spending to \((i, j)\) by \((i^*, j^*)\). By definition

\[
\|V_n(u_i, x_i, j)\| \leq \sum_{i \leq m/4} \sum_{j < m'/4} \|V_n(u_i, x_i, j)\|.
\]

Moreover, for \(1 < i < m/4, 1 < j < m'/4\), we have by (27)

\[
A_{i^*, j^*}(\hat{\delta}) \leq A_{i^*, j^*}(\hat{\delta}) + \exp\left\{\frac{-1}{n} \sum_{i \leq m/4} \sum_{j < m'/4} \|V_n(u_i, x_i, j)\| \right\}.
\]

Substituting the last two inequalities into (28) and observing that

\[
\exp\left\{\frac{-1}{n} \sum_{i \leq m/4} \sum_{j < m'/4} \|V_n(u_i, x_i, j)\| \right\} < \exp\left\{\frac{-1}{n} \sum_{i \leq m/4} \sum_{j < m'/4} \|V_n(u_i, x_i, j)\| \right\}.
\]

holds trivially, we can establish (1). The other inequalities of the lemma are proved analogously.

**References**


**Extension of Source Coding Theorems for Block Codes to Sliding-Block Codes**

JOHN C. KIEFFER

*Abstract:* For arbitrary alphabets and single-letter fidelity criteria, two theorems are given which allow any fixed-rate or variable-rate source coding theorem for block codes to be extended to sliding-block codes. Applications are given to universal coding and to the coding of a stationary nonergodic source.

**I. BLOCK AND SLIDING BLOCK CODES**

Let \((A, \mathcal{A})\) and \((\hat{A}, \hat{\mathcal{A}})\) be measurable spaces. We assume \(\hat{\mathcal{A}}\) contains all singletons \(\{x\}\), \(x \in \hat{A}\). Let \((A^\omega, \mathcal{A}^\omega)\) be the measurable space consisting of \(A^\omega\), the set of all doubly infinite sequences \(x=(x_i)_{i=-\infty}^{\infty}\) from \(A\), and \(\mathcal{A}^\omega\), the usual product \(\sigma\)-field of subsets of \(A^\omega\). The space \((\hat{A}^\omega, \hat{\mathcal{A}}^\omega)\) is defined in similar fashion. If \(x=(x_i)\) is a finite or infinite sequence, let \(x(m) = (x_1, x_2, \ldots, x_m)\) and \(x(m) = (x_1, x_2, \ldots, x_{m-1})\). If \(\{W_i\}\) is a finite or infinite sequence of measurable functions, we similarly define \(W^m = (W_1, \ldots, W_k)\) and \(W^m = (W_1, \ldots, W_{m-1})\). However, we frequently write a random vector \((W_1, \ldots, W_n)\) as \(W^m\) rather than \(W^m\). Unless otherwise specified, for each integer \(i\), \(X_i\) denotes the map from \(A^n \rightarrow A\) such that \(X_i(x) = x_i, x \in A^n\). \(T_A\) denotes the shift transformation from \(A^n \rightarrow A^n\). We have \(T_A(x) = x_{i+1}\), \(x \in A^n\); \(i\) an integer. Similarly, \(T_A\) denotes the shift on \(A^\omega\).
We sometimes denote $T_A$ or $T_B$ by $T$ if the context makes clear what space the shift operates on. By a source with alphabet $A$, we mean a probability measure on $\mathbb{A}^\infty$. The source $\mu$ is stationary if $T$ preserves $\mu$ and ergodic if it is stationary, and every $T$-invariant measurable subset has $\mu$-measure 0 or 1. Let $\mathcal{P}$ be the set of all stationary sources.

Let $\mu: A^\times A^\rightarrow [0, \infty)$ be a jointly measurable function. For each $n=1, 2, \cdots$, let $\rho_n: A^n \times A^n \rightarrow [0, \infty)$ be the map such that, if $x=(x_1, \ldots, x_n), y=(y_1, \ldots, y_n) \in A^n$, then $\rho_n(x, y)=n^{-1} \sum_{i=1}^{n} \mu(x_i, y_i)$. We now define block and sliding-block codes used for fixed rate coding of the sources in $\mathcal{P}$ with respect to the single-letter fidelity criterion $\rho_n$.

A map $\phi: A^\infty \rightarrow A^\infty$ is a block code if there exists a positive integer $t$, a finite set $B \subseteq A^t$, and a measurable map $\psi: A^t \rightarrow B$ such that

$$\phi(x)_{n+1}^{n+t} = \psi(x_{n+1}^{n+t}), \quad x \in A^\infty, \quad i \in Z,$$

where $Z$ denotes the set of integers. We call $t$ the order of the code. The rate $r(\psi)$ of the foregoing block code $\phi$ is defined to be $t^{-1} \log |\phi(A^t)|$. (All logarithms in this paper are to base 2; also if $S$ is a set, $|S|$ denotes the cardinality of $S$.) The average distortion $\bar{\rho}(\phi, \mu)$ resulting from using $\phi$ to code $\mu \in \mathcal{P}$ is

$$\int_A \rho(x, \phi(x)) \, d\mu(x).$$

A map $\psi: A^2 \rightarrow A^\infty$ is a sliding-block code if there exists a positive integer $t$, a finite set $B \subseteq A^t$, and a measurable map $\phi: A^t \rightarrow B$ such that

$$\phi(x)_{n+1}^{n+t} = \psi(x_{n+1}^{n+t}), \quad x \in A^\infty, \quad i \in Z.$$

For the foregoing sliding-block code $\psi$ and for each $n=1, 2, \cdots$, let $M_n(\psi)=|\{ \psi(x)_{n+1}^{n+t} \mid x \in A^n \}|$. The rate $r(\psi)$ of $\psi$ is defined to be $n^{-1} \log M_n(\psi)$. (Since $\log M_n+\log M_{n+1} < \log M_n(\psi) + \log M_{n-1}(\psi)$, the limit exists; see [7, p. 112]). The average distortion $\bar{\rho}(\psi, \mu)$ resulting from using $\psi$ to code $\mu \in \mathcal{P}$ is

$$\int_A \rho(x, \psi(x)) \, d\mu(x).$$

We now state the first of the two main theorems. The proof is given in Section V.

**Theorem 1:** a) Given a block code $\phi$ and $\epsilon > 0$, there exists a sliding-block code $\psi$ such that $r(\psi) < r(\phi) + \epsilon$ and $\bar{\rho}(\psi, \mu) < \bar{\rho}(\phi, \mu)$ for every $\mu \in \mathcal{P}$.

b) Given a sliding-block code $\psi$ and $\epsilon > 0$, there exists a block code $\phi$ such that $r(\phi) < r(\psi) + \epsilon$ and $\bar{\rho}(\psi, \mu) < \bar{\rho}(\phi, \mu)$ for every $\mu \in \mathcal{P}$.

Theorem 1 can be used to extend any fixed-rate source coding theorem for block codes to sliding-block codes and vice versa. This means one no longer has to give a separate proof to extend each possible block coding theorem to sliding-block codes. We give applications in the next section.

**II. APPLICATIONS TO FIXED-RATE CODING**

If $\mu \in \mathcal{P}$ and $R > 0$, define $\delta_{(\mu, R)}=\inf \{ \bar{\rho}(\phi, \mu) \mid \phi$ a block code, $r(\phi) < R \}$. The quantity $\delta_{(\mu, R)}$ is the optimum distortion obtainable via block coding of $\mu$ at rate $R$. (A formula for $\delta_{(\mu, R)}$ in terms of the ergodic components of $\mu$ is given in [10].) Define $\delta_{(\mu, R)}=\inf \{ \bar{\rho}(\psi, \mu) \mid \psi$ a sliding-block code, $r(\psi) < R \}$. The quantity $\delta_{(\mu, R)}$ is the optimum distortion obtainable via sliding-block coding of $\mu$ at rate $R$. An immediate consequence of Theorem 1 is that $\delta_{(\mu, R)} = \delta_{(\mu, R)}$ for every $\mu \in \mathcal{P}$. Thus sliding-block and block coding of a single source at a fixed rate are equivalent. (This result had been shown for special cases in [6], [15], and [20]).

Let $\mathcal{M}$ be a subfamily of $\mathcal{P}$. We say weak universal fixed-rate block coding of $\mathcal{M}$ at rate $R$ is possible if there exists a sequence of block codes $\{ \phi_n \}_{n=1}^{\infty}$ such that the following two conditions hold:

$$r(\phi_n) < R, \quad n=1, 2, \cdots; \quad \lim_{n \rightarrow \infty} \bar{\rho}(\phi_n, \mu) = \delta_{(\mu, R)}, \quad \mu \in \mathcal{M}. \quad (2.2)$$

We say weak universal fixed-rate sliding-block coding of $\mathcal{M}$ at rate $R$ is possible if (2.1)–(2.2) hold for some sequence $\{ \phi_n \}$ of sliding-block codes.

Theorem 1 implies that weak universal fixed-rate block coding of $\mathcal{M}$ is possible if and only if weak universal fixed-rate sliding-block coding of $\mathcal{M}$ is possible. In [4], necessary and sufficient conditions were given in order for weak universal fixed-rate block coding of $\mathcal{M}$ at all rates to be possible. Thus these conditions are also necessary and sufficient for weak universal fixed-rate sliding-block coding of $\mathcal{M}$ at all rates to be possible.

We say strong universal fixed-rate block coding of $\mathcal{M}$ at the rate $R$ is possible if for any $\epsilon > 0$, there exists a block code $\phi$ such that the following two conditions hold:

$$r(\phi) < R + \epsilon; \quad \bar{\rho}(\phi, \mu) < \delta_{(\mu, R)} + \epsilon, \quad \mu \in \mathcal{M}. \quad (2.3)$$

We say strong universal fixed-rate sliding-block coding of $\mathcal{M}$ at rate $R$ is possible if for any $\epsilon > 0$, there exists a sliding-block code $\psi$ such that (2.3)–(2.4) hold. Theorem 1 implies that strong universal fixed-rate block coding of $\mathcal{M}$ is possible if and only if strong universal fixed-rate sliding-block coding of $\mathcal{M}$ is possible. For example, if $A=A^t$ and $\rho$ is a metric, then in [17] it was shown that strong universal fixed-rate block coding of $\mathcal{M}$ is possible if $\mathcal{M}$ is totally bounded in the $\rho$-metric. Thus this same result holds true for sliding-block coding.

**III. VARIABLE-RATE CODES**

Variable-rate block source coding with respect to the single letter fidelity criterion $\rho_n$ was considered in [1], [2], [3], and [4]. We consider variable-rate sliding-block source coding here.

Let $\{0, 1\}^*$ be the set of all binary sequences of finite length. A variable-rate block code is a quadruple $C_\psi=\langle \psi, A, n, \tau \rangle$, where $A$ is a finite subset of $A^\infty$, $\phi: A^m \rightarrow A^\infty$ is a block code of order $n$, and $\tau: A^\infty \rightarrow \{0, 1\}^*$ is a one-to-one map such that $\tau(A^\infty)$ is a prefix code. (This definition is slightly different from but equivalent to that given in [11]). The code $C_\psi$ is used in two steps: on the first step, $\phi$ is used to block code a sequence in $A^\infty$ into a sequence in $A^\infty$; on the second step, $\tau$ is used to block encode the resulting sequence in $A^\infty$ into a stream of binary symbols.
The second step of the coding is noiseless— all the distortion arises in the first step. The average distortion \( \bar{d}(C, \mu) \) resulting from using the foregoing \( C \) to encode source sequences produced by a source \( \mu \) is 
\[
\int_{A} d(x) \, d\mu(x).
\]
The average rate \( \bar{r}(C, \mu) \) is 
\[
\int_{A} r(x)/u(x) \, d\mu(x),
\]
where \( I \) denotes length.

Suppose we modify this two-step encoding procedure so that the first step involves coding in a sliding-block [6] manner. We thereby get a code we will call a variable-rate sliding-block code. More precisely, a variable-rate sliding-block code \( C \) is a quadruple \( (\psi, A, n, \tau) \), where \( A, n, \tau \) are as before and \( \psi \) is now a sliding-block code. The average distortion \( \bar{d}(C, \mu) \) for each \( \mu \in \mathcal{P} \) is 
\[
\int_{A} d(x) \, d\mu(x).
\]
The average rate \( \bar{r}(C, \mu) \) is 
\[
\int_{A} r(x)/u(x) \, d\mu(x).
\]

At this time, we make a slight change in our definitions which will be more convenient for proofs. If \( B \) is a finite subset of \( \mathbb{A}^n \) we call \( \sigma : B \to \{1, 2, \ldots \} \) a length function if \( \sum_{x \in B} 2^{-\sigma(x)} < 1 \). If \( \tau : B \to \{0, 1\}^* \) is a one-to-one map and \( \tau(B) \) is a prefix code, then it is known [1, ch. 3] that \( \sigma : \mathbb{A}^n \to \{1, 2, \ldots \} \) is a length function, where \( \sigma(x) = I(\tau(x)) \), \( x \in B \). Conversely, given a length function \( \sigma : B \to \{1, 2, \ldots \} \), there is a one-to-one \( \tau : B \to \{0, 1\}^* \) such that \( \tau(B) \) is a prefix code and \( I(\tau(x)) = \sigma(x), x \in B \). Hence, without loss of generality, we will regard a variable-rate block code as a quadruple \( C_\sigma = (\psi, A, n, \sigma) \) where \( \psi, A, n \) are as before but now \( \sigma : \mathbb{A}^n \to \{1, 2, \ldots \} \) and \( I(\sigma(x)) = \sigma(x) \). Similarly, a variable-rate sliding-block code \( C \) is a quadruple \( (\psi, A, n, \sigma) \) where \( \sigma \) is now a length function and \( \sigma : \mathbb{A}^n \to \{1, 2, \ldots \} \) and \( \bar{r}(C, \mu) \) is 
\[
\int_{A} r(x)/u(x) \, d\mu(x).
\]

Here is the second of the two main results of the paper. See Section V for the proof.

**Theorem 2.** a) Given a variable-rate block code \( C_\sigma \) and \( \epsilon > 0 \), there exists a variable-rate sliding-block code \( C_\sigma \) such that
\[
\bar{d}(C_\sigma, \mu) < \bar{d}(C, \mu), \quad \mu \in \mathcal{P}
\]
\[
\bar{r}(C_\sigma, \mu) < \bar{r}(C, \mu) + \epsilon, \quad \mu \in \mathcal{P}.
\]

b) Given a variable-rate sliding-block code \( C_\sigma \) and \( \epsilon > 0 \), there exists a variable-rate block code \( C_\mu \) such that
\[
\bar{d}(C_\mu, \mu) < \bar{d}(C_\sigma, \mu), \quad \mu \in \mathcal{P},
\]
\[
\bar{r}(C_\mu, \mu) < \bar{r}(C_\sigma, \mu) + \epsilon, \quad \mu \in \mathcal{P}.
\]

This theorem allows us to immediately extend any variable-rate source coding theorem for block codes to sliding-block codes and vice-versa. Applications are given in the next section.

**IV. APPLICATIONS TO VARIABLE-RATE CODING**

For each \( \mu \in \mathcal{P} \), let \( R_{\mu}(\cdot) \) denote the rate-distortion function for \( \mu \) relative to \( \left(\mu_n\right) \). If \( \mathcal{M} \subset \mathcal{P} \) is a family of ergodic sources and \( \{D_\mu : \mu \in \mathcal{M}\} \) are nonnegative numbers, we say weak universal variable-rate block coding of \( \mathcal{M} \) at the levels \( \{D_\mu\} \) is possible if there exists a sequence of variable-rate block codes \( \{C_\sigma^{(n)}\} \) such that
\[
\limsup_{n \to \infty} \bar{d}(C_\sigma^{(n)}, \mu) < D_\mu, \quad \mu \in \mathcal{M}, \quad (4.1)
\]
\[
\limsup_{n \to \infty} \bar{r}(C_\sigma^{(n)}, \mu) < R_\mu(D_\mu), \quad \mu \in \mathcal{M}. \quad (4.2)
\]
We say weak universal variable-rate sliding-block coding of \( \mathcal{M} \) at the levels \( \{D_\mu\} \) is possible if there exists a sequence of variable-rate sliding-block codes \( \{C_\sigma^{(n)}\} \) such that \((4.1)-(4.2)\) hold.

We see from Theorem 2 that weak universal variable-rate block coding of \( \mathcal{M} \) at the levels \( \{D_\mu\} \) is possible if and only if weak universal variable-rate block coding is possible.

We say strong universal variable-rate block coding of \( \mathcal{M} \) at the levels \( \{D_\mu\} \) is possible if for any \( \epsilon > 0 \) there exists a variable-rate block code \( C_\epsilon \) such that
\[
\bar{d}(C_\epsilon, \mu) < D_\mu + \epsilon, \quad \mu \in \mathcal{P}, \quad (4.3)
\]
\[
\bar{r}(C_\epsilon, \mu) < R_\mu(D_\mu) + \epsilon, \quad \mu \in \mathcal{P}. \quad (4.4)
\]
We say strong universal variable-rate sliding-block coding of \( \mathcal{M} \) at the levels \( \{D_\mu\} \) is possible if for any \( \epsilon > 0 \) there exists a variable-rate sliding-block code \( C_\epsilon \) such that \((4.3)-(4.4)\) hold.

We see from Theorem 2 that strong universal variable-rate block coding of \( \mathcal{M} \) at the levels \( \{D_\mu\} \) is possible if and only if strong universal variable-rate sliding-block coding is possible.

We now state a few variable-rate sliding-block source coding theorems which can be obtained with the help of Theorem 2.

First, we say a family \( \mathcal{M} \subset \mathcal{P} \) is \( \left(\mu_n\right)\)-separable if there is a countable set \( \mathcal{B} \) of block codes such that for any \( \mu \in \mathcal{M}, \epsilon > 0 \) and block code \( \phi \), there exists a \( \phi \in \mathcal{B} \) such that \( r(\phi) < r(\mu) + \epsilon \) and \( \bar{d}(\phi, \mu) < \bar{d}(\phi, \mu) + \epsilon \). The restriction that \( \mathcal{M} \) be \( \left(\mu_n\right)\)-separable is not very stringent; many examples of such families are given in [4] and [22]. In fact, in order for weak universal fixed-rate block coding to be possible for \( \mathcal{M} \), it is necessary and sufficient for \( \mathcal{M} \) to be \( \left(\mu_n\right)\)-separable; see [4].

The following is an extension of [3, th. 3] and [4, cor. 1 of th. 4] from block codes to sliding-block codes. It follows immediately from an application of Theorem 2 to [4, cor. 1 to th. 4]. (In the following, \( E_{\mu} \) denotes expectation with respect to the measure \( \mu \).)

**Theorem 3.** Let \( \mathcal{M} \) be a \( \left(\mu_n\right)\)-separable family of ergodic sources. Suppose there exists \( a \in \mathcal{A} \) such that \( E_{\mu} \rho(X_\mu, a^*) < \infty, \mu \in \mathcal{M} \). Let \( D > 0 \) be given satisfying \( \sup_{\mu \in \mathcal{M}} R_\mu(D) < \infty \). Also, suppose for each \( \mu \in \mathcal{M} \) there exists \( D_\mu^* < D \) with \( R_\mu(D_\mu^*) < \infty \). Then weak universal variable-rate sliding-block coding of \( \mathcal{M} \) at the levels \( \{D_\mu : \mu \in \mathcal{M}\} \) is possible, where \( D_\mu = D \) for all \( \mu \in \mathcal{M} \).

A measurable space \( (\Lambda, \mathcal{F}) \) is standard if \( \Lambda \) is a Borel subset of a complete separable metric space and \( \mathcal{F} \) is the collection of Borel subsets of \( \Lambda \). Given a measurable space \( (\Lambda, \mathcal{F}) \), we will say a family \( \{\mu_\theta : \theta \in \Lambda\} \) of sources with alphabet \( A \) is \( \mathcal{F}\)-regular if the mapping \( \theta \mapsto \mu_\theta \) is
one-to-one and if for each \( E \in \mathcal{E} \) the map \( \theta \mapsto \mu_\theta(E) \) is \( \mathcal{F} \)-measurable.

If \( f \) is a function whose domain is an interval on the real line and whose range space is the real line, \( f(x^+) \) will denote the right limit of \( f \) at \( x \).

The following is an extension of [4, cor. 2 of th. 4] from block codes to sliding-block codes. It will be proved in Appendix B.

**Theorem 4:** Let \((\Lambda, \mathcal{G}, \mu)\) be a probability space, where \((\Lambda, \mathcal{G})\) is standard. Let \((A, \mathcal{A})\) be standard and let \(\{\mu_\theta : \theta \in \Lambda\}\) be a \(\mathcal{G}\)-regular \(\{\mu_\theta\}\)-separable family of ergodic sources with alphabet \(A\). Suppose \(\theta \mapsto D_\theta\) is a measurable map from \(\Lambda\) to \([0, \infty)\). Suppose there exists \(\alpha^* \in A\) such that \(E_\alpha \rho(X_0, \alpha^*) < \infty, \theta \in \Lambda\). Then, given \(\varepsilon > 0\), there exists \(E \in \mathcal{E}\) with \(\mu(E) > 1 - \varepsilon\) and a variable-rate sliding-block code \(c\) such that

\[
\bar{\rho}(\xi, \mu_\theta) < D_\theta + \varepsilon, \quad \theta \in \Theta,
\]

and

\[
\bar{r}(\xi, \mu_\theta) < R_\mu(D_\theta^*) + \varepsilon, \quad \xi \in E.
\]

The following extends part of [8] from finite alphabets to infinite alphabets. We give the proof in Appendix B.

**Theorem 5:** Let the hypotheses of Theorem 4 hold except that instead of assuming \(E_\alpha \rho(X_0, \alpha^*) < \infty\) for all \(\alpha\) we assume \(\int E_\alpha \rho(X_0, \alpha^*) d\mu(\theta) < \infty\). Let \(\mu\) be the stationary source such that \(\mu(E) = \int \mu_\theta(E) d\mu(\theta), E \in \mathcal{E}\). Then, given \(\varepsilon > 0\), there exists \(E \in \mathcal{E}\) with \(\mu(E) > 1 - \varepsilon\) and a variable-rate sliding-block code \(c\) such that

\[
\bar{\rho}(\xi, \mu_\theta) < D_\theta + \varepsilon, \quad \theta \in \Theta,
\]

and

\[
\bar{r}(\xi, \mu_\theta) < R_\mu(D_\theta^*) + \varepsilon, \quad \xi \in E.
\]

The block coding version of the following was given in [9] for finite alphabets. We give the extension to infinite alphabets in Appendix C.

**Theorem 6:** Let \((A, \mathcal{A})\) be standard and let \(\rho\) be a stationary source with alphabet \(A\). Suppose there exists \(\alpha^* \in A\) such that \(E_\alpha \rho(X_0, \alpha^*) < \infty\). Let \(R_\rho(D) < \infty\). Then, given \(\varepsilon > 0\), there exists a variable-rate sliding-block code \(c\) such that

\[
\bar{\rho}(\xi, \mu_\theta) < D + \varepsilon, \quad \theta \in \Theta,
\]

and

\[
\bar{r}(\xi, \mu_\theta) < R_\rho(D) + \varepsilon.
\]

For our remaining theorem, we assume \(A = \hat{A}\) and that \(\rho\) is a metric on \(A\). (We take \(\mathcal{A}\) to be the collection of Borel subsets of \(A\) under the metric \(\rho\).) As defined in [5], we have the \(\rho\)-metric defined for pairs of sources.

The block coding version of the following was proved in [3] for finite alphabets. We give the extension to infinite alphabets in Appendix D.

**Theorem 7:** Let \(A = \hat{A}\) and let \(\rho\) be a metric. Let \(\mathcal{M}\) be a family of ergodic sources with alphabet \(A\) which is totally bounded in the \(\bar{\rho}\)-metric. Suppose \(\sup_{\alpha \in \mathcal{M}} R_\rho(D) < \infty\) and that there exists \(\alpha^* \in \hat{A}\) such that \(E_\alpha \rho(X_0, \alpha^*) < \infty, \mu \in \mathcal{M}\). Then strong universal variable-rate sliding-block coding of \(\mathcal{M}\) at the levels \(\{D_\mu : \mu \in \mathcal{M}\}\) is possible, where \(D_\mu = D\) for all \(\mu \in \mathcal{M}\).

**V. PROOFS OF MAIN THEOREMS**

First some notation. If \(S^\infty = \{0, 1, 2, \ldots\}\) is a sequence space and \(j \in \mathbb{N}\), we say \(f\) is finite-dimensional (f.d.) if for some \(k = 1, 2, \ldots, j \in \mathbb{N}\), and \(\rho(x) = f(x) \) whenever \(\rho(x^{(k)}) = \rho(x^{(j)})\). We say \(E \subset S^\infty\) is f.d. if \(I_E : S^\infty \to \{0, 1\}\), the indicator function of \(E\), is f.d.

If \((S^\infty_1, S^\infty_2)\) and \((S^\infty_3, S^\infty_4)\) are two sequence spaces, a channel \([S_1, S_2, \rho]\) is a family \(\rho = \{\rho(x) : x \in S^\infty_1\}\) if probability measures on \(S^\infty_2\), such that for each \(E \in \mathcal{E}\), the map \(x \mapsto \rho(E|x)\) from \(S^\infty_1 \to [0, 1]\) is \(S^\infty_1\)-measurable. We say a channel \([S_1, S_2, \rho]\) is f.d. if for each \(E \in \mathcal{E}\), the map \(x \mapsto \rho(E|x)\) is f.d. The channel \([S_1, S_2, \rho]\) is stationary if \(\rho(E|x) = \rho(T_1E|T_1x), x \in S^\infty_2, E \in \mathcal{E}\), where \(T_1, T_2\) are the shifts on \(S^\infty_2, S^\infty_3\), respectively. A measurable map \(f: S^\infty_3 \to S^\infty_4\) is f.d. if for each \(i \in \mathbb{N}\) there exists \(j \in \mathbb{N}\) and a positive integer \(k\) such that \(f(x) = f(x^{(k)})\), where \(x^{(k)} = x^{(j)}\).

Choose \(\delta > 0\) so that \(\delta < \varepsilon/2\). By Lemma A5 of Appendix A, there exists a finite set \(A \subset \hat{A}\), a stationary f.d. channel \([A, A', \rho]\), a positive integer \(t\) and a length function \(\sigma: A' \to \{1, 2, \ldots\}\) such that

\[
\int_{A'} \left[ \int_{A} \rho(x, y_0) d\nu(y|x) \right] d\mu(x) < \bar{\rho}(C_\mu, \mu), \quad \mu \in \mathcal{M},
\]

\[
\int_{A'} \left[ \int_{A} t^{-1} \rho(y_0^{-1}) d\nu(y|x) \right] d\mu(x) < \bar{r}(C_\mu, \mu) + \delta, \quad \mu \in \mathcal{M}.
\]

By Lemma A6 of Appendix A, there exists a positive integer \(N > t\) and for each \(k > N\) an f.d. measurable map \(\sigma_k: A^\infty \to \hat{A}^\infty\) such that

\[
\sum_{i=0}^{k-1} \rho(x, \alpha_i(x)) < (1 + \delta) \sum_{i=0}^{k-1} g(T^i(x)), \quad x \in A^\infty, \quad k > N.
\]

(5.7)

Define \(g, h: A^\infty \to [0, \infty)\) so that

\[
g(x) = \int_{A^\infty} \rho(x, y_0) d\nu(y|x), \quad x \in A^\infty.
\]

(5.5)

\[
h(x) = \int_{A^\infty} \sigma(x_0^{-1}) d\nu(y|x), \quad x \in A^\infty.
\]

(5.6)

By Lemma A5 of Appendix A, there exist a positive integer \(N > t\) and for each \(k > N\) an f.d. measurable map \(\alpha_k: A^\infty \to \hat{A}^\infty\) such that

\[
\sum_{i=0}^{k-1} \rho(x, \alpha_i(x)) < (1 + \delta) \sum_{i=0}^{k-1} g(T^i(x)) + \delta, \quad x \in A^\infty, \quad k > N.
\]

(5.8)
We may assume $N$ is so large that
\[-N^{-1} \log(N^{-1}) - (1 - N^{-1}) \log(1 - N^{-1}) < \delta, \quad (5.9)\]
\[-N^{-1} < \delta, \quad (5.10)\]
\[1 + \left[ 2t \log|\mathcal{A}| \right] + \left[ \log t \right] < N \delta. \quad (5.11)\]
Let $\tau : A \to \mathcal{A}$ be a measurable map such that $\rho(x, \tau(x)) = \min_{y \in \mathcal{A}} \rho(x, y)$. We note for later use that
\[\rho(x_i, \tau(x_i)) < g(T^i x), \quad x \in A^\omega, \quad i \in \mathbb{Z}. \quad (5.12)\]
From Lemma A1 of Appendix A, we may choose a finite-dimensional set $D \in \mathcal{G}_\omega$, a positive integer $L > N$, and for each $J > L$ a subset $S_J$ of $\mathcal{A}_J$ with no more than $2^J$ elements such that
\[
\text{if } x \in A^\omega \text{ and } T^i x \notin D \text{ for all } j \text{ in the range } i+1 < j < i+N-1; \quad (5.13)
\]
\[
\text{if } J > L \text{ and } x \in A^\omega \text{ and } T^i x \notin D \text{ for all } i \text{ in the range } k < i < k+J-2, \text{ then } \left( \tau(x_{k-1}), \tau(x_k), \ldots, \tau(x_{k+J-2}) \right) \in S_J. \quad (5.14)
\]
We define a sliding-block code $\psi : A^\omega \to \mathcal{A}_\omega$. Fix $x \in A^\omega$. We proceed to define $\psi(x)$. Let $\vartheta$ be the collection of all intervals $[m, n) \subset \mathbb{Z}$ such that
\[T^m x \in D, \quad T^n x \in D, \quad \text{and } T^i x \notin D \text{ for all } m < i < n, \quad (5.15)\]
\[n - m < L. \quad (5.16)\]
If $[m, n) \in \vartheta$, define $\psi(x)_n^{-1} = \alpha_{n-m}(T^m x)^0_0^{-m-1}$. If $i \in \mathbb{Z}$ and $i \notin \bigcup \{[m, n) : [m, n) \in \vartheta \}$, define $\psi(x)_i = \tau(x_i)$.
From (5.7)–(5.8) we see that if $[m, n) \in \vartheta$,
\[
\sum_{i=m}^{n-1} \rho(x_i, \psi(x)_i) < (1 + \delta) \sum_{i=m}^{n-1} g(T^i x), \quad (5.17)
\]
\[
\sum_{i=m}^{n-1} \sigma(\psi(x)_i) < \sum_{i=m}^{n-1} \left\{ h(T^i x) + \delta \right\}. \quad (5.18)
\]
For each $j > 1$, define a length function $\sigma_j : \mathcal{A}^j \to \{1, 2, \cdots\}$ so that
\[\sigma_j(y) = \left[ j \log|\mathcal{A}| \right] + 1, \quad j > 1, \quad y \in \mathcal{A}^j, \quad (5.19)\]
\[\sigma_j(y) = 1 + \left[ 2t \log|\mathcal{A}| \right] + \left[ \log t \right] + \min_{1 \leq i <, \left[ j \log|\mathcal{A}| \right] + 1, \quad j > 1, \quad y \in \mathcal{A}^j, \quad (5.19)\]
\[\sigma_j(y) = \left[ j \delta \right] + 1, \quad y \in \mathcal{A}^j, \quad (5.20)\]
\[\sigma_j(y) = \left[ j \delta \right] + 2, \quad y \in \mathcal{A}^j, \quad (5.20)\]
Choose $J > L$ a multiple of $N$ so large that
\[\left\lfloor \log \left( \sum_{i=0}^{JN^{-1} - 1} \left( \frac{\delta}{J} \right)^i \right) \right\rfloor < J \delta, \quad (5.22)\]
\[8 + 2 \left\lfloor \log|\mathcal{A}| \right\rfloor < J \delta. \quad (5.23)\]
(Since
\[
\lim_{k \to \infty} (kN)^{-1} \left\lfloor \log \left( \sum_{i=0}^{kN^{-1} - 1} \left( \frac{\delta}{kN} \right)^i \right) \right\rfloor = -N^{-1} \log(N^{-1}) - (1 - N^{-1}) \log(1 - N^{-1}) < \delta \quad (5.9)
\]
\[= -N^{-1} \log(N^{-1}) - (1 - N^{-1}) \log(1 - N^{-1}) < \delta \quad (5.9)
\]
\[\text{by (5.9) and } [23, \text{ eq. (4.7.5)}], \text{ this can be done.} \]
Choose a length function $\sigma^* : \mathcal{A}^* \to \{1, 2, \cdots\}$ so that for any increasing sequence $1 = n_0 < \cdots < n_r < n_{r+1} = J + 1$ with $n_k - n_{k-1} > N$, $k = 2, \cdots, r$:
\[\sigma^*(y) < \sum_{k=1}^{r+1} \sigma_{n_k - n_{k-1} - 1}(y(n_k^{-1}) - 1) + \left[ \log \left( \sum_{i=0}^{n_{r+1} - 1} \left( \frac{\delta}{J} \right)^i \right) \right], \quad y \in \mathcal{A}^j. \quad (5.24)\]
We show $\rho(\psi, \tau) = \rho(\mathcal{C}, \mu) < (1 + \delta) E_\mu \rho$, $\mu \in \mathcal{G}$. \quad (5.25)
But by (5.5) and (5.3), $E_\mu \rho < \bar{\rho}(\mathcal{C}, \mu)$, and so (5.1) follows from (5.25).
Determine integers $1 = n_0 < n_1 < \cdots < n_r < n_{r+1} = J + 1$, where $\{1 < i < J : T^i x \in D\} = \{n_1, \cdots, n_r\}$. From (5.10)–(5.11), (5.18)–(5.23), and the fact that $r < JN^{-1} + 1$, we obtain
\[\sigma^*(y) < J \delta + \sum_{k=1}^{r+1} \sigma_{n_k - n_{k-1} - 1}(y(n_k^{-1}) - 1) + \delta \sum_{k=1}^{r+1} \left\{ \left( \frac{n_k}{JN - 1} \right) + 1 \right\} \quad (5.25)\]
\[+ 2 \left\lfloor \log|\mathcal{A}| \right\rfloor + 2 \quad (5.25)\]
\[\leq J \delta + 2 \left\lfloor \log|\mathcal{A}| \right\rfloor \quad (5.25)\]
\[+ \sum_{k=1}^{r+1} \left\{ \left( \frac{n_k}{JN - 1} \right) + 1 \right\} \quad (5.25)\]
\[\leq J \delta + 2 \left\lfloor \log|\mathcal{A}| \right\rfloor \quad (5.25)\]
\[+ \sum_{k=1}^{r+1} \left\{ \left( \frac{n_k}{JN - 1} \right) + 1 \right\} \quad (5.25)\]
\[\leq J \delta + 2 \left\lfloor \log|\mathcal{A}| \right\rfloor \quad (5.25)\]
\[+ \sum_{k=1}^{r+1} \left\{ \left( \frac{n_k}{JN - 1} \right) + 1 \right\} \quad (5.25)\]
\[\leq J \delta + 2 \left\lfloor \log|\mathcal{A}| \right\rfloor \quad (5.25)\]
Taking the expected value, we get (5.2), since \( t^{-1}E_{\mu}h \leq \tilde{r}(C_0, \mu) + \delta \) by (5.6) and (5.4).

**Proof of Theorem 2b):** Let the sliding-block code \( C_s \) and \( \epsilon > 0 \) be given. We wish to find a block code \( C_b \) such that

\[
\tilde{r}(C_s, \mu) \leq \tilde{r}(C_b, \mu) + \epsilon, \quad \mu \in \mathcal{F}.
\]  

Say \( C_s = (\psi, \tilde{A}, N', \sigma) \). Now (5.31) reduces to

\[
\tilde{r}(C_s, \mu) \leq \tilde{r}(C_b, \mu) + \epsilon/2, \quad \mu \in \mathcal{F}.
\]  

From (5.30) and (5.32), \( \sup_{\mu \in \mathcal{F}} \tilde{r}(C_s, \mu) < \tilde{r}(\phi) + \epsilon \). Hence \( \tilde{r}(\psi, \nu) + \epsilon \) by Lemma A4 of Appendix A.

Theorem 1b) follows from Theorem 2b) in similar fashion.

**APPENDIX A**

**Lemma A1:** Let \( \tilde{A} \subset \tilde{A} \) be finite. Let \( \tau: \tilde{A} \rightarrow \tilde{A} \) be a given measurable function. Let a positive integer \( N \) and \( \delta > 0 \) be given.

Then there exists an f.d. \( D' \in \tilde{A}^\infty \), an integer \( L > n \), and for each \( J > L \) a subset \( \tilde{S}_J \) of \( \tilde{A}^L \) with no more than \( 2^L \) elements such that if \( x \in \tilde{A}^\infty \) and \( T'x \in D' \) then \( T'Ix \in \tilde{D}' \) for all \( i \) in the range \( k + 1 < i < n + 1 \),

\[
\text{if } J > L \text{ and } x \in \tilde{A}^\infty \text{ and } T'x \in D', \text{ then } T'Ix \in \tilde{D}' \text{ for all } i \text{ in the range } k + 1 < i < n + 1.
\]  

**Proof:** We can assume \( n \) is so large that \( n^{-1} \log|\tilde{A}| < \delta/2 \). Let \( N = n！ \) and \( \tilde{C} = \tilde{A}^N \) be the set of all blocks of form \( b \in \tilde{A}^n \) such that \( b \in \tilde{A}^{n+k} \) appears a total of \( n \) times. Let \( S = \{x \in \tilde{C}: x \in \tilde{A}^\infty \} \). We note the following property of \( S \).

If \( b \in S \) and \( x \in \tilde{A}^\infty \) and \( x_{j+k-1} \neq b_j \) for all \( j, \) then \( x_{j+k} \neq b_k \) for all \( i < j < i + n - 1 \).

Let \( b_1, b_2, \ldots, b_k \) be an enumeration of the elements of \( S \). Let \( D \) be the subset of \( \tilde{A}^\infty \) defined by

\[
D = \bigcup_{m \in \mathbb{N}} \bigcap_{i=1}^{m-1} \bigcap_{k=1}^{n} \{x \in \tilde{A}^\infty: x_0 \neq b_{m-1}, x_{j+k-n} \neq b_k \}.
\]  

From (A3) we obtain the property:

\[
\text{if } x \in \tilde{A}^\infty \text{ and } T_1 \in D, \text{ then } T_2x \in \tilde{D}' \text{ for } i \text{ in the range } k + 1 < i < n + 1.
\]  

From the way in which \( D \) was defined we obtain the following.

If \( x \in \tilde{A}^\infty \) and \( T_2 \in \tilde{D}' \) for all \( i \) in the range \( k < i < k + L \) then \( x_{j+k-n} \neq b_k \) for all \( i \) in the range \( k + (r-1)n < i < k + L - (r-1)n \).

Since \( N^{-1} \log|\tilde{A}| = n^{-1} \log|\tilde{A}| < \delta/2 \), it follows from (A6) that we may choose a positive integer \( L \) so large that

\[
\bigcup_{k=1}^{r} \bigcup_{i=k}^{r-1} \{x \in \tilde{A}^\infty: T_2x \in D, i + 1 < j < i + n - 1 \} < 2^L.
\]  

From (A5) and (A7), \( J > L \) follows from (A6).

**Lemma A2:** Suppose for every \( A, \tilde{A} \) and \( \phi: A \times \tilde{A} \rightarrow [0, \infty) \) the following is true.
For any variable-rate block code $C_\delta$ and $\delta > 0$ there exists a variable-rate sliding-block code $C_\delta$ such that $p(C_\delta, \mu) < (1 + \delta) \bar{p}(C_\delta, \mu)$ and $H(C_\delta, \mu) < \bar{H}(C_\delta, \mu) + \delta, \mu \in \mathfrak{P}$.

(A8)

Then Theorem 2a) follows from (A8).

**Proof:** Let $C_\delta = (\Phi, \mu, N, \sigma)$ and $\epsilon > 0$ be given. We can assume $N$ is so large that $(N + 1)^{-1} \log |A| < \epsilon/2$. Let $r: A^\infty \to A^\infty$ be a code block of order 1 such that $p(x_i, \tau(x_i)) = \min_{\mu, \pi, \rho} \rho(x_i, y_i)$ for all $y_i$. Let $\phi: A^\infty \to A^\infty$ be a block code of order $N + 1$ such that

$$\phi(x)_i = \phi(x)_i, \quad 1 \leq i \leq N$$

$$(N + 1)^{-1} \left[ N \sigma \phi(x) + \alpha(x_{N+1}, \tau(x_{N+1})) \right].$$

But $\alpha(x_{N+1}, \tau(x_{N+1})) = 0$, and so

$$\bar{a}(C_\delta, \mu) = (N + 1)^{-1} N \bar{a}(C_\delta, \mu), \quad \mu \in \mathfrak{P}. \quad \text{(A9)}$$

Now $(N + 1)^{-1} \bar{a}(\phi(x)N^{N+1}) = (N + 1)^{-1} \bar{a}(\phi(x)N^{N+1}) + \epsilon/2$. Hence

$$\bar{r}(C_\delta, \mu) < \bar{r}(C_\delta, \mu) + \epsilon/2, \quad \mu \in \mathfrak{P}. \quad \text{(A10)}$$

Choose $\delta > 0$ so small that $\delta < \epsilon/2$ and $(1 + \delta)(N + 1)^{-1} N < 1$. Applying (A8) to the fidelity criterion $(\alpha)$ we obtain $C_\delta = (\Phi, \mu, N, \sigma)$ such that

$$\alpha(C_\delta, \mu) < (1 + \delta) \bar{a}(C_\delta, \mu), \quad \mu \in \mathfrak{P}, \quad \text{(A11)}$$

$$\bar{r}(C_\delta, \mu) < \bar{r}(C_\delta, \mu) + \epsilon, \quad \mu \in \mathfrak{P}. \quad \text{(A12)}$$

We show that

$$\bar{p}(C_\delta, \mu) < \bar{p}(C_\delta, \mu), \quad \mu \in \mathfrak{P}, \quad \text{(A13)}$$

$$\bar{r}(C_\delta, \mu) < \bar{r}(C_\delta, \mu) + \epsilon, \quad \mu \in \mathfrak{P}. \quad \text{(A14)}$$

Now (A14) follows from (A12) and (A10). To show (A13), fix $\mu \in \mathfrak{P}$. If $\bar{p}(C_\delta, \mu) = \infty$, there is nothing to prove. Assume $\bar{p}(C_\delta, \mu) < \infty$. Then

$$\bar{p}(C_\delta, \mu) < \infty.$$ 

From (A9) and (A11) we have

$$\bar{a}(C_\delta, \mu) < \bar{a}(C_\delta, \mu). \quad \text{(A15)}$$

By definition of $\alpha$,

$$\alpha(x_0, \psi(x_0)) = \rho(x_0, \psi(x_0)) - \rho(x_0, \tau(x_0)).$$

$$\alpha(x_0, \psi(x_0)) = \min_{\mu, \pi} \rho(x_0, \psi(x_0)) - \rho(x_0, \tau(x_0)).$$

Integrating, we see that

$$\bar{a}(C_\delta, \mu) = \bar{a}(C_\delta, \mu) - \bar{p}(C_\delta, \mu) - \bar{r}(C_\delta, \mu).$$

Thus (A15) reduces to (A13).

**Lemma A3:** Let $f: A^\infty \to (-\infty, \infty)$ be a bounded measurable f.d. function. Suppose $E_{\mu, f} > 0$ for all $\mu \in \mathfrak{P}$. Given $\epsilon > 0$, there exists a positive integer $N$ such that for all $n > N$,

$$n^{n-1} \sum_{i=0}^{n-1} f(T^i) > \epsilon$$

throughout $A^\infty$.

**Proof:** If $g: A^\infty \to (-\infty, \infty)$ is finite-dimensional, and $k$ is a positive integer, we say $g$ is $k$-dimensional if for some $j \in \mathbb{Z}$, $g(x) = g(x)$ whenever $x_j^{(i)} = g_j^{(i)}$. We let $k$ be the dimension of the given f.d. function $f: A^\infty \to (-\infty, \infty)$. Fix $\epsilon > 0$ and let $M = \sup_{x \in \mathbb{Z}} f(x)$. Choose $N$ so large that $N^{-1} M < \epsilon$. Fix $n > N$ and let $\tilde{f} = f - f(1 - \Sigma_{i=0}^{n-1} f(T^i)$. Now $\tilde{f}$ is f.d. with dimension $n + k - 1$. It will thus follow that $\tilde{f} \to -\epsilon$ everywhere if $\tilde{f}(h) \to -\epsilon$ for every $h \in A^\infty$ satisfying $T^{n+k-1} = h$. Fix $b \in A^\infty$ satisfying $T^{n+k-1} = b$. Let $\mu \in \mathfrak{P}$ be the measure such that $\mu(E) = (n + k - 1)^{-1} \Sigma_{i=0}^{n-1} f(T^i)$. Now $E_{\mu, f} = E_{\mu, \tilde{f}} > 0$ by assumption. Consequently,

$$n^{n-1} \sum_{i=0}^{n-1} f(T^i) = n^{n-1} \sum_{i=0}^{n-1} f(T^i) + \epsilon$$

$$= n^{n-1} \sum_{i=0}^{n-1} f(T^i) > 0.$$

Hence

$$f(T^i) > \epsilon$$

for all $n > N$. Hence

$$\epsilon \leq \tilde{f}(b) > \epsilon \quad \text{or} \quad \tilde{f}(b) = \epsilon.$$
Proof: Let \( \mathcal{C}_g = (\phi, \tilde{A}, n, \sigma) \). Let \( \{A, \tilde{A}, \mu\} \) be the f.d. stationary channel such that

\[
q(E|x) = n^{-1} \sum_{i=1}^{n} I_{g} (T \phi(T^{-i} x)), \quad E \in \tilde{A}^\infty, \ x \in A^\infty.
\]

(\( \tilde{A}^\infty \) denotes the product field for \( \tilde{A}^\infty, \tilde{A} \) being the set of all subsets of \( A \)). Then, if \( \mu \in \mathcal{C}_g \),

\[
\int_{A^n} \left[ \int_{A^n} \rho(x_0, y_0) d\nu(y|x) \right] d\mu(x)
\]

is defined (A17). Let \( t \) be a multiple of \( n \) so large that

\[
\left[ 2n \log |\tilde{A}| \right] + [\log n] < t/2, \quad r^{-1} \text{ max } \sigma' < \epsilon/2.
\]

Let \( \sigma : \tilde{A} \rightarrow (1, 2, \cdots) \) be the length function

\[
\sigma(y) = \min_{0 < j < n-1} \sum_{1 < s < j + n} \sigma(y_{(s)})
\]

Then, if \( \mu \in \mathcal{C}_g \),

\[
\int_{A^n} \left[ \int_{A^n} \rho(x_0, y_0) d\nu(y|x) \right] d\mu(x)
\]

For \( 1 < i < n \),

\[
\int_{A^n} \sigma(\phi(x)^{i+1}_+) d\mu(x)
\]

\[
\int_{A^n} \rho(\sigma(\phi(x)^{n+i}_+)) d\mu(x)
\]

The integer \( r \) must satisfy

\[
(r+2)J > k.
\]

Choose a positive integer \( J \) so large that

\[
2M/J < \delta.
\]

We now are able to specify \( k \). We choose \( k \) any positive integer large enough so that

\[
J/k < \delta,
\]

Partition \( [0, k) \) into subintervals \( [0, J), [J, 2J), \cdots, [rJ, (r+1)J) \). The integer \( r \) must satisfy \( (r+2)J > k \). Define

\[
I_s = [sJ, (s+1)J), \quad 0 < s < r.
\]

Fix \( x \in A^\infty \). Let \( \lambda_f \) be the probability measure on \( \tilde{A}^\infty \) such that \( \lambda_f(\phi)^{(i)} \) is independent and for each \( i \in \mathbb{Z} \) the distribution of \( Y_{(i)}^{(i)} \) under \( \lambda_f \) and under \( \mu(x) \) are the same. For \( \beta > 0 \) we need the estimate

\[
\lambda \left[ \sum_{i=0}^{k-1} f(Y_{(i)}^{(i)}) \right] > \sum_{i=0}^{k-1} \left[ E_{[0, k)} f(Y_{(i)}^{(i)}) + \beta \right] > k^{-2} \beta^{-2} C^2 \left[ 2Jk + k^{2} (1 - (1 - \delta)^{4}) \right].
\]
To see this, first observe that by Chebyshev’s inequality the left side of (A25) is no bigger than
\[ k^{-2} \beta^2 \sum_{i,j \in D} \text{COV}_b(\gamma_i(M), \gamma_j(M)). \]  
(A26)

Let \( D = \cup_i \cup_j I_i^j \). Since each covariance term is no bigger than \( C^2 \), we can upper bound (A26) by
\[ k^{-2} \beta^2 \sum_{i,j \in D} \text{COV}_b(\gamma_i(M), \gamma_j(M)) + C^2(2k^2 - |D|^2). \]  
(A27)

Now \( |D| \geq (r + 1)(J - 2M) \geq (kJ - 1)(J - 2M) \). We have \( kJ - 1 \geq (1 - \delta)kJ - 1 \) by (A24). Also, \( J \geq 2M \) by (A23). Hence \( |D| \geq (1 - \delta)^2 kJ \). Note that if \( i, j \in D \) and \( |i - j| > J \), then \( \gamma_i(M), \gamma_j(M) \) are independent and so \( \text{COV}_b(\gamma_i(M), \gamma_j(M)) = 0 \). There are no more than \( 2Jk \) pairs \((i, j)\) such that \( i, j \in D \) and \( |i - j| < J \). Thus (A27) is upper bounded by the right side of (A25).

Now if \( i \in D \), \( \gamma_i(M) \) has the same distribution under \( \lambda \) as under \( f(\cdot|x) \). By the stationarity of \( \nu \), the \( \nu(\cdot|x) \) expectation of \( \gamma_i(M) \) is \( h(T'x) \). Hence
\[ \lambda \gamma_i(M) = h(T'x), \quad i \in D. \]

Consequently
\[ \sum_{i=0}^{k-1} E_{\lambda} f^s(\gamma_i(M)) - \sum_{i=0}^{k-1} h(T'x) < 2C(k - |D|) \leq 2Ck \left( 1 - (1 - \delta)^2 \right) \leq 4Ck \delta k < ek/2. \]

This gives
\[ \lambda \left[ \sum_{i=0}^{k-1} f^s(\gamma_i(M)) > \sum_{i=0}^{k-1} h(T'x) + \epsilon \right] \leq \lambda \left[ \sum_{i=0}^{k-1} f^s(\gamma_i(M)) > \sum_{i=0}^{k-1} \left( E_{\lambda} f^s(\gamma_i(M)) + \epsilon/2 \right) \right] < k^{-2}(2/2)C^2 \left( 2Jk - k(1 - (1 - \delta)^2) \right) < \epsilon/1 + \epsilon, \]
by (A22), (A24), and (A25). Now for all \( i \),
\[ \lambda \left[ p(x_i, Y_i) > (1 + \epsilon) \sum_{i=0}^{k-1} h(T'x) \right] < \frac{1}{1 + \epsilon}. \]  
(A28)

From (A28)–(A29), we see that \( \lambda(F'_x) < \epsilon/(1 + \epsilon) + 1/(1 + \epsilon) = 1 \). Hence \( \lambda(F'_x) > 0 \) and we are done.

**Lemma A7**: Let \( f : \mathbb{A} \to [0, \infty] \) be \( \mathbb{B} \)-measurable. Let \( \mu \) be a stationary source with alphabet \( \mathbb{A} \). Then
\[ E_\mu \left[ \lim \sup_{n \to \infty} \sum_{i=1}^{n} f(T^i) \right] = E_\mu \left[ \lim \inf_{n \to \infty} \sum_{i=1}^{n} f(T^i) \right] = E_\mu[f]. \]

**Proof**: If \( E_\mu[f] < \infty \), the result follows from the ergodic theorem. Thus we suppose \( E_\mu[f] = \infty \). Since
\[ E_\mu \left[ \lim \sup_{n \to \infty} \sum_{i=1}^{n} f(T^i) \right] > E_\mu \left[ \lim \inf_{n \to \infty} \sum_{i=1}^{n} f(T^i) \right], \]
it suffices to show
\[ E_\mu \left[ \lim \inf_{n \to \infty} \sum_{i=1}^{n} f(T^i) \right] = \infty. \]

Let \( \{f_k\} \) be an increasing sequence of bounded functions which converges pointwise to \( f \). For each \( k \),
\[ E_\mu \left[ \lim \inf_{n \to \infty} \sum_{i=1}^{n} f_k(T^i) \right] > E_\mu \left[ \lim \inf_{n \to \infty} \sum_{i=1}^{n} f(T^i) \right] = E_\mu[f]. \]

By the monotone convergence theorem \( E_\mu[f_k] = \infty \) as \( k \to \infty \).

**APPENDIX B**

From now on, we frequently regard a block code as a finite subset \( B \) of \( \mathbb{A}^n \) for some \( n \), rather than as a map \( \phi : \mathbb{A}^\infty \to \mathbb{A}^\infty \). The justification for this is that such a \( B \) yields a block code \( \phi : \mathbb{A}^\infty \to \mathbb{A}^\infty \) of order \( n \) satisfying \( \rho_b(x_1^n, \phi(x_1^n)) = \inf_{B \in B} \rho_b(x_1^n, y) \). For a block code \( B \subset \mathbb{A}^n \), we define its rate \( r(B) \) to be \( n^{-1} \log |B| \), and for each probability measure \( \mu \) on \( \mathbb{A}^\infty \), the average distortion \( \rho_b(B) \) is defined to be \( f_{1/2} \rho_b(x_1^n, B) d\mu(x) \), where for each \( x \in \mathbb{A}^n \), \( \rho_b(x_1^n, B) = \inf_B \rho_b(x_1^n, y) \).

The proof of the following lemma involves only a slight modification in the proof of [4, th. 4] so we omit the proof.

**Lemma B1**: Given a family \( \mathfrak{G} \) of ergodic sources with alphabet \( \mathbb{A} \); an element \( \alpha^* \in \mathfrak{A} \) such that \( E_\alpha \rho(X_0, \alpha^*) < \infty \), \( \rho \in \mathfrak{G} \); and families of nonnegative numbers \( \{D_\alpha : \mu \in \mathfrak{G}\} \) and \( \{R_\mu : \mu \in \mathfrak{G}\} \).

Suppose \( \sup_{\mu \in \mathfrak{G}} R_\mu < \infty \) and that, for each \( n = 1, 2, \ldots \), there exists a measurable function \( Y_{1/n} : \mathbb{A}^n \to [0, \infty] \) such that, for each \( \mu \in \mathfrak{G} \), the sequence \( \{Y_{1/n}(X_1^n)\} \) converges as \( n \to \infty \) in \( L_1(\mu) \) to \( D_\mu \). Suppose there exists a countable set \( \mathcal{B} \) of block codes such that for any \( \epsilon > 0 \) and \( \mu \in \mathfrak{G} \) there exists \( B \in \mathcal{B} \) with \( \rho_b(B) < D_\mu + \epsilon \).

Then, for each \( n = 1, 2, \ldots \), there exists a variable-rate block code \( C_n = (\phi_n, A_n, n, o_n) \) such that
a) \( \rho(\mathcal{C}_n, \mu) < E_\mu \rho(X_0, \alpha^*), \quad \mu \in \mathfrak{G} \);

b) \( n^{-1} \max_{\mu \in \mathfrak{G}} o_n < \sup_{\mu \in \mathfrak{G}} R_\mu \); 5;

c) \( \lim \sup_{n \to \infty} \rho(\mathcal{C}_n, \mu) < D_\mu, \mu \in \mathfrak{G} \);

d) \( \lim \inf_{n \to \infty} \rho(\mathcal{C}_n, \mu) > R_\mu, \mu \in \mathfrak{G} \).

Let \( (\Omega, \mathcal{F}) \) be a measurable space and let \( T : \Omega \to \Omega \) be a measurable transformation. Unless otherwise specified, if \( \Omega \) is a sequence space \( B^\infty \) we require \( T \) to be the shift \( T_B \); if \( \Omega = B^\infty \times C^\infty \), the product of two sequence spaces, we require \( T \) to be \( T_B \times T_C \). If \( E, F \) are sets and \( U : E \to E, V : F \to F \) are maps, then \( U \times V \) denotes the map \( E \times F \to E \times F \) such that \( (U \times V)(x, y) = (U(x), V(y)) \), \( x \in E, y \in F \). If \( P \) is a probability measure on \( \mathcal{F} \) stationary with respect to \( T \). We define an ergodic decomposition of \( P \) (relative to \( T \)) to be a family \( \{P_x : x \in \Omega\} \) of probability measures on \( \mathcal{F} \) such that
a) \( P_x \) is stationary and ergodic with respect to \( T, x \in \Omega \);

b) \( P_x = P_{T_x}, x \in \Omega \);

c) for each \( E \in \mathcal{F} \), the map \( x \to P_x(E) \) from \( \Omega \to [0, 1] \) is \( \mathcal{F} \)-measurable, and

d) \( P(E \cap F) = \int F(P_x(dP(x)), E, F \in \mathcal{F}, T^{-1}F = F \).
Letting $\mathcal{F}$ be the sub-$\sigma$-field of $\mathcal{F}$ consisting of all $F \in \mathcal{F}$ such that $T^{-1}F = F$, we see that for each $E \in \mathcal{F}$, the map $x \rightarrow P(E)$ from $\Omega \rightarrow [0,1]$ is just $P(E | \mathcal{F})$, the conditional probability function for $E$ with respect to $\mathcal{F}$. Thus it is easy to see from a theorem on regular conditional probabilities [19, th. 6.6.51, that $P(E | \mathcal{F})$ is essentially unique; i.e., if $\{P_1\}$ and $\{P_2\}$ are two ergodic decompositions of $P$, then for $P$-almost all $x \in \Omega$, $P_1 = P_2$.

We assume from now on that the space $(A, \mathcal{A})$ is standard.

**Lemma B2:** Let $(A, \mathcal{A})$ be standard and $\mu$ be stationary. Let $(\mu_x : x \in A^\omega)$ be an ergodic decomposition of $\mu$. Suppose, for each $x \in A^\omega$, nonnegative numbers $\rho(x)$ and $D(\mu_x)$ are defined which depend on $x$ only through $\mu_x$. Let the maps $x \rightarrow D(\mu_x)$ and $x \rightarrow R(\mu_x)$ be measurable. Assume there exists a countable class $\mathcal{G}$ of block codes such that for any $\epsilon > 0$ and any $x$, there exists $B \in \mathcal{G}$ with $R(B) < R(\mu_x) + \epsilon$ and $\rho(B) < D(\mu_x) + \epsilon$. Suppose there exists a $\tau \in \mathcal{A}$ such that $\mu(\tau)$ is smaller than $\rho$. Letting $\mathcal{G}_0$ be an ergodic decomposition of $\mu$. Suppose, for each $\epsilon > 0$, there exists a variable-rate sliding-block code $\mathcal{G}$ and a set $W \in \mathcal{F}$ with $\mu(W) > 1 - \epsilon$ such that

$$r(\mathcal{C}, \mu_x) \leq R(\mu_x) \mu(dx) + \epsilon,$$

$$\rho(\mathcal{C}, \mu_x) \leq D(\mu_x) \mu(dx) + \epsilon,$$

$$\rho(\mathcal{C}, \mu_x) \leq R(\mu_x) \mu(dx) + \epsilon,$$

$$\rho(\mathcal{C}, \mu_x) \leq D(\mu_x) \mu(dx) + \epsilon,$$

where

$$\rho(\mathcal{C}, \mu_x) = \int_{W^c} \rho(\mathcal{C}, \mu_x) \mu(dx) + \int_{W \cap \tau} \rho(\mathcal{C}, \mu_x) \mu(dx) + \mu(\tau) \rho(\mathcal{C}, \mu_x) \mu(dx).$$

Now, let

$$r(\mathcal{C}, \mu_x) \leq R(\mu_x) \mu(dx) + \epsilon,$$

$$\rho(\mathcal{C}, \mu_x) \leq D(\mu_x) \mu(dx) + \epsilon,$$

Expressions (B6), (B8), (B9), and (B10) yield (B3), (B4), (B1), and (B2), respectively, if $M$ is chosen large enough so that $\mu(W)_{i+1}$ is sufficiently small, and $W$, $W'$ are chosen so that $\mu(W \cap W')$ is sufficiently small.

We introduce the following notation. Let $\mathcal{S}$ denote the class of all ergodic sources with alphabet $A$. We let $\mathcal{S'(S)}$ denote the smallest $\sigma$-field of subsets of $\mathcal{S}$ such that for each $E \in \mathcal{S}$, the map $\mu \rightarrow \mu(E)$ defined on $\mathcal{S}$ is measurable. If $(A, \mathcal{A})$ is standard, then $\mathcal{S'(S)}$ is standard [13].

**Lemma B3:** Let $(A, \mathcal{A})$ be standard and $\mu \in \mathcal{S'(S)}$ be $(\rho_\tau)$-separable. For each $\mu \in \mathcal{S}$, let $D_\mu > 0$ be defined. Suppose the map $\mu \rightarrow D_\mu$ on $\mathcal{S}$ is $\mathcal{S'(S)}$-measurable, where $\mathcal{S'(S)} = \{ (\mathcal{S}, \mathcal{S'(S)}) \}$. Suppose also that $D_\mu(A^\omega) < \infty$, $\mu \in \mathcal{S'}$. Then the map $\mu \rightarrow D_\mu(A^\omega)$ on $\mathcal{S'(S)}$ is $\mathcal{S'(S)}$-measurable.

**Proof:** Let $\mathcal{S}$ be a countable collection of block codes which serves in the definition of $(\rho_\tau)$-separability. If $\alpha > 0$, then

$$\{ \mu \in \mathcal{S} : \mathcal{H}(D_\mu) < \alpha \} \cap \bigcup_{\alpha < \alpha} \{ \mu \in \mathcal{S} : \mathcal{H}(D_\mu) < \alpha \}$$

is a set in $\mathcal{S'(S)}$.

**Proof of Theorem 4:** The fact that $\{ \mu_\tau : \tau \in \Lambda \}$ is $\mathcal{S}$-regular is equivalent to saying that the map $\theta \rightarrow \mu_\theta$ is a one-to-one measurable mapping from the standard space $(\Lambda, \mathcal{S})$ to the standard space $(\mathcal{S}, \mathcal{S'(S)})$. From the theory of standard spaces [13], this implies that the image $\mathcal{G}$ of $\Lambda$ under the map $\theta \rightarrow \mu_\theta$ is in $\mathcal{S'(S)}$ and that the map $\theta \rightarrow \mu_\theta$ is an isomorphism of the measurable spaces $(\Lambda, \mathcal{S})$ and $(\mathcal{S'}, \mathcal{S'(S)})$. We let $\tau : \mathcal{S'} \rightarrow \Lambda$ be the inverse of the map $\theta \rightarrow \mu_\theta$. Let $\rho$ be the stationary source with alphabet $A$ such that $\mu(E) = \int_p \rho(\mathcal{C}, \mu_\theta) d\theta$, $E \in \mathcal{S'}$. Choose $\{ \mu_x : x \in A^\omega \}$ an ergodic decomposition of $\mu$. We may assume each $\mu_x \in \mathcal{S'}$. For each $\rho \in \mathcal{S}$, define $D(\rho) = D_{\rho_\tau}$ and $R(\rho) = R_\rho(D_\rho)$. By [18, th. 7.2.4] and the $(\rho_\tau)$-separability of $\mathcal{S}$, there is a countable class $\mathcal{G}$ of block codes such that for any $x$ and any $\alpha > 0$ there exists $\mathcal{G} \in \mathcal{S}$ such that $\rho(B) < D_\rho + \alpha$ and $R(B) < R_\rho(D_\rho) + \alpha$. The map $\theta \rightarrow D_\mu$ is measurable by assumption, so using the isomorphism between $\Lambda$ and $\mathcal{S'}$, we conclude $x \rightarrow D_\mu(x)$ is measurable. Similarly, the measurability of $\theta \rightarrow R_\rho(D_\rho)$ by Lemma B3 implies the measurability of $x \rightarrow R_\rho(D_\rho)$. We may suppose $\int_A \rho(x, a^*) \mu(dx) < \infty$. (Otherwise, pick $E \in \mathcal{S}$ such that $\int_p \rho(X, a^*) \mu(dx) < \infty$ and $\rho(E) > 1 - \epsilon/2$. Letting $p'$ be the measure on $\mathcal{S}$ such that $p'(\mathcal{C}) = p'((\mathcal{C} \cap \mathcal{F}) \cup \mathcal{F})$, suppose we find $E' \in \mathcal{S}$ such that $p'(E') > 1 - \epsilon/2$ and a sliding-block code $\mathcal{C}$ such that $\rho(\mathcal{C}, \mu) < D_\rho + \epsilon$ and $r(\mathcal{C}, \mu) < R_\rho(D_\rho) + \epsilon$. Then, $p'(E') > p'(E) > 1 - \epsilon$ and we are done.) By Lemma B2, there exists a sliding-block code $\mathcal{C}$ such that

$$\mu(\mathcal{C}, \mu) = \int_{W^c} \rho(\mathcal{C}, \mu_x) \mu(dx) + \int_{W \cap \tau} \rho(\mathcal{C}, \mu_x) \mu(dx) + \mu(\tau) \rho(\mathcal{C}, \mu_x) \mu(dx),$$

Now, (B11) implies $\int_A \int_p \rho(x, a^*) \mu(dx) < \infty$. (Otherwise, pick $E \in \mathcal{S}$ such that $\int_p \rho(x, a^*) \mu(dx) < \infty$ and $\rho(E) > 1 - \epsilon/2$. Letting $p'$ be the measure on $\mathcal{S}$ such that $p'(\mathcal{C}) = p'((\mathcal{C} \cap \mathcal{F}) \cup \mathcal{F})$, suppose we find $E' \in \mathcal{S}$ such that $p'(E') > 1 - \epsilon/2$ and a sliding-block code $\mathcal{C}$ such that $\rho(\mathcal{C}, \mu) < D_\rho + \epsilon$ and $r(\mathcal{C}, \mu) < R_\rho(D_\rho) + \epsilon$. Then, $p'(E') > p'(E) > 1 - \epsilon$ and we are done.) By Lemma B2, there exists a sliding-block code $\mathcal{C}$ such that

$$\mu(x \in A^\omega : \rho(\mathcal{C}, \mu_x) < D_\rho + \epsilon, r(\mathcal{C}, \mu_x) < R_\rho(D_\rho) + \epsilon) > 1 - \epsilon.$$
Let \( \mu' \) be the measure on \( S \) such that
\[
\mu'(A) = \mu(\theta \in \Lambda : \theta \in A), \quad A \in \mathfrak{G}(\Sigma). \tag{B14}
\]

By a similar argument to that following (B12), we observe
\[
\int_E \mu'_{\theta}(E) = f(E) = \text{sup}_{\theta \in \Lambda} \int_{\mathfrak{B}(\Sigma)} f(\theta) \mu'_{\theta}(\cdot), \quad E \in \mathfrak{B}(\Sigma),
\]
and since \( \mu(E) = \int_\Sigma \mu'_{\theta}(E) d\mu(\theta) \), we obtain
\[
\mu(E) = \int_{\mathfrak{B}(\Sigma)} f(E) d\mu(\theta), \quad E \in \mathfrak{B}(\Sigma). \tag{B15}
\]

Combining (B15) and (B13), we see from uniqueness of \( \mu^* \) that \( \mu' = \mu^* \).

Let \( \Sigma = \{v \in S' : \theta \in A_{\infty} \}, \mu^* : S' \to \Lambda, \) and \( (D(x) : x \in \Sigma) \) be measurable. By (B11) and (B12), \( \mu^*(\theta) \mid \mu \in \mathfrak{A}_x \), \( \mu^*(\theta) = 1 - \epsilon \). Set \( E = \{(\theta \in \Lambda : \theta \in \Sigma) \} \). Then, by (B14), \( \mu(E) = \mu^*(\theta) > 1 - \epsilon \). Since \( \mu(\Sigma, \mu^*), \mu^* < \mu + \epsilon \), \( \mu^* = \mu \).

Proof of Theorem 5: Define \( S'_0 = (\mu^*, \mu') \), \( \mu^* : S' \to \Lambda, \) and \( (D(x) : x \in \Sigma) \) as in the proof of Theorem 4. By Lemma B2, we obtain \( S \) such that
\[
\rho(\Sigma, \mu) \leq \int_{\mathfrak{B}(\Sigma)} f(D(x)) d\mu(x) + \epsilon.
\]

Now \( \int_{\mathfrak{B}(\Sigma)} f(D(x)) d\mu(x) = \int_{\mathfrak{B}(\Sigma)} f(D(x)) d\mu'(x) \) follows from (B12). Hence
\[
\int_{\mathfrak{B}(\Sigma)} f(D(x)) d\mu(x) < \int_{\mathfrak{B}(\Sigma)} f(D(x)) d\mu'(x).
\]

Therefore, \( \mu^* \) is a stationary source with alphabet \( S_0^{\infty} \)."
By the definition of rate-distortion function [18], we now choose \( m \) so that there is a \( m \) ary DMC \([A, A, v]\) such that
\[
\sum_{x \in A^m} p(x) I_{\mu_x}(X^m, Y^m) < R_x(D) + \epsilon/2,
\]
and
\[
E_{\mu_x} p_m(X^m, Y^m) < D.
\]

Let \( U=\{U_i\} \) and \( V=\{V_i\} \) be the processes defined on \( A^\infty \times A^\infty \) (relative to \( T_{A} \times T_{A} \)) such that \( U_i = Y_{(m+1)}^m, V_i = Y_{(m+1)}^m, i \in \mathbb{Z} \).

Choose a family \( \{\lambda_x: x \in A^\infty\} \) of probability measures on \( \mathbb{R}^\infty \) such that
\[\begin{align*}
&\text{a) } \lambda_x \text{ is stationary and ergodic relative to } T^m, x \in A^\infty, \\
&\text{b) } \lambda_{Tx} = \lambda_x, x \in A^\infty, \\
&\text{c) the map } x + A, E \rightarrow \mathbb{R}^\infty \text{ is } T^m \text{-measurable,} \\
&\text{d) } p(E \cap F) = \int p(E) \, d\mu(E), \text{for } E \in \mathbb{R}^\infty, F \text{ } T^m \text{-invariant,} \\
&\text{e) } \lambda_x(\{x_i: i \in \mathbb{Z}\}) < \infty, x \in A^\infty. \\
\end{align*}\]

We can choose \( \{\lambda_x\} \) satisfying a)-d) by ergodic decomposition theory [11, 12]. To see that e) can be made to hold, note that if \( E, F \in \mathbb{R}^\infty \) and \( T_{A}E = F \), then
\[\int p(E \cap F) \, d\mu(E) = \int p(E) \, d\mu(E), \text{for } E \in \mathbb{R}^\infty, F \text{ } T^m \text{-invariant.}\]

This forces \( p(x \in A^\infty; \lambda_x = \lambda_{Tx} = 1, \) and so redefining \( \{\lambda_x\} \) on a set of \( \mu \)-measure 0 we get e). Similarly we can assume \( E_{\lambda_x}(X_0, a^*) < \infty \) for all components \( \lambda_x \) since
\[\int E_{\lambda_x}(X_0, a^*) \, d\mu(x) = E_{\lambda_x}(X_0, a^*) < \infty. \]

Now for each \( x \in A^\infty \), define
\[\int p(E \cap F) \, d\mu(E) = \int p(E) \, d\mu(E), \text{for } E \in \mathbb{R}^\infty, F \text{ } T^m \text{-invariant.}\]

This forces \( p(x \in A^\infty; \lambda_x = \lambda_{Tx} = 1, \) and so redefining \( \{\lambda_x\} \) on a set of \( \mu \)-measure 0 we get e). Similarly we can assume \( E_{\lambda_x}(X_0, a^*) < \infty \) for all components \( \lambda_x \) since
\[\int E_{\lambda_x}(X_0, a^*) \, d\mu(x) = E_{\lambda_x}(X_0, a^*) < \infty. \]

Now for each \( x \in A^\infty \), define
\[\int p(E \cap F) \, d\mu(E) = \int p(E) \, d\mu(E), \text{for } E \in \mathbb{R}^\infty, F \text{ } T^m \text{-invariant.}\]

Let \( \delta > 0 \) be arbitrary. Since \( R(\lambda_x) \) and \( D(\lambda_x) \) are finite for \( \mu \)-almost all \( x \), then for some positive integer \( k \) we may choose disjoint \( T^m \)-invariant subsets \( A_1, A_2, \ldots, A_k \) of \( A^\infty \) of positive \( \mu \)-measure such that
\[\mu(A_1 \cup \cdots \cup A_k) > 1 - \delta/m,\]
\[|D(\lambda_x) - D(\lambda_y)| < \delta, \quad x, y \in A_i, i = 1, \ldots, k.\]

Define \( T^m \)-stationary probability measures \( \mu_1, \mu_2, \ldots, \mu_k \) on \( A^\infty \) as follows:
\[\mu_i(E) = p(E \cap A_i) / \mu(A_i), \quad E \in \mathbb{R}^\infty, i = 1, \ldots, k.\]

From d) we have that \( f \mu_\psi = \int f(A) \, d\mu(A) \) for every measurable \( f: A^\infty \rightarrow [0, \infty) \). Thus, if \( F \in \mathbb{R}^\infty \times \mathbb{R}^\infty \),
\[p(\mu_\psi(F) | x) = \int p(F | x) \, d\lambda_x(s) \, d\mu_\psi(x),\]

where \( F_x = \{y: (x, y) \in F\}. \)
that
\[ r(B) < R(\mu_x) + 4\delta \]
and
\[ \bar{\rho}_x(B) < D(\mu_x) + 4\delta, \]
where
\[ \mu_x = m^{-1} \sum_{i=0}^{m-1} \lambda \tau_x, \]
\[ R(\mu_x) = m^{-1} \sum_{i=0}^{m-1} R(\lambda \tau_x), \]
\[ D(\mu_x) = m^{-1} \sum_{i=0}^{m-1} D(\lambda \tau_x). \]

Since \( \delta \) is arbitrary, there is a set \( W^* \) of \( \mu \)-measure one and a countable set \( \mathcal{G} \) of block codes such that for any \( \eta > 0 \) and \( x \in W^* \), there exists \( B \in \mathcal{G} \) such that
\[ r(B) < D(\mu_x) + \eta \]
and
\[ \bar{\rho}_x(B) < R(\mu_x) + \eta. \]

Note that each member of \( \{\mu_x: x \in A^\infty\} \) is an ergodic source with alphabet \( A \), also:

i) \( \mu_x = \mu_{\tau_x}, x \in A^\infty \)
ii) the map \( x \to \mu_{\tau_x}(E) \) is \( A^\infty \)-measurable, \( E \in \mathcal{B} \)
iii) \( \mu(E \cap F) = \int \mu_{\tau_x}(E) \mu(x), E, F \in \mathcal{B} \), \( F \)-invariant.

By Lemma B2, there is a variable-rate sliding-block code \( C \) such that
\[ \bar{\rho}(C, \mu) < \int R(\mu_x) \mu(x) + \epsilon/2, \]
and
\[ \bar{\rho}(C, \mu) < \int D(\mu_x) \mu(x) + \epsilon/2. \]

But
\[ \int R(\mu_x) \mu(x) = \int m^{-1} \sum_{i=0}^{m-1} R(\lambda \tau_x) \mu(x) = \int R(\lambda_x) \mu(x) \]
\[ < R_\mu(D) + \epsilon/2. \]

Similarly,
\[ \int D(\mu_x) \mu(x) = \int D(\lambda_x) \mu(x) < D. \]

Thus (4.9) and (4.10) hold and the proof of Theorem 6 is complete.

**APPENDIX D**

We now assume that \( A = \hat{A} \) and that \( \rho \) is a metric on \( A \). (We take \( \mathcal{B} = \mathcal{B}^\infty \) the collection of Borel sets.) If \( \mu, \nu \) are stationary sources with alphabet \( A \), let \( \mathcal{F}(\mu, \nu) \) be the collection of all probability measures \( P \) on \( \hat{A}^\infty \times \hat{A}^\infty \) stationary with respect to \( T_x \times T_{\tau_x} \) such that the joint distribution of the \( \{X_i\} \) under \( P \) is \( \mu \) and the joint distribution of the \( \{Y_i\} \) under \( P \) is \( \nu \).

The \( \bar{\rho} \)-distance \( [5] \) between \( \mu \) and \( \nu \) is defined to be
\[ \inf \{ E_P[\rho(X_i, Y_i): P \in \mathcal{F}(\mu, \nu)] \}. \]
As is well-known, \( \bar{\rho} \) is a metric on the family of all stationary sources with alphabet \( A \), where we allow \( \infty \) as a value for the metric.

**Lemma D1:** Let \( n \) be a positive integer and let \( \alpha > 0 \) and \( D > 0 \) be given. Then for any finite subset \( B \) of \( A^n \), and any pair of stationary sources \( \mu, \nu \),
\[ \mu\left[ \rho_x(\bar{X}^n | B) > D + \delta \right] < \alpha^{-1} \bar{\rho}(\mu, \nu) + \nu\left[ \rho_x(\bar{X}^n | B) > D + \delta \right]; \]
\[ \int \left[ \rho_x(\bar{X}^n | B) > D + \delta \right] \rho_x(\bar{X}^n | B) d\mu < \bar{\rho}(\mu, \nu) (\alpha^{-1} D + 2) + \int \left[ \rho_x(\bar{X}^n | B) > D + \delta \right] \rho_x(\bar{X}^n | B) d\nu + \alpha. \]

**Proof:** Fix \( P \in \mathcal{F}(\mu, \nu) \). We have
\[ \rho_x(\bar{X}^n | B) < \rho_x(\bar{Y}^n | B) + \rho_x(\bar{X}^n, \bar{Y}^n). \]

This implies
\[ P\left[ \rho_x(\bar{X}^n | B) > D + \alpha \right] < P\left[ \rho_x(\bar{Y}^n | B) > D + \alpha \right] + P\left[ \rho_x(\bar{X}^n, \bar{Y}^n) > \alpha \right], \]
because the set on the left side is contained in the union of the two events on the right side. By Chebyshev's inequality,
\[ P\left[ \rho_x(\bar{Y}^n | B) > D + \alpha \right] < \alpha^{-1} P\left[ \rho_x(\bar{Y}^n, \bar{Y}^n) = \alpha^{-1} P\left[ \rho(X_0, Y_0) \right] \right] \]
but
\[ r(B) < D + 2\alpha \]
and
\[ \bar{\rho}(\mu, \nu) < \alpha^{-1} \rho_x(\bar{X}^n, \bar{Y}^n). \]

Taking the infimum over \( \mathcal{F}(\mu, \nu) \), we get (D1).

Again fix \( P \in \mathcal{F}(\mu, \nu) \). By (D3),
\[ \rho_x(\bar{X}^n | B) \int \rho_x(\bar{X}^n, \bar{Y}^n) > D + \alpha \]
\[ + P\left[ \rho_x(\bar{Y}^n | B) > D + \alpha \right] + P\left[ \rho_x(\bar{X}^n, \bar{Y}^n) > \alpha \right]. \]

Taking the expected value,
\[ \int \rho_x(\bar{X}^n | B) \left[ \rho_x(\bar{X}^n, \bar{Y}^n) > \alpha \right] + \rho_x(\bar{Y}^n | B) > D + \alpha \]
\[ = \alpha + P\left[ \rho_x(\bar{X}^n, \bar{Y}^n) > D + \alpha \right] \]
\[ + P\left[ \rho_x(\bar{Y}^n | B) > D + \alpha \right] \rho_x(\bar{Y}^n | B) d\nu. \]

Applying Chebyshev's inequality as before and taking the infimum over \( \mathcal{F}(\mu, \nu) \), we get (D2).

**Proof of Theorem 7:** Fix \( \epsilon > 0 \). Let \( \delta > 0 \) be so small that
\[ 6\delta + (M + 2\delta)(2\delta) < \epsilon, \]
where \( M = \sup_{p \in \mathcal{F}(\mu, \nu)} R_p(D) \). By Lemma D1, choose \( \beta > 0 \) so small that \( \bar{\rho}(\mu, \nu) < \beta \) implies for every \( n \) and every finite subset \( B \) of \( A^n \) that
\[ \mu\left[ \rho_x(\bar{X}^n | B) > D + \delta \right] < 6\delta + \nu\left[ \rho_x(\bar{X}^n | B) > D + \delta \right]; \]
\[ \int \left[ \rho_x(\bar{X}^n | B) > D + \delta \right] \rho_x(\bar{X}^n | B) d\mu < 2\delta \]
\[ + \int \left[ \rho_x(\bar{X}^n | B) > D + \delta \right] \rho_x(\bar{X}^n | B) d\nu. \]

Partition \( \mathcal{B} \) into sets \( E_1, E_2, \ldots, E_k \) such that
\[ |R_p(D) - R_p(D)| < \delta, \]
\[ \bar{\rho}(\mu, \nu) < \beta, \quad \mu, \nu \in \mathcal{B}_i, \quad i = 1, \ldots, k. \]

Choose \( \mu_i \in \mathcal{B}_i, i = 1, \ldots, k. \)

By [4] we may choose block codes \( B_1, B_2, \ldots, B_k \subset A^n \) where \( n \) is so large that
\[ 2n^{-1} + n^{-1} |\log k| < \delta \]
and for \( i = 1, \ldots, k, \)
\[ \mu\left[ \rho_x(\bar{X}^n | B_i) > D + \delta \right] < \delta; \]
\[ r(B_i) < R_p(D) + \delta; \]
\[ \int \left[ \rho_x(\bar{X}^n | B_i) > D + \delta \right] \rho_x(\bar{X}^n | B_i) d\mu_i < \delta. \]

Thus
\[ \int \left[ \rho_x(\bar{X}^n | B_i) > D + \delta \right] \rho_x(\bar{X}^n | B_i) d\nu_i. \]
We now define a certain variable-rate block code $\mathcal{C}_n = (\hat{\mathcal{A}}, \hat{\mathcal{A}}, n, \sigma)$. $\mathcal{A}$ is any finite subset of $\mathcal{A}$ such that $\mathcal{A}^n \supset B_1 \cup \cdots \cup B_k$. $\sigma: \mathcal{E}^n \rightarrow \{1, 2, \cdots\}$ is a length function such that $\sigma \leq |\log |B_i|| + |\log k| + 1$ on $B_i$, for $i = 1, \cdots, k$. (D10)

For each $i = 1, \cdots, k$, define $\phi_i: \mathcal{A}^n \rightarrow B_i$ so that

$$p_{\phi_i}(x, \phi_i(x)) = p_{\sigma}(x | B_i).$$

Define $\phi: \mathcal{A}^n \rightarrow \hat{\mathcal{A}}$ so that:

a) if $\min p_{\phi_i}(x, \phi_i(x)) > D + 2\delta$, then $\phi(x) = \phi_i(x)$, where $i'$ is the least $i$ in the range $1 < i < k$ such that

$$p_{\phi_i}(x, \phi_i(x)) \geq p_{\sigma}(x, \phi_i(x)).$$

b) otherwise $\phi(x) = \phi_i(x)$, where $i'$ is the least $i$ in the range $1 < i < k$ such that

$$p_{\phi_i}(x, \phi_i(x)) < D + 2\delta$$

and $r(B_j) = \min \{ r(B_j) : p_{\phi_i}(x, \phi_i(x)) < D + 2\delta, 1 < j < k \}$.

Then $\hat{\mathcal{A}}^n \rightarrow \hat{\mathcal{A}}^n$ is the block code of order $n$ such that $\phi_i'(x)^n = \phi_i'(x^n)$. Note the following.

If $p_{\phi_i}(x, \phi_i(x)) > D + 2\delta$, then $p_{\phi_i}(x, \phi_i(x)) = \min p_{\phi_i}(x, \phi_i(x))$. (D11)

If $p_{\phi_i}(x, \phi_i(x)) < D + 2\delta$, then $p_{\phi_i}(x, \phi_i(x)) < \sigma(x)$. (D12)

We have by (D10) and the choice of $n$ that

$$n^{-1} \max \sigma < M + 2\delta.$$ (D13)

If $\mu \in \mathcal{E}_n$, then by (D11), (D5), and (D9)

$$\int_{\{\mu(\hat{\mathcal{A}}^n, \phi(\hat{\mathcal{A}}^n)) > D + 2\delta \}} p_{\phi}(\hat{\mathcal{A}}^n, \phi(\hat{\mathcal{A}}^n)) \, d\mu
\leq \int_{\{\rho(\hat{\mathcal{A}}^n) > D + 2\delta \}} p_{\phi}(\hat{\mathcal{A}}^n, \phi(\hat{\mathcal{A}}^n)) \, d\mu
< 2\delta + \int_{\{\rho(\hat{\mathcal{A}}^n) > D + \delta \}} p_{\phi}(\hat{\mathcal{A}}^n, \phi(\hat{\mathcal{A}}^n)) \, d\mu_i
< \delta.$$ (D14)

If $\mu \in \mathcal{E}_n$, by (D12), (D13), (D8), (D11), and (D4),

$$\mathcal{R}(\mathcal{C}_n, \mu) = \int_{\{\rho(\hat{\mathcal{A}}^n, \phi(\hat{\mathcal{A}}^n)) < D + 2\delta \}} n^{-1} \rho(\hat{\mathcal{A}}^n, \phi(\hat{\mathcal{A}}^n)) \, d\mu
+ \int_{\{\rho(\hat{\mathcal{A}}^n, \phi(\hat{\mathcal{A}}^n)) > D + 2\delta \}} n^{-1} \rho(\hat{\mathcal{A}}^n, \phi(\hat{\mathcal{A}}^n)) \, d\mu
< n^{-1} \left( \left[ \log |B_i| \right] + \left[ \log k \right] + 1 \right)
+ (M + 2\delta) \mu \left( \rho(\hat{\mathcal{A}}^n, \phi(\hat{\mathcal{A}}^n)) > D + 2\delta \right)
< R_{\phi}(D) + 2\delta + (M + 2\delta) \left( \delta + \mu \left( \rho(\hat{\mathcal{A}}^n | B_i) > D + \delta \right) \right)
< R_{\phi}(D) + 3\delta + (M + 2\delta)(2\delta).$$

Thus for every $\mu \in \mathcal{E}_n$,

$$\hat{\mathcal{R}}(\mathcal{C}_n, \mu) < D + 5\delta,$$

and

$$\hat{\mathcal{R}}(\mathcal{C}_n, \mu) < R_{\phi}(D) + 3\delta + (M + 2\delta)(2\delta).$$

Applying Theorem 2, we can get a variable-rate sliding-block code $\mathcal{C}_n$ such that for every $\mu \in \mathcal{E}_n$,

$$\hat{\mathcal{R}}(\mathcal{C}_n, \mu) < R_{\phi}(D) + \epsilon,$$

$$\hat{\mathcal{P}}(\mathcal{C}_n, \mu) < D + \epsilon.$$

### References


