Problem 1

On the Achievable Rate

![Diagram of two parallel channels](https://example.com/diagram.png)

The second channel, $W_2$, can be considered to consist of two parallel channels $W_1$ as shown in the figure above, i.e., one use of $W_2$ corresponds to two independent uses of $W_1$.

a) Consider a sequence of codes $\{C_1\}_n$ of length $n$ and achievable rate $R$ for $W_1$. Now consider a sequence of codes $\{C_2\}_n$ for $W_2$ whose codewords are of length $n$ and consists of codewords from $\{C_1\}_n$ in the following way:

$$C_2 = \left\{ c_2 = \begin{bmatrix} c_1^{(u)} \\ c_1^{(l)} \end{bmatrix} : c_1^{(u)}, c_1^{(l)} \in C_1 \right\},$$

where $c_1^{(u)}$ and $c_1^{(l)}$ are row vectors. There are $2^{nR}$ codewords in $\mathcal{C}_1$, and hence we can have $2^{nR} \cdot 2^{nR} = 2^{2nR}$ codewords in $\mathcal{C}_2$ if we choose the two components independently. Thus the rate of $\mathcal{C}_2$ is $(\log 2^{2nR})/n = 2R$. The receiver observes the sequence $y$ of length $2 \cdot n$, which consists of the $n$-length sequences $y_1$ and $y_2$ of the two parallel channels with law $W_1$, i.e.,

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

The decoder for the channel $W_2$ divides $y$ into $y_1$ and $y_2$ and decodes these sequences independently using the decoder for channel $W_1$. Since the rate $R$ is achievable on $W_1$ both $y_1$ and $y_2$ can be decoded with a probability of error which tends to 0 as $n \to \infty$ and therefore the probability of decoding an error from $y$ on $W_2$ also tends to 0. Thus, the rate $2R$ is achievable.
b) The main idea lies in making two consecutive uses of the same channel with law $W_1$ instead of transmitting over two parallel independent channels of law $W_1$. As we are using the channel twice as often as before, the rate is going to be divided by two.

Consider a sequence of rate-$R$ codes $\{C_2\}_n$ of length $n$ for $W_2$:

$$C_2 = \{c_2 = [c_{21}, c_{22}, \ldots, c_{2n}] : c_{2i} \in \mathcal{X} \times \mathcal{X}, \forall i \in \{1, \ldots, n\}\}$$

and corresponding decoding functions $\phi_{2,n} : (\mathcal{Y} \times \mathcal{Y})^n \to \{1, 2, \ldots, 2^{nR}\}$ such that the error probability $P_e$ of $C_2$ tends to zero as $n \to \infty$. As in Part a) we can think of every codeword $c_2 \in C_2$ as a $2 \times n$ matrix with entries in $\mathcal{X}$ and denote the row containing the first component of each symbol as $c_2^{(u)}$ and the row containing the second component as $c_2^{(l)}$.

We construct a sequence of rate-$\frac{R}{2}$ codes $\{C_1\}_{2n}$ of length $2n$ and decoding functions $\phi_{1,2n} : \mathcal{Y}^{2n} \to \{1, 2, \ldots, 2^{nR}\}$ by choosing every codeword in $C_1$ to correspond to a codeword $c_2$ in $C_2$, i.e.,

$$C_1 = \{c_1 = [c_2^{(u)} || c_2^{(l)}] : c_2 \in C_2\},$$

and using the same decoder as for the code $C_2$

$$\phi_{1,2n}(y_1||y_2) = \phi_{2,n}(y_1 \times y_2).$$

Then the maximum error probability $P_e$ of $C_1$ tends to zero as $n \to \infty$, as the probability of an error is the same as for the code $C_2$. Thus we achieve the rate $R_1 = \frac{nR}{2n} = \frac{R}{2}$ on the channel $W_1$.

**Problem 2**  
**An Additive Noise Channel**

The channel output is $Y = X + Z$, where $X \in \{0, 1\}$ and $Z \in \{0, a\}$. We must distinguish various cases depending on the value of $a$:

- $a = 0$: In this case $Y = X$, and therefore

$$C = \max_{P_X(\cdot)} I(X; Y) = \max_{P_X(\cdot)} H(X) = 1 \text{ bit},$$

which is achieved by a uniform distribution $P_X(\cdot)$. Hence the capacity is 1 bit per transmission.

- $a \notin \{0, \pm 1\}$: In this case $Y$ has four possible values. If $Y$ is 0 or $a$, we know that $X = 0$. If $Y$ is 1 or $1 + a$, we know that $X = 1$. Hence $H(X|Y) = 0$, and therefore

$$C = \max_{P_X(\cdot)} I(X; Y) = \max_{P_X(\cdot)} H(X) = 1 \text{ bit},$$

which is achieved by a uniform distribution on the input $X$.

- $a = 1$: In this case $Y$ has three possible output values: 0, 1, and 2. The channel looks as follows:
One sees that the channel is equivalent to a binary erasure channel with erasure probability $\alpha = \frac{1}{2}$. The capacity of the binary erasure channel is $C = 1 - \alpha = \frac{1}{2}$ bit per transmission, which is achieved by a uniform distribution on the input $X$.

- $a = -1$: This is similar to the case when $a = 1$: $Y$ also can take on three different values: $-1$, $0$, and $1$, where now $0$ is the “erasure” output. We again have a BEC and the capacity is also $C = \frac{1}{2}$ bit per transmission, achieved by a uniform distribution.

### Problem 3

**Z-Channel**

Remember that the Z-channel looks as shown in Figure 1. First we express $I(X;Y)$, the mutual information between the input and output of the Z-channel, as a function of $p = \Pr[X = 1]$:

\[
\begin{align*}
H(Y|X = 0) &= 0; \\
H(Y|X = 1) &= H_b\left(\frac{1}{2}\right) = 1 \text{ bit}; \\
\implies H(Y) &= \Pr[Y = 0] \cdot 0 + \Pr[Y = 1] \cdot 1 = p \text{ bits}; \\
\Pr[Y = 0] &= \frac{1}{2} \cdot \Pr[X = 1] + 1 \cdot \Pr[X = 0] = \frac{1}{2}p + 1 - p = 1 - \frac{1}{2}p; \\
\Pr[Y = 1] &= \frac{1}{2} \Pr[X = 1] = \frac{1}{2}p; \\
\implies H(Y) &= H_b\left(\frac{p}{2}\right); \\
\implies I(X;Y) &= H(Y) - H(Y|X) = H_b\left(\frac{p}{2}\right) - p \text{ bits}.
\end{align*}
\]

Since $I(X;Y) = 0$ when $p = 0$ and $p = 1$, the maximum mutual information is obtained for some value of $p$ such that $0 < p < 1$. Using elementary calculus, we determine that

\[
\frac{d}{dp} I(X;Y) = \frac{1}{2} \log_2 \frac{2 - p}{p} - 1,
\]
which is equal to zero for \( p = \frac{2}{5} \). (It is reasonable that \( \Pr[X = 1] < \frac{1}{2} \) because \( X = 1 \) is the noisy input to the channel.) So the capacity of the Z-channel in bits is \( H_b(\frac{1}{3}) - \frac{2}{5} \approx 0.722 - 0.4 = 0.322 \) bits per channel use.

**Problem 4**

**Capacity of a Sum Channel**

a) Let \( S \in \{1, \ldots, \nu\} \) be the chance variable representing the selected channel, i.e., \( S \sim q \) where \( q \) is the probability vector \((q_1, \ldots, q_\nu)\).

We have

\[
I(X; Y) = I(X, S; Y) = I(S; Y) + I(X; Y | S)
\]

(input alphabets are disjoint: \( X \) determines \( S \))

\[
= H(S) - H(S|Y) + I(X; Y | S)
\]

(chain rule)

\[
= H(S) + I(X; Y | S)
\]

(output alphabets are disjoint: \( Y \) determines \( S \))

\[
= H(q) + \sum_{i=1}^{\nu} q_i I(X; Y | S = i)
\]

\[
\leq H(q) + \sum_{i=1}^{\nu} q_i C_i.
\]

Equality in the last expression is achieved when conditional on \( S = i \), \( X \) is distributed according to \( p_i^* \), a capacity-achieving input distribution of the \( i \)th channel. The only variable in the last expression is \( q \). Thus, \( C \) is equal to the entropy of the channel selection plus the average of the channel capacities for the channel selection probabilities that achieve capacity \( C \).

Continuing with the last expression we get

\[
I(X; Y) \leq - \sum_{i=1}^{\nu} q_i \log_2 q_i + \sum_{i=1}^{\nu} q_i C_i
\]

\[
= - \sum_{i=1}^{\nu} q_i \log_2 q_i + \sum_{i=1}^{\nu} q_i \log_2 2^{C_i}
\]

\[
= - \sum_{i=1}^{\nu} q_i \log_2 \frac{q_i}{2^{C_i}}
\]

\[
= - \sum_{i=1}^{\nu} q_i \log_2 \frac{q_i/\alpha}{2^{C_i}/\alpha}
\]

(where \( \alpha \triangleq \sum_{i=1}^{\nu} 2^{C_i} \))

\[
= - \sum_{i=1}^{\nu} q_i \log_2 \frac{q_i}{t_i} + \sum_{i=1}^{\nu} q_i \log_2 \frac{t_i}{\alpha}
\]

(where \( t_i \triangleq \frac{2^{C_i}}{\alpha} \))

\[
= - D(q\|t) + \log_2 \alpha
\]

\[
\leq \log_2 \alpha
\]

(because \( D(q\|t) \geq 0 \))

\[
= \log_2 \left( \sum_{i=1}^{\nu} 2^{C_i} \right).
\]

Here the second inequality is achieved with equality if \( q = t \).

Note that this upper bound on the mutual information does not depend on the input anymore and it can be achieved for the optimal choice \( q^* = t \) and if conditional on \( S = i \), \( X \sim p_i^* \). Hence, it must be the capacity of the channel

\[
C = \max_{p} I(X; Y) = \log_2 \left( \sum_{i=1}^{\nu} 2^{C_i} \right).
\]

© Amos Lapidoth, 2014/2015
and the capacity-achieving input distribution is

\[ p^\ast (x) = t_i \cdot p_i^\ast (x) = \frac{p_i^\ast (x) \cdot 2^{C_i}}{\sum_{j=1}^{\nu} 2^{C_j}} \quad \text{if } x \in \mathcal{X}_i, \quad i = 1, \ldots, \nu, \]

where \( \mathcal{X}_i \) is the input alphabet of the \( i \)-th channel and \( p_i^\ast \) is the capacity-achieving input distribution of the \( i \)-th channel.

Just for fun, we will verify that the Karush–Kuhn–Tucker conditions hold for our input distribution \( p^\ast (x) \). Firstly, let \( x \) be fixed in such a way that \( p^\ast (x) > 0 \) and \( x \in \mathcal{X}_i \) for some \( i \). We denote by \( r^\ast (y) \) the output distribution corresponding to \( p^\ast (x) \). Then for \( x \in \mathcal{X}_i \) and \( y \in \mathcal{Y}_i \) we have \( p^\ast (x) = p_i^\ast (x) \cdot t_i \) and \( r^\ast (y) = r_i^\ast (y) \cdot t_i \), where \( r_i^\ast (y) \) is the output distribution of the \( i \)-th channel when the input distribution is \( p_i^\ast (y) \). Hence,

\[
\mathcal{D}(W(\cdot|x)||r^\ast (\cdot)) = \sum_{y \in \mathcal{Y}_i} W(y|x) \log_2 \frac{W(y|x)}{r^\ast (y)} \\
= \sum_{y \in \mathcal{Y}_i} W(y|x) \log_2 \frac{W(y|x)}{r_i^\ast (y) \cdot t_i} \\
= \sum_{y \in \mathcal{Y}_i} W(y|x) \log_2 \frac{W(y|x)}{r_i^\ast (y)} + \log_2 \frac{1}{t_i} \\
= C_i + \log_2 \frac{1}{t_i} \\
= C_i + \log_2 \left( \frac{\sum_{j=1}^{\nu} 2^{C_j}}{2^{C_i}} \right) \\
= C_i + \log_2 \left( \sum_{j=1}^{\nu} 2^{C_j} \right) - \log_2 2^{C_i} \\
= \log_2 \left( \sum_{j=1}^{\nu} 2^{C_j} \right) \\
= C,
\]

where (1) follows because we have fixed \( x \in \mathcal{X}_i \), which means that \( y \in \mathcal{Y}_i \); and where (2) follows from the Karush–Kuhn–Tucker conditions for the \( i \)-th channel because \( p_i^\ast (x) > 0 \) and \( p_i^\ast (\cdot) \) is the capacity-achieving input distribution for the \( i \)-th channel.

Next, let \( x \) be such that \( p^\ast (x) = 0 \) and \( x \in \mathcal{X}_i \) for some \( i \). Then the derivation up to before (2) remains the same, but then the Karush–Kuhn–Tucker condition for the \( i \)-th channel will yield an inequality because \( p_i^\ast (x) = 0 \) (and \( p_i^\ast (\cdot) \) is the capacity-achieving input distribution for the \( i \)-th channel):

\[
\mathcal{D}(W(\cdot|x)||r^\ast (\cdot)) = \sum_{y \in \mathcal{Y}_i} W(y|x) \log_2 \frac{W(y|x)}{r_i^\ast (y)} + \log_2 \frac{1}{t_i} \\
\leq C_i + \log_2 \frac{1}{t_i} \\
= C.
\]

Thus the KKT conditions are verified.

b) The capacity of the BSC is \( 1 - H_b(\epsilon) \) bits and the capacity of the other channel is zero. Thus by a) the capacity of the sum channel is

\[ C = \log \left( 2^{1-H_b(\epsilon)} + 2^0 \right) = \log \left( 1 + 2^{1-H_b(\epsilon)} \right). \]