**Institut für Nachrichtentechnik**
**Prof. Dr.-Ing. Gerhard Bauch**

Technische Universität Hamburg-Harburg

Exam WS 2015/2016

# Information Theory and Coding

| Name: ………………………………………………………….. | Immatriculation number: …………………………… | |
|---|---|---|
| | Points | of |
| Problem 1 | | |
| Problem 2 | | |
| Total points | | |
| Grade | | |

- The following aids are allowed in this exam:
  - 2 Din A4 sheets, handwritten on both sides (4 pages in total)
  - calculator (non-programmable)
  - Pens
- Other aids are not allowed.
- Please use a separate solution sheet for each task.
- Write your name and matriculation number on each solution sheet.
- An arrow next to a question means that this part of the task can be solved independently of the previous questions.
- Please do not write with pencils and do not use a red pen.
- The duration of the exam is 90 minutes.
- The exam consists of **5** pages (including this cover page).
- Switch off your cell phones!

Good luck!

**Problem 1:** Huffman Code and Relative Entropy

Consider a 4-ary random variable $X$ with realizations $x \in \{a, b, c, d\}$. The true probability distribution $p_X(x)$ is given in the following table:

| $x$ | $p_X(x)$ |
|---|---|
| $a$ | $1/2$ |
| $b$ | $1/4$ |
| $c$ | $1/8$ |
| $d$ | $1/8$ |

⇨ a) Determine the entropy $H(X)$ of the random variable $X$.

⇨ b) Construct a Huffman-Code for binary encoding of the realizations $x$.
Make sure that all steps in the construction of the Huffman code are clearly given.
Represent the mapping of realizations $x$ to codewords $\mathbf{c}$ of the Huffman code in a table.

c) Determine the average codeword length of the Huffman code from b).

Construction of a Huffman code requires knowledge of the statistics of the random variable $X$. Often, the probability distribution $p_X(x)$ of the random variable $X$ is not exactly known but has to be estimated. Assume, that a probability distribution $q_X(x)$ according to the following table has been estimated instead of the true probability distribution $p_X(x)$:

| $x$ | $q_X(x)$ |
|---|---|
| $a$ | $1/8$ |
| $b$ | $1/8$ |
| $c$ | $1/4$ |
| $d$ | $1/2$ |

⇨ d) Construct a Huffman code for $X$ based on the estimated probabilities $q_X(x)$.
Represent the mapping of realizations $x$ to codewords $\mathbf{w}$ of the Huffman code in a table.

e) Determine the average codeword length of the Huffman code from d) and compare to your result from c).

The distance of two probability distributions $p_X(x)$ and $q_X(x)$ can be measured by the *relative entropy*

$$D(p\|q) = \sum_x p_X(x) \log_2 \frac{p_X(x)}{q_X(x)},$$

which is also called the *Kullback Leibler distance.*

f) Determine the relative entropy $D(p||q)$ of the probability distributions $p_X(x)$ and $q_X(x)$.

g) Interpret the meaning of the relative entropy for compression in view of your results for problems a)-e).

h) Derive the relation between $p_X(x)$ and $q_X(x)$ for which the relative entropy is zero, i.e. $D(p||q) = 0$.

i) Is the relative entropy in general a true distance in the sense that it is symmetric, i.e. does $D(p||q) = D(q||p)$ hold for arbitrary probability distributions $p_X(x)$ and $q_X(x)$ ?

j) Show, that the mutual information $I(X, Y)$ between two random variables $X$ and $Y$ is the relative entropy between the joint probability distribution $p_{X,Y}(x,y)$ and the product probability distribution $p_X(x)p_Y(y)$.

## **Problem 2:** Shortening of Linear Block Codes

Sometimes, it is not possible to find a code of suitable codeword length or information word length for a given application. Therefore, it is desirable to construct a suitable code by shortening or extension of a known code. We will investigate the strategy of code shortening for the example of a $(7, 4, 3)_2$ Hamming code. The parity check matrix **H** of the Hamming code is given by

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

⇨ a) Determine the codeword length $N$ of the Hamming code.

⇨ b) Determine the code rate $R$ of the Hamming code.

⇨ c) Determine the minimum Hamming distance $d_{\min}$ of the Hamming code.

⇨ d) Determine the number of parity checks which are defined in the parity check matrix **H**.

⇨ e) Derive a systematic generator matrix **G** from the parity check matrix **H**. The information bits shall appear in original order as the first bits of the codeword. Make sure that each step in your derivation is clearly given.

The set of codewords of the $(7,4,3)_2$ Hamming code is summarized in the following table:

| | | | |
|---|---|---|---|
| *0000000* | *0100101* | *1000011* | *1100110* |
| *0001111* | *0101010* | *1001100* | *1101001* |
| *0010110* | *0110011* | *1010101* | *1110000* |
| *0011001* | *0111100* | *1011010* | *1111111* |

In the following problems f)-j), we consider code shortening. The codeword length and the information word length of the original code are denoted by $N$ and $K$, respectively. For shortening of the code, the shortened code consists only of a subset of the codewords in the table above. For the shortened code, we choose only those codewords, which have $M$ leading zeros. The $M$ leading zeros are then deleted in order to obtain codewords of length $N$-$M$.

⇨ f) Determine the information word length of the shortened code depending on $K$ and $M$.

g) Determine the code rate $R_s$ of the shortened code depending on $N$, $K$ and $M$.

⇨ h) What is the impact of code shortening on the minimum Hamming distance. Give clear reasons for your answer.

⇨ i) How can a parity check matrix $\mathbf{H}_s$ of the shortened code be obtained from the parity check matrix $\mathbf{H}$ of the original code ? What are the dimensions of the parity check matrix $\mathbf{H}_s$ of the shortened code depending on $N$, $K$ and $M$ ?

⇨ j) Consider shortening of the $(7,4,3)_2$ Hamming code by $M=3$. Determine the codewords of the resulting code. Which type of code is obtained ?

⇨ k) Code shortening can also be done in a fundamentally different way than described in this task. Explain, how code shortening is typically done for convolutional codes. Make sure that your explanation can be clearly understood and use complete sentences in your answer.