Information Theory

Lecture 9
Channel Coding

STEFAN HÖST

Why is error correction possible?

Example

"Educafion is wzat rempins aftqr onx hay porgotten uhat kne has lehrned in sctool."

Albert Einstein

- Despite of 11 erroneous letters the text in this example can still be well understood
- Language contains redundancy which allows the reader to decode the errors
- Shannon showed that English prose has a redundancy of more than 50%

Parity check bit

Detection of an error can be achieved by simple addition of a parity bit $x_n = \sum_{i=1}^{n-1} x_i \mod 2$

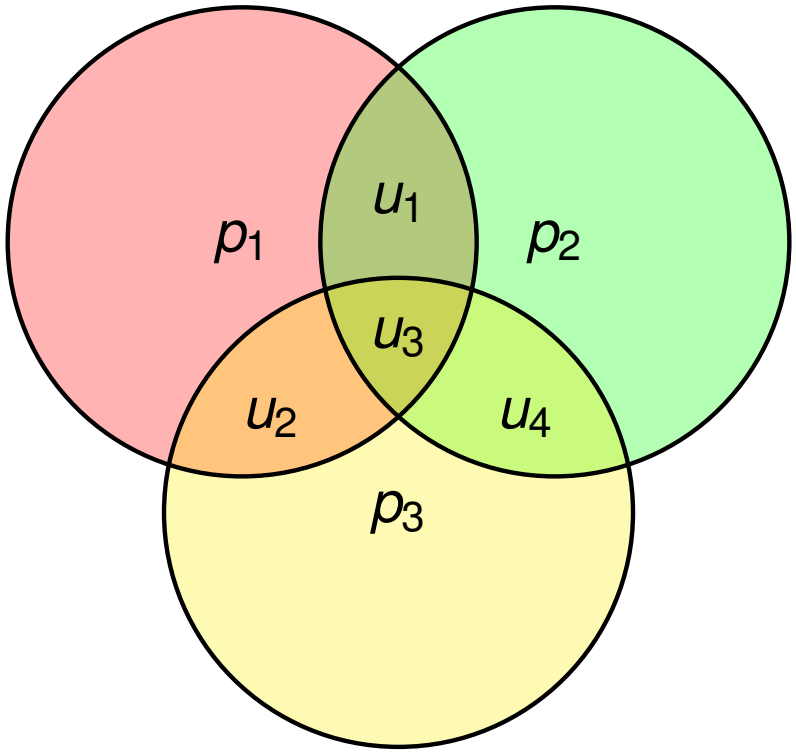
$$1\,0\,1\,0\,0\,1\,0 \quad \rightarrow \quad 1\,0\,1\,0\,0\,1\,0\,1$$

A single error can be detected by $\sum_{i=1}^{n} x_i \neq 0 \mod 2$.

Hamming, 1947

> *"Damn it, if the machine can detect an error, why can't it locate the position of the error and correct it?"*

Hamming's idea: Use several parity-checks



$$
\begin{aligned}
p_1 &= u_1 + u_2 + u_3 \\
p_2 &= u_1 \quad\quad + u_3 + u_4 \\
p_3 &= \quad\quad u_2 + u_3 + u_4
\end{aligned}
$$

Hamming codes are able to correct arbitrary single errors.

An experiment

1. **Source:** Pick one of 16 messages
2. **Encoder:** Select codeword from table
3. **Channel:** Flip one arbitray symbol
4. **Decoder:** Choose closest codeword and outputs corresponding message

Assume that the following vector is received:

1001101

Which message was transmitted?

message	codeword
0000	0000000
0001	1110001
0010	1100010
0011	0010011
0100	1010100
0101	0100101
0110	0110110
0111	1000111
1000	0111000
1001	**1001001**
1010	1011010
1011	0101011
1100	1101100
1101	0011101
1110	0001110
1111	1111111

Channel Codes

Definition

Code An (M, n) binary code consists of M codewords of length n. Its rate is

$$R = \frac{\log M}{n} = \frac{k}{n}$$

where k is the information word length.

Code example

The previous Hamming code with $M = 16$ codewords of length $n = 7$ is a $(16, 7)$ code. Its rate is $R = \frac{4}{7}$.

Linear codes

Definition

A code $\mathcal{B}$ is linear if for every pair of codewords, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, the sum (or difference) is also a codeword,

$$\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{B} \Rightarrow \boldsymbol{x}_i + \boldsymbol{x}_j \in \mathcal{B} \mod 2$$

All-zero codeword

From the definition we see that the all-zero vector is a codeword, $\mathbf{0} \in \mathcal{B}$, in a linear code.

Linear codes

Linear code space

A binary linear code $\mathcal{B}$, with rate $R = \frac{k}{n}$, is a k-dimensional subspace of $\mathbb{F}_2^n$. A codeword is a linear combination of k linearly independendent codewords, $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k$, where $\boldsymbol{g}_i \in \mathcal{B}$.
Encoding:

$$\boldsymbol{x} = u_1\boldsymbol{g}_1 + \cdots + u_k\boldsymbol{g}_k = (u_1 \ldots u_k) \begin{pmatrix} \boldsymbol{g}_1 \\ \vdots \\ \boldsymbol{g}_k \end{pmatrix}$$

$$= (u_1 \ldots u_k) \begin{pmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & & \vdots \\ g_{k1} & \cdots & g_{kn} \end{pmatrix} = \boldsymbol{u}G$$

where $\boldsymbol{u}$ is the information word and G the generator matrix

Generator matrix

A generator matrix for the $(M, n) = (16, 7)$ Hamming code is

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Notice the mapping is not the same as before.

message	codeword
0000	0000000
0001	0001110
0010	0010011
0011	0011101
0100	0100101
0101	0101011
0110	0110110
0111	0111000
1000	1000111
1001	1001001
1010	1010100
1011	1011010
1100	1100010
1101	1101100
1110	1110001
1111	1111111

LUND UNIVERSITY

Hamming distance

Definition

The Hamming distance between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, $d_H(\boldsymbol{x}, \boldsymbol{y})$, is the number of positions in which they differ.

The Hamming weight of a vector $\boldsymbol{x}$, $w_H(\boldsymbol{x})$, is the number of non-zero positions.

For binary vectors

$$d_H(\boldsymbol{x}, \boldsymbol{y}) = w_H(\boldsymbol{x} - \boldsymbol{y}) = w_H(\boldsymbol{x} + \boldsymbol{y})$$

Decoding

Decoding criteria

To minimise the probability of error

- Maximum a posteriori (MAP) decoder

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}\in\mathcal{B}} \{ P(\boldsymbol{x}|\boldsymbol{y}) \}$$

- Equivalently, for equally likely codewords,
  Maximum likelihood (ML) decoder

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}\in\mathcal{B}} \{ P(\boldsymbol{y}|\boldsymbol{x}) \}$$

- Equivalently, for BSC,
  Minimum distance (MD) decoder

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathcal{B}} \{ d_H(\boldsymbol{y}, \boldsymbol{x}) \}$$

Minimum distance

Definition

The minimum distance for a code is the minimum Hamming distance between two different codewords,

$$d_{\min} = \min_{\substack{\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{B} \\ \boldsymbol{x}_1 \neq \boldsymbol{x}_2}} \left\{ d_H(\boldsymbol{x}_1, \boldsymbol{x}_2) \right\}$$

For a linear code the minimum distance can be derived as

$$d_{\min} = \min_{\substack{\boldsymbol{x} \in \mathcal{B} \\ \boldsymbol{x} \neq \mathbf{0}}} \left\{ w_H(\boldsymbol{x}) \right\}$$

Minimum distance

codeword	w_H
0000000	0
0001110	3
0010011	3
0011101	4
0100101	3
0101011	4
0110110	4
0111000	3
1000111	4
1001001	3
1010100	3
1011010	4
1100010	3
1101100	4
1110001	4
1111111	7

For the $(16, 7)$ Hamming code

$$d_{\min} = \min_{\boldsymbol{y} \in \mathcal{B} \setminus \mathbf{0}} \{ w_H(\boldsymbol{y}) \}$$

Error detection and correction

Theorem

- *It is always possible to detect an error* $\boldsymbol{e}$ *if*

$$w_H(\boldsymbol{e}) \leq d_{\min} - 1$$

- *It is always possible to correct an error* $\boldsymbol{e}$ *if*

$$w_H(\boldsymbol{e}) \leq \frac{d_{\min} - 1}{2}$$

Parity check matrix

Definition

The generator matrix G spans a k-dimensional subspace of $\mathbb{F}_2^n$.
Let the parity check matrix H be defined by its null-space,

$$GH^T = 0$$

Then $\boldsymbol{x}$ is a codeword in $\mathcal{B}$ if and only if $\boldsymbol{x}H^T = 0$.

Example (Hamming code)

Use all non-zero vectors of length $7 - 4 = 3$

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} I & P^T \end{pmatrix}$$

Hamming code

Example (Hamming code)

The code

$$\mathcal{B}_H = \{0000000, 0001110, 0010011, 0011101, 0100101, 0101011, 0110110, 0111000, 1000111, 1001001, 1010100, 1011010, 1100010, 1101100, 1110001, 1111111\}$$

is an $(M, n) = (16, 7)$ code with rate $R = \frac{4}{7}$ and $d_{\min} = 3$

Hamming code

Example

A generator matrix for the Hamming code

$$G = \begin{pmatrix} P & I \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

where P is the first three columns of G and I the unity matrix of size 4.

Hamming code

Example

The mapping between information words and codewords obtained by G gives:

u	x	u	x
0000	0000000	1000	0111000
0001	1110001	1001	1001001
0010	1100010	1010	1011010
0011	0010011	1011	0101011
0100	1010100	1100	1101100
0101	0100101	1101	0011101
0110	0110110	1110	0001110
0111	1000111	1111	1111111

Syndrome decoder

Definition

Let $\boldsymbol{e}$ be an error pattern on a BSC. Then the received vector is $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}$. The syndrom is $\boldsymbol{s} = \boldsymbol{y}H^T = \boldsymbol{x}H^T + \boldsymbol{e}H^T = \boldsymbol{e}H^T$

Syndrome decoding (MD)

- Form a list with the most probable (least Hamming weight) error patterns for each possible syndrome.
- Derive the syndrome of the received vector and use the list to estimate the error pattern, $\hat{\boldsymbol{e}}$.
- Estimate the transmitted codeword as $\hat{\boldsymbol{x}} = \boldsymbol{y} + \hat{\boldsymbol{e}}$.
- Derive the estimated information word $\hat{\boldsymbol{x}} \rightarrow \hat{\boldsymbol{u}}$

Hamming code

Example

Syndrome list for (16,7) Hamming code

$\boldsymbol{e}$	$\boldsymbol{s} = \boldsymbol{e}H^T$	$\boldsymbol{e}$	$\boldsymbol{s} = \boldsymbol{e}H^T$
0000000	000	0001000	011
1000000	100	0000100	101
0100000	010	0000010	110
0010000	001	0000001	111

Back to our experiment:

Assuming $\mathbf{y} = (1001101)$, which message was transmitted?

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$$

We compute the syndrome

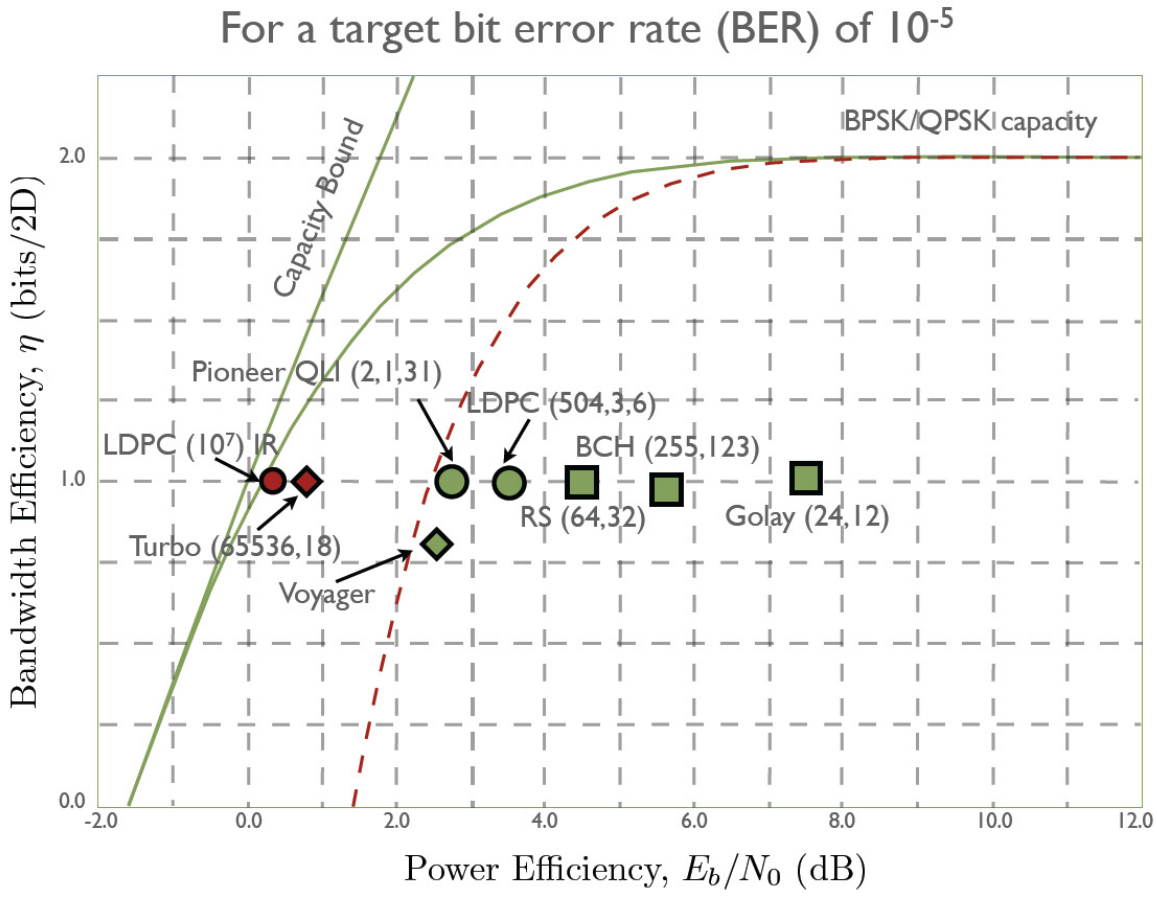
$$\mathbf{s} = (1001101) \cdot \mathbf{H}^T = (101)$$

and conclude that $\hat{\mathbf{e}} = (0000100)$. Then

$$\hat{\mathbf{x}} = \mathbf{y} - \hat{\mathbf{e}} = \mathbf{y} + \hat{\mathbf{e}} = (1001001)$$

such that $\hat{\mathbf{u}} = (1001)$.

Capacity, the coding theory challenge



Source: D.J. Costello, Jr., "Modern Coding Theory", Lecture at the Third Canadian Summer School on Communications and Information Theory, Banff, Alberta, Canada, August 19, 2008

Application examples

Coding for audio compact disc (CD):

- protection against dirt and scratches on surface
- Reed-Solomon codes are used for error detection and error correction

Coding in mobile communications:

- fading leads to burst errors of several hundreds of bits
- GSM system uses convolutional codes with interleaving
- UMTS / LTE: more modern turbo codes with iterative decoding
- WiFi (802.11n/ac/ax): low-density parity-check (LDPC) codes

Coding in deep space communications:

- Voyager mission to Jupiter, Saturn, Uranus, and Neptune
- concatenation of Reed-Solomon codes and convolutional codes
- Modern CCSDS standard: LDPC codes (Consultative Committee for Space Data Systems)

Application examples

2D Barcodes:

- Quick response codes (QR codes) use Reed-Solomon error correction
- different levels of error protection are possible

Data centers: (Google, DropBox, etc.)

- storage of massive data is currently a very hot topic
- efficient handling of failures is required
  $\rightarrow$ on the fly replacement of damaged hard disks
- how can the data be efficiently reproduced?
  $\rightarrow$ error control coding over several disks