

## Überblick zu Kapitel 1 des Buches „Digitalsignalübertragung“

Das Kapitel 1 führt in das weite Gebiet der **Digitalsignalübertragung** ein, wobei einige vereinfachende Annahmen getroffen werden: ein redundanzfreies binäres Sendesignal, keine Impulsinterferenzen. Obwohl die Beschreibung vorwiegend im Basisband erfolgt, lassen sich die Ergebnisse meist auch auf die digitalen Trägerfrequenzsysteme (Kapitel 1.5) übertragen. Im Einzelnen werden behandelt:

- der prinzipielle *Aufbau* und die *Komponenten* eines Basisbandübertragungssystems,
- die Definitionen von *Bitfehlerwahrscheinlichkeit* und *Bitfehlerhäufigkeit* (BER),
- die Eigenschaften der *Nyquistsysteme*, die eine impulsinterferenzfreie Übertragung erlauben,
- die *Optimierung* der binären Basisbandsysteme bei Leistungs- und Spitzenwertbegrenzung,
- die Verallgemeinerung der Ergebnisse auf Trägerfrequenzsysteme,
- die weitgehend gemeinsame Beschreibung von ASK, BPSK und 4-QAM.

**Geeignete Literatur:** [BBC87] – [Hän97] – [Hag02] – [Han04] – [Hay88] – [Kam04] – [Lük04] – [Pro01] – [PS04] – [Söd00] – [ST85] – [TS87]

Die grundlegende Theorie wird auf 46 Seiten dargelegt. Außerdem beinhaltet dieses Kapitel noch 59 Grafiken, zehn Aufgaben und sieben Zusatzaufgaben mit insgesamt 90 Teilaufgaben, sowie zwölf Lernvideos (LV) und zehn Interaktionsmodule (IM). Unter anderem:

- **Gesetze der Fouriertransformation** (LV, Grundlagen, 2-tlg.: Dauer 5:57 – 5:55)
- **Klassische Definition der Wahrscheinlichkeit** (LV, Grundlagen, Dauer 5:20)
- **Eigenschaften des Übertragungskanal** (LV zu Kap. 1.1, Dauer 5:50)
- **Komplementäre Gaußsche Fehlerfunktionen** (IM zu Kapitel 1.1 ff.)
- **Das Bernouillische Gesetz der großen Zahlen** (LV zu Kapitel 1.2, Dauer 4:25)
- **Gegenüberstellung Binomial- und Poissonverteilung** (IM zu Kapitel 1.2)
- **Symbolfehlerwahrscheinlichkeit von Digitalsystemen** (IM zu Kapitel 1.2)
- **Zur Verdeutlichung der grafischen Faltung** (IM zu Kapitel 1.2)
- **Zur Verdeutlichung des Matched-Filters** (IM zu Kapitel 1.2)
- **Frequenzgang und zugehörige Impulsantwort** (IM zu Kapitel 1.3)
- **Der AWGN-Kanal – Teil 2** (LV zu Kapitel 1.4, Dauer 5:15)
- **Eigenschaften von TP- und BP-Signalen** (LV zu Kapitel 1.5, Dauer 5:16)
- **Harmonische Schwingungen** (LV zu Kapitel 1.5, 2-tlg.: Dauer 4:33 – 6:15)

Weitere Informationen zum Thema sowie Aufgaben, Simulationen und Programmierübungen finden Sie in den Kapiteln 13 bis 16 des Praktikums *Simulationsmethoden in der Nachrichtentechnik* [Söd00], das über viele Jahre für Studierende der Elektrotechnik und Informationstechnik an der TU München von Prof. Günter Söder angeboten wurde. Dieses Praktikum des Lehrstuhls für Nachrichtentechnik (LNT) basiert auf den 24 DOS-Programmen des Lehrsoftwarepakets *LNTsim*.

**Lehrsoftwarepaket LNTsim** (Zip-Version, mehr als 50 MB)

**Praktikumsanleitung – Teil B** (PDF-Version incl. Kapitel 13 – 16, ca. 9.5 MB)

## Vereinfachtes Systemmodell

Im gesamten ersten Kapitel wird für das Digitalsystem von folgendem Blockschaltbild ausgegangen:



Im Vergleich zu einem **analogen Übertragungssystem** (siehe Buch „Modulationsverfahren“) erkennt man in diesem vereinfachten Systemmodell folgende Gemeinsamkeiten und Unterschiede:

- Das Blockschaltbild ist in beiden Fällen in genau gleicher Weise aufgebaut – bestehend aus Quelle, Sender, Kanal, Empfänger und Sinke – und auch die Signale werden gleich bezeichnet.
- Auch beim digitalen Übertragungssystem ist das Empfangssignal  $r(t)$  aufgrund der Störungen zeit- und wertkontinuierlich. Das Sendesignal  $s(t)$  kann zeit- und wertdiskret sein, muss aber nicht.
- Im Unterschied zum Buch „Modulationsverfahren“ sind aber nun das Quellensignal  $q(t)$  und das Sinkensignal  $v(t)$  stets Digitalsignale. Sie sind dementsprechend sowohl zeit- als auch wertdiskret.
- Alle Informationen über  $q(t)$  und  $v(t)$  können somit auch durch die **Quellensymbolfolge**  $\langle q_v \rangle$  und die **Sinkensymbolfolge**  $\langle v_v \rangle$  gemeinsam mit der Symboldauer  $T$  ausgedrückt werden.
- Ein Digitalempfänger unterscheidet sich grundsätzlich vom Empfänger eines Analogsystems, da er zusätzlich eine **Entscheidungskomponente** zur Gewinnung des digitalen Sinkensignals  $v(t)$  aus dem analogen Empfangssignals  $r(t)$  beinhalten muss.
- In den ersten drei Kapiteln dieses Buches betrachten wir die **digitale Basisbandübertragung**, was besagt, dass das Nachrichtensignal  $q(t)$  ohne vorherige Frequenzumsetzung (Modulation mit einer Trägerschwingung) übertragen wird.
- Deshalb sind hier  $s(t)$  und  $r(t)$  Tiefpass-Signale und auch für den Kanal (inklusive der Störungen) muss stets von einer Tiefpass-Charakteristik ausgegangen werden.

Nachfolgend werden die Eigenschaften der einzelnen Systemkomponenten detailliert beschrieben, wobei die idealisierenden Voraussetzungen für Kapitel 1 geeignet berücksichtigt werden.

## Beschreibungsgrößen der digitalen Quelle (1)

Die **digitale Quelle** erzeugt die Quellensymbolfolge  $\langle q_\nu \rangle$ , die möglichst fehlerfrei zur Senke übertragen werden soll. Im Allgemeinen entstammt jedes Symbol der zeitlichen Folge  $\langle q_\nu \rangle$  mit  $\nu = 1, 2, \dots$  einem Symbolvorrat  $\{q_\mu\}$  mit  $\mu = 1, \dots, M$ , wobei  $M$  als Quellensymbolumfang oder auch als **Stufenzahl** bezeichnet wird. Für das erste Kapitel dieses Buches wird von folgenden Voraussetzungen ausgegangen:

- Die Quelle ist **binär** ( $M = 2$ ) und die beiden möglichen Symbole sind **L** („Low“) und **H** („High“).
- Die Quellensymbole sind **statistisch unabhängig** voneinander, das heißt, die Wahrscheinlichkeit  $\Pr(q_\nu = q_\mu)$ , dass das  $\nu$ -te Symbol der Folge  $\langle q_\nu \rangle$  gleich dem  $\mu$ -ten Symbol des Symbolvorrates  $\{q_\mu\}$  ist, hängt nicht von  $\nu$  ab.
- Aufgrund dieser zwei Annahmen wird die digitale Quelle durch die **Symbolwahrscheinlichkeiten**  $p_L = \Pr(q_\nu = \mathbf{L})$  und  $p_H = \Pr(q_\nu = \mathbf{H}) = 1 - p_L$  vollständig beschrieben.
- Gilt weiterhin  $p_L = p_H = 0.5$ , so ist die Quelle **redundanzfrei**. Meist – jedoch nicht immer – wird im Kapitel 1 eine solche vorausgesetzt.
- Der zeitliche Abstand zweier Symbole sei  $T$ . Man bezeichnet diese Größe als die **Symboldauer** und den Kehrwert als die **Symbolrate**  $R = 1/T$ . Bei Binärquellen ( $M = 2$ ) nennt man diese Größen auch Bitdauer bzw. Bitrate.
- Bei **systemtheoretischer Betrachtungsweise** der digitalen Basisbandübertragung beschreibt man das Quellensignal am besten durch eine Folge gewichteter und verschobener Diracimpulse:

$$q(t) = \sum_{(\nu)} a_\nu \cdot \delta(t - \nu \cdot T).$$

- Hierbei bezeichnet man  $a_\nu$  als die **Amplitudenkoeffizienten**. Im Falle der **binären** unipolaren Digitalsignalübertragung gilt:

$$a_\nu = \begin{cases} 1 & \text{für } q_\nu = \mathbf{H}, \\ 0 & \text{für } q_\nu = \mathbf{L}. \end{cases}$$

Entsprechend gilt bei einem **bipolaren** (oder antipodischen) System:

$$a_\nu = \begin{cases} +1 & \text{für } q_\nu = \mathbf{H}, \\ -1 & \text{für } q_\nu = \mathbf{L}. \end{cases}$$

Die nachfolgende Beschreibung erfolgt meist für diesen zweiten Fall.

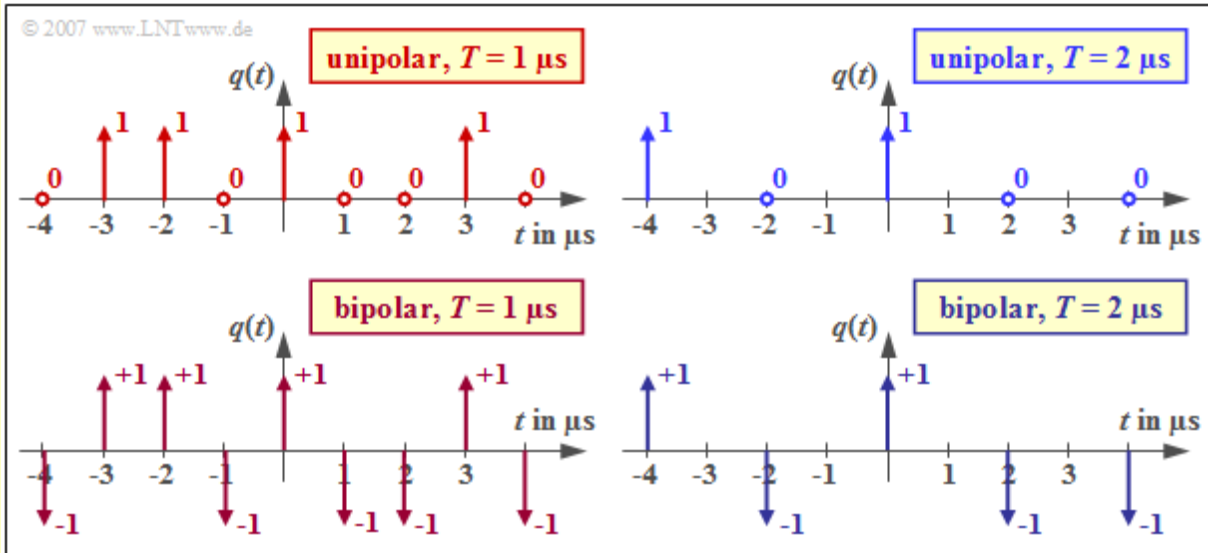
**Hinweis zur Nomenklatur:** In der Literatur wird meist das Symbol **H** mit **0** bezeichnet. Bei unipolarer Signalisierung wird dann das Symbol **0** durch den Amplitudenkoeffizienten „1“ und das Symbol **L** durch den Zahlenwert „0“ dargestellt. Um diesen unschönen Sachverhalt zu vermeiden, wird in *LNTwww* das Symbol **0** mit **H** bezeichnet, wobei „High“ den Sachverhalt richtig ausdrückt.

## Beschreibungsgrößen der digitalen Quelle (2)

**Beispiel:** Die Grafik zeigt vier binäre diracförmige Quellsignale im Bereich von  $-4 \mu\text{s}$  bis  $+4 \mu\text{s}$ , wobei jeweils die folgende Quellensymbolfolge zugrundeliegt:

$$\langle q_\nu \rangle = \langle \dots, \text{L}, \text{H}, \text{H}, \text{L}, \text{H}, \text{L}, \text{L}, \text{H}, \text{L}, \dots \rangle.$$

Das mittlere, etwas abgesetzte Symbol bezieht sich jeweils auf den Zeitpunkt  $t = 0$ .



Die zwei oberen Signale eignen sich zur Beschreibung unipolarer Systeme, die unteren für die bipolare (antipodische) Digitalsignalübertragung. Für die jeweils linken Grafiken ist  $T = 1 \mu\text{s}$  vorausgesetzt. Für die beiden rechten gilt dagegen  $T = 2 \mu\text{s}$  und damit die halbe Symbolrate.

## Kenngrößen des digitalen Senders (1)

Der Sender eines digitalen Übertragungssystems hat die Aufgabe, aus dem (diracförmigen) Quellensignal ein geeignetes Sendesignal  $s(t)$  zu erzeugen, das die Nachricht der Quelle vollständig beinhaltet und an die Eigenschaften von Übertragungskanal, Störungen sowie aller technischen Empfangseinrichtungen angepasst ist. Außerdem sorgt der Sender für die Bereitstellung einer hinreichend großen Sendeleistung.

Als Beschreibungsgröße für den Sender verwenden wir den **Sendegrundimpuls**  $g_s(t)$ . Aufgrund der Definition des Quellensignals  $q(t)$  als Summe von gewichteten und verschobenen Diracfunktionen lässt sich das Sendesignal mit den Amplitudenkoeffizienten  $a_\nu$  in folgender Weise darstellen:

$$s(t) = q(t) \star g_s(t) = \sum_{(\nu)} a_\nu \cdot g_s(t - \nu \cdot T).$$

Häufig wird der Sendegrundimpuls  $g_s(t)$  als rechteckförmig mit

- der Impulshöhe  $s_0 = g_s(t = 0)$  und
- der (absoluten) Impulsdauer  $T_S$

angenommen. Gilt  $T_S < T$ , so spricht man von einem **RZ-Impuls** („return-to-zero“), bei  $T_S = T$  von einem **NRZ-Impuls** („non-return-to-zero“).

Bei anderem Sendegrundimpuls, z.B. einem Gauß-, Trapez-,  $\cos^2$ - oder Wurzel-Nyquist-Impuls, verwendet man als Beschreibungsparameter anstelle der absoluten Impulsdauer  $T_S$  meist die über das flächengleiche Rechteck definierte **äquivalente Impulsdauer**:

$$\Delta t_S = \frac{\int_{-\infty}^{+\infty} g_s(t) dt}{\text{Max}[g_s(t)]} \leq T_S.$$

Nur bei rechteckförmigem Sendegrundimpuls gilt  $\Delta t_S = T_S$ .

Unterscheidet sich die Amplitude des Sendegrundimpulses  $g_s(t)$  vom Maximalwert  $s_0$  des Sendesignals  $s(t)$ , so bezeichnen wir die Impulsamplitude mit  $A_S$ . Dies trifft zum Beispiel beim Gaußimpuls zu.

Das folgende Interaktionsmodul zeigt einige geläufige Grundimpulse und die dazugehörigen Spektren:

### Impulse und deren Spektren

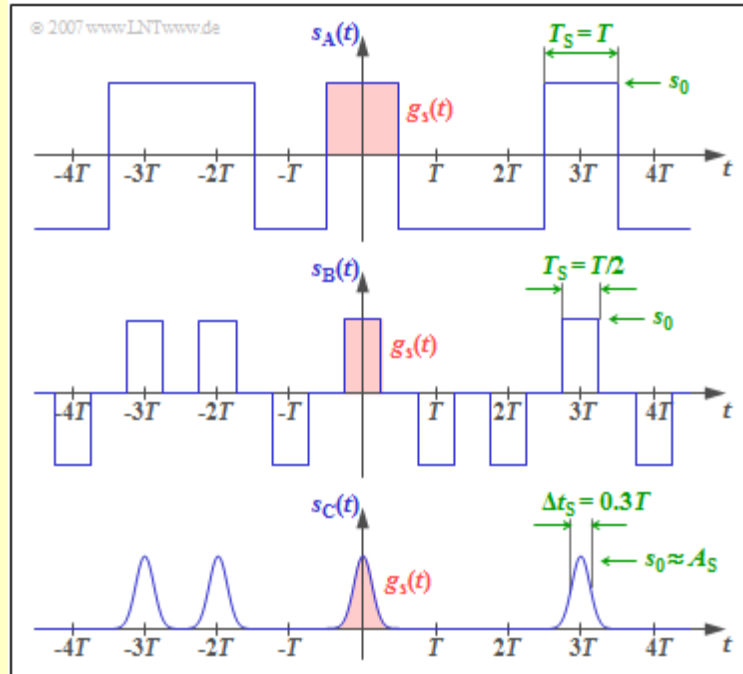
## Kenngrößen des digitalen Senders (2)

**Beispiel:** Die folgende Grafik zeigt drei Sendesignale. Zugrunde liegt stets die Quellensymbolfolge

$$\langle q_\nu \rangle = \langle \dots, \text{L}, \text{H}, \text{H}, \text{L}, \text{H}, \text{L}, \text{L}, \text{H}, \text{L}, \dots \rangle.$$

Die Grafik berücksichtigt

- ein bipolares Sendesignal  $s_A(t)$  mit NRZ-Rechteckimpulsen,
- ein bipolares Sendesignal  $s_B(t)$  mit RZ-Rechteckimpulsen, und
- ein unipolares Sendesignal  $s_C(t)$  mit Gaußimpulsen.



Bei den folgenden Beschreibungen wird meist das bipolare NRZ-Rechtecksignal  $s_A(t)$  vorausgesetzt. Die Dauer  $T_S$  des in der Grafik rot eingezeichneten Sendegrundimpulses  $g_s(t)$  ist hier gleich dem Abstand  $T$  zweier aufeinanderfolgender Impulse. Aus den weiteren Skizzen erkennt man:

- Beim RZ-Sendesignal  $s_B(t)$  unterscheidet sich die Impulsdauer  $T_S$  vom Impulsabstand  $T$ . Die Skizze gilt für das Tastverhältnis  $T_S/T = 0.5$ . Obwohl  $s_B(t)$  ebenfalls ein Binärsignal ist, gibt es hier drei mögliche Signalwerte, nämlich  $+s_0$ ,  $-s_0$  und  $0$ . Von Vorteil ist, dass sich auch bei einer langen **H**- oder **L**-Folge kein Gleichsignal ergibt, wodurch die Taktsynchronisierung einfacher wird. Nachteilig bei RZ-Signalisierung ist das breitere Spektrum sowie die niedrigere Energie pro Symbol, was zu einer höheren Bitfehlerrate führt.
- Das Signal  $s_C(t)$  ist unipolar und verwendet einen gaußförmigen Grundimpuls  $g_s(t)$ . Ein solches Signal findet man zum Beispiel bei optischen Systemen mit Intensitätsmodulation, da ein Laser oder eine LED prinzipiell keine negativen Impulse erzeugen kann und ein Rechteckimpuls technologisch schwieriger zu erreichen ist als die Gaußform.
- Im Falle eines „echten Gaußimpulses“ gilt für die absolute Impulsdauer stets  $T_S \rightarrow \infty$ . Die (normierte) äquivalente Impulsdauer ist hier mit  $\Delta t_S/T = 0.3$  relativ klein gewählt, so dass der Maximalwert  $s_0$  des Sendesignals etwa gleich der Impulsamplitude  $A_S$  ist. Bei breiteren Gaußimpulsen überlappen sich diese; die Näherung  $s_0 \approx A_S$  trifft in diesem Fall nicht mehr zu.

## Übertragungskanal und Störungen (1)

Der Übertragungskanal umfasst alle Einrichtungen, die zwischen dem Sender und dem Empfänger liegen. Hauptbestandteil ist das **Übertragungsmedium**, das zum Beispiel eine symmetrische Doppelleitung, ein Koaxialkabel, ein Lichtwellenleiter (eine Glasfaser) oder ein Funkfeld sein kann. Daneben beinhaltet der Übertragungskanal verschiedene aus Betriebsgründen notwendige Einrichtungen wie Stromversorgung, Blitzschutz und Fehlerortung.

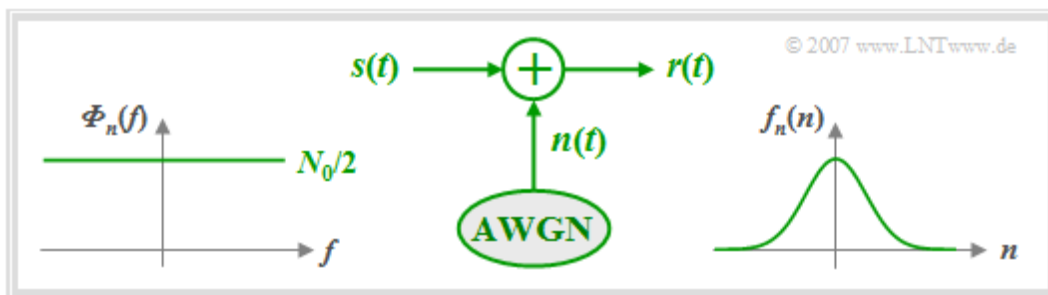
Im allgemeinsten Fall müssen folgende physikalischen Effekte berücksichtigt werden:

- Die Übertragungseigenschaften können **zeitabhängig** sein, insbesondere bei sich bewegendem Sender und/oder Empfänger, wie es im Buch „Mobilkommunikation“ im Detail beschrieben wird. In diesem Grundlagenbuch wird der Kanal stets als linear und zeitinvariant (LZI) angenommen.
- Die Eigenschaften des LZI-Kanals können **frequenzabhängig** sein, gekennzeichnet durch den Frequenzgang  $H_K(f)$ . Bei leitungsgebundener Übertragung gilt stets  $H_K(f) \neq \text{const.}$  und es kommt zu Verzerrungen, wie in **Kapitel 3.1** behandelt.
- Dem Nutzsignal überlagern sich **stochastische Störungen**  $n(t)$ , zum Beispiel das unvermeidbare thermische Rauschen, Impulsstörungen und Nebensprechstörungen anderer Teilnehmer.

Für Kapitel 1 wird stets  $H_K(f) = 1$  vorausgesetzt, das heißt, dass die beiden erstgenannten Punkte vorerst ausgeschlossen werden. Somit gilt im Folgenden für das Signal am Kanalausgang stets:

$$r(t) = s(t) + n(t).$$

Die einfachste realistische Annahme für den Übertragungskanal eines Nachrichtenübertragungssystems ist **Additive White Gaussian Noise**, wie bereits im **Kapitel 3.5** des Buches „Stochastische Signaltheorie“ und im **Kapitel 1.2** des Buches „Modulationsverfahren“ ausgeführt wurde.



Auf der nächsten Seite wird dieses Modell im Detail erklärt.



## Übertragungskanal und Störungen (2)

Das AWGN-Modell lässt sich wie folgt zusammenfassen:

- Der Buchstabe „N“ weist darauf hin, dass durch das AWGN-Modell ausschließlich **Rauschen** („Noise“) berücksichtigt wird. Verzerrungen werden durch dieses einfache Modell nicht erfasst.
- Obwohl Rauschstörungen im Allgemeinen durch eine Vielzahl von Rauschquellen entlang der gesamten Übertragungsstrecke hervorgerufen werden, können diese bei linearen Systemen durch einen einzigen **additiven** Rauschterm am Kanalausgang berücksichtigt werden (Buchstabe „A“).
- Das Rauschen beinhaltet alle Frequenzen gleichermaßen; es besitzt ein konstantes, **weißes** („W“) **Leistungsdichtespektrum** (LDS) und eine diracförmige **Autokorrelationsfunktion** (AKF):

$$\Phi_n(f) = \frac{N_0}{2} \quad \bullet \text{---} \circ \quad \varphi_n(\tau) = \frac{N_0}{2} \cdot \delta(\tau).$$

Der Faktor 1/2 in diesen Gleichungen berücksichtigt die zweiseitige Spektraldarstellung.

- Beispielsweise gilt bei thermischem Rauschen für die physikalische Rauschleistungsdichte (das heißt: einseitige Betrachtungsweise) mit der Rauschzahl  $F \geq 1$  und der absoluten Temperatur  $\theta$ :

$$N_0 = F \cdot k_B \cdot \theta, \quad k_B = 1.38 \cdot 10^{-23} \text{ Ws/K (Boltzmann – Konstante)}.$$

- Bei echt weißem Rauschen würde sich eine unendliche große Leistung ergeben. Deshalb ist stets eine Bandbegrenzung auf  $B$  zu berücksichtigen, und es gilt für die wirksame Rauschleistung:

$$N = \sigma_n^2 = N_0 \cdot B.$$

- Das Rauschsignal  $n(t)$  besitzt eine **Gaußsche** Wahrscheinlichkeitsdichtefunktion (kurz: WDF), was durch den Buchstaben „G“ zum Ausdruck gebracht wird:

$$f_n(n) = \frac{1}{\sqrt{2\pi} \cdot \sigma_n} \cdot e^{-n^2/(2\sigma_n^2)}.$$

Wir möchten Sie hier gerne auf ein dreiteiliges Lernvideo aus dem Buch „Stochastische Signaltheorie“ hinweisen, in dem die Eigenschaften des AWGN-Kanals ausführlich beschrieben werden:

**Der AWGN-Kanal – Teil 1** (Dauer 6:00)

**Der AWGN-Kanal – Teil 2** (Dauer 5:15)

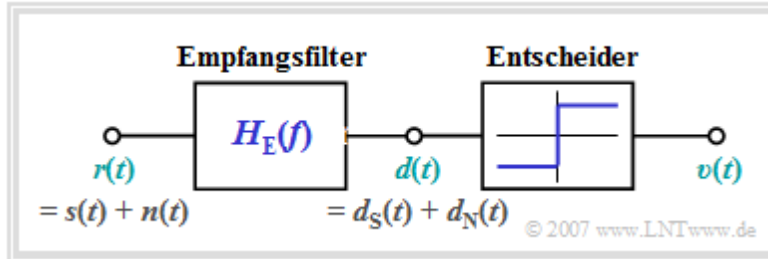
**Der AWGN-Kanal – Teil 3** (Dauer 6:15)



## Empfangsfilter und Schwellenwertentscheider (1)

Der einfachste Empfänger bei Binärübertragung über den AWGN-Kanal besteht aus

- einem **Empfangsfilter** mit dem Frequenzgang  $H_E(f)$  und
- einem **Schwellenwertentscheider** zur Gewinnung des Binärsignals.



Diese Empfängerstruktur ist wie folgt zu begründen:

- Das Signal  $d(t)$  nach dem Empfangsfilter kann zumindest gedanklich wie folgt aufgeteilt werden: Der Anteil  $d_S(t)$  ist auf das Nutzsignal  $s(t)$  zurückzuführen, der Anteil  $d_N(t)$  auf das Rauschen  $n(t)$ . Die beiden Indizes stehen hierbei für **S: Signal** und **N: Noise**. Mit der Impulsantwort  $h_E(t)$  als die Fourierrechtransformierte des Frequenzgangs  $H_E(f)$  gilt:

$$\begin{aligned}d_S(t) &= s(t) \star h_E(t), \\d_N(t) &= n(t) \star h_E(t).\end{aligned}$$

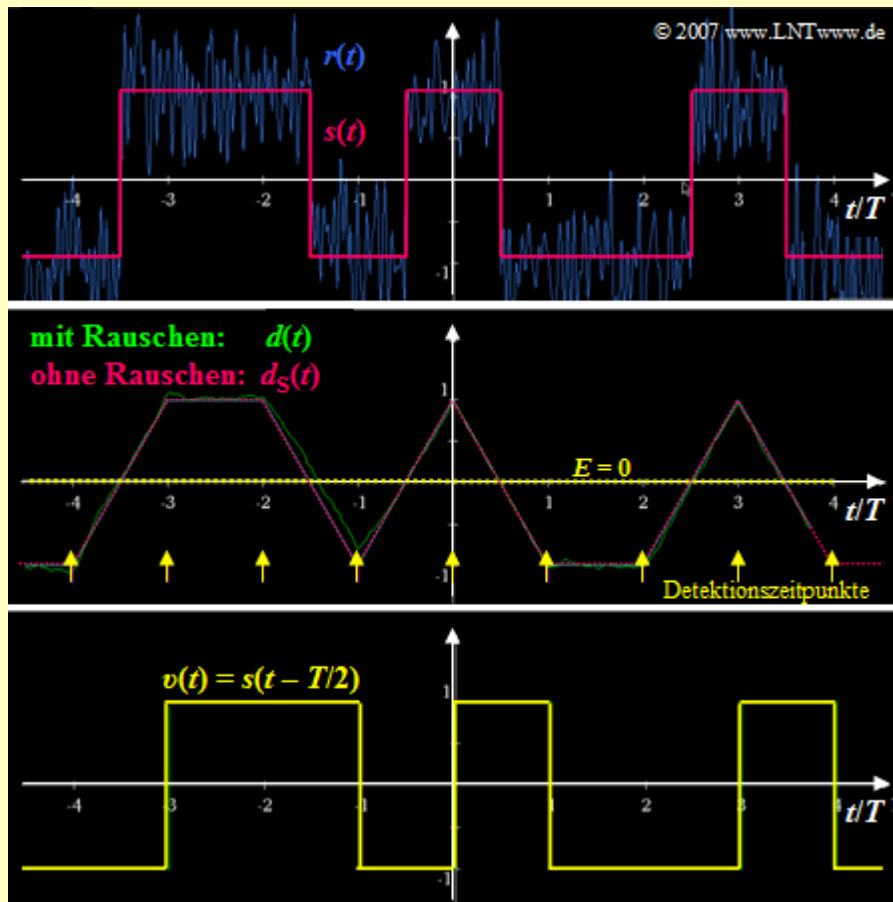
- Das weiße Rauschen  $n(t)$  am Empfängereingang besitzt theoretisch eine unendliche große Leistung (praktisch: eine unnötig große Leistung). Durch den Tiefpass mit dem Frequenzgang  $H_E(f)$  wird diese auf den quadratischen Erwartungswert des Detektionsstörsignals („Varianz“) begrenzt:

$$\sigma_d^2 = E[d_N(t)^2].$$

- Allerdings ist zu beachten, dass der Tiefpass  $H_E(f)$  nicht nur das Störsignal  $n(t)$ , sondern auch das Nutzsignal  $s(t)$  verändert. Dadurch werden die einzelnen Sendeimpulse verbreitert und in ihrer Amplitude vermindert. Nach den Voraussetzungen für dieses Kapitel muss sichergestellt werden, dass es nicht zu Impulsinterferenzen kommt.
- Aufgabe des Entscheiders ist es, aus dem wert- und zeitkontinuierlichen Detektionssignal  $d(t)$  das wert- und zeitdiskrete Sinkensignal  $v(t)$  zu erzeugen, das die Nachricht des Sendesignals  $s(t)$  „möglichst gut“ wiedergeben sollte. Die Funktionsweise des (binären) Schwellenwertentscheiders wird im Beispiel auf der nächsten Seite beschrieben.

## Empfangsfilter und Schwellenwertentscheider (2)

**Beispiel:** Die obere Grafik zeigt rot das rechteckförmige, auf  $\pm 1$  normierte Sendesignal  $s(t)$ , das von additivem Rauschen  $n(t)$  überlagert ist. Blau dargestellt ist das Empfangssignal  $r(t) = s(t) + n(t)$ .

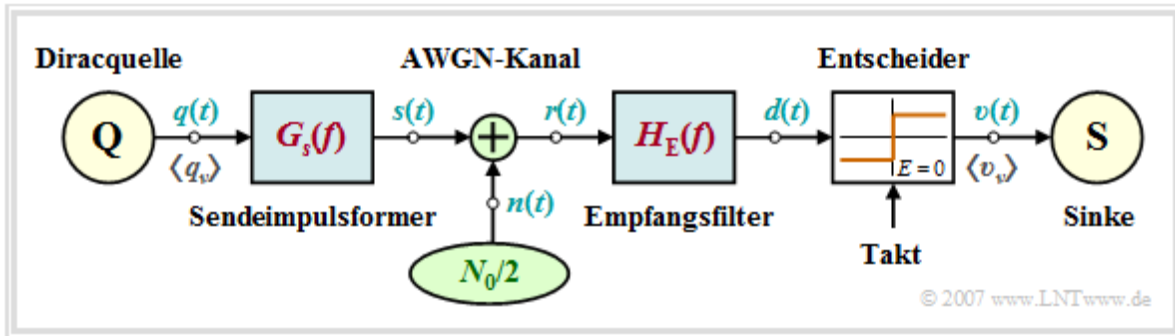


Zu dieser Grafik ist weiter anzumerken:

- Nach dem Empfangsfilter mit rechteckförmiger Impulsantwort der Dauer  $T$  ergibt sich das im mittleren Bild dargestellte Signal  $d(t)$ . Der Anteil  $d_S(t)$ , der ausschließlich auf das Sendesignal  $s(t)$  zurückgeht, hat in diesem Sonderfall den in der mittleren Grafik rot gepunktet dargestellten, abschnittsweise linearen Verlauf. Die Differenz  $d(t) - d_S(t)$  ist der Rauschanteil  $d_N(t)$ , der vom AWGN-Term  $n(t)$  herrührt.
- Der anschließende Schwellenwertentscheider wertet das Detektionssignal  $d(t)$  aus. Dazu vergleicht er die Detektionsabstastwerte zu den äquidistanten Detektionszeitpunkten – in der Grafik durch gelbe Pfeile markiert – mit dem Schwellenwert  $E = 0$  und setzt entsprechend das Sinkensignal  $v(t)$  im Bereich  $v \cdot T \dots (v + 1) \cdot T$  auf +1 oder -1, je nachdem, ob der Detektionsabstastwert  $d(vT)$  größer oder kleiner ist als die Entscheiderschwelle  $E$ .
- Trifft wie im dargestellten Beispiel der Entscheider stets die richtige Entscheidung, so ist sein Ausgangssignal  $v(t) = s(t - T/2)$ . Die Laufzeit von einer halben Symboldauer ( $T/2$ ) ist darauf zurückzuführen, dass das Detektionssignal  $d(t)$  stets in Symbolmitte entschieden wird, die Bereitstellung des Sinkensignals  $v(t)$  aber aus Kausalitätsgründen erst danach erfolgen kann.

## Ersatzschaltbild und Voraussetzungen für Kapitel 1

Für die weiteren Abschnitte dieses ersten Kapitels wird das folgende Ersatzschaltbild zugrunde gelegt:



Wenn nicht explizit etwas anderes angegeben ist, gelten die nachfolgend aufgeführten Voraussetzungen:

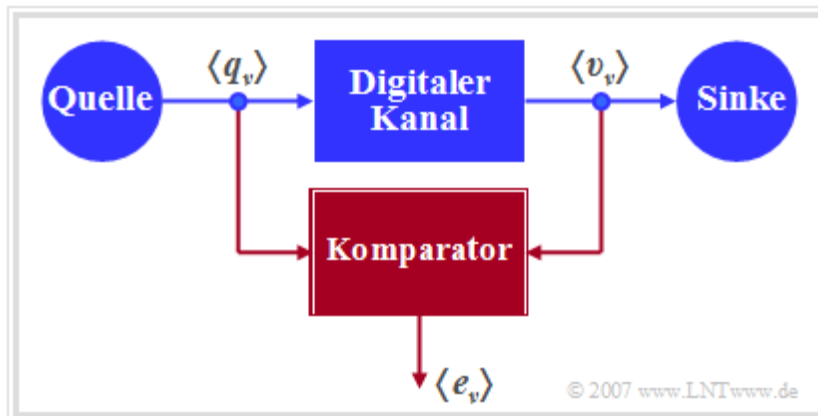
- Die Übertragung erfolgt binär, bipolar und redundanzfrei mit der Bitrate  $R = 1/T$ . Die mehrstufige und/oder redundante Digitalübertragung wird im **Kapitel 2** behandelt.
- Das Sendesignal  $s(t)$  ist zu allen Zeiten  $t$  gleich  $\pm s_0$ , das heißt: Der Sendegrundimpuls  $g_s(t)$  ist NRZ-rechteckförmig mit Amplitude  $s_0$  und Dauer  $T$ . Die Spektralfunktion lautet:

$$G_s(f) = s_0 \cdot T \cdot \text{si}(\pi f T) \quad \text{mit} \quad \text{si}(x) = \frac{\sin(x)}{x}.$$

- Für das Empfangssignal gelte  $r(t) = s(t) + n(t)$ , wobei der AWGN-Term  $n(t)$  durch die konstante einseitige (physikalische) Rauschleistungsdichte  $N_0$  gekennzeichnet ist. Der Kanalfrequenzgang ist somit stets  $H_K(f) = 1$  und muss nicht weiter berücksichtigt werden.
- Das Empfangsfilter mit dem Frequenzgang  $H_E(f)$  und der Impulsantwort  $h_E(t) = \mathcal{F}^{-1}[H_E(f)]$  ist optimal an den Sendegrundimpuls  $g_s(t)$  angepasst, so dass Impulsinterferenzen keine Rolle spielen. Impulsinterferenzbehaftete Systeme werden in **Kapitel 3** ausführlich behandelt.
- Die Parameter des (binären) Schwellenwertentscheiders sind optimal gewählt. Aufgrund der oben aufgelisteten Voraussetzungen (u.a. bipolare Signalisierung) ist die optimale Entscheidungsschwelle  $E = 0$  und die optimalen Detektionszeitpunkte liegen bei  $v \cdot T$ .

## Definition der Bitfehlerwahrscheinlichkeit

Die Grafik zeigt ein sehr einfaches, aber allgemeingültiges Modell eines binären Übertragungssystems.



Dieses lässt sich wie folgt charakterisieren:

- Die Quelle und die Senke werden durch die beiden Binärfolgen  $\langle q_v \rangle$  und  $\langle v_v \rangle$  beschrieben.
- Das gesamte Übertragungssystem – bestehend aus Sender, Übertragungskanal inklusive Störungen und Empfänger – wird als „Black Box“ mit binärem Ein- und Ausgang betrachtet.
- Dieser „Digitale Kanal“ wird allein durch die Fehlerfolge  $\langle e_v \rangle$  charakterisiert. Bei fehlerfreier Übertragung des  $v$ -ten Bits ( $v_v = q_v$ ) gilt  $e_v = 0$ , andernfalls ( $v_v \neq q_v$ ) wird  $e_v = 1$  gesetzt.

**Definition:** Die (mittlere) **Bitfehlerwahrscheinlichkeit** ist bei einem Binärsystem wie folgt gegeben:

$$p_B = E[\Pr(v_v \neq q_v)] = \overline{\Pr(v_v \neq q_v)} = \lim_{N \rightarrow \infty} \frac{1}{N} \cdot \sum_{v=1}^N \Pr(v_v \neq q_v).$$

Diese statistische Größe ist das wichtigste Beurteilungskriterium eines jeden Digitalsystems.

Die Berechnung als Erwartungswert  $E[\dots]$  gemäß dem ersten Teil der obigen Gleichung entspricht einer Scharmittelung über die Verfälschungswahrscheinlichkeit  $\Pr(v_v \neq q_v)$  des  $v$ -ten Symbols, während die überstreichende Linie im rechten Teil eine Zeitmittelung kennzeichnet. Beide Berechnungsarten führen – unter der gerechtfertigten Annahme ergodischer Prozesse – zum gleichen Ergebnis, wie im **Kapitel 4** des Buches „Stochastische Signaltheorie“ gezeigt wurde.

Auch aus der Fehlerfolge  $\langle e_v \rangle$  lässt sich die Bitfehlerwahrscheinlichkeit als Erwartungswert bestimmen, wobei zu berücksichtigen ist, dass  $e_v$  nur die Werte 0 und 1 annehmen kann:

$$p_B = E[\Pr(e_v = 1)] = E[e_v].$$

Die obige Definition der Bitfehlerwahrscheinlichkeit gilt unabhängig davon, ob es statistische Bindungen innerhalb der Fehlerfolge  $\langle e_v \rangle$  gibt oder nicht. Je nachdem ist der Aufwand zur Berechnung von  $p_B$  unterschiedlich groß und bei einer Systemsimulation müssen unterschiedliche digitale Kanalmodelle herangezogen werden.

Im **Kapitel 5** wird gezeigt, dass das sog. BSC-Modell (*Binary Symmetrical Channel*) statistisch unabhängige Fehler liefert, während für die Beschreibung von Bündelfehlerkanälen auf die Modelle von Gilbert-Elliott [Gil60] und McCullough [McC68] zurückgegriffen werden muss.

## Definition der Bitfehlerquote (1)

Die Bitfehlerwahrscheinlichkeit  $p_B$  eignet sich zum Beispiel gut für die Konzipierung und Optimierung von Digitalsystemen. Diese ist eine **Apriori-Kenngröße**, die eine Vorhersage über das Fehlverhalten eines Nachrichtensystems erlaubt, ohne dass dieses bereits realisiert sein muss.

Dagegen muss zur messtechnischen Erfassung der Qualität eines realisierten Systems oder bei einer Systemsimulation auf die Bitfehlerquote übergegangen werden, die durch den Vergleich von Quellen- und Sinkensymbolfolge ermittelt wird. Diese ist somit eine **Aposteriori-Kenngröße** des Systems.

**Definition:** Die **Bitfehlerquote** (englisch: *Bit Error Rate*, BER) ist das Verhältnis aus der Anzahl  $n_B(N)$  der aufgetretenen Bitfehler ( $v_v \neq q_v$ ) und der Anzahl  $N$  der insgesamt übertragenen Symbole:

$$h_B(N) = \frac{n_B(N)}{N}.$$

Im Sinne der Wahrscheinlichkeitsrechnung stellt die Bitfehlerquote eine relative Häufigkeit dar; sie wird demzufolge auch **Bitfehlerhäufigkeit** genannt.

Die Schreibweise  $h_B(N)$  soll deutlich machen, dass die per Messung oder durch Simulation ermittelte Bitfehlerquote signifikant von dem Parameter  $N$  – also der Anzahl der insgesamt übertragenen oder simulierten Symbole – abhängt. Nach den elementaren Gesetzen der Wahrscheinlichkeitsrechnung stimmt nur im Grenzfall  $N \rightarrow \infty$  die Aposteriori-Kenngröße  $h_B$  mit der Apriori-Kenngröße  $p_B$  exakt überein.

Der Zusammenhang zwischen Wahrscheinlichkeit und relativer Häufigkeit wird in einem Lernvideo zum Buch „Stochastische Signaltheorie“ verdeutlicht:

**Das Bernoullische Gesetz der großen Zahlen** (Dateigröße: 1.97 MB, Dauer: 4:25)

Für die nachfolgende Herleitung wird das **BSC-Modell** zugrunde gelegt, das in **Kapitel 5.2** im Detail beschrieben wird. Jedes einzelne Bit wird mit der Wahrscheinlichkeit  $p = \Pr(v_v \neq q_v) = \Pr(e_v = 1)$  verfälscht, unabhängig von den Fehlerwahrscheinlichkeiten der benachbarten Symbole. Die (mittlere) Bitfehlerwahrscheinlichkeit  $p_B$  ist somit ebenfalls gleich  $p$ .

## Definition der Bitfehlerquote (2)

Nun soll abgeschätzt werden, wie genau die Bitfehlerwahrscheinlichkeit  $p_B = p$  beim BSC-Modell durch die Bitfehlerquote  $h_B$  approximiert wird. Dies geschieht in mehreren Schritten:

- Die Anzahl der Bitfehler bei der Übertragung von  $N$  Symbolen ist eine diskrete Zufallsgröße:

$$n_B(N) = \sum_{\nu=1}^N e_{\nu} \in \{0, 1, \dots, N\}.$$

- Bei statistisch unabhängigen Fehlern (BSC) ist  $n_B$  **binominalverteilt**. Demzufolge gilt:

$$m_{n_B} = N \cdot p_B, \quad \sigma_{n_B} = \sqrt{N \cdot p_B \cdot (1 - p_B)}.$$

- Für Mittelwert und Streuung der Bitfehlerquote  $h_B = n_B/N$  gilt deshalb:

$$m_{h_B} = \frac{m_{n_B}}{N} = p_B, \quad \sigma_{h_B} = \frac{\sigma_{n_B}}{N} = \sqrt{\frac{p_B \cdot (1 - p_B)}{N}}.$$

- Nach **Moivre** und **Laplace** lässt sich die Binominalverteilung in eine Gaußverteilung überführen:

$$f_{h_B}(h_B) \approx \frac{1}{\sqrt{2\pi} \cdot \sigma_{h_B}} \cdot \exp\left[-\frac{(h_B - p_B)^2}{2 \cdot \sigma_{h_B}^2}\right].$$

- Mit dem **Gaußschen Fehlerintegral**  $Q(x)$  lässt sich somit die Wahrscheinlichkeit  $p_{\varepsilon}$  berechnen, dass die per Simulation/Messung über  $N$  Symbole ermittelte Bitfehlerquote  $h_B(N)$  betragsmäßig um weniger als einen Wert  $\varepsilon$  von der tatsächlichen Bitfehlerwahrscheinlichkeit  $p_B$  abweicht:

$$p_{\varepsilon} = \Pr(|h_B(N) - p_B| < \varepsilon) = 1 - 2 \cdot Q\left(\frac{\varepsilon}{\sigma_{h_B}}\right) = 1 - 2 \cdot Q\left(\frac{\varepsilon \cdot \sqrt{N}}{\sqrt{p_B \cdot (1 - p_B)}}\right).$$

Dieses Ergebnis ist wie folgt zu interpretieren: Wenn man unendlich viele Versuchsreihen über jeweils  $N$  Symbole durchführt, ist der Mittelwert  $m_{h_B}$  tatsächlich gleich der gesuchten Fehlerwahrscheinlichkeit  $p_B$ . Bei einer einzigen Versuchsreihe wird man dagegen nur eine Näherung erhalten, wobei die jeweilige Abweichung vom Sollwert bei mehreren Versuchsreihen gaußverteilt ist.

**Beispiel:** Die Bitfehlerwahrscheinlichkeit betrage  $p_B = 10^{-3}$  und es ist bekannt, dass die Bitfehler statistisch unabhängig sind. Macht man nun sehr viele Versuchsreihen mit jeweils  $N = 10^5$  Symbolen, so werden die jeweiligen Ergebnisse  $h_B$  entsprechend einer Gaußverteilung um den Sollwert  $10^{-3}$  variieren. Die Streuung beträgt dabei

$$\sigma_{h_B} = \sqrt{p_B \cdot (1 - p_B)/N} \approx 10^{-4}.$$

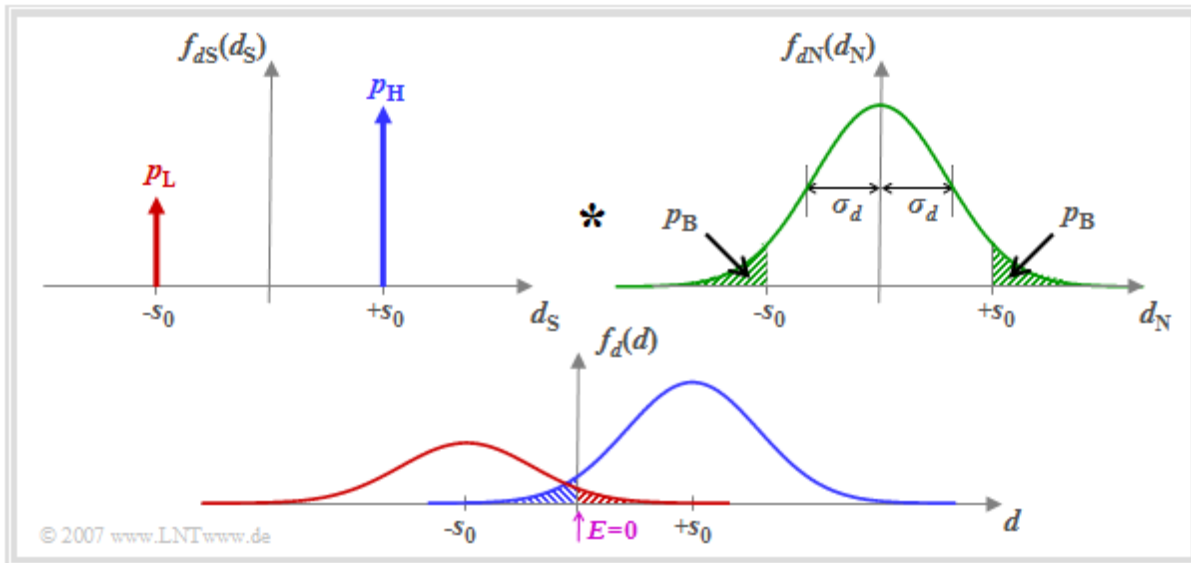
Die Wahrscheinlichkeit, dass die relative Häufigkeit einen Wert zwischen  $0.9 \cdot 10^{-3}$  und  $1.1 \cdot 10^{-3}$  ( $\varepsilon = 10^{-4}$ ) haben wird, ist somit gleich  $p_{\varepsilon} = 1 - 2 \cdot Q(\varepsilon/\sigma_{h_B}) = 1 - 2 \cdot Q(1) \approx 68.4\%$ . Soll diese Wahrscheinlichkeit (Genauigkeit) auf 95% gesteigert werden, so müsste  $N$  auf 400 000 erhöht werden.

## Fehlerwahrscheinlichkeit bei Gaußischem Rauschen (1)

Entsprechend den **Voraussetzungen** zu diesem Kapitel gehen wir davon aus, dass das Detektionssignal zu den Detektionszeitpunkten wie folgt dargestellt werden kann:

$$d(\nu T) = d_S(\nu T) + d_N(\nu T).$$

Der Nutzanteil wird durch die Wahrscheinlichkeitsdichtefunktion (WDF)  $f_{d_S}(d_S)$  beschrieben, wobei wir hier von unterschiedlichen Auftretswahrscheinlichkeiten  $p_L = \Pr(d_S = -s_0)$ ,  $p_H = \Pr(d_S = +s_0) = 1 - p_L$  ausgehen. Die WDF  $f_{d_N}(d_N)$  der Störkomponente ist gaußförmig und besitzt die Streuung  $\sigma_d$ .



Die WDF  $f_d(d)$  der Detektionsabstastwerte  $d(\nu T)$  ergibt sich unter der Voraussetzung, dass  $d_S(\nu T)$  und  $d_N(\nu T)$  statistisch unabhängig voneinander sind („signalunabhängiges Rauschen“), als Faltungsprodukt:

$$f_d(d) = f_{d_S}(d_S) * f_{d_N}(d_N).$$

Der Schwellenwertentscheider mit der Schwelle  $E = 0$  trifft dann eine falsche Entscheidung, wenn

- das Symbol **L** gesendet wurde ( $d_S = -s_0$ ) und  $d > 0$  ist (rote schraffierte Fläche), oder
- das Symbol **H** gesendet wurde ( $d_S = +s_0$ ) und  $d < 0$  ist (blaue schraffierte Fläche).

Da die Flächen der zwei Gaußkurven zusammen 1 ergeben, gibt die Summe aus der rot und der blau schraffierten Fläche die Bitfehlerwahrscheinlichkeit  $p_B$  an. Die beiden grün schraffierten Flächen in der oberen Wahrscheinlichkeitsdichtefunktion  $f_{d_N}(d_N)$  sind – jede für sich – ebenfalls gleich  $p_B$ .

Die anhand der Grafik veranschaulichten Ergebnisse sollen nun formelmäßig hergeleitet werden. Es gilt:

$$p_B = p_L \cdot \Pr(v_\nu = \mathbf{H} \mid q_\nu = \mathbf{L}) + p_H \cdot \Pr(v_\nu = \mathbf{L} \mid q_\nu = \mathbf{H}).$$

Hierbei sind  $p_L$  und  $p_H$  die Quellensymbolwahrscheinlichkeiten, während die jeweils zweiten, bedingten Wahrscheinlichkeiten  $\Pr(v_\nu \mid q_\nu)$  die Verfälschungen durch den AWGN-Kanal beschreiben. Aus der Entscheidungsregel des Schwellenwertentscheiders (mit Schwelle  $E = 0$ ) ergibt sich auch:

$$\begin{aligned} p_B &= p_L \cdot \Pr(d(\nu T) > 0) + p_H \cdot \Pr(d(\nu T) < 0) = \\ &= p_L \cdot \Pr(d_N(\nu T) > s_0) + p_H \cdot \Pr(d_N(\nu T) < -s_0). \end{aligned}$$

Die Herleitung wird auf der nächsten Seite fortgesetzt.



## Fehlerwahrscheinlichkeit bei Gaußischem Rauschen (2)

Die weitere Herleitung soll nun Schritt für Schritt erfolgen. Ausgegangen wird von der Gleichung:

$$\begin{aligned} p_B &= p_L \cdot \Pr(d(\nu T) > 0) + p_H \cdot \Pr(d(\nu T) < 0) = \\ &= p_L \cdot \Pr(d_N(\nu T) > s_0) + p_H \cdot \Pr(d_N(\nu T) < -s_0). \end{aligned}$$

- Die beiden Überschreitungswahrscheinlichkeiten in obiger Gleichung sind aufgrund der Symmetrie der Gaußschen WDF  $f_{d_N}(d_N)$  gleich und es gilt:

$$p_B = (p_L + p_H) \cdot \Pr(d_N(\nu T) > s_0) = \Pr(d_N(\nu T) > s_0).$$

Das bedeutet, dass die Bitfehlerwahrscheinlichkeit bei einem Binärsystem mit der Schwelle  $E = 0$  nicht von den Symbolwahrscheinlichkeiten  $p_L$  und  $p_H = 1 - p_L$  abhängt.

- Die Wahrscheinlichkeit, dass der gaußverteilte Stör- bzw. Rauschterm mit Streuung  $\sigma_d$  größer ist als die NRZ-Sendeimpulsamplitude  $s_0$ , ergibt sich zu

$$p_B = \int_{s_0}^{+\infty} f_{d_N}(d_N) dd_N = \frac{1}{\sqrt{2\pi} \cdot \sigma_d} \int_{s_0}^{+\infty} \exp\left[-\frac{d_N^2}{2\sigma_d^2}\right] dd_N.$$

- Unter Verwendung des **komplementären Gaußschen Fehlerintegrals** lautet das Ergebnis:

$$p_B = Q\left(\frac{s_0}{\sigma_d}\right) \quad \text{mit} \quad Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-u^2/2} du.$$

- Häufig – insbesondere in der englischsprachigen Literatur – wird anstelle der Funktion  $Q(x)$  die vergleichbare **komplementäre Error Function**  $\text{erfc}(x)$  verwendet. Mit dieser gilt:

$$p_B = 1/2 \cdot \text{erfc}\left(\frac{s_0}{\sqrt{2} \cdot \sigma_d}\right) \quad \text{mit} \quad \text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-u^2} du.$$

Beide Funktionen findet man in Formelsammlungen in tabellarischer Form. Nachfolgend finden Sie ein Interaktionsmodul zur Berechnung der Funktionswerte von  $Q(x)$  und  $1/2 \cdot \text{erfc}(x)$ :

### Komplementäre Gaußsche Fehlerfunktionen

Zur Auswertung obiger Gleichungen ist nach Ansicht des Autors bei der Basisbandübertragung die  $Q$ -Funktion besser geeignet. Aber das ist natürlich Geschmacksache.

**Beispiel:** Für das Folgende wird vorausgesetzt, dass Tabellen zur Verfügung stehen, in denen das Argument der Gaußschen Fehlerfunktionen im Abstand 0.1 aufgelistet sind.

Mit  $s_0/\sigma_d = 4$  erhält man für die Bitfehlerwahrscheinlichkeit:

$$\begin{aligned} p_B &= Q(4) = 0.317 \cdot 10^{-4}, \\ p_B &= 1/2 \cdot \text{erfc}(4/\sqrt{2}) = 1/2 \cdot \text{erfc}(2.828) \approx 1/2 \cdot \text{erfc}(2.8) = 0.375 \cdot 10^{-4}. \end{aligned}$$

Richtig ist der erste Wert. Bei der zweiten Berechnungsart muss man runden oder – noch besser – interpolieren, was aufgrund der starken Nichtlinearität dieser Funktion sehr schwierig ist.

Außerhalb von Übungsbeispielen wird  $s_0/\sigma_d$  in der Regel einen „krummen“ Wert besitzen. In diesem Fall bietet „ $Q(x)$ “ natürlich keinen Vorteil gegenüber „ $\text{erfc}(x)$ “.

## Optimaler Binärempfänger - Realisierung mit Matched-Filter (1)

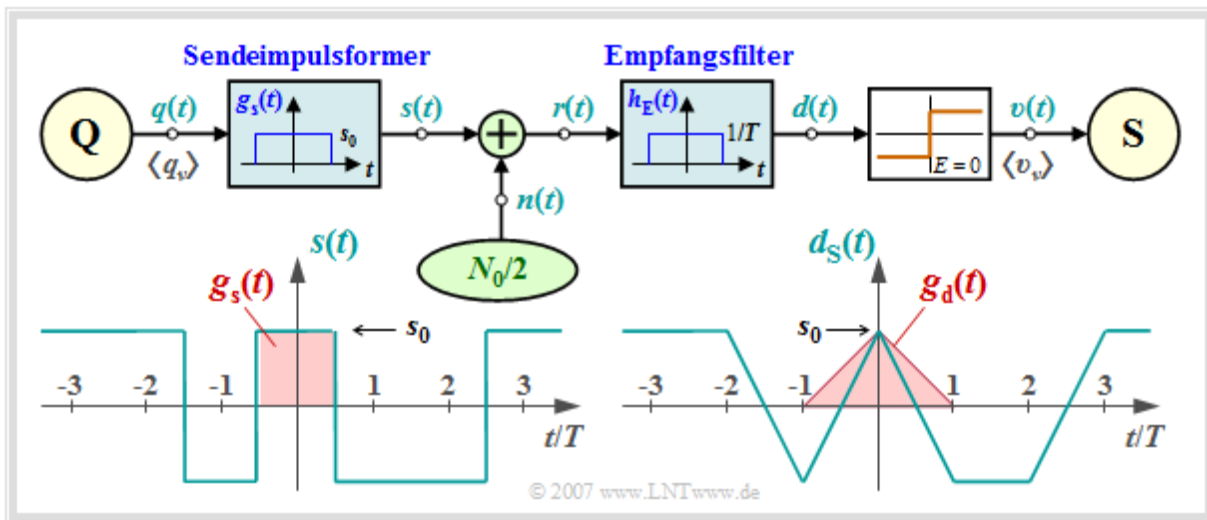
Für das Folgende wird von den im Kapitel 1.1 genannten **Voraussetzungen** ausgegangen. Dann kann man für den Frequenzgang und die Impulsantwort des Empfängerfilters ansetzen:

$$H_E(f) = \text{si}(\pi f T) \quad \bullet \text{---} \circ \quad h_E(t) = \begin{cases} 1/T & \text{für } |t| < T/2, \\ 1/(2T) & \text{für } |t| = T/2, \\ 0 & \text{für } |t| > T/2. \end{cases}$$

Aufgrund der Linearität kann für das **Detektionsnutzsignal** geschrieben werden:

$$d_S(t) = \sum_{(\nu)} a_\nu \cdot g_d(t - \nu \cdot T) \quad \text{mit} \quad g_d(t) = g_s(t) \star h_E(t).$$

Die Faltung zweier Rechtecke gleicher Breite  $T$  ergibt einen dreieckförmigen **Detektionsgrundimpuls**  $g_d(t)$  mit  $g_d(t=0) = s_0$ . Wegen  $g_d(|t| \geq T) = 0$  ist das System impulsinterferenzfrei; es gilt  $d_S(\nu T) = \pm s_0$ .



Die Varianz des Detektionsstörsignals  $d_N(t)$  – also die **Detektionsstörleistung** – lautet:

$$\sigma_d^2 = \frac{N_0}{2} \cdot \int_{-\infty}^{+\infty} |H_E(f)|^2 df = \frac{N_0}{2} \cdot \int_{-\infty}^{+\infty} \text{si}^2(\pi f T) df = \frac{N_0}{2T}.$$

Damit ergeben sich für die **Bitfehlerwahrscheinlichkeit** folgende Gleichungen:

$$p_B = Q \left( \sqrt{\frac{2 \cdot s_0^2 \cdot T}{N_0}} \right) = Q(\sqrt{\rho_d}),$$

$$p_B = 1/2 \cdot \text{erfc} \left( \sqrt{s_0^2 \cdot T / N_0} \right) = 1/2 \cdot \text{erfc} \left( \sqrt{\rho_d / 2} \right).$$

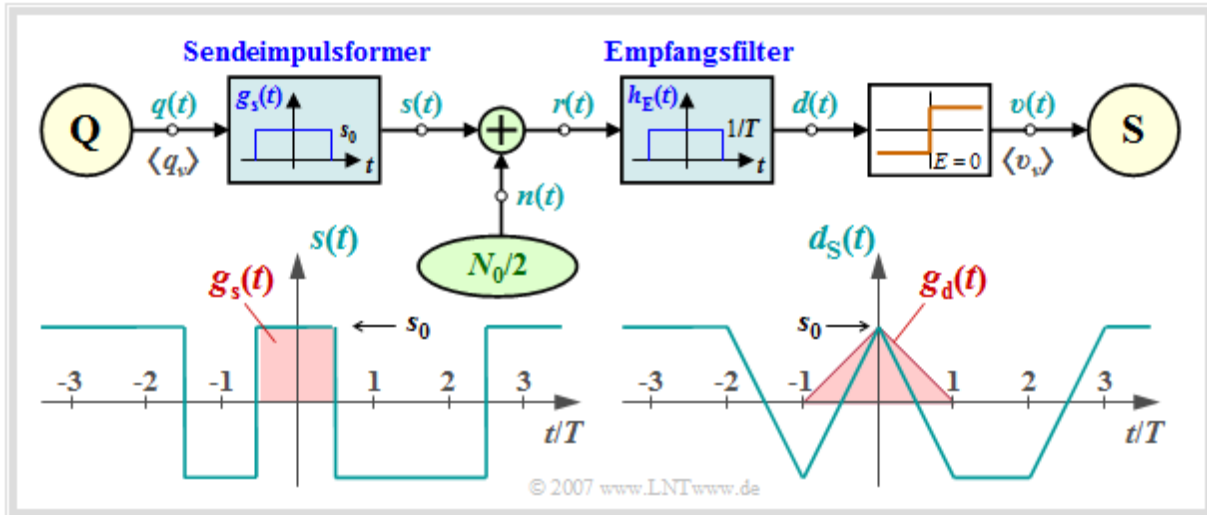
$\rho_d$  ist das momentane **Signal-zu-Stör-Leistungsverhältnis** (SNR) des Detektionssignals  $d(t)$  zu den Zeitpunkten  $\nu T$ . Wir nennen es im Folgenden kurz „Detektions-SNR“. Es gilt die Definition:

$$\rho_d = \frac{d_S^2(\nu T)}{E[d_N^2(\nu T)]} = \frac{s_0^2}{\sigma_d^2}.$$

## Optimaler Binärempfänger - Realisierung mit Matched-Filter (2)

Ein Vergleich der Ergebnisse der letzten Seite mit **Kapitel 5.4** von Buch „Stochastische Signaltheorie“ zeigt, dass das Empfangsfilter  $H_E(f)$  ein an den Sendegrundimpuls  $g_s(t)$  angepasstes **Matched-Filter** ist:

$$H_E(f) = H_{MF}(f) = K_{MF} \cdot G_s^*(f).$$



Gegenüber der Seite **Matched-Filter-Optimierung** sind hier folgende Modifikationen berücksichtigt:

- Die Matched-Filter-Konstante ist hier zu  $K_{MF} = 1/(s_0 \cdot T)$  gesetzt. Damit ist der Frequenzgang  $H_{MF}(f)$  dimensionslos.
- Der im allgemeinen frei wählbare Detektionszeitpunkt ist hier zu  $T_D = 0$  gewählt. Damit ergibt sich allerdings ein akausales Filter.
- Das Detektions-SNR kann für jeden beliebigen Sendegrundimpuls  $g_s(t)$  mit Spektrum  $G_s(f)$  wie folgt dargestellt werden, wobei sich die rechte Identität aus dem Parsevalschen Theorem ergibt:

$$\rho_d = \frac{2 \cdot E_B}{N_0} \quad \text{mit} \quad E_B = \int_{-\infty}^{+\infty} g_s^2(t) dt = \int_{-\infty}^{+\infty} |G_s(f)|^2 df.$$

- $E_B$  wird oft als **Energie pro Bit** bezeichnet und  $E_B/N_0$  – fälschlicherweise – als SNR. Bei binärer Basisbandübertragung unterscheidet sich  $E_B/N_0$  vom tatsächlichen SNR  $\rho_d$  um den Faktor 2.
- Die auf der letzten Seite hergeleitete Bitfehlerwahrscheinlichkeit kann somit auch in der folgenden Weise geschrieben werden:

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

Die Eigenschaften des Matched-Filters werden in folgendem Interaktionsmodul behandelt:

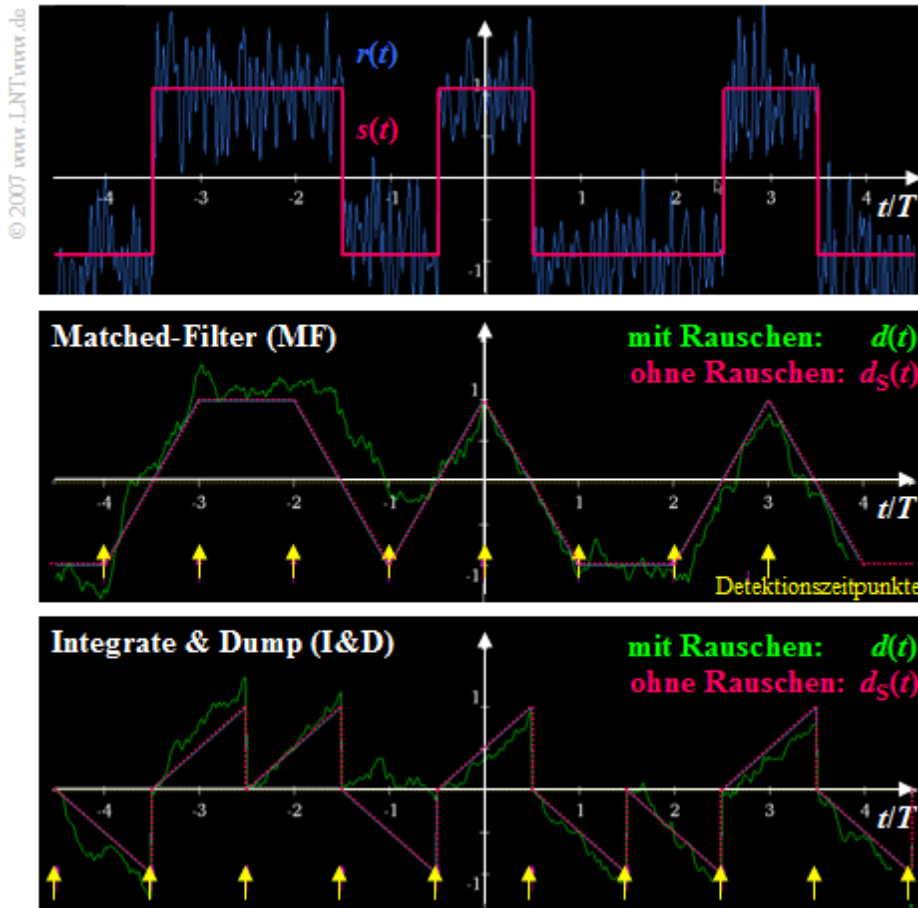
### Zur Verdeutlichung des Matched-Filters

## Optimaler Binärempfänger – Realisierungsform Integrate & Dump

Bei rechteckförmigen NRZ-Sendeimpulsen kann das Matched-Filter auch als Integrator (jeweils über eine Symboldauer  $T$ ) realisiert werden. Damit gilt für das Detektionssignal zu den Detektionszeitpunkten:

$$d(\nu \cdot T + T/2) = \frac{1}{T} \cdot \int_{\nu \cdot T - T/2}^{\nu \cdot T + T/2} r(t) dt.$$

Die folgende Grafik verdeutlicht die Unterschiede bei der Realisierung des optimalen Binärempfängers mit Matched-Filter (MF)  $\Rightarrow$  mittlere Skizze bzw. als „Integrate & Dump“ (I&D)  $\Rightarrow$  untere Skizze.



Man erkennt aus diesen Signalverläufen:

- Das Detektionssignal  $d_S(t)$  ist zu den Detektionszeitpunkten in beiden Fällen gleich  $\pm s_0$ .
- Die unterschiedlichen Detektionszeitpunkte sind darauf zurückzuführen, dass das Matched-Filter im Gegensatz zu „Integrate & Dump“ als akausal angesetzt wurde (siehe letzte Seite).
- Beim MF-Empfänger ist die Varianz des Detektionsstörsignals zu allen Zeiten  $t$  gleich:

$$E[d_N^2(t)] = \sigma_d^2 = \text{const.}$$

Dagegen nimmt beim I&D-Empfänger die Varianz von Symbolanfang bis Symbolende zu.

- Zu den Detektionszeitpunkten  $\nu T$  ist die Detektionsstörleistung in beiden Fällen gleich, so dass sich die genau gleiche Bitfehlerwahrscheinlichkeit ergibt. Mit  $E_B = s_0^2 \cdot T$  gilt nämlich:

$$\sigma_d^2 = \frac{N_0}{2} \cdot \int_{-\infty}^{+\infty} \text{si}^2(\pi fT) df = \frac{N_0}{2T} \Rightarrow p_B = Q\left(\sqrt{s_0^2/\sigma_d^2}\right) = Q\left(\sqrt{2 \cdot E_B/N_0}\right).$$

## Interpretation des optimalen Empfängers

In diesem Abschnitt wurde gezeigt, dass mit einem Empfänger, bestehend aus linearem Empfangsfilter und nichtlinearem Entscheider, die kleinstmögliche Bitfehlerwahrscheinlichkeit zu erreichen ist:

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

Die sich ergebende Konfiguration ist ein Sonderfall des sog. **Maximum-A-posteriori-Empfängers** (MAP), der im **Kapitel 3.7** dieses Buches behandelt wird.

Für die Gültigkeit obiger Gleichung müssen allerdings eine Reihe von Voraussetzungen erfüllt sein:

- Das Sendesignal  $s(t)$  ist binär sowie bipolar (antipodisch) und weist pro Bit die (mittlere) Energie  $E_B$  auf. Die (mittlere) Sendeleistung ist somit  $E_B/T$ .
- Es liegt ein AWGN-Kanal (*Additive White Gaussian Noise*) mit der konstanten (einseitigen) Rauschleistungsdichte  $N_0$  vor.
- Das Empfangsfilter  $H_E(f)$  ist bestmöglich an das Sendegrundimpulsspektrum  $G_s(f)$  entsprechend dem Matched-Filter-Kriterium angepasst.
- Der Entscheider (Schwellenwert, Detektionszeitpunkte) sei optimal. Eine kausale Realisierung des Matched-Filters kann man durch Verschiebung des Detektionszeitpunktes ausgleichen.
- Obige Gleichung gilt unabhängig vom Sendegrundimpuls  $g_s(t)$ . Allein die für die Übertragung eines Binärsymbols aufgewendete Energie  $E_B$  ist entscheidend für die Fehlerwahrscheinlichkeit.
- Voraussetzung für die Anwendbarkeit obiger Gleichung ist, dass die Detektion eines Symbols nicht durch andere Symbole beeinträchtigt wird. Solche Impulsinterferenzen vergrößern  $p_B$  enorm.
- Ist die absolute Sendeimpulsdauer  $T_S$  kleiner oder gleich dem Symbolabstand  $T$ , so ist obige Gleichung bei Erfüllung des Matched-Filter-Kriteriums immer anwendbar.
- Die Gleichung gilt auch für Nyquistsysteme, bei denen zwar  $T_S > T$  gilt, es aber aufgrund von äquidistanten Nulldurchgängen des Grundimpulses  $g_d(t)$  nicht zu Impulsinterferenzen kommt.

## Erstes Nyquistkriterium im Zeitbereich

Für dieses Kapitel wurde vorausgesetzt, dass die Detektion eines Symbols nicht durch Nachbarimpulse beeinträchtigt werden soll. Dies erreicht man durch die Detektion des Signals

$$d(t) = \sum_{(\nu)} a_{\nu} \cdot g_d(t - \nu T)$$

zu den Zeitpunkten  $\nu T$  immer dann, wenn der Detektionsgrundimpuls  $g_d(t)$

- auf den Bereich  $|t| < T$  beschränkt ist, was für das **Kapitel 1.2** vorausgesetzt wurde, oder
- äquidistante Nulldurchgänge zu den Zeitpunkten  $\nu T$  aufweist.

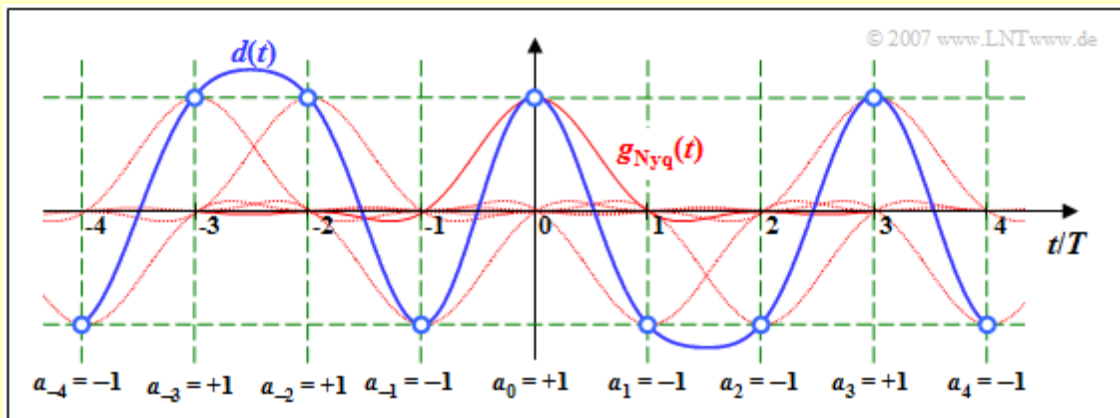
Aus Gründen einer möglichst einfachen Darstellung wird im Kapitel 1.3 das Detektionsstörsignal  $d_N(t)$  als vernachlässigbar klein angenommen.

**Definition:** Man bezeichnet einen Detektionsgrundimpuls mit den Eigenschaften

$$g_d(t = \nu T) = 0 \quad \text{für } \nu = \pm 1, \pm 2, \pm 3, \dots$$

als **Nyquistimpuls**  $g_{\text{Nyq}}(t)$ , benannt nach dem Physiker **Harry Nyquist**.

**Beispiel:** Die Grafik zeigt das Detektionssignal  $d(t)$  eines solchen Nyquistsystems. Rot gepunktet sind die (gewichteten und verschobenen) Nyquistimpulse  $a_{\nu} \cdot g_{\text{Nyq}}(t - \nu T)$  eingezeichnet.



Zu den Detektionszeitpunkten gilt  $d(\nu T) = a_{\nu} \cdot g_{\text{Nyq}}(0)$ , wie aus den blauen Kreisen und dem grünen Raster hervorgeht. Die Nachläufer der vorangegangenen Impulse ( $\nu < 0$ ) sowie die Vorläufer der nachfolgenden Impulse ( $\nu > 0$ ) beeinflussen beim Nyquistsystem die Detektion des Symbols  $a_0$  nicht.

Der Vollständigkeit halber sei erwähnt, dass für diese Grafik der Detektionsgrundimpuls

$$g_{\text{Nyq}}(t) = g_0 \cdot \text{si}\left(\frac{\pi \cdot t}{T}\right) \cdot \text{si}\left(\frac{\pi \cdot t}{2 \cdot T}\right)$$

mit trapezförmigem Spektrum und dem Rolloff-Faktor  $r = 0.5$  zugrunde liegt. Dieser wurde bereits im **Kapitel 3** des Buches „Signaldarstellung“ behandelt.

## Erstes Nyquistkriterium im Frequenzbereich

Harry Nyquist hat die Bedingung für eine impulsinterferenzfreie Detektion nicht nur für den Zeitbereich formuliert, sondern 1928 auch das entsprechende Kriterium im Frequenzbereich angegeben.

**Erstes Nyquistkriterium:** Erfüllt das Spektrum  $G_d(f)$  des Detektionsgrundimpulses die Bedingung

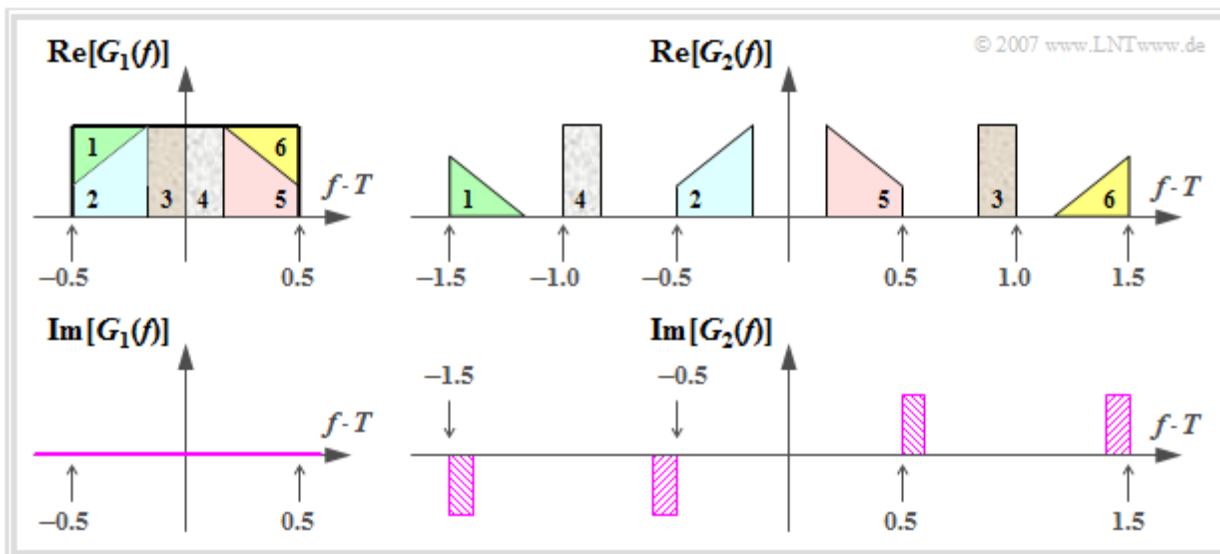
$$\sum_{k=-\infty}^{+\infty} G_d \left( f - \frac{k}{T} \right) = g_0 \cdot T = \text{const.},$$

so ist  $g_d(t)$  ein Nyquistimpuls mit äquidistanten Nulldurchgängen zu den Zeitpunkten  $vT$  ( $v \neq 0$ ) und der Amplitude  $g_d(t=0) = g_0$ . *Hinweis:* Sie finden den Beweis auf **der nächsten Seite**.

Die nachfolgende Grafik zeigt zwei Nyquistspektren. Das Spektrum

$$G_1(f) = \begin{cases} g_0 \cdot T & \text{für } |f| < 1/(2T), \\ 0 & \text{für } |f| > 1/(2T) \end{cases}$$

erfüllt offensichtlich die oben formulierte Bedingung und zwar mit der kleinstmöglichen Bandbreite. Der dazugehörige Nyquistimpuls  $g_1(t) = g_0 \cdot \text{si}(\pi t/T)$  klingt sehr langsam ab, nämlich asymptotisch mit  $1/t$ .



Der rechts oben dargestellte Realteil des Spektrums  $G_2(f)$  wurde aus dem Rechteckspektrum  $G_1(f)$  durch Verschiebung von Teilstücken um  $1/T$  nach rechts oder links konstruiert. Wegen

$$\sum_{k=-\infty}^{+\infty} \text{Re} \left[ G_2 \left( f - \frac{k}{T} \right) \right] = g_0 \cdot T, \quad \sum_{k=-\infty}^{+\infty} \text{Im} \left[ G_2 \left( f - \frac{k}{T} \right) \right] = 0$$

handelt es sich bei  $G_2(f)$  ebenfalls um ein Nyquistspektrum. Beim Imaginärteil heben sich die jeweils gleich schraffierten Anteile, die jeweils um  $2/T$  auseinander liegen, auf. Die Angabe des dazugehörigen Nyquistimpulses  $g_2(t)$  ist allerdings sehr kompliziert.



## Beweis des ersten Nyquistkriteriums

1. Wir gehen von der ersten Nyquistbedingung im Zeitbereich aus:

$$g_{\text{Nyq}}(\nu T) = \begin{cases} g_0 & \text{für } \nu = 0, \\ 0 & \text{für } \nu \neq 0. \end{cases}$$

2. Aus dem zweiten Fourierintegral erhält man somit für  $\nu \neq 0$ :

$$g_{\text{Nyq}}(\nu T) = \int_{-\infty}^{+\infty} G_{\text{Nyq}}(f) \cdot e^{j2\pi f \nu T} df = 0.$$

3. Zerlegt man das Fourierintegral in Teilintegrale der Breite  $1/T$ , so lauten die Bedingungsgleichungen:

$$\sum_{k=-\infty}^{+\infty} \int_{(k-1/2)/T}^{(k+1/2)/T} G_{\text{Nyq}}(f) \cdot e^{j2\pi f \nu T} df = 0.$$

4. Mit der Substitution  $f' = f + k/T$  folgt daraus:

$$\sum_{k=-\infty}^{+\infty} \int_{-1/(2T)}^{1/(2T)} G_{\text{Nyq}}\left(f' - \frac{k}{T}\right) \cdot e^{j2\pi \cdot (f' - k/T) \cdot \nu T} df' = 0.$$

5. Für alle ganzzahligen Werte von  $k$  und  $\nu$  gilt:

$$e^{-j2\pi k \nu} = 1 \quad \Rightarrow \quad \sum_{k=-\infty}^{+\infty} \int_{-1/(2T)}^{1/(2T)} G_{\text{Nyq}}\left(f' - \frac{k}{T}\right) \cdot e^{j2\pi f' \nu T} df' = 0.$$

6. Durch Vertauschen von Summation und Integration sowie Umbenennen von  $f'$  in  $f$  folgt weiter:

$$\int_{-1/(2T)}^{1/(2T)} \sum_{k=-\infty}^{+\infty} G_{\text{Nyq}}\left(f - \frac{k}{T}\right) \cdot e^{j2\pi f \nu T} df = 0.$$

7. Diese Forderung ist für alle  $\nu \neq 0$  nur dann zu erfüllen, wenn die unendliche Summe unabhängig von  $f$  ist, also einen konstanten Wert besitzt:

$$\sum_{k=-\infty}^{+\infty} G_{\text{Nyq}}\left(f - \frac{k}{T}\right) = K_{\text{Nyq}}.$$

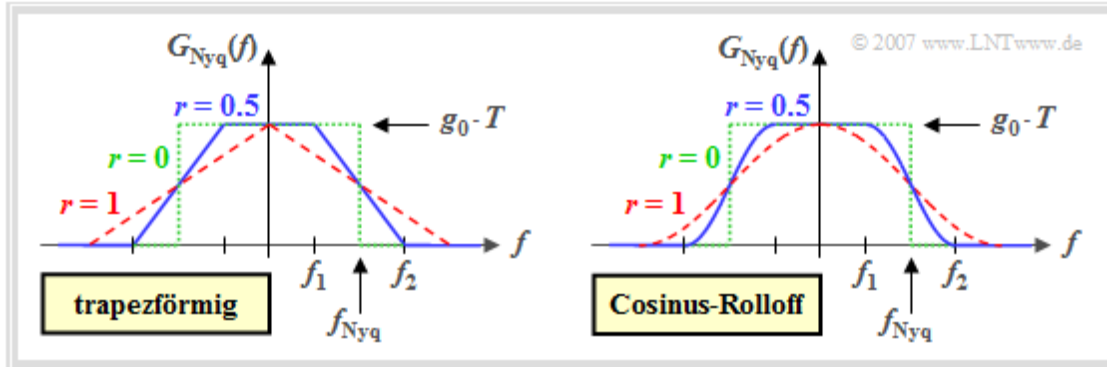
8. Aus der vorletzten Gleichung erhält man gleichzeitig für  $\nu = 0$ :

$$\int_{-1/(2T)}^{1/(2T)} \sum_{k=-\infty}^{+\infty} G_{\text{Nyq}}\left(f - \frac{k}{T}\right) df = K_{\text{Nyq}} \cdot \frac{1}{T} = g_0 \quad \Rightarrow \quad K_{\text{Nyq}} = g_0 \cdot T.$$

q.e.d.

## 1/T-Nyquistspektren (1)

Eine besondere Bedeutung für die Digitalsignalübertragung besitzen solche Nyquistspektren, die auf den Frequenzbereich  $-1/T \leq f \leq +1/T$  beschränkt und zusammenhängend sind. Die Grafik zeigt mit der **Trapez-Charakteristik** und der **Cosinus-Rolloff-Charakteristik** zwei diesbezügliche Varianten.



Für beide Nyquistspektren gilt in gleicher Weise:

- Der Flankenabfall erfolgt zwischen den zwei Eckfrequenzen  $f_1$  und  $f_2$  punktsymmetrisch um die **Nyquistfrequenz**  $f_{Nyq} = 1/(2T) = (f_1 + f_2)/2$ . Das heißt, dass für  $0 \leq f \leq f_{Nyq}$  gilt:

$$G_{Nyq}(f_{Nyq} + f) + G_{Nyq}(f_{Nyq} - f) = g_0 \cdot T.$$

- $G_{Nyq}(f)$  ist für alle Frequenzen  $|f| \leq f_1$  konstant gleich  $g_0 \cdot T$  und für  $|f| \geq f_2$  identisch 0. Im Bereich zwischen  $f_1$  und  $f_2$  gilt:

$$\frac{G_{Nyq}(f)}{g_0 \cdot T} = \begin{cases} \frac{f_2 - |f|}{f_2 - f_1} & \text{beim Trapez,} \\ \cos^2\left(\frac{\pi}{2} \cdot \frac{f_2 - |f|}{f_2 - f_1}\right) & \text{beim Cosinus - Rolloff.} \end{cases}$$

- Zur Parametrisierung der Flankensteilheit verwenden wir in beiden Fällen den **Rolloff-Faktor**, der Werte zwischen 0 und 1 (einschließlich dieser Grenzen) annehmen kann:

$$r = \frac{f_2 - f_1}{f_2 + f_1}.$$

- Für  $r = 0$  ( $f_1 = f_2 = f_{Nyq}$ ) ergibt sich das Rechteck-Nyquistspektrum, während der Rolloff-Faktor  $r = 1$  ( $f_1 = 0, f_2 = 2f_{Nyq}$ ) ein dreieckförmiges bzw.  $\cos^2$ -Spektrum angibt – je nachdem, von welcher der beiden oben abgebildeten Grundstrukturen man ausgeht.

**Hinweis:** In der Literatur wird der Rolloff-Faktor auch oft mit  $\alpha$  („alpha“) bezeichnet.

## 1/T–Nyquistspektren (2)

Betrachten wir nun die Nyquistimpulse. Beim trapezförmigem Spektrum mit Rolloff-Faktor  $r$  erhält man:

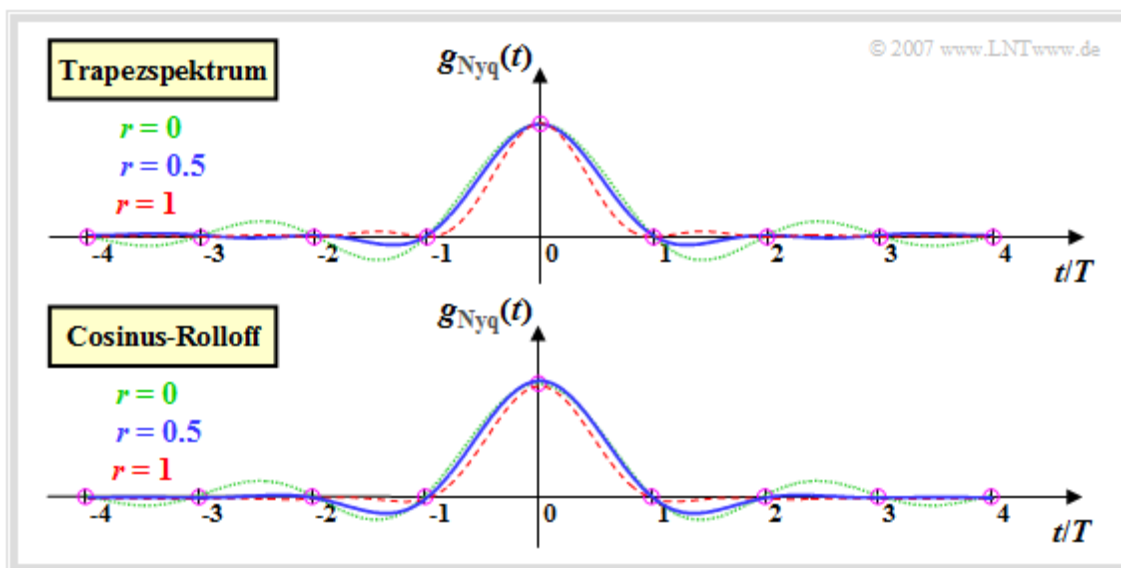
$$g_{\text{Trapez}}(t) = g_0 \cdot \text{si}\left(\frac{\pi \cdot t}{T}\right) \cdot \text{si}\left(\frac{\pi \cdot r \cdot t}{T}\right) \quad \text{mit} \quad \text{si}(x) = \sin(x)/x.$$

Dagegen liefert die Fourierrücktransformation des Cosinus-Rolloff-Spektrums:

$$g_{\text{CRO}}(t) = g_0 \cdot \text{si}\left(\frac{\pi \cdot t}{T}\right) \cdot \frac{\cos(\pi \cdot r \cdot t/T)}{1 - (2 \cdot r \cdot t/T)^2} \quad \text{mit} \quad \text{si}(x) = \sin(x)/x.$$

Diese beiden Nyquistimpulse kann man im nachfolgend genannten Interaktionsmodul mit der Einstellung  $\Delta f \cdot T = 1$  betrachten und sich dabei den Einfluss des Rolloff-Faktors verdeutlichen:

### Tiefpässe im Frequenz- und Zeitbereich

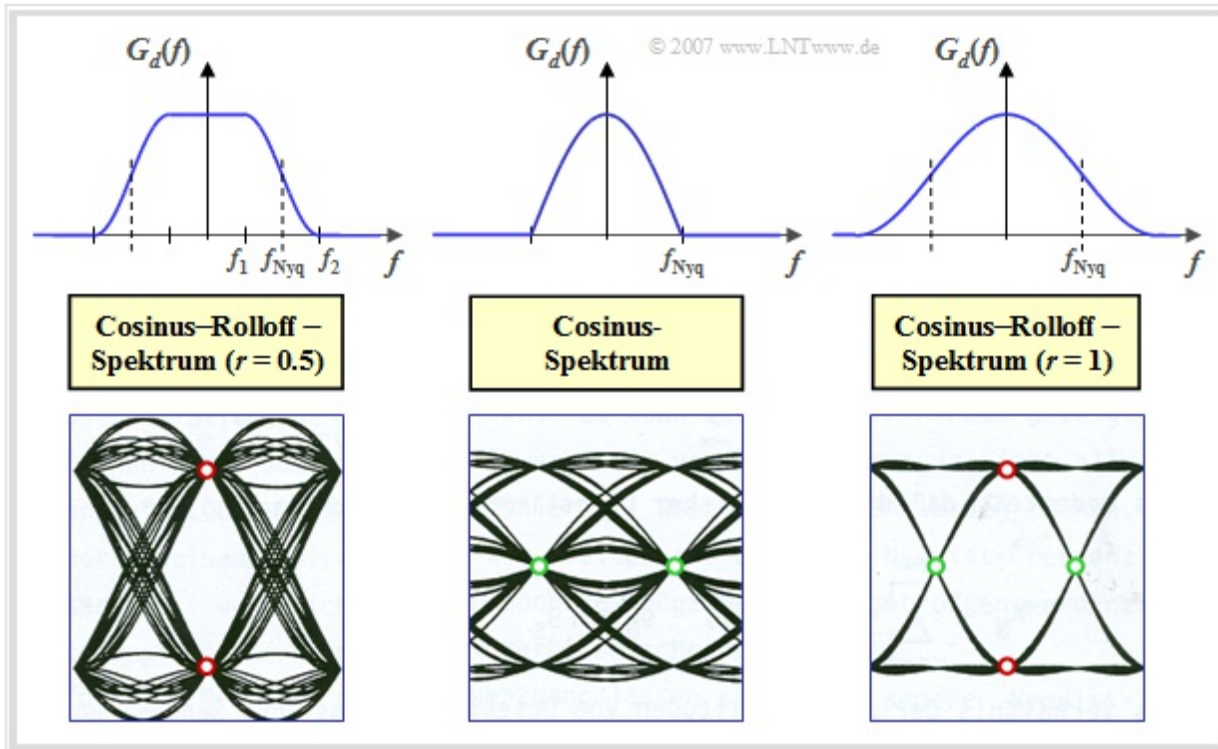


Die obere Grafik zeigt den Nyquistimpuls mit Trapezspektrum für verschiedene Rolloff-Faktoren. Unten ist der entsprechende Zeitverlauf für das Cosinus-Rolloff-Spektrum dargestellt. Man erkennt:

- Je kleiner der Rolloff-Faktor  $r$  ist, desto langsamer erfolgt der Abfall des Nyquistimpulses. Diese Aussage trifft sowohl für das Trapez- als auch für das Cosinus-Rolloff-Spektrum zu.
- Im Grenzfall  $r \rightarrow 0$  ergibt sich in beiden Fällen das rechteckförmige Nyquistspektrum und der si-förmige Nyquistimpuls, der asymptotisch mit  $1/t$  abklingt (grüne Kurven).
- Bei einem mittleren Rolloff ( $r \approx 0.5$ ) sind die ersten Überschwinger beim Trapezspektrum geringer als beim CRO-Spektrum, da bei gegebenem  $r$  die Nyquistflanke flacher verläuft (blaue Kurven).
- Mit dem Rolloff-Faktor  $r = 1$  wird im Frequenzbereich aus dem Trapez ein Dreieck und aus dem CRO-Spektrum das Cosinus-Quadrat-Spektrum (rote Kurven).
- Mit  $r = 1$  erfolgt der asymptotische Abfall der oberen Zeitfunktion (gemäß dem Trapezspektrum) mit  $1/t^2$  und der Abfall der unteren Zeitfunktion (gemäß dem CRO-Spektrum) mit  $1/t^3$ .

## Zweites Nyquistkriterium (1)

Vor der exakten mathematischen Definition soll anhand von Grafiken veranschaulicht werden, welche Bedeutung das **zweite Nyquistkriterium** zur Bewertung eines Digitalsystems besitzt.



- Die linke Grafik zeigt im Vorgriff auf das **Kapitel 3.2** das Augendiagramm eines Nyquistsystems mit Cosinus-Rolloff-Charakteristik, wobei der Rolloff-Faktor  $r = 0.5$  gewählt wurde. Da hier das erste Nyquistkriterium erfüllt ist (es besteht eine Punktsymmetrie um  $f_{Nyq}$ ), ergibt sich für die vertikale Augenöffnung zum Zeitpunkt  $t = 0$  der größtmögliche Wert  $2 \cdot g_d(0)$ . Alle Augenlinien gehen zum Zeitpunkt  $t = 0$  durch einen der beiden rot markierten Punkte.
- Das mittlere Spektrum weist keine Symmetrie bezüglich des Flankenabfalls auf, so dass hier das erste Nyquistkriterium nicht erfüllt ist – im Gegensatz zum zweiten. Alle Augenlinien schneiden hier die Zeitachse zu den selben Zeiten (markiert durch die grünen Punkte), was beispielsweise die Taktwiedergewinnung mittels einer PLL (*Phase-Locked Loop*) erleichtert. Die **horizontale Augenöffnung** ist bei Erfüllung des zweiten Nyquistkriteriums maximal gleich der Symboldauer  $T$ .
- Das rechte Augendiagramm verdeutlicht, dass beim CRO-Spektrum mit  $r = 1$  – und nur bei diesem – sowohl das erste als auch das zweite Nyquistkriterium erfüllt werden. Der Nyquistimpuls

$$g_d(t) = g_0 \cdot \frac{\pi}{4} \cdot \text{si} \left( \frac{\pi \cdot t}{T} \right) \cdot \left[ \text{si} \left( \pi \cdot \left( \frac{t}{T} + \frac{1}{2} \right) \right) + \text{si} \left( \pi \cdot \left( \frac{t}{T} - \frac{1}{2} \right) \right) \right]$$

weist hier die erforderlichen Nulldurchgänge bei  $t = \pm T, t = \pm 1.5T, t = \pm 2T, t = \pm 2.5T, \dots$  auf, nicht jedoch bei  $\pm 0.5T$ . Die Impulsamplitude ist  $g_d(t = 0) = g_0$ .

## Zweites Nyquistkriterium (2)

Ein Impuls  $g_{\text{Nyq-2}}(t)$ , der das zweite Nyquistkriterium erfüllt, muss Nulldurchgänge bei  $\pm 1.5T$ ,  $\pm 2.5T$ ,  $\pm 3.5T$ , ... besitzen. Dagegen weist ein Nyquist-1-Impuls Nulldurchgänge bei  $\pm T$ ,  $\pm 2T$ ,  $\pm 3T$  auf. Deshalb lässt sich ein Nyquist-2-Impuls immer als Summe zweier (eventuell unterschiedlicher) und um  $\pm T/2$  verschobener Nyquist-1-Impulse darstellen. Geht man von einem Impuls  $g_{\text{Nyq-1}}(t)$  aus, so gilt:

$$g_{\text{Nyq-2}}(t) = g_{\text{Nyq-1}}(t + T/2) + g_{\text{Nyq-1}}(t - T/2).$$

Im Frequenzbereich lautet das zweite Nyquistkriterium (siehe [ST85]):

$$\sum_{k=-\infty}^{+\infty} \frac{G_d(f - \frac{k}{T})}{\cos(\pi \cdot f \cdot T - k \cdot \pi)} = \text{const.}$$

**Beispiel:** Ausgehend vom Impuls  $g(t) = g_0 \cdot \text{si}(\pi \cdot t/T)$ , der das erste Nyquistkriterium erfüllt, lautet der dazugehörige Nyquist-2-Impuls:

$$\begin{aligned} g_{\text{Nyq-2}}(t) &= g_0 \cdot \left[ \text{si}\left(\pi \cdot \frac{t + T/2}{T}\right) + \text{si}\left(\pi \cdot \frac{t - T/2}{T}\right) \right] = \\ &= \frac{2 \cdot g_0}{\pi} \cdot \frac{\cos(\pi \cdot t/T)}{1 - (2 \cdot t/T)^2}. \end{aligned}$$

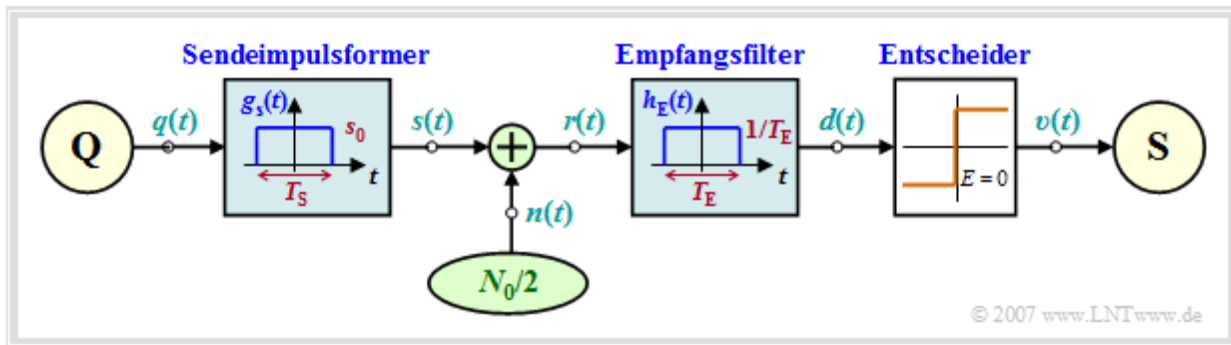
Aufgrund der Begrenzung des Spektrums  $G_{\text{Nyq-1}}(f)$  auf den Bereich  $|f| \leq f_{\text{Nyq}} = 1/(2T)$  beschränkt sich in obiger Gleichung die Summe auf den Term mit  $k = 0$  und man erhält:

$$G_{\text{Nyq-2}}(f) = \begin{cases} g_0 \cdot T \cdot \cos\left(\frac{\pi \cdot f}{2 \cdot f_{\text{Nyq}}}\right) & \text{für } |f| < f_{\text{Nyq}}, \\ 0 & \text{sonst.} \end{cases}$$

Dieser Frequenzverlauf und das dazugehörige Augendiagramm ist in der Grafik auf der letzten Seite in der mittleren Spalte oben skizziert. Man erkennt deutlich die Erfüllung des zweiten Nyquistkriteriums.

## Voraussetzungen und Optimierungskriterium

Für dieses Kapitel 1.4 gilt das folgende Blockschaltbild:



Wenn nicht explizit anders angegeben, wird im Folgenden von folgenden Voraussetzungen ausgegangen:

- Die Übertragung erfolgt binär, bipolar und redundanzfrei; mehrstufige und/oder codierte Systeme werden im Kapitel 2 behandelt. Der Abstand aufeinander folgender Symbole beträgt  $T$ . Die (äquivalente) Bitrate ist bei den hier getroffenen Voraussetzungen gleich  $R = 1/T$ .
- Der Sendegrundimpuls  $g_s(t)$  ist rechteckförmig und weist die Amplitude  $s_0$  sowie die Impulsdauer  $T_S \leq T$  auf. Stimmt die Sendepulsdauer  $T_S$  mit der Symboldauer  $T$  überein, so spricht man von NRZ-Rechteckimpulsen. Im Fall  $T_S/T < 1$  liegt ein RZ-Format vor.
- Als Übertragungskanal wird das AWGN-Modell mit der (einseitigen) Rauschleistungsdichte  $N_0$  verwendet, so dass für das Empfangssignal  $r(t) = s(t) + n(t)$  gilt. Die für systemtheoretische Untersuchungen besser geeignete zweiseitige Rauschleistungsdichte beträgt somit  $N_0/2$ .
- Die Impulsantwort  $h_E(t)$  des Empfangsfilters ist ebenfalls rechteckförmig, allerdings mit der Höhe  $1/T_E$  und der Breite  $T_E$ . Daraus folgt für den Gleichsignalübertragungsfaktor  $H_E(f=0) = 1$ . Nur im Sonderfall  $T_E = T_S$  kann man  $H_E(f)$  als Matched-Filter bezeichnen.
- Um Impulsinterferenzen auszuschließen, muss bei der Optimierung der beiden Systemparameter  $T_S$  bzw.  $T_E$  stets die Randbedingung  $T_S + T_E \leq 2T$  eingehalten werden. Impulsinterferenzen werden erst im **Kapitel 3** betrachtet.
- Zur Gewinnung der Sinkensymbolfolge wird ein einfacher Schwellenwertentscheider mit optimaler Entscheidungsschwelle  $E = 0$  und optimalen Detektionszeitpunkten (bei  $vT$ ) verwendet.

Unter **Systemoptimierung** verstehen wir hier, die Parameter  $T_S$  und  $T_E$  von Sendegrundimpuls und Empfangsfilter-Impulsantwort so zu bestimmen, dass die Bitfehlerwahrscheinlichkeit den kleinstmöglichen Wert annimmt.

## Leistungs- und Spitzenwertbegrenzung (1)

Die Optimierung der Systemgrößen wird entscheidend dadurch beeinflusst, ob als Nebenbedingung der Optimierung Leistungsbegrenzung oder Spitzenwertbegrenzung des Sendesignals gefordert wird.

**Definition:** Unter **Leistungsbegrenzung** versteht man, dass die (mittlere) Sendeleistung  $P_S$  einen vorgegebenen Maximalwert  $P_{S, \max}$  nicht überschreiten darf:

$$P_S = E[s(t)^2] = \overline{s(t)^2} \leq P_{S, \max}.$$

Um die minimale Fehlerwahrscheinlichkeit zu erzielen, wird man natürlich die mittlere Sendeleistung  $P_S$  im erlaubten Bereich möglichst groß wählen. Deshalb wird im Folgenden stets  $P_S = P_{S, \max}$  gesetzt.

Die Frage, ob als Nebenbedingung der Optimierung tatsächlich von Leistungsbegrenzung ausgegangen werden kann, hängt von den technischen Randbedingungen ab. Diese Annahme ist insbesondere bei Funkübertragungssystemen gerechtfertigt, unter Anderem deshalb, weil die als „Elektrosmog“ bekannte Beeinträchtigung von Mensch und Tier in starkem Maße von der (mittleren) Strahlungsleistung abhängt.

Anzumerken ist, dass ein Funkübertragungssystem natürlich nicht im Basisband arbeitet. Die hier am Beispiel der Basisbandübertragung definierten Beschreibungsgrößen werden aber im **Kapitel 4** dieses Buches dahingehend modifiziert, dass sie auch für digitale Trägerfrequenzsysteme anwendbar sind.

**Definition:** Von **Spitzenwertbegrenzung** spricht man immer dann, wenn der Aussteuerbereich der Sendeeinrichtung begrenzt ist. Bei bipolarer Signalisierung lautet die entsprechende Bedingung:

$$|s(t)| \leq s_0 \quad \text{für alle } t.$$

Off verwendet man anstelle von Spitzenwertbegrenzung auch den Begriff **Amplitudengrenzung**, der aber den Sachverhalt nicht ganz richtig wiedergibt.

Natürlich wird auch bei Spitzenwertbegrenzung die Leistung begrenzt, aber nicht die mittlere, sondern die Spitzenleistung. Die Nebenbedingung „Spitzenwertbegrenzung“ ist zum Beispiel dann sinnvoll und sogar notwendig, wenn

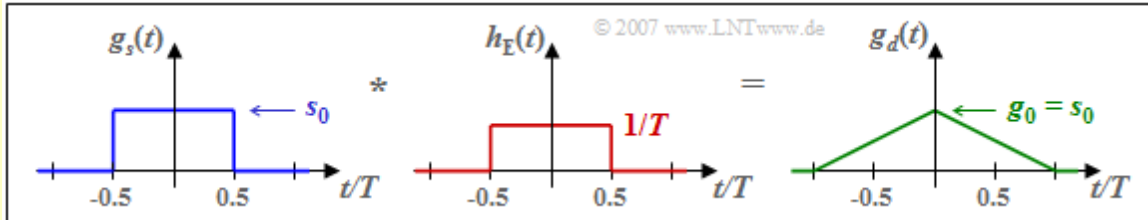
- der Aussteuerbereich des Senders wegen Nichtlinearitäten von Bauelementen und Endverstärkern beschränkt ist, oder
- die Nebensprechstörung zu keiner Zeit einen gewissen Wert nicht überschreiten darf. Hierauf ist insbesondere bei der Kommunikation über Zweidrahtleitungen zu achten.



## Leistungs- und Spitzenwertbegrenzung (2)

**Beispiel:** Sendegrundimpuls  $g_s(t)$  und Empfangsfilter-Impulsantwort  $h_E(t)$  seien rechteckförmig. Die Amplitude  $g_0$  des Ausgangsimpulses stimmt stets mit der Eingangsimpulsamplitude  $s_0$  überein.

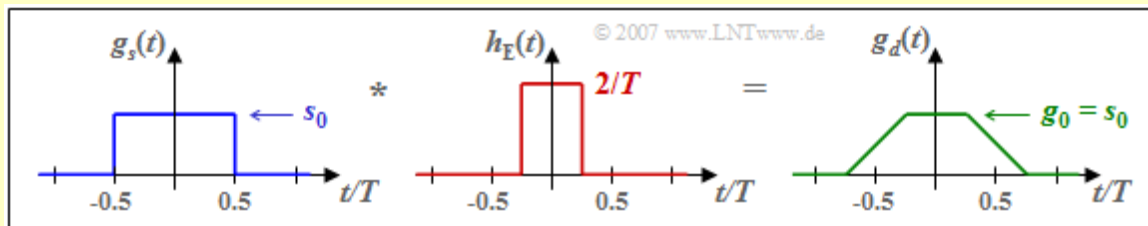
- Beim **System A** ( $T_S = T, T_E = T$ ) handelt es sich um die Matched-Filter-Realisierung mit NRZ-Rechteckimpulsen.



Die Rauschleistung ist  $\sigma_d^2 = N_0/(2T)$ . Die Bitfehlerwahrscheinlichkeit ergibt sich somit unter Berücksichtigung von  $s_0^2 \cdot T = E_B$  („Energie pro Bit“) wie folgt:

$$p_B = Q(g_0/\sigma_d) = Q\left(\sqrt{2 \cdot s_0^2 \cdot T/N_0}\right) = Q\left(\sqrt{2 \cdot E_B/N_0}\right).$$

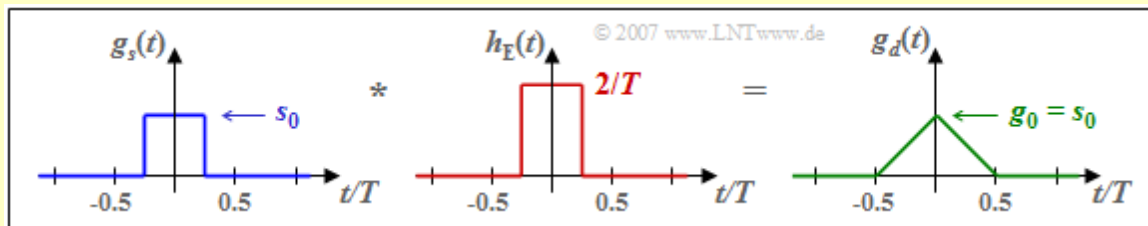
- Beim **System B** ( $T_S = T, T_E = T/2$ ) ist das Empfangsfilter  $h_E(t)$  nicht an den Sendegrundimpuls  $g_s(t)$  angepasst. Deshalb ergibt sich ein trapezförmiger Detektionsgrundimpuls  $g_d(t)$ :



Die Rauschleistung ist gegenüber **System A** doppelt so groß  $\Rightarrow \sigma_d^2 = N_0/T$ , und es ergibt sich deshalb eine deutlich größere Bitfehlerwahrscheinlichkeit:

$$p_B = Q(g_0/\sigma_d) = Q\left(\sqrt{s_0^2 \cdot T/N_0}\right) = Q\left(\sqrt{E_B/N_0}\right).$$

- Beim **System C** ( $T_S = T/2, T_E = T/2$ ) ist wie bei **System A** die Matched-Filter-Bedingung erfüllt, allerdings bei RZ-Rechteckimpulsen mit dem Tastverhältnis 1/2.

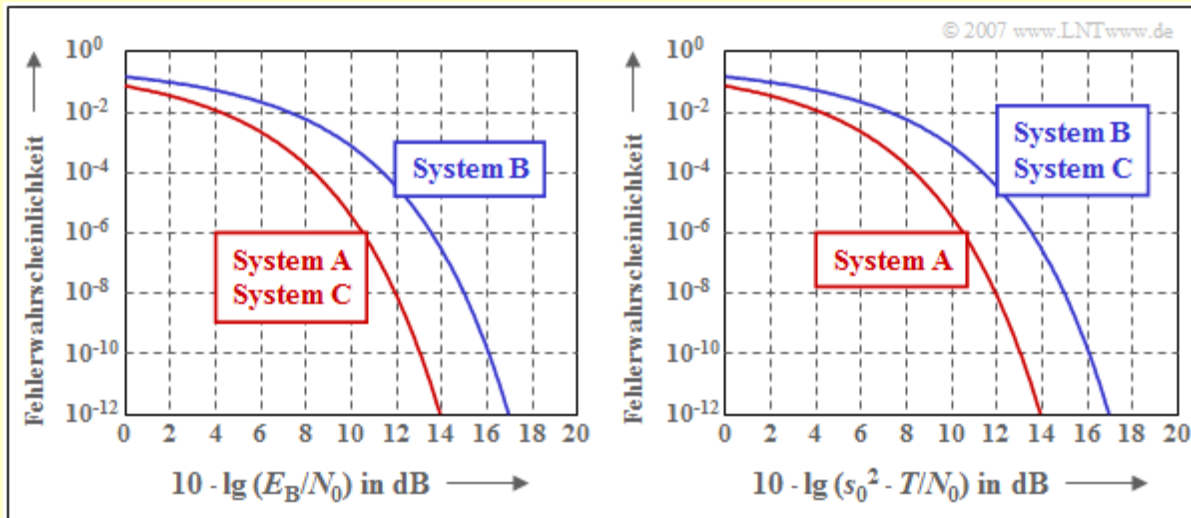


Die Rauschleistung ist so groß wie bei **System B**  $\Rightarrow \sigma_d^2 = N_0/T$ . Bei der zweiten Gleichung ist berücksichtigt, dass die Energie pro Bit jetzt nur noch halb so groß ist  $\Rightarrow E_B = 1/2 \cdot s_0^2 \cdot T$ :

$$p_B = Q(g_0/\sigma_d) = Q\left(\sqrt{s_0^2 \cdot T/N_0}\right) = Q\left(\sqrt{2 \cdot E_B/N_0}\right).$$

## Leistungs- und Spitzenwertbegrenzung (3)

**Fortsetzung des Beispiels:** Die Grafik zeigt die Bitfehlerwahrscheinlichkeit  $p_B$  in Abhängigkeit vom Verhältnis  $E_B/N_0$  (linkes Diagramm) bzw. als Funktion von  $s_0^2 \cdot T/N_0$  (rechtes Diagramm). Es handelt sich dabei um die Interpretation der oben hergeleiteten Ergebnisse.



Diese beiden Diagramme in doppelt-logarithmischer Darstellung sind wie folgt zu interpretieren:

- Die linke Grafik vergleicht die Systeme bei gleicher mittlerer Leistung ( $P_s$ ) bzw. bei konstanter Energie pro Bit ( $E_B$ ). Da der Abszissenwert zusätzlich auf  $N_0$  bezogen ist, gibt  $p_B(E_B/N_0)$  den Sachverhalt auch für unterschiedliche Rauschleistungsdichten ( $N_0$ ) richtig wieder.
- Bei Leistungsbegrenzung sind die Konfigurationen A und C gleichwertig und stellen jeweils das Optimum dar. Wie auf den nächsten Seiten hergeleitet wird, liegt ein bei Leistungsbegrenzung optimales System immer dann vor, wenn  $g_s(t)$  und  $h_E(t)$  formgleich sind (Matched-Filter). Die kleinere Leistung von System C wird durch die hier gewählte Abszisse ausgeglichen.
- Dagegen wird bei System B die Matched-Filter-Bedingung nicht eingehalten ( $T_E \neq T_S$ ) und die Fehlerwahrscheinlichkeitskurve liegt nun um 3 dB rechts von der durch die Systeme A und C vorgegebenen Grenzkurve.
- Die rechte Grafik beschreibt das Optimierungsergebnis bei Spitzenwertbegrenzung, was an der Abszissenbeschriftung zu erkennen ist. Der Kurvenzug A (NRZ-Impuls, Matched-Filter) gibt auch hier die Grenzkurve an, die von keinem anderen System unterschritten werden kann.
- Die Kurve B in der rechten Grafik hat den genau gleichen Verlauf wie in der linken Darstellung, da wiederum NRZ-Sendeimpulse verwendet werden. Der Abstand von 3 dB zur Grenzkurve ist wieder auf die Nichteinhaltung der Matched-Filter-Bedingung zurückzuführen.
- Im Gegensatz zur linken Grafik liegt nun auch das Matched-Filter-System C um 3 dB rechts vom Optimum. Der Grund für diese Degradation ist, dass bei gleichem Spitzenwert (gleicher Spitzenleistung) das System C nur die halbe mittlere Leistung wie das System A bereitstellt.

## Systemoptimierung bei Leistungsbegrenzung (1)

Die Minimierung der Bitfehlerwahrscheinlichkeit

$$p_B = Q(\sqrt{\rho_d}) \Rightarrow \text{Minimum}$$

kann aufgrund des monotonen Funktionsverlaufs der komplementären Gaußschen Fehlerfunktion  $Q(x)$  auf die Maximierung des Signal-zu-Rausch-Leistungsverhältnisses vor dem Schwellenwertentscheider (Detektions-SNR) zurückgeführt werden:

$$\rho_d = g_0^2 / \sigma_d^2 \Rightarrow \text{Maximum.}$$

Hierbei steht  $g_0$  als Abkürzung für  $g_d(t=0)$ . Gleichzeitig muss sichergestellt werden, dass

- der Detektionsgrundimpuls  $g_d(t) = g_s(t) * h_E(t)$  das erste Nyquistkriterium erfüllt, und
- die Energie des Sendegrundimpulses  $g_s(t)$  einen vorgegebenen Wert  $E_B$  nicht überschreitet.

In den vorangegangenen Abschnitten wurde bereits mehrfach erwähnt, dass beim AWGN-Kanal für das optimale System unter der Nebenbedingung der Leistungsbegrenzung gilt:

$$p_{B, \min} = Q(\sqrt{\rho_{d, \max|L}}) \quad \text{mit} \quad \rho_{d, \max|L} = 2 \cdot E_B / N_0.$$

**Definition:** Der **Systemwirkungsgrad bei Leistungsbegrenzung** einer vorliegenden Konfiguration ist der Quotient aus dem tatsächlichen und dem größtmöglichen Detektions-SNR:

$$\eta_L = \frac{\rho_d}{\rho_{d, \max|L}} = \frac{g_0^2 / \sigma_d^2}{2 \cdot E_B / N_0}.$$

Hierbei sind folgende Systemgrößen verwendet:

- $g_0 = g_d(t=0)$  gibt die Amplitude des betrachteten Nyquistimpulses an.
- $E_B$  beschreibt die Energie des Sendegrundimpulses.
- $N_0$  ist die (einseitige) AWGN-Rauschleistungsdichte.
- $\sigma_d^2$  bezeichnet die Detektionsstörleistung für das gegebene Empfangsfilter.

Nachfolgend wird gezeigt, dass für den Systemwirkungsgrad stets  $\eta_L \leq 1$  gilt.

## Systemoptimierung bei Leistungsbegrenzung (2)

Der Beweis erfolgt im Frequenzbereich. Aus Darstellungsgründen normieren wir den Sendegrundimpuls:

$$h_S(t) = \frac{g_s(t)}{g_0 \cdot T} \quad \circ \longrightarrow \bullet \quad H_S(f) = \frac{G_s(f)}{g_0 \cdot T}.$$

Damit hat  $h_S(t)$  die Einheit 1/s und  $H_S(f)$  ist dimensionslos. Für die einzelnen Systemgrößen folgt daraus:

- Aufgrund des ersten Nyquistkriteriums muss gelten:

$$G_d(f) = G_s(f) \cdot H_E(f) = G_{\text{Nyq}}(f)$$

$$\Rightarrow H_S(f) \cdot H_E(f) = H_{\text{Nyq}}(f) = \frac{G_{\text{Nyq}}(f)}{g_0 \cdot T}.$$

- Die Amplitude des Detektionsgrundimpulses ist gleich

$$g_d(t=0) = g_0 \cdot T \cdot \int_{-\infty}^{+\infty} H_{\text{Nyq}}(f) \, df = g_0.$$

- Die Energie des Sendegrundimpulses ist wie folgt gegeben:

$$E_B = g_0^2 \cdot T^2 \cdot \int_{-\infty}^{+\infty} |H_S(f)|^2 \, df.$$

- Die Detektionsstörleistung lautet:

$$\sigma_d^2 = \frac{N_0}{2} \cdot \int_{-\infty}^{+\infty} |H_E(f)|^2 \, df = \frac{N_0}{2} \cdot \int_{-\infty}^{+\infty} \frac{|H_{\text{Nyq}}(f)|^2}{|H_S(f)|^2} \, df.$$

Setzt man diese Teilergebnisse in die Gleichung für den Systemwirkungsgrad ein, so erhält man:

$$\eta_L = \left[ T \cdot \int_{-\infty}^{+\infty} |H_S(f)|^2 \, df \cdot T \cdot \int_{-\infty}^{+\infty} \frac{|H_{\text{Nyq}}(f)|^2}{|H_S(f)|^2} \, df \right]^{-1}.$$

Wendet man auf den Ausdruck in der Klammer die Schwartzsche Ungleichung **[BS01]**

$$T \cdot \int_{-\infty}^{+\infty} |H_1(f)|^2 \, df \cdot T \cdot \int_{-\infty}^{+\infty} |H_2(f)|^2 \, df \geq \left[ T \cdot \int_{-\infty}^{+\infty} |H_1(f) \cdot H_2(f)| \, df \right]^2$$

$$\Rightarrow T \cdot \int_{-\infty}^{+\infty} |H_S(f)|^2 \, df \cdot T \cdot \int_{-\infty}^{+\infty} \frac{|H_{\text{Nyq}}(f)|^2}{|H_S(f)|^2} \, df \geq \left[ T \cdot \int_{-\infty}^{+\infty} H_{\text{Nyq}}(f) \, df \right]^2 = 1$$

an, so erkennt man, dass stets  $\eta_L \leq 1$  gilt. Die Auswertung zeigt, dass für

$$H_{S,\text{opt}}(f) = \gamma \cdot \sqrt{H_{\text{Nyq}}(f)}$$

in obiger Ungleichung unabhängig vom Parameter  $\gamma$  das Gleichheitszeichen gilt:

$$\gamma^2 \cdot T \cdot \int_{-\infty}^{+\infty} H_{\text{Nyq}}(f) \, df \cdot \frac{1}{\gamma^2} \cdot T \cdot \int_{-\infty}^{+\infty} H_{\text{Nyq}}(f) \, df = \left[ T \cdot \int_{-\infty}^{+\infty} H_{\text{Nyq}}(f) \, df \right]^2 = 1$$

$$\Rightarrow \eta_L = 1.$$

Im Folgenden wird meist vereinfachend  $\gamma = 1$  gesetzt.

## Wurzel–Nyquist–Systeme

Das wesentliche Ergebnis der Berechnungen auf den letzten Seiten war, dass beim optimalen Binärsystem unter der Nebenbedingung **Leistungsbegrenzung**

- die Impulsantwort  $h_E(t)$  des Empfangsfilters formgleich mit dem Sendegrundimpuls  $g_s(t)$  zu wählen ist (gleiches gilt für die zugehörigen Spektren), und
- der Detektionsgrundimpuls  $g_d(t) = g_s(t) * h_E(t)$  die erste Nyquistbedingung erfüllen muss.

Sind sowohl  $g_s(t)$  als auch  $h_E(t)$  rechteckförmig mit  $T_S = T_E \leq T$ , so werden beide Bedingungen erfüllt. Nachteilig für diese Konfiguration ist allerdings der große Bandbreitenbedarf aufgrund der nur langsam abfallenden, si-förmigen Spektralfunktionen  $G_s(f)$  und  $H_E(f)$ .

Geht man von einem Nyquistspektrum mit Cosinus-Rolloff-Flanke (und Rolloff-Faktor  $r$ ) aus,

$$G_d(f) = G_s(f) \cdot H_E(f) = g_0 \cdot T \cdot H_{CRO}(f)$$

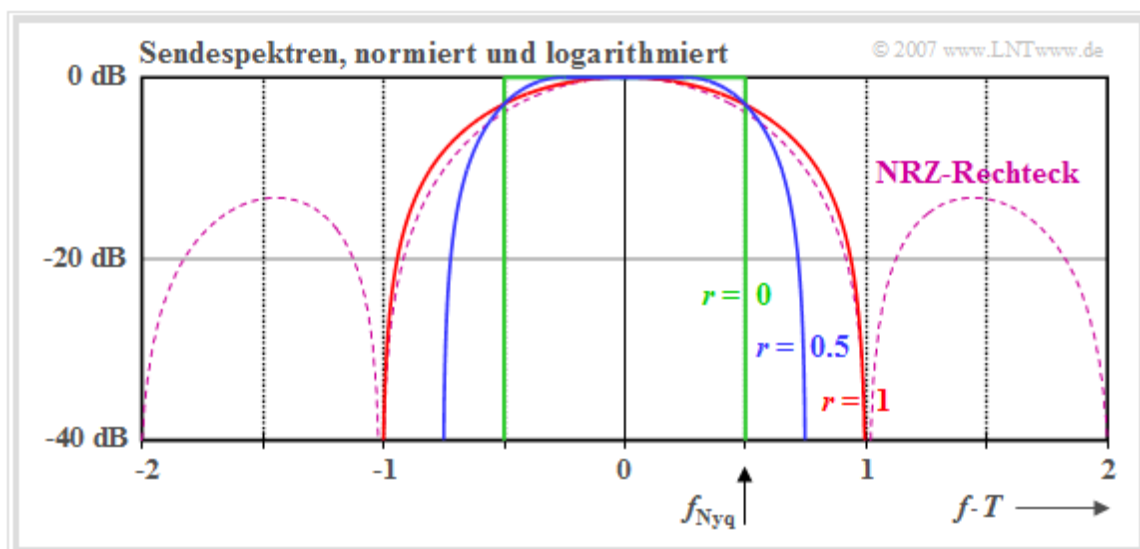
$$\Rightarrow G_s(f) = g_0 \cdot T \cdot \sqrt{H_{CRO}(f)}, \quad H_E(f) = \sqrt{H_{CRO}(f)},$$

so ergeben sich günstigere Spektraleigenschaften und ein geringerer Bandbreitenbedarf.

Die folgende Grafik zeigt die normierten Sendespektren  $G_s(f)/(g_0T)$  in logarithmierter Darstellung für die drei Rolloff-Faktoren

- $r = 0$  (grüne Kurve),
- $r = 0.5$  (blaue Kurve), und
- $r = 1$  (rote Kurve).

Die Spektralfunktion  $G_s(f)/(g_0T)$ , die sich bei einem rechteckförmigen NRZ-Sendegrundimpuls ergibt, ist gestrichelt und violett eingezeichnet.



Anzumerken ist, dass der Bandbreitenbedarf bei der Basisbandübertragung nur eine untergeordnete Rolle spielt. Die Grafik gilt jedoch auch für die Trägerfrequenzsysteme entsprechend **Kapitel 4** bei Darstellung im äquivalenten Tiefpassbereich. Bei diesen Systemen spielt die Bandbreite eine wichtige Rolle. Jedes zusätzliches Hertz an Bandbreite kann sehr teuer sein.

## Systemoptimierung bei Spitzenwertbegrenzung (1)

Ist das (bipolare) Sendesignal auf  $\pm s_0$  begrenzt, so gilt für die minimale Bitfehlerwahrscheinlichkeit:

$$p_{B, \min} = Q\left(\sqrt{\rho_{d, \max|A}}\right) \quad \text{mit} \quad \rho_{d, \max|A} = 2 \cdot s_0^2 \cdot T/N_0.$$

Der Buchstabe **A** steht hierbei für Amplitudenbegrenzung (oder Spitzenwertbegrenzung). Es gibt nur ein einziges System, das diese minimale Fehlerwahrscheinlichkeit erreicht, nämlich eine Konfiguration mit NRZ-Rechteck-Sendegrundimpuls und daran angepasstem Empfangsfilter.

**Definition:** Der **Systemwirkungsgrad bei Amplitudenbegrenzung** (Spitzenwertbegrenzung) lautet:

$$\eta_A = \frac{\rho_d}{\rho_{d, \max|A}} = \frac{g_0^2/\sigma_d^2}{2 \cdot s_0^2 \cdot T/N_0}.$$

Hierbei sind folgende Systemgrößen verwendet:

- $g_0 = g_d(t=0)$  gibt die Amplitude des betrachteten Nyquistimpulses an.
- $s_0$  stellt den maximalen Betrag des bipolaren Sendesignals dar.
- $N_0$  ist die (einseitige) Rauschleistungsdichte.
- $\sigma_d^2$  bezeichnet die Detektionsstörleistung.

Dieser Wirkungsgrad unterscheidet sich vom Systemwirkungsgrad  $\eta_L$  bei Leistungsbegrenzung dadurch, dass nun im Nenner  $s_0^2 \cdot T$  anstelle von  $E_B$  steht. Es besteht folgender Zusammenhang:

$$\eta_A = \frac{E_B}{s_0^2 \cdot T} \cdot \eta_L = \eta_L / C_S^2.$$

Hierbei bezeichnet der **Crestfaktor** das Verhältnis von Maximalwert  $s_0$  und Effektivwert  $s_{\text{eff}}$  von  $s(t)$ :

$$C_S = \frac{s_0}{\sqrt{E_B/T}} = \frac{\text{Max}[s(t)]}{\sqrt{E[s^2(t)]}} = s_0/s_{\text{eff}}.$$

$s_{\text{eff}}$  ist gleich der Wurzel aus der Signalleistung ( $E_B/T$ ).

**Beispiel:** Wir betrachten wie im **vorherigen Beispiel** drei unterschiedliche Konfigurationen mit jeweils rechteckförmigen Zeitfunktionen  $g_s(t)$  und  $h_E(t)$  und geben hierfür die Systemwirkungsgrade an:

System A:  $\rho_d = 2 \cdot E_B/N_0 = 2 \cdot s_0^2 \cdot T/N_0 \Rightarrow \eta_L = 1, \eta_A = 1.$

System B:  $\rho_d = E_B/N_0 = s_0^2 \cdot T/N_0 \Rightarrow \eta_L = 0.5, \eta_A = 0.5.$

System C:  $\rho_d = 2 \cdot E_B/N_0 = s_0^2 \cdot T/N_0 \Rightarrow \eta_L = 1, \eta_A = 0.5.$

Man erkennt, dass beim **System B** beide Systemwirkungsgrade aufgrund der fehlenden Anpassung ( $T_E \neq T_S$ ) nur jeweils 0.5 sind. Beim **System C** hat zwar der Systemwirkungsgrad  $\eta_L$  wegen  $T_E = T_S$  den Maximalwert. Dagegen ist  $\eta_A$  nur 0.5, da der RZ-Impuls nicht die maximale Energie besitzt, die aufgrund der Spitzenwertbegrenzung erlaubt wäre. Der Crestfaktor hat hier den Wert „Wurzel aus 2“.



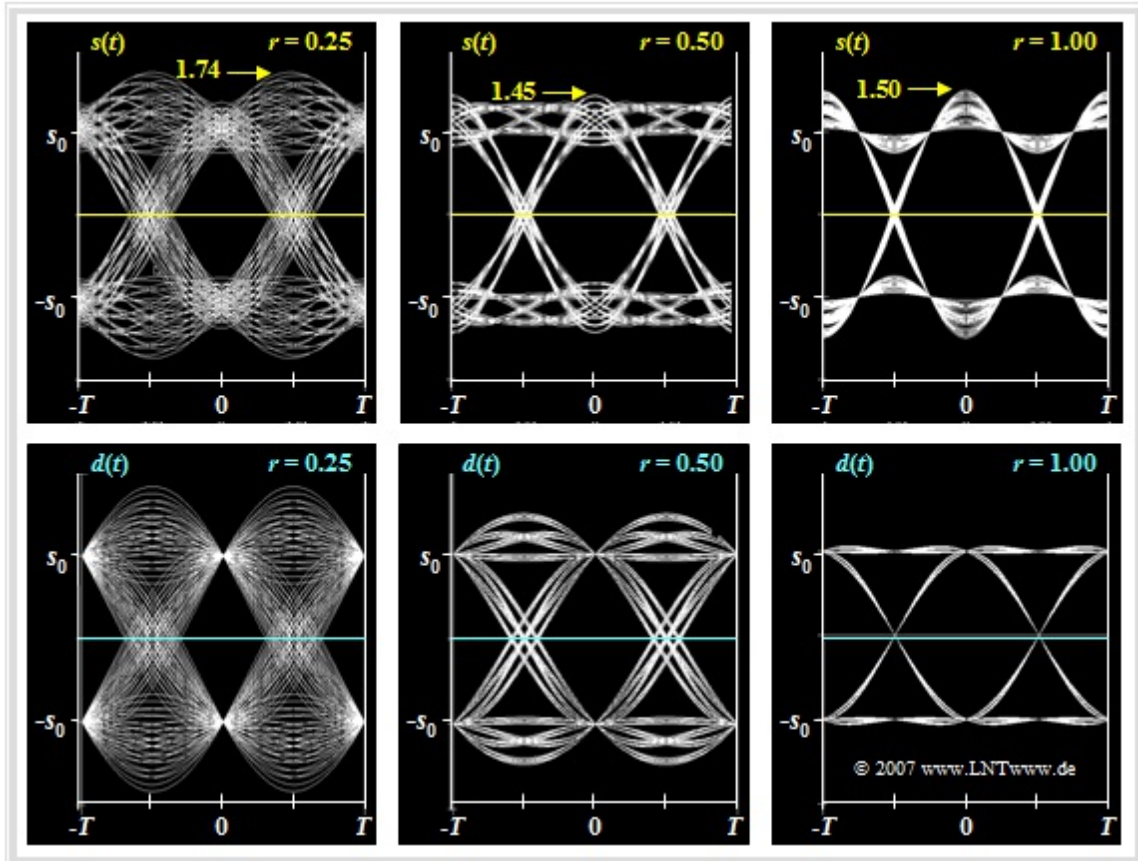
## Systemoptimierung bei Spitzenwertbegrenzung (2)

Nun betrachten wir eine **Wurzel–Nyquist–Konfiguration** mit Cosinus–Rolloff–Gesamtfrequenzgang:

$$G_s(f) = g_0 \cdot T \cdot \sqrt{H_{\text{CRO}}(f)}, \quad H_E(f) = \sqrt{H_{\text{CRO}}(f)}$$

$$\Rightarrow G_d(f) = g_0 \cdot T \cdot H_{\text{CRO}}(f) = G_{\text{Nyq}}(f).$$

Die Grafik zeigt die Augendiagramme am Sender (oben) und am Empfänger (unten), jeweils für die Rolloff–Faktoren  $r = 0.25$ ,  $r = 0.50$  und  $r = 1$ . Es sei daran erinnert, dass eine solche Konfiguration unter der Nebenbedingung der Leistungsbegrenzung unabhängig vom Rolloff–Faktor  $r$  optimal ist.



Man erkennt aus dieser Darstellung:

- Der Sendepuls  $g_s(t)$  erfüllt nicht die Nyquistbedingung: Das Auge am Sender (obere Bildreihe) ist nicht vollständig geöffnet; der Maximalwert des Sendesignals ist größer als sein Effektivwert.
- Der Crestfaktor  $C_S = s_0/s_{\text{eff}}$  wird mit kleinerem  $r$  größer und damit der Wirkungsgrad  $\eta_A$  kleiner. Für  $r = 0.5$  ergibt sich  $C_S \approx 1.45$  und damit  $\eta_A \approx 0.47$ . Das Detektions–SNR ist dann um den Betrag  $10 \cdot \lg \eta_A \approx 3.2$  dB geringer als bei der Rechteck–Rechteck–Konfiguration.
- Im (theoretischen) Grenzfalle  $r = 0$  gilt sogar  $C_S \rightarrow \infty$  und  $\eta_A \rightarrow 0$ . Der Sendegrundimpuls  $g_s(t)$  fällt hier noch langsamer als mit  $1/t$  ab, und es gilt:

$$\max_t \{s(t)\} = \max_t \left[ \sum_{(\nu)} a_\nu \cdot g_s(t - \nu \cdot T) \right] \rightarrow \infty.$$

- Begrenzt man das Sendesignal  $s(t)$  durch einen gegen 0 gehenden Gewichtungsfaktor auf einen endlichen Maximalwert  $s_0$ , so führt dies zu einem geschlossenem Auge vor dem Entscheider.



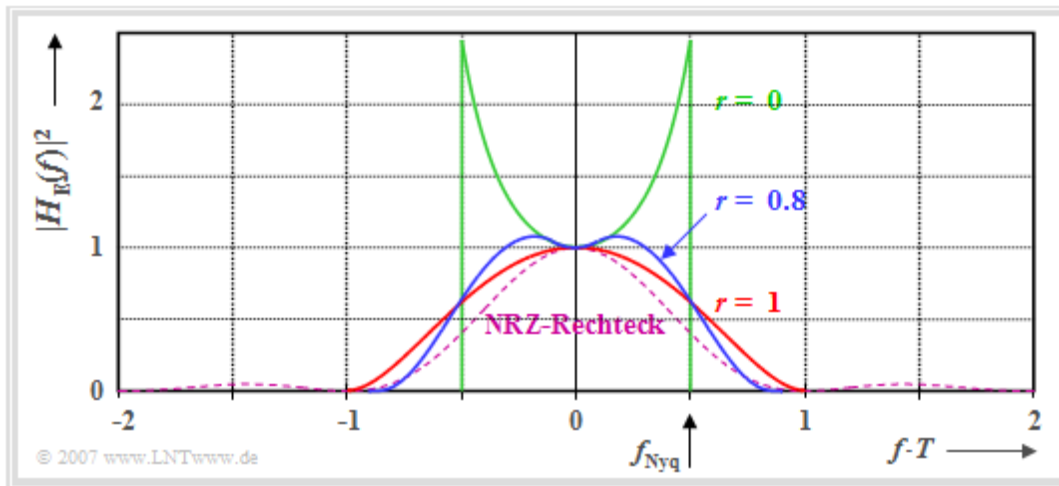
## Optimierung des Rolloff-Faktors bei Spitzenwertbegrenzung

Es wird von folgenden Voraussetzungen ausgegangen:

- Der Sendegrundimpuls  $g_s(t)$  sei NRZ-rechteckförmig; bei Spitzenwertbegrenzung ist dies optimal.
- Der Gesamtfrequenzgang  $H_S(f) \cdot H_E(f) = H_{\text{Nyq}}(f)$  erfülle die Nyquistbedingung und werde durch einen Cosinus-Rolloff-Tiefpass  $H_{\text{CRO}}(f)$  beschrieben.
- Da die Impulsamplitude  $g_0$  unabhängig vom Rolloff-Faktor  $r$  ist, lässt sich die SNR-Maximierung hier auf die Minimierung der Rauschleistung vor dem Entscheider zurückführen:

$$\sigma_d^2 = \frac{N_0}{2} \cdot \int_{-\infty}^{+\infty} |H_E(f)|^2 df \Rightarrow \text{Minimum, wobei } H_E(f) = \frac{H_{\text{CRO}}(f)}{\text{si}(\pi f T)}.$$

Die Grafik zeigt die Leistungsübertragungsfunktion  $|H_E(f)|^2$  für drei verschiedene Rolloff-Faktoren. Die Flächen unter diesen Kurven sind jeweils ein Maß für die Rauschleistung vor dem Entscheider.



Man erkennt aus dieser Darstellung:

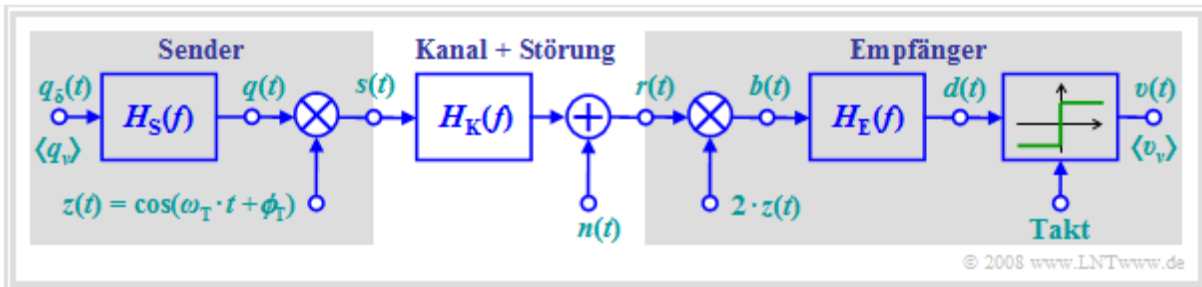
- Der Rolloff-Faktor  $r = 0$  (Rechteck) führt trotz des sehr schmalbandigen Empfangsfilters nur zum Wirkungsgrad  $\eta_A = 0.65$ , da  $H_E(f)$  wegen der si-Funktion im Nenner mit wachsendem  $f$  ansteigt.
- $r = 1$  bewirkt zwar ein doppelt so breites Nyquistspektrum, führt aber zu keiner Anhebung. Da die Fläche unter der roten Kurve kleiner ist als die grüne, ergibt sich ein besserer Wert:  $\eta_A = 0.88$ .
- Der größte Systemwirkungsgrad ergibt sich für  $r_{\text{opt}} \approx 0.8$  (flaches Maximum) mit  $\eta_A = 0.89$ . Hierfür erreicht man den bestmöglichen Kompromiss zwischen Bandbreite und Überhöhung.
- Durch Vergleich mit dem optimalen Frequenzgang  $H_E(f) = \text{si}(\pi f T)$  bei Spitzenwertbegrenzung, der zum Ergebnis  $\sigma_d^2 = N_0/(2T) \Rightarrow \eta_A = 1$  führt, erhält man für den Systemwirkungsgrad:

$$\eta_A = \left[ T \cdot \int_{-\infty}^{+\infty} |H_E(f)|^2 df \right]^{-1}.$$

- Das bedeutet: Das beste Cosinus-Rolloff-Nyquistspektrum mit  $r_{\text{opt}} = 0.8$  (blaue Kurve) ist gegenüber dem Matched-Filter (violett-gestrichelte Kurve) um ca. 0.5 dB schlechter, da die Fläche unter der blauen Kurve um ca. 12% größer ist als die Fläche unter der violetten Kurve.

## Gemeinsames Blockschaltbild für ASK und BPSK

Im **Kapitel 4.2** des Buches „Modulationsverfahren“ wurden die digitalen Trägerfrequenzsysteme **ASK** (*Amplitude Shift Keying*) und **BPSK** (*Binary Phase Shift Keying*) bereits ausführlich beschrieben. In diesem Kapitel wird nun die **Fehlerwahrscheinlichkeit** dieser Systeme berechnet, wobei von dem folgenden gemeinsamen Blockschaltbild ausgegangen wird:



Wie im Kapitel 4.2 des Buches „Modulationsverfahren“ gelten auch hier folgende Voraussetzungen:

- Die Demodulation geschieht stets **kohärent**. Das heißt: Beim Empfänger wird ein Trägersignal mit gleicher Frequenz wie beim Sender zugesetzt, aber mit doppelter Amplitude. Der Phasenversatz sei zunächst  $\Delta\phi_T = 0$ .
- Bei BPSK wird von den bipolaren Amplitudenkoeffizienten  $a_v \in \{-1, +1\}$  ausgegangen und die Entscheiderschwelle liegt bei  $E = 0$ . Dagegen gilt bei ASK  $a_v \in \{0, 1\}$ . Die Entscheiderschwelle  $E$  ist für diesen unipolaren Fall bestmöglich zu wählen.
- Wir betrachten stets den **AWGN-Kanal**, das heißt, dass für den Kanalfrequenzgang  $H_K(f) = 1$  gilt und  $n(t)$  weißes Gaußsches Rauschen mit der (einseitigen) Rauschleistungsdichte  $N_0$  darstellt.
- Die Entzerrung linearer Kanalverzerrungen – also der Fall  $H_K(f) \neq \text{const.}$  – ist in gleicher Weise möglich wie bei der Basisbandübertragung. Hierzu sei auf das **Kapitel 3.3** verwiesen.

## Fehlerwahrscheinlichkeit des optimalen BPSK–Systems (1)

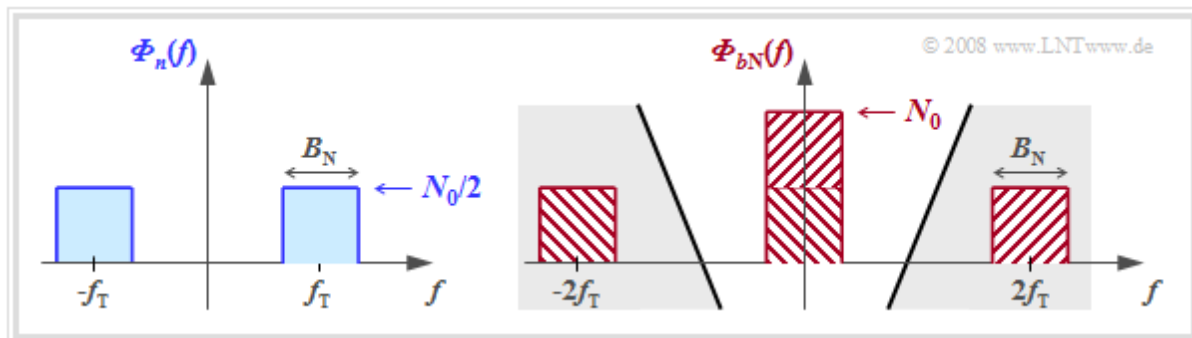
Wir gehen zunächst von einem bipolaren rechteckförmigen Quellensignal  $q(t)$  mit der Amplitude  $\pm s_0$  aus. Dessen normiertes Spektrum lautet:  $H_S(f) = \text{si}(\pi f T)$ . Ebenso wie bei der **Basisbandübertragung** ergibt sich die kleinstmögliche Bitfehlerwahrscheinlichkeit für das Empfangsfilter  $H_E(f) = H_S^*(f) = \text{si}(\pi f T)$ .

Die **Signalverläufe** des BPSK–Systems mit Matched–Filter–Empfänger zeigen:

- Das Detektionsnutzsignal  $d_S(t)$  – also ohne Rauschanteil – ist zu allen Detektionszeitpunkten  $\nu T$  stets  $\pm s_0$ , wobei das Vorzeichen durch die Amplitudenkoeffizienten  $a_\nu$  festgelegt sind.
- Wie beim vergleichbaren Basisbandsystem beträgt die **Fehlerwahrscheinlichkeit**  $p_B = Q(s_0/\sigma_d)$ , wobei  $Q(x)$  das **komplementäre Gaußsche Fehlerintegral** bezeichnet.
- Unterschiedlich zum Basisbandsystem ist jedoch die Rauschleistung. Der Rauschanteil  $b_N(t)$  ergibt sich durch die Multiplikation des Bandpassrauschens  $n(t)$  mit dem Träger  $2 \cdot \cos(2\pi f_T t)$  und besitzt die Rauschleistungsdichte

$$\Phi_{bN}(f) = \Phi_n(f) \star [1^2 \cdot \delta(f - f_T) + 1^2 \cdot \delta(f + f_T)].$$

- Die nachfolgende Grafik verdeutlicht diese Gleichung am Beispiel von bandbegrenztem weißem Rauschen mit der Bandbreite  $B_N$ . Während  $\Phi_n(f = f_T)$  gleich  $N_0/2$  gilt, ist  $\Phi_{bN}(f = 0) = N_0$ . Die Anteile um  $\pm 2f_T$  werden durch das nachfolgende Empfangsfilter  $H_E(f)$  eliminiert und spielen für die weiteren Betrachtungen keine Rolle.

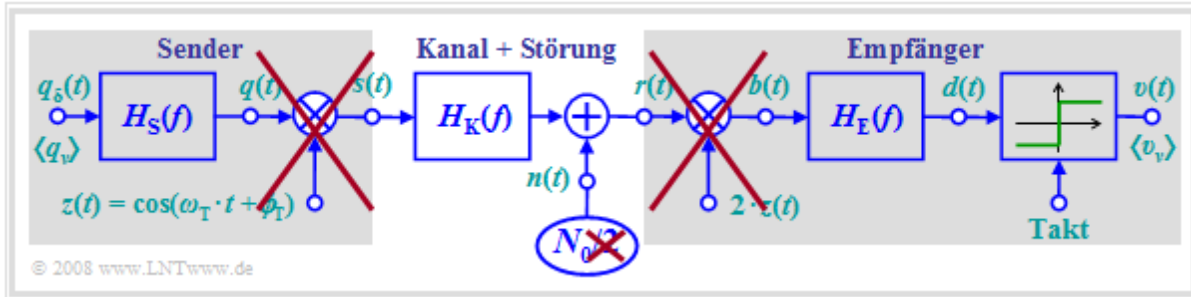


- Bei echt weißem Rauschen gilt mit dem Grenzübergang  $B_N \rightarrow \infty$ :

$$\Phi_n(f) = N_0/2, \quad \Phi_{bN}(f) = N_0.$$

## Fehlerwahrscheinlichkeit des optimalen BPSK–Systems (2)

Die gerade durchgeführten Betrachtungen zeigen, dass man zur Berechnung der Fehlerwahrscheinlichkeit des BPSK–Systems auf die beiden Multiplikationen mit  $z(t)$  und  $2 \cdot z(t)$  verzichten kann, wenn man die Rauschleistung verdoppelt.



Damit ergibt sich für die **Rauschleistung** vor dem Entscheider bei AWGN–Rauschen:

$$\sigma_d^2 = N_0 \cdot \int_{-\infty}^{+\infty} \text{si}^2(\pi f T_B) df = N_0 / T_B,$$

also der doppelte Wert als bei der Basisbandübertragung. *Hinweis:* Um später einen Vergleich mit der QAM zu ermöglichen, wurde hier die Symboldauer  $T$  durch die Bitdauer  $T_B$  ersetzt. Bei BPSK (und auch bei ASK) sind  $T$  und  $T_B$  gleich.

Damit lautet die BPSK–Fehlerwahrscheinlichkeit mit den zwei üblichen **Gaußschen Fehlerfunktionen**:

$$p_B = Q\left(\sqrt{\frac{s_0^2 \cdot T_B}{N_0}}\right) = 1/2 \cdot \text{erfc}\left(\sqrt{\frac{s_0^2 \cdot T_B}{2 \cdot N_0}}\right).$$

Berücksichtigt man weiter, dass die bei BPSK aufgewandte Energie pro Bit

$$E_B = 1/2 \cdot s_0^2 \cdot T_B$$

beträgt, so kann diese Gleichung wie folgt umgeformt werden:

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

Es ergibt sich somit genau die gleiche Formel wie bei der **Basisbandübertragung**, bei der jedoch für die „Energie pro Bit“  $E_B = s_0^2 \cdot T_B$  zu verwenden war und nicht wie hier  $E_B = 1/2 \cdot s_0^2 \cdot T_B$ .

*Anmerkung:* Diese letzte Gleichung gilt nicht nur bei Rechteck–Quellensignal  $\Rightarrow H_S(f) = \text{si}(\pi f t)$ , sondern für jedes beliebige  $H_S(f)$ , solange

- das Empfangsfilter  $H_E(f) = H_S^*(f)$  exakt an den Sender angepasst ist,
- das Produkt  $H_S(f) \cdot H_E(f)$  das erste Nyquistkriterium erfüllt.

## Fehlerwahrscheinlichkeit des optimalen ASK-Systems

Wir betrachten nun ein **ASK-System** bei gleichen Voraussetzungen wie das BPSK-System. Hier

- sind alle Detektionsnutzsignalwerte  $d_S(vT)$  entweder 0 oder  $s_0$ ,
- ist dementsprechend der Abstand von der Schwelle  $E = s_0/2$  jeweils  $s_0/2$ ,
- ist der Rauscheffektivwert  $\sigma_d$  genau so groß wie bei BPSK,
- ist die Energie pro Bit nur halb so groß wie bei BPSK:

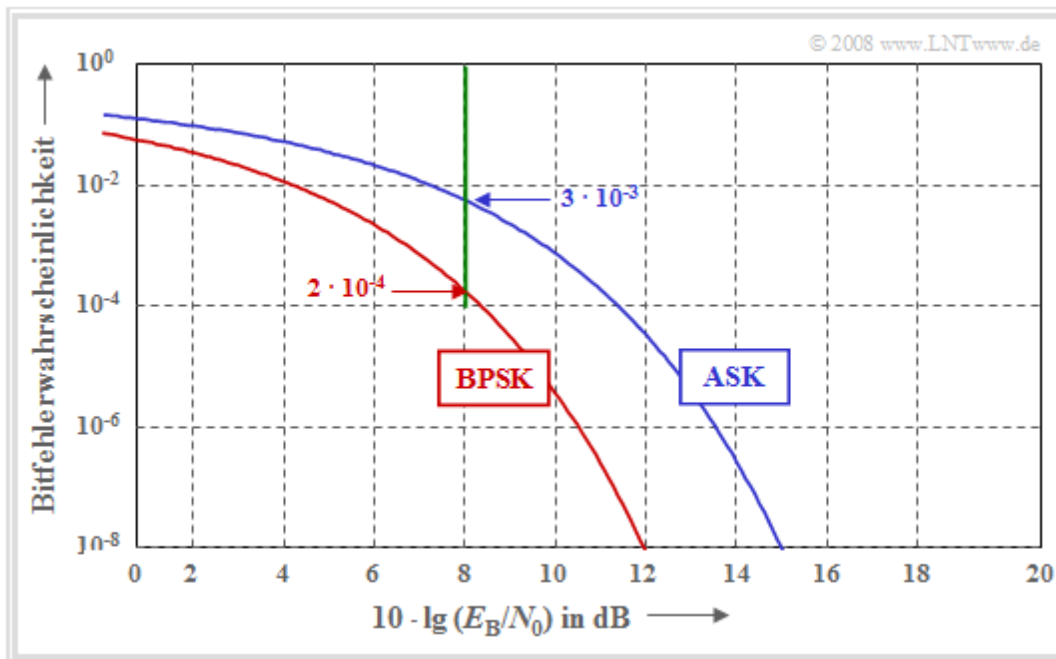
$$E_B = 1/4 \cdot s_0^2 \cdot T_B.$$

Damit lauten die entsprechenden Gleichungen für die Bitfehlerwahrscheinlichkeit:

$$p_B = Q\left(\frac{s_0/2}{\sigma_d}\right) = Q\left(\sqrt{\frac{s_0^2 \cdot T_B}{4 \cdot N_0}}\right),$$

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

Die Grafik zeigt die Fehlerwahrscheinlichkeiten von ASK und BPSK abhängig vom Quotienten  $E_B/N_0$ .



Man erkennt aus dieser doppelt-logarithmischer Darstellung:

- Die ASK-Kurve liegt um 3 dB rechts von der BPSK-Kurve. Für die Fehlerwahrscheinlichkeit  $10^{-8}$  benötigt man bei BPSK etwa  $10 \cdot \lg(E_B/N_0) = 12$  dB, bei ASK dagegen ca. 15 dB.
- Der Systemvergleich bei dem festen Abszissenwert  $10 \cdot \lg(E_B/N_0) = 8$  dB liefert die beiden Bitfehlerwahrscheinlichkeiten  $2 \cdot 10^{-4}$  (für BPSK) bzw.  $6 \cdot 10^{-3}$  (für ASK).

## Fehlerwahrscheinlichkeit bei 4-QAM und 4-PSK (1)

Im **Kapitel 4.3** des Buches „Modulationsverfahren“ wurde die Quadraturamplitudenmodulation (QAM) ausführlich beschrieben. Dem **Blockschaltbild** ist zu entnehmen:

- Die 4-QAM kann durch zwei zueinander **orthogonale BPSK-Systeme** mit Cosinus- bzw. Minus-Sinus-Träger dargestellt werden.
- Das binäre Quellensignal  $q(t)$  mit der Bitdauer  $T_B \Rightarrow$  Bitrate  $R_B$  wird in zwei Teilsignale  $q_I(t)$  und  $q_Q(t)$  mit jeweils halber Rate aufgespaltet (Seriell-Parallel-Wandlung). Die Symboldauer von  $q_I(t)$  bzw.  $q_Q(t)$  beträgt jeweils  $T = 2 \cdot T_B$ , die Symbolrate jeweils  $R_B/2$ .
- Die Amplituden der beiden orthogonalen Trägersignale sind um den Faktor „Wurzel aus 2“ kleiner gewählt als bei der BPSK, so dass die Hüllkurve des **Sendesignals**  $s(t)$  wiederum  $s_0$  beträgt.

Die Fehlerwahrscheinlichkeit der QAM ist die gleiche wie die der zwei orthogonalen BPSK-Systemen. Wegen der kleineren Signalamplitude und der niedrigeren Symbolrate gilt:

$$p_B = Q\left(\frac{s_0/\sqrt{2}}{\sigma_d}\right) \quad \text{mit} \quad \sigma_d^2 = \frac{N_0}{2 \cdot T_B}$$
$$\Rightarrow p_B = Q\left(\sqrt{\frac{s_0^2}{2} \cdot \frac{2 \cdot T_B}{N_0}}\right) = Q\left(\sqrt{2 \cdot E_B/N_0}\right).$$

Das heißt: Obwohl mit der 4-QAM gegenüber der BPSK die doppelte Informationsmenge übertragen werden kann, ergibt sich in Abhängigkeit von  $E_B/N_0$  die genau gleiche Bitfehlerwahrscheinlichkeit. Berücksichtigt ist hierbei, dass auch bei der 4-QAM für die mittlere Energie pro Bit gilt:

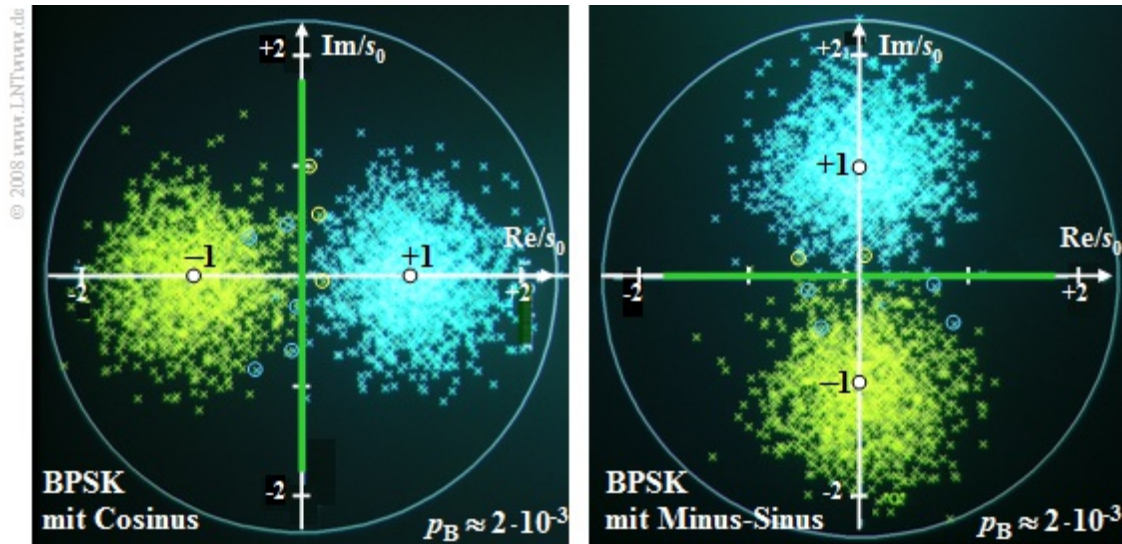
$$E_B = 1/2 \cdot s_0^2 \cdot T_B.$$

Da sich die **quaternäre Phasenmodulation** (4-PSK) von der 4-QAM nur um eine Phasenverdrehung von  $45^\circ$  unterscheidet, ergibt sich auch für diese bei Berücksichtigung geeigneter Entscheidungsgebiete die gleiche Fehlerwahrscheinlichkeit.



## Fehlerwahrscheinlichkeit bei 4-QAM und 4-PSK (2)

Die obere Grafik zeigt die Phasendiagramme der BPSK mit Cosinus- bzw. Minus-Sinusträger, jeweils für  $10 \cdot \lg E_B/N_0 = 6 \text{ dB}$ . *Hinweis:* In der linken (bzw. rechten) Grafik erkennt man Bitfehler durch gelbe Kreuze rechts (bzw. oberhalb) der Entscheidungsschwelle. Zur besseren Kenntlichmachung sind Bitfehler durch Kreise hervorgehoben, zum Beispiel gelbe Punkte rechts bzw. oberhalb der Schwelle.



Der Abstand der Nutzabtastwerte ohne Rauschen (markiert durch die weißen Punkte) von der jeweiligen Entscheidungsschwelle (grün markiert) beträgt jeweils  $s_0$ . Die Varianz der Detektionsabtastwerte – erkennbar an den Punktwolken – ist gleich

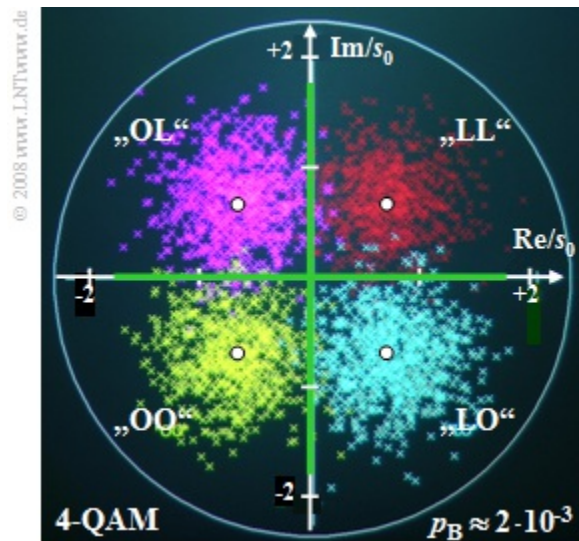
$$\sigma_d^2 = \frac{N_0}{T_B} = \frac{s_0^2/2}{E_B/N_0}, \quad \text{wegen } E_B = 1/2 \cdot s_0^2 \cdot T_B.$$

Mit  $10 \cdot \lg E_B/N_0 = 6 \text{ dB} \Rightarrow E_B/N_0 = 10^{0.6} \approx 4$  ergibt sich daraus:

$$\begin{aligned} \sigma_d^2/s_0^2 &= [2 \cdot 10^{0.6}]^{-1} \approx 0.125 \Rightarrow \sigma_d/s_0 \approx 0.35 \\ \Rightarrow p_B &= Q(s_0/\sigma_d) = Q(\sqrt{2 \cdot 10^{0.6}}) = Q(2.8) \approx 2 \cdot 10^{-3}. \end{aligned}$$

Das untere Bild gilt für die 4-QAM, die man als zwei orthogonale BPSK-Systeme mit Cosinus- und Minus-Sinusträger auffassen kann.

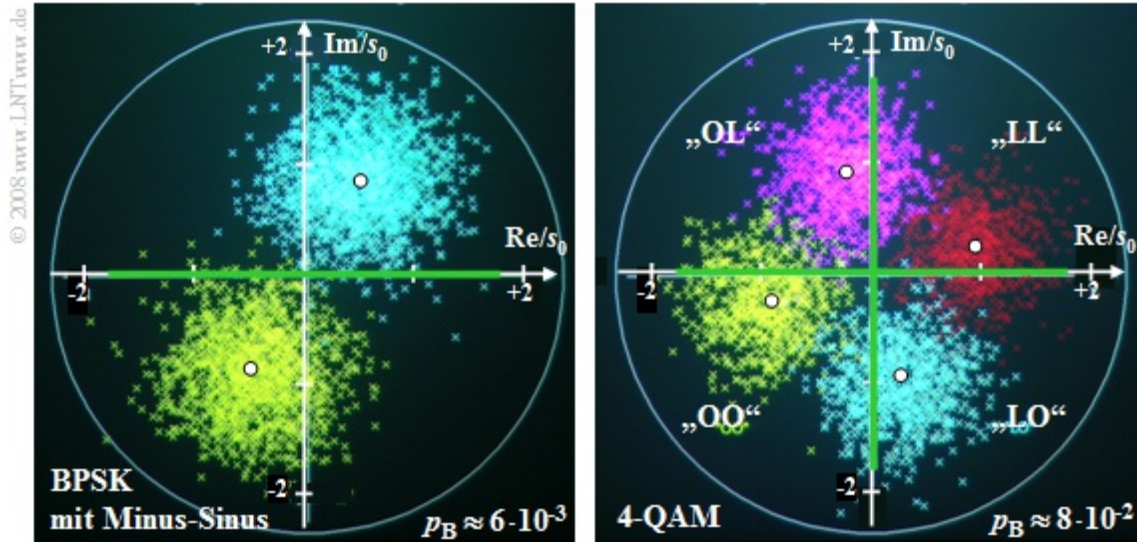
- Der Abstand der nunmehr vier Nutzabtastwerte ohne Rauschen (markiert durch weiße Punkte) vom Ursprung ist wieder  $s_0$ .
- Der Abstand zu den Entscheidungsschwellen ist nun allerdings um den Faktor „Wurzel aus 2“ geringer als bei BPSK, aber auch der Rauscheffektivwert  $\sigma_d$  ist um den gleichen Faktor kleiner.
- Die Bitfehlerwahrscheinlichkeit für 4-QAM und BPSK ist somit exakt gleich:  $p_B = 2 \cdot 10^{-3}$ .





## Phasenversatz zwischen Sender und Empfänger

Voraussetzung für die Gültigkeit der bisherigen Gleichungen ist eine strenge Synchronität zwischen den bei Sender und Empfänger zugesetzten Trägersignalen. Nun wird ein Phasenversatz  $\Delta\phi_T$  zwischen den beiden Trägersignalen angenommen, während weiterhin von Frequenzsynchronität ausgegangen wird.



Die Grafik zeigt die Phasendiagramme für  $\Delta\phi_T = 30^\circ$ . Man erkennt:

- Sowohl bei BPSK (links) als auch bei der 4-QAM (rechts) bewirkt ein Phasenversatz um  $\Delta\phi_T$  eine **Drehung** des Phasendiagramms.
- Bei BPSK bewirkt der Phasenversatz ein um  $\cos(\Delta\phi_T)$  kleineres Nutzsignal. Den gleichen Effekt haben wir bereits beim **Synchrondemodulator** eines analogen Übertragungssystems festgestellt.
- Demzufolge wird auch der Abstand des Detektionsnutzsignals von der Entscheiderschwelle um den gleichen Faktor geringer, was zu einer höheren Fehlerwahrscheinlichkeit führt:

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

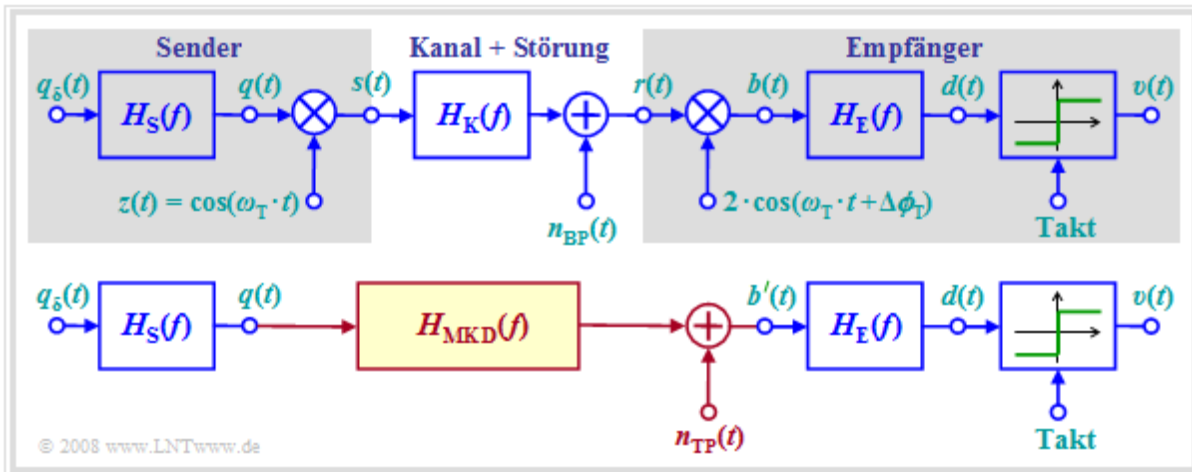
Mit den hier zugrundeliegenden Zahlenwerten ( $10 \cdot \lg E_B/N_0 = 6 \text{ dB}$ ,  $\Delta\phi_T = 30^\circ$ ) erhöht sich die Fehlerwahrscheinlichkeit der BPSK von 0.2% auf etwa 0.6%. Dagegen wird bei der 4-QAM (rechtes Diagramm) die Fehlerwahrscheinlichkeit mit  $\phi_T = 30^\circ$  nahezu um den Faktor 40 größer ( $p_B \approx 8\%$ ). Allgemein gilt hier für die 4-QAM, falls  $|\Delta\phi_T| < 45^\circ$ , wie in **Aufgabe A1.9** gezeigt werden soll:

$$p_B = \frac{1}{2} \cdot Q\left(\sqrt{\frac{2 \cdot E_B}{N_0}} \cdot \frac{\cos(45^\circ)}{\cos(45^\circ + \Delta\phi_T)}\right) + \frac{1}{2} \cdot Q\left(\sqrt{\frac{2 \cdot E_B}{N_0}} \cdot \frac{\cos(45^\circ)}{\cos(45^\circ - \Delta\phi_T)}\right).$$

**Fazit:** Obwohl man mit der 4-QAM über den gleichen Kanal die doppelte Information wie bei BPSK übertragen kann, weisen bei idealen Bedingungen beide Systeme die gleiche Übertragungsqualität (Fehlerwahrscheinlichkeit) auf. Bei nicht idealen Bedingungen – zum Beispiel einem Phasenversatz – steigt allerdings die Fehlerwahrscheinlichkeit der 4-QAM sehr viel stärker an als bei der BPSK.

## Basisbandmodell für ASK und BPSK (1)

Die Grafik zeigt oben nochmals das Gesamtschaltbild eines Trägerfrequenzsystems mit kohärenter Demodulation, das für ASK (unipolare Amplitudenkoeffizienten) und BPSK (bipolare Koeffizienten) in gleicher Weise gültig ist. Durch die Multiplikation mit dem Trägersignal  $z(t)$  wird das Spektrum  $Q(f)$  des Quellensignals – und dementsprechend auch das Leistungsdichtespektrum  $\Phi_q(f)$  – um die Trägerfrequenz  $\pm f_T$  beidseitig verschoben. Nach dem Kanal wird diese Verschiebung durch den Synchrondemodulator wieder rückgängig gemacht.



Geht man vom **äquivalenten Basisbandmodell** entsprechend der unteren Grafik aus, so lässt sich die Berechnung der Signale nach dem Demodulator vereinfachen:

- Man kürzt quasi den Einfluss von Modulator und Demodulator und ersetzt den Bandpasskanal mit dem Frequenzgang  $H_K(f)$  durch eine geeignete Tiefpass-Übertragungsfunktion  $H_{MKD}(f)$ , wobei der Index für „Modulator–Kanal– Demodulator“ steht.
- Unter Berücksichtigung einer Phasendifferenz  $\Delta\phi_T$  zwischen den Trägersignalen von Sender und Empfänger erhält man für die resultierende Übertragungsfunktion:

$$H_{MKD}(f) = \frac{1}{2} \cdot [e^{-j \cdot \Delta\phi_T} \cdot H_K(f - f_T) + e^{j \cdot \Delta\phi_T} \cdot H_K(f + f_T)] .$$

- Bei einem reellen und um die Trägerfrequenz  $f_T$  symmetrischen Kanalfrequenzgang  $H_K(f)$  – also falls  $H_K(f_T - f) = H_K(f_T + f)$  gilt – kann man diese Gleichung wie folgt vereinfachen:

$$H_{MKD}(f) = \frac{\cos(\Delta\phi_T)}{2} \cdot [H_K(f - f_T) + H_K(f + f_T)] .$$

- Die Signale  $b'(t)$  im unteren Bild sowie  $b(t)$  nach dem Demodulator des Bandpass-Systems (oberes Bild) sind somit bis auf die  $\pm 2f_T$ -Anteile identisch. Diese Anteile werden jedoch durch das Empfangsfilter  $H_E(f)$  eliminiert.

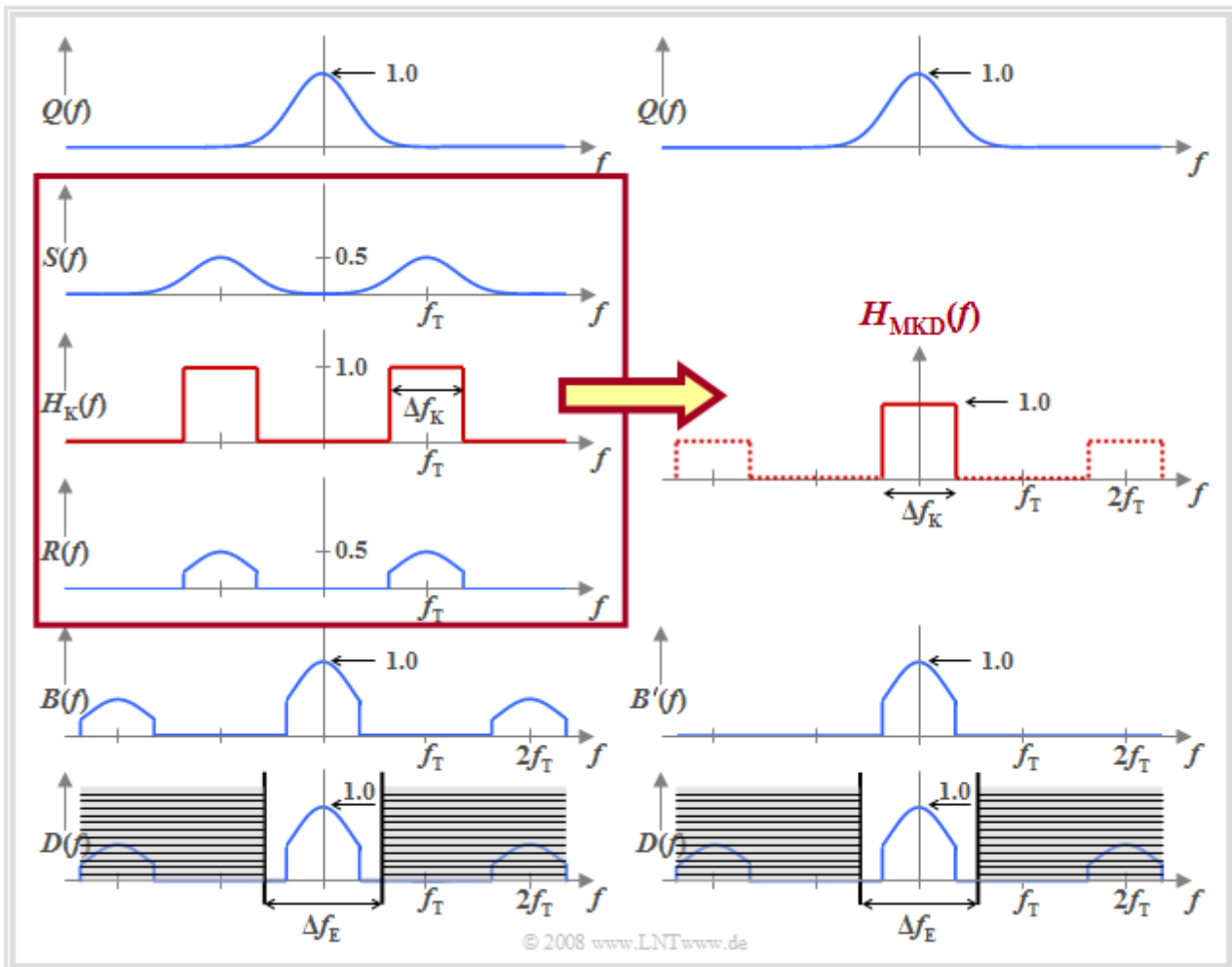
Die Bitfehlerwahrscheinlichkeiten von ASK und BPSK können somit auch mit dem einfacheren Basisbandmodell berechnet werden, und zwar auch dann, wenn ein verzerrender Kanal  $H_K(f)$  vorliegt. Zu beachten ist, dass auch das Rauschsignal  $n(t)$  in den Tiefpassbereich transformiert werden muss. Bei **weißem Rauschen** muss hierzu  $\Phi_n(f) = N_0/2$  durch  $\Phi_{n,TP}(f) = N_0$  ersetzt werden.

## Basisbandmodell für ASK und BPSK (2)

Die Grafik verdeutlicht das Basisbandmodell anhand der Amplitudenspektren, wobei vereinfachend

- ein gaußförmiges  $Q(f)$ ,
- BPSK-Modulation,
- ein rechteckförmiger Bandpasskanal  $H_K(f)$ ,
- phasensynchrone Demodulation, und
- ein ebenfalls rechteckförmiges Empfangsfilter  $H_E(f)$  mit  $\Delta f_E > \Delta f_K$

vorausgesetzt werden.



Man erkennt:

- Das Spektrum  $D(f)$  wird durch das äquivalente Basisbandmodell richtig wiedergegeben, obwohl sich die Spektren  $B(f)$  bzw.  $B'(f)$  um die doppelte Trägerfrequenz unterscheiden.
- Die resultierende Übertragungsfunktion  $H_{MKD}(f)$  berücksichtigt auch die Bandbegrenzung durch den Kanal, der in diesem Beispiel als rechteckförmig um die Trägerfrequenz angenommen wurde.