The Annals of Probability
1980, Vol. 8, No. 6, 1085–1092

ABSTRACT ALPHABET SLIDING-BLOCK ENTROPY COMPRESSION CODING WITH A FIDELITY CRITERION[1]

BY JAMES G. DUNHAM

Washington University

The existing sliding-block entropy compression coding with a fidelity criterion theorem is generalized to abstract source and source reproduction alphabets, stationary and ergodic sources and nonmetric distortion measures. This establishes the sliding-block entropy compression coding theorem in the same generality as the corresponding block entropy compression coding theorem.

1. Introduction. The goal of the theory of entropy compression coding with a fidelity criterion is to code a source into a minimally distorted reproduction of the source whose mean entropy rate is constrained. Since Shannon (1959) the theory has focused almost exclusively on block codes—coding structures which process consecutive nonoverlapping blocks of source symbols into source reproduction symbols. The fundamental theorems show that the optimum performance theoretically attainable (OPTA) for block codes is given by the distortion-rate function. Thus performance arbitrarily close to the distortion-rate function can be achieved by sufficiently long and complex block codes.

Gray, Neuhoff and Ornstein (1975) introduced a new class of nonblock coding structures called sliding-block codes—coding structures which process consecutive overlapping blocks of source symbols into source reproduction symbols. A theorem on sliding-block entropy compression coding with a fidelity criterion was established which equated the OPTA function for sliding-block codes with the distortion-rate function for finite-alphabet, stationary, ergodic and aperiodic sources with bounded single-letter fidelity criteria. Using quantization techniques, the theorem was extended to situations where the source and source reproduction alphabet were the same complete separable metric space and the distortion measure was the possibly unbounded metric of the underlying metric space. Successively simpler proof have been given by Gray (1975), Gray and Ornstein (1976), Shields and Neuhoff (1977) and Davisson and Gray (1978), but they did not establish the theorem in any greater generality.

In this paper the existing sliding-block entropy compression coding with a fidelity criterion theorem is generalized to abstract source and source reproduction alphabets, stationary and ergodic sources and nonmetric distortion measures. Thus the sliding-block compression coding theorem is established in the same generality as the corresponding block entropy compression coding theorem (Berger (1971)).

Furthermore, it is shown that nearly optimal performance can be achieved by sufficiently long constraint length sliding-block codes which yield a finite alphabet source reproduction process. By combining a sliding-block entropy compression code with a joint noiseless-source/noisy-channel code of Gray and Ornstein (1976), the source can be transmitted over a noisy channel to a user. With appropriate restrictions on the distortion measure, a joint source-channel coding theorm (information transmission theorem) can be established which equates the OPTA for sliding-block communication systems with the distortion-rate function evaluated at the channel capacity. Thus under appropriate restrictions on the distortion measure, the ultimate performance which can be achieved by blocking and sliding-block communication systems is the same.

Received June 9, 1978; revised July 10, 1979.

[1] This research was supported in part by NSF Grants ENG 76-02276 and ENG 7910001. The author was formerly with Stanford University.

AMS 1970 *subject classification.* Primary 94A15; secondary 94A05.

Key words and phrases. Entropy compression coding with a fidelity criterion, source coding, information theory.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Probability.* www.jstor.org

2. Notation and definitions. Let $(A, \mathscr{A})$ be an abstract measurable space where the abstract space A is called the source alphabet. Define the product measurable space $(A^\infty, \mathscr{A}^\infty) = \times_{k=-\infty}^{\infty} (A_k, \mathscr{A}_k)$ where $(A_k, \mathscr{A}_k) = (A, \mathscr{A})$, all k; A^∞ is the space of all doubly infinite sequences $x = (\cdots, x_{-1}, x_0, x_1, \cdots)$ from A; and $\mathscr{A}^\infty$ is the usual product σ-algebra. For each integer n, let $(A^n, \mathscr{A}^n) = \times_{k=1}^{n} (A_k, \mathscr{A}_k)$ and a typical element of A^n is denoted by $x^n = (x_1, \cdots, x_n)$. Let $X_n^l: A^\infty \to A^l$ be a vector coordinate function defined by $X_n^l(x) = x_n^l = (x_n, \cdots, x_{n+l-1})$, all $x \in A$. When $l = 1$, this is simply the nth scalar coordinate function X_n. Let $T: A^\infty \to A^\infty$ be the usual shift transformation defined by $X_k(Tx) = x_{k+1}$. Finally for integers $m \geq n$ let $X_{(m,n)}: A^\infty \to A^{(n-m+1)}$ be another vector coordinate function defined by $X_{(m,n)}(x) = x_{(m,n)} = (x_m, \cdots, x_n)$, in which the actual coordinates of the vector are emphasized.

We denote by $[A, \mu, X]$ the discrete-time *source* with underlying alphabet A, probability measure μ on $(A^\infty, \mathscr{A}^\infty)$ and name X. If A is a finite alphabet, then $[A, \mu, X]$ is called a finite alphabet source. A source is stationary (ergodic) if μ is stationary (ergodic) with respect to T. A source is aperiodic if $\mu(\{x \in A^\infty \mid T^n x = x;\ n = 1, 2, \cdots\}) = 0$ and periodic if $\mu(\{x \in A^\infty \mid T^n x = x;\ n = 1, 2, \cdots\}) = 1$. If a source is ergodic, then it is either periodic or aperiodic.

Denote the cardinality of a set F by $\| F \|$. A measurable set F is called a *finite dimensional* set if, for some integer N, $F = \cdots \times A_{-N-1} \times F' \times A_{N+1} \times \cdots$ where $F' \in \times_{k=-N}^{N} \mathscr{A}_k$. Let $R = \{R_k\}_{k=1}^{M}$ be a *partition* of A^∞, that is, the atoms $R_k \in \mathscr{A}^\infty$ are disjoint and $\bigcup_{k=1}^{M} R_k = A^\infty$. If $\| R \| = M < \infty$, then R is called a finite partition. If each atom of R is a finite dimensional set, then R is called a finite dimensional partition. If R and S are two partitions, their *join* is

$$R \vee S = \{R_i \cap S_j \mid R_i \in R, S_j \in S\}.$$

Define the entropy of a partition R by $H(R) = -\sum_{k=1}^{\|R\|} \mu(R_k) \log \mu(R_k)$ where all logarithms are to the base 2. For a source $[A, \mu, X]$, define the *mean entropy rate* by

$$\bar{H}(X) = \limsup_{n\to\infty} n^{-1} H(X^n)$$

where

$$H(X^n) = \sup_{R \in \mathscr{R}(n)} H(R)$$

and $\mathscr{R}(n)$ denotes the set of all finite partitions of $(A^n, \mathscr{A}^n)$. Define the *Kolmogorov-Sinai invariant* of ergodic theory by

$$\bar{\mathscr{H}}(X) = \sup_{R \in \mathscr{R}} H^\infty(R)$$

where

$$H^\infty(R) = \limsup_{n\to\infty} n^{-1} H\left(\bigvee_{k=0}^{n-1} T^k R\right)$$

and $\mathscr{R}$ denotes the set of all finite partitions of $(A^\infty, \mathscr{A}^\infty)$. It is clear for a finite alphabet source $[A, \mu, X]$ that $\bar{H}(X) = \bar{\mathscr{H}}(X)$. In general $\bar{\mathscr{H}}(X) \leq \bar{H}(X)$ and the inequality may be strict (Pinsker (1964)).

Let $\hat{A}, \hat{\mathscr{A}})$, be an abstract measurable space where the abstract space $\hat{A}$ is called the source reproduction alphabet. For any integer L, a *sliding-block code* of constraint length $2L$ is any measurable function $f^{(L)}: A^{2L+1} \to \hat{A}$. Let $\mathscr{C}_S$ denote the class of all sliding-block codes with a finite or infinite constraint length and let $\mathscr{C}'_S$ denote the set of all sliding-block codes with a finite constraint length and $\| f^{(L)}(A^{2L+1}) \| < \infty$. The source reproduction process $\{\hat{X}_k\}_{k=-\infty}^{\infty}$ defined by $\hat{X}_k = f^{(L)}(X_{(k-L,k+L)}) = f^{(L)}(X_{k-L}, \cdots, X_k, \cdots, X_{k+L})$. Connecting a stationary (ergodic) source $[A, \mu, X]$ to a sliding-block code $f^{(L)}$ induces a stationary (ergodic) pair process $[A \times \hat{A}, p, (X, \hat{X})]$ where the probability measure p is defined for $E \in \mathscr{A}^\infty$, $F \in \hat{\mathscr{A}}^\infty$ *by* $p(E \times F) = \mu(E \cap f^{(L)-1}(F))$ where $f^{(L)-1}(F)$ denotes the inverse image of the set F.

We note that a sliding-block code $f^{(L)}$ is equivalent to the partition $R_f(L) = \{R_{\hat{a}}\}_{\hat{a}\in\hat{A}}$ *of* A^∞ where $R_{\hat{a}}$ $\{x \in A^\infty \mid f^{(L)}(x) = \hat{a}\}$ and that a partition $R = \{R_k\}_{k=1}^{M}$ is equivalent to the sliding-block code defined by $f_R^{(\infty)}(x) = k$ if $x \in R_k$ where $\hat{A} = \{1, 2, \cdots, M\}$. In particular, note that $\mathscr{C}'_S$ is equivalent to the set of all finite, finite dimensional partitions of A^∞. For a partition R, if the source reproduction process defined by $f_R^{(\infty)}$ is aperiodic (periodic), then R is called an aperiodic (periodic) partition.

Let $\rho: A \times \hat{A} \to [0, \infty]$ be a measurable function, called a *per-letter distortion measure*, which specifies the cost in reproducing the source letter x by the source reproduction letter $\hat{x}$. A sliding-block code has *expected distortion*

$$\rho(f^{(L)}) = E_\mu[\rho(X_0, f^{(L)}(X_{(-L,L)}))]$$

where E_μ denotes the expectation over μ and mean entropy rate $\bar{H}(f^{(L)}) = \bar{H}(\hat{X})$. If the source is ergodic and $\rho(f^{(L)})$ is finite, then it is easily shown that $\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \rho(X_k, \hat{X}_k) = \rho(f^{(L)})$ almost everywhere with respect to μ.

The goal of entropy compression coding is to produce a source reproduction process whose mean entropy rate is less than some fixed number and which well approximates the original source in the ρ sense. Therefore, we define the OPTA function using sliding-block codes with a mean entropy rate constraint by

$$\delta_S(R) = \inf_{f^{(L)} \in \mathscr{C}_S : \bar{H}(f^{(L)}) \le R} \rho(f^{(L)})$$

$$\delta'_S(R) = \inf_{f^{(L)} \in \mathscr{C}'_S : \bar{H}(f^{(L)}) \le R} \rho(f^{(L)}).$$

Since $\mathscr{C}'_S \subset \mathscr{C}_S$, $\delta'_S(R) \ge \delta_S(R)$.

The following theorem shows that the Shannon (1959) distortion rate function is a general lower bound to the OPTA functions $\delta_S(R)$ and $\delta'_S(R)$. It was first established by Gray, Neuhoff and Omura (1975) for finite alphabet source and source reproduction spaces and generalized to abstract source and source reproduction spaces by Dunham (1979).

CONVERSE THEOREM. *Let $[A, \mu, X]$ be a stationary source having a distortion-rate function $D(R)$ with respect to a per-letter distortion measure ρ. Then*

$$\delta'_S(R) \ge \delta_S(R) \ge D(R)$$

for all rates R, except possibly at the rate $R_0 = \inf\{R \ge 0 \mid D(R) < \infty\}$.

3. The sliding-block entropy compression coding theorem. The basic result of this paper is the following theorem.

THEOREM 1. *Let $[A, \mu, X]$ be an abstract alphabet, stationary and ergodic source having a distortion-rate function $D(R)$ with respect to a per-letter distortion measure ρ. If there exists a reference letter $a^* \in \hat{A}$ such that*

$$\int \rho(X_0(x), a^*)\, d\mu(x) < \infty,$$

then

$$\delta'_S(R) = \delta_S(R) = D(R)$$

for all rates $R \in (0, \infty)$.

The basic idea for proving the positive half of the theorem is to construct a good sliding-block code from a good block code. The fundamental tool required for this construction is a variant of the strong Rohlin theorem (Shields (1973), Halmos (1956)) given by the following theorem.

VARIANT OF THE STRONG ROHLIN THEOREM. *Let $[A, \mu, X]$ be a stationary, ergodic and aperiodic process with a finite, finite dimensional and aperiodic partition P. Given $\gamma, \eta \ge 0$ and an integer L, there exists a set $F \in \mathscr{A}^\infty$ called the base where the following holds:*

(1) *$F, TF, \cdots, T^{L-1}F$ are disjoint;*

(2) $\mu(\cup_{k=0}^{L-1} T^k F) \ge 1 - \gamma$;

(3) *$\sum_{i=1}^{\|P\|} |\mu(P_i) - \mu_F(P_i)| \le \eta$, where $\mu_F(G) \equiv \mu(G \cap F)/\mu(F)$; and*

(4) *F is a finite dimensional set.*

PROOF. The theorem follows easily from Theorem 3 of Shields and Neuhoff (1977) by letting $F = \{x \in A^\infty \mid X_0^{-L}(x)$ is an N-cell$\}$ and choosing δ small enough for the partition $\vee_{k=0}^{L-1} T^k P$.

COMMENT. The collection of sets $\{T^k F \mid k = 0, \cdots, L-1\}$ together with the partition P will be called a (γ, L, P, η)—gadget or simply a gadget. The set G which is the complement of the set $\cup_{k=0}^{L-1} T^k F$ will be called the garbage set. Property (2) states that almost the entire measure of the space is contained in the gadget. Property (3) states that the distribution of the partition P on the base will be almost the same as the distribution of the partition P through the entire space. Property (4) states that the base set F is determined by the finite number of coordinates $X_{(-N,N)}$ for a sufficiently large integer N. This version differs from the strong form of the Rohlin theorem in that the strong Rohlin theorem requires $\mu(P_i) = \mu_F(P_i)$ for all $i = 1, \cdots, \|P\|$; but the base set F is not required to be finite dimensional.

4. Proof of main result. Let $[A, \mu, X]$ be a stationary and ergodic source having a distortion-rate function $D(R)$ with respect to a per-letter distortion measure ρ. Let $a^* \in \hat{A}$ be a reference letter where

$$\int \rho(X_0(x), a^*)\, d\mu(x) < \infty.$$

The converse theorem shows that $\delta_S'(R) \geq \delta_S(R) \geq D(R)$ for all $R \in (0, \infty)$ and it remains to show that $\delta_S'(R) \leq D(R)$.

First, consider the case where $\mathscr{H}(X) = 0$. Fix a rate $R \in (0, \infty)$ and an $\epsilon > 0$. The conditions of the theorem imply that the theorem on block source coding with a fidelity criterion holds (Berger (1971) Theorem 7.2.4). Therefore, there exists for sufficiently large block length L a block code $f_B^{(L)}: A^L \to \hat{A}^L$ where $L^{-1}\log\|f_B^{(L)}(A^L)\| \leq R$ and

$$E_\mu[\rho_L(X_0^L, f_B^{(L)}(X_0^L))] \leq D(R) + \epsilon$$

where $\rho_L(x^L, \hat{x}^L) = L^{-1} \sum_{k=1}^{L-1} \rho(x_k, \hat{x}_k)$. Let B denote the set consisting of all distinct source reproduction symbols which can be produced by the block source coder $f_B^{(L)}$ and note that $\|B\| < \infty$. Let f_S be a sliding-block code defined by mapping the source symbol X_0 into a source reproduction symbol $\hat{X}_0$ in B of minimal distortion, that is, f_S is a block code of block length 1 formed by the set of source reproduction symbols B, and it is clear that $f_S \in \mathscr{C}_S'$. Computing the mean entropy rate, it follows from Pinsker (1964) Theorem 7.2.1(7) and page 76 that

$$\bar{H}(f_S) = \bar{H}(\hat{X}) = \mathscr{H}(\hat{X}) = \mathscr{H}(X) = 0 \leq R.$$

Computing the expected distortion, the stationarity of the source and the fact that $f_B^{(L)}(A^L) \subseteq B^L$ yields

$$\rho(f_S) = E_\mu[\rho(X_0, f_S(X_0))] = E_\mu[L^{-1} \textstyle\sum_{k=1}^{L} \rho(X_k, f_S(X_k))]$$

$$\leq E_\mu[\rho_L(X_0^L, f_B^{(L)}(X_0^L))] \leq D(R) + \epsilon.$$

Since ϵ is arbitrary, this then implies that $\delta_S'(R) \leq D(R)$ for all $R \in [0, \infty]$.

Next consider the case where $\mathscr{H}(X) > 0$. To simplify the proof, we shall break it up into a number of steps.

STEP 1. Fix an $\epsilon > 0$ and chose a rate $R \in (0, \infty)$. By the continuity of the distortion-rate function for $R > 0$, there exists a sufficiently small $\delta > 0$ such that

$$D(R - \delta) \leq D(R) + \epsilon/4.$$

Fix a δ to satisfy the above and we may also assume that

$$\delta < \epsilon/4.$$

The conditions of the theorem imply that the theorem on block source coding with a fidelity criterion holds (Berger (1971) Theorem 7.2.4). Therefore, given $\delta > 0$ and rate $R - \delta$, there exists for sufficiently large L a block code $f_B^{(L)}: A^L \to \hat{A}^L$ where $L^{-1}\log \| f_B^{(L)}(A^L) \| \le R - \delta$ and

(1) $$E_{\mu^L}[\rho_L(X_0^L, f_B^{(L)}(X_0^L))] \le D(R - \delta) + \delta \le D(R) + \epsilon/2.$$

Fix a length L to satisfy the above and we may also assume that L is large enough so that

$$h_b(L^{-1}) \le \delta$$

where $h_b(L^{-1}) \equiv -L^{-1}\log L^{-1} - (1 - L^{-1})\log(1 - L^{-1})$.

STEP 2. Consider the composition of the mappings $X_0^L: A^\infty \to A^L$ and $f_B^{(L)}: A^L \to \hat{A}^L$, $f_B^{(L)}(X_0^L)$, that is, we apply the block code to the source symbols X_0^L. Since the rate R is finite, $M \equiv \| f_B^{(L)}(A^L) \| \le 2^{LR} < \infty$ and let $\hat{\mathbf{x}}_1, \cdots, \hat{\mathbf{x}}_M$ represent these M block code words, each consisting of L source reproduction letters. For $k = 1, \cdots, M$, let $E_k \equiv [f_B^{(L)}(X_0^L)]^{-1}(\hat{\mathbf{x}}_k) = \{x \in A^\infty \mid f_B^{(L)}(X_0^L(x)) = \hat{\mathbf{x}}_k\}$, the inverse image of the source reproduction word $\hat{\mathbf{x}}_k$, and clearly the partition $E \equiv \{E_k \mid k = 1, \cdots, M\}$ is a finite, finite dimensional partition of A^∞. In fact, the σ-algebra generated by the partition E is the smallest σ-algebra such that $f_B^{(L)}(X_0^L)$ is measurable.

Suppose for some $\eta > 0$ and $\gamma > 0$ we were to use the variant of the strong Rohlin to construct a (γ, L, E, η)-gadget and then imbed the block code $f_B^{(L)}$ on the gadget as described by Gray and Ornstein (1976). If we were to look at the set $E_k \cap F$ on the base of the gadget, then, for all $x \in E_k \cap F$, we have that $f_B^{(L)}(X_0^L(x)) = \hat{\mathbf{x}}_k$; that is, the sequence of names of points in the column located on the gadget directly above the base set $E_k \cap F$ is mapped by the block code $f_B^{(L)}$ into the reproduction block of symbols $\hat{\mathbf{x}}_k$. By property 3 of the variant of the strong Rohlin theorem $|\mu_F(E_k) - \mu(E_k)| \le \eta$, that is, each block reproduction word $\hat{\mathbf{x}}_k$ is produced with nearly the correct probability. But we note that there is no control over the distortion $\rho_L(X^L(x), \hat{\mathbf{x}}_k)$. Since a general distortion measure ρ may be unbounded, a poor approximation on one atom of the partition may cause an unbounded expected distortion for the sliding-block code.

We now further refine the atom E_k so that the distortion $\rho_L(X_0^L(x), \hat{\mathbf{x}}_k)$ can be precisely controlled for each point $x \in E_k \cap F$. Decompose $[0, \infty]$ into a countable number of consecutive nonoverlapping intervals $[(j-1)\delta, j\delta)$ of length δ and let the set E_k^j be the inverse image of the jth interval under $\rho_L(X_0^L(x), \hat{\mathbf{x}}_k)$. Then the atom E_k has been refined by a countable partition $E^k \equiv \{E_k^j\}_{j=1}^\infty$ where for any $x \in E_k^j \cap F$

(2) $$(j-1)\delta \le \rho_L(X_0^L(x), \hat{\mathbf{x}}_k) < j\delta.$$

In order to later use the variant of the strong Rohlin theorem, it will be necessary to obtain a finite partitioning of E_k.

Since a^* is a reference letter and since ρ is nonnegative, the function $\rho(X_0(x), a^*)$ is an absolutely integrable function. By the continuity theorem for Lebesgue integrals of absolutely integrable functions (Hewitt and Stromberg (1965), Theorem 13.34), there exists a $\gamma > 0$ such that if $E \in \mathscr{A}^\infty$ and $\mu(E) < \gamma$, then

(3) $$\int_E \rho(X_0(x), a^*)\, d\mu(x) < \epsilon/16L.$$

Fix a γ to satisfy the above and we may also assume that

$$\gamma \le \epsilon/2LR \log L.$$

Since $\sum_{j=1}^\infty \mu(E_k^j) = \mu(E_k) \le 1$, there is a finite number n_k such that

(4) $$\textstyle\sum_{j=n_k+1}^\infty \mu(E_k^j) < \gamma/M.$$

Let $G_k = \cup_{j=n_k+1}^\infty E_k^j$. Consider the finite partitioning of E_k consisting of $\{G_k, E_k^j \mid j = 1, \cdots, n_k\}$. From (2), we see that the distortion $\rho_L(X_0^L(x), \hat{\mathbf{x}}_k)$ can be controlled within an amount δ for $x \in E_k^j \cap F$ and $j = 1, \cdots, n_k$. For $x \in G_k \cap F$, if we were to now relabel this part of the gadget with the sequence of L reference symbols $\mathbf{a}^* \equiv (a^*, \cdots, a^*)$ instead of $\hat{\mathbf{x}}_k$, the expected distortion of $\rho_L(X_0^L, \mathbf{a}^*)$ would be small, independent of the actual value of $\rho(X_0^L(x), \hat{\mathbf{x}}_k)$ since $\mu(G_k \cap T^i F) < \gamma$ for $i = 0, \cdots, L-1$.

For each $k = 1, \cdots, M$, choose an n_k and G_k as described above. Letting $G_0 = \cup_{k=1}^M G_k$, it follows from (4) that

$$\mu(G_0) \le \gamma. \tag{5}$$

Finally define a finite, finite dimensional partition Q by

$$Q = \{G_0, E_k^j \mid k = 1, \cdots, M; j = 1, \cdots, n_k\}. \tag{6}$$

STEP 3. Since $\bar{\mathscr{H}}(x) > 0$ and μ is ergodic, the process $[A, \mu, X]$ is aperiodic. However, the partition Q may not be an aperiodic partition of $[A, \mu, X]$. We now further refine the partition Q to insure that it is aperiodic.

By the definition of the Kolomogorov-Sinai invariant $\bar{\mathscr{H}}$, there exists a finite partition R of A^∞ where $\mathscr{H}^\infty(R) > 0$. Then as in step 3 of the proof of Lemma 1 in Gray, Neuhoff and Ornstein (1975), R can be approximated by a finite, finite dimensional partition P where $\mathscr{H}^\infty(R)$ is arbitrarily close to $\mathscr{H}^\infty(R)$. Thus a partition P can be chosen where $\mathscr{H}^\infty(P) > 0$ and clearly P is an aperiodic partition. It is then easily shown that $P \vee Q$ is a finite, finite dimensional aperiodic partition.

Fix an $\eta > 0$ where $\eta < \gamma$ and

$$\eta \textstyle\sum_{k=1}^M \sum_{j=1}^{n_k} j\delta \le \epsilon/8.$$

Using the variant of the strong Rohlin theorem construct a $(\gamma, L, P \vee Q, \eta)$-gadget which has a finite dimensional base set F. Restricting attention to the Q partition on the base F, we now specify a sliding-block coder f_S by

$$\begin{aligned} f_S(x) &= a^* && \text{if } x \in G; \\ &= a^* && \text{if for some } 0 \le i < L-1;\ T^{-i}x \in G_0 \cap F \\ &= \hat{x} && \text{if for some } 0 \le i < L-1, \text{ some } 1 \le k \le M \text{ and some } 1 \le j \le n_k; \\ & && T^{-i}x \in E_k^j \cap F \text{ and the } i\text{th symbol of } \hat{\mathbf{x}}_k \text{ is } \hat{x}. \end{aligned} \tag{7}$$

In words, we label each column on the gadget directly above the set $E_k^j \cap F$ with the block of reproduction symbols $\hat{\mathbf{a}}_k$, we label the column directly above $G_0 \cap F$ with the block of reference symbols $\mathbf{a}^*$ and we label the garbage set G with the reference symbol a^*. For any $x \in \cup_{k=1}^M \cup_{j=1}^{n_k} E_k^j \cap F$, our sliding-block source coder acts as if we were applying the block code $f_B^{(L)}$ directly to the corresponding source vector $X_0^L(x)$.

Since the finite number of sets used in the definition of the sliding-block code f_S, (7), are all finite dimensional sets, it follows that f_S is a sliding-block source coder with constraint length $2N$ for a sufficiently large integer N. We also observe that the largest number of possible reproduction symbols which the sliding-block code f_S can output is

$$\|f_S(A^{2N+1})\| \le 1 + L\|f_B^{(L)}(A^L)\| \le L2^{LR} + 2 < \infty. \tag{8}$$

Therefore, $f_S \in \mathscr{C}_S'$.

STEP 4. We now compute bounds for the expected distortion $\rho(f_S)$. The definitions of the partition Q and the base set F imply that

$$\begin{aligned} \rho(f_S) &= \int_{A^\infty} \rho(X_0, f_S(X_{(-N,N)}))\, d\mu \\ &= \textstyle\sum_{k=1}^M \sum_{j=1}^{n_k} \sum_{i=0}^{L-1} \displaystyle\int_{T^i(E_k^j \cap F)} \rho(X_0, f_S(X_{(-N,N)}))\, d\mu \\ &\quad + \textstyle\sum_{i=0}^{L-1} \displaystyle\int_{T^i(G_0 \cap F)} \rho(X_0, f_S(X_{(-N,N)}))\, d\mu + \int_G \rho(X_0, f_S(X_{(-N,N)}))\, d\mu. \end{aligned} \tag{9}$$

Examining the first term of (9), the stationarity of μ and the definition of f_S, (7), imply for any $1 \le k \le M$ and $1 \le j \le n_k$ that

$$(10) \qquad \sum \int_{T^i(E_k^j \cap F)} \rho(X_0, f_S(X_{(-N,N)}))\, d\mu = L \int_{E_k^j \cap F} \rho_L(X_0^L, \hat{\mathbf{x}}_k)\, d\mu \le Lj\delta\mu(E_k^j \cap F)$$

where the last inequality follows from the definition of the set E_k^j and (2). Using (10), the fact that $\mu(F) > 0$ and property 3 of the variant of the strong Rohlin theorem,

$$(11) \qquad \begin{aligned} \sum_{k=1}^M \sum_{j=1}^{n_k} \sum_{i=0}^{L-1} \int_{T^i(E_k^j \cap F)} \rho(X_0, f_S(X_{(-N,N)}))\, d\mu &= L \sum_{k=1}^M \sum_{j=1}^{n_k} j\delta\mu_F(E_k^j)\mu(F) \\ &\le \sum_{k=1}^M \sum_{j=1}^{n_k} j\delta\mu(E_k^j) + \eta \sum_{k=1}^M \sum_{j=1}^{n_k} j\delta \end{aligned}$$

where the last inequality follows from the fact that $\mu(F) \le L^{-1}$. For the first term of (11) the definition of the set E_k^j and (2) show that

$$(12) \qquad \begin{aligned} \sum_{k=1}^M \sum_{j=1}^{n_k} j\delta\mu(E_k^j) &\le \sum_{k=1}^M \sum_{j=1}^{\infty} j\delta\mu(E_k^j) \\ &\le \sum_{k=1}^M \sum_{j=1}^{\infty} \int_{E_k^j} [\rho(X_0^L, f_B^{(L)}(X_0^L)) + \delta]\, d\mu^L \le D(R) + 3\epsilon/4 \end{aligned}$$

where the last inequality follows from the choice of the block code $f_B^{(L)}$ and the choice of δ. By the choice of η, the second term of (11) is bounded by $\epsilon/8$ and combining this with (12) yields

$$(13) \qquad \sum_{k=1}^M \sum_{j=1}^{n_k} \sum_{i=0}^{L-1} \int_{T^i(E_k^j \cap F)} \rho(X_0, f_S(X_{(-N,N)}))\, d\mu \le D(R) + 7\epsilon/8.$$

Examining the second term of (9) the definition of f_S, (7), shows that

$$(14) \qquad \begin{aligned} &\sum_{i=0}^{L-1} \int_{T^i(G_0 \cap F)} \rho(X_0, f_S(X_{(-N,N)}))\, d\mu \\ &= \sum_{i=0}^{L-1} \int_{T^i(G_0 \cap F)} \rho(X_0, a^*)\, d\mu \\ &\le \sum_{i=0}^{L-1} \epsilon/16L = \epsilon/16 \end{aligned}$$

where the last inequality follows from the fact that $\mu(T^i(G_0 \cap F)) \le \mu(T^iG_0) = \mu(G_0) < \gamma$ and relation (3).

Examining the last term of (9) the definition of f_S, (7), shows

$$(15) \qquad \int_G \rho(X_0, f_S(X_{(-N,N)}))\, d\mu = \int_G \rho(X_0, a^*)\, d\mu \le \epsilon/16L \le \epsilon/16$$

where the last inequality follows because $\mu(G) < \gamma$ by the construction of the (γ, L, $P \vee Q$, η)-gadget and relation (3).

Using the bounds (13), (14) and (15) in (9) yields

$$(16) \qquad \rho(f_S) \le D(R) + \epsilon.$$

STEP 5. We now compute the mean entropy rate of our distorted reproduction of the source. The conditions of Lemma 2.2 of Gray, Ornstein and Dobrushin (1980) are satisfied and, therefore,

$$(17) \qquad \bar{H}(\hat{X}) \le L^{-1}\log \| f_B^{(L)}(A^L) \| + h_b(L^{-1}) \le R - \delta + \delta = R$$

where the last inequality follows from the choice of the block code book $f_B^{(L)}$ and the block length L.

STEP 6. In Steps 1–3, a sliding-block code $f_S \in \mathscr{C}_S$ has been constructed. Then (16) and

(17) together imply that $\delta_S(R) \leq D(R) + \epsilon$ and since ϵ is arbitrary, $\delta_S(R) \leq D(R)$ for all rates $R \in (0, \infty)$, completing the proof of the theorem.

REFERENCES

[1] Berger, T. (1971). *Rate Distortion Theory.* Prentice-Hall, Englewood Cliffs, N.J.
[2] Davission, L. D. and Gray, R. M. (1978). A simplified proof of the sliding-block source coding theorem and its universal extension. *IEEE Trans. Information Theory.* To appear.
[3] Dunham, J. G. (1979). Process definitions of distortion-rate functions and OPTA functions. *Information and Control.* To appear.
[4] Gray, R. M. (1975). Sliding-block source coding. *IEEE Trans. Information Theory* **21** 375–368.
[5] Gray, R. M. and Ornstein, D. S. (1976). Sliding-block joint source/noisy-channel coding theorems. *IEEE Trans. Information Theory* **22** 682–690.
[6] Gray, R. M., Neuhoff, D. L. and Omura, J. K. (1975). Process definitions of distortion-rate functions and source coding theorems. *IEEE Trans. Information Theory* **21,** 525–532.
[7] Gray, R. M., Neuhoff, D. L. and Ornstein, D. S. (1975). Nonblock source coding with a fidelity criterion. *Ann. Probability* **3** 478–491.
[8] Gray, R. M., Ornstein, D. S. and Dobrushin, R. L. (1980). Block synchronization, sliding-block coding, invulnerable sources, and zero error codes for discrete noisy channels. *Ann. Probability* **8** 639–674.
[9] Halmos, P. (1956). *Lectures on Ergodic Theory.* Chelsa, New York.
[10] Hewitt, E. and Stromberg, K. (1969). *Real and Abstract Analysis, 2nd printing, corrected.* Springer, New York.
[11] Pinsker, M. S. (1964). *Information and Information Stability of Random Variables and Processes.* Holden-Day, San Francisco.
[12] Shannon, C. E. (1959). Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record* **4** 142–163.
[13] Shields, P. C. (1973). *The Theory of Bernoulli Shifts.* Univ. Chicago Press.
[14] Shields, P. C. and Neuhoff, D. L. (1977). Block and sliding-block source coding. *IEEE Trans. Information Theory* **23** 211–215.

Department of Electrical Engineering
Washington University
St. Louis, Missouri 63130