

Überblick zu Kapitel 4 von „Einführung in die Informationstheorie“

Im Kapitel 4 werden die bisher für den wertdiskreten Fall definierten informationstheoretischen Größen derart adaptiert, dass sie auch für wertkontinuierliche Zufallsgrößen angewandt werden können. Aus der Entropie $H(X)$ für die wertdiskrete Zufallsgröße X wird so zum Beispiel im wertkontinuierlichen Fall die *differentielle Entropie* $h(X)$. Während $H(X)$ die Unsicherheit hinsichtlich der diskreten Zufallsgröße X angibt, kann $h(X)$ einer kontinuierlichen Zufallsgröße nicht in dieser Weise interpretiert werden.

Viele der im dritten Kapitel für die herkömmliche Entropie hergeleiteten Zusammenhänge gelten auch für die differentielle Entropie. So kann auch für wertkontinuierliche Zufallsgrößen X und Y die differentielle Verbundentropie $h(XY)$ angegeben werden und ebenso die bedingte differentielle Entropie $h(X|Y)$.

Im Einzelnen werden im Kapitel 4 behandelt:

- die Besonderheiten *wertkontinuierlicher Zufallsgrößen*,
- die Definition und Berechnung der *differentiellen Entropie* $h(X)$ sowie deren Eigenschaften,
- die *Transinformation* zwischen zwei wertkontinuierlichen Zufallsgrößen,
- die *Kapazität des AWGN-Kanals* und mehrerer solcher paralleler Gaußkanäle,
- das *Kanalcodierungstheorem*, eines der Highlights der Shannonschen Informationstheorie,
- die Kanalkapazität für wertdiskrete Eingangssignale (BPSK, QPSK).

Geeignete Literatur: [AM90] – [Bla87] – [CT06] – [Fan61] – [For72] – [Gal68] – [Har28] – [Joh92b] – [Kra13] – [McE77] – [Meck09] – [PS02] – [Sha48] – [WZ73]

Die grundlegende Theorie wird auf 29 Seiten dargelegt. Außerdem beinhaltet das Kapitel 76 Grafiken, zehn Aufgaben und sieben Zusatzaufgaben mit insgesamt 82 Teilaufgaben, sowie sechs Lernvideos und fünf Interaktionsmodule.

Zusammenstellung der **Lernvideos** (LV) zu den Grundlagen und zu Kapitel 4:

- **Wahrscheinlichkeit & WDF** (LV zu Kap. 4.1 – 2-tlg: Dauer 5:30 – 6:35)
- **Zusammenhang WDF – VTF** (LV zu Kap. 4.1 – 2-teilig: Dauer 6:40 – 3:20)
- **Momente von diskreten Zufallsgrößen** (LV zu Kap. 4.1 – Dauer 6:32)
- **Der AWGN-Kanal – Teil 1** (LV zu Kap. 4.2 und 4.3 – Dauer 6:00)
- **Der AWGN-Kanal – Teil 2** (LV zu Kap. 4.2 und 4.3 – Dauer 5:15)
- **Der AWGN-Kanal – Teil 3** (LV zu Kap. 4.2 und 4.3 – Dauer 6:15)

Zusammenstellung der **Interaktionsmodule** (IM) zu den Grundlagen und zu Kapitel 4:

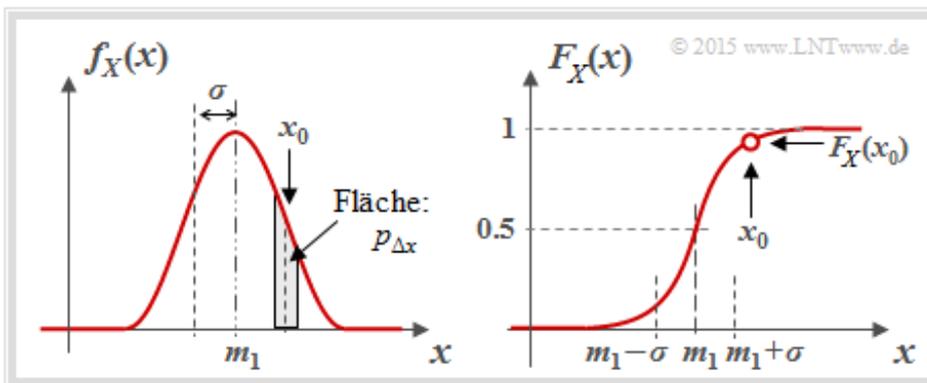
- **WDF, VTF und Momente spezieller Verteilungen** (IM zu Kap. 4.1)
- **WDF und VTF bei Gaußschen 2D-Zufallsgrößen** (IM zu Kap. 4.2)
- **Abtastung analoger Signale und Signalrekonstruktion** (IM zu Kap. 4.3)
- **Komplementäre Gaußsche Fehlerfunktionen** (IM zu Kap. 4.3)
- **M-PSK und Union Bound** (IM zu Kap. 4.3)

Eigenschaften wertkontinuierlicher Zufallsgrößen (1)

Bisher wurden stets *wertdiskrete Zufallsgrößen* der Form $X = \{x_1, x_2, \dots, x_\mu, \dots, x_M\}$ betrachtet, die aus informationstheoretischer Sicht vollständig durch ihre **Wahrscheinlichkeitsfunktion** (englisch: *Probability Mass Function*, PMF) $P_X(X)$ charakterisiert werden:

$$P_X(X) = [p_1, p_2, \dots, p_\mu, \dots, p_M] \quad \text{mit} \quad p_\mu = P_X(x_\mu) = \Pr(X = x_\mu).$$

Eine **wertkontinuierliche Zufallsgröße** kann dagegen – zumindest in endlichen Intervallen – jeden beliebigen Wert annehmen. Aufgrund des nicht abzählbaren Wertevorrats ist in diesem Fall die Beschreibung durch eine Wahrscheinlichkeitsfunktion nicht möglich oder zumindest nicht sinnvoll: Es ergäbe sich nämlich $M \rightarrow \infty$ sowie $p_1 \rightarrow 0, p_2 \rightarrow 0$, usw.



**Nomenklaturhinweise
zu WDF und VTF**

Man verwendet zur Beschreibung wertkontinuierlicher Zufallsgrößen gemäß den Definitionen im Buch „Stochastische Signaltheorie“ gleichermaßen (beachten Sie die Einträge in der Grafik):

- **Wahrscheinlichkeitsdichtefunktion** (WDF, englisch: *Probability Density Function*, PDF):

$$f_X(x_0) = \lim_{\Delta x \rightarrow 0} \frac{p_{\Delta x}}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Pr\{x_0 - \Delta x/2 \leq X \leq x_0 + \Delta x/2\}}{\Delta x}.$$

In Worten: Der WDF-Wert bei x_0 gibt die Wahrscheinlichkeit $p_{\Delta x}$ an, dass die Zufallsgröße X in einem (unendlich kleinen) Intervall der Breite Δx um x_0 liegt, dividiert durch Δx .

- **Mittelwert** (Moment erster Ordnung, englisch: *Mean Value* bzw. *Expectation Value*):

$$m_1 = E[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx.$$

- **Varianz** (Zentralmoment zweiter Ordnung, englisch: *Variance*):

$$\sigma^2 = E[(X - m_1)^2] = \int_{-\infty}^{+\infty} (x - m_1)^2 \cdot f_X(x - m_1) dx.$$

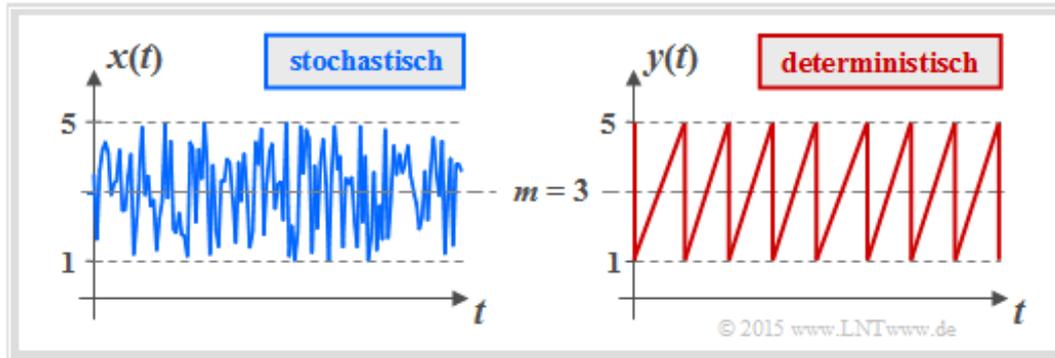
- **Verteilungsfunktion** (VTF, englisch: *Cumulative Distribution Function*, CDF):

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi = \Pr(X \leq x).$$

Beachten Sie, dass sowohl die WDF-Fläche als auch der VTF-Endwert stets gleich 1 sind.

Eigenschaften wertkontinuierlicher Zufallsgrößen (2)

Wir betrachten nun mit der **Gleichverteilung** einen wichtigen Sonderfall. Die Grafik zeigt den Verlauf zweier gleichverteilter Größen, die alle Werte zwischen 1 und 5 (Mittelwert $m_1 = 3$) mit gleicher Wahrscheinlichkeit annehmen können. Links ist das Ergebnis eines Zufallsprozesses dargestellt, rechts ein deterministisches Signal („Sägezahn“) mit gleicher Amplitudenverteilung.



Die *Wahrscheinlichkeitsdichtefunktion* der Gleichverteilung hat den unten skizzierten Verlauf:

$$f_X(x) = \begin{cases} (x_{\max} - x_{\min})^{-1} & \text{für } x_{\min} < x < x_{\max}, \\ (x_{\max} - x_{\min})^{-1}/2 & \text{für } x = x_{\min} \text{ und } x = x_{\max}, \\ 0 & \text{für } x > x_{\max}. \end{cases}$$

Es ergeben sich hier für den *Mittelwert* $m_1 = E[X]$ und die *Varianz* $\sigma^2 = E[(X - m_1)^2]$ folgende Gleichungen:

$$m_1 = \frac{x_{\max} + x_{\min}}{2}, \quad \sigma^2 = \frac{(x_{\max} - x_{\min})^2}{12}.$$

Unten dargestellt ist die *Verteilungsfunktion* (VTF):

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi = \Pr(X \leq x).$$

Diese ist für $x \leq x_{\min}$ identisch 0, steigt danach linear an und erreicht bei $x = x_{\max}$ den VTF-Endwert 1.

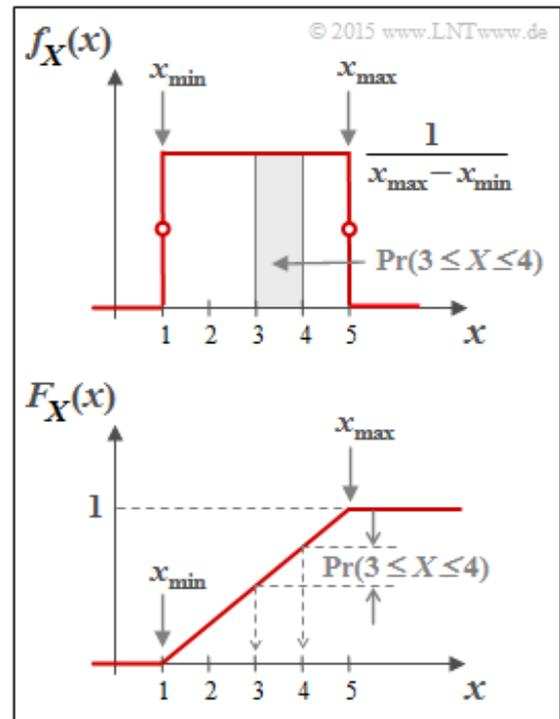
Die Wahrscheinlichkeit, dass die Zufallsgröße X einen Wert zwischen 3 und 4 annimmt, kann sowohl aus der WDF als auch aus der VTF ermittelt werden:

$$\Pr(3 \leq X \leq 4) = \int_3^4 f_X(\xi) d\xi = 0.25,$$

$$\Pr(3 \leq X \leq 4) = F_X(4) - F_X(3) = 0.25.$$

Weiterhin ist zu beachten:

- Das Ergebnis $X = 0$ ist bei dieser Zufallsgröße ausgeschlossen $\Rightarrow \Pr(X = 0) = 0$.
- Das Ergebnis $X = 4$ ist dagegen durchaus möglich. Trotzdem gilt auch hier $\Pr(X = 4) = 0$.



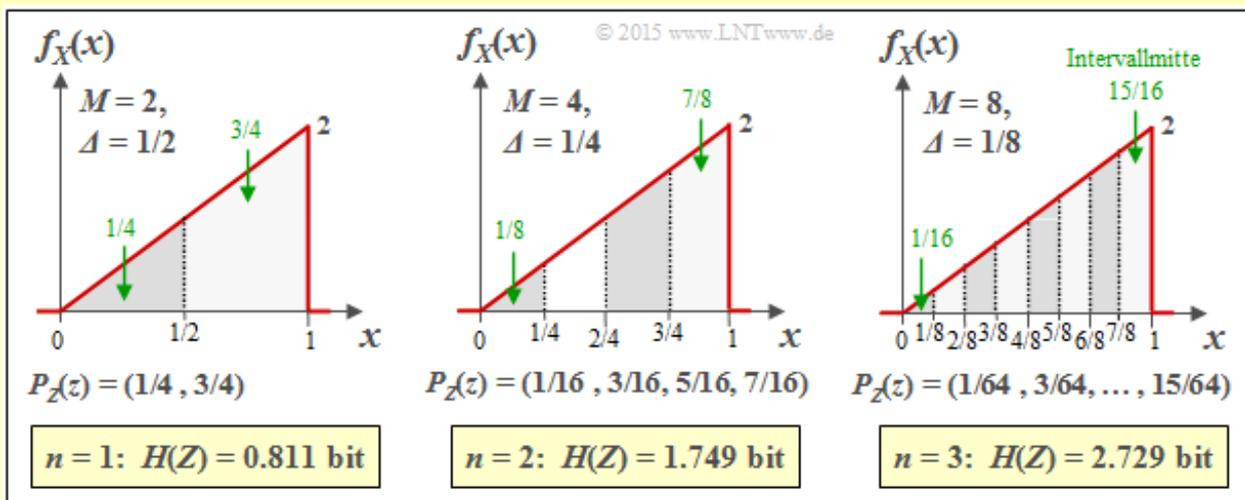
Entropie wertkontinuierlicher Zufallsgrößen nach Quantisierung (1)

Wir betrachten nun eine wertkontinuierliche Zufallsgröße X im Bereich von 0 bis 1.

- Wir quantisieren die kontinuierliche Zufallsgröße X , um die bisherige Entropieberechnung weiter anwenden zu können. Die so entstehende diskrete (quantisierte) Größe nennen wir Z .
- Die Quantisierungsstufenzahl sei M , so dass jedes Quantisierungsintervall μ bei der vorliegenden WDF die Breite $\Delta = 1/M$ aufweist. Die Intervallmitten bezeichnen wir mit x_μ .
- Die Wahrscheinlichkeit $p_\mu = \Pr(Z = z_\mu)$ bezüglich Z ist gleich der Wahrscheinlichkeit, dass die kontinuierliche Zufallsgröße X einen Wert zwischen $x_\mu - \Delta/2$ und $x_\mu + \Delta/2$ besitzt.
- Zunächst setzen wir $M = 2$ und verdoppeln anschließend M in jeder Iteration. Dadurch wird die Quantisierung zunehmend feiner. Im n -ten Versuch gilt dann $M = 2^n$ und $\Delta = 2^{-n}$.

Beispiel: Die Grafik zeigt die Ergebnisse der ersten drei Versuche für eine dreieckförmige WDF (zwischen 0 und 1):

- $n = 1 \Rightarrow M = 2 \Rightarrow \Delta = 1/2$: $H(Z) = 0.811$ bit,
- $n = 2 \Rightarrow M = 4 \Rightarrow \Delta = 1/4$: $H(Z) = 1.749$ bit,
- $n = 3 \Rightarrow M = 8 \Rightarrow \Delta = 1/8$: $H(Z) = 2.729$ bit.



Zudem können der Grafik noch folgende Größen entnommen werden, zum Beispiel für $\Delta = 1/8$:

- Die Intervallmitten liegen bei $x_1 = 1/16, x_2 = 3/16, \dots, x_8 = 15/16 \Rightarrow x_\mu = \Delta \cdot (\mu - 1/2)$.
- Die Intervallflächen ergeben sich zu $p_\mu = \Delta \cdot f_X(x_\mu) \Rightarrow p_8 = 1/8 \cdot (7/8+1)/2 = 15/64$.
- Damit erhält man $P_Z(Z) = (1/64, 3/64, 5/64, 7/64, 9/64, 11/64, 13/64, 15/64)$.

Die Ergebnisse dieses Experiments interpretieren wir wie folgt:

- Die Entropie $H(Z)$ nimmt mit steigendem M immer mehr zu.
- Der Grenzwert von $H(Z)$ für $M \rightarrow \infty \Rightarrow \Delta \rightarrow 0$ ist unendlich.
- Damit ist auch die Entropie $H(X)$ der wertkontinuierlichen Zufallsgröße X unendlich groß.
- Daraus folgt: Die bisherige Entropie-Definition versagt hier.

Entropie wertkontinuierlicher Zufallsgrößen nach Quantisierung (2)

Zur Verifizierung unseres empirischen Ergebnisses gehen wir von folgender Gleichung aus:

$$H(Z) = \sum_{\mu=1}^M p_{\mu} \cdot \log_2 \frac{1}{p_{\mu}} = \sum_{\mu=1}^M \Delta \cdot f_X(x_{\mu}) \cdot \log_2 \frac{1}{\Delta \cdot f_X(x_{\mu})}.$$

- Wir spalten nun $H(Z) = S_1 + S_2$ in zwei Summen auf:

$$S_1 = \log_2 \frac{1}{\Delta} \cdot \sum_{\mu=1}^M \Delta \cdot f_X(x_{\mu}) \approx -\log_2 \Delta,$$

$$S_2 = \sum_{\mu=1}^M f_X(x_{\mu}) \cdot \log_2 \frac{1}{f_X(x_{\mu})} \cdot \Delta \approx \int_0^1 f_X(x) \cdot \log_2 \frac{1}{f_X(x)} dx.$$

Die Näherung $S_1 \approx -\log_2 \Delta$ gilt exakt nur im Grenzfalle $\Delta \rightarrow 0$. Die angegebene Näherung für S_2 gilt ebenfalls nur für kleine $\Delta \rightarrow dx$, so dass man die Summe durch das Integral ersetzen kann.

Verallgemeinerung: Nähert man die wertkontinuierliche Zufallsgröße X mit der WDF $f_X(x)$ durch eine wertdiskrete Zufallsgröße Z an, indem man eine (feine) Quantisierung mit der Intervallbreite Δ durchführt, so erhält man für die Entropie der Zufallsgröße Z :

$$H(Z) \approx -\log_2 \Delta + \int_{\text{supp}(f_X)} f_X(x) \cdot \log_2 \frac{1}{f_X(x)} dx = -\log_2 \Delta + h(X) \quad [\text{in bit}].$$

Das Integral beschreibt die **differentielle Entropie** $h(X)$ der wertkontinuierlichen Zufallsgröße X . Für den Sonderfall $\Delta = 1/M = 2^{-n}$ kann die obige Gleichung auch wie folgt geschrieben werden:

$$H(Z) = n + h(X) \quad [\text{in bit}].$$

- Im Grenzfalle $\Delta \rightarrow 0 \Rightarrow M \rightarrow \infty \Rightarrow n \rightarrow \infty$ ist auch die Entropie der wertkontinuierlichen Zufallsgröße unendlich groß: $H(X) \rightarrow \infty$.
- Auch bei kleinerem n stellt diese Gleichung lediglich eine Näherung für $H(Z)$ dar, wobei die differentielle Entropie $h(X)$ der wertkontinuierlichen Größe als Korrekturfaktor dient.

Beispiel: Wir betrachten wie im **letzten Beispiel** eine Dreieck-WDF (zwischen 0 und 1). Deren differentielle Entropie ergibt sich zu $h(X) = -0.279$ bit – siehe **Aufgabe A4.2**. In der Tabelle ist die Entropie $H(Z)$ der mit n Bit quantisierten Größe Z angegeben. Man erkennt bereits für $n = 3$ eine gute Übereinstimmung zwischen der Näherung (untere Zeile) und der exakten Berechnung.

$H(Z)$ in bit	$n = 1$ ($M = 2$)	$n = 2$ ($M = 4$)	$n = 3$ ($M = 8$)	...	$n = 10$ ($M = 1024$)
Exakte Berechnung	0.811	1.749	2.729		—
Näherung: $n + h(X)$	0.721	1.721	2.721		9.721

© 2015 www.LNTwww.de

Definition und Eigenschaften der differentiellen Entropie (1)

Definition: Die **differentielle Entropie** $h(X)$ einer wertkontinuierlichen Zufallsgröße X lautet mit der Wahrscheinlichkeitsdichtefunktion $f_X(x)$:

$$h(X) = - \int_{\text{supp}(f_X)} f_X(x) \cdot \log [f_X(x)] dx \quad \text{mit} \quad \text{supp}(f_X) = \{x : f_X(x) > 0\}.$$

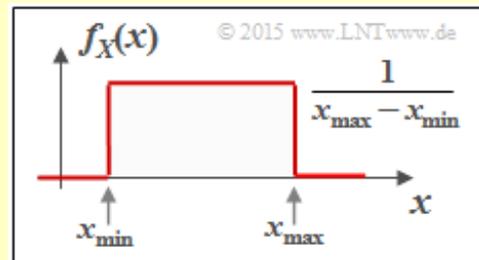
Hinzugefügt werden muss jeweils eine Pseudo-Einheit:

- „nat“ bei Verwendung von „ln“ \Rightarrow natürlicher Logarithmus,
- „bit“ bei Verwendung von „log₂“ \Rightarrow Logarithmus dualis.

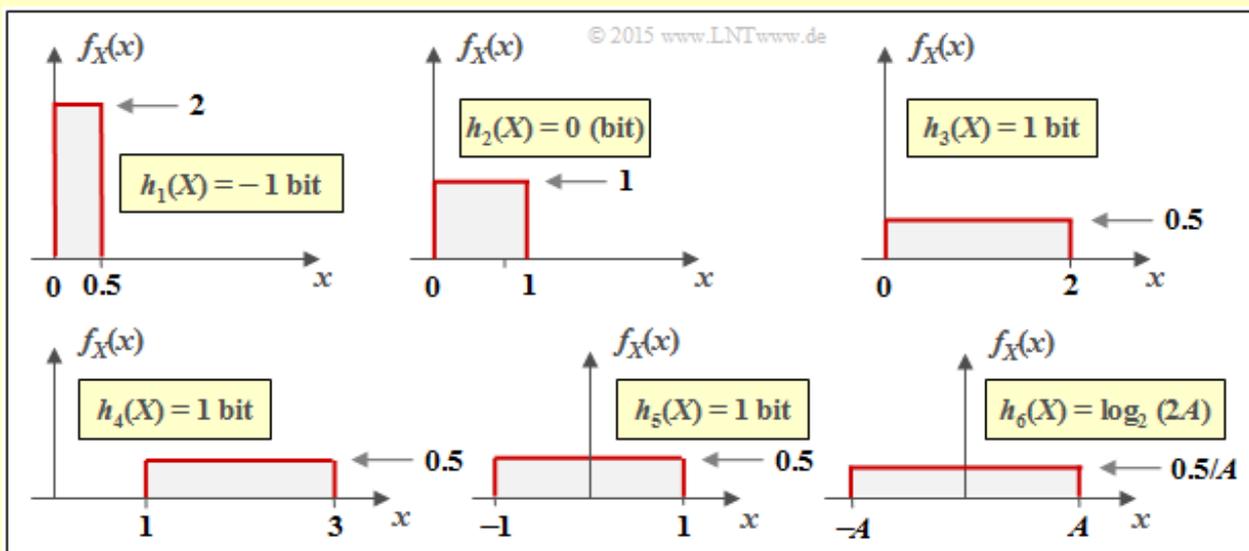
Während für die (herkömmliche) Entropie einer wertdiskreten Zufallsgröße X stets $H(X) \geq 0$ gilt, kann die differentielle Entropie $h(X)$ einer wertkontinuierlichen Zufallsgröße auch negativ sein. Daraus ist bereits ersichtlich, dass $h(X)$ im Gegensatz zu $H(X)$ nicht als „Unsicherheit“ interpretiert werden kann.

Beispiel: Die Grafik zeigt die Wahrscheinlichkeitsdichte einer zwischen x_{\min} und x_{\max} gleichverteilten Zufallsgröße X . Für deren differentielle Entropie erhält man in „nat“:

$$h(X) = - \int_{x_{\min}}^{x_{\max}} \frac{1}{x_{\max} - x_{\min}} \cdot \ln \left[\frac{1}{x_{\max} - x_{\min}} \right] dx = \dots = \ln [x_{\max} - x_{\min}] \cdot \left[\frac{1}{x_{\max} - x_{\min}} \right]_{x_{\min}}^{x_{\max}} = \ln [x_{\max} - x_{\min}].$$



Die Gleichung für die differentielle Entropie in „bit“ lautet: $h(X) = \log_2 [x_{\max} - x_{\min}]$.



Die Grafik zeigt anhand einiger Beispiele die numerische Auswertung des obigen Ergebnisses. Auf der nächsten Seite wird auf die Größen $h_1(X), \dots, h_6(X)$ näher eingegangen.

Definition und Eigenschaften der differentiellen Entropie (2)

Aus den **Skizzen des letzten Beispiels** lassen sich wichtige Eigenschaften der differentiellen Entropie $h(X)$ ablesen:

- Die differentielle Entropie wird durch eine WDF-Verschiebung (um k) nicht verändert:

$$h(X + k) = h(X) \quad \Rightarrow \quad h_3(X) = h_4(X) = h_5(X).$$

- $h(X)$ ändert sich durch Stauchung/Spreizung der WDF um den Faktor $k \neq 0$ wie folgt:

$$h(k \cdot X) = h(X) + \log_2 |k| \quad \Rightarrow \quad h_6(X) = h_5(AX) = h_5(X) + \log_2(A) = \log_2(2A).$$

Des Weiteren gelten viele der in **Kapitel 3** für den wertdiskreten Fall hergeleitete Gleichungen auch für wertkontinuierliche Zufallsgrößen. Aus der folgenden Zusammenstellung erkennt man, dass oft nur das „ H “ durch ein „ h “ sowie die PMF durch die entsprechende WDF zu ersetzen ist.

- Bedingte differentielle Entropie** (englisch: *Conditional Differential Entropy*):

$$H(X|Y) = E \left[\log \frac{1}{P_{X|Y}(X|Y)} \right] = \sum_{(x,y) \in \text{supp}(P_{XY})} P_{XY}(x,y) \cdot \log \frac{1}{P_{X|Y}(x|y)}$$

$$\Rightarrow h(X|Y) = E \left[\log \frac{1}{f_{X|Y}(X|Y)} \right] = \iint_{(x,y) \in \text{supp}(f_{XY})} f_{XY}(x,y) \cdot \log \frac{1}{f_{X|Y}(x|y)} dx dy.$$

- Differentielle Verbundentropie** (englisch: *Joint Differential Entropy*):

$$H(XY) = E \left[\log \frac{1}{P_{XY}(X,Y)} \right] = \sum_{(x,y) \in \text{supp}(P_{XY})} P_{XY}(x,y) \cdot \log \frac{1}{P_{XY}(x,y)}$$

$$\Rightarrow h(XY) = E \left[\log \frac{1}{f_{XY}(X,Y)} \right] = \iint_{(x,y) \in \text{supp}(f_{XY})} f_{XY}(x,y) \cdot \log \frac{1}{f_{XY}(x,y)} dx dy.$$

- Kettenregel** der differentiellen Entropie:

$$H(X_1 X_2 \dots X_n) = \sum_{i=1}^n H(X_i | X_1 X_2 \dots X_{i-1}) \leq \sum_{i=1}^n H(X_i)$$

$$\Rightarrow h(X_1 X_2 \dots X_n) = \sum_{i=1}^n h(X_i | X_1 X_2 \dots X_{i-1}) \leq \sum_{i=1}^n h(X_i).$$

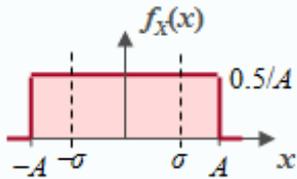
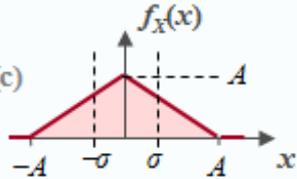
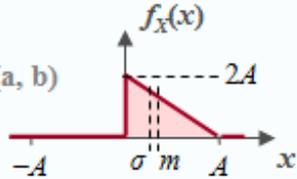
- Kullback–Leibler–Distanz** zwischen den Zufallsgrößen X und Y :

$$D(P_X || P_Y) = E \left[\log \frac{P_X(X)}{P_Y(X)} \right] = \sum_{x \in \text{supp}(P_X)} P_X(x) \cdot \log \frac{P_X(x)}{P_Y(x)} \geq 0$$

$$\Rightarrow D(f_X || f_Y) = E \left[\log \frac{f_X(X)}{f_Y(X)} \right] = \int_{x \in \text{supp}(f_X)} f_X(x) \cdot \log \frac{f_X(x)}{f_Y(x)} dx \geq 0.$$

Differentielle Entropie einiger spitzenwertbegrenzter Zufallsgrößen

Die Tabelle zeigt die Ergebnisse für drei beispielhafte Wahrscheinlichkeitsdichtefunktionen $f_X(x)$. Diese sind alle spitzenwertbegrenzt, das heißt, es gilt jeweils $|X| \leq A$.

<p>Gleichverteilung</p> <p>$h(X)$: siehe Seite 3a</p> <p>Varianz: $\sigma^2 = A^2/3$</p> 	<p>$h(X) = \log(2 \cdot A)$ $\Rightarrow \Gamma_A = 2.000$</p> <p>$h(X) = 1/2 \cdot \log(12 \cdot \sigma^2)$ $\Rightarrow \Gamma_L = 12.00$</p>
<p>Dreieckverteilung</p> <p>$h(X)$: siehe Aufgabe A4.2(c)</p> <p>$m_1 = 0, m_2 = A^2/6$ \rightarrow Varianz: $\sigma^2 = A^2/6$</p> 	<p>$h(X) = \log(e^{1/2} \cdot A)$ $\Rightarrow \Gamma_A = e^{1/2} = 1.649$</p> <p>$h(X) = 1/2 \cdot \log(6e \cdot \sigma^2)$ $\Rightarrow \Gamma_L = 6e = 16.31$</p>
<p>Einseitiges Dreieck</p> <p>$h(X)$: siehe Aufgabe A4.2(a, b)</p> <p>$m_1 = A/3, m_2 = A^2/6$ \rightarrow Varianz: $\sigma^2 = A^2/18$</p> 	<p>$h(X) = \log(e^{1/2}/2 \cdot A)$ $\Rightarrow \Gamma_A = e^{1/2}/2 = 0.824$</p> <p>$h(X) = 1/2 \cdot \log(4.5e \cdot \sigma^2)$ $\Rightarrow \Gamma_L = 4.5e = 12.23$</p>

WDF-Skizze nicht maßstabsgetreu!

© 2015 www.LNTwww.de

Bei *Spitzenwertbegrenzung* kann man die differentielle Entropie stets wie folgt darstellen:

$$h(X) = \log(\Gamma_A \cdot A).$$

Das Argument $\Gamma_A \cdot A$ ist unabhängig davon, welchen Logarithmus man verwendet. Anzufügen ist

- bei Verwendung von „ln“ ist die Pseudo-Einheit „nat“,
- bei Verwendung von „log₂“ ist die Pseudo-Einheit „bit“.

Theorem: Unter der Nebenbedingung **Spitzenwertbegrenzung** (englisch: *Peak Constraint*) \Rightarrow also WDF $f_X(x) = 0$ für $|x| > A$ – führt die **Gleichverteilung** zur maximalen differentielle Entropie:

$$h_{\max}(X) = \log(2A).$$

Beweis

Das Theorem bedeutet gleichzeitig, dass bei jeder anderen spitzenwertbegrenzten WDF (außer der Gleichverteilung) der Kennparameter Γ_A kleiner als 2 sein wird.

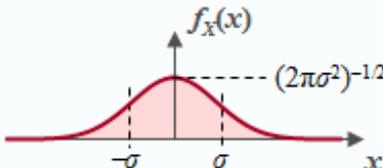
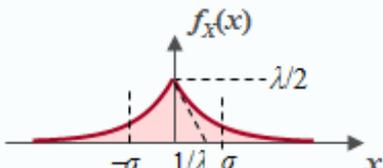
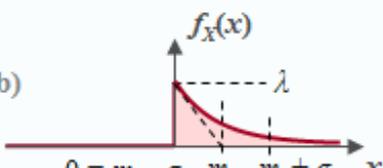
- Für die symmetrische Dreieckverteilung ergibt sich nach obiger Tabelle $\Gamma_A = e^{1/2} \approx 1.649$.
- Beim einseitigen Dreieck (zwischen 0 und A) ist demgegenüber Γ_A nur halb so groß.
- Auch für jedes andere Dreieck (Breite A, Spitze beliebig zwischen 0 und A) gilt $\Gamma_A \approx 0.824$.

Die jeweils zweite $h(X)$ -Angabe und die Kenngröße Γ_L eignet sich dagegen für den Vergleich von Zufallsgrößen bei Leistungsbegrenzung – siehe **nächste Seite**. Unter dieser Nebenbedingung ist die symmetrische Dreieckverteilung ($\Gamma_L \approx 16.31$) besser als die Gleichverteilung ($\Gamma_L = 12$).

Differentielle Entropie einiger leistungsbegrenzter Zufallsgrößen

Die differentielle Entropie $h(X)$ für drei beispielhafte Dichtefunktionen $f_X(x)$, die alle die gleiche Varianz $\sigma^2 = E[(X - m_x)^2] \Rightarrow$ Streuung σ aufweisen, sind der folgenden Tabelle zu entnehmen:

- *Gaußverteilung* \Rightarrow siehe Buch „Stochastische Signaltheorie“, **Kapitel 3.5**,
- *Laplaceverteilung* \Rightarrow siehe Buch „Stochastische Signaltheorie“, **Kapitel 3.6**,
- *Exponentialverteilung* \Rightarrow siehe Buch „Stochastische Signaltheorie“, **Kapitel 3.6**.

<p>Gaußverteilung</p> <p>$h(X)$: siehe Aufgabe A4.3 Streuung σ beim WDF– Wendepunkt ablesbar</p>	 <p style="font-size: small;">A plot of the Gaussian probability density function $f_X(x)$ versus x. The curve is symmetric and bell-shaped, centered at $x=0$. The standard deviation σ is indicated by dashed lines from the x-axis to the points where the curve's slope is zero (the inflection points). The peak of the curve is labeled with the value $(2\pi\sigma^2)^{-1/2}$.</p>	<p>$h(X) = 1/2 \cdot \log(2\pi e \cdot \sigma^2)$</p> <p>$\Rightarrow \Gamma_L = 2\pi e \approx 17.08$</p>
<p>Laplaceverteilung</p> <p>$h(X)$: siehe Aufgabe A4.2(c) Varianz σ^2 ist um den Faktor 2 größer als $1/\lambda^2$</p>	 <p style="font-size: small;">A plot of the Laplace probability density function $f_X(x)$ versus x. The curve is symmetric and has a sharp peak at $x=0$. The standard deviation σ is indicated by dashed lines from the x-axis to the points where the curve's slope is zero. The peak height is labeled $\lambda/2$.</p>	<p>$h(X) = 1/2 \cdot \log(2e^2 \cdot \sigma^2)$</p> <p>$\Rightarrow \Gamma_L = 2e^2 \approx 14.78$</p>
<p>Exponentialverteilung</p> <p>$h(X)$: siehe Aufgabe A4.2(a, b) $m_1 = 1/\lambda = \sigma$ (Streuung) Varianz: $\sigma^2 = 1/\lambda^2$</p>	 <p style="font-size: small;">A plot of the exponential probability density function $f_X(x)$ versus x. The curve starts at the origin and decays exponentially. The mean $m_1 = 1/\lambda$ is indicated by a dashed line from the x-axis to the point where the curve's slope is $-\lambda$. The x-axis is marked with $0 = m_1 - \sigma$, m_1, and $m_1 + \sigma$.</p>	<p>$h(X) = 1/2 \cdot \log(e^2 \cdot \sigma^2)$</p> <p>$\Rightarrow \Gamma_L = e^2 \approx 7.39$</p>

WDF–Skizze nicht maßstabsgetreu !

© 2015 www.LNTwww.de

Die differentielle Entropie lässt sich bei allen diesen Beispielen als

$$h(X) = 1/2 \cdot \log(\Gamma_L \cdot \sigma^2)$$

darstellen. Das Ergebnis unterscheidet sich nur durch die Pseudo–Einheit „nat“ bei Verwendung von „ln“ bzw. „bit“ bei Verwendung von „log₂“.

Theorem: Unter der Nebenbedingung der **Leistungsbegrenzung** (englisch: *Power Constraint*) führt die **Gaußverteilung**

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(x - m_1)^2}{2\sigma^2}\right]$$

unabhängig vom Mittelwert m_1 zur maximalen differentiellen Entropie:

$$h(X) = 1/2 \cdot \log(2\pi e \cdot \sigma^2).$$

Beweis

Dies bedeutet gleichzeitig, dass für jede andere WDF als die Gaußverteilung $\Gamma_L < 2\pi e \approx 17.08$ gelten muss. Beispielsweise ergibt sich der Kennwert $\Gamma_L = 6e \approx 16.31$ für die Dreieckverteilung, $\Gamma_L = 2e^2 \approx 14.78$ für die Laplaceverteilung und $\Gamma_L = 12$ für die Gleichverteilung.

WDF–Herleitung für maximale differentielle Entropie (1)

Beweis für Spitzenwertbegrenzung $\Rightarrow |X| \leq A$:

Unter der Nebenbedingung des Spitzenwertbegrenzung gilt für die differentielle Entropie:

$$h(X) = \int_{-A}^{+A} f_X(x) \cdot \log \frac{1}{f_X(x)} dx.$$

Von allen möglichen Wahrscheinlichkeitsdichtefunktionen $f_X(x)$, die die Bedingung

$$\int_{-A}^{+A} f_X(x) dx = 1$$

erfüllen, ist nun diejenige Funktion $g_X(x)$ gesucht, die zur maximalen differentiellen Entropie $h(X)$ führt.

Zur Herleitung benutzen wir das Verfahren der **Lagrange–Multiplikatoren**:

- Wir definieren die Lagrange–Kenngröße L in der Weise, dass darin sowohl $h(X)$ als auch die Nebenbedingung $|X| \leq A$ enthalten sind:

$$L = \int_{-A}^{+A} f_X(x) \cdot \log \frac{1}{f_X(x)} dx + \lambda \cdot \int_{-A}^{+A} f_X(x) dx.$$

- Wir setzen allgemein $f_X(x) = g_X(x) + \varepsilon \cdot \varepsilon_X(x)$, wobei $\varepsilon_X(x)$ eine beliebige Funktion darstellt, mit der Einschränkung, dass die WDF–Fläche gleich 1 sein muss. Damit erhalten wir:

$$L = \int_{-A}^{+A} [g_X(x) + \varepsilon \cdot \varepsilon_X(x)] \cdot \log \frac{1}{g_X(x) + \varepsilon \cdot \varepsilon_X(x)} dx + \lambda \cdot \int_{-A}^{+A} [g_X(x) + \varepsilon \cdot \varepsilon_X(x)] dx.$$

- Die bestmögliche Funktion ergibt sich dann, wenn es für $\varepsilon = 0$ eine stationäre Lösung gibt:

$$\left[\frac{dL}{d\varepsilon} \right]_{\varepsilon=0} = \int_{-A}^{+A} \varepsilon_X(x) \cdot \left[\log \frac{1}{g_X(x)} - 1 \right] dx + \lambda \cdot \int_{-A}^{+A} \varepsilon_X(x) dx \stackrel{!}{=} 0.$$

- Diese Bedingungsgleichung ist unabhängig von ε_X nur dann zu erfüllen, wenn gilt:

$$\log \frac{1}{g_X(x)} - 1 + \lambda = 0 \quad \forall x \in [-A, +A] \quad \Rightarrow \quad g_X(x) = \text{const.} \quad \forall x \in [-A, +A].$$

Resümee: Die maximale differentielle Entropie ergibt sich unter der **Nebenbedingung $|X| \leq A$** für die **gleichverteilte Zufallsgröße** (englisch: *Uniform PDF*):

$$h_{\max}(X) = \log(\Gamma_A \cdot A) = \log(2A) \quad \Rightarrow \quad \Gamma_A = 2.$$

Jede andere Zufallsgröße mit der WDF–Eigenschaft $f_X(|x| > A) = 0$ führt zu einer kleineren differentiellen Entropie, gekennzeichnet durch den Parameter $\Gamma_A < 2$.

WDF–Herleitung für maximale differentielle Entropie (2)

Beweis für Leistungsbegrenzung $\Rightarrow E[|X - m_1|^2] \leq \sigma^2$:

Vorneweg zur Begriffserklärung: Eigentlich wird nicht die Leistung \Rightarrow das **zweite Moment** m_2 begrenzt, sondern das **zweite Zentralmoment** \Rightarrow Varianz $\mu_2 = \sigma^2$. Lassen wir nur mittelwertfreie Zufallsgrößen zu, so umgehen wir das Problem. Damit lautet die Laplace–Kenngröße:

$$L = \int_{-\infty}^{+\infty} f_X(x) \cdot \log \frac{1}{f_X(x)} dx + \lambda_1 \cdot \int_{-\infty}^{+\infty} f_X(x) dx + \lambda_2 \cdot \int_{-\infty}^{+\infty} x^2 \cdot f_X(x) dx.$$

Nach ähnlichem Vorgehen wie im Fall der **Spitzenwertbegrenzung** erhält man das Ergebnis, dass die „bestmögliche“ WDF $g_X(x)$ proportional zu $\exp(-\lambda_2 \cdot x^2)$ sein muss \Rightarrow **Gaußverteilung**:

$$g_X(x) = 1/\sqrt{2\pi\sigma^2} \cdot \exp(-x^2/(2\sigma^2)).$$

Wir verwenden hier aber für den expliziten Beweis zur Abwechslung die **Kullback–Leibler–Distanz** zwischen einer geeigneten allgemeinen WDF $f_X(x)$ und der Gauß–WDF $g_X(x)$:

$$D(f_X || g_X) = \int_{-\infty}^{+\infty} f_X(x) \cdot \ln \frac{f_X(x)}{g_X(x)} dx = -h(X) - I_2$$
$$\Rightarrow I_2 = \int_{-\infty}^{+\infty} f_X(x) \cdot \ln g_X(x) dx.$$

Zur Vereinfachung wurde hier der natürliche Logarithmus verwendet. Damit erhalten wir:

$$I_2 = -\frac{1}{2} \cdot \ln(2\pi\sigma^2) \cdot \int_{-\infty}^{+\infty} f_X(x) dx - \frac{1}{2\sigma^2} \cdot \int_{-\infty}^{+\infty} x^2 \cdot f_X(x) dx.$$

Das erste Integral ist definitionsgemäß gleich 1 und das zweite Integral ergibt σ^2 :

$$I_2 = -1/2 \cdot \ln(2\pi\sigma^2) - 1/2 \cdot [\ln(e)] = -1/2 \cdot \ln(2\pi e \cdot \sigma^2)$$
$$\Rightarrow D(f_X || g_X) = -h(X) - I_2 = -h(X) + 1/2 \cdot \ln(2\pi e \cdot \sigma^2).$$

Da auch bei wertkontinuierlichen Zufallsgrößen die Kullback–Leibler–Distanz größer oder gleich 0 ist, erhält man nach Verallgemeinerung ($\ln \Rightarrow \log$):

$$h(X) \leq 1/2 \cdot \log(2\pi e \cdot \sigma^2).$$

Das Gleichzeichen gilt nur, wenn die Zufallsgröße X gaußverteilt ist.

Resümee: Die maximale differentielle Entropie unter der **Nebenbedingung** $|X - m_1|^2 \leq \sigma^2$ ergibt sich für die **Gaußverteilung** (englisch: *Gaussian PDF*) unabhängig vom Mittelwert m_1 :

$$h_{\max}(X) = 1/2 \cdot \log(\Gamma_L \cdot \sigma^2) = 1/2 \cdot \log(2\pi e \cdot \sigma^2) \Rightarrow \Gamma_L = 2\pi e.$$

Jede andere wertkontinuierliche Zufallsgröße X mit Varianz $E[|X - m_1|^2] \leq \sigma^2$ führt zu einer kleineren differentiellen Entropie, gekennzeichnet durch die Kenngröße $\Gamma_L < 2\pi e$.

Transinformation zwischen wertkontinuierlichen Zufallsgrößen (1)

Im **Kapitel 3.3** wurde die Transinformation (englisch: *Mutual Information*) zwischen den beiden wertdiskreten Zufallsgrößen X und Y unter Anderem in folgender Form angegeben:

$$I(X; Y) = \sum_{(x,y) \in \text{supp}(P_{XY})} P_{XY}(x, y) \cdot \log \frac{P_{XY}(x, y)}{P_X(x) \cdot P_Y(y)}.$$

Diese Gleichung entspricht gleichzeitig der **Kullback–Leibler–Distanz** (kurz KLD) zwischen der Verbundwahrscheinlichkeitsfunktion P_{XY} und dem Produkt der beiden Einzel-PMFs P_X und P_Y :

$$I(X; Y) = D(P_{XY} || P_X \cdot P_Y).$$

Um daraus die Transinformation $I(X; Y)$ zwischen zwei wertkontinuierlichen Zufallsgrößen X und Y abzuleiten, geht man wie folgt vor (Hochkommata weisen auf quantisierte Größen hin):

- Man quantisiert die Zufallsgrößen X und Y (mit den Quantisierungsintervallen Δ_x und Δ_y) und erhält so die Wahrscheinlichkeitsfunktionen $P_{X'}$ und $P_{Y'}$.
- Die „Vektoren“ $P_{X'}$ und $P_{Y'}$ werden nach den Grenzübergängen $\Delta_x \rightarrow 0$, $\Delta_y \rightarrow 0$ unendlich lang, und auch die Verbund-PMF $P_{X'Y'}$ ist in der Fläche unendlich weit ausgedehnt.
- Durch diese Grenzübergänge ergeben sich die Wahrscheinlichkeitsdichtefunktionen der drei kontinuierlichen Zufallsgrößen entsprechend den folgenden Gleichungen:

$$f_X(x_\mu) = \frac{P_{X'}(x_\mu)}{\Delta_x}, \quad f_Y(y_\mu) = \frac{P_{Y'}(y_\mu)}{\Delta_y}, \quad f_{XY}(x_\mu, y_\mu) = \frac{P_{X'Y'}(x_\mu, y_\mu)}{\Delta_x \cdot \Delta_y}.$$

- Aus der Doppelsumme in der obigen Gleichung wird nach der Umbenennung $\Delta_x \rightarrow d_x$ bzw. $\Delta_y \rightarrow d_y$ die für wertkontinuierliche Zufallsgrößen gültige Gleichung:

$$I(X; Y) = \iint_{(x,y) \in \text{supp}(f_{XY})} f_{XY}(x, y) \cdot \log \frac{f_{XY}(x, y)}{f_X(x) \cdot f_Y(y)} dx dy.$$

Durch Aufspaltung dieses Doppelintegrals lässt für die Transinformation auch schreiben:

$$I(X; Y) = h(X) + h(Y) - h(XY).$$

Verwendet ist hierbei die *differentielle Verbund-Entropie*

$$h(XY) = - \iint_{(x,y) \in \text{supp}(f_{XY})} f_{XY}(x, y) \cdot \log [f_{XY}(x, y)] dx dy$$

sowie die beiden *differentiellen Einzel-Entropien*

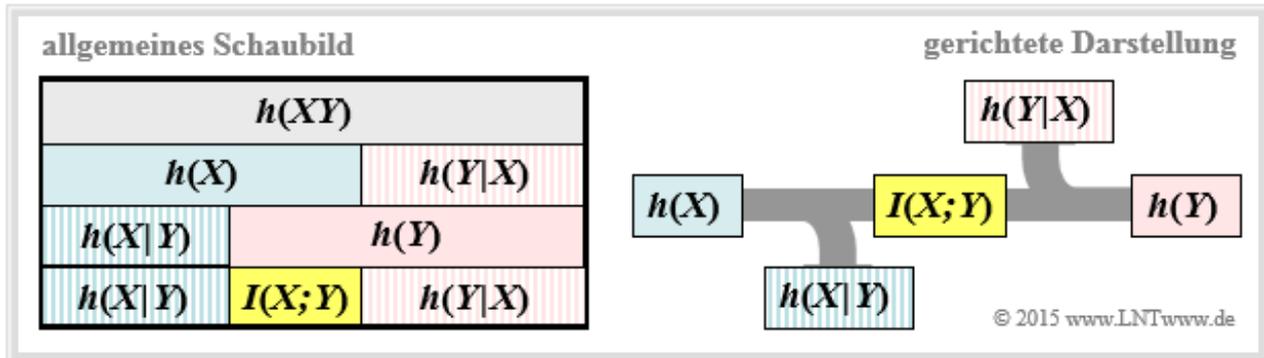
$$h(X) = - \int_{x \in \text{supp}(f_X)} f_X(x) \cdot \log [f_X(x)] dx, \quad h(Y) = - \int_{y \in \text{supp}(f_Y)} f_Y(y) \cdot \log [f_Y(y)] dy.$$

Transinformation zwischen wertkontinuierlichen Zufallsgrößen (2)

Wir gehen weiter von der wertkontinuierlichen Transinformationsgleichung aus:

$$I(X; Y) = h(X) + h(Y) - h(XY).$$

Diese Darstellung findet sich auch im folgenden Schaubild (linke Grafik, über alle Zeilen).



Daraus erkennt man, dass die Transinformation auch noch wie folgt dargestellt werden kann:

$$I(X; Y) = h(Y) - h(Y|X) = h(X) - h(X|Y).$$

Diese fundamentalen informationstheoretischen Zusammenhänge kann man auch aus der rechten Grafik ablesen. Diese gerichtete Darstellung ist für Nachrichtenübertragungssysteme besonders geeignet. Die abfließende bzw. zufließende differentielle Entropie kennzeichnet

- die **Äquivokation** (englisch: *Equivocation*):

$$h(X|Y) = - \iint_{(x,y) \in \text{supp}(f_{XY})} f_{XY}(x, y) \cdot \log [f_{X|Y}(x|y)] \, dx \, dy,$$

- die **Irrelevanz** (englisch: *Irrelevance*):

$$h(Y|X) = - \iint_{(x,y) \in \text{supp}(f_{XY})} f_{XY}(x, y) \cdot \log [f_{Y|X}(y|x)] \, dx \, dy.$$

Auf die Bedeutung dieser beiden informationstheoretischen Größen wird in **Aufgabe Z4.5** noch genauer eingegangen. Vergleicht man die grafischen Darstellungen der Transinformation bei

- wertdiskreten Zufallsgrößen im **Kapitel 3.3** und
- wertkontinuierlichen Zufallsgrößen entsprechend obiger Grafik,

so erkennt man als einziges Unterscheidungsmerkmal, dass jedes „ H “ (Entropie; größer/gleich Null) durch ein „ h “ (differentielle Entropie, kann positiv, negativ oder 0 sein) ersetzt wurde. Ansonsten ist die Transinformation in beiden Darstellungen gleich und es gilt stets $I(X; Y) \geq 0$.

Im Folgenden verwenden wir meist den *Logarithmus dualis* \Rightarrow „ \log_2 “ und erhalten somit die Transinformation in „bit“.

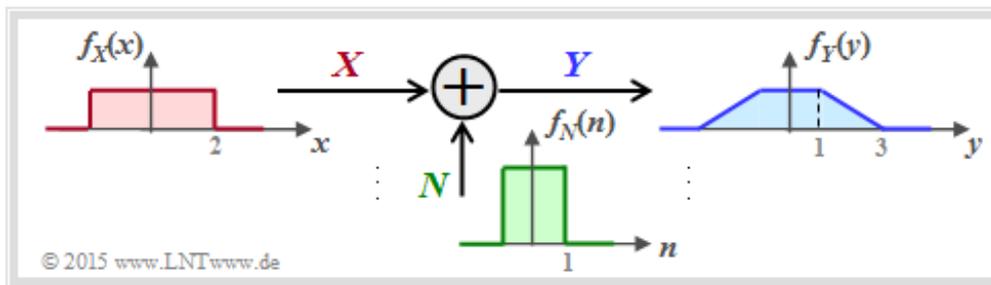
Transinformationsberechnung bei additiver Störung

Wir betrachten nun ein sehr einfaches Modell der Nachrichtenübertragung.

- Die Zufallsgröße X steht für das (mittelwertfreie) Sendesignal und ist durch die WDF $f_X(x)$ und die Varianz σ_X^2 gekennzeichnet. Die Sendeleistung P_X ist identisch mit σ_X^2 .
- Die additive Störung N ist durch die WDF $f_N(n)$ und die Störleistung $P_N = \sigma_N^2$ gegeben. Da X und N als statistisch unabhängig angenommen werden, gilt $E[X \cdot N] = E[X] \cdot E[N] = 0$.
- Das Empfangssignal ist $Y = X + N$. Die Ausgangs-WDF $f_Y(y)$ ist mit der **Faltungsoption** berechenbar $\Rightarrow f_Y(y) = f_X(x) * f_N(n)$ und für die Empfangsleistung (Varianz) gilt:

$$P_Y = \sigma_Y^2 = E[Y^2] = E[(X + N)^2] = E[X^2] + E[N^2] = \sigma_X^2 + \sigma_N^2 = P_X + P_N.$$

Die in der nachfolgenden Grafik eingezeichneten Dichtefunktionen (rechteck- bzw. trapezförmig) sollen nur den Rechengang verdeutlichen und haben keine praktische Relevanz.



Zur Berechnung der Transinformation zwischen dem Eingang X und dem Ausgang Y gibt es entsprechend dem **Schaubild** auf der vorherigen Seite drei Möglichkeiten:

$I(X, Y) = h(X) + h(Y) - h(XY)$:

Die beiden ersten Terme sind aus $f_X(x)$ bzw. $f_Y(y)$ in einfacher Weise berechenbar. Problematisch ist die differentielle Verbundentropie $h(XY)$. Hierzu benötigt man die 2D-Verbund-WDF $f_{XY}(x, y)$, die meist nicht direkt gegeben ist.

$I(X, Y) = h(Y) - h(Y|X)$:

$h(Y|X)$ bezeichnet die differentielle Streuentropie. Es gilt $h(Y|X) = h(X + N|X) = h(N)$, so dass $I(X; Y)$ bei Kenntnis von $f_X(x)$ und $f_N(n)$ über die Gleichung $f_Y(y) = f_X(x) * f_N(n)$ sehr einfach zu berechnen ist.

$I(X, Y) = h(X) - h(X|Y)$:

Nach dieser Gleichung benötigt man allerdings die differentielle Rückschlusentropie $h(X|Y)$, die schwieriger angebar ist als $h(Y|X)$.

Resümee: Im Folgenden schreiben wir für die **Transinformation** zwischen dem Eingang X und dem Ausgang Y eines **Nachrichtenübertragungssystems** bei additiver und unkorrelierter Störung N :

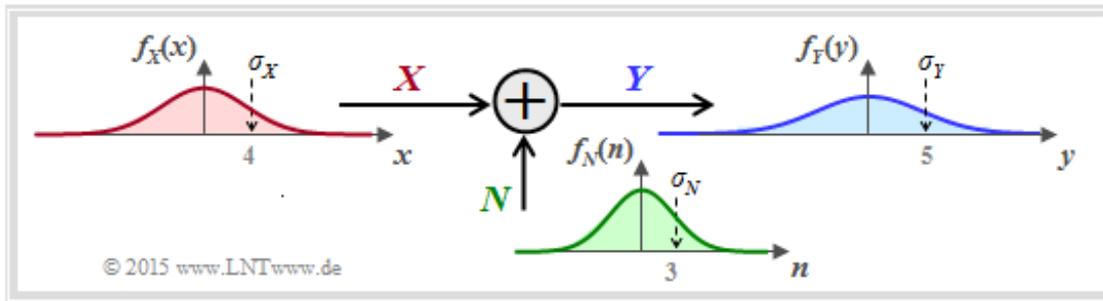
$$I(X; Y) = h(Y) - h(N) = - \int_{y \in \text{supp}(f_Y)} f_Y(y) \cdot \log [f_Y(y)] dy + \int_{n \in \text{supp}(f_N)} f_N(n) \cdot \log [f_N(n)] dn.$$

Kanalkapazität des AWGN-Kanals

Spezifiziert man im bisherigen **Systemmodell** die Wahrscheinlichkeitsdichtefunktion der Störung (bzw. des Rauschens) zu

$$f_N(n) = \frac{1}{\sqrt{2\pi\sigma_N^2}} \cdot \exp\left[-\frac{n^2}{2\sigma_N^2}\right] \quad \text{mit} \quad \sigma_N^2 = P_N,$$

so erhalten wir das folgende Modell zur Berechnung der Kanalkapazität des **AWGN-Kanals** (*Additive White Gaussian Noise*).



Die **Kanalkapazität** C gibt die maximale Transinformation $I(X; Y)$ zwischen der Eingangsgröße X und der Ausgangsgröße Y des AWGN-Kanals an. Die Maximierung bezieht sich dabei auf die bestmögliche Eingangs-WDF. Somit gilt unter der Nebenbedingung der **Leistungsbegrenzung**:

$$C = \max_{f_X: E[X^2] \leq P_X} I(X; Y) = -h(N) + \max_{f_X: E[X^2] \leq P_X} h(Y).$$

Hier ist bereits berücksichtigt, dass sich die Maximierung allein auf $h(Y) \Rightarrow$ WDF $f_Y(y)$ bezieht. Bei gegebener Störleistung P_N ist nämlich $h(N) = 1/2 \cdot \log_2(2\pi e \cdot P_N)$ eine Konstante.

- Das Maximum für $h(Y)$ erhält man für eine Gaußsche WDF $f_Y(y)$, wobei $P_Y = P_X + P_N$ zu setzen ist, siehe **Kapitel 4.1** $\Rightarrow \text{Max}[h(Y)] = 1/2 \cdot \log_2[2\pi e \cdot (P_X + P_N)]$.
- Die Ausgangs-WDF $f_Y(y) = f_X(x) * f_N(n)$ ist aber nur dann gaußförmig, wenn sowohl $f_X(x)$ als auch $f_N(n)$ Gaußfunktionen sind. Ein plakativer Merkspruch zur Faltungsoperation lautet nämlich: „*Gauß bleibt Gauß, und Nicht-Gauß wird nie (exakt) Gauß*“.

Resümee: Beim **AWGN-Kanal** \Rightarrow Gaußsche WDF $f_N(n)$ ergibt sich die **Kanalkapazität** C genau dann, wenn die Eingangs-WDF $f_X(x)$ ebenfalls gaußförmig ist:

$$C_{\text{AWGN}} = h_{\text{max}}(Y) - h(N) = 1/2 \cdot \log_2 \frac{P_Y}{P_N} = 1/2 \cdot \log_2 \left(1 + \frac{P_X}{P_N}\right).$$

Die folgende Tabelle zeigt die AWGN-Kanalkapazität für verschiedene Quotienten P_X/P_N .

P_X/P_N	0	5	10	15	20	25	30
C_{AWGN} in bit	0	1.292	1.730	2.000	2.196	2.350	2.477

© 2015 www.LNTwww.de

Parallele Gaußsche Kanäle (1)

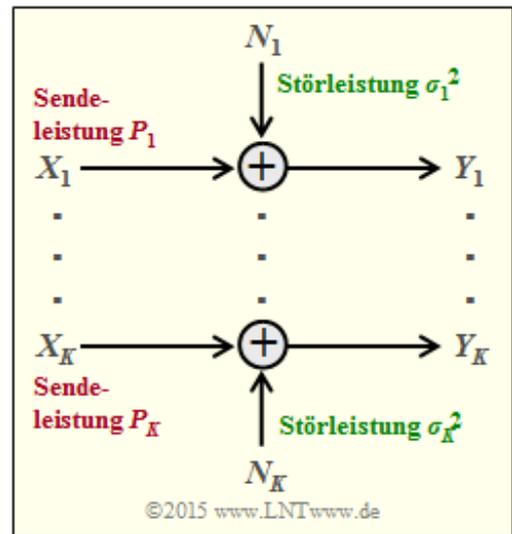
Wir betrachten nun K parallele Gaußkanäle entsprechend der nebenstehenden Grafik

- von $X_1 \rightarrow Y_1, \dots, \text{von } X_k \rightarrow Y_k, \dots, \text{von } X_K \rightarrow Y_K$.
- Die Sendeleistungen in den K Kanälen nennen wir $P_1 = E[X_1^2], \dots, P_k = E[X_k^2], \dots, P_K = E[X_K^2]$.
- Die K Störleistungen $\sigma_1^2, \dots, \sigma_k^2, \dots, \sigma_K^2$ können ebenfalls unterschiedlich sein.

Gesucht ist nun die maximale Transinformation

$$I(X_1, \dots, X_K; Y_1, \dots, Y_K)$$

- zwischen den K Eingangsgrößen X_1, \dots, X_K , sowie
- den K Ausgangsgrößen Y_1, \dots, Y_K ,



die wir als *Gesamt-Kanalkapazität* dieser AWGN-Konfiguration bezeichnen. Dabei gehen wir von Leistungsbegrenzung des Gesamtsystems aus, das heißt, dass die Summe aller Leistungen P_k in den K Einzelkanälen den vorgegebenen Wert P_X nicht überschreiten darf:

$$P_1 + \dots + P_K = \sum_{k=1}^K E[X_k^2] \leq P_X.$$

Unter der nur wenig einschränkenden Annahme unabhängiger Störquellen N_1, \dots, N_K kann für die Transinformation nach einigen Zwischenschritten geschrieben werden:

$$I(X_1, \dots, X_K; Y_1, \dots, Y_K) = h(Y_1, \dots, Y_K) - \sum_{k=1}^K h(N_k).$$

Dafür kann folgende obere Schranke angegeben werden:

$$I(X_1, \dots, X_K; Y_1, \dots, Y_K) \leq \sum_{k=1}^K [h(Y_k) - h(N_k)] \leq 1/2 \cdot \sum_{k=1}^K \log_2 \left(1 + \frac{P_k}{\sigma_k^2} \right).$$

- Das Gleichheitszeichen (Identität) gilt bei mittelwertfreien Gaußschen Eingangsgrößen X_k sowie bei statistisch voneinander unabhängigen Störungen N_k .
- Man kommt von dieser Gleichung zur *maximalen Transinformation* \Rightarrow *Kanalkapazität*, wenn man die gesamte Sendeleistung P_X unter Berücksichtigung der unterschiedlichen Störungen in den einzelnen Kanälen (σ_k^2) bestmöglich aufteilt.
- Dieses Optimierungsproblem lässt sich wieder mit dem Verfahren der *LagrangeMultiplikatoren* elegant lösen. Das Beispiel auf der nächsten Seite erläutert nur das Ergebnis.

Parallele Gaußsche Kanäle (2)

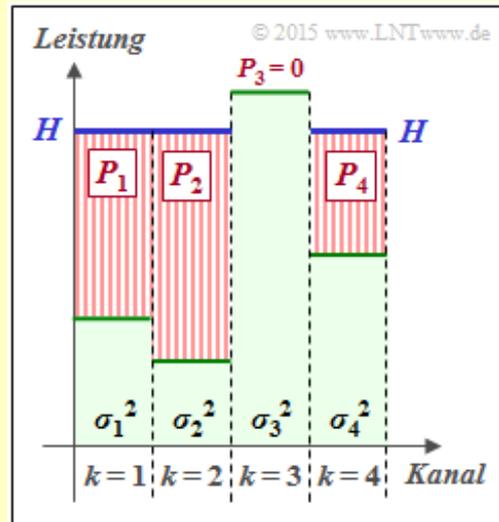
Beispiel: Wir betrachten $K = 4$ parallele Gaußkanäle mit unterschiedlichen Störleistungen ($\sigma_1^2, \dots, \sigma_4^2$) gemäß der nebenstehenden Abbildung. Gesucht ist die bestmögliche Aufteilung der Sendeleistung auf die vier Kanäle.

Würde man dieses Profil langsam mit Wasser auffüllen, so würde das Wasser zunächst nur in den Kanal 2 fließen. Gießt man weiter, so sammelt sich auch im Kanal 1 etwas Wasser an und später auch im Kanal 4.

Die eingezeichnete „Wasserhöhe“ H beschreibt genau den Zeitpunkt, zu dem die Summe ($P_1 + P_2 + P_4$) der insgesamt zur Verfügung stehenden Sendeleistung P_X entspricht.

Die optimale Leistungsaufteilung für dieses Beispiel ergibt $P_2 > P_1 > P_4$ sowie $P_3 = 0$. Erst bei größerer Sendeleistung P_X würde auch dem dritten Kanal eine kleine Leistung P_3 zugewiesen.

Man bezeichnet dieses Allokationsverfahren als **Water-Filling-Algorithmus**.



Werden alle K Gaußkanäle in gleicher Weise gestört $\Rightarrow \sigma_1^2 = \dots = \sigma_K^2 = P_N$, so sollte man natürlich die gesamte zur Verfügung stehende Sendeleistung P_X gleichmäßig auf alle Kanäle verteilen: $P_k = P_X/K$. Für die Gesamtkapazität erhält man dann:

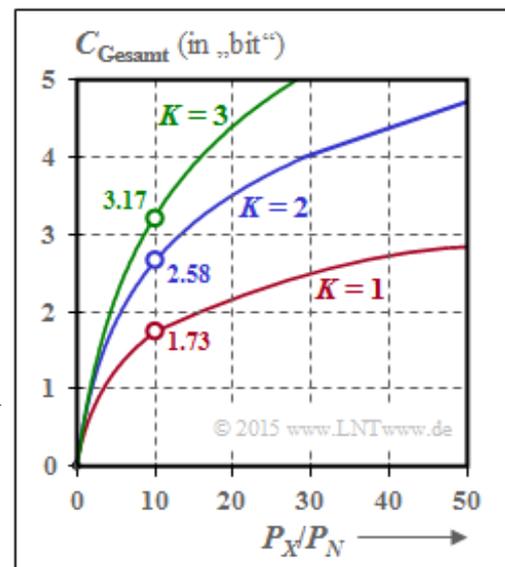
$$C_{\text{Gesamt}} = \frac{K}{2} \cdot \log_2 \left(1 + \frac{P_X}{K \cdot P_N} \right).$$

Die rechte Grafik zeigt die Gesamtkapazität als Funktion des Quotienten P_X/P_N für $K = 1, K = 2$ und $K = 3$.

- Bei $P_X/P_N = 10 \Rightarrow 10 \cdot \lg(P_X/P_N) = 10$ dB wird die Gesamtkapazität um ca. 50% größer, wenn man die Gesamtleistung auf zwei Kanäle gleichmäßig aufteilt.
- Im Grenzfall $P_X/P_N \rightarrow \infty$ nimmt die Gesamtkapazität um den Faktor K zu \Rightarrow Verdoppelung mit $K = 2$.

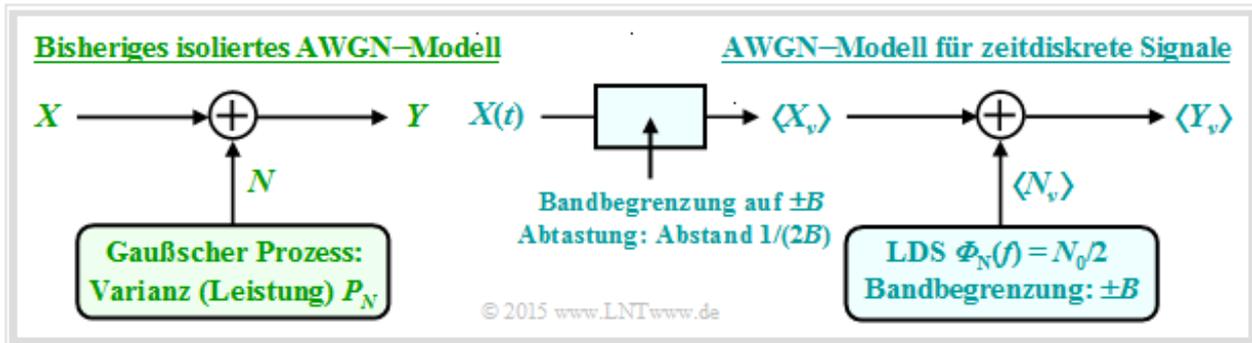
Die beiden identischen und voneinander unabhängigen Kanäle kann man auf unterschiedliche Weise realisieren, zum Beispiel durch Zeit-, Frequenz- oder Raummultiplexverfahren.

Der Fall $K = 2$ lässt sich aber auch durch die Verwendung orthogonaler Basisfunktionen wie „Cosinus“ und „Sinus“ verwirklichen wie zum Beispiel bei der **Quadratur-Amplitudenmodulation (QAM)** oder einer **mehrstufigen Phasenmodulation** wie QPSK oder 8-PSK.



AWGN-Modell für zeitdiskrete bandbegrenzte Signale (1)

Am Ende von Kapitel 4.2 wurde das AWGN-Modell entsprechend der linken Grafik verwendet, gekennzeichnet durch die beiden Zufallsgrößen X und Y am Eingang und Ausgang sowie die stochastische Störung N als das Ergebnis eines mittelwertfreien Gaußschen Zufallsprozesses \Rightarrow „Weißes Rauschen“ mit der Varianz σ_N^2 . Die Störleistung P_N ist ebenfalls gleich σ_N^2 .

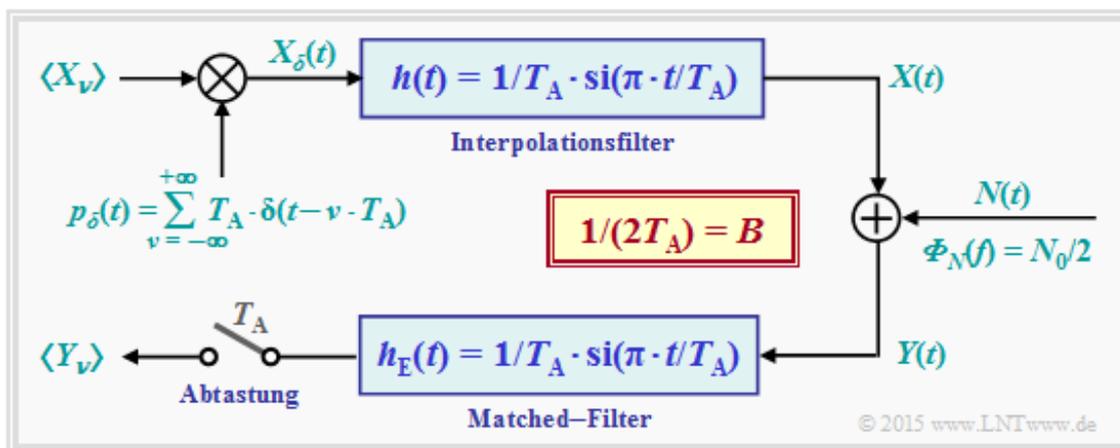


Die maximale Transformation $I(X; Y)$ zwischen Eingang und Ausgang \Rightarrow Kanalkapazität C ergibt sich dann, wenn eine Gaußsche Eingangs-WDF $f_X(x)$ vorliegt. Mit der Sendeleistung $P_X = \sigma_X^2$ (Varianz der Zufallsgröße X) lautet die Kanalkapazitätsgleichung:

$$C = 1/2 \cdot \log_2 (1 + P_X/P_N).$$

Nun beschreiben wir das AWGN-Kanalmodell gemäß dem rechts skizzierten Fall, dass am Kanaleingang die Folge $\langle X_v \rangle$ anliegt, wobei der Abstand zwischen aufeinander folgenden Werten T_A beträgt. Diese Folge ist das zeitdiskrete Äquivalent des zeitkontinuierlichen Signals $X(t)$ nach Bandbegrenzung und Abtastung.

Der Zusammenhang zwischen beiden Modellen kann anhand der folgenden Grafik hergestellt werden, die auf der nächsten Seite noch genauer beschrieben wird.



Die wesentlichen Erkenntnisse vorneweg:

- Beim rechten Modell gilt zu den Abtastzeitpunkten $v \cdot T_A$ genau der gleiche Zusammenhang $Y_v = X_v + N_v$ wie beim bisherigen (linken) Modell.
- Die Störkomponente N_v ist nun durch (auf $\pm B$) bandbegrenztetes Weißes Rauschen mit zweiseitiger Leistungsdichte $\Phi_N(f) = N_0/2$ zu modellieren, wobei $B = 1/(2T_A)$ gelten muss \Rightarrow

„Abtasttheorem“.

AWGN-Modell für zeitdiskrete bandbegrenzte Signale (2)

Beim Modell gemäß der **unteren Grafik** auf der letzten Seite gehen wir von einer unendlichen Folge $\langle X_\nu \rangle$ von Gaußschen Zufallsgrößen aus, die einem **Diracpuls** $p_\delta(t)$ eingepreßt werden. Das resultierende zeitdiskrete Signal lautet somit:

$$X_\delta(t) = T_A \cdot \sum_{\nu=-\infty}^{+\infty} X_\nu \cdot \delta(t - \nu \cdot T_A).$$

Der Abstand aller (gewichteten) Diracfunktionen ist einheitlich T_A .

Durch das Interpolationsfilter mit der Impulsantwort $h(t)$ sowie dem Frequenzgang $H(f)$, wobei

$$h(t) = 1/T_A \cdot \text{si}(\pi \cdot t/T_A) \quad \longleftrightarrow \quad H(f) = \begin{cases} 1 & \text{für } |f| \leq B, \\ 0 & \text{für } |f| > B, \end{cases} \quad B = \frac{1}{T_A}$$

gelten muss, entsteht das zeitkontinuierliche Signal $X(t)$ mit folgenden Eigenschaften:

- Die Abtastwerte $X(\nu \cdot T_A)$ sind für alle ganzzahligen ν identisch mit den Eingangswerten X_ν , was mit den äquidistanten Nullstellen der **Spaltfunktion** $\Rightarrow \text{si}(x) = \sin(x)/x$ begründet werden kann.
- Gemäß dem **Abtasttheorem** ist $X(t)$ auf den Spektralbereich $\pm B$ ideal bandbegrenzt, wie die obige Rechnung gezeigt hat \Rightarrow rechteckförmiger Frequenzgang $H(f)$ der einseitigen Bandbreite B .

Nach der Addition der Störung $N(t)$ mit der (zweiseitigen) Leistungsdichte $\Phi_N(f) = N_0/2$ folgt das Matched-Filter mit si -förmiger Impulsantwort. Für die Störleistung am MF-Ausgang erhält man:

$$P_N = E[N_\nu^2] = \frac{N_0}{2T_A} = N_0 \cdot B.$$

Beweis: Mit $B = 1/(2T_A)$ erhält man für die Impulsantwort $h_E(t)$ und die Spektralfunktion $H_E(f)$:

$$h_E(t) = 2B \cdot \text{si}(2\pi \cdot B \cdot t) \quad \longleftrightarrow \quad H_E(f) = \begin{cases} 1 & \text{für } |f| \leq B, \\ 0 & \text{für } |f| > B. \end{cases}$$

Daraus folgt entsprechend den Erkenntnissen der **Stochastischen Systemtheorie**:

$$P_N = \int_{-\infty}^{+\infty} \Phi_N(f) \cdot |H_E(f)|^2 df = \int_{-B}^{+B} \Phi_N(f) df = \frac{N_0}{2} \cdot 2B = N_0 \cdot B.$$

- Tastet man das MF-Ausgangssignal in äquidistanten Abständen T_A ab, so ergibt sich für die Zeitpunkte $\nu \cdot T_A$ die gleiche Konstellation wie bisher: $Y_\nu = X_\nu + N_\nu$.
- Der Störanteil N_ν im zeitdiskreten Ausgangssignal Y_ν ist somit „bandbegrenzt“ und „weiß“. Die Kanalkapazitätsgleichung muss somit nur geringfügig angepasst werden;

$$C = \frac{1}{2} \cdot \log_2 \left(1 + \frac{P_X}{N_0 \cdot B} \right) = \frac{1}{2} \cdot \log_2 \left(1 + \frac{2 \cdot P_X \cdot T_A}{N_0} \right) = \frac{1}{2} \cdot \log_2 \left(1 + \frac{2 \cdot E_S}{N_0} \right).$$

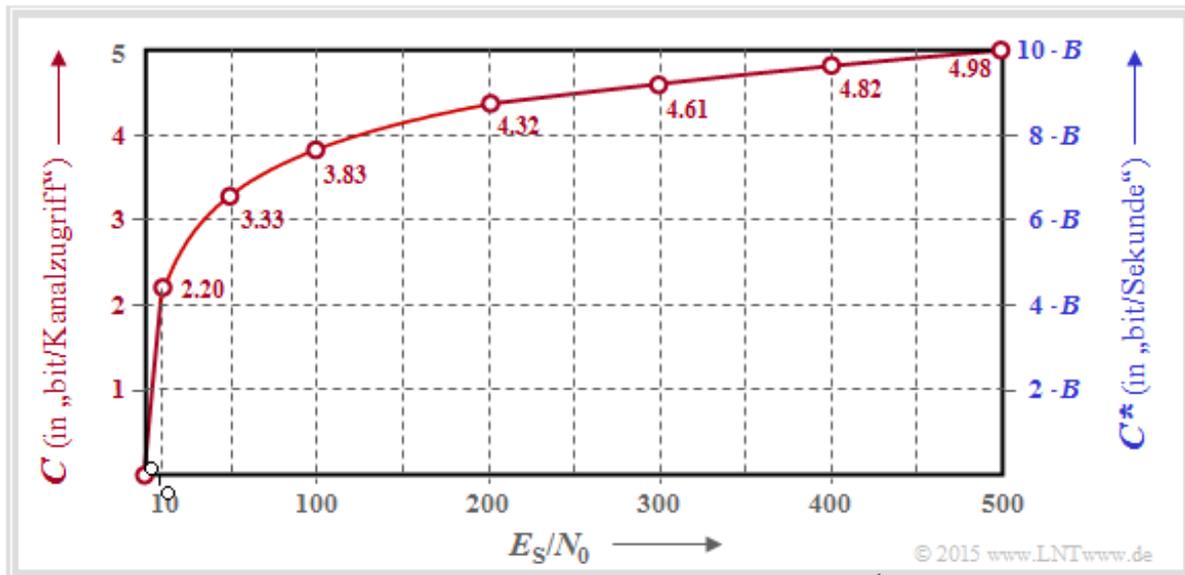
E_S ist die Sende-Energie innerhalb einer Symboldauer $T_A \Rightarrow$ **Energie pro Symbol**.

Die Kanalkapazität C als Funktion von E_S/N_0

Die obere Grafik zeigt den Verlauf der AWGN-Kanalkapazität in Abhängigkeit des Quotienten E_S/N_0 , wobei die linke Koordinatenachse und die roten Beschriftungen gültig sind:

$$C = \frac{1}{2} \cdot \log_2 \left(1 + \frac{2 \cdot E_S}{N_0} \right) \quad \text{Einheit: bit/Kanalzugriff (englisch: bit/channel use).}$$

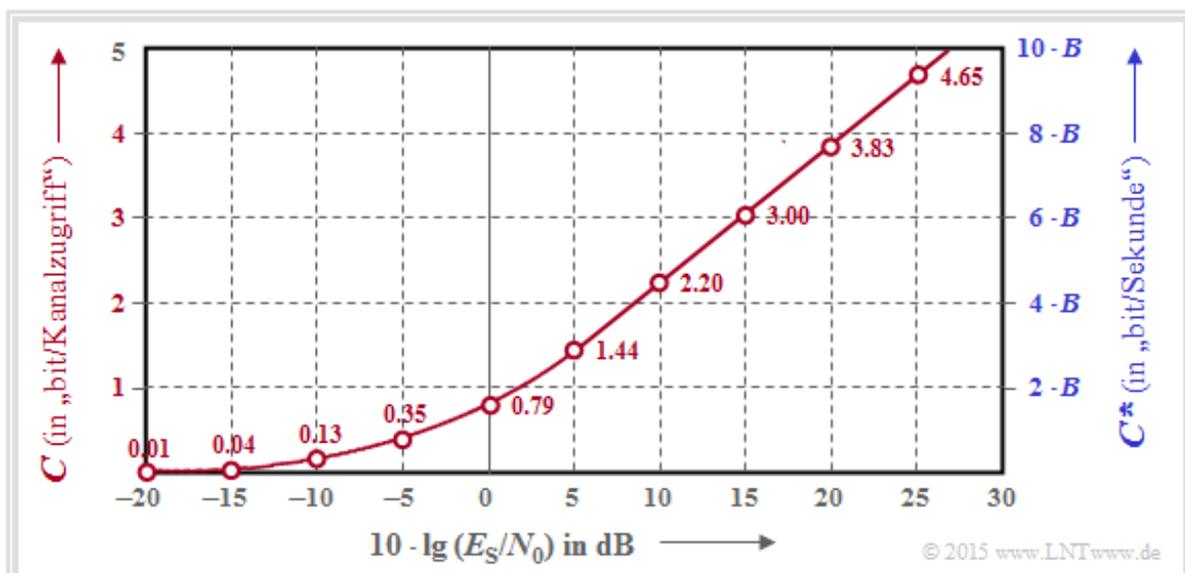
Die Einheit wird manchmal auch mit „bit/Quellensymbol“ oder kurz „bit/Symbol“ bezeichnet.



Die rechte (blaue) Achsenbeschriftung berücksichtigt die Beziehung $B = 1/(2T_A)$ und liefert somit eine obere Schranke für die Bitrate eines Digitalsystems, die bei diesem AWGN-Kanal noch möglich ist.

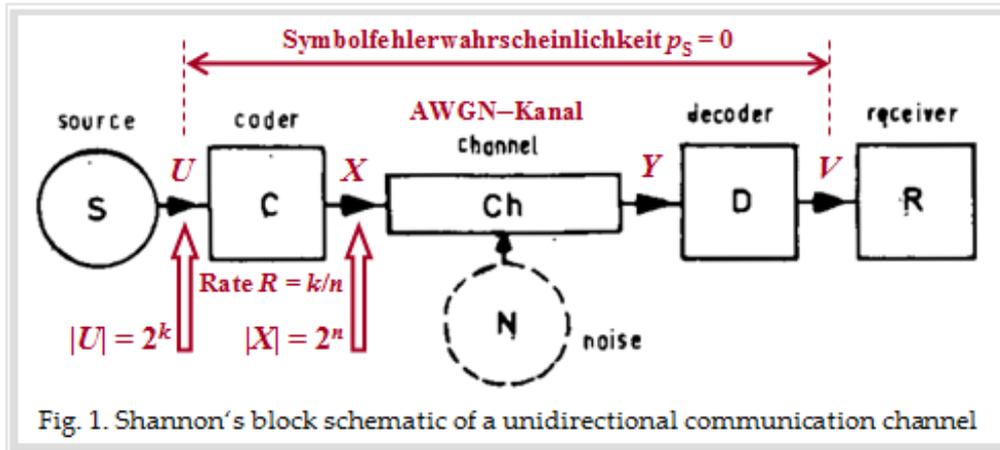
$$C^* = \frac{C}{T_A} = B \cdot \log_2 \left(1 + \frac{2 \cdot E_S}{N_0} \right) \quad \text{Einheit: bit/Sekunde.}$$

Meist gibt man den Quotienten aus *Symbolenergie* (E_S) und AWGN-Rauschleistungsdichte (N_0) in logarithmischer Form an. Die untere Grafik zeigt die Kanalkapazitäten C bzw. C^* als Funktion von $10 \cdot \lg(E_S/N_0)$ im Bereich von -20 dB bis $+30$ dB. Ab etwa 10 dB ergibt sich ein (nahezu) linearer Verlauf.



Systemmodell zur Interpretation der AWGN-Kanalkapazität

Um das **Kanalcodierungstheorem** im Zusammenhang mit dem AWGN-Kanal besprechen zu können, benötigen wir noch eine Codiervorrichtung, die informationstheoretisch vollständig durch die Coderate R gekennzeichnet wird.



Die Grafik beschreibt das von Shannon betrachtete Nachrichtensystem mit den Blöcken Quelle, Coder, (AWGN-)Kanal, Decoder und Empfänger. Im Hintergrund erkennt man ein Originalbild aus einem Shannon-Aufsatz zu diesem Thema. Rot eingezeichnet sind einige Bezeichnungen und Erläuterungen für den folgenden Text:

- Das Quellensymbol U entstammt einem Alphabet mit $M_U = |U| = 2^k$ Symbolen und kann durch k gleichwahrscheinliche statistisch unabhängige Binärsymbole repräsentiert werden.
- Das Alphabet des Codesymbols X hat den Symbolumfang $M_X = |X| = 2^n$, wobei sich n aus der Coderate $R = k/n$ ergibt. Für $R = 1$ gilt somit $n = k$.
- Der Fall $n > k$ führt zu einer Coderate $R < 1$ und aus $n < k$ folgt für die Coderate $R > 1$.

Das **Kanalcodierungstheorem** besagt, dass es (mindestens) einen Code der Rate R gibt, der zur Symbolfehlerwahrscheinlichkeit $p_S = \Pr(V \neq U) = 0$ führt, falls folgende Bedingungen erfüllt sind:

- Die Coderate R ist nicht größer als die Kanalkapazität C .
- Ein solcher geeigneter Code ist unendlich lang: $n \rightarrow \infty$, das heißt, dass die Zufallsgröße X am Kanaleingang wertkontinuierlich ist. Gleiches gilt für U sowie für die Zufallsgrößen Y und V nach dem AWGN-Kanal.
- Wegen $n \rightarrow \infty$ ist auch tatsächlich eine Gaußverteilung $f_X(x)$ am Kanaleingang möglich, die der bisherigen Berechnung der AWGN-Kanalkapazität stets zugrunde gelegt wurde:

$$C = \frac{1}{2} \cdot \log_2 \left(1 + \frac{2 \cdot E_S}{N_0} \right) \quad \text{Einheit: bit/Kanalzugriff (englisch: bit/channel use).}$$

- Für einen Systemvergleich ist die Energie pro Symbol (E_S) ungeeignet. Ein Vergleich sollte vielmehr auf der Energie E_B pro Informationsbit basieren. Mit $E_B = E_S/R$ gilt somit auch:

$$C = \frac{1}{2} \cdot \log_2 \left(1 + \frac{2 \cdot R \cdot E_B}{N_0} \right) \quad \text{Einheit: bit/Kanalzugriff (englisch: bit/channel use).}$$

Diese beiden Gleichungen werden auf der nächsten Seite diskutiert.

Die Kanalkapazität C als Funktion von E_B/N_0

Die folgende Grafik zeigt die AWGN-Kanalkapazität C als Funktion von

- $10 \cdot \lg(E_S/N_0) \Rightarrow$ roter Kurvenverlauf:

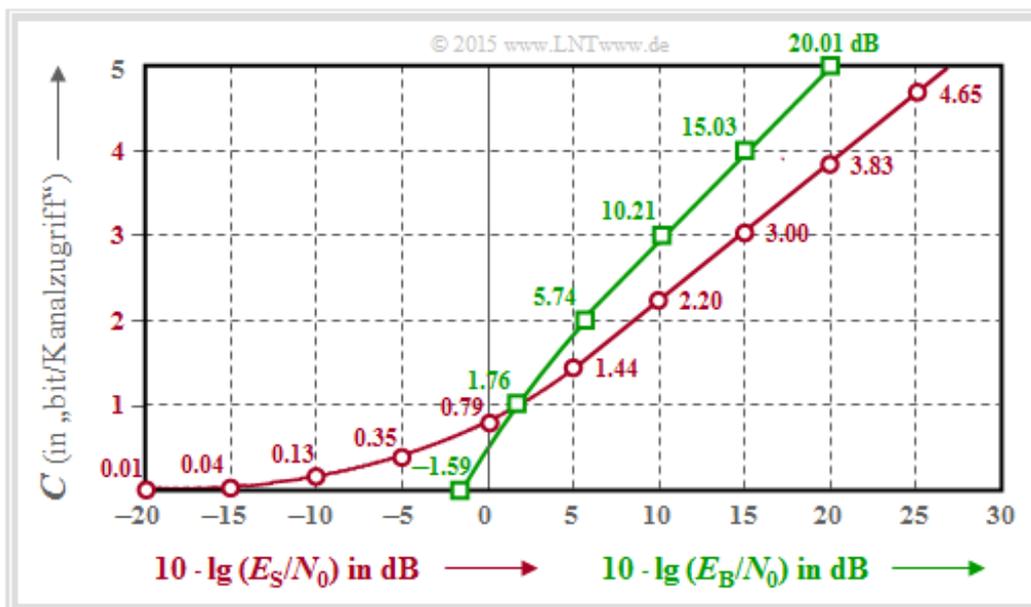
$$C = \frac{1}{2} \cdot \log_2 \left(1 + \frac{2 \cdot E_S}{N_0} \right) \quad \text{Einheit: bit/Kanalzugriff (oder: bit/Symbol).}$$

Rote Zahlen: Kapazität C in „bit/Symbol“ für $10 \cdot \lg(E_S/N_0) = -20$ dB, -15 dB, ..., $+30$ dB.

- $10 \cdot \lg(E_B/N_0) \Rightarrow$ grüner Kurvenverlauf:

$$C = \frac{1}{2} \cdot \log_2 \left(1 + \frac{2 \cdot R \cdot E_B}{N_0} \right) \quad \text{Einheit: bit/Kanalzugriff (oder: bit/Symbol).}$$

Grüne Zahlen: Erforderliches $10 \cdot \lg(E_B/N_0)$ in „dB“ für $C = 0, 1, \dots, 5$ in „bit/Symbol“.

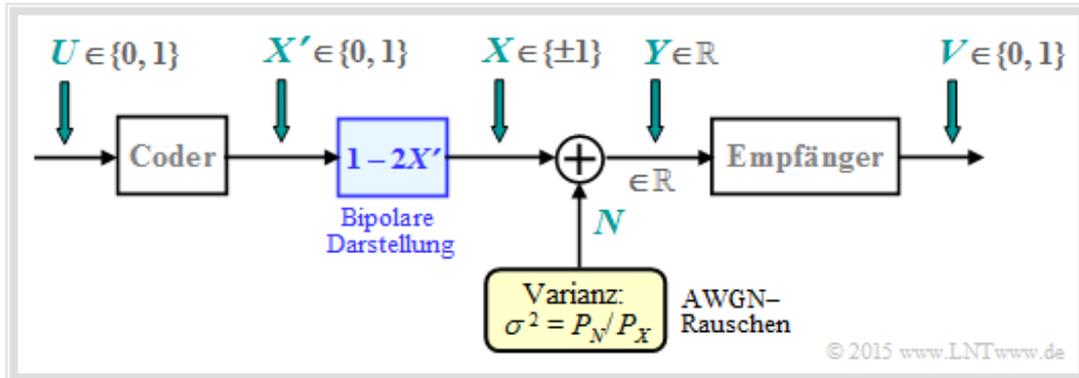


Die $C(E_B/N_0)$ -Berechnung finden Sie in der **Aufgabe A4.8** und der zugehörigen **Musterlösung**. Im Folgenden interpretieren wir das Ergebnis im Vergleich zur $C(E_S/N_0)$ -Kurve:

- Wegen $E_S = R \cdot E_B$ liegt der Schnittpunkt beider Kurven bei $C (=R) = 1$ [bit/Symbol]. Erforderlich sind dazu $10 \cdot \lg(E_S/N_0) = 1.76$ dB bzw. $10 \cdot \lg(E_B/N_0) = 1.76$ dB.
- Im Bereich $C > 1$ liegt die grüne Kurve stets über der roten. Beispielsweise ergibt sich für $10 \cdot \lg(E_B/N_0) = 20$ dB die Kanalkapazität $C \approx 5$, für $10 \cdot \lg(E_S/N_0) = 20$ dB nur $C = 3.83$.
- Ein Vergleich in horizontaler Richtung zeigt, dass die Kanalkapazität $C = 3$ bit/Symbol schon mit $10 \cdot \lg(E_B/N_0) \approx 10$ dB erreichbar ist, man aber $10 \cdot \lg(E_S/N_0) \approx 15$ dB benötigt.
- Im Bereich $C < 1$ liegt die rote Kurve stets über der grünen. Für $E_S/N_0 > 0$ gilt auch $C > 0$. Bei logarithmischer Abszisse reicht somit die rote Kurve bis ins „Minus-Unendliche“.
- Dagegen endet die grüne Kurve bei $E_B/N_0 = \ln(2) = 0.693 \Rightarrow 10 \cdot \lg(E_B/N_0) = -1.59$ dB \Rightarrow absolute Grenze für die (fehlerfreie) Übertragung über den AWGN-Kanal.

AWGN-Kanalkapazität für binäre Eingangssignale (1)

Auf den bisherigen Seiten des Kapitels 4.3 wurde stets entsprechend der Shannon-Theorie von einem gaußverteiltern und damit wertkontinuierlichem AWGN-Eingang X ausgegangen. Nun betrachten wir den binären Fall und werden somit der Überschrift „AWGN-Kanalkapazität bei wertdiskretem Eingang“ dieses Kapitels gerecht.



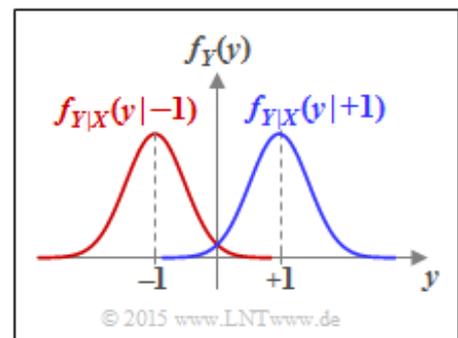
Die Grafik zeigt das zugrundeliegende Blockschaltbild für **Binary Phase Shift Keying (BPSK)** mit binärem Eingang U und ebenfalls binärem Ausgang V . Durch eine bestmögliche Codierung soll erreicht werden, dass die Fehlerwahrscheinlichkeit $\Pr(V \neq U)$ verschwindend klein wird.

- Der Coderausgang ist gekennzeichnet durch die binäre Zufallsgröße $X' = \{0, 1\} \Rightarrow M_{X'} = 2$, während der Ausgang Y des AWGN-Kanals weiterhin wertkontinuierlich ist: $M_Y \rightarrow \infty$.
- Durch das Mapping $X = 1 - 2X'$ kommt man von der unipolaren Darstellung zu der für BPSK besser geeigneten bipolaren (antipodalen) Beschreibung: $X' = 0 \rightarrow X = +1$; $X' = 1 \rightarrow X = -1$.
- Der AWGN-Kanal ist hier durch die beiden bedingten Wahrscheinlichkeitsdichtefunktionen charakterisiert:

$$f_{Y|X}(y|X = +1) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(y-1)^2}{2\sigma^2}\right],$$

$$f_{Y|X}(y|X = -1) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(y+1)^2}{2\sigma^2}\right].$$

In Kurzform: $f_{Y|X}(y|+1)$ bzw. $f_{Y|X}(y|-1)$.



- Da hier das Nutzsinal X auf ± 1 normiert ist \Rightarrow Leistung 1 anstelle von P_X , muss die Varianz des AWGN-Rauschens N in gleicher Weise normiert werden: $\sigma^2 = P_N/P_X$.
- Der Empfänger trifft aus der reellwertigen Zufallsgröße Y (am AWGN-Kanalausgang) eine **Maximum-Likelihood-Entscheidung**. Der Empfängerausgang V ist binär (0 oder 1).

Ausgehend von diesem Modell wird auf der nächsten Seite die Kanalkapazität berechnet.

AWGN-Kanalkapazität für binäre Eingangssignale (2)

Die Kanalkapazität des AWGN-Kanals unter der Nebenbedingung einer binären Eingangsgröße X lautet allgemein unter Berücksichtigung von $\Pr(X = -1) = 1 - \Pr(X = +1)$:

$$C_{\text{BPSK}} = \max_{\Pr(X=+1)} I(X; Y).$$

Aufgrund des symmetrischen Kanals ist offensichtlich, dass die Eingangswahrscheinlichkeiten

$$\Pr(X = +1) = \Pr(X = -1) = 0.5$$

zum Optimum führen werden. Gemäß **Kapitel 4.2** gibt es mehrere Berechnungsmöglichkeiten:

$$C_{\text{BPSK}} = h(X) + h(Y) - h(XY),$$

$$C_{\text{BPSK}} = h(Y) - h(Y|X),$$

$$C_{\text{BPSK}} = h(X) - h(X|Y).$$

Alle Ergebnisse sind noch um die Pseudo-Einheit „bit“ zu ergänzen. Wir wählen hier die mittlere Gleichung:

- Die hierfür benötigte bedingte differentielle Entropie ist gleich

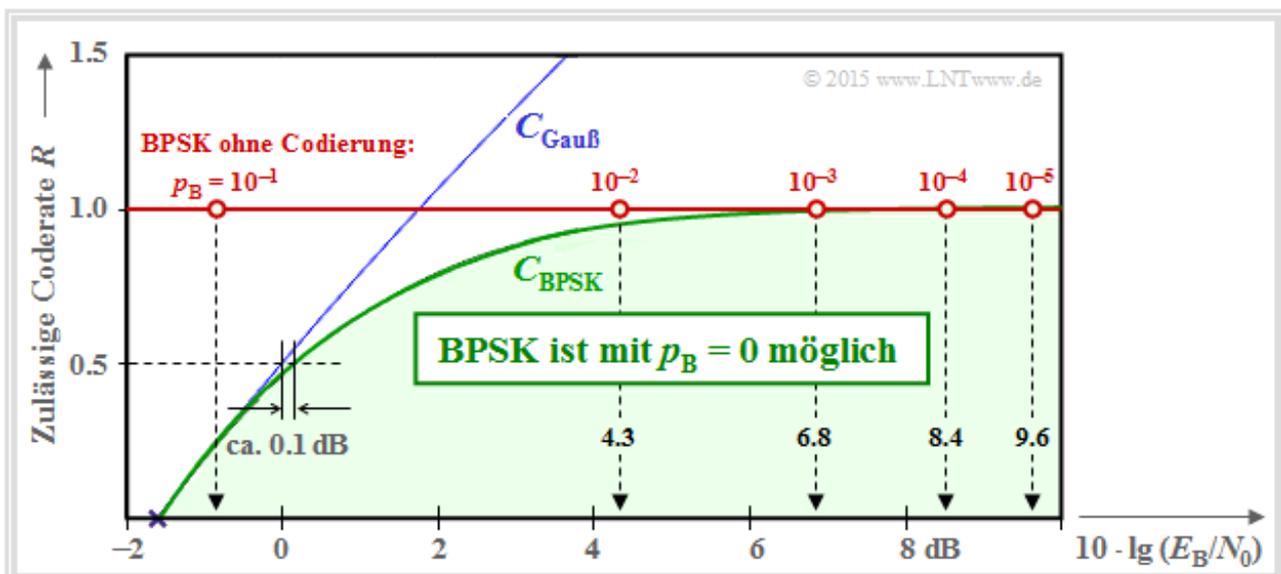
$$h(Y|X) = h(N) = 1/2 \cdot \log_2(2\pi e \cdot \sigma^2).$$

- Die differentielle Entropie $h(Y)$ ist vollständig durch die WDF $f_Y(y)$ gegeben. Mit den vorne definierten und skizzierten bedingten Wahrscheinlichkeitsdichtefunktionen erhält man:

$$f_Y(y) = \frac{1}{2} \cdot [f_{Y|X}(y|X = -1) + f_{Y|X}(y|X = +1)]$$

$$\Rightarrow h(Y) = - \int_{y \in \text{supp}(f_Y)} f_Y(y) \cdot \log_2[f_Y(y)] dy.$$

Es ist offensichtlich, dass $h(Y)$ nur durch numerische Integration ermittelt werden kann, insbesondere, wenn man berücksichtigt, dass sich im Überlappungsbereich $f_Y(y)$ aus der Summe der beiden bedingten Gauß-Funktionen ergibt.

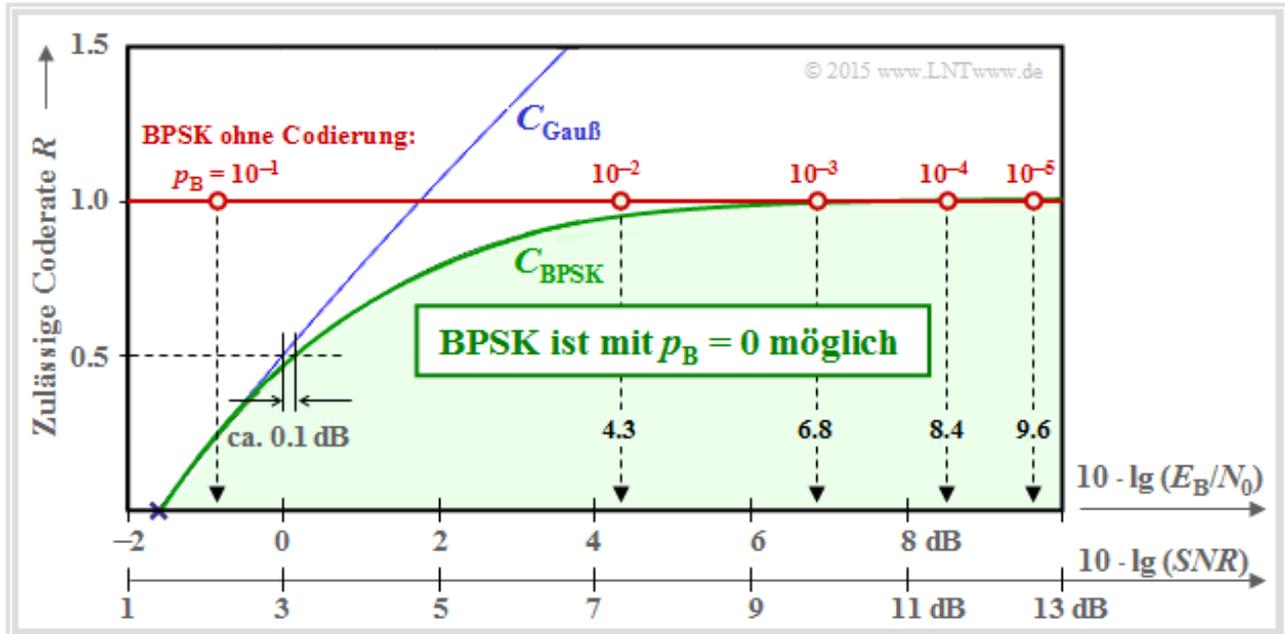


Das skizzierte Ergebnis wird auf der nächsten Seite diskutiert.

AWGN-Kanalkapazität für binäre Eingangssignale (3)

In der folgenden Grafik sind über der Abszisse $10 \cdot \lg(E_B/N_0)$ drei Kurven dargestellt:

- die Kanalkapazität $C_{\text{Gauß}}$, gültig für eine Gaußsche Eingangsgröße $X \Rightarrow M_X \rightarrow \infty$,
- die Kanalkapazität C_{BPSK} für die Zufallsgröße $X = (+1, -1)$, sowie
- die mit „BPSK ohne Codierung“ bezeichnete Horizontale.



Diese Kurvenverläufe sind wie folgt zu interpretieren:

- Die grüne Kurve C_{BPSK} gibt die maximal zulässige Coderate R einer BPSK an, bei der für das gegebene E_B/N_0 durch bestmögliche Codierung die Bitfehlerwahrscheinlichkeit $p_B = 0$ möglich ist.
- Für alle BPSK-Systeme mit den Koordinaten $(10 \cdot \lg E_B/N_0, R)$ im „grünen Bereich“ ist $p_B = 0$ prinzipiell erreichbar. Aufgabe der Nachrichtentechniker ist es, hierfür geeignete Codes zu finden.
- Die BPSK-Kurve liegt stets unter der absoluten Shannon-Grenzkurve $C_{\text{Gauß}}$ für $M_X \rightarrow \infty$. Im unteren Bereich gilt $C_{\text{BPSK}} \approx C_{\text{Gauß}}$. Zum Beispiel muss ein BPSK-System mit $R = 1/2$ nur ein um 0.1 dB größeres E_B/N_0 bereitstellen, als es die (absolute) Kanalkapazität $C_{\text{Gauß}}$ fordert.
- Ist E_B/N_0 endlich, so gilt stets $C_{\text{BPSK}} < 1 \Rightarrow$ siehe **Aufgabe Z4.9**. Eine BPSK mit $R = 1$ (und somit ohne Codierung) wird stets eine Bitfehlerwahrscheinlichkeit $p_B > 0$ zur Folge haben.
- Die Fehlerwahrscheinlichkeiten eines solchen BPSK-Systems ohne Codierung ($R = 1$) sind auf der roten Horizontalen angegeben. Um $p_B \leq 10^{-5}$ zu erreichen, benötigt man mindestens $10 \cdot \lg(E_B/N_0) = 9.6$ dB.

Die Wahrscheinlichkeiten ergeben sich gemäß **Kapitel 1.5** im Buch „Digitalsignalübertragung“ zu

$$p_B = Q\left(\sqrt{SNR}\right) \quad \text{mit} \quad SNR = 2 \cdot E_B/N_0.$$

Hinweis: In obiger Grafik ist $10 \cdot \lg(SNR)$ als zweite, zusätzliche Abszissenachse eingezeichnet. Die

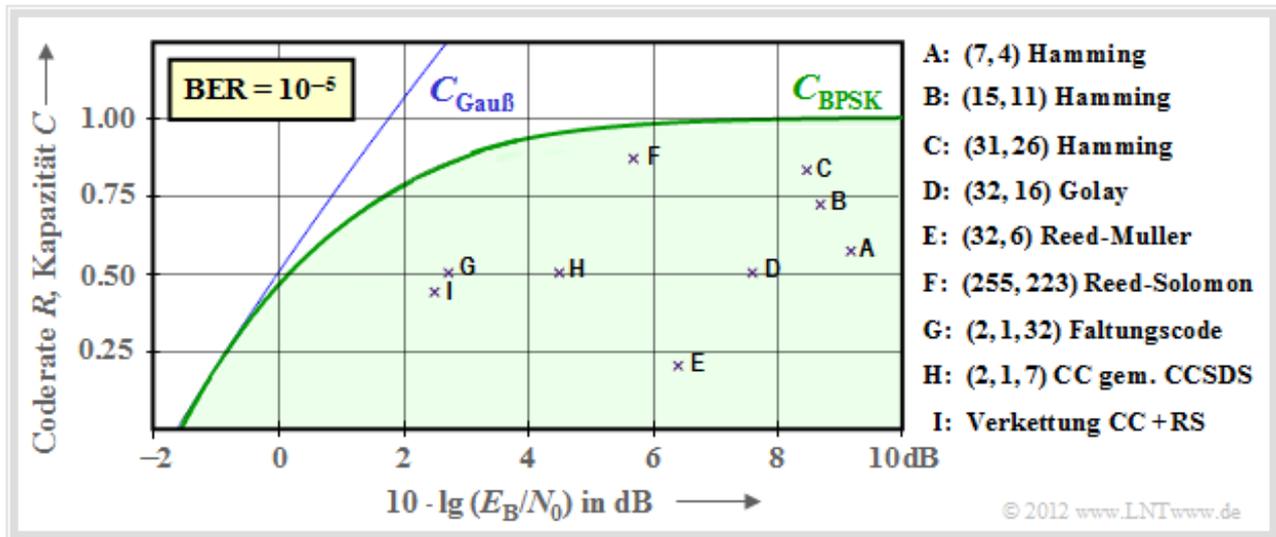
Funktion $Q(x)$ bezeichnet man als die **komplementäre Gaußsche Fehlerfunktion**.

Vergleich zwischen Theorie und Praxis (1)

Anhand zweier Grafiken soll gezeigt werden, in wie weit sich etablierte Kanalcodes der BPSK-Kanalkapazität (grüne Kurve) annähern. Als Ordinate aufgetragen ist die Rate $R = k/n$ dieser Codes bzw. die Kapazität C (wenn noch die Pseudo-Einheit „bit/Kanalzugriff“ hinzugefügt wird). Vorausgesetzt ist:

- der AWGN-Kanal, gekennzeichnet durch $10 \cdot \lg(E_B/N_0)$ in dB, und
- für die durch Kreuze markierten realisierten Codes eine Bitfehlerrate (BER) von 10^{-5} .

Zu beachten ist, dass die Kanalkapazitätskurven stets für $n \rightarrow \infty$ und $BER = 0$ gelten. Würde man diese strenge Forderung „fehlerfrei“ auch an die betrachteten Kanalcodes endlicher Codelänge n anlegen, so wäre hierfür stets $10 \cdot \lg(E_B/N_0) \rightarrow \infty$ erforderlich. Dies ist aber ein eher akademisches Problem, das für die Praxis weniger Bedeutung hat. Für $BER = 10^{-10}$ ergäbe sich eine qualitativ ähnliche Grafik.



Es folgen einige Erläuterungen zu den Daten, die der Vorlesung [Liv10] entnommen wurden. Die folgenden Links beziehen sich oft auf das LNTwww-Buch **Einführung in die Kanalcodierung**.

- Die Punkte **A**, **B** und **C** markieren **Hamming-Codes** der Raten $R = 4/7 \approx 0.57$, $R \approx 0.73$ bzw. $R \approx 0.84$. Für $BER = 10^{-5}$ benötigen diese sehr frühen Codes (aus dem Jahr 1950) alle $10 \cdot \lg(E_B/N_0) > 8$ dB.
- Die Markierung **D** kennzeichnet den binären **Golay-Code** mit der Rate $1/2$ und der Punkt **E** einen **Reed-Muller-Code**. Dieser sehr niederrangige Code kam bereits 1971 bei der Raumsonde Mariner 9 zum Einsatz.
- Die **Reed-Solomon-Codes** (RS-Codes, ca. 1960) sind eine Klasse zyklischer Blockcodes. **F** markiert einen RS-Code der Rate $223/255 > 0.9$ und einem erforderlichen $E_B/N_0 < 6$ dB.
- Die Punkte **G** und **H** bezeichnen zwei **Faltungscodes** (englisch: *Convolutional Codes*, CC) mittlerer Rate. Der Code **G** wurde schon 1972 bei der Pioneer10-Mission eingesetzt.
- Die Kanalcodierung der Voyager-Mission Ende der 1970er Jahre ist mit **I** markiert. Es handelt sich um die Verkettung eines (2, 1, 7)-Faltungscodes mit einem RS-Code.

Anzumerken ist, dass bei den Faltungscodes der dritte Kennungsparameter eine andere Bedeutung hat

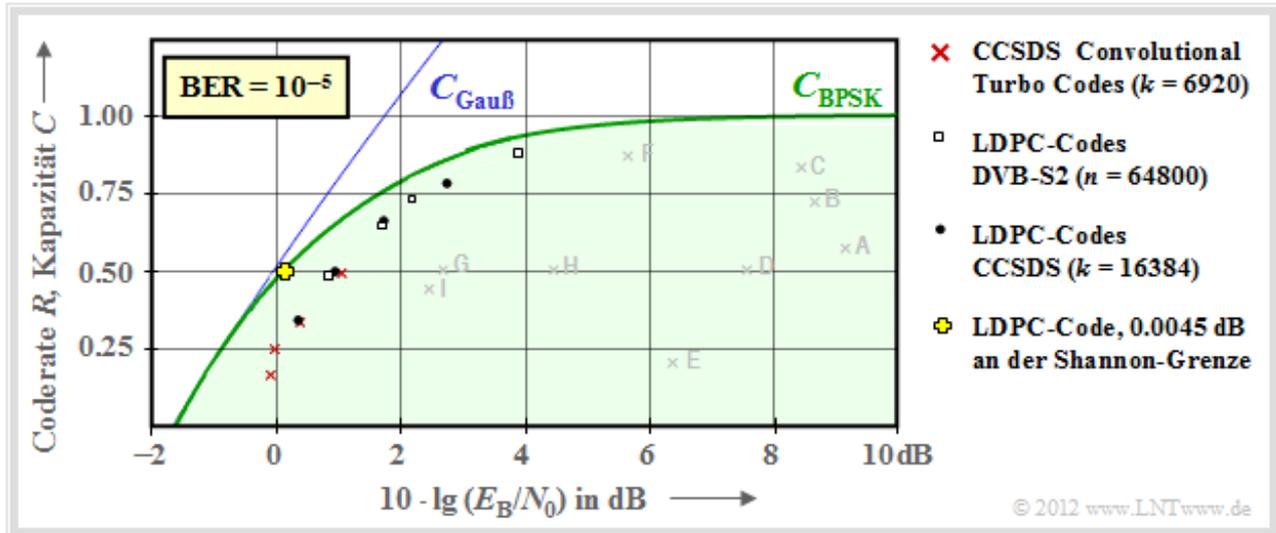
als bei den Blockcodes. $(2, 1, 32)$ weist beispielsweise auf das Memory $m = 32$ hin.

Auf der nächsten Seite folgen noch die Kenndaten von Systemen mit iterativer Decodierung.

Vergleich zwischen Theorie und Praxis (2)

Die frühen Kanalcodes der letzten Seite liegen noch relativ weit von der Kanalkapazitätskurve entfernt. Dies war wahrscheinlich auch ein Grund, warum dem Autor die auch große praktische Bedeutung der Informationstheorie verschlossen blieb, als er diese Anfang der 1970er Jahre im Studium kennenlernte.

Diese Sichtweise hat sich deutlich verändert, als in den 1990er Jahren sehr lange Kanalcodes zusammen mit iterativer Decodierung aufkamen. Die neuen Markierungspunkte liegen näher an der Kapazitätskurve.



Hier noch einige Erläuterungen zu dieser Grafik:

- Rote Kreuze markieren sog. **Turbo-Codes** nach CCSDS (*Consultative Committee for Space Data Systems*) mit jeweils $k = 6920$ Informationsbits und unterschiedlichen Codelängen $n = k/R$. Diese von **Claude Berrou** um 1990 erfundenen Codes können iterativ decodiert werden. Die (roten) Markierungen liegen jeweils weniger als 1 dB von der Shannon-Grenze entfernt.
- Ähnlich verhalten sich die **LDPC-Codes** (*Low Density Parity-check Codes*) mit konstanter Codelänge $n = 64800 \Rightarrow$ weiße Rechtecke). Sie werden seit 2006 bei DVB-S2 (*Digital Video Broadcast over Satellite*) eingesetzt und eignen sich aufgrund der spärlichen Einsen-Belegung der Prüfmatrix sehr gut für die iterative Decodierung mittels **Faktor-Graphen** und **Exit Charts**.
- Schwarze Punkte markieren die von CCSDS spezifizierten LDPC-Codes mit konstanter Anzahl an Informationsbits ($k = 16384$) und variabler Codewortlänge $n = k/R$. Diese Codeklasse erfordert ein ähnliches E_B/N_0 wie die roten Kreuze und die weißen Rechtecke.

Um die Jahrhundertwende hatten viele Forscher den Ehrgeiz, sich der Shannon-Grenze bis auf Bruchteile von einem dB anzunähern. Das gelbe Kreuz markiert ein derartiges Ergebnis (0.0045 dB) von **Chung et al.** aus dem Jahr 2001. Verwendet wurde ein irregulärer LDPC-Code mit Rate 1/2 und Codelänge 10^7 .

An dieser Stelle soll nochmals die Brillanz und der Weitblick von **Claude E. Shannon** hervorgehoben werden. Er hat 1948 eine bis dahin nicht bekannte Theorie entwickelt, mit der die Möglichkeiten, aber auch die Grenzen der Digitalsignalübertragung aufgezeigt werden. Zu dieser Zeit waren die ersten Überlegungen zur digitalen Nachrichtenübertragung gerade mal zehn Jahre alt \Rightarrow **Pulscode modulation** (Alec Reeves, 1938) und selbst der Taschenrechner kam erst mehr als 20 Jahre später. Shannon's Arbeiten zeigen uns, dass man auch ohne gigantische Computer Großes leisten kann.

Kanalkapazität des komplexen AWGN-Kanals

Höherstufige Modulationsverfahren wie

- **M-QAM** \Rightarrow Quadraturamplitudenmodulation; $M \geq 4$ quadratische Signalraumpunkte
- **M-PSK** \Rightarrow $M \geq 4$ Signalraumpunkte in kreisförmiger Anordnung

können jeweils durch eine Inphase- und eine Quadraturkomponente dargestellt werden. Die beiden Komponenten lassen sich im **äquivalenten Tiefpassbereich** auch als *Realteil* bzw. *Imaginärteil* eines komplexen Rauschterms N beschreiben.

Alle oben genannten Verfahren sind zweidimensional. Der (komplexe) AWGN-Kanal stellt somit $K = 2$ voneinander unabhängige Gaußkanäle zur Verfügung. Entsprechend **Kapitel 4.2** ergibt sich deshalb für die Kapazität dieses Kanals:

$$C_{\text{Gauß, komplex}} = C_{\text{Gesamt}}(K = 2) = \log_2 \left(1 + \frac{P_X/2}{\sigma^2} \right).$$

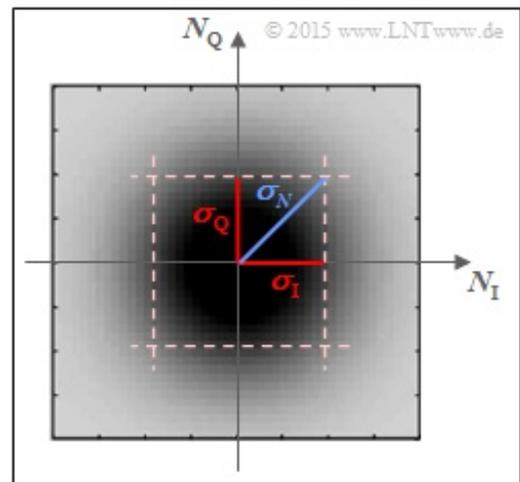
- Die gesamte Nutzleistung von Inphase- und Quadraturkomponente wird mit P_X bezeichnet.
- Dagegen bezieht sich die Varianz σ^2 der Störung nur auf eine Dimension: $\sigma^2 = \sigma_I^2 = \sigma_Q^2$.

Die rechte Abbildung zeigt die 2D-WDF $f_N(n)$ des Gaußschen Rauschprozesses N über den beiden Achsen

- N_I (Inphase-Anteil, Realteil) und
- N_Q (Quadraturanteil, Imaginärteil).

Dunklere Bereiche der rotationssymmetrischen WDF $f_N(n)$ um den Nullpunkt weisen auf mehr Störanteile hin. Für die Varianz des komplexen Gaußschen Rauschens N gilt aufgrund der Rotationsinvarianz ($\sigma_R = \sigma_I$) folgender Zusammenhang:

$$\sigma_N^2 = \sigma_I^2 + \sigma_Q^2 = 2 \cdot \sigma^2.$$



Damit lässt sich die Kanalkapazität auch wie folgt ausdrücken:

$$C_{\text{Gauß, komplex}} = \log_2 \left(1 + \frac{P_X}{\sigma_N^2} \right) = \log_2 (1 + SNR).$$

Diese Gleichung wird auf der nächsten Seite numerisch ausgewertet. Bereits aus dieser Gleichung ist zu ersehen, dass für das Signal-zu-Störleistungsverhältnis gilt:

$$SNR = P_X / \sigma_N^2.$$

Maximale Coderate für QAM-Strukturen

In der Grafik ist die Kanalkapazität des komplexen AWGN-Kanals als rote Kurve dargestellt:

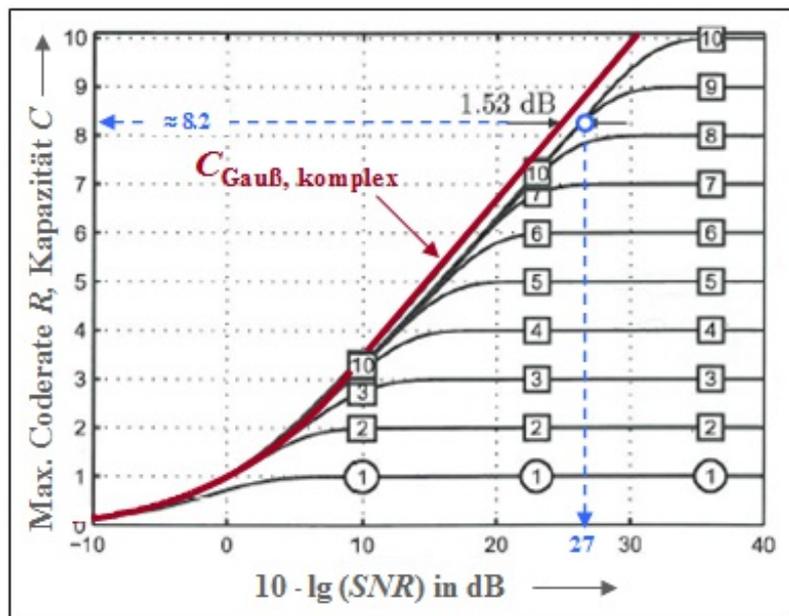
$$C_{\text{Gauß, komplex}} = \log_2(1 + SNR).$$

Die Einheit dieser Kanalkapazität ist wieder „bit/Kanalzugriff“ oder „bit/Quellensymbol“. Als Abszisse ist der Signal-zu-Störleistungsverhältnis $10 \cdot \log(SNR)$ mit $SNR = P_X/\sigma_N^2$ aufgetragen.

Die rote Kurve basiert entsprechend der Shannon-Theorie wieder auf einer Gaußverteilung $f_X(x)$ am Eingang. Zusätzlich eingezeichnet sind zehn weitere Kapazitätskurven für wertdiskreten Eingang:

- die BPSK (mit „1“ markiert),
- die M -QAM, ($M = 2^2, \dots, 2^{10}$).

Diese Grafik wurde der Dissertation [Göb10] entnommen. Wir danken unserem ehemaligen Kollegen am LNT, Dr.-Ing. **Bernhard Göbel**, für sein Einverständnis, diese Abbildung verwenden zu dürfen, sowie für seine Unterstützung unseres Lerntutorials.



Man erkennt aus dieser Darstellung:

- Die BPSK-Kurve sowie alle M -QAM-Kurven liegen rechts von der Shannon-Grenzkurve. Bei kleinem SNR sind alle Kurven von der roten Kurve fast nicht mehr zu unterscheiden.
- Der Endwert aller Kurven für wertdiskrete Eingangssignale ist $\log_2(M)$. Für $SNR \rightarrow \infty$ erhält man beispielsweise $C_{\text{BPSK}} = 1$ bit/Symbol sowie $C_{4\text{-QAM}} = C_{\text{QPSK}} = 2$ bit/Symbol.
- Die blauen Markierungen zeigen, dass eine 2^{10} -QAM mit $10 \cdot \lg(SNR) \approx 27$ dB eine Coderate von $R \approx 8,2$ ermöglicht. Der Abstand zur Shannon-Kurve beträgt hier 1.53 dB.
- Man spricht hier von einem *Shaping Gain* von $10 \cdot \lg(\pi/6) = 1.53$ dB. Diese Verbesserung lässt sich erzielen, wenn man die Lage der 32^2 quadratisch angeordneten Signalraumpunkte so ändern würde, dass sich eine gaußähnliche Eingangs-WDF ergibt \Rightarrow *Signal Shaping*.

In der **Aufgabe A4.10** werden die AWGN-Kapazitätskurven von BPSK und QPSK diskutiert:

- Ausgehend von der Abszisse $10 \cdot \lg(E_B/N_0)$ mit der Energie E_B pro Informationsbit kommt man zur QPSK-Kurve durch Verdopplung der BPSK-Kurve:

$$C_{\text{QPSK}}(10 \cdot \lg(E_B/N_0)) = 2 \cdot C_{\text{BPSK}}(10 \cdot \lg(E_B/N_0)).$$

- Vergleicht man aber BPSK und QPSK bei gleicher Energie pro Informationssymbol (E_S), so gilt:

$$C_{\text{QPSK}}(10 \cdot \lg E_S/N_0) = 2 \cdot C_{\text{BPSK}}(10 \cdot \lg E_S/N_0 - 3 \text{ dB}).$$

Hierbei ist berücksichtigt, dass bei QPSK die Energie in einer Dimension nur $E_S/2$ beträgt.