Introduction to Information and Coding theory 30 June 2013
(Dr. Permuter Haim and Mr. Ziv Goldfeld)

Partial solution to Final Exam
Total time for the exam: 3 hours!

1) **True or False** (30 points):
Copy each relation to your notebook and write **true** or **false**. Then, if it's true, prove it. If it is false give a counterexample or prove that the opposite is true.

a) Let X be a continuous random variable. Then the following holds

$$I(X;X) = h(X).$$

Solution: False. $I(X;X) = \infty$ when X is continuous (show it.). Note that $h(X|X)$ is undefined since given X the r.v. X is discrete and not continuous.

b) Let X, Y, Z be three random variables that satisfies $H(X,Y) = H(X) + H(Y)$ and $H(Y,Z) = H(Z) + H(Y)$. Then the following holds

$$H(X,Y,Z) = H(X) + H(Y) + H(Z).$$

Solution: False. Consider X, Y be Binary Bern($\frac{1}{2}$) and $Z = X + Y$.

c) For any X, Y, Z and the deterministic function f, g

$$I(X;Y|Z) = I(X, f(X,Y); Y, g(Y,Z)|Z).$$

Solution: False. Let (X,Y,Z) be a triplet of random variables such that $H(Y|X,Z) > 0$. Taking $f(X,Y) = Y$ and $g(Y,Z) = 0$ yields $I(X, f(X,Y); Y, g(Y,Z)|Z) = I(X,Y;Y|Z) = h(Y|Z)$, which is strictly larger than $I(X;Y|Z)$.

d) $H(X|Z)$ is concave in $P_{X|Z}$ for fixed P_Z.
Solution: True. $H(X|Z=z)$ is concave in $P_{X|Z=z}$ and $H(X|Z)$ is a linear combination of $H(X|Z=z)$ with coefficients $P(z)$.

e) Let $P(y|x)$ characterize a channel with Binary input and let $P(y|x=1) = P(y|x=0)$ for all $y \in \mathcal{Y}$. The capacity of this channel is 0.
Solution: True. For any $P(x)$ we have $P(y) = P(y|x=1) = P(y|x=0)$. (Show it.), hence $H(Y) = H(Y|X)$.

2) **Two antennas with Gaussian noise** (20 points): In this question we consider a point-to-point discrete memoryless channel (DMC) in which the transmitter and the receiver both have two antennas, illustrated in Fig. 1. This channel is defined by two input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ and a channel transition matrix $P_{Y_1Y_2|X_1X_2}$. A message M is randomly and uniformly chosen from the message set $\mathcal{M} = \{1, 2, \ldots, 2^{nR}\}$ and is to be transmitted from the encoder to the decoder in a lossless manner (as defined in class).

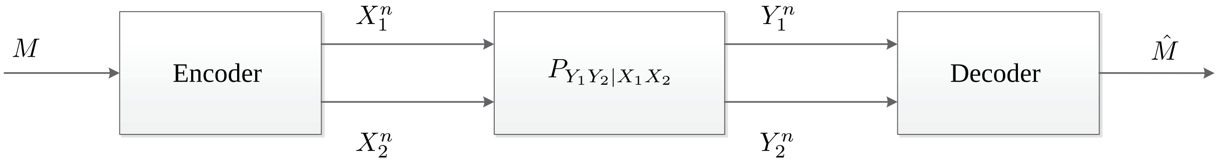


Fig. 1. Two antenna point-to-point DMC.

a) What is the capacity of the channel?

Now, consider the following Gaussian two antenna point-to-point DMC illustrated in Fig. 2

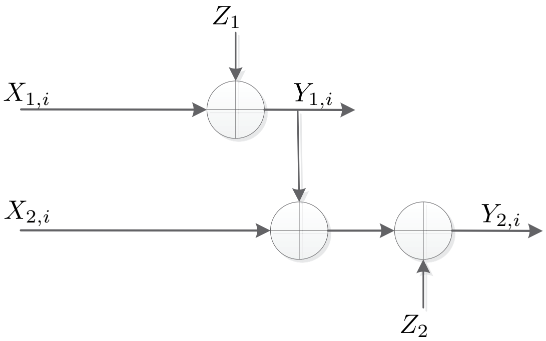


Fig. 2. A Gaussian two antenna point-to-point DMC.

The outputs of the channel for every time $i \in \{1, \dots, n\}$ are give by,

$$
\begin{aligned}
Y_{1,i} &= X_{1,i} + Z_1, && (1)\\
Y_{2,i} &= X_{1,i} + X_{2,i} + Z_1 + Z_2, && (2)
\end{aligned}
$$

where (Z_1, Z_2) are two independent (of each other and of everything else) Gaussian random variable distributed according to $Z_1 \sim \mathcal{N}(0, N_1)$ and $Z_1 \sim \mathcal{N}(0, N_2)$. The input signals are bound to an average power constraints,

$$
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_{1,i}^2\right] \leq P_1 \quad ; \quad \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_{2,i}^2\right] \leq P_2. \tag{3}
$$

b) Find the capacity of the Gaussian channel in terms of the provided parameters and state the joint distribution of (X_1, X_2) that achieves it.

Solution:

a) Let us denote the input pair (X_1^n, X_2^n) by $\widetilde{X}^n$ and the output pair (Y_1^n, Y_2^n) by $\widetilde{Y}^n$. An equivalent channel to the one considered in this question is the point-to-point DMC for which $(\widetilde{X}^n, \widetilde{Y}^n)$ serve as the channel's input and output sequences, respectively, and the channel transition matrix is $P_{\widetilde{Y}|\widetilde{X}}$. Recalling that the point-to-point channel capacity is given by $\max_{P_{\widetilde{X}}} I(\widetilde{X}; \widetilde{Y})$, and substituting $\widetilde{X}^n = (X_1^n, X_2^n)$ and $\widetilde{Y}^n = (Y_1^n, Y_2^n)$ we obtain:

$$
C = \max_{P_{X_1 X_2}} I(X_1, X_2; Y_1, Y_2). \tag{4}
$$

b) First now that Y_2 can be rewritten as $Y_2 = Y_1 + X_2 + Z_2$. Now, we upper bound the capacity as:

$$
\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2 | Y_1)\\
&= h(Y_1) - h(Y_1 | X_1, X_2) + h(Y_2 | Y_1) - h(Y_2 | X_1, X_2, Y_1)\\
&\overset{(a)}{=} h(Y_1) - h(Z_1) + h(Y_1 + X_2 + Z_2 | Y_1) - h(Y_1 + X_2 + Z_2 | X_1, X_2, Y_1)\\
&\overset{(b)}{=} h(Y_1) - h(Z_1) + h(X_2 + Z_2 | Y_1) - h(Z_2)\\
&\overset{(c)}{\leq} h(Y_1) - h(Z_1) + h(X_2 + Z_2) - h(Z_2)\\
&= h(Y_1) - \frac{1}{2}\log(2\pi e N_1) + h(Y_2) - \frac{1}{2}\log(2\pi e N_2)\\
&\overset{(d)}{\leq} \frac{1}{2}\log\big(2\pi e(P_1 + N_1)\big) - \frac{1}{2}\log(2\pi e N_1) + \frac{1}{2}\log\big(2\pi e(P_2 + N_2)\big) - \frac{1}{2}\log(2\pi e N_2)\\
&= \frac{1}{2}\log\Big(\prod_{i=1}^{2}(P_i + N_i)\Big)
\end{aligned}
$$

where:
(a) follows from the definitions of Y_1 and the fact that Z_1 is independent of X_1;
(b) follows from the fact that Z_2 is independent of (X_1, X_2, Z_1) and therefore it is independent of (X_1, X_2, Y_1);
(c) follows from the fact that conditioning reduces entropy;
(d) follows by the maximum of differential entropy property.

This upper bound is achieved by choosing (X_1, X_2) to be jointly Gaussian RVs with the following distribution,

$$
\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}\right) \tag{5}
$$

This distribution achieves (c) with an equality since by this choice we get that $Y_1 = X_1 + Z_1$ and $X_2 + Z_2$ are independent. Whereas (d) is achieved with equality since by this choice Y_1 and Y_2 are Gaussian RVs with variances $P_1 + N_1$ and $P_2 + N_2$, respectively (which achieves the maximum of entropy)

3) **Riddle** (20 points): In a magic trick, there are three participants: the magician, an assistant, and a volunteer. The assistant, who claims to have paranormal abilities, is in a soundproof room. The magician has a deck of 100 cards. A different number from 1 to 100 is written on each card (in other words, they are numbered from 1 to 100). The magician asks a volunteer from the crowd to pick six cards. Then, the cards are shown to the crowd. The volunteer keeps one of the cards. The magician arranges the five cards that are left in some order. Now the assistant comes to the stage looks at the five cards and announces the number of the card kept by the volunteer!

a) The magician and the assistant are experts in information theory. How did they preform the trick? (Hint: One can used the 5 remaining cards to encode a message).

b) Can the magician and the assistant preform this trick with more then 100 cards? If yes, explain how many. If the answer is no, explain.

Solsution: The five cards are uniquely identified by their numbers (low to high). There are $5! = 120$ possible orderings for the five cards, which is more than enough to encode the number on the sixth card.

The actual numbers written on the white cards don't matter, let's call them, in increasing order, $C_1, C_2, , C_5$. Since there are $5!$ permutations of $1, 2, , 5$, enough to encode all the numbers from 1 to 120. The code (for 120) could go as follows. If C_1 is the top card, then the hidden card is in the range 1 to 24; if it is C_2, the hidden card is 25 to 48, and so on. Then the range identified by the top card is narrowed down by considering the next card. And so on. With some practice the "reading" would be quick.

The magician and the assistant can indeed perform the magic with a deck of more than 100 card. A deck of 125 cards is the maximum possible (without flipping or rotating the cards). This is since, as mentioned above, there are $5! = 120$ possible orderings for the five cards, plus the 5 cards that the assistant already knows, thus making a total of 125 cards.

4) **network coding** (30 points): Consider the network shown in Fig.3. This network consists of one source node S, two destination nodes D_1, D_2, 12 another nodes and a XOR gate (the most left circle). Every edge (i, j) in the network represents a directional noiseless link from node i to j, that can transmit 1 bit per second.

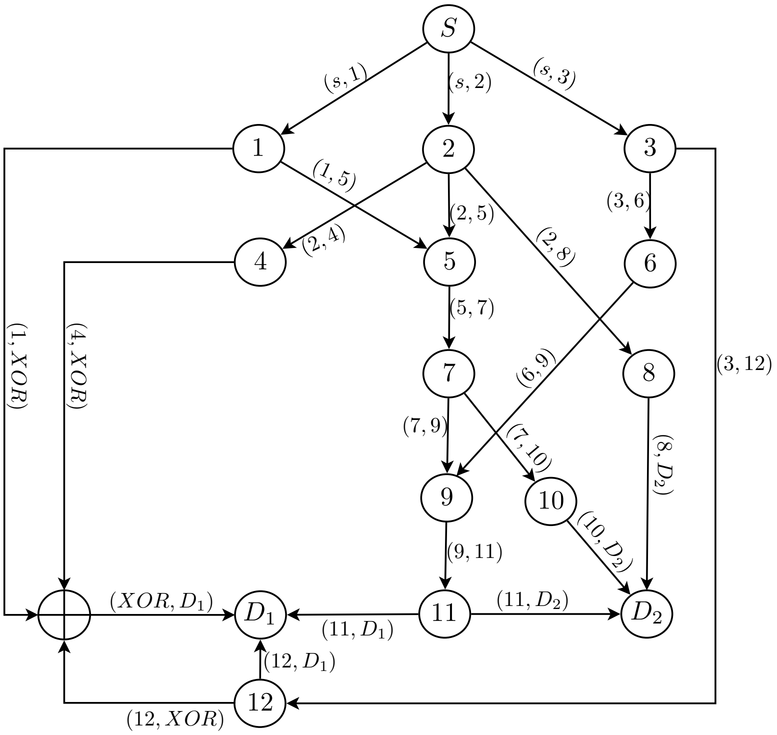


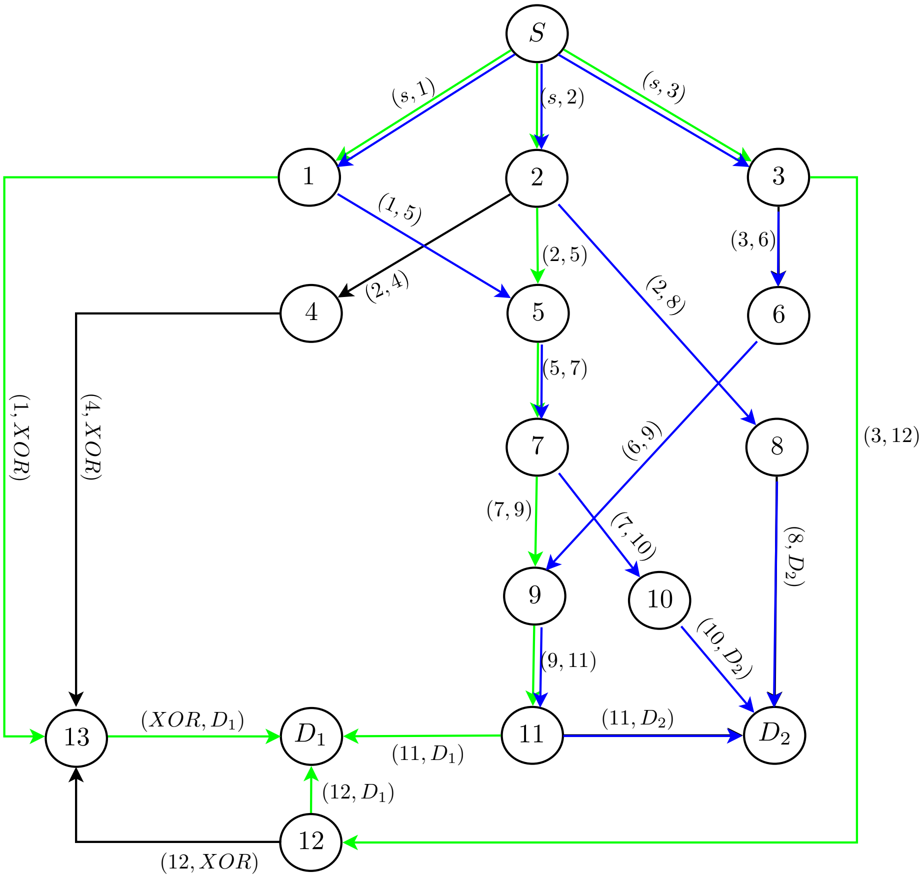
Fig. 3. A network with one source node, two destination nodes, another 12 nodes and a XOR gate

In subquestions a-c the XOR gate is replaced by an ordinary node 13 like the other nodes $1-12$ (it doesn't have to do a XOR operation on the received bits).

a) For every destination node, find how many edge-disjoint paths exist between the source node and that destination node?
b) We want to transmit data from the source node to the destination node D_i $(i = 1$ or $2)$. The other destination node doesn't have to receive the data. What is the maximal transmission rate R_i (in bits per second) that can be achieved in that case? Find for $i = 1, 2$.
c) Can we achieve these rates by simple routing scheme, where every node sends on its output links only bits there were received at its input links.
d) Design a linear network code that would allow the source node to transmit data to both destination nodes D_1, D_2 at the rates R_1, R_2. (The XOR node behave in this subsection as XOR, namely the output is a XOR of all inputs.) Draw the network and on each edge write the linear function that is applied.
e) Write the transfer matrix for each destination as a function of the source.

Solution

a) There are three edge-disjoint paths from the source node S to both destination nodes D_1, D_2. They are shown in the figure below. The paths from S to D_1 are shown in green, while the paths to D_2 are shown in blue.



b) For every destination D_i, we know that the maximal transmission rate R_i that can be achieved is equal to the amount of edge-disjoint paths from the source node S to D_i. We've seen that their amount is 3 (for both D_1 and D_2). Since every link in the network allows to send at a rate of $1[bps]$, we have

$$R_1 = R_2 = 3[bps]$$

We now want to transmit data from the source node to both destination nodes D_1, D_2 at the rates R_1, R_2 that were just found.

c) We'll prove it by contradiction. Suppose the rates $R_1 = R_2 = 3[bps]$ can be achieved by simple routing scheme. In that case, the source node S will send three bits per second on its output links. Suppose the bit X_1 is sent on the link $(s,1)$, the bit X_2 is sent on the link $(s,2)$ and the bit X_3 is sent on the link $(s,3)$. Every node i (for $i = 1,2,3$) will send the bits X_i on its output links (assumption of simple routing scheme). Note that there is only one path that can lead the bit X_3 to node D_2, which is

$$P_{D_2}(3) = \{(3,6),(6,9),(9,11),(11,D_2)\}$$

Therefore every bit that will be sent on these links must be X_3.Moreover, the bit that will be sent on the link $(11, D_1)$ must be also X_3 since thats the bit node 11 receives. Node 4 will send the bit X_2 to node 13 (the node that replaces the XOR gate) which in turn will need to decide which bit to send on the link (XOR, D_1): X_1 or X_2. In any case, the node D_1 won't get both X_1 and X_2. This proves that the rates $R_1 = R_2 = 3[bps]$ can't be achieved in that way.

d) We'll use finite field to solve this question. We'll identify every two bits as a scalar from the field $F_{2^2} = \{0,1,2,3\}$. Every node will send a scalar every two seconds on their output links. The sent scalars are functions of the received scalars. The code is as follows:

i) The source node S sends 3 scalars X_1, X_2, X_3 on the links $(s,1),(s,2),(s,3)$ respectively.

ii) Node 1 gets the scalar X_1 and sends it on the link $(1,5)$.

iii) Node 2 gets the scalar X_2 and sends it on the links $(2,4),(2,5),(2,8)$.

iv) Node 3 gets the scalar X_3 and sends it on the links $(3,6),(3,12)$.

v) Nodes 1,4 and 12 send the scalars X_1, X_2, X_3 to the XOR gate, which in turn will send their XOR result to D_1 on the link (XOR, D_1). Note that the bitwise XOR result of three scalars is their summation in the finite field F_{2^2}.

vi) Node 5 gets the scalars X_1, X_2 and sends the scalar W on the link $(5,7)$, where

$$W = X_1 + 2X_2$$

vii) Node 6 gets the scalar X_3 and sends it on the link $(6,9)$.

viii) Node 7 gets the scalar W and sends it on the links $(7,9),(7,10)$.

ix) Node 8 gets the scalar X_2 and sends it on the link $(8, D_2)$.

x) Node 9 gets the scalars W, X_3 and sends the scalar Z the link $(7, D_2)$, where

$$Z = W + X_3$$

xi) Node 10 gets the scalar W and sends it on the link $(10, D_2)$.

xii) Node 11 gets the scalar Z and sends it on the links $(11, D_1), (11, D_2)$.

xiii) Node 12 gets the scalar X_3 and sends it on the link $(12, D_1)$.

xiv) The destination D_1 receives the following scalars:

$$Y_1 = X_1 + 2X_2 + X_3, Y_2 = X_3, Y_3 = X_1 + X_2 + X_3$$

and it finds X_1, X_2, X_3 by solving a set of linear equations:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

These equations have a unique solution since the determinant of the matrix isn't zero:

$$\begin{vmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{vmatrix} = (-1) \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = 3$$

xv) The destination D_2 receives the following scalars:

$$Y_1 = X_2, Y_2 = X_1 + 2X_2, Y_3 = X_1 + 2X_2 + X_3$$

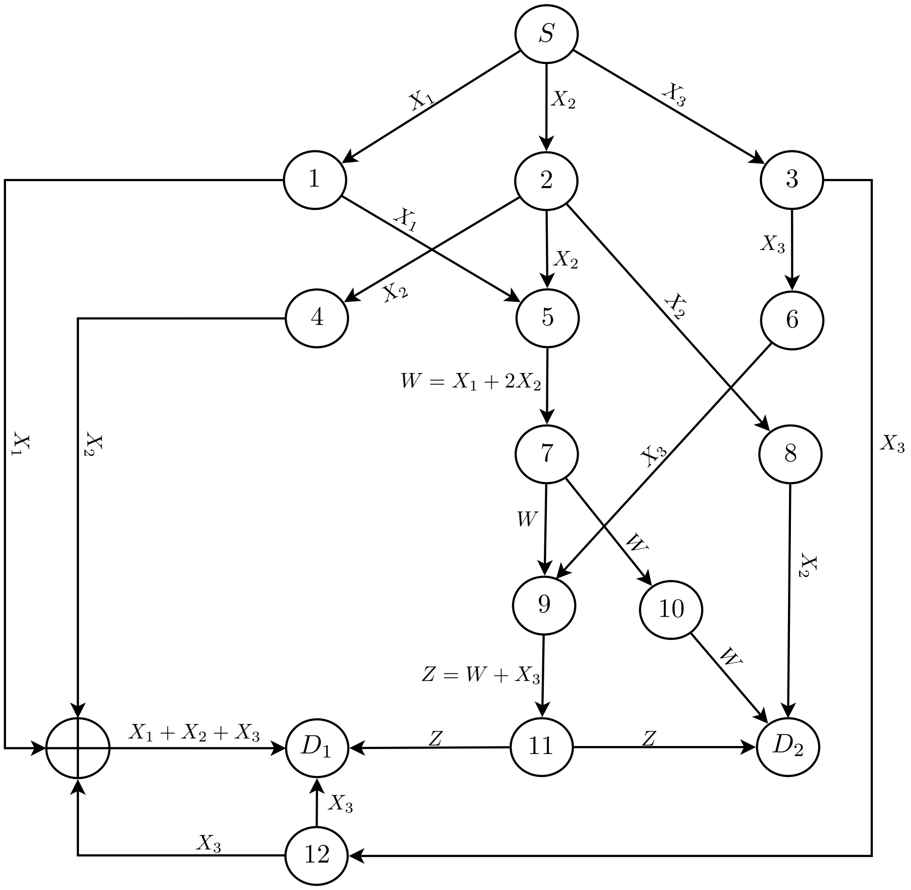
and it finds X_1, X_2, X_3 by solving a set of linear equations:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

These equations have a unique solution:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

A diagram of the network code is shown in the figure below



Remark: There exists more than one solution. You may work on $\mathbb{F}_2$ field, transmit on $(S, 3)$, X_3, on $(5, 7)$, X_1 and on (9,11) $X_1 + X_3$. This yields three equations that are independent.

Good Luck!