Problem 1 – Source Coding (20 Points)

(1) **(14 Points)** Consider symbol-wise source coding on a finite alphabet $\mathcal{X} = \{a, b, \dots\}$ (of unspecified size), with symbol distribution P_X, code lengths $\ell(a), \ell(b), \dots$, and an average code length of $L(C)$.

(i) Does there exist a source P_X for which the value of $\ell(a)$ for Shannon-Fano coding exceeds that of Huffman coding by at least 5? Explain.

(ii) Does there exist a source P_X for which the value of $L(C)$ for Shannon-Fano coding exceeds that of Huffman coding by at least 5? Explain.

(iii) Give an example of a source P_X such that Huffman coding applied to pairs (i.e., applied to the alphabet $\mathcal{X}^2$ and distribution $P_{X_1X_2}(x_1, x_2) = P_X(x_1)P_X(x_2)$) gives a strictly smaller average length per $\mathcal{X}$-symbol compared to Huffman coding applied to P_X alone. Explain.

(b) **(6 Points)** Consider a discrete memoryless source with per-symbol distribution P_X and block length n. Suppose that someone claims to come up with a "better" notion of a typical set (different from the typical set definition in the lecture), denoted $\mathcal{T}_n^*$, such that $\mathbb{P}[\mathbf{X} \in \mathcal{T}_n^*] \to 1$ as $n \to \infty$ and $|\mathcal{T}_n^*| \leq 2^{nA}$ for some constant value of $A < H(X)$ (where A and P_X do not vary as n increases, and where $\mathbf{X}$ is distributed according to $\prod_{i=1}^n P_X(x_i)$ as usual). Is this possible? Explain briefly.

Problem 2 – Discrete and Continuous Information Measures (25 Points)

(a) **(16 Points)** Answer the following:

(i) Write down a chain of inequalities between $H(X)$, $H(X,Y)$, and $I(X;Y)$, ordering them from smallest to largest. Briefly explain each inequality (2 in total).

(ii) Consider comparing $I(X;Y)$ to $H(X|Y)$ for an arbitrary joint distribution P_{XY}. Can we say that $I(X;Y) \geq H(X|Y)$ always, that $I(X;Y) \leq H(X|Y)$ always, or that both $I(X;Y) > H(X|Y)$ and $I(X;Y) < H(X|Y)$ are possible? Explain.

(iii) Prove that $I(X,Z;Y,Z) = H(Z) + I(X;Y|Z)$ for any random variables (X,Y,Z).

(iv) Describe a joint distribution on (X,Y,Z) such that X, Y, and Z are binary-valued (0 or 1), $H(X) = H(Y) = H(Z) = 1$, $I(X;Y) = I(X;Z) = I(Y;Z) = 0$, and $H(X,Y,Z) = 2$. Briefly explain why all these entropy and mutual information values are attained under your choice of joint distribution.

(b) **(9 Points)** Suppose that someone has already computed the differential entropy $h(\tilde{U}+\tilde{V})$ in the case that $\tilde{U}$ and $\tilde{V}$ are independent random variables distributed uniformly in the interval $[0,1]$; let ξ denote this value of $h(\tilde{U}+\tilde{V})$. (Its precise value is not needed for the purpose of answering this question.)

Let U and V be independent random variables distributed uniformly in the interval $[0,4]$, and define $Z = U + V$. Compute the differential entropies $h(U)$, $h(Z)$, and $h(Z|U)$, and the mutual information $I(U;Z)$, leaving your answers in terms of the quantity ξ introduced above as appropriate.

Problem 3 – Practical Codes (25 Points)

(a) **(10 Points)** Consider the following code with 4 codewords: $\{000000, 001111, 111100, 111111\}$.

(i) Explain why this is not a linear code.

(ii) What is the minimum distance of this code? Explain briefly.

(iii) Replace one of the 4 codewords by a new codeword such that the resulting code (different from the one above) is a linear code.

(b) **(8 Points)** Write down (i) the generator matrix $\mathbf{G}$, (ii) the parity check matrix $\mathbf{H}$, and (iii) the code rate, for a systematic code with five information bits $(u_1, u_2, u_3, u_4, u_5)$, and three parity check bits taking the form

$$u_1 \oplus u_2 \oplus u_3$$
$$u_1 \oplus u_3 \oplus u_5$$
$$u_2 \oplus u_4.$$

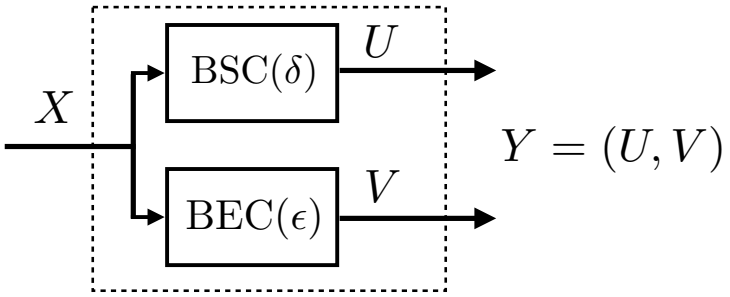
Explanations are not required.

(c) **(7 Points)** In the lecture, we introduced the Hamming code with 16 binary codewords of length 7, and showed that the code is guaranteed to correct up to one error (i.e., given a received string containing no bit flips or a single bit flip compared to some codeword, it's guaranteed that this codeword can be uniquely recovered).

Prove that it is *impossible* to find a code with 17 or more binary codewords of length 7 while being guaranteed to correct up to one error, even if the code may be non-linear.

Problem 4 – Channel Coding (30 Points)

(a) **(20 Points)** Consider the following setup in which $X \in \{0,1\}$ is passed through a binary symmetric channel (BSC) to get $U \in \{0,1\}$, and X is also passed through a binary erasure channel (BEC) to get $V \in \{0,1,e\}$, where e is the erasure symbol. (Assume that the BSC and BEC outputs are conditionally independent given X.) Then, the overall output is a pair consisting of both resulting values, i.e., $Y = (U,V)$.



Overall Channel

Consider choosing $P_X(1) = p$ and $P_X(0) = 1-p$ for some $p \in (0,1)$, and answer the following with your answers depending on p and/or δ and/or ϵ as needed:

(i) Explain why $I(X;U|V=0)$ and $I(X;U|V=1)$ are both zero.

(ii) Prove that the joint conditional distribution of (X,U) given $V=e$ is the same as the unconditional joint distribution of (X,U)

(iii) Using the previous parts or otherwise, show that $I(X;U|V) = \epsilon \cdot I(X;U)$.

(iv) Using the previous parts or otherwise, find $I(X;V)$ and $I(X;U|V)$ when $p = \frac{1}{2}$.

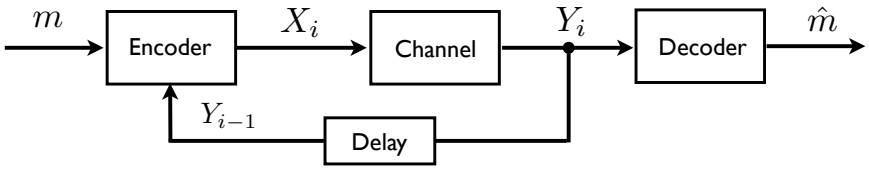
(v) Using the previous parts or otherwise, find the capacity of the overall channel $P_{Y|X}$ (and explain why this is the capacity).

(Note: You may make use of the binary entropy function $H_2(\cdot)$ as usual. You may also use any known facts/properties from the lecture regarding the BSC and/or the BEC and/or their associated mutual information terms.)

(More space for answering on the next page)

(More space for answering Question 4(a))

(b) **(10 Points – Advanced)** Consider the following setup of communication with feedback:



Formally, the setup described is as follows:

- As usual, the message m is uniformly distributed over the set $\{1, \ldots, M\}$.
- Different from the lecture, the encoder may choose the next input as a function of all previous outputs: $X_i = f_i(Y_1, \ldots, Y_{i-1})$ for some deterministic function f_i.
- The channel is assumed to be a discrete memoryless channel, with the memorylessness assumption now being that Y_i is conditionally independent of the message and previous inputs/outputs (i.e., $(m, X_1, \ldots, X_{i-1}, Y_1, \ldots, Y_{i-1})$) given the input X_i. The corresponding conditional distribution is written as $P_{Y|X}$, as usual.
- The decoder outputs the estimate $\hat{m}$ after all outputs $Y_1, \ldots, Y_n$ have been received.

We also define $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ as usual.

Defining the channel capacity with feedback, C_{F}, in an analogous manner to the lecture, it turns out that $C_{\mathrm{F}} = \max_{P_X} I(X;Y)$, i.e., feedback does not increase the capacity. The converse proof is mostly the same as the non-feedback case, but we no longer necessarily have the Markov chain relation $m \to \mathbf{X} \to \mathbf{Y} \to \hat{m}$, so it is now more difficult to show that $I(m;\hat{m}) \leq \sum_{i=1}^{n} I(X_i;Y_i)$. This question works through filling in this missing step:

(i) Explain why $I(m;\hat{m}) \leq I(m;\mathbf{Y})$.

(ii) Prove that $I(m;\mathbf{Y}) \leq \sum_{i=1}^{n} I(X_i;Y_i)$, carefully explaining all steps.

(Note: Marks will <u>not</u> be awarded for copying the same steps as the non-feedback case.)

(More space for answering on the next page)

(More space for answering Question 4(b))

[Use this page for any extra working if you run out of space. You must clearly write "See final pages" for any question continued here, and here you must clearly indicate each exact question and part (e.g., 2(c)).]

[Use this page for any extra working if you run out of space. You must clearly write "See final pages" for any question continued here, and here you must clearly indicate each exact question and part (e.g., 2(c)).]

END OF PAPER