EECS 290S: Network Information Flow **Fall 2008**

Lecture 18 — October 30

Lecturer: Anant Sahai and David Tse *Scribe: Changho Suh*

In this lecture, we studied two types of one-to-many channels: (1) *compound* channels where the message is common to all the receivers (e.g., TV broadcast); (2) *broadcast* channels where each receiver wants an independent message (e.g., cellular downlink). Typically we refer to the second case when saying the broadcast channel.

18.1 Compound Channels

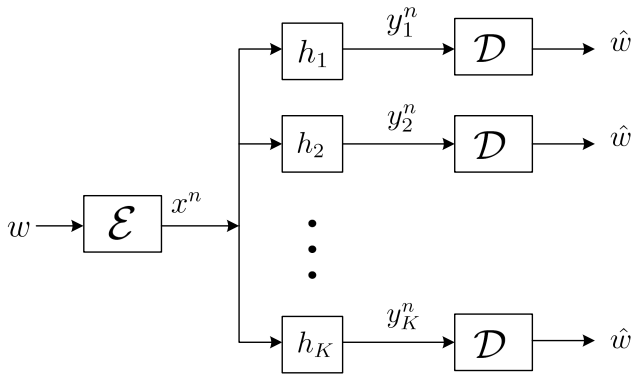

Figure 18.1. The Compound Channel

First we consider compound channels which have common messages to all the receivers. The best-well known example is TV broadcast. Notice that compound channels are different from broadcast channels with independent messages to different receivers.

Fig. 18.1 illustrates the encoding and decoding operations for the compound channel. Since the message is *common*, we have only one message w. From w, the encoder generates codewords $x^n(w)$. This passes through different channels $(h_1, h_2, \cdots, h_K)$ and we get $y_1^n, y_2^n, \cdots, y_K^n$ in each receiver, respectively. Since the message is the same, all the receivers also have a common decoder $\mathcal{D}$.

Suppose that the channels are non-varying and known to each receiver. Then, the trivial achievable scheme is to generate codewords based on the worst channel. Therefore, we can achieve

$$R = \max_{p(x)} \min_{k=1,\cdots,K} I(X; Y_k). \tag{18.1}$$

Notice that it is not $\min_k \max_{p(x)} I(X;Y_k)$ since the channel is set by nature. In fact, this achievable rate is indeed capacity. The converse proof is trivial. For any given distribution $p(x)$, we have a bunch of K inequalities: $R \le I(X;Y_k)$ for all $k = 1, \cdots, K$, which implies that $R \le \min_k I(X;Y_k)$.

In the above, we assume that channels are non-varying. However, if the channels are time-varying, then we have a problem in computing the capacity. To handle this case, we typically consider ϵ-*outage* capacity which is the maximum achievable rate that allows small error probability, say ϵ.

18.2 Broadcast Channels

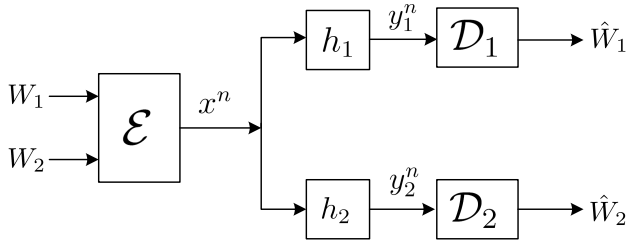

Figure 18.2. The Broadcast Channel

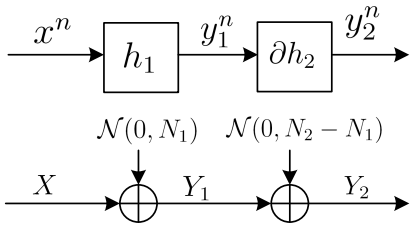

Figure 18.3. The Degraded Broadcast Channel: the Gaussian Channel Example

Fig. 18.2 describes the two-user broadcast channel. There are two independent messages W_1 and W_2. The encoder generates a codeword $x^n(w_1, w_2)$ based on these two messages. This passes through channels h_1 and h_2, so each receiver takes y_1^n and y_2^n. Each receiver decodes its desired message $\hat{W}_1$ and $\hat{W}_2$, respectively.

For the general channel, unfortunately, the capacity region has been open for several decades. However, we have exact capacity results for some special cases. One case is the *degraded* broadcast channel. Fig. 18.3 illustrates the *physically* degraded broadcast channel. Notice that y_2^n is a physically degraded signal of y_1^n. The important thing is that the capacity result is not limited to the physically degraded channel, but also holds for the *stochastically*

degraded channel, which should be distinguished from the physically degraded one. Let us clarify the difference between these two.

Definition 1. *A broadcast channel is said to be physically degraded if*

$$p(y_1, y_2|x) = p(y_1|x)p(y_2|y_1). \tag{18.2}$$

Definition 2. *A broadcast channel is said to be stochastically degraded if its conditional marginal distributions are compatible with those of a physically degraded one; that is, if there exists a distribution $p'(y_2|y_1)$ s.t.*

$$p(y_2|x) = \sum_{y_1} p(y_1|x)p'(y_2|y_1). \tag{18.3}$$

Notice that the capacity result also holds for a stochastically degraded channel since the capacity of a broadcast channel depends *only* on the conditional marginal distributions. Therefore, it is enough to stick to the physically degraded channel as depicted in Fig. 18.3.

18.2.1 Gaussian Broadcast Channels

To gain some insights into the achievable scheme of a degraded channel, we first consider a simple case of the AWGN channel. Fortunately all *scalar* Gaussian broadcast channels belong to the class of degraded broadcast channels. In this case, we can model the channel as follows:

$$\begin{aligned} Y_1 &= X + Z_1, \\ Y_2 &= X + Z_2 = Y_1 + \tilde{Z}_2, \end{aligned} \tag{18.4}$$

where $E[X^2] \leq P$, $Z_1 \sim \mathcal{N}(0, N_1)$, $\tilde{Z}_2 \sim \mathcal{N}(0, N_2 - N_1)$, and Z_1,$\tilde{Z}_2$ are independent. Here, without loss of generality, we assume Y_1 is less noisy than Y_2, i.e., $N_2 > N_1$. Due to the independence of Z_1 and $\tilde{Z}_2$, we can easily check that noise variance of $Z_2 = Z_1 + \tilde{Z}_2$ is N_2. The main idea of the achievable scheme is *superposition* encoding: first generating cloud-center codewords; and then *adding* finer codewords on top of that. Specifically, we assign $(1-\alpha)P$ to the cloud-center message to generate a Gaussian codeword $x_2^n(w_2)$ for user 2. We then assign remaining power αP to the finer message to generate another Gaussian codeword $x_1^n(w_1)$ for user 1. Finally we *add* them up to generate the encoded signal $x^n(w_1, w_2) = x_1^n(w_1) + x_2^n(w_2)$. The decoding operation is very simple. User 2 decodes its own message w_2 just by treating user 1 signal as noise. Since all the codewords are Gaussian, user 2 can decode w_2 if

$$R_2 < \frac{1}{2} \log \left(1 + \frac{(1-\alpha)P}{\alpha P + N_2} \right). \tag{18.5}$$

Since user 1 is better than user 2, it can always decode w_2 to subtract it from the received signal, so that it can obtain the non-interfered signal. Therefore, user 1 can decode w_1 if

$$R_1 < \frac{1}{2} \log \left(1 + \frac{\alpha P}{N_1} \right). \tag{18.6}$$

In fact, this rate region ranging for different $0 < \alpha < 1$ is the capacity region. The converse proof is deferred to the next lecture.

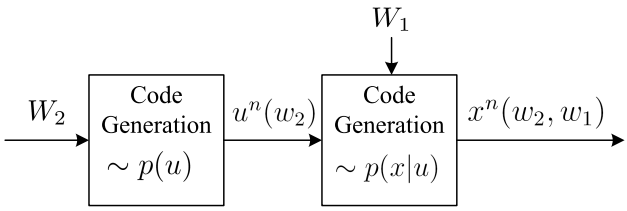

Figure 18.4. The Codeword Generation for the Degraded Broadcast Channel

18.2.2 Degraded Broadcast Channels

Now come back to the general case of the degraded channels. Similar to the Gaussian case, the achievability idea is also *successive* encoding, i.e., we generate *cloud*-center codewords and then finer codewords. In the general case, however, we cannot just *add* two codewords as the Gaussian case. Then what do we have to change? Fig. 18.4 describes the detailed encoding scheme. Unlike the Gaussian case, we introduce an auxiliary random variable U to indicate cloud-center codewords. First fix the joint distribution $p(u, x) = p(u)p(x|u)$ which is our choice. Next from the message $w_2 \in \{1, \cdots, 2^{nR_2}\}$, we generate $u^n(w_2)$ according to $p(u)$. For each codeword $u^n(w_2)$ and $w_1 \in \{1, \cdots, 2^{nR_1}\}$, we generate codeword $x^n(w_1, w_2)$ according to $p(x|u)$. Note that unlike the Gaussian case, we generate $x^n(w_1, w_2)$ for each cloud-center codeword $u^n(w_2)$ instead of just adding two independent codewords. Since $u^n(w_2)$ is intended for user 2, user 2 can decode its message if

$$R_2 < I(U; Y_2). \tag{18.7}$$

Under this condition, user 1 can automatically decode w_2 due to the degradedness assumption. Therefore, user 1 can decode its message w_1 if

$$R_1 < I(X; Y_1|U). \tag{18.8}$$

We omit the detailed achievability proof involving sophisticated calculation of probability error. Refer to Cover-Thomas book for details [1]. Instead we focus on the converse proof in this lecture because the converse proof was deferred to the exercise problem in Cover-Thomas book.

Now let us prove the converse. The following proof is based on Gallager's proof [2]. The main trick is how to introduce an auxiliary random variable U. We start with the upper

bound for R_1.

$$\begin{aligned}
nR_1 &= H(W_1) \\
&\overset{(a)}{=} H(W_1|W_2) \\
&\overset{(b)}{=} I(W_1; Y_1^n|W_2) + H(W_1|Y_1^n, W_2) \\
&\overset{(c)}{\leq} I(W_1; Y_1^n|W_2) + n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n} [H(Y_{1i}|Y_1^{i-1}, W_2) - H(Y_{1i}|Y_1^{i-1}, W_2, W_1)] + n\epsilon_n \\
&\overset{(e)}{=} \sum_{i=1}^{n} [H(Y_{1i}|Y_1^{i-1}, W_2) - H(Y_{1i}|Y_1^{i-1}, W_2, W_1, X_i)] + n\epsilon_n \\
&\overset{(f)}{=} \sum_{i=1}^{n} [H(Y_{1i}|U_i) - H(Y_{1i}|U_i, W_1, X_i)] + n\epsilon_n \\
&\overset{(g)}{=} \sum_{i=1}^{n} [H(Y_{1i}|U_i) - H(Y_{1i}|U_i, X_i)] + n\epsilon_n \\
&= \sum_{i=1}^{n} I(X_i; Y_{1i}|U_i) + n\epsilon_n
\end{aligned} \tag{18.9}$$

where (a) follows from the independence of W_1 and W_2; (b) follows from letting $Y_1^n = (Y_{11}, \cdots, Y_{1n})$; (c) follows from the Fano's inequality; (d) follows from the chain rule; (e) follows from the fact that X_i is a function of W_1 and W_2; (f) follows from letting $U_i = (W_2, Y_1^{i-1})$; and (g) follows from the *memoryless* channel. In fact, we can shorten the procedures by using *data processing inequality* to directly introduce X_i. However, for completeness, we showed all the detailed steps. Here the tricky part is to introduce an auxiliary random variable U_i as follows:

$$U_i = (W_2, Y_1^{i-1}). \tag{18.10}$$

This is a typical trick that can be used in the converse proof when needing an auxiliary random variable.

Next consider the upper bound of R_2.

$$\begin{aligned}
nR_2 &= H(W_2) \\
&\overset{(a)}{\leq} I(W_2; Y_2^n) + n\epsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} [H(Y_{2i}|Y_2^{i-1}) - H(Y_{2i}|Y_2^{i-1}, W_2)] + n\epsilon_n \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} [H(Y_{2i}|Y_2^{i-1}) - H(Y_{2i}|Y_2^{i-1}, W_2, Y_1^{i-1})] + n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n} [H(Y_{2i}|Y_2^{i-1}) - H(Y_{2i}|W_2, Y_1^{i-1})] + n\epsilon_n \\
&\overset{(e)}{\leq} \sum_{i=1}^{n} [H(Y_{2i}) - H(Y_{2i}|U_i)] + n\epsilon_n \\
&= \sum_{i=1}^{n} I(U_i; Y_{2i}) + n\epsilon_n
\end{aligned} \tag{18.11}$$

where (a) follows from the Fano's inequality; (b) follows from the chain rule; (c) follows from the fact that conditioning reduces entropy; (d) follows from the fact that Y_2^{i-1} is a degraded signal of Y_1^{i-1}; and (e) follows from the fact that conditioning reduces entropy and $U_i = (W_2, Y_1^{i-1})$.

We can rewrite (18.9) and (18.11) as

$$\begin{aligned}
R_1 &\leq \sum_{i=1}^{n} \frac{1}{n} I(X_i; Y_{1i}|U_i) + \epsilon_n, \\
R_2 &\leq \sum_{i=1}^{n} \frac{1}{n} I(U_i; Y_{2i}) + \epsilon_n.
\end{aligned} \tag{18.12}$$

Now let Q be uniform over $\{1, 2, \cdots, n\}$. We define

$$\begin{aligned}
X' &= X_Q, \\
Y_1' &= Y_{1Q}, \\
Y_2' &= Y_{2Q}, \\
U' &= U_Q,
\end{aligned} \tag{18.13}$$

for some distribution $p(q)p(u|q)p(x|u,q)p(y_1, y_2|x)$. This implies that we define X', Y_1', Y_2', U' such that

$$\begin{aligned}
p(X' = x|Q = q) &= p(X_q = x), \\
p(Y_1' = y|Q = q) &= p(Y_{1q} = y), \\
p(Y_2' = y|Q = q) &= p(Y_{2q} = y), \\
p(U' = u|Q = q) &= p(U_q = u).
\end{aligned} \tag{18.14}$$

Then, we get

$$\begin{aligned} R_1 &\le I(X';Y_1'|U',Q)+\epsilon_n, \\ R_2 &\le I(U';Y_2'|Q)+\epsilon_n. \end{aligned} \tag{18.15}$$

We are not satisfied with this form because we still have time-sharing random variable Q. However, we can remove it by redefining $U''=(U',Q)$. From this and $\epsilon_n \to 0$ ((R_1,R_2) is achievable), we can argue that the region (18.15) is equal to the convex closure of regions of the form

$$\begin{aligned} R_1 &\le I(X';Y_1'|U''), \\ R_2 &\le I(U'';Y_2'), \end{aligned} \tag{18.16}$$

for some joint distribution $p(u'')p(x'|u'')p(y_1',y_2'|x')$. Cleaning up prime notation, we finally get

$$\begin{aligned} R_1 &\le I(X;Y_1|U), \\ R_2 &\le I(U;Y_2), \end{aligned} \tag{18.17}$$

for some joint distribution $p(u)p(x|u)p(y_1,y_2|x)$.

Now the only remaining part is to prove the cardinality bound of U. The main idea is to use standard methods from convex set theory. Intuitively, we can come up with the following necessary condition from the Markov-chain relationship $U-X-Y_1-Y_2$:

$$|\mathcal{U}| \le \min\{|\mathcal{X}|,|\mathcal{Y}_1|,|\mathcal{Y}_2|\}. \tag{18.18}$$

Indeed this is also sufficient. We omit a rigorous proof in this lecture. Refer to [3] (p.11-12) for the detailed proof.

Bibliography

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Academic Press, 2nd ed., July 2006.

[2] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Peredachi Inf.*, pp. 3–14, 1974.

[3] M. Salehi, "Cardinality bounds on auxiliary variables in multiple-user theory via the method of Ahlswede and Korner," *Technical Report 33, Stanford Univ.*, Aug. 1978.