

Überblick zu Kapitel 4 des Buches „Digitalsignalübertragung“

Das Kapitel 4 liefert eine abstrahierte Beschreibung der Digitalsignalübertragung, die auf so genannten *Basisfunktionen* und *Signalraumkonstellationen* aufbaut. Dadurch ist es möglich, sehr unterschiedliche Konfigurationen – zum Beispiel Bandpass-Systeme und solche für das Basisband – in einheitlicher Form zu behandeln. Der jeweils optimale Empfänger besitzt in allen Fällen die gleiche Struktur.

Im Einzelnen werden behandelt:

- die Bedeutung von *Basisfunktionen* und deren Auffinden nach dem *Gram–Schmidt–Verfahren*,
- die Struktur des *optimalen Empfängers* für die Basisbandübertragung,
- das *Theorem der Irrelevanz* und dessen Bedeutung für die Herleitung optimaler Detektoren,
- der optimale Empfänger für den *AWGN–Kanal* und *Implementierungsaspekte*,
- die Systembeschreibung durch komplexes bzw. *N*–dimensionales *Gaußsches Rauschen*,
- die *Fehlerwahrscheinlichkeitsberechnung* und –*approximation* bei sonst idealen Bedingungen,
- die Anwendung der Signalraumbeschreibung auf *Trägerfrequenzsysteme*,
- die unterschiedlichen Ergebnisse für *OOK*, *M–ASK*, *M–PSK*, *M–QAM* und *M–FSK*,
- die unterschiedlichen Ergebnisse für *kohärente* bzw. *nichtkohärente* Demodulation.

Viele der in diesem Kapitel hergeleiteten Ergebnisse wurden bereits in anderen Büchern oder Kapiteln angegeben. Im Mittelpunkt der folgenden Beschreibungen stehen also weniger diese Ergebnisse als die Vorgehensweise, diese in einer möglichst strukturierten Weise zu finden.

Geeignete Literatur: [CT06] – [Gal08] – [Gra06] – [Kam04] – [Kra10] – [KöZ08] – [PM96] – [Pro01] – [PS02] – [PS04] – [Sha48] – [WJ65]

Die grundlegende Theorie wird auf 56 Bildschirmseiten dargelegt. Außerdem beinhaltet dieses Kapitel 84 Grafiken, 19 Aufgaben und neun Zusatzaufgaben mit insgesamt 146 Teilaufgaben, sowie sieben Lernvideos (LV) und sieben Interaktionsmodule (IM), unter anderem:

- **AWGN–Kanal – Teil 1** (LV, Grundlagen – Dauer 6:00)
- **Momente von diskreten Zufallsgrößen** (LV, Kapitel 4.1 – Dauer 6:30)
- **Statistische Abhängigkeit/Unabhängigkeit** (LV, Kapitel 4.2 – Dauer 11:40)
- **Wahrscheinlichkeit und WDF** (LV, Kapitel 4.2 – Dauer 12:05)
- **Mengentheoretische Begriffe und Gesetze** (LV, Kapitel 4.3 – Dauer 12:20)
- **Komplementäre Gaußsche Fehlerfunktionen** (IM, Grundlagen)
- **Gram–Schmidt–Verfahren** (IM, Kapitel 4.1)
- **Zweidimensionale Gauß–Verteilung** (IM, Kapitel 4.2)
- **Zweidimensionale Laplace–Verteilung** (IM, Kapitel 4.2)
- **Optimale Entscheidungsregionen** (IM, Kapitel 4.3)
- **Mehrstufige PSK und Union–Bound** (IM, Kapitel 4.4)
- **Nichtkohärentes On–Off–Keying** (IM, Kapitel 4.5)

Zur Nomenklatur im vierten Kapitel (1)

Nahezu alle Ergebnisse dieses Kapitels wurden bereits in früheren Abschnitten hergeleitet. Grundlegend neu ist jedoch die Herangehensweise:

- Im Buch „Modulationsverfahren“ sowie in den ersten drei Kapiteln dieses Buches wurden bereits bei den Herleitungen die spezifischen Systemeigenschaften berücksichtigt – zum Beispiel, ob die Übertragung des Digitalsignals im Basisband erfolgt oder ob eine digitale Amplituden-, Frequenz- oder Phasenmodulation vorliegt.
- Hier sollen nun die Systeme dahingehend abstrahiert werden, dass sie einheitlich behandelt werden können. Der jeweils optimale Empfänger besitzt in allen Fällen die gleiche Struktur, und die Fehlerwahrscheinlichkeit lässt sich auch für nichtgaußverteiltes Rauschen angeben.

Anzumerken ist, dass sich durch diese eher globale Vorgehensweise gewisse Systemunzulänglichkeiten nicht oder nur sehr ungenau erfassen lassen, wie beispielsweise

- der Einfluss eines nichtoptimalen Empfangsfilters auf die Fehlerwahrscheinlichkeit,
- ein falscher Schwellenwert (Schwellendrift) oder
- Phasenjitter (Schwankungen der Abtastzeitpunkte).

Insbesondere bei Vorhandensein von Impulsinterferenzen sollte also weiterhin entsprechend **Kapitel 3** vorgegangen werden.

Die Beschreibung basiert auf dem Skript **[KöZ08]** von **Ralf Kötter** und **Georg Zeitler**, das sich stark an das Buch **[WJ65]** anlehnt. **Gerhard Kramer**, Lehrstuhlinhaber des LNT seit 2010, behandelt in seiner NT2-Vorlesung **[Kra10]** die gleiche Thematik mit sehr ähnlicher Nomenklatur.

Um unseren eigenen Studenten an der TU München das Lesen nicht unnötig zu erschweren, halten wir uns weitestgehend an diese Nomenklatur, auch wenn diese von anderen *LNTwww*-Kapiteln abweicht.

Zur Nomenklatur im vierten Kapitel (2)

Gegenüber den anderen Kapiteln in *LNTwww* ergeben sich hier folgende Nomenklaturunterschiede:

- Die zu übertragende **Nachricht** ist ein ganzzahliger Wert $m \in \{m_i\}$ mit $i = 0, \dots, M - 1$, wobei M den Symbolumfang angibt. Wenn es die Beschreibung vereinfacht, wird $i = 1, \dots, M$ induziert.
- Das Ergebnis des Entscheidungsprozesses beim Empfänger ist ebenfalls ein Integerwert mit dem gleichen Symbolalphabet wie beim Sender. Man bezeichnet dieses Ergebnis auch als **Schätzwert**:
 $\hat{m} \in \{m_i\}, i = 0, 1, \dots, M - 1$ (bzw. $i = 1, 2, \dots, M$).

- Die **Symbolfehlerwahrscheinlichkeit** wird meist wie folgt bezeichnet:

$$\Pr(\mathcal{E}) = \Pr(\hat{m} \neq m) = 1 - \Pr(\mathcal{C}), \quad \text{Komplementärereignis: } \Pr(\mathcal{C}) = \Pr(\hat{m} = m).$$

Im Fließtext wird aufgrund des durch HTML eingeschränkten Zeichensatzes „Pr(Symbolfehler)“ oder auch „ p_S “ verwendet.

- Bei einer **Wahrscheinlichkeitsdichtefunktion** (WDF) wird nun entsprechend $p_r(\rho)$ zwischen der Zufallsgröße (r) und der Realisierung (ρ) unterschieden. Bisher wurde für eine WDF die Bezeichnung „ $f_r(r)$ “ verwendet – siehe **Kapitel 3.1** im Buch „Stochastische Signaltheorie“.

- Mit der Schreibweise $p_r(\rho)$ geben r und ρ Skalare an. Sind dagegen Zufallsgröße und Realisierung **Vektoren** (geeigneter Länge), so wird dies durch Fettschrift ausgedrückt:

$p_r(\boldsymbol{\rho})$ mit den Vektoren \boldsymbol{r} und $\boldsymbol{\rho}$.

- Um Verwechslungen mit Energiewerten zu vermeiden, heißt nun der *Schwellenwert* G anstelle von E und wird in diesem Kapitel vorwiegend als **Entscheidungsgrenze** bezeichnet.
- Ausgehend von den beiden reellen und energiebegrenzten Zeitfunktionen $x(t)$ und $y(t)$ erhält man für das **innere Produkt**

$$\langle x(t), y(t) \rangle = \int_{-\infty}^{+\infty} x(t) \cdot y(t) dt,$$

und für die **2-Norm** (oder kurz Norm):

$$\|x(t)\| = \sqrt{\langle x(t), x(t) \rangle}.$$

Gegenüber dem Skript **[KöZ08]** unterscheidet sich die Bezeichnungsweise hier wie folgt:

- Die **Wahrscheinlichkeit** des Ereignisses „E“ ist $\Pr(„E“)$; in [KöZ08] wird hier $P(„E“)$ verwendet. Diese Nomenklaturänderung wurde auch deshalb vorgenommen, da Wahrscheinlichkeiten und Leistungen in manchen Gleichungen gemeinsam vorkommen.
- **Bandpass-Signale** werden weiterhin mit Index „BP“ gekennzeichnet und nicht wie in [KöZ08] mit einer Tilde. Das entsprechende Tiefpass-Signal ist (meist) mit dem Index „TP“ versehen.

Orthonormale Basisfunktionen (1)

Wir gehen in diesem Kapitel von einem Satz $\{s_i(t)\}$ möglicher Sendesignale aus, die den möglichen Nachrichten m_i eindeutig zugeordnet werden können. Mit $i = 1, \dots, M$ gilt:

$$m \in \{m_i\}, s(t) \in \{s_i(t)\}: m = m_i \Leftrightarrow s(t) = s_i(t).$$

Weiter setzen wir für das Folgende voraus, dass die M Signale $s_i(t)$ **energiebegrenzt** sind, was meist gleichzeitig bedeutet, dass sie nur von endlicher Dauer sind.

Satz: Eine jede Menge $\{s_1(t), \dots, s_M(t)\}$ energiebegrenzter Signale lässt sich in $N \leq M$ orthonormale **Basisfunktionen** $\varphi_1(t), \dots, \varphi_N(t)$ entwickeln, wobei gilt:

$$s_i(t) = \sum_{j=1}^N s_{ij} \cdot \varphi_j(t), \quad i = 1, \dots, M, \quad j = 1, \dots, N.$$

Jeweils zwei Basisfunktionen $\varphi_j(t)$ und $\varphi_k(t)$ müssen **orthonormal** zueinander sein, das heißt, es muss gelten (δ_{jk} nennt man das Kronecker-Symbol):

$$\langle \varphi_j(t), \varphi_k(t) \rangle = \int_{-\infty}^{+\infty} \varphi_j(t) \cdot \varphi_k(t) dt = \delta_{jk} = \begin{cases} 1 & \text{falls } j = k \\ 0 & \text{falls } j \neq k \end{cases}.$$

Der Parameter N gibt dabei an, wieviele Basisfunktionen $\varphi_j(t)$ benötigt werden, um die M möglichen Sendesignale darzustellen. Mit anderen Worten: N ist die **Dimension des Vektorraums**, der von den M Signalen aufgespannt wird. Dabei gilt:

- Ist $N = M$, so sind alle Sendesignale zueinander orthogonal. Sie sind nicht notwendigerweise orthonormal, das heißt, die Energien $E_i = \langle s_i(t), s_i(t) \rangle$ können durchaus ungleich 1 sein.
- $N < M$ ergibt sich, wenn mindestens ein Signal $s_i(t)$ als Linearkombination von Basisfunktionen $\varphi_j(t)$ dargestellt werden kann, die sich aus anderen Signalen $s_j(t) \neq s_i(t)$ ergeben haben.

Orthonormale Basisfunktionen (2)

Beispiel: Wir betrachten $M = 3$ energiebegrenzte Signale gemäß der Grafik. Man erkennt sofort, dass

- $s_1(t)$ und $s_2(t)$ zueinander orthogonal sind,
- die Energie $E_1 = A^2 \cdot T = E$ ist und $E_2 = E/4$ gilt,
- $\varphi_1(t)$ und $\varphi_2(t)$ jeweils formgleich mit $s_1(t)$ bzw. $s_2(t)$ sind und beide die Energie 1 besitzen:

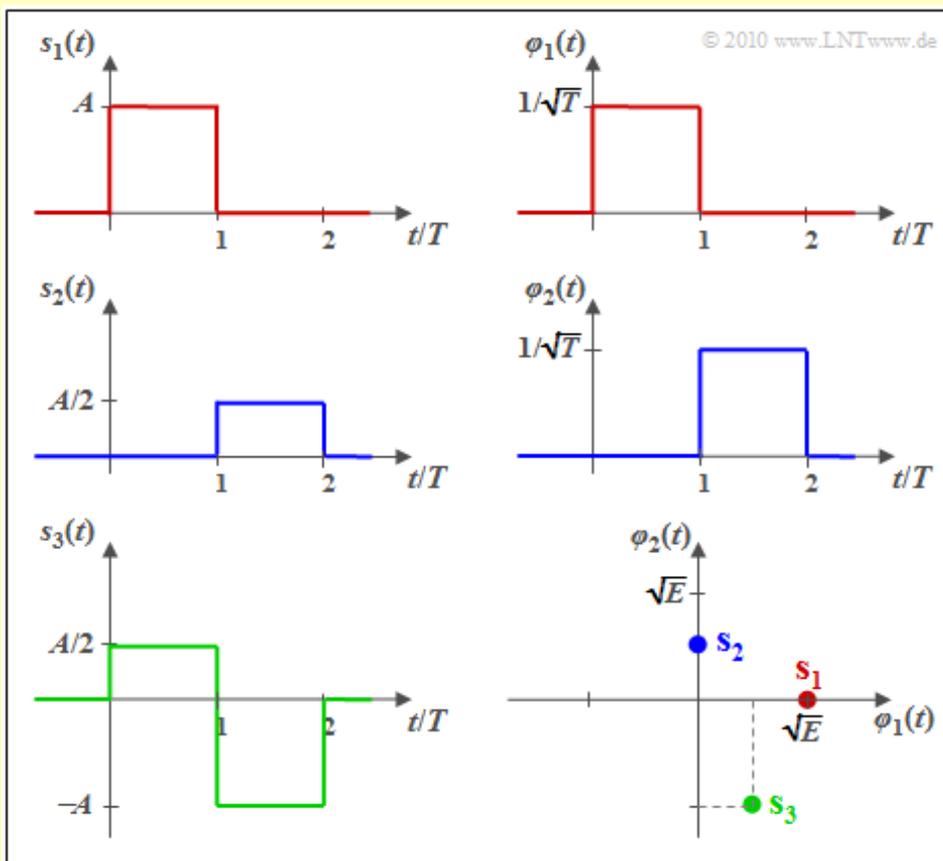
$$\varphi_1(t) = \frac{s_1(t)}{\sqrt{E_1}} = \frac{s_1(t)}{\sqrt{A^2 \cdot T}} = \frac{1}{\sqrt{T}} \cdot \frac{s_1(t)}{A} \Rightarrow s_1(t) = s_{11} \cdot \varphi_1(t), \quad s_{11} = \sqrt{E},$$

$$\varphi_2(t) = \frac{s_2(t)}{\sqrt{E_2}} = \frac{s_2(t)}{\sqrt{(A/2)^2 \cdot T}} = \frac{1}{\sqrt{T}} \cdot \frac{s_2(t)}{A/2} \Rightarrow s_2(t) = s_{21} \cdot \varphi_2(t), \quad s_{21} = \frac{\sqrt{E}}{2}.$$

- $s_3(t)$ durch die Basisfunktionen $\varphi_1(t)$ und $\varphi_2(t)$ ausgedrückt werden kann:

$$s_3(t) = s_{31} \cdot \varphi_1(t) + s_{32} \cdot \varphi_2(t),$$

$$s_{31} = A/2 \cdot \sqrt{T} = \sqrt{E}/2, \quad s_{32} = -A \cdot \sqrt{T} = -\sqrt{E}.$$



Im rechten unteren Bild sind die Signale in einer 2D-Darstellung mit den Basisfunktionen $\varphi_1(t)$ und $\varphi_2(t)$ als Achsen dargestellt, wobei $E = A^2 \cdot T$ gilt und der Zusammenhang zu den anderen Grafiken durch die Farbgebung zu erkennen ist. Die vektoriellen Repräsentanten der Signale $s_1(t)$, $s_2(t)$ und $s_3(t)$ in diesem zweidimensionalen Vektorraum lassen sich daraus wie folgt ablesen:

$$\mathbf{s}_1 = (\sqrt{E}, 0), \quad \mathbf{s}_2 = (0, \sqrt{E}/2), \quad \mathbf{s}_3 = (\sqrt{E}/2, -\sqrt{E}).$$

Das Verfahren nach Gram-Schmidt (1)

Im Beispiel auf der letzten Seite war die Angabe der beiden orthonormalen Basisfunktionen $\varphi_1(t)$ und $\varphi_2(t)$ sehr einfach, da diese formgleich mit $s_1(t)$ und $s_2(t)$ waren. Das Gram-Schmidt-Verfahren findet die Basisfunktionen $\varphi_1(t), \dots, \varphi_N(t)$ für beliebig vorgebbare Signale $s_1(t), \dots, s_M(t)$, und zwar wie folgt:

- Die erste Basisfunktion $\varphi_1(t)$ ist formgleich mit $s_1(t)$. Es gilt:

$$\varphi_1(t) = \frac{s_1(t)}{\sqrt{E_1}} = \frac{s_1(t)}{\|s_1(t)\|} \Rightarrow \|\varphi_1(t)\| = 1, \quad s_{11} = \|s_1(t)\|, \quad s_{1j} = 0 \text{ für } j \geq 2.$$

- Es wird nun angenommen, dass aus den Signalen $s_1(t), \dots, s_{k-1}(t)$ bereits die Basisfunktionen $\varphi_1(t), \dots, \varphi_{n-1}(t)$ berechnet wurden ($n \leq k$). Dann berechnen wir mittels $s_k(t)$ die Hilfsfunktion

$$\theta_k(t) = s_k(t) - \sum_{j=1}^{n-1} s_{kj} \cdot \varphi_j(t) \quad \text{mit} \quad s_{kj} = \langle s_k(t), \varphi_j(t) \rangle, \quad j = 1, \dots, n-1.$$

- Ist $\theta_k(t) \equiv 0 \Rightarrow \|\theta_k(t)\| = 0$, so liefert $s_k(t)$ keine neue Basisfunktion. Vielmehr lässt sich dann $s_k(t)$ durch die $n-1$ bereits vorher gefundenen Basisfunktionen $\varphi_1(t), \dots, \varphi_{n-1}(t)$ ausdrücken:

$$s_k(t) = \sum_{j=1}^{n-1} s_{kj} \cdot \varphi_j(t).$$

- Eine neue Basisfunktion (nämlich die n -te) ergibt sich, falls $\|\theta_k(t)\| \neq 0$ ist:

$$\varphi_n(t) = \frac{\theta_k(t)}{\|\theta_k(t)\|} \Rightarrow \|\varphi_n(t)\| = 1.$$

Diese Prozedur kann fortgesetzt werden, bis alle M Signale berücksichtigt wurden. Danach hat man alle $N \leq M$ orthonormalen Basisfunktionen $\varphi_j(t)$ gefunden. Der Sonderfall $N = M$ ergibt sich nur dann, wenn alle M Signale linear voneinander unabhängig sind.

Auf der nächsten Seite wird das Gram-Schmidt-Verfahren an einem einfachen Beispiel verdeutlicht. Wir verweisen auch auf das folgende Interaktionsmodul:

Gram-Schmidt-Verfahren

Das Verfahren nach Gram-Schmidt (2)

Beispiel: Wir betrachten die $M = 4$ energiebegrenzten Signale $s_1(t), \dots, s_4(t)$ entsprechend der Grafik. Zur Vereinfachung der Berechnungen ist hier sowohl die Amplitude als auch die Zeit normiert. Man erkennt:

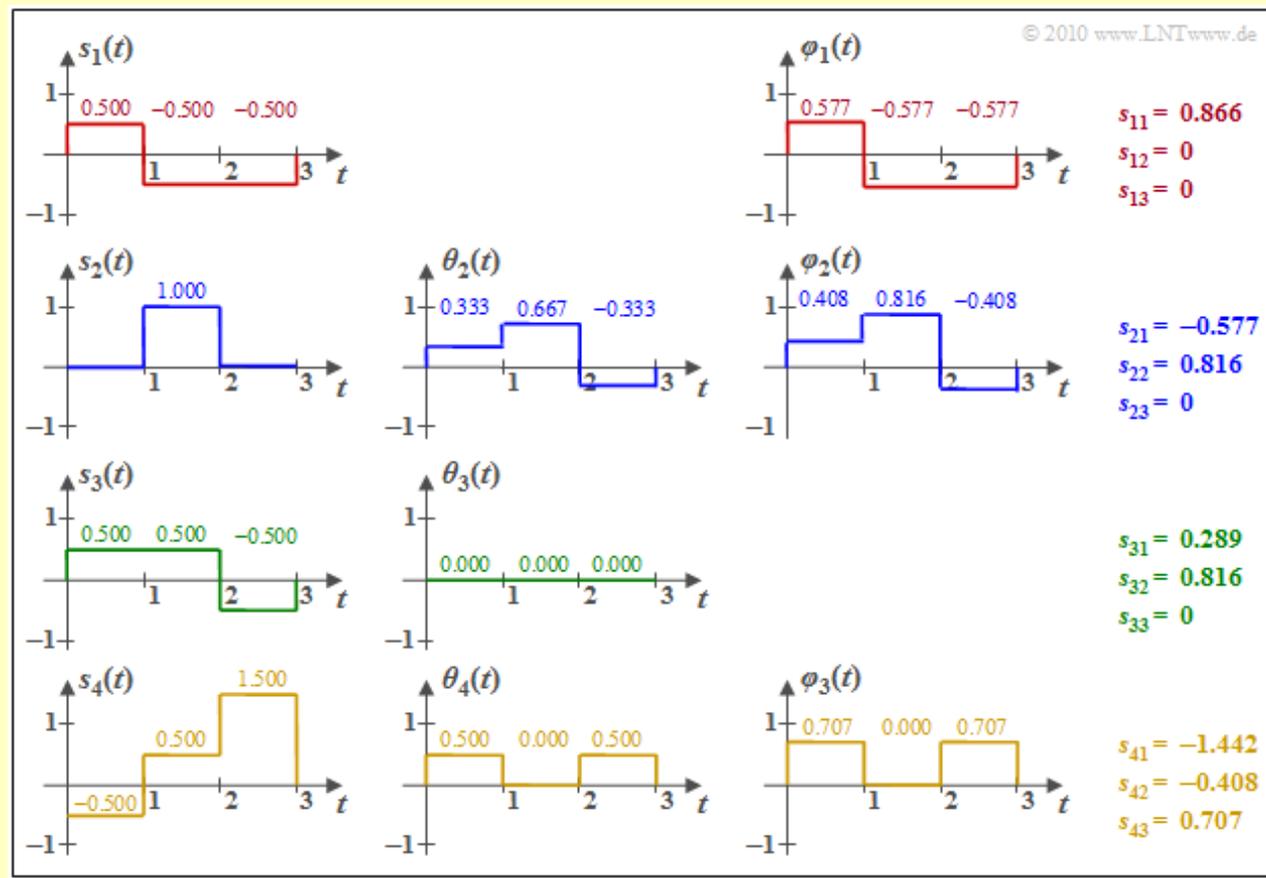
- Die Basisfunktion $\varphi_1(t)$ ist formgleich mit $s_1(t)$. Wegen $E_1 = \|s_1(t)\|^2 = 3 \cdot 0.5^2 = 0.75$ ergibt sich $s_{11} = \|s_1(t)\| = 0.866$. $\varphi_1(t)$ selbst besitzt abschnittsweise die Werte $\pm 0.5/0.866 = \pm 0.577$.
- Zur Berechnung der Hilfsfunktion $\theta_2(t)$ berechnen wir

$$\begin{aligned} s_{21} &= \langle s_2(t), \varphi_1(t) \rangle = 0 \cdot (+0.577) + 1 \cdot (-0.577) + 0 \cdot (-0.577) = -0.577 \\ \Rightarrow \theta_2(t) &= s_2(t) - s_{21} \cdot \varphi_1(t) = (0.333, 0.667, -0.333) \\ \Rightarrow \|\theta_2(t)\|^2 &= (1/3)^2 + (2/3)^2 + (-1/3)^2 = 0.667 \\ \Rightarrow s_{22} &= \sqrt{0.667} = 0.816, \quad \varphi_2(t) = \theta_2(t)/s_{22} = (0.408, 0.816, -0.408). \end{aligned}$$

- Die inneren Produkte zwischen $s_3(t)$ mit $\varphi_1(t)$ bzw. $\varphi_2(t)$ liefern folgende Ergebnisse:

$$\begin{aligned} s_{31} &= \langle s_3(t), \varphi_1(t) \rangle = 0.5 \cdot (+0.577) + 0.5 \cdot (-0.577) - 0.5 \cdot (-0.577) = 0.289 \\ s_{32} &= \langle s_3(t), \varphi_2(t) \rangle = 0.5 \cdot (+0.408) + 0.5 \cdot (+0.816) - 0.5 \cdot (-0.408) = 0.816 \\ \Rightarrow \theta_3(t) &= s_3(t) - 0.289 \cdot \varphi_1(t) - 0.816 \cdot \varphi_2(t) = 0. \end{aligned}$$

Das bedeutet: Die grüne Funktion $s_3(t)$ liefert keine neue Basisfunktion $\varphi_3(t)$, im Gegensatz zur Funktion $s_4(t)$. Die numerischen Ergebnisse hierfür können der Grafik entnommen werden.



Basisfunktionen komplexer Zeitsignale

In der Nachrichtentechnik hat man es oft mit komplexen Zeitfunktionen zu tun,

- nicht etwa, weil es komplexe Signale in der Realität gibt, sondern
- weil die Beschreibung eines BP-Signals im äquivalenten TP-Bereich zu komplexen Signalen führt.

Die Bestimmung der $N \leq M$ komplexwertigen Basisfunktionen $\xi_k(t)$ aus den M komplexen Signalen $s_i(t)$ kann ebenfalls mit dem **Gram-Schmidt-Verfahren** erfolgen, doch ist nun zu berücksichtigen, dass das innere Produkt zweier komplexer Signale $x(t)$ und $y(t)$ wie folgt zu berechnen ist:

$$\langle x(t), y(t) \rangle = \int_{-\infty}^{+\infty} x(t) \cdot y^*(t) dt.$$

Die entsprechenden Gleichungen lauten nun mit $i = 1, \dots, M$ und $k = 1, \dots, N$:

$$s_i(t) = \sum_{k=1}^N s_{ik} \cdot \xi_k(t), \quad s_i(t) \in \mathcal{C}, \quad s_{ik} \in \mathcal{C}, \quad \xi_k(t) \in \mathcal{C},$$

$$\langle \xi_k(t), \xi_j(t) \rangle = \int_{-\infty}^{+\infty} \xi_k(t) \cdot \xi_j^*(t) dt = \delta_{ik} = \begin{cases} 1 & \text{falls } k = j \\ 0 & \text{falls } k \neq j \end{cases}.$$

Natürlich lässt sich jede komplexe Größe auch durch zwei reelle Größen – nämlich durch den Realteil und den Imaginärteil – ausdrücken. Somit erhält man hier folgende Gleichungen:

$$\begin{aligned} s_i(t) &= s_{Ii}(t) + j \cdot s_{Qi}(t), \quad s_{Ii}(t) = \operatorname{Re}[s_i(t)], \quad s_{Qi}(t) = \operatorname{Im}[s_i(t)], \\ \xi_k(t) &= \varphi_k(t) + j \cdot \psi_k(t), \quad \varphi_k(t) = \operatorname{Re}[\xi_k(t)], \quad \psi_k(t) = \operatorname{Im}[\xi_k(t)], \\ s_{ik} &= s_{Iik} + j \cdot s_{Qik}, \quad s_{Iik} = \operatorname{Re}[s_{ik}], \quad s_{Qik} = \operatorname{Im}[s_{ik}], \\ s_{Iik} &= \operatorname{Re}[\langle s_i(t), \varphi_k(t) \rangle], \quad s_{Qik} = \operatorname{Re}[\langle s_i(t), j \cdot \psi_k(t) \rangle]. \end{aligned}$$

Die Nomenklatur ergibt sich aus der Hauptanwendung für komplexe Basisfunktionen, nämlich der **Quadratur-Amplitudenmodulation** (QAM). Der Index „I“ steht für Inphasekomponente und gibt den Realteil an, während die Quadraturkomponente (Imaginärteil) mit dem Index „Q“ gekennzeichnet ist.

Um Verwechslungen mit der imaginären Einheit zu vermeiden, sind hier die komplexen Basisfunktionen $\xi_k(t)$ mit „k“ induziert und nicht mit „j“.

Dimension der Basisfunktionen (1)

Bei der Basisbandübertragung sind die möglichen Sendesignale (Betrachtung nur einer Symboldauer)

$$s_i(t) = a_i \cdot g_s(t), \quad i = 0, \dots, M - 1,$$

wobei $g_s(t)$ den **Sendegrundimpuls** angibt und die a_i in **Kapitel 1** und **Kapitel 2** als die möglichen **Amplitudenkoeffizienten** bezeichnet wurden. Anzumerken ist, dass im bisherigen Kapitel 4.1 für die Laufvariable i die Werte 1 bis M vorausgesetzt wurden und nicht wie hier 0 bis $M - 1$.

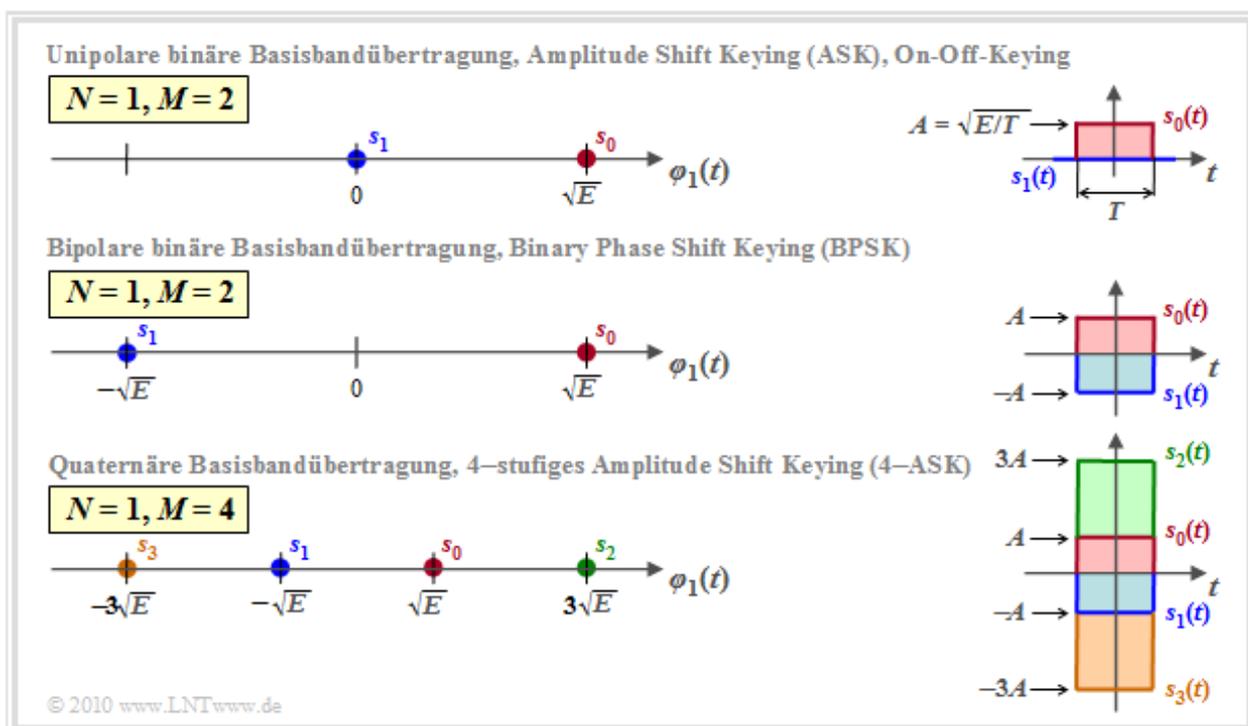
Nach der Beschreibung dieses Kapitels handelt es sich unabhängig von der Stufenzahl M um ein eindimensionales Modulationsverfahren ($N = 1$), wobei bei der Basisbandübertragung

- die Basisfunktion $\varphi_1(t)$ gleich dem energienormierten Sendegrundimpuls $g_s(t)$ ist:

$$\varphi_1(t) = g_s(t) / \sqrt{E_{gs}} \quad \text{mit} \quad E_{gs} = \int_{-\infty}^{+\infty} g_s^2(t) dt,$$

- die dimensionslosen Amplitudenkoeffizienten a_i in die Signalraumpunkte s_i umgerechnet werden können, die die Einheit „Wurzel aus Energie“ aufweisen.

Die Grafik zeigt die Signalraumkonstellationen für die binäre unipolare (oben), die binäre bipolare (Mitte) sowie die quaternäre bipolare (unten) Basisbandübertragung. Rechts sind am Beispiel „Rechteckimpuls“ die zwei bzw. vier möglichen Sendesignale $s_i(t)$ angegeben. Man kann daraus auch den Zusammenhang zwischen Signalenergie E und Impulsamplitude A erkennen. Die jeweils linken Darstellungen auf der φ_1 -Achse gelten aber unabhängig von der Form des Sendegrundimpulses $g_s(t)$, nicht nur für Rechtecke.



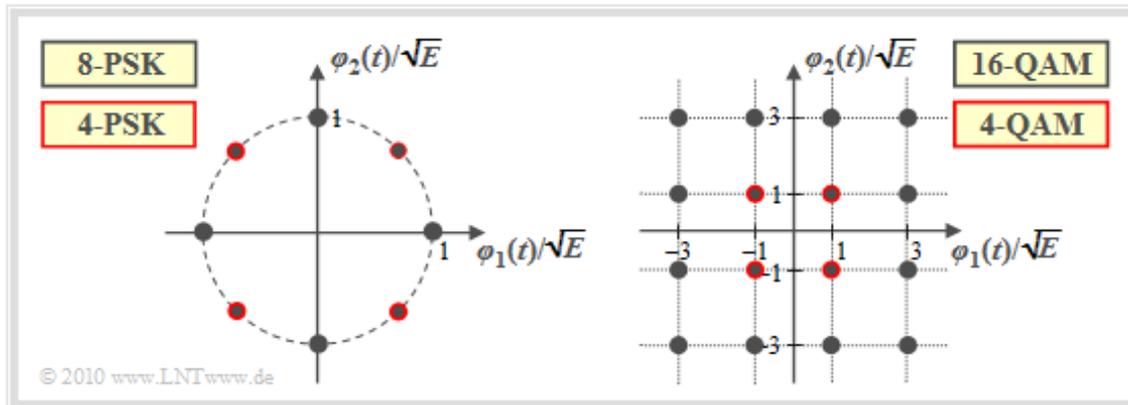
- Die Grafik beschreibt gleichzeitig die eindimensionalen Trägerfrequenzsysteme On-Off-Keying (oben), BPSK bzw. 2-ASK (Mitte) und 4-ASK (unten).
- Die Signale $s_i(t)$ und die Basisfunktion $\varphi_1(t)$ beziehen sich dann auf den äquivalenten TP-Bereich. Im BP-Bereich ist $\varphi_1(t)$ eine auf den Zeitbereich $0 \leq t \leq T$ begrenzte harmonische Schwingung.

Dimension der Basisfunktionen (2)

Zu den zweidimensionalen Modulationsverfahren ($N = 2$) gehören

- M -stufiges *Phase Shift Keying* (M -PSK),
- Quadratur-Amplitudenmodulation (4-QAM, 16-QAM, 64-QAM, ...),
- binäres (orthogonales) *Frequency Shift Keying* (2-FSK).

Allgemein ist bei orthogonaler FSK die Anzahl N der Basisfunktionen $\varphi_k(t)$ gleich der Anzahl M der möglichen Sendesignale $s_i(t)$. $N = 2$ ist deshalb nur für $M = 2$ möglich.



Die linke Grafik zeigt die 8-PSK-Konstellation. Beschränkt man sich auf die rot umrandeten Punkte, so liegt eine 4-PSK (*Quaternary Phase Shift Keying*, QPSK) vor.

Die rechte Grafik bezieht sich auf die 16-QAM beziehungsweise – wenn man nur die rot umrandeten Signalraumpunkte betrachtet – auf die 4-QAM. Ein Vergleich der beiden Bilder zeigt, dass die 4-QAM mit der QPSK bei entsprechender Achsenskalierung identisch ist.

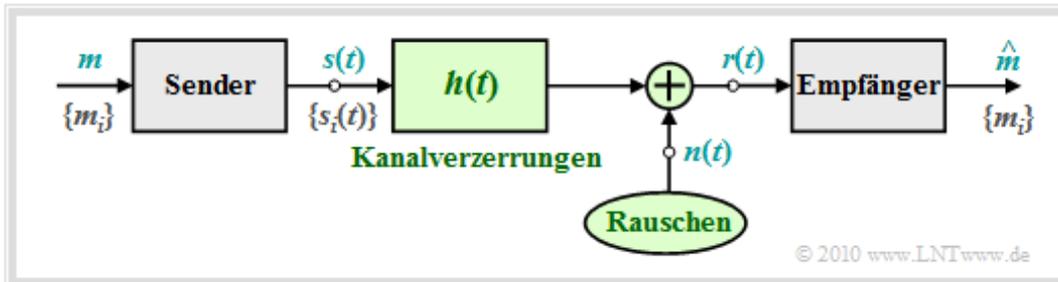
Die Grafiken beschreiben die Modulationsverfahren sowohl im Bandpass- als auch im äquivalenten Tiefpassbereich:

- Bei der Betrachtung als Bandpass-System sind die Basisfunktionen $\varphi_1(t)$ und $\varphi_2(t)$ cosinusförmig bzw. (minus-)sinusförmig – vergleiche hierzu **Aufgabe A4.2**.
- Dagegen ist nach der Transformation der QAM-Systeme in den äquivalenten Tiefpassbereich die Basisfunktion $\varphi_1(t)$ gleich dem energienormierten (Energie 1) Sendegrundimpuls $g_s(t)$, während $\varphi_2(t) = j \cdot \varphi_1(t)$ zu setzen ist. Sie finden Näheres hierzu in der **Aufgabe Z4.2**.

Blockschaltbild und Voraussetzungen

In diesem Kapitel wird die Struktur des optimalen Empfängers eines digitalen Übertragungssystems sehr allgemein hergeleitet, wobei

- das Modulationsverfahren und weitere Systemdetails nicht weiter spezifiziert werden,
- von den Basisfunktionen und der Signalraumdarstellung gemäß **Kapitel 4.1** ausgegangen wird.



Zum obigen Blockschaltbild ist anzumerken:

- Der Symbolumfang der Quelle beträgt M und der Symbolvorrat ist $\{m_i\}$ mit $i = 0, \dots, M - 1$. Die zugehörigen Symbolwahrscheinlichkeiten $\Pr(m = m_i)$ seien auch dem Empfänger bekannt.
- Zur Nachrichtenübertragung stehen M verschiedene Signalformen $s_i(t)$ zur Verfügung, wobei für die Laufvariable ebenfalls die Indizierung $i = 0, \dots, M - 1$ gelten soll.
- Es besteht eine feste Beziehung zwischen den Nachrichten $\{m_i\}$ und den Signalen $\{s_i(t)\}$. Wird die Nachricht $m = m_i$ übertragen, so ist das Sendesignal $s(t) = s_i(t)$.
- Lineare Kanalverzerrungen sind in der obigen Grafik durch die Impulsantwort $h(t)$ berücksichtigt. Außerdem ist ein (irgendwie geartetes) Rauschen $n(t)$ wirksam.
- Mit diesen beiden die Übertragung störenden Effekten lässt sich das am Empfänger ankommende Signal $r(t)$ in folgender Weise angeben:

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

- Aufgabe des (optimalen) Empfängers ist es, anhand seines Eingangssignals $r(t)$ herauszufinden, welche der M möglichen Nachrichten m_i – bzw. welches der Signale $s_i(t)$ – gesendet wurde.
- Der vom Empfänger gefundene Schätzwert für m wird in Gleichungen durch ein „Circonflexe“ (^) gekennzeichnet. Im Fließtext (HTML–Zeichensatz) ist diese Darstellung leider nicht möglich.
- Man spricht von einem optimalen Empfänger, wenn die Symbolfehlerwahrscheinlichkeit den für die Randbedingungen kleinstmöglichen Wert annimmt:

$$p_s = \Pr(\mathcal{E}) = \Pr(\hat{m} \neq m) \Rightarrow \text{Minimum}.$$

Hinweis: Im Folgenden wird meist $r(t) = s(t) + n(t)$ vorausgesetzt, was bedeutet, dass $h(t) = \delta(t)$ als verzerrungsfrei angenommen wird. Andernfalls könnten wir die Signale $s_i(t)$ als $s'_i(t) = s_i(t) * h(t)$ neu definieren, also die deterministischen Kanalverzerrungen dem Sendesignal beaufschlagen.

Fundamentaler Ansatz zum optimalen Empfängerentwurf (1)

Gegenüber dem auf der vorherigen Seite gezeigten **Blockschaltbild** führen wir nun einige wesentliche Verallgemeinerungen durch:

- Der Übertragungskanal wird durch die **bedingte Wahrscheinlichkeitsdichtefunktion** $p_{r(t)|s(t)}$ beschrieben, welche die Anhängigkeit des Empfangssignals $r(t)$ vom Sendesignal $s(t)$ festlegt.
- Wurde nun ein ganz bestimmtes Signal $r(t) = \rho(t)$ empfangen, so hat der Empfänger die Aufgabe, anhand dieses Signals $\rho(t)$ sowie der M bedingten Wahrscheinlichkeitsdichtefunktionen

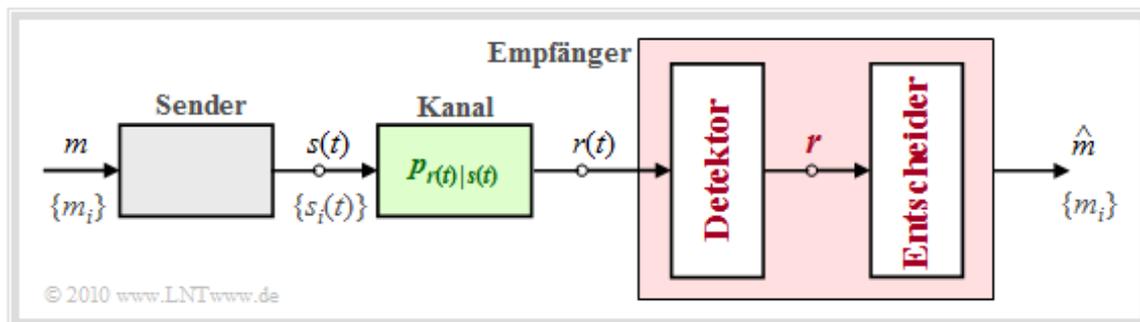
$$p_{r(t)|s(t)}(\rho(t)|s_i(t)) \quad \text{mit } i = 0, \dots, M - 1$$

unter Berücksichtigung aller möglichen Sendesignale $s_i(t)$ und deren Auftretswahrscheinlichkeiten $\Pr(m = m_i)$ herauszufinden, welche der möglichen Nachrichten (m_i) bzw. welches der möglichen Signale ($s_i(t)$) am wahrscheinlichsten gesendet wurde.

- Die **Schätzung** des optimalen Empfängers ist also ganz allgemein bestimmt durch die Gleichung

$$\hat{m} = \arg \max_i p_{s(t)|r(t)}(s_i(t)|\rho(t)) = \arg \max_i p_{m|r(t)}(m_i|\rho(t)),$$
 wobei wieder berücksichtigt ist, dass die gesendete Nachricht $m = m_i$ und das gesendete Signal $s(t) = s_i(t)$ eineindeutig ineinander übergeführt werden können.

In anderen Worten: Der optimale Empfänger betrachtet diejenige Nachricht m_i als die gesendete, wenn die bedingte Wahrscheinlichkeitsdichtefunktion $p_{m|r(t)}$ für das anliegende Empfangssignal $\rho(t)$ sowie unter der Annahme $m = m_i$ den größtmöglichen Wert annimmt.



Bevor wir die obige Entscheidungsregel näher diskutieren, soll der optimale Empfänger entsprechend der Grafik noch in zwei Funktionsblöcke aufgeteilt werden:

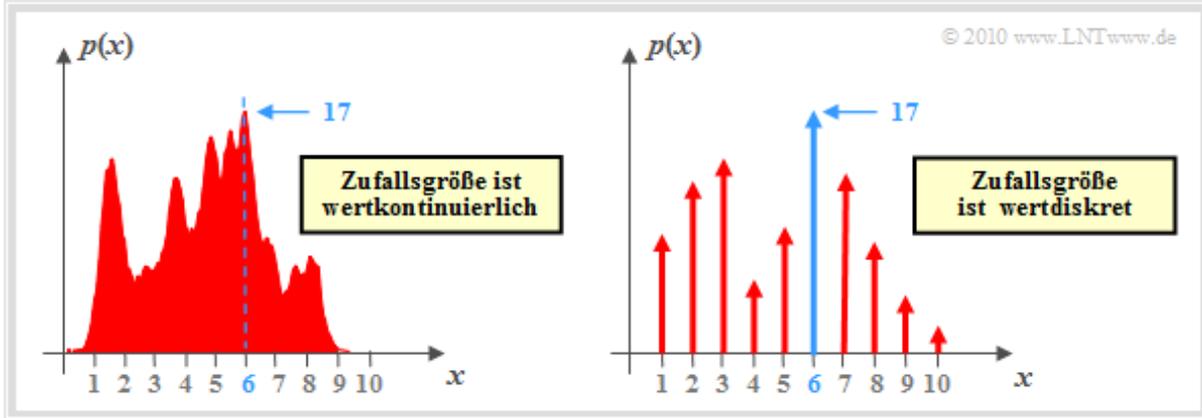
- Der **Detektor** nimmt am Empfangssignal $r(t)$ verschiedene Messungen vor und fasst diese im Vektor \mathbf{r} zusammen. Bei K Messungen entspricht \mathbf{r} einem Punkt im K -dimensionalen Vektorraum.
- Der **Entscheider** bildet abhängig von diesem Vektor den Schätzwert. Bei einem gegebenen Vektor $\mathbf{r} = \boldsymbol{\rho}$ lautet dabei die Entscheidungsregel:

$$\hat{m} = \arg \max_i P_{m|\mathbf{r}}(m_i|\boldsymbol{\rho}).$$

Im Gegensatz zur oberen Gleichung tritt nun in der Entscheidungsregel eine bedingte Wahrscheinlichkeit $P_{m|\mathbf{r}}$ anstelle der bedingten Wahrscheinlichkeitsdichtefunktion (WDF) $p_{m|r(t)}$ auf. Beachten Sie bitte die Groß- bzw. Kleinschreibung für die unterschiedlichen Bedeutungen.

Fundamentaler Ansatz zum optimalen Empfängerentwurf (2)

Wir betrachten nun die Funktion $y = \arg \max p(x)$, wobei $p(x)$ die Wahrscheinlichkeitsdichtefunktion (WDF) einer wertkontinuierlichen oder wertdiskreten Zufallsgröße x beschreibt. Im zweiten Fall besteht die WDF aus einer Summe von Diracfunktionen mit den Wahrscheinlichkeiten als Impulsgewichte.



Die Grafik zeigt beispielhafte Funktionen. In beiden Fällen liegt das WDF–Maximum (17) bei $x = 6$:

$$\max_i p(x) = 17, \quad y = \arg \max_i p(x) = 6.$$

Man nennt die (bedingten) Wahrscheinlichkeiten in der Gleichung

$$\hat{m} = \arg \max_i P_{m|r}(m_i|\rho)$$

auch **a–Posteriori–Wahrscheinlichkeiten**. Mit dem **Satz von Bayes** kann hierfür geschrieben werden:

$$P_{m|r}(m_i|\rho) = \frac{\Pr(m_i) \cdot p_{r|m}(\rho|m_i)}{p_r(\rho)}.$$

Da der Term im Nenner für alle m_i gleich ist, muss er für die Entscheidung nicht weiter berücksichtigt werden. Damit erhält man die folgenden Regeln:

Satz 1: Die **Entscheidungsregel** des optimalen Empfängers, auch bekannt als **MAP–Empfänger** (die Abkürzung steht für Maximum–a–posteriori), lautet:

$$\hat{m}_{\text{MAP}} = \arg \max_i P_{m|r}(m_i|\rho) = \arg \max_i [\Pr(m_i) \cdot p_{r|m}(\rho|m_i)].$$

Der Vorteil dieser Gleichung ist, dass die die Vorwärtsrichtung des Kanals beschreibende bedingte WDF $p_{r|m}$ („Ausgang unter der Bedingung Eingang“) verwendet werden kann. Dagegen verwendet die erste Gleichung die Rückschlusswahrscheinlichkeiten $P_{m|r}$ („Eingang unter der Bedingung Ausgang“).

Satz 2: Ein **Maximum–Likelihood–Empfänger** (ML–Empfänger) verwendet die Entscheidungsregel

$$\hat{m}_{\text{ML}} = \arg \max_i p_{r|m}(\rho|m_i).$$

Bei diesem werden die möglicherweise unterschiedlichen Auftrittswahrscheinlichkeiten $\Pr(m_i)$ für den Entscheidungsprozess nicht herangezogen, zum Beispiel, weil sie dem Empfänger nicht bekannt sind.

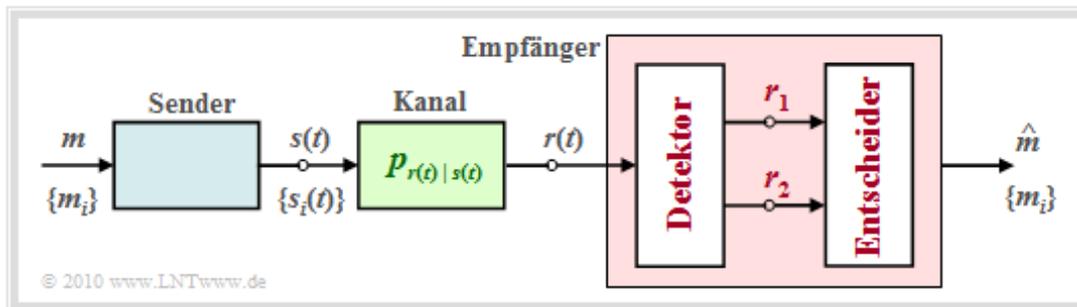
Hinweis: Im **Kapitel 3.7** finden Sie eine andere Herleitung. Allgemein gilt: Bei gleichwahrscheinlichen Nachrichten $\{m_i\} \Rightarrow \Pr(m_i) = 1/M$ ist der ML–Empfänger gleichwertig mit dem MAP–Empfänger:

$$\hat{m}_{\text{MAP}} = \hat{m}_{\text{ML}} = \arg \max_i p_{r|m}(\rho|m_i).$$

Das Theorem der Irrelevanz (1)

Zu beachten ist, dass der auf der letzten Seite beschriebene Empfänger nur dann optimal ist, wenn auch der Detektor bestmöglich implementiert ist, das heißt, wenn durch den Übergang vom kontinuierlichen Signal $r(t)$ zum Vektor \mathbf{r} keine Information verloren geht.

Um die Frage zu klären, welche und wieviele Messungen am Empfangssignal $r(t)$ durchzuführen sind, um Optimalität zu garantieren, ist das *Theorem der Irrelevanz* hilfreich. Dazu betrachten wir den nachfolgend skizzierten Empfänger, dessen Detektor aus dem Empfangssignal $r(t)$ die zwei Vektoren \mathbf{r}_1 und \mathbf{r}_2 ableitet und dem Entscheider zur Verfügung stellt. \mathbf{r}_1 und \mathbf{r}_2 stehen mit der Nachricht $m \in \{m_i\}$ über die Verbundwahrscheinlichkeitsdichte $p_{\mathbf{r}_1, \mathbf{r}_2|m}$ in Zusammenhang.



Die Entscheidungsregel des MAP–Empfängers lautet mit Anpassung an dieses Beispiel:

$$\begin{aligned} \hat{m}_{\text{MAP}} &= \arg \max_i [\Pr(m_i) \cdot p_{\mathbf{r}_1, \mathbf{r}_2|m}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2|m_i)] = \\ &= \arg \max_i [\Pr(m_i) \cdot p_{\mathbf{r}_1|m}(\boldsymbol{\rho}_1|m_i) \cdot p_{\mathbf{r}_2|\mathbf{r}_1, m}(\boldsymbol{\rho}_2|\boldsymbol{\rho}_1, m_i)]. \end{aligned}$$

Hierzu ist anzumerken:

- Die Vektoren \mathbf{r}_1 und \mathbf{r}_2 sind Zufallsgrößen. Ihre Realisierungen werden hier und im Folgenden mit $\boldsymbol{\rho}_1$ und $\boldsymbol{\rho}_2$ bezeichnet. Zur Hervorhebung sind alle Vektoren in der Grafik rot eingetragen.
- Die Voraussetzungen für die Anwendung des „Theorems der Irrelevanz“ sind die gleichen wie die an eine **Markovkette** erster Ordnung. Die Zufallsvariablen x, y, z formen dann eine solche, falls die Verteilung von z bei gegebenem y unabhängig von x ist:

$$p(x, y, z) = p(x) \cdot p(y|x) \cdot p(z|y) \quad \text{anstelle von} \quad p(x, y, z) = p(x) \cdot p(y|x) \cdot p(z|x, y).$$

- Der optimale Empfänger muss im allgemeinen Fall beide Vektoren \mathbf{r}_1 und \mathbf{r}_2 auswerten, da in obiger Entscheidungsregel beide Verbundwahrscheinlichkeitsdichten $p_{\mathbf{r}_1|m}$ und $p_{\mathbf{r}_2|\mathbf{r}_1, m}$ auftreten.
- Dagegen kann der Empfänger ohne Informationseinbuße die zweite Messung vernachlässigen, falls \mathbf{r}_2 bei gegebenem \mathbf{r}_1 unabhängig von der Nachricht m ist:

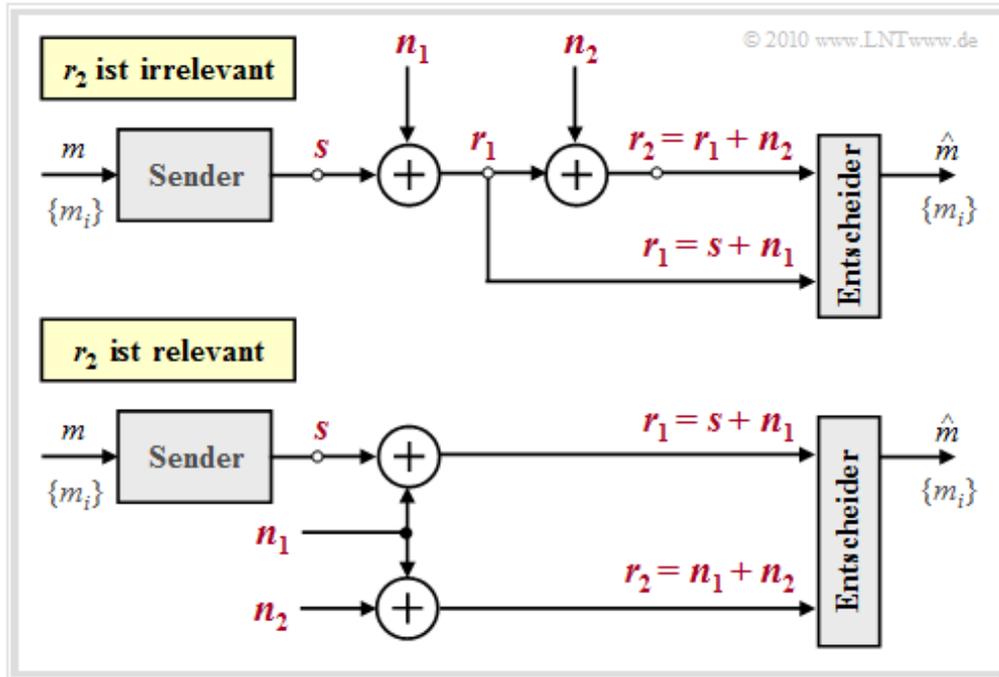
$$p_{\mathbf{r}_2|\mathbf{r}_1, m}(\boldsymbol{\rho}_2|\boldsymbol{\rho}_1, m_i) = p_{\mathbf{r}_2|\mathbf{r}_1}(\boldsymbol{\rho}_2|\boldsymbol{\rho}_1).$$

- In diesem Fall lässt sich die Entscheidungsregel weiter vereinfachen:

$$\begin{aligned} \hat{m}_{\text{MAP}} &= \arg \max_i [\Pr(m_i) \cdot p_{\mathbf{r}_1|m}(\boldsymbol{\rho}_1|m_i) \cdot p_{\mathbf{r}_2|\mathbf{r}_1, m}(\boldsymbol{\rho}_2|\boldsymbol{\rho}_1, m_i)] = \\ &= \arg \max_i [\Pr(m_i) \cdot p_{\mathbf{r}_1|m}(\boldsymbol{\rho}_1|m_i) \cdot p_{\mathbf{r}_2|\mathbf{r}_1}(\boldsymbol{\rho}_2|\boldsymbol{\rho}_1)] = \\ &= \arg \max_i [\Pr(m_i) \cdot p_{\mathbf{r}_1|m}(\boldsymbol{\rho}_1|m_i)]. \end{aligned}$$

Das Theorem der Irrelevanz (2)

Betrachten wir zur Verdeutlichung des soeben vorgestellten Theorems der Irrelevanz zwei verschiedene Systemkonfigurationen mit jeweils zwei Rauschtermen n_1 und n_2 . Anmerkung: Alle vektoriellen Größen sind rot eingezeichnet und s , n_1 und n_2 seien jeweils unabhängig voneinander.



Die Analyse dieser beiden Anordnungen liefert folgende Ergebnisse:

- Der Entscheider muss in beiden Fällen die Komponente $r_1 = s + n_1$ berücksichtigen, da nur diese die Information über das Nutzsignal s und damit über die gesendete Nachricht m liefert.
- Bei der oberen Konfiguration enthält r_2 keine Information über m , die nicht bereits von r_1 geliefert wurde. Vielmehr ist $r_2 = r_1 + n_2$ nur eine verrauschte Version von r_1 und hängt nur vom Rauschen n_2 ab, sobald r_1 bekannt ist $\Rightarrow r_2$ ist irrelevant:

$$p_{r_2|r_1, m}(\rho_2|\rho_1, m_i) = p_{r_2|r_1}(\rho_2|\rho_1) = p_{n_2}(\rho_2 - \rho_1).$$

- Bei der unteren Konfiguration ist dagegen $r_2 = n_1 + n_2$ für den Empfänger hilfreich, da dadurch dem Empfänger ein Schätzwert für den Rauschterm n_1 geliefert wird $\Rightarrow r_2$ sollte nicht verworfen werden. Formal lässt sich dieses Resultat wie folgt ausdrücken:

$$\begin{aligned} p_{r_2|r_1, m}(\rho_2|\rho_1, m_i) &= p_{r_2|n_1, m}(\rho_2|\rho_1 - s_i, m_i) = \\ &= p_{n_2|n_1, m}(\rho_2 - \rho_1 + s_i|\rho_1 - s_i, m_i) = \\ &= p_{n_2}(\rho_2 - \rho_1 + s_i). \end{aligned}$$

Da nun im Argument dieser Funktion die Nachricht (s_i) erscheint, ist r_2 nicht irrelevant.

Einige Eigenschaften des AWGN-Kanals (1)

Um weitere Aussagen über die Art der optimalen Messungen des Vektors \mathbf{r} machen zu können, ist es notwendig, die den Kanal charakterisierende (bedingte) Wahrscheinlichkeitsdichtefunktion $p_{\mathbf{r}(t)|s(t)}$ weiter zu spezifizieren. Im Folgenden wird die Kommunikation über den **AWGN-Kanal** betrachtet, dessen wichtigste Eigenschaften hier nochmals kurz zusammengestellt werden:

- Das Ausgangssignal des AWGN-Kanals ist $r(t) = s(t) + n(t)$, wobei $s(t)$ das Sendesignal angibt und $n(t)$ durch einen Gaußschen Rauschprozess dargestellt wird.
- Ein Zufallsprozess $\{n(t)\}$ ist gaußisch, falls die Elemente der k -dimensionalen Zufallsvariablen $\{n(t_1), \dots, n(t_k)\}$ gemeinsam gaußverteilt sind („*Jointly Gaussian*“).
- Der Mittelwert des AWGN-Rauschens ist $E[n(t)] = 0$. Außerdem ist $n(t)$ weiß, was bedeutet, dass das **Leistungsdichtespektrum** (LDS) für alle Frequenzen (von $-\infty$ bis $+\infty$) konstant ist:

$$\Phi_n(f) = N_0/2.$$

- Nach dem **Wiener-Chintchine-Theorem** ergibt sich die **Autokorrelationsfunktion** (AKF) als die **Fourierrücktransformierte** von $\Phi_n(f)$:

$$\varphi_n(\tau) = E[n(t) \cdot n(t + \tau)] = N_0/2 \cdot \delta(\tau)$$

$$\Rightarrow E[n(t) \cdot n(t + \tau)] = \begin{cases} \rightarrow \infty & \text{für } \tau = 0, \\ 0 & \text{für } \tau \neq 0, \end{cases}$$

- N_0 gibt dabei die physikalische (nur für $f \geq 0$ definierte) Rauschleistungsdichte an. Der konstante LDS-Wert ($N_0/2$) und das Gewicht der Diracfunktion in der AKF (ebenfalls $N_0/2$) ergibt sich allein durch die zweiseitige Betrachtungsweise.

Weitere Informationen zum AWGN-Kanal liefert das Lernvideo **Der AWGN-Kanal – Teil 2**.

Einige Eigenschaften des AWGN-Kanals (2)

Aus dem vorletzten Statement auf der letzten Seite geht hervor, dass

- reines AWGN-Rauschen $n(t)$ stets eine unendliche Varianz (Leistung) aufweist: $\sigma_n^2 \rightarrow \infty$,
- in der Realität demzufolge nur gefiltertes Rauschen $n'(t) = n(t) * h_n(t)$ auftreten kann.

Mit der Impulsantwort $h_n(t)$ und dem Frequenzgang $H_n(f) = F[h_n(t)]$ gelten dann folgende Gleichungen:

$$\begin{aligned} E[n'(t)] &= E[n(t)] = 0, \\ \Phi_{n'}(f) &= N_0/2 \cdot |H_n(f)|^2, \\ \varphi_{n'}(\tau) &= N_0/2 \cdot [h_n(\tau) \star h_n(-\tau)], \\ \sigma_n^2 &= \varphi_{n'}(\tau = 0) = N_0/2 \cdot \int_{-\infty}^{+\infty} h_n^2(t) dt = N_0/2 \cdot \langle h_n(t), h_n(t) \rangle = \\ &= \int_{-\infty}^{+\infty} \Phi_{n'}(f) df = N_0/2 \cdot \int_{-\infty}^{+\infty} |H_n(f)|^2 df. \end{aligned}$$

Im Folgenden beinhaltet $n(t)$ stets implizit eine **Bandbegrenzung**; auf die Schreibweise $n'(t)$ wird also zukünftig verzichtet.

Ähnlich wie das Sendesignal $s(t)$ lässt sich auch der Rauschprozess $n(t)$ als gewichtete Summe von orthonormalen Basisfunktionen $\varphi_j(t)$ schreiben. Im Gegensatz zu $s(t)$ ist allerdings nun eine Beschränkung auf eine endliche Anzahl an Basisfunktionen nicht möglich. Vielmehr gilt bei rein stochastischen Größen

$$n(t) = \lim_{N \rightarrow \infty} \sum_{j=1}^N n_j \cdot \varphi_j(t),$$

wobei der Koeffizient n_j durch die Projektion von $n(t)$ auf die Basisfunktion $\varphi_j(t)$ bestimmt ist:

$$n_j = \langle n(t), \varphi_j(t) \rangle .$$

Hinweis: Um eine Verwechslung mit den Basisfunktionen $\varphi_j(t)$ zu vermeiden, wird im Folgenden die AKF $\varphi_n(\tau)$ des Rauschprozesses stets nur noch als der Erwartungswert $E[n(t) \cdot n(t + \tau)]$ ausgedrückt.

Optimaler Empfänger für den AWGN-Kanal (1)

Auch das Empfangssignal $r(t) = s(t) + n(t)$ lässt sich in bekannter Weise in Basisfunktionen zerlegen:

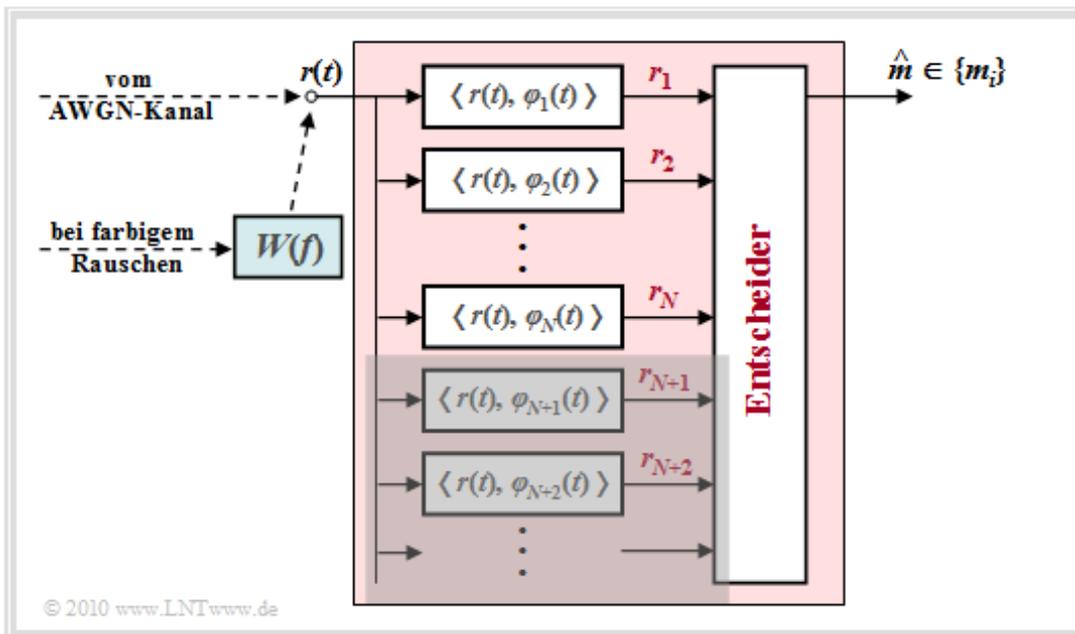
$$r(t) = \sum_{j=1}^{\infty} r_j \cdot \varphi_j(t).$$

Zu berücksichtigen ist:

- Die M möglichen Sendesignale $\{s_i(t)\}$ spannen einen Signalraum mit insgesamt N Basisfunktionen $\varphi_1(t), \dots, \varphi_N(t)$ auf.
- Diese N Basisfunktionen $\varphi_j(t)$ werden gleichzeitig zur Beschreibung des Rauschsignals $n(t)$ und des Empfangssignals $r(t)$ verwendet.
- Zur vollständigen Charakterisierung von $n(t)$ bzw. $r(t)$ werden nun aber darüber hinaus noch unendlich viele weitere Basisfunktionen $\varphi_{N+1}(t), \varphi_{N+2}(t), \dots$ benötigt.
- Damit ergeben sich die Koeffizienten des Empfangssignals $r(t)$ gemäß folgender Gleichung, wobei berücksichtigt ist, dass die Signale $s_i(t)$ und das Rauschen $n(t)$ voneinander unabhängig sind:

$$r_j = \langle r(t), \varphi_j(t) \rangle = \begin{cases} \langle s_i(t), \varphi_j(t) \rangle + \langle n(t), \varphi_j(t) \rangle = s_{ij} + n_j & j = 1, 2, \dots, N, \\ \langle n(t), \varphi_j(t) \rangle = n_j & j > N. \end{cases}$$

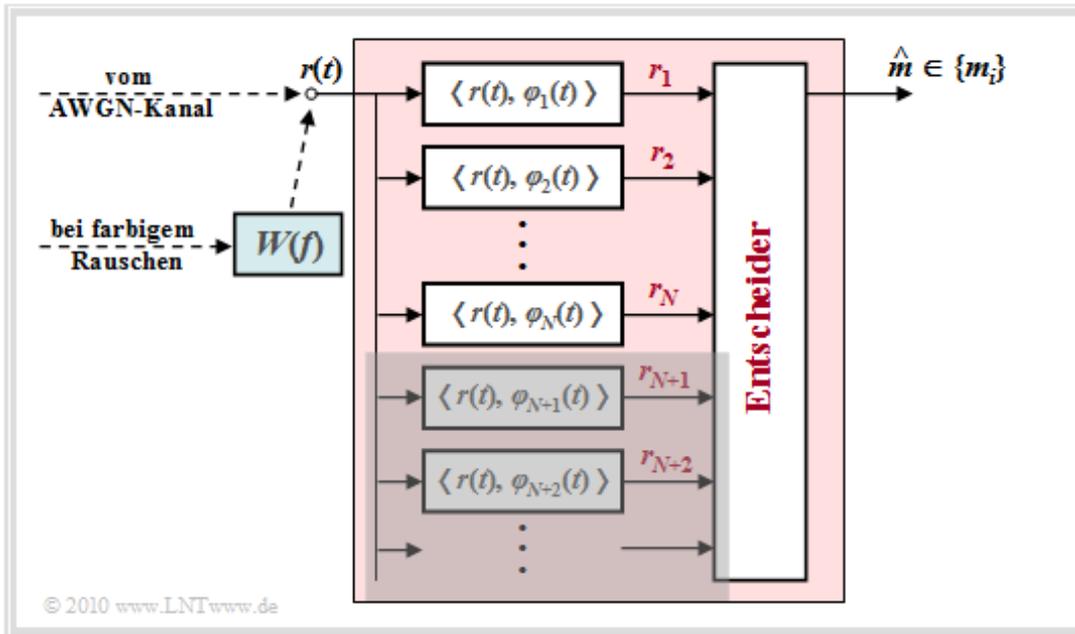
Somit ergibt sich für den optimalen Empfänger die folgende Struktur.



Die Bildbeschreibung folgt auf der nächsten Seite.

Optimaler Empfänger für den AWGN-Kanal (2)

Betrachten wir zunächst wieder den AWGN-Kanal. Hier kann auf das Vorfilter mit dem Frequenzgang $W(f)$ verzichtet werden, das in der Grafik für farbiges Rauschen vorgesehen ist.



Der Detektor des optimalen Empfängers bildet die Koeffizienten $r_j = \langle r(t), \varphi_j(t) \rangle$ und reicht diese an den Entscheider weiter. Basiert die Entscheidung auf sämtlichen – also unendlich vielen – Koeffizienten r_j , so ist die Wahrscheinlichkeit für eine Fehlentscheidung minimal und der Empfänger optimal.

Die reellwertigen Koeffizienten r_j wurden auf der letzten Seite wie folgt berechnet:

$$r_j = \begin{cases} s_{ij} + n_j & j = 1, 2, \dots, N, \\ n_j & j > N. \end{cases}$$

Nach dem **Theorem der Irrelevanz** lässt sich zeigen, dass für additives weißes Gaußsches Rauschen

- die Optimalität nicht herabgesetzt wird, wenn man die nicht von der Nachricht (s_{ij}) abhängigen Koeffizienten r_{N+1}, r_{N+2}, \dots nicht in den Entscheidungsprozess einbindet, und somit
- der Detektor nur die Projektionen des Empfangssignals $r(t)$ auf die N durch das Nutzsignal $s(t)$ vorgegebenen Basisfunktionen $\varphi_1(t), \dots, \varphi_N(t)$ bilden muss.

In der Grafik ist diese signifikante Vereinfachung durch die graue Hinterlegung angedeutet.

Im Fall von **farbigem Rauschen** \Rightarrow Leistungsdichtespektrum $\Phi_n(f) \neq \text{const.}$ ist lediglich zusätzlich ein Vorfilter mit dem Amplitudengang

$$|W(f)| = \frac{1}{\sqrt{\Phi_n(f)}}$$

erforderlich. Man nennt dieses Filter auch „*Whitening Filter*“, da die Rauschleistungsdichte am Ausgang wieder konstant, also „weiß“ ist. Genaueres hierzu finden Sie im **Kapitel 5.4** des Buches „Stochastische Signaltheorie“.

Implementierungsaspekte

Wesentliche Bestandteile des optimalen Empfängers sind die Berechnungen der inneren Produkte gemäß den Gleichungen $r_j = \langle r(t), \varphi_j(t) \rangle$. Diese können auf verschiedene Art und Weise implementiert werden:

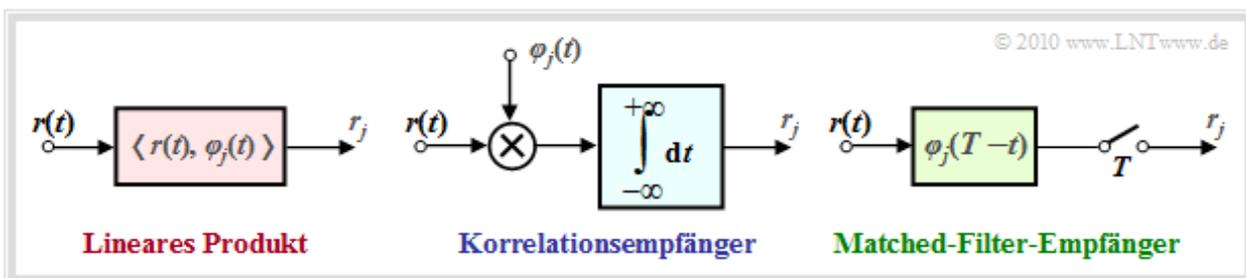
- Beim **Korrelationsempfänger** (Näheres zu dieser Implementierung finden Sie im **Kapitel 3.7**) werden die inneren Produkte direkt entsprechend der Definition mit analogen Multiplizierern und Integratoren realisiert:

$$r_j = \int_{-\infty}^{+\infty} r(t) \cdot \varphi_j(t) dt.$$

- Der **Matched-Filter-Empfänger**, der bereits im **Kapitel 1.2** dieses Buches hergeleitet wurde, erzielt mit einem linearen Filter mit der Impulsantwort $h_j(t) = \varphi_j(T - t)$ und anschließender Abtastung zum Zeitpunkt $t = T$ das gleiche Ergebnis:

$$r_j = \int_{-\infty}^{+\infty} r(\tau) \cdot h_j(t - \tau) d\tau = \int_{-\infty}^{+\infty} r(\tau) \cdot \varphi_j(T - t + \tau) d\tau$$

$$\Rightarrow r_j(t = T) = \int_{-\infty}^{+\infty} r(\tau) \cdot \varphi_j(\tau) d\tau = r_j.$$

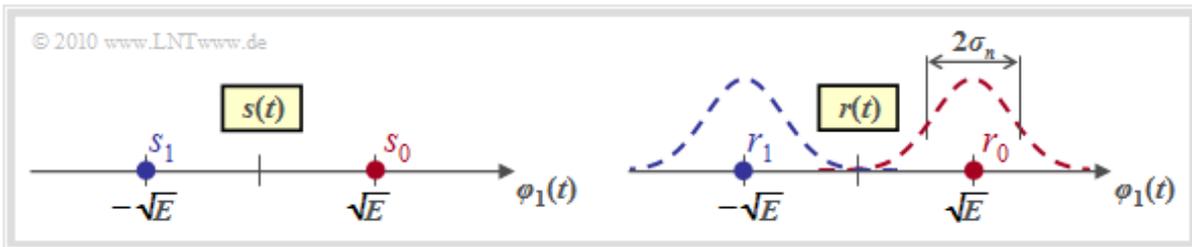


Die Abbildung zeigt die beiden möglichen Realisierungsformen des optimalen Detektors. Bevor wir uns im folgenden Kapitel der optimalen Gestaltung des Entscheiders und der Berechnung und Annäherung der Fehlerwahrscheinlichkeit zuwenden, erfolgt zunächst eine für den AWGN-Kanal gültige statistische Analyse der Entscheidungsgrößen r_j .

Wahrscheinlichkeitsdichtefunktion der Empfangswerte

Betrachten wir nochmals den optimalen Binärempfänger für die bipolare Basisbandübertragung über den AWGN-Kanal, wobei wir von der für Kapitel 4 gültigen Beschreibungsform ausgehen. Bei $N = 1$ und $M = 2$ ergibt sich für das Sendesignal die in der linken Grafik dargestellte Signalraumkonstellation

- mit nur einer Basisfunktion $\varphi_1(t)$, wegen $N = 1$,
- mit den beiden Signalraumpunkten $s_i \in \{s_0, s_1\}$, wegen $M = 2$.



Für das Signal am Ausgang des AWGN-Kanals,

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

ergibt sich im rauschfreien Fall $\Rightarrow r(t) = s(t)$ die genau gleiche Konstellation; die Signalraumpunkte liegen somit bei

$$r_0 = s_0 = \sqrt{E}, \quad r_1 = s_1 = -\sqrt{E}.$$

Bei Berücksichtigung des (bandbegrenzten) AWGN-Rauschens $n(t)$ überlagern sich den beiden Punkten r_0 und r_1 jeweils Gaußkurven mit der Varianz $\sigma_n^2 \Rightarrow$ Streuung σ_n (siehe rechte Grafik). Die WDF der Rauschkomponente n lautet dabei:

$$p_n(n) = \frac{1}{\sqrt{2\pi} \cdot \sigma_n} \cdot \exp \left[-\frac{n^2}{2\sigma_n^2} \right].$$

Für die bedingte Wahrscheinlichkeitsdichte, dass der Empfangswert ρ anliegt, wenn s_i gesendet wurde, ergibt sich dann folgender Ausdruck:

$$p_{r|s}(\rho|s_i) = \frac{1}{\sqrt{2\pi} \cdot \sigma_n} \cdot \exp \left[-\frac{(\rho - s_i)^2}{2\sigma_n^2} \right].$$

Zu den Einheiten der hier aufgeführten Größen ist zu bemerken:

- $r_0 = s_0, r_1 = s_1$ sowie n sind jeweils Skalare mit der Einheit „Wurzel aus Energie“.
- Damit ist offensichtlich, dass σ_n ebenfalls die Einheit „Wurzel aus Energie“ besitzt und σ_n^2 eine Energie darstellt.
- Beim AWGN-Kanal ist die Rauschvarianz $\sigma_n^2 = N_0/2$. Diese ist also ebenfalls eine physikalische Größe mit der Einheit $W/Hz = Ws$.

Die hier angesprochene Thematik wird in der **Aufgabe A4.6** an Beispielen verdeutlicht.

N–dimensionales Gaußsches Rauschen (1)

Liegt ein N –dimensionales Modulationsverfahren vor, das heißt, es gilt mit $0 \leq i \leq M-1$ und $1 \leq j \leq N$:

$$s_i(t) = \sum_{j=1}^N s_{ij} \cdot \varphi_j(t) = s_{i1} \cdot \varphi_1(t) + s_{i2} \cdot \varphi_2(t) + \dots + s_{iN} \cdot \varphi_N(t)$$

$$\Rightarrow \mathbf{s}_i = (s_{i1}, s_{i2}, \dots, s_{iN}),$$

so muss der Rauschvektor \mathbf{n} ebenfalls mit der Dimension N angesetzt werden, und das gleiche gilt auch für den Empfangsvektor \mathbf{r} :

$$\mathbf{n} = (n_1, n_2, \dots, n_N), \quad \mathbf{r} = (r_1, r_2, \dots, r_N).$$

Die Wahrscheinlichkeitsdichtefunktion (WDF) lautet dann für den AWGN–Kanal mit der Realisierung $\boldsymbol{\eta}$ des Rauschsignals

$$p_{\mathbf{n}}(\boldsymbol{\eta}) = \frac{1}{(\sqrt{2\pi} \cdot \sigma_n)^N} \cdot \exp \left[-\frac{\|\boldsymbol{\eta}\|^2}{2\sigma_n^2} \right],$$

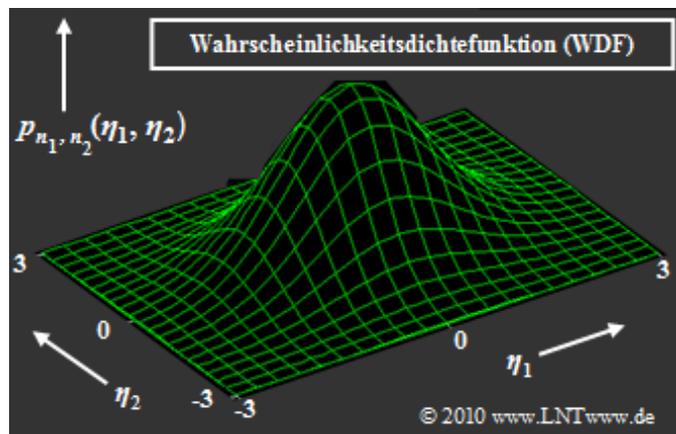
und für die bedingte WDF in der ML–Entscheidungsregel ist anzusetzen:

$$p_{\mathbf{r}|\mathbf{s}}(\boldsymbol{\rho}|\mathbf{s}_i) = p_{\mathbf{n}|\mathbf{s}}(\boldsymbol{\rho} - \mathbf{s}_i|\mathbf{s}_i) =$$

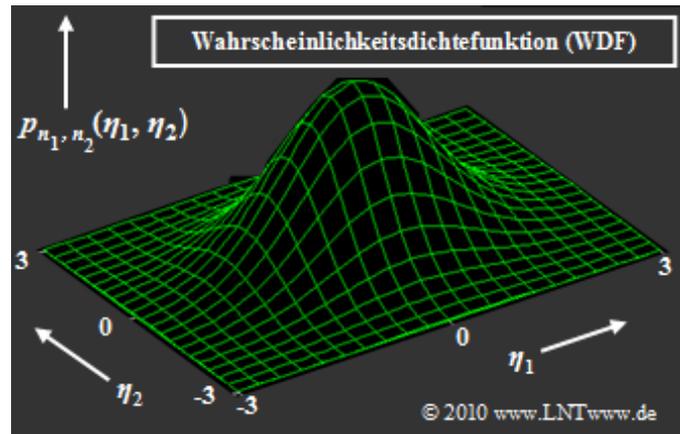
$$= \frac{1}{(\sqrt{2\pi} \cdot \sigma_n)^2} \cdot \exp \left[-\frac{\|\boldsymbol{\rho} - \mathbf{s}_i\|^2}{2\sigma_n^2} \right].$$

Die Gleichung ergibt sich aus der allgemeinen Darstellung der N –dimensionalen Gaußschen WDF in **Kapitel 4.7** des Buches „Stochastische Signaltheorie“ unter der Voraussetzung, dass die Komponenten unkorreliert (und somit statistisch unabhängig) sind. $\|\boldsymbol{\eta}\|$ bezeichnet man als die Norm (Länge) des Vektors $\boldsymbol{\eta}$.

Dargestellt ist die zweidimensionale Gauß–WDF $p_{\mathbf{n}}(\boldsymbol{\eta})$ der 2D–Zufallsgröße $\mathbf{n} = (n_1, n_2)$. Die Bildbeschreibung folgt auf der nächsten Seite.



N–dimensionales Gaußsches Rauschen (2)



Zu der dargestellten 2D–WDF ist anzumerken:

- Die Dichtefunktion bezieht sich auf die Zufallsgröße $\mathbf{n} = (n_1, n_2)$. Realisierungen hiervon werden mit $\boldsymbol{\eta} = (\eta_1, \eta_2)$ bezeichnet. Die Gleichung der dargestellten Glockenkurve lautet:

$$p_{n_1, n_2}(\eta_1, \eta_2) = \frac{1}{(\sqrt{2\pi} \cdot \sigma_n)^2} \cdot \exp \left[-\frac{\eta_1^2 + \eta_2^2}{2\sigma_n^2} \right].$$

- Das Maximum dieser Funktion liegt bei $\eta_1 = \eta_2 = 0$ und hat den Wert $2\pi \cdot \sigma_n^2$. Mit $\sigma_n^2 = N_0/2$ lässt sich die 2D–WDF in Vektorform auch wie folgt schreiben:

$$p_{\mathbf{n}}(\boldsymbol{\eta}) = \frac{1}{\pi \cdot N_0} \cdot \exp \left[-\frac{\|\boldsymbol{\eta}\|^2}{N_0} \right].$$

- Diese rotationssymmetrische WDF eignet sich zum Beispiel für die Beschreibung/Untersuchung eines **linearen zweidimensionalen** Modulationsverfahrens wie QAM, M –PSK oder 2–FSK.
- Oft werden zweidimensionale reelle Zufallsgrößen aber auch eindimensional–komplex dargestellt, meist in der Form $n(t) = n_I(t) + j \cdot n_Q(t)$. Die beiden Komponenten bezeichnet man dann als *Inphaseanteil* n_I und *Quadraturanteil* n_Q des Rauschens.
- Die Wahrscheinlichkeitsdichtefunktion hängt nur vom Betrag $|n(t)|$ der Rauschvariablen ab und nicht von Winkel $\arg n(t)$. Das heißt: Komplexes Rauschen ist zirkulär symmetrisch (siehe Grafik).
- Zirkulär symmetrisch bedeutet auch, dass Inphasekomponente n_I und Quadraturkomponente n_Q die gleiche Verteilung aufweisen und damit auch gleiche Varianz (Streuung) besitzen:

$$\begin{aligned} E[n_I^2(t)] &= E[n_Q^2(t)] = \sigma_n^2, \\ E[n(t) \cdot n^*(t)] &= E[n_I^2(t)] + E[n_Q^2(t)] = 2\sigma_n^2. \end{aligned}$$

Abschließend noch einige **Beschreibungsformen für Gaußsche Zufallsgrößen**:

$$\begin{aligned} x &= \mathcal{N}(\mu, \sigma^2): \text{ reelle gaussverteilte ZG, Mittelwert } \mu, \text{ Varianz } \sigma^2, \\ y &= \mathcal{CN}(\mu, \sigma^2): \text{ komplexe gaussverteilte ZG.} \end{aligned}$$

Optimale Entscheidung bei binärer Übertragung (1)

Wir gehen hier von einem Übertragungssystem aus, das wie folgt charakterisiert werden kann: $\mathbf{r} = \mathbf{s} + \mathbf{n}$:

- Der das Übertragungssystem vollständig beschreibende Vektorraum wird von $N = 2$ zueinander orthogonalen Basisfunktionen $\varphi_1(t)$ und $\varphi_2(t)$ aufgespannt.
- Demzufolge ist auch die Wahrscheinlichkeitsdichtefunktion des additiven und weißen Gaußschen Rauschens zweidimensional anzusetzen, gekennzeichnet durch den Vektor $\mathbf{n} = (n_1, n_2)$.
- Es gibt nur zwei mögliche Sendesignale ($M = 2$), die durch die beiden Vektoren $\mathbf{s}_0 = (s_{01}, s_{02})$ und $\mathbf{s}_1 = (s_{11}, s_{12})$ beschrieben werden:

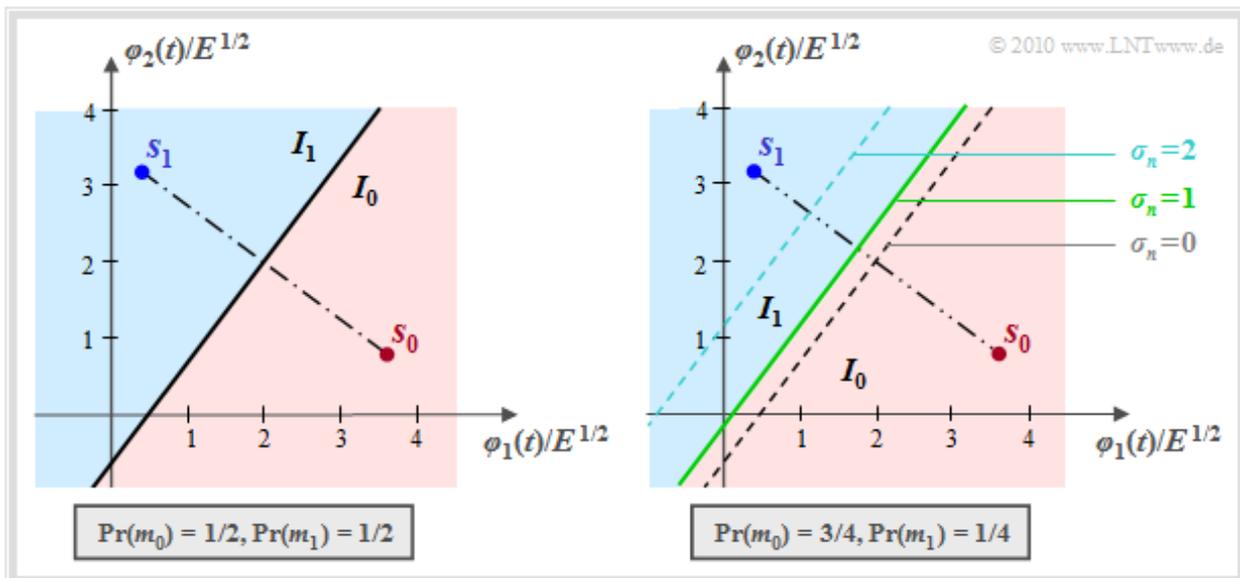
$$s_0(t) = s_{01} \cdot \varphi_1(t) + s_{02} \cdot \varphi_2(t),$$

$$s_1(t) = s_{11} \cdot \varphi_1(t) + s_{12} \cdot \varphi_2(t).$$

- Die beiden Nachrichten $m_0 \Leftrightarrow \mathbf{s}_0$ und $m_1 \Leftrightarrow \mathbf{s}_1$ sind nicht notwendigermaßen gleichwahrscheinlich.
- Aufgabe des Entscheiders ist es nun, für den gegebenen Empfangsvektor \mathbf{r} einen Schätzwert nach der **MAP-Entscheidungsregel** anzugeben. Diese lautet im vorliegenden Fall:

$$\hat{m} = \arg \max_i [\Pr(m_i) \cdot p_{\mathbf{r}|m}(\boldsymbol{\rho}|m_i)] \in \{m_i\} \quad \text{mit} \quad \mathbf{r} = \boldsymbol{\rho} = (\rho_1, \rho_2).$$

Im hier betrachteten Sonderfall $N = 2$ und $M = 2$ partitioniert der Entscheider den zweidimensionalen Raum in die zwei disjunkten Gebiete I_0 und I_1 , wie in der nachfolgenden Grafik verdeutlicht. Liegt der Empfangswert in I_0 , so wird als Schätzwert m_0 ausgegeben, andernfalls m_1 .



Die Herleitung und Bildbeschreibung folgt auf der nächsten Seite.

Optimale Entscheidung bei binärer Übertragung (2)

Beim AWGN-Kanal und $M = 2$ lautet somit die Entscheidungsregel: Man entscheide sich immer dann für die Nachricht m_0 , falls folgende Bedingung erfüllt ist:

$$\Pr(m_0) \cdot \exp \left[-\frac{1}{2\sigma_n^2} \cdot \|\boldsymbol{\rho} - \mathbf{s}_0\|^2 \right] > \Pr(m_1) \cdot \exp \left[-\frac{1}{2\sigma_n^2} \cdot \|\boldsymbol{\rho} - \mathbf{s}_1\|^2 \right].$$

Die Grenzlinie zwischen den beiden Entscheidungsregionen I_0 und I_1 erhält man, wenn man in obiger Gleichung das Größerzeichen durch das Gleichheitszeichen ersetzt und die Gleichung etwas umformt:

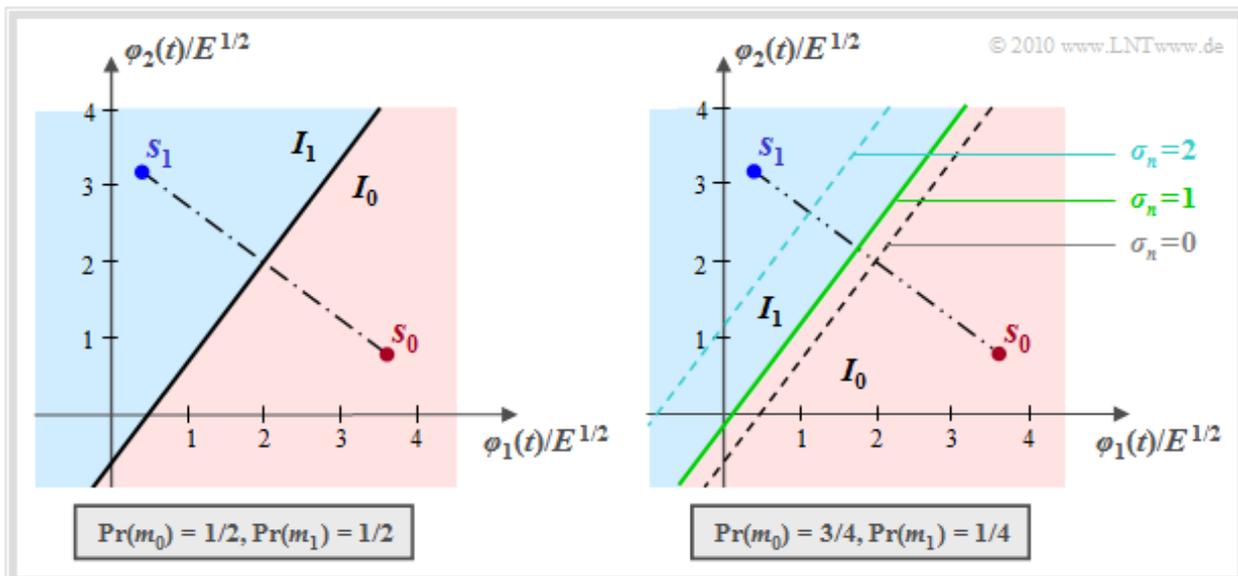
$$\begin{aligned} \|\boldsymbol{\rho} - \mathbf{s}_0\|^2 - 2\sigma_n^2 \cdot \ln [\Pr(m_0)] &= \|\boldsymbol{\rho} - \mathbf{s}_1\|^2 - 2\sigma_n^2 \cdot \ln [\Pr(m_1)] \\ \Rightarrow \|\mathbf{s}_1\|^2 - \|\mathbf{s}_0\|^2 + 2\sigma_n^2 \cdot \ln \frac{\Pr(m_0)}{\Pr(m_1)} &= 2 \cdot \boldsymbol{\rho}^T \cdot (\mathbf{s}_1 - \mathbf{s}_0). \end{aligned}$$

Aus dieser Gleichung erkennt man:

- Die Grenzkurve zwischen den Regionen I_0 und I_1 ist eine Gerade, da die Bestimmungsgleichung linear im Empfangsvektor $\boldsymbol{\rho} = (\rho_1, \rho_2)$ ist.
- Bei gleichwahrscheinlichen Symbolen verläuft die Grenze genau in der Mitte zwischen \mathbf{s}_0 und \mathbf{s}_1 und um 90° verdreht gegenüber der Verbindungslinie zwischen den Sendepunkten (linke Grafik):

$$\|\mathbf{s}_1\|^2 - \|\mathbf{s}_0\|^2 = 2 \cdot \boldsymbol{\rho}^T \cdot (\mathbf{s}_1 - \mathbf{s}_0).$$

- Für $\Pr(m_0) > \Pr(m_1)$ ist die Entscheidungsgrenze in Richtung des unwahrscheinlicheren Symbols (\mathbf{s}_1) verschoben, und zwar um so mehr, je größer die AWGN-Streuung σ_n ist.



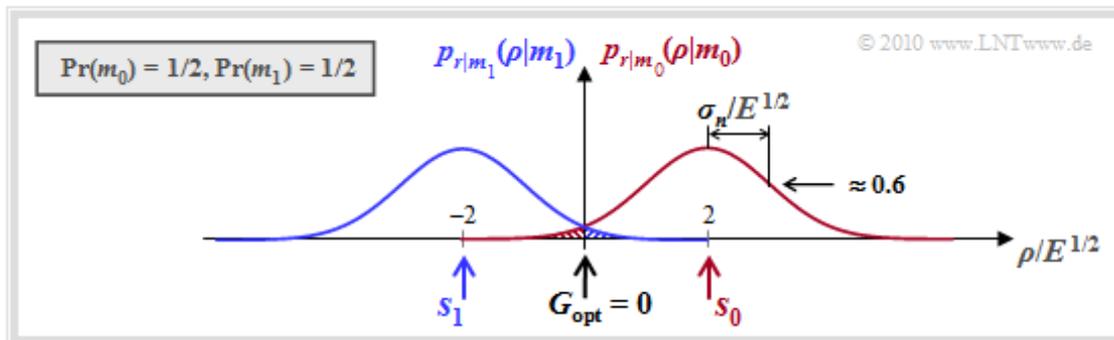
Die grün-durchgezogene Entscheidungsgrenze im rechten Bild sowie die Entscheidungsregionen I_0 (rot) und I_1 (blau) gelten für die Streuung $\sigma_n = 1$ und die gestrichelten Grenzlinien für $\sigma_n = 0$ bzw. $\sigma_n = 2$.

Gleichwahrscheinliche Binärsymbole – Fehlerwahrscheinlichkeit (1)

Wir gehen weiterhin von einem Binärsystem aus ($M = 2$), betrachten aber nun den einfachen Fall, dass dieses durch eine einzige Basisfunktion beschrieben werden kann ($N = 1$). Die Fehlerwahrscheinlichkeit hierfür wurde bereits in **Kapitel 1.2** berechnet.

Mit der für Kapitel 4 gewählten Nomenklatur und Darstellungsform ergibt sich folgende Konstellation:

- Der Empfangswert $r = s + n$ – nunmehr ein Skalar – setzt sich aus dem Sendesignal $s \in \{s_0, s_1\}$ und dem Rauschterm n zusammen. Die Abszisse ρ bezeichnet eine Realisierung von r .
- Die Abszisse ist auf die Bezugsgröße $E^{1/2}$ normiert, wobei die Normierungsenergie E keine herausgehobene physikalische Bedeutung hat.
- Der Rauschterm n ist gaußverteilt mit dem Mittelwert 0 und der Varianz σ_n^2 . Die Wurzel aus der Varianz (σ_n) wird als Effektivwert oder Streuung bezeichnet.
- Die Entscheidungsgrenze G unterteilt den gesamten Wertebereich von r in die beiden Teilbereiche I_0 (in dem unter anderem s_0 liegt) und I_1 (mit dem Signalwert s_1).
- Ist $\rho > G$, so liefert der Entscheider den Schätzwert m_0 , andernfalls m_1 . Hierbei ist vorausgesetzt, dass die Nachricht m_i mit dem Sendesignal s_i eineindeutig zusammenhängt: $m_i \Leftrightarrow s_i$.



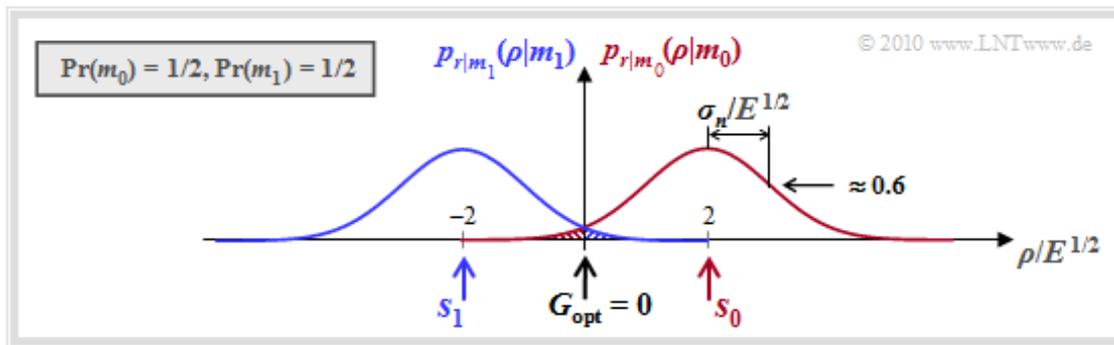
Die Grafik zeigt die bedingten (eindimensionalen) Wahrscheinlichkeitsdichtefunktionen $p_{r|m_0}$ und $p_{r|m_1}$ für den hier betrachteten AWGN-Kanal, wobei gleiche Symbolwahrscheinlichkeiten vorausgesetzt sind: $\Pr(m_0) = \Pr(m_1) = 0.5$. Dementsprechend ist die (optimale) Entscheidungsgrenze $G = 0$.

Man erkennt aus dieser Darstellung:

- Ist $m = m_0$ und damit $s = s_0 = 2 \cdot E^{1/2}$, so kommt es nur dann zu einer Fehlentscheidung, wenn η , die Realisierung der Rauschgröße n , kleiner ist als $-2 \cdot E^{1/2}$.
- In diesem Fall ist $\rho < 0$, wobei ρ eine Realisierung des Empfangswertes r bezeichnet.

Die Bildbeschreibung wird auf der nächsten Seite fortgesetzt.

Gleichwahrscheinliche Binärsymbole – Fehlerwahrscheinlichkeit (2)



Kommen wir nun zur Berechnung der Fehlerwahrscheinlichkeit:

- Bei AWGN-Rauschen mit dem Effektivwert (Streuung) σ_n erhält man in diesem Fall, wie bereits in **Kapitel 1.2** mit anderer Nomenklatur berechnet wurde:

$$\begin{aligned} \Pr(\mathcal{E}|m_0) &= \int_{-\infty}^{G=0} p_{r|m_0}(\rho|m_0) d\rho = \int_{-\infty}^{-s_0} p_n(\eta|m_0) d\eta = \\ &= \int_{-\infty}^{-s_0} p_n(\eta) d\eta = \int_{s_0}^{\infty} p_n(\eta) d\eta = Q(s_0/\sigma_n). \end{aligned}$$

- Bei der Herleitung der Gleichung wurde berücksichtigt, dass das AWGN-Rauschen η unabhängig vom Signal (m_0 oder m_1) ist und eine symmetrische WDF besitzt. Verwendet wurde zudem das komplementäre Gaußsche Fehlerintegral

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

- Entsprechend gilt für $m = m_1 \Leftrightarrow s = s_1 = -2 \cdot E^{1/2}$:

$$\Pr(\mathcal{E}|m_1) = \int_0^{\infty} p_{r|m_1}(\rho|m_1) d\rho = \int_{-s_1}^{\infty} p_n(\eta) d\eta = Q(-s_1/\sigma_n).$$

- Mit dem Abstand $d = s_1 - s_0$ der zwei Signalraumpunkte lassen sich die beiden Ergebnisse zusammenfassen, wobei noch $\Pr(m_0) + \Pr(m_1) = 1$ zu berücksichtigen ist:

$$\begin{aligned} \Pr(\mathcal{E}|m_0) &= \Pr(\mathcal{E}|m_1) = Q(d/(2\sigma_n)) \\ \Rightarrow \Pr(\mathcal{E}) &= \Pr(m_0) \cdot \Pr(\mathcal{E}|m_0) + \Pr(m_1) \cdot \Pr(\mathcal{E}|m_1) = \\ &= [\Pr(m_0) + \Pr(m_1)] \cdot Q(d/(2\sigma_n)) = Q(d/(2\sigma_n)). \end{aligned}$$

Diese Gleichung gilt unter der Voraussetzung $G = 0$ ganz allgemein, also auch für $\Pr(m_0) \neq \Pr(m_1)$. Bei nicht gleichwahrscheinlichen Symbolen lässt sich allerdings die Symbolfehlerwahrscheinlichkeit durch eine andere Entscheidungsgrenze verkleinern.

Hinweis: Die hier genannte Gleichung gilt auch dann, wenn die Signalraumpunkte keine Skalare sind, sondern durch die Vektoren \mathbf{s}_0 und \mathbf{s}_1 beschrieben werden. Der Abstand d ergibt sich dann als die Norm des Differenzvektors:

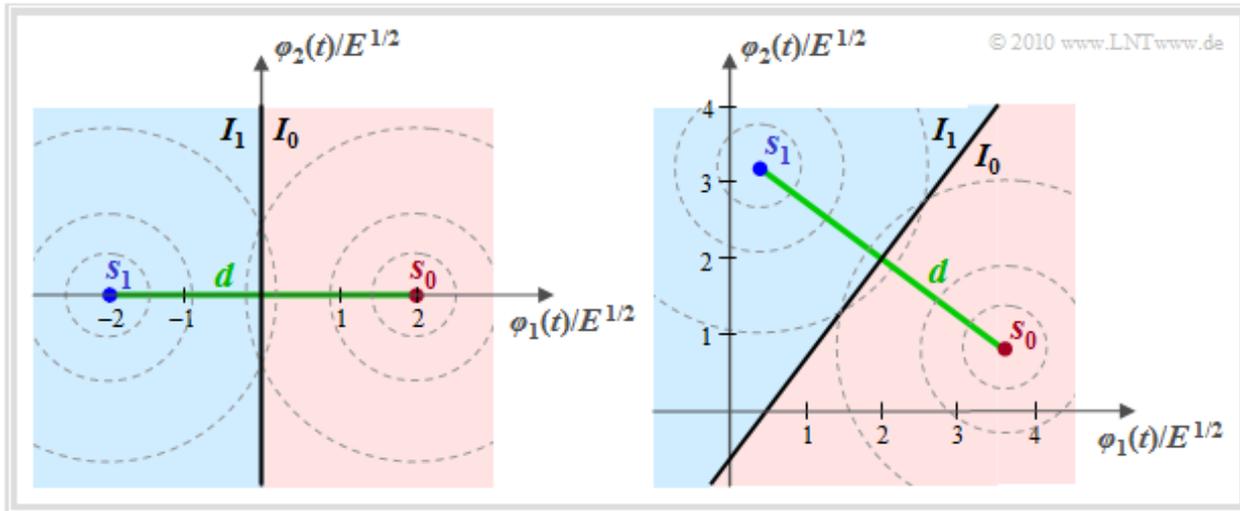
$$d = \|\mathbf{s}_1 - \mathbf{s}_0\|.$$

Gleichwahrscheinliche Binärsymbole – Fehlerwahrscheinlichkeit (3)

Betrachten wir nun nochmals die Signalraumkonstellation von der **ersten Seite** dieses Kapitels mit den Werten $\mathbf{s}_0/E^{1/2} = (3.6, 0.8)$ und $\mathbf{s}_1/E^{1/2} = (0.4, 3.2)$. Hier beträgt der Abstand der Signalraumpunkte

$$d = \|\mathbf{s}_1 - \mathbf{s}_0\| = \sqrt{E \cdot (0.4 - 3.6)^2 + E \cdot (3.2 - 0.8)^2} = 4 \cdot \sqrt{E},$$

also der genau gleiche Wert wie für $\mathbf{s}_0/E^{1/2} = (2, 0)$ und $\mathbf{s}_1/E^{1/2} = (-2, 0)$. Die AWGN-Rauschvarianz beträgt jeweils $\sigma_n^2 = N_0/2$.



Die Abbildungen zeigen diese beiden Konstellationen und lassen folgende Gemeinsamkeiten bzw. Unterschiede erkennen:

- Wie bereits gesagt, sind sowohl der Abstand der Signalpunkte von der Entscheidungsgeraden ($d/2 = 2 \cdot E^{1/2}$) als auch der AWGN-Kennwert σ_n in beiden Fällen gleich.
- Daraus folgt: Die beiden Anordnungen führen zur gleichen Fehlerwahrscheinlichkeit, wenn man den Parameter E (eine Art Normierungsenergie) konstant lässt:

$$\Pr(\text{Symbolfehler}) = \Pr(\mathcal{E}) = Q(d/(2\sigma_n)) .$$
- Bei gegebener *mittlerer Energie pro Symbol* (E_s) ist jedoch die linke Konstellation ($E_s = 4 \cdot E$) der rechten ($E_s = 24 \cdot E$) deutlich überlegen: Die gleiche Fehlerwahrscheinlichkeit ergibt sich mit weniger Energie.

Auf diesen Sachverhalt wird in der **Aufgabe Z4.6** noch im Detail eingegangen. Die Kreise in obiger Grafik veranschaulichen die zirkuläre Symmetrie von 2D-AWGN-Rauschen.

Nicht gleichwahrscheinliche Binärsymbole – Schwellenoptimierung (1)

Gilt $\Pr(m_0) \neq \Pr(m_1)$, so kann man durch eine Verschiebung der Entscheidungsgrenze G eine etwas kleinere Fehlerwahrscheinlichkeit erreichen. Die nachfolgenden Ergebnisse werden ausführlich in der Musterlösung zur **Aufgabe A4.7** hergeleitet:

- Bei ungleichen Symbolwahrscheinlichkeiten liegt die optimale Entscheidungsgrenze G_{opt} zwischen den Regionen I_0 und I_1 näher beim unwahrscheinlicheren Symbol.
- Die normierte optimale Verschiebung gegenüber der Grenze $G = 0$ bei gleichwahrscheinlichen Symbolen beträgt

$$\gamma_{\text{opt}} = \frac{G_{\text{opt}}}{s_0} = 2 \cdot \frac{\sigma_n^2}{d^2} \cdot \ln \frac{\Pr(m_1)}{\Pr(m_0)}.$$

- Die Fehlerwahrscheinlichkeit ist dann gleich

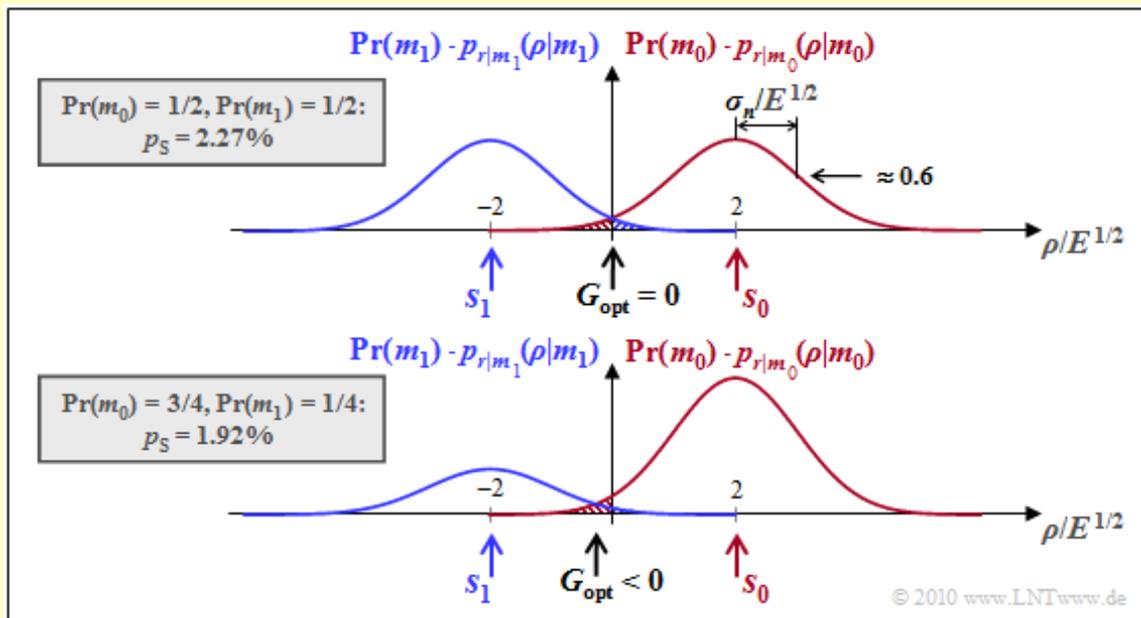
$$\Pr(\mathcal{E}) = \Pr(m_0) \cdot Q[d/(2\sigma_n) \cdot (1 - \gamma_{\text{opt}})] + \Pr(m_1) \cdot Q[d/(2\sigma_n) \cdot (1 + \gamma_{\text{opt}})].$$

Beispiel: Für das Folgende gelte

$$\mathbf{s}_0 = (2 \cdot \sqrt{E}, 0), \quad \mathbf{s}_1 = (-2 \cdot \sqrt{E}, 0), \quad \Rightarrow \quad d = 2 \cdot \sqrt{E}, \quad \sigma_n = \sqrt{E}.$$

Bei gleichwahrscheinlichen Symbolen ergibt sich die optimale Entscheidungsgrenze zu $G_{\text{opt}} = 0$. Damit erhält man für die Fehlerwahrscheinlichkeit:

$$\Pr(\mathcal{E}) = Q(d/(2\sigma_n)) = Q(2) \approx 2.26\%.$$



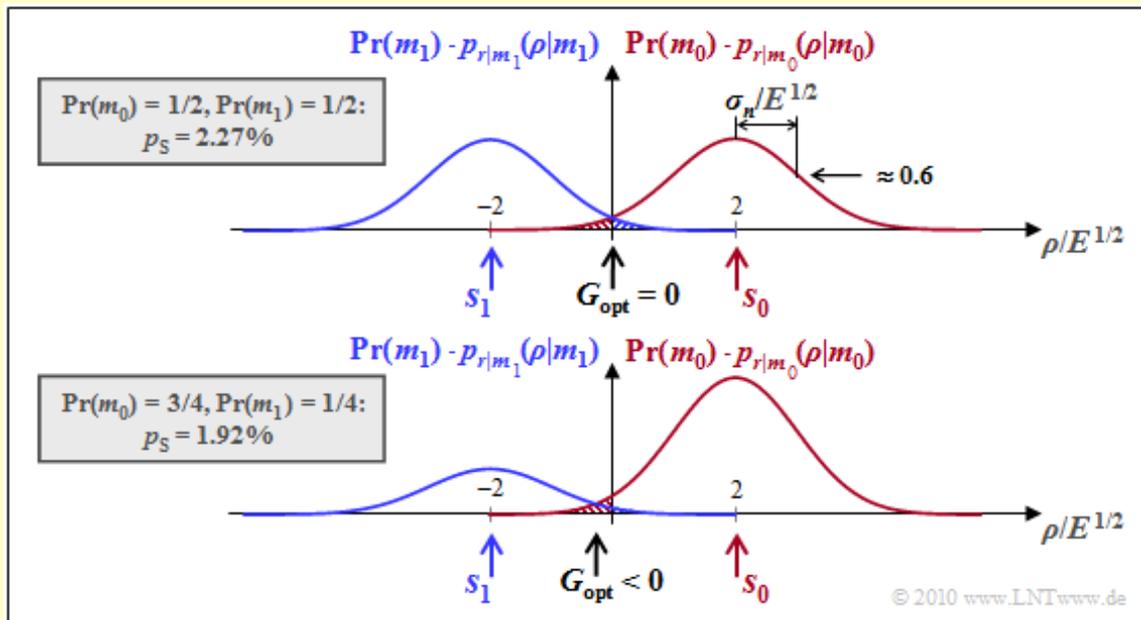
Die Beschreibung der unteren Grafik folgt auf der nächsten Seite.

Nicht gleichwahrscheinliche Binärsymbole – Schwellenoptimierung (2)

Fortsetzung des Beispiels:

Wir betrachten nun ungleiche Symbolwahrscheinlichkeiten, wie für das untere Bild vorausgesetzt:

$$\Pr(m_0) = 3/4, \quad \Pr(m_1) = 1/4.$$



Die weiteren Systemgrößen seien gegenüber der oberen Grafik unverändert:

$$\mathbf{s}_0 = (2 \cdot \sqrt{E}, 0), \quad \mathbf{s}_1 = (-2 \cdot \sqrt{E}, 0), \quad \Rightarrow \quad d = 2 \cdot \sqrt{E}, \quad \sigma_n = \sqrt{E}.$$

In diesem Fall beträgt der optimale (normierte) Verschiebungsfaktor

$$\gamma = 2 \cdot \frac{\sigma_n^2}{d^2} \cdot \ln \frac{\Pr(m_1)}{\Pr(m_0)} = 2 \cdot \frac{E}{16 \cdot E} \cdot \ln \frac{1/4}{3/4} \approx -0.14,$$

was einer Verschiebung um 14% hin zum unwahrscheinlicheren Symbol \mathbf{s}_1 (also nach links) bedeutet. Dadurch wird die Fehlerwahrscheinlichkeit geringfügig kleiner als bei gleichwahrscheinlichen Symbolen:

$$\begin{aligned} \Pr(\mathcal{E}) &= 0.75 \cdot Q(2 \cdot 1.14) + 0.25 \cdot Q(2 \cdot 0.86) = \\ &= 0.75 \cdot 0.0113 + 0.25 \cdot 0.0427 \approx 1.92\%. \end{aligned}$$

Man erkennt aus diesen Zahlenwerten: Durch die Schwellenverschiebung wird nun zwar das Symbol \mathbf{s}_1 stärker verfälscht, das wahrscheinlichere Symbol \mathbf{s}_0 jedoch überproportional weniger.

Das Ergebnis sollte nicht zu Fehlinterpretationen führen. Im unsymmetrischen Fall $\Rightarrow \Pr(m_0) \neq \Pr(m_1)$ ergibt sich zwar eine kleinere Fehlerwahrscheinlichkeit als für $\Pr(m_0) = \Pr(m_1) = 0.5$, aber mit jedem Symbol kann auch nur weniger Information übertragen werden, bei den gewählten Zahlenwerten 0.81 bit/Symbol statt 1 bit/Symbol. Aus informationstheoretischer Sicht ist $\Pr(m_0) = \Pr(m_1)$ optimal.

Anmerkung: Bei $\Pr(m_0) \neq \Pr(m_1)$ müssen nun die absoluten Wahrscheinlichkeitsdichtefunktionen $\Pr(m_i) \cdot p_{r|m_i}(\rho | m_i)$ betrachtet werden. Der formale Parameter ρ gibt dabei wieder eine Realisierung der AWGN-Zufallsgröße $r = s + n$ an. Im Folgenden wird dieser Sachverhalt berücksichtigt.

Entscheidungsregionen im nichtbinären Fall ($M > 2$)

Allgemein partitionieren die Entscheidungsregionen I_i den N -dimensionalen reellen Raum in M zueinander disjunkte Gebiete. I_i ist definiert als die Menge aller Punkte, die zum Schätzwert m_i führen:

$$\rho \in I_i \iff \hat{m} = m_i, \text{ wobei}$$

$$I_i = \{ \rho \in \mathcal{R}^N \mid \Pr(m_i) \cdot p_{\mathbf{r} | m}(\rho | m_i) > \Pr(m_k) \cdot p_{\mathbf{r} | m}(\rho | m_k) \quad \forall k \neq i \}.$$

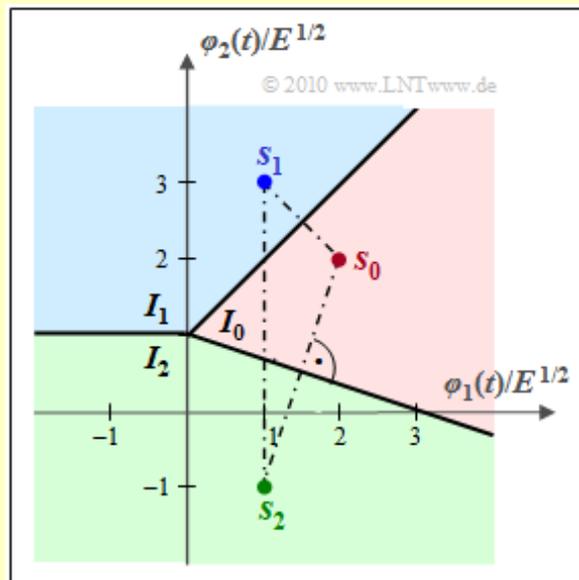
Die Form der Entscheidungsregionen I_i ($i = 0, \dots, M - 1$) im N -dimensionalen Raum hängen von den bedingten Wahrscheinlichkeitsdichtefunktionen $p_{\mathbf{r} | m}$ ab, also vom betrachteten Kanal. In vielen Fällen – so auch beim AWGN-Kanal – sind die Entscheidungsgrenzen zwischen je zwei Signalpunkten Gerade, was die weiteren Betrachtungen deutlich vereinfacht.

Beispiel: Die Grafik zeigt die Entscheidungsregionen I_0, I_1 und I_2 für ein Übertragungssystem mit den Parametern $N = 2$ und $M = 3$. Die normierten Sendevektoren sind dabei

$$\mathbf{s}_0 = (2, 2), \quad \mathbf{s}_1 = (1, 3), \quad \mathbf{s}_2 = (1, -1).$$

Es sind nun zwei Fälle zu unterscheiden:

- Bei gleichen Symbolwahrscheinlichkeiten, $\Pr(m_0) = \Pr(m_1) = \Pr(m_2) = 1/3$, verlaufen die Grenzen zwischen jeweils zwei Regionen stets geradlinig, mittig und rechtwinklig zu den Verbindungsgeraden.
- Bei ungleichen Symbolwahrscheinlichkeiten sind die Entscheidungsgrenzen dagegen jeweils in Richtung des unwahrscheinlicheren Symbols (parallel) zu verschieben, und zwar umso weiter, je größer die AWGN-Streuung σ_n ist.



Nachdem die Entscheidungsregionen I_i festliegen, kann man die Symbolfehlerwahrscheinlichkeit des Gesamtsystems berechnen. Auf den nächsten Seiten benutzen wir folgende Bezeichnungen, wobei wir aufgrund der Einschränkungen durch den verwendeten HTML-Zeichensatz im Fließtext manchmal andere Namen als in Gleichungen verwenden müssen:

- **Symbolfehlerwahrscheinlichkeit:**

$$\Pr(\mathcal{E}) = \Pr(\text{Symbolfehler}),$$

- Wahrscheinlichkeit für **korrekte Entscheidung:**

$$\Pr(\mathcal{C}) = 1 - \Pr(\mathcal{E}) = \Pr(\text{korrekte Entscheidung}),$$

- **Bedingte Wahrscheinlichkeit** einer korrekten Entscheidung unter der Bedingung $m = m_i$:

$$\Pr(\mathcal{C} | m_i) = 1 - \Pr(\mathcal{E} | m_i).$$

Fehlerwahrscheinlichkeitsberechnung im nichtbinären Fall (1)

Mit den Definitionen der letzten Seite gilt für die Wahrscheinlichkeit einer korrekten Entscheidung:

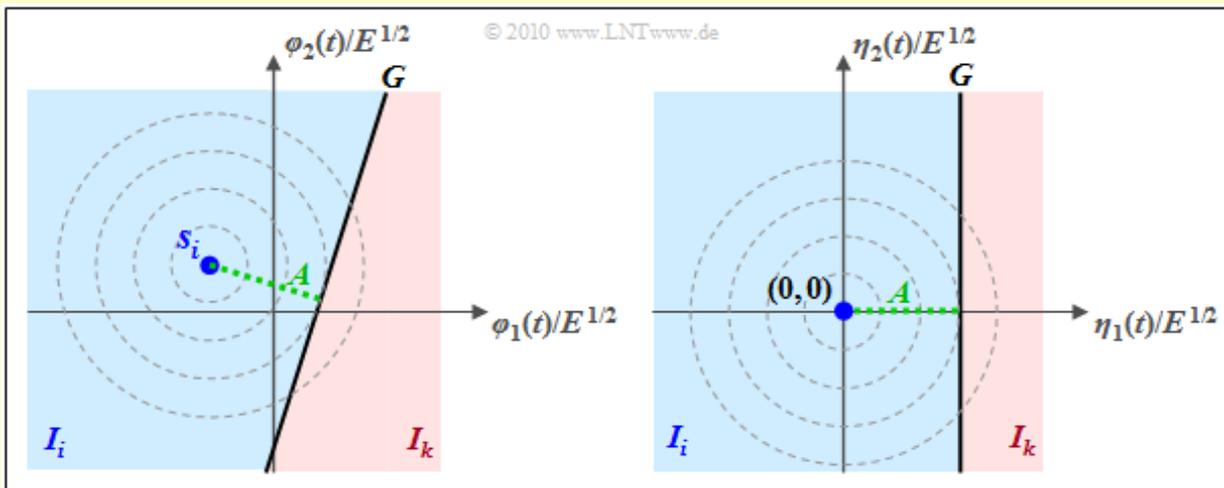
$$\begin{aligned} \Pr(\mathcal{C}) &= \sum_{i=0}^{M-1} \Pr(m_i) \cdot \Pr(\mathcal{C} | m_i) = \sum_{i=0}^{M-1} \Pr(m_i) \cdot \Pr(\mathbf{r} \in I_i | m_i) = \\ &= \sum_{i=0}^{M-1} \Pr(m_i) \cdot \int_{I_i} p_{\mathbf{r}|m}(\boldsymbol{\rho} | m_i) d\boldsymbol{\rho}. \end{aligned}$$

Für den AWGN-Kanal gilt dabei entsprechend **Kapitel 4.2**:

$$\Pr(\mathcal{C} | m_i) = 1 - \Pr(\mathcal{E} | m_i) = \frac{1}{(\sqrt{2\pi} \cdot \sigma_n)^N} \cdot \int_{I_i} \exp \left[-\frac{1}{2\sigma_n^2} \cdot \|\boldsymbol{\rho} - \mathbf{s}_i\|^2 \right] d\boldsymbol{\rho}.$$

Dieses Integral muss im allgemeinen Fall numerisch berechnet werden. Nur bei einigen wenigen, einfach beschreibbaren Entscheidungsregionen $\{I_i\}$ ist eine analytische Lösung möglich.

Beispiel: Beim AWGN-Kanal liegt eine 2D-Gaußglocke um den Sendepunkt \mathbf{s}_i , in der linken Grafik erkennbar an den konzentrischen Höhenlinien. Etwas willkürlich ist zudem die Entscheidungsgerade G eingezeichnet. Rechts dargestellt ist in einem anderen Koordinatensystem (verschoben und gedreht) allein die WDF der Rauschkomponente.



Die Grafik lässt sich wie folgt interpretieren:

- Die Wahrscheinlichkeit, dass der Empfangsvektor nicht in das Gebiet I_i fällt, sondern in das rot hinterlegte Gebiet I_k , ist $Q(A/\sigma_n)$. A ist der Abstand zwischen \mathbf{s}_i und G und σ_n der Effektivwert (Wurzel aus der Varianz) des AWGN-Rauschens. $Q(x)$ ist die Gaußsche Fehlerfunktion.
- Entsprechend ist die Wahrscheinlichkeit für das Ereignis $r \in I_i$ gleich dem Komplementärwert

$$\Pr(\mathcal{C} | m_i) = \Pr(\mathbf{r} \in I_i | m_i) = 1 - Q(A/\sigma_n).$$

Fehlerwahrscheinlichkeitsberechnung im nichtbinären Fall (2)

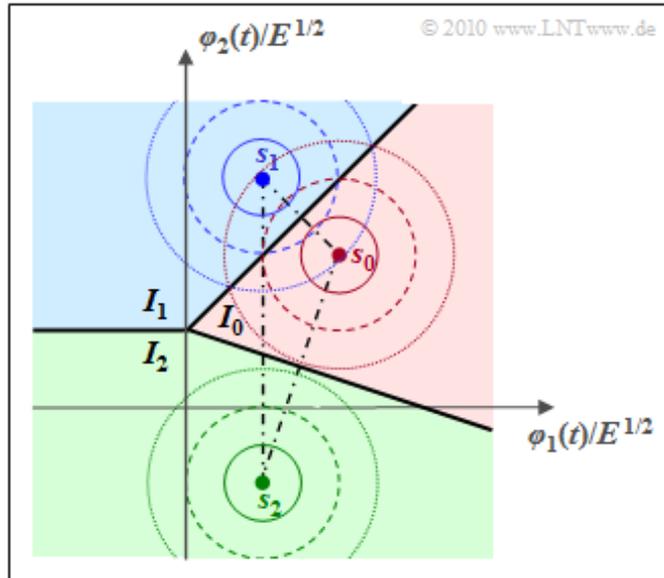
Wir betrachten nun die auf der letzten Seite angegebenen Gleichungen

$$\Pr(\mathcal{C}) = \sum_{i=0}^{M-1} \Pr(m_i) \cdot \Pr(\mathcal{C} | m_i) \quad \text{mit} \quad \Pr(\mathcal{C} | m_i) = \int_{I_i} p_{\mathbf{r}|m}(\boldsymbol{\rho} | m_i) d\boldsymbol{\rho}$$

etwas genauer, wobei wir wieder von zwei Basisfunktionen ($N = 2$) und den drei Signalraumpunkten \mathbf{s}_0 , \mathbf{s}_1 und \mathbf{s}_2 (also $M = 3$) ausgehen. Die Entscheidungsregionen I_0 , I_1 und I_2 sind **bestmöglich** gewählt. Das AWGN-Rauschen ist in der Skizze durch jeweils drei kreisförmige Höhenlinien angedeutet.

Man erkennt aus dieser Darstellung:

- Unter der Voraussetzung, dass $m = m_i \Leftrightarrow \mathbf{s} = \mathbf{s}_i$ gesendet wurde, wird nur dann eine richtige Entscheidung getroffen, wenn der Empfangswert \mathbf{r} in der Region I_i liegt.
- Die Wahrscheinlichkeit $\Pr(\mathbf{r} \in I_i | m_2)$ für eine ist für $i = 2$ (weitaus) am größten \Rightarrow richtige Entscheidung. $\Pr(\mathbf{r} \in I_0 | m_2)$ ist deutlich kleiner. Nahezu vernachlässigbar ist $\Pr(\mathbf{r} \in I_1 | m_2)$.



- Die Verfälschungswahrscheinlichkeiten für $m = m_0$ bzw. $m = m_1$ lauten:

$$\begin{aligned} \Pr(\mathcal{E} | m_0) &= \Pr(\mathbf{r} \in I_1 | m_0) + \Pr(\mathbf{r} \in I_2 | m_0), \\ \Pr(\mathcal{E} | m_1) &= \Pr(\mathbf{r} \in I_0 | m_1) + \Pr(\mathbf{r} \in I_2 | m_1). \end{aligned}$$

- Die größte Verfälschungswahrscheinlichkeit ergibt sich für $m = m_0$. Wegen

$$\Pr(\mathbf{r} \in I_1 | m_0) \approx \Pr(\mathbf{r} \in I_0 | m_1), \quad \Pr(\mathbf{r} \in I_2 | m_0) \gg \Pr(\mathbf{r} \in I_2 | m_1)$$

gelten folgende Relationen:

$$\Pr(\mathcal{E} | m_0) > \Pr(\mathcal{E} | m_1) > \Pr(\mathcal{E} | m_2).$$

Diese Ergebnisse lassen sich wie folgt zusammenfassen:

- Zur Berechnung der (mittleren) Fehlerwahrscheinlichkeit muss auch bei gleichwahrscheinlichen Symbolen allgemein über alle M Terme gemittelt werden. Ausnahme: Symmetrische Anordnung.
- Im Fall gleichwahrscheinlicher Symbole kann $\Pr(m_i) = 1/M$ vor die Summation gezogen werden, was allerdings den Rechengang nicht sonderlich vereinfacht.

Union Bound - Obere Schranke für die Fehlerwahrscheinlichkeit (1)

Bei beliebigen Werten von M gilt für die Verfälschungswahrscheinlichkeit unter der Voraussetzung, dass die Nachricht m_i (bzw. das Signal \mathbf{s}_i) gesendet wurde:

$$\Pr(\mathcal{E} | m_i) = \Pr \left[\bigcup_{k \neq i} \mathcal{E}_{ik} \right], \quad \mathcal{E}_{ik} : \mathbf{r} \text{ liegt näher bei } \mathbf{s}_k \text{ als beim Sollwert } \mathbf{s}_i.$$

Für diesen Ausdruck lässt sich mit einer Booleschen Ungleichung, der so genannten **Union Bound**, eine obere Schranke angeben:

$$\Pr(\mathcal{E} | m_i) \leq \sum_{k=0, k \neq i}^{M-1} \Pr(\mathcal{E}_{ik}) = \sum_{k=0, k \neq i}^{M-1} Q [d_{ik}/(2\sigma_n)].$$

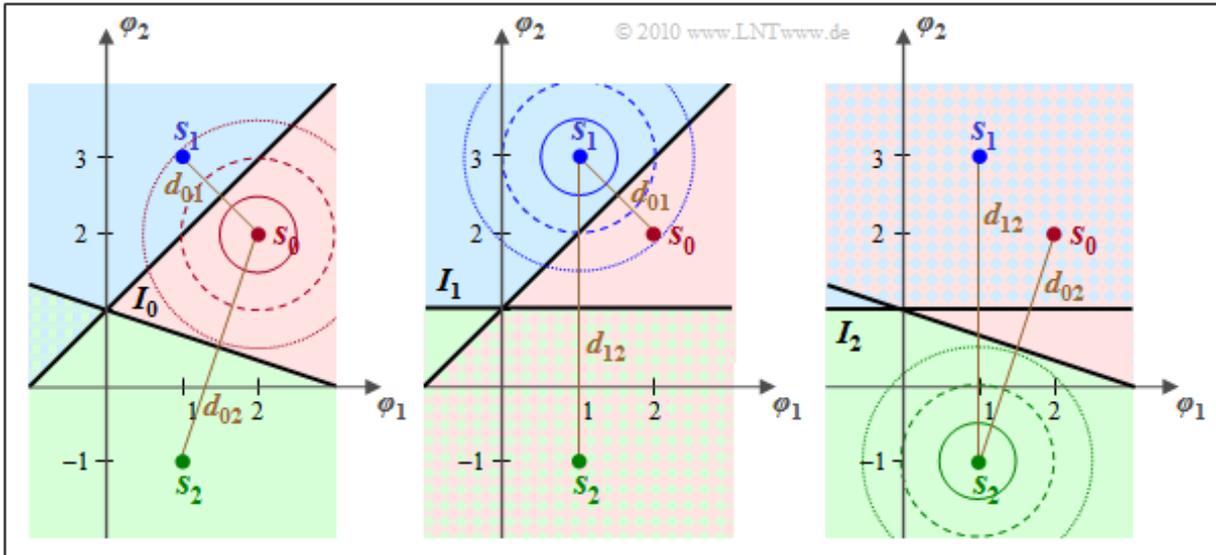
Hierzu ist anzumerken:

- Es ist $d_{ik} = \|\mathbf{s}_i - \mathbf{s}_k\|$ der Abstand der Signalraumpunkte \mathbf{s}_i und \mathbf{s}_k , und σ_n gibt den Effektivwert des AWGN-Rauschens an.
- Die Union Bound ist nur bei gleichwahrscheinlichen Symbolen ($\Pr(m_i) = 1/M$) anwendbar. Auch dann muss zur Berechnung der (mittleren) Fehlerwahrscheinlichkeit über alle m_i gemittelt werden.

Auf der nächsten Seite folgt ein Beispiel zur Anwendung der Union Bound. Auf der **übernächsten Seite** wird die hier angegebene *Union Bound* weiter vereinfacht.

Union Bound - Obere Schranke für die Fehlerwahrscheinlichkeit (2)

Beispiel: Die Grafik verdeutlicht die **Union Bound** am Beispiel $M = 3$ mit gleichwahrscheinlichen Symbolen: $\Pr(m_0) = \Pr(m_1) = \Pr(m_2) = 1/3$.



Zu diesen Darstellungen ist anzumerken:

- Für die Symbolfehlerwahrscheinlichkeit gilt:

$$\Pr(\mathcal{E}) = 1 - \Pr(\mathcal{C}), \quad \Pr(\mathcal{C}) = 1/3 \cdot [\Pr(\mathcal{C} | m_0) + \Pr(\mathcal{C} | m_1) + \Pr(\mathcal{C} | m_2)] .$$

- Der erste Term im Klammerausdruck unter der Voraussetzung $m = m_0 \Leftrightarrow \mathbf{s} = \mathbf{s}_0$ ist in der linken Grafik visualisiert. Dieser Term $\Pr(\mathbf{r} \in I_0 | m_0)$ beschreibt die rot ausgefüllte Region I_0 .
- Die Komplementärregion „ $\mathbf{r} \notin I_0 | m_0$ “ ist in der linken Grafik entweder blau oder grün oder blau–grün schraffiert markiert. Es gilt:

$$\begin{aligned} \Pr(\mathcal{C} | m_0) &= 1 - \Pr(\mathcal{E} | m_0), \quad \Pr(\mathcal{E} | m_0) = \Pr(\mathbf{r} \in I_1 \cup \mathbf{r} \in I_2 | m_0) \leq \\ &\leq \Pr(\mathbf{r} \in I_1 | m_0) + \Pr(\mathbf{r} \in I_2 | m_0) = Q[d_{01}/(2\sigma_n)] + Q[d_{02}/(2\sigma_n)] . \end{aligned}$$

- Das „ \leq “-Zeichen berücksichtigt hier, dass die blau–grün schraffierte Fläche sowohl zum Gebiet „ $\mathbf{r} \in I_1$ “ als auch zum Gebiet „ $\mathbf{r} \in I_2$ “ gehört, so dass die Summe einen zu großen Wert liefert. Das heißt: Die Union Bound liefert stets eine obere Schranke.
- Die mittlere Grafik verdeutlicht die Berechnung der Union Bound unter der Voraussetzung, dass $m_1 \Leftrightarrow \mathbf{s}_1$ gesendet wurde. Für das rechte Bild ist $m = m_2 \Leftrightarrow \mathbf{s} = \mathbf{s}_2$ zugrundegelegt.

Union Bound - Obere Schranke für die Fehlerwahrscheinlichkeit (3)

Die Abschätzung nach der „Union Bound“ lässt sich verbessern, indem man nur solche Signalraumpunkte berücksichtigt, die direkte Nachbarn des aktuellen Sendevektors \mathbf{s}_i sind:

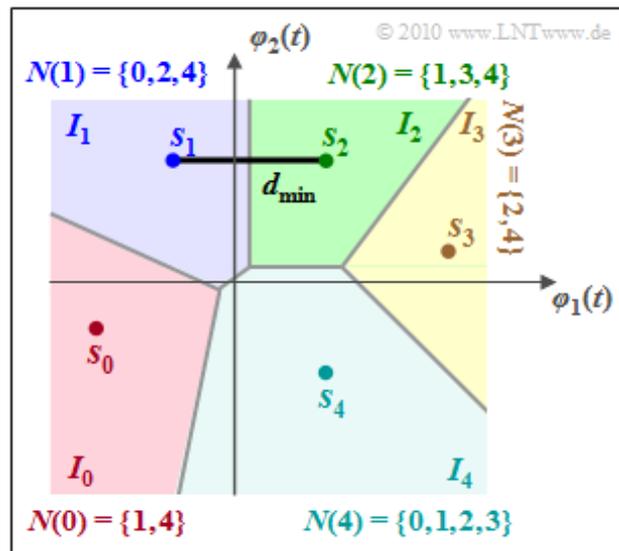
$$\Pr(\mathcal{E} | m_i) = \sum_{k=0, k \neq i}^{M-1} Q[d_{ik}/(2\sigma_n)] \Rightarrow \Pr(\mathcal{E} | m_i) = \sum_{k=0, k \in N(i)}^{M-1} Q[d_{ik}/(2\sigma_n)] .$$

Dazu definieren wir die Nachbarn von \mathbf{s}_i als

$$N(i) = \{k \in \{i = 0, 1, 2, \dots, M - 1\} \mid I_i \text{ grenzt direkt an } I_k\} .$$

Die Grafik verdeutlicht diese Definition am Beispiel $M = 5$. Die Regionen I_0 und I_3 haben jeweils nur zwei direkte Nachbarn, während I_4 an alle anderen Entscheidungsregionen angrenzt.

Durch die Einführung der Nachbarmengen $N(i)$ wird die Qualität der Union Bound-Approximation verbessert, das heißt, die Schranke liegt dann näher an der tatsächlichen Fehlerwahrscheinlichkeit, wird also nach unten verschoben.



Eine weitere und häufig verwendete Schranke benutzt nur den minimalen Abstand d_{\min} zwischen zwei Signalpunkten. Im obigen Beispiel tritt dieser zwischen \mathbf{s}_1 und \mathbf{s}_2 auf.

Dann gilt für gleichwahrscheinliche Symbole $\Rightarrow \Pr(m_i) = 1/M$ die folgende Abschätzung:

$$\begin{aligned} \Pr(\mathcal{E}) &\leq \sum_{i=0}^{M-1} \left[\Pr(m_i) \cdot \sum_{k \neq i} Q[d_{ik}/(2\sigma_n)] \right] \leq \frac{1}{M} \cdot \sum_{i=0}^{M-1} \left[\sum_{k \neq i} Q[d_{\min}/(2\sigma_n)] \right] = \\ &= \sum_{k \neq i} Q[d_{\min}/(2\sigma_n)] = (M - 1) \cdot Q[d_{\min}/(2\sigma_n)] . \end{aligned}$$

Hierzu ist anzumerken:

- Diese Schranke ist auch für große M -Werte sehr einfach zu berechnen. Bei vielen Anwendungen ergibt sich jedoch damit ein viel zu großer Wert für die Fehlerwahrscheinlichkeit.
- Die Schranke ist nur dann gleich der tatsächlichen Fehlerwahrscheinlichkeit, wenn alle Regionen an alle anderen direkt angrenzen und die Distanzen aller M Signalpunkte gleich d_{\min} sind.
- Im Sonderfall $M = 2$ sind diese beiden Voraussetzungen häufig erfüllt, so dass dann die hier angegebene Schranke exakt mit der tatsächlichen Fehlerwahrscheinlichkeit übereinstimmt.

Signalraumdarstellung der linearen Modulation (1)

Im bisherigen Kapitel 4 wurde die Struktur des optimalen Empfängers und die Signaldarstellung mittels Basisfunktionen am Beispiel der Basisbandübertragung behandelt. Mit der gleichen Systematik und der gleichen Einheitlichkeit sollen nun auch Bandpass-Systeme betrachtet werden, die bereits in früheren Büchern bzw. Kapiteln beschrieben wurden, nämlich

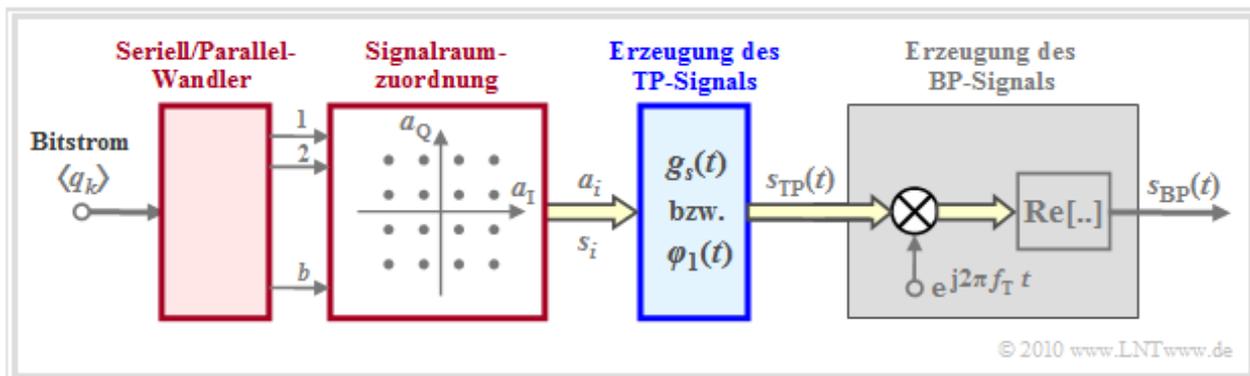
- im **Kapitel 4** des Buches „Modulationsverfahren“,
- im **Kapitel 1.5** des vorliegenden Buches.

Dabei beschränken wir uns im Kapitel 4.4 auf

- **lineare Modulationsverfahren**, und
- **kohärente Demodulation**.

Das bedeutet, dass am Empfänger das beim Sender zugesetzte Trägersignal hinsichtlich Frequenz und Phase exakt bekannt sein muss.

In diesem Fall kann das gesamte Übertragungssystem im **äquivalenten Tiefpassbereich** beschrieben werden und der Zusammenhang zur Basisbandübertragung ist noch offensichtlicher zu erkennen als bei Betrachtung der Bandpass-Signale. Es ergibt sich somit das folgende Modell, das auf der nächsten Seite im Detail beschrieben wird.



Die Beschreibung der nichtlinearen Modulationsverfahren und der nichtkohärenten Demodulation folgt im **Kapitel 4.5**.

Signalraumdarstellung der linearen Modulation (2)

Zum Übertragungsmodell der letzten Seite ist Folgendes zu bemerken:

- Aus dem ankommenden Bitstrom $\langle q_k \rangle \in \{0, L\}$ werden je b Datenbits seriell/parallel gewandelt. Diese Ausgangsbits ergeben die Nachricht $m \in \{m_0, \dots, m_{M-1}\}$, wobei $M = 2^b$ die Stufenzahl angibt. Für das Folgende wird die Nachricht $m = m_i$ vorausgesetzt.
- In der **Signalraumzuordnung** wird jeder Nachricht m_i ein komplexer Amplitudenkoeffizient $a_i = a_{Ii} + j \cdot a_{Qi}$ zugeordnet, dessen Realteil die Inphasekomponente und dessen Imaginärteil die Quadraturkomponente des späteren Sendesignals formen wird.

- Am Ausgang des blau markierten Blockes **Erzeugung des TP-Signals** liegt das (im allgemeinen) komplexwertige Tiefpass-Signal

$$s_{\text{TP}}(t) \Big|_{m=m_i} = a_i \cdot g_s(t) = a_{Ii} \cdot g_s(t) + j \cdot a_{Qi} \cdot g_s(t)$$

vor, wobei $g_s(t)$ vorerst ebenso wie $s_{\text{TP}}(t)$ auf den Bereich $0 \leq t < T$ beschränkt sein soll und der Index i wiederum einen Hinweis auf die gesendete Nachricht m_i liefert.

- Durch Energienormierung kommt man vom Sendegrundimpuls $g_s(t)$ zur Basisfunktion

$$\varphi_1(t) = g_s(t) / \sqrt{E_{gs}} \quad \text{mit} \quad E_{gs} = \int_0^T g_s(t)^2 dt$$

$$\Rightarrow s_{\text{TP}}(t) \Big|_{m=m_i} = s_{Ii} \cdot \varphi_1(t) + s_{Qi} \cdot j \cdot \varphi_1(t).$$

- Während die Koeffizienten a_{Ii} und a_{Qi} dimensionslos sind, weisen die neuen Koeffizienten s_{Ii} und s_{Qi} die Einheit „Wurzel aus Energie“ auf – siehe auch **Kapitel 4.1**:

$$s_{Ii} = \sqrt{E_{gs}} \cdot a_{Ii}, \quad s_{Qi} = \sqrt{E_{gs}} \cdot a_{Qi}.$$

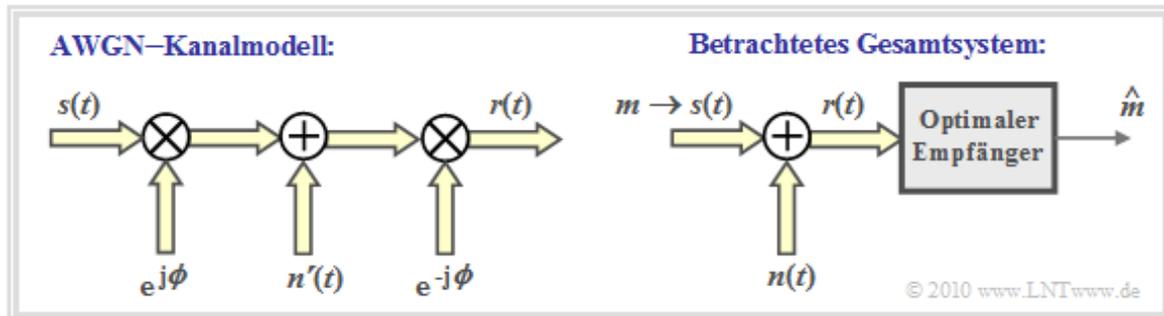
- Die obere Gleichung zeigt weiter, dass das hier betrachtete System im äquivalenten TP-Bereich durch je eine reelle Basisfunktion $\varphi_1(t)$ und eine rein imaginäre Basisfunktion $\psi_1(t) = j \cdot \varphi_1(t)$ oder durch eine einzige komplexe Basisfunktion $\xi_1(t)$ vollständig beschrieben wird.
- Der grau hinterlegte Teil des Blockschaltbildes zeigt das Modell zur Erzeugung des BP-Signals $s_{\text{BP}}(t)$, zuerst die Multiplikation des TP-Signals $s_{\text{TP}}(t)$ mit dem komplexen Drehzeiger $\exp(j2\pi f_T t)$ – auf diese Weise ergibt sich das analytische Signal $s_+(t)$ – und anschließend die Realteilbildung.
- Die beiden Basisfunktionen des Bandpass-Signals $s_{\text{BP}}(t)$ ergeben sich hier als energienormierte und auf den Bereich $0 \leq t \leq T$ zeitbegrenzte Cosinus- bzw. Minus-Sinus-Schwingungen.

Im Folgenden beschränken wir uns aber auf die Darstellung des äquivalenten Tiefpass-Signals $s_{\text{TP}}(t)$.

Hinweis: Im skizzierten Modell sind komplexe Größen durch einen gelb gefüllten Doppelpfeil markiert. Diese Vereinbarung soll auch für alle nachfolgenden Grafiken gelten.

Kohärente Demodulation und optimaler Empfänger

Im Folgenden gehen wir stets vom äquivalenten Tiefpass-Signalen aus, wenn nicht ausdrücklich etwas anderes angegeben ist. Insbesondere stellen die Signale $s(t)$ und $r(t)$ in der Grafik Tiefpass-Signale und sind somit im Allgemeinen komplex. Auf den Zusatz „TP“ wird im Weiteren verzichtet.



Zu dieser Abbildung ist zu bemerken:

- Die Phasenlaufzeit des Kanals (also eine mit der Frequenz linear ansteigende Phasenfunktion) wird im Tiefpassbereich durch den zeitunabhängigen Drehfaktor $\exp(j\phi)$ ausgedrückt.
- Das Signal $n'(t)$ beschreibt einen komplexen weißen Gaußschen Zufallsprozess im TP-Bereich, wie im **Kapitel 4.2** angegeben. Das Hochkomma wurde angefügt, um später beim Gesamtsystem mit $n(t)$ arbeiten zu können.
- Der Empfänger kennt die Kanalphase ϕ und korrigiert diese durch den konjugiert-komplexen Drehfaktor $\exp(-j\phi)$. Damit lautet das Empfangssignal im äquivalenten Tiefpassbereich:

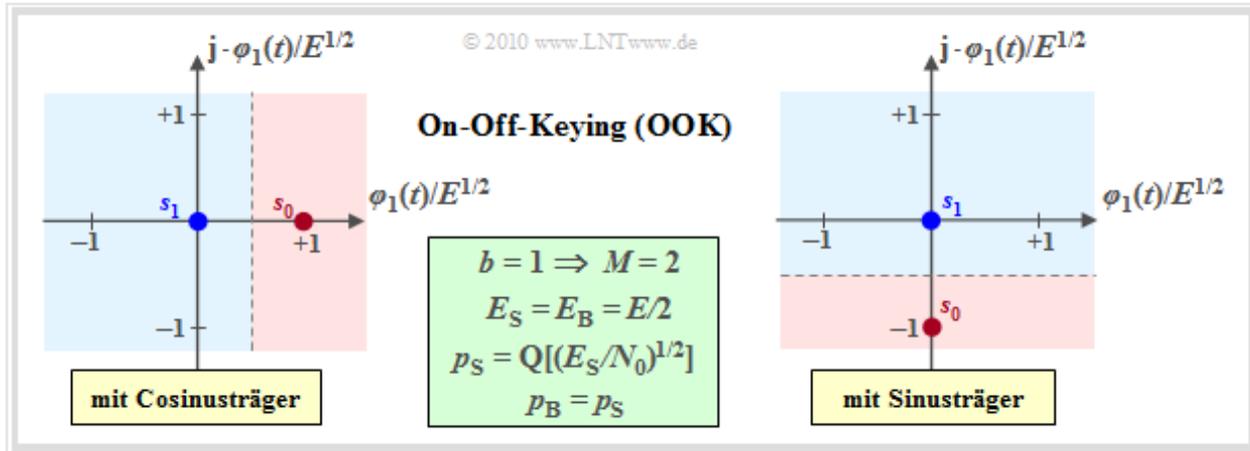
$$r(t) = s(t) + n'(t) \cdot e^{j\phi} = s(t) + n(t).$$
- Durch die Phasendrehung ändert sich an den Eigenschaften des zirkular symmetrischen Rauschens nichts. Das heißt, $n(t) = n'(t) \cdot \exp(-j\phi)$ hat genau gleiche statistische Eigenschaften wie $n'(t)$.

Die linke Grafik im obigen Bild verdeutlicht die soeben beschriebenen Sachverhalte. Die rechte Grafik zeigt das Gesamtsystem, wie es für den Rest von Kapitel 4 verwendet wird. Nach dem AWGN-Kanal folgt ein optimaler Empfänger gemäß **Kapitel 4.2**. Ein Symbolfehler kann wie folgt beschrieben werden:

$$m = m_i \cap \hat{m} \neq m_i.$$

On-Off-Keying bzw. 2-ASK

Das einfachste digitale Modulationsverfahren ist *On-Off-Keying* (**OOK**), das bereits im **Kapitel 4.2** des Buches „Modulationsverfahren“ anhand seiner Bandpass-Signale ausführlich beschrieben wurde. Dort wurde dieses Verfahren teilweise auch *Amplitude Shift Keying* (**ASK**) genannt.



Dieses Verfahren kann wie folgt charakterisiert werden:

- OOK ist ein eindimensionales Modulationsverfahren ($N = 1$) mit $s_{li} \in \{0, E^{1/2}\}$ und $s_{Qi} = 0$ bzw. $s_{li} = 0$ und $s_{Qi} \in \{0, -E^{1/2}\}$. Abkürzend gilt $E = E_{gs}$. Die erste Kombination beschreibt ein cosinusförmiges Trägersignal, die zweite Kombination einen sinusförmigen Träger.
- Jedes Bit wird einem Binärsymbol zugeordnet ($b = 1, M = 2$); man benötigt keinen S/P-Wandler. Bei gleichwahrscheinlichen Symbolen, was für das Folgende stets vorausgesetzt wird, ist sowohl die *mittlere Energie pro Symbol* (E_S) als auch die *mittlere Energie pro Bit* (E_B) gleich $E/2$.
- Der optimale OOK-Empfänger projiziert quasi das komplexwertige Empfangssignal $r(t)$ auf die Basisfunktion $\varphi_1(t)$, wenn man von der linken Skizze (Cosinusträger) ausgeht.
- Wegen $N = 1$ kann das Rauschen eindimensional angesetzt werden mit der Varianz $\sigma_n^2 = N_0/2$. Mit den Aussagen von **Kapitel 4.3** erhält man für die mittlere *Symbolfehlerwahrscheinlichkeit*:

$$p_S = \Pr(\mathcal{E}) = Q\left(\frac{d/2}{\sigma_n}\right) = Q\left(\sqrt{\frac{E}{2N_0}}\right) = Q\left(\sqrt{E_S/N_0}\right).$$

- Da jedes Bit genau auf ein Symbol abgebildet wird, ist die mittlere *Bitfehlerwahrscheinlichkeit* p_B genau so groß:

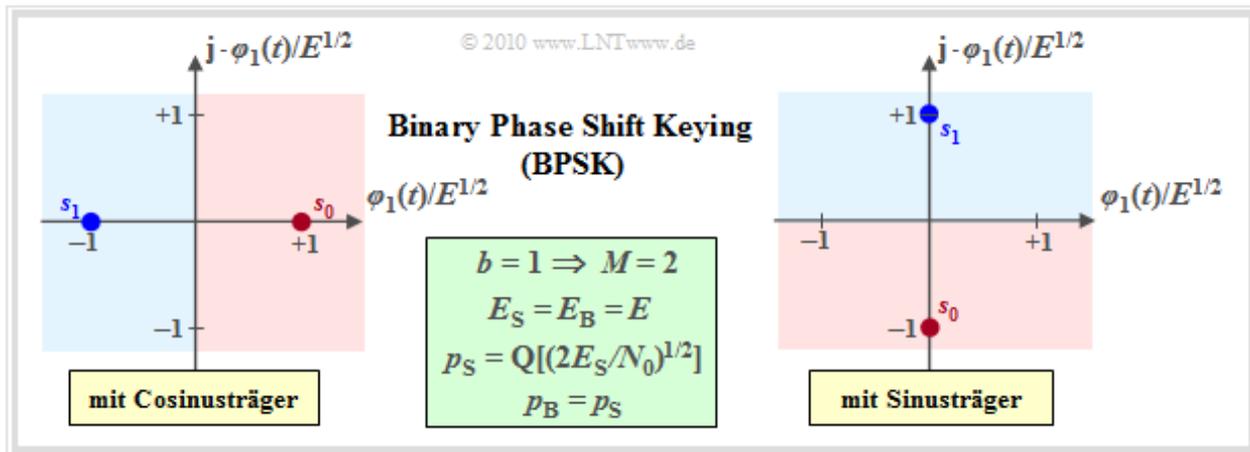
$$p_B = Q\left(\sqrt{E_S/N_0}\right) = Q\left(\sqrt{E_B/N_0}\right).$$

Binary Phase Shift Keying (BPSK)

Das sehr oft angewandte Verfahren *Binary Phase Shift Keying* (BPSK), das bereits im **Kapitel 4.2** des Buches „Modulationsverfahren“ anhand der Bandpass-Signale (typisch: Phasensprünge) ausführlich beschrieben wurde, unterscheidet sich von *On-Off-Keying* durch eine konstante Hüllkurve.

Für die Signalraumpunkte gilt stets $\mathbf{s}_1 = -\mathbf{s}_0$. Sie lauten beispielsweise:

- $s_{li} \in \{\pm E^{1/2}\}$, $s_{Qi} = 0$ bei cosinusförmigem Träger,
- $s_{Qi} \in \{\pm E^{1/2}\}$, $s_{li} = 0$ bei sinusförmigem Träger.



Anhand der in der Grafik angegebenen Gleichungen (grün hinterlegtes Feld) erkennt man die folgenden Verbesserungen gegenüber On-Off-Keying (OOK):

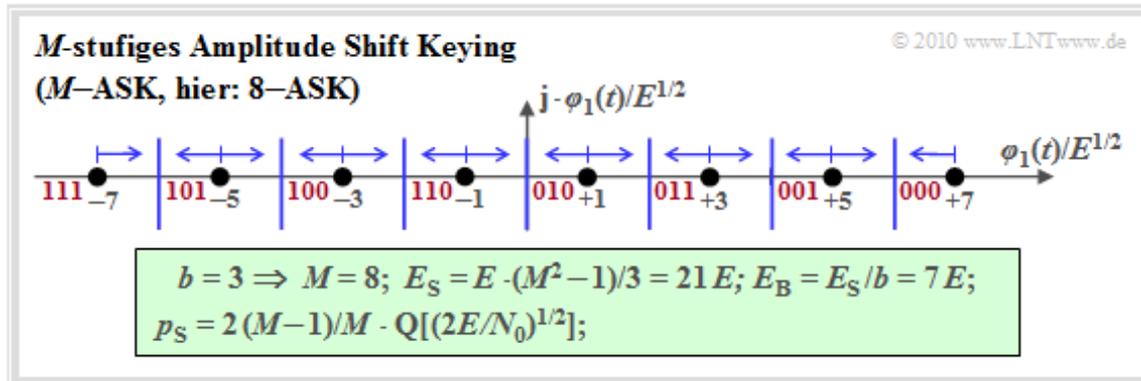
- Bei gegebener Normierungsenergie E ist der Abstand zwischen \mathbf{s}_1 und \mathbf{s}_0 doppelt so groß. Damit erhält man für die Fehlerwahrscheinlichkeit:

$$p_S = \Pr(\mathcal{E}) = Q\left(\frac{d/2}{\sigma_n}\right) = Q\left(\sqrt{2E/N_0}\right) = Q\left(\sqrt{2E_S/N_0}\right).$$

- In dieser Gleichung ist ebenfalls berücksichtigt, dass nun $E_S = E_B = E$ gilt, das heißt, dass nun die mittleren Energien pro Symbol bzw. pro Bit doppelt so groß sind als bei OOK.
- Die BPSK-Fehlerwahrscheinlichkeit ist durch den Faktor 2 unter der Wurzel im Argument der Q-Funktion merklich geringer als bei On-Off-Keying, wenn E_S und N_0 nicht verändert werden.
- Anders ausgedrückt: BPSK benötigt bei gleichem N_0 nur die halbe Symbolenergie E_S , um die gleiche Fehlerwahrscheinlichkeit wie OOK zu erzielen. Der logarithmische Gewinn beträgt 3 dB.

M-stufiges Amplitude Shift Keying (M-ASK)

In Analogie zur **M-stufigen Basisbandübertragung** betrachten wir nun eine **M-stufige ASK**, dessen Tiefpass-Signalraumkonstellation für die Parameter $b = 3 \Rightarrow M = 8 \Rightarrow 8\text{-ASK}$ wie folgt aussieht.



Der Name **M-ASK** ist nicht ganz zutreffend. Vielmehr handelt es sich um ein *kombiniertes ASK/PSK-Verfahren*, da sich zum Beispiel die beiden innersten Signalraumpunkte (± 1) nicht in der Amplitude (Hüllkurve) unterscheiden, sondern nur durch die Phase (0° bzw. 180°). Weiter ist anzumerken:

- Die *mittlere Energie pro Symbol* kann für dieses eindimensionale Verfahren unter Ausnutzung der Symmetrie wie folgt berechnet werden:

$$E_S = \frac{2}{M} \cdot \sum_{k=1}^{M/2} (2k - 1)^2 \cdot E = \frac{M^2 - 1}{3} \cdot E.$$

- Da jedes der M Symbole $b = \log_2(M)$ Bit darstellt, erhält man für die *mittlere Energie pro Bit*:

$$E_B = \frac{E_S}{b} = \frac{E_S}{\log_2(M)} = \frac{M^2 - 1}{3 \cdot \log_2(M)} \cdot E \Rightarrow M = 8: E_S/E = 21, E_B/E = 7.$$

- Die Wahrscheinlichkeit, dass eines der beiden äußeren Symbole aufgrund von AWGN-Rauschen verfälscht wird, ist gleich

$$\Pr(\mathcal{E} \mid \text{äusseres Symbol}) = Q\left(\sqrt{2E/N_0}\right).$$

- Die Verfälschungswahrscheinlichkeit der $M-2$ inneren Symbole ist doppelt so groß, da hier sowohl rechts als auch links andere Entscheidungsregionen angrenzen. Durch Mittelung erhält man die *mittlere Symbolfehlerwahrscheinlichkeit*:

$$\begin{aligned} p_S = \Pr(\mathcal{E}) &= \frac{1}{M} \cdot \left[2 \cdot 1 \cdot Q\left(\sqrt{2E/N_0}\right) + (M - 2) \cdot 2 \cdot Q\left(\sqrt{2E/N_0}\right) \right] = \\ &= \frac{2 \cdot (M - 1)}{M} \cdot Q\left(\sqrt{2E/N_0}\right) = \frac{2 \cdot (M - 1)}{M} \cdot Q\left(\sqrt{\frac{6 \cdot E_S}{(M^2 - 1) \cdot N_0}}\right). \end{aligned}$$

- Bei Verwendung des **Graycodes** (benachbarte Symbole unterscheiden sich jeweils um ein Bit) ist die *Bitfehlerwahrscheinlichkeit* näherungsweise um den Faktor $b = \log_2(M)$ kleiner als p_S :

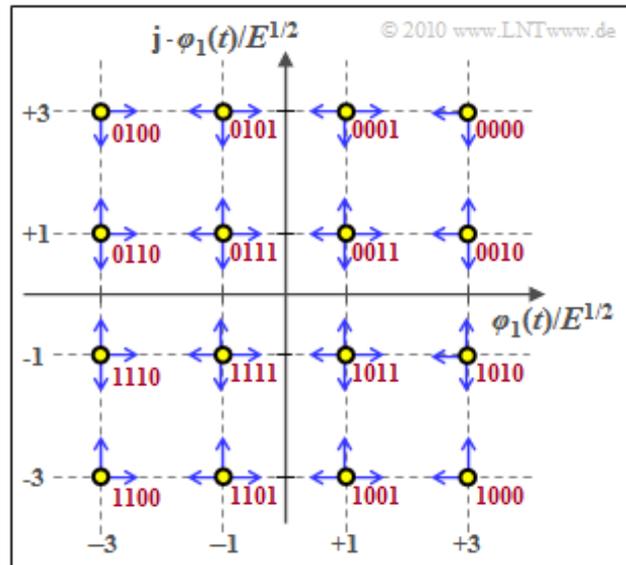
$$p_B \approx \frac{p_S}{b} = \frac{2 \cdot (M - 1)}{M \cdot \log_2(M)} \cdot Q\left(\sqrt{6 \cdot \log_2(M) / (M^2 - 1) \cdot E_B / N_0}\right).$$

Quadraturamplitudenmodulation (M-QAM)

Die **Quadraturamplitudenmodulation (M-QAM)** ergibt sich durch je eine M-ASK für Inphase- und Quadraturkomponente $\Rightarrow M^2$ Signalraumpunkte.

Durch jedes Symbol werden nun $b = \log_2(M)$ Binärzeichen (Bit) dargestellt. Die Grafik zeigt den Sonderfall $M = 16 \Rightarrow b = 4$. Rot eingezeichnet ist die Bitzuordnung nach der Graycodierung.

Die *mittlere Energie* pro Symbol (E_S) bzw. pro Bit (E_B) kann man aus dem Ergebnis für die M-ASK einfach berechnen (beachten Sie in dieser Gleichung den Unterschied zwischen einer Energie E und dem Erwartungswert $E[\dots]$):



$$E_S = E[|s_i|^2] = E[|s_{Ii}|^2] + E[|s_{Qi}|^2] = 2 \cdot E[|s_{Ii}|^2] = 2 \cdot \frac{M_I^2 - 1}{3} \cdot E = 2/3 \cdot (M - 1) \cdot E, \quad E_B = \frac{2 \cdot (M - 1)}{3 \cdot \log_2(M)} \cdot E.$$

Daneben hat die M-stufige Quadraturamplitudenmodulation folgende Eigenschaften:

- Als obere Schranke für die Symbolfehlerwahrscheinlichkeit kann die **Union Bound** herangezogen werden, wobei zu beachten ist, dass ein inneres Symbol in vier Richtungen verfälscht werden kann:

$$p_S = \Pr(\mathcal{E}) \leq \begin{cases} 4 \cdot p & \text{für } M \geq 16, \\ 2 \cdot p & \text{für } M = 4, \end{cases} \quad \text{mit } p = Q\left(\sqrt{2E/N_0}\right).$$

- Berücksichtigt man, dass nur die $(M_I - 2)^2$ inneren Punkte in vier Richtungen verfälscht werden, die vier Eckpunkte dagegen nur in zwei und die restlichen in drei Richtungen (blaue Pfeile in der Grafik), so erhält man mit $M = M_I^2$ die bessere Näherung

$$\begin{aligned} p_S &\approx 1/M \cdot [(M_I - 2)^2 \cdot 4p + 4 \cdot 2p + 4 \cdot (M_I - 2) \cdot 3p] = \\ &= p/M \cdot [4 \cdot M - 16 \cdot \sqrt{M} + 16 + 8 + 12 \cdot \sqrt{M} - 24] = \\ &= 4p/M \cdot [M - \sqrt{M}] = 4p \cdot [1 - 1/\sqrt{M}] \\ \Rightarrow M = 16: p_S &\approx 3 \cdot p = 3 \cdot Q\left(\sqrt{2E/N_0}\right) = 3 \cdot Q\left(\sqrt{1/5 \cdot E_S/N_0}\right). \end{aligned}$$

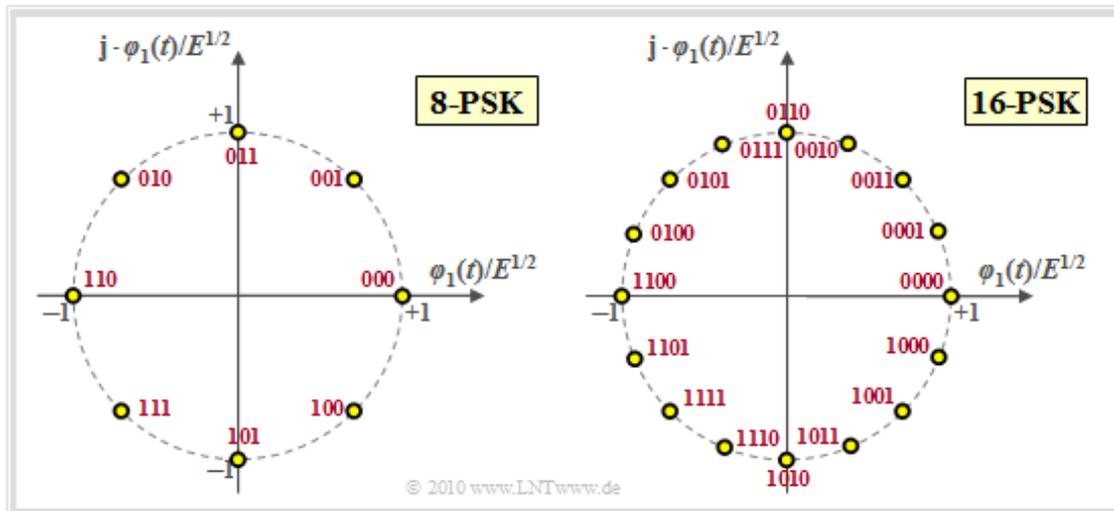
Allgemein gilt $E_B = E_S/\log_2(M)$ und bei Graycodierung zusätzlich $p_B = p_S/\log_2(M)$. Damit erhält man für die *mittlere Bitfehlerwahrscheinlichkeit*:

$$p_B \approx 4 \cdot (1 - 1/\sqrt{M})/\log_2(M) \cdot Q\left(\sqrt{3 \cdot \log_2(M)/(M - 1) \cdot E_B/N_0}\right).$$

Anmerkung: Die beiden Näherungen gelten für $M \geq 16$ exakt, wenn – wie für die obere Grafik vorausgesetzt – keine „diagonalen Verfälschungen“ auftreten können. Der Sonderfall 4-QAM (ohne innere Symbole) wird in der **Aufgabe A4.13** behandelt.

Mehrstufiges Phase-Shift Keying (M-PSK)

Bei mehrstufiger Phasenmodulation, wobei die Stufenzahl M in der Praxis meist eine Zweierpotenz ist, liegen alle Signalraumpunkte auf einem Kreis mit Radius $E^{1/2}$ gleichmäßig verteilt. Damit gilt für die mittlere Symbolenergie $E_S = E$ und für die *mittlere Energie pro Bit* $\Rightarrow E_B = E_S/b = E/\log_2(M)$.



Für Inphase- und die Quadraturkomponente der Signalraumpunkte s_i gilt allgemein ($i = 0, \dots, M-1$):

$$s_{Ii} = \cos(2\pi i/M + \phi_{\text{off}}), \quad s_{Qi} = \sin(2\pi i/M + \phi_{\text{off}}) \Rightarrow \|s_i\| = \sqrt{s_{Ii}^2 + s_{Qi}^2} = 1.$$

Der Phasenoffset ϕ_{off} ist in obiger Grafik jeweils zu 0 gesetzt. Die 4-PSK mit $\phi_{\text{off}} = \pi/4$ (45°) ist identisch mit der 4-QAM. Der Abstand zwischen zwei benachbarten Punkten ist in allen Fällen gleich:

$$d_{\min} = d_{0,1} = d_{1,2} = \dots = d_{M-1,0} = 2 \cdot \sqrt{E} \cdot \sin(\pi/M).$$

$$\Rightarrow d_{\min}/E^{0.5} = 2^{0.5} \approx 1.414 \text{ (für } M=4\text{)}, \approx 0.765 \text{ (für } M=8\text{)} \text{ und } \approx 0.390 \text{ (für } M=16\text{)}.$$

Die obere Schranke für die AWGN-Symbolfehlerwahrscheinlichkeit nach der **Union Bound** liefert:

$$p_S = \Pr(\mathcal{E}) \leq 2 \cdot Q\left(\sin(\pi/M) \cdot \sqrt{2E_S/N_0}\right) = p_{UB}.$$

Für $M = 2$ (BPSK) erhält man daraus die Abschätzung $p_S \leq p_{UB} = 2 \cdot Q[(2E_S/N_0)^{1/2}]$. Ein Vergleich mit der auf der **BPSK-Seite** angegebenen Gleichung $p_S = Q[(2E_S/N_0)^{1/2}]$ zeigt, dass in diesem Sonderfall die „Union Bound“ als obere Schranke den doppelten Wert liefert. Je größer M ist, umso genauer nähert p_{UB} die Symbolfehlerwahrscheinlichkeit p_S an. Das Interaktionsmodul **Mehrstufige PSK und Union Bound** gibt auch die exakte, durch Simulation gewonnene Fehlerwahrscheinlichkeit an.

Die Schranke für die Bitfehlerwahrscheinlichkeit lautet (Graycode \Rightarrow rote Beschriftung vorausgesetzt):

$$p_B \leq \frac{2}{\log_2(M)} \cdot Q\left(\sqrt{\log_2(M)} \cdot \sin(\pi/M) \cdot \sqrt{2E_B/N_0}\right).$$

Diese Schranke muss man allerdings nur für $M > 4$ anwenden. Für $M = 2$ (BPSK) und $M = 4$ (wegen der Identität zwischen 4-PSK und 4-QAM) kann man die Bitfehlerwahrscheinlichkeit direkt angeben:

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

Binary Frequency Shift Keying \Rightarrow 2-FSK (1)

Auch diese Modulationsart mit Parameter $b = 1 \Rightarrow M = 2$ wurde bereits im **Kapitel 4.4** des Buches „Modulationsverfahren“ anhand der Bandpass-Signale ausführlich beschrieben. Die beiden möglichen Signalformen werden im Bereich $0 \leq t \leq T$ durch zwei unterschiedliche Frequenzen dargestellt:

$$\begin{aligned}s_{\text{BP}0}(t) &= A \cdot \cos(2\pi \cdot (f_T + \Delta f_A) \cdot t), \\ s_{\text{BP}1}(t) &= A \cdot \cos(2\pi \cdot (f_T - \Delta f_A) \cdot t).\end{aligned}$$

Hierbei bezeichnet f_T die Trägerfrequenz und Δf_A den (einseitigen) Frequenzhub. Die mittlere Energie pro Symbol bzw. pro Bit ist jeweils gleich:

$$E_S = E_B = E = \frac{A^2 \cdot T}{2}.$$

Hier soll nun die FSK im äquivalenten Tiefpass-Signalraum betrachtet werden. Hier gilt:

$$\begin{aligned}s_{\text{TP}0}(t) &= \sqrt{E/T} \cdot e^{+j \cdot 2\pi \cdot \Delta f_A \cdot t}, \quad 0 \leq t \leq T, \\ s_{\text{TP}1}(t) &= \sqrt{E/T} \cdot e^{-j \cdot 2\pi \cdot \Delta f_A \cdot t}, \quad 0 \leq t \leq T,\end{aligned}$$

und für das innere Produkt erhält man

$$\begin{aligned}\langle s_{\text{TP}0}(t) \cdot s_{\text{TP}1}(t) \rangle &= \int_0^T s_{\text{TP}0}(t) \cdot s_{\text{TP}1}^*(t) dt = A^2 \cdot \int_0^T e^{j \cdot 4\pi \cdot \Delta f_A \cdot t} dt = \\ &= \frac{A^2}{j \cdot 4\pi \cdot \Delta f_A} \cdot [e^{j \cdot 4\pi \cdot \Delta f_A \cdot T} - 1].\end{aligned}$$

Die beiden Signale sind dann orthogonal, wenn dieses innere Produkt gleich 0 ist. Also muss gelten:

$$h = 2 \cdot \Delta f_A \cdot T = 1, 2, 3, \dots$$

Diese Größe bezeichnet man als den **Modulationsindex**.

Binary Frequency Shift Keying \Rightarrow 2-FSK (2)

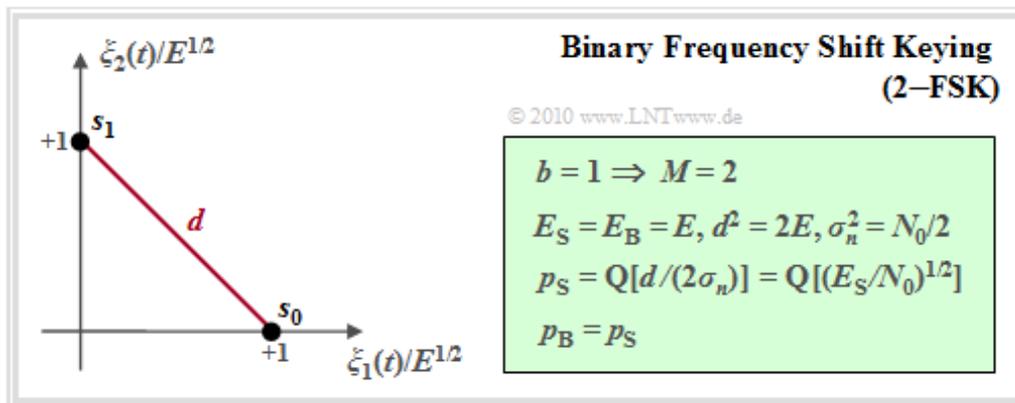
Der *Modulationsindex* $h = 2 \cdot \Delta f_A \cdot T$ gibt das Verhältnis zwischen dem gesamten (also beideseitigen) Frequenzhub ($2 \cdot \Delta f_A$) und der Symbolrate ($1/T$) an. Setzt man h als ganzzahlig voraus, so lassen sich die beiden Tiefpass-Signale in der Form

$$\begin{aligned} s_{\text{TP0}}(t) &= \sqrt{E} \cdot \xi_1(t), \\ s_{\text{TP1}}(t) &= \sqrt{E} \cdot \xi_2(t) \end{aligned}$$

mit den komplexen Basisfunktionen

$$\begin{aligned} \xi_1(t) &= \sqrt{1/T} \cdot e^{+j \cdot \pi \cdot h \cdot t/T}, \quad 0 \leq t \leq T, \\ \xi_2(t) &= \sqrt{1/T} \cdot e^{-j \cdot \pi \cdot h \cdot t/T}, \quad 0 \leq t \leq T \end{aligned}$$

darstellen, und es ergibt sich die nachfolgend skizzierte Signalraumdarstellung der binären FSK.



Daraus folgt:

- Bei ganzzahligem Modulationsindex h sind die beiden Tiefpass-Signale $s_{\text{TP0}}(t)$ und $s_{\text{TP1}}(t)$ der binären FSK zueinander orthogonal.
- Damit ergibt sich für die Symbolfehlerwahrscheinlichkeit (Herleitung in der Grafik):

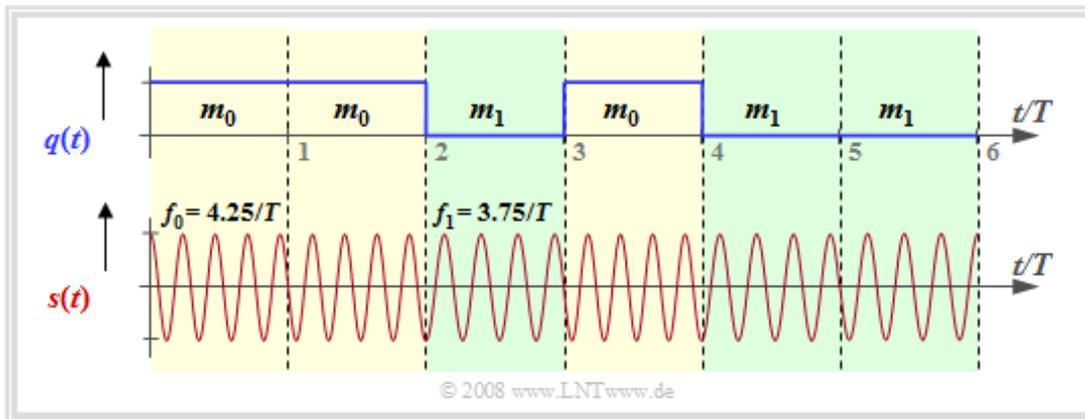
$$p_S = \text{Pr}(\mathcal{E}) = Q\left(\sqrt{E_S/N_0}\right).$$

- Die Bitfehlerwahrscheinlichkeit hat den gleichen Wert: $p_B = p_S$.

Hinweis: Im Gegensatz zur Darstellung in [KöZ08] ist hier der Frequenzhub Δf_A einseitig definiert. Deshalb unterscheiden sich die Gleichungen teilweise um den Faktor 2. Arbeitet man jedoch mit dem Modulationsindex h , so gibt es keine Unterschiede.

Minimum Shift Keying (MSK)

Unter **Minimum Shift Keying** (MSK) versteht man ein binäres FSK-System mit *Modulationsindex* $h = 0.5 \Rightarrow$ Frequenzhub $\Delta f_A = 1/(2T)$. Die Grafik zeigt ein MSK-Signal für die Trägerfrequenz $f_T = 4/T$. Die beiden Frequenzen innerhalb des Sendesignals sind $f_0 = f_T + 1/(4T)$ zur Darstellung der Nachricht m_0 (gelbe Hinterlegung) sowie $f_1 = f_T - 1/(4T) \Rightarrow$ Nachricht m_1 (grüne Hinterlegung). In dieser Grafik ist auch eine kontinuierliche Phasen Anpassung bei den Übergängen berücksichtigt, um die Signalbandbreite zu verringern. Man spricht dann von **Continuous Phase Modulation** (CPM).



Ohne diese Phasen Anpassung lauten die beiden Bandpass-Signalförmern:

$$s_{BP0}(t) = \sqrt{2E/T} \cdot \cos(2\pi f_0 t), \quad 0 \leq t \leq T,$$

$$s_{BP1}(t) = \sqrt{2E/T} \cdot \cos(2\pi f_1 t), \quad 0 \leq t \leq T.$$

Bildet man das innere Produkt der Bandpass-Signale, so erhält man mit $f_\Delta = f_0 - f_1$ und $f_\Sigma = f_0 + f_1$:

$$\begin{aligned} \langle s_{BP0}(t) \cdot s_{BP1}(t) \rangle &= 2E/T \cdot \int_0^T \cos(2\pi f_0 t) \cdot \cos(2\pi f_1 t) dt = \\ &= E/T \cdot \int_0^T \cos(2\pi f_\Delta t) dt + E/T \cdot \int_0^T \cos(2\pi f_\Sigma t) dt \\ &= E/T \cdot \int_0^T \cos(\pi \cdot t/T) dt + E/T \cdot \int_0^T \cos(2\pi \cdot 2f_T \cdot t) dt. \end{aligned}$$

Das erste Integral ist 0 (Integral über „Cosinus“ von 0 bis π). Für $f_T \gg 1/T$, was man in der Praxis voraussetzen kann, verschwindet auch das zweite Integral. Damit erhält man für das innere Produkt:

$$\langle s_{BP0}(t) \cdot s_{BP1}(t) \rangle = 0.$$

Damit ist gezeigt, dass für den Modulationsindex $h = 0.5$ (also MSK) und allen Vielfachen hiervon die beiden Bandpass-Signale orthogonal sind. Mit den neuen reellen Basisfunktionen

$$\varphi_1(t) = \sqrt{2/T} \cdot \cos(2\pi f_0 t), \quad 0 \leq t \leq T,$$

$$\varphi_2(t) = \sqrt{2/T} \cdot \cos(2\pi f_1 t), \quad 0 \leq t \leq T$$

erhält man die genau gleiche Signalraumkonstellation wie für geradzahliges $h (= 1, 2, 3, \dots)$, und es ergibt sich somit auch die gleiche Fehlerwahrscheinlichkeit:

$$p_S = \Pr(\mathcal{E}) = Q\left(\sqrt{E_S/N_0}\right) = p_B.$$

Rayleigh- und Riceverteilung (1)

Die für eine kohärente Demodulation erforderliche Schätzung des Phasenwinkels aus dem ankommenden Signal ist bei vielen Anwendungen nicht oder nur eingeschränkt möglich. So führt die Bewegung eines Mobilteilnehmers mit hoher Geschwindigkeit zu sehr schnellen zeitlichen Änderungen des Phasenwinkels ϕ , was dessen ausreichend genaue Bestimmung erschwert oder gar verhindert.

Diese Tatsache führt zu den **nichtkohärenten Demodulationsverfahren** mit dem Vorteil reduzierter Komplexität, allerdings mit erhöhter Verfälschungswahrscheinlichkeit. Bei der Herleitung der Gleichungen stößt man auf zwei Wahrscheinlichkeitsdichtefunktionen, die hier vorneweg angegeben werden:

- Die **Rayleighverteilung** erhält man für die WDF der Zufallsgröße y mit Realisierung η , die sich aus den beiden gaußverteilten und statistisch unabhängigen Komponenten u und v (beide mit der gleichen Streuung σ_n) wie folgt ergibt:

$$y = \sqrt{u^2 + v^2} \Rightarrow p_y(\eta) = \eta/\sigma_n^2 \cdot \exp[-\eta^2/(2\sigma_n^2)] .$$

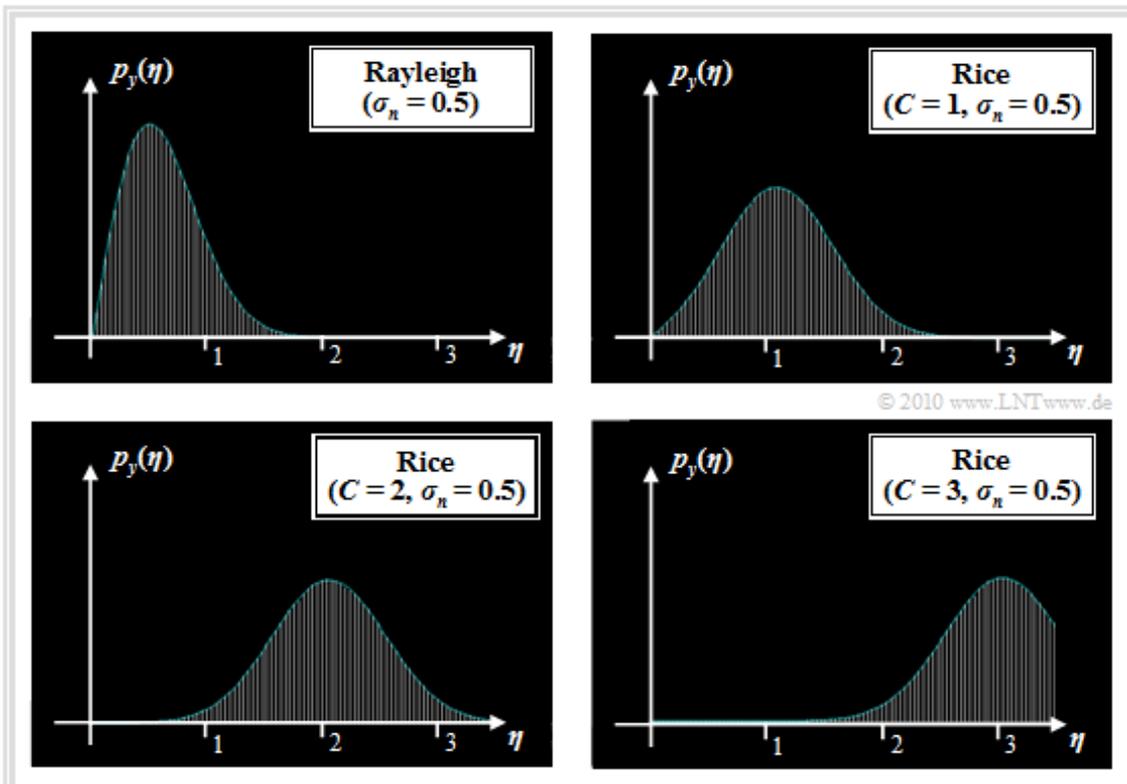
- Die **Riceverteilung** erhält man unter sonst gleichen Randbedingungen für den Fall, dass bei einer der Komponenten (entweder u oder v) noch eine Konstante C addiert wird:

$$y = \sqrt{(u + C)^2 + v^2} \Rightarrow p_y(\eta) = \eta/\sigma_n^2 \cdot \exp[-(\eta^2 + C^2)/(2\sigma_n^2)] \cdot I_0[\eta \cdot C/\sigma_n^2] .$$

Für die Riceverteilung benötigt man die *modifizierte Besselfunktion nullter Ordnung*, deren Definition und Reihenentwicklung wie folgt lauten:

$$I_0(x) = \frac{1}{2\pi} \cdot \int_0^{2\pi} e^{-x \cdot \cos(\alpha)} d\alpha \approx \sum_{k=0}^{\infty} \frac{(x/2)^{2k}}{k! \cdot \Gamma(k + 1)} .$$

Die Grafik zeigt Rayleigh- und Rice-Wahrscheinlichkeitsdichtefunktionen. Die Bildbeschreibung folgt auf der nächsten Seite.



Rayleigh- und Riceverteilung (2)

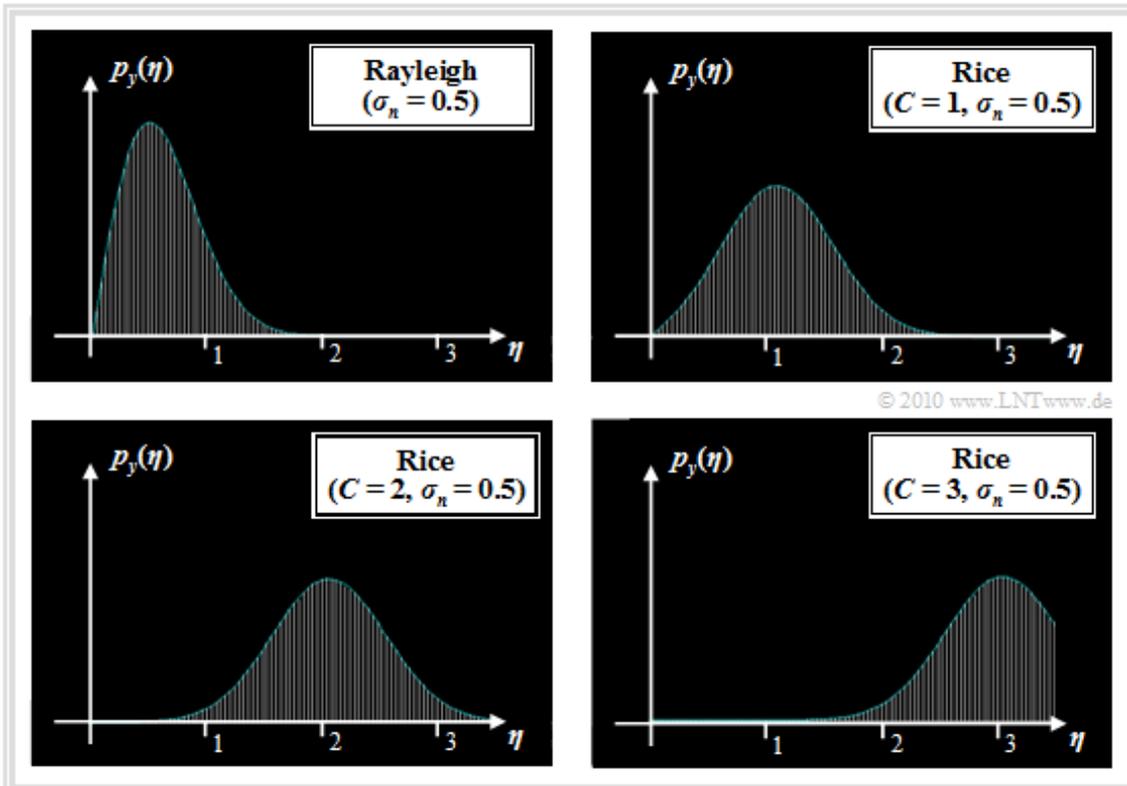
Die Grafik am Seitenende zeigt nochmals die Wahrscheinlichkeitsdichtefunktionen von Rayleigh- und Riceverteilung. Zu dieser Darstellung ist anzumerken:

- Die **Riceverteilung** ist durch die beiden Parameter C und σ_n bestimmt. Mit $C = 0$ ist die Rice-WDF identisch mit der **Rayleigh-WDF**.
- Die Rayleigh-WDF mit größerem σ_n ist formgleich mit der gezeichneten Kurve ($\sigma_n = 0.5$), jedoch im Verhältnis der Streuungen breiter und niedriger.
- σ_n gibt die Streuungen der beiden gaußverteilten Zufallsgrößen u und v an (beide haben gleiche Streuung) und nicht die Streuung der rayleighverteilten Zufallsgröße y . Für diese gilt vielmehr:

$$\sigma_y = \sigma_n \cdot \sqrt{2 - \pi/2} \approx 0.655 \cdot \sigma_n.$$

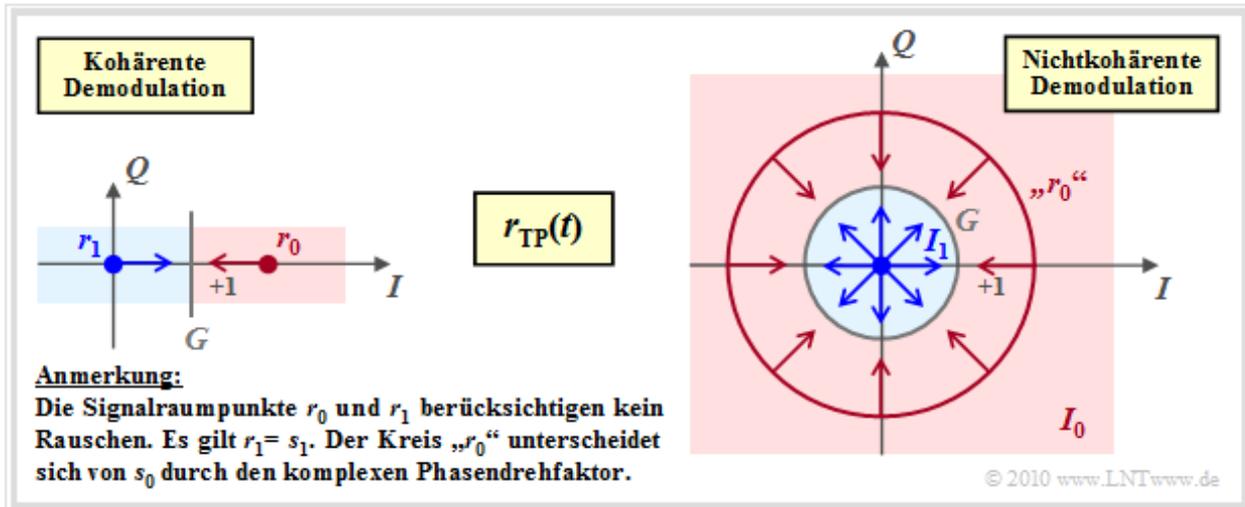
- Die Rayleighverteilung ist extrem unsymmetrisch, erkennbar am (relativ) großen Wert für das **Zentralmoment 3. Ordnung**: $\mu_3/\sigma_y^3 \approx 0.27$.
- Die Riceverteilung ist um so symmetrischer, je größer das Verhältnis C/σ_n von deterministischer und stochastischer Komponente ist. Für $C/\sigma_n \geq 4$ ist μ_3 nahezu 0.
- Weiterhin ist zu erkennen, dass sich die Riceverteilung (mit den Parametern C und σ_n) immer mehr einer Gaußverteilung mit Mittelwert C und Streuung σ_n annähert, je größer der Quotient C/σ_n ist:

$$p_y(\eta) \approx \frac{1}{\sqrt{2\pi} \cdot \sigma_n} \cdot \exp \left[-\frac{(\eta - C)^2}{2\sigma_n^2} \right] \Rightarrow m_y = C, \sigma_y = \sigma_n.$$



Nichtkohärente Demodulation von On-Off-Keying (1)

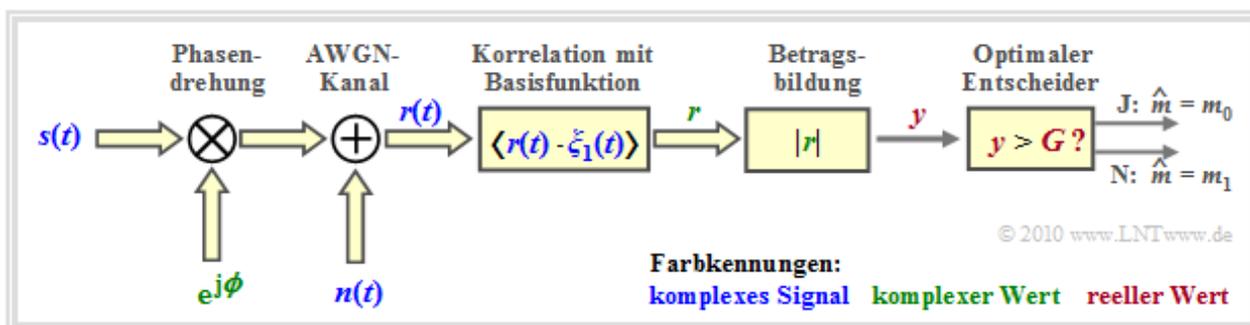
Wir betrachten **On-Off-Keying** im äquivalenten Tiefpassbereich. Bei kohärenter Demodulation (linke Grafik) ist die Signalraumkonstellation des Empfangssignals gleich der des Sendesignals und besteht aus zwei Punkten. Die Entscheidungsgrenze G liegt in der Mitte zwischen diesen Punkten r_0 und r_1 . Die Pfeile markieren Rauschvektoren, die eventuell zu Übertragungsfehlern führen.



Dagegen gilt bei nichtkohärenter Demodulation:

- Der Punkt $r_1 = s_1 = 0$ bleibt weiter erhalten.
- Dagegen kann $r_0 = s_0 \cdot \exp(j\phi)$ auf jeden Punkt des Kreises um s_0 liegen, da ϕ unbekannt ist.
- Der Entscheidungsprozess unter Berücksichtigung des AWGN-Rauschens ist nun 2-dimensional zu interpretieren, wie es durch die Pfeile in der rechten Grafik angedeutet ist.
- Das Entscheidungsgebiet I_1 ist nun ein Kreis, dessen Radius G ein optimierbarer Parameter ist. Das Entscheidungsgebiet I_0 liegt außerhalb dieses Kreises.

Damit liegt die Struktur des optimalen OOK-Empfängers (im äquivalenten Tiefpassbereich) fest.



Nichtkohärente Demodulation von On–Off–Keying (2)

Entsprechend der **Grafik** auf der letzten Seite gilt:

- Das Eingangssignal $\mathbf{r}(t) = \mathbf{s}(t) \cdot \exp(j\phi) + \mathbf{n}(t)$ ist aufgrund des Phasenwinkels ϕ und wegen des komplexen Rauschterms im allgemeinen komplex. Alle komplexen Signale sind blau beschriftet.
- Erforderlich ist demzufolge nun die Korrelation zwischen dem komplexen Empfangssignal $\mathbf{r}(t)$ und einer komplexen Basisfunktion $\xi_1(t)$.
- Das Ergebnis ist der (komplexe) Detektorwert \mathbf{r} , woraus als reelle Entscheidungsgröße der Betrag $y = |\mathbf{r}|$ gebildet wird.
- Ist der Entscheidungswert $y > G$, so wird als Schätzwert m_0 ausgegeben, andernfalls m_1 . Somit ergibt sich für die Fehlerwahrscheinlichkeit bei gleichwahrscheinlichen Symbolen:

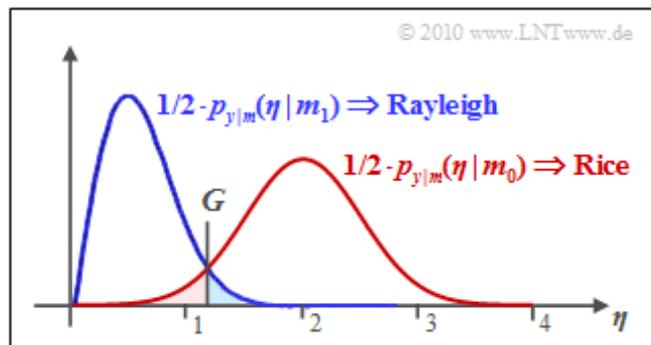
$$p_S = \Pr(\mathcal{E}) = 1/2 \cdot \int_0^G p_{y|m}(\eta|m_0) d\eta + 1/2 \cdot \int_G^\infty p_{y|m}(\eta|m_1) d\eta.$$

- Aufgrund der Rice–WDF $p_{y|m}(\eta|m_0)$ und der Rayleigh–WDF $p_{y|m}(\eta|m_1)$ kann allerdings diese Wahrscheinlichkeit nur numerisch berechnet werden. Die optimale Entscheidungsgrenze G ist vorher als die Lösung der folgenden Gleichung zu bestimmen:

$$p_{y|m}(G|m_0) = p_{y|m}(G|m_1).$$

Die Grafik zeigt das Ergebnis dieser Gleichung für $\sigma_n = 0.5$ und $C = 2$, wobei die (rote) Rice–WDF durch eine Gauß–WDF mit Mittelwert C und Streuung σ_n approximiert ist. Man erkennt daraus:

- Die optimale Entscheidungsgrenze (hier: $G \approx 1.25$) ergibt sich aus dem Schnittpunkt der beiden Kurven.
- Die Symbolfehlerwahrscheinlichkeit p_S ist die Summe der beiden farblich markierten Flächen. Im Beispiel ergibt sich $p_S \approx 5\%$.



Die Fehlerwahrscheinlichkeit für andere Werte von C und σ_n sowie die optimale Entscheidungsgrenze G können Sie mit dem Berechnungstool **Nichtkohärentes On–Off–Keying** bestimmen.

Nichtkohärente Demodulation von binärer FSK (1)

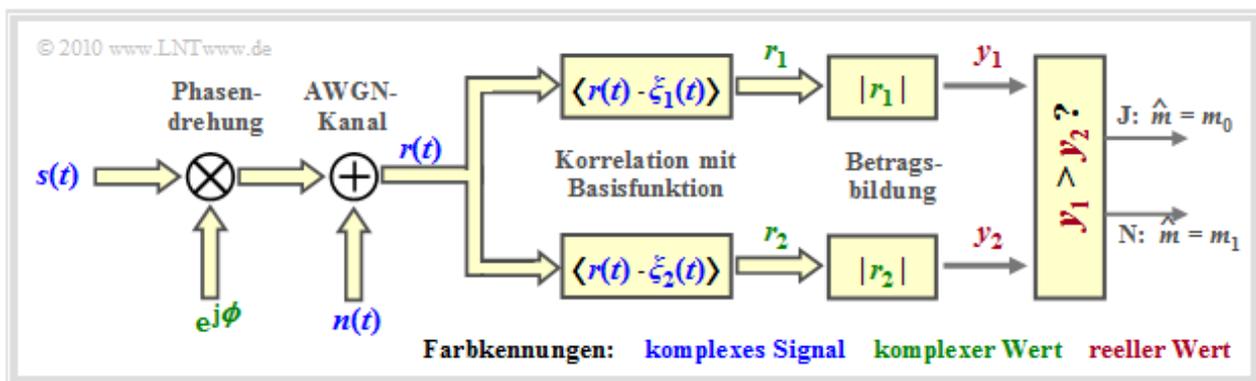
Wie schon im **Kapitel 4.4** gezeigt, lässt sich *binäres Frequency Shift Keying* (BFSK) im äquivalenten Tiefpassbereich durch die Basisfunktionen

$$\begin{aligned} \xi_1(t) &= \sqrt{1/T} \cdot e^{+j \cdot \pi \cdot h \cdot t/T}, \quad 0 \leq t \leq T, \\ \xi_2(t) &= \sqrt{1/T} \cdot e^{-j \cdot \pi \cdot h \cdot t/T}, \quad 0 \leq t \leq T \end{aligned}$$

darstellen. Um Orthogonalität zwischen diesen beiden komplexen Basisfunktionen zu erreichen, muss der **Modulationsindex** h ganzzahlig sein:

$$\langle \xi_1(t) \cdot \xi_2(t) \rangle = 0 \Rightarrow h = 2 \cdot \Delta f_A \cdot T = 1, 2, 3, \dots$$

Die Grafik zeigt die Struktur zur nichtkohärenten orthogonalen Demodulation der binären FSK.



Im rauschfreien Fall $\Rightarrow n(t) = 0$ gilt für die beiden Ausgänge der Korrelatoren:

$$\begin{aligned} r_1 &= \langle r(t) \cdot \xi_1(t) \rangle = 0, \quad \text{falls } m = m_1, \\ r_2 &= \langle r(t) \cdot \xi_2(t) \rangle = 0, \quad \text{falls } m = m_0. \end{aligned}$$

Nach jeweiliger Betragbildung $\Rightarrow y_1 = |r_1|, y_2 = |r_2|$ ist dann folgende Entscheidungsregel anwendbar:

$$\hat{m} = \begin{cases} m_0 & \text{falls } y_1 > y_2, \\ m_1 & \text{falls } y_1 < y_2. \end{cases}$$

Zur einfacheren Realisierung des Entscheiders kann auch die Differenz $y_1 - y_2$ mit der Entscheidungsgrenze $G = 0$ ausgewertet werden.

Im Folgenden wird die Fehlerwahrscheinlichkeit unter der Annahme berechnet, dass $m = m_0$ gesendet wurde. Unter der weiteren Voraussetzung gleichwahrscheinlicher binärer Nachrichten m_0 und m_1 ist die absolute Fehlerwahrscheinlichkeit genau so groß:

$$\Pr(\mathcal{E}) = \Pr(\mathcal{E} | m_0).$$

Mit $m = m_0$ ergeben sich für die komplexen Korrelationsausgangswerte r_i und deren Beträge y_i :

$$\begin{aligned} r_1 &= \sqrt{E} \cdot e^{j\phi} + n_1 \Rightarrow y_1 = |r_1| \text{ ist riceverteilt,} \\ r_2 &= n_2 \Rightarrow y_2 = |r_2| \text{ ist rayleighverteilt.} \end{aligned}$$

Hierbei steht E für die *Symbolenergie* E_S und die *Bitenergie* E_B gleichermaßen (wegen $M = 2$), und n_1 und n_2 sind unkorrelierte komplexe Rauschgrößen mit Mittelwert 0 und Varianz $2\sigma_n^2$.

Nichtkohärente Demodulation von binärer FSK (2)

Somit lautet die Verbundwahrscheinlichkeitsdichtefunktion:

$$p_{y_1, y_2 | m}(\eta_1, \eta_2 | m_0) = p_{y_1 | m}(\eta_1 | m_0) \cdot p_{y_2 | m}(\eta_2 | m_0),$$

$$p_{y_1 | m}(\eta_1 | m_0) = \eta_1 / \sigma_n^2 \cdot e^{-(\eta_1^2 + E)/(2\sigma_n^2)} \cdot I_0 \left[\eta_1 \cdot \sqrt{E} / \sigma_n^2 \right],$$

$$p_{y_2 | m}(\eta_2 | m_0) = \eta_2 / \sigma_n^2 \cdot e^{-\eta_2^2 / (2\sigma_n^2)}.$$

Die Fehlerwahrscheinlichkeit ergibt sich allgemein wie folgt:

$$\begin{aligned} \Pr(\mathcal{E}) &= \int_0^\infty \int_{\eta_1}^\infty p_{y_1, y_2 | m}(\eta_1, \eta_2 | m_0) \, d\eta_2 \, d\eta_1 = \\ &= \int_0^\infty p_{y_1 | m}(\eta_1 | m_0) \cdot \int_{\eta_1}^\infty p_{y_2 | m}(\eta_2 | m_0) \, d\eta_2 \, d\eta_1. \end{aligned}$$

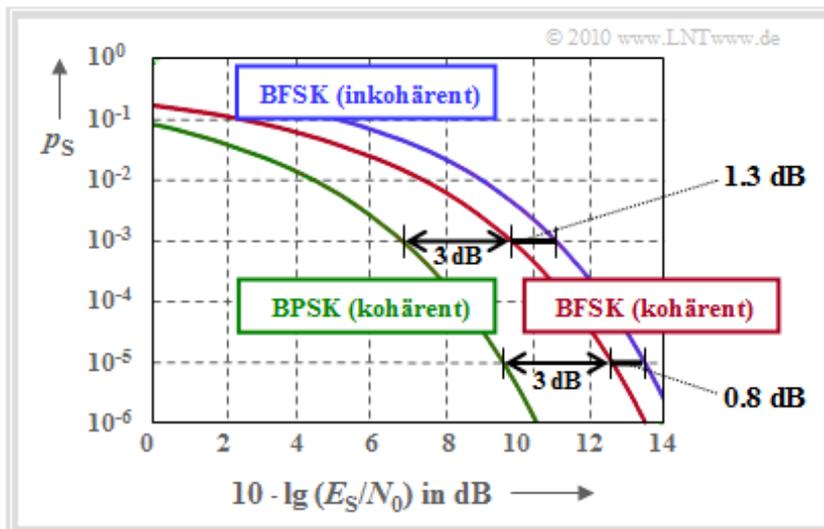
Nach einigen mathematischen Umformungen erhält man für die **nichtkohärente Demodulation** der binären FSK das überraschend einfache Ergebnis (Herleitung auf der nächsten Seite):

$$p_S = \Pr(\mathcal{E}) = 1/2 \cdot e^{-E_S / (2N_0)}.$$

Zum Vergleich sei nochmals das Ergebnis für die **kohärente Demodulation** angegeben:

$$p_S = \Pr(\mathcal{E}) = Q(\sqrt{E_S / N_0}).$$

Die Grafik stellt die Fehlerwahrscheinlichkeitskurven beider Demodulationsverfahren in Abhängigkeit des AWGN-Quotienten E_S/N_0 vergleichend gegenüber.



Man erkennt:

- Die nichtkohärente FSK benötigt gegenüber der kohärenten FSK bei $p_S = 10^{-5}$ ein um 0.8 dB größeres E_S/N_0 . Bei $p_S = 10^{-3}$ beträgt der Abstand sogar 1.3 dB.
- Dagegen beträgt der Abstand zwischen der kohärenten binären FSK von der kohärenten BPSK unabhängig von der Fehlerwahrscheinlichkeit gleich 3 dB.

Herleitung: Fehlerwahrscheinlichkeit von nichtkohärenter BFSK

Das auf der letzten Seite vorweg genommene Ergebnis soll nun in einigen Rechenschritten hergeleitet werden. Wir gehen dabei von den folgenden Gleichungen aus:

$$\Pr(\mathcal{E}) = \int_0^\infty p_{y_1|m}(\eta_1 | m_0) \cdot \int_{\eta_1}^\infty p_{y_2|m}(\eta_2 | m_0) d\eta_2 d\eta_1,$$

$$p_{y_1|m}(\eta_1 | m_0) = \eta_1 / \sigma_n^2 \cdot e^{-(\eta_1^2 + E)/(2\sigma_n^2)} \cdot I_0 \left[\eta_1 \cdot \sqrt{E} / \sigma_n^2 \right],$$

$$p_{y_2|m}(\eta_2 | m_0) = \eta_2 / \sigma_n^2 \cdot e^{-\eta_2^2 / (2\sigma_n^2)}.$$

(1) Das innere Integral gibt die Wahrscheinlichkeit an, dass die rayleighverteilte Zufallsgröße η_2 größer ist als η_1 – siehe **Musterlösung zur Aufgabe ZA.17**:

$$\begin{aligned} \int_{\eta_1}^\infty p_{y_2|m}(\eta_2 | m_0) d\eta_2 &= e^{-\eta_1^2 / (2\sigma_n^2)} \\ \Rightarrow \Pr(\mathcal{E}) &= \int_0^\infty \eta_1 / \sigma_n^2 \cdot e^{-(2\eta_1^2 + E)/(2\sigma_n^2)} \cdot I_0 \left[\eta_1 \cdot \sqrt{E} / \sigma_n^2 \right] d\eta_1. \end{aligned}$$

(2) Mit den (willkürlichen) Substitutionen $C_0^2 = E/4$ und $\sigma_0^2 = \sigma_n^2/2$ erhält man daraus:

$$\begin{aligned} \Pr(\mathcal{E}) &= \int_0^\infty \frac{\eta_1}{2 \cdot \sigma_0^2} \cdot \exp \left[-\frac{2\eta_1^2 + 4C_0^2}{4\sigma_0^2} \right] \cdot I_0 \left[\frac{\eta_1 \cdot 2C_0}{2\sigma_0^2} \right] d\eta_1 = \\ &= \int_0^\infty \frac{\eta_1}{2 \cdot \sigma_0^2} \cdot \exp \left[-\frac{\eta_1^2 + 2C_0^2}{2\sigma_0^2} \right] \cdot I_0 \left[\frac{\eta_1 \cdot C_0}{\sigma_0^2} \right] d\eta_1. \end{aligned}$$

(3) Durch Verschieben von Anteilen vor das Integral gelingt es, dass der Integrand wieder eine Riceverteilung beschreibt:

$$\Pr(\mathcal{E}) = 1/2 \cdot \exp \left[-\frac{C_0^2}{2\sigma_0^2} \right] \cdot \int_0^\infty \frac{\eta_1}{\sigma_0^2} \cdot \exp \left[-\frac{\eta_1^2 + C_0^2}{2\sigma_0^2} \right] \cdot I_0 \left[\frac{\eta_1 \cdot C_0}{\sigma_0^2} \right] d\eta_1.$$

(4) Der Integrand beschreibt nun die Rice-WDF. Das Integral über das gesamte Definitionsgebiet von 0 bis $+\infty$ ergibt wie bei jeder WDF den Wert 1, so dass gilt:

$$\Pr(\mathcal{E}) = 1/2 \cdot \exp \left[-\frac{C_0^2}{2\sigma_0^2} \right].$$

(5) Mit $C_0^2 = E/4$, $\sigma_0^2 = \sigma_n^2/2$ und der allgemein gültigen Beziehung $\sigma_n^2 = N_0/2$ erhält man:

$$\Pr(\mathcal{E}) = 1/2 \cdot \exp \left[-\frac{C_0^2}{2\sigma_0^2} \right] = 1/2 \cdot \exp \left[-\frac{E_S/4}{N_0/2} \right]$$

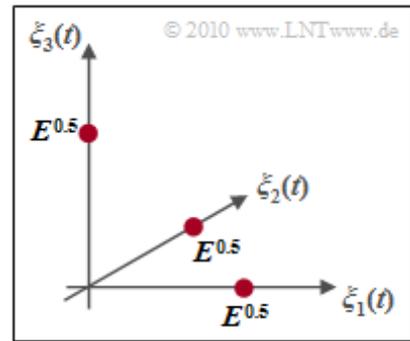
$$p_S = \Pr(\mathcal{E}) = 1/2 \cdot e^{-E_S/(2N_0)} \quad \mathbf{q.e.d.}$$

E_S gibt hierbei mittlere Signalenergie pro Symbol an, die bei FSK gleich der Normierungsenergie E ist.

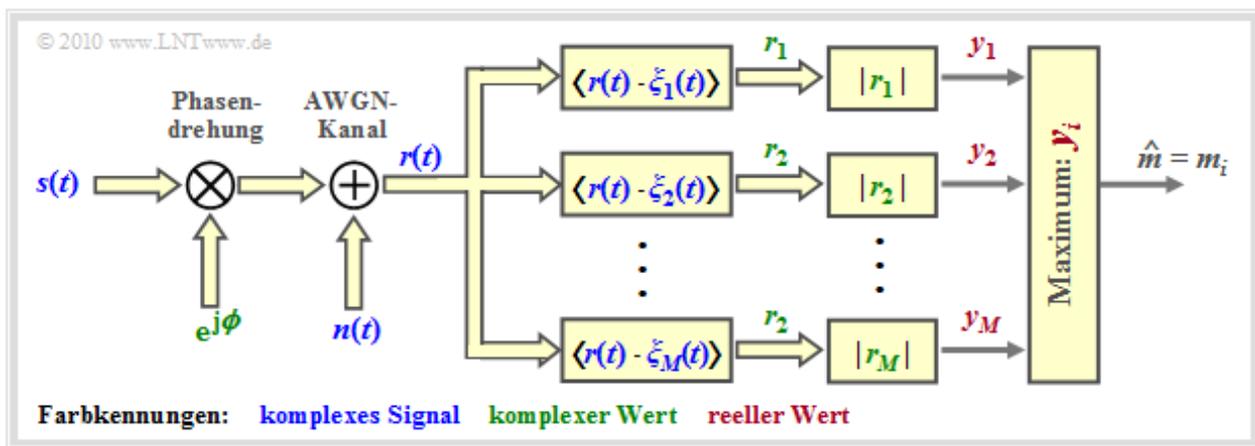
Nichtkohärente Demodulation von mehrstufiger FSK (1)

Wir betrachten nun die Nachrichtenmenge $\{m_0, m_1, \dots, m_{M-1}\}$ und bezeichnen M als Stufenzahl.

- Voraussetzung für die Anwendung des Modulationsverfahrens „*Frequency Shift Keying*“ und zugleich eines nichtkohärenten Demodulators ist ein ganzzahliger Modulationsindex h .
- In diesem Fall ist die M -stufige FSK orthogonal und es ergibt sich eine Signalraumkonstellation, wie in der oberen Grafik für den Sonderfall $M = 3$ dargestellt.



Der nichtkohärente Demodulator ist nachfolgend skizziert. Gegenüber der **Empfängerstruktur für binäre FSK** unterscheidet sich dieser Empfänger lediglich durch M Zweige anstelle von nur zweien, welche die Vergleichswerte y_1, y_2, \dots, y_M liefern.



Zur Berechnung der Fehlerwahrscheinlichkeit gehen wir wieder von der Annahme aus, dass m_0 gesendet wurde. Das bedeutet, dass die Entscheidung richtig ist, wenn y_1 der größte Detektionsausgangswert ist:

$$\begin{aligned} \Pr(\mathcal{C}) &= \Pr[(y_2 < y_1) \cap (y_3 < y_1) \cap \dots \cap (y_M < y_1) | m = m_0] = \\ &= \Pr\left[\bigcap_{k=2}^M (y_k < y_1) | m = m_0\right], \quad \Pr(\mathcal{E}) = 1 - \Pr(\mathcal{C}). \end{aligned}$$

Die Fehlerwahrscheinlichkeitsberechnung wird auf der nächsten Seite fortgesetzt.

Nichtkohärente Demodulation von mehrstufiger FSK (2)

Bitte beachten Sie: Es gibt gewisse Analogien zur Herleitung der **BFSK-Fehlerwahrscheinlichkeit**.

(1) Mit der bedingten Wahrscheinlichkeitsdichte $p_{y_1|m}(\eta_1|m_0)$ erhält man:

$$\Pr(\mathcal{C}) = \int_0^\infty \Pr \left[\bigcap_{k=2}^M (y_k < y_1) \mid y_1 = \eta_1, m = m_0 \right] \cdot p_{y_1|m}(\eta_1 | m_0) \, d\eta_1.$$

(2) Die Entscheidungswerte y_2, y_3, \dots, y_M sind bei gegebenem y_1 statistisch unabhängig. Deshalb gilt:

$$\Pr(\mathcal{C}) = \int_0^\infty \{ \Pr [(y_2 < y_1) \mid y_1 = \eta_1, m = m_0] \}^{M-1} \cdot p_{y_1|m}(\eta_1 | m_0) \, d\eta_1.$$

(3) Der ausgewählte Wert y_2 konditioniert auf m_0 besitzt eine Rayleighverteilung mit Parameter σ_n^2 :

$$\begin{aligned} \Pr [(y_2 < y_1) \mid y_1 = \eta_1, m = m_0] &= \int_0^{\eta_1} p_{y_2|m}(\eta_2 | m_0) \, d\eta_2 = \\ &= 1 - \exp \left[-\eta_1^2 / (2\sigma_n^2) \right] = 1 - a \quad (\text{Abkürzung}). \end{aligned}$$

(4) Gesucht ist nun der Ausdruck $(1 - a)^{M-1}$, für den gilt:

$$\begin{aligned} (1 - a)^{M-1} &= \sum_{i=0}^{M-1} (-1)^i \cdot \binom{M-1}{i} \cdot (-1)^i \cdot a^i = \\ &= \sum_{i=0}^{M-1} (-1)^i \cdot \binom{M-1}{i} \cdot \exp \left[-\frac{i \cdot \eta_1^2}{2\sigma_n^2} \right]. \end{aligned}$$

(5) Weiterhin besitzt y_1 konditioniert auf $m = m_0$ eine Riceverteilung. Die Wahrscheinlichkeit für eine korrekte Entscheidung lässt sich somit in folgende Form bringen:

$$\begin{aligned} \Pr(\mathcal{C}) &= \sum_{i=0}^{M-1} (-1)^i \cdot \binom{M-1}{i} \cdot \\ &\cdot \int_0^\infty \exp \left[-\frac{i \cdot \eta_1^2}{2\sigma_n^2} \right] \cdot \frac{\eta_1}{\sigma_n^2} \cdot \exp \left[-\frac{\eta_1^2 + E_S}{2\sigma_n^2} \right] \cdot I_0 \left[\frac{\eta_1 \cdot \sqrt{E_S}}{\sigma_n^2} \right] \, d\eta_1. \end{aligned}$$

(6) Durch Substitutionen gelingt es, den Integranden entsprechend der Riceverteilung zu gestalten. Da sich jede Wahrscheinlichkeitsdichte zu 1 integriert, erhält man:

$$\Pr(\mathcal{C}) = \sum_{i=0}^{M-1} (-1)^i \cdot \binom{M-1}{i} \cdot \frac{1}{i+1} \cdot \exp \left[-\frac{i \cdot E_S}{(i+1) \cdot N_0} \right], \quad E_S = E_B \cdot \log_2(M).$$

(7) Der Sonderfall $M = 2$ führt zum genau gleichen Ergebnis, wie für die binäre FSK berechnet:

$$\begin{aligned} \Pr(\mathcal{C}) &= (-1)^0 \cdot \binom{2-1}{0} \cdot \frac{1}{0+1} \cdot \exp \left[-\frac{0 \cdot E_S}{(0+1) \cdot N_0} \right] + \\ &+ (-1)^1 \cdot \binom{2-1}{1} \cdot \frac{1}{1+1} \cdot \exp \left[-\frac{1 \cdot E_S}{(1+1) \cdot N_0} \right] = \\ &= 1 - 1/2 \cdot e^{-E_S/(2N_0)} \quad \Rightarrow \quad \Pr(\mathcal{E}) = 1/2 \cdot e^{-E_S/(2N_0)}. \end{aligned}$$