EINFÜHRUNG IN DIE INFORMATIONS- UND CODIERUNGSTHEORIE

mit Anmerkungen zur

KRYPTOLOGIE

Notizen zur Vorlesung von
Dr.-Ing. Wolfgang Sauer-Greff

Technische Universität Kaiserslautern

Oktober 2012

Niederschrift: T. Divivier
Dr. C. Zhao
U. Ruby
J. Eckert
U. Botzenhardt

sauer@eit.uni-kl.de

I. INFORMATIONSTHEORIE

Inhaltsverzeichnis

1 Begriffe, Übersicht

1.1 Literatur

Lehrbücher

- **K. Wesolowski:** "Introduction to Digital Communication Systems", Wiley, 2009
- **H. Rohling:** "Einführung in die Informations- und Codierungstheorie", Teubner, 1995
- **B. Friedrichs:** "Kanalcodierung", Springer, 1996
- **W. Heise, P. Quattrocchi:** "Informations- und Codierungstheorie", Springer, 1995
- **R. Johannesson:** "Informationstheorie", Addison-Wesley, 1992

Historische Aufsätze zur Informationstheorie

- **W. Nyquist:** "Certain Factors Affecting Telegraph Speed", BSTJ, 1924
- **R. Hartley:** "Transmission of Information", BSTJ, 1928
- **C. Shannon:** "A Mathematical Theory of Communication", BSTJ, 1948
  BSTJ = Bell System Technical Journal

Verbindliche Begriffsdefinitionen: DIN 44300, DIN 44301

1.2 Begriffe

1.2.1 Information-Nachricht

Definitionen:

Information ist eine, durch die empfangene **Nachricht** erhaltene Kenntnis, die es dem Empfänger erlaubt, seinen **Wissenszustand** zu erhöhen.
Information ist für den Empfänger
– *potenziell*, da er sie nicht nutzen muss,
– *relativ*, da sie abhängig von seinem Wissenszustand ist.

Nachricht: Zeichen oder kontinuierliche Funktion
(im weiteren: Zeichen ≡ Nachricht (message))

Zeichen: Element aus einer zur Darstellung von Information vereinbarten endlichen Menge (= Zeichenvorrat)

- Generierung und Empfang von Nachrichten sind zum Informationstransfer (Kommunikation) geeignet.
  - Nachricht ist Informationsträger, physikalische Darstellung als Signal
  - Signalübertragung ist mit Energietransfer verbunden ($\Rightarrow$ max. Lichtgeschwindigkeit)
- Interpretationsrahmen für die Kommunikation = Vereinbarung zur Darstellung von Information
  - hier: von der Semantik losgelöster Informationsgehalt von Nachrichten (Neuigkeitswert)
  - Informationstheorie benutzt Methoden der Stochastik.

Nachricht (message)

Element aus dem Nachrichtenvorrat (= Zeichenvorrat) der Nachrichtenquelle

- kontinuierlich
- diskret, endlich: $\mathbf{A} = \{a_1, a_2, \ldots, a_q\}$; $\rightarrow q$-närer Nachrichtenvorrat

  z.B.: $q = 2$: binärer Zeichenvorrat; $(\{0,1\}; \{L,H\}; \{-,+\} \ldots)$

  (elementare) Nachricht $a_i \in \mathbf{A}, \quad i = 1, \ldots, q$

  Nachrichtenfolge $\underline{a}$ der Länge n: $\underline{a} = [a(t = 1T), a(t = 2T), \ldots, a(t = nT)]$

  Nachrichtenblock $\vec{a}$ der Länge n: $\vec{a} = (a_1, a_2, \ldots, a_n)^T$

Begriffsanalogien:

Datum (Verarbeitung)	$\equiv$	Nachricht (Übertragung)	$\equiv$	Zeichen
Datenfolge	$\equiv$	Nachrichtenfolge	$\equiv$	Zeichenfolge
Symbol	$:=$	Zeichen mit Bedeutung		
Signal	$:=$	physikal. Darstellung der Nachricht		

1.3 Übersicht

Kommunikationssystem

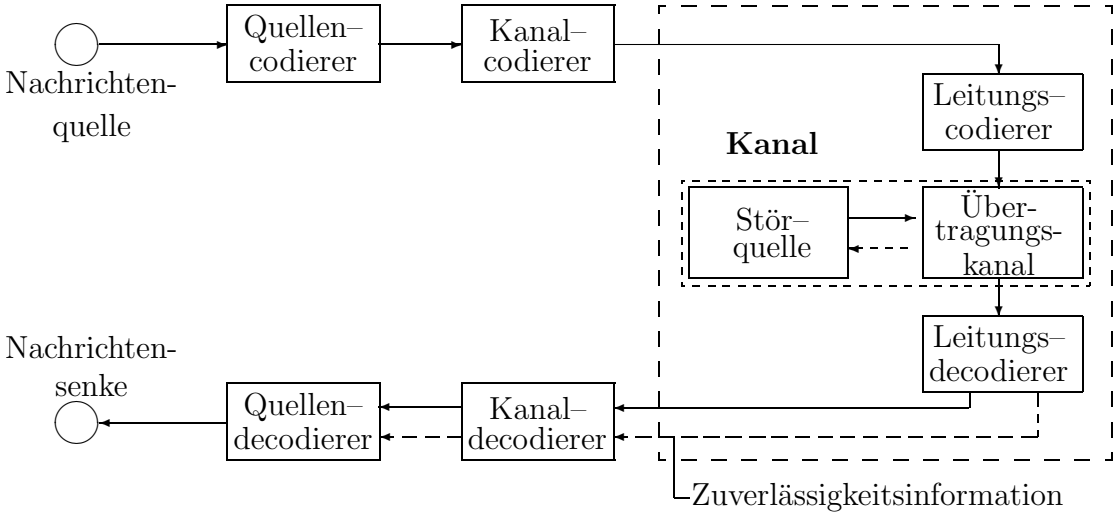


Bild 1.1: Kommunikationssystem

Nachrichtenquelle (message source)

Hier: Stochastisches Modell ohne Semantik

- Zufällige Auswahl aus Zeichenvorrat $\mathbf{A} = \{a_i;\ i = 1, \ldots, q\}$
  mit bekannten Wahrscheinlichkeiten $\vec{P} = \big(P(a_i);\ i = 1, \ldots, q\big)$;
  unabhängig vom Zeitpunkt $\Rightarrow$ stationärer, diskreter, stochastischer Prozess

- Nachrichten-Folge ist Musterfunktion eines stochastischen Prozesses

  – **gedächtnislose Quelle** = statistisch unabhängige Zeichen:
  $P\big(a(i) \mid a(j)\big) = P\big(a(i)\big)\,; \qquad P\big(a(i), a(j)\big) = P\big(a(i)\big) \cdot P\big(a(j)\big)$

  – **gedächtnisbehaftete Quelle** $\longrightarrow$ Markoff-Quelle

Signalisiertempo: Zeiteinheit T_q je generiertem Zeichen

Bsp.: Compact Disc (CD) mit Stereo-Musik
- 2 Kanäle mit je 20 kHz Bandbreite $\rightarrow$ 44,1 kHz Abtastrate
- Signalisiertempo: 88.200 $\frac{\text{Zeichen}}{\text{s}}$
- bei 16 bit Quantisierung $\rightarrow$ Zeichenvorrat: $q = 2^{16} = 65536$
- Datenrate $\approx 1,4$ Mbit/s

Quellencodierer (source coder)

- Zuordung: Nachrichten(folge) → Zeichenfolge
- Datenkompression: Eliminieren von Redundanz (verlustlos) und Irrelevanz (verlustbehaftet)

  Bsp.: Lauflängencodierung: "*aaaaaaaa*" → "8{*a*}"
  Linear Predictive Coding: Sprache (PCM) 64 kbit/s → 9,6 kbit/s
  MUSICAM: Audio (CD) 1,4 Mbit/s → 256 kbit/s
  MPEG: Video (HDTV) 200 Mbit/s → 6 Mbit/s

- Geschwindigkeitsoptimierung: häufige Nachricht → kurze Zeichenfolge
  seltene Nachricht → lange Zeichenfolge

  Bsp.: Morse-Alphabet, Huffman-Codierung, Lempel-Ziv-Algorithmus (.zip)

Leitungscodierer (line coding) **/ Modulation**

Anpassung Signalform an Kanalrestriktionen: Zeichen → elementare Signalform;
z.B. NRZ-Datensignal, spektrale Formung (Filter, Scrambler), Modulation

Leitungsdecodierer (line decoding, detection) **/ Demodulation**

Detektion der ggf. modulierten empfangenen Signale und Rückwandlung in Zeichen;
z.B. Optimalempfänger (Matched Filter + Schwellwertentscheider; Viterbi-Algorithmus)

Kanal (channel)

Bsp.: Teilnehmeranschlussleitung, Mobilfunkkanal, extraterrestrische Funkverbindung, Magnetband, CD

Umfasst in der Informationstheorie den Übertragungskanal und die Störung, jeweils mit spezifischem Leitungscodierer/ -decodierer

Störung (noise)

- Stochastische Kanalstörung (Wahrscheinlichkeitsdichte; Leistungsdichtespektrum; multiplikativ / additiv)
- Verzerrungen (gedächtnisbehaftet und/oder nichtlinear): abhängig vom übertragenen Signal

Anschaulich: Kanal ist charakterisiert durch:

- endliche Menge zulässiger Zeichen (Amplitudenstufen) (maximale Amplitude) $A_{\max}$

- „Zeiteinheit“ T: Zeitspanne zur Aufnahme zulässiger Zeichen (z.B. eine Sekunde)
- endliche Übertragungsgeschwindigkeit (Rate) $R_{\max}$ = maximal zulässige Anzahl von Zeichen je Zeiteinheit T.

$\longrightarrow$ Kanalkapazität

Die Übertragungsdauer $T_{\ddot{u}}$ einer Nachrichtenmenge hängt von $A_{\max}$ und $R_{\max}$ ab, s.Bild 1.2.

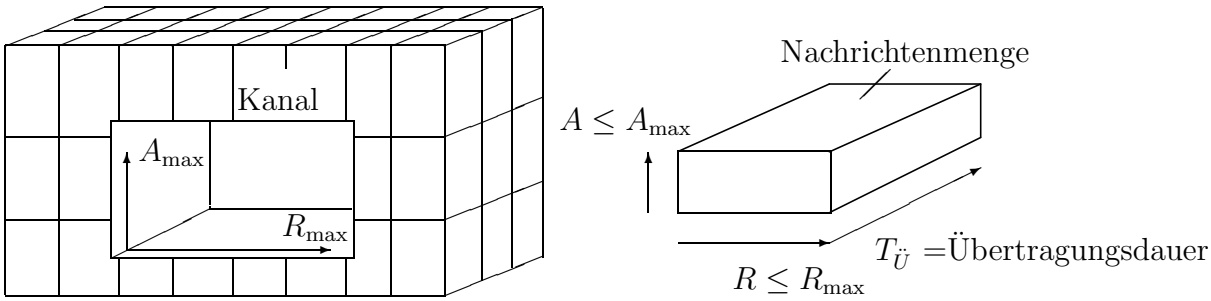


Bild 1.2: Zur Veranschaulichung von Kanalkapazität und Übertragungsdauer

Informationstheoretisch:

Äquivokation: Informationsverlust durch Kanal
Irrelevanz: durch Kanal hinzugefügte Information
Transinformation: Nutzinformationsfluss

$\longrightarrow$*Kanalkapazität* C = maximal möglicher mittlerer Nutzinformationsfluss/T, s. Bild 1.3.

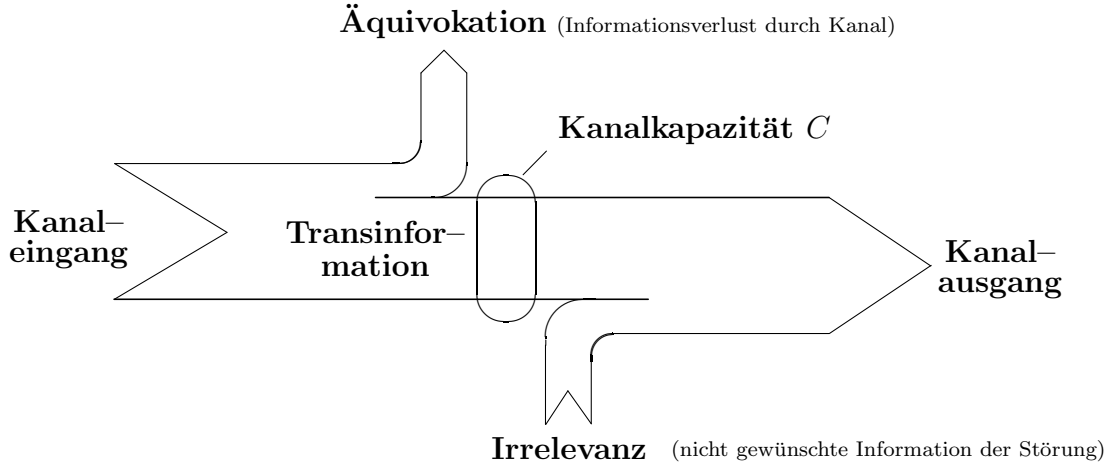


Bild 1.3: Zusammenhang zwischen Äquivokation, Irrelevanz und Transinformation

Kanalcodierer (channnel code)

Zuverlässige Nachrichtenübertragung trotz Kanalstörung
Idee: gezielt Redundanz hinzufügen = Kontrollinformation zur:

- Fehlererkennung: nur bestimmte Zeichenfolgen (= Codewörter CW) sind zulässig; bei unzulässigen CW: neu anfordern (Automatic Repeat on Request: ARQ)
- Fehlerkorrektur: Decodierer kann bis zu t-Fehler korrigieren = das nächstliegende zulässige CW ausgeben (Forward Error Correction: FEC)

Nachteil: Verlust an Übertragungsgeschwindigkeit

Verfahren: Betrachte „Blöcke“ aus k Info-tragenden Zeichen + r Prüfzeichen:
übertragen werden $n = k + r$-lange Zeichen = Codewörter,
Coderate: $R := \frac{k}{n}$.
Blockcode = blockweise Übertragung ohne Codierergedächtnis
Faltungscodes = fortlaufende Übertragung mit Coderate $R = \frac{k}{n}$,
Codierer hat Gedächtnis

Kanalcodierungssatz von Shannon:

Wenn $R < C$, dann existiert eine Codierung, so dass beliebig sichere Übertragung möglich ist.
Keine fehlerfreie Übertragung für $R > C$.

Wie? $\longrightarrow$ Codierungstheorie

Weitere Maßnahmen bei Bündelfehlern:
Kanal ist temporär für aufeinanderfolgende CW gestört (Abschattung bei Mobilfunk, Kratzer in CD):
Aufbrechen von Bündelfehler in Einzelfehler durch senderseitiges Blockinterleaving (Verflechtung) und empfangsseitiges Deinterleaving, s. Bild 1.4.
(Bei CD: (32,28) Reed-Solomon-Code (RS), Faltungsinterleaving über 28 Blöcke,
(28,24) RS-Code; beide RS-Codes können 2 Bytefehler erkennen und korrigieren und 4 erkannte Bytefehler korrigieren; Gesamtcoderate $R = \frac{3}{4}$)

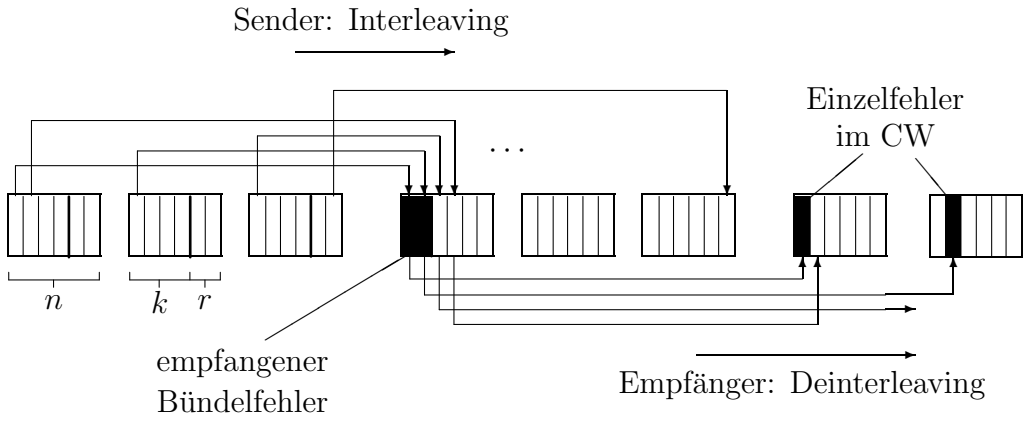


Bild 1.4: Interleaving

Kanaldecodierer (channel decoder)

Rekonstruktion der k Info-Zeichen aus n empfangenen und gestörten Signalen mit Hilfe der r redundanten Zeichen und dem Codiergesetz:

1.) richtige Rekonstruktion bis max. t Übertragungsfehler

2.) keine Rekonstruktion möglich, aber Fehler erkannt

3.) falsche Rekonstruktion und/oder Fehler nicht erkannt → Decodierfehler; besser: Zuverlässigkeitsinformation (reliability information) ausgeben

Quellendecodierer (source decoder)

- Rückübersetzung in Nachrichten für die Nachrichtensenke (z.B. LPC-Vocoder)
- Rekonstuktionssstategien bei unzuverlässigen Zeichen (z.B. Interpolation bei CD)

Problem, das hier nicht betrachtet wird: Synchronisation

- Taktsynchronisation (Abtastzeitpunkt, Taktfrequenz)
- Blocksynchronisation (Codewortanfang)
- Rahmensynchronisation (Protokoll, Interpretationsrahmen)

2 Nachrichtenquelle und Entropie

2.1 Informationsgehalt

Informationsgehalt $I(a_i)$ einer Nachricht $a_i \equiv$ Seltenheitswert

- $I(a_i)$ ist um so größer, je seltener a_i gesendet wird.
  - Sicheres Ereignis: $P(a_i) = 1 \quad \Rightarrow I(a_i) = 0$
- Forderungen:
  1.) $I(a_i) = f[P(a_i)]; \quad f$: stetig, $P(a_i) \neq 0$
  2.) $P(a_i) < P(a_j) \Rightarrow I(a_i) > I(a_j)$;
  $P(a_i) = P(a_j) \Rightarrow I(a_i) = I(a_j)$
  3.) Wenn a_i, a_j unabhängige Nachrichten mit $P(a_i, a_j) = P(a_i) \cdot P(a_j)$, dann $I(a_i, a_j) = I(a_i) + I(a_j)$

Def.: *Informationsgehalt* (Entscheidungsgehalt) einer Nachricht (R. Hartley) : Sei a_i eine mit der Wahrscheinlichkeit $P(a_i)$ gesendete Nachricht. Dann besitzt die Nachricht a_i bezüglich einer Basis r den Informationsgehalt

$$\boxed{I_r(a_i) := -\log_r P(a_i) = \log_r \frac{1}{P(a_i)}}$$

Maßeinheiten:

$$\begin{aligned} r = 2 \quad &: \quad I_2(a_i) \quad = \quad \mathrm{ld}\frac{1}{P(a_i)} \quad [\text{bit}] \\ r = \mathrm{e} \quad &: \quad I_{\mathrm{e}}(a_i) \quad = \quad \ln \frac{1}{P(a_i)} \quad [\text{nat}] \\ r = 10 \quad &: \quad I_{10}(a_i) \quad = \quad \lg \frac{1}{P(a_i)} \quad [\text{Hartley}] \end{aligned}$$

Anmerkung: Die Maßeinheit des Informationsgehalts legt die Basis des Logarithmus fest; allgemein: $I(a) = -\log P(a)$.

Aufgabe : Erfüllt die Definition die Forderungen 1-3 ?

1.) Da $0 < P(a_i) \leq 1 \Rightarrow x > 0$. $\log x$ ist stetige, monoton wachsende Funktion für $x > 0$.

2.) $P(a_i) < P(a_j) \Leftrightarrow \dfrac{1}{P(a_i)} > \dfrac{1}{P(a_j)} \overset{\log}{\Rightarrow} \log \dfrac{1}{P(a_i)} > \log \dfrac{1}{P(a_j)} \Leftrightarrow I(a_i) > I(a_j)$

3.) $I(a_i, a_j) = \log \frac{1}{P(a_i, a_j)}\Big|_{\text{stat. unabh.}} = \log \frac{1}{P(a_i) \cdot P(a_j)} = \log \frac{1}{P(a_i)} + \log \frac{1}{P(a_j)}$

$= I(a_i) + I(a_j).$ □

Eigenschaften des Logarithmus:

- $\log_r x = \frac{\log_b x}{\log_b r}$
- $\log x + \log y = \log(x \cdot y)$
- $a \cdot \log x = \log x^a$
- $\frac{d}{dx} \ln x = \frac{1}{x}$ für $x > 0$
- $\lim_{x \to 0}(x \cdot \log x) = 0$
- $\ln x \leq x - 1$

Zu zeigen: $f(x) = \ln x - (x - 1) \leq 0$

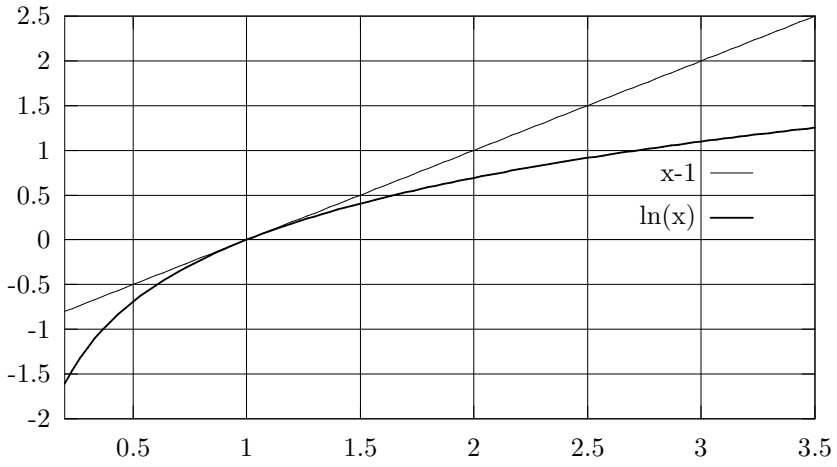


Bild 2.1: Zur Veranschaulichung von $x - 1$ und $\ln x$

Beweis: $f'(x) = \frac{1}{x} - 1 \Rightarrow f'(x = 1) = 0$ ist Extremum.

Da $\lim_{x \to 0} f(x) = \lim_{x \to \infty} f(x) \to -\infty \Rightarrow f(x) \leq 0.$ □

Allgemein: • $\log_r x \leq \frac{(x-1)}{\ln r}$

Bsp.: Informationsgehalt Würfelwurf:

$$P(a_i) = \frac{1}{6} \to I_2(a_i)/[\text{bit}] = \text{ld}\, \frac{1}{\frac{1}{6}} = \text{ld}\, 6 \approx 2.6$$

2.2 Entropie

Mittlerer Informationsgehalt der Nachrichten einer Quelle = **Entropie**.

2.2.1 Gedächtnislose Quelle

$\mathbf{A} = \{a_1, \ldots, a_q\}$; $(P(a_i), \ldots, P(a_q))$; $\quad a_i$: stat. unabhängig, $\sum_{i=1}^{q} P(a_i) = 1$.

Def.: Entropie einer gedächtnislosen q-nären Quelle:

$$\boxed{H_r(A) := \mathrm{E}[I_r(a_i)] = \sum_{i=1}^{q} P(a_i) \cdot I_r(a_i) = \sum_{i=1}^{q} P(a_i) \cdot \log_r \frac{1}{P(a_i)}}$$

$r = 2$: [bit]
$r = \mathrm{e}$: [nat]
$r = 10$: [Hartley]

Eigenschaften:

- $H_r(A) \geq 0$;
  $H_r(A) = 0 \;\Leftrightarrow\; P(a_k) = 1,\; P(a_{i \neq k}) = 0, \quad 1 \leq k \leq q$;
  d.h. keine Unsicherheit über die Nachrichtenquelle

- $0 \leq H_r(A) \leq \log_r q$

  Zu zeigen: $H_r(A) - \log_r q \leq 0$

  Beweis:

$$\begin{aligned}
\sum_{i=1}^{q} P(a_i) \cdot \log_r \frac{1}{P(a_i)} - \underbrace{\sum_{i=1}^{q} P(a_i)}_{=1} \cdot \log_r q & \\
&= \sum_{i=1}^{q} P(a_i) \log_r \frac{1}{q \cdot P(a_i)} \\
&\leq \sum_{i=1}^{q} P(a_i) \frac{\frac{1}{q \cdot P(a_i)} - 1}{\ln r} \\
&= \frac{1}{\ln r} \Big(\underbrace{\sum_{i=1}^{q} \frac{P(a_i)}{q \cdot P(a_i)}}_{=1} - \underbrace{\sum_{i=1}^{q} P(a_i)}_{=1} \Big) = 0 \quad \text{qed.}
\end{aligned}$$

- $H_r(A) = \log_r q \;\Leftrightarrow\; 1 = \dfrac{1}{q \cdot P(a_i)} \;\forall\, i \;\Leftrightarrow\; P(a_i) = \dfrac{1}{q};\; \forall\, i = 1, \ldots, q$

Die Entropie einer Nachrichtenquelle ist maximal, wenn alle Nachrichten mit der gleichen Wahrscheinlichkeit $P(a_i) = \dfrac{1}{q}\; \forall\, i = 1, \ldots, q$ gesendet werden.

Bsp.: Binärquelle: $q = 2 \rightarrow \mathbf{A} = \{+, -\}; \quad P(+) = p, \quad P(-) = 1 - p$

Entropiefunktion: $\boxed{\mathrm{H}_r(p) = p \cdot \log_r \frac{1}{p} + (1-p) \cdot \log_r \frac{1}{1-p}}$

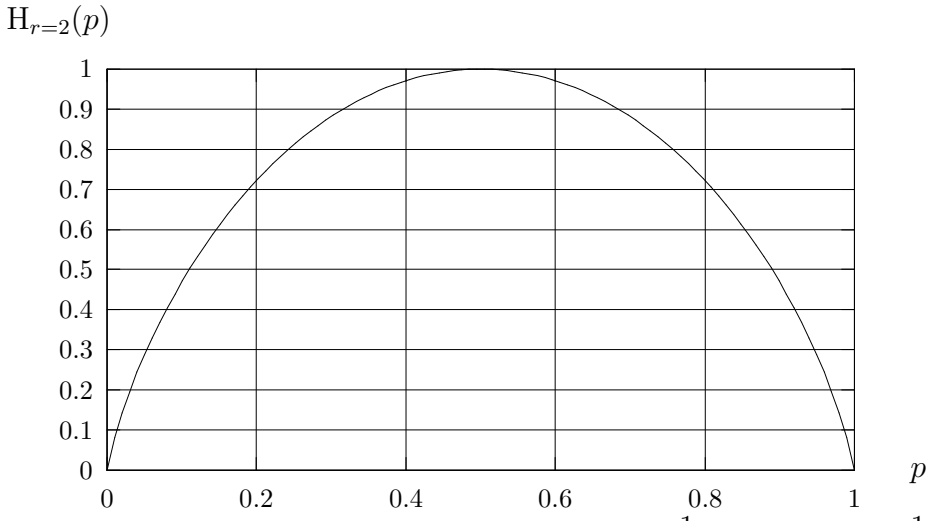


Bild 2.2: Entropiefunktion $\mathrm{H}_{r=2}(p) = p \cdot \mathrm{ld}\frac{1}{p} + (1-p) \cdot \mathrm{ld}\frac{1}{1-p}$

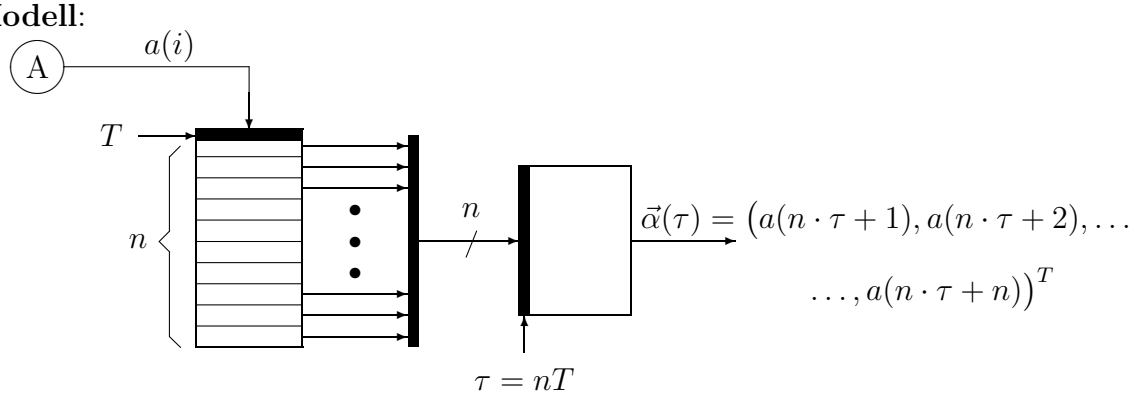
2.2.2 n-te Quellerweiterung

Nachrichtenblöcke: = n-te Erweiterung einer gedächtnislosen Quelle

Gegeben: gedächtnislose q-näre Quelle: $\mathbf{A} = \{a_1, a_2, a_2, \ldots, a_q\}$ und $P(a_j)$, $j = 1, \ldots, q$

Bilde: Blöcke (Teilfolgen) der Länge n
$\rightarrow \mathbf{A}^n$(n-te Erweiterung von A): gedächtnislose q^n-näre Quelle mit
$\vec{\alpha} = (a^{(1)}, \ldots, a^{(n)}); \vec{\alpha} \in \mathbf{A}^n = (\vec{\alpha}^{(1)}, \ldots, \vec{\alpha}^{(q^n)})$ und $P(\vec{\alpha}^{(k)})$; $k = 1 \ldots q^n$

Modell:



$$P(\vec{\alpha}_k) = P(a_{k_1}, a_{k_2}, \ldots, a_{k_n}) = P(a_{k_1}) \cdot P(a_{k_2}) \cdots P(a_{k_n}) = \prod_{i=1}^{n} P(a_{k_i})$$

Bild 2.3: Modell der n-ten Erweiterung einer gedächtnislosen Quelle

Entropie von A^n:

$$
\begin{aligned}
H(A^n) &= \sum_{\vec{\alpha}\in A^n} P(\vec{\alpha}_k)\cdot \log\frac{1}{P(\vec{\alpha}_k)} \\
&= \sum_{k_1=1}^{q}\cdots\sum_{k_n=1}^{q} P(a_{k_1})\cdots P(a_{k_n})\cdot\big(\log\frac{1}{P(a_{k_1})}+\cdots+\log\frac{1}{P(a_{k_n})}\big) \\
&= \sum_{k_1=1}^{q} P(a_{k_1})\cdot\log\frac{1}{P(a_{k_1})}\cdot\underbrace{\sum_{k_2=1}^{q} P(a_{k_2})\cdots\underbrace{\sum_{k_n=1}^{q} P(a_{k_n})}_{=1}}_{=1} \quad+\cdots \\
&\quad \ldots+\sum_{k_n=1}^{q} P(a_{k_n})\log\frac{1}{P(a_{k_n})}\cdot\underbrace{\sum_{k_1=1}^{q} P(a_{k_1})\cdots\underbrace{\sum_{k_{n-1}=1}^{q} P(a_{k_{n-1}})}_{=1}}_{=1} \\
&= \underbrace{\sum_{k_1=1}^{q} P(a_{k_1})\cdot\log\frac{1}{P(a_{k_1})}}_{=H(A)}+\cdots+\underbrace{\sum_{k_n=1}^{q} P(a_{k_n})\cdot\log\frac{1}{P(a_{k_n})}}_{=H(A)}
\end{aligned}
$$

$$\boxed{H(A^n)=n\cdot H(A)}$$

Die Entropie der n-ten Erweiterung der gedächtnislosen Quelle A^n ist n mal höher als die Entropie der Quelle A. A^n hat einen Nachrichtenvorrat aus q^n Elementen.

Bsp.: Quelle A: $\mathbf{A}=\{-,0,+\}$; $P(-)=\frac{1}{2}$; $P(+)=P(0)=\frac{1}{4}$.
Betrachte A^2:

$$
\begin{aligned}
H_2(A) &= \frac{1}{2}\mathrm{ld}\,2+\frac{1}{4}\mathrm{ld}\,4+\frac{1}{4}\mathrm{ld}\,4=\frac{3}{2} \\
H_2(A^2) &= \frac{1}{4}\mathrm{ld}\,4+4\cdot\frac{1}{8}\mathrm{ld}\,8+4\cdot\frac{1}{16}\mathrm{ld}\,16=\frac{2}{4}+\frac{12}{8}+\frac{16}{16}=3=2\cdot H_2(A)
\end{aligned}
$$

$\mathbf{A}^2$:	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
entspricht	$--$	-0	$-+$	$0-$	00	$0+$	$+-$	$+0$	$++$
$P(\alpha_i)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$

2.2.3 Markoff-Quelle (Quelle mit Gedächtnis)

Bei einer q-nären Markoff-Quelle M mit der Einflusslänge $L+1$ (Rückgrifftiefe = Gedächtnistiefe L) hängt die Wahrscheinlichkeit der Nachricht zum Zeitpunkt $i, a(i) \in \mathbf{A}$ von den letzten L gesendeten Nachrichten ab:

$$P\Big(a(i)|\underbrace{\big(a(i-1), a(i-2), \ldots, a(i-L)\big)}_{\vec{z}(i)}\Big) = P\big(a(i)|\vec{z}(i)\big)$$

Zustand der Quelle: $\vec{z}(i) = (a(i-1), \ldots, a(i-L))$; es gibt q^L verschiedene Zustände.

Folgezustand: $\vec{z}(i+1) = (a(i), a(i-1), \cdots, a(i-L+1))$
$\rightarrow$ gesendete Nachricht bedingt neuen Zustand

Äquivalente Beschreibung: $P\left(a(i)|\vec{z}(i)\right) = P\left(\vec{z}(i+1)|\vec{z}(i)\right)$

Modell:

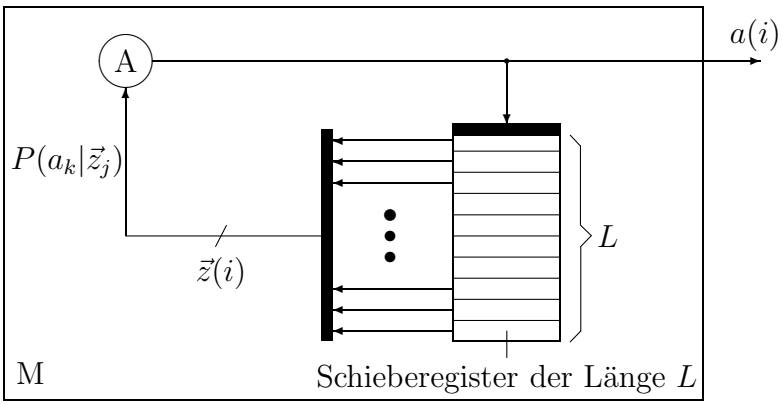


Bild 2.4: Modell einer Markoff-Quelle

Entropie einer Markoff-Quelle:
Informationsgehalt für einen Zustand $\vec{z}_j$: $I(a_k|\vec{z}_j) = \log \dfrac{1}{P(a_k|\vec{z}_j)}$

Mittlerer Info-Gehalt für einen Zustand $\vec{z}_j$:

$$H(A|\vec{z}_j) = \mathrm{E}_{a_k}[I(a_k|\vec{z}_j)] = \sum_{k=1}^{q} P(a_k|\vec{z}_j) \cdot I(a_k|\vec{z}_j)$$

Mittlerer Info-Gehalt der Markoff-Quelle über alle Zustände $\vec{z}_j$: $H(A) = \mathrm{E}_{\vec{z}_j}[H(A|\vec{z}_j)]$

$$\begin{aligned} H(A) &= \sum_{j=1}^{q^L} P(\vec{z}_j) \cdot \sum_{k=1}^{q} P(a_k|\vec{z}_j) \cdot I(a_k|\vec{z}_j) = \sum_{k=1}^{q} \sum_{j=1}^{q^L} P(a_k, \vec{z}_j) \cdot I(a_k|\vec{z}_j) \\ &= \sum_{a(i)} \sum_{a(i-1)} \cdots \sum_{a(i-L)} \Bigg(P\left(a(i), a(i-1), \ldots, a(i-L)\right) \cdot \\ &\qquad \log \frac{1}{P\left(a(i)|\left(a(i-1), a(i-2), \ldots a(i-L)\right)\right)} \Bigg) \end{aligned}$$

Beschreibung : Zustandsgraph mit Übergangswahrscheinlichkeiten

Bsp.: $q = 2,\ L = 2,\ \mathbf{A} = \{-, +\}$
Es gelte: $P\Big(a(i)|\big(a(i-1), a(i-2)\big)\Big)$

$$\begin{array}{rcccccl} P(-|--) &=& P(+|++) &=& 0,8 \\ P(+|--) &=& P(-|++) &=& 0,2 \\ P(-|-+) &=& P(+|-+) &=& \\ && = P(-|+-) &=& P(+|+-) &=& 0,5 \end{array}$$

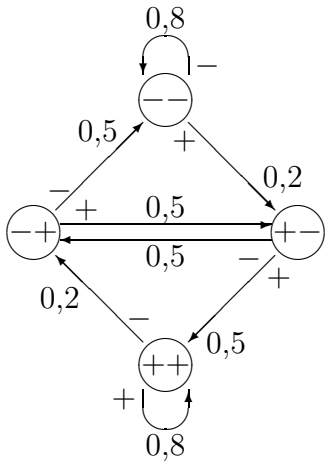


Bild 2.5: Zustandsübergangswahrscheinlichkeiten

Frage: Mit welcher Wahrscheinlichkeit befindet sich die Quelle im Zustand $\vec{z}_j$?
Annahme: stationäre Zustandsverteilung; d.h. $P(\vec{z}(i)) = P\big(\vec{z}(i+1)\big)$

$$\left.\begin{array}{rcl} P\underbrace{(--)}_{\vec{z}(i+1)} &=& \overbrace{P(-|--)}^{P(--|--)} \cdot P\underbrace{(--)}_{\vec{z}(i)} + \overbrace{P(-|-+)}^{P(--|-+)} \cdot P\underbrace{(-+)}_{\vec{z}(i)} \\ &\vdots& \\ P(++) &=& P(+|++) \cdot P(++) + P(+|+-) \cdot P(+-) \end{array}\right\} \begin{array}{l} 4\,\text{Gleichungen,} \\ \text{nur 3 lin.} \\ \text{unabhängig!} \end{array}$$

$$P(++) + P(+-) + P(-+) + P(--) = 1$$

$\Bigg\}$ 4 linear unabhängige Gleichungen, i.a. lösbar

Allgemein:

Zustandsverteilung:

$$\vec{P}(i) := \left[P\big(\vec{z}_1(i)\big), P\big(\big(\vec{z}_2(i)\big), \ldots, P\big(\vec{z}_{q^L}(i)\big)\right]$$

Markoff-Matrix:

$$[M] := \begin{bmatrix} P(\vec{z}_1|\vec{z}_1), & P(\vec{z}_2|\vec{z}_1), \ldots & P(\vec{z}_{q^L}|\vec{z}_1) \\ \vdots & & \vdots \\ P(\vec{z}_1|\vec{z}_{q^L}), & \ldots & P(\vec{z}_{q^L}|\vec{z}_{q^L}) \end{bmatrix}$$

Anfangsverteilung: $\vec{P}(0)$

Es gilt: $\vec{P}(i) = \vec{P}(i-1) \cdot [M] = \vec{P}(0) \cdot [M]^i$, da $\vec{P}(1) = \vec{P}(0) \cdot [M]$

$$\begin{aligned} \vec{P}(2) &= \vec{P}(1) \cdot [M] = \vec{P}(0) \cdot [M] \cdot [M] \\ &= \vec{P}(0) \cdot [M]^2 \\ &\text{usw.} \end{aligned}$$

Ergodische Markoff-Quelle:

Unabhängig von $\vec{P}(0)$ stellt sich eine stationäre Grenzverteilung $\vec{P}^\infty$ ein, d.h.

$$\begin{aligned} \vec{P}^\infty &= \lim_{i\to\infty} \vec{P}(0) \cdot [M]^i \ \ \forall\, \vec{P}(0) \\ \text{und folglich auch } \ \vec{P}^\infty &= \vec{P}^\infty \cdot [M]. \end{aligned}$$

Es gilt:

$$\lim_{i\to\infty} [M]^i = \begin{bmatrix} P_1^\infty & P_2^\infty & \ldots & P_{q^L}^\infty \\ \vdots & & & \vdots \\ P_1^\infty & \ldots & \ldots & P_{q^L}^\infty \end{bmatrix}.$$

Zum obigen Beispiel:

$$\begin{array}{lllllll} P(--|--) & = & P(++|++) & = & 0,8 \\ P(+-|--) & = & P(-+|++) & = & 0,2 \\ P(--|-+) & = & P(+-|-+) & = & P(-+|+-) \\ & & & = & P(++|+-) = 0,5\,, \end{array}$$

für die übrigen Wahrscheinlichkeiten gilt: $P\big(z(i+1)|z(i)\big) = 0$.

Somit ist:

$$\begin{aligned} \vec{P} &= \big(P(--), P(-+), P(+-), P(++)\big) \\ [M] &= \begin{bmatrix} 0,8 & 0 & 0,2 & 0 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,5 & 0 & 0,5 \\ 0 & 0,2 & 0 & 0,8 \end{bmatrix}. \end{aligned}$$

Stationäre Zustandsverteilung $\vec{P}^{\infty} = \vec{P}^{\infty} \cdot [M]$ mit $\sum\limits_{\forall\, i} P^{\infty}(z_i) = 1$ liefert:

$$\vec{P}^{\infty} = \left(\frac{5}{14}, \frac{2}{14}, \frac{2}{14}, \frac{5}{14}\right) \quad \square$$

Entropie der Markoff-Quelle:

$$\begin{aligned} H(A)/bit &= \sum_{\vec{z}\in\{--,-+,+-,++\}} P(\vec{z}) \cdot \sum_{a\in\{-,+\}} P(a|\vec{z}) \cdot \mathrm{ld}\frac{1}{\mathrm{P(a|\vec{z})}} \\ &= \frac{5}{14}\left(0,8 \cdot \mathrm{ld}\frac{1}{0,8} + 0,2 \cdot \mathrm{ld}\frac{1}{0,2}\right) + \frac{2}{14}\left(0,5 \cdot \mathrm{ld}\frac{1}{0,5} + 0,5 \cdot \mathrm{ld}\frac{1}{0,5}\right) + \cdots \\ &\quad \cdots + \frac{2}{14}\left(0,5 \cdot \mathrm{ld}\frac{1}{0,5} + 0,5 \cdot \mathrm{ld}\frac{1}{0,5}\right) + \frac{5}{14}\left(0,2 \cdot \mathrm{ld}\frac{1}{0,2} + 0,8 \cdot \mathrm{ld}\frac{1}{0,8}\right) \\ &= \frac{1}{7}\Big(5 \cdot \mathrm{H}_2(0,2) + 2 \cdot \mathrm{H}_2(0,5)\Big) \\ &\approx 0,8 \end{aligned}$$

3 Quellencodierung

3.1 Codierung

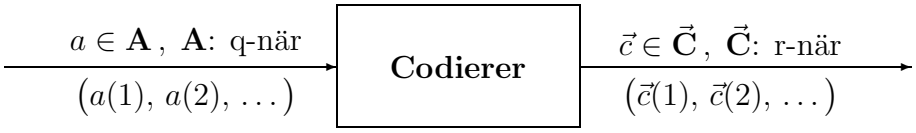


Bild 3.1: Modell eines Codierers

3.1.1 Definition Codierung

Sei $\mathbf{A}= \{a_1, \ldots, a_q\}$ eine q-näre Quelle und $\mathbf{C}= \{c_1, \ldots, c_r\}$ eine r-näre Symbolmenge. Durch die Codierung $\mathbf{C}$ wird jedem $a_i \in \mathbf{A}$ eine Sequenz von $c_k \in \mathbf{C}$, dem Codewort (CW), $\vec{c}_j \in \vec{\mathbf{C}}$, zugeordnet:

$$\mathbf{C}: \; a_i \in \mathbf{A} \longmapsto \vec{c}_j \in \vec{\mathbf{C}} \backslash \emptyset$$

Beispiel: $q = 4$

	$\mathbf{C}^1$	$\mathbf{C}^2$	$\mathbf{C}^3$	$\mathbf{C}^4$	$\mathbf{C}^5$	$\mathbf{C}^6$
a_1	$-$	$--$	$-$	$-$	$-$	$-$
a_2	$+-$	$-+$	$+--$	$+-$	$-+$	$+-$
a_3	$--$	$+-$	$++-$	$++-$	$-++$	$++-$
a_4	$-+$	$++$	$++$	$+++-$	$+++$	$+++$
Blockcode	nein	ja	nein	nein	nein	nein
eindeutig decodierbar	nein	ja	nein	ja	ja	ja
$\sum_{i=1}^{n}(\frac{1}{2})^{l_i}$	$1\frac{1}{4} \geq 1$	1	1 (!?)	$\frac{15}{16} \leq 1$	1	1
sofort decodierbar	nein	ja, Blockcode	nein	ja, Kommacode	nein	ja

- $\mathbf{C}$ ist ein *Blockcode* der Länge n, wenn alle CW die gleiche Komponentenanzahl besitzen, d.h. $\vec{c}_j \in \vec{\mathbf{C}} = \mathbf{C}^n$. (Sonst: Code variabler Länge, siehe Beispiel)
- Kann aus jeder Folge der CW: $(\vec{c}(1), \vec{c}(2), \ldots)$ eindeutig auf die Ursprungsfolge $\big(a(1), a(2), \ldots\big)$ geschlossen werden, so heißt der Code *eindeutig decodierbar*.
  Im Beispiel:

$\mathbf{C}^1$: CW : $\overbrace{-}\;\overbrace{+-}\;\overbrace{--}\;\overbrace{-}$ nicht eindeutig

$\mathbf{C}^3$: CW : $\overbrace{++}\;\overbrace{-}$ nicht eindeutig

sonst : eindeutig

→ Im weiteren nur noch eindeutig decodierbare Codes!

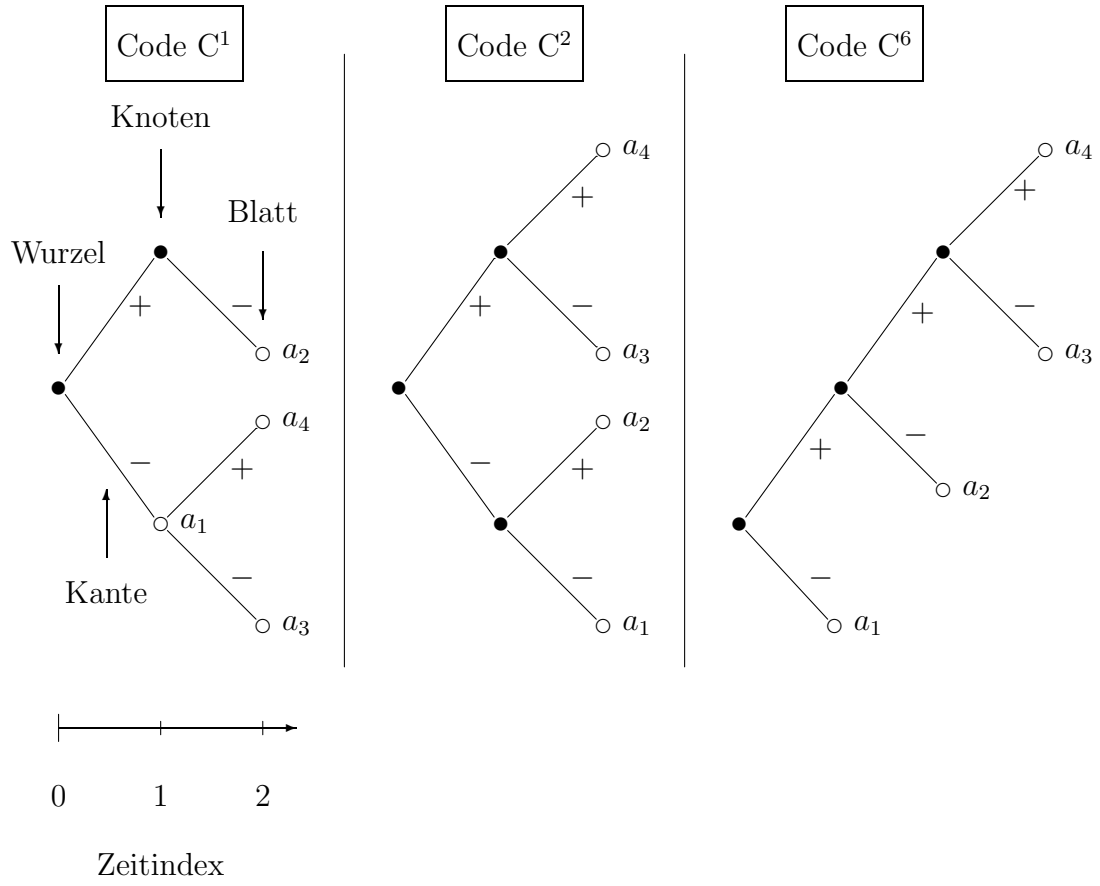


Bild 3.2: Darstellung von Codes im Codebaum

3.1.2 Notwendige Bedingung für eindeutige Decodierbarkeit

Sei $\mathbf{C}$ ein r-närer Code mit den CW $\vec{c}_j \in \vec{\mathbf{C}},\ j = 1, \ldots, u$ und den Codewortlängen l_j (= Anzahl der Komponenten der CW). Wenn $\mathbf{C}$ eindeutig decodierbar ist, dann gilt:

$$\sum_{j=1}^{u} \left(\frac{1}{r}\right)^{l_j} \leq 1 \qquad \text{(Ungleichung von Kraft und McMillan).}$$

Beweis: Betrachte $m \in \mathbf{N}$ verschiedene CW: $\vec{c}(1), \vec{c}(2), \ldots \vec{c}(m)$.
Sei $z_{l,m}$ die Anzahl der CW der Länge l in dieser CW-Folge $(\vec{c}(1), \vec{c}(2), \ldots \vec{c}(m))$.
Es gibt r^l verschiedene CW der Länge l. Da eindeutig decodierbar, kann jedes der verschiedenen CW der Länge l höchstens je einem in der Folge der m CW zugeordnet sein. $\Rightarrow z_{l,m} \leq r^l$

Betrachte:

$$\begin{aligned}
\left(\sum_{j=1}^{u}\left(\frac{1}{r}\right)^{l_j}\right)^m &= \overbrace{\sum_{j_1=1}^{u}\cdots\sum_{j_m=1}^{u}}^{m}\left(\frac{1}{r}\right)^{\overbrace{(l_{j1}+l_{j2}+l_{j3}+\cdots+l_{jm})}^{l}} \\
&= \sum_{l=m\cdot l_1}^{m\cdot l_u}\underbrace{z_{l,m}}_{\leq r^l}\left(\frac{1}{r}\right)^l \\
&\leq \sum_{l=m\cdot l_1}^{m\cdot l_u} r^l\cdot\left(\frac{1}{r}\right)^l = \sum_{l=m\cdot l_1}^{m\cdot l_u} 1 = 1+m\cdot l_u - m\cdot l_1 \leq m\cdot l_u \\
\Rightarrow \sum_{j=1}^{u}\left(\frac{1}{r}\right)^{l_j} &\leq m^{\frac{1}{m}}\cdot l_u^{\frac{1}{m}} = \exp\left(\frac{1}{m}\ln m\right)\cdot\exp\left(\frac{1}{m}\ln l_u\right) = \exp\left(\frac{1}{m}(\ln m+\ln l_u)\right)
\end{aligned}$$

Da auch für $m\to\infty$:

$$\sum_{j=1}^{n}\left(\frac{1}{r}\right)^{l_j} \leq \lim_{m\to\infty}\exp\left(\frac{1}{m}(\ln m+\ln l_u)\right) = \exp\overbrace{\left(\lim_{m\to\infty}\frac{1}{m}(\ln m+\ln l_u)\right)}^{\to 0} = 1 \quad \square$$

Bei Binärcodes ($r=2$) : $\displaystyle\sum_{j=1}^{u}\left(\frac{1}{2}\right)^{l_j}\leq 1$

3.1.3 Sofort decodierbare Codes

Sofort decodierbare Codes werden auch "Präfix–Codes" oder "Instantaneous Codes" genannt.

Def.: Ein Code heißt *sofort decodierbar*, wenn aus der laufenden CW-Folge vom Folgenbeginn an wortweise decodiert werden kann, ohne nachfolgende CW beachten zu müssen.

$\vec{\alpha}=(\alpha_1,\alpha_2,\alpha_3,\ldots,\alpha_m)$ ist ein Präfix von $\vec{\beta}=(\beta_1,\beta_2,\ldots,\beta_n)$ mit $m<n$, wenn $\alpha_i=\beta_i\;\forall\, i=1,\ldots,m$.

Im Beispiel: $\vec{\mathbf{C}}^1$: $a_1\widehat{=}(-)$ ist Präfix von $a_3\widehat{=}(--)$ und $a_4\widehat{=}(-+)$

Fano-Bedingung
Ein Code ist sofort decodierbar, wenn kein Codewort das Präfix eines anderen ist.
(CW sind Blätter im Codebaum.)

Gilt offensichtlich für • Blockcodes (im Beispiel: $\vec{\mathbf{C}}^2$) und
• Komma-Codes (es gibt ein CW-Schlusszeichen, im Beispiel $\vec{\mathbf{C}}^4$).

Im Beispiel: $\vec{\mathbf{C}}^1, \vec{\mathbf{C}}^3$ und $\vec{\mathbf{C}}^5$ sind nicht sofort decodierbar.

Ein eindeutig sofort decodierbarer Code mit den Codewortlängen $l_j,\ j = 1, \ldots, u$ existiert genau dann, wenn gilt

$$\boxed{\sum_{j=1}^{u} \left(\frac{1}{r}\right)^{l_j} \leq 1.}$$

D.h., die Ungleichung von Kraft u. McMillan ist notwendig und hinreichend für die Existenz eines sofort decodierbaren eindeutigen Codes.

Beachte: Der Satz bezieht sich nicht auf eine spezielle Codierung (Festlegung von Codeworten). Erfüllt ein Code die Ungleichung wie im Beispiel $\mathbf{C}^3$, so kann die gewählte Codierung trotzdem nicht eindeutig / sofort decodierbar sein. Es muss aber eine eindeutig sofort decodierbare Codierung mit dieser CW–Längenverteilung geben (im Beispiel $\mathbf{C}^6$).

Zu jedem eindeutig decodierbaren Code gibt es einen sofort decodierbaren Code derselben Ordnung r, derselben Codewortzahl m und derselben Codewortlängen l_i, $i = 1, \ldots, u$.
Im Beispiel: $\mathbf{C}^3$, $\mathbf{C}^5$ und $\mathbf{C}^6$.

3.2 Effizienz einer Quellencodierung

Betrachte: Quelle A: a_i mit $P(a_i) = P_i$, q-när.
Quellcode $\mathbf{C}$: $a_i \longmapsto \vec{c}_i$: CW mit $l_i(C)$ Komponenten, r-när.

Mittlere Codewortlänge: $L(C) = \mathrm{E}[l_i(C)]$: mittlere Anzahl der CW zur Codierung der Nachrichten der Quelle $\mathbf{A}$

$$L(C) = \sum_{i=1}^{q} P_i \cdot l_i(C)$$

Def.: Eine Quellencodierung $\mathbf{C}$ heißt *kompakt* (optimal), wenn keine andere Codierung $\mathbf{C'} \neq \mathbf{C}$ existiert, so dass $L(C') < L(C)$.

3.2.1 Mittlere Codewortlänge

Ermittlung der minimalen mittleren Länge für sofort decodierbare eindeutige Codes aus der Quellentropie $H(A)$

Hilfssatz: Sei $\sum_{i=1}^{q} P_i = 1, \sum_{i=1}^{q} Q_i = 1,$ dann gilt: $\sum_{i=1}^{q} P_i \cdot \log_r \frac{1}{P_i} \leq \sum_{i=1}^{q} P_i \cdot \log_r \frac{1}{Q_i}$.

Zu zeigen: $\sum_{i=1}^{q} P_i \cdot \log_r \frac{Q_i}{P_i} \leq 0$

Beweis: es gilt $\log_r x \leq \frac{x-1}{\ln r}$

$$\sum_{i=1}^{q} P_i \cdot \log_r \frac{Q_i}{P_i} \leq \frac{1}{\ln r} \cdot \sum_{i=1}^{q} P_i \cdot \left(\frac{Q_i}{P_i} - 1 \right) = \frac{1}{\ln r} \cdot \left(\sum_{i=1}^{q} Q_i - \sum_{i=1}^{q} P_i \right) = 0.$$

Gleichheit gilt dann, wenn $P_i = Q_i \ \forall\ i$ □

Entropie der Quelle A:

Sei $Q_i := \left(\frac{1}{r}\right)^{l_i} \cdot \left(\sum_{j=1}^{q} \left(\frac{1}{r}\right)^{l_j} \right)^{-1}$.

Dann ist

$$\begin{aligned} H_r(A) &= \sum_{i=1}^{q} P_i \cdot \log_r \frac{1}{P_i} \leq \sum_{i=1}^{q} P_i \cdot \log_r \frac{1}{Q_i} = \sum_{i=1}^{q} P_i \cdot \log_r \frac{\sum_{j=1}^{q} r^{-l_j}}{r^{-l_i}} \\ &\leq \sum_{i=1}^{q} P_i \cdot \log_r r^{l_i} + \sum_{i=1}^{q} P_i \cdot \log_r \sum_{j=1}^{q} r^{-l_j} = \underbrace{\sum_{i=1}^{q} l_i \cdot P_i}_{=L} \cdot \underbrace{\log_r r}_{=1} + \underbrace{\log_r \sum_{j=1}^{q} r^{-l_j}}_{\leq 0} \end{aligned}$$

$\Rightarrow\ H_r(A) \leq L$.

Gleichheit (d.h. optimal), wenn:

1.) $\sum_{j=1}^{q} r^{-l_j} = 1$ und

2.) $P_i = Q_i = (\frac{1}{r})^{l_i} \ \forall\ i$

$\Longrightarrow\ l_i = \log_r \frac{1}{P_i}$;
da $l_i \in \mathbf{N} \Rightarrow$ nur wenn $P_i = (\frac{1}{r})^{l_i}$ gilt Gleichheit.

Wenn $P_i \neq (\frac{1}{r})^{l_i}$, dann nächst größere ganze Zahl: $\log_r \frac{1}{P_i} \leq l_i \leq 1 + \log_r \frac{1}{P_i}$, wobei dann nicht gewährleistet ist, dass ein kompakter Code entsteht.

Man erhält somit für die mittlere Länge:

$$\begin{array}{ccccccc} \sum_{i=1}^{q} P_i \cdot \log_r \frac{1}{P_i} & \leq & \sum_{i=1}^{q} P_i \cdot l_i & \leq & \sum_{i=1}^{q} P_i & + & \sum_{i=1}^{q} P_i \cdot \log_r \frac{1}{P_i} \\ H_r(A) & \leq & L & \leq & 1 & + & H_r(A) \end{array}$$

Anmerkung: Wenn $(\frac{1}{r})^{l_i} \leq P_i$, ist die Ungleichung von Kraft und McMillan

$$\sum_{i=1}^{q} \left(\frac{1}{r}\right)^{l_i} \leq \sum_{i=1}^{q} P_i = 1$$

stets erfüllt, d.h. es gibt kompakte sofort decodierbare eindeutige Codes.

Effizienzsteigerung:
durch Codierung von Nachrichtenblöcken (n-te Erweiterung $A^n : \alpha_j \in \mathbf{A}^n$)
Mittlere Länge der Codierung bei n-ter Erweiterung: $L_n = \sum_{j=1}^{q^n} P(\alpha_j) \cdot \lambda_j$,
wobei λ_j: Länge des zu α_j gehörenden CW.

$$\begin{array}{ccccc} H_r(A^n) & \leq & L_n & \leq & H_r(A^n) + 1 \\ n \cdot H_r(A) & \leq & L_n & \leq & n \cdot H_r(A) + 1 \\ H_r(A) & \leq & \frac{L_n}{n} & \leq & H_r(A) + \frac{1}{n} \end{array}$$

$\frac{L_n}{n}$: mittlere Komponentenanzahl der CW bei n Nachrichtenblöcken.

Ergebnis: $\lim_{n\to\infty} \frac{L_n}{n} = H_r(A)$; d.h. mittlere CW-Länge strebt gegen Quellentropie, wenn Nachrichtenblöcke unendlich lang.

Folgerung:

Quellcodierungssatz von Shannon:
Die mittlere Anzahl von r-nären Codesymbolen, die zur sofort decodierbaren eindeutigen Codierung der Nachrichten einer gedächtnislosen Quelle A benötigt werden, kann nie kleiner sein als die Entropie $H_r(A)$ der Quelle A.

3.2.2 Effizienz (Datenkompression) eines Codes C

Die Effizienz eines Codes ist der Bruchteil der Quellenentropie, der im Mittel auf eine Komponente eines CW's entfällt. → Kompakte Codes haben maximale Effizienz.

$$\eta(C) = \frac{H_r(A)}{L(C)}$$

Redunanz: $red := \frac{L - H_r(A)}{L} = 1 - \eta$

Beachte: Shannon: für größere Blöcke A^n, $n \to \infty \Rightarrow \eta = 1$ und $red = 0$

Bsp.: Quelle A: $\mathbf{A} = \{+, -\};\ P(+) = 0,1,\ P(-) = 0,9$

– Entropie: $H_2(A) = 0,1 \cdot \mathrm{ld}\, 10 + 0,9 \cdot \mathrm{ld}\, \frac{10}{9} \approx 0,47$

- triviale Codierung $\mathbf{C}^1 :\ \mathbf{C} = \{0, 1\},\ r = 2$

a_i	c_i
$+$	1
$-$	0

} kompakter Code bei trivialer Codierung

– mittlere CW-Länge: $L(C^1) = 1 \cdot 0,1 + 1 \cdot 0,9 = 1$

– Effizienz: $\eta(C^1) = \dfrac{H(A)}{L(C^1)} = \dfrac{0,47}{1} \ \widehat{=}\ 47\%$

- 2-te Erweiterung: A^2

– Effizienz Blockcode der Länge 2: $\eta(C_{2Block}) = \dfrac{2 \cdot H(A)}{2 \cdot L(C^1)} = \ \widehat{=}\ 47\%$

– Blockweise Codierung variabler CW-Länge:

C^2 :

j	$\vec{a}_j$	P_j	l_j
1	$--$	0,81	1
2	$-+$	0,09	4
3	$+-$	0,09	4
4	$++$	0,01	7

} Code ist nicht kompakt! (siehe C^3)

Optimale CW-Länge l_j für die Einzelwahrscheinlickeiten P_j :
$\mathrm{ld}\, \frac{1}{P_j} \le l_j \le \mathrm{ld}\, \frac{1}{P_j} + 1,\ l_j \in \mathbf{N}$

$$\begin{aligned} P_1 = 0,81 \quad &\Rightarrow \quad \mathrm{ld}\, \tfrac{1}{0,81} \le 1 \quad &\Rightarrow \quad l_1 = 1 \\ P_2 = P_3 = 0,09 \quad &\Rightarrow \quad \mathrm{ld}\, \tfrac{100}{9} \approx 3,45 \quad &\Rightarrow \quad l_{2,3} = 4 \\ P_4 = 0,01 \quad &\Rightarrow \quad \mathrm{ld}\, 100 \approx 6,64 \quad &\Rightarrow \quad l_4 = 7 \end{aligned}$$

– $L(C^2) = \sum\limits_j l_j \cdot P_j = 1 \cdot 0,81 + 2 \cdot (4 \cdot 0,09) + 7 \cdot 0,01 \approx 1,6$

– Effizienz: $\eta(C^2) = \dfrac{2 \cdot H(A)}{L(C^2)} = \dfrac{2 \cdot 0,47}{1,6} \ \widehat{=}\ 58\%$

– "Kompakter Code" $\mathbf{C}^3$:

j	P_j	$\vec{c}_j$
1	0,81	0
2	0,09	10
3	0,09	110
4	0,01	111

– $L(C^3) = 1 \cdot 0,81 + 2 \cdot 0,09 + 3 \cdot 0,09 + 3 \cdot 0,01 = 1,29$

– Effizienz: $\eta(C^3) = \dfrac{2 \cdot H(A)}{L(C^3)} \ \widehat{=}\ 73\%$

- Weitere Effizienzsteigerung durch höhere Quellenerweiterung: $n = 3, 4, \ldots$

3.3 Huffman-Algorithmus

Der Huffman-Algorithmus (1952) dient zur Konstruktion kompakter Codes.

Gegeben: gedächtnislose Quelle $\mathbf{A}$: $\left(\mathbf{A} = \{a_1, a_2, \ldots a_q\}, \vec{P} = (P_1, \ldots, P_q)\right) \mathrel{\widehat{=}} A^{(0)}$

Gesucht: kompakte Quellencodierung, **hier binär**, $r = 2, \mathbf{C} = \{0, 1\}$

(Literatur für r-näre Quellen: Heise, Quattrocchi, S.133 ff.)

Idee:

I) rekursive Konstruktion von Ersatzquellen $A^{(0)}, A^{(1)}, \ldots A^{(q-2)}$ mit immer weniger Superzeichen; $A^{(q-2)}$ ist dann binäre Quelle.

II) $A^{(q-2)}$ besitzt eine triviale kompakte Codierung $\rightarrow$ auch für $A^{(q-3)}$ gibt es eine triviale kompakte Codierung, usw.

Algorithmus:

I) 1.) Starte mit $j = 0$ und $A^{(0)} = (\mathbf{A}, \vec{P})$

2.) Ordne die Quelle $A^{(j)}$ nach fallenden Zeichenwahrscheinlichkeiten: $P_1^{(j)} \leq P_2^{(j)} \leq \ldots \leq P_{q-j}^{(j)}$.

3.) Fasse die $r = 2$ am wenigsten wahrscheinlichen Zeichen $a_{q-j}^{(j)}$ und $a_{q-j-1}^{(j)}$ der Quelle $A^{(j)}$ zu einem Superzeichen $a_k^{(j+1)}$ mit der Wahrscheinlichkeit $P_{q-j-1}^{(j+1)}$ $= P_{(q-j)}^{j} + P_{q-j-1}^{(j)}$ in der Ersatzquelle $A^{(j+1)}$ zusammen.

4.) Wiederhole 2.) und 3.) $q-2$ mal, setze $j := j + 1$.

5.) Ordne die $r = 2$ Zeichen der Quelle $A^{(q-2)}$ nach fallenden Wahrscheinlichkeiten, $j = q - 2$. $\rightarrow$ Die Ersatzquelle $A^{(q-r)}$ enthält genau $r = 2$ Zeichen; eine kompakte Codierung für $A^{(q-r)}$ ist trivial.

II) Stufenweiser Aufbau einer kompakten, sofort decodierbaren, eindeutigen Codierung

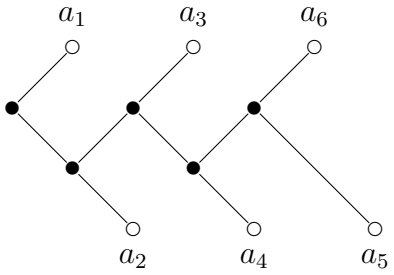
6.) Die Zeichen $a_1^{(q-2)}$ und $a_2^{(q-2)}$ werden (bei $r = 2$) mit "0" bzw. "1" codiert, $j := j - 1$. Jedes Zeichen von $A^{(j)}$ außer $a_{q-j}^{(j)}$ und $a_{q-j-1}^{(j)}$ erhält die gleiche Codierung wie bei der Ersatzquelle $A^{(j+1)}$. $a_{q-j}^{(j)}$ bzw. $a_{q-j-1}^{(j)}$ wird das CW, verlängert um "0" bzw. "1", zugeordnet, mit dem $a_{q-j-1}^{(j+1)}$ in $A^{(j+1)}$ codiert war.

7.) Wiederhole 6.) bis $j = 0$; $j := j - 1$.

Bsp.: $\mathbf{A} = \{a_1, a_2, \ldots, a_6\},\ \vec{P} = (0,4;\ 0,3;\ 0,1;\ 0,1;\ 0,06;\ 0,04),\ H(A) = 2,14$

	$P^{(0)}$	$\vec{c}^{(0)}$	$P^{(1)}$	$\vec{c}^{(1)}$	$P^{(2)}$	$\vec{c}^{(2)}$	$P^{(3)}$	$\vec{c}^{(3)}$	$P^{(4)}$	$\vec{c}^{(4)}$
$a_1^{(j)}$	$0,4$	1	$0,4$	1	$0,4$	1	$0,4$	1	$\underline{0,6}$	0
$a_2^{(j)}$	$0,3$	00	$0,3$	00	$0,3$	00	$\mathbf{0,3}$	00	$0,4$	1
$a_3^{(j)}$	$0,1$	011	$0,1$	011	$\underline{\mathbf{0,2}}$	010	$\underline{\mathbf{0,3}}$	01		
$a_4^{(j)}$	$0,1$	0100	$\mathbf{0,1}$	0100	$\mathbf{0,1}$	011				
$a_5^{(j)}$	$\mathbf{0,06}$	01010	$\underline{\mathbf{0,1}}$	0101						
$a_6^{(j)}$	$\mathbf{0,04}$	01011								

Codebaum:

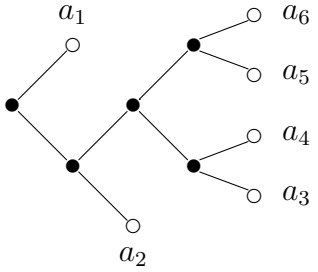


$L = 2,2;\ \eta = \dfrac{H(A)}{L} = 0,97;\ red = 3\%$

alternativ:

	$P^{(0)}$	$\vec{c}^{(0)}$	$P^{(1)}$	$\vec{c}^{(1)}$	$P^{(2)}$	$\vec{c}^{(2)}$	$P^{(3)}$	$\vec{c}^{(3)}$	$P^{(4)}$	$\vec{c}^{(4)}$
$a_1^{(j)}$	$0,4$	1	$0,4$	1	$0,4$	1	$0,4$	1	$\underline{0,6}$	0
$a_2^{(j)}$	$0,3$	00	$0,3$	00	$0,3$	00	$\mathbf{0,3}$	00	$0,4$	1
$a_3^{(j)}$	$0,1$	0100	$\underline{0,1}$	011	$\underline{\mathbf{0,2}}$	010	$\underline{\mathbf{0,3}}$	01		
$a_4^{(j)}$	$0,1$	0101	$\mathbf{0,1}$	0100	$\mathbf{0,1}$	011				
$a_5^{(j)}$	$\mathbf{0,06}$	0110	$\mathbf{0,1}$	0101						
$a_6^{(j)}$	$\mathbf{0,04}$	0111								

Codebaum:



Alternatives Verfahren

Problem bei Huffman–Algorithmus:

- die Auftrittswahrscheinlichkeiten der Symbole müssen bekannt sein
- keine Berücksichtigung von Symbolblöcken

Lempel–Ziv–Algorithmus (1977): adaptive Erstellung eines "Codewortbuches"

- bei englischen Texten Kompressionsfaktor bei Huffman: 43 %
  bei Lempel–Ziv: 55 %

Eingangs-Sequenz	-	+	- -	++	- -+	- - -	++ -	+ -	...
Codebuch-Index	0	1	2	3	4	5	6	7	8
Codebucherstellung	$\downarrow$	$\downarrow$	$\downarrow\downarrow$	$\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow$	$\rightarrow$
komprimierte Sequenz	0	1	00	11	21	20	30	10	...

Typisches Codebuch $\approx 2^{12}$ Einträge

4 Kanal

Der Kanal sei gedächtnislos, stationär, gestört, zeit- und wertediskret, in Abschnitt 5.4 wertekontinuierlich.

4.1 Stochastisches Kanalmodell

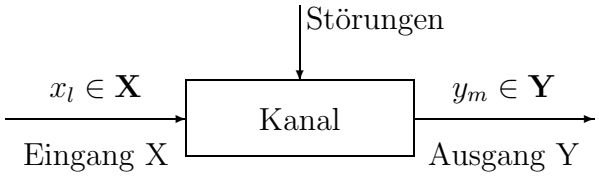


Bild 4.1: Modell des stochastischen Kanals

Eingang X: $x_l \in \mathbf{X} = \{x_1, x_2, \ldots x_j\}$; j-wertiges Alphabet mit Wahrscheinlichkeit $P(x_l) := P(X = x_l)$

Ausgang Y: $y_m \in \mathbf{Y} = \{y_1, y_2, \ldots y_k\}$; k-wertiges Alphabet mit Wahrscheinlichkeit $P(y_m) := P(Y = y_m)$

- Kanal generiert zum Zeitpunkt i aus jedem $x(i)$ ein $y(i)$.
- Kanalstörung: $x \longmapsto y$ ist stochastisch mit Übergangswahrscheinlichkeiten: $P(y_m|x_l) := P(Y = y_m|X = x_l), \quad l = 1, \ldots, j;\ m = 1, \ldots, k.$
- diskreter Kanal: $\mathbf{X}$ und $\mathbf{Y}$ sind endliche Mengen: $(j < \infty,\ k < \infty)$.
- stationärer Kanal: $P(y_m|x_l)$ sind unabhängig vom Zeitpunkt i.
- gedächtnisloser Kanal: $P(y_m|x_l)$ sind unabhängig von vorangegangenen Ein- und Ausgangssymbolen.

Hinweis: $\mathbf{X}$ und $\mathbf{Y}$ müssen nicht identisch sein, oft jedoch $k \geq j$

Bsp.:

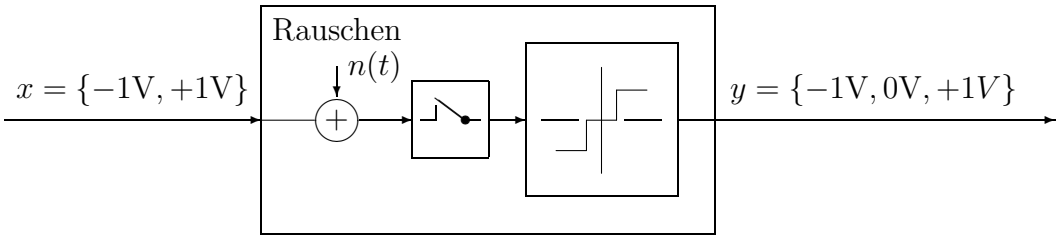


Bild 4.2: Modell eines zeit- und wertediskreten, gestörten Kanals

→ Kanal ist durch die Übergangswahrscheinlichkeiten $P(y_m|x_l),\ l = 1, \ldots, j;$ $m = 1, \ldots, k$ vollständig beschrieben:

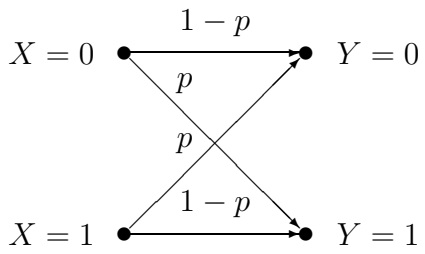
- stochastische Eingangswerte: $\vec{P}_x = \big(P(x_1), \ldots P(x_j)\big);\ \sum_{l=1}^{j} P(x_l) = 1$
- stochastische Ausgangswerte: $\vec{P}_y = \big(P(y_1), \ldots P(y_k)\big);\ \sum_{m=1}^{k} P(y_m) = 1$
- stochastische Kanalmatrix $[P_\mathrm{K}]$ mit $\sum_{m=1}^{k} P(y_m|x_j) = 1\ \forall\, j$

$$[P_\mathrm{K}] = \begin{bmatrix} P(y_1|x_1) & \cdots & P(y_k|x_1) \\ \vdots & & \vdots \\ P(y_1|x_j) & \cdots & P(y_k|x_j) \end{bmatrix}$$

Es gilt: $P(y_m) = \sum_{l=1}^{j} P(y_m|x_l) \cdot P(x_l)\ \Rightarrow \vec{P}_y = \vec{P}_x \cdot [P_\mathrm{K}]\ ;\ \vec{P}_x, \vec{P}_y$: Zeilenvektoren

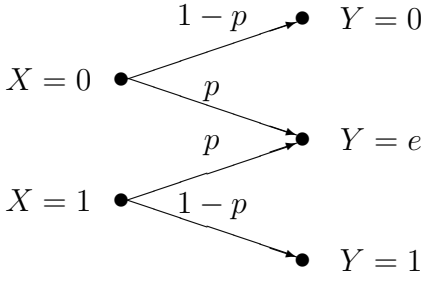
Bsp.: 1.) Binär Symmetrischer Kanal (BSC):

$$\left.\begin{array}{rcrcl} P(y_2|x_1) & = & P(y_1|x_2) & = & p \\ P(y_1|x_1) & = & P(y_2|x_2) & = & (1-p) \end{array}\right\} \Longrightarrow [P_\mathrm{K}] = \begin{bmatrix} (1-p) & p \\ p & (1-p) \end{bmatrix}$$



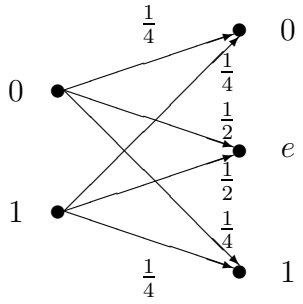
2.) Binär auslöschender Kanal: $\mathbf{Y} = \{0, e, 1\};\ e :=$ Fehlerzeichen (Erasure)

$$[P_\mathrm{K}] = \begin{bmatrix} (1-p) & p & 0 \\ 0 & p & (1-p) \end{bmatrix}$$



3.) Total gestörter Kanal:
Durch Beobachtung von Y keine Zusatzinformation gewinnbar.
Bsp.:

$$\left.\begin{array}{rcrcr} P(0|0) & = & P(0|1) & = & \frac{1}{4} \\ P(e|0) & = & P(e|1) & = & \frac{1}{2} \\ P(1|0) & = & P(1|1) & = & \frac{1}{4} \end{array}\right\} \quad [P_{\mathrm{K}}] = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$



Allgemein:

$$\vec{P}_y = \vec{P}_x \cdot [P_{\mathrm{K}}] \ = \vec{\text{konst}} \ \ \forall \ \ \vec{P}_x \Rightarrow \ \ [P_{\mathrm{K}}] = \begin{bmatrix} P(y_1) & \dots & P(y_k) \\ \vdots & & \vdots \\ P(y_1) & \dots & P(y_k) \end{bmatrix}$$

und $P(x, y) = P(x) \cdot P(y)$.

4.) Kaputter Kanal: Kurzschluss auf $y_m \in \mathbf{Y}$

$$[P_{\mathrm{K}}] = \begin{bmatrix} 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 \end{bmatrix}$$

5.) Ungestörter Kanal
Es gilt:

$$\mathbf{X} = \mathbf{Y}, \ \ [P_{\mathrm{K}}] = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

In $[P_{\mathrm{K}}]$ enthält genau jede Spalte eine Eins (Permutationsmatrix).

Beachte: $[P_{\mathrm{K}}]$ ist statistische Aussage, da nur für Wahrscheinlichkeiten $\vec{P}_x$ und $\vec{P}_y$.

4.2 Informationstransport über Kanäle

Gegeben: Quelle X: $(\mathbf{X}, \vec{P}_x)$, $H(X)$,
Kanal: Kanalmatrix $[P_K]$,
Kanalausgang Y: $(\mathbf{Y}, \vec{P}_y)$, $H(Y)$.

Gesucht:
- Wieviel Information transportiert der Kanal im Mittel (Transinformation)?
- Wieviel Information geht verloren (Äquivokation)?
- Wieviel Information fügt der Kanal hinzu (Irrelevanz)?

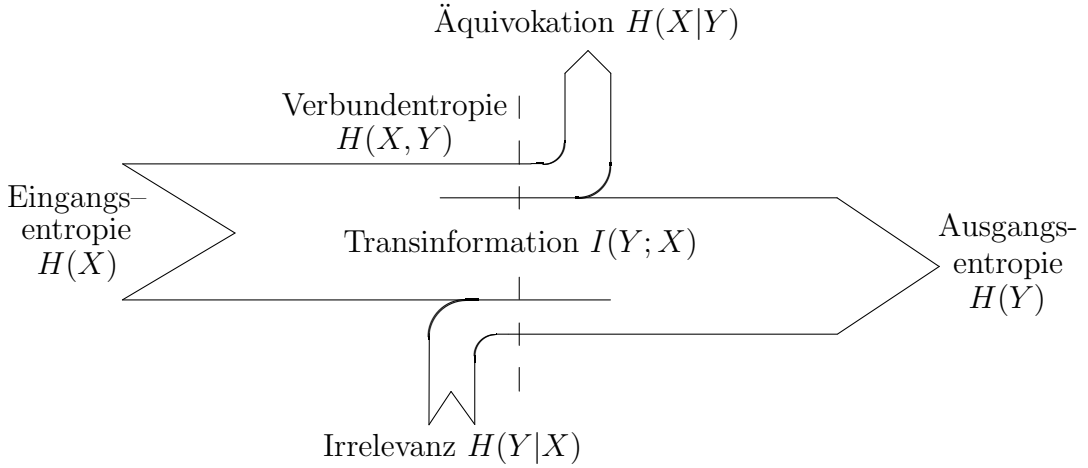


Bild 4.3: Zusammenhang zwischen Äquivokation, Irrelevanz und Transinformation

4.2.1 Irrelevanz

Def.: *Bedingter Informationsgehalt* $I(y|x)$: Gehalt der durch Beobachtung von y erhaltene Information, wenn bekannt ist, dass x gesendet wurde.
$\widehat{=}$ durch Kanal vorgetäuschte Information.

Bedingte Entropie (x wurde gesendet):

$$H(Y|x) := \mathrm{E}_y\left[I(y|x)\right] = \sum_{y\in\mathbf{Y}} \underbrace{P(y|x)}_{\text{aus } [P_\mathrm{K}]} \cdot \log \frac{1}{P(y|x)}$$

Mittlere bedingte Entropie (über alle Quellsymbole $x \in \mathbf{X}$ gemittelt):

$$H(Y|X) := \mathrm{E}_x\left[H(Y|x)\right] = \sum_{x\in\mathbf{X}} P(x) \cdot \sum_{y\in\mathbf{Y}} P(y|x) \cdot \log \frac{1}{P(y|x)}$$

$$H(Y|X) = \sum_{x\in\mathbf{X}} \sum_{y\in\mathbf{Y}} \underbrace{P(y|x)\cdot P(x)}_{P(x,y)} \cdot \log \frac{1}{P(y|x)}$$

$H(Y|X)$: Irrelevanz (Streuentropie)
$\widehat{=}$ durch Kanal hinzugefügte Entropie
$\rightarrow$ Kanalcodierung soll Irrelevanz entlarven.

4.2.2 Äquivokation

Def.: $I(x|y)$: Gehalt der durch Beobachtung von x erhaltenen Informationen, wenn die empfangene Nachricht y bekannt ist;
$\widehat{=}$ bei der Übertragung verlorene Information

$$H(X|Y) := \sum_{x\in\mathbf{X}}\sum_{y\in\mathbf{Y}} P(x|y)\cdot P(y)\cdot \log\frac{1}{P(x|y)}$$

Rückschlusswahrscheinlichkeit (Satz von Bayes):

$$P(x|y) = \frac{P(y|x)\cdot P(x)}{P(y)} = \frac{P(y|x)\cdot P(x)}{\sum\limits_{x'\in\mathbf{X}} P(y|x')P(x')}$$

$H(X|Y)$: Äquivokation (Rückschlussentropie)
$\widehat{=}$ durch Kanal verzehrte Entropie
$\rightarrow$ Kanalcodierung muß diese durch Redundanz absichern.

4.2.3 Verbund-Entropie

Def.: *Verbund-Informationsgehalt* $I(x,y)$: Durch gleichzeitige Beobachtung von x und y erhaltene Informationen.

$H(X,Y)$: Verbund-Entropie:

$$\begin{aligned} H(X,Y) &:= \sum_{x\in\mathbf{X}}\sum_{y\in\mathbf{Y}} P(x,y)\cdot\log\frac{1}{P(x,y)} = \sum_x\sum_y P(x,y)\cdot\log\frac{1}{P(y|x)\cdot P(x)} \\ &= \underbrace{\sum_x\sum_y P(x,y)\cdot\log\frac{1}{P(y|x)}}_{H(Y|X)} + \underbrace{\sum_x\left(\sum_y P(x,y)\right)\cdot\log\frac{1}{P(x)}}_{\sum_x P(x)\cdot\log\frac{1}{P(x)}=H(x)} \\ &= H(Y|X)+H(X) \end{aligned}$$

Es gilt der „Erhaltungssatz der Entropie“ (rechfertigt Bild 4.3):

$$H(X|Y)+H(Y) = H(Y,X) = H(X,Y) = H(Y|X)+H(X)$$

4.2.4 **Transinformation** (Mutual Information)

Def.: *Transinformationsgehalt* $I(x;y)$: Informationsgehalt der empfangenen Nachrichten, die vom gesendeten x herrührt.
(„Die Information x, die den Kanal überlebt.“)

Anschauliche Definition:

$$
\begin{aligned}
I(X;Y) &:= H(X) - H(X|Y) \\
&= \sum_{x\in\mathbf{X}} P(x)\cdot\log\frac{1}{P(x)} - \sum_{x\in\mathbf{X}}\sum_{y\in\mathbf{Y}} P(x,y)\cdot\log\frac{1}{P(x|y)} \\
&= \sum_{x\in\mathbf{X}} \underbrace{\left(\sum_{y\in\mathbf{Y}} P(x,y)\right)}_{P(x)} \cdot\log\frac{1}{P(x)} - \sum_{x\in\mathbf{X}}\sum_{y\in\mathbf{Y}} P(x,y)\log\frac{1}{P(x|y)} \\
&= \sum_{x\in\mathbf{X}}\sum_{y\in\mathbf{Y}} P(x,y)\cdot \underbrace{\log\frac{P(x|y)}{P(x)}}_{I(x;y):=\log\frac{P(x|y)}{P(x)}} \\
&= \sum_{x\in\mathbf{X}}\sum_{y\in\mathbf{Y}} P(x,y)\cdot\log\frac{P(x,y)}{P(x)\cdot P(y)} = \ldots \quad \text{da } P(x,y) = P(x|y)\cdot P(y) \\
&= H(Y) - H(Y|X) = I(Y;X)
\end{aligned}
$$

Transinformationsgehalt $I(x;y) := \log\dfrac{P(x|y)}{P(x)} = \log\dfrac{P(x,y)}{P(x)\cdot P(y)} = \log\dfrac{P(y|x)}{P(y)}$

$I(X;Y)$: Transinformation (mittlere wechselseitige Information)
$\widehat{=}$ Ausgabeentropie, die ausschließlich von der Kanaleingabe herrührt

Die Transinformation $I(X;Y) = f([P_\mathrm{K}], \vec{P}_x)$ hängt von der Wahrscheinlichkeitsverteilung der Quelle $\vec{P}_x$ und der Kanalmatrix $[P_\mathrm{K}]$ ab.

Abschätzung: Nach Abschnitt 3.2.1 gilt: $\sum_n P_n \log\frac{Q_n}{P_n} \le 0$
$\Longrightarrow \sum_n P_n\cdot\log\left(\frac{P_n}{Q_n}\right) \ge 0$, mit $\sum_n P_n = \sum_n Q_n = 1$;
daher gilt auch für die Doppelsumme: $\sum_x\sum_y P(x,y)\cdot\log\frac{P(x,y)}{P(x)\cdot P(y)} \ge 0$.

Gleichheit tritt genau dann auf, wenn $P_n = Q_n$, d.h. $P(x,y) = P(x)\cdot P(y)$,
also wenn der Kanal total gestört ist.

Daraus folgt für $I(X;Y)$:

$$\boxed{\begin{array}{rcl} I(X;Y) & = & H(X) - H(X|Y) \\ & = & H(Y) - H(Y|X) \\ & = & I(Y;X) \geq 0 \end{array}}$$

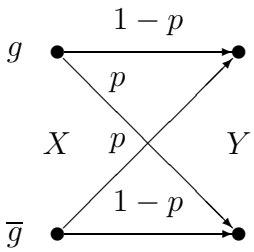
und $I(X;Y) = 0$, wenn Kanal total gestört.

Beachte: Bei q-närer Quelle und bei einem Kanal mit $j = k = q$ gilt:

$$0 \leq I_q(X;Y) = \underbrace{H_q(Y)}_{\leq \log_q q = 1} - \underbrace{H_q(Y|X)}_{\geq 0} \leq 1 \quad \left[\begin{array}{c} \text{bit} \\ \text{nat} \\ \text{Hartley} \end{array}\right]$$

4.2.5 Beispiel

BSC mit $p = \frac{3}{16},\ \overline{p} = (1-p) = \frac{13}{16}$
Quelle: $\{-,+\};\ P(-) = \frac{1}{10} = g;\ P(+) = \frac{9}{10} = \overline{g}$



- Entropie: $H(X) = g \cdot \text{ ld } \frac{1}{g} + \overline{g} \cdot \text{ ld } \frac{1}{\overline{g}} = \text{H}(g) = \frac{1}{10} \text{ ld } 10 + \frac{9}{10} \text{ ld } \frac{10}{9} \approx 0,47$ [bit]
- Irrelevanz: $H(Y|X) = \sum_x \sum_y P(y|x) \cdot P(x) \cdot \text{ld } \frac{1}{P(y|x)}$
  $= p \cdot g \cdot \text{ld } \frac{1}{p} + p \cdot \overline{g} \cdot \text{ld } \frac{1}{p} + \overline{p} \cdot g \cdot \text{ld } \frac{1}{\overline{p}} + \overline{p} \cdot \overline{g} \cdot \text{ld } \frac{1}{\overline{p}}$
  $= (g + \overline{g}) \cdot (p \cdot \text{ld } \frac{1}{p} + \overline{p} \cdot \text{ld } \frac{1}{\overline{p}})$
  $= \text{H(p)} = \ldots \approx 0,7$ [bit] (unabhängig von g!)
- Ausgang: $P(Y = -) = g \cdot \overline{p} + \overline{g} \cdot p = g + p - 2gp = 0,25$
  $P(Y = +) = 1 - g - p + 2gp = 0,75$
- Ausgangsentropie: $H(Y) = \text{H}(g + p - 2gp) \approx 0,81$ [bit]
- Transinfomation: $I(X;Y) = H(Y) - H(Y|X) \approx 0,11$ [bit]
- Äquivokation: $H(X|Y) = H(X) - I(X;Y) = H(X) - H(Y) + H(Y|X)$
  $\approx 0,36$ [bit]
- Verbundentropie: $H(X,Y) = H(X) + H(Y|X) = \text{H}(g) + \text{H}(p) \approx 1,17$ [bit]

4.2.6 "Hauptsatz der Nachrichtenübertragung"

Hintereinanderschaltung von Kanälen:
Die durch Äquivokation verlorene Transinformation kann im Mittel nicht regeneriert werden, d.h.:

$$I(X;Y) \geq I(X;Z)$$

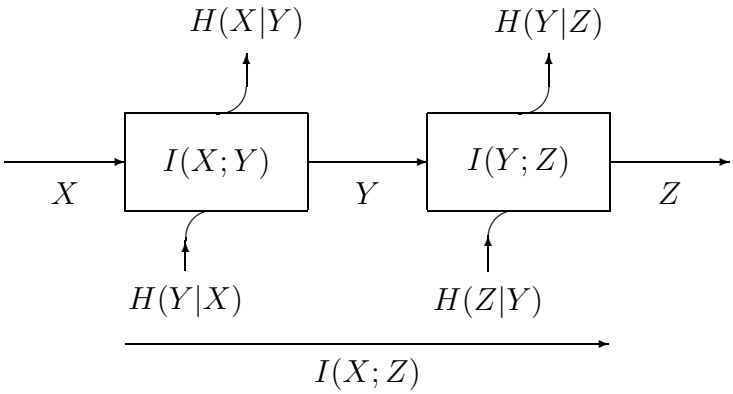


Bild 4.4: Transinformation bei hintereinandergeschalteten Kanälen

5 Kanalcodierungssatz

Mit welcher Rate kann über einen Kanal fehlerfrei Information übertragen werden?

5.1 Kanalkapazität

Def.: Die *Kanalkapazität* ist ein Maß für den maximal möglichen mittleren Informationsfluss.
$\widehat{=}$ maximale Transinformation durch einen Kanal mit der Kanalmatrix $[P_K]$

$$\boxed{C = \max_{\vec{P}_x} I(X;Y)} \text{ mit: } \vec{P}_x\{0 \le P(x) \le 1 \; \forall \; x \in \mathbf{X} \text{ und } \sum_{x \in \mathbf{X}} P(x) = 1\}$$

Maximumsuche über alle möglichen Zeichen-Wahrscheinlichkeitsverteilungen.
Quellcode beeinflusst $\vec{P}(x)$!
$\rightarrow$ Optimierungsproblem mit Randbedingungen (i.d.R. aufwändige numerische Suche)

5.1.1 Spezialfälle

1.) $C = 0 \;\rightarrow\; I(X;Y) = 0 \, \forall \, \vec{P}_x \;\Rightarrow\; P(x,y) = P(x) \cdot P(y) \;\; \forall \, x \in \mathbf{X}, y \in \mathbf{Y}$
$\rightarrow$ Die Kanalkapazität ist genau dann Null, wenn der Kanal total gestört ist.

2.) $C = 1 \;\rightarrow\; H_q(Y|X) = 0 \;\Rightarrow\; \sum_X \sum_Y P(y|x) \cdot P(x) \cdot \log \dfrac{1}{P(y|x)} = 0$
$\Rightarrow \; P(y|x) = 0 \text{ oder } 1 \;\; \forall \, x; y$

d.h. $[P_K] = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & & \vdots \end{bmatrix}$ Jede Spalte enthält genau eine Eins.
$\rightarrow$ ungestörter Kanal

$\rightarrow$ Die Kanalkapazität ist genau dann maximal, wenn der Kanal ungestört ist.

3.) BSC, gedächtnislos, $\vec{P}_x = \big(g, (1-g)\big) \;\rightarrow$ siehe Beispiel 4.2.5

$$C = \max_{\vec{P}_x} I(X;Y) = \max_{0 \le g \le 1} H(Y) - H(Y|X) = \max_{0 \le g \le 1} \underbrace{\mathrm{H}(p+g-2pg)}_{\le \mathrm{H}(\frac{1}{2}) = 1;\; g_{\mathrm{opt}} = \frac{1}{2}} - \mathrm{H}(p)$$

$$C = 1 - \mathrm{H}(p)$$

Die Transinformation erreicht die Kanalkapazität nur, wenn
$P(\text{“+”}) = P(\text{“-”}) = 0,5$.

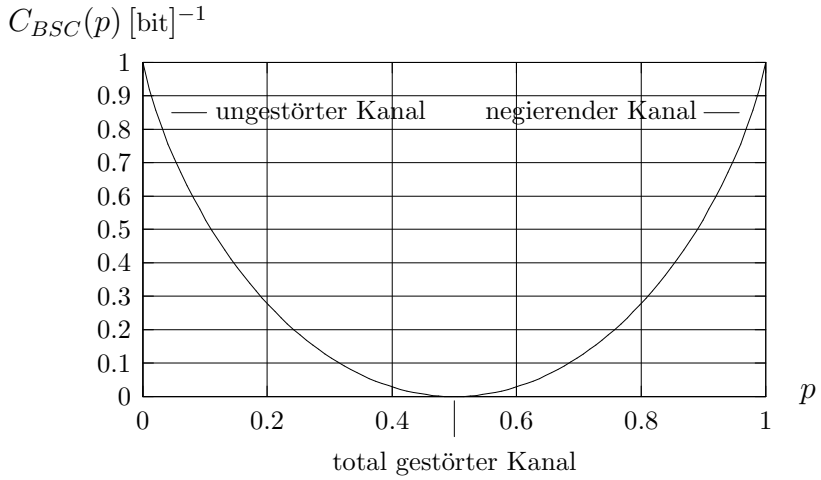


Bild 5.1: Kanalkapazität des gedächtnislosen BSC

5.1.2 Kanalerweiterung

Blockweise Nachrichtenübertragung

- Gedächtnislose Quelle, j-när $\Rightarrow$ $x(i)$ statistisch unabhängig; $H(X)$
  n-te Erweiterung: $\vec{x} = \big(x(1), \ldots x(n)\big) \in \vec{\mathbf{X}} = \mathbf{X}^n, \quad H(X^n) = n \cdot H(X)$

- Kanal: P_K, Ausgang Y: k-när
  n-te Erweiterung: $\vec{y} = \big(y(1), \ldots y(n)\big) \in \vec{\mathbf{Y}} = \mathbf{Y}^n$

  Übergangswahrscheinlichkeiten:

$$\begin{array}{rcl} P(\vec{y}|\vec{x}) & = & P\big(y(1), \ldots y(n)|\vec{x}\big) \\ \text{da } x(i) \text{ stat. unabhängig} \Rightarrow & = & P\big(y(1)|x(1)\ldots x(n)\big) \cdot P\big(y(2)|\vec{x}\big) \cdots P\big(y(n)|\vec{x}\big)\big) \\ \text{und gedächtnisloser Kanal} \Rightarrow & = & P\big(y(1)|x(1)\big) \cdot P\big(y(2)|x(2)\big) \cdots P\big(y(n)|x(n)\big) \end{array}$$

  $\Rightarrow$ Kanalmatrix der n-ten Erweiterung:
  $\left[P_K^{(n)}\right] = (j^n \times k^n)$-Matrix mit $P(\vec{y}|\vec{x}) = \prod_{i=1}^{n} P\big(y(i)|x(i)\big)$

- Kanalkapazität der n-ten Kanalerweiterung $C^{(n)} = n \cdot C$

Bsp.: BSC mit p; Blocklänge n
Wenn sich $\vec{x}$ und $\vec{y}$ an $0 \le d_H \le n$ Stellen unterscheiden, gilt:

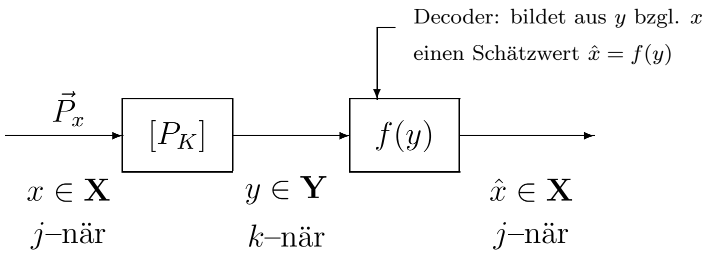
$$P(\vec{y}|\vec{x}) = \underbrace{p^{d_H}}_{\text{falsch}} \cdot \underbrace{(1-p)^{n-d_H}}_{\text{richtig}} \quad \forall\, \vec{x} \in \vec{\mathbf{X}},\ \vec{y} \in \vec{\mathbf{Y}} = \vec{\mathbf{X}}$$

z.B.:

$$\left.\begin{array}{rcl} \vec{x} & = & 00101 \\ \vec{y} & = & 00011 \\ \hline & & \cdot\cdot\,\text{xx}\cdot \end{array}\right\} \Rightarrow d_H = 2$$

5.2 Decoder-Strategie bei gestörten Kanälen

Gegeben: gestörtes Übertragungssystem (ggf. mit Kanalcodierer), $[P_{\mathrm{K}}]$ sei konst.



Entscheidungsregel $\widehat{x} = f(y): \; y \in \mathbf{Y} \xrightarrow{\text{f}} \widehat{x} \in \mathbf{X}$
$\rightarrow$ Es gibt j^k verschiedene Entscheidungsregeln $f(\cdot)$.

5.2.1 Entscheidungsregel für minimale Fehlerwahrscheinlichkeit

$$P_E = 1 - P_{\text{richtig}} = 1 - \mathrm{E}_x\big[P\big(f(y) = x|y\big)\big]$$

$$\min(P_E) \longrightarrow \max(P_{\text{richtig}}) \longrightarrow \max_{f(\cdot)} \Big(P\big(f(y) = x|y\big)\Big) \;\forall\; y \in \mathbf{Y}$$

$\Rightarrow$ Wähle den Schätzwert $\widehat{x} = f(y)$ so, dass $P\,\underbrace{(\widehat{x} = x|y)}_{\substack{\text{richtige}\\ \text{Entscheidung}}} \geq P(\tilde{x} = x|y)\;\forall\,\tilde{x} \in \mathbf{X}$

$$P(\widehat{x} = x|y) \cdot P(y) \overset{\text{Bayes}}{=\!=} P(y|\widehat{x}) \cdot P(\widehat{x}) \xRightarrow{P(y)\neq 0}$$

$$\boxed{P(y|\widehat{x}) \cdot P(\widehat{x}) \geq P(y|\tilde{x}) \cdot P(\tilde{x})\;\forall\,\tilde{x} \in \mathbf{X}}$$

$\rightarrow$ Maximum-A- Posteriori-(MAP)-Entscheidungsregel: (Rückschlusswahrscheinlichkeit), d.h. wähle $\widehat{x} \in \mathbf{X}$ so, dass am Kanalausgang nach Senden von $\widehat{x}$ mit größter Wahrscheinlichkeit das tatsächlich empfangene y anstehen würde, gewichtet mit der Wahrscheinlichkeit $P(\widehat{x})$.

$\rightarrow$ Maximum-Likelihood-(ML)- Entscheidungsregel:
Annahme: $P(x) = \text{konst}\;\forall\,x \in \mathbf{X}$ (oder wenn $P(x)$ unbekannt)

$\Longrightarrow\; P(y|\widehat{x}) \geq P(y|\tilde{x}) \;\;\forall\,\tilde{x} \in \mathbf{X}$, wobei $P(y|x)$ gegeben durch $[P_{\mathrm{K}}]$

Bsp.: $j = 2,\; k = 3;\; \mathbf{X} = \{-,+\},\; \mathbf{Y} = \{-1, 0, +1\}$

$$[P_K] = \begin{bmatrix} \underline{0,5} & \underline{0,4} & 0,1 \\ 0,2 & 0,3 & \underline{0,5} \end{bmatrix} \xRightarrow{\text{ML–Entscheidungsregel}} \begin{array}{rcl} f(-1) &=& - \\ f(0) &=& - \\ f(+1) &=& + \end{array}$$

5.2.2 Blockweise Übertragung

(n-te Quell- und Kanalerweiterung)
Abstandsmaß (Metrik) zwischen zwei Zeichen:

$$d_H(x,y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$$

Abstand von zwei binären Folgen (Blöcken, Codeworten) der Länge n
= Hamming-Abstand zwischen $\vec{x}$ und $\vec{y}$: $d_H(\vec{x},\vec{y}) := \sum_{i=1}^{n} d_H\big(x(i),y(i)\big),\ 0 \le d_H \le n$
Bsp.:

$$\left.\begin{array}{rcl} \vec{x} & = & 00101 \\ \vec{y} & = & 00011 \\ \hline & & \cdot\cdot\,\mathrm{xx}\cdot \end{array}\right\} \Rightarrow d_H(\vec{x},\vec{y}) = 2$$

Anwendung bei BSC: → siehe Kanalmatrix $\left[P_K^{(n)}\right]$ bei n-ter Kanalerweiterung, siehe Abschnitt 5.1.2

$$P(\vec{y}\,|\vec{x}\,) = p^{d_H(\vec{x},\vec{y})} \cdot (1-p)^{n-d_H(\vec{x},\vec{y})} \ \ \forall\ \vec{x} \in \vec{\mathbf{X}},\ \vec{y} \in \vec{\mathbf{Y}}$$

ML-Entscheidungsregel bei BSC und Blocklänge n:
Mit $d_H = d_H(\vec{y},\widehat{\vec{x}}\,)$ gilt:

$$P(\vec{y}\,|\widehat{\vec{x}}\,) = p^{d_H} \cdot (1-p)^{n-d_H} = (p/1-p)^{d_H} \cdot (1-p)^n$$

Für $p < \frac{1}{2}$ gilt:

$$(1-p)^n = (1-p)(1-p)^{n-1} > p^1(1-p)^{n-1} > \cdots > p^{d_H}(1-p)^{n-d_H} > \cdots > p^n$$

→ Wähle $\widehat{\vec{x}}$ so, dass d_H minimal, d.h.: $\boxed{\widehat{\vec{x}} = \arg\min_{\vec{x}\in\vec{\mathbf{X}}} d_H(\vec{x},\vec{y})}$

⇒ Der im Sinne der ML-Strategie optimale Entscheider wählt ein $\widehat{\vec{x}}$, das sich in den wenigsten Stellen von $\vec{y}$ unterscheidet.

Veranschaulichung: n-dimensionaler Vektorraum mit Hamming-Metrik:

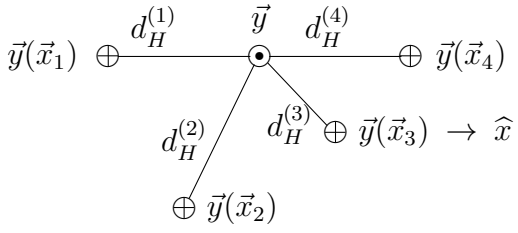


Bild 5.2: Zur Veranschaulichung der ML–Entscheidungsregel

Bsp.: Repetitions-Blockcode bei BSC mit Fehlerwahrscheinlichkeit p
Jede binäre Nachricht wird durch einen Block gleicher Zeichen der Länge n (ungerade) übertragen.
z.B.: $a = 0 \;\rightarrow\; \vec{x}_0 = (\underbrace{0 \ldots 0}_{n});\;\; a = 1 \;\rightarrow\; \vec{x}_1 = (\underbrace{1 \ldots 1}_{n})$

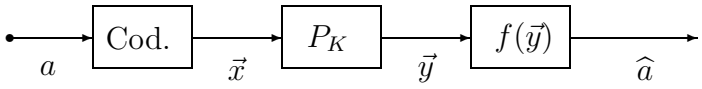


Bild 5.3: System mit Kanalcodierer und Decodierer

Rate: $R = \dfrac{k}{n} = \dfrac{1 \leftarrow \text{Info-Länge}}{n \leftarrow \text{Block-Länge}}$

ML − Entscheidungsregel : $\widehat{a} = \begin{cases} 1: & \text{wenn } d_H(\vec{y}, \vec{x}_0) > d_H(\vec{y}, \vec{x}_1) \\ 0: & \text{sonst} \end{cases}$

Fehlerwahrscheinlichkeiten:

Bsp.: $n = 3 \rightarrow P_E(p, n) = P(\text{2 oder mehr Übertragungsfehler}) = p^3 + \binom{3}{2} \cdot (1-p) \cdot p^2$
allgemein:

$$P_E(p, n) = \sum_{d = \frac{n+1}{2}}^{n} \binom{n}{d} \cdot p^d (1-p)^{n-d}$$

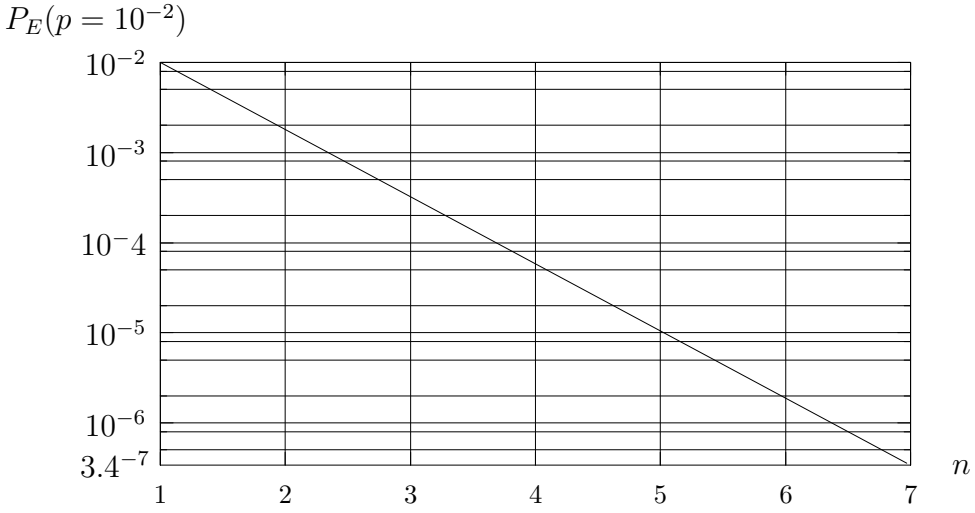


Bild 5.4: Fehlerwahrscheinlickeit P_E in Abhängigkeit von der Blocklänge n

Bsp.: $p = 0.01 \;\rightarrow\; \mathrm{H}(p) = 0,08\,\text{Bit} \;\rightarrow\; C = 0,92\,\text{Bit}$
(s. auch Bild 5.4 und Bild 5.5)

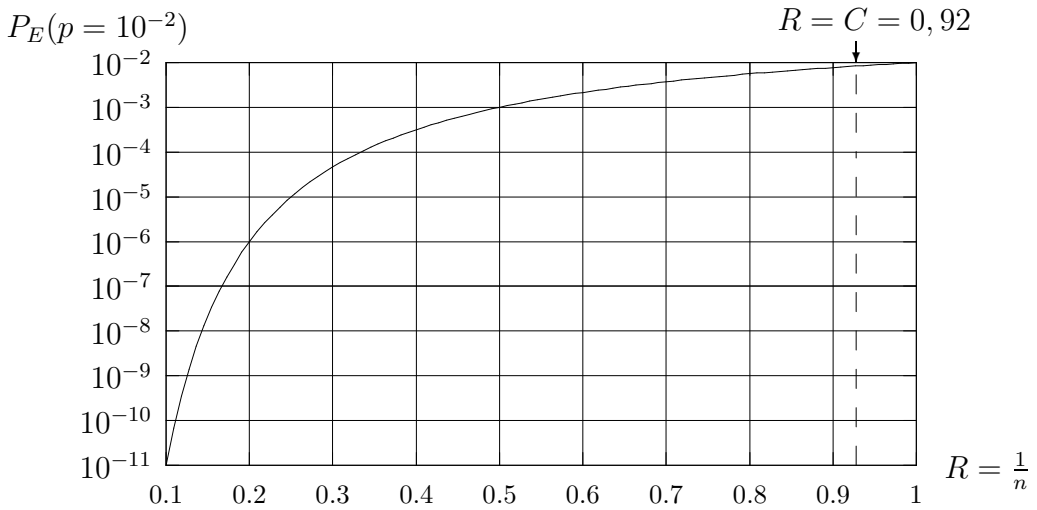


Bild 5.5: Fehlerwahrscheinlichkeit P_E in Abhängigkeit von der Rate R

→ Bei Repetitionscodes mit der Rate $R = C$ ist nur eine sehr hohe Restfehlerwahrscheinlichkeit realisierbar. ⇒ "schlechter" Code

5.3 Kanalcodierungssatz von Shannon

Frage: Gibt es eine maximale Code-Rate, ab der keine Codierung mehr existiert, die eine fehlerfreie Übertragung erlaubt?

Antwort: Shannons Kanalcodierungssatz für BSC

Gegeben sei ein BSC mit Fehlerwahrscheinlichkeit p und Kanalkapazität $C = 1-\mathrm{H}(p)$ sowie ein Blockcode der Länge n. Sei ε eine beliebig kleine positive Zahl und $M := 2^{n\cdot(C-\varepsilon)}$. Dann gibt es bei einer hinreichend großen Blocklänge n unter den 2^n möglichen Blöcken eine Teilmenge von M Blöcken derart, dass ihre Decodierfehlerwahrscheinlicheit eine von ε abhängige Schranke unterschreitet.

Folgerung: → Die M Blöcke können $\lfloor n \cdot C - \varepsilon \rfloor$Bit (= ganzzahliger Anteil) übertragen.

→ Rate $R = \dfrac{k}{n} = \dfrac{\lfloor n \cdot C - \varepsilon \rfloor}{n} \approx C - \varepsilon \quad \Rightarrow \quad R < C/[bit]$

Zur fehlerfreien Übertragung darf die Kanalkapazität nie kleiner sein als die Coderate.

Verallgemeinerung: (mit Quellencodierungssatz) Sei X eine gedächtnislose q-näre Quelle mit der Entropie $H(X)$ und einer Quellsymbolrate von $\frac{1}{T_x}$. Der gedächtnislose diskrete q-näre Kanal mit der Kanalkapazität C habe eine Sendesymbolrate von $\frac{1}{T_c}$. Wenn gilt:

$$\frac{H(X)}{T_x} < \frac{C}{T_c},$$

dann existieren für hinreichend große Blocklängen Codes, die eine beliebig kleine Restfehlerwahrscheinlichkeit ermöglichen.

Umkehrung: Wenn $C < \frac{T_c}{T_x} \cdot H(X)$, dann gibt es keine Codierung, die eine beliebig kleine Restfehlerwahrscheinlichkeit ermöglicht.

Bsp.: BSC mit Fehlerwahrscheinlichkeit p

- $p = 0,01 \;\rightarrow\; \mathrm{H}(p) = 0,08\,\mathrm{bit} \;\rightarrow\; C = 0,92\,\mathrm{bit}$
  $\rightarrow$ Es gibt eine Codierung mit $R < 0,92$ für restfehlerfreie Übertragung.
- $p = \dfrac{3}{16} \;\rightarrow\; \mathrm{H}(p) = 0,7\,\mathrm{bit} \;\rightarrow\; C = 0,3\,\mathrm{bit} \;\rightarrow\; R < 0,3\,\mathrm{bit}$
  $T_c = 10^{-3}\frac{s}{\text{Zeichen}} \Rightarrow$ Der Kanal kann bei geeigneter Kanal- und Quellencodierung einen Quell-Entropiestrom von 300 $\frac{\text{bit}}{\text{s}}$ mit beliebig kleiner Restfehlerwahrscheinlichkeit übertragen.

5.4 Kanalkapazität wertekontinuierlicher Kanäle

5.4.1 Wertekontinuierliche Nachrichten

$x \in \mathbf{R}$: kontinuierlicher Wertevorrat, evt. Intervall reller Zahlen
$p_{\mathbf{x}}(x)$: Wahrscheinlichkeitsdichtefunktion (= differenzielle Wahrscheinlichkeit)

$$P(x_0 - \triangle x \le \mathbf{x} \le x_0) = \int_{x_0-\triangle x}^{x_0} p_{\mathbf{x}}(\xi)\, d\xi \approx p_{\mathbf{x}}(x_0) \cdot \triangle x \quad , \triangle x \to 0$$

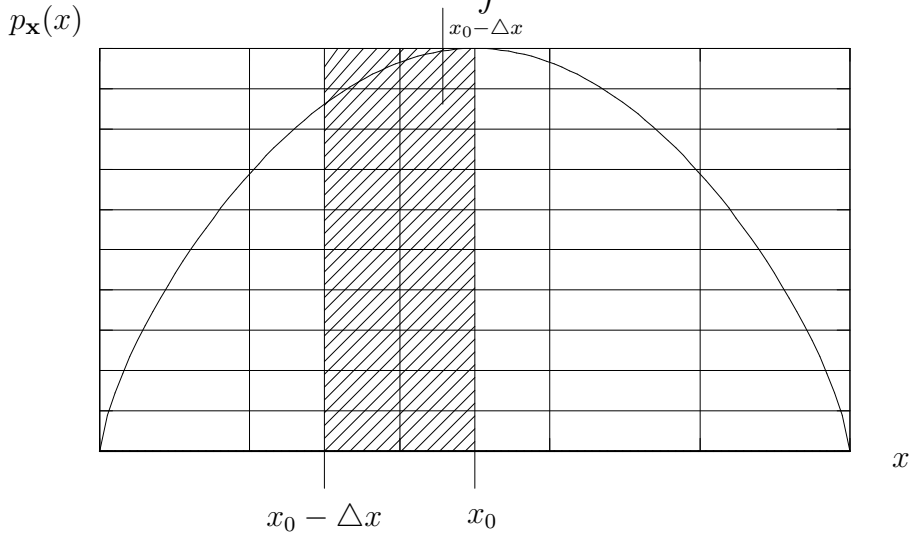


Bild 5.6: Zur Approximation von $P(x_0 - \triangle x \le \mathbf{x} \le x_0)$

Quelle: kontinuierlicher stochastischer Prozess mit
- Mittelwert $\overline{x} = \mathrm{E}[\mathbf{x}]$ und
- Varianz $\sigma_{\mathbf{x}}^2 = \mathrm{E}[(\mathbf{x} - \overline{x})^2]$ = mittlere Wechselleistung

Entropie: $H(X) \to \infty$, da mit $x \in \mathbf{R}$ beliebig viel Information darstellbar.

Übergang: diskret $\to$ kontinuierlich:
Symbole $x_k := k \cdot \triangle x,\ k = 0, \pm 1, \pm 2, \ldots;\ \triangle x \to 0$
$\Rightarrow P(x_k) = p_{\mathbf{x}}(x_k) \cdot \triangle x$

$$
\begin{aligned}
H(X)/\,[\text{bit}] &= \sum_{k=-\infty}^{\infty} p_{\mathbf{x}}(x_k) \cdot \triangle x \cdot \mathrm{ld}\,\frac{1}{p_{\mathbf{x}}(x_k) \cdot \triangle x} \\
&= \sum_{k} p_{\mathbf{x}}(x_k) \cdot \mathrm{ld}\,\frac{1}{p_{\mathbf{x}}(x_k)} \cdot \triangle x + \underbrace{\sum_{k} P(x_k)}_{=1} \cdot \mathrm{ld}\,\frac{1}{\triangle x} \\
&\overset{\triangle x \to 0}{=\!=\!=} \underbrace{\int_{-\infty}^{\infty} p_{\mathbf{x}}(\xi) \cdot \mathrm{ld}\,\frac{1}{p_{\mathbf{x}}(\xi)}\, d\xi}_{\text{differenzielle Entropie}} + \lim_{\triangle x \to 0} \underbrace{\mathrm{ld}\,\frac{1}{\triangle x}}_{\substack{\text{Konst. f. alle} \\ \text{Entropie–} \\ \text{berechnungen}}}
\end{aligned}
$$

Approximation:

$$
H(X) = h(X) + \underbrace{K(\triangle x)}_{\to\infty \text{ für } \triangle x \to 0}
$$

Formale Definition:

- differenzieller Informationsgehalt: $i(\mathbf{x}) := \mathrm{ld}\frac{1}{p_{\mathbf{x}}(x)} \cdot [\text{bit}]$
- differenzielle Entropie:

$$
h(X)\,/[\text{bit}] = \mathrm{E}\big[i(\mathbf{x})\,/[\text{bit}]\big] = \int_{-\infty}^{\infty} p_{\mathbf{x}}(\xi)\, \mathrm{ld}\,\frac{1}{p_{\mathbf{x}}(\xi)}\, d\xi
$$

Frage: Für welchen $p_{\mathbf{x}}(x)$ wird $h(X)$ maximal?

$$
\Rightarrow \text{ Nebenbedingungen } \quad 1.)\ \int_{-\infty}^{\infty} p_{\mathbf{x}}(\xi)\, d\xi = 1
$$

$$
2.)\ \int_{-\infty}^{\infty} (\xi - \overline{x})^2 p_{\mathbf{x}}(\xi)\, d\xi = \sigma_{\mathbf{x}}^2 < \infty
$$

Ergebnis: (über Variationsrechnung)

$$p_{\mathbf{x}}(x) = \frac{1}{\sqrt{2\pi}\sigma_{\mathbf{x}}}\,\mathrm{e}^{-\dfrac{(x-\overline{x})^2}{2\cdot\sigma_{\mathbf{x}}^2}} \qquad \rightarrow \text{Gaußverteilte Zufallsvariable, Gaußprozess}$$

$$\begin{aligned}\Longrightarrow \max_{p_{\mathbf{x}}(x)}\{h(X)\,/[\text{bit}]\} &= \int_{-\infty}^{\infty} p_{\mathbf{x}}(\xi)\,\mathrm{ld}\,(\sqrt{2\pi}\sigma_{\mathbf{x}})\,d\xi + \int_{-\infty}^{\infty} p_{\mathbf{x}}(\xi)\frac{(\xi-\overline{x})^2}{2\cdot\sigma_{\mathbf{x}}^2}\,\mathrm{ld}\,\mathrm{e}\,d\xi \\ &= \frac{1}{2}\,\mathrm{ld}\,(2\pi\sigma_{\mathbf{x}}^2) + \frac{\sigma_{\mathbf{x}}^2}{2\sigma_{\mathbf{x}}^2}\mathrm{ld}\,\mathrm{e} \;=\; \frac{1}{2}\,\mathrm{ld}\,(2\pi\mathrm{e}\sigma_{\mathbf{x}}^2)\;\;(\text{ist unabhängig von } \overline{x}\;!)\end{aligned}$$

Bedingte differenzielle Entropie:

$$\begin{aligned}\text{differenzielle Irrelevanz:}\quad h(Y|X)\,/[\text{bit}] &:= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{\mathbf{x},\mathbf{y}}(\xi,\eta)\cdot\mathrm{ld}\,\frac{1}{p_{\mathbf{y}|\mathbf{x}}(\eta|\xi)}\,d\eta\,d\xi \\ \text{analog zu diff. Entropie:}\qquad H(Y|X) &:= h(Y|X) + K(\triangle x)\end{aligned}$$

Transinformation bei wertkontinuierlichen Kanälen:

$$\begin{aligned}I(X;Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= h(Y) + K(\triangle x) - \big(h(Y|X) + K(\triangle x)\big) \\ &= h(Y) - h(Y|X) = h(X) - h(X|Y)\end{aligned}$$

5.4.2 Wertekontinuierlicher, bandbegrenzter Gaußkanal

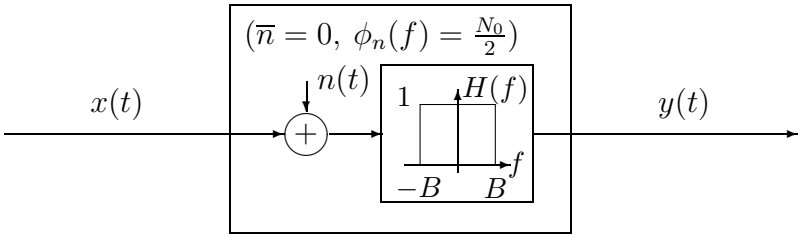


Bild 5.7: Modell des wertekontinuierlichen, bandbegrenzten Gaußkanals

Quelle X : Zufallsprozess mit $p_{\mathbf{x}}(x)$, leistungsbegrenzt

$$\lim_{T\to\infty}\frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} x^2(t)\,dt \;= \mathrm{E}[\mathbf{x}^2] = \;\; P \;<\; \infty$$

- Da der Kanal bandbegrenzt ist, muss auch $x(t)$ bandbegrenzt sein, sonst Informationsverlust.
- Da $x(t)$ bandbegrenzt, ist gesamte Information in Abtastwerten enthalten: $x(i) := x(i \cdot \tau),\ \tau \leq \frac{1}{2 \cdot B}$; Signalleistung: P .
- Auch Kanalausgang ist ohne Informationsverlust abtastbar: $y(i) = y(i \cdot \tau)$.
- Rauschabtastwerte $n(i)$ sind gaußverteilt, unkorreliert mit $\sigma_{\mathbf{n}}^2 = \frac{N_0}{2} \cdot 2 \cdot B = N_0 \cdot B$.
- Kanalgleichung $y(i) = x(i) + n(i)$:
  gedächtnisloser, zeitdiskreter, wertekontinuierlicher Gaußkanal (s. Bild 5.8) = Additive White Gaussian Noise (AWGN) Kanal

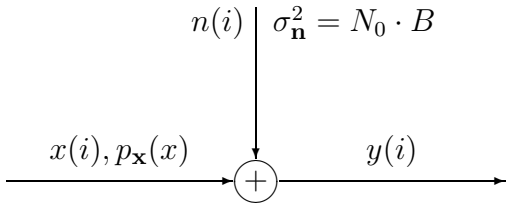


Bild 5.8: Gedächtnisloser, zeitdiskreter, wertekontinuierlicher Gaußkanal

$$p_{\mathrm{y|x}}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_{\mathbf{n}}} \, \mathrm{e}^{-\frac{(y-x)^2}{2\sigma_n^2}} = \frac{1}{\sqrt{2\pi}\sigma_{\mathbf{n}}} \, \mathrm{e}^{-\frac{n^2}{2\sigma_n^2}}$$

Differenzielle Irrelevanz:

$$\begin{aligned} h(Y|X) \,/\, [\text{bit}] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{\mathbf{x},\mathbf{y}}(\xi,\eta) \cdot \left(\mathrm{ld}\,(\sqrt{2\pi} \cdot \sigma_{\mathbf{n}}) + \frac{(\eta-\xi)^2}{2\sigma_{\mathbf{n}}^2} \,\mathrm{ld}\,\mathrm{e} \right) d\xi \, d\eta \\ &= \frac{1}{2} \,\mathrm{ld}\,(2\pi\sigma_{\mathbf{n}}^2) \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{\mathbf{x},\mathbf{y}}(\xi,\eta) \, d\xi \, d\eta}_{1} + \frac{\mathrm{ld}\,\mathrm{e}}{2\sigma_{\mathbf{n}}^2} \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\eta-\xi)^2 p_{\mathbf{x},\mathbf{y}}(\xi,\eta) \, d\xi \, d\eta}_{\sigma_{\mathbf{n}}^2} \\ &= \frac{1}{2} \,\mathrm{ld}\,(2\pi\mathrm{e}\sigma_{\mathbf{n}}^2) \end{aligned}$$

Kanalkapazität pro Kanalzugriff (= pro Abtastwert):

$$C_s = \max_{p_{\mathbf{x}}(x)} \left\{ I(X;Y); \mathrm{E}[\mathbf{x}^2] = P \right\} = \max_{p_{\mathbf{y}}(y)} \left\{ h(y); P \right\} - h(Y|X)$$

- $h(Y)$ ist maximal, wenn y gaußverteilt (s.o.) $\rightarrow h(Y) = \frac{1}{2} \,\mathrm{ld}\,(2\pi\mathrm{e}\sigma_{\mathbf{y}}^2)$ [bit]
- $\sum$ gaußverteilter Z.V. ist gaußverteilt $\Rightarrow \mathbf{x}$ muss gaußverteilt sein.
  Es gilt: Summenleistung = Summe der Einzelleistungen $\Rightarrow \sigma_{\mathbf{y}}^2 = \sigma_{\mathbf{n}}^2 + P$

$$
\begin{aligned}
\rightarrow h(Y)\,/\,[\text{bit}] &= \frac{1}{2}\,\text{ld}\left(2\pi\text{e}\,(P+\sigma_{\mathbf{n}}^2)\right)\\
C_s\,/\,[\text{bit}] &= \frac{1}{2}\,\text{ld}\left(2\pi\text{e}\,(P+\sigma_{\mathbf{n}}^2)\right) - \frac{1}{2}\,\text{ld}\left(2\pi\text{e}\,\sigma_{\mathbf{n}}^2\right)\\
&= \frac{1}{2}\,\text{ld}\left(1+\underbrace{\frac{P}{\sigma_{\mathbf{n}}^2}}_{\substack{\text{Signal-Stör-}\\ \text{Verhältnis (SNR)}}}\right) = \frac{1}{2}\,\text{ld}\left(1+\underbrace{\frac{P}{N_0\cdot B}}_{\text{SNR}}\right)
\end{aligned}
$$

5.4.3 Kanalkapazität pro Zeiteinheit T

Kanalkapazität C in [bit/s];
Symboldauer: $\tau \Rightarrow$ Nyquistrate $\frac{1}{\tau}$, Kanalbandbreite: $B = \frac{1}{2\cdot\tau}$ in [Hz]
Anzahl der Symbole je Zeiteinheit T: $n = \dfrac{T}{\tau} = 2\cdot B\cdot T$

Kanalkapazität:

$$
C := \frac{n\cdot C_s}{T} = \frac{2\cdot B\cdot T}{2\cdot T}\,\text{ld}\,\left(1+\frac{P}{N_0\cdot B}\right)\cdot[\text{bit}] \;=\; \frac{B}{[\text{Hz}]}\cdot\,\text{ld}\,\left(1+\frac{P}{N_0\cdot B}\right)\cdot\left[\tfrac{\text{bit}}{\text{s}}\right]
$$

Spektrale Effizienz:

$$
\begin{aligned}
\frac{C}{B} &= \text{ld}\,\left(1+\frac{P}{\sigma_{\mathbf{n}}^2}\right)\,\left[\tfrac{\text{bit}}{\text{s}\cdot\text{Hz}}\right]\\
\frac{C}{B}\left[\tfrac{\text{bit}}{\text{s}\cdot\text{Hz}}\right]^{-1} &= \text{ld}\,\left(1+10^{\frac{\text{SNR}}{10\,\text{dB}}}\right) \approx \frac{\text{SNR}}{10\,\text{dB}}\,\text{ld}\,10 = \frac{\text{SNR}}{3\,\text{dB}} \qquad \text{für SNR} \gg 10\text{dB}
\end{aligned}
$$

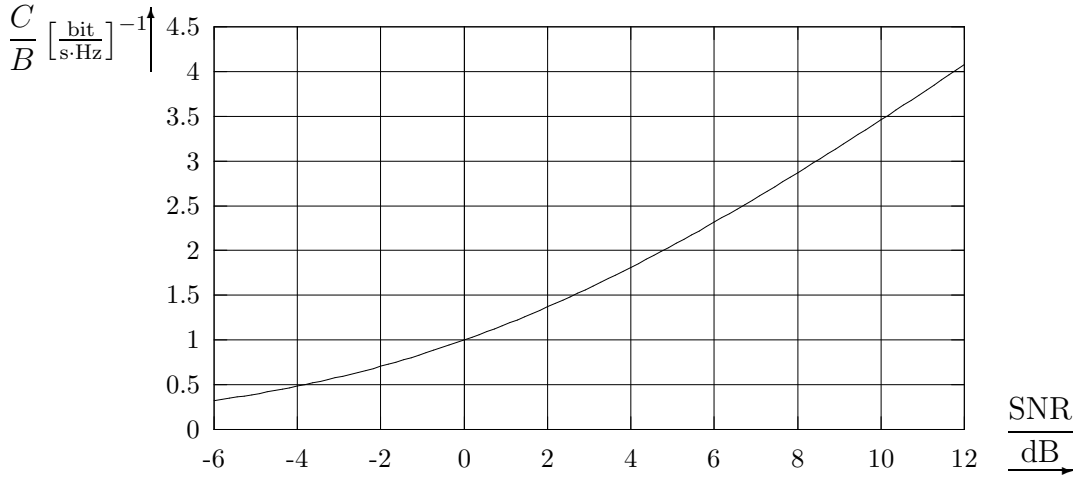


Bild 5.9: Spektrale Effizienz $\frac{C}{B}$ in Abhängigkeit vom Signal-Stör-Verhältnis SNR

Satz von Shannon:
Ein auf die Bandbreite B begrenzter gedächtnisloser Kanal, der durch additives weißes Gaußrauschen der Leistung $N_0 \cdot B$ gestört wird, hat pro Zeiteinheit die Kanalkapazität:

$$C = B \cdot \mathrm{ld}\left(1 + \frac{P}{N_0 \cdot B}\right) \cdot [\mathrm{bit}]\,,$$

wobei für die maximale Transinformation das übertragene Signal mit der endlichen mittleren Leistung P eine gaußsche Wahrscheinlichkeitsdichte aufzuweisen hat.

Bandbreite $\to \infty$: Kanal mit $P = \mathrm{const}$, $N_0 = \mathrm{const}$ (z.B. Funkkanal), $C = f(B)$.

$$\begin{aligned}\lim_{B\to\infty} C\,/\,[\mathrm{bit}] &= \lim_{B\to\infty} \left.\frac{\mathrm{ld}\left(1+\frac{P}{N_0\cdot B}\right)}{\frac{1}{B}}\right|_{\text{l'Hôpital}} \\ &= \lim_{B\to\infty} \mathrm{ld\,e}\,\frac{\frac{1}{\left(1+\frac{P}{N_0\cdot B}\right)}\cdot\frac{P}{N_0}\cdot\frac{-1}{B^2}}{-\frac{1}{B^2}} = \frac{P}{N_0}\cdot\underbrace{\mathrm{ld\,e}}_{\approx 1{,}44\,\widehat{=}\,1{,}6\mathrm{dB}}\end{aligned}$$

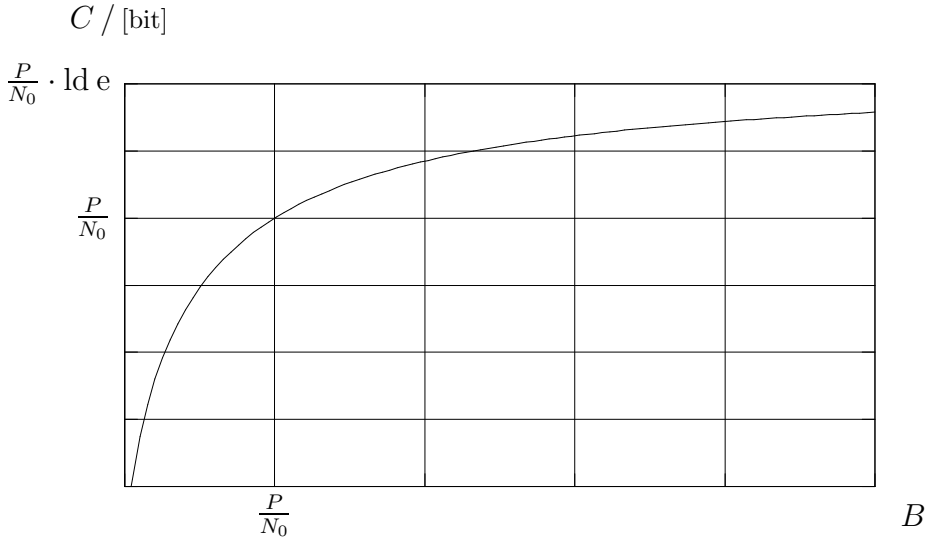


Bild 5.10: Kanalkapazität C abhängig von der Bandbreite B bei konstantem N_0

Bsp.: Telefonkanal mit $B = (3,4 - 0,3)\,\mathrm{kHz} = 3,1\,\mathrm{kHz}$; $P_{\mathrm{empf}} = 50\,(\mathrm{mV})^2/\Omega$; $N_0 = 0,005\,\frac{(\mathrm{mV})^2/\Omega}{\mathrm{kHz}}$

$$\begin{aligned}\frac{P_{\mathrm{empf}}}{B\cdot N_0} &= \frac{50\cdot 10^{-6}\mathrm{W}}{3,1\,\mathrm{kHz}\cdot 5\cdot 10^{-9}\frac{\mathrm{W}}{\mathrm{kHz}}} = 3.2\cdot 10^3 \;\widehat{=}\; 35\,\mathrm{dB} \\ \Longrightarrow \frac{C}{B} &\approx 11.7\,\tfrac{\mathrm{bit}}{\mathrm{s\cdot Hz}} \Longrightarrow C \approx 36\,\tfrac{\mathrm{kbit}}{\mathrm{s}}\end{aligned}$$

Anmerkung: Mit V.34-Modem (Trellis-codierte Modulation + Preshaping + Entzerrer) sind $33,6\ \frac{\text{kbit}}{\text{s}}$ möglich.
$\lim\limits_{B\to\infty} C \approx 14,4\ \frac{\text{Mbit}}{\text{s}}$ Kanalkapazität, wenn keine Bandbreitenbeschränkung.

Bsp.: Funkkanal mit SNR $= -10\,\text{dB}$ zulässig, geforderte Kapazität $C = 2,4\ \frac{\text{kbit}}{\text{s}}$

$$\Rightarrow \frac{P}{\sigma_{\mathbf{n}}^2} = 10^{-1} \Rightarrow \frac{C}{B} = \text{ld}\,1,1\,[\text{bit}] \Rightarrow B = \frac{C}{\text{ld}\,1,1\,[\text{bit}]} \approx 17,454\,\text{kHz erforderlich!}$$

Leistungs-Bandbreite-Grenzkurve für AWGN

- Übertragung mit der Rate: $R_m[\frac{\text{bit}}{\text{s}}]$, fehlerfrei wenn $R_m < C$
- Signalenergie pro Bit := E_b $\Rightarrow$ mittlere Signalleistung $P = E_b \cdot R_m$

$\longrightarrow$ Grenzkurve der spektralen Effizienz für fehlerfreie Übertragung: $R_m = C$

$$\Rightarrow \frac{R_m}{B \cdot [\text{bit}]} = \text{ld}\left(1 + \frac{E_b}{N_0} \cdot \frac{R_m}{B}\right) \longrightarrow \frac{2^{\frac{R_m}{B}} - 1}{\frac{R_m}{B}} = \frac{E_b}{N_0}$$

(s. auch Bild 5.11)

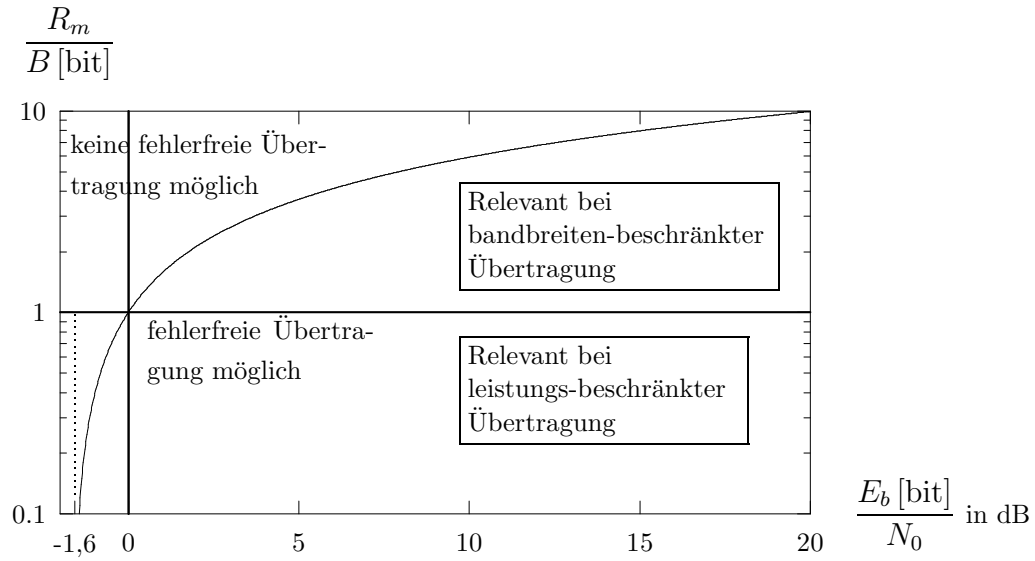


Bild 5.11: Leistungs-Bandbreite-Grenzkurve eines AWGN-Kanals