The Annals of Probability
1980, Vol. 8, No. 5, 942–961

ON THE TRANSMISSION OF BERNOULLI SOURCES OVER STATIONARY CHANNELS[1]

JOHN C. KIEFFER

University of Missouri-Rolla

For a discrete-time finite-alphabet stationary channel ν satisfying a weak continuity requirement, it is shown that there are capacities $C_s(\nu)$ and $C_b(\nu)$ which have the following operational significance. A Bernoulli source μ is transmissible over ν via sliding-block coding if and only if the entropy $H(\mu)$ of μ is no greater than $C_s(\nu)$; μ is transmissible via block coding if and only if $H(\mu)$ is no greater than $C_b(\nu)$. The weak continuity requirement is satisfied for the $\bar{d}$-continuous channels of Gray-Ornstein as well as other channels. An example of a channel is given to show that the case $C_s(\nu) \neq C_b(\nu)$ can occur.

1. Introduction. In this introductory section we present some notation, definitions, and then the main result to be proved.

If a letter S denotes a finite set, the corresponding script letter $\mathcal{S}$ will denote the set of all subsets of S. We let $(S^\infty, \mathcal{S}^\infty)$ denote the measurable space consisting of S^∞, the set of all doubly-infinite sequences $x = (x_i)_{i=-\infty}^{\infty}$ from S, and $\mathcal{S}^\infty$, the usual product σ-field of subsets of S^∞. The shift transformation from $S^\infty \to S^\infty$ is denoted by T_S; we have

$$(T_S x)_i = x_{i+1}, \qquad x \in S^\infty, i \in Z,$$

where Z denotes the set of integers. (T_S will be denoted by T whenever S is understood.) If x is a finite or infinite sequence of elements from S, x_i will denote the ith coordinate of x, and for $m \leqslant n$, x_m^n will denote $(x_m, x_{m+1}, \cdots, x_n)$. If μ is a measure on $\mathcal{S}^\infty$, and N is a positive integer, μ^N will denote the measure on S^N such that

$$\mu^N(\mathbf{x}) = \mu\{x \in S^\infty : x_1^N = \mathbf{x}\}, \qquad \mathbf{x} \in S^N.$$

A set $E \in \mathcal{S}^\infty$ is a finite-dimensional cylinder set if for some $n = 1, 2, \cdots,$ and $k \in Z$, there exists $E' \subset S^n$ such that

$$E = \{x \in S^\infty : x_k^{k+n-1} \in E'\}.$$

If S is a finite set, $|S|$ denotes the cardinality of S. All logarithms in the paper are to base 2. If $(\Omega, \mathcal{F})$ is a measurable space and E an event in $\mathcal{F}$, E^c denotes the complementary event

$$E^c = \{\omega \in \Omega : \omega \notin E\}.$$

Received October 16, 1978; revised May 16, 1979.

[1]Partially supported by NSF Grant 76-02276-A01.

AMS 1970 *subject classifications.* Primary 94A15.

Key words and phrases. Stationary channels, sliding-block codes, block codes, Bernoulli sources.

If V_i is a function from $\Omega_i \to S_i, i = 1, \cdots, n$, $V_1 \times \cdots \times V_n$ denotes the function from $\Omega_1 \times \cdots \times \Omega_n \to S_1 \times \cdots \times S_n$ such that

$$(V_1 \times \cdots \times V_n)(\omega_1, \cdots, \omega_n) = (V_1(\omega_1), \cdots, V_n(\omega_n)),$$

$(\omega_1, \cdots, \omega_n) \in \Omega_1 \times \cdots \times \Omega_n$. If V_i is a function from $\Omega \to S_i, i = m, m + 1, \cdots, n$, V_m^n denotes the function from $\Omega \to S_m \times \cdots \times S_n$ such that

$$V_m^n(\omega) = (V_m(\omega), \cdots, V_n(\omega)), \qquad \omega \in \Omega.$$

If V is a function from Ω to S^N or S^∞, V_i denotes the map from Ω to S such that

$$V_i(\omega) = V(\omega)_i, \qquad \omega \in \Omega.$$

If V is a measurable function from a probability space Ω to a measurable space $(F, \mathfrak{F})$, P^V will denote the probability measure on $\mathfrak{F}$ such that

$$P^V(E) = \text{Prob}[V \in E], \qquad E \in \mathfrak{F}.$$

A measurable map X from a measurable space Ω_1 to a measurable space Ω_2 is called a process if $\Omega_2 = A^\infty$ for some finite set A. We will call a process from $\Omega \to A^\infty$ an A-valued process defined on Ω.

If A is a finite set, we place on A the discrete topology and on A^∞ the resulting product topology. Then if B, C are finite sets we endow $B^\infty \times C^\infty$ with the product topology arising from the topologies on B^∞, C^∞. For each $N = 1, 2, \cdots$, let $\mathcal{P}_N(A)$ denote the set of all probability measures on $\mathcal{Q}^\infty$ stationary with respect to T_A^N, and let $\mathcal{P}_N(B, C)$ denote the set of all probability measures on $\mathcal{B}^\infty \times \mathcal{C}^\infty$ stationary with respect to $T_B^N \times T_C^N$. We assign the weak topology to each of the spaces $\mathcal{P}_N(A), \mathcal{P}_N(B, C)$. (The weak topology on $\mathcal{P}_N(A)$ is the weakest topology such that for each continuous $f: A^\infty \to (-\infty, \infty)$, the map $\mu \to \int_{A^\infty} f\, d\mu$ is continuous on $\mathcal{P}_N(A)$. In a similar way one defines the weak topology on $\mathcal{P}_N(B, C)$.)

CODES. A code ϕ is a measurable map from $A^\infty \to B^\infty$, for certain finite sets A, B. The code ϕ is *stationary* if $\phi \cdot T_A = T_B \cdot \phi$. A stationary code $\phi: A^\infty \to B^\infty$ is a *sliding-block code* if for some positive integer M there exists a map $\phi_M: A^{2M+1} \to B$ such that

$$\phi(x)_i = \phi_M(x_{i-M}^{i+M}), \qquad x \in A^\infty, i \in Z.$$

A code $\phi: A^\infty \to B^\infty$ is a *block code* of order N if there exists $\phi_N: A^N \to B^N$ such that

$$\phi(x)_{iN+1}^{iN+N} = \phi_N(x_{iN+1}^{iN+N}), \qquad x \in A^\infty, i \in Z.$$

SOURCES. A *source* is a pair $[A, \mu]$ where A is a finite set, called the source alphabet, and μ is a probability measure on $\mathcal{Q}^\infty$. If $\mu \in \mathcal{P}_1(A), [A, \mu]$ is a *stationary source*. If $\mu \in \mathcal{P}_N(A), [A, \mu]$ is an *N-stationary* source. If $[A, \mu]$ is stationary and μ is ergodic with respect to T_A, we call $[A, \mu]$ an *ergodic source*. We call $[A, \mu]$ *aperiodic* if $\mu\{x\} = 0$, for all $x \in A^\infty$. If $[A, \mu]$ is a stationary source, $H(\mu)$ will denote the entropy of the source. If $[A, \mu]$ is a source and $\phi : A^\infty \to B^\infty$ is a code, $[B, \mu^\phi]$ denotes the source where

$$\mu^\phi(E) = \mu(\phi \in E), \qquad E \in \mathcal{B}^\infty.$$

We call a stationary source $[A_1, \mu_1]$ a *factor* of the stationary source $[A_2, \mu_2]$ if there is a stationary code $\phi : A_2^\infty \to A_1^\infty$ such that $\mu_1 = \mu_2^\phi$. We define a stationary source $[A, \mu]$ to be *memoryless* if μ is a product measure. We define a stationary source $[A, \mu]$ to be a *Bernoulli source* (which we abbreviate as B-*source*) if it is a factor of some memoryless source. The B-sources form a wide class of sources. For example, any stationary source $[A, \mu]$ where μ is mixing and N-Markovian relative to T_A is a B-source [11]. Thus the set of μ on $\mathcal{A}^\infty$ for which $[A, \mu]$ is a B-source is a dense subset of $\mathcal{P}_1(A)$ (with respect to weak topology). There are various characterizations of B-sources known; $[A, \mu]$ is a B-source if and only if T_A is a finitely determined transformation on $(A^\infty, \mathcal{A}^\infty, \mu)$ [11], or if and only if T_A is very weakly Bernoulli [11], or if and only if μ is almost block independent [13].

CHANNELS. A *channel* is a triple $[B, C, \nu]$ where B, C are finite sets and $\nu = \{\nu_x : x \in B^\infty\}$ is a family of probability measures on $\mathcal{C}^\infty$ such that for each $E \in \mathcal{C}^\infty$, the map $x \to \nu_x(E)$ is a measurable map from $B^\infty \to [0, 1]$. B is called the input alphabet of the channel, and C is called the output alphabet. The channel $[B, C, \nu]$ is said to be *stationary* if

$$\nu_{Tx}(TE) = \nu_x(E), \qquad x \in B^\infty, E \in \mathcal{C}^\infty.$$

If $[B, \mu]$ is a source, and $[B, C, \nu]$ is a channel, $\mu\nu$ denotes the probability measure on $\mathcal{B}^\infty \times \mathcal{C}^\infty$ such that

$$\mu\nu(E \times F) = \int_E \nu_x(F)\, d\mu(x), \qquad E \in \mathcal{B}^\infty, F \in \mathcal{C}^\infty.$$

Note that if $[B, C, \nu]$ is stationary and $[B, \mu]$ is N-stationary, then $\mu\nu \in \mathcal{P}_N(B, C)$. We define a stationary channel $[B, C, \nu]$ to be *weakly continuous* if the map $\mu \to \mu\nu$ from $\mathcal{P}_1(B) \to \mathcal{P}_1(B, C)$ is continuous. We define a stationary channel $[B, C, \nu]$ to be *totally weakly continuous* if for every $N = 1, 2, \cdots$, the map $\mu \to \mu\nu$ from $\mathcal{P}_N(B) \to \mathcal{P}_N(B, C)$ is continuous. Some examples of channels which are totally weakly continuous are:

(a) Stationary channels $[B, C, \nu]$ for which the map $x \to \nu_x(E)$ from $B^\infty \to [0, 1]$ is continuous, for each finite-dimensional cylinder set $E \in \mathcal{C}^\infty$.

(b) $\bar{d}$-continuous stationary channels (as defined in [2]).

We omit the easy proof that the channels in (a) are totally weakly continuous. We prove that the channels in (b) are totally weakly continuous in the Appendix. The $\bar{d}$-continuous channels are the most general class of stationary channels for which coding theorems of information theory have been proved [2, 3].

THE SOURCE-CHANNEL HOOKUP. We define a sequence U, X, Y, V of processes defined on some probability space to be a *hookup* of the source $[A, \mu]$ to the channel $[B, C, \nu]$ if:

(a) U is A-valued, X is B-valued, Y is C-valued, V is A-valued;

(b) U, X, Y, V form a Markov chain;
(c) $P^U = \mu$;
(d) $P^X \nu = P^{(X,Y)}$.

Intuitively, the process U is a model for the source $[A, \mu]$. If $u \in A^\infty$ is the value observed for U, some random experiment $\mathfrak{S}_u$ is performed the outcome of which is a value x of X. The sequence $x \in B^\infty$ is then transmitted over the channel yielding a random $y \in C^\infty$ (the value of Y) according to the distribution ν_x. A random experiment $\mathfrak{S}_y$ is then performed, the outcome of which is v, the value of V. The hookups of most interest are those for which the value x is determined uniquely once u is known, and v is determined uniquely once y is known. That is, there are codes $\phi_1 : A^\infty \to B^\infty$ and $\phi_2 : C^\infty \to A^\infty$ such that

$$x = \phi_1(u), \qquad v = \phi_2(y).$$

These codes ϕ_1, ϕ_2 are called the encoder and decoder, respectively.

SLIDING-BLOCK TRANSMISSIBILITY. We say a stationary source $[A, \mu]$ is *sliding-block transmissible* over a stationary channel $[B, C, \nu]$ if for each $\varepsilon > 0$, a hookup U, X, Y, V of $[A, \mu]$ to $[B, C, \nu]$ may be found such that

(a) there is a sliding-block code $\phi_1 : A^\infty \to B^\infty$ such that $X = \phi_1(U)$;
(b) there is a sliding-block code $\phi_2 : C^\infty \to A^\infty$ such that $V = \phi_2(y)$;
(c) $\text{Prob}[U_0 \neq V_0] < \varepsilon$.

We define the sliding-block capacity $C_s(\nu)$ of the stationary channel $[B, C, \nu]$ as follows:

$$C_s(\nu) = \sup\{H(\mu) : [A, \mu] \text{ ergodic and sliding-block transmissible}\}.$$

We state the first of the two main theorems of this paper.

THEOREM 1. *Let $[B, C, \nu]$ be a weakly continuous stationary channel. Then, a* B-*source $[A, \mu]$ is sliding-block transmissible over $[B, C, \nu]$ if and only if $H(\mu) \leqslant C_s(\nu)$.*

For special types of $\bar{d}$-continuous channels, sliding-block coding theorems had been obtained for the class of all aperiodic ergodic sources [3], [6]. By restricting ourselves to the subclass of B-sources, we are able to dispense with the $\bar{d}$-continuity requirement on the channel.

BLOCK TRANSMISSIBILITY. Let $[A, \mu]$ be a stationary source and $[B, C, \nu]$ a stationary channel.

We say $[A, \mu]$ is *block transmissible* over $[B, C, \nu]$ if for each $\varepsilon > 0$, there exists a positive integer N and a hookup U, X, Y, V for which:

(i) $X = \phi_1(U)$ for some block code $\phi_1 : A^\infty \to B^\infty$ of order N;
(ii) $V = \phi_2(Y)$ for some block code $\phi_2 : C^\infty \to A^\infty$ of order N;
(iii) $N^{-1}\sum_{i=1}^{N} \text{Prob}[U_i \neq V_i] < \varepsilon$.

We define the block capacity $C_b(\nu)$ of $[B, C, \nu]$ as follows:

$$C_b(\nu) = \sup\{H(\mu) : [A, \mu] \text{ ergodic and block transmissible}\}.$$

The second of our two main theorems is now stated.

THEOREM 2. *Let $[B, C, \nu]$ be a totally weakly continuous stationary channel. Then a* B-*source $[A, \mu]$ is block transmissible over $[B, C, \nu]$ if and only if $H(\mu) \leqslant C_b(\nu)$.*

For a stationary $\bar{d}$-continuous channel, a coding theorem had been obtained for the class of all ergodic sources [2]. The preceding theorem implies that the assumption of $\bar{d}$-continuity is not necessary if one is only interested in transmitting B-sources.

We now point out a stronger notion of block transmissibility used by many information theorists. We say $[A, \mu]$ is *strongly block transmissible* over $[B, C, \nu]$ if, for each $\varepsilon > 0$, there exists N and a hookup U, X, Y, V for which (i), (ii) above hold and in place (iii) we have

(iii′) $\text{Prob}[U_1^N \neq V_1^N] < \varepsilon$.

The corresponding capacity $C_s'(\nu)$ is given by:

$$C_s'(\nu) = \sup\{H(\mu) : [A, \mu] \text{ ergodic and strongly block transmissible}\}.$$

For $\bar{d}$-continuous channels, Theorem 11 of Section 2 will show that one obtains the same capacity and the same coding theorem no matter which of the two types of transmissibility is used. (There can exist a source which is block transmissible and not strongly block transmissible [8], but the only way this can happen is if the entropy of the source equals the capacity of the channel.) For non-$\bar{d}$-continuous channels, it is an open problem whether the two types of block transmissibility are equivalent in the sense just described. Of the two types, we emphasize block transmissibility, for the simple reason that the techniques we use do not appear to apply to strong block transmissibility.

In Section 2, we prove some sliding-block and block coding theorems, including Theorems 1 and 2. In Section 3, we present an example of a channel for which $C_s(\nu) = C_b(\nu)$ fails. In Section 4, we present some open problems.

2. Block and sliding-block coding theorems. If defined on some probability space we have for each $i = 1, \cdots, n$, an A_i-valued process V_i, we say $(V_1, \cdots, V_n)$ is jointly stationary if $P^{(V_1, \cdots, V_n)}$ is $T_{A_1} \times \cdots \times T_{A_n}$ stationary. We say $(V_1, \cdots, V_n)$ is jointly N-stationary if $P^{(V_1, \cdots, V_n)}$ is $T_{A_1}^N \times \cdots \times T_{A_n}^N$ stationary.

If $[B, C, \nu]$ is a stationary channel, let $\mathfrak{M}_s(\nu)$ denote the set of all stationary sources sliding-block transmissible over $[B, C, \nu]$, and let $\mathfrak{M}_b(\nu)$ denote the set of all stationary sources block-transmissible over $[B, C, \nu]$.

THEOREM 3. *Let $[B, C, \nu]$ be a weakly continuous stationary channel. Let $[A, \mu]$ be a stationary aperiodic source. Then $[A, \mu] \in \mathfrak{M}_s(\nu)$ if for each $\varepsilon > 0$, there is a hookup U, X, Y, V such that*:

(a) *(U, X, Y, V) is jointly stationary*;

(b) $\text{Prob}[U_0 \neq V_0] < \varepsilon$.

REMARK. What this theorem says is that in the definition of sliding-block transmissibility, it makes no difference whether the codings $U \to X$ and $Y \to V$ are deterministic or random.

PROOF. We prove a little more. Let U, X, Y, V be a jointly-stationary hookup such that $E\rho(U_0, V_0) < \varepsilon$ where $\rho: A \times A \to [0, \infty)$ is a metric. We will show there exists a hookup $\tilde{U}, \tilde{X}, \tilde{Y}, \tilde{V}$ such that $\tilde{X}$ is a sliding-block coding of $\tilde{U}$, $\tilde{V}$ is a sliding-block coding of $\tilde{Y}$, and $E\rho(\tilde{U}_0, \tilde{V}_0) < 4\varepsilon$.

First, by Lemma 3 of the Appendix there is a sliding-block code $\phi_2: C^\infty \to A^\infty$ such that $E\rho(U_0, \phi_2(Y)_0) < \varepsilon$. Let $\psi: A \times B^\infty \to [0, \infty)$ be the map such that

$$\psi(u, x) = \Sigma_{v \in A} \nu_x\{y \in C^\infty : \phi_2(Y)_0 = v\}\rho(u, v), \quad u \in A, x \in B^\infty. \tag{2.1}$$

Then for any hookup U', X', Y', V' with $V' = \phi_2(Y')$, we have

$$E\rho(U_0', V_0') = \int_{A^\infty \times B^\infty} \psi(u_0, x)\, dP^{(U', X')}(u, x).$$

We also observe the following property of ψ:

$$\psi(u_1, x) \leqslant \psi(u_2, x) + \rho(u_1, u_2), \quad u_1, u_2 \in A, x \in B^\infty. \tag{2.2}$$

Again, using Lemma 3 of the Appendix, choose a sliding-block code $g: B^\infty \to A^\infty$ such that $E\rho(U_0, g(X)_0) < \varepsilon$. Since ν is weakly continuous it follows that the map

$$\lambda \to \int_{B^\infty} \psi(g(x)_0, x)\, d\lambda(x)$$

is a continuous map from $\mathcal{P}_1(B) \to [0, 1]$. Thus we may find $\delta > 0$ and a positive integer N such that if $\lambda_1, \lambda_2 \in \mathcal{P}_1(B)$ and

$$\Sigma_{x \in B^N} |\lambda_1^N(x) - \lambda_2^N(x)| < \delta$$

then

$$\left|\int_{B^\infty} \psi(g(x)_0, x)\, d\lambda_1(x) - \int_{B^\infty} \psi(g(x)_0, x)\, d\lambda_2(x)\right| < \varepsilon.$$

By Lemma 6 of the Appendix there exists a sliding-block code $\phi_1: A^\infty \to B^\infty$ such that

$$E\rho(U_0, g(\phi_1(U))_0) < \varepsilon$$

and

$$\Sigma_{x \in B^N} \left|P^{\phi_1(U)_1^N}(x) - P^{X_1^N}(x)\right| < \delta.$$

Consider a hookup $\tilde{U}, \tilde{X}, \tilde{Y}, \tilde{V}$ where $\tilde{X} = \phi_1(\tilde{U})$, $\tilde{V} = \phi_2(\tilde{Y})$. We have

$$\begin{aligned}
E\rho(\tilde{U}_0, \tilde{V}_0) &= \int_{A^\infty \times B^\infty} \psi(u_0, x)\, dP^{(\tilde{U}, \tilde{X})}(u, x) \\
&= \int_{A^\infty} \psi(u_0, \phi_1(u))\, d\mu(u) \leqslant \int_{A^\infty} \psi(g(\phi_1(u))_0, \phi_1(u))\, d\mu(u) \\
&\quad + \int_{A^\infty} \rho(g(\phi_1(u))_0, u_0)\, d\mu(u) \leqslant \int_{A^\infty} \psi(g(x)_0, x)\, dP^{\phi_1(U)}(x) + \varepsilon \\
&\leqslant \int_{A^\infty} \psi(g(x)_0, x)\, dP^X(x) + 2\varepsilon = \int_{A^\infty \times B^\infty} \psi(g(x)_0, x)\, dP^{(U,X)}(u, x) + 2\varepsilon \\
&\leqslant \int_{A^\infty \times B^\infty} \psi(u_0, x)\, dP^{(U,X)}(u, x) + \int_{A^\infty \times B^\infty} \rho(u_0, g(x)_0)\, dP^{(U,X)}(u, x) + 2\varepsilon \\
&< 4\varepsilon.
\end{aligned}$$

THEOREM 4. *Let $[B, C, \nu]$ be a stationary weakly-continuous channel. Then $\mathfrak{M}_s(\nu) \subset \mathfrak{M}_b(\nu)$.*

PROOF. Let $\rho : A \times A \to [0, \infty)$ be a metric. Let $[A, \mu]$ be a stationary source. Let $\phi_1 : A^\infty \to B^\infty$ and $\phi_2 : C^\infty \to A^\infty$ be sliding-block codes and U, X, Y, V a hookup of $[A, \mu]$ to $[B, C, \nu]$ such that

(a) $X = \phi_1(U)$, $V = \phi_2(Y)$;

(b) $E\rho(U_0, V_0) < \varepsilon$.

We will construct for some N block codes $\hat{\phi}_1 : A^\infty \to B^\infty$ and $\hat{\phi}_2 : C^\infty \to A^\infty$ of order N and a hookup $\hat{U}, \hat{X}, \hat{Y}, \hat{V}$ such that

(c) $\hat{X} = \hat{\phi}_1(\hat{U})$, $\hat{V} = \hat{\phi}_2(\hat{Y})$;

(d) $E[N^{-1}\Sigma_{i=1}^N \rho(\hat{U}_i, \hat{V}_i)] < 5\varepsilon$.

For each $N = 1, 2, \cdots$, let $\rho_N : A^N \times A^N \to [0, \infty)$ be the map such that

$$\rho_N(x,y) = N^{-1}\Sigma_{i=1}^N \rho(x_i, y_i), \qquad x, y \in A^N.$$

Let $\psi : A \times B^\infty \to [0, \infty)$ be the function defined in (2.1). Then for any hookup $\tilde{U}, \tilde{X}, \tilde{Y}, \tilde{V}$ with $\tilde{V} = \phi_2(\tilde{Y})$,

$$E\rho_{K-L+1}(\tilde{U}_L^K, \tilde{V}_L^K) = \int_{A^\infty \times B^\infty}\Big[(K-L+1)^{-1}\Sigma_{i=L}^K \psi(u_i, T^i x)\Big]\, dP^{(\tilde{U}, \tilde{X})}(u, x), \qquad L \leqslant K.$$

Pick a positive integer M such that there exist maps $\phi_1' : A^{2M+1} \to B$ and $\phi_2' : C^{2M+1} \to A$ such that

$$\phi_1(u)_i = \phi_1'(u_{i-M}^{i+M}), \quad \phi_2(y) = \phi_2'(y_{i-M}^{i+M}), \qquad u \in A^\infty, y \in C^\infty, i \in Z.$$

For each $N > 2M + 1$, let $\phi_1^{(N)} : A^\infty \to B^\infty$ and $\phi_2^{(N)} : C^\infty \to A^\infty$ be block codes of order N such that:

$$\phi_1^{(N)}(u)_i = \phi_1(u)_i, \qquad M + 1 \leqslant i \leqslant N - M, u \in A^\infty$$

$$\phi_2^{(N)}(y)_i = \phi_2(y)_i, \qquad M + 1 \leqslant i \leqslant N - M, y \in C^\infty.$$

It is easy to see that

$$N^{-1}\Sigma_{i=1}^N P^{\phi_1^{(N)}(U)} \cdot T^{-i} \to \mu^\phi$$

in the topology on $\mathcal{P}_1(B)$.

Now by Lemma 3 of the Appendix pick a sliding-block coder $g : B^\infty \to A^\infty$ such that $E\rho(U_0, g(X)_0) < \varepsilon$. We observe, using (2.2), that

$$\begin{aligned}\int_{B^\infty}\psi(g(x)_0, x)\, dP^X(x) &= \int_{A^\infty}\psi(g(\phi_1(u))_0, \phi_1(u))\, d\mu(u) \\ &\leqslant \int_{A^\infty}\psi(u_0, \phi_1(u))\, d\mu(u) + \int_{A^\infty}\rho(u_0, g(\phi_1(u)_0))\, d\mu(u) \\ &< 2\varepsilon.\end{aligned}$$

Since the map $\lambda \to \int\psi(g(x)_0, x)\, d\lambda(x)$ is a continuous map from $\mathcal{P}_1(B) \to [0, \infty)$, we may choose N so large that

(e) $\int\psi(g(x)_0, x)\, d\big(N^{-1}\Sigma_{i=1}^N P^{\phi_1^{(N)}(U)} \cdot T^{-i}\big)(x) < 2\varepsilon$;

(f) $\sup_{u \in A^\infty}\rho_N\big(g(\phi_1^{(N)}(u))_1^N, g(\phi_1(u))_1^N\big) < \varepsilon$;

(g) $\sup_{y \in C^\infty}\rho_N\big(\phi_2^{(N)}(y)_1^N, \phi_2(y)_1^N\big) < \varepsilon$.

Let $\tilde{U}, \tilde{X}, \tilde{Y}, \tilde{V}$ be a hookup such that $\tilde{X} = \phi_1^{(N)}(\tilde{U})$, $\tilde{V} = \phi_2(\tilde{Y})$. Then, from (e), (f),

$$
\begin{aligned}
E\rho_N(\tilde{U}_1^N, \tilde{V}_1^N) &= \int_{A^\infty \times B^\infty} N^{-1}\sum_{i=1}^N \psi(u_i, T^i x)\, dP^{(\tilde{U},\tilde{X})}(u,x) \\
&\leq \int_{A^\infty \times B^\infty} N^{-1}\sum_{i=1}^N \psi\big(g(\phi_1^{(N)}(u))_i, T^i x\big)\, dP^{(\tilde{U},\tilde{X})}(u,x) \\
&\quad + \int_{A^\infty} \rho_N\big(g(\phi_1^{(N)}(u))_1^N, g(\phi_1(u))_1^N\big)\, d\mu(u) + \int_{A^\infty} \rho_N\big(u_1^N, g(\phi_1(u))_1^N\big)\, d\mu(u) \\
&< \int_{B^\infty} N^{-1}\sum_{i=1}^N \psi(g(x)_i, T^i x)\, dP^{\tilde{X}}(x) + 2\varepsilon \\
&= \int_{B^\infty} \psi(g(x)_0, x)\, d\Big(N^{-1}\sum_{i=1}^N P^{\tilde{X}}\cdot T^{-i}\Big)(x) + 2\varepsilon < 4\varepsilon.
\end{aligned}
$$

Letting $\hat{U}, \hat{X}, \hat{Y}, \hat{V}$ be the processes such that $(\hat{U}, \hat{X}, \hat{Y}) = (\tilde{U}, \tilde{X}, \tilde{Y})$ and $\hat{V} = \phi_2^{(N)}(\tilde{Y})$, we get from (g)

$$E\rho_N(\hat{U}_1^N, \hat{V}_1^N) \leq E\rho_N(\tilde{U}_1^N, \tilde{V}_1^N) + E\rho_N(\tilde{V}_1^N, \hat{V}_1^N) < 5\varepsilon.$$

THEOREM 5. *Let $[B, C, \nu]$ be a totally weakly continuous stationary channel. Let $[A, \mu]$ be a stationary aperiodic source. Then $[A, \mu] \in \mathfrak{M}_b(\nu)$ if for each $\varepsilon > 0$, there exists a hookup U, X, Y, V and a positive integer N such that*

(a) *(U, X, Y, V) is jointly N-stationary;*

(b) $N^{-1}\sum_{i=1}^N \mathrm{Prob}[U_i \neq V_i] < \varepsilon$.

PROOF. If S is a finite set and N a positive integer, let $\alpha_S^{(N)}$ denote the isomorphism of the measurable space $((S^N)^\infty, (\mathcal{S}^N)^\infty)$ onto the measurable space $(S^\infty, \mathcal{S}^\infty)$ such that if $x \in (S^N)^\infty$ then

$$\alpha_S^{(N)}(x) = y \in S^\infty,$$

where

$$y_{iN+1}^{iN+N} = x_i, \qquad i \in Z.$$

The source $[A, \mu]$ and channel $[B, C, \nu]$ then induce a source $[A^N, \mu^{(N)}]$ and channel $[B^N, C^N, \nu^{(N)}]$ where

$$\mu^{(N)}(E) = \mu(\alpha_A^{(N)}(E)), \qquad E \in (\mathcal{A}^N)^\infty.$$

$$\nu_x^{(N)}(E) = \nu_{\alpha_B^{(N)}(x)}(\alpha_C^{(N)}(E)), \qquad x \in (B^N)^\infty, E \in (\mathcal{C}^N)^\infty.$$

Note that $[B, C, \nu]$ totally weakly continuous implies $[B^N, C^N, \nu^{(N)}]$ weakly continuous. Given a hookup of $[A, \mu]$ to $[B, C, \nu]$ such that (a), (b) hold, we get a jointly stationary hookup $U^{(N)}, X^{(N)}, Y^{(N)}, V^{(N)}$ of $[A^N, \mu^{(N)}]$ to $[B^N, C^N, \nu^{(N)}]$ such that $E\rho(U^{(N)}, V^{(N)}) < \varepsilon$, where $\rho : A^N \times B^N \to [0, \infty)$ is the metric such that

$$\rho(x,y) = N^{-1}|\{1 \leq i \leq N : x_i \neq y_i\}|.$$

By the proofs of Theorems 3 and 4 we can find block codes $\phi_1 : (A^N)^\infty \to (B^N)^\infty$ and $\phi_2 : (C^N)^\infty \to (A^N)^\infty$ of some order K and a hookup $\hat{U}^{(N)}, \hat{X}^{(N)}, \hat{Y}^{(N)}, \hat{V}^{(N)}$ of

$[A^N, \mu^{(N)}]$ to $[B^N, C^N, \nu^{(N)}]$ such that

(c) $\hat{X}^{(N)} = \phi_1(\hat{U}^{(N)})$, $\hat{V}^{(N)} = \phi_2(\hat{Y}^{(N)})$;

(d) $E[K^{-1}\Sigma_{i=1}^K \rho(\hat{U}_i^{(N)}, \hat{V}_i^{(N)})] < 20\varepsilon$.

This implies there are block codes $\phi_1' : A^\infty \to B^\infty$ and $\phi_2' : C^\infty \to A^\infty$ of order NK and a hookup $\hat{U}, \hat{X}, \hat{Y}, \hat{V}$ of $[A, \mu]$ to $[B, C, \nu]$ such that

$$(NK)^{-1}\Sigma_{i=1}^{NK}\text{Prob}\left[\hat{U}_i \neq \hat{V}_i\right] < 20\varepsilon.$$

In the following, if $[A, \mu_1]$ and $[A, \mu_2]$ are two stationary sources on the same alphabet, let $\bar{d}(\mu_1, \mu_2)$ denote the $\bar{d}$-distance between μ_1, μ_2 [10]. We need the fact that $\bar{d}(\mu_1, \mu_2) < \varepsilon$ if and only if there exist jointly stationary A-valued processes X, Y on some probability space such that the distribution of X is μ_1, the distribution of Y is μ_2, and $\text{Prob}[X_0 \neq Y_0)] < \varepsilon$.

Theorem 6. *Let $[B, C, \nu]$ be a weakly continuous stationary channel. Let $\{[A, \mu_n]\}_{n=1}^\infty$ be a sequence of sources in $\mathfrak{M}_s(\nu)$. Let $[A, \mu]$ be an aperiodic source such that $\bar{d}(\mu_n, \mu) \to 0$. Then $[A, \mu] \in \mathfrak{M}_s(\nu)$.*

Proof. Fix $\varepsilon > 0$. Pick n so large that $\bar{d}(\mu_n, \mu) < \varepsilon/2$. Pick a jointly-stationary hookup U, X, Y, V of $[A, \mu_n]$ to $[B, C, \nu]$ such that $\text{Prob}[U_0 \neq V_0] < \varepsilon/2$. Pick a jointly-stationary sequence of processes W', U', X', Y', V' such that

(a) W', U', X', Y', V' forms a Markov chain;

(b) (U', X', Y', V') and (U, X, Y, V) have same distribution;

(c) W' has distribution μ and $\text{Prob}[W_0' \neq U_0'] < \varepsilon/2$.

Then W', X', Y', V' is a jointly-stationary hookup of $[A, \mu]$ to $[B, C, \nu]$ and $\text{Prob}[W_0' \neq V_0'] < \varepsilon$. By Theorem 3, $[A, \mu] \in \mathfrak{M}_s(\nu)$.

We omit the proof of the following theorem since it is similar to the proof of Theorem 6.

Theorem 7. *Let $[B, C, \nu]$ be a totally weakly continuous stationary channel. Let $\{[A, \mu_n]\}_{n=1}^\infty$ be a sequence of sources in $\mathfrak{M}_b(\nu)$. Let $[A, \mu]$ be an aperiodic source such that $\bar{d}(\mu_n, \mu) \to 0$. Then $[A, \mu] \in \mathfrak{M}_b(\nu)$.*

Theorem 8. *Let $[B, C, \nu]$ be a weakly continuous stationary channel. Let $[A_1, \mu_1] \in \mathfrak{M}_s(\nu)$. Let $[A_2, \mu_2]$ be an aperiodic factor of $[A_1, \mu_1]$. Then $[A_2, \mu_2] \in \mathfrak{M}_s(\nu)$.*

Note. The requirement that $[A_2, \mu_2]$ be aperiodic cannot be removed. In [6] examples of $\bar{d}$-continuous channels are given which show this.

Proof of Theorem 8. Fix $\varepsilon > 0$. Since $[A_2, \mu_2]$ is a factor of $[A_1, \mu_1]$ we may pick jointly-stationary processes W, U and a sliding-block code $\phi : A_1^\infty \to A_2^\infty$ such that the distribution of W is μ_2, the distribution of U is μ_1, and $\text{Prob}[W_0 \neq \phi(U)_0] < \varepsilon/2$. Pick a positive integer M such that

$$\phi(x)_i \neq \phi(y)_i$$

implies

$$x_{i-M}^{i+M} \neq y_{i-M}^{i+M}, \qquad i \in Z, x, y \in A_1^\infty.$$

Pick a jointly-stationary Markov chain W', U', X', Y', V' such that

(a) (W', U') and (W, U) have same distribution;

(b) U', X', Y', V' is a hookup of $[A_1, \mu_1]$ to $[B, C, \nu]$ such that $\mathrm{Prob}[U'_0 \neq V'_0] < \varepsilon(4M + 2)^{-1}$.

Then $W', X', Y', \phi(V')$ is a jointly-stationary hookup of $[A_2, \mu_2]$ to $[B, C, \nu]$ and

$$\mathrm{Prob}\big[W'_0 \neq \phi(V')_0\big] \leqslant \mathrm{Prob}\big[W'_0 \neq \phi(U')_0\big] + \mathrm{Prob}\big[\phi(U')_0 \neq \phi(V')_0\big]$$
$$< \varepsilon/2 + \Sigma_{i=-M}^{M}\mathrm{Prob}\big[U'_i \neq V'_i\big] < \varepsilon.$$

Now apply Theorem 3.

The proof of the following theorem is similar to that of Theorem 8 so we omit it.

THEOREM 9. *Let* $[B, C, \nu]$ *be a totally weakly continuous stationary channel. Let* $[A, \mu] \in \mathfrak{M}_b(\nu)$. *Then every aperiodic factor of* $[A, \mu]$ *is in* $\mathfrak{M}_b(\nu)$.

Our next theorem will enable us to obtain an upper bound to $C_s(\nu)$ and $C_b(\nu)$ which is computable for many channels. First, some notation and definitions. If X and Y are discrete random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, $\overline{H}(X|Y)$ will denote the conditional entropy of X given Y [14]. If (X, Y) are jointly N-stationary processes on $(\Omega, \mathcal{F}, P)$, let $h(X)$, $h(X|Y)$, $i(X, Y)$ denote the functions in $L^1(P)$ such that

$$h(X)(\omega) = \lim_{n\to\infty} - n^{-1}\log P(X_1 = X_1(\omega), \cdots, X_n = X_n(\omega)), \text{ almost surely;}$$
$$h(X|Y)(\omega) = \lim_{n\to\infty} - n^{-1}\log P(X_1 = X_1(\omega), \cdots, X_n$$
$$= X_n(\omega)|Y_1 = Y_1(\omega), \cdots, Y_n = Y_n(\omega)), \text{ almost surely;}$$

and

$$i(X, Y) = h(X) - h(X|Y).$$

(We write $i_P(X, Y)$ for $i(X, Y)$ to emphasize the underlying probability measure P, if necessary.) Then, $H(X)$, the entropy of the process X, is $E[h(X)]$, and $H(X|Y)$, the conditional entropy of the process X given the process Y, is $E[h(X|Y)]$.

Following [2], define $C^*(\nu)$, the information quantile capacity of the stationary channel $[B, C, \nu]$, as follows:

$$C^*(\nu) = \lim_{\lambda\to 0^+} C^*(\nu, \lambda), \quad \text{where for } \lambda > 0,$$
$$C^*(\nu, \lambda) = \sup_{\mu\in\mathcal{P}_1(B)}\sup\{r : \mu\nu\big[i_{\mu\nu}(X, Y) \leqslant r\big] < \lambda,$$

X, Y in this case being the projections from $B^\infty \times C^\infty$ to B^∞, C^∞ respectively. (We remark for later use that in calculating $C^*(\nu, \lambda)$, the outer supremum over $\mathcal{P}_1(B)$ may be replaced by a supremum over $\mathcal{P}_N(B), N = 1, 2, \cdots$. See [4], Lemma 3.)

THEOREM 10. *Let* $[B, C, \nu]$ *be a stationary channel. If* $[A, \mu]$ *is an ergodic source in* $C_s(\nu)$ *or* $C_b(\nu)$, *then* $H(\mu) \leqslant C^*(\nu)$.

PROOF. Assume $[A, \mu] \in C_b(\nu)$. Fix ε such that $0 < \varepsilon < \frac{1}{2}$. For some N find block codes $\phi_1 : A^\infty \to B^\infty$ and $\phi_2 : C^\infty \to A^\infty$ of order N and a hookup U, X, Y, V of

$[A, \mu]$ to $[B, C, \nu]$ defined on $(\Omega, \mathcal{F}, P)$ such that

(a) $X = \phi_1(U)$, $V = \phi_2(Y)$;

(b) $N^{-1}\Sigma_{i=1}^N P[U_i \neq V_i] < \varepsilon$.

Let $q : [0, 1) \to [0, \infty)$ be the function

$$q(x) = -x \log x - (1 - x)\log(1 - x) + x \log|A|, \quad 0 < x < 1$$
$$q(0) = 0.$$

By Fano's inequality [14], Theorem 3.7.1, and the concavity of q,

$$N^{-1}\overline{H}(U_1^N | V_1^N) \leqslant N^{-1}\Sigma_{i=1}^N \overline{H}(U_i | V_i) \leqslant N^{-1}\Sigma_{i=1}^N q(P(U_i \neq V_i))$$
$$\leqslant q(N^{-1}\Sigma_{i=1}^N P(U_i \neq V_i)) \leqslant q(\varepsilon).$$

The preceding implies, as shown by Gray and Ornstein in the proof of Theorem 5.2 of [2],

(c) $H(X | Y) \leqslant q(\varepsilon)$, and

(d) $H(U) - H(X) \leqslant q(\varepsilon)$.

Now Lemma 3.1 of [15] implies that the essential supremum of $h(X)$ is the maximum of the entropies of the T_B^N-ergodic components of P^X. These components are encodings of the at most N T_A^N-ergodic components of μ, each of which has entropy equal to that of μ. Since encoding does not increase entropy, we see that

(e) $P[h(X) \leqslant H(U)] = 1$.

From (d) and (e), we have

$$\text{(f)} \quad P\left[h(X) \leqslant H(U) - q(\varepsilon)^{\frac{1}{2}}\right] = P\left[H(U) - h(X) \geqslant q(\varepsilon)^{\frac{1}{2}}\right]$$
$$\leqslant q(\varepsilon)^{-\frac{1}{2}}(H(U) - H(X)) \leqslant q(\varepsilon)^{\frac{1}{2}}.$$

From (c) we have

(g) $P[h(X | Y) \geqslant q(\varepsilon)^{\frac{1}{2}}] \leqslant q(\varepsilon)^{\frac{1}{2}}$.

From (f) and (g), we obtain

$$P\left[i(X, Y) \leqslant H(U) - 2q(\varepsilon)^{\frac{1}{2}}\right] \leqslant 2q(\varepsilon)^{\frac{1}{2}}.$$

Since $P^X \in \mathcal{P}_N(B)$, we have by the remark following the definition of $C^*(\nu)$ that

$$C^*(\nu, 3q(\varepsilon)^{\frac{1}{2}}) \geqslant H(\mu) - 2q(\varepsilon)^{\frac{1}{2}}.$$

Letting $\varepsilon \to 0$, we get $C^*(\nu) \geqslant H(\mu)$. Similarly, one can show this inequality if $[A, \mu] \in C_s(\nu)$.

THEOREM 11. *Let $[B, C, \nu]$ be a stationary $\bar{d}$-continuous channel. Then $C_b(\nu) = C_b'(\nu) = C^*(\nu)$. If $[A, \mu]$ is an ergodic source, the following hold:*

(A) *If $H(\mu) < C^*(\nu)$, $[A, \mu]$ is strongly block transmissible. If $H(\mu) > C^*(\nu)$, $[A, \mu]$ is not strongly block transmissible.*

(B) *If $H(\mu) < C^*(\nu)$, $[A, \mu]$ is block transmissible. If $H(\mu) > C^*(\nu)$, $[A, \mu]$ is not block transmissible.*

PROOF. Part (A) was proved in [2]. Part (B) follows from Part (A) and Theorem 10. $C_b(\nu) = C_b'(\nu) = C^*(\nu)$ follows from (A), (B).

PROOF OF THEOREM 1. Let $[B, C, \nu]$ be a weakly continuous stationary channel. Let $[A, \mu]$ be a B-source with $H(\mu) < C_s(\nu)$. Pick an ergodic source $[A', \mu']$ such that $[A', \mu'] \in \mathfrak{M}_s(\nu)$ and $H(\mu') > H(\mu)$. Then $[A, \mu]$ is a factor of $[A', \mu']$ ([11]), and so $[A, \mu] \in \mathfrak{M}_s(\nu)$. (Treat the case where $[A, \mu]$ is not aperiodic separately; in that case, μ assigns measure one to some constant sequence in A^∞ and thus it is trivial to see that $[A, \mu]$ is transmissible.) To complete the proof, if $[A, \mu]$ is a B-source with $H(\mu) = C_s(\nu) > 0$, we need to show $[A, \mu] \in \mathfrak{M}_s(\nu)$.

It is clear from the results of [5] that there is a sequence of stationary sources $\{[A, \mu_n]\}$ such that $\mu_n \to \mu$ in the weak topology, $H(\mu_n) \to H(\mu)$, and for each n, μ_n is a finite-order Markovian measure mixing with respect to T_A. From [11], each $[A, \mu_n]$ is a B-source and $\bar{d}(\mu_n, \mu) \to 0$. We have $[A, \mu_n] \in \mathfrak{M}_s(\nu)$ for each n by the first part of the proof since $H(\mu_n) < C_s(\nu)$. Hence $[A, \mu] \in \mathfrak{M}_s(\nu)$ by Theorem 6. The proof of Theorem 2 is similar.

3. An example. We see from Theorem 4 that $C_s(\nu) \leq C_b(\nu)$. For certain types of $\bar{d}$-continuous channels [3, 6], it is known that $C_s(\nu) = C_b(\nu)$. We present an example of a $\bar{d}$-continuous channel for which $C_s(\nu) < C_b(\nu)$.

First, we need the following lemma.

LEMMA 1. *Let $[B, C, \nu]$ be a stationary channel. Suppose $\mathfrak{M}_s(\nu)$ contains an aperiodic source. Then for each $\varepsilon > 0$, there exists $W \in \mathcal{C}^\infty$ and $x \in B^\infty$ such that $\nu_x(W) < \varepsilon$ and $\nu_{Tx}(W^c) < \varepsilon$.*

PROOF. Fix $\varepsilon > 0$. Let $\delta > 0$ be a small positive number to be chosen later. Let $[A, \mu]$ be an aperiodic source in $\mathfrak{M}_s(\nu)$. By Rohlin's theorem [12] pick $E \in \mathcal{A}^\infty$ such that

(a) $E \cap T^{-1}E = \varnothing$;

(b) $\mu(E \cup T^{-1}E) > 1 - \delta$.

We may suppose that for some N, E is of form

$$E = \{u \in A^\infty : u_1^N \in E'\}$$

for some subset E' of A^N. Pick sliding-block codes $\phi_1 : A^\infty \to B^\infty$ and $\phi_2 : C^\infty \to A^\infty$ and a hookup U, X, Y, V of $[A, \mu]$ to $[B, C, \nu]$ such that $X = \phi_1(U)$, $V = \phi_2(Y)$, and $\text{Prob}[U_0 \neq V_0] < \delta/N$. By Lemma 3 of the Appendix, pick a sliding-block code $f: B^\infty \to A^\infty$ such that $\text{Prob}[U_0 \neq f(X)_0] < \delta/N$. Let $\hat{X} = f(X)$. Then, letting $F = f^{-1}E$, $G = \phi_2^{-1}E$,

$$\mu^{\phi_1}\nu\big[(F \times G) \cup (T^{-1}F \times T^{-1}G)\big] \geq 2\,\text{Prob}\big[U_1^N = \hat{X}_1^N, U_1^N = V_1^N, U \in E\big] > 1 - 5\delta.$$

Let

$$E' = (F \times G) \cup (T^{-1}F \times T^{-1}G), E'' = (F \times T^{-1}G) \cup (T^{-1}F \times T^{-2}G).$$

Then $E' \cap E'' = \phi$, and

$$\int_{B^\infty} \nu_{T^{-1}x}(E''_x)\, d\mu^{\phi_1}(x) = \int_F \nu_{T^{-1}x}(T^{-1}G)\, d\mu^{\phi_1}(x)$$
$$+ \int_{T^{-1}F} \nu_{T^{-1}x}(T^{-2}G)\, d\mu^{\phi_1}(x) = \int_F \nu_x(G)\, d\mu^{\phi_1}(x) + \int_{T^{-1}F} \nu_x(T^{-1}G)\, d\mu^{\phi_1}(x)$$
$$= \mu^{\phi_1}\nu(E') > 1 - 5\delta,$$

where
$$E''_x = \{y \in C^\infty : (x,y) \in E''\}.$$

We thus have

(c) $\int_{B^\infty} \nu_{T^{-1}x}(E''_x)\, d\mu^{\phi_1}(x) > 1 - 5\delta$;

(d) $\int_{B^\infty} \nu_x(E'_x)\, d\mu^{\phi_1}(x) > 1 - 5\delta$.

Chebyshev's inequality applied to (c) gives

$$\nu_{T^{-1}x}(E''_x) > 1 - (5\delta)^{\frac{1}{2}}$$

for a set of x of μ^{ϕ_1} probability $> 1 - (5\delta)^{\frac{1}{2}}$. Similarly, from (d) we get that $\nu_x(E'_x) > 1 - (5\delta)^{\frac{1}{2}}$ for a set of x of μ^{ϕ_1} probability $> 1 - (5\delta)^{\frac{1}{2}}$. If $1 - (5\delta)^{\frac{1}{2}} > 1/2$, there must exist $x_1 \in B^\infty$ such that

$$\nu_{T^{-1}x_1}(E''_{x_1}) > 1 - (5\delta)^{\frac{1}{2}} \text{ and } \nu_{x_1}(E'_{x_1}) > 1 - (5\delta)^{\frac{1}{2}}.$$

Setting $W = (E''_{x_1})^c$, we get, since $E''_{x_1} \cap E'_{x_1} = \varnothing$,

$$\nu_{T^{-1}x_1}(W) < (5\delta)^{\frac{1}{2}}, \nu_{x_1}(W^c) < (5\delta)^{\frac{1}{2}}.$$

Hence, choosing δ so that $(5\delta)^{\frac{1}{2}} < \min(\frac{1}{2}, \varepsilon)$, there exists $x \in B^\infty$ such that $\nu_x(W) < \varepsilon$ and $\nu_{Tx}(W^c) < \varepsilon$.

Now we construct our example. Let $[B, C, \nu]$ be the channel such that $B = C = \{0, 1\}$ and

$$\nu_x(E) = \tfrac{1}{2} I_E(x) + \tfrac{1}{2} I_E(Tx), \qquad x \in B^\infty, E \in \mathcal{C}^\infty,$$

where I_E denotes the indicator function (characteristic function) of the set E. Thus an input x when transmitted over the channel yields either itself or the shift of itself, each with probability $\frac{1}{2}$. This channel is stationary and has *finite input memory*; that is, there exists a positive integer M such that for $n = 1, 2, \cdots$, and $k \in Z$, if $E \subset C^n$ and $x, \hat{x}$ are sequences in B^∞ with

$$x_{k-M}^{k+M+n-1} = \hat{x}_{k-M}^{k+M+n-1}$$

then

$$\nu_x\{y \in C^\infty : y_k^{k+n-1} \in E\} = \nu_{\hat{x}}\{y \in C^\infty : y_k^{k+n-1} \in E\}.$$

Such channels are known to be $\bar{d}$-continuous [2].

Now for every $E \in \mathcal{C}^\infty$,

$$\nu_x(E) + \nu_{Tx}(E^c) \geqslant \tfrac{1}{2} I_E(Tx) + \tfrac{1}{2} I_{E^c}(Tx) = \tfrac{1}{2}.$$

Hence, by Lemma 1, $\mathfrak{M}_s(\nu)$ contains no aperiodic sources. This implies

$$C_s(\nu) = 0.$$

However, it is an easy task, which we leave to the reader, to construct a block encoder and decoder for transmitting an ergodic source of entropy $< (\log 2)/2$. Thus $C_b(\nu) \geqslant (\log 2)/2$. Or, one can simply observe that ν is obtained by averaging two ergodic channels of capacity log 2; hence by a formula of Nedoma (see [9] or [4]),

$$C_b(\nu) \geqslant \left[(\log 2)^{-1} + (\log 2)^{-1}\right]^{-1} = (\log 2)/2.$$

4. Open problems. Can one get coding theorems for the class of all ergodic aperiodic sources, not just the B-sources? Can one obtain formulas for $C_s(\nu)$ and $C_b(\nu)$? Specifically, does $C_b(\nu) = C^*(\nu)$ hold in general? Can one find a necessary and sufficient condition in order that $C_s(\nu) = C_b(\nu)$? Specifically, motivated by the example given in previous section, is the only way that $C_s(\nu)$ can fail to equal $C_b(\nu)$ is if $C_s(\nu) = 0$? Can one find a necessary and sufficient condition for a channel to be weakly continuous (totally weakly continuous)? Is every weakly continuous channel totally weakly continuous? When does $C_b(\nu) = C_b'(\nu)$? Is there a coding theorem analogous to Theorem 2 for the strong type of block transmissibility?

APPENDIX

A stationary channel $[B, C, \nu]$ is $\bar{d}$-continuous if for any $\varepsilon > 0$ there exists a positive integer N_0 such that for any $N \geqslant N_0$ and $x, \hat{x} \in B^\infty$ with $x_1^N = \hat{x}_1^N$, we may find C^N-valued random functions $Y, \hat{Y}$ defined on some probability space so that

(a) the distribution of Y is ν_x^N;

(b) the distribution of $\hat{Y}$ is $\nu_{\hat{x}}^N$;

(c) $E\rho_N^C(Y, \hat{Y}) < \varepsilon$.

(If S is finite, ρ_N^S is the map from $S^N \times S^N \to [0, 1]$ such that

$$\rho_N(y,\hat{y}) = N^{-1}|\{1 \leqslant i \leqslant N : y_i \neq \hat{y}_i\}|, y, \hat{y} \in S^N.)$$

LEMMA 2. *Let $[B, C, \nu]$ be a stationary $\bar{d}$-continuous channel. Then $[B, C, \nu]$ is totally weakly continuous.*

PROOF. It suffices to show that $[B, C, \nu]$ is weakly continuous. (For if $[B, C, \nu]$ is $\bar{d}$-continuous it is easy to see that $[B^N, C^N, \nu^{(N)}]$ is $\bar{d}$-continuous, $N = 1, 2, \cdots$. The channel $[B, C, \nu]$ is totally weakly continuous if and only if each channel $[B^N, C^N, \nu^{(N)}]$ is weakly continuous, $N = 1, 2, \cdots$.) Fix finite dimensional cylinder sets $E \in \mathcal{B}^\infty$, $F \in \mathcal{C}^\infty$. It suffices to show that the map $\mu \to \mu\nu(E \times F)$ is continuous from $\mathcal{P}_1(B) \to [0, \infty)$. Let $A = \{0, 1, 2\}$. Let $f : B^\infty \to A^\infty$ and $g : C^\infty \to A^\infty$ be the sliding-block codes such that

$$\begin{aligned} f(x)_0 &= 0, && x \in E \\ &= 1, && \text{otherwise.} \\ g(y)_0 &= 0, && y \in F \\ &= 2, && \text{otherwise.} \end{aligned}$$

Let X, Y be the maps from $B^\infty \times C^\infty$ to B^∞, C^∞ respectively, such that

$$X(x,y) = x,\ Y(x,y) = y, \qquad x \in B^\infty, y \in C^\infty.$$

Let $U = f(X)$, $V = g(Y)$. Then if $\mu \in \mathcal{P}_1(B)$,

$$\mu\nu[U_0 \neq V_0] = 1 - \mu\nu(E \times F).$$

Thus all we have to do is show that the map $\mu \to \mu\nu[U_0 \neq V_0]$ from $\mathcal{P}_1(B) \to [0, \infty)$ is continuous.

Pick M such that $x_{i-M}^{i+M} = \hat{x}_{i-M}^{i+M}$ implies

$$f(x)_i = f(\hat{x})_i, \qquad x, \hat{x} \in B^\infty, i \in Z,$$

and $y_{i-M}^{i+M} = \hat{y}_{i-M}^{i+M}$ implies

$$g(y)_i = g(\hat{y})_i, \qquad y, \hat{y} \in C^\infty, i \in Z.$$

Fix $\varepsilon > 0$.

Pick N so large that:

(a) if $x, \hat{x} \in B^\infty$ and $x_{1-M}^{N+M} = \hat{x}_{1-M}^{N+M}$ then there are C^{N+2M}-valued random functions $Y, \hat{Y}$ such that

$$\mathrm{Prob}[Y = \mathbf{y}] = \nu_x\{y \in C^\infty : y_{1-M}^{N+M} = \mathbf{y}\}, \quad \mathbf{y} \in C^{N+2M},$$

$$\mathrm{Prob}[\hat{Y} = \mathbf{y}] = \nu_{\hat{x}}\{y \in C^\infty : y_{1-M}^{N+M} = \mathbf{y}\}, \quad \mathbf{y} \in C^{N+2M},$$

and $E\rho_{N+2M}^C(Y, \hat{Y}) < \varepsilon/(2M + 1)$.

(b) $2M/N < 1$.

Let $\mu_1, \mu_2 \in \mathcal{P}_1(B)$ satisfy

$$\Sigma_{\mathbf{x} \in B^{N+2M}} |\mu_1^{N+2M}(\mathbf{x}) - \mu_2^{N+2M}(\mathbf{x})| < \varepsilon.$$

We show $|\mu_1\nu[U_0 \neq V_0] - \mu_2\nu[U_0 \neq V_0]| < 5\varepsilon$. We have

$$|\mu_1\nu[U_0 \neq V_0] - \mu_2\nu[U_0 \neq V_0]| = |\textstyle\int_{B^\infty}\psi\, d\mu_1 - \int_{B^\infty}\psi\, d\mu_2|,$$

where $\psi : B^\infty \to [0, 1]$ is the map such that

$$\psi(x) = \textstyle\int_{C^\infty}\rho_N^A\big(f(x)_1^N, g(y)_1^N\big)\, d\nu_x(y), \qquad x \in B^\infty.$$

Pick maps $f^* : B^{N+2M} \to A^N$ and $g^* : C^{N+2M} \to A^N$ such that

$$f(x)_1^N = f^*(x_{1-M}^{N+M}),\quad g(y)_1^N = g^*(y_{1-M}^{N+M}),$$

$$x \in B^\infty, y \in C^\infty.$$

Let $\mathbf{x} \in B^{N+2M}$. Let $x, \hat{x} \in B^\infty$ satisfy $x_{1-M}^{N+M} = \hat{x}_{1-M}^{N+M} = \mathbf{x}$. Let $Y, \hat{Y}$ be the C^{N+2M}-valued random functions corresponding to $x, \hat{x}$ given by (a).

We have

$$|\psi(x) - \psi(\hat{x})| = |E[\rho_N^A(f^*(\mathbf{x}), g^*(Y))] - E[\rho_N^A(f^*(\mathbf{x}), g^*(\hat{Y}))]|$$

$$\leqslant E\rho_N^A\big(g^*(Y), g^*(\hat{Y})\big).$$

A simple calculation shows $\rho_N^A(g^*(Y), g^*(\hat{Y})) \leqslant N^{-1}(2M + 1)(N + 2M)\rho_{N+2M}^C(Y, \hat{Y})$. Taking the expected value we see using (a), (b) that $|\psi(x) - \psi(\hat{x})| \leqslant 2\varepsilon$. Thus, letting $E_{\mathbf{x}} = \{x \in B^\infty : x_{1-M}^{N+M} = \mathbf{x}\}$, we have shown that the oscillation of ψ over $E_{\mathbf{x}}$ is $\leqslant 2\varepsilon$. Hence,

$$|\int_{E_{\mathbf{x}}}\psi\, d\mu_1 - \int_{E_{\mathbf{x}}}\psi\, d\mu_2| \leqslant |\int_{E_{\mathbf{x}}}(\psi - \sup_{E_{\mathbf{x}}}\psi)\, d\mu_1 - \int_{E_{\mathbf{x}}}(\psi - \sup_{E_{\mathbf{x}}}\psi)\, d\mu_2|$$

$$+ |\sup_{E_{\mathbf{x}}}\psi||\mu_1(E_{\mathbf{x}}) - \mu_2(E_{\mathbf{x}})| \leqslant 2\varepsilon(\mu_1(E_{\mathbf{x}}) + \mu_2(E_{\mathbf{x}})) + |\mu_1(E_{\mathbf{x}}) - \mu_2(E_{\mathbf{x}})|.$$

Summing over $\mathbf{x} \in B^{N+2M}$, we get $|\int\psi\, d\mu - \int\psi\, d\mu_2| < 5\varepsilon$.

LEMMA 3. *Let A be a finite set and $(\Omega_1, \mathfrak{F}_1)$ a measurable space. Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let X, Y, Z be measurable functions defined on this space such that*

(a) *X is A-valued, Y is Ω_1-valued, Z is A-valued;*

(b) *X, Y, Z form a Markov chain.*

Then there exists a measurable map $f: \Omega_1 \to A$ such that $E[\rho(X, f(Y))] \leqslant E[\rho(X, Z)]$, where ρ is a map from $A \times A \to [0, \infty)$.

PROOF. Let $\{P_\omega : \omega \in \Omega_1\}$ be a family of probability measures on A which serve as regular conditional probabilities of X given Y; that is,

(c) The map $\omega \to P_\omega(E)$ from $\Omega_1 \to [0, 1]$ is measurable, $E \subset A$;

(d) $P[X \in E, Y \in F] = \int_F P_\omega(E)\, dP^Y(\omega)$, $E \subset A, F \in \mathfrak{F}_1$.

Let $\psi : \Omega_1 \times A \to [0, \infty)$ be the function such that

$$\psi(\omega, x) = \int_A \rho(x', x)\, dP_\omega(x'), \qquad x \in A, \omega \in \Omega_1.$$

Because of the Markov assumption (b),

$$E[\rho(X, Z)] = \int_{\Omega_1 \times A}\psi(y, z)\, dP^{(Y,Z)}(y, z).$$

Define $f: \Omega_1 \to A$ so that $\psi(\omega, f(\omega)) = \min_{a \in A}\psi(\omega, a)$, $\omega \in \Omega$. Then

$$E[\rho(X, f(Y))] = \int_{\Omega_1}\psi(y, f(y))\, dP^Y(y) = \int_{\Omega_1 \times A}\psi(y, f(y))\, dP^{(Y,Z)}(y, z)$$

$$\leqslant \int_{\Omega_1 \times A}\psi(y, z)\, dP^{(Y,Z)}(y, z) = E[\rho(X, Z)].$$

For the following, we introduce some notation. If A, B are sets and τ is a relation on $B \times A$ then

$$\tau[x] = \{y \in A : x\tau y\}, \qquad x \in B.$$

If $S \subset A$,

$$\tau^{-1}[S] = \{x \in B : x\tau y \text{ for some } y \in S\}.$$

LEMMA 4. *Let A be a finite set. Let $\{p_S : S \subset A\}$ and $\{\lambda_i : i \in A\}$ be sets of nonnegative numbers such that*

(a) $\sum_{\{S' \subset A : S' \cap S \neq \phi\}} p_{S'} \geqslant \sum_{i \in S}\lambda_i$, $S \subset A$.

Then there exists a set $\{\alpha_i^S : i \in A, S \subset A\}$ of nonnegative numbers such that

(b) $\sum_{\{i \in A : i \in S\}}\alpha_i^S \leqslant p_S$, $S \subset A$.

(c) $\sum_{\{S \subset A : i \in S\}}\alpha_i^S = \lambda_i$, $i \in A$.

PROOF. Choose finite nonatomic measure spaces $(\Omega_1, \mathfrak{F}_1, m_1)$ and $(\Omega_2, \mathfrak{F}_2, m_2)$ so that $(\Omega_1, \mathfrak{F}_1)$, $(\Omega_2, \mathfrak{F}_2)$ are standard Borel spaces and there exists a partition $P = \{E_S : S \subset A\}$ of Ω_1 and a partition $Q = \{F_i : i \in A\}$ of Ω_2 with $m_1(E_S) = P_S$,

$S \subset A$, and $m_2(F_i) = \lambda_i$, $i \in A$. Define a relation τ on $P \times Q$ as follows: $E_S \tau F_i$ if and only if $i \in S$.

Then (a) reduces to:

(d) $m_1\left[\ \cup_{E \in \tau^{-1}[\mathbb{S}]} E\ \right] \geqslant m_2\left[\ \cup_{F \in \mathbb{S}} F\ \right], \mathbb{S} \subset Q$.

Thus by the measure marriage lemma of [7], page 23, there exists a one-to-one measure-preserving map $f: \Omega_2 \to \Omega_1$ which preserves the relation. This means

(e) $f(F_i) \subset \cup_{\{S \subset A : i \in S\}} E_S, i \in A$.

Set

$$\alpha_i^S = m_1(f(F_i) \cap E_S), \qquad i \in A, S \subset A.$$

Now (e) implies that $\alpha_i^S = 0$ if $i \notin S$. Thus

$$\Sigma_{\{S \subset A : i \in S\}} \alpha_i^S = \Sigma_{S \subset A} \alpha_i^S = m_1(f(F_i)) = m_2(F_i) = \lambda_i,$$

and so (c) holds. Inequality (b) holds automatically from the way the $\{\alpha_i^S\}$ were defined.

LEMMA 5. *Let $(B, \mathcal{B}, m)$, $(A, \mathcal{Q}, \lambda)$ be finite measure spaces, where A, B are finite sets. Let τ be a relation on $B \times A$. Suppose*

(a) $m[\tau^{-1}[S]] \geqslant \lambda(S), S \subset A$.

Then there exist disjoint subsets $\{F_i : i \in A\}$ of B such that

(b) $m(\cup_{i \in A} F_i) \geqslant \Sigma_{i \in A} \lambda(i) - |A| 2^{|A|} \max_{b \in B} m(b)$.

(c) $|m(F_i) - \lambda(i)| \leqslant 2^{|A|} \max_{b \in B} m(b), i \in A$.

(d) $\tau^{-1}\{i\} \supset F_i, i \in A$.

PROOF. Let $E_S = \{x \in B : \tau[x] = S\}$, $S \subset A$. Let $p_S = m(E_S)$, $S \subset A$. Then (a) gives

(e) $\Sigma_{\{S' \subset A : S' \cap S \neq \phi\}} p_{S'} \geqslant \lambda(S), S \subset A$.

Hence, by Lemma 4, there are nonnegative numbers $\{\alpha_i^S : i \in A, S \subset A\}$ such that

(f) $\Sigma_{\{S \subset A : i \in S\}} \alpha_i^S = \lambda(i), i \in A$.

(g) $\Sigma_{\{i \in A : i \in S\}} \alpha_i^S \leqslant m(E_S), S \subset A$.

Because of (g) we may choose sets $\{E_S^i : i \in A, S \subset A\}$ such that

(h) $E_{S'}^{i'} \cap E_S^i = \varnothing$ if $i \neq i'$ or $S \neq S'$.

(i) $E_S^i \subset E_S, i \in S$.

(j) $|m(E_S^i) - \alpha_i^S| \leqslant \max_{b \in B} m(b), i \in S, S \neq \varnothing, S \subset A$.

Define

$$F_i = \cup_{\{S \subset A : i \in S\}} E_S^i, \qquad i \in A.$$

The $\{F_i\}$ are disjoint by (h). From (i), we get (d). Also,

$$|m(F_i) - \lambda(i)| = |\Sigma_{S \supset \{i\}} m(E_S^i) - \Sigma_{S \supset \{i\}} \alpha_i^S| \leqslant 2^{|A|} \max_{b \in B} m(b),$$

by (f) and (j). Thus (c) holds. (b) follows from (c).

LEMMA 6. *Let A, B be finite sets. Let X, Y be a pair of jointly stationary processes on some probability space $(\Omega, \mathcal{F}, P)$, where X is A-valued and aperiodic, and Y is B-valued. Then there is a sequence $\{\phi_n\}_{n=1}^{\infty}$ of sliding-block codes from $A^\infty \to B^\infty$ such that $P^{(X, \phi_n(X))} \to P^{(X, Y)}$ in the weak topology on $\mathcal{P}_1(A, B)$.*

PROOF. We can assume $\Omega = A^\infty \times B^\infty$, $\mathcal{F} = \mathcal{A}^\infty \times \mathcal{B}^\infty$, and that X, Y are the projections from $A^\infty \times B^\infty$ to A^∞, B^∞, respectively. Let $\delta > 0$ and a positive integer N be given. It suffices to construct a stationary code $\phi : A^\infty \to B^\infty$ such that

$$\Sigma_{(x,y)\in A^N\times B^N}|P^{(X_1^N,\phi(X)_1^N)}(x,y) - P^{(X_1^N,Y_1^N)}(x,y)| < \delta.$$

(For, by [1] Theorem 3.1 the stationary code ϕ may be replaced by a sliding-block code which does this also.)

We fix $0 < \eta < 1$ to be specified later. There is a family of probability measures $\{P_\omega : \omega \in A^\infty \times B^\infty\}$ on $\mathcal{A}^\infty \times \mathcal{B}^\infty$ such that

(a) The map $\omega \to P_\omega(E)$ from $A^\infty \times B^\infty \to [0, 1]$ is measurable, $E \in \mathcal{A}^\infty \times \mathcal{B}^\infty$.

(b) $P(E) = \int_{A^\infty\times B^\infty} P_\omega(E)\,dP(\omega), E \in \mathcal{A}^\infty \times \mathcal{B}^\infty$.

(c) $\lim_{n\to\infty} n^{-1}\Sigma_{i=1}^n I_E(T^i\omega) = P_\omega(E)$ for P-almost all $\omega \in A^\infty \times B^\infty$, for each $E \in \mathcal{A}^\infty \times \mathcal{B}^\infty$.

Pick a measurable partition $A_1, \cdots, A_j$ of $A^\infty \times B^\infty$ such that

(d) $|P_{\omega_1}[(X_1^N, Y_1^N) = b] - P_{\omega_2}[(X_1^N, Y_1^N) = b]| < \eta$, $\omega_1, \omega_2 \in A_i$, $i = 1, \cdots, j$; $b \in A^N \times B^N$.

Let $w_i = P(A_i)$, $i = 1, \cdots, j$.

Let $\alpha_1, \cdots, \alpha_j$ be the probability measures on $A^N \times B^N$ such that

$$\alpha_i(b) = w_i^{-1}\int_{A_i} P_\omega\big[(X_1^N, Y_1^N) = b\big]\,dP(\omega), \qquad b \in A^N \times B^N.$$

(If $w_i = 0$, just define α_i to be any probability measure on $A^N \times B^N$.) Note that

$$P^{(X_1^N, Y_1^N)} = \Sigma_{i=1}^j w_i\alpha_i.$$

Also, from (d),

(e) $|P_\omega[(X_1^N, Y_1^N) = b] - \alpha_i(b)| < \eta$ for P-almost every $\omega \in A_i$, $b \in A^N \times B^N$, $i = 1, \cdots, j$.

For $M = 1, 2, \cdots$, let $A_1^{(M)}, A_2^{(M)}, \cdots, A_j^{(M)}$ be a partition of $A^M \times B^M$ such that

(f) $P[\{(X_1^M, Y_1^M) \in A_i^{(M)}\}\Delta A_i] \to 0$ as $M \to \infty$, $i = 1, \cdots, j$.

If $(x,y) \in A^M \times B^M$, where $M > N$, and α is a probability measure on $A^N \times B^N$, we say (x,y) is η-typical of α if

$$|(M - N + 1)^{-1}\Sigma_{i=1}^{M-N+1} I_{\{b\}}(x_i^{i+N-1}, y_i^{i+N-1}) - \alpha(b)| < \eta, \qquad b \in A^N \times B^N.$$

Because of (c), (e) and (f), the following hold for M sufficiently large:

(g) $P[\cup_{i=1}^j\{(X_1^M, Y_1^M)$ is η-typical of α_i and $(X_1^M, Y_1^M) \in A_i^{(M)}\}] > 1 - \eta/j$;

(h) $|P[(X_1^M, Y_1^M) \in A_i^{(M)}] - w_i| < \eta/j, i = 1, \cdots, j$.

Fix such a large M. (Later on in the proof we will make further restrictions on M.)

Let

$$E_i = \{(x,y) \in A^M \times B^M : (x,y) \text{ is } \eta\text{-typical of } \alpha_i, (x,y) \in A_i^{(M)}\}, \qquad i = 1, \cdots, j.$$

Define a relation τ on $A^M \times \{1, \cdots, j\}$ as follows: if $x \in A^M$ and $1 \leqslant i \leqslant j$, then

$x\tau i$ if and only if there exists $y \in B^M$ such that $(x,y) \in E_i$. If $G \subset \{1,\cdots,j\}$, then

$$\begin{aligned} P^{X_1^M}[\tau^{-1}[G]] &\geqslant P^{X_1^M}\{x \in A^M : P[\cup_{i\in G}\{(X_1^M, Y_1^M) \in E_i\} | X = x] > 0\} \\ &\geqslant \int_{A^M} P[\cup_{i\in G}\{(X_1^M, Y_1^M) \in E_i\} | X_1^M = x]\, dP^{X_1^M}(x) \\ &= \Sigma_{i\in G} P[(X_1^M, Y_1^M) \in E_i]. \end{aligned}$$

By Lemma 5, there are disjoint subsets $\{F_i : i = 1,\cdots,j\}$ of A^M such that (letting $\varepsilon_M = \max_{x\in A^M} P^{X_1^M}(x)$)

(i) $P^{X_1^M}(\cup_{i=1}^j F_i) \geqslant \Sigma_{i=1}^j P[(X_1^M, Y_1^M) \in E_i)] - j2^j\varepsilon_M > 1 - \eta - j2^j\varepsilon_M$;

(j) $|P^{X_1^M}(F_i) - P[(X_1^M, Y_1^M) \in E_i]| \leqslant 2^j\varepsilon_M, i = 1,\cdots,j$;

(k) $\tau^{-1}\{i\} \supset F_i, i = 1,\cdots,j$.

From (k) we may choose a function $f: A^M \to B^M$ such that $(x,f(x)) \in E_i$ if $x \in F_i, i = 1,\cdots,j$. Note that for $i = 1,\cdots,j$,

$$\begin{aligned} &|P[(X_1^M, Y_1^M) \in E_i] - P[(X_1^M, Y_1^M) \in A_i^{(M)}]| \\ &\qquad \leqslant \Sigma_{i'=1}^j P[(X_1^M, Y_1^M) \in A_{i'}^{(M)}] - P[(X_1^M, Y_1^M) \in E_{i'}] < \eta/j, \text{by(g)}. \end{aligned}$$

We thus have, from the preceding and (h), (j),

(l) If $x \in F_i$, $(x,f(x))$ is η-typical of α_i, $i = 1,\cdots,j$.

(m) $|P^{X_1^M}(F_i) - w_i| \leqslant 2^j\varepsilon_M + 2\eta/j, i = 1,\cdots,j$.

Since X is aperiodic we may pick $W \in \mathcal{Q}^\infty$ such that

(n) $W, TW,\cdots, T^{M-1}W$ are disjoint;

(o) $|P[X \in W, X_0^{M-1} \in S] - P[X \in W]P[X_0^{M-1} \in S]| \leqslant \eta/Mj, S \subset A^M$;

(p) $P^X[W \cup TW \cup \cdots \cup T^{M-1}W] > 1 - \eta$.

(W can be found using the argument on page 23 of [12].)

Define $\phi : A^\infty \to B^\infty$ to be any stationary code such that

$$\phi(x)_i^{i+M-1} = f(x_i^{i+M-1}), \qquad T^i x \in W, x \in A^\infty, i \in Z.$$

If $b \in A^N \times B^N$, we have

(q) $P[(X_1^N, \phi(X)_1^N) = b] \leqslant \eta + \Sigma_{k=1}^j\{(\alpha_k(b) + \eta)(M - N + 1) + N - 1\}P[X \in W, X_0^{M-1} \in F_k] + MP[\{X_0^{M-1} \notin F_1 \cup \cdots \cup F_j\} \cap \{X \in W\}]$.

To see this, if $x \in A^\infty$, write

$$\psi(x) = \lim_{n\to\infty} n^{-1}\Sigma_{i=1}^n I_{\{b\}}(x_i^{i+N-1}, \phi(x)_i^{i+N-1}),$$

if limit exists. (It exists almost everywhere $[P^X]$.) Now

$$E[\psi(X)] = P[(X_1^N, \phi(X)_1^N) = b].$$

In the summation from $i = 1$ to $i = n$ defining $\psi(x)$, divide the summation into three parts. First, sum over subintervals $[s, s + M - 1]$ of $[1,n]$ where $T^s x \in W$, $x_s^{s+M-1} \in F_k$ for some $1 \leqslant k \leqslant j$. Then let $n \to \infty$, and take the expected value. We get a contribution to $E[\psi(X)]$ no bigger than the middle term on the righthand side of (q). Second, sum over subintervals $[s, s + M - 1]$ with $T^s x \in W$, $x_s^{s+M-1} \notin F_1 \cup \cdots \cup F_k$. This gives rise to a term in $E[\psi(X)]$ no bigger than the third and last term on the righthand side of (q). Finally, sum over all $i \in [1,n]$ lying in no subinterval $[s,s+M-1]$ with $T^s x \in W$. This gives a contribution to $E[\psi(X)]$ at most η, the first term on the righthand side of (q).

Using (o), (m) and (i), it is a simple matter to show that the righthand side of (q) is no bigger than

$$8\eta + 2j2^j\varepsilon_M + 2N/M + \Sigma_{k=1}^{j} w_k \alpha_k(b).$$

By similar reasoning to that used in (q), we get the lower bound

(r) $P[(X_1^N, \phi(X_1^N)) = b] \geqslant \Sigma_{k=1}^{j}(\alpha_k(b) - \eta)(M - N + 1)P[X_0^{M-1} \in F_k, X \in W]$.

Using (o), (p), and (m) to lower bound the right-hand side of (r) and combining this with our upper bound, we wind up with, as the reader may easily verify,

$$|P[(X_1^N, \phi(X_1^M)) = b] - P[(X_1^N, Y_1^N) = b]| \leqslant 8\eta + 2j2^j\varepsilon_M + 2N/M.$$

Since X is aperiodic, $\varepsilon_M \to 0$ as $M \to \infty$. Hence if M is chosen large enough and η small enough, we will have

$$\Sigma_{b \in A^N \times B^N} |P^{(X_1^N, \phi(X)_1^N)}(b) - P^{(X_1^N, Y_1^N)}(b)| < \delta.$$

REFERENCES

[1] Gray, R. M. (1975). Sliding-block source coding. *IEEE Trans. Information Theory* **21** 357-368.

[2] Gray, R. M. and Ornstein, D. S. (1979). Block coding for discrete stationary $\bar{d}$-continuous noisy channels. *IEEE Trans. Information Theory*. **25** 292-306..

[3] 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*. **4** 639-674..

[4] Kieffer, J. C. (1974a). A general formula for the capacity of stationary nonanticipatory channels. *Information and Control* **26** 381-391.

[5] Kieffer, J. C. (1974b). On the approximation of stationary measures by periodic and ergodic measures. *Ann. Probability* **2** 530-534.

[6] Kieffer, J. C. (1977). On sliding block coding for transmission of a source over a stationary nonanticipatory channel. *Information and Control* **35** 1-19.

[7] Moser, J., Phillips, E. and Varadhan, S. (1975). *Ergodic Theory, A Seminar*. Courant Institute of Mathematical Sciences, New York Univ.

[8] Nedoma, J. (1957). The capacity of a discrete channel. *Trans. First Prague Conf. Information Theory*. 143-182, Prague.

[9] Nedoma, J. (1960). On nonergodic channels. *Trans. Second Prague Conf. Information Theory*. 363-395, Prague.

[10] Ornstein, D. S. (1973). An application of ergodic theory to probability theory. *Ann. Probability*, **1** 43-58.

[11] Ornstein, D. S. (1974). *Ergodic Theory, Randomness, and Dynamical Systems*, Yale Univ. Press, New Haven, Conn.

[12] Shields, P. C. (1973). *The Theory of Bernoulli Shifts*. Univ. Chicago Press.

[13] Shields, P. C. (1979). Almost block independence. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **49** 119-123.

[14] Ash, R. B. (1965). *Information Theory*. Wiley, New York.

[15] Parthasarathy, K. R. (1963). Effective entropy rate and transmission of information through channels with additive random noise. *Sankhyā, Ser. A* **25** 75-84.

Department of Mathematics
University of Missouri-Rolla
Rolla, Missouri 65401