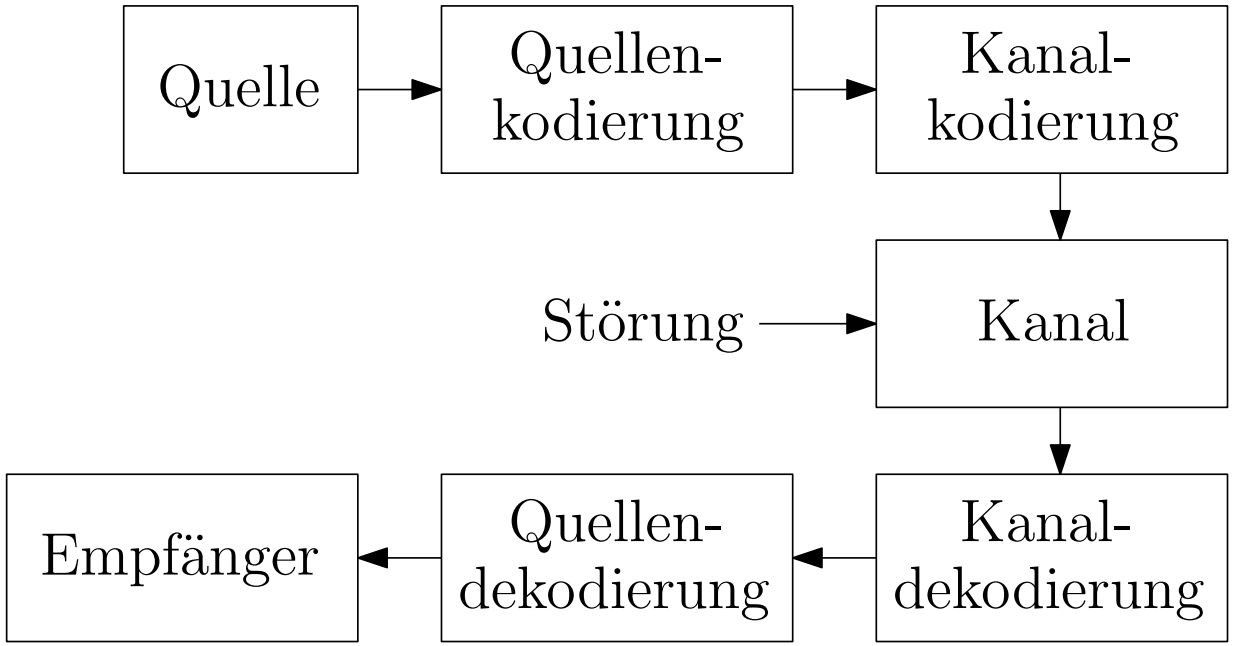
Karlsruher Institut für Technologie

Theoretische Grundlagen der Informatik
Vorlesung am 4. Februar 2020

INSTITUT FÜR THEORETISCHE INFORMATIK

Kodierung zum Schutz gegen Übertragungsfehler



Kodierung zum Schutz gegen Übertragungsfehler

- Qualität einer digitalen Übertragung wird häufig als gemessene Bitfehlerquote bzw. Bitfehlerwahrscheinlichkeit angegeben.
- Beherrschung von Übertragungsfehlern:
  - Fehlerkorrektur (beim Empfänger)
  - Fehlererkennung und Wiederholungsanforderung
- Trade-off: Wahrscheinlichkeit unentdeckter Fehler vs. Datendurchsatz

Paritätscodes – Einfach binär

Quelle: Wikipedia

- Paritätscode der RS232-Schnittstelle
- Neunpoliges Kabel ermöglicht die parallele Übertragung von 8 Bits.
- Dabei werden nur sieben Bits $b_1, \ldots, b_7$ für die Nachrichtenübertragung genutzt.
- Das achte Bit b_8 wird Paritätsbit genannt.

Paritätscodes – Einfach binär

- Paritätscode der RS232-Schnittstelle
- Neunpoliges Kabel ermöglicht die parallele Übertragung von 8 Bits.
- Dabei werden nur sieben Bits $b_1, \ldots, b_7$ für die Nachrichtenübertragung genutzt.
- Das achte Bit b_8 wird Paritätsbit genannt.

Wiederholung XOR-Verknüpfung $\oplus$

$\oplus$	0	1
0	0	1
1	1	0

- Es wird b_8 so gesendet, dass

$$b_8 = b_1 \oplus b_2 \oplus b_3 \oplus b_4 \oplus b_5 \oplus b_6 \oplus b_7.$$

Paritätscodes – Einfach binär

Wiederholung XOR-Verknüpfung $\oplus$

$\oplus$	0	1
0	0	1
1	1	0

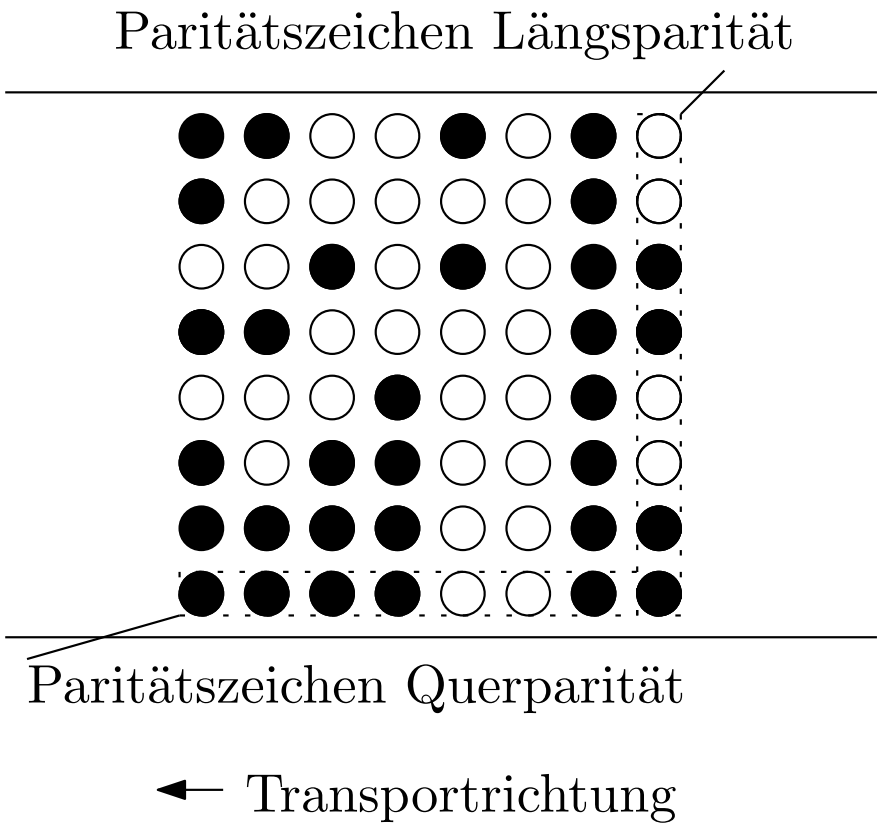
- Es wird b_8 so gesendet, dass

$$b_8 = b_1 \oplus b_2 \oplus b_3 \oplus b_4 \oplus b_5 \oplus b_6 \oplus b_7 .$$

- Mit dem Paritätscode werden einfache Fehler im Codewort erkannt.
- Gleichzeitige Übertragung von 2 Fehlern wird nicht erkannt.
- Falls ein Fehler erkannt wird, kann die ursprüngliche Nachricht nicht rekonstruiert werden: Die Fehlerstelle ist unbekannt.

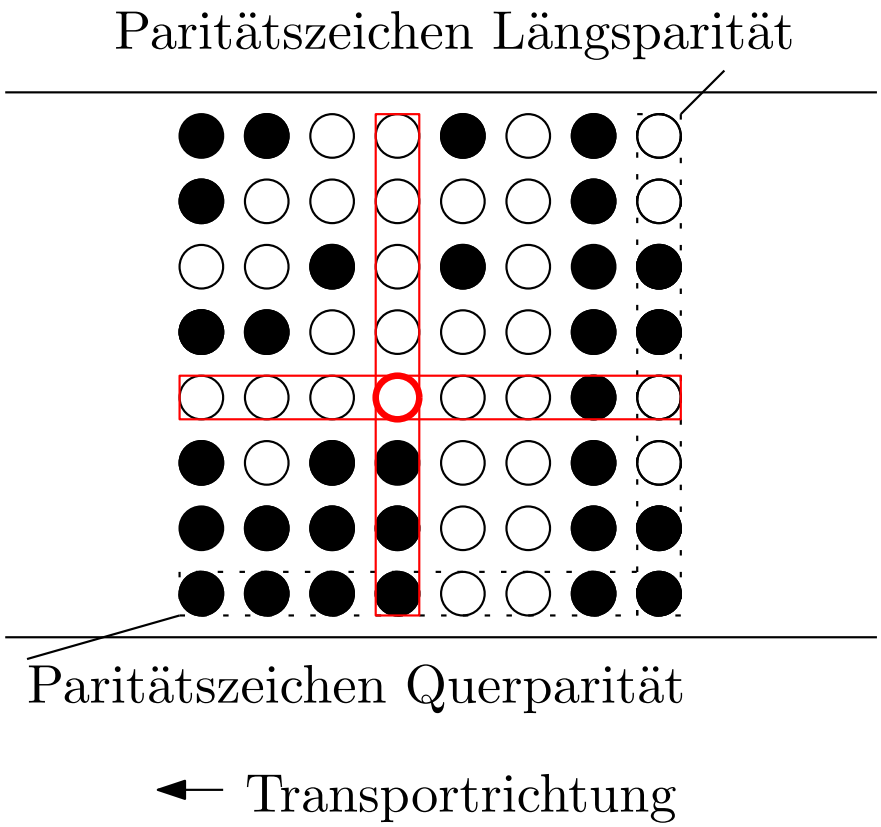
Kreuzsicherung

- Dient zum Schutz gegen Doppelfehler
- Erklärung am Beispiel Lochkarte



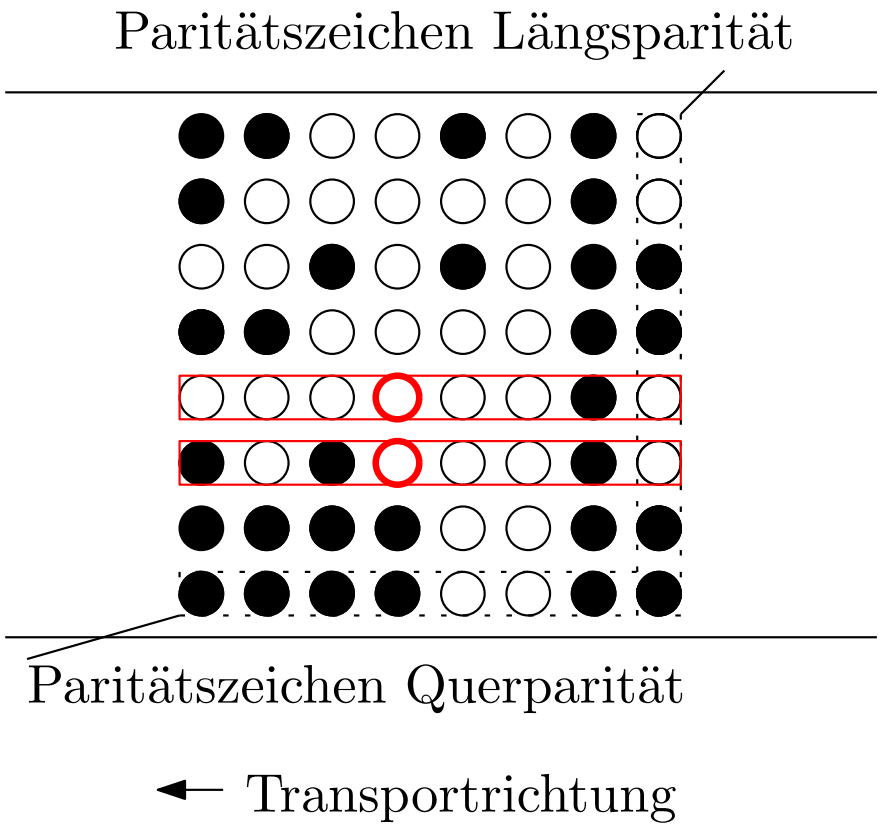
Kreuzsicherung

- Alle 1,2,3-fachen Fehler sind erkennbar
- Ab 4 Fehlern nicht zwingend erkennbar



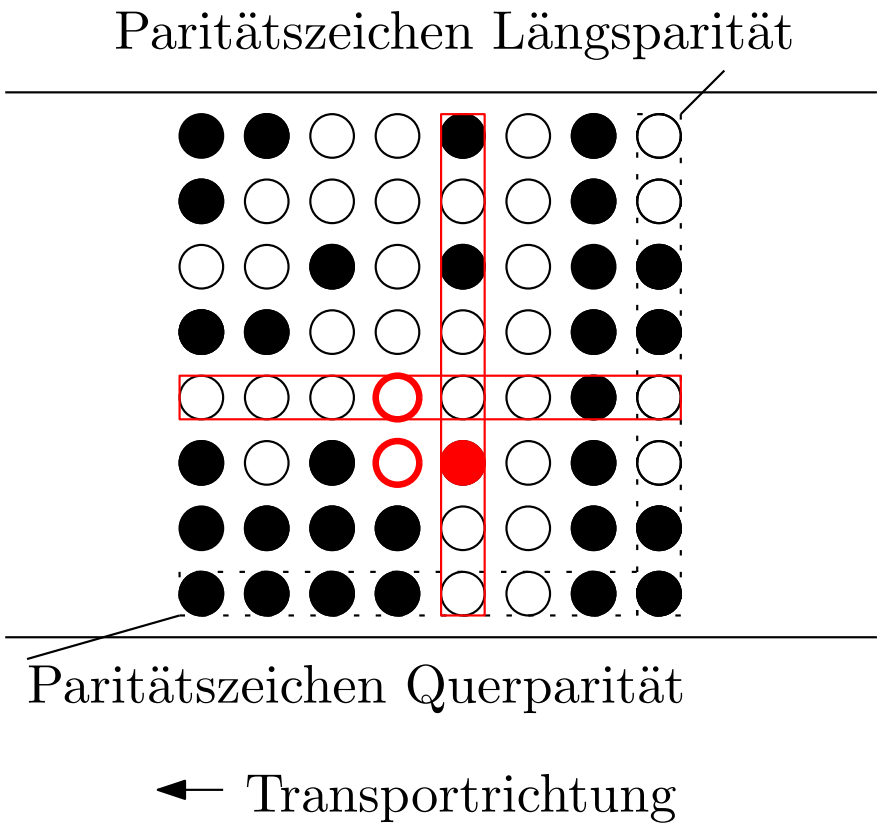
Kreuzsicherung

- Alle 1,2,3-fachen Fehler sind erkennbar
- Ab 4 Fehlern nicht zwingend erkennbar



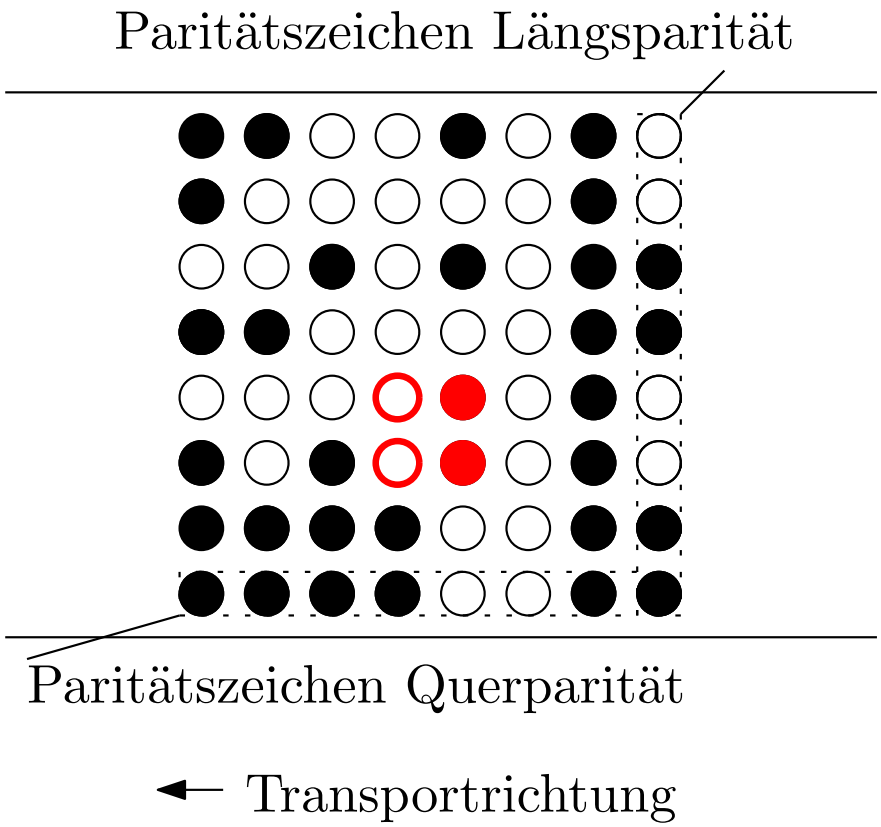
Kreuzsicherung

- Alle 1,2,3-fachen Fehler sind erkennbar
- Ab 4 Fehlern nicht zwingend erkennbar



Kreuzsicherung

- Alle 1,2,3-fachen Fehler sind erkennbar
- Ab 4 Fehlern nicht zwingend erkennbar



Paritätscodes

Paritätscodes

- Gegeben ein Alphabet $\Sigma = \{1, 2, \ldots, q-1\}$.
- Ein Code der Länge n zur Basis q sei eine Menge von Folgen mit Elementen aus Σ.
- Einzelne Folgen werden Codewörter genannt.

Ein Paritätscode liegt vor, wenn für jedes Codewort $a_1 a_2 \cdots a_n$

$$(a_1 + a_2 + \cdots + a_n) \mod q = 0$$

gilt.

Satz:
Jeder Paritätscode erkennt Einzelfehler.

Beweis

- Sei $a_1 \cdots a_n$ das ursprüngliche Codewort.
- Annahme: Das i-te Element wurde fehlerhaft als $\widetilde{a}_i$ übertragen.
- Ein Übertragungsfehler liegt vor, wenn

$$(a_1 + a_2 + \cdots + \widetilde{a}_i + \cdots + a_n) \mod q = 0.$$

- Für das ursprüngliche Codewort gilt:

$$(a_1 + a_2 + \cdots + a_n) \mod q = 0$$

Also:

$$\begin{aligned} 0 &= (a_1 + a_2 + \cdots + \widetilde{a}_i + \cdots + a_n) \mod q \\ &= (a_1 + a_2 + \cdots + a_n) \mod q \end{aligned}$$

Beweis

$$\begin{aligned} 0 &= (a_1 + a_2 + \cdots + \widetilde{a}_i + \cdots + a_n) \mod q \\ &= (a_1 + a_2 + \cdots + a_n) \mod q \end{aligned}$$

Damit:

$$\begin{aligned} 0 &= (a_1 + a_2 + \cdots + \widetilde{a}_i + \cdots + a_n) \mod q - \\ &\quad (a_1 + a_2 + \cdots + a_n) \mod q \\ &= (\widetilde{a}_i - a_i) \mod q \end{aligned}$$

Um dies zu erfüllen, muss $(\widetilde{a}_i - a_i)$ durch q teilbar sein. Weil aber

$$0 \leq \widetilde{a}_i, a_i < q,$$

folgt

$$0 \leq |\widetilde{a}_i - a_i| < q\,.$$

Paritätscodes gegen Vertauschungsfehler

- Häufige Fehlerart bei manueller Eingabe: Vertauschungsfehler
- Diese werden von gewöhnlichen Paritätscodes nicht erkannt.
- Sei wieder $a_1 \cdots a_n$ ein Codewort.
- **Paritätscode mit Gewichten:** Wir führen zusätzlich ganzzahlige Gewichte $w_1, \ldots, w_{n-1}$ ein, sodass

$$(w_1 \cdot a_1 + w_2 \cdot a_2 + \cdots + w_{n-1} \cdot a_{n-1} + a_n) \mod q = 0$$

  gilt.
- Zusatzbedingung: Alle Gewichte w_i müssen teilerfremd zu q sein.

Paritätscodes gegen Vertauschungsfehler

- **Paritätscode mit Gewichten:** Wir führen zusätzlich ganzzahlige Gewichte $w_1, \ldots, w_{n-1}$ ein, sodass

$$(w_1 \cdot a_1 + w_2 \cdot a_2 + \cdots + w_{n-1} \cdot a_{n-1} + a_n) \mod q = 0$$

  gilt.
- Zusatzbedingung: Alle Gewichte w_i müssen teilerfremd zu q sein.

Satz:
Jeder Paritätscode mit Gewichten erkennt Einzelfehler.

Beweis: Analog zu normalen Paritätscodes kann gezeigt werden, dass für jeden Einzelfehler $a_i \to \widetilde{a}_i$ gilt:

$$w_i \cdot (\widetilde{a}_i - a_i) \mod q = 0$$

Paritätscodes gegen Vertauschungsfehler

- **Paritätscode mit Gewichten:** Wir führen zusätzlich ganzzahlige Gewichte $w_1, \ldots, w_{n-1}$ ein, sodass

$$(w_1 \cdot a_1 + w_2 \cdot a_2 + \cdots + w_{n-1} \cdot a_{n-1} + a_n) \mod q = 0$$

  gilt.
- Zusatzbedingung: Alle Gewichte w_i müssen teilerfremd zu q sein.

Satz:
Ein Paritätscode mit Gewichten erkennt die Vertauschung an den Stellen i und j, falls die Zahl $w_i - w_j$ teilerfremd zu q ist.

Beweis: Analog zu oben: Vertauschungsfehler wird nicht erkannt, falls

$$[(w_i a_i + w_j a_j) - (w_j a_i + w_i a_j)] \mod q = [(w_i - w_j)(a_i - a_j)] \mod q = 0\,.$$

Bsp: ISBN-10

Beschreibung ISBN-10 (oben)

- Alphabet $\Sigma = \{0, 1, \ldots, 9\}$
- Basis $q = 11$ (Primzahl!)
- Länge $n = 10$
- Paritätscode
- Für Code $a_1 \cdots a_{10}$ berechnet sich die Prüfziffer a_{10} aus

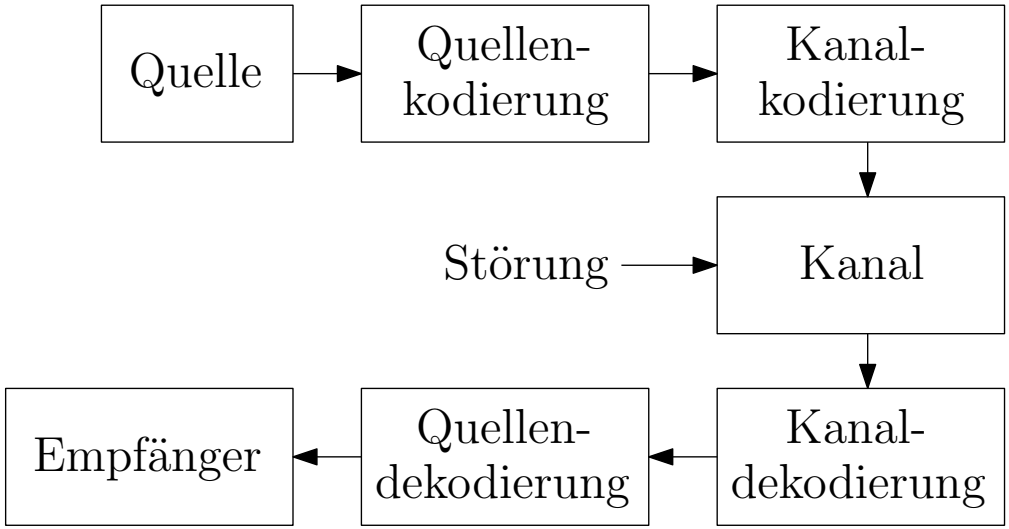
$$(10a_1 + 9a_2 + 8a_3 + \cdots + 2a_9 + a_{10}) \mod 11 = 0 .$$

- $10a_1 + \cdots + 2a_9 = 275 = 11 \cdot 25$
  $\Rightarrow a_{10} = 0$

Bsp: ISBN-10

- ISBN: International Standard Book Number (ISO Standard 2108)
- ISBN-10 war bis 2006 übliche Kodierung
- Seit 2007 EAN-13 (European Article Number)

Block-Codes



- Man unterscheidet verschiedene Arten von Kanal-Codes.
- **Block-Codes:** Hier betrachtet man Codeworte fester Länge. Aufeinanderfolgende Blöcke werden unabhängig voneinander kodiert.
- **Faltungs-Codes:** Codeworte können beliebig lang sein. Die Zeichen sind vom Vorgeschehen abhängig.
- In TGI befassen wir uns ausschließlich mit Block-Codes.

Hamming-Distanz und Fehlerkorrektur

Hamming-Distanz
Für $x, y \in \{0,1\}^n$ ist

$$d(x,y) := \#\{i \mid i = 1,\ldots,n, x_i \neq y_i\}$$

die Hamming-Distanz zwischen x und y.

Anschaulich: Die Hamming-Distanz zwischen x und y ist die Anzahl der Zeichen in x, die sich von denen in y unterscheiden.

Es sei $B_r(x)$ die Menge aller Worte y mit $d(x,y) \leq r$.

Anschaulich: B_r ist eine Kugel (die Hamming-Kugel) um x mit Radius r.

Hamming-Distanz und Fehlerkorrektur

Hamming-Distanz
Für $x, y \in \{0,1\}^n$ ist

$$d(x,y) := \#\{i \mid i = 1,\ldots,n, x_i \neq y_i\}$$

die Hamming-Distanz zwischen x und y.

Anschaulich: Die Hamming-Distanz zwischen x und y ist die Anzahl der Zeichen in x, die sich von denen in y unterscheiden.

Maximum-Likelihood-Decoding
Sei eine Kodierung C gegeben und y ein empfangenes Wort. Dekodiere y als dasjenige Codewort $x \in C$, für das $d(x,y)$ minimal wird.

Block-Codes

Block-Code

- Gegeben ist ein endliches Alphabet Σ.
- Ein Block-Code ist eine Teilmenge $C \subseteq \Sigma^n$ für ein $n \in \mathbb{N}$.
- Falls $\#C = 1$, so heißt C trivial, da es nur ein Codewort gibt.

Minimaldistanz

Die Minimaldistanz eines nichttrivialen Block-Codes C ist

$$m(C) := \min_{c_1, c_2 \in C, c_1 \neq c_2} d(c_1, c_2) \ .$$

Beispiel

Zeichen	Codeworte C
A	00000
H	10011
L	11100
O	01111

Maximum-Likelihood-Decoding

- Empfänger dekodiert 00010 als A, weil $d(00010, 00000) = 1$.
- Minimaldistanz ist $m(C) = \min\{3, 3, 4, 4, 3, 3\} = 3$.
- Es werden 2 Fehler erkannt.
- Es kann 1 Fehler korrigiert werden.

Ziel:

- Finde Codeworte C in $\{0, 1\}^n$ mit $m(C)$ groß, also paarweise großer Hamming-Distanz.

Block-Codes

Satz:
Ein Block-Code C mit Minimaldistanz $m(C) = d$ kann bis zu $d - 1$ Fehler erkennen und bis zu $\left\lfloor \frac{d-1}{2} \right\rfloor$ Fehler korrigieren.

Beweisskizze