The Annals of Probability
1975, Vol. 3, No. 3, 478-491

NONBLOCK SOURCE CODING WITH A FIDELITY CRITERION[1]

BY ROBERT M. GRAY, DAVID L. NEUHOFF
AND DONALD S. ORNSTEIN

Stanford University, University of Michigan
and Stanford University

A new nonblock source coding (data compression) technique is introduced and a source coding theorem is proved using recently developed techniques from ergodic theory. The existence theorem is valid for all stationary aperiodic sources (e.g., ergodic sources) with finite alphabets and all ergodic sources with separable alphabets and is proved without Shannon-style random coding arguments. The coding technique and the optimal performance bounds are compared and contrasted with Shannon block coding techniques.

1. Introduction. Since Shannon's (1948), (1959) original development of the theory of source coding subject to a fidelity criterion, the theory has dealt almost exclusively with block coding, i.e., "compressing" a source by mapping consecutive nonoverlapping blocks of source data into an allowed codebook containing a constrained number of reproduction blocks or codewords. The fundamental theorems relating optimal source code performance with an information theoretic minimization—the rate-distortion function—have been notoriously difficult to prove in general cases, involving complex random coding arguments coupled with the decomposition of sequences of n-tuples from ergodic sources into ergodic modes as in Gallager (1968) and Berger (1971), and the decomposition of stationary sources into ergodic subsources as in Gray and Davisson (1974).

In many situations, block coding structures are exceedingly difficult to implement and numerous existing algorithms for real-world data compression such as the interpolating and predictive compression schemes described by Davisson (1968) do not have a block structure and hence, cannot be studied using the Shannon formulation.

In this paper, a new source coding technique dubbed "sliding-block source coding" is introduced and the relevant source coding theorem proved. The coding technique is derived from the work of Ornstein (1973) and the theorem is proved using recently developed techniques of ergodic theory as described, e.g., in Shields (1973). In particular, the proof of the coding theorem is based on a simple geometric picture of stationary aperiodic processes (such as ergodic processes) due to Rohlin as described by Shields (1973), and a generalization by

Received May 2, 1974.

[1] This work was supported by NSF Grants GK-31630 and GJ 776, and by the JSEP program at Stanford under U.S. Navy Grant N00014-67-A-0112-0044.

AMS 1970 *subject classifications.* Primary 60G35, 94A15; Secondary, 94A05.

Key words and phrases. Source coding with a fidelity criterion information and ergodic theory.

Gray, Neuhoff, and Shields (1974) of a distance between random processes developed by Ornstein (1973). Somewhat surprisingly, the theorem proof involves no traditional Shannon-style random coding arguments.

As a simple example of the coding technique, consider the source coder of Fig. 1. The binary source data is shifted each time-unit by one letter through

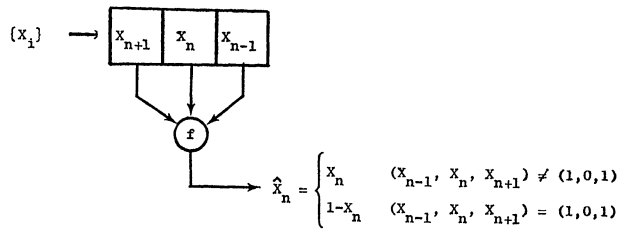

FIG. 1. A simple sliding-block code.

the shift register and at each time the encoded letter representing a reproduction of the center source letter is output. In the example, the reproduction letter agrees with the central letter in the shift register unless the shift register contains the pattern "101," in which case the reproduction bit is the complement of the center source bit. If the source is a binary independent, identically distributed sequence with equiprobable zeros and ones, then the average error rate between X_n and $\hat{X}_n$ is easily seen to be $\frac{1}{8}$. "Compression" is achieved in that the sequence $\{\hat{X}_n\}$ has a reduced entropy rate, shown by computer evaluation to be less than 0.7 bit per symbol. The reproduction process can therefore be transmitted reliably over a channel with reduced capacity. Note that compression here is in the sense of entropy rate reduction and not of "redundancy removal" since the reproduction sequence has memory. "Compression" in the sense of actually sending fewer binary digits can be achieved by following $\{\hat{X}_n\}$ by a block-to-variable length noiseless source coder such as a Huffman coder as described, for example, in Gallager (1968). We note that such noiseless coders are fairly easily implemented and often used in real systems.

The structure of sliding block encoders resembles that of non-linear convolutional channel encoders suggesting that possibly both operations of "compression" and reliable communication over a noisy channel can be performed by a single joint source-channel encoder as first observed in the noiseless case by Koshelev (1973) and Hellman (1974). The observation that sliding-block source encoders insert redundancy while reducing entropy rate supports this conjecture since, hopefully, the redundancy can be inserted in a controlled way to combat channel noise. We therefore feel that the techniques used here to study the source coding theorem with a noiseless channel may prove useful in reformulating noisy channel and joint source-channel coding theorems.

Modern ergodic theory shares with information theory much of its origins in the work of Shannon (1948). We hope that the methods and results described herein may contribute to the realization of Krengel's (1973) prediction that

"the pendulum may now swing back again" from abstract ergodic theory to the Shannon theory of communication.

2. Notation and definitions. In this section we introduce the notation and definitions necessary to formulate the problem and state the coding theorems. Additional notation and definitions required only in the proofs will be provided when needed.

Let A be the source *alphabet* or space of possible outputs at any given time. The alphabet A will be assumed to be either finite or a separable complete metric space. Let $\mathscr{B}$ denote a σ-field of subsets of A: the class of all subsets of A if A is finite, the Borel σ-field if A is a metric space. Let $\Sigma = A^\infty$ denote the sequence space of all possible doubly infinite sequences drawn from A; i.e., if $x \in \Sigma$, then $x = (\cdots, x_{-1}, x_0, x_1, \cdots)$, $x_i \in A$, all i. Let $X_n : \Sigma \to A$ denote the coordinate function $X_n(x) = x_n$. Let T denote the shift operation on Σ, i.e., $X_n(Tx) = x_{n+1}$. Let $S = \mathscr{B}^\infty$ denote the smallest σ-field containing all cylinders of the form $\{x : x_i \in B_i;\ m \leqq i \leqq n\}$, where $B_i \in \mathscr{B}$, $m \leqq i \leqq n$, for all finite integers n and m. If A is finite then S is also generated by the thin cylinders of the form $\{x : x_i = a_i;\ m \leqq i \leqq n\}$ where $a_i \in A$, $m \leqq i \leqq n$. Let μ be a stationary measure on (Σ, S), i.e., $\mu(TB) = \mu(B)$, all $B \in S$. The sequence of random variables $\{X_n\}_{n=-\infty}^{\infty}$ defined on the probability space (Σ, S, μ) is then a (directly-given) discrete-time stationary random process and is called the *source*. The source is sometimes denoted by $[A, \mu]$. Let μ^N denote the restriction of μ to $(A^N, \mathscr{B}^N)$.

The source is aperiodic if $\mu(\{x : T^n x = x\}) = 0$ for each n. A source is ergodic if $TB = B$ implies that $\mu(B) = 0$ or $\mu(B) = 1$, i.e., the only events left unchanged by shifting are trivial. All nontrivial ergodic processes are aperiodic.

Let $H(X^N)$ denote the entropy of the random vector $X^N = (X_0, \cdots X_{N-1})$, i.e., if $(A^N, \mathscr{B}^N, \mu^N)$ is atomic then

$$H(X^N) = -\sum_{x^N \in A^N} \mu^N(x^N) \log \mu^N(x^N)$$

where μ^N also denotes the probability mass function for the vector X^N. If $(A^N, \mathscr{B}^N, \mu^N)$ is not atomic, then $H(X^N) = \infty$. The *entropy rate* $H(X)$ or $H([A, \mu])$ or H_μ of a process is defined by $H(X) = \lim_{N\to\infty} N^{-1}H(X^N)$. The various equivalent notations will prove useful.

Let $\hat{A}$ be the *available reproducing alphabet*, i.e., the set of allowable reproduction letters. For any integer N, a *sliding-block source encoder* of blocklength $2N + 1$ is any function $f^{(N)} : A^{2N+1} \to \hat{A}$. The reproduction process $\{\hat{X}_n\}_{n=-\infty}^{\infty}$ is defined by $\hat{X}_n = f^{(N)}(X_{n-N}, \cdots, X_n, \cdots, X_{n+N})$. If X_n is ergodic or only stationary, then $\hat{X}_n$ is ergodic or stationary, respectively. As in the Shannon theory, we consider the source coding problem as separated from the channel coding problem, or equivalently, we assume a noiseless channel with input and output alphabet $\hat{A}$. Hence, the source decoder is simply an identity transformation.

Let ρ be a nonnegative distortion measure on $(A \cup \hat{A}) \times (A \cup \hat{A})$. If A is finite, then ρ may be any finite-valued distortion measure. If A is not discrete,

but is a metric space, we take ρ to be the metric (metric distortion measure) and assume that $A \cup \hat{A}$ is a separable complete metric space under ρ. A sliding-block source code $f^{(N)}$ has *average distortion*

$$\rho(f^{(N)}) = E_\mu\{\rho(X_0, f^{(N)}(X_{-N}, \cdots, X_N))\},$$

where E_μ denotes expectation over μ, and *entropy rate* $H(f^{(N)}) =_\Delta H(\hat{X})$, where $H(\hat{X})$ is the entropy rate of the induced reproduction process.

It is easily shown that if the source is ergodic, then the induced joint process $\{X_n, \hat{X}_n\}_{n=-\infty}^{\infty}$ is also ergodic and, hence, $\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \rho(X_k, \hat{X}_k) = \rho(f^{(N)})$ almost everywhere. This desirable convergence of time average distortion to its expected value does not follow immediately in the traditional non-stationary block-coding formulation.

The object of source coding is to produce a reproduction process having an entropy rate less then some fixed number (possibly due to later channel capacity or storage constraints) such that the "compressed" reproduction well approximates the original process in the ρ sense. We therefore define for rate $R > 0$ and blocklength $2N + 1$ the optimal performance theoretically attainable (abbreviated OPTA) using sliding-block codes by

$$\delta(R, N) = \inf_{f^{(N)}: H(f^{(N)}) \leq R} \rho(f^{(N)}).$$

As in the usual Shannon approach, we are interested in the optimal performance over all blocklengths and hence we define $\delta(R) = \inf_N \delta(R, N)$. It is easily shown that the limit of $\delta(R, N)$ as $N \to \infty$ exists and equals the infimum.

Unlike the block coding case, it is easy to consider the infinite blocklength case, i.e., let $f^{(\infty)}: A^\infty \to \hat{A}$ be an infinite blocklength sliding-block code and define

$$\delta^*(R) = \inf_{f^{(\infty)}: H(f^{(\infty)}) \leq R} \rho(f^{(\infty)}).$$

Clearly, $\delta(R) \geqq \delta^*(R)$, but we shall later see that $\delta(R) = \delta^*(R)$, i.e., that finite blocklength codes perform in the limit of long blocklength as well as a code allowed to view the entire source sample function. We note that such infinite codes were suggested by Krengel (1973) in his discussion of Ornstein's (1973) paper.

In another departure from the usual approach, the basic theorem will not involve Shannon's (1959) distortion-rate function (DRF) explicitly. It is shown by Gray, Neuhoff, and Omura (1975), however, that the optimal performance bound is given by the usual DRF when the source is ergodic. Instead of the DRF, we here use the concept of the $\bar{\rho}$ distance between processes. Given the source $\{X_n\}$ defined on the probability space (Σ, S, μ) and any stationary random process $\{Y_n\}$ defined in a similar manner on a probability space $(\hat{\Sigma} = \hat{A}^\infty, \hat{S}, \nu)$, the $\bar{\rho}$ distance $\bar{\rho}(X, Y)$ or $\bar{\rho}([A, \mu], [\hat{A}, \nu])$ can be defined as follows:

$$\bar{\rho}([A, \mu], [\hat{A}, \nu]) = \inf_{p \in P_{\mu,\nu}} E_p[\rho(X_0, Y_0)]$$

$P_{\mu,\nu} = \{\text{all stationary measures } p \text{ on } (\Sigma \times \hat{\Sigma}, S \times \hat{S}) : p(\Sigma \times G) = \nu(G), \forall\, G \in \hat{S},$

$p(B \times \hat{\Sigma}) = \mu(B), \forall B \in S\}$, that is, $P_{\mu,\nu}$ is the class of all measures on random sequences of pairs such that one coordinate is probabilistically identical to $\{X_n\}$ and the other coordinate is similarly identical to $\{Y_n\}$. $\bar{\rho}$ measures how well two processes can be matched up in an average ρ sense at a given time if the two processes are stochastically linked in an optimal stationary manner. Several equivalent definitions for $\bar{\rho}$ and some properties and applications are given in Gray, Neuhoff, and Shields (1974). In particular, if ρ is a metric, then so is $\bar{\rho}$.

3. The sliding-block source coding theorem. In this section we state and discuss the two basic source coding theorems, the lemmas used to prove the theorems, and some related results.

The basic results of this paper are the following.

THEOREM 1. *If $[A, \mu]$ is an aperiodic random process and if A and $\hat{A}$ are finite and ρ is an arbitrary finite-valued distortion measure, then*

$$\delta(R) = \inf_{[\hat{A},\nu]: H_\nu \leqq R} \bar{\rho}([A, \mu], [\hat{A}, \nu]) . \tag{3.1}$$

THEOREM 2. *If $[A, \mu]$ is an ergodic random process and if A and $\hat{A}$ are such that $A \cup \hat{A}$ is a separable complete complete metric space under a finite-valued metric ρ and if there exists a reference letter $a_0 \in A$ such that*

$$E_\mu \rho(X_0, a_0) \leqq \rho^* < \infty , \tag{3.2}$$

then (3.1) *holds.*

Note that the "negative" side of the coding theorem

$$\delta(R) \geqq \inf_{[\hat{A},\nu]: H_\nu \leqq R} \bar{\rho}([A, \mu], [\hat{A}, \nu])$$

is obvious in both cases since an f^N approximately yielding $\delta(R)$ to within ε produces a process ν with $H_\nu \leqq R$ and $\bar{\rho}([A, \mu], [\hat{A}, \nu]) \leqq \delta(R) + \varepsilon$.

The first theorem is not a special case of the second since, even though A and $\hat{A}$ are more restricted in Theorem 1, ρ need not be a metric and the source need not be ergodic.

Roughly speaking, both theorems state that the optimal attainable distortion using a sliding-block source coder is given by the minimal $\bar{\rho}$ distance between the source and any process $[\hat{A}, \nu]$ with the desired entropy rate. The theorems follow from the following lemmas.

LEMMA 1. *Given two finite alphabet stationary aperiodic processes $[A, \mu]$ and $[\hat{A}, \nu]$, a (finite-valued) distortion measure ρ, and $\delta > 0$, there exists for $N = N(\delta)$ sufficiently large a sliding-block code $f^{(N)}$ for $[A, \mu]$ such that*

$$\begin{aligned} \rho(f^{(N)}) &\leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + \delta \\ H(f^{(N)}) &\leqq H([\hat{A}, \nu]) + \delta . \end{aligned} \tag{3.3}$$

LEMMA 2. *Given two stationary ergodic processes $[A, \mu]$ and $[\hat{A}, \nu]$ such that*

(i) *$A \cup \hat{A}$ is a separable metric space under a (finite-valued) metric ρ, and*
(ii) *there exists a reference letter $a_0 \in A$ satisfying* (3.2),

then for any $\delta > 0$, *there exists for* $N = N(\delta)$ *sufficiently large a sliding-block code* $f^{(N)}$ *satisfying* (3.3).

The lemmas state roughly that the optimal behavior between the coordinates of a random process of pairs can be mimicked arbitrarily closely by a deterministic sliding-block encoding of one process. The theorems follow from the lemmas simply by optimal choice of $[\hat{A}, \nu]$. Lemma 1 and Theorem 1 are proved in the next section and demonstrate the basic approach. Lemma 2, and hence Theorem 2, follow from Lemma 1 via straightforward two-step encoding procedures involving quantization as in Gray *et al* (1974), (1975), and hence their proofs are relegated to an appendix. The two-step procedure, however, requires ergodicity.

The results described have several interesting similarities and differences with the usual Shannon-style block source coding theorems as stated, e.g., in Berger (1971), and Gray, Neuhoff, and Shields (1975). In each case, a deterministic optimum is related to a stochastic optimum. In the Shannon case, the stochastic optimum (the distortion-rate function) is in terms of a limit of optimizations over finite dimensional vectors with a constraint on the average mutual information between randomly chosen source and reproduction vectors. Here, the stochastic optimum is directly in terms of processes and is in terms of the entropy rate of the reproduction. This viewpoint resembles a conjecture of Dick (1973) and is discussed in some depth in Gray, Neuhoff, and Omura (1975).

The usual proofs involve the random generation of an ensemble of code-books and prove that there exists at least one code in the ensemble that works nearly optimally since the average over the ensemble is close to the optimal. In the proof of the sliding-block coding theorem, no such ensemble of randomly generated codes is used. Instead, the encoder is constructed based on a geometric picture of random processes called a gadget. The aperiodicity (or ergodicity) is used only to construct such a gadget and the proof then consists of copying one gadget representing a joint process onto another representing the source. Hence the ergodic theorem or law of large numbers are not used and there is no average over an ensemble of codes. The encoder $f^{(N)}$ is given by a mathematical construction, but it is likely difficult construction in a practical sense and the source coding theorems should be considered only as existence theorems. It is hoped, however, that the existence results will stimulate research on potentially tractable implementable sliding-block source codes, and on the construction of entropy reducing functions with good fidelity.

The block source coding theorem relates a corresponding $\delta(R)$ function to the distortion rate function $D(R)$ as defined, for example, in Gray, Neuhoff, and Shields (1975). The potential use of this result is that the DRF can usually be evaluated via computer (Blahut (1972)), while the direct evaluation of $\delta(R)$ is usually intractable. It is shown by Gray, Neuhoff, and Omura (1974) that for ergodic sources,

$$\inf_{[\hat{A},\nu]:H_\nu \leq R} \bar{\rho}([A, \mu], [\hat{A}, \nu]) = D(R)$$

so that the optimal performance using sliding-block codes is (not surprisingly) the same as that using block codes and hence the Blahut (1972) algorithm is also applicable to sliding-block performance bounds.

Contained in the proof of the previous coding theorems is the following corollary, also proved in the next section

COROLLARY 1.

$$\delta^*(R) = \delta(R) = \lim_{N\to\infty} \delta(R, N) .$$

In words, an infinite blocklength sliding-block code can be approximated arbitrarily well by a finite blocklength sliding-block code of sufficiently large blocklength.

4. Proof of the finite alphabet source coding theorem. Let $\{X_n\}$ be a random process defined as previously on a sequence probability space (Σ, S, μ) with a finite alphabet $A = \{a_1, \cdots, a_K\}$. Let $P = \{P_1, \cdots, P_K\}$ be the partition of Σ according to the zero coordinate, that is, the *atoms* P_k are given by $P_k = \{x: x \in \Sigma, X_0(x) = x_0 = a_k\}$. Note that $X_n(x) = X_0(T^n x) = a_k$ if $T^n x \in P_k$ or, equivalently, $x \in T^{-n}P_k$. Thus $T^{-n}P =_\Delta \{T^{-n}P_1, \cdots, T^{-n}P_K\}$ partitions Σ according to the output at time n. Given two partitions $P = \{P_1, \cdots, P_K\}$ and $Q = \{Q_1, \cdots, Q_J\}$ of the same space, the *join* $P \vee Q$ is the partition of Σ with atoms $P_k \cap Q_j$; $k = 1, \cdots, K, j = 1, \cdots, J$. Thus $\bigvee_{i=0}^{n-1} T^{-i}P$ partitions Σ according to the outputs at times zero through $n-1$, i.e., the atoms of $\bigvee_{i=0}^{n-1} T^{-i}P$ are simply all the disjoint n-dimensional thin cylinders of the form $\{x: x_i = a_{k_i}, 0 \leqq i \leqq n-1\} = \bigcap_{i=0}^{n-1} T^{-i}P_{k_i}$. The notation $\bigvee_{i=0}^{n-1} T^{-i}P$ also denotes the σ-field generated by the atoms; i.e., the class of all unions of atoms of the partition $\bigvee_{i=0}^{n-1} T^{-i}P$.

The *distribution* $d(\bigvee_{i=0}^{n-1} T^{-i}P)$ is defined as the vector having as entries the measure of the atoms of $\bigvee_0^{n-1} T^{-i}P$ in lexicographical order, that is,

$$d(\bigvee_{i=0}^{n-1} T^{-1}P) =_\Delta \{\mu(\bigcap_{i=0}^{n-1} T^{-i}P_{k_i}); \text{ all } k^n = (k_0, \cdots, k_{n-1}) \in \{1, \cdots, K\}^n\} .$$

Given any event F, the partition $\bigvee_0^{n-1} T^{-1}P/F$ is the partition on F induced by $\bigvee_0^{n-1} T^{-i}P$, i.e., it has as atoms all sets of the form $F \cap (\bigcap_{i=0}^{n-1} T^{-i}P_{k_i})$. If $\mu(F) > 0$, the conditional distribution $d(\bigvee_0^{n-1} T^{-i}P/F)$ is the vector of conditional probabilities $\mu(\bigcap_{i=0}^{n-1} T^{-i}P_{k_i} \mid F) = \mu(F \cap (\bigcap_{i=0}^{n-1} T^{-i}P_{k_i}))/\mu(F)$.

The key to the proof of Lemma 1 is the following strong form of Rohlin's Theorem that gives a simple geometric picture of the behavior of stationary aperiodic processes over finite time (Shields (1973) pages 16–17, 22–24, 64):

ROHLIN'S THEOREM. *Given a stationary aperiodic process* (Σ, S, μ), *a positive integer* n, *and any* $\varepsilon > 0$, *then there exists an event* F *such that* $F, TF, \cdots, T^{n-1}F$ *are disjoint sets,* $\mu(\bigcup_{i=0}^{n-1} T^i F) \geqq 1 - \varepsilon$, *and*

$$d(\bigvee_{i=0}^{n-1} T^{-i}P) = d(\bigvee_{i=0}^{n-1} T^{-i}P/F) .$$

Intuitively, the theorem states that any stationary aperiodic random process has a similar structure over any finite time in the sense that the sequence space

Σ can be "carved up" into n disjoint sets that are all shifts of a *base* F, the union of these sets has almost all of the probability measure, and the base reflects the n-tuple distribution of the entire space, i.e., given any n-tuple $(a_{k_0}, \cdots, a_{k_1})$, $\mu(\{x: x_i = a_{k_i}, 0 \leqq i \leqq n-1\}) = \mu(\{x: x_i = a_{k_i}, 0 \leqq i \leqq n-1\} \mid F)$. In other words, the base F is independent of all n-dimensional cylinders $\{x: x_i = a_{k_i}, 0 \leqq i \leqq n-1\}$. The collection $\{T^i F; i = 0, \cdots, n-1\}$ together with the partition P is called a *gadget* or *ε-gadget* (T, n, F, P).

PROOF OF LEMMA 1. Given the two stationary aperiodic processes $[A, \mu]$ and $[\hat{A}, \nu]$, and $\delta > 0$, let p be a stationary measure on $(\Sigma \times \hat{\Sigma}, S \times \hat{S})$ approximately yielding the infimum defining $\bar{\rho}([A, \mu], [\hat{A}, \nu])$, i.e.,

$$p(\Sigma \times B) = \nu(B), \qquad \forall B \in \hat{S}, \tag{4.1a}$$

$$p(B \times \Sigma) = \mu(B), \qquad \forall B \in S, \tag{4.1b}$$

$$E_p[\rho(X_0, Y_0)] \leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + \delta/3. \tag{4.1c}$$

This is possible by the definition of $\bar{\rho}$. The $\delta/3$ is required since the infimum might not be a minimum. Since $[A, \mu]$ and $[\hat{A}, \nu]$ are aperiodic, the joint process $[A \times \hat{A}, p]$ must also be aperiodic.

Given the probability space $(\Sigma \times \hat{\Sigma}, S \times \hat{S}, p)$, let U denote the shift on $\Sigma \times \hat{\Sigma}$, i.e., if $z = (x, y) \in \Sigma \times \hat{\Sigma}$, then U shifts the pair sequence z: $(U(x, y))_n = (x_{n+1}, y_{n+1})$. Let $W = \{W_k; k = 1, \cdots, K\}$ denote the partition of $\Sigma \times \hat{\Sigma}$ according to the X coordinate of the zeroth letter, i.e., $W_k = \{(x, y): x_0 = a_k\}$. Similarly, define the partition $V = \{V_j, j = 1, \cdots, J\}$ by $V_j = \{(x, y): y_0 = b_j\}$ where $b_j \in \hat{A} =_\Delta \{b_1, \cdots, b_J\}$. The partition $W \vee V$ therefore partitions the space $\Sigma \times \hat{\Sigma}$ according to the output at time zero of each coordinate. The average distortion between the coordinate processes is

$$\begin{aligned} \rho(W, V) &=_\Delta E_p[\rho(X_0, Y_0)] = \textstyle\sum_{k=1}^{K} \sum_{j=1}^{J} \rho(a_k, b_j) p(W_k \cap V_j) \\ &\leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + \delta/3, \end{aligned} \tag{4.2}$$

and the entropy rate of the $[\hat{A}, \nu]$ process can be written as

$$H([\hat{A}, \nu]) =_\Delta H(U, V) = \lim_{n \to \infty} n^{-1} H(\textstyle\bigvee_0^{n-1} U^i V),$$

where

$$H(\textstyle\bigvee_0^{n-1} U^i V) = -\sum_{\text{all atoms } G \text{ of } \bigvee_0^{n-1} U^i V} \mu(G) \log \mu(G).$$

Given $0 < \delta < e^{-1}$, choose an integer l large enough such that

$$|l^{-1} H(\textstyle\bigvee_0^{l-1} U^i V) - H(U, V)| \leqq \delta/3$$

and chose ε small enough and n large enough so that

$$KJ\rho_M \varepsilon \leqq \delta/3 \tag{4.3a}$$

$$J^l(\varepsilon + (l-1)n^{-1})^{\frac{1}{2}} \leqq \delta/3 \tag{4.3b}$$

where $\rho_M =_\Delta \max_{k,j} \rho(a_k, b_j)$ was assumed finite.

Step 1. Use the Rohlin Theorem to construct an ε-gadget $(U, n, \bar{F}, W \vee V)$ on the joint sequence probability space $(\Sigma \times \hat{\Sigma}, S \times \hat{S}, p)$ i.e., $U^i\bar{F}$, $i = 0, \cdots, n-1$, are disjoint sets, $p(\bigcup_0^{n-1} U^i\bar{F}) \geqq 1 - \varepsilon$, and

$$d(\bigvee_0^{n-1} U^{-i}(W \vee V)/\bar{F}) = d(\bigvee_0^{n-1} U^{-1}(W \vee V)) . \tag{4.4}$$

Note that this implies that

$$d(\bigvee_0^{n-1} U^{-i}V/\bar{F}) = d(\bigvee_0^{n-1} U^{-i}V) . \tag{4.5}$$

In a similar manner, construct an ε-gadget $(T, n, F. P)$ on the source space (Σ, S, μ) so that T^iF, $i = 0, \cdots, n-1$, are disjoint, $\mu(\bigcup_0^{n-1} T^iF) \geqq 1 - \varepsilon$, and

$$d(\bigvee_0^{n-1} T^{-i}P/F) = d(\bigvee_0^{n-1} T^{-i}P) . \tag{4.6}$$

By definition of the $\bar{\rho}$ metric, p on $(\Sigma \times \hat{\Sigma}, S \times \hat{S})$ induces the original source measure μ on (Σ, S), and therefore $d(\bigvee_0^{n-1} T^{-i}P) = d(\bigvee_0^{n-1} U^{-i}W)$ and

$$d(\bigvee_0^{n-1} U^{-i}W/\bar{F}) = d(\bigvee_0^{n-1} T^{-i}P/F) . \tag{4.7}$$

When two gadgets satisfy (4.7), i.e., when the distribution of all n-tuples on the bases are equal, the gadgets are said to be *isomorphic* and we write $(U, n, \bar{F}, W) \sim (T, n, F, P)$. From Lemma 4.4 of Shields (1973), since $(U, n, \bar{F}, W) \sim (T, n, F, P)$ and V is a partition of $\Sigma \times \hat{\Sigma}$, then there is a partition $Q = \{Q_1, \cdots, Q_J\}$ of $\bigcup_0^{n-1} T^iF$ such that $(T, n, F, P \vee Q) \sim (U, n, \bar{F}, W \vee V)$, i.e., such that

$$\begin{aligned} d(\bigvee_0^{n-1} T^{-i}(P \vee Q)/F) &= d(\bigvee_0^{n-1} U^{-i}(W \vee V)/\bar{F}) \\ &= d(\bigvee_0^{n-1} U^{-i}(W \vee V)) . \end{aligned}$$

Note that this implies that

$$d(\bigvee_0^{n-1} T^{-i}Q/F) = d(\bigvee_0^{n-1} U^{-i}V/\bar{F}) \tag{4.8}$$

so that $(T, n, F, Q) \sim (U, n, \bar{F}, V)$.

Step 2. Extend the partition Q of $\bigcup_0^{n-1} T^iF$ to Σ in any manner. From Lemma A of the Appendix with $l = 1$, (4.8) implies that

$$|\mu(P_k \cap Q_j) - p(W_k \cap V_j)| \leqq \varepsilon \qquad \text{all } k, j$$

and therefore

$$\begin{aligned} \rho(P, Q) &= \textstyle\sum_{k,j} \rho(a_k, b_j)\mu(P_k \cap Q_j) \\ &\leqq \textstyle\sum_{k,j} \rho(a_k, b_j)p(W_k \cap V_j) + KJ\varepsilon \rho_M \\ &\leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + 2\delta/3 . \end{aligned}$$

Application of Lemma A to (T, n, F, Q) and the isomorphic $(U, n, \bar{F}, V)$ yields

$$\begin{aligned} l^{-1}H(\bigvee_0^{l-1} T^{-i}Q) &\leqq l^{-1}H(\bigvee_0^{l-1} U^{-i}V) + J^l(\varepsilon + (l-1)n^{-1})^{\frac{1}{2}} \\ &\leqq H([\hat{A}, \nu]) + 2\delta/3 . \end{aligned}$$

Comment. This proves the lemma for $f^{(\infty)}$ since $f^{(\infty)}$ is equivalent to the partition Q by defining $f^{(\infty)}(x) = b_j$ iff $x \in Q_j$ and therefore

$$H(f^{(\infty)}) \leqq l^{-1}H(\bigvee_0^{l-1} T^{-i}Q) \leqq H([\hat{A}, \nu]) + 2\delta/3 \tag{4.9a}$$

$$\rho(f^{(\infty)}) = \textstyle\sum_{k,j} \rho(a_k, b_j)\mu(P_k \cap Q_j) \leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + 2\delta/3 . \tag{4.9b}$$

Step 3. Since the σ-field S is generated by the cylinders, given any $\delta' > 0$, there exists an N sufficiently large and a partition $\bar{Q} \in \bigvee_{-N}^{N} T^{-i}P$ such that

$$|\bar{Q} - Q| =_{\Delta} \sum_{j=1}^{J} \mu(\bar{Q}_j \triangle Q_j) \leqq \delta'$$

where $\triangle$ denotes symmetric distance. This follows since each atom in Q can be approximated arbitrarily closely by a cylinder set (generator of the σ-field) or from Shields' (1973) Lemma 10.1. From Shields (1973) Lemma 8.2, δ' can be chosen sufficiently small and thus N sufficiently large to ensure that

$$|\lim_{n\to\infty} n^{-1}H(\bigvee_{0}^{n-1} T^{-i}Q) - \lim_{n\to\infty} n^{-1}H(\bigvee_{0}^{n-1} T^{-i}\bar{Q})| \leqq \delta/3 .$$

In addition, since

$$\rho(P, \bar{Q}) = \sum_{k=1}^{K} \sum_{j=1}^{J} \rho(a_k, b_j)\mu(P_k \cap \bar{Q}_j)$$

and since $P_k \cap \bar{Q}_j \subseteq (P_k \cap Q_j) \cup (Q_j \triangle \bar{Q}_j)$, so that $\mu(P_k \cap \bar{Q}_j) \leqq \mu(P_k \cap Q_j) + \mu(Q_j \triangle \bar{Q}_j)$, we have that

$$\begin{aligned}\rho(P, \bar{Q}) &\leqq \sum_{k=1}^{K} \sum_{j=1}^{J} \rho(a_k, b_j)\mu(P_k \cap Q_j) + \rho_M|Q - \bar{Q}| \\ &\leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + 2\delta/3 + \rho_M\delta' .\end{aligned}$$

Choosing δ' so that in addition $\rho_M\delta' \leqq \delta/3$, we have that

$$\begin{aligned}\lim_{n\to\infty} n^{-1}H(\bigvee_{0}^{n-1} T^{-i}\bar{Q}) &\leqq \lim_{n\to\infty} n^{-1}H(\bigvee_{0}^{n-1} T^{-i}Q) + \delta/3 \\ &\leqq l^{-1}H(\bigvee_{0}^{l-1} T^{-i}Q) + \delta/3 \leqq H([\hat{A}, \nu]) + \delta ,\end{aligned}$$

and

$$\rho(P, \bar{Q}) \leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + \delta .$$

Since $\bar{Q} \in \bigvee_{-N}^{N} T^{-i}P$, defining the function $f^{(N)}$ by

$$f^{(N)}(a_{k_{-N}}, \ldots, a_{k_0}, \ldots, a_{k_N}) = b_j \qquad \text{iff} \quad \bigcap_{-N}^{N} T^{-i}P_{k_i} \subseteq \bar{Q}_j$$

will yield

$$\begin{aligned}\rho(f^{(N)}) &\leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + \delta \\ H(f^{(N)}) &\leqq H([\hat{A}, \nu]) + \delta\end{aligned}$$

completing the proof of the lemma.

Proof of Theorem 1. Let $[\hat{A}, \nu]$ approximately yield the infimum of (3.1) i.e., given $\delta > 0$

$$\begin{aligned}\bar{\rho}([A, \mu], [\hat{A}, \nu]) &\leqq \inf_{[\hat{A}, \nu']: H_{\nu'} \leqq R} \bar{\rho}([A, \mu], [\hat{A}, \nu']) + \delta \\ H([\hat{A}, \nu]) &\leqq R .\end{aligned}$$

Application of Lemma 1 to the above implies that there exists for sufficiently large N a function $f^{(N)}$ such that

$$\begin{aligned}\rho(f^{(N)}) &\leqq \inf_{[\hat{A}, \nu]: H_\nu \leqq R} \bar{\rho}([A, \mu], [\hat{A}, \nu]) + 3\delta/2 \\ H(f^{(N)}) &\leqq R + \delta .\end{aligned}$$

Since δ is arbitrary, the theorem is proved.

Using the above $[\hat{A}, \nu]$, (4.9) proves the corollary.

Acknowledgment. The authors gratefully acknowledge the many helpful comments and criticisms of Professor Lee D. Davisson of the University of Southern California. The specific simple example of a sliding-block source coder is due to James Dunham of Stanford University.

APPENDIX

LEMMA A. *Given two spaces* (Σ, S, μ) *and* $(\bar{\Sigma}, \bar{S}, \bar{\mu})$, *and two* ε*-gadgets* (T, F, n, P) *on* (Σ, S, μ) *and* $(\bar{T}, \bar{F}, n, \bar{P})$ *such that*

(a) $$\|P\| = \|\bar{P}\| = K$$

(b) $$d(\bigvee_0^{n-1} \bar{T}^{-i}\bar{P}/\bar{F}) = d(\bigvee_0^{n-1} \bar{T}^{-i}\bar{P})$$

(c) $$(T, F, n, P) \sim (\bar{T}, \bar{F}, n, \bar{P})$$

then

(i) *For any* $l \leqq n$ *and any* l*-tuple* $(k_0, \cdots, k_{l-1})$,

$$|\mu(\bigcup_0^{l-1} T^{-i}P_{k_i}) - \bar{\mu}(\bigcap_0^{l-1} \bar{T}^{-i}\bar{P}_{k_i})| \leqq \varepsilon + (l-1)n^{-1}$$

(ii) *If* $\varepsilon + (l-1)/n \leqq e^{-1}$, *then*

$$|H(\bigvee_0^{l-1} T^{-i}P) - H(\bigvee_0^{l-1} \bar{T}^{-i}\bar{P})| \leqq K^l(\varepsilon + (l-1)/n)^{\frac{1}{2}}.$$

PROOF. Equations (b) and (c) imply that

$$\mu(\bigcap_0^{n-1} T^{-i}P_{k_i}/F) = \bar{\mu}(\bigcap_0^{n-1} \bar{T}^{-i}\bar{P}_{k_i}) \qquad \text{any } k^n = (k_0, \cdots, k_{n-1}).$$

Given a fixed $\bar{k}^l = (\bar{k}_0, \cdots, \bar{k}_{l-1})$,

$$\mu(\bigcap_0^{l-1} T^{-i}P_{\bar{k}_i}) = \sum_{j=0}^{n-1} \mu(\bigcap_0^{l-1} T^{-i}P_{\bar{k}_i} \mid T^jF)\mu(T^jF) + \mu(\bigcap_0^{l-1} T^{-i}P_{\bar{k}_i} \mid \Sigma - \bigcap_0^{n-1} T^jF)\mu(\Sigma - \bigcap_0^{n-1} T^jF).$$

Since μ is stationary,

$$\mu(\bigcap_{i=0}^{l-1} T^{-i}P_{\bar{k}_i} \mid T^jF) = \mu(\bigcap_{i=0}^{l-1} T^{-(i+j)}P_{\bar{k}_i} \mid F) = \mu(\bigcap_{i=j}^{j+l-1} T^{-i}P_{\bar{k}_{i-j}} \mid F).$$

If $j \geqq 0$ and $j + l - 1 \leqq n - 1$, then

$$\begin{aligned}\mu(\bigcap_{i=j}^{j+l-1} T^{-i}P_{\bar{k}_{i-j}} \mid F) &= \sum_{k^n : k_i = \bar{k}_{i-j};\, j \leqq i \leqq j+l-1} \mu(\bigcap_{i=0}^{n-1} T^{-i}P_{k_i} \mid F) \\ &= \sum_{k^n : k_i = \bar{k}_{i-j};\, j \leqq i \leqq j+l-1} \bar{\mu}(\bigcap_{i=0}^{n-1} \bar{T}^{-i}\bar{P}_{k_i}) \\ &= \bar{\mu}(\bigcap_{i=j}^{j+l-1} \bar{T}^{-i}\bar{P}_{\bar{k}_{i-j}}) \\ &= \bar{\mu}(\bigcap_{i=0}^{l-1} \bar{T}^{-i}\bar{P}_{k_i}); \qquad 0 \leqq j \leqq n-l\end{aligned}$$

independent of j. Thus, we have

$$\begin{aligned}\mu(\bigcap_0^{l-1} T^{-i}P_{\bar{k}_i}) &\leqq \sum_{j=0}^{n-l} \mu(\bigcap_0^{l-1} T^{-i}P_{\bar{k}} \mid T^jF)\mu(T^jF) \\ &\quad + \sum_{j=n-l+1}^{n-1} \mu(\bigcap_0^{l-1} T^{-i}P_{\bar{k}_i} \mid T^jF)\mu(T^jF) + \varepsilon \\ &\leqq (n-l+1)n^{-1}\bar{\mu}(\bigcap_0^{l-1} \bar{T}^{-i}\bar{P}_{\bar{k}_i}) + (l-1)n^{-1} + \varepsilon \\ &\leqq \bar{\mu}(\bigcap_0^{l-1} \bar{T}^{-i}\bar{P}_{\bar{k}_i}) + (l-1)n^{-1} + \varepsilon\end{aligned}$$

and

$$\mu(\bigcap_0^{l-1} T^{-i}P_{\bar{k}_i}) \geqq \bar{\mu}(\bigcap_0^{l-1} \bar{T}^{-i}\bar{P}_{\bar{k}_i})(n - l + 1)n^{-1}(1 - \varepsilon)$$
$$\geqq \bar{\mu}(\bigcap_0^{l-1} \bar{T}^{-1}\bar{P}_{\bar{k}_i}) - (l - 1)n^{-1} - \varepsilon,$$

proving (i).

From Blackwell, Brieman, and Thomasian (1959), if $|p - p'| \leqq \varepsilon < e^{-1}$, then $|p \log p - p' \log p'| \leqq \varepsilon^{\frac{1}{2}}$ and hence

$$|H(\bigvee_0^{l-1} T^{-i}P) - H(\bigvee_0^{l-1} \bar{T}^{-i}\bar{P})| \leqq K^l(\varepsilon + (l - 1)n^{-1})^{\frac{1}{2}}.$$

PROOF OF THEOREM 2. Parallel to the proof of Theorem 1, we first prove Lemma 2.

PROOF OF LEMMA 2. As in the finite case, let p be a stationary ergodic measure on $(\Sigma \times \hat{\Sigma}, S \times \hat{S})$ approximately yielding the infimum defining $\bar{\rho}([A, \mu], [\hat{A}, \nu])$, i.e., (4.1) is satisfied. Since A is separable under ρ, we can construct a countable partition $\{G_i\}_{i=1}^{\infty}$ of A with maximal diameter $\delta/6$. Let a_i be any element of G_i and relabel, if necessary, so that the reference letter a_0 is in G_0. Since $E_\mu\{\rho(X_0, a_0)\} \leqq \rho^* < \infty$, then as in Gray and Davisson (1974) the alphabet A can be quantized as follows: Choose $K = K(\delta)$ such that

$$\sum_{i=K}^{\infty} E_\mu\{\rho(X_0, a_0)I_{G_i}\} \leqq \delta/6$$

where I_{G_i} is the indicator function of G_i. Define the quantized alphabet $\tilde{A} = \{a_0, \cdots, a_{K-1}\} \subseteq A$ and the quantizer function $q: A \to \tilde{A}$ by

$$q(x_0) = a_k \quad \text{if} \quad x_0 \in G_k, \qquad k \leqq K - 1$$
$$= a_0 \quad \text{otherwise}$$

and note

$$\begin{aligned} E_\mu\{\rho(X_0), q(X_0)\} &= \int_A d\mu^1(x)\rho(x, q(x)) \\ &= \int_{K-1, x \in \bigcup_0 G_k} d\mu^1(x)\rho(x, q(x)) \\ &\quad + \int_{K-1, x \notin \bigcup_0 G_k} d\mu^1(x)\rho(x, a_0) \\ &\leqq \delta/3. \end{aligned} \tag{A.1}$$

We next similarly quantize the alphabet $\hat{A}$.

$$\begin{aligned} E_\nu\{\rho(Y_0, b_0)\} &= E_p\{\rho(Y_0, b_0)\} \\ &\leqq E_p\{\rho(Y_0, X_0) + \rho(X_0, a_0) + \rho(a_0, b_0)\} \\ &\leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + \delta/3 + \rho^* + \rho(a_0, b_0) \end{aligned}$$

so that if $\bar{\rho}([A, \mu], [\hat{A}, \nu]) < \infty$ (otherwise the lemma is trivial) we can construct a quantized alphabet $\tilde{B} = \{b_0, \cdots, b_{J-1}\}$, where $J = J(\delta)$, and a quantizer $\hat{q}: \hat{A} \to \tilde{B}$ such that

$$E_\nu\{\rho(Y_0, \hat{q}(Y_0))\} \leqq \delta/3. \tag{A.2}$$

Let $[\tilde{A}, \tilde{\mu}]$ be the resulting quantized source $\{q(X_n)\}$ and let $[\tilde{B}, \tilde{\nu}]$ be the resulting reproduction process $\{\hat{q}(Y_n)\}$ induced by $[A, \mu]$ and $[\hat{A}, \nu]$, respectively. Note that (A.1) and (A.2) imply that

$$\bar{\rho}([A, \mu], [\tilde{A}, \tilde{\mu}]) \leqq \delta/3, \qquad \bar{\rho}([\hat{A}, \nu], [\tilde{B}, \tilde{\nu}]) \leqq \delta/3,$$

and therefore since $\bar{\rho}$ is a metric

$$\bar{\rho}([\tilde{A}, \tilde{\mu}], [\tilde{B}, \tilde{\nu}]) \leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + 2\delta/3 .$$

Note that since $[\tilde{B}, \tilde{\nu}]$ is a quantized version of $[\hat{A}, \nu]$,

$$H([\tilde{B}, \tilde{\nu}]) \leqq H([\hat{A}, \nu]) \leqq R .$$

Quantizing an ergodic source yields an ergodic process. Aperiodicity, however may not be inherited. Application of Lemma 1 with $\delta/3$ to $[\tilde{A}, \tilde{\mu}]$ and $[\tilde{B}, \tilde{\nu}]$ yields a code $\tilde{f}^{(N)} : \tilde{A}^{2N+1} \to \tilde{B}$ such that $\rho_{\tilde{\mu}}(\tilde{f}^{(N)}) \leqq \bar{\rho}([\tilde{A}, \tilde{\mu}], [\tilde{B}, \tilde{\nu}]) + \delta/3$ and $H_{\tilde{\mu}}(\tilde{f}^{(N)}) \leqq R + \delta$. This in turn implies a code $f^{(N)} : A^{2N+1} \to \hat{A}$ given by $f^{(N)}(x_{-N}, \cdots, x_N) =_{\Delta} \tilde{f}^{(N)}(q(x_{-N}), \cdots, q(x_N))$ such that

$$\rho_{\mu}(f^{(N)}) = \rho_{\tilde{\mu}}(\tilde{f}^{(N)}) \leqq \bar{\rho}([A, \mu], [\hat{A}, \nu]) + \delta .$$
$$H_{\mu}(f^{(N)}) = H_{\tilde{\mu}}(\tilde{f}^{(N)}) \leqq R + \delta$$

completing the proof of the lemma.

Theorem 2 follows from Lemma 2 exactly as Theorem 1 follows from Lemma 1.

REFERENCES

[1] BERGER, T. (1971). *Rate Distortion Theory*. Prentice-Hall, Englewood Cliffs, N.J.

[2] BLACKWELL, D., BREIMAN, L. and THOMASIAN, A. (1959). The capacity of a class of channels. *Ann. Math. Statist.* **30** 1229–1241.

[3] BLAHUT, R. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Trans. Information Theory* **18** 460–473.

[4] DAVISSON, L. D. (1968). The theoretical analysis of data compression systems. *Proc. IEEE* **56** 176–186.

[5] DICK, R. (1973). Ph. D. research, Cornell Univ.

[6] GALLAGER, R. G. (1968). *Information Theory and Reliable Communication*. Wiley, New York, Chapter 9.

[7] GRAY, R. M. and DAVISSON, L. D. (1974). Source coding theorems without the ergodic assumption. *IEEE Trans. Information Theory* **20** 502–516.

[8] GRAY, R. M., NEUHOFF, D. L. and OMURA, J. K. (1975). Process definitions of distortion-rate functions and source coding theorems. To appear in *IEEE Trans. Information Theory* **21**.

[9] GRAY, R. M., NEUHOFF, D. L. and SHIELDS, P. C. (1975). A generalization of Ornstein's $\bar{d}$ distance with applications to information theory. *Ann. Probability* **3** 315–328.

[10] HELLMAN, M. (1975). Convolutional source coding. To appear in *IEEE Trans. Information Theory* **21**.

[11] KOSHELEV, V. (1973). Direct sequential encoding and decoding for discrete sources. *IEEE Trans. Information Theory* **19** 340–343.

[12] KRENGEL, U. (1973). Discussion on Professor Ornstein's paper. *Ann. Probability* **1** 61–62.

[13] NEUHOFF, D. L., GRAY, R. M. and DAVISSON, L. D. (1975). Fixed rate universal block source coding with a fidelity criterion. To appear in *IEEE Trans. Information Theory* **21**.

[14] ORNSTEIN, D. S. (1973). An application of ergodic theory to probability theory. *Ann. Probability* **1** 43–58.

[15] SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27** 379–423.

[16] SHANNON, C. E. (1959). Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record*, Part 4, 142–163.

[17] SHIELDS, P. C. (1973). *The Theory of Bernouilli Shifts.* Univ. of Chicago Press.

ROBERT GRAY
STANFORD ELECTRONICS LABORATORY
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DONALD S. ORNSTEIN
DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DAVID L. NEUHOFF
DEPARTMENT OF ELECTRICAL AND COMPUTER SCIENCE
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN

Note added in proof. J. Feldman of the University of California, Berkeley has pointed out that the proof of Theorem 1 is not quite complete since the rate is $R + \delta$ rather than R. The δ can be removed as in the block coding case by using the continuity of the DRF.