

Technische Informatik I

Vorlesung 10: Kodierungstheorie I

Joachim Schmidt
jschmidt@techfak.uni-bielefeld.de

Überblick

- Motivation: Wozu Kodierungstheorie?
- Begriffsdefinitionen
- Übertragung durch gestörte Kanäle
- Hamming-Distanz
- Parität, Parity Check, CRC
- Huffman-Kodierung
- Lempel-Ziv-Kodierung
- Shannon'sches Theorem
- Beispiele
- Ausblick

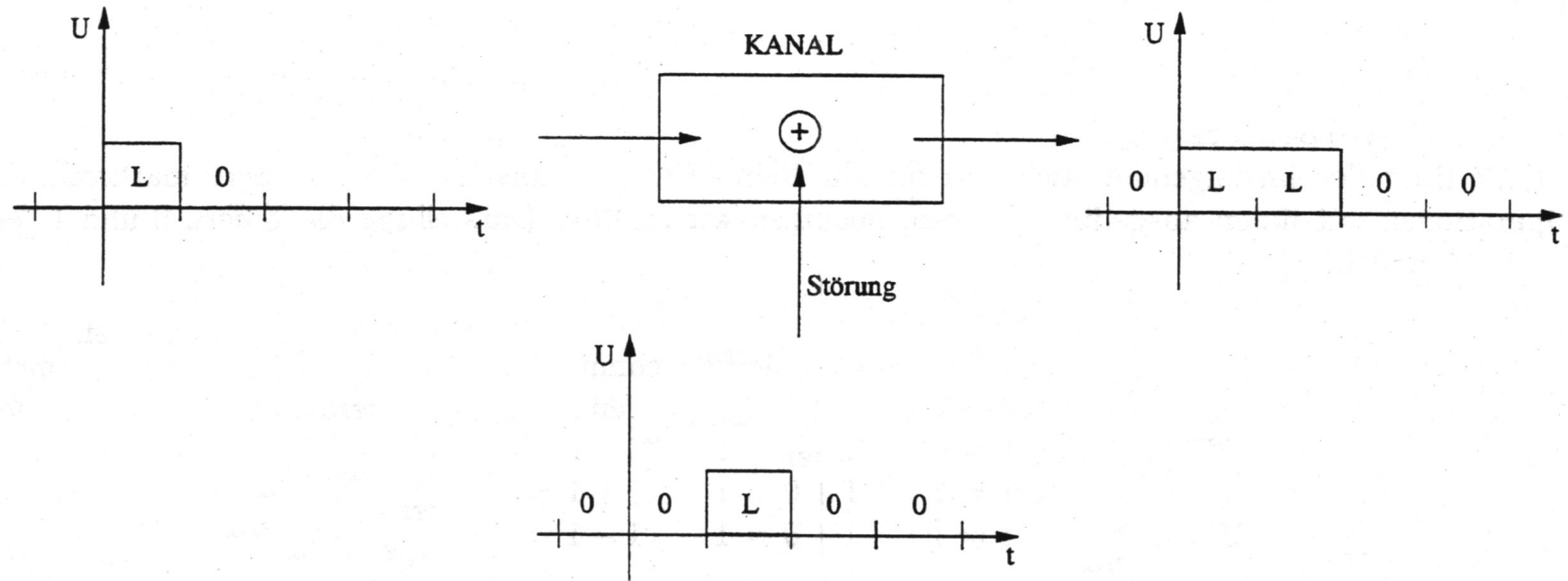
Wozu Kodierungstheorie?

- Wir hatten bereits:
Repräsentation von Informationen
- Gewünscht:
Möglichst günstige Repräsentation für:
 - Möglichst sparsame Darstellung zur Speicherung, nicht Verlustbehaftet
 - Sichere Übertragung von Informationen auf gestörten Kanälen
- Unabdingbares Instrument in der modernen Kommunikationstechnik

Begriffe

- Datenquelle:
 - Informationsliefernde Einheit / „Sender“
 - Mit (bekannten) statistischen Eigenschaften
- Kanal:
 - Übertragungsweg für Informationen
 - Digital, d.h. kleinste Informationseinheit = 1 Bit
 - Fehlerbehaftet → sog. Bitfehler
- Datensenke:
 - Informationsempfangende Einheit / „Empfänger“
 - Soll die Information genau so erhalten, wie sie gesendet wurde

Kanal mit Fehlereinstreuung

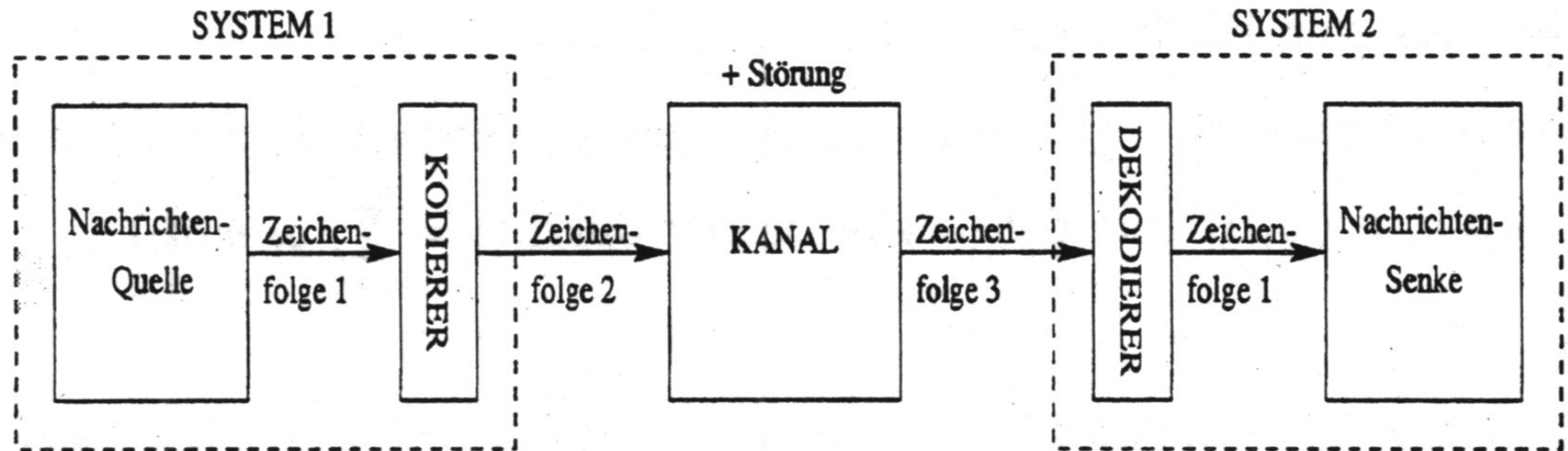


- Der Empfänger kann „echte“ Information und eingestreuten Fehler nicht unterscheiden

Begriffe

- Code:
 - Die darzustellende Information in einer bestimmten Repräsentation
 - Zeichenfolge
- Kodierung:
 - Übergang von einem Repräsentationssystem (mit gewissen Eigenschaften) in ein anderes (mit anderen Eigenschaften)

Sichere Übertragung



- Ist die Nachricht zu sehr gestört, kann es zu Fehldekodierungen kommen
- Wissen über die Fehleranfälligkeit des Kanals erforderlich
- Anpassung der eingesetzten Codierung an den jeweiligen Kanal

Redundanz

- Bezeichnung für das Vorhandensein von weglassbaren Elementen in einer Nachricht, die keine zusätzlichen Informationen liefern, sondern lediglich die beabsichtigte Grundinformation stützen.
- Fünf \leftrightarrow 5
- o i e u g s e o i e

Entropie

- Physik: Wärmelehre
- Größe des Nachrichtengehaltes einer nach statistischen Gesetzen gesteuerten Nachrichtenquelle
- Mittlerer Informationsgehalt der Zeichen eines bestimmten Zeichenvorrates
- Maß für den Grad der Ungewißheit über das Senden eines Zeichens aus einer Datenquelle

Definitionen

- Auftrittswahrscheinlichkeit p_a

Eines Zeichens a ist die Anzahl n der Auftritte des Zeichens in einer zufälligen Zeichenfolge der Länge N

$$p_a = \lim_{N \rightarrow \infty} \frac{n}{N} \text{ mit } N \rightarrow \infty$$

- Informationsgehalt I_a

Eines Zeichens a ist der Kehrwert seiner Auftrittswahrscheinlichkeit p_a

$$I_a = \text{ld} \frac{1}{p_a} = -\text{ld} p_a \text{ in bit}$$

Definitionen

- Entropie

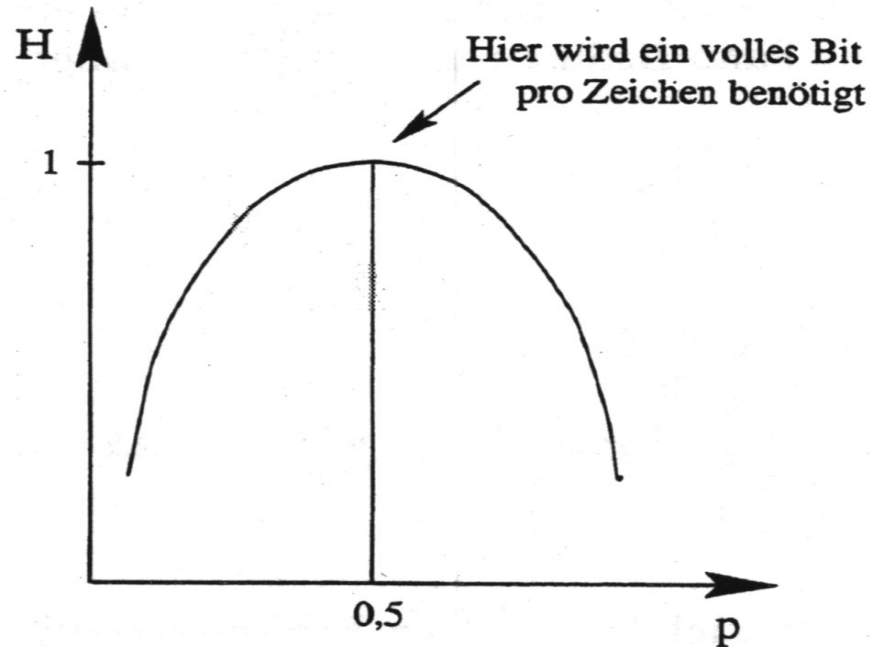
Einer Quelle ist der mittlere Informationsgehalt

$$H = E\{I_i\} = - \sum_{i=1}^M p_i \text{ld } p_i$$

Je höher die Entropie der Quelle, desto größer ist der Übertragungswert der von ihr ausgegebenen Nachrichten. Die Entropie erreicht ihr Maximum, wenn alle Zeichen gleich wahrscheinlich sind:

$$H_0 = I_i = \text{ld } M$$

Entropie einer Einzelbitquelle



- p : Wahrscheinlichkeit für das Auftreten von 1
- H : Entropie der Quelle
- Hohe Entropie \rightarrow geringes Wissen über nächstes Zeichen

Hammingabstand

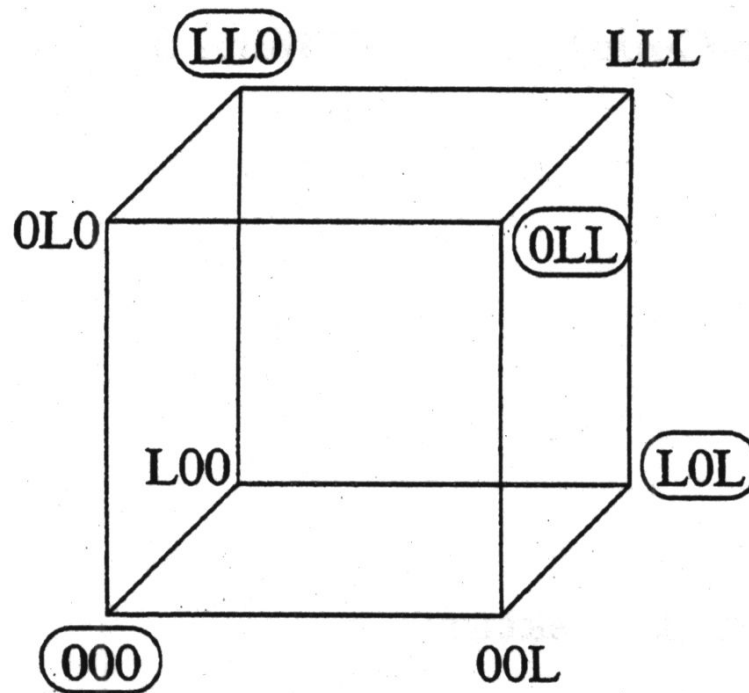
- Anzahl der (Bit-)Stellen, in denen sich zwei Wörter voneinander unterscheiden
- Direkter Vergleich, keine Einfügungen oder Löschungen erlaubt


$$h : \mathbf{2}^n \times \mathbf{2}^n \rightarrow N \text{ mit } h(\langle a_1 \dots a_n \rangle, \langle b_1 \dots b_n \rangle)$$

$$= \sum_{i=1}^n d_i, \text{ mit } d_i = \begin{cases} 1 & \text{falls } a_i \neq b_i \\ 0 & \text{sonst} \end{cases}$$

- h ('Kanne', 'Tanne') = 1
- h ('11001001', '01010001') = 3

Codewürfel



 = wird benutzt

Codewürfel

- Nicht der gesamte zur Verfügung stehende Wortraum wird genutzt
- Tritt ein beliebiger Einzelbitfehler auf entsteht ein „Illegales Codewort“
- Einzelbitfehler werden erkannt
- Stelle des Fehlers nicht erkennbar
- Daher keine Korrektur möglich

Parität (Parity)

- Sicherung der Nachricht durch gezieltes Einfügen zusätzlicher Redundanz
- Unterscheidung in Nutzstellen und Prüfstellen
- Mögliche Vorschrift:

Ergänze Wort so, daß es eine gerade Anzahl Einsen enthält. (Even Parity)

a	$C_P(a)$	Paritätsbit
0	00	0
1	0L	L
2	L0	L
3	LL	0

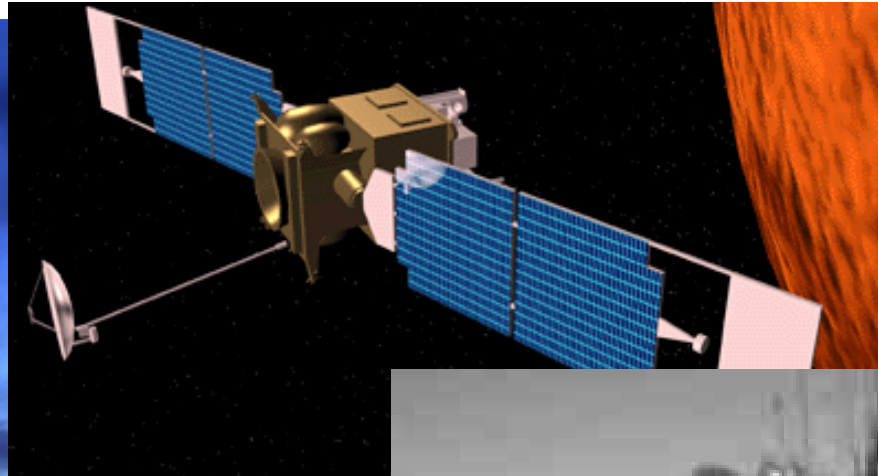
Prüfziffern

- Verschiedenste Berechnungsvorschriften
- Ergänzung bestehender Ziffernfolgen
- Dadurch Sicherung gegen Fehler
- Keine Korrektur möglich
- Beispiele:
 - ISBN-Nummern (Bücher)
 - EAN 13-Code (Lebensmittel)

ISBN 3-928444-04-2



Beispiele - NASA



- Mariner, Voyager
- Deep Space Network
- Pathfinder
- Global Surveyor

Beispiele – Audio CD



- Zwei „verschachtelte“ Codes
- Bis zu 4000 aufeinanderfolgende Fehler korrigierbar (ca 2.5 mm Spurlänge)
- Audio CD-Spieler können durch Interpolation des Signals noch stärkere Fehler umgehen

Weitere Beispiele

- Netzwerke (Ethernet, WLAN)
- Mobilfunk (Handys)
- Digitales Fernsehen (DVB)
- Digitales Radio (DAB)

Nächste Woche

- Kodierungstheorie II:
 - Huffman-Kodierung
 - Lempel-Ziv-Kodierung
 - Shannon'sches Theorem