

Grundlagen

■ Inhalt

◆ Signale

- ✦ Analog/Digital Signale
- ✦ Zeit/Frequenzbereich
- ✦ Übertragungssystem
- ✦ Signalisierungs-/Datenrate

◆ Information

- ✦ Wahrscheinlichkeiten, Informationsgehalt, Entropie

◆ Codierung

- ✦ Diskretisierung, Codierung
- ✦ Kompression (Dynamik, Datenratenreduktion)

Terminologie

■ Signal

- ◆ physikalische Repräsentation von Informationen (x)
- ◆ hat Informationsinhalt, der durch informationstragende Parameter dargestellt wird (Werte/Dynamik, Zeitinformation,...)
- ◆ Störgrößen (z) können Nutzsignal beeinflussen (Rauschen, Übertragungsfehler, ...)

$$\mathbf{y} = \mathbf{O}(\mathbf{x}, \mathbf{z})$$

Terminologie

■ Kontinuierliche/diskrete Signale

◆ Quantisierung des Informationsparameters

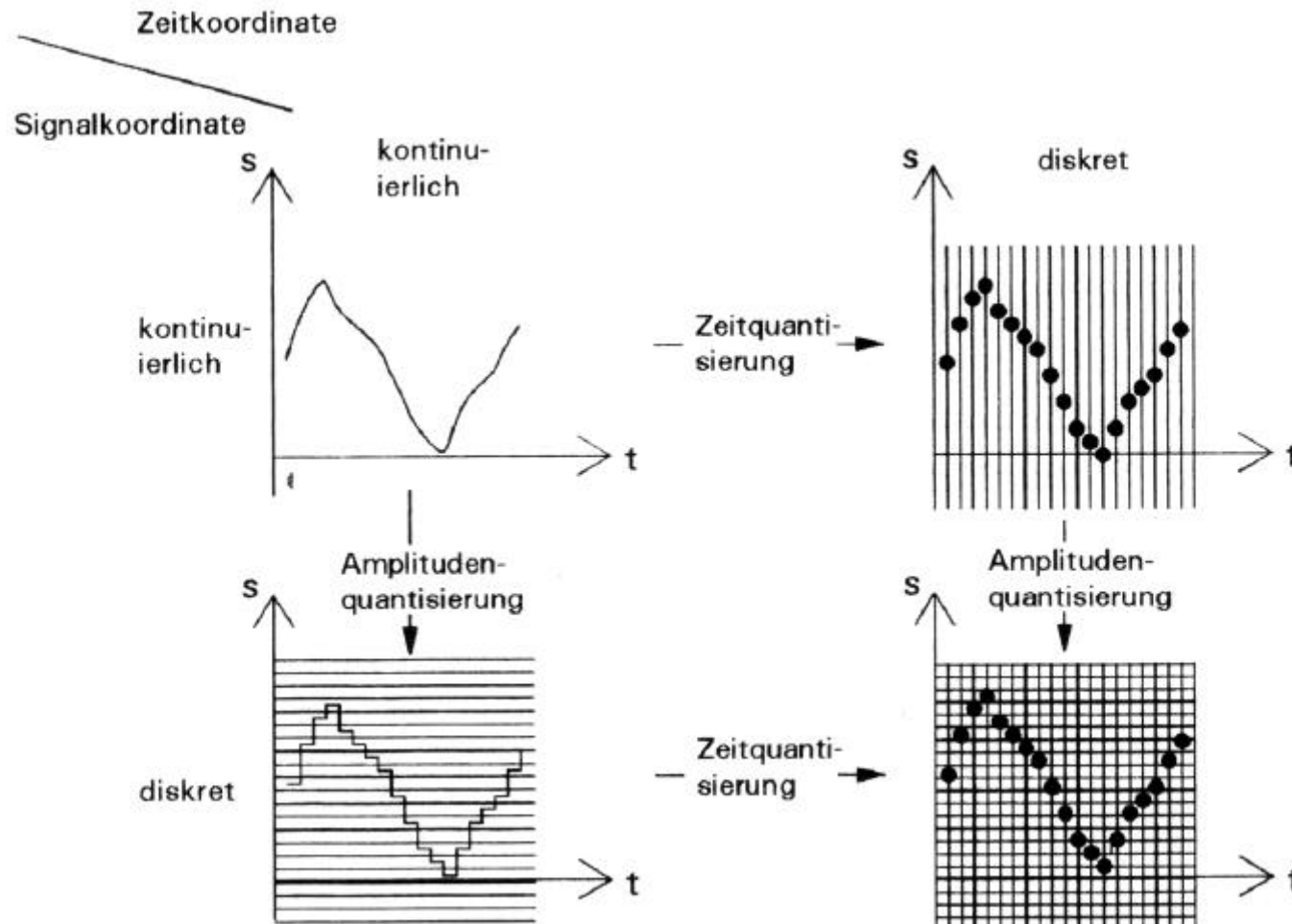
- ✦ Signal kann jeden beliebigen Wert annehmen: „analoges Signal“
- ✦ Signal kann nur bestimmte Werte annehmen: „diskretes Signal“
- ✦ Sonderfall: Signal kann nur 2 Werte annehmen: „Digital/Binärsignal“

◆ Quantisierung des Zeitparameters

- ✦ Informationsparameter kann sich jederzeit ändern: „zeitkontinuierliches Signal“
- ✦ Informationsparameter kann sich nur zu bestimmten (meist äquidistanten) Zeitpunkten ändern: „zeitdiskretes/diskontinuierliches“ Signal

Terminologie

■ Kontinuierliche/diskrete Signale



Terminologie

■ periodische Signale

◆ charakterisiert durch

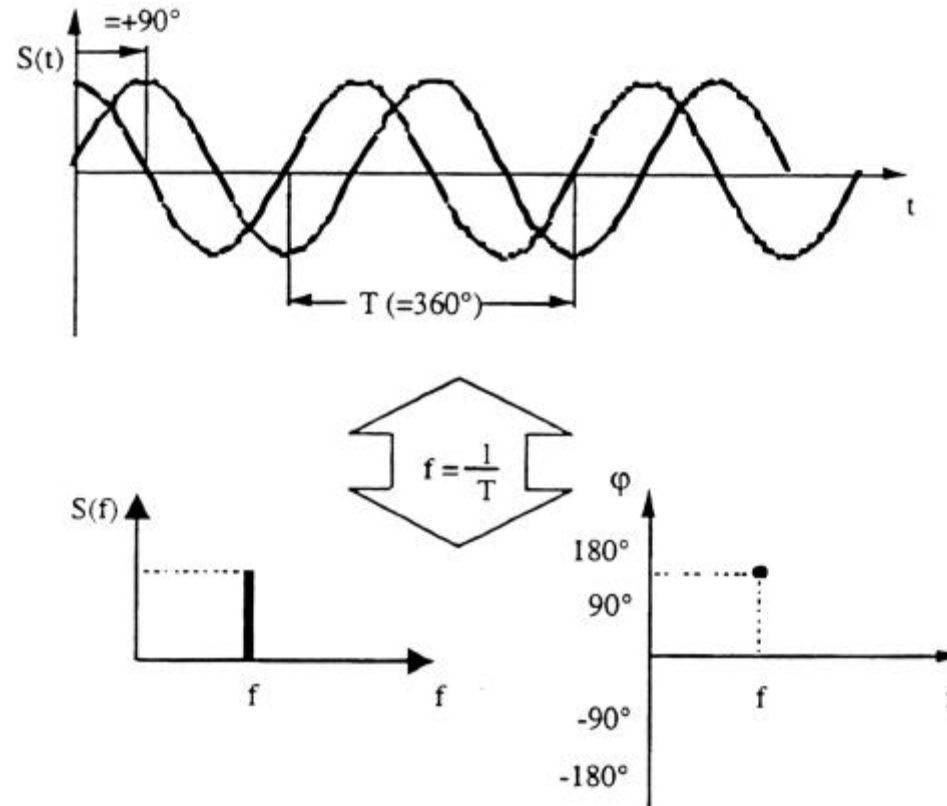
- ✦ Periodendauer T
- ✦ Frequenz $f = 1/T$
- ✦ Amplitude $S(t)$
- ✦ Phase φ

◆ Darstellungsmöglichkeiten

- ✦ im Zeitbereich $S(t)$
 - nicht sehr kompakt (redundante Infos), mit Fourier-Transformation →
- ✦ im Frequenzbereich $S(f)$, $\varphi(f)$
 - erleichtert ibs. Beschreibung von überlagerten Signalen →
 - „Spektrum“

Terminologie

■ periodische Signale - 2

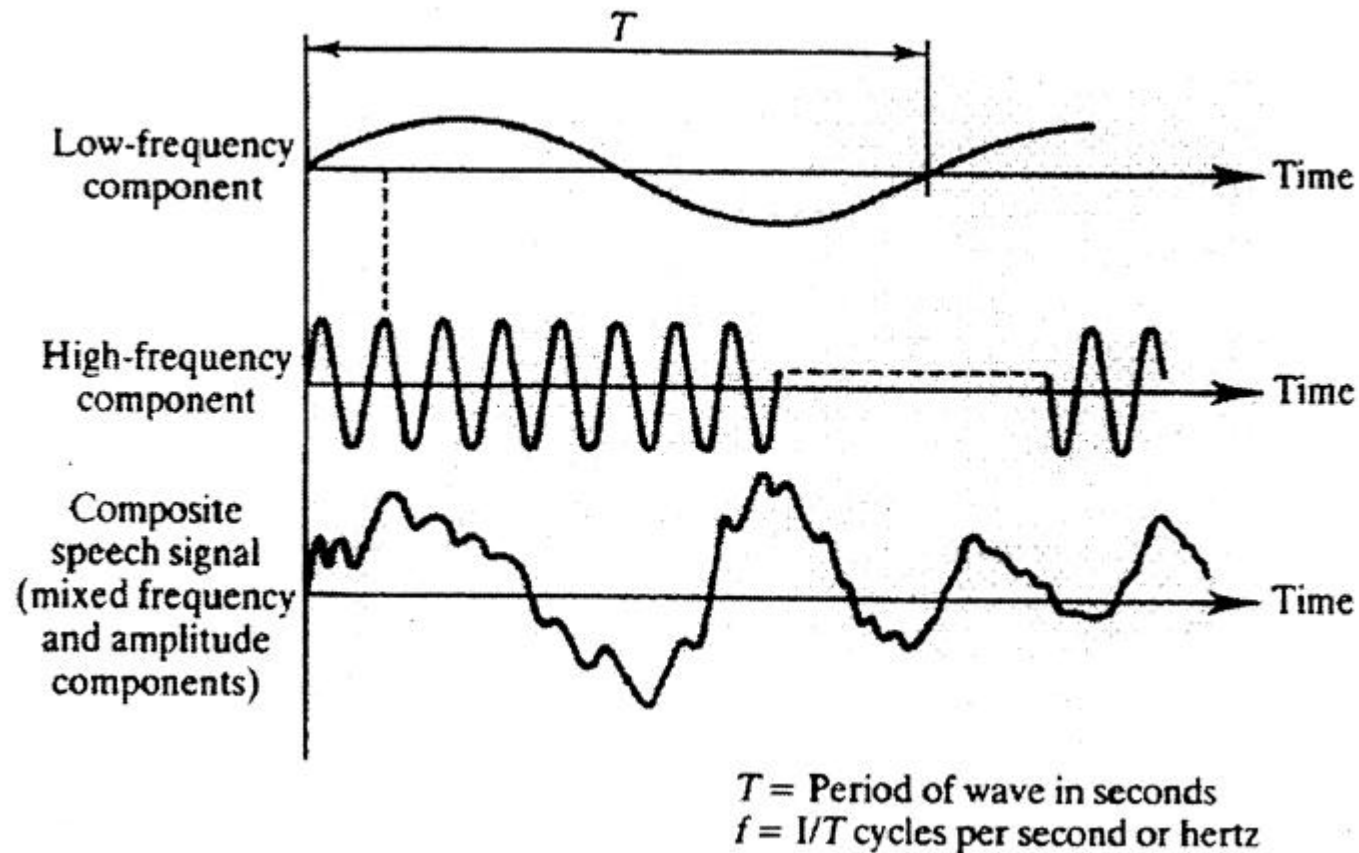


◆ Amplitude $S(f)$, Phase $\varphi(f)$

Terminologie

■ periodische Signale - 3

- ✦ Unterschied zwischen periodischem Signal und z.B. Sprachsignal



Terminologie

■ Bandbreite

- ◆ beschreibt Beschränkung des Frequenzspektrums (Unter/Obergrenze)
- ◆ jedes reale Übertragungssystem hat gewissen Frequenzgang/ Bandbreite (meist sog. Bandpaß-Eigenschaften)
- ◆ Bandbreitenbegrenzung durch
 - ✦ physikalische Eigenschaften des Übertragungsmediums
 - ✦ Eigenschaften der eingesetzten elektronischen Komponenten

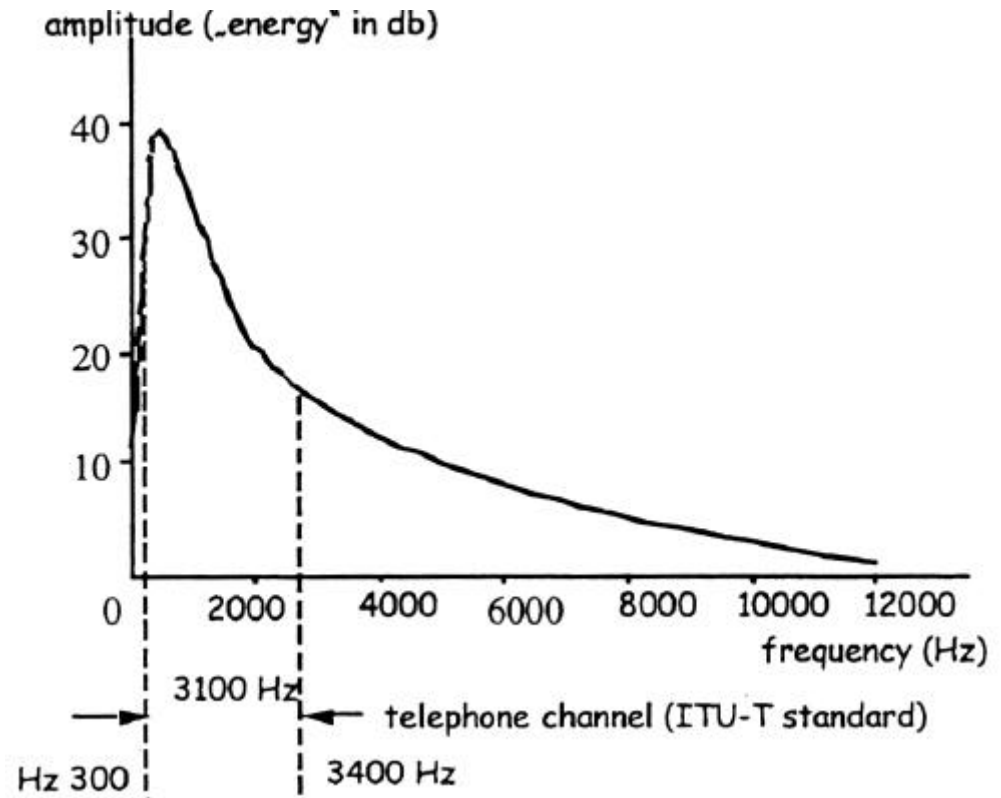
■ Konsequenzen

- ◆ es können keine beliebigen Signale übertragen werden (weder was Datenrate noch Frequenzbereich angeht)
- ◆ zu übertragendes Signal muß so Codiert werden, daß optimale Übertragungsrate bei gegebener Bandbreite erreicht werden kann

Terminologie

■ Beispiel Telefonie-Kanal

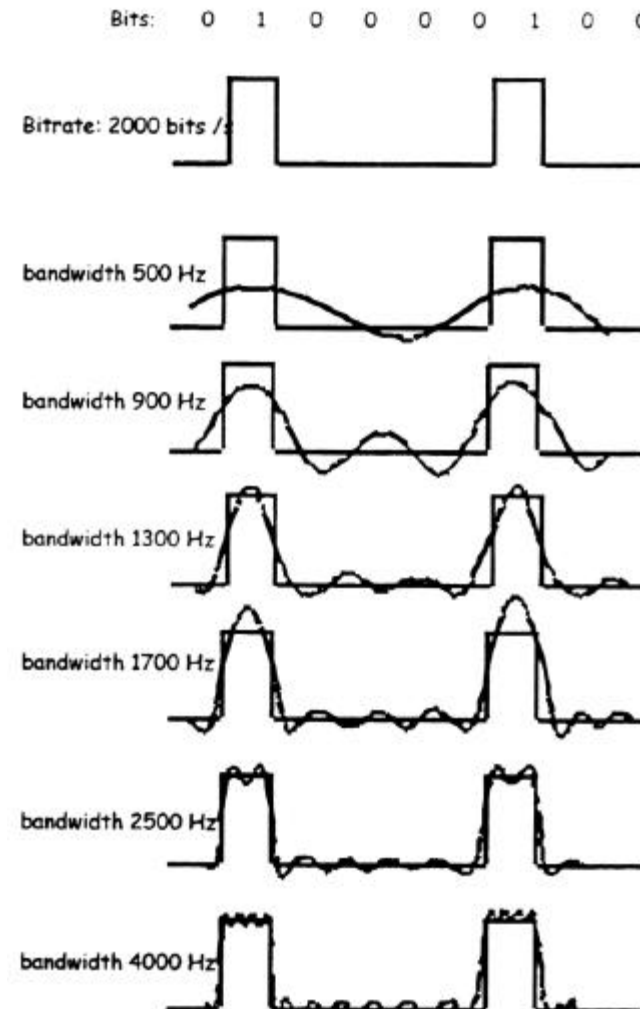
- ◆ angepaßt an spektrale Zusammensetzung der Sprache
 - ✦ untere Grenzfrequenz 300Hz
 - ✦ obere Grenzfrequenz 3,4 kHz
 - ✦ Bandbreite: 3,1 kHz



Terminologie

■ Beispiel Übertragung Digitalsignal

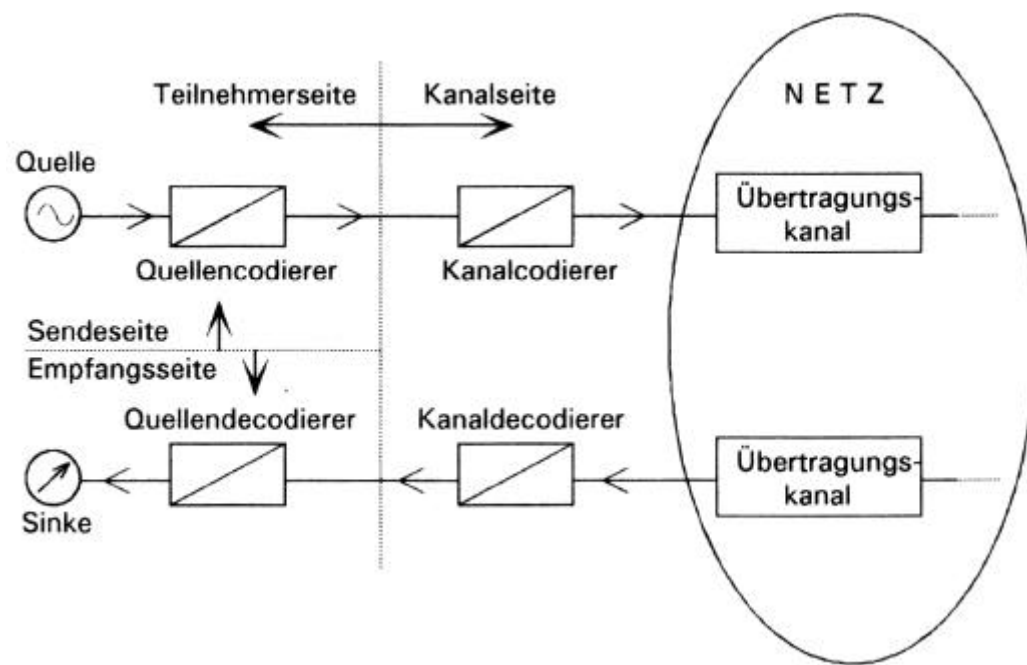
- ◆ je größer Bandbreite um so größeres Spektrum kann übertragen werden und desto exakter kann Originalsignal wiedergegeben werden



Übertragungssysteme

■ Übertragungssysteme

- ◆ bestehen aus Sender - Übertragungskanal - Empfänger



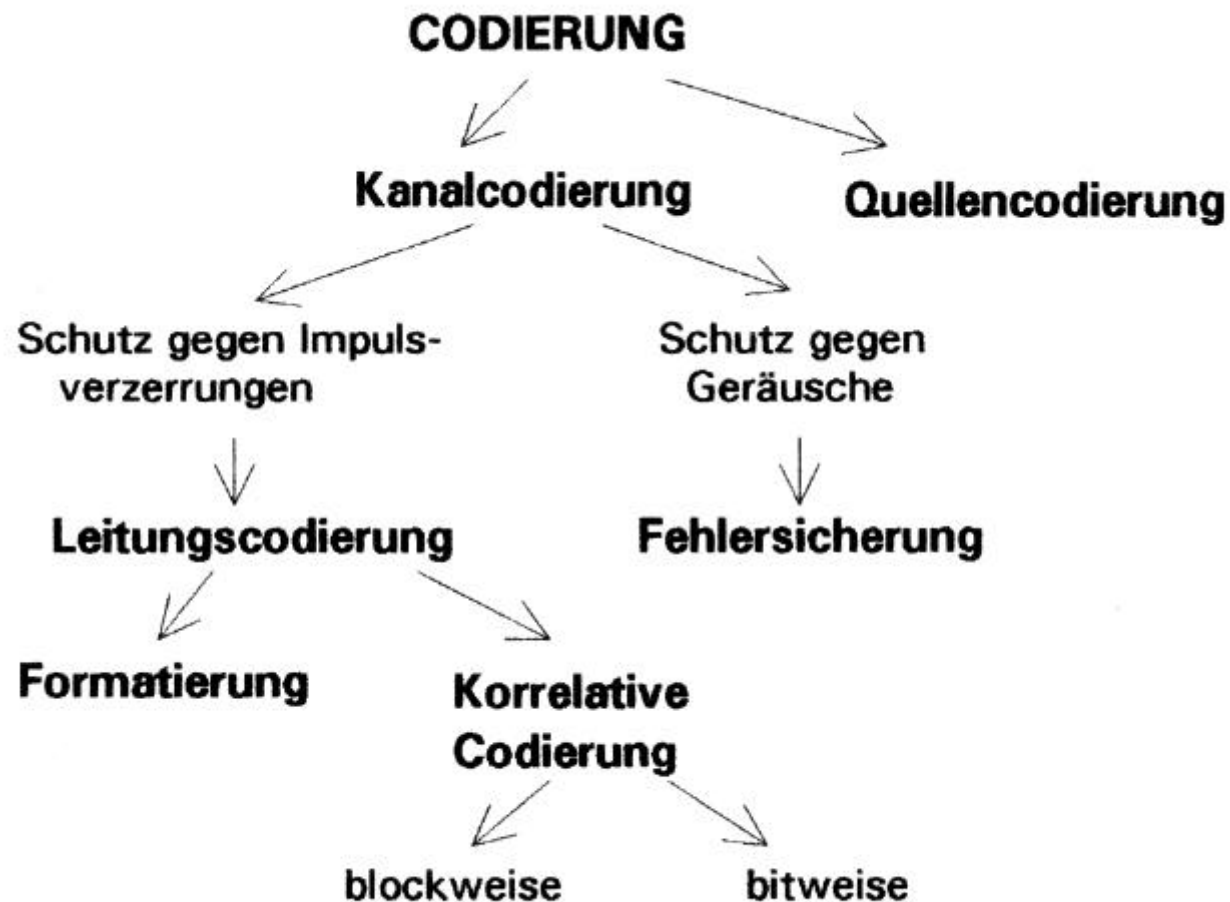
Übertragungssysteme

■ Quellen-/Kanalcodierung/-decodierung

- ◆ Sendersignale müssen entsprechend umgewandelt (codiert) werden und Übertragungskanal angepaßt werden (moduliert)
- ◆ Empfängerseitig müssen Signale demoduliert und (möglichst) wieder in Ursprungsform gebracht werden (decodiert)
 - ✦ jeder Kommunikationsteilnehmer hat meist beide Funktionen im System (Codec und Modem)
- ◆ die Trennung in Quellencodierer/-decodierer und Kanalcodierer/-decodierer erlaubt die Auswahl des optimalen Codecs für die zu übertragende Nutzsignalart unabhängig vom verwendeten Übertragungskanal
 - ✦ Quellencodierung dient dabei der Reduktion von Redundanz und Irrelevanz
 - ✦ Kanalcodierung paßt Signal an (fehlerbehafteten) Übertragungskanal durch Zufügung von Redundanz an

Übertragungssysteme

■ Quellen-/Kanalcodierung/-decodierung - 2



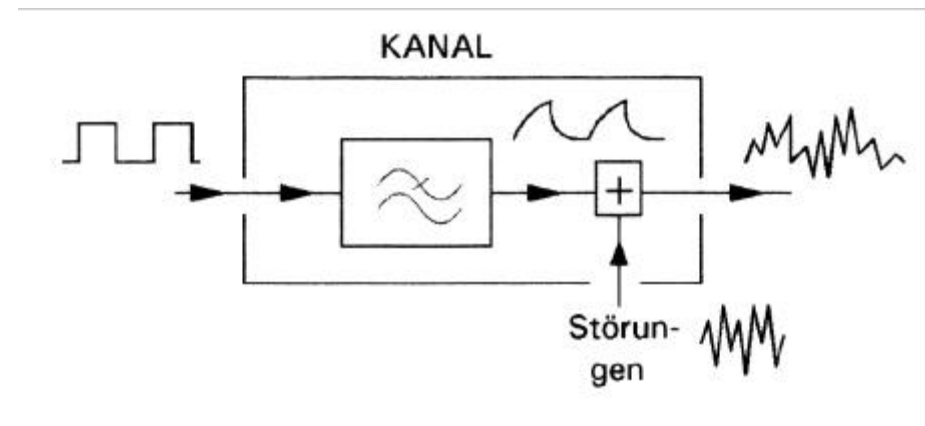
Übertragungssysteme

■ Übertragungskanal besteht aus

- ✦ physikalisches Medium (Glas, Kupfer) oder Funk
- ✦ elektronisches Equipment (Verstärker, Regeneratoren)
- ✦ Steckverbindungen, ...

◆ Kanal beeinflusst durchlaufendes Signal durch

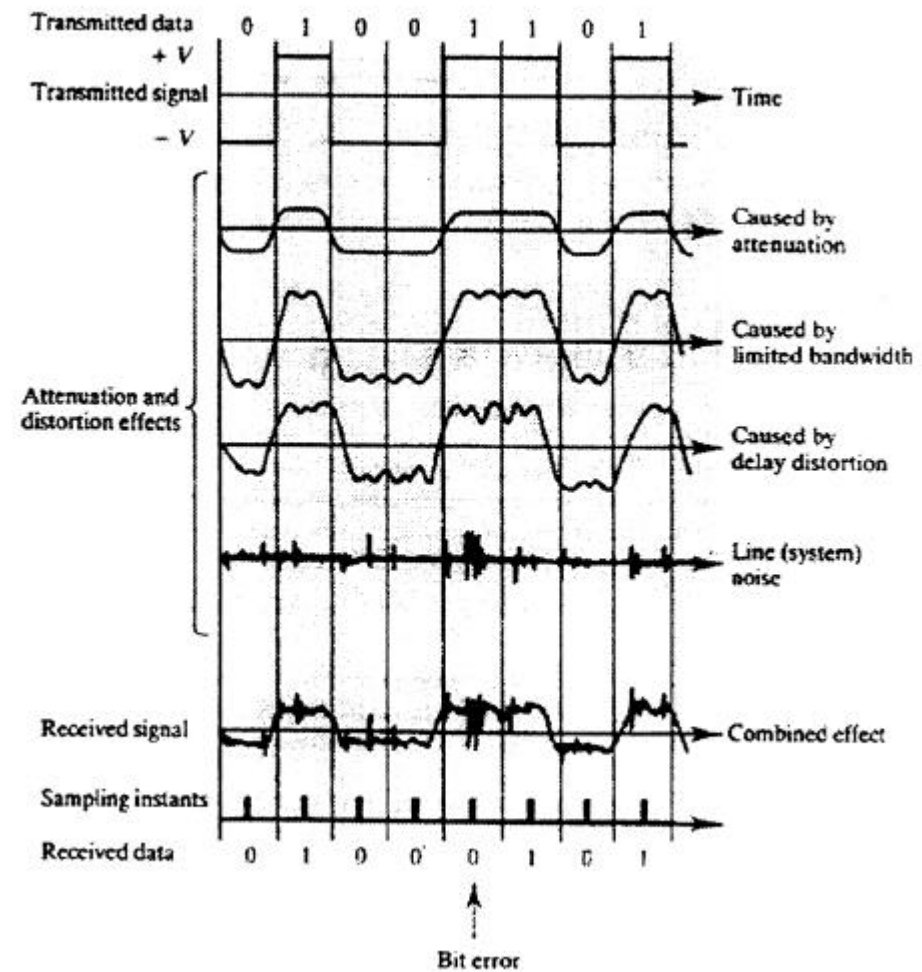
- ✦ Bandbreitenbeschränkungen (Impulsverzerrungen)
- ✦ Signalabschwächung (Dämpfung)
- ✦ Unterschiedliche Laufzeiten von Signalanteilen (Gruppenlaufzeiten)
- ✦ Störsignale
 - Impulsstörungen
 - Sinusstörungen
 - Nebensprechen
 - Rauschen der Bauelemente
 - atmosphärische Störungen (Funk)



Übertragungssysteme

■ Übertragungskanal - 2

- ◆ durch Störeinflüsse
 - ✦ Einschränkung der maximal übertragbaren Datenmenge
 - ✦ Übertragungsfehler



Übertragungssysteme

■ Nyquist (1924)

- ◆ störungsfreier Kanal mit Bandbreite H
- ◆ Signal mit V diskreten Werten

$$\text{max. Datenrate} = 2 H \log_2 V$$

- ◆ z.B: $H=3000\text{Hz}$, Binärsignal ($V=2$), max. Datenrate = 6000 bit/s
- ◆ störungsfreier Kanal in der Realität aber nicht anzutreffen
- ◆ Erweiterung um Störeinflüsse notwendig
- ◆ Begriff des Signal/Rausch-Abstandes (S/R , SNR)
 - ✦ Verhältnis des Nutzsignalpegels zum Störsignalpegel

Übertragungssysteme

■ Shannon (1948)

- ◆ Kanal mit Bandbreite H und statistischem Rauschen
- ◆ Signal-Rauschabstand SNR bekannt
- ◆ Signal mit V diskreten Werten

$$\text{max. Datenrate} = H \log_2 (1 + \text{SNR})$$

- ◆ z.B: Modem für Telefonleitung ($H=3100\text{Hz}$, $\text{SNR}=48\text{dB}$)

$$\text{SNR [dB]} = 10 \cdot \log_{10}(S/R)$$

$$48 = 10 \cdot \log_{10}(S/R) \rightarrow S/R \sim 60000$$

$$\begin{aligned} \text{max. Daten rate} &= 3100 * \log_2 (1 + 60000) \\ &= \sim 49.000 \text{ bit/s} \end{aligned}$$

Übertragungssysteme

■ Signalisierungs- und Datenrate

- ◆ Signalisierungsrate v_s [Baud]
 - ✦ Anzahl der übertragenen Symbole pro Sekunde (wie oft am Medium ein Signalwechsel stattfindet)

- ◆ Datenrate v_B [bit/s]
 - ✦ Anzahl der Bits die pro Sekunde übertragen werden

- ◆ Entscheidend
 - ✦ wieviele Bits werden pro Symbol (Signalwert) übertragen
 - ✦ daraus ergibt sich das Modulationsverfahren für das Übertragungsmedium (Baseband, QAM, PSK, ...)

Übertragungssysteme

■ Signalisierungs- und Datenrate

◆ Binärsignale $v_B = v_S$

- ✦ wenn zusätzlich noch Synchronisationsinfos (Clock,...) enthalten

$$v_B < v_S$$

◆ Modulationsverfahren $v_B > v_S$

- ✦ pro Symbol wird Information aus n möglichen Werte übertragen

$$v_B = v_S \lg(n)$$

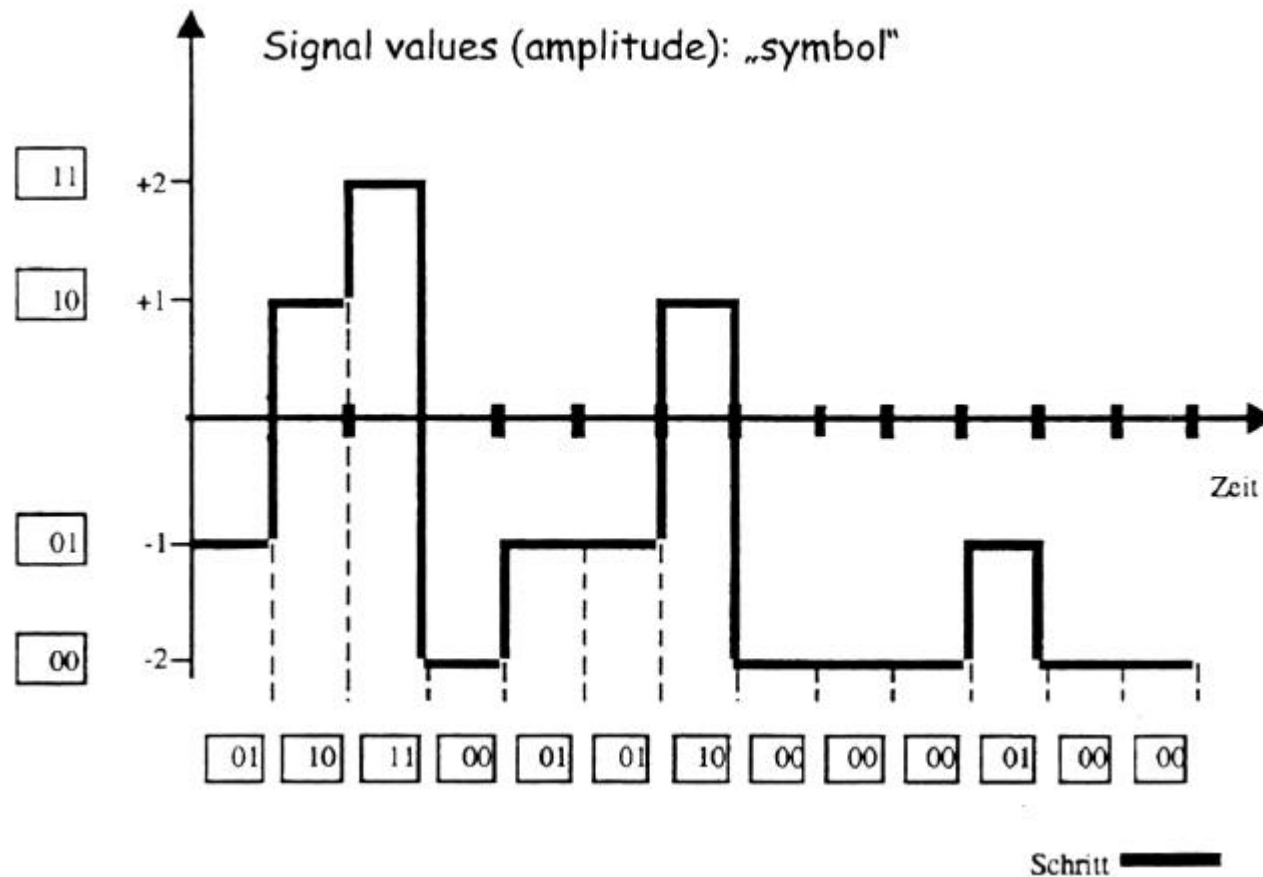
- ✦ z.B:

$n = 2$	Binärsignal
$n = 3$	ternäres Signal
$n = 4$	quaternäres Signal (DIBIT, QPSK-Modulator)
$n = 16$	z.B. 16-QAM Modulator

Übertragungssysteme

■ Signalisierungs- und Datenrate - 3

- ◆ Bsp: DIBIT



Übertragungssysteme

■ Synchronisation

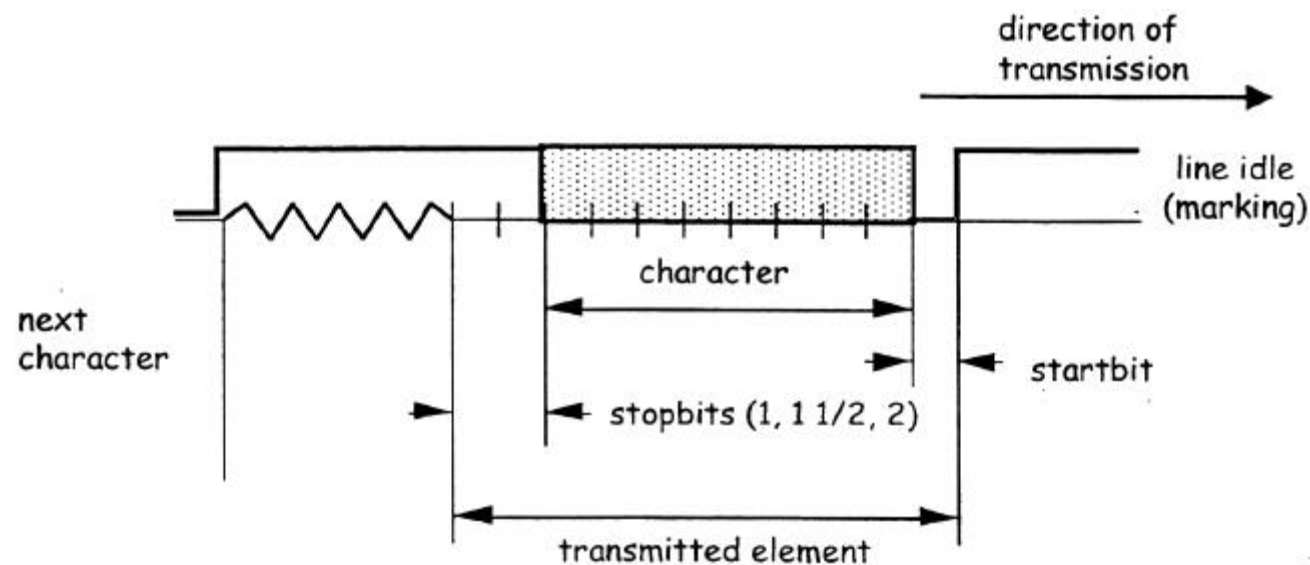
- ◆ um auf Empfängerseite die übertragene Info wiederzugewinnen muß Empfänger wissen, wann welche Daten von Sender ankommen („sample time“)
- ◆ Empfänger muß sich auf Sendesignal „synchronisieren“
- ◆ mehrere Möglichkeiten
 - ✦ Sender-/Empfänger haben idente-synchrone Clocks (nur lokal möglich)
 - ✦ extra Übertragungskanal für Synchronisations-Informationen (z.B. eigene Clock-Leitung)
 - ✦ Synchronisationsinformation führt die Nutzdaten an
 - asynchrone Übertragung (Start/Stop-Bits)
 - synchrone Übertragung (Blockweise Synchronisation)
 - ✦ Synchronisationsinfo ist im Nutzsignal enthalten

Übertragungssysteme

■ Synchronisation

◆ asynchrone Übertragung

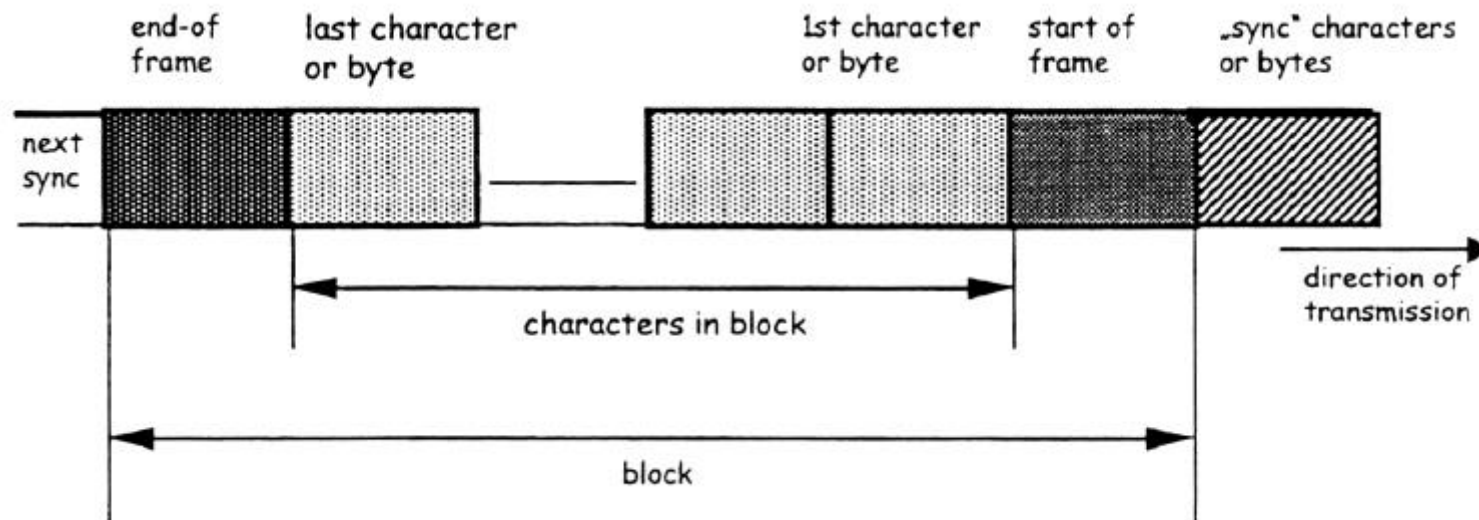
- ✦ Empfänger synchronisiert sich auf Sendeclock durch Startbit
- ✦ Symbolende durch Stopbit(s) markiert



Übertragungssysteme

■ Synchronisation

- ◆ (block)synchrone Übertragung
 - ✦ Empfänger synchronisiert sich auf Sendeclock durch Sync-Zeichen gefolgt von SOF (Start Of Frame) Symbol
 - ✦ Blockende durch EOF (End Of Frame) markiert



Übertragungssysteme

■ Kanalcodierung

- ◆ so zu wählen, daß dem Übertragungsmedium entsprechend Symbole übertragen werden
- ◆ wichtig ist Clock-Regenerierung (Synchronisation) auf Empfängerseite
- ◆ 2 grundlegende Methoden
 - ✦ Basisbandcodierung
 - Signal wird direkt über Medium übertragen, z.B.
 - Laser ein/aus
 - Spannung High/Low (RS232, RS485,...)
 - Strom ein/aus (current loop, z.B. 20mA TTY)
 - ✦ Modulation
 - periodisches Trägersignal wird mit Symbolen „moduliert“/verändert

Übertragungssysteme

■ Basisbandcodierung

- ◆ einfach zu implementieren
- ◆ Digitalsignal steuert direkt Übertragungselektronik

- ◆ Probleme:
 - ✦ Pegel zur 0/1 Erkennung sehr störanfällig (geg. Rauschen)
 - ✦ Wunsch/Forderung $v_B > v_S$ schlecht erfüllbar
 - ✦ Taktrückgewinnung auf Empfängerseite (lange 0/1 Folgen)
 - ✦ Gleichspannungsoffset $\leq 0V$
 - ✦ schlecht geeignet für analoge Übertragungstrecken (Realität)

- ◆ Lösung
 - ✦ Basisbandencoder um zusätzliche Infos zur z.B. Taktregenerierung in Signal einzufügen
 - ✦ Flanken um 0/1 Bits zu erkennen (bessere Störfestigkeit)

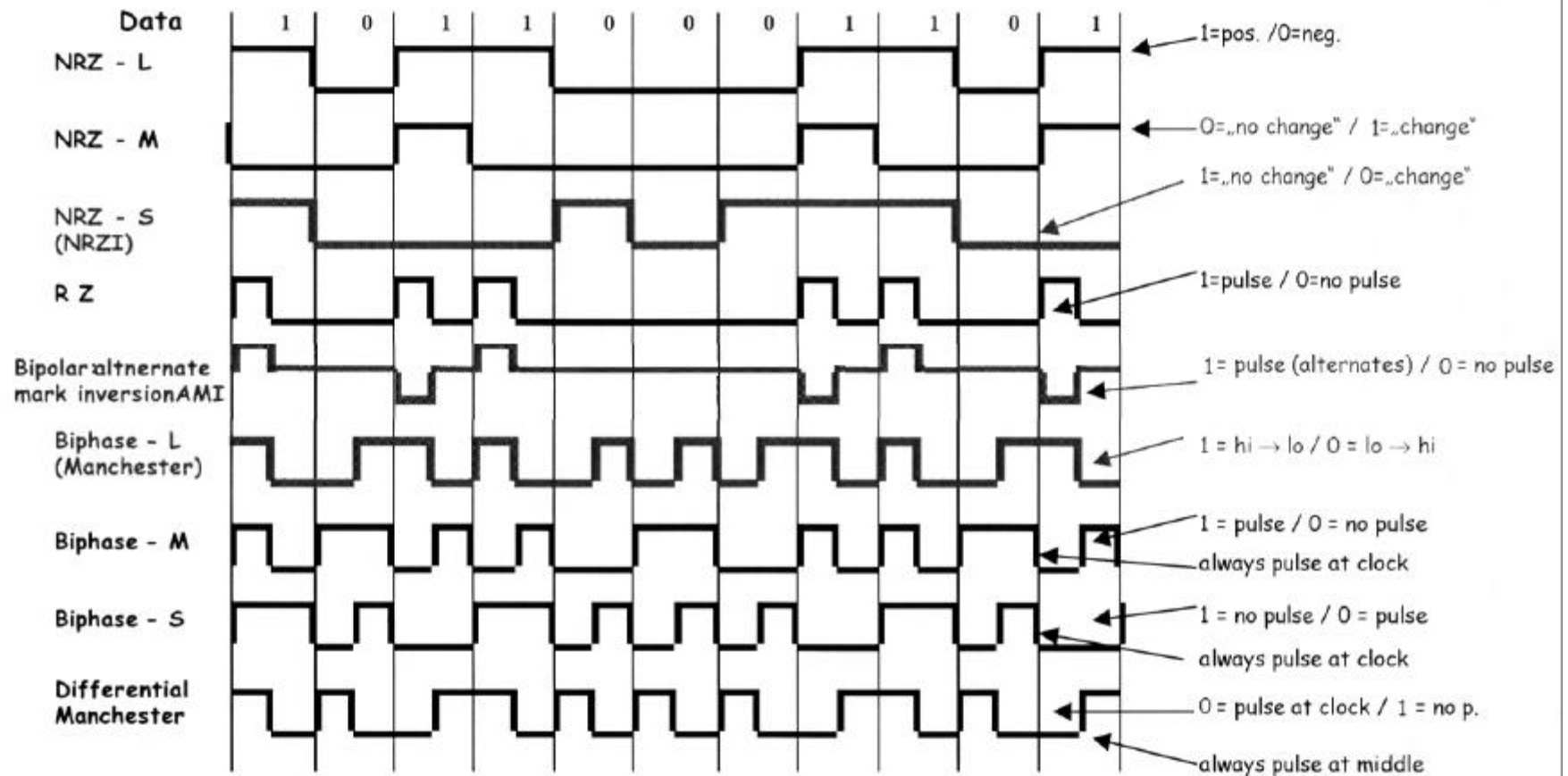
Übertragungssysteme

■ Basisbandcodierung

- ◆ NRZ-L (Non-Return to Zero-Level)
 - ✦ fixe Spannungspegel für 0/1 Bits
- ◆ NRZ-M/S oder NRZ-I (Non-Return to Zero Inverted)
 - ✦ Pegelwechsel nur bei 0 oder 1 Bit
- ◆ RZ (Return to Zero)
 - ✦ 1-Bits durch Impuls dargestellt
- ◆ Biphas-L
 - ✦ 0/1 Information durch Pegelwechsel in Taktmitte
- ◆ Biphas-M/S
 - ✦ 0/1 Information durch Wechsel/kein Wechsel in Taktmitte, Pegelwechsel immer zur Taktzeit
- ◆ Differential Manchester
 - ✦ 0/1 Information durch Wechsel/kein Wechsel zur Taktzeit, Pegelwechsel immer in Taktmitte

Übertragungssysteme

■ Basisbandcodierung



Übertragungssysteme

■ Basisbandcodierung - 4b/5b

- ✦ 4bit Nutzdaten zu 5bit Symbol (80% Effizienz)
- ✦ 0/1 Wechsel spätestens alle 4 Bits
- ✦ z.B: FDDI; Linespeed: 125Mbit/s (100Mbit/s netto)

abbrev.	symbol	data/meaning
Q	00000	Quiet
I	11111	Idle
H	00100	Halt (Forced Break)
J	11000	1 st of sequential SD-Pair
K	10001	2 nd of sequential SD-Pair
0	11110	0000
1	01001	0001
2	10100	0010
3	10101	0011
4	01010	0100
5	01011	0101
6	01110	0110
7	01111	0111
8	10010	1000
9	10011	1001
A	10110	1010
B	10111	1011
C	11010	1100
D	11011	1101
E	11100	1110
F	11101	1111
T	01101	ED (for Data Stream)
R	00111	Logical ZERO (reset)
S	11001	Logical ONE (set)

Übertragungssysteme

■ Basisbandcodierung - 4B/3T

- ✦ 4bit Nutzdaten in 3 ternären Symbolen codiert
- ✦ Zustandswechsel in Codierungstabelle, kein DC-Anteil
- ✦ z.B: ISDN: 192kbit/s Nutzdatenrate, 144kSymbole/s

data	S1	Sn	S2	Sn	S3	Sn	S4	Sn
0001	0-+	1	0-+	2	0-+	3	0-+	4
0111	-0+	1	-0+	2	-0+	3	-0+	4
0100	-+0	1	-+0	2	-+0	3	-+0	4
0010	+ -0	1	+ -0	2	+ -0	3	+ -0	4
1001	+0-	1	+0-	2	+0-	3	+0-	4
1110	0+-	1	0+-	2	0+-	3	0+-	4
1001	+++	2	+++	3	+++	4	---	1
0011	00+	2	00+	3	00+	4	--0	2
1101	0+0	2	0+0	3	0+0	4	-0-	2
1000	+00	2	+00	3	+00	4	0--	2
0110	---	2	---	3	---	2	---	3
1010	++-	2	++-	3	++-	2	++-	3
1111	++0	3	00-	1	00-	2	00-	3
0000	+0+	3	0-0	1	0-0	2	0-0	3
0101	0++	4	-00	1	-00	2	-00	3
1100	+++	4	++-	1	+-	2	+-	3

Übertragungssysteme

■ Modulation

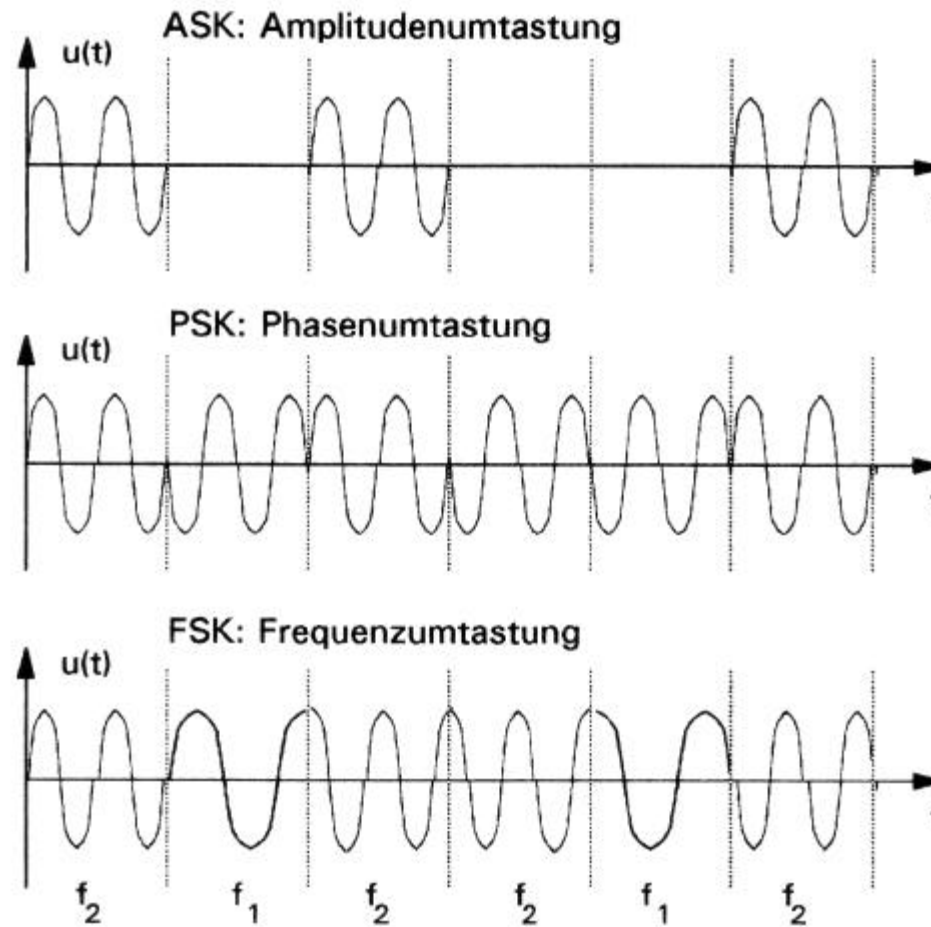
- ◆ für Übertragungskanäle mit geringer Bandbreite oder mit Bandpaßcharakteristik
- ◆ Sinusträger wird mit digitalem Signal moduliert

- ◆ 3 Modulationsarten
 - ✦ Amplitudenmodulation A
 - ✦ Frequenzmodulation Δf
 - ✦ Phasenmodulation φ

$$u(t) = A \sin(2\pi(f + \Delta f)t + \varphi)$$

Übertragungssysteme

■ Modulation



xSK = ... Shift Keying

Übertragungssysteme

■ Fehlersicherung

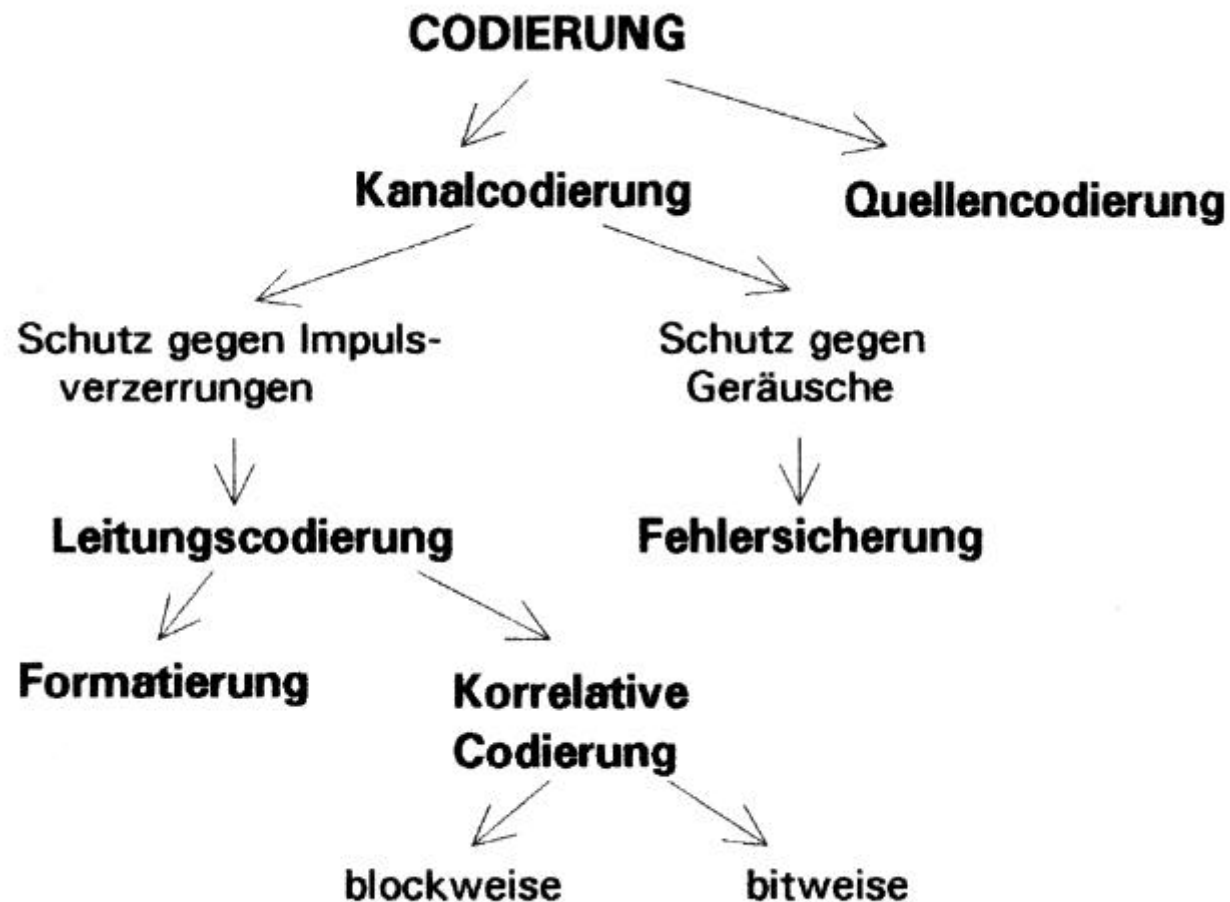
- ◆ Zufügen von Redundanz mittels Codierer um Übertragungsfehler auf der Empfängerseite erkennen und korrigieren zu können

■ prinzipiell 2 Arten von Codierung möglich

- ◆ Blockcodes
 - ✦ fixe Symbolgrößen (z.B. Byte mit 8 Bits)
 - ✦ an Symbole werden zusätzliche Bits mit Fehlerkorrekturinformationen angehängt (z.B. Parität)
- ◆ Sequentielle/Faltungscodes
 - ✦ kontinuierlicher Bitstrom wird umcodiert, sodaß Fehlererkennung/korrektur möglich wird

Übertragungssysteme

■ Quellen-/Kanalcodierung/-decodierung



Quellencodierung

■ Quellencodierung/-decodierung

- ◆ dient der Umwandlung von Signalquellen in eine (binäre) Darstellung, die über Übertragungskanal transportiert werden kann

- ◆ unterschiedliche Quellen mit unterschiedlichen Eigenschaften
 - ✦ Audio (Ton/Sprache)
 - ✦ Bilddaten/Bewegtbilder (Video)

- ◆ Quellencodierung soll dabei
 - ✦ Redundanz/Irrelevanz aus Quellendaten entfernen
 - ✦ dadurch Leistungs/Bandbreitengewinn erzielen
 - ✦ auf der Decodierseite jedoch wieder (möglichst exakt) das Originalsignal herstellen können
 - ✦ Art der Codierung/Reduktion hängt dabei maßgeblich vom entsprechenden Sinnesorgan ab

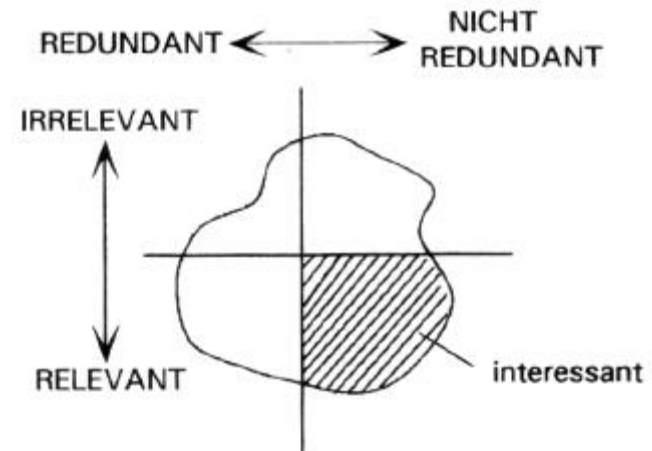
Quellencodierung

■ Nachrichtenebene

- ◆ Nachricht kann in 4 unterschiedliche Teile zerlegt werden
 - ✦ irrelevanter (unwichtiger)
 - ✦ relevanter (wichtiger)
 - ✦ redundanter (bereits bekannter)
 - ✦ nicht redundanter (unbekannter)

- ◆ für Empfänger ist nur der nicht redundante, relevante Teil interessant (bringt Neuigkeit)

- ◆ Quellencodierung sollte eine Nachrichtenreduktion so durchführen, daß Irrelevanz und Redundanz möglichst vollständig eliminiert wird, bring Einsparung an
 - ✦ Bandbreite
 - ✦ Sendeleistung
 - ✦ Speicherbedarf / Übertragungszeit



Quellencodierung

■ Nachrichtenebene - Bsp. Irrelevanz

- ◆ Nachrichten werden von Auge/Ohr wahrgenommen
- ◆ psycho-physiologische Eigenschaften dieser Sinnesorgane bewirken nur beschränkte Aufnahmefähigkeit
- ◆ gewisse Nachrichtenanteile werden nicht wahrgenommen → Irrelevanz → Datenreduktion

- ◆ Auge
 - ✦ S/W Auflösung deutlich höher wie Farbauflösung
 - ✦ flüssige Bewegungen ab ca. 6 Bilder/s erkennbar
 - ✦ Auflösung der Netzhaut bei starrer Projektion ca. 600 Zeilen
 - ✦ unempfindlich gegen Bildstörungen und Bildaussetzer

Quellencodierung

■ Nachrichtenebene - Irrelevanz

◆ Ohr

- ✦ Frequenzbereich ca. 15Hz bis 15kHz
- ✦ für gut verständliche Sprachübertragung 300Hz-3,4kHz
- ✦ keine Auswertung der Phasenlage von Signalen
- ✦ Hörschwelle muß für Wahrnehmung von Signalen überschritten werden
- ✦ Verdeckungseffekt bewirkt, daß lauter Ton benachbarte Töne unhörbar macht (abhängig vom Schallpegel und Frequenzabstand)
- ✦ sehr empfindlich gegenüber Frequenzschwankungen und Aussetzern

- ◆ Qualitative Aussagen werden in der Shannon'schen Informationstheorie (1948) systematisiert und für technischen optimalen Entwurf von Übertragungssystemen quantifizierbar gemacht.

Informationstheorie

■ formale Beschreibung der Nachrichtenquelle

- ◆ Quelle X hat Nachrichtenvorrat $X=\{x_1,\dots,x_n\}$
- ◆ *Entscheidungsgehalt* einer Quelle mit n Zeichen ist definiert mit

$$H_0 = \lg(n) \text{ [bit]}$$

- ◆ Bsp:
Quelle hat 40 alphanumerische Zeichen (A..Z, ÄÖÜ, 0...9 und Leerzeichen)

$$H_0 = \lg(40) = 5,32 \text{ bit}$$

Quelle hat 2 Zeichen (0 und 1)

$$H_0 = \lg(2) = 1 \text{ bit}$$

Informationstheorie

■ formale Beschreibung der Nachrichtenquelle

- ◆ noch keine Aussage, ob Quelle auch den Zeichenvorrat wirklich ausschöpft, oder ob jeder Zeichen gleichwahrscheinlich auftritt
- ◆ jedem Zeichen aus Nachrichtenvorrat $X=\{x_1,\dots,x_n\}$ wird nun zusätzlich die Auftrittswahrscheinlichkeit $p(x_1), \dots, p(x_n)$ zugeordnet
- ◆ *Informationsgehalt* eines Zeichens x_k ist dann definiert mit

$$I(x_k) = \text{ld}\left(\frac{1}{p(x_k)}\right) = -\text{ld}(p(x_k)) \quad [\text{bit}]$$

◆ Bsp:

Fußgänger an Weggabelung, 2 Möglichkeiten, $p_1=p_2=0,5$

$$I(x_1)=I(x_2)=\text{ld } 1/0,5 = \text{ld } 2 = 1 \text{ bit}$$

wenn aber 1 Weg fix ($p_1=1, p_2=0$) gibt's keine Unsicherheit und damit auch keine Information $I(x_1) = \text{ld } 1/1 = \text{ld } 1 = 0 \text{ bit}$

Informationstheorie

■ formale Beschreibung der Nachrichtenquelle

- ◆ Maßzahl für mittleren Informationsgehalt einer Quelle?
- ◆ jedes Zeichen aus Nachrichtenvorrat $X=\{x_1,\dots,x_n\}$ hat wieder die Auftrittswahrscheinlichkeit $p(x_1), \dots, p(x_n)$ zugeordnet
- ◆ *mittlerer Informationsgehalt* einer Quelle (*Entropie*) ist dann definiert mit

$$H(X) = \sum_{k=1}^N p(x_k) I(x_k) = \sum_{k=1}^N p(x_k) \text{ld}\left(\frac{1}{p(x_k)}\right) = -\sum_{k=1}^N p(x_k) \text{ld}(p(x_k))$$

- ◆ Bsp:

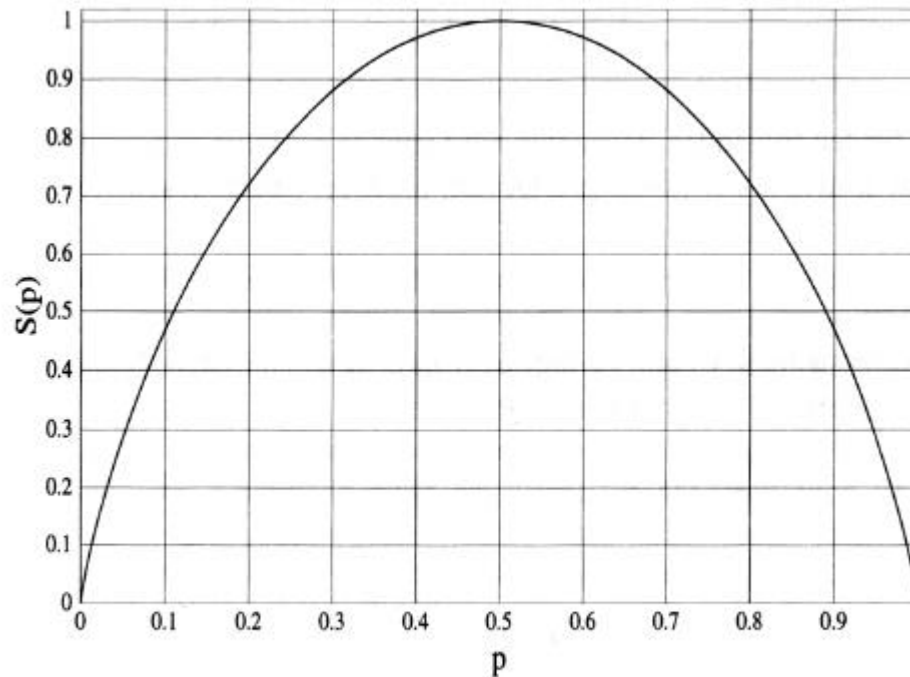
alle Zeichen haben gleiche Auftrittswahrscheinlichkeit $p(x_i)=1/N$

$$H(X) = \sum_{k=1}^N \frac{1}{N} \text{ld}(N) = \text{ld}(N)$$

Entscheidungsgehalt H_0 = mittlerer Informationsgehalt $H(X)$

Informationstheorie

- ◆ Bsp: Quelle mit 2 Zeichen x_1 und x_2 (Binärquelle)
 $p(x_1)=p$ und $p(x_2)=1-p$
 $H(X)=-p \log(p) - (1-p) \log(1-p) = S(p)$ „Shannon Funktion“



Entropie der binären Quelle (Shannon-Funktion $S(p)$)

Informationstheorie

■ formale Beschreibung der Nachrichtenquelle

◆ Schlußfolgerung

- ✦ mittlere Informationsgehalt (Entropie) $H(X)$ ist am größten, wenn alle Zeichen gleiche Auftrittswahrscheinlichkeit haben und entspricht dann dem Entscheidungsgehalt H_0

◆ Die real existierende Differenz zwischen Entscheidungsgehalt H_0 und Entropie $H(X)$ ist „Redundanz“

$$\Delta H = H_0 - H(X)$$

◆ Bsp: Shannon-Funktion

$$H_0 = \log_2 2 = 1 \quad \text{damit} \quad \Delta H = 1 - H(X) = 1 - S(p)$$

„Redundanz ist am größten, wenn Entscheidung sowieso vorweggenommen ist“

Informationstheorie

■ Anwendung in der Quellencodierung

- ◆ Codierung der Quelldaten so wählen, daß
 - ✦ häufig auftretende Datenwerte mit kurzen Codewörtern und
 - ✦ selten auftretende Datenwerte mit längeren Codewörtern codiert werden

- ◆ man erhält damit einen sogenannten „redundanzmindernden Code“

- ◆ Verschiedene Codierungsverfahren um diese Codes zu erhalten
 - ✦ Shannon-Fano-Codierung
 - ✦ Huffman-Codierung
 - ✦ aber auch Morse-Code

Informationstheorie

- Anwendung in der Quellencodierung
 - ◆ Beispiel für Codierung von englischen Texten

Zeichen	Wahrsch.	ASCII-Code	Morse-Code	Huffman-Code
Leerz.	0.1859	00100000	space	000
A	0.0642	01000001	01	0100
B	0.0127	01000010	1000	01111111
C	0.0218	01000011	1010	11111
D	0.0317	01000100	100	01011
E	0.1031	01000101	0	101
F	0.0208	01000110	0010	001100
G	0.0152	01000111	110	011101
H	0.0467	01001000	0000	1110
I	0.0575	01001001	00	1000
J	0.0008	01001010	0111	0111001110
K	0.0049	01001011	101	01110010
L	0.0321	01001100	0100	01010
M	0.0198	01001101	11	001101
N	0.0574	01001110	10	1001
O	0.0632	01001111	111	0110
P	0.0152	01010000	0110	011110
Q	0.0008	01010001	1101	0111001101
R	0.0484	01010010	010	1101
S	0.0514	01010011	000	1100
T	0.0796	01010100	1	0010
U	0.0228	01010101	001	11110
V	0.0083	01010110	0001	0111000
W	0.0175	01010111	011	001110
X	0.0013	01011000	1001	0111001100
Y	0.0164	01011001	1011	001111
Z	0.0005	01011010	1100	0111001111

Informationstheorie

■ Huffman-Codierung

- ◆ optimalen, redundanzreduzierten Code erzeugen
- ◆ Codewörter sollen dabei in Abhängigkeit von der Auftrittswahrscheinlichkeit unterschiedlich lange sein

Gegeben:

N Codewörter x_1, \dots, x_N nach Auftrittswahrscheinlichkeit sortiert mit Wahrscheinlichkeiten $p(x_1) \geq p(x_2) \geq \dots \geq p(x_N)$

Zeichen mit kleinster Wahrscheinlichkeit x_{N-1} und x_N haben dabei eine Codewortlänge von $L(x_{N-1}) = L(x_N)$

Es läßt sich zeigen, daß wenn man diese beiden Zeichen im Entscheidungsbaum mit den Binärwerten 0 und 1 belegt und dann zu einem gemeinsamen Zeichen zusammenfasst (mit den entsprechend addierten Wahrscheinlichkeiten), und dieses „neue“ Zeichen wieder in der Liste der verbleibenden Zeichen nach Auftrittswahrscheinlichkeit einsortiert und rekursiv nach dem gleichen Verfahren fortführt, automatisch ein redundanzreduzierter Code entsteht.

Informationstheorie

■ Bsp Huffman

9 Codewörter mit gegebenen Auftrittswahrscheinlichkeiten

Code	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
$p(x)$	0,22	0,19	0,15	0,12	0,08	0,07	0,07	0,06	0,04

gesucht: Huffman Code dazu

1. Schritt: Codewörter mit geringsten $p(x_i)$ mit 0/1 versehen und p 's zusammenfassen und einsortieren

0: x_8 , 1: x_9 , $p(x_8)+p(x_9)=0,10$

Code	X_1	X_2	X_3	X_4	X_{8+9}	X_5	X_6	X_7	
$p(x)$	0,22	0,19	0,15	0,12	0,10	0,08	0,07	0,07	

Informationstheorie

■ Bsp Huffman - 2

Code	X ₁	X ₂	X ₃	X ₄	X ₈₊₉	X ₅	X ₆	X ₇	
p(x)	0,22	0,19	0,15	0,12	0,10	0,08	0,07	0,07	

2. Schritt: Codewörter mit geringsten $p(x_i)$ mit 0/1 versehen und p's zusammenfassen und einsortieren

0: x₆, 1: x₇, $p(x_6)+p(x_7)=0,14$

Code	X ₁	X ₂	X ₃	X ₆₊₇	X ₄	X ₈₊₉	X ₅		
p(x)	0,22	0,19	0,15	0,14	0,12	0,10	0,08		

Informationstheorie

■ Bsp Huffman - 3

Code	X_1	X_2	X_3	X_{6+7}	X_4	X_{8+9}	X_5		
$p(x)$	0,22	0,19	0,15	0,14	0,12	0,10	0,08		

3. Schritt: Codewörter mit geringsten $p(x_i)$ mit 0/1 versehen und p's zusammenfassen und einsortieren

0: x_{8+9} , 1: x_5 , $p(x_{8+9})+p(x_5)=0,18$ damit: $X_8: 00$ $X_9: 01$

Code	X_1	X_2	X_{8+9+5}	X_3	X_{6+7}	X_4			
$p(x)$	0,22	0,19	0,18	0,15	0,14	0,12			

u.s.w.

Informationstheorie

■ Bsp Huffman - 4

gesamte Tabelle
mit Code-Generierung

wie sich zeigen läßt,
erhält man durch
schrittweises
Aufsummieren der
Wahrscheinlichkeiten der
letzten beiden Codeworte
automatisch die mittlere
Codewortlänge

1	2	3	4	5	6	7	8	9	0.1
0.22	0.19	0.15	0.12	0.08	0.07	0.07	0.06	0.04	
1	2	3	4	8	9	5	6	7	
				0	1				
0.22	0.19	0.15	0.12	0.1		0.08	0.07	0.07	0.14
1	2	3	6	7	4	8	9	5	
			0	1		0	1		
0.22	0.19	0.15	0.14		0.12	0.1		0.08	0.18
1	2	8	9	5	3	6	7	4	
		00	01	1		0	1		
0.22	0.19	0.18			0.15	0.14		0.12	0.26
6	7	4	1	2	8	9	5	3	
00	01	1			00	01	1		
0.26			0.22	0.19	0.18			0.15	0.33
8	9	5	3	6	7	4	1	2	
000	001	01	1	00	01	1			
0.33				0.26			0.22	0.19	0.41
1	2	8	9	5	3	6	7	4	
0	1	000	001	01	1	00	01	1	
0.41		0.33				0.26			0.59
8	9	5	3	6	7	4	1	2	
0000	0001	001	01	100	101	11	0	1	
0.59							0.41		1.0
8	9	5	3	6	7	4	1	2	3.01
00000	00001	0001	001	0100	0101	011	10	11	= L

Informationstheorie

■ Huffman-Codierung Zusammenfassung

- ◆ Huffman liefert mit einfachem, rekursiven Verfahren für N Codewörter in N Schritten einen redundanzminimierten Code

■ Huffman Code hat „Präfixeigenschaft“

- ◆ kein Codewort stellt den Beginn eines anderen Codewortes dar
 - auf Empfängerseite ist jedes Codewort eindeutig decodierbar
 - keine Pausezeichen zwischen den Codewörtern notwendig
- ◆ auch ASCII-Code hat Präfixeigenschaft
- ◆ Morse-Code nicht (benötigt Pausenzeichen)

Quellencodierung

■ Quellencodierung - forts.

◆ bisher besprochen

- ✦ verlustfreie Codierung der Quelldaten durch
 - geeignete Codewahl (z.B. Präfixcodierung)
 - Redundanzreduzierung

◆ bisher nicht berücksichtigt

- ✦ Eigenschaften der menschl. Sinnesorgane und mögliche Ansätze zur Datenreduktion

■ → verlustbehafteten Kompressionsverfahren