

Written Exam in  
**Data compression**  
**TSBK08**

26th August 2023 8:00 - 12:00

<b>Location:</b>	TER3
<b>Examiner:</b>	Harald Nautsch
<b>Teacher:</b>	Harald Nautsch,
<b>Department:</b>	ISY
<b>Module:</b>	TEN1
<b>Number of problems:</b>	7
<b>Number of pages:</b>	3
<b>Permitted equipment:</b>	Calculator, general English dictionaries
<b>Other:</b>	Answers can be given in English or in Swedish. The teacher will only be available on phone during the exam.
<b>Grades:</b>	0-13 U 14-19 3 20-25 4 26-30 5

- 1
- a) Formulate Kraft-McMillan's inequality. (1 p)
  - b) Explain what an instantaneous code is. (1 p)
  - c) Explain what a Golomb code is and what type of probability distribution it is good for. (2 p)
  - d) Explain how prediction with partial match (ppm) coding works. (2 p)
  - e) Explain what universal coding is and give an example of such a coding method. (2 p)

2 Let  $H(X)$  be the entropy of the random variable  $X$ . Show that

$$0 \leq H(X) \leq \log L$$

where  $L$  is the size of the alphabet.

(4 p)

3 A second order Markov source  $X_i$  with alphabet  $\mathcal{A} = \{a, b\}$  is given by the transition probabilities  $p(x_i|x_{i-1}x_{i-2})$  below (note the symbol order)

$$\begin{aligned} p(a|aa) &= 0.8, & p(b|aa) &= 0.2 \\ p(a|ab) &= 0.6, & p(b|ab) &= 0.4 \\ p(a|ba) &= 0.3, & p(b|ba) &= 0.7 \\ p(a|bb) &= 0.1, & p(b|bb) &= 0.9 \end{aligned}$$

a) Calculate the entropies  $H(X_i)$ ,  $H(X_i|X_{i-1})$  and  $H(X_i|X_{i-1}, X_{i-2})$  for the source. (3 p)

b) Construct a Huffman code for triples of symbols and calculate the rate of the code (in bits/symbol). (3 p)

- 4 A memoryless source has the alphabet  $\mathcal{A} = \{x, y, z\}$  with symbol probabilities  $p(x) = 0.6$ ,  $p(y) = 0.3$  and  $p(z) = 0.1$ . Show how arithmetic coding works by coding the sequence

$yxxzxx$

Give both the corresponding interval and the codeword. You can assume that all calculations are exact.

(4 p)

- 5 A source has the alphabet  $\{a, b, c, d, e, f, g, h\}$ . A sequence from the source is coded using LZW and gives the following index sequence:

1, 0, 8, 10, 0, 7, 4, 13, 15, 0, 6, 3, 17, 19, ...

The starting dictionary is:

index	sequence	index	sequence
0	$a$	4	$e$
1	$b$	5	$f$
2	$c$	6	$g$
3	$d$	7	$h$

Decode the index sequence. Also give the resulting dictionary.

(3 p)

- 6 A source has the alphabet  $\mathcal{A} = \{a, b, c, d\}$ .

Code the sequence

$cabcbbcad$

using BWT (Burrows-Wheeler's transform) followed by mtf (move-to-front coding). The block size should be 10.

(3 p)

- 7 Calculate the differential entropy of the distribution with probability density function  $f(x)$

$$f(x) = \begin{cases} 2 - 2x & ; 0 \leq x \leq 1 \\ 0 & ; \text{otherwise} \end{cases}$$

(2 p)