Information Theory

Lecture 10

- Network Information Theory (CT15); a focus on channel capacity results
  - The (two-user) multiple access channel (15.3)
  - The (two-user) broadcast channel (15.6)
  - The relay channel (15.7)
  - Some remarks on general multiterminal channels (15.10)

Joint Typicality

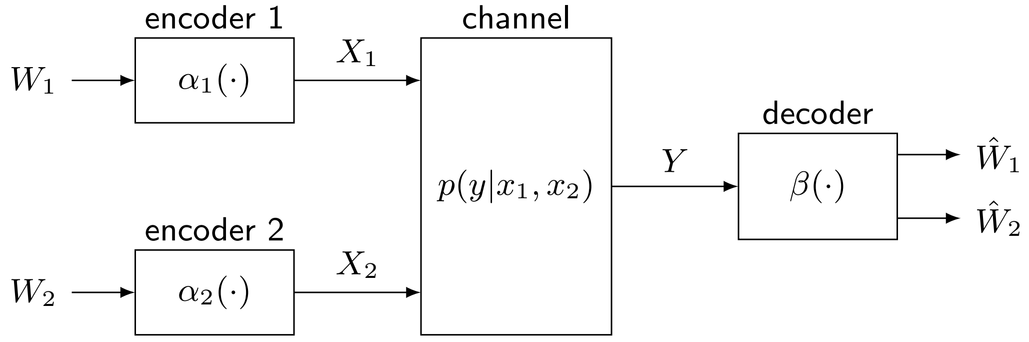
- Extension of previous results to an arbitrary number of variables (most basic defs here, many additional results in CT)
- Notation
  - For any k-tuple $x_1^k = (x_1, x_2, \dots, x_k) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k$ and subset of indices $\mathcal{S} \subseteq \{1, 2, \dots, k\}$ let $x_\mathcal{S} = (x_i)_{i \in \mathcal{S}}$
  - Assume $\mathbf{x}_i \in \mathcal{X}_i^n$, any i, and let $\mathbf{x}_\mathcal{S}$ be a matrix with $\mathbf{x}_i$ as rows for $i \in \mathcal{S}$. Let the $|\mathcal{S}|$-tuple $\mathbf{x}_{\mathcal{S},j}$ be the jth column of $\mathbf{x}_\mathcal{S}$.
  - As in CT, $a_n \doteq 2^{n(c \pm \varepsilon)}$ means

$$\left| \frac{1}{n} \log a_n - c \right| < \varepsilon,$$

for all sufficiently large n

- For random variables X_1^k with joint distribution $p(x_1^k)$:
  Generate $\mathbf{X}_\mathcal{S}$ via n independent copies of $\mathbf{x}_{\mathcal{S},j}$, $j = 1, \ldots, n$. Then,
$$\Pr(\mathbf{X}_\mathcal{S} = \mathbf{x}_\mathcal{S}) = \prod_{j=1}^{n} p(\mathbf{x}_{\mathcal{S},j}) \triangleq p(\mathbf{x}_\mathcal{S})$$
- For $\mathcal{S} \subseteq \{1, 2, \ldots, k\}$, define the set of ε-typical n-sequences
$$A_\varepsilon^{(n)}(\mathcal{S}) = \left\{ \mathbf{x}_\mathcal{S} : \Pr(\mathbf{X}_{\mathcal{S}'} = \mathbf{x}_{\mathcal{S}'}) \doteq 2^{-n[H(\mathbf{X}_{\mathcal{S}'}) \pm \varepsilon]}, \ \forall \mathcal{S}' \subseteq \mathcal{S} \right\}$$
- Then, for any $\varepsilon > 0$, sufficiently large n, and $\mathcal{S} \subseteq \{1, \ldots, k\}$,
$$P(A_\varepsilon^{(n)}(\mathcal{S})) \geq 1 - \varepsilon$$
$$p(\mathbf{x}_\mathcal{S}) \doteq 2^{-n[H(\mathbf{X}_\mathcal{S}) \pm \varepsilon]} \quad \text{if} \quad \mathbf{x}_\mathcal{S} \in A_\varepsilon^{(n)}(\mathcal{S})$$
$$|A_\varepsilon^{(n)}(\mathcal{S})| \doteq 2^{n[H(\mathbf{X}_\mathcal{S}) \pm 2\varepsilon]}$$

The Multiple Access Channel



- Two "users" communicating over a common channel.
  (The generalization to more than two is straightforward.)

Coding:

- Memoryless pmf (or pdf):

$$p(y|x_1, x_2),\ x_1 \in \mathcal{X}_1,\ x_2 \in \mathcal{X}_2,\ y \in \mathcal{Y}$$

- Data: $W_1 \in \mathcal{I}_1 = \{1, \ldots, M_1\}$ and $W_2 \in \mathcal{I}_2 = \{1, \ldots, M_2\}$
  - Assume W_1 and W_2 uniformly distributed and independent
- Encoders: $\alpha_1 : \mathcal{I}_1 \to \mathcal{X}_1^n$ and $\alpha_2 : \mathcal{I}_2 \to \mathcal{X}_2^n$
- Rates: $R_1 = \frac{1}{n} \log M_1$ and $R_2 = \frac{1}{n} \log M_2$
- Decoder: $\beta : \mathcal{Y}^n \to \mathcal{I}_1 \times \mathcal{I}_2,\ \beta(Y^n) = (\hat{W}_1, \hat{W}_2)$
- Error probability: $P_e^{(n)} = \Pr\left((\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\right)$

Capacity:

We have two (or more) rates, R_1 and R_2
$\Longrightarrow$ cannot consider one maximum achievable rate
$\Longrightarrow$ study sets of achievable rate-pairs (R_1, R_2)
$\Longrightarrow$ trade-off between R_1 and R_2

- Achievable rate-pair: (R_1, R_2) is achievable if $(\alpha_1, \alpha_2, \beta)$ exists such that $P_e^{(n)} \to 0$ as $n \to \infty$
- **Capacity region**:

  The closure of the set of all achievable rate-pairs (R_1, R_2)

Capacity Region for the Multiple Access Channel

- Fix $\pi(x_1, x_2) = p_1(x_1)p_2(x_2)$ on $\mathcal{X}_1$ and $\mathcal{X}_2$. Draw $\{X_1^n(i) : i \in \mathcal{I}_1\}$ and $\{X_2^n(j) : j \in \mathcal{I}_2\}$ in an i.i.d. manner according to p_1 and p_2.
- Symmetry of codebook generation $\implies$

$$
\begin{aligned}
P_e^{(n)} &= \Pr\big((\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\big) \\
&= \Pr\big\{(\hat{W}_1, \hat{W}_2) \neq (1,1) \big| (W_1, W_2) = (1,1)\big\}
\end{aligned}
$$

where the second "Pr" is with respect to the channel and the random codebook design.

- Also

$$
\begin{aligned}
\Pr\big((\hat{W}_1, \hat{W}_2) \neq (1,1)\big) \; &= \; \Pr(\hat{W}_1 \neq 1, \hat{W}_2 \neq 1) \\
&\quad + \Pr(\hat{W}_1 \neq 1, \hat{W}_2 = 1) + \Pr(\hat{W}_1 = 1, \hat{W}_2 \neq 1) \\
&= P_{12}^{(n)} + P_1^{(n)} + P_2^{(n)}
\end{aligned}
$$

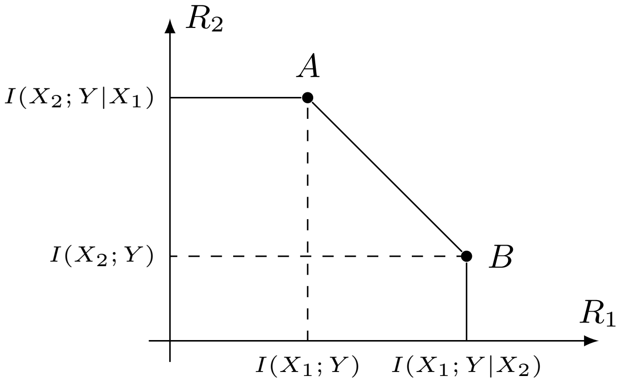
conditioned that $(W_1, W_2) = (1,1)$ everywhere.

- Joint typicality decoding, declare $(\hat{W}_1, \hat{W}_2) = (1,1)$ if

$$
\big(X_1^n(i), X_2^n(j), Y^n\big) \in \mathcal{A}_\varepsilon^{(n)}
$$

only for $i = j = 1 \Rightarrow$

$$
\begin{aligned}
P_{12}^{(n)} &\leq 2^{n[R_1 + R_2 - I(X_1, X_2; Y) + 4\varepsilon]} \\
P_1^{(n)} &\leq 2^{n[R_1 - I(X_1; Y | X_2) + 3\varepsilon]} \\
P_2^{(n)} &\leq 2^{n[R_2 - I(X_2; Y | X_1) + 3\varepsilon]}
\end{aligned}
$$



- Hence, for a fixed $\pi(x_1, x_2) = p_1(x_1)p_2(x_2)$ the capacity region contains at least all pairs (R_1, R_2) in the set Π defined by

$$\begin{aligned} R_1 &< I(X_1; Y|X_2) \\ R_2 &< I(X_2; Y|X_1) \\ R_1 + R_2 &< I(X_1, X_2; Y) \end{aligned}$$

- The corner points
  - Consider the point 'A'

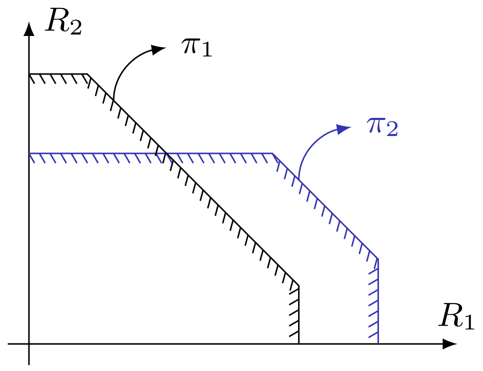
$$\left.\begin{aligned} R_1 &= I(X_1; Y) \\ R_2 &= I(X_2; Y|X_1) \end{aligned}\right\} R_1 + R_2 = I(X_1, X_2; Y)$$

  - User 1 ignores the presence of user 2 $\Rightarrow$ $R_1 = I(X_1; Y)$
  - Decode user 1's codeword $\Rightarrow$ User 2 sees an equivalent channel with input X_2^n and output (Y^n, X_1^n) $\Rightarrow$

$$\begin{aligned} R_2 &= I(X_2; Y, X_1) \\ &= I(X_2; Y|X_1) + I(X_1; X_2) \\ &= I(X_2; Y|X_1) \end{aligned}$$

  - The above can be repeated with $1 \leftrightarrow 2$ and $A \leftrightarrow B$
  - Points on the line A–B can be achieved by time sharing

- Each particular choice of distribution π gives an achievable region Π; for two different π's,



- Fixed $\pi \implies \Pi$ is convex.
Varying $\pi \implies \Pi$ can be non-convex.
However all rates on a line connecting two achievable rate-pairs are achievable by time-sharing.

- The **capacity region** for the **multiple access channel** is the closure of the convex hull of the set of points defined by the three inequalities

$$\begin{aligned} R_1 &< I(X_1; Y | X_2) \\ R_2 &< I(X_2; Y | X_1) \\ R_1 + R_2 &< I(X_1, X_2; Y) \end{aligned}$$

over all possible product distributions $p_1(x_1)p_2(x_2)$ for (X_1, X_2).

- **Proof**: Achievability proof based on jointly typical sequences (as shown before) and a "time-sharing variable".

Converse proof based on Fano's inequality and the independence of X_1^n and X_2^n (since they are functions of independent messages).

Example: A Gaussian Channel

- Bandlimited AWGN channel with two additive users

$$Y(t) = X_1(t) + X_2(t) + Z(t).$$

The noise $Z(t)$ is zero-mean Gaussian with power spectral density $N_0/2$, and $X_1(t)$ and $X_2(t)$ are subject to the power constraints P_1 and P_2, respectively. The available bandwidth is W.

- The capacity of the corresponding single-user channel (with power constraint P) is

$$W \cdot \mathsf{C}\left(\frac{P}{WN_0}\right) \quad \text{[bits/second]}$$

where

$$\mathsf{C}(x) = \log(1 + x).$$

- **Time-Division Multiple-Access (TDMA)**:
Let user 1 use all of the bandwidth with power P_1/α a fraction $\alpha \in [0, 1]$ of time, and let user 2 use all the bandwidth with power $P_2/(1-\alpha)$ the remaining fraction $1-\alpha$ of time. The achievable rates then are

$$R_1 < W \cdot \alpha\, \mathsf{C}\left(\frac{P_1/\alpha}{WN_0}\right) \qquad R_2 < W \cdot (1-\alpha)\, \mathsf{C}\left(\frac{P_2/(1-\alpha)}{WN_0}\right)$$

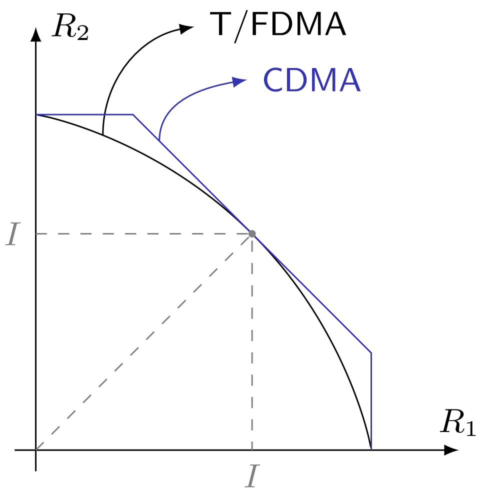
- **Frequency-Division Multiple-Access (FDMA)**:
Let user 1 transmit with power P_1 using a fraction α of the available bandwidth W, and let user two transmit with power P_2 the remaining fraction $(1-\alpha)W$. The achievable rates are

$$R_1 < \alpha W \cdot \mathsf{C}\left(\frac{P_1}{\alpha W N_0}\right) \qquad R_2 < (1-\alpha)W \cdot \mathsf{C}\left(\frac{P_1}{(1-\alpha)WN_0}\right)$$

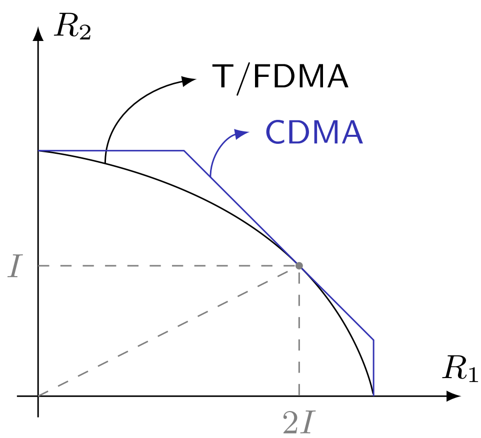
- TDMA and FDMA are equivalent from a capacity perspective!

- **Code-Division Multiple-Access (CDMA)**:
  Defined, in our context, as all schemes that can be implemented to achieve the rates in the true capacity region

$$R_1 \le \quad W \cdot \mathsf{C}\left(\frac{P_1}{WN_0}\right) \quad = W \log\left(1 + \frac{P_1}{WN_0}\right)$$
$$R_2 \le \quad W \cdot \mathsf{C}\left(\frac{P_2}{WN_0}\right) \quad = W \log\left(1 + \frac{P_2}{WN_0}\right)$$
$$R_1 + R_2 \le \quad W \cdot \mathsf{C}\left(\frac{P_1 + P_2}{WN_0}\right) = W \log\left(1 + \frac{P_1 + P_2}{WN_0}\right)$$



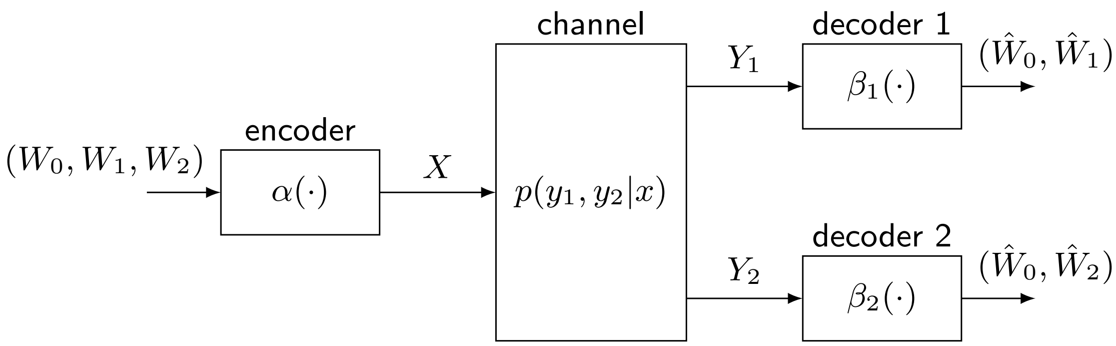
Capacity region for $P_1 = P_2$



Capacity region for $P_1 = 2P_2$

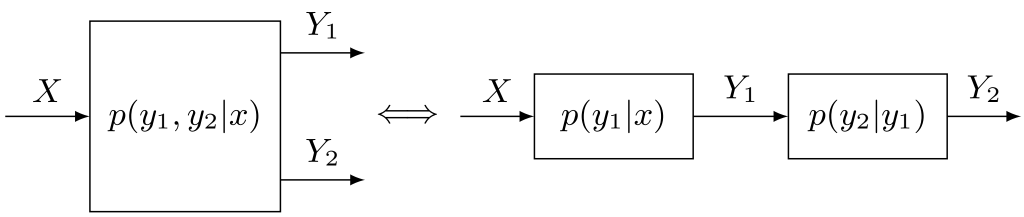
Note that T/FDMA is only optimal when $\dfrac{\alpha}{1-\alpha} = \dfrac{P_1}{P_2}$.

The Broadcast Channel



- One transmitter, several receivers
- Message W_0 is a public message for both receivers, whereas W_1 and W_2 are private messages

The Degraded Broadcast Channel

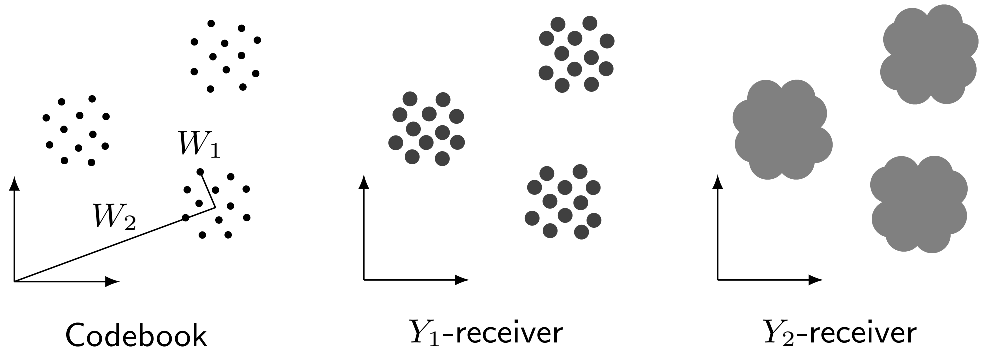


- A broadcast channel is degraded if it can be split as in the figure. That is, Y_2 is a "noisier" version of X than Y_1,

$$p(y_1, y_2|x) = p(y_2|y_1)p(y_1|x).$$

- The Gaussian and the binary symmetric broadcast channels are degraded (see the examples in CT).

Superposition Coding for the Degraded Broadcast Channel



- Assume there is no common information (for simplicity). Let W_2 chose a cloud of possible W_1-codewords.
- The Y_1-receiver sees all codewords, whereas the Y_2-receiver is only able to distinguish between clouds.

- The **capacity region** of the **degraded broadcast channel** (with no common information), is the closure of the convex hull of all rates satisfying

$$R_2 < I(U; Y_2)$$
$$R_1 < I(X; Y_1|U)$$

for some distribution $p_1(u)p_2(x|u)$.

- **Proof**: Choose W_2-codewords i.i.d. according to $p_1(u)$, and for each one, choose W_1-codewords i.i.d. according to $p_2(x|u)$.

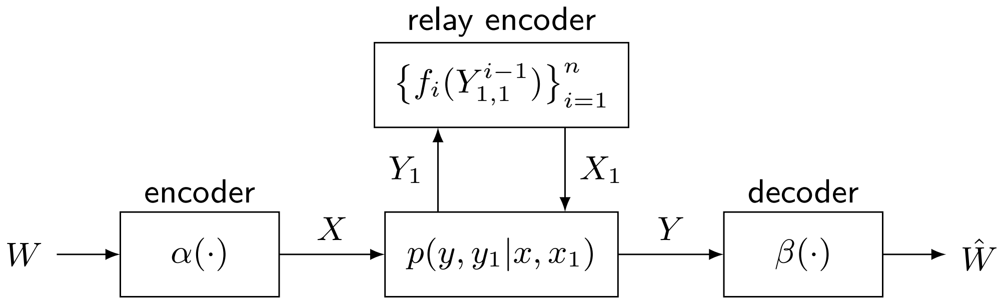
  The overall channel from W_2 to Y_2 (the clouds) can be made error-free as long as $R_2 < I(U; Y_2)$, and conditioned on W_2 the channel from W_1 to Y_1 can be made error-free as long as $R_1 < I(X; Y_1|U)$.

  Converse proved in Problem 15.11 (based on Fano, as usual).

- The **capacity region** of the **degraded broadcast channel** (with **common information**): If the pair $(R_1 = a, R_2 = b)$ is achievable for independent messages, as before, the triple $(R_0, R_1 = a, R_2 = b - R_0)$ is achievable with common information at rate R_0 (as long as $R_0 < b$).

- Since the better receiver can decode both W_1 and W_2, part of the W_2-message can be made to include common information!

The Relay Channel



- One sender, one receiver, and one intermediate node

- The problem does not define the set of relay functions $\{f_i(\cdot)\}_{i=1}^n$. The relay's strategy might be to decode the message, or compress its channel observation, or amplify it and retransmit it, or ...

- Capacity is not known in general. Some known bounds:
  - Cut-set upper bound: The relay is assumed to be co-located with the transmitter or with the receiver.

$$R \le \max_{p(x,x_1)} \min\{\, I(X, X_1; Y),\ I(X; Y, Y_1|X_1)\,\}$$

  - Decode-and-forward lower bound:

$$R \le \max_{p(x,x_1)} \min\{\, I(X, X_1; Y),\ I(X; Y_1|X_1)\,\}$$

Proof: Split transmission in b blocks.

Choose $2^{n\tilde{R}}$ codewords i.i.d. $\sim p(x_1)$, and for each one, choose 2^{nR} codewords i.i.d. $\sim p(x|x_1)$ and distribute them in $2^{n\tilde{R}}$ bins.

The relay can decode the message if $R < I(X; Y_1|X_1)$ and then it sends the bin index in the next block. The receiver can decode the bin index if $\tilde{R} < I(X_1; Y)$ and, knowing this index, it can decode the message from the previous block if $R - \tilde{R} < I(X; Y|X_1)$.

  - These bounds coincide if the relay channel is degraded:

$$p(y, y_1|x, x_1) = p(y_1|x, x_1)p(y|y_1, x_1)$$

General Multiterminal Systems

- M different nodes, each transmitting X_m and receiving Y_m. The message from node i to node j is $W_{i,j}$ with rate $R_{i,j}$. The channel between nodes is $p(y_1, \ldots, y_M|x_1, \ldots, x_M)$.
- Although significant progress in recent years, still only a few general results are known.
- One of them is El Gamal's cut-set bound: If the rates $\{R_{i,j}\}$ are achievable there exists a $p(x_1, \ldots, x_m)$ such that

$$\sum_{i\in\mathcal{S}, j\in\mathcal{S}^c} R_{i,j} < I(X^{(\mathcal{S})}; Y^{(\mathcal{S}^c)}|X^{(\mathcal{S}^c)})$$

for all $\mathcal{S} \subset \{1, \ldots, M\}$.

- The source–channel separation principle:

  For general multiterminal networks the source and channel codes cannot be designed separately without loss. Essentially the source code needs to know the channel to provide optimal *dependencies* between channel input variables.

- Feedback:

  Feedback can increase the capacity of a multi-terminal channel, since it can help transmitters to "cooperate" to reduce interference.