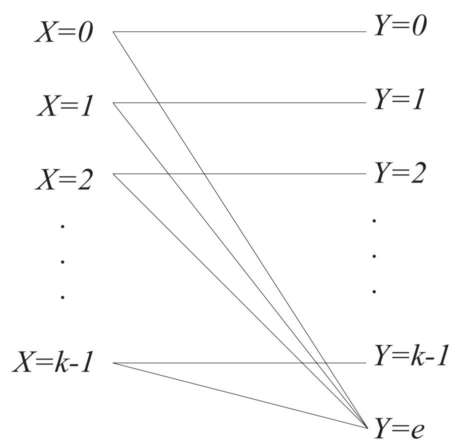
**16:332:542** **Information Theory and Coding** **Final Examination**

May 8, 2002

**SOLUTION**

This is an 180 minute exam. Please answer the following four questions in the notebooks provided. You are permitted to look at the Cover& Thomas text but not other materials. Make sure that you have included your name, ID number (last 4 digits only) and signature in each book used (*5 points*). Read each question carefully. All statements must be justified. Computations should be simplified as much as possible.

1. *20 points* Consider the following $k$ input and $k+1$ output discrete erasure channel:




(a) For a given input distribution $p(x)$, what is the mutual information $I(X;Y)$? (Express your answer in terms of $H(X)$)

---

We can solve this directly via

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) - H(X|Y=e)P\{Y=e\} - \sum_{y\neq e} H(X|Y=y)P\{Y=y\}. \end{aligned}$$

The event $Y=e$ tells us nothing about $X$ since

$$P\{X=x|Y=e\} = \frac{P\{Y=e|X=x\}P\{X=x\}}{P\{Y=e\}} = P\{X=x\}.$$

Thus, $H(X|Y=e) = H(X)$. Further, given $Y=y\neq e$, we know exactly what $X$ was sent across the channel. Hence, for $y\neq e$, $H(X|Y=y)=0$. Combining these facts,

$$I(X;Y) = H(X) - H(X)P\{Y=e\} = (1-\epsilon)H(X).$$

---

(b) Define the random variable $E$ as

$$E = \begin{cases} 1 & Y=e, \\ 0 & \text{otherwise}. \end{cases}$$

What are $H(Y,E)$ and $H(Y,E|X)$?

---

Since $E$ is a deterministic function of $Y$, $H(E|Y)=0$. Thus,

$$H(Y,E) = H(Y) + H(E|Y) = H(Y).$$

Furthermore,

$$\begin{aligned} H(Y, E) &= H(E) + H(Y|E) \\ &= H(\epsilon, 1-\epsilon) + H(Y|E=1)P\{E=1\} + H(Y|E=0)P\{E=0\}. \end{aligned}$$

Since $Y = e$ when $E = 1$, $H(Y|E = 1) = 0$. Also, when $E = 0$, $Y = X$, so $H(Y|E = 0) = H(X)$. Thus,

$$H(Y) = H(Y, E) = H(\epsilon, 1-\epsilon) + H(X)(1-\epsilon).$$

For the same reason, $H(E|X, Y) = 0$ and

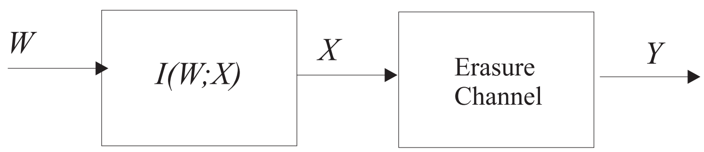
$$H(Y, E|X) = H(E|X) + H(E|X, Y) = H(E|X).$$

Since $E$ is a Bernoulli $\epsilon$ random variable for any $X = x$.

$$H(Y, E|X) = H(E|X) = \sum_x p(x)H(E|X = x) = \sum_x p(x)H(\epsilon, 1-\epsilon) = H(\epsilon, 1-\epsilon).$$

---

(c) Suppose an arbitrary $j$ input, $k$ output channel from $W$ to $X$ is followed in cascade by the $X, Y$ erasure channel from part (a) as follows:




What is $I(W; Y)$? Your answer should be expressed in terms of $I(W; X)$. Hint: consider the auxiliary random variable $E$.

---

We know that $I(W; Y) = H(Y) - H(Y|W)$ and $H(Y) = (1-\epsilon)H(X)$. Since $E$ is a deterministic function of $Y$,

$$H(Y|W) = H(Y, E|W) = H(E|W) + H(Y|W, E).$$

For any event $W = w$, $E$ is a Bernoulli $(\epsilon)$ random variable so $H(E|W = w) = H(\epsilon, 1-\epsilon)$. For the second term, we observe that given $E = 1$, $Y = e$, no matter what the value of $W$. Hence, $H(Y|W, E = 1) = 0$. On the other hand, if $E = 0$, then $Y = X$ and $H(Y|W, E = 0) = H(X|W)$. It follows that

$$H(Y|W, E) = P\{E = 0\}H(Y|W, E = 0) + P\{E = 1\}H(Y|W, E = 1) \quad = (1-\epsilon)H(X|W).$$

and

$$H(Y|W) = H(E|W) + H(Y|W, E) = H(\epsilon, 1-\epsilon) + (1-\epsilon)H(X|W).$$

Finally,

$$\begin{aligned} I(W; Y) &= H(Y) - H(Y|W) \\ &= H(\epsilon, 1-\epsilon) + H(X)(1-\epsilon) - [H(\epsilon, 1-\epsilon) + (1-\epsilon)H(X|W)] \\ &= (1-\epsilon)I(X; W). \end{aligned}$$

---

2. Consider a channel consisting of two parallel AWGN channels with inputs $X_1, X_2$ and outputs

$$\begin{aligned} Y_1 &= X_1 + Z_1, \\ Y_2 &= X_2 + Z_2. \end{aligned}$$

The noises $Z_1$ and $Z_2$ are independent and have variances $N_1$ and $N_2$ with $N_1 < N_2$. However, we are constrained to use the same symbol on both channels, i.e. $X_1 = X_2 = X$, where $X$ is constrained to have power $E\left[X^2\right] = P$.

(a) Suppose at the receiver, we combine the outputs to produce $Y = Y_1 + Y_2$? What is the capacity $C_1$ of channel with input $X$ and output $Y$? What type of signaling achieves this capacity?

---

This was easy,

$$Y = Y_1 + Y_2 = X_1 + X_2 + (Z_1 + Z_2) = 2X + Z,$$

where $Z = Z_1 + Z_2$. This is an AWGN channel with received power $E\left[(2X)^2\right] = 4P$ and noise power $\sigma_Z^2 = \sigma_{Z_1}^2 + \sigma_{Z_2}^2 = N_1 + N_2$. The capacity of the $X$ to $Y$ channel is the Shannon capacity

$$C_1 = \frac{1}{2} \log \left( 1 + \frac{4P}{N_1 + N_2} \right).$$

Any rate $R < C_1$ can be achieved using a $(n, 2^{nR})$ Gaussian codebook for sufficiently large $n$. The fact that the received power is $4P$ rather than $2P$ may seem surprising. The typical case where $X_1 + X_2$ has power $2P$ occurs when $X_1$ and $X_2$ are uncorrelated. When $X_1 = X_2 = X$, then $X_1 + X_2$ has power $4P$. When transmitting the same signal over two channels in a communications system, this benefit is realized only when the two copies of the received signal are combined coherently.

---

(b) Suppose the receiver can view both outputs $Y_1$ and $Y_2$. What is the capacity $C_2$ of this system? Does the optimal signaling change from part (a)?

---

In this case, we have a vector channel $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$ of the form

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} X + \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}.$$

We can follow the derivation of mutual information for the colored Gaussian noise channel, except here the inputs $(X_1, X_2)$ are completely correlated since $X_1 = X_2 = X$ and the noises are uncorrelated. In particular $(X_1, X_2)$ have covariance matrix

$$\mathbf{K}_X = E\left[\mathbf{X}\mathbf{X}'\right] = E\left[\begin{bmatrix} 1 \\ 1 \end{bmatrix} X X \begin{bmatrix} 1 & 1 \end{bmatrix}\right] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} E\left[X^2\right] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} P.$$

In addition, since noises $Z_1$ and $Z_2$ are independent, and also independent of the input $X$,

$$\mathbf{K}_Z = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix}, \qquad \mathbf{K}_Y = \mathbf{K}_X + \mathbf{K}_Z = \begin{bmatrix} P + N_1 & P \\ P & P + N_2. \end{bmatrix}$$

For the mutual information,

$$\begin{aligned} I(X; Y_1, Y_2) = I(X_1, X_2; Y_1, Y_2) &= h(Y_1, Y_2) - h(Y_1, Y_2 | X_1, X_2) \\ &= h(Y_1, Y_2) - h(Z_1, Z_2). \end{aligned}$$

The mutual information is maximized when the differential entropy $h(Y_1, Y_2)$ is maximized, which occurs when $Y_1, Y_2$ are Gaussian. In this case,

$$\begin{aligned} C_2 = I(X; Y_1, Y_2) &= \frac{1}{2} \log \left( (2\pi e)^2 |\mathbf{K}_Y| \right) - \frac{1}{2} \log \left( (2\pi e)^2 |\mathbf{K}_Z| \right) \\ &= \frac{1}{2} \log \frac{|\mathbf{K}_Y|}{|\mathbf{K}_Z|} \\ &= \frac{1}{2} \log \frac{(P + N_1)(P + N_2) - P^2}{N_1 N_2} \\ &= \frac{1}{2} \log \left( 1 + \frac{P(N_1 + N_1)}{N_1 N_2} \right) \\ &= \frac{1}{2} \log \left( 1 + \frac{P}{N_1} + \frac{P}{N_2} \right). \end{aligned}$$

Because $C_1$ is derived by the processing $Y_1+Y_2$, we should be able to show that $C_1 < C_2$. Equivalently, if we write $C_1 = (1/2)\log(1+\gamma_1)$ and $C_2 = (1/2)\log(1+\gamma_2)$, then we should be able to show $\gamma_2 > \gamma_1$. In fact,

$$\gamma_2 - \gamma_1 = \frac{P}{N_1} + \frac{P}{N_2} - \frac{4P}{N_1 + N_2} = \frac{P(N_1 - N_2)^2}{N_1 N_2 (N_1 + N_2)}$$

After the next part, it won't be surprising that $N_1 = N_2$ implies $C_1 = C_2$.

---

(c) Suppose the receiver must combine the two received signals to produce $Y' = \alpha Y_1 + (1-\alpha)Y_2$ where $0 \le \alpha \le 1$. However, as the receiver designer, you can choose the best $\alpha$ for combining. What is the capacity $C'$ of this system with input $X$ and output $Y'$? Is there a loss in capacity relative to $C_2$?

---

We observe that for $X_1 = X_2 = X$, the output $Y'$ is

$$\begin{aligned} Y' &= \alpha Y_1 + (1-\alpha)Y_2 \\ &= \alpha(X + Z_1) + (1-\alpha)(X + Z_2) \\ &= X + \alpha Z_1 + (1-\alpha) Z_2 \\ &= X + Z. \end{aligned}$$

Hence, we have an AWGN channel with signal power $E\left[X^2\right] = P$ and noise power

$$N = E\left[Z^2\right] = \sigma_Z^2 = \alpha^2 \sigma_{Z_1}^2 + (1-\alpha)^2 \sigma_{Z_2}^2 = \alpha^2 N_1 + (1-\alpha)^2 N_2.$$

The capacity of the system is maximized by choosing $\alpha = \alpha^*$ so that $N$ is minimized. Setting $dN/d\alpha|_{\alpha=\alpha^*} = 0$ yields

$$2\alpha^* N_1 - 2(1-\alpha^*)N_2 = 0 \qquad \Longrightarrow \qquad \alpha^* = \frac{N_2}{N_1 + N_2}.$$

With the optimal $\alpha$, the noise variance is

$$N = \left(\frac{N_2}{N_1 + N_2}\right)^2 N_1 + \left(\frac{N_1}{N_1 + N_2}\right)^2 N_2 = \frac{N_1 N_2}{N_1 + N_2}.$$

Lastly, using the optimal $\alpha$, the capacity of the system is

$$C' = \frac{1}{2}\log\left(1 + \frac{P}{N}\right) = \frac{1}{2}\log\left(1 + \frac{P}{N_1} + \frac{P}{N_2}\right).$$

We see that $C' = C_2$. That is, there is no loss in capacity by using optimal linear combining of the two signals. The reason for this is because in both case, the two signals are constrained to carry the same information signal. If we were allowed to transmit independent $X_1$ and $X_2$, linear combining would result in a significant loss in capacity.

---

(d) Suppose the transmitter, is still constrained to transmit the same signal on both channels, but can choose how much power to use on each channel. That is, for constants $a$ and $b$, $X_1 = aX$ and $X_2 = bX$. Subject to a constraint that the total transmitted power is bounded by $2P$, what are the optimal $a$ and $b$ and corresponding capacity $C'$ of the system with outputs $(Y_1, Y_2)$?

---

In this case, the constraint on total transmitted power is that $E\left[X_1^2 + X_2^2\right] = a^2 P + b^2 P \le 2P$ or $a^2 + b^2 \le 2$. In this case, it is helpful to view this as a vector system $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$ where

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} X + \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}.$$

The input has covariance

$$\mathbf{K}_X = E[\mathbf{X}\mathbf{X}'] = E\left[\begin{bmatrix} a \\ b \end{bmatrix} XX \begin{bmatrix} a & b \end{bmatrix}\right] = \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix} P.$$

As always, since $\mathbf{Z}$ has independent components,

$$\mathbf{K}_Z = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix}, \qquad \mathbf{K}_Y = \mathbf{K}_X + \mathbf{K}_Z = \begin{bmatrix} a^2P + N_1 & abP \\ abP & b^2P + N_2 \end{bmatrix}.$$

As in part (b), capacity is acheived with Gaussian signaling, yielding In this case,

$$\begin{aligned} C'' = I(X_1, X_2; Y_1, Y_2) &= \frac{1}{2}\log\frac{|\mathbf{K}_Y|}{|\mathbf{K}_Z|} \\ &= \frac{1}{2}\log\frac{(a^2P + N_1)(b^2P + N_2) - (abP)^2}{N_1N_2} \\ &= \frac{1}{2}\log\left(1 + a^2\frac{P}{N_1} + b^2\frac{P}{N_2}\right). \end{aligned}$$

Hence, we want to maximize

$$S = a^2\frac{P}{N_1} + b^2\frac{P}{N_2}$$

subject to $a^2 + b^2 = 2$. Making the substitution $b^2 = 2 - a^2$, we have

$$S = \frac{2P}{N_2} + a^2\left(\frac{P}{N_1} - \frac{P}{N_2}\right)$$

where we must have $0 \le a^2 \le 2$. Since $N_1 < N_2$, $P/N_1 > P/N_2$, implying we want to choose $a^2$ as large as possible. In this case, $a^2 = 2$ and $b^2 = 0$, yielding $S = 2P/N_1$ and

$$C'' = \frac{1}{2}\log\left(1 + \frac{2P}{N_1}\right).$$

We see that $C'' > C_2$. When we are constrained to use the same signal in two orthogonal channels, its best to put all the power into the channel with less noise.

(e) **Added during the exam** Can a higher capacity be achieved using an arbitary linear combination $\hat{Y} = aY_1 + bY_2$ rather than (as we did in part (c)) $a = \lambda$, $b = 1 - \lambda$, with $0 \le \lambda \le 1$.

For the optimal $a$ and $b$, let $\hat{Y} = aY_1 + bY_2$. Similarly, for the optimal $\lambda$, let $Y' = \lambda Y_1 + (1-\lambda)Y_2$. By the data processing theorem, $I(X; Y_1, Y_2) \ge I(X; \hat{Y})$. Also, sicne we could have chosen $a = \lambda$ and $b = 1 - \lambda$, using the optimal $a$ and $b$ implies $I(X; \hat{Y}) \ge I(X; Y')$. Thus

$$I(X; Y_1, Y_2) \ge I(X; \hat{Y}) \ge I(X; Y')$$

Finally, in parts (b) and (c), we showed $I(X; Y_1, Y_2) = I(X; Y')$. Hence, there is no advantage in choosing arbitrary $a$ and $b$.

To understand why this is true, for any $a$ and $b$, we can write

$$\hat{Y} = aY_1 + bY_2 = (a+b)X + aZ_1 + bZ_2 = (a+b)[X + \frac{a}{a+b}Z_1 + \frac{b}{a+b}Z_2].$$

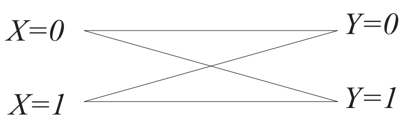
We note that if $a + b = 0$, then $\hat{Y}$ contains no signal components, which is clearly suboptimal. When $a + b \ne 0$, we can define $\lambda = a/(a+b)$ so that

$$\hat{Y} = (a+b)[X + \lambda Z_1 + (1-\lambda)Z_2] = (a+b)\tilde{Y}$$

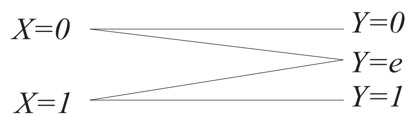
where $\tilde{Y} = X + \lambda Z_1 + (1-\lambda)Z_2$. Since amplifying the received signal by a constant cannot increase capacity, $I(X;\hat{Y}) = I(X;\tilde{Y})$. Perhaps the only remaining question is why the constraint $0 \le \lambda \le 1$ does not limit capacity. The reason is that for $\lambda < 0$, or $\lambda > 1$ and $1-\lambda < 0$, we are actually throwing away signal power and then renormalizing the received power; the result is enhanced noise power.

---

3. Consider the binary symmetric channel and the binary erasure channel shown below:




BSC




BEC

(a) Find the capacity $C_{\text{BSC}}(\epsilon)$ of the BSC and $C_E(\delta)$ of the erasure channel.

---

I forgot that the exam was open book when I tossed in this part. Straight from the text,

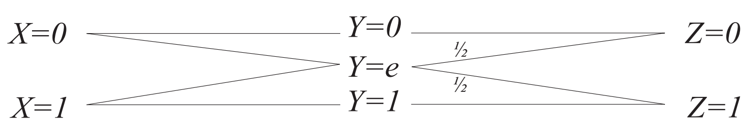
$$C_{\text{BSC}}(\epsilon) = 1 - H(\epsilon), \qquad C_E(\delta) = 1 - \delta.$$

---

(b) When $\delta = \epsilon$, use the data processing theorem to prove that the BEC has higher capacity than the BSC.

---

Perhaps I should have given a better hint for this. The idea of the hint was that the data processing inequality says that the capacity of a cascade of two channels is less than capacity of the first channel. If the first channel is the BEC, we can cascade a second channel to create a composite BSC. then the BSC has lower capacity. The cascade structure




is a BSC with crossover probability $\delta/2$. For this cascade, the data processing inequality says $I(X;Z) \le I(X;Y)$. If $I(X;Z)$ is maximized by input distribution $p_1(x)$ while $p_2(x)$ maximizes $I(X;Y)$, then

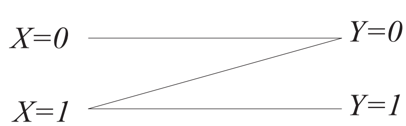
$$C_{\text{BSC}}(\delta/2) = I(X;Z)|_{p(x)=p_1(x)} \le I(X;Y)|_{p(x)=p_1(x)} \le I(X;Y)|_{p(x)=p_2(x)} = C_E(\delta).$$

Finally, for small $\delta$, we know that

$$C_{\text{BSC}}(\delta) \le C_{\text{BSC}}(\delta/2) \le C_E(\delta).$$

---

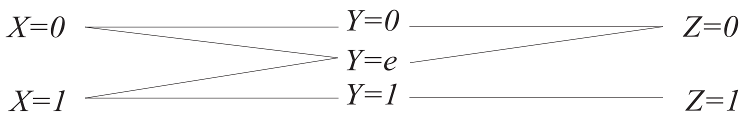
(c) Consider the $Z$-channel:




Use the data processing theorem to find an upper bound and a lower bound to the capacity $C_Z(\alpha)$ of the $z$ channel. Express these bounds in terms of $C_{\text{BSC}}(\cdot)$ and $C_E(\cdot)$.

---

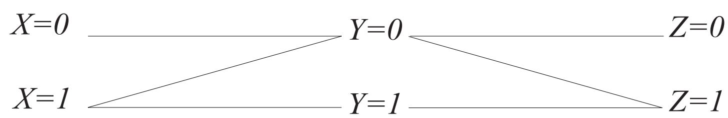
To find an upper bound to capacity of the $Z$-channel, consider the following concatentation to the binary erasure channel:




The $X \to Z$ channel is a $Z$-channel with crossover probability $\delta$. For this cascade, the dataprocessing inequality says $I(X;Z) \le I(X;Y)$ for any input distribution. If $I(X;Z)$ is maximized by input distribution $p_1(x)$ while $p_2(x)$ maximizes $I(X;Y)$, then

$$C_Z(\delta) = I(X;Z)|_{p(x)=p_1(x)} \le I(X;Y)|_{p(x)=p_1(x)} \le I(X;Y)|_{p(x)=p_2(x)} = C_E(\delta).$$

For a lower bound to the capacity of the $Z$ channel, consider the following concatenation of two $Z$-channels




By choosing $\beta = \alpha/(1+\alpha)$, the concatenation is a BSC with crossover probability $\epsilon = \alpha/(1+\alpha)$. The $X \to Z$ channel is a BSC with crossover probability $\alpha/(1+\alpha)$. For this cascade, the dataprocessing inequality says $I(X;Z) \le I(X;Y)$ for any input distribution. If $I(X;Z)$ is maximized by input distribution $p_1(x)$ while $p_2(x)$ maximizes $I(X;Y)$, then

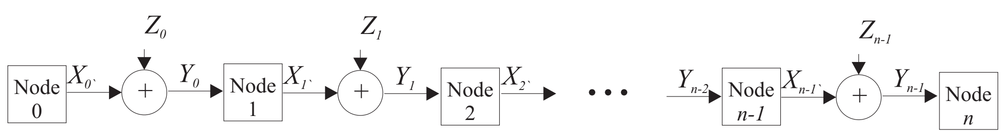
$$C_{\mathrm{BSC}}(\alpha/(1+\alpha)) = I(X;Z)|_{p(x)=p_1(x)} \le I(X;Y)|_{p(x)=p_1(x)} \le I(X;Y)|_{p(x)=p_2(x)} = C_Z(\alpha).$$

Putting the two bounds together, we have

$$C_{\mathrm{BSC}}(\alpha/(1+\alpha)) \le C_Z(\alpha) \le C_E(\alpha) = 1-\alpha.$$

---

4. Suppose we have a wireless network with $n$ hops. For $i = 0, 1, \ldots, n-1$, node $i$ transmits coded messages at rate $R_i$ to node $i+1$ over an AWGN channel with noise variance $N_i$:




Assume each node transmits in an orthogonal channel. Node $i$ decodes messages the transmission of node $i-1$ and forwards to node $i+1$. Note that the nodes may or may not use different coding strategies. Node $i$ transmits at power $P_i$ and the multihop system is subject to the constraint $\sum_{i=0}^{n-1} P_i = P$.

(a) In terms of $P_i$ and $N_i$, what is the capacity $C_i$ of the channel $i$ from node $i$ to node $i+1$? (Yes, this is a gift.)

---

Each channel $i$ is just an AWGN channel with capacity

$$C_i = \frac{1}{2}\log\left(1 + \frac{P_i}{N_i}\right).$$

---

(b) For a given set of powers $P_0, \ldots, P_n$, what is the capacity of $C$ of the multihop communication system from node 0 to node $n$? Express your answer in terms of $C_i$. Explain your answer in terms of the end-to-end data rate $R$.

---

On channel $i$, reliable communication at rate $R$ is possible iff $R < C_i$. Hence, end-to-end reliable communication is possible iff $R < C = \min_i C_i$.

---

(c) What is the optimal power allocation $P_0, \ldots, P_{n-1}$? What is the corresponding channel capacity $C$?

---

For a given set of powers $\mathbf{P} = (P_0, \ldots, P_{n-1})$,

$$C(\mathbf{P}) = \min_i \frac{1}{2} \log \left(1 + \frac{P_i}{N_i}\right) = \frac{1}{2} \log \left(1 + \min_i \frac{P_i}{N_i}\right).$$

Let $\gamma = \min_i P_i/N_i$. The optimal solution has

$$\gamma = \frac{P_0}{N_0} = \frac{P_1}{N_1} = \cdots = \frac{P_{n-1}}{N_{n-1}}$$

To see this, suppose there exists $j$ such that $P_j/N_j > 0$. In that case, we can decrease $P_j$ by $\epsilon > 0$ and then we can raise the power by some $\delta > 0$ on all channels $i$ such that $P_i/N_i = \gamma$. This will increase capacity. This implies $P_i = \gamma N_i$. The constraint $\sum_{i=0}^{n-1} E = P$ implies

$$\sum_{i=1}^{n-1} \gamma N_i = P \quad \Longrightarrow \quad \gamma = \frac{P}{N_0 + \cdots + N_{n-1}}$$

Finally,

$$C = \frac{1}{2} \log(1 + \gamma) = \frac{1}{2} \log \left(1 + \frac{P}{N_0 + N_1 + \cdots + N_{n-1}}\right).$$

---