Capacity for Classes of Broadcast Channels with Receiver Side Information

Gerhard Kramer
Communications and Statistical Sciences Department
Bell Labs, Alcatel-Lucent
Murray Hill, NJ 07974
gkr@bell-labs.com

Shlomo Shamai (Shitz)
Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
sshlomo@ee.technion.ac.il

***Abstract*— Methods are developed that achieve all rate points inside the capacity region of two-receiver memoryless broadcast channels where each receiver knows the message it need not decode. The capacity result generalizes to channels with per-letter average cost constraints, e.g., additive white Gaussian noise channels with power constraints. The result further generalizes to cases where the receivers have only partial message information and where one receiver must decode both messages.**

I. Introduction

Consider a memoryless broadcast channel $P_{Y_1Y_2|X}$ as depicted in Fig. 1. The encoder maps the messages W_1 and W_2 to the string $X^n = X_1, X_2, \ldots, X_n$ whose symbols are input to the channel. Message W_k has B_k bits so the rates are $R_k = B_k/n$ bits/use, $k = 1, 2$. Receiver 1 observes Y_1^n and the message side information W_2, and similarly for receiver 2. Receiver k emits the estimate $\hat{W}_k$, $k = 1, 2$. The *capacity region* $\mathcal{C}$ is the closure of the set of (R_1, R_2) for which there exist, for sufficiently large n, an encoder and decoders so that

$$\Pr[\{\hat{W}_1 \neq W_1\} \cup \{\hat{W}_2 \neq W_2\}] < \epsilon \qquad (1)$$

where ϵ is an arbitrary positive number.

The above problem arises in several contexts.

- A two-way relay channel where nodes 1 and 2 exchange messages via a relay node [10]. If the relay decodes both messages, then it can *broadcast* a common codeword to both receivers that each know their own message.
- The two-way relay channel is often used to demonstrate benefits of network coding in a wireless setting (see e.g. [6]). For example, if $R_1 = R_2$ then the relay can transmit an encoded version of $W_1 \oplus W_2$ where the XOR is performed bit-by-bit.
- The standard network coding example for a butterfly network [1, Fig. 7(b)] can be viewed as a special case of Fig. 1 if the broadcast channel has memory.

The capacity region of the network in Fig. 1 is known and is a special case of a more general result in [11, Thm. 6] (see also [7, Sec. 4.2.8], [8], [14], [15]). One of our aims is to prove the coding theorem in several ways different than in [11]. We are motivated to do this because different code structures have different performance versus complexity tradeoffs under realistic constraints, e.g., finite block lengths and modulation constraints. For example, one of our approaches is a random-coding counterpart of the nested lattice codes of [4], [5]. Our approaches are also related to the nested and algebraic superposition codes in [12], [13].

More generally, we are interested in cases where the receivers have only partial information about the source messages. Such problems arise in the following situations.

- The receivers cooperate by exchanging parts of the messages they each must decode. For example, node 1 might decode its message first and then transmit part of this message to node 2 before node 2 decodes.
- The butterfly network mentioned above, but where the links carrying "side-information" have limited capacity.

We derive the capacity region for such problems when one receiver must decode both messages. This type of model is often referred to as having "degraded message sets".

This document is organized as follows. In Section II, we review the capacity region of the model in Fig. 1. In Section III, we generalize the result to channels with cost constraints, including additive white Gaussian noise (AWGN) channels. In Sections IV and V, we generalize the results to cases where the receivers have partial side information and the message sets are degraded. Section VI concludes the paper.

II. A Capacity Result

Theorem 1: The capacity region of the network of Fig. 1 is

$$\mathcal{C} = \bigcup_{P_X} \{(R_1, R_2) : 0 \leq R_k \leq I(X; Y_k),\ k = 1, 2\}. \qquad (2)$$

Proof: Theorem 1 is a special case of [11, Thm. 6] by choosing the random variables appropriately and mapping the messages onto the source strings. In particular, in [11, Fig. 1] we choose X^n to be a vector $(X_1, X_2)^n$ where X_1 and X_2 are independent and X_k has entropy R_k bits, $k = 1, 2$. We further choose $Y_1 = X_2$, $Y_2 = X_1$, and $m = n$. The model in [11, Fig. 1] is then the same as the model Fig. 1 up to introducing invertible mappings between W_k and X_k^n, $k = 1, 2$. ■

Remark 1: The converse in [11] follows by cut-set arguments, while achievability follows by using a codebook with $2^{n(R_1+R_2)}$ codewords $x^n(w_1, w_2)$, $w_k = 1, 2, \ldots, 2^{nR_k}$, $k = 1, 2$, where every letter $x_i(w_1, w_2)$, $i = 1, 2, \ldots, n$, is chosen independently via P_X. Each receiver uses a typical sequence decoder given its channel output and side information.

1-4244-1564-0/07/$25.00 ©2007 IEEE

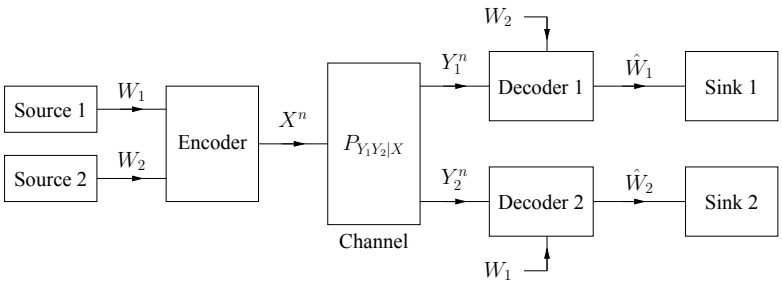


Fig. 1. A broadcast channel with receiver side information.

The relation of Theorem 1 to [11, Thm. 6] was pointed out in [7, Sec. 4.2.8] and [8]; Theorem 1 also appeared in [14], [15]. We prove the coding theorem differently than in [11]. However, to introduce notation, we first prove the converse.

Fano's inequality [2, Sec. 2.11] implies that reliable communication requires

$$\begin{aligned} nR_1 &\le I(W_1;\hat{W}_1) \\ &\le I(W_1;Y_1^n\,W_2) \\ &= I(W_1;Y_1^n|W_2) \\ &= \sum_{i=1}^{n} H(Y_{1,i}|Y_1^{i-1}W_2) - H(Y_{1,i}|X_i) \\ &\le \sum_{i=1}^{n} I(X_i;Y_{1,i}) \\ &= nI(X_T;Y_{1,T}|T) \end{aligned} \tag{3}$$

where the second step follows by the data-processing inequality [2, p. 32], $Y_{1,i}$ is the ith letter of Y_1^n, and T is a "time-sharing" random variable that is uniform over $1,2,\ldots,n$, i.e., we have

$$P_{TX_TY_{1,T}Y_{2,T}}(i,a,b,c) = \frac{1}{n}P_{X_i}(a)P_{Y_1Y_2|X}(b,c|a). \tag{4}$$

We define $X = X_T$, $Y_1 = Y_{1,T}$, and $Y_2 = Y_{2,T}$, and use (3) and (4) to bound

$$R_1 \le I\left(X;Y_1\right) \tag{5}$$

where

$$P_{XY_1Y_2}(a,b,c) = \left[\frac{1}{n}\sum_{i=1}^{n} P_{X_i}(a)\right] P_{Y_1Y_2|X}(b,c|a). \tag{6}$$

By symmetry, we have

$$R_2 \le I\left(X;Y_2\right) \tag{7}$$

which completes the converse.

For the coding theorem, we use an approach motivated by nested lattice codes [4], [5]; other approaches are outlined in Appendix A. Generate codewords of length $n = ML$ by using an L-letter coding argument. Consider a distribution P_X with domain $\mathcal{X}$ and the typical set (see [9])

$$\begin{aligned} T_\delta^L(P_X) = \big\{x^L : &\left|N(a|x^L)/L - P_X(a)\right| \\ &\le \delta \cdot P_X(a),\ \text{all } a \in \mathcal{X}\big\} \end{aligned} \tag{8}$$

where $N(a|x^L)$ is the number of times the letter a occurs in x^L. Let X_1^L and X_2^L be independent random variables that are uniformly distributed over $T_\delta^L(P_X)$. Generate codewords

$$x_1^n(w_1) = x_{1,1}^L(w_1), x_{1,2}^L(w_1), \ldots, x_{1,M}^L(w_1) \tag{9}$$

by choosing every $x_{1,m}^L(w_1)$ independently using $P_{X_1^L}$ for $m = 1,2,\ldots,M$ and $w_1 = 1,2,\ldots,2^{nR_1}$. Generate codewords $x_2^n(w_2)$, $w_2 = 1,2,\ldots,2^{nR_2}$, in the same way with $P_{X_2^L}$.

Let $|\mathcal{S}|$ be the cardinality of the set $\mathcal{S}$. We map every x^L in $T_\delta^L(P_X)$ to a unique integer $\mathrm{Int}(x^L)$ in the set $\{0,1,2,\ldots,|T_\delta^L(P_X)|-1\}$. The encoder uses superposition coding and transmits $x^n = x_1^L, x_2^L, \ldots, x_M^L$ where

$$\begin{aligned} x_m^L = \mathrm{Int}^{-1}\big(&\mathrm{Int}(x_{1,m}^L(w_1)) \\ &+ \mathrm{Int}(x_{2,m}^L(w_2)) \mod |T_\delta^L(P_X)|\big) \end{aligned} \tag{10}$$

for $m = 1,2,\ldots,M$. The "mod" refers to integer modulo addition. We will abuse notation and write (10) as

$$x_m^L = x_{1,m}^L + x_{2,m}^L \mod |T_\delta^L(P_X)|. \tag{11}$$

Receiver 1 observes y_1^n, computes $x_2^n(w_2)$, and tries to find a w_1 such that

$$(x_1^n(w_1), x_2^n(w_2), y_1^n) \in T_{\delta_2}^M\left(P_{X_1^L X_2^L Y_1^L}\right) \tag{12}$$

where $\delta_2 > \delta$ and $P_{X_1^L X_2^L Y_1^L}$ is the distribution resulting when

$$X^L = X_1^L + X_2^L \mod |T_\delta^L(P_X)| \tag{13}$$

is input to the channel $P_{Y_1Y_2|X}$. Standard typicality arguments (cf. [2, Ch. 14]) show that w_1 is decoded reliably if M is large, δ_2 is small, and

$$\begin{aligned} LR_1 &< I\left(X_1^L;Y_1^L|X_2^L\right) \\ &= H\left(Y_1^L|X_2^L\right) - H\left(Y_1^L|X^L\right). \end{aligned} \tag{14}$$

A key point now is that, for all a^L and b^L, over the ensemble of all codebooks we have

$$P_{X^L|X_2^L}(a^L|b^L) = P_{X^L}(a^L) = |T_\delta^L(P_X)|^{-1} \tag{15}$$

so that X^L and X_2^L are independent. Furthermore, we have the Markov chain $X_2^L - X^L - Y_1^L$ so that $H(Y_1^L|X_2^L) = H(Y_1^L)$. The bound (14) is thus

$$LR_1 < I\left(X^L;Y_1^L\right) \tag{16}$$

where X^L is uniform over $T_\delta^L(P_X)$. But for $\delta_2 > \delta$ we have

$$\begin{aligned} & I\left(X^L; Y_1^L\right) \\ & \stackrel{(a)}{=} \sum_{a^L \in T_\delta^L(P_X)} \sum_{b^L} P_{X^L Y_1^L}(a^L, b^L) \log \frac{P_{X^L|Y_1^L}(a^L|b^L)}{P_{X^L}(a^L)} \\ & \stackrel{(b)}{\geq} (1 - \epsilon_{\delta,\delta_2}(L)) \log \left. \frac{P_{X^L|Y_1^L}(a^L|b^L)}{P_{X^L}(a^L)} \right|_{(a^L,b^L) \in T_{\delta_2}^L(P_{XY_1})} \\ & \stackrel{(c)}{\geq} (1 - \epsilon_{\delta,\delta_2}(L)) L \left[I(X;Y_1) - 2\delta_2 H(X) \right] \end{aligned} \tag{17}$$

where (a) follows because $x^L \in T_\delta^L(P_X)$, (b) follows by Lemma 3 in Appendix B, and (c) follows by Lemmas 1 and 2 in Appendix B and because $H(X|Y_1) \leq H(X)$. One can thus replace (16) by the more restrictive

$$R_1 < (1 - \epsilon_{\delta,\delta_2}(L)) \left[I(X;Y_1) - 2\delta_2 H(X) \right] \tag{18}$$

where $\epsilon_{\delta,\delta_2}(L) \to 0$ as $L \to \infty$. We thus have the $k = 1$ bound in (2) by letting $L \to \infty$ and $\delta, \delta_2 \to 0$. By symmetry, we also have the $k = 2$ bound in (2). This proves Theorem 1.

Example 1: Consider the deterministic broadcast channel with $\mathcal{X} = \{0, 1, 2, 3\}$, $Y_1 = X$, and $Y_2 = (X \mod 3)$. We find that the interesting boundary points of $\mathcal{C}$ are given by

$$R_1 = 1 + h(p) \tag{19}$$

$$R_2 = p + h(p) \tag{20}$$

where $h(p) = -p \log(p) - (1-p)\log(1-p)$ is the binary entropy function and $1/2 \leq p \leq 2/3$.

III. Cost Constraints and Gaussian Channels

Suppose X^n, Y_1^n, and Y_2^n have the cost constraints

$$\sum_{i=1}^{n} \frac{1}{n} \mathrm{E}[s_j(X_i, Y_{1,i}, Y_{2,i})] \leq S_j, \quad j = 1, 2, \ldots, J \tag{21}$$

where each function s_j evaluates to a non-negative real number. We will require that, for all j, there is a letter $a_j \in \mathcal{X}$ for which $\mathrm{E}[s_j(a_j, Y_1, Y_2)] \leq S_j$ (this requirement ensures that the cost constraints can be met). One can use the above methods to show that the capacity region of the network in Fig. 1 with the constraints (21) is given by (2), but where the union is taken over the P_X satisfying

$$\mathrm{E}[s_j(X, Y_1, Y_2)] \leq S_j, \quad j = 1, 2, \ldots, J. \tag{22}$$

This result, in turn, suggests that one can determine the capacity region of the Gaussian broadcast channel with

$$Y_k = X + Z_k, \quad k = 1, 2 \tag{23}$$

$$\sum_{i=1}^{n} \frac{1}{n} \mathrm{E}[X_i^2] \leq P \tag{24}$$

where X is real, and Z_1 and Z_2 are independent, Gaussian random variables with variances N_1 and N_2, respectively. In fact, we have the following result.

Theorem 2: The capacity region of the Gaussian broadcast channel in Fig. 1 with (23) and (24) is

$$\mathcal{C} = \left\{ (R_1, R_2) : 0 \leq R_k \leq \frac{1}{2} \log\left(1 + \frac{P}{N_k}\right), \ k = 1, 2 \right\}. \tag{25}$$

Proof: The region (25) is clearly an outer bound on $\mathcal{C}$. For the coding theorem, we could simply invoke (2) and the maximum entropy theorem [2, p. 234]. Instead, to relate to our theory in Sec. II, we use modulo-lattice addition (MLA) as described in [4], [5]. Consider a lattice Λ in the set $\mathbb{R}^L$ of real vectors of length L. Let $\mathcal{V}$ be the fundamental Voronoi region of Λ which means, e.g., that every $x^L \in \mathbb{R}^L$ can be uniquely written as (see [4, Sec. II])

$$x^L = \lambda + r^L$$

where $\lambda \in \Lambda$ and $r^L \in \mathcal{V}$. We write $r^L = x^L \mod \Lambda$.

Let X_1^L and X_2^L be independent random variables that are uniform over $\mathcal{V}$. Generate codewords $x_1^n(w_1)$ as in (9) by choosing the $x_{1,m}^L(w_1)$ using $P_{X_1^L}$ for $m = 1, 2, \ldots, M$, $w_1 = 1, 2, \ldots, 2^{nR_1}$. Similarly generate codewords $x_2^n(w_2)$ using $P_{X_2^L}$. The encoder transmits

$$x_m^L = x_{1,m}^L + x_{2,m}^L \mod \Lambda \tag{26}$$

for $m = 1, 2, \ldots, M$, which is similar to (11). Alternatively, one can write (26) as $X^L = X_1^L + X_2^L \mod \Lambda$, which is similar to (13). The key point now is that X^L and X_2^L (and X_L and X_1^L) are independent, as in (15) and [4, Lemma 1].

Receiver 1 observes y_1^n and computes

$$\tilde{y}_{1,m}^L = \alpha_1 \, y_{1,m}^L - x_{2,m}^L \mod \Lambda \tag{27}$$

for $m = 1, 2, \ldots, M$ where $\alpha_1 = P/(P + N_1)$. One can now follow the same steps as in the proof of [4, Theorem 1] and argue that for large L there are Λ such that R_1 approaches $\log(1 + P/N_1)/2$. By symmetry, one can simultaneously make R_2 approach $\log(1 + P/N_2)/2$. ■

Remark 2: The signal x_2^n serves as a dither signal for x_1^n, and vice versa (cf. [4, Sec. III]). Observe also that the superposition coding (11) is a natural discrete-channel counterpart of MLA superposition coding for AWGN channels.

IV. Partial Side Information and Degraded Message Sets

Let $W_k = (W_k', W_k'')$ where W_k' and W_k'' are independent with entropies nR_k' and nR_k'', respectively, i.e., we have $R_k = R_k' + R_k''$. Consider Fig. 1, but where decoders 1 and 2 receive only W_2' and W_1', respectively, as side information. We determine the capacity region for the case where decoder 1 must decode both W_1 and W_2, while decoder 2 must decode only W_2. This problem generalizes the previous problem since decoder 1 in Fig. 1 may as well be required to decode both messages.

Theorem 3: The capacity region of the network in Fig. 1 where decoder 1 has inputs Y_1^n and W_2', decoder 2 has inputs

Y_2^n and W_1', decoder 1 must reliably decode W_1 and W_2, and decoder 2 must reliably decode W_2 is

$$\mathcal{C} = \bigcup \{(R_1, R_2) : 0 \le R_1, 0 \le R_2, \left.\begin{array}{l} R_2 \le I(U;Y_2) \\ R_1 + R_2 \le R_2' + I(X;Y_1) \\ R_1 + R_2 \le R_1' + I(X;Y_1|U) + I(U;Y_2) \end{array}\right\} \quad (28)$$

where the union is over $P_{UXY_1Y_2}$ that factor as $P_{UX}P_{Y_1Y_2|X}$.

Proof: Let the estimate of W_1 be $\hat{W}_1$, and the estimates of W_2 at decoders 1 and 2 be $\hat{W}_2(1)$ and $\hat{W}_2(2)$, respectively. The "common" rate can be viewed as being $R_1'+R_2$ since both receivers "decode" W_1' and W_2. We choose P_{UX} and generate $2^{n(R_1'+R_2)}$ codewords $u^n(w_1', w_2)$ via P_U. For every pair (w_1', w_2), we then generate $2^{nR_1''}$ codewords $x^n(w_1', w_1'', w_2)$ via $P_{X|U}(\cdot|u_i(w_1', w_2))$, $i = 1, 2, \ldots, n$. Given (w_1, w_2), the encoder transmits $x^n(w_1, w_2)$ (cf. [3, p. 360].)

Given y_1^n and w_2', receiver 1 tries to find a $(\tilde{w}_1', \tilde{w}_1'', \tilde{w}_2'')$ such that

$$(u_n(\tilde{w}_1', w_2', \tilde{w}_2''), x^n(\tilde{w}_1', \tilde{w}_1'', w_2', \tilde{w}_2''), y_1^n) \in T_\delta^n(P_{UXY_1}). \quad (29)$$

Similarly, given y_2^n and w_1', receiver 2 tries to find a $\tilde{w}_2$ such that

$$(u_n(w_1', \tilde{w}_2), y_2^n) \in T_\delta^n(P_{UY_2}). \quad (30)$$

The usual typicality arguments guarantee that both decoders put out reliable estimates as long as n is large and

$$R_1 + R_2'' < I(X;Y_1) \quad (31)$$
$$R_1'' < I(X;Y_1|U) \quad (32)$$
$$R_2 < I(U;Y_2). \quad (33)$$

Next, observe that if (R_1', R_2, R_1'') is achievable, then so is $(R_1', R_2 - \Delta, R_1'' + \Delta)$ where $0 \le \Delta \le R_2$. Inserting this into the above equations and eliminating the variable Δ, we find that all rate points inside $\mathcal{C}$ in (28) are achievable.

To prove the converse, let

$$Y_{2,i+1}^n = Y_{2,i+1}, Y_{2,i+2}, \ldots, Y_n \quad (34)$$
$$U_i = [W_1', W_2, Y_1^{i-1}, Y_{2,i+1}^n] \quad (35)$$

and note that $U_i - X_i - (Y_{1,i}, Y_{2,i})$ forms a Markov chain. Using Fano's inequality, reliable communication requires

$$\begin{aligned} nR_2 &\le I(W_2;\hat{W}_2(2)) \\ &\le I(W_2;Y_2^n\, W_1') \\ &= I(W_2;Y_2^n|W_1') \\ &= \sum_{i=1}^n H(Y_{2,i}|W_1'Y_{2,i+1}^n) - H(Y_{2,i}|W_1'W_2Y_{2,i+1}^n) \\ &\le \sum_{i=1}^n H(Y_{2,i}) - H(Y_{2,i}|W_1'W_2Y_1^{i-1}Y_{2,i+1}^n) \\ &= nI(U_T;Y_{2,T}|T) \\ &\le nI\,(U;Y_2) \end{aligned} \quad (36)$$

where T is again a time-sharing random variable that is uniform over $1, 2, \ldots, n$, and where we define $U = (T, U_T)$, $X = X_T$, $Y_1 = Y_{1,T}$, and $Y_2 = Y_{2,T}$. We thus have

$$P_{UXY_1Y_2}((i,a),b,c,d) = \frac{1}{n}P_{U_iX_i}(a,b)P_{Y_1Y_2|X}(c,d|b). \quad (37)$$

We further have

$$\begin{aligned} n(R_1+R_2) &\le I(W_1W_2;\hat{W}_1\hat{W}_2(1)) \\ &\le I(W_1W_2;Y_1^n\,W_2') \\ &= I(W_1W_2;W_2') + I(W_1W_2;Y_1^n|W_2') \\ &= H(W_2') + I(W_1W_2;Y_1^n|W_2') \\ &= \sum_{i=1}^n R_2' + H(Y_{1,i}|W_2'Y_1^{i-1}) - H(Y_{1,i}|X_i) \\ &\le \sum_{i=1}^n R_2' + I(X_i;Y_{1,i}) \\ &\le n\,[R_2' + I\,(X;Y_1)]\,. \end{aligned} \quad (38)$$

Similarly, we have

$$\begin{aligned} &n(R_1+R_2) \\ &\le I(W_1;\hat{W}_1) + I(W_2;\hat{W}_2(2)) \\ &\le I(W_1;Y_1^nW_1'W_2) + I(W_2;Y_2^nW_1') \\ &= I(W_1;W_1'W_2) + I(W_1;Y_1^n|W_1'W_2) + I(W_2;Y_2^n|W_1'). \end{aligned} \quad (39)$$

We compute $I(W_1;W_1'W_2) = H(W_1') = nR_1'$ and bound

$$\begin{aligned} &I(W_1;Y_1^n|W_1'W_2) \\ &= \sum_{i=1}^n I(W_1;Y_{1,i}|W_1'W_2Y_1^{i-1}) \\ &\le \sum_{i=1}^n I(W_1Y_{2,i+1}^n;Y_{1,i}|W_1'W_2Y_1^{i-1}) \\ &= \sum_{i=1}^n I(Y_{2,i+1}^n;Y_{1,i}|W_1'W_2Y_1^{i-1}) + I(W_1;Y_{1,i}|U_i). \end{aligned} \quad (40)$$

Note that $I(W_1;Y_{1,i}|U_i) = I(X_i;Y_{1,i}|U_i)$. We also have

$$\begin{aligned} &I(W_2;Y_2^n|W_1') \\ &= \sum_{i=1}^n I(W_2;Y_{2,i}|W_1'Y_{2,i+1}^n) \\ &= \sum_{i=1}^n I(W_2Y_1^{i-1};Y_{2,i}|W_1'Y_{2,i+1}^n) \\ &\quad - I(Y_1^{i-1};Y_{2,i}|W_1'W_2Y_{2,i+1}^n) \\ &\le \sum_{i=1}^n I(U_i;Y_{2,i}) - I(Y_1^{i-1};Y_{2,i}|W_1'W_2Y_{2,i+1}^n). \end{aligned} \quad (41)$$

We can now apply a standard identity (see [3, p. 409])

$$\sum_{i=1}^n I(Y_{2,i+1}^n;Y_{1,i}|AY_1^{i-1}) = \sum_{i=1}^n I(Y_1^{i-1};Y_{2,i}|AY_{2,i+1}^n) \quad (42)$$

for any random variable A, and simplify (39) as

$$\begin{aligned} & n(R_1+R_2) \\ & \le \sum_{i=1}^{n} R_1' + I(X_{1,i};Y_{1,i}|U_i) + I(U_i;Y_{2,i}) \\ & \le n\left[R_1' + I\left(X;Y_1|U\right) + I\left(U;Y_2\right)\right] \end{aligned} \tag{43}$$

Combining (36)-(38) and (43) completes the converse. ■

Example 2: Consider the following special cases.

- $R_1' = R_1$, $R_2' = R_2$: the bounds in (28) give

$$R_2 \le I(U;Y_2), \quad R_1 \le I(X;Y_1)$$

  and we recover Theorem 1 by choosing $U = X$.
- $R_1' = R_1$, $R_2' = 0$: the bounds in (28) give

$$R_2 \le I(U;Y_2), \quad R_1 + R_2 \le I(X;Y_1)$$

  and the best choice is again $U = X$.
- $R_1' = 0$, $R_2' = R_2$: the bounds in (28) give

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

- $R_1' = R_2' = 0$: the bounds in (28) give

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

  We thereby recover the capacity region of the broadcast channel with degraded message sets [3, p. 360].

V. Partial Side Information, Degraded Message Sets, and Gaussian Channels

Consider again the Gaussian channel of (23) and (24) but with the setup of Sec. IV. We prove the following.

Theorem 4: For Gaussian broadcast channels with (23) and (24) and the setup of Sec. IV, the capacity region (28) is

$$\begin{aligned} \mathcal{C} = \bigcup \{(R_1,R_2) : 0 \le R_1, 0 \le R_2, \\ R_2 \le \tfrac{1}{2}\log\left(1+\tfrac{\alpha P}{(1-\alpha)P+N_2}\right) \\ R_1 + R_2 \le R_2' + \tfrac{1}{2}\log\left(1+\tfrac{P}{N_1}\right) \\ R_1 + R_2 \le R_1' + \tfrac{1}{2}\log\left(1+\tfrac{(1-\alpha)P}{N_1}\right) \\ + \tfrac{1}{2}\log\left(1+\tfrac{\alpha P}{(1-\alpha)P+N_2}\right) \end{aligned} \Bigg\} \tag{44}$$

where the union is over α with $0 \le \alpha \le 1$. Moreover, if $N_2 \le N_1$ then $\alpha = 1$ is best for all rate pairs.

The proof follows by using standard steps that we repeat for completeness. For the coding theorem, we choose $X = U + V$ where U and V are independent, Gaussian random variables with zero mean and respective variances αP and $(1-\alpha)P$ for some α satisfying $0 \le \alpha \le 1$. We compute

$$I(U;Y_2) = \frac{1}{2}\log\left(1+\frac{\alpha P}{(1-\alpha)P+N_2}\right) \tag{45}$$

$$I(X;Y_1) = \frac{1}{2}\log\left(1+P/N_1\right) \tag{46}$$

$$I(X;Y_1|U) = \frac{1}{2}\log\left(1+(1-\alpha)P/N_1\right). \tag{47}$$

Inserting these expressions into (28), we obtain (44). One can easily check that $\alpha = 1$ is best if $N_2 \le N_1$.

For the converse, consider

$$\begin{aligned} I(X;Y_1) &= h(Y_1) - h(Y_1|X) \\ &\le \frac{1}{2}\log\left(2\pi e(P+N_1)\right) - \frac{1}{2}\log\left(2\pi e N_1\right) \\ &= \frac{1}{2}\log\left(1+P/N_1\right) \end{aligned} \tag{48}$$

where the inequality follows by the maximum entropy theorem. We similarly have

$$\begin{aligned} I(U;Y_2) &= h(Y_2) - h(Y_2|U) \\ &\le \frac{1}{2}\log\left(2\pi e(P+N_2)\right) - h(Y_2|U). \end{aligned} \tag{49}$$

We further have

$$\begin{aligned} & \frac{1}{2}\log(2\pi e N_2) = h(Y_2|X) \\ & \le h(Y_2|U) \le h(Y_2) \le \frac{1}{2}\log\left(2\pi e(P+N_2)\right) \end{aligned} \tag{50}$$

so there must exist an α, $0 \le \alpha \le 1$, such that

$$h(Y_2|U) = \frac{1}{2}\log\left(2\pi e[(1-\alpha)P+N_2]\right). \tag{51}$$

Consider first the case $N_2 \ge N_1$, and observe that we can write $Z_2 = Z_2' + Z_2''$ where Z_2' and Z_2'' are independent and have variances N_1 and $N_2 - N_1$, respectively. We use the entropy power inequality [2, p. 496] to bound

$$\begin{aligned} 2^{2h(Y_2|U)} &= 2^{2h(Y_1+Z_2''|U)} \\ &\ge 2^{2h(Y_1|U)} + 2^{2h(Z_2''|U)} \\ &= 2^{2h(Y_1|U)} + 2\pi e(N_2-N_1). \end{aligned} \tag{52}$$

Using (51), we therefore have

$$\begin{aligned} I(X;Y_1|U) &= h(Y_1|U) - h(Y_1|X) \\ &\le \frac{1}{2}\log\left(1+(1-\alpha)P/N_1\right). \end{aligned} \tag{53}$$

Combining (48), (49), (51), and (53) completes the converse for $N_2 \ge N_1$.

Finally, consider $N_2 \le N_1$ and write $Z_1 = Z_1' + Z_1''$ where Z_1' and Z_1'' are independent and have variances N_2 and $N_1 - N_2$, respectively. We have

$$\begin{aligned} I(X;Y_1|U) &\le I(X;Y_1 Z_1''|U) \\ &= I(X;Y_1|U Z_1'') \\ &= I(X;Y_2|U Z_1'') \\ &= I(X;Y_2|U) \end{aligned} \tag{54}$$

where the second and fourth steps follow because Z_1'' is independent of (U,X,Z_2). The last bound in (28) thus implies

$$\begin{aligned} R_1 + R_2 &\le R_1' + I(X;Y_2) \\ &\le R_1' + \frac{1}{2}\log\left(1+P/N_2\right). \end{aligned} \tag{55}$$

Combining (55) with (48), (49), and (51), we have a capacity region outer bound. But we achieve all points inside this outer bound by choosing $\alpha = 1$ in (45)-(47).

VI. Conclusions

We developed capacity theorems and coding strategies for broadcast channels with side information. The capacity result of Sec. IV might generalize to models with correlated sources, in a similar way that the results for the model of Fig. 1 do.

Appendix A
Alternative Proofs of Theorem 1

We prove the direct part of Theorem 1 in a slightly more direct way than in Sec. II. Suppose that $R_1 \leq R_2$. Generate 2^{nR_2} codewords $x^n(w)$, $w = 1, 2, \ldots, 2^{nR_2}$, by choosing every symbol $x_i(w)$, $i = 1, 2, \ldots, n$, independently using P_X. Next, map w_k to a unique integer $\mathrm{Int}(w_k)$ in $\{0, 1, \ldots, 2^{nR_2} - 1\}$ for $k = 1, 2$. The encoder transmits

$$x^n \left(\mathrm{Int}^{-1} \left(\mathrm{Int}(w_1) + \mathrm{Int}(w_2) \mod 2^{nR_2} \right) \right). \tag{56}$$

We abuse notation and write this as

$$x^n \left(w_1 + w_2 \mod 2^{nR_2} \right). \tag{57}$$

Receiver 1 observes y_1^n and w_2, and tries to find a w_1 such that

$$\left(x^n(w_1 + w_2 \mod 2^{nR_2}), y_1^n \right) \in T_\delta^M \left(P_{XY_1} \right). \tag{58}$$

Receiver 2 decodes similarly. Standard typicality arguments (cf. [2, Ch. 14]) show that w_1 and w_2 are decoded reliably if n is large, δ is small, and

$$R_k < I(X; Y_k), \quad k = 1, 2. \tag{59}$$

By symmetry, we have the same rate bounds if $R_1 \geq R_2$.

One can, in fact, perform random coding in several other ways. For example, for $R_1 \leq R_2$ one could encode w_1 with a binary code of rate R_1/R_2, then XOR the resulting codeword $x_1^{nR_2}(w_1)$ with w_2, and finally encode the result to $x^n(x_1^{nR_2}(w_1) \oplus w_2)$. Each of these methods has its own advantages. For example, the last approach protects the w_1 bits more than the w_2 bits since w_1 is encoded twice.

Appendix B
Properties of Typical Sequences

Define the set of δ-*typical sequences* with respect to P_X as in (8). Define the *support set* of X to be $\mathcal{S}_X = \{a \in \mathcal{X} : P_X(a) > 0\}$. Let μ_X be the smallest positive probability of P_X. The set of *jointly* δ-typical sequences with respect to P_{XY} is

$$T_\delta^L(P_{XY}) = \{ (x^L, y^L) : |N(a, b|x^L, y^L)/L - P_{XY}(a, b)| \leq \delta \cdot P_{XY}(a, b) \text{ all } i, j \} \tag{60}$$

where $N(a, b|x^L, y^L)$ is the number of times the pair (a, b) occurs in the sequence $(x_1, y_1), (x_2, y_2), \ldots, (x_L, y_L)$. Define $\mathcal{S}_{XY} = \{(a, b) \in \mathcal{X} \times \mathcal{Y} : P_{XY}(a, b) > 0\}$ and let μ_{XY} be the smallest positive probability of P_{XY}.

The following properties of typical sequences and sets are proved in [9, Appendix]. The logarithms are to the base 2.

Lemma 1: Let $0 < \delta \leq 1$ and

$$\epsilon_\delta(L) = 2|\mathcal{S}_X| e^{-\delta^2 \cdot \mu_X \cdot L/3}. \tag{61}$$

We have

$$(1 - \epsilon_\delta(L)) \cdot 2^{(1-\delta)H(X)L} \leq |T_\delta^L(P_X)| \leq 2^{(1+\delta)H(X)L}. \tag{62}$$

Lemma 2: For all $(x^L, y^L) \in T_\delta^L(P_{XY})$ we have

$$2^{-(1+\delta)H(Y|X)L} \leq \prod_{i=1}^{L} P_{Y|X}(y_i|x_i) \leq 2^{-(1-\delta)H(Y|X)L}. \tag{63}$$

Lemma 3: Let $0 < \delta_1 < \delta_2 \leq 1$ and

$$\epsilon_{\delta_1, \delta_2}(L) = 2|\mathcal{S}_{XY}| e^{-\frac{(\delta_2 - \delta_1)^2}{1+\delta_1} \cdot \mu_{XY} \cdot L/3}. \tag{64}$$

Suppose $x^L \in T_{\delta_1}^L(P_X)$ is input to $P_{Y|X}$ that puts out Y^L. We have

$$\Pr[(x^L, Y^L) \in T_{\delta_2}^L(P_{XY}) | X^L = x^L] \geq 1 - \epsilon_{\delta_1, \delta_2}(L). \tag{65}$$

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*, Budapest: Akadémiai Kiadó, 1981.

[4] U. Erez and R. Zamir, "Achieving 1/2log(1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[5] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.

[6] C. Fragouli, J. Widmer, and J.-Y. Le Boudec, "Network coding: an instant primer," *ACM SIGCOM Comp. Commun. Review*, vol. 36, no.1, pp. 63–68, Jan. 2006.

[7] G. Kramer, I. Marić, and R. D. Yates, *Cooperative Communications*. Foundations and Trends in Networking. Hanover, MA: now Publishers, Inc., vol. 1, no. 3-4, June 2007.

[8] T. Oechtering, I. Bjelakovic, C. Schnurr, and H. Boche, "Broadcast capacity region of two-phase bidirectional relaying," http://arxiv.org/abs/cs/0703078v1, March 2007.

[9] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 903–917, March 2001.

[10] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," *Proc. 2006 IEEE Int. Symp. Inf. Theory*, Seattle, WA, July 2006, pp. 1668–1672.

[11] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1469–1482, April 2006.

[12] L. Xiao, T. E. Fuja, J. Kliewer, and D. J. Costello, Jr., "Nested codes with multiple interpretations," *Proc. 2006 Conf. Inf. Sci. Sys. (CISS)*, Princeton, NJ, March 22-24, 2006.

[13] L. Xiao, T. E. Fuja, J. Kliewer, and D. J. Costello, Jr., "Cooperative diversity based on code superposition," *Proc. 2006 Int. Symp. Inf. Theory (ISIT)*, Seattle, WA, July 9-14, 2006.

[14] L.-L. Xie, Presentation at 2007 IEEE Commun. Theory Workshop (CTW), Sedona, AZ, May 20-23, 2007.

[15] L.-L. Xie, "Network coding and random binning for multi-user channels," *2007 Can. Workshop Inf. Theory (CWIT)*, Edmonton, AB, Canada, June 6-8, 2007.