IMPERIAL COLLEGE LONDON

EE4-40
EE9-SO20

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2020

MSc and EEE/EIE PART IV: MEng and ACGI

INFORMATION THEORY

Tuesday, 12 May 10:00 am

Time allowed: 3:00 hours

There are FOUR questions on this paper.

Answer ALL questions.

All questions carry equal marks

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible First Marker(s) : C. Ling
Second Marker(s) : D. Gunduz

Information for students

Notation:

(a) Random variables are shown in Tahoma font. x, $\mathbf{x}$, $\mathbf{X}$ denote a random scalar, vector and matrix respectively.

(b) The size of a set A is denoted by $|A|$.

(c) By default, the logarithm is to the base 2.

(d) $\oplus$ denotes the exclusive-or operation, or modulo-2 addition.

(e) "i.i.d." means "independent identically distributed".

(f) $H(\cdot)$ is the entropy function.

(g) $C(x) = \frac{1}{2}\log_2(1+x)$ is the capacity function for the Gaussian channel in bits/channel use.

The Questions

1. Basics of information theory.

a)

i) Write down the formula of mutual information $I(X;Y)$. Give an interpretation of mutual information. [3]

ii) What are the maximum and minimum of $I(X;Y)$ over the input distribution? When are they achieved? [3]

iii) For $\mathbf{p} = [\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}]$ and $\mathbf{q} = [\frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{2}]$, compute the relative entropy $D(\mathbf{p}||\mathbf{q})$ and $D(\mathbf{q}||\mathbf{p})$. [3]

iv) Based on iii) say if the relative entropy is a distance or not. [3]

v) Express mutual information $I(X;Y)$ in terms of relative entropy. [3]

b)

i) Determine the stationary distribution and the entropy rate, $H(X)$, of a Markov chain with two states, 0 and 1 and transition matrix

$$T = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}.$$

[8]

ii) Find the value of p that maximizes $H(X)$. [2]

2. Source coding and typicality.

a) Lempel-Ziv coding. Give the LZ78 parsing and encoding of the following sequence:

$$110000110101000000110101$$

[Note: For this question, you will see less than 15 phrases; so ALWAYS use four bits to represent the location of a phrase. Do not worry about how to save such bits.]

[10]

b) Suppose the joint distribution $p_{xy}(x,y)$ is given by

$p_{xy}(x,y)$	$y = 0$	$y = 1$
$x = 0$	0.4	0.1
$x = 1$	0.1	0.4

$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ are drawn i.i.d from the above distribution. Now suppose $n = 10$ and $\varepsilon = 0.1$.

i) Of the 2^n possible sequences $\mathbf{x}$ of length n, how many of them are in the typical set $A_\varepsilon^{(n)}(\mathbf{x}) = \{\mathbf{x} : \left| -n^{-1}\log_2 p_x(\mathbf{x}) - H(x) \right| < \varepsilon\}$? [3]

ii) Of the 2^n possible sequences $\mathbf{y}$ of length n, how many of them are in the typical set $A_\varepsilon^{(n)}(\mathbf{y}) = \{\mathbf{y} : \left| -n^{-1}\log_2 p_y(\mathbf{y}) - H(y) \right| < \varepsilon\}$? [2]

iii) Define the jointly typical set

$$J_\varepsilon^{(n)} = \Big\{ \mathbf{x}, \mathbf{y} : \left| -n^{-1}\log_2 p_x(\mathbf{x}) - H(x) \right| < \varepsilon,$$
$$\left| -n^{-1}\log_2 p_y(\mathbf{y}) - H(y) \right| < \varepsilon,$$
$$\left| -n^{-1}\log_2 p_{xy}(\mathbf{x}, \mathbf{y}) - H(x, y) \right| < \varepsilon \Big\}$$

Determine the size and probability of the jointly typical set $J_\varepsilon^{(n)}$.

[10]

3. Channel coding and capacity.

a) Consider the binary erasure channel (BEC) shown in Fig. 3.1.

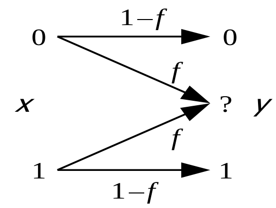

Fig. 3.1. Binary erasure channel (BEC).

Justify each step of the following derivation:

$$
\begin{aligned}
I(x;y) &\overset{(1)}{=} H(x) - H(x \mid y) \\
&\overset{(2)}{=} H(x) - p(y=0)\times 0 - p(y=?)H(x) - p(y=1)\times 0 \\
&\overset{(3)}{=} H(x) - H(x)f = (1-f)H(x) \\
&\overset{(4)}{\le} 1-f \\
&\overset{(5)}{\Rightarrow} C = 1-f
\end{aligned}
$$

[5]

b) Let W be a BEC with $f = 0.75$. Consider channel combining and splitting of polar codes in the following figure:

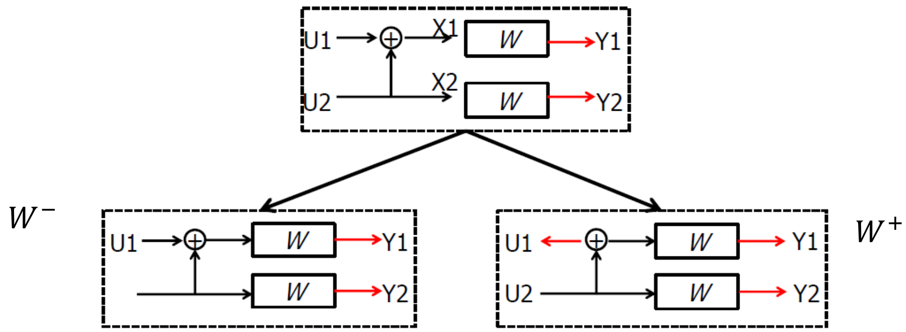

Fig. 3.2. Channel combining and splitting of polar codes.

Compute the capacity of the two sub-channels W^-, W^+.

[10]

c) Consider the diagram of three-level polarization:

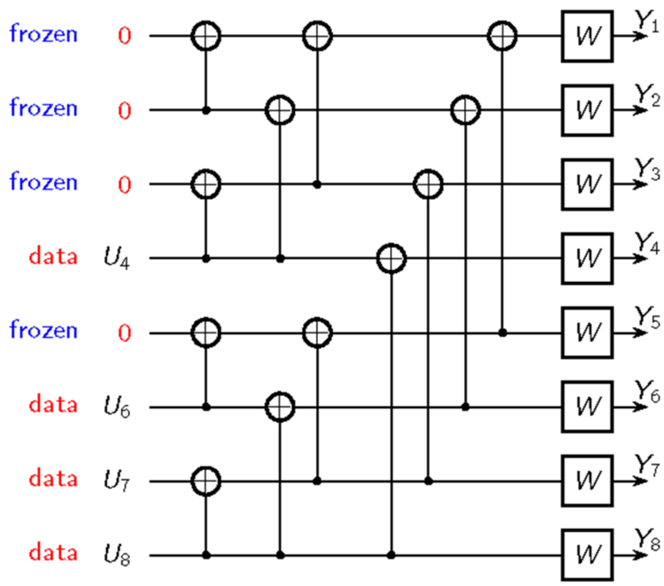

Fig. 3.3. Three-level polarization.

Set the frozen bits to 0 as shown above. Let data bits be $(U_4, U_6, U_7, U_8) = (1,1,1,1)$. Find the codeword $(X_1, X_2, \dots, X_8)$ of this polar code.

[10]

4. Continuous channels.

a) Parallel channels and waterfilling. Consider the following three parallel Gaussian channels

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$$

where

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} \right)$$

with a power constraint $E(x_1^2 + x_2^2 + x_3^2) \le 3P$. Assume that $\sigma_1^2 \ge \sigma_2^2 \ge \sigma_3^2$. At what power does the channel behave like

i) a single channel with noise variance σ_3^2 ? Find the channel capacity. [5]

ii) a pair of channels with noise variances σ_3^2 and σ_2^2 ? Find the channel capacity. [5]

iii) three channels with noise variances σ_3^2 , σ_2^2 , and σ_1^2 ? Find the channel capacity. [5]

b) Channel with uniformly distributed noise. Consider an additive channel whose input $X \in \{0, \pm 1\}$ and whose output $Y = X + Z$ where Z is noise uniformly distributed over interval [–1, 1]. Calculate the capacity of this channel. [10]