National University of Singapore
Department of Electrical & Computer Engineering

Examination for

EE5139 Information Theory for Communication Systems

(Semester I, 2019/20)
November/December 2019

Time Allowed: 3.0 hours

INSTRUCTIONS FOR CANDIDATES:

- This paper contains **FIVE (5)** questions, printed on **SEVEN (7)** pages.
- The total number of marks is 100.
- Answer all questions.
- Programmable calculators are NOT allowed.
- Electronic communicating devices MUST be turned off and inaccessible throughout the examination. They CANNOT be used as calculators, timers or clocks.
- You are allowed to bring ONE (1) HANDWRITTEN A4 size sheet. You may write on both sides of the sheet. No other material is allowed.
- All logarithms (log) are to the base 2.
- Throughout we let $H_{\mathrm{b}}(c) = -c\log_2 c - (1-c)\log_2(1-c)$ be the *binary entropy function.*

Question 1 (Total 20 points)

The following are TRUE/FALSE questions. No explanations are needed. Please just write TRUE or FALSE. Please indicate the part of the question ((a), (b), etc.) in your answer script clearly.

1. **(2 points)** We always have $I(X;Y|Z) \leq I(X;Y)$.

2. **(2 points)** We always have $D(P\|Q) = D(Q\|P)$.

3. **(2 points)** The typical set $A_\epsilon^{(n)}$ is the smallest set of sequences of length n which has total probability greater than $1-\epsilon$.

4. **(2 points)** There is a binary prefix-free variable-length source code with lengths $(1, 2, 3, 3)$.

5. **(2 points)** It is possible to reliably communicate a source X with entropy $H(X) = 0.75$ bits per source symbol over an additive white Gaussian noise channel with SNR 1.

6. **(2 points)** The capacity of a channel with input alphabet $\mathcal{X} = \{a, b, c\}$ and output alphabet $\mathcal{Y} = \{d, e, f, g\}$ can be $\log 5$ bits per channel use.

7. **(2 points)** Let $h(X)$ be the differential entropy of X. Then $h(X + c) = h(X)$ for any constant c.

8. **(2 points)** Let $h(X)$ be the differential entropy of X. Then $h(cX) = h(X)$ for any constant c.

9. **(2 points)** If X and Y are real-valued independent random variables, then $h(X + Y) \geq h(X)$.

10. **(2 points)** If $U_1 \in \{0, 1\}$ and $U_2 \in \{0, 1\}$ are i.i.d. Bernoulli-1/2 bits and $X_1 = U_1 \oplus U_2$ ($\oplus$ is modulo-2 addition) and $X_2 = U_2$, then $I(U_1, U_2; Y) = I(X_1, X_2; Y)$ for any random variable Y.

Question 2 (Total 20 points)

1. **(10 points)** Find the capacity of the binary erasure channel (BEC) with erasure probability $\beta \in [0, 1]$. See Fig. 1.

   Show your working carefully. You may assume that the capacity of a DMC is

$$C = \max_{P_X} I(X; Y).$$

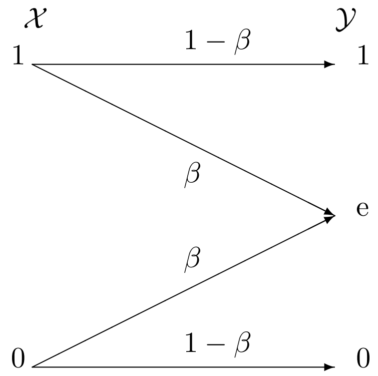

Figure 1: A BEC with erasure probability β.

2. **(10 points)** We have a parallel Gaussian channels consisting of three constituent Gaussian channels with noise variances

$$\sigma_1^2 = 2, \qquad \sigma_2^2 = 4, \qquad \sigma_3^2 = 8.$$

   Assume that the common power constraint is

$$\mathbb{E}\left[\sum_{j=1}^{3} X_j^2\right] \leq 10$$

   Find the optimal power allocation for the three channels and the resulting capacity of the parallel Gaussian channel.

Question 3 (Total 20 points)

Consider a binary memoryless channel with state whose input-output relation is described as follows. The received channel output Y is given by

$$Y = X \oplus S \oplus Z$$

where X is the channel input, S is the channel state, and Z is the channel noise. The random variables X, S, and Z take values in $\{0, 1\}$ and $\oplus$ denotes modulo-2 addition. Furthermore $S \sim \mathrm{Bern}(q)$ and $Z \sim \mathrm{Bern}(p)$ for $p, q \in [0, 1/2]$ are independent and jointly independent of the channel input X. When we employ n-block encoding and decoding, we have for each $1 \le i \le n$,

$$Y_i = X_i \oplus S_i \oplus Z_i$$

where $S^n = (S_1, \ldots, S_n)$ are i.i.d. $\mathrm{Bern}(q)$ and $Z^n = (Z_1, \ldots, Z_n)$ are i.i.d. $\mathrm{Bern}(p)$ and where S^n and Z^n are independent, and jointly independent of the channel input sequence $X^n = (X_1, \ldots, X_n)$.

In this question, you may assume that the capacity of a binary symmetric channel with crossover probability α is

$$C(\mathrm{BSC}(\alpha)) = 1 - H_{\mathrm{b}}(\alpha).$$

Find the capacity of the channel in the following cases:

3(a) **(4 points)** Neither the encoder nor the decoder knows the state sequence. See Fig. 2.

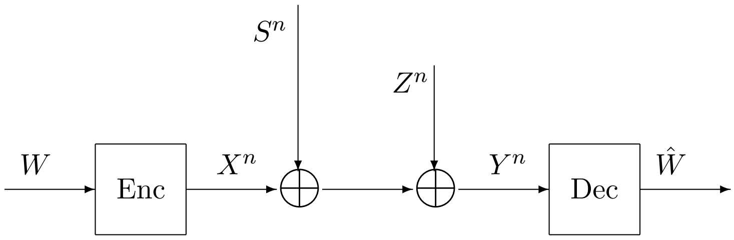

Figure 2: In part (a), neither link is closed

3(b) **(8 points)** The state sequence is known *only* to the decoder but *not* the encoder, i.e., the n-block decoder gets to base its decision on *both* Y^n and S^n. See Fig. 3.

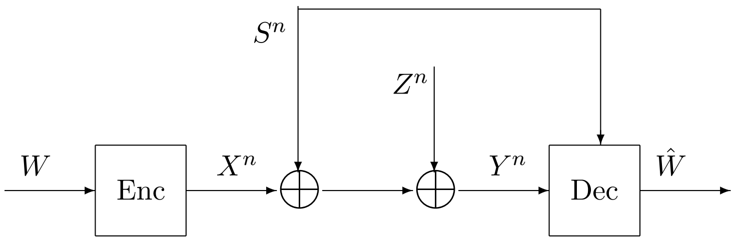

Figure 3: In part (b), the link from S^n to Dec is closed;

(Question continued on next page)

3(c) **(8 points)** The state sequence is known to *both* the encoder and the decoder, i.e., the n-block encoder and decoder know S^n prior to communication. See Fig. 4.

Hint: The capacity expression involves maximizing over a certain conditional probability.

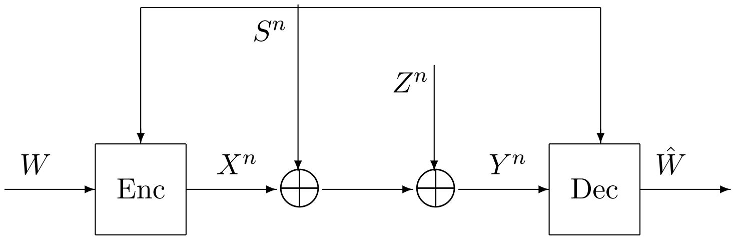

Figure 4: In part (c), the links from S^n to Enc and Dec are closed

Question 4 (Total 20 points)

In this question, we would like to determine the *minimum distance* of a randomly chosen code. Some notation: The *binary divergence* between Bern(α) and Bern(β) is

$$D(\alpha\|\beta) = \alpha \log \frac{\alpha}{\beta} + (1-\alpha) \log \frac{1-\alpha}{1-\beta}.$$

The *Hamming distance* between two binary vectors $a^n, b^n \in \{0,1\}^n$ is

$$d_{\mathrm{H}}(a^n, b^n) = \sum_{i=1}^{n} \mathbf{1}\{a_i \neq b_i\}.$$

The *Gilbert-Varshamov (GV) distance* at rate R, denoted as $\delta_{\mathrm{GV}}(R)$ is defined as the root $\delta \leq 0.5$ of the equation

$$D(\delta\|0.5) = R.$$

A random binary code $\mathcal{C}$ of length n and rate R is a collection of $M = 2^{nR}$ random codewords (vectors) each of length n, i.e.,

$$\mathcal{C} = \{X^n(1), X^n(2), \dots, X^n(M)\}.$$

Assume that each random codeword $X^n(m)$ has components $X_1(m), X_2(m), \dots, X_n(m)$ that are chosen independently from the uniform distribution on $\{0,1\}$, i.e.,

$$\Pr(X^n(m) = x^n) = 2^{-n}, \qquad \forall\, x^n \in \{0,1\}^n$$

4(a) **(6 points)** Show that the probability that the Hamming distance of $X^n(1)$ from the all zeros vector is $d \in \{0, 1, \dots, n\}$ is given by

$$\Pr(d_{\mathrm{H}}(X^n(1), 0^n) = d) \doteq 2^{-nD(\delta\|0.5)} \qquad (*)$$

where $\delta = d/n$.

4(b) **(2 points)** Let a^n be an arbitrary vector in $\{0,1\}^n$. Provide a convincing argument that for any $1 \leq m \leq M$, the probability $\Pr(d_{\mathrm{H}}(X^n(m), a^n) = d)$ is equal to the quantity in (*) in part (a).

4(c) **(5 points)** Let $N_{\mathcal{C}}(d)$ be the number of pairs of random vectors $X^n(m)$ and $X^n(m')$ for $m \neq m'$ that are at distance d, i.e.,

$$N_{\mathcal{C}}(d) := \sum_{m=1}^{M-1} \sum_{m'=1}^{m-1} \mathbf{1}\{d_{\mathrm{H}}(X^n(m), X^n(m')) = d\}$$

Show that

$$\mathbb{E}\left[N_{\mathcal{C}}(d)\right] \doteq 2^{n[2R - D(\delta\|0.5)]}.$$

4(d) **(7 points)** The *relative minimum distance* $\delta_{\min}(\mathcal{C})$ of the code is defined to be the smallest δ such that $N_{\mathcal{C}}(\delta n) \neq 0$. Show using Markov's inequality that for $0 \leq R < 0.5$ and any $\epsilon > 0$, the probability that the random code of length n and rate R has relative minimum distance less than $\delta_{\mathrm{GV}}(2R) - \epsilon$ tends to zero exponentially fast.

Question 5 (Total 20 points)

In class, we considered the binary hypothesis testing problem

$$H_0 : X^n \overset{i.i.d.}{\sim} P_0, \qquad H_1 : X^n \overset{i.i.d.}{\sim} P_1.$$

For an acceptance region $\mathcal{A} \subset \mathcal{X}^n$ for hypothesis H_1, we define the type-I and type-II error probabilities as follows:

$$\alpha_n(\mathcal{A}) = \Pr(X^n \in \mathcal{A}|H_0) = P_0^n(\mathcal{A}), \qquad \beta_n(\mathcal{A}) = \Pr(X^n \in \mathcal{A}^c|H_1) = P_1^n(\mathcal{A}^c).$$

For a fixed $\epsilon \in (0, 1)$, we considered

$$\beta_n(\epsilon) := \min_{\mathcal{A}\subset\mathcal{X}^n} \{\beta_n(\mathcal{A}) : \alpha_n(\mathcal{A}) \le \epsilon\}$$

and showed that its exponential rate of decay is $D(P_0\|P_1)$. We now demand a more stringent condition on the type-I error. Instead of a non-vanishing error probability ϵ, we demand that it decays exponentially fast with exponent $\lambda > 0$. Specifically, we would like to quantify

$$\widetilde{\beta}_n(\lambda) := \min_{\mathcal{A}\subset\mathcal{X}^n} \left\{\beta_n(\mathcal{A}) : \alpha_n(\mathcal{A}) \le 2^{-n\lambda}\right\}$$

5(a) **(1 point)** Let $\mathcal{T}_Q \subset \mathcal{X}^n$ be the type class of $Q \in \mathcal{P}_n(\mathcal{X})$ and let

$$\widetilde{\lambda} := \lambda + \frac{|\mathcal{X}| \log(n + 1)}{n}.$$

Let the set $\mathcal{A}$ be defined as

$$\mathcal{A} := \bigcup_{Q\in\mathcal{P}_n(\mathcal{X}):D(Q\|P_0)\ge\widetilde{\lambda}} \mathcal{T}_Q$$

Write out explicitly the test, in terms of the type of X^n, that we use when the acceptance region $\mathcal{A}$ is employed.

5(b) **(6 points)** By using the method of types, show that the set $\mathcal{A}$ satisfies the constraint

$$\alpha_n(\mathcal{A}) \le 2^{-n\lambda}.$$

5(c) **(6 points)** By using the set $\mathcal{A}$ defined in part (a) and the method of types, find a lower bound for

$$\lim_{n\to\infty} -\frac{1}{n} \log \widetilde{\beta}_n(\lambda)$$

5(d) **(4 points)** By taking $\lambda = \frac{1}{n} \log \frac{1}{\epsilon} \downarrow 0$, show that the exponential rate of decay of the type-II error when the type-I error is bounded above by ϵ is given by $D(P_0\|P_1)$.

5(e) **(3 points)** By taking $\lambda \uparrow D(P_1\|P_0)$, what can we say about the exponential rate of decay of the type-II error probability in (c)?

END OF PAPER