellationen

Die Grafik zeigt weitere Signalraumkonstellationen, wobei links wieder die 4-QAM nach der bisherigen Beschreibung dargestellt ist.



Die Konstellation rechts daneben kennzeichnet eine vierstufige Phasenmodulation (4-PSK bzw. QPSK – *Quaternary Phase Shift Keying*). Ein Vergleich der beiden linken Diagramme zeigt:

- Die hier als QPSK bezeichnete Variante verwendet die Phasenlagen 0° , 90° , 180° und 270° . Man erkennt aus den eingezeichneten Entscheidungsgebieten, dass hier die Detektion nicht auf zwei Binärentscheidungen zurückgeführt werden kann.
- Auch die 4-QAM kann als eine vierstufige Phasenmodulation mit den möglichen Phasenlagen $\pm 45^\circ$ und $\pm 135^\circ$ aufgefasst werden. Gegenüber dem zweiten Phasendiagramm ergibt sich eine Drehung um 45° ($\pi/4$), so dass die 4-QAM oft auch als $\pi/4$ -QPSK bezeichnet wird.
- Ähnlich wie man bei der BPSK durch Vorcodierung zur **DPSK** kommt, kann auch die 4-PSK zur 4-DPSK erweitert und dadurch die Demodulation erleichtert werden. Diese wird zum Beispiel bei der Datenübertragung über Telefonkanäle gemäß der CCITT-Empfehlung V26 angewendet (Trägerfrequenz 1800 Hz, Datenrate 2400 bit/s).

Die beiden rechten Diagramme zeigen höherstufige Modulationsverfahren:

- Die 8-PSK bzw. 8-DPSK erlaubt entsprechend der CCITT-Empfehlung V27 beim Telefonkanal eine Datenrate von bis zu 4800 bit/s.
- Die Empfehlung V29 sieht mit der 16-ASK/PSK eine hybride Modulationsform vor, die bei ausgewählten, fest verschalteten Leitungen eine Datenrate von 9600 bit/s ermöglicht.

Nyquist- und Wurzel-Nyquist-QAM-Systeme (1)

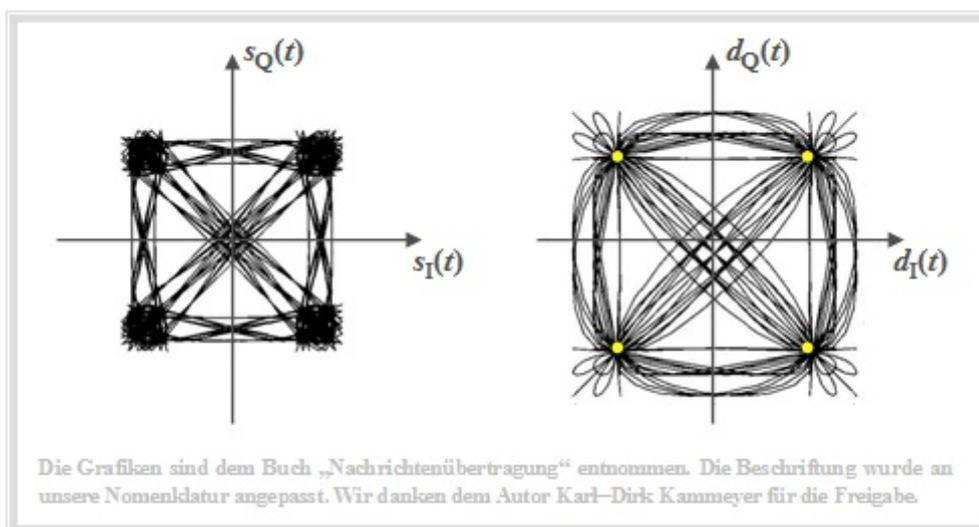
Bisher wurde im Kapitel 4.3 aus Darstellungsgründen stets vom rechteckförmigen Sendegrundimpuls ausgegangen. In der Praxis verwendet man aber meist eine Wurzel-Nyquist-Charakteristik entsprechend dem **Kapitel 1.4** im Buch „Digitalsignalübertragung“. In aller Kürze lassen sich diese Systeme wie folgt darstellen:

- Der Empfängerfrequenzgang $H_E(f)$ wird hier formgleich mit dem normierten Sendepulspektrum $H_S(f)$ gewählt, was unter der Nebenbedingung der Leistungsbegrenzung (das heißt: konstante mittlere Sendeleistung) zur kleinsten Fehlerwahrscheinlichkeit führt.
- Der Gesamtfrequenzgang $H_{Nyq}(f) = H_S(f) \cdot H_E(f)$ erfüllt das **erste Nyquistkriterium**, so dass es beim Empfänger zu keinen Impulsinterferenzen kommt. $H_S(f)$ und $H_E(f)$ haben somit jeweils eine Wurzel-Nyquist-Charakteristik.
- Für den Frequenzgang $H_{Nyq}(f)$ verwendet man einen **Cosinus-Rolloff-Tiefpass** $H_{CRO}(f)$ mit der äquivalenten Bandbreite $\Delta f_{CRO} = 1/T$ und frei wählbarem Rolloff-Faktor ($0 \leq r \leq 1$).

Mit dem nachfolgenden Interaktionsmodul können Sie sich den Frequenzgang und die Impulsantwort dieses Tiefpasses verdeutlichen:

Frequenzgang und zugehörige Impulsantwort

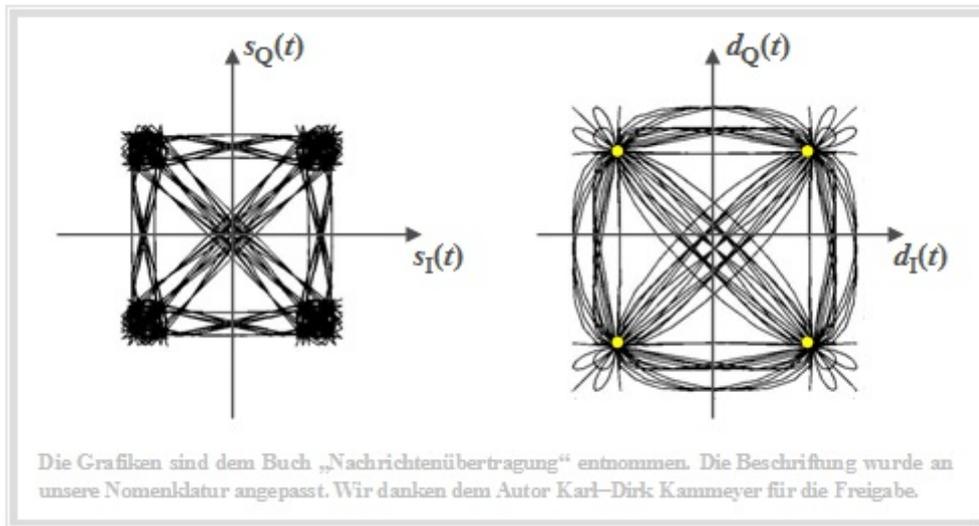
Der Vorteil dieser **Wurzel-Nyquist-Systeme** ist die deutlich kleinere Bandbreite $(1 + r)/T$ gegenüber der bisher betrachteten Konfiguration mit rechteckförmigem $g_s(t)$ und rechteckförmigem $h_E(t)$, dessen Spektrum (theoretisch) unendlich weit ausgedehnt ist. Hinsichtlich **Fehlerwahrscheinlichkeit** ändert sich gegenüber der Rechteck-Rechteck-Konfiguration nichts, da der Grundimpuls $g_d(t)$ vor dem Entscheider äquidistante Nulldurchgänge aufweist und somit Impulsinterferenzen vermieden werden.



Die Grafik zeigt die Phasendiagramme für diesen Fall, die dem Buch **[Kam04]** entnommen sind. Der Rolloff-Faktor beträgt $r = 0.5$. Die Bildbeschreibung folgt auf der nächsten Seite.

Nyquist- und Wurzel-Nyquist-QAM-Systeme (2)

Die Grafik zeigt die Phasendiagramme für diesen Fall, die dem Buch [Kam04] entnommen sind. Der Roloff-Faktor beträgt $r = 0.5$.



Man erkennt aus diesen Darstellungen:

- Das rechte Diagramm zeigt in 2D-Darstellung die Detektionssignale im I- und Q-Zweig nach den Wurzel-Nyquist-Empfangsfiltern. Die dazugehörigen Spektren besitzen jeweils cosinusförmige Flanken um die Nyquistfrequenz $f_{\text{Nyq}} = 1/(2T)$.
- Zu den Detektionszeitpunkten sind im Phasendiagramm nur die vier gelb eingezeichneten Punkte möglich. Die Übergänge dazwischen sind vielfältig.
- Links dargestellt sind die beiden Sendesignale im äquivalenten Tiefpassbereich, $s_I(t) = \text{Re}[s_{\text{TP}}(t)]$ und $s_Q(t) = \text{Im}[s_{\text{TP}}(t)]$. Aufgrund der Wurzel-Nyquist-Spektralformung gibt es beim Sender Impulsinterferenzen, das heißt, dass das äquivalente Tiefpass-Signal $s_{\text{TP}}(t) = s_I(t) + j \cdot s_Q(t)$ auch zu den Detektionszeitpunkten nicht auf vier Punkte beschränkt ist.
- Der Betrag $|s_{\text{TP}}(t)|$ – also der Abstand vom Nullpunkt – gibt die Hüllkurve des 4-QAM-Signals an. Man erkennt aus dem linken Diagramm deutlich, dass es insbesondere bei Phasenänderungen um π starke Amplitudeneinbrüche gibt, da $s_{\text{TP}}(t)$ dann auch (komplexe) Werte nahe 0 annimmt.

Offset-Quadraturamplitudenmodulation

Ausgehend von den Gleichungen für die 4-QAM (bzw. $\pi/4$ -QPSK) kommt man zur **Offset-4-QAM**, die wir vereinfachend mit Offset-QPSK (kurz: O-QPSK) bezeichnen. Für das Sendesignal gilt hier:

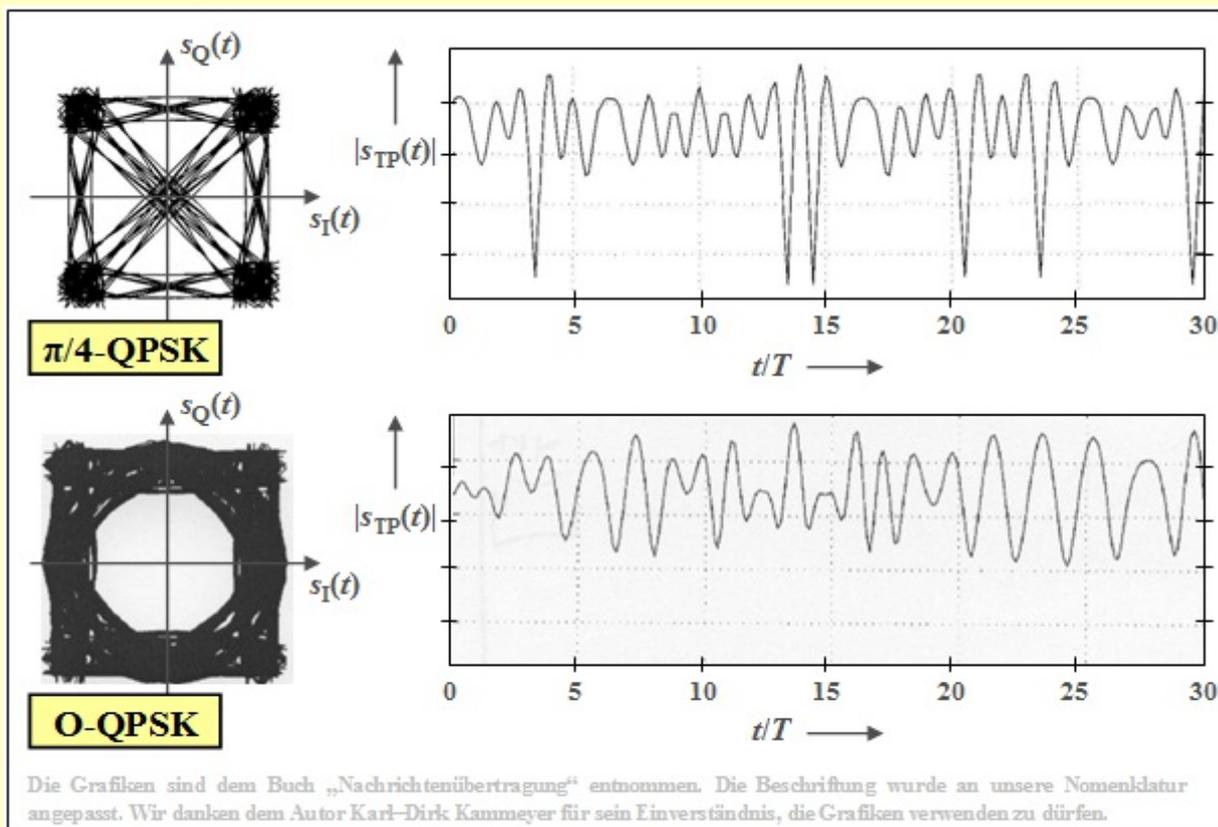
$$s(t) = s_I(t) \cdot \cos(2\pi f_T t) - s_Q(t) \cdot \sin(2\pi f_T t),$$

$$\text{mit } s_I(t) = \sum_{\nu=-\infty}^{+\infty} a_{I\nu} \cdot g_s(t - \nu \cdot T), \quad s_Q(t) = \sum_{\nu=-\infty}^{+\infty} a_{Q\nu} \cdot g_s(t - T/2 - \nu \cdot T).$$

Der einzige, aber entscheidende Unterschied ist die zeitliche Verschiebung der Quadraturkomponente gegenüber der Inphasekomponente um eine halbe Symboldauer ($T/2$). Dies hat den Vorteil,

- dass die Phasenfunktion nicht durch 0 geht, und
- dass dadurch die Hüllkurve $|s_{TP}(t)|$ auch deutlich weniger schwankt.

Beispiel: Die Grafik zeigt oben nochmals das Phasendiagramm für die 4-QAM ($\pi/4$ -QPSK) und rechts daneben einen beispielhaften Verlauf der Hüllkurve, wobei wie auf der letzten Seite ein Wurzel-Nyquist-Sendespektrum mit dem Rolloff-Faktor $r = 0.5$ zugrunde liegt.



Die beiden unteren Bilder zeigen, dass die Offset-QPSK hinsichtlich der Einhüllenden deutlich bessere Eigenschaften (weniger gravierende Signaleinbrüche) besitzt.

Alle Eigenschaften von 4-QAM/QPSK und Offset-QPSK können Sie sich mit einem Interaktionsmodul verdeutlichen, wobei als Grundimpuls ein Rechteck, ein cosinusförmiger Verlauf, ein Nyquistimpuls oder ein Wurzel-Nyquistimpuls wählbar ist:

Quaternary Phase Shift Keying (QPSK und O-QPSK)

Eigenschaften nichtlinearer Verfahren (1)

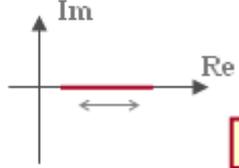
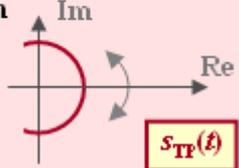
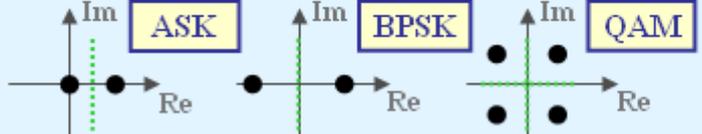
Die Gesamtheit aller Modulationsverfahren lassen sich alternativ wie folgt klassifizieren:

- Amplituden-, Phasen- und Frequenzmodulation (AM, PM, FM),
- analoge und digitale Modulationsverfahren,
- lineare und nichtlineare Modulationsverfahren.

Hinsichtlich des letzten Unterscheidungsmerkmals soll gelten **[Klo01]**:

Definition: Ein **lineares Modulationsverfahren** liegt vor, wenn eine beliebige Linearkombination von Signalen am Modulatoreingang zu einer entsprechenden Linearkombination an dessen Ausgang führt. Andernfalls spricht man von einem nichtlinearen Modulationsverfahren.

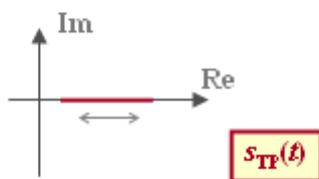
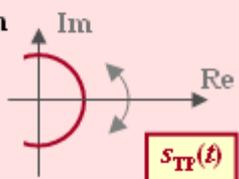
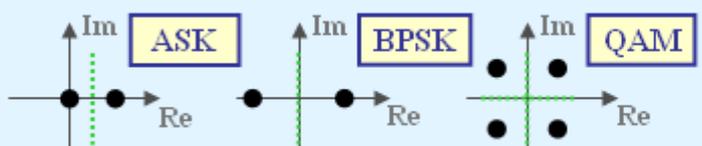
Die Grafik zeigt einige der Unterschiede hinsichtlich der oben angegebenen Klassifizierungen.

	AM	PM	FM
Analoge Modulation $q(t)$ ist ein Analogsignal	<ul style="list-style-type: none"> • Lineare Modulation • Ortskurve: Horizontale 	<ul style="list-style-type: none"> • Viele Gemeinsamkeiten zwischen PM und FM ⇒ Winkelmodulation (WM) • Nichtlineare Modulationsverfahren • Ortskurve ist ein Kreisbogen • Bessel-Spektrum, falls $q(t)$ nur eine Frequenz beinhaltet 	
Digitale Modulation $q(t)$ ist ein Digitalsignal	ASK (OOK)	BPSK, 4-QAM	FSK
	<ul style="list-style-type: none"> • Gemeinsamkeiten zwischen ASK, BPSK und 4-QAM • Lineare Modulationsverfahren • Ortskurve besteht aus Punkten 		<ul style="list-style-type: none"> • Nichtlineare Modulation <li style="border: 1px solid blue; padding: 2px;">$e^{j \cdot (a+b)} \neq e^{j \cdot a} + e^{j \cdot b}$ • schwieriger handhabbar • Spektraldarstellung ist oft kompliziert

© 2008 www.LNTwww.de

Zu **Beginn** von Kapitel 4 wurde bereits darauf hingewiesen, dass der wesentliche Unterschied zwischen einem analogen und einem digitalen Modulationsverfahren darin besteht, dass beim ersten ein analoges Quellensignal $q(t)$ anliegt und beim zweiten ein Digitalsignal. Bei genauerer Betrachtung wird man jedoch feststellen, dass es zwischen diesen Verfahren noch einige Unterschiede mehr gibt. Darauf wird auf der nächsten Seite genauer eingegangen.

Eigenschaften nichtlinearer Verfahren (2)

	AM	PM	FM
Analoge Modulation $q(t)$ ist ein Analogsignal	<ul style="list-style-type: none"> • Lineare Modulation • Ortskurve: Horizontale 	<ul style="list-style-type: none"> • Viele Gemeinsamkeiten zwischen PM und FM \Rightarrow Winkelmodulation (WM) • Nichtlineare Modulationsverfahren • Ortskurve ist ein Kreisbogen • Bessel-Spektrum, falls $q(t)$ nur eine Frequenz beinhaltet 	
Digitale Modulation $q(t)$ ist ein Digitalsignal	ASK (OOK)	BPSK, 4-QAM	FSK
	<ul style="list-style-type: none"> • Gemeinsamkeiten zwischen ASK, BPSK und 4-QAM • Lineare Modulationsverfahren • Ortskurve besteht aus Punkten 		<ul style="list-style-type: none"> • Nichtlineare Modulation <div style="border: 1px solid blue; padding: 5px; margin: 5px 0;"> $e^{j \cdot (a+b)} \neq e^{j \cdot a} + e^{j \cdot b}$ </div> <ul style="list-style-type: none"> • schwieriger handhabbar • Spektraldarstellung ist oft kompliziert

© 2008 www.LNTwww.de

- Die **analoge Amplitudenmodulation** ist ein lineares Modulationsverfahren. Die Ortskurve – also das äquivalente Tiefpass-Signal dargestellt in der komplexen Ebene – ist eine Gerade.
- Zwischen analoger PM und FM gibt es viele Gemeinsamkeiten \Rightarrow gemeinsame Beschreibung als **Winkelmodulation**. Die Ortskurve ist ein Kreisbogen. Bei einer harmonischen Schwingung gibt es ein Linienspektrum $S(f)$ bei Vielfachen der Nachrichtenfrequenz f_N um die Trägerfrequenz.
- Die **digitale Amplitudenmodulation**, die man entweder als *Amplitude Shift Keying* (ASK) oder als *On-Off-Keying* (OOK) bezeichnet, ist ebenfalls linear. Die Ortskurve besteht im binären Fall nur noch aus zwei Punkten.
- Da sich die **binäre Phasenmodulation** (BPSK) als ASK mit bipolaren Amplitudenkoeffizienten darstellen lässt, ist diese ebenfalls linear. Die Form des BPSK-Leistungsdichtespektrums wird wesentlich durch das Betragsquadratspektrum $|G_s(f)|^2$ des Sendegrundimpulses bestimmt.
- Das bedeutet aber auch: Die Spektralfunktion der BPSK ist kontinuierlich in f . Würde man die BPSK als (analoge) Phasenmodulation mit digitalem Quellensignal $q(t)$ betrachten, so müssten zur Berechnung von $\Phi_s(f)$ unendlich viele Bessel-Linienspektren miteinander gefaltet werden, wenn man $Q(f)$ als unendliche Summe von Einzelfrequenzen darstellt.
- Da die **4-QAM** auch als Summe zweier zueinander orthogonaler und damit quasi-unabhängiger BPSK-Systeme beschrieben werden kann, stellt auch diese ein lineares Modulationsverfahren dar. Gleiches gilt für die höherstufigen QAM-Verfahren wie 16-QAM, 64-QAM usw..
- Eine höherstufige PSK, zum Beispiel die 8-PSK, ist nur in Sonderfällen linear, siehe **[Klo01]**. Die digitale Frequenzmodulation (*Frequency Shift Keying*, FSK) ist dagegen stets nichtlinear. Dieses Verfahren wird nachfolgend beschrieben, wobei wir uns auf die binäre FSK beschränken.

FSK – Frequency Shift Keying (1)

Wir gehen hier aus vom Sendesignal der analogen Frequenzmodulation aus,

$$s(t) = s_0 \cdot \cos(\psi(t)) \quad \text{mit} \quad \psi(t) = 2\pi f_T t + K_{\text{FM}} \cdot \int q(t) dt$$

und dem rechteckförmigen Binärsignal mit $a_\nu \in \{+1, -1\} \Rightarrow$ bipolare Signalisierung:

$$q(t) = \sum_{\nu=-\infty}^{+\infty} a_\nu \cdot g_s(t - \nu \cdot T) \quad \text{mit} \quad g_s(t) = \begin{cases} A & \text{für } 0 < t < T, \\ 0 & \text{für sonst.} \end{cases}$$

Fasst man die Amplitude A und die Modulatorkonstante K_{FM} zum Frequenzhub (Definition siehe unten)

$$\Delta f_A = \frac{A \cdot K_{\text{FM}}}{2\pi}$$

zusammen, so lautet das **FSK-Sendesignal** im ν -ten Zeitintervall:

$$s(t) = s_0 \cdot \cos(2\pi \cdot t \cdot (f_T + a_\nu \cdot \Delta f_A)).$$

Dieses lässt sich mit den beiden möglichen Signalfrequenzen

$$f_{+1} = f_T + \Delta f_A, \quad f_{-1} = f_T - \Delta f_A$$

auch in folgender Form schreiben:

$$s(t) = \begin{cases} s_0 \cdot \cos(2\pi \cdot f_{+1} \cdot t) & \text{für } a_\nu = +1, \\ s_0 \cdot \cos(2\pi \cdot f_{-1} \cdot t) & \text{für } a_\nu = -1. \end{cases}$$

Zu jedem Zeitpunkt tritt also stets nur eine der beiden Frequenzen f_{+1} und f_{-1} auf. Die Trägerfrequenz f_T selbst kommt im Signal nicht vor.

Definition: Der **Frequenzhub** Δf_A ist in gleicher Weise definiert wie bei der analogen FM, nämlich als die maximale Abweichung der Augenblicksfrequenz $f_A(t)$ von der Trägerfrequenz f_T . Häufig wird der Frequenzhub in der Literatur auch mit Δf bzw. F bezeichnet.

Eine weitere wichtige Beschreibungsgröße ist in diesem Zusammenhang der Modulationsindex, der ebenfalls bereits bei der **analogen Frequenzmodulation** als $\eta = \Delta f_A / f_N$ definiert wurde. Bei der FSK ist eine etwas andere Definition erforderlich, was durch einen anderen Kennbuchstaben berücksichtigt wird: $\eta \Rightarrow h$.

Definition: Bei der digitalen Frequenzmodulation (FSK) bezeichnet man als den **Modulationsindex** h das Verhältnis aus dem Gesamtfrequenzhub und der Symbolrate $1/T$:

$$h = \frac{2 \cdot \Delta f_A}{1/T} = 2 \cdot \Delta f_A \cdot T.$$

Manchmal wird in der Fachliteratur h auch als **Phasenhub** bezeichnet.

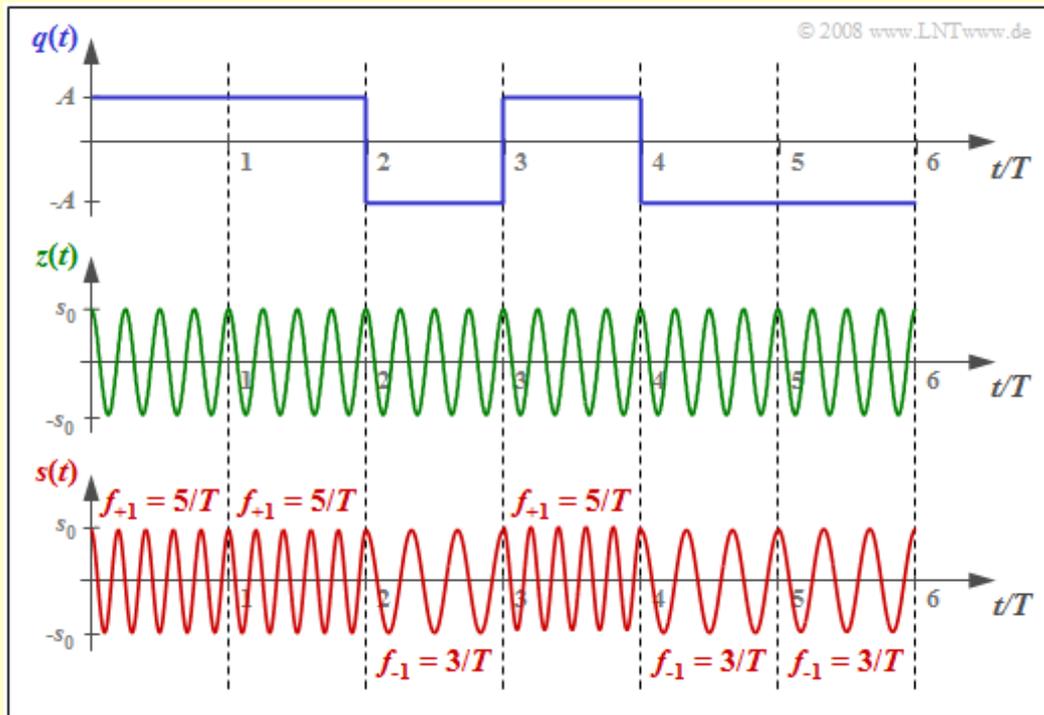
FSK – Frequency Shift Keying (2)

Beispiel: Die Grafik zeigt unten das FSK–Sendesignal $s(t)$ für

- das oben skizzierte binäre Quellensignal $q(t)$, und
- das darunter gezeichnete Trägersignal $z(t)$ mit vier Schwingungen pro Symboldauer ($f_T \cdot T = 4$).

Der zugrundeliegende Frequenzhub ist $\Delta f_A = 1/T$, was dem Modulationsindex $h = 2$ entspricht. Die beiden möglichen Frequenzen sind somit

$$f_{+1} = 5/T, \quad f_{-1} = 3/T.$$

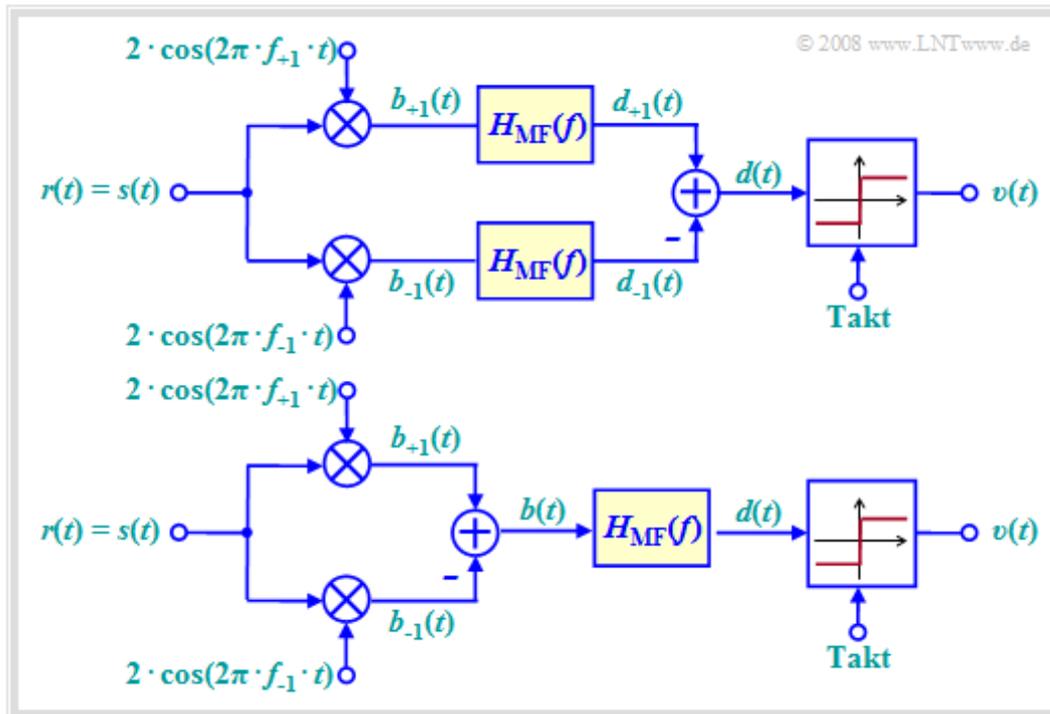


Bei einem FSK–Übertragungssystem mit der Bitrate 1 Mbit/s ($T = 1 \mu\text{s}$) und der Amplitude $A = 1 \text{ V}$ des Quellensignals müsste somit die folgende FM–Konstante verwendet werden:

$$K_{\text{FM}} = \frac{2\pi \cdot \Delta f_A}{A} = \frac{2\pi}{A \cdot T} \approx 6.28 \cdot 10^6 \text{ V}^{-1}\text{s}^{-1}.$$

Kohärente Demodulation der FSK

Die folgende Grafik zeigt den bestmöglichen Demodulator für binäre FSK, der kohärent arbeitet und demzufolge auch Kenntnis über die Phasenlage des FSK-Signals benötigt. Im Blockschaltbild ist dies berücksichtigt, indem das Empfangssignal $r(t)$ identisch mit dem Sendesignal $s(t)$ angenommen wurde – siehe **Signalverläufe** auf der vorherigen Seite.



Dieser Demodulator arbeitet nach folgendem Prinzip:

- Es handelt sich hierbei um einen Maximum-Likelihood-Empfänger (ML) in der Realisierungsform mit Matched-Filter. Dieses Filter mit dem Frequenzgang $H_{MF}(f)$ kann bei dem vorausgesetzten rechteckförmigen Sendegrundimpuls $g_s(t)$ auch als Integrator realisiert werden.
- Die Signale $b_{+1}(t)$ bzw. $b_{-1}(t)$ vor den Matched-Filtern ergeben sich durch die phasenrichtige Multiplikation mit den Schwingungen der Frequenz f_{+1} bzw. f_{-1} .
- Der ML-Empfänger entscheidet sich bekanntlich für den Zweig (das Symbol) mit der größeren „Metrik“, wobei das nachgeschaltete Matched-Filter zu berücksichtigen ist. Das heißt: Gilt $d_{+1}(\nu \cdot T) > d_{-1}(\nu \cdot T) \Rightarrow d(\nu \cdot T) = d_{+1}(\nu \cdot T) - d_{-1}(\nu \cdot T) > 0$, so wurde wahrscheinlich $a_\nu = +1$ gesendet.
- Das obere Blockschaltbild wurde zum besseren Verständnis so gezeichnet. Natürlich kann man die Matched-Filterung aber auch auf die rechte Seite der Differenzbildung verschieben, wie im unteren Bild dargestellt. Damit muss nur noch ein Filter realisiert werden.

In der **Aufgabe A4.12** wird dieser FSK-Demodulator ausführlich behandelt. Auf dem entsprechenden Angabenblatt sehen Sie auch die Signalverläufe.

Fehlerwahrscheinlichkeit der orthogonalen FSK

Man spricht von **orthogonaler FSK**, wenn

- der Modulationsindex h ein ganzzahliges Vielfaches von 0.5 ist, und damit
- der Frequenzhub Δf_A ein ganzzahliges Vielfaches von $0.25/T$.

Beim kohärenten Demodulator ist der Korrelationskoeffizient zwischen $d_{+1}(T_D)$ und $d_{-1}(T_D)$ zu allen Detektionszeitpunkten gleich 0. Der Betrag $|d(T_D)|$ – also der Abstand der Detektionsabstastwerte von der Schwelle – ist somit konstant. Es treten keine Impulsinterferenzen auf.

Die Fehlerwahrscheinlichkeit ergibt sich zu:

$$p_B = Q\left(\sqrt{\frac{E_B}{N_0}}\right) = \frac{1}{2} \cdot \operatorname{erfc}\left(\sqrt{\frac{E_B}{2 \cdot N_0}}\right),$$

wenn man von den folgenden Voraussetzungen ausgeht:

- orthogonale FSK,
- AWGN-Kanal (gekennzeichnet durch den Quotienten E_B/N_0), und
- der hier beschriebenen kohärenten Demodulation.

Dies entspricht einer Degradation von 3 dB gegenüber der **BPSK**, weil

- zwar der kohärente FSK-Demodulator bezüglich des Nutzsignals das gleiche Ergebnis liefert,
- auch die Rauschleistungen in den beiden Zweigen genau so groß sind wie bei der BPSK,
- es aber wegen der Subtraktion zu einer Verdopplung der Gesamtrauschleistung kommt.

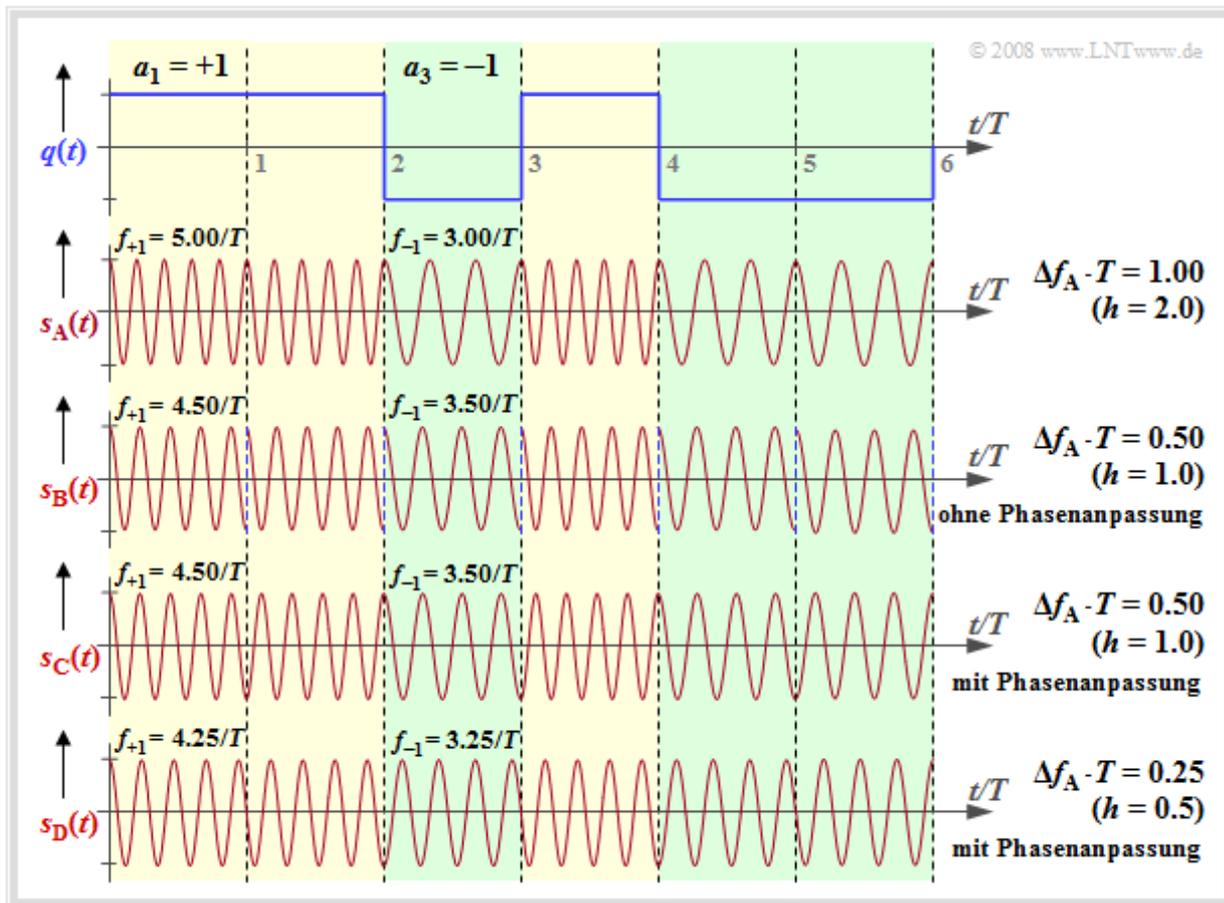
Während aber bei der BPSK eine nichtkohärente Demodulation auf keinen Fall möglich ist, gibt es auch einen **nichtkohärenten FSK-Demodulator**, allerdings mit etwas erhöhter Fehlerwahrscheinlichkeit:

$$p_B = 1/2 \cdot e^{-E_B/(2N_0)}.$$

Die Herleitung dieser Gleichung erfolgt im **Kapitel 4.5** des Buches „Digitalsignalübertragung“.

Binäre FSK mit kontinuierlicher Phasenanpassung

Wir betrachten weiter die orthogonale FSK. Die Grafik zeigt oben das Quellensignal $q(t)$ und das FSK-Signal $s_A(t)$ mit dem Frequenzhub $\Delta f_A = 1/T \Rightarrow$ Modulationsindex $h = 2 \cdot \Delta f_A \cdot T = 2$.



Zu den weiteren Signalverläufen ist Folgendes anzumerken:

- $s_B(t)$ verwendet die Momentanfrequenzen $f_{+1} = 4.5/T$ und $f_{-1} = 3.5/T \Rightarrow \Delta f_A \cdot T = 0.5 \Rightarrow h = 1$. Auch diese FSK ist orthogonal, da $h = 1$ ein Vielfaches von 0.5 ist. Bei kleinerem h ist aber die Bandbreiteneffizienz besser \Rightarrow das Spektrum $S_B(f)$ ist weniger breit als das Spektrum $S_A(f)$.
- Allerdings erkennt man im Signal $s_B(t)$ an jeder Symbolgrenze einen Phasensprung um π , wodurch sich wieder eine Verbreiterung des Spektrums ergibt. Solche Phasensprünge lassen sich durch Phasenanpassung vermeiden. Man spricht dann von **Continuous Phase Modulation** (CPM).
- Auch beim CPM-Signal $s_C(t)$ gilt $f_{+1} = 4.5/T, f_{-1} = 3.5/T$ und $h = 1$. Im Bereich von $0 \dots T$ wird der Koeffizient $a_1 = +1$ mit $\cos(2\pi \cdot f_{+1} \cdot t)$ repräsentiert, im Bereich $T \dots 2T$ dagegen der ebenfalls positive Koeffizient $a_2 = +1$ durch die um π verschobene Funktion $-\cos(2\pi \cdot f_{+1} \cdot t)$ dargestellt.
- Der Modulationsindex $h = 0.5$ von Signal $s_D(t)$ ist der kleinstmögliche Wert, der eine orthogonale FSK ermöglicht \Rightarrow Bezeichnung **Minimum Shift Keying** (MSK). Bei MSK sind bei jeder Symbolgrenze – je nach den vorherigen Symbolen – vier unterschiedliche Anfangsphasen möglich.

Zur Verdeutlichung des hier dargelegten Sachverhalts gibt es das Interaktionsmodul

Continuous Phase Modulation (CPM)

MSK – Minimum Shift Keying

Die Grafik zeigt das Blockschaltbild zur Erzeugung einer MSK-Modulation und typische Signalverläufe an verschiedenen Punkten des MSK-Senders. Man erkennt

- das digitale Nachrichtensignal am Punkt 1 (Quelle), bestehend aus einer Folge von Diracimpulsen im Abstand T , gewichtet mit den Amplitudenkoeffizienten $a_\nu \in \{-1, +1\}$:

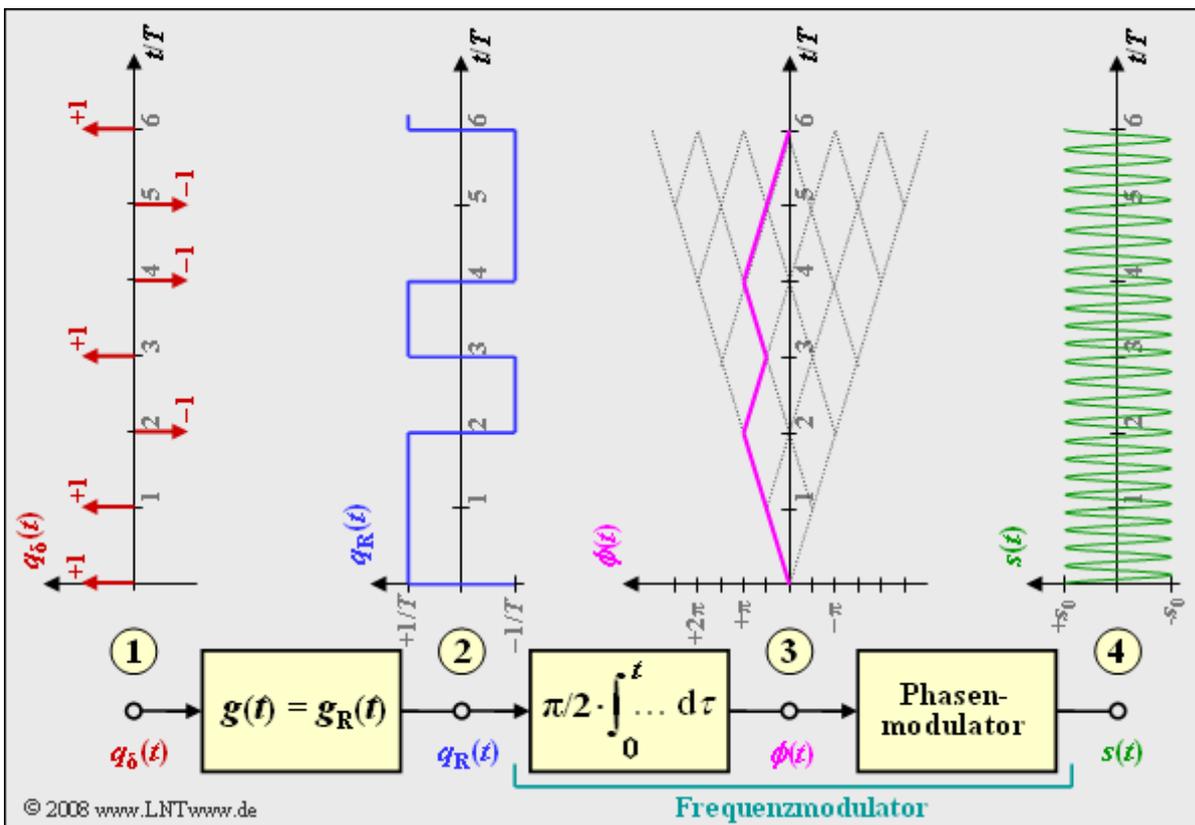
$$q_\delta(t) = \sum_{\nu=-\infty}^{+\infty} a_\nu \cdot \delta(t - \nu \cdot T),$$

- das Rechtecksignal $q_R(t)$ am Punkt 2 nach Faltung mit dem Rechteckimpuls $g(t)$ der Dauer T und der Höhe $1/T$ (die Amplitude wurde aus Kompatibilitätsgründen zu späteren Seiten so gewählt):

$$q_R(t) = \sum_{\nu=-\infty}^{+\infty} a_\nu \cdot g(t - \nu \cdot T),$$

- den herkömmlichen Frequenzmodulator, der sich entsprechend den Angaben in **Kapitel 3.2** als Integrator und nachgeschaltetem Phasenmodulator realisieren lässt. Für das Signal am Punkt 3 gilt:

$$\phi(t) = \frac{\pi}{2} \cdot \int_0^t q_R(\tau) \, d\tau.$$

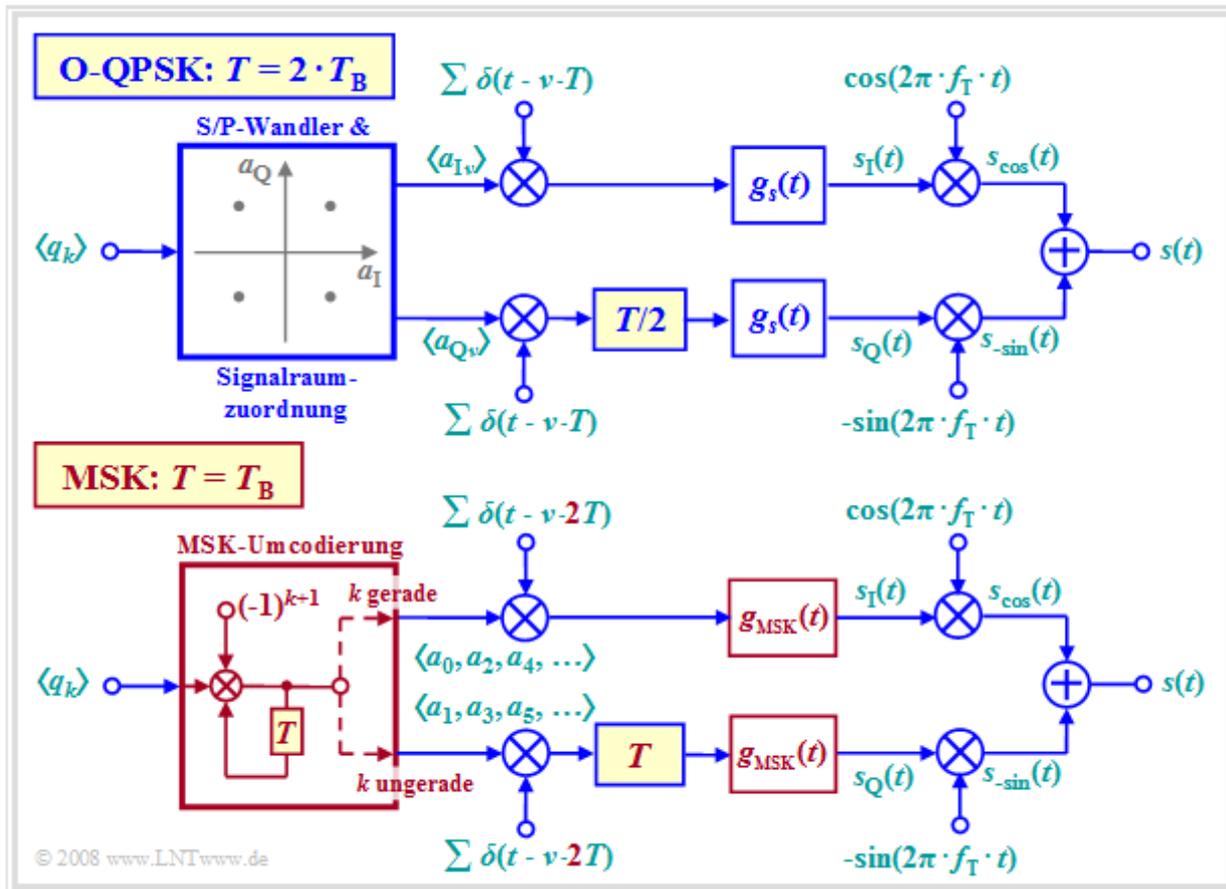


Die Phasenwerte bei Vielfachen der Symboldauer T sind Vielfache von $\pi/2$, wobei der für MSK gültige Modulationsindex $h = 0.5$ berücksichtigt ist. Der Phasenverlauf ist linear. Daraus ergibt sich das MSK-Signal am Punkt 4 des Blockschaltbildes zu

$$s(t) = s_0 \cdot \cos(2\pi f_T t + \phi(t)) = s_0 \cdot \cos(2\pi \cdot t \cdot (f_T + a_\nu \cdot \Delta f_A)).$$

Realisierung der MSK als Offset-QPSK (1)

Durch einen modifizierten Betrieb von **Offset Quaternary Phase Shift Keying** (O-QPSK) gemäß der unteren Grafik lässt sich *Minimum Shift Keying* (MSK) realisieren.



Gegenüber dem herkömmlichen Betrieb (obere Grafik) sind folgende Modifikationen zu berücksichtigen, die in der unteren Grafik rot hervorgehoben sind:

- Die Symboldauer T der MSK ist gleich der Bitdauer T_B des binären Eingangssignals, während bei der originären O-QPSK $T = 2 T_B$ gilt.
- Anstelle der Seriell-Parallel-Wandlung und Signalraumzuordnung müssen nun die Quellsymbole umcodiert entsprechend $a_k = (-1)^{k+1} \cdot a_{k-1} \cdot q_k$ werden.
- Alle Amplitudenkoeffizienten a_k mit geradzahligem Index (a_0, a_2, \dots) werden dem Diracpuls im oberen Zweig eingepreßt, während a_1, a_3, \dots im unteren Zweig übertragen werden.
- Der Abstand der einzelnen Diracimpulse beträgt nun $2T$ anstelle von T und der Versatz (Offset) im Quadraturzweig ist nicht mehr $T/2$, sondern T . In beiden Fällen ist aber der Offset gleich T_B .
- Während beim herkömmlichen O-QPSK-Betrieb jeder beliebige Grundimpuls $g_s(t)$ möglich ist, zum Beispiel ein Rechteck- oder ein Wurzel-Nyquist-Impuls, gibt es für den MSK-Betrieb nur einen einzigen geeigneten Grundimpuls. Dieser erstreckt sich über zwei Symboldauern:

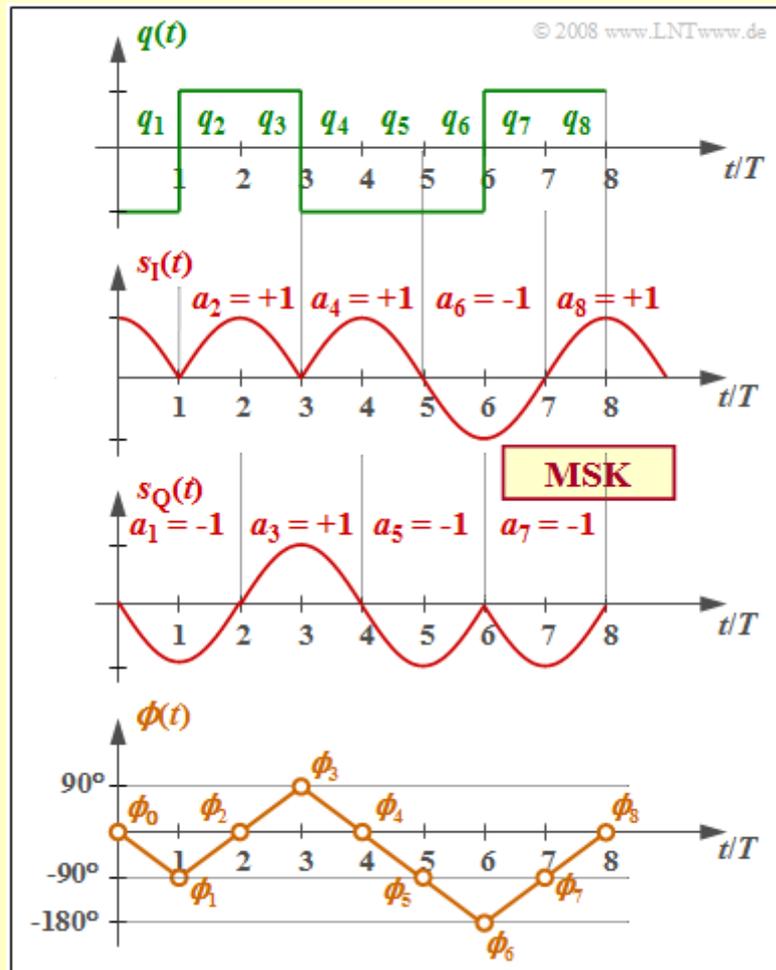
$$g_{MSK}(t) = \begin{cases} s_0 \cdot \cos\left(\frac{\pi \cdot t}{2 \cdot T}\right) & \text{für } -T \leq t \leq +T, \\ 0 & \text{für sonst.} \end{cases}$$

Realisierung der MSK als Offset-QPSK (2)

Beispiel: Die Grafik zeigt

- oben das binäre bipolare Quellensignal $q(t)$,
- in der Mitte die äquivalenten TP-Signale $s_I(t)$ und $s_Q(t)$ im I- und Q-Zweig, sowie
- unten den Phasenverlauf $\phi(t)$ des gesamten MSK-Sendesignals $s(t)$.

Die Umcodierung $a_k = (-1)^{k+1} \cdot a_{k-1} \cdot q_k$ ist bereits berücksichtigt, ebenso der MSK-Grundimpuls.



Man erkennt aus dem Vergleich des obersten und des untersten Diagramms:

- Der MSK-Phasenverlauf $\phi(t)$ ist abschnittsweise linear und steigt bzw. fällt innerhalb einer Symboldauer um 90° ($\pi/2$), je nachdem, ob gerade $q_k = +1$ oder $q_k = -1$ anliegt.
- Das zugehörige Sendesignal $s(t)$ beinhaltet abschnittsweise die beiden Frequenzen $f_T \pm 1/(4T)$. Es hat den prinzipiell gleichen Verlauf wie das Signal $s_D(t)$ auf Seite 5 dieses Kapitels.

Diese Form der MSK-Realisierung können Sie mit dem nachfolgenden Interaktionsmodul darstellen, wenn Sie die Einstellungen O-QPSK, MSK-Zuordnung und Cosinusimpuls wählen:

Quaternary Phase Shift Keying (4-QAM/QPSK bzw. O-QPSK)

Allgemeingültige Beschreibung der CPM (1)

Wir gehen wieder davon aus, dass die Quelle durch die Amplitudenkoeffizienten a_ν charakterisiert wird. Diese können sowohl binär (± 1) als auch M -stufig sein. Sie sind aber stets bipolar zu betrachten.

Die **Phasenfunktion** $\phi(t)$ kann bei *Continuous Phase Modulation* (CPM) allgemein in folgender Form dargestellt werden (h bezeichnet wieder den Modulationsindex):

$$\phi(t) = \pi \cdot h \cdot \int_{-\infty}^t \sum_{\nu=-\infty}^{+\infty} a_\nu \cdot g(\tau - \nu \cdot T) d\tau.$$

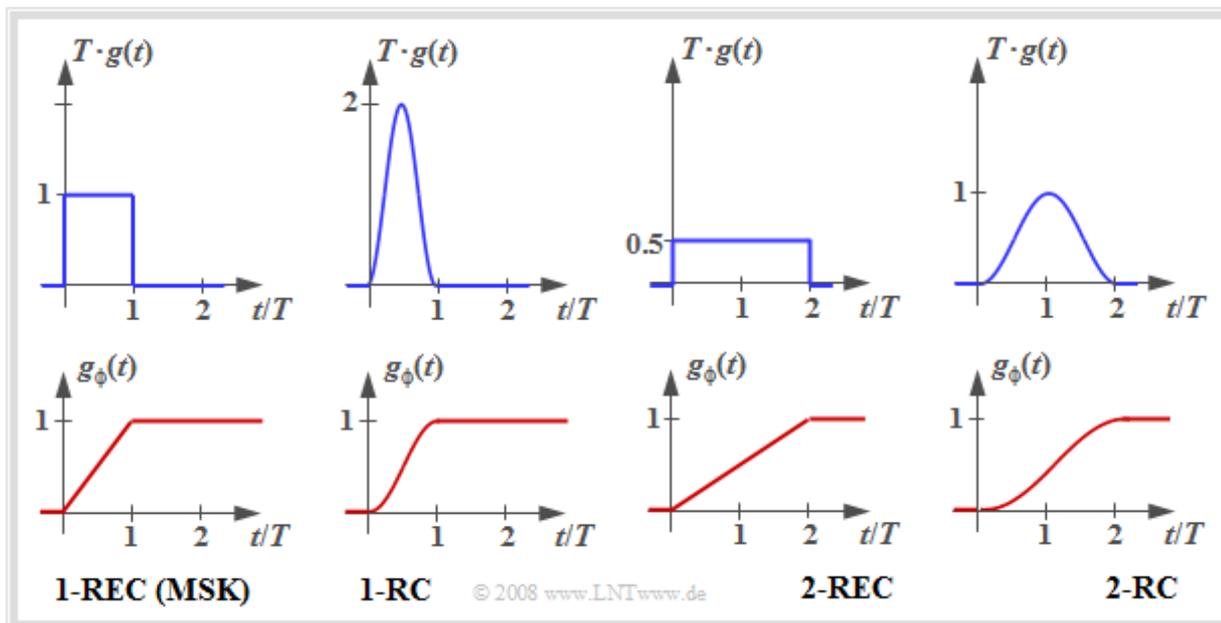
In dieser Darstellung bezeichnet $g(t)$ den **Frequenzimpuls**, der folgende Bedingung erfüllen muss:

$$\int_{-\infty}^{+\infty} g(t) dt = 1.$$

Mit dem **Phasenimpuls** $g_\phi(t)$ gilt aber auch der folgende Zusammenhang:

$$\phi(t) = \pi \cdot h \cdot \sum_{\nu=-\infty}^{+\infty} a_\nu \cdot g_\phi(t - \nu \cdot T), \text{ wobei } g_\phi(t) = \int_{-\infty}^t g(\tau) d\tau.$$

Durch geeignete Wahl der Impulse $g(t)$ bzw. $g_\phi(t)$ lassen sich viele CPM-Varianten realisieren. Einige davon sind in der nachfolgenden Grafik dargestellt.



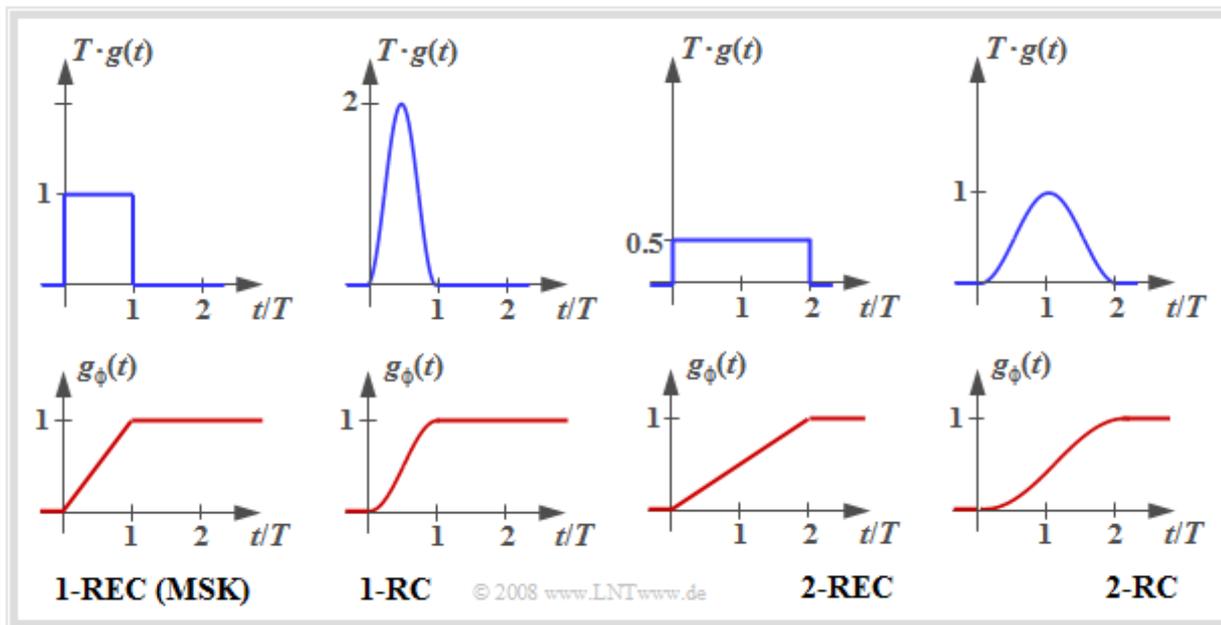
Die Bildbeschreibung folgt auf der nächsten Seite.

Allgemeingültige Beschreibung der CPM (2)

Die folgende Grafik zeigt jeweils

- oben den CPM–Frequenzimpuls $g(t)$,
- unten den CPM–Phasenimpuls $g_\phi(t)$.

Die Bezeichnung „1–REC“ für die Darstellungen der linken Spalte gibt an, dass der Frequenzimpuls sich über eine Symboldauer erstreckt und rechteckförmig ist. Diese Grafiken beschreiben also die MSK.



Die weiteren in der Grafik berücksichtigten CPM–Varianten wurden mit dem Ziel entworfen, die bereits kleine Bandbreite des MSK–Signals asymptotisch weiter zu verringern:

- Bei „1–RC“ ergibt sich allein durch den „weicheren“ **Raised–Cosine–Impuls** $g(t)$ gegenüber dem Rechteck ein schmaleres Leistungsdichtespektrum.
- Bei „2–RC“ und „2–REC“ handelt es sich um **Partial–Response–Impulse**, die sich jeweils über zwei Symboldauern erstrecken. Hierdurch wird der Phasenverlauf ebenfalls weicher. Gleichzeitig wird dadurch aber die Demodulation und Detektion erschwert, da in das Datensignal am Eingang des Frequenzmodulators gezielte Pseudomehrstufigkeiten eingebracht werden.

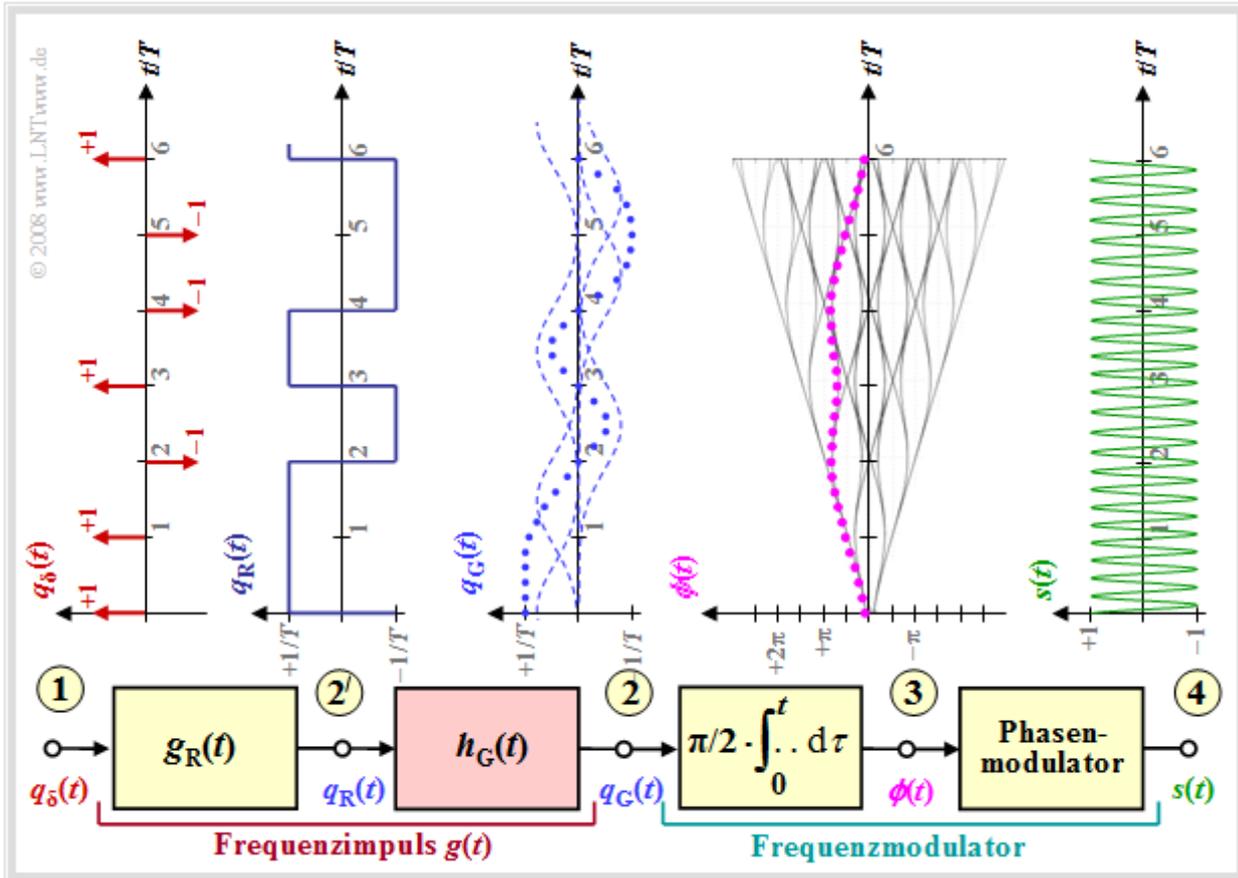
Die Berechnung der CPM–Verfahren im Spektralbereich ist im allgemeinen kompliziert. Hier sei auf die Fachliteratur verwiesen, zum Beispiel auf das empfehlenswerte Buch **[Kam04]**. Nur der Sonderfall „MSK“ führt zu einfach handhabbaren Gleichungen, wie in der **Aufgabe A4.14** gezeigt werden wird.

Kurze Zusammenfassung: Die **Continuous Phase Modulation (CPM)** ist keine Phasenmodulation, sondern stellt eine **nichtlineare digitale Frequenzmodulation (FSK)** dar, mit dem Ziel

- eine **konstante Betragseinhüllende** zu garantieren, und
- einen **stetigen Phasenverlauf** zu ermöglichen (Phasensprünge verbreitern das Spektrum).

GMSK – Gaussian Minimum Shift Keying

Ein Vorteil der MSK gegenüber anderen Verfahren ist der geringere Bandbreitenbedarf. Durch eine geringfügige Modifikation hin zum **Gaussian Minimum Shift Keying** – abgekürzt GMSK– wird das Spektrum weiter verschmälert. GMSK wird zum Beispiel beim Mobilfunkstandard GSM angewendet.



Man erkennt aus dem Grafik, dass der Frequenzimpuls $g(t)$ nun nicht mehr rechteckförmig ist wie $g_R(t)$, sondern flachere Flanken aufweist. Dadurch ergibt sich ein weicherer Phasenverlauf am Punkt 3 als beim MSK–Verfahren (siehe letzte Seite), bei dem $\phi(t)$ symbolweise linear ansteigt bzw. abfällt.

Man erreicht die sanfteren GMSK–Phasenübergänge durch einen **Gaußtieffpass**. Dessen Frequenzgang und Impulsantwort lauten mit der systemtheoretischen Grenzfrequenz f_G :

$$H_G(f) = e^{-\pi \cdot \left(\frac{f}{2f_G}\right)^2} \quad \bullet \text{---} \circ \quad h_G(t) = 2f_G \cdot e^{-\pi \cdot (2f_G \cdot t)^2} .$$

Der resultierende Frequenzimpuls $g(t)$ am Punkt 2 ergibt sich aus der Faltung von $g_R(t)$ und $h_G(t)$. Das Signal $s(t)$ am Punkt 4 weist bei GMSK nun nicht mehr abschnittsweise (je Symboldauer) eine konstante Frequenz auf wie bei MSK, auch wenn dies aus obiger Grafik mit bloßem Auge schwer zu erkennen ist.

Beispiel: Beim GSM–Verfahren ist die 3dB–Grenzfrequenz zu $f_{3dB} = 0.3/T$ spezifiziert, wobei zwischen der systemtheoretischen und der 3dB–Grenzfrequenz folgender Zusammenhang besteht:

$$H_G(f = f_{3dB}) = e^{-\pi \cdot (f_{3dB}/2f_G)^2} = 1/\sqrt{2} \Rightarrow f_{3dB} = f_G \cdot \sqrt{4/\pi \cdot \ln\sqrt{2}} \approx 2/3 \cdot f_G .$$

Aus $f_{3dB} = 0.3/T$ folgt damit auch $f_G \cdot T \approx 0.45$.