Proof (Part ii): Since $f(t)$ is a nondecreasing function of t, (1) and (2) yield

$$E[d_n(X,\tilde{X})] \geq \frac{f(n)}{n}\left(E\sum_{t=0}^{n-1}(1-\delta(X_t,\tilde{X}_t))\right)$$
$$\geq \frac{f(n)}{n}E[1-\delta(X,\tilde{X})]$$
$$= P_e f(n)/n. \tag{5}$$

For equally probable source vectors, Theorem 1-ii lower-bounds P_e and thus

$$E[d_n(X,\tilde{X})] \geq \frac{f(n)}{n}\exp\left[-n[E_{sp}(R-o_1(n))+o_2(n)]\right]. \tag{6}$$

Part ii easily follows from (6).

Proof (Part i): Since $[1-\delta(x,\tilde{x})]^2 = [1-\delta(x,\tilde{x})]$, applying the Schwartz inequality to (1) yields

$$E[d_n(X,\tilde{X})] \leq \left(n^{-1}\sum_{t=0}^{n-1}[f(t+n)]^2\right)^{1/2}\left(E\left[n^{-1}\sum_{t=0}^{n-1}(1-\delta(X_t,\tilde{X}_t))\right]\right)^{1/2}. \tag{7}$$

Since $[n^{-1}\sum_{t=0}^{n-1}(1-\delta(x_t,\tilde{x}_t))] \leq [1-\delta(x,\tilde{x})]$, Theorem 1-*i* and (7) yields

$$E[d_n(X,\tilde{X})] \leq \left(n^{-1}\sum_{t=0}^{n-1}[f(t+n)]^2\right)^{1/2}(\exp[-nE_r(R)])^{1/2}. \tag{8}$$

Part i easily follows from (8) and Theorem 1-iii.

Thus, $f(t)$ can grow exponentially (but not too fast) and still have $E[d_n(X,\tilde{X})] \to 0$. In the following theorem it is interesting to see that a similar property does not hold for $C < H(X)$.

Theorem 3: Assume that the source letters are independent and identically distributed and that

$$d_n(x,\tilde{x}) \geq n^{-1}\sum_{t=0}^{n-1} f(t)\rho(x_t,\tilde{x}_t) \tag{9}$$

where $\rho(x,\tilde{x})$ is nonnegative and $f(t)$ is a nondecreasing function of t such that $\lim_{t\to\infty} f(t) = \infty$. Let $R_\rho(D)$ be the rate distortion function with respect to $\rho(x,\tilde{x})$. If the channel capacity $C < R_\rho(0) \triangleq \lim_{D\to\infty} R_\rho(D)$, then for every sequence of block length n codes

$$\lim_{n\to\infty} E[d_n(X,\tilde{X})] = \infty. \tag{10}$$

Note: Usually $R_\rho(0) = H(X)$. Thus this theorem shows that for any channel of capacity $C < H(X)$, infinite block length codes perform very poorly no matter how slowly $f(t) \to \infty$.

Proof: Let

$$R_n(D) = \min_{[D_t: n^{-1}\sum_{t=0}^{n-1} D_t = D]}\left[n^{-1}\sum_{t=0}^{n-1} R^t(D_t)\right] \tag{11}$$

where $R^t(D_t) = R_\rho(D_t/f(t))$. $R_n(D)$ may be thought of as the rate-distortion function for the product of n sources with sum distortion measure where the tth source has distortion measure. For equally probable source vectors, Theorem 1-ii lower-$f(t)\rho(x,x)$. (See Berger [4, sect. 2.8].) Therefore, if $C < R/(D)$, then $E[d/(X,\tilde{X})] > D$, for all block length n codes. Since $f(t)$ is nondecreasing and $R!(D)$ is convex upward and nonincreasing, it follows that for any N

$$R_n(D) \geq \min_{[D_t: n^{-1}\sum_{t=0}^{n-1} D_t = D]}\left[n^{-1}\sum_{t=N}^{n-1} R_\rho(D_t/f(N))\right]$$
$$= \frac{n-N}{n}R_\rho\left[\left(\frac{n}{n-N}\right)\left(\frac{D}{f(N)}\right)\right]. \tag{12}$$

Therefore, $\lim_{n\to\infty} R_n(D) \geq R_\rho(D/f(N))$. Since N is an arbitrarily large number and $f(t) \to \infty$, we have $\lim_{n\to\infty} R_n(D) = R_\rho(0)$. The theorem is proven since for any $D > 0$ and sufficiently large n, we have $R_n(D) > C$.

References

[1] S. I. Krich and T. Berger, "Coding for a delay-dependent fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 77–85, Jan. 1974.
[2] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Inform. Contr.*, vol. 10, pp. 65–103 (Part I), p. 522 (Part II), 1967.
[3] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.
[4] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Englewood Cliffs, N.J.: Prentice-Hall, 1971.
[5] S. I. Krich, "Coding for a time varying fidelity criterion," Ph.D. dissertation, Dep. Elec. Eng., Cornell Univ., Ithaca, N.Y., 1972.

A Simple Converse for Broadcast Channels with Additive White Gaussian Noise

PATRICK P. BERGMANS

Abstract—Sets of achievable rates for the additive white Gaussian noise broadcast channel have been found by Cover [1] for channels with two outputs and generalized by Bergmans [2] to channels with any number of outputs. In this correspondence, we establish a simple converse showing the optimality of these sets of achievable rates. The proof is made simple by use of special properties of the Gaussian channel.

The set of achievable rates for discrete-time additive white Gaussian noise (AWGN) broadcast channels with noise powers $N_1 < N_2 \cdots N_N$ in the different links and average input power S is given by [1], [2]

$$R_i \leq \frac{1}{2}\ln\left(1+\frac{\alpha_i S}{N_i + \sum_{j<i}\alpha_j S}\right), \qquad i = 1,\cdots,N \tag{1}$$

with $\alpha_i \geq 0, \sum \alpha_i = 1$.

In this correspondence, we shall use the equivalent notation

$$R_i \leq g(N_i + \beta_i S) - g(N_i + \beta_{i-1}S), \qquad i = 1,\cdots,N \tag{2}$$

with

$$g(S) \triangleq \tfrac{1}{2}\ln(2\pi e S) \tag{3}$$

and

$$0 = \beta_0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_N = 1. \tag{4}$$

We shall need the following lemmas.

Lemma 1: Let X be a random n-tuple such that $X \in R^n$, $H(X) \geq nv$, and let $Y = X + Z$ where Z is a Gaussian noise vector independent of X with independent zero-mean components of variance N. Then

$$H(Y) \geq ng(N + g^{-1}(v)). \tag{5}$$

Manuscript received March 12, 1973; revised September 13, 1973. This work was supported in part by the National Science Foundation under Grant GK-32240.
The author is with the School of Electrical Engineering, Cornell University, Ithaca, N.Y. 14850.

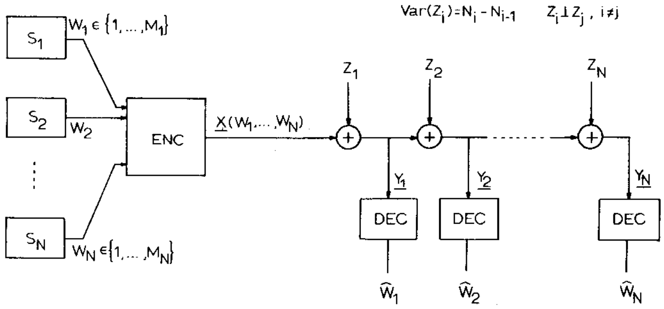

Fig. 1.

Lemma II: Consider the ensemble (X,Y,W), such that $X \in R^n$, $W \in I_1 \times I_2 \times \cdots \times I_k$ (a product of index sets), $H(X \mid W) \geq nv$, and let $Y = X + Z$, with Z as in Lemma I. Then

$$H(Y \mid W) \geq ng(N + g^{-1}(v)). \tag{6}$$

Lemma I is a straightforward consequence of an inequality of Shannon on the entropy power of the sum of two ensembles [3, p. 636], [4]. It is the counterpart for the Gaussian case of the results obtained by Wyner and Ziv for the binary case [5]. Lemma II is just the conditional version of Lemma I, and can be established easily using Lemma I, the convexity of $g(N + g^{-1}(v))$ in v and Jensen's inequality.

We now state the converse to the coding theorem.

Theorem: No point $(R_1 \cdots R_N)$ such that

$$R_i \geq g(N_i + \beta_i S) - g(N_i + \beta_{i-1}S), \qquad i = 1,\cdots,N$$
$$R_j = g(N_j + \beta_j S) - g(N_j + B_{j-1}S) + \delta, \qquad \text{some } j, \delta > 0 \tag{7}$$

is achievable, where the β_j are as in (4).

Proof (by Contradiction): We suppose that the rates of (7) are achievable. We refer to Fig. 1 for the meaning of the various quantities. The outputs of the sources are independent and equiprobable.

If the rates of (7) are achievable, the probability of decoding error for each receiver can be upper bounded by an arbitrarily small λ (for sufficiently large n)

$$\Pr [\hat{W}_i \neq W_i \mid Y_i] < \lambda, \qquad i = 1,\cdots,N. \tag{8}$$

This implies (Fano's inequality)

$$H(W_i \mid Y_i) \leq h(\lambda) + \lambda \log (M_i - 1), \qquad i = 1,\cdots,N. \tag{9}$$

Using standard information-theoretic arguments and the independence of the W_i, we find, from (9)

$$\log M_i \leq I(Y_i; W_i \mid W_{i+1} \cdots W_N) + \varepsilon_i(\lambda)$$
$$= H(Y_i \mid W_{i+1} \cdots W_N) - H(Y \mid W_i \cdots W_N) + \varepsilon_i(\lambda) \tag{10}$$

where $\varepsilon_i(\lambda) \to 0$ as $\lambda \to 0$.

From (7)

$$\log M_i \geq ng(N_i + \beta_i S) - ng(N_i + \beta_{i-1}S). \tag{11}$$

Next, using Lemma II, (10) (an upper bound for $\log M_i$) and (11) (a lower bound for $\log M_i$), we show in the Appendix

$$H(Y_i \mid W_{i+1} \cdots W_N) \geq ng(N_i + \beta_i S) - \eta_i(\lambda) \tag{12}$$

where $\eta_i(\lambda) \to 0$ as $\lambda \to 0$, and where $n\delta$ should be added to the right side of (12), for $i \geq j$ (where j was defined in the statement of the theorem). It follows from (12), with $i = N \geq j$, that

$$H(Y_N) \geq ng(N_N + S) - \eta_N(\lambda) + n\delta. \tag{13}$$

Since λ, and hence $\eta_N(\lambda)$, can be made arbitrarily small, we may conclude that, for sufficiently large n, $H(Y_N) \geq ng(N_N + S) + n\delta$. This, however, is impossible, since $\operatorname{var}(Y_N) \leq N_N + S$ and, therefore, $H(Y_N) \leq ng(N_N + S)$. The desired contradiction has been obtained, so the theorem is proved.

Appendix

Proof of Equation (12)

Letting $i = 1$ in (10) and (11), we find

$$H(Y_1 \mid W_2 \cdots W_N) - H(Y_1 \mid W_1 \cdots W_N) + \varepsilon_1(\lambda) \geq ng(N_1 + \beta_1 S) - ng(N_1)$$

or

$$H(Y_1 \mid W_2 \cdots W_N) \geq ng(N_1 + \beta_1 S) - \eta_1(\lambda) \tag{14}$$

since

$$H(Y_1 \mid W_1 \cdots W_N) = H(Y_1 \mid X) = ng(N_1).$$

The rest of the proof is by recursion. We assume that (12) is true. Writing (10) and (11) for $i + 1$, we find

$$H(Y_{i+1} \mid W_{i+2} \cdots W_N) \geq ng(N_{i+1} + \beta_{i+1}S) - ng(N_{i+1} + \beta_i S) + H(Y_{i+1} \mid W_{i+1} \cdots W_N) - \varepsilon_{i+1}(\lambda). \tag{15}$$

From (12)

$$H(Y_i \mid W_{i+1} \cdots W_N) \geq n \left[g(N_i + \beta_i S) - \frac{\eta_i(\lambda)}{n} \right]. \tag{16}$$

Since Y_{i+1} is obtained by adding Gaussian noise of variance $N_{i+1} - N_i$ to Y_i, we have, using Lemma II

$$H(Y_{i+1} \mid W_{i+1} \cdots W_N)$$
$$\geq ng \left[(N_{i+1} - N_i) + g^{-1} \left(g(N_i + \beta_i S) - \frac{\eta_i(\lambda)}{n} \right) \right]$$
$$= ng \left[(N_{i+1} - N_i) + g^{-1} g(N_i + \beta_i S) - \frac{\zeta_i(\lambda)}{n} \right]$$
$$= ng(N_{i+1} + \beta_i S) - \xi_i(\lambda) \tag{17}$$

where $\zeta(_i\lambda)$, $\xi_i(\lambda) \to 0$ as $\lambda \to 0$. Substitution of (17) in (15) yields

$$H(Y_{i+1} \mid W_{i+2} \cdots W_N) \geq ng(N_{i+1} + \beta_{i+1}S) - \eta_{i+1}(\lambda) \tag{18}$$

with $\eta_{i+1}(\lambda) = \varepsilon_{i+1}(\lambda) + \xi_i(\lambda) \to 0$ as $\lambda \to 0$. Equation (18) establishes the recursion. Finally, for $i \geq j$, $n\delta$ should be added to the right side of (18) because of the presence of δ in (7) for $i = j$, and, hence, of $n\delta$ in (11).

References

[1] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[2] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197–207, Mar. 1973.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 (pt. I), pp. 623–656 (pt. II), 1948.

[4] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 267–271, Apr. 1965.

[5] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 769–772, Nov. 1973.