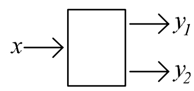
6.441 Transmission of Information May 9, 2006

Lecture 22

Lecturer: Madhu Sudan *Scribe: Jonathan Rohrs*

1 Broadcast Channels



This channel broadcasts messages to two receivers, and is characterized by the joint (marginal) probability mass function:

$$P_{(y_1, y_2 | x)}$$

and alphabets for the inputs and two outputs.

1.1 Independent Information

Suppose one wishes to use such a channel to transmit some message ($W_1 \in \{1, \ldots, 2^{nR_1}\}$) to one receiver, and some other message ($W_2 \in \{1, \ldots, 2^{nR_2}\}$) to the other. This is done as follows

1. An encoder (E) is used to encode the two messages as one word:

$$(W_1, W_2) \xrightarrow{E} X^n$$

2. The channel (Ch) produces its two outputs words according to is probability distribution ($P_{(y_1, y_2|x)}$)

$$X^n \xrightarrow{Ch} (Y_1^n, Y_2^n)$$

3. Two decoders (D_1 and D_2) are used to decode the respective output words, producing the decoded messages ($\hat{W}_1$ and $\hat{W}_2$)

$$Y_1^n \xrightarrow{D_1} \hat{W}_1$$

$$Y_2^n \xrightarrow{D_2} \hat{W}_2$$

The two receivers are not allowed to collude in decoding their messages.

1.2 Achievable Transmission Rates

As usual, we are interested in the possible transmission rates. We say that the *rate pair* (R_1, R_2) is *achievable* if there exists an encoder (E) and two decoders (D_1 and D_2) such that for any two messages (W_1 and W_2), the probability of decoding error goes to zero with n. That is,

$$P_{err} = \Pr(\,(W_1, W_2) \neq (\hat{W}_1, \hat{W}_2)\,) \longrightarrow 0$$

Unfortunately, useful results for this general case do not exist. Instead, we investigate special cases, namely degraded channels.

1.3 Stochastic Equivalence

We say two broadcast channels are *stochastically equivalent*

$$P_{(y_1,y_2)|x} =_{stoc} P'_{(y_1,y_2)|x}$$

if they share the same marginal distributions:

$$P_{y_1|x} =_{stoc} P'_{y_1|x}$$

$$P_{y_2|x} =_{stoc} P'_{y_2|x}$$

Achievable rate pairs depend only on the marginal distribution of the channel. This means that if two channels are stochastically equivalent:

$$P =_{stoc} P'$$

then they share the same set of achievable rate pairs:

$$(R_1, R_2) \text{ is achievable by } P \Longleftrightarrow (R_1, R_2) \text{ is achievable by } P'$$

This can be useful, in that it sometimes allows one to find rate results for a broadcast channel by studying another, simpler but equivalent, channel instead.

1.4 Degraded Broadcast Channels

Degraded channels are a special case of broadcast channels. We study them in large part because they give clean results where results for general broadcast channels have not been found.

1.4.1 Definitions

Degraded broadcast channels come in two types:

- *Physically* degraded broadcast channels are those that form a Markov chain:

  $$X \to Y_1 \to Y_2$$

  It is as if the channel degrades the signal between the source and the first receiver, and then degrades it some more before the second receiver.

- *Stochastically* degraded broadcast channels are those that are stochastically equivalent to a physically degraded broadcast channel.

  The definitions of stochastic equivalence and physically degraded channels (together) imply that a channel (P) is stochastically degraded if there exists a distribution (P') such that

  $$p(y_2|x) = \sum_{y_1} p(y_1|x) \cdot p'(y_2|y_1)$$

Typically, one will prove rate results in terms of physically degraded channels, but the same results will apply to stochastically degraded channels, since they share the same set of achievable rate pairs.

2 Coding Theorem for Degraded Channel

Assume (without loss of generality) that we are given a physically degraded channel.

2.1 The Theorem

$\exists U$ s.t. $U \to X \to Y_1 \to Y_2$ and

$$\left.\begin{array}{c} R_2 \le I(U;Y_2) \\ \text{and} \\ R_1 \le I(X;Y_1|U) \end{array}\right\} \Longrightarrow (R_1, R_2) \text{ is achievable}$$

2.2 The Converse

$$(R_1, R_2) \text{ is achievable } \Longrightarrow \left\{\begin{array}{c} \exists U \text{ s.t. } U \to X \to Y_1 \to Y_2 \\ \text{and} \\ R_2 \le I(U;Y_2) \\ \text{and} \\ R_1 \le I(X;Y_1|U) \end{array}\right.$$

Furthermore, U takes values from a set of size

$$|\Omega_U| \le \min(\,|\Omega_X|,\, |\Omega_{Y_1}|,\, |\Omega_{Y_2}|\,)$$

2.3 Theorem Proof

Given some (W_1, W_2), use the following scheme for encoding:

1. Map $W_2 \to U^n = (U_1, U_2, \ldots, U_n)$ s.t.

   $$U_i \sim P_U$$

   independent for each W_2 and i.

2. Map $(U^n, W_1) \to X^n = (X_1, X_2, \ldots, X_n)$ s.t.

   $$X_i \sim P_{X|U}$$

   independent for each W_1 and i.

Then use this scheme for decoding:

- Receiver 2: If there exists a unique $\hat{W}_2$ s.t.

  $$(\,U(\hat{W}_2),\, Y_2^n\,) \text{ is jointly typical}$$

  then output $\hat{W}_2$. Otherwise `error`.

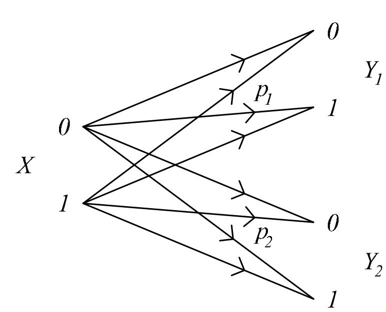
- Receiver 1: If there exists a unique $(\hat{W}_1, \hat{W}_2)$ s.t.

  $$(\,U(\hat{W}_2),\, X(U(\hat{W}_2), \hat{W}_1)\,,\, Y_1^n\,) \text{ is jointly typical}$$

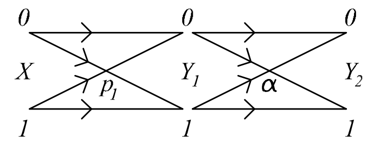
  then output $\hat{W}_1$. Otherwise `error`.

3 Example

Consider a pair of symmetric binary channels with bit-flip parameters p_1 and p_2, forming a broadcast channel as shown below:



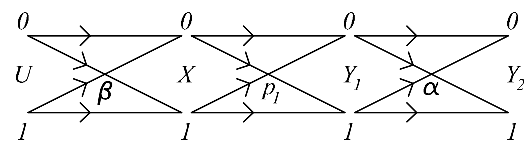
Since the bit-flips are independent, this does *not* form a physically degraded channel. However, it is stochastically equivalent to the physically degraded channel:



if α is chosen so that

$$p_2 = p_1 \star \alpha = p_1(1-\alpha) + \alpha(1-p_1)$$

For the purposes of applying the coding theorem for degraded channels, add an additional variable $U \sim Bern(\frac{1}{2})$ and a third symmetric binary channel stage (with parameter β):



In summary, alternately stated (using modulo 2 arithmetic),

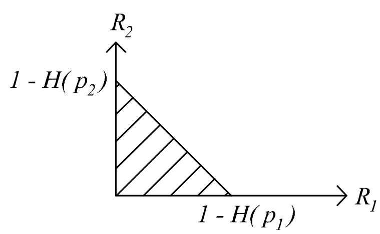
$$\begin{array}{lll} U \sim Bern(\frac{1}{2}) & & \\ X = U + Z_1 & \text{where} & Z_1 \sim Bern(\beta) \\ Y_1 = X + Z_2 & \text{where} & Z_2 \sim Bern(p_1) \\ Y_2 = Y_1 + Z_3 & \text{where} & Z_3 \sim Bern(\alpha) \end{array}$$

Applying the theorem: For fixed β, p_1, and p_2, we can achieve rates

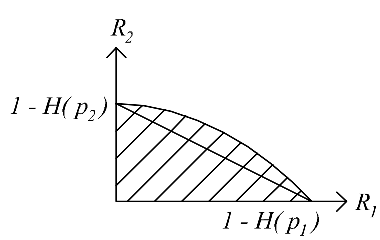
$$\begin{array}{l} R_1 = H(\beta \star p_1) - H(p_1) \\ R_2 = 1 - H(\beta \star p_2) \end{array}$$

If $p_1 = p_2$, then

$$R_2 = 1 - H(p_1) - R_1$$

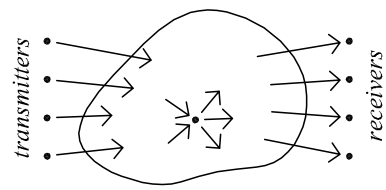


This is equivalent to time-sharing. By contrast, if $p_1 < p_2$ (and $R_1, R_2 \neq 0$), one can gain compared to time-sharing.



4 Network Information Theory

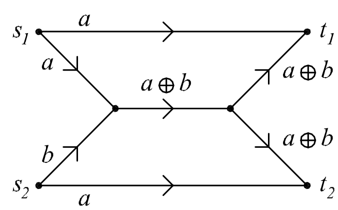
In general, a network may have many transmitters, receivers, and nodes (which both transmit and receive):



nodes: receive and transmit

This may operate in a number of ways.

- *Store and Forward*: Nodes are not allowed to do calculations. Network capacity can be found in exponential time.
- *Recompute and Redistribute*: Nodes can compute as well. This may allow for an increase in capacity. For example:



If each edge has capacity of 1, then the network shown can achieve a rate pair $(1, 1)$. If computation was not allowed (i.e. node could not calculate $a \oplus b$), one could only achieve $(\frac{1}{2}, \frac{1}{2})$.

Network information theory is an active area of research. See Li and Yeung for more information on what is known and what is still open.