

SOME TOPOLOGIES ON THE SET OF DISCRETE STATIONARY CHANNELS

JOHN C. KIEFFER

Some topologies are defined on the set \mathfrak{S} of all discrete stationary channels with given finite input and output alphabets, which are weaker than the topology of Neuhoff and Shields arising from the \bar{d} concept of channel distance. The closure of various subsets of \mathfrak{S} with respect to certain of these topologies on \mathfrak{S} are determined. For example, a topology on \mathfrak{S} is introduced with respect to which the set of weakly continuous channels (the most general class of channels for which coding theorems of information theory have been obtained) is the closure in \mathfrak{S} of the set of channels with input finite memory and anticipation. As a by-product, results are obtained on simulating channels by block codes and on constructing sliding-block codes from block codes using sets called coding sets.

I. Introduction.

NOTATION. A, B will denote finite sets to be fixed for the rest of the paper; \mathcal{A}, \mathcal{B} will denote the sets of all subsets of A, B , respectively. $(A^\infty, \mathcal{A}^\infty)$ denotes the measurable space consisting of A^∞ , the set of all doubly-infinite sequences $x = (x_i)_{i=-\infty}^\infty$ from A , and \mathcal{A}^∞ , the usual product σ -field of subsets of A^∞ . We define $(B^\infty, \mathcal{B}^\infty)$ similarly.

Let Z be the set of all integers. If $x \in A^\infty$, and $i \in Z$, x_i denotes the i th coordinate of x . If $m, n \in Z$ and $m \leq n$, x_m^n denotes the $(n - m + 1)$ -tuple (y_1, \dots, y_{n-m+1}) in A^{n-m+1} such that $y_i = x_{m+i-1}$, $1 \leq i \leq n - m + 1$. Similarly if $x \in A^k$, where k is a positive integer, and $1 \leq i \leq k$, x_i denotes the i th coordinate of x , and if $1 \leq m \leq n \leq k$, x_m^n denotes $(y_1, \dots, y_{n-m+1}) = (x_m, \dots, x_n)$. One defines similar notations for $x \in B^\infty$ or B^k , $k \geq 1$.

$\{\tilde{X}_i; i \in Z\}$ denotes the family of projections from $A^\infty \rightarrow A$; that is, $\tilde{X}_i(x) = x_i$, $i \in Z$, $x \in A^\infty$. Similarly, $\{\tilde{Y}_i; i \in Z\}$ denotes the family of projections from $B^\infty \rightarrow B$. $\{X_i; i \in Z\}$ are the maps from $A^\infty \times B^\infty \rightarrow A$ and $\{Y_i; i \in Z\}$ the maps from $A^\infty \times B^\infty \rightarrow B$ such that $X_i(x, y) = \tilde{X}_i(x)$, $Y_i(x, y) = \tilde{Y}_i(y)$, $x \in A^\infty$, $y \in B^\infty$, $i \in Z$. If $m, n \in Z$ and $m \leq n$, \tilde{X}_m^n denotes the map $(\tilde{X}_m, \dots, \tilde{X}_n)$ from $A^\infty \rightarrow A^{n-m+1}$. Similarly, we define $\tilde{Y}_m^n, X_m^n, Y_m^n$, except we denote $X_1^n, Y_1^n, \tilde{X}_1^n, \tilde{Y}_1^n$ respectively by $X^n, Y^n, \tilde{X}^n, \tilde{Y}^n$.

R denotes the real line. If S is a finite set, $|S|$ denotes the cardinality of S . If (Ω, \mathfrak{T}) is a measurable space and $W \in \mathfrak{T}$, W^c denotes the complement of W ; that is, $W^c = \{\omega \in \Omega: \omega \notin W\}$. The map $I_W: \Omega \rightarrow R$ is the indicator function of W —the function equal to 1 on W and 0 on W^c .

A function $f: A^\infty \rightarrow R$ is finite-dimensional (f.d.) if f is a function of \tilde{X}_m^n for some $m \leq n$, $m, n \in Z$. Similarly, we define a f.d. function $f: B^\infty \rightarrow R$. A function $f: A^\infty \times B^\infty \rightarrow R$ is f.d. if f is a function of (X_m^n, Y_m^n) for some $m \leq n$, $m, n \in Z$. A subset E of A^∞ , B^∞ , or $A^\infty \times B^\infty$ is said to be f.d. if I_E is f.d.

$T_A: A^\infty \rightarrow A^\infty$ and $T_B: B^\infty \rightarrow B^\infty$ denote the two-sided shift transformations. (That is, $T_A x = \tilde{x}$, where $\tilde{x}_i = x_{i+1}$, $i \in Z$; similarly for T_B .) We will sometimes denote T_A or T_B by T , when the context makes clear what the domain is.

Sources. \mathfrak{M} denotes the family of all probability measures on \mathcal{A}^∞ stationary with respect to T_A . We call the elements of \mathfrak{M} *stationary sources*. \mathfrak{M}' denotes the family of all probability measures on $\mathcal{A}^\infty \times \mathcal{B}^\infty$. On \mathfrak{M} , \mathfrak{M}' we place the *weak topology* [19]; for example, the weak topology on \mathfrak{M} is the unique metrizable topology such that if $\{\mu_n\}_{n=1}^\infty \subset \mathfrak{M}$ and $\mu \in \mathfrak{M}$, then $\mu_n \rightarrow \mu$ in the weak topology if and only if $\mu_n(E) \rightarrow \mu(E)$ for every f.d. $E \subset A^\infty$. \mathfrak{M} , \mathfrak{M}' are compact topological spaces.

A sequence $x \in A^\infty$ is *periodic* if $T^N x = x$ for some positive integer N . If x is periodic, let N be the smallest positive integer N such that $T^N x = x$. N is called the *period* of x . We let S_p denote the set of all the periodic sequences in A^∞ . We say $\mu \in \mathfrak{M}$ is *periodic* if $\mu(S_p) = 1$. We say a periodic $\mu \in \mathfrak{M}$ has *order* N if μ places probability one on the set of all periodic sequences of period $\leq N$. We let \mathfrak{M}_p denote the subset of \mathfrak{M} consisting of the periodic sources. We say $\mu \in \mathfrak{M}$ is *aperiodic* if $\mu(S_p) = 0$. \mathfrak{M}_a will denote the subset of \mathfrak{M} consisting of aperiodic sources. We let $\mathfrak{E} \subset \mathfrak{M}$ denote the family of all *ergodic sources*; that is, $\mu \in \mathfrak{E}$ if and only if μ is ergodic relative to T_A .

Channels. We say ν is a *channel* if for some pair of measurable spaces $(\Omega_1, \mathfrak{T}_1)$, $(\Omega_2, \mathfrak{T}_2)$, ν is a family $\nu = \{\nu_x: x \in \Omega_1\}$ of probability measures on \mathfrak{T}_2 such that for each $E \in \mathfrak{T}_2$, the map $x \rightarrow \nu_x(E)$ from $\Omega_1 \rightarrow [0, 1]$ is \mathfrak{T}_1 -measurable. We call Ω_1 the *input space* of the channel ν , and Ω_2 the *output space*. We will be dealing with *discrete finite-alphabet channels*; that

is, channels whose input and output spaces are spaces of doubly-infinite sequences drawn from finite sets.

If M is a positive integer, a channel ν with input space $(A^\infty, \mathcal{Q}^\infty)$ and output space $(B^\infty, \mathfrak{B}^\infty)$ is said to be M -stationary if $\nu_{T_A^M x}(T_B^M E) = \nu_x(E)$, $x \in A^\infty$, $E \in \mathfrak{B}^\infty$. We call a 1-stationary channel simply a *stationary channel*. Let \mathfrak{S} denote the set of all stationary channels with input space $(A^\infty, \mathcal{Q}^\infty)$ and output space $(B^\infty, \mathfrak{B}^\infty)$. Following [3], a channel $\nu \in \mathfrak{S}$ is said to have *input finite memory and anticipation* if for some positive integer M , the following holds: For each $k = 1, 2, \dots$, and $E \subset B^{2k+1}$, the map $x \rightarrow \nu_x(\tilde{Y}_{-k}^k \in E)$ is a f.d. function from $A^\infty \rightarrow [0, 1]$, depending on x only through x_{-k-M}^{k+M} . We let \mathfrak{S}_f denote the subset of \mathfrak{S} consisting of all such channels.

By a *stationary code*, we mean a measurable map $\psi: A^\infty \rightarrow B^\infty$ such that $\psi \cdot T_A = T_B \cdot \psi$. Given a stationary code $\psi: A^\infty \rightarrow B^\infty$, let ν^ψ denote the channel in \mathfrak{S} such that $\nu_x^\psi(E) = I_E(\psi(x))$, $E \in \mathfrak{B}^\infty$, $x \in A^\infty$. We let \mathfrak{S}_d denote the subset of \mathfrak{S} such that $\mathfrak{S}_d = \{\nu^\psi: \psi \text{ a stationary code}\}$.

The elements of \mathfrak{S}_d are called *deterministic channels*. We let \mathfrak{S}_{df} denote $\mathfrak{S}_d \cap \mathfrak{S}_f$. A stationary code $\psi: A^\infty \rightarrow B^\infty$ is called a *sliding-block code* ([2], [20]) if for some positive integer k , there exists $\psi': A^{2k+1} \rightarrow B$ such that $\psi(x)_i = \psi'(x_{i-k}^{i+k})$, $x \in A^\infty$, $i \in \mathbb{Z}$. It is easy to see that $\mathfrak{S}_{df} = \{\nu^\psi: \psi \text{ a sliding-block code}\}$.

Given a channel ν with input space $(A^\infty, \mathcal{Q}^\infty)$ and output space $(B^\infty, \mathfrak{B}^\infty)$, and a probability measure μ on \mathcal{Q}^∞ , we let $\mu\nu$ denote the probability measure on $\mathcal{Q}^\infty \times \mathfrak{B}^\infty$ such that $\mu\nu(E \times F) = \int_E \nu_x(F) d\mu(x)$, $E \in \mathcal{Q}^\infty$, $F \in \mathfrak{B}^\infty$. We define $\nu \in \mathfrak{S}$ to be *weakly continuous* if the map $\mu \rightarrow \mu\nu$ from $\mathfrak{M} \rightarrow \mathfrak{M}'$ is continuous. We let \mathfrak{S}_c denote the subset of \mathfrak{S} consisting of the weakly continuous channels. \mathfrak{S}_c is the most general class of channels for which coding theorems of information theory have been obtained [7] [9] [10] [11]. In §III of this paper, we provide a characterization of such channels.

Channel Topologies. In [14] Neuhoff and Shields introduced the \bar{d} concept of channel distance, which is a pseudometric on \mathfrak{S} , and determined the closure with respect to the resulting topology on \mathfrak{S} of a simple class of channels called primitive channels; subsequently, in [15] [16] they determined the closure of some interesting subclasses of the class

of primitive channels. In the same spirit, we define here some topologies on \mathfrak{S} weaker than theirs and then devote the rest of the paper to investigating the closure of $\mathfrak{S}_f, \mathfrak{S}_d, \mathfrak{S}_{df}$ under these topologies.

Let \mathfrak{D} be a subset of \mathfrak{M} . We define $\mathfrak{T}_p(\mathfrak{D})$ to be the topology on \mathfrak{S} whose basis consists of the sets of form $E(\mathcal{G}, F, \nu_0, \epsilon) = \{\nu \in \mathfrak{S} : \sup_{E \in \mathcal{G}, \mu \in F} |\mu\nu(E) - \mu\nu_0(E)| < \epsilon\}$, for all $\epsilon > 0, \nu_0 \in \mathfrak{S}$, finite collections \mathcal{G} of f.d. sets from $\mathcal{A}^\infty \times \mathcal{B}^\infty$, and finite subsets F of \mathfrak{D} . If $\{\nu_n\}_{n=1}^\infty \subset \mathfrak{S}$ and $\nu \in \mathfrak{S}$, then $\nu_n \rightarrow \nu$ in the $\mathfrak{T}_p(\mathfrak{D})$ topology if and only if $\mu\nu_n \rightarrow \mu\nu$ for every $\mu \in \mathfrak{D}$, or equivalently, if and only if $\mu\nu_n(E) \rightarrow \mu\nu(E)$ for every $\mu \in \mathfrak{D}$ and f.d. set E from $\mathcal{A}^\infty \times \mathcal{B}^\infty$. Consequently, we call the topology $\mathfrak{T}_p(\mathfrak{D})$ the *topology of pointwise convergence on \mathfrak{D}* .

We define $\mathfrak{T}_u(\mathfrak{D})$ to be the topology on \mathfrak{S} whose basis consists of sets of form $E(\mathcal{G}, \nu_0, \epsilon) = \{\nu \in \mathfrak{S} : \sup_{E \in \mathcal{G}, \mu \in \mathfrak{D}} |\mu\nu(E) - \mu\nu_0(E)| < \epsilon\}$, for all $\epsilon > 0, \nu_0 \in \mathfrak{S}$, and finite collections \mathcal{G} of f.d. sets from $\mathcal{A}^\infty \times \mathcal{B}^\infty$. It is easy to see that $\nu_n \rightarrow \nu$ in the $\mathfrak{T}_u(\mathfrak{D})$ topology if and only if for each f.d. set E in $\mathcal{A}^\infty \times \mathcal{B}^\infty, \mu\nu_n(E) \rightarrow \mu\nu(E)$ uniformly in $\mu \in \mathfrak{D}$. Accordingly we call $\mathfrak{T}_u(\mathfrak{D})$ the *topology of uniform convergence on \mathfrak{D}* .

If $\mathfrak{S}' \subset \mathfrak{S}$ and \mathfrak{T} is a topology on \mathfrak{S} , let $\mathcal{C}[\mathfrak{S}' | \mathfrak{T}]$ denote the closure of \mathfrak{S}' relative to the topology \mathfrak{T} . Let $\mathcal{C}_s[\mathfrak{S}' | \mathfrak{T}]$ denote the *sequential closure* of \mathfrak{S}' relative to \mathfrak{T} ; that is, $\nu \in \mathcal{C}_s[\mathfrak{S}' | \mathfrak{T}]$ if and only if there is a sequence $\{\nu_n\}_{n=1}^\infty$ from \mathfrak{S}' such that $\nu_n \rightarrow \nu$ in the \mathfrak{T} -topology. In general $\mathcal{C}_s[\mathfrak{S}' | \mathfrak{T}] \subset \mathcal{C}[\mathfrak{S}' | \mathfrak{T}]$, with equality if \mathfrak{T} is first countable. The uniform topologies $\mathfrak{T}_u(\mathfrak{D})$ are first countable and so, $\mathcal{C}_s[\mathfrak{S}' | \mathfrak{T}_u(\mathfrak{D})] = \mathcal{C}[\mathfrak{S}' | \mathfrak{T}_u(\mathfrak{D})]$. The topologies $\mathfrak{T}_p(\mathfrak{D})$ may not be first countable if \mathfrak{D} is not countable, and so it is possible one may have $\mathcal{C}_s[\mathfrak{S}' | \mathfrak{T}_p(\mathfrak{D})] \neq \mathcal{C}[\mathfrak{S}' | \mathfrak{T}_p(\mathfrak{D})]$ for such topologies.

Coding Sets. If N is a positive integer, $E \subset A^\infty$ is called a *coding set of order N* if E is f.d. and does not contain any periodic sequence of period $\leq N$. $E \subset A^\infty$ is called a *coding set* if E is a coding set of order N for some N . Coding sets will play a role in constructing sliding-block codes, as we will see in §V.

Let $\mathfrak{D} \subset \mathfrak{M}$. We say \mathfrak{D} is *concentrated on coding sets* if there is a sequence of coding sets $\{F_N\}_{N=1}^\infty$ such that for each N, F_N is of order N , and for each $\mu \in \mathfrak{D}, \lim_{N \rightarrow \infty} \mu(F_N) = 1$.

We say \mathfrak{D} is *uniformly concentrated on coding sets* if for each $\varepsilon > 0$ and positive integer N , there exists a coding set of order N such that $\mu(F_N) > 1 - \varepsilon$ for all $\mu \in \mathfrak{D}$.

It is important to know when \mathfrak{D} is concentrated on coding sets or uniformly concentrated on coding sets in view of results to be obtained later in the paper (see §II, P3 and P4).

II. Statement of main results. The main results to be obtained in this paper are the following

Closure Properties of $\mathfrak{S}_f, \mathfrak{S}_d, \mathfrak{S}_{df}$

P1. $\mathcal{C}[\mathfrak{S}_f | \mathfrak{T}_p(\mathfrak{N})] = \mathfrak{S}$.

P2. $\mathcal{C}[\mathfrak{S}_f | \mathfrak{T}_u(\mathfrak{N})] = \mathfrak{S}_c$.

P3. $\mathcal{C}[\mathfrak{S}_{df} | \mathfrak{T}_u(\mathfrak{D})] \supset \mathfrak{S}_c$ if $\mathfrak{D} \subset \mathfrak{N}$ is uniformly concentrated on coding sets.

P4. $\mathcal{C}[\mathfrak{S}_{df} | \mathfrak{T}_p(\mathfrak{D})] \supset \mathfrak{S}_c$ if $\mathfrak{D} \subset \mathfrak{N}$ is concentrated on coding sets.

P5. $\mathcal{C}[\mathfrak{S}_{df} | \mathfrak{T}_p(\mathfrak{N}_a)] = \mathfrak{S}$.

P6. $\mathcal{C}[\mathfrak{S}_d | \mathfrak{T}_u(\mathfrak{N}_a)] = \mathfrak{S}$.

Properties P1, P5, P6 show that $\mathfrak{S}_f, \mathfrak{S}_d, \mathfrak{S}_{df}$ are dense in \mathfrak{S} if one chooses the right topology on \mathfrak{S} .

Property P2 gives a characterization of weakly continuous channels. Properties P3, P4 allow one to approximate weakly continuous channels by channels from \mathfrak{S}_{df} .

Property P1 is proved at the end of this section. Property P2 is proved in §III. Properties 3 and 4 follow from Theorem 2 of §V. Property P5 follows from Properties P1 and P3, as indicated at the end of §VI. Property P6 follows from Property P5 with the help of Lemma A3 of the Appendix.

Proof of P1. Let $\nu \in \mathfrak{S}$, $\varepsilon > 0$, a positive integer j , and a finite subset $\{\mu_1, \dots, \mu_n\}$ of \mathfrak{N} be specified. P1 follows provided we can find $\hat{\nu} \in \mathfrak{S}_f$ such that

$$(2.1) \quad \sup_{a \in A^j, b \in B^j} \sup_{1 \leq i \leq n} |\mu_i \hat{\nu}[X^j = a, Y^j = b] - \mu_i \nu[X^j = a, Y^j = b]| < \varepsilon.$$

Choose $k > j$ so large that $2j/k < \varepsilon/2$. Choose $\delta > 0$ so small that $n\delta |A|^k |B|^k < \varepsilon/2$. Let $\mu = n^{-1}(\mu_1 + \dots + \mu_n)$. Since f.d. functions are dense in $L^1(\mu)$, there is a channel ν' with input space A^∞ and output space

B^k such that for each $b \in B^k$ the map $x \rightarrow \nu'_x(b)$ from $A^\infty \rightarrow [0, 1]$ is f.d. and

$$(2.2) \quad \int_{A^\infty} |\nu'_x(b) - \nu_x(\tilde{Y}^k = b)| d\mu(x) < \delta, \quad b \in B^k.$$

By Lemma A2 of the Appendix, there exists $\hat{\nu} \in \mathfrak{S}_f$ such that for all $a \in A^j, b \in B^j$

$$(2.3) \quad \begin{aligned} & |\mu_i \hat{\nu}[X^j = a, Y^j = b] - \mu_i \nu[X^j = a, Y^j = b]| \\ & \leq 2j/k + \sum_{a' \in A^k, b' \in B^k} \left| \int_{\{\tilde{X}^k = a'\}} \{ \nu'_x(b') - \nu_x(\tilde{Y}^k = b') \} d\mu_i(x) \right|, \\ & \hspace{25em} i = 1, \dots, n. \end{aligned}$$

By (2.2), we have for $a' \in A^k, b' \in B^k$, and all i ,

$$\begin{aligned} & \left| \int_{\{\tilde{X}^k = a'\}} \{ \nu'_x(b') - \nu_x(\tilde{Y}^k = b') \} d\mu_i(x) \right| \\ & \leq n \int_{A^\infty} |\nu'_x(b') - \nu_x(\tilde{Y}^k = b')| d\mu(x) < n\delta. \end{aligned}$$

Hence the right-hand side of inequality (2.3) is upper bounded by $2j/k + |A|^k |B|^k n\delta < \varepsilon$. (2.1) now follows.

III. A Characterization of weakly continuous channels. In this section we prove P2 which gives the following characterization of weakly continuous channels: A channel $\nu \in \mathfrak{S}$ is weakly continuous if and only if there is a sequence $\{\nu_n\} \subset \mathfrak{S}_f$ such that $\nu_n \rightarrow \nu$ in the $\mathfrak{T}_u(\mathfrak{M})$ topology. We accomplish parts of the proof by means of the two lemmas which follow.

LEMMA 1. *Let $W \subset A^\infty$ be f.d. Let $f: A^\infty \rightarrow [0, 1]$ be a measurable function such that the map $\mu \rightarrow \int_W f d\mu$ from $\mathfrak{M} \rightarrow [0, 1]$ is continuous. Then, given $\varepsilon > 0$, there is a f.d. function $g: A^\infty \rightarrow R$ such that*

$$(3.1) \quad \left| \int_W g d\mu - \int_W f d\mu \right| < \varepsilon, \quad \mu \in \mathfrak{M}.$$

Proof. We place the discrete topology on A and the product topology on A^∞ , making A^∞ a compact topological space. Let $C(A^\infty)$ be the Banach space of all continuous functions from $A^\infty \rightarrow R$, under the supremum norm. By the Stone-Weierstrass Theorem [1, Theorem A7.5], the f.d. functions are dense in $C(A^\infty)$. Thus we need only obtain a $g \in C(A^\infty)$ such that (3.1) holds.

Let \mathcal{L} be the space of all finite signed measures on \mathcal{Q}^∞ , with the weak topology. If $\mu \in \mathcal{L}$, let $\mu_W \in \mathcal{L}$ be the measure such that

$$\mu_W(E) = \mu(E \cap W), \quad E \in \mathcal{Q}^\infty.$$

Let $\mathcal{L}_W = \{\mu_W: \mu \in \mathcal{L}\}$, $\mathfrak{N}_W = \{\mu_W: \mu \in \mathfrak{N}\}$. \mathcal{L}_W , with the topology it inherits as a subspace of \mathcal{L} , is a locally convex space, and \mathfrak{N}_W is a compact convex subset. Hence any continuous affine functional on \mathfrak{N}_W is the uniform limit on \mathfrak{N}_W of continuous linear functionals on \mathcal{L}_W [13, Theorem T.6, p. 221]. Thus there is a continuous linear functional $G: \mathcal{L}_W \rightarrow R$ such that

$$(3.2) \quad \left| G(\mu) - \int_W f d\mu \right| < \varepsilon, \quad \mu \in \mathfrak{N}_W.$$

Let $G^*: \mathcal{L} \rightarrow R$ be the linear functional such that

$$(3.3) \quad G^*(\mu) = G(\mu_W), \quad \mu \in \mathcal{L}.$$

G^* is continuous since it is the composition of the continuous map $\mu \rightarrow \mu_W$ from $\mathcal{L} \rightarrow \mathcal{L}_W$ with G . Now by the Riesz Representation Theorem [1, Thm. 4.3.13], \mathcal{L} can be identified with the dual space of $C(A^\infty)$ under the weak-star topology. By [1, Thm. 3.5.17], since G^* is a weak-star continuous functional, there must be $g \in C(A^\infty)$ such that

$$(3.4) \quad G^*(\mu) = \int_{A^\infty} g d\mu, \quad \mu \in \mathcal{L}.$$

If $\mu \in \mathfrak{N}$, then $\mu_W \in \mathfrak{N}_W$ and so by (3.2)–(3.4)

$$(3.5) \quad \left| G(\mu_W) - \int_W f d\mu_W \right| = \left| \int_{A^\infty} g d\mu - \int_W f d\mu_W \right| < \varepsilon.$$

Replacing μ in (3.5) by μ_W gives

$$(3.6) \quad \left| \int_{A^\infty} g d\mu_W - \int_W f d(\mu_W)_W \right| < \varepsilon, \quad \mu \in \mathfrak{N}.$$

But $\int g d\mu_W = \int_W g d\mu$ and $\int_W f d(\mu_W)_W = \int_W f d\mu$, so (3.1) follows from (3.6).

LEMMA 2. *Let $W \subset A^\infty$ be f.d. Let $\delta > 0$. Let f_1, \dots, f_n be f.d. functions from $A^\infty \rightarrow R$ such that*

- (a) $-\delta/n \leq \int_W f_i d\mu \leq \mu(W) + \delta/n, i = 1, \dots, n, \mu \in \mathfrak{N}.$
- (b) $\mu(W) - \delta \leq \int_W (\sum_{i=1}^n f_i) d\mu \leq \mu(W) + \delta, \mu \in \mathfrak{N}.$

Then there are f.d. functions $g_1, \dots, g_n: A^\infty \rightarrow R$ such that

- (a') $|\int_W f_i d\mu - \int_W g_i d\mu| < 6\delta, \mu \in \mathfrak{N}, i = 1, \dots, n.$
- (b') $\sum_{i=1}^n g_i = 1$ throughout $A^\infty.$
- (c') $g_i \geq 0$ throughout $A^\infty, i = 1, \dots, n.$

Proof. From (a), (b) and Lemma A1 of the Appendix we see that for k sufficiently large,

$$(3.7) \quad k^{-1} \sum_{j=0}^{k-1} [(f_i - 1)I_W] \cdot T^j \leq 2\delta/n, \quad i = 1, \dots, n.$$

$$(3.8) \quad k^{-1} \sum_{j=0}^{k-1} (f_i I_W) \cdot T^j \geq -2\delta/n, \quad i = 1, \dots, n.$$

$$(3.9) \quad -2\delta \leq k^{-1} \sum_{j=0}^{k-1} \left[\left(1 - \sum_{i=1}^n f_i \right) I_W \right] \cdot T^j \leq 2\delta.$$

Let $\tilde{W} = \{x \in W: T^i x \in W \text{ for infinitely many } i > 0\}$. By Poincare's Recurrence Theorem [4],

$$(3.10) \quad \mu(\tilde{W}) = \mu(W), \quad \mu \in \mathfrak{N}.$$

Define $\phi: \tilde{W} \rightarrow \{1, 2, \dots\}$ to be the map such that

$$\phi(x) = \inf\{i > 0: T^i x \in W\}, \quad x \in \tilde{W}.$$

Thus, ϕ is the time of first return to W . Define $\tilde{T}: \tilde{W} \rightarrow \tilde{W}$ as follows:

$$\tilde{T}x = T^{\phi(x)}x, \quad x \in \tilde{W}.$$

As is well-known [5], \tilde{T} is measure-preserving; that is

$$(3.11) \quad \int_{\tilde{W}} f \cdot \tilde{T} d\mu = \int_{\tilde{W}} f d\mu,$$

for every measurable $f: \tilde{W} \rightarrow [0, \infty)$, and every $\mu \in \mathfrak{N}$. Also, by Kac's Recurrence Theorem [4],

$$(3.12) \quad \int_{\tilde{W}} \phi d\mu \leq 1, \quad \mu \in \mathfrak{N}.$$

For any $f: A^\infty \rightarrow R$ it is easily seen that

$$\sum_{i=0}^{k-1} f(\tilde{T}^i x) = \sum_{i=0}^{\phi(x) + \dots + \phi(\tilde{T}^{k-1}x)} (fI_W)(T^i x), \quad x \in \tilde{W}.$$

Applying this to (3.7)–(3.9), it is clear that we may fix a positive integer k so that for every $x \in \tilde{W}$,

$$(3.13) \quad \sum_{j=0}^{k-1} [f_i - 1](\tilde{T}^j x) \leq 2\delta/n \sum_{j=0}^{k-1} \phi(\tilde{T}^j x), \quad i = 1, \dots, n.$$

$$(3.14) \quad \sum_{j=0}^{k-1} f_i(\tilde{T}^j x) \geq -2\delta/n \sum_{j=0}^{k-1} \phi(\tilde{T}^j x), \quad i = 1, \dots, n.$$

$$(3.15) \quad -2\delta \sum_{j=0}^{k-1} \phi(\tilde{T}^j x) \leq \sum_{j=0}^{k-1} \left[1 - \sum_{i=1}^n f_i \right](\tilde{T}^j x) \leq 2\delta \sum_{j=0}^{k-1} \phi(\tilde{T}^j x).$$

Let $f: \tilde{W} \rightarrow R$ be the function such that $f(x) = k^{-1} \sum_{j=0}^{k-1} \phi(\tilde{T}^j x)$, $x \in \tilde{W}$. For each $i = 1, \dots, n$, let $h_i: \tilde{W} \rightarrow R$ be the function such that $h_i(x) = k^{-1} \sum_{j=0}^{k-1} f_i(\tilde{T}^j x)$, $x \in \tilde{W}$. Then (3.13)–(3.15) reduce to

$$(3.16) \quad h_i \leq 1 + 2\delta f/n, \quad i = 1, \dots, n.$$

$$(3.17) \quad h_i \geq -2\delta f/n, \quad i = 1, \dots, n.$$

$$(3.18) \quad 1 - 2\delta f \leq \sum_{i=1}^n h_i \leq 1 + 2\delta f.$$

For each $i = 1, \dots, n$, let $h_i^*: \tilde{W} \rightarrow R$ be the function such that

$$h_i^* = \frac{h_i + 2\delta n^{-1}f}{\sum_{i'=1}^n (h_{i'} + 2\delta n^{-1}f)}.$$

Then,

$$(3.19) \quad h_i^* \geq 0 \text{ for each } i \quad \text{and} \quad \sum_{i'=1}^n h_{i'}^* = 1.$$

Now by (3.18),

$$1 \leq \sum_{i=1}^n (h_i + 2\delta n^{-1}f) \leq 1 + 4\delta f,$$

and so for each i

$$(3.20) \quad \frac{h_i}{1 + 4\delta f} \leq h_i^* \leq h_i + 2\delta n^{-1}f.$$

Now by (3.16), (3.17),

$$(3.21) \quad |h_i| \leq 1 + 4\delta f \quad \text{for each } i.$$

Hence by (3.12), (3.21) we have for all i and all $\mu \in \mathfrak{N}$,

$$\left| \int_{\tilde{W}} h_i d\mu - \int_{\tilde{W}} \frac{h_i}{1 + 4\delta f} d\mu \right| \leq 4\delta \int_{\tilde{W}} \frac{|h_i|f}{1 + 4\delta f} d\mu \leq 4\delta \int_{\tilde{W}} f d\mu \leq 4\delta.$$

From this and (3.20), we get

$$-4\delta + \int_{\tilde{W}} h_i d\mu \leq \int_{\tilde{W}} h_i^* d\mu \leq \int_{\tilde{W}} h_i d\mu + 2\delta.$$

But $\int_{\tilde{W}} h_i d\mu = \int_{\tilde{W}} \tilde{f}_i d\mu$ by (3.11) and so

$$(3.22) \quad \left| \int_{\tilde{W}} h_i^* d\mu - \int_{\tilde{W}} \tilde{f}_i d\mu \right| \leq 4\delta \quad \text{for all } i, \mu.$$

Now the $\{h_i^*\}$ may not be f.d., so we modify them appropriately. If $1 \leq r \leq s$, define

$$E_{rs} = \{x \in W: T^i x \in W \text{ for at least } r \text{ indices } i \text{ in the range } 1 \leq i \leq s\}.$$

We may fix positive integers t and r so large that for all $s \geq r$,

$$(3.23) \quad h_i^* \text{ is constant on } \tilde{W} \cap E_{rs} \cap \{\tilde{X}_{-t}^s = a\}, \\ a \in A^{s+t+1}, i = 1, \dots, n.$$

From (3.10), we have

$$\mu(W \cap E_{rs}^c) \downarrow 0 \text{ as } s \rightarrow \infty, \quad \mu \in \mathfrak{N}.$$

Since $W \cap E_{rs}^c$ is f.d., the map $\mu \rightarrow \mu(W \cap E_{rs}^c)$ from $\mathfrak{N} \rightarrow [0, 1]$ is continuous. Hence, by Dini's Theorem [1, p. 181], we may fix s so large that

$$\mu(W \cap E_{rs}^c) < \delta, \quad \mu \in \mathfrak{N}.$$

Now by (3.19), (3.23) we may select f.d. $g_1, \dots, g_n: A^\infty \rightarrow R$ such that (b'), (c') hold and

$$g_i = h_i^* \quad \text{on } \tilde{W} \cap E_{rs}, \quad i = 1, \dots, n.$$

Then for all i, μ ,

$$\left| \int_{\tilde{W}} g_i d\mu - \int_{\tilde{W}} h_i^* d\mu \right| \leq 2\mu(\tilde{W} \cap E_{rs}^c) < 2\delta.$$

This, coupled with (3.22), gives (a').

Proof of P2. Suppose ν is in the $\mathfrak{T}_u(\mathfrak{N})$ closure of \mathfrak{S}_f . Since $\mathfrak{T}_u(\mathfrak{N})$ is first countable, this implies the existence of a sequence $\{\nu_n\} \subset \mathfrak{S}_f$ such that

$\nu_n \rightarrow \nu$ in the $\mathfrak{T}_u(\mathfrak{M})$ topology. Fix a f.d. set $E \subset A^\infty \times B^\infty$. Let $\Phi: \mathfrak{M} \rightarrow R$ be the map $\Phi(\mu) = \mu\nu(E)$, $\mu \in \mathfrak{M}$. It will follow that $\nu \in \mathcal{S}_c$ if Φ is continuous. For each n , let $\Phi_n: \mathfrak{M} \rightarrow R$ be the map such that $\Phi_n(\mu) = \mu\nu_n(E)$, $\mu \in \mathfrak{M}$. Then $\Phi_n \rightarrow \Phi$ uniformly on \mathfrak{M} . Since each Φ_n is continuous, Φ must be continuous.

Conversely, suppose $\nu \in \mathcal{S}_c$. Fix $\epsilon > 0$ and a positive integer j . We will have ν in the $\mathfrak{T}_u(\mathfrak{M})$ closure of \mathcal{S}_j if we can find $\hat{\nu} \in \mathcal{S}_j$ such that

$$(3.24) \quad \sup_{a \in A^j, b \in B^j} \sup_{\mu \in \mathfrak{M}} |\mu\hat{\nu}[X^j = a, Y^j = b] - \mu\nu[X^j = a, Y^j = b]| < \epsilon.$$

Fix $k > j$ so large that $2j/k < \epsilon/2$. Choose $\delta > 0$ so that $7\delta|A|^k|B|^k < \epsilon/2$.

By Lemma 1 pick for each $a \in A^k$ f.d. functions $\{f_a^b: b \in B^k\}$ such that for all b, μ

$$(3.25) \quad \left| \int_{\{\tilde{X}^k = a\}} f_a^b(x) d\mu(x) - \int_{\{\tilde{X}^k = a\}} \nu_x(\tilde{Y}^k = b) d\mu(x) \right| < \delta/|B|^k.$$

To ease the notation, denote the set $\{\tilde{X}^k = a\}$ by $[a]$. Then from (3.25) we obtain for $a \in A^k, b \in B^k, \mu \in \mathfrak{M}$,

$$\begin{aligned} -\delta/|B|^k &\leq \int_{[a]} f_a^b d\mu \leq \mu[a] + \delta/|B|^k, \\ \mu[a] - \delta &\leq \int_{[a]} \left(\sum_{b \in B^k} f_a^b \right) d\mu \leq \mu[a] + \delta. \end{aligned}$$

Thus, by Lemma 2, we may select for each $a \in A^k$ f.d. functions $\{g_a^b: b \in B^k\}$ such that

$$(3.26) \quad \left| \int_{[a]} g_a^b d\mu - \int_{[a]} f_a^b d\mu \right| < 6\delta, \quad \text{for all } b, \mu;$$

$$(3.27) \quad g_a^b \geq 0, \quad \text{for all } b;$$

$$(3.28) \quad \sum_{b \in B^k} g_a^b = 1.$$

Let ν' be the channel with input space A^∞ and output space B^k such that

$$(3.29) \quad \nu'_x(b) = g_a^b(x), \quad x \in [a], b \in B^k.$$

By Lemma A2 of the Appendix, there exists $\hat{\nu} \in \mathfrak{S}_f$ such that for all $\mu \in \mathfrak{N}$, $a \in A^j$, $b \in B^j$,

$$(3.30) \quad \left| \mu \nu [X^j = a, Y^j = b] - \mu \hat{\nu} [X^j = a, Y^j = b] \right| \\ \leq 2j/k + \sum_{a' \in A^k, b' \in B^k} \left| \int_{\{\tilde{X}^k = a'\}} \{ \nu'_x(b') - \nu_x(\tilde{Y}^k = b') \} d\mu(x) \right|.$$

Now by (3.29), (3.26), (3.25)

$$\left| \int_{\{\tilde{X}^k = a'\}} \{ \nu'_x(b') - \nu_x(\tilde{Y}^k = b') \} d\mu(x) \right| < 7\delta, \quad \text{for all } a', b', \mu.$$

Hence the right-hand side of (3.30) is upper bounded by $2j/k + 7\delta |A|^k |B|^k < \varepsilon$. This gives (3.24).

IV. Simulating a channel by block codes. A *block code* is a map $\phi: A^N \rightarrow B^N$ for some N . In this section we show that block codes can be used to simulate certain channels. We now make precise what we mean by this.

In the rest of the paper, if u, v are two elements drawn from the same set, we define $\delta(u, v)$ to be 0 if $u \neq v$ and 1 if $u = v$.

Let K, M be positive integers and ν a channel with input space $A^{2M+2K+1}$ and output space B^{2K+1} . If $N > 2K + 2M$, $\varepsilon > 0$, and $(x, y) \in A^N \times B^N$, we say (x, y) is ε -typical of ν if

$$\left| \sum_{j=K+M+1}^{N-K-M} \delta(x_{j-K-M}^{j+K+M}, a) \delta(y_{j-K}^{j+K} = b) - \nu_a(b) \sum_{j=K+M+1}^{N-K-M} \delta(x_{j-K-M}^{j+K+M}, a) \right| \\ < (N - 2K - 2M)\varepsilon, \quad \text{for all } a \in A^{2K+2M+1}, b \in B^{2K+1}.$$

We say ν can be *simulated by block codes* if for each $\varepsilon > 0$, there exists a positive integer N_ε such that for every $N \geq N_\varepsilon$ there is a block code $\phi: A^N \rightarrow B^N$ satisfying the requirement that $(x, \phi(x))$ is ε -typical of ν , $x \in A^N$.

Before giving the main result of this section on the simulation of channels, we need to present some more notation.

If S is a finite set and $\pi = \{\pi(i, j): (i, j) \in S \times S\}$ is a stochastic matrix, $\alpha(\pi)$ denotes the number $\max_{i_1, i_2 \in S} \sum_{j \in S} |\pi_{i_1 j} - \pi_{i_2 j}|$. If X is a random variable defined on the probability space $(\Omega, \mathfrak{F}, \lambda)$, $E_\lambda X$ denotes the expected value of X . If Y is another random variable on this space,

$\text{COV}_\lambda(X, Y)$ denotes the covariance of the pair (X, Y) . Suppose now λ is a probability measure on \mathfrak{B}^∞ such that $\lambda(E) > 0$ for every f.d. set $E \in \mathcal{Q}^\infty$. If $m, n \in \mathbb{Z}$ and $m \leq n$, define $\lambda(\tilde{Y}_m, \dots, \tilde{Y}_n)$ to be the function from $B^\infty \rightarrow [0, 1]$ such that $\lambda(\tilde{Y}_m, \dots, \tilde{Y}_n)(y) = \lambda(\tilde{Y}_m = \tilde{Y}_m(y), \dots, \tilde{Y}_n = \tilde{Y}_n(y))$, $y \in B^\infty$; define the function $\lambda(\tilde{Y}_{n+1} | \tilde{Y}_m, \dots, \tilde{Y}_n)$ to be the ratio

$$\lambda(\tilde{Y}_m, \dots, \tilde{Y}_{n+1}) / \lambda(\tilde{Y}_m, \dots, \tilde{Y}_n).$$

THEOREM 1. *Let K, M be positive integers and ν a channel with input space $A^{2K+2M+1}$ and output space B^{2K+1} . Suppose there exists $\nu' \in \mathfrak{S}$ such that $\nu'_x(\tilde{Y}_{-K}^K = b) = \nu_{x_{-K-M}^{K+M}}(b)$, $x \in A^\infty$, $b \in B^{2K+1}$. Then ν can be simulated by block codes.*

Proof. It is not hard to see that it suffices to prove the theorem under the assumption that $\nu_x(E) > 0$ for all $x \in A^\infty$ and all f.d. sets $E \in \mathfrak{B}^\infty$. Fix $\varepsilon > 0$. Choose $\delta > 0$ so that $\delta < \varepsilon$ and $2|B|^{2K+1}|A|^{2K+2M+1}\delta^2 < 1$. Let $\hat{\nu} \in \mathfrak{S}$ be the channel such that

$$\begin{aligned} \hat{\nu}_x(\tilde{Y}_{j-K}^{j+K}) &= \nu'_x(\tilde{Y}_{j-K}^{j+K}), \quad j \in \mathbb{Z}, x \in A^\infty; \\ \hat{\nu}_x(\tilde{Y}_r^s) &= \nu'_x(\tilde{Y}_r^{r+2K}) \prod_{i=r+2K+1}^s \nu'_x(\tilde{Y}_i | \tilde{Y}_{i-2K}^{i-1}), \\ &\quad r, s \in \mathbb{Z}, s - r > 2K, x \in A^\infty. \end{aligned}$$

For each $i \in \mathbb{Z}$, $x \in A^\infty$, let $M^{x,i}$ be the stochastic matrix $M^{x,i} = \{M^{x,i}(b_1, b_2) : b_1, b_2 \in B^{2K+1}\}$ such that

$$M^{x,i}(\tilde{Y}_i^{i+2K}, \tilde{Y}_{i+1}^{i+2K+1}) = \nu'_x(\tilde{Y}_{i+2K+1} | \tilde{Y}_i^{i+2K}).$$

For each $x \in A^\infty$, $\{\tilde{Y}_i^{i+2K}\}_{i=-\infty}^\infty$ is a Markov chain under $\hat{\nu}_x$ such that the conditional distribution of \tilde{Y}_{i+1}^{i+2K+1} given \tilde{Y}_i^{i+2K} is given by $M^{x,i}$, $i \in \mathbb{Z}$. The set of matrices $S = \{M^{x,i} : x \in A^\infty, i \in \mathbb{Z}\}$ is finite and any finite product of elements from S is regular. Thus, by a result of Wolfowitz [21] there is a positive integer J such that if $n \geq J$ and $\pi_1, \dots, \pi_n \in S$ then $\alpha(\pi_1 \cdots \pi_n) < \delta^4$. Choose N_0 so large that $N_0 > 2K + 2M$ and $2J/(N_0 - 2K - 2M) < \delta^4$. Fix $N \geq N_0$. We want to show there is a block code $\phi: A^N \rightarrow B^N$ “ ε -typical” of ν . Fix $\tilde{x} \in A^N$. Our task is to define $\phi(\tilde{x})$ so that $(\tilde{x}, \phi(\tilde{x}))$ is ε -typical of ν . Fix $x \in A^\infty$ so that $x_1^N = \tilde{x}$. By definition of $\hat{\nu}$, the choice of J , and Lemma A4 of the Appendix,

$$(4.1) \quad E_{\hat{\nu}_x} [I_{\{\tilde{Y}_{i-K}^{i+K}=b\}}] = \nu_{x_{i-K-M}^{i+K+M}}(b), \quad i \in \mathbb{Z}, b \in B^{2K+1}.$$

$$(4.2) \quad \text{COV}_{\hat{\nu}_x} [I_{\{\tilde{Y}_{i-K}^{i+K}=b\}}, I_{\{\tilde{Y}_{j-K}^{j+K}=b\}}] < \delta^4,$$

$$i, j \in \mathbb{Z}, |i - j| \geq J, b \in B^{2K+1}.$$

For each $a \in A^{2K+2M+1}$, let $R_a = \{K + M + 1 \leq j \leq N - K - M: \tilde{x}_{j-K-M}^{j+K+M} = a\}$. Letting $N_1 = N - 2K - 2M$, we have, using Chebyshev's Inequality and (4.1), (4.2),

$$\begin{aligned} & \hat{p}_x[(\tilde{x}, \tilde{Y}^N) \text{ is not } \delta\text{-typical of } \nu] \\ & \leq \sum_{a \in A^{2M+2K+1}, b \in B^{2K+1}} \hat{p}_x \left[\left| \sum_{i \in R_a} I_{\{\tilde{Y}_{i-K}^{i+K}=b\}} - \nu_a(b) \right| |R_a| \geq \delta N_1 \right] \\ & \leq \sum_{a,b} \delta^{-2} N_1^{-2} \sum_{i,j \in R_a} \text{COV}_{\hat{p}_x} [I_{\{\tilde{Y}_{i-K}^{i+K}=b\}}, I_{\{\tilde{Y}_{j-K}^{j+K}=b\}}] \\ & \leq \sum_{a,b} \delta^{-2} N_1^{-2} [2N_1 J + N_1^2 \delta^4] < 2 |A|^{2K+2M+1} |B|^{2K+1} \delta^2 < 1. \end{aligned}$$

In the preceding, the covariance terms were first summed over i, j with $|i - j| < J$, yielding the upper bound $2N_1 J$ by Lemma A4; the remaining terms have upper bound $N_1^2 \delta^4$ by (4.2).

Hence, we may find some $\tilde{y} \in B^N$ such that (\tilde{x}, \tilde{y}) is δ -typical and therefore ε -typical of ν . Define $\phi(\tilde{x}) = \tilde{y}$.

V. Building finite codes using coding sets. In this section we investigate the possibility of building a sliding-block code ψ so that ν^ψ is “close to” a given weakly continuous channel. We describe briefly how this is done before going into details. A finite sequence of disjoint sets $\{F_0, F_1, \dots, F_n\}$ from \mathcal{A}^∞ is called a *Rohlin-Kakutani tower* if $T^i F_0 = F_i$, $1 \leq i \leq n$. Roughly speaking, to approximate a given channel ν we first simulate the channel with a block code. We then choose an appropriate coding set and use it to build a sufficiently long Rohlin-Kakutani tower. The block code is then “embedded” in the tower (as for example in [20]) to obtain the desired sliding-block code ψ . The extent to which ν^ψ mimics ν depends only on how good a coding set can be found. Later on, in the next section, we provide a complete answer to the question of when a good coding set can be found.

By means of the following lemma, we show how a coding set can be used to construct a Rohlin-Kakutani tower.

LEMMA 3. *Let N be a positive integer. Let $F \subset A^\infty$ be a coding set of order N . There exists a f.d. set $E \subset A^\infty$ such that*

- (a) $E, TE, \dots, T^N E$ are disjoint.
- (b) For some positive integer L , $F \subset \bigcup_{i=0}^L T^i E$.

Proof. Find $i \in Z$, a positive integer k , and $F' \subset A^k$ such that $F = \{\tilde{X}_{i+1}^{i+k} \in F'\}$. Now $F' = \{b_1, \dots, b_r\}$, say. Each b_i , $1 \leq i \leq r$, has the following property:

$$(5.1) \quad \begin{aligned} &\text{There is no } x \in A^\infty \text{ such that } x_1^k = b_i \text{ and } T^j(x)_1^k = b_i \\ &\text{for some } j \text{ in the range } 1 \leq j \leq N. \end{aligned}$$

For, if (5.1) fails, it is not hard to see that F must contain a periodic sequence of period $\leq N$. Define

$$E = \bigcup_{m=1}^r \left\{ \tilde{X}_{i+1}^{i+k} = b_m, \tilde{X}_{i+j+k}^{i+j+k} \neq b_s; 1 \leq s \leq m-1, -N \leq j \leq N \right\}.$$

From (5.1) we deduce (a). Suppose for some $s \in Z$ and $M > 0$ we have $T^j x \notin E$, for all j in the range $s \leq j \leq s + M$. It follows that $T^j x \notin F$ for all j in the range $s + (r-1)N \leq j \leq s + M - (r-1)N$. Then, replacing E by some shift $T^m E$ of itself if necessary, we get (b).

Here is the main result of this section.

THEOREM 2. *Let $\nu \in \mathfrak{S}_c$. Let a positive integer K and $\epsilon > 0$ be given. Then there exists a positive integer J such that:*

For each coding set F of order J there exists a sliding-block $\psi: A^\infty \rightarrow B^\infty$ for which

$$(5.2) \quad \sum_{a \in A^{2K+1}, b \in B^{2K+1}} \left| \mu \nu^\psi [X_{-K}^K = a, Y_{-K}^K = b] - \mu \nu [X_{-K}^K = a, Y_{-K}^K = b] \right| \leq \epsilon + 2\mu(F^c), \quad \mu \in \mathfrak{M}.$$

Proof. In view of P2, we may assume $\nu \in \mathfrak{S}_f$. Thus we can find a positive integer M and a channel $\hat{\nu}$ with input space $A^{2K+2M+1}$ and output space B^{2K+1} such that

$$\nu_x(\tilde{Y}_{-K}^K = b) = \hat{\nu}_{x_{-K-M}^{K+M}}(b), \quad b \in B^{2K+1}, x \in A^\infty.$$

Choose $\delta > 0$ so small that

$$(5.3) \quad 3 |A|^{2K+2M+1} |B|^{2K+1} \delta < \epsilon.$$

In view of Theorem 1, we may pick $N > 2K + 2M$ and a block code $\phi: A^N \rightarrow B^N$ such that

$$(5.4) \quad (x, \phi(x)) \text{ is } \delta\text{-typical of } \hat{\nu}, \quad x \in A^N.$$

We assume N is large enough so that

$$(5.5) \quad 2N^{-1}(2K + 2M) < \delta.$$

Let J be a multiple of N so large that

$$(5.6) \quad 2N/J < \delta.$$

Let F be any coding set of order J . By Lemma 3, pick a f.d. set E and $L > J$ such that

$$(5.7) \quad E, TE, \dots, T^J E \text{ are disjoint,}$$

$$(5.8) \quad F \subset \bigcup_{i=0}^L T^i E.$$

Let $\psi: A^\infty \rightarrow B^\infty$ be a sliding-block code satisfying the following property:

If $x \in A^\infty$, $i \in \mathbb{Z}$, and $T^i x \in E$, then $\psi(x)_{i+sN}^{i+sN+N-1} = \phi(x_{i+sN}^{i+sN+N-1})$ for all s in the range $0 \leq s \leq r - 1$, where $r = r(x, i)$ is the largest positive integer for which both (a)–(b) which follow hold.

(a) $T^s x \notin E$ for all s in the range $i < s \leq i + rN - 1$.

(b) $rN - 1 \leq L$.

(Note that r is at least 1 by (5.7).) Roughly speaking, we code onward from each coordinate i where $T^i x \in E$ (that is, where E is “windowed”) using the block code ϕ , stopping when the next coordinate where E is windowed is reached or when no more than L coordinates of x have been coded, whichever comes first.

Since both sides of inequality (5.2) are continuous functions of $\mu \in \mathfrak{M}$ and \mathfrak{E} is dense in \mathfrak{M} [18], we need only show (5.2) for $\mu \in \mathfrak{E}$. Fix $\mu \in \mathfrak{E}$. We may assume $\mu(F) > 0$. (Otherwise (5.2) trivially holds.) Then, from (5.8), $\mu(E) > 0$. Fix $x \in A^\infty$ a regular point of μ . (This means $\lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f(T^i x) = \int_{A^\infty} f d\mu$, for every f.d. function $f: A^\infty \rightarrow R$.) Suppose $i, j \in \mathbb{Z}$, $i < j$, are given so that $T^i x \in E$, $T^j x \in E$, and $T^s x \notin E$ for all s in the range $i < s < j$. Then, we have for all $a \in A^{2K+2M+1}$, $b \in B^{2K+1}$

$$(5.9) \quad \left| \sum_{s=i}^{j-1} \delta(x_{s-K-M}^{s+K+M}, a) \delta(\psi(x)_{s-K}^{s+K}, b) - \hat{\nu}_a(b) \sum_{s=i}^{j-1} \delta(x_{s-k-M}^{s+K+M}, a) \right|$$

$$\leq (j - i)\delta + 2N^{-1}(2K + 2M)(j - i) + 2N$$

$$+ \sum_{s=i}^{j-1} I_{F^c}(T^s x) \delta(x_{s-K-M}^{s+K+M}, a) \delta(\psi(x)_{s-K}^{s+K}, b)$$

$$+ \hat{\nu}_a(b) \sum_{s=i}^{j-1} I_{F^c}(T^s x) \delta(x_{s-K-M}^{s+K+M}, a).$$

The first three terms on the right-hand side of inequality (5.9) arise from summing over indices s which are no bigger than $i + rN + N - 1$; summing over the remaining indices, if any, we get the last two terms on the right-hand side of (5.9), since such an index s is at a distance from i greater than L and so $T^s x \notin F$ by (5.8).

Now $j - i \geq J$ by (5.7) and so by (5.6),

$$(5.10) \quad 2N < J\delta \leq (j - i)\delta.$$

Since $\mu(E) > 0$, we have $T^i x \in E$ for infinitely many $i > 0$, say $i_1 < i_2 < \dots$. We obtain from (5.5), (5.9), (5.10) that for $t = 2, 3, \dots$,

$$\begin{aligned} & \left| (i_t - i_1)^{-1} \sum_{s=i_1}^{i_t-1} \delta(x_{s-K-M}^{s+K+M}, a) \delta(\psi(x)_{s-K}^{s+K}, b) \right. \\ & \quad \left. - \hat{\nu}_a(b) (i_t - i_1)^{-1} \sum_{s=i_1}^{i_t-1} \delta(x_{s-K-M}^{s+K+M}, a) \right| \\ & \leq 3\delta + (i_t - i_1)^{-1} \sum_{s=i_1}^{i_t-1} I_{F^c}(T^s x) \delta(x_{s-K-M}^{s+K+M}, a) \delta(\psi(x)_{s-K}^{s+K}, b) \\ & \quad + \hat{\nu}_a(b) (i_t - i_1)^{-1} \sum_{s=i_1}^{i_t-1} I_{F^c}(T^s x) \delta(x_{s-K-M}^{s+K+M}, a). \end{aligned}$$

Letting $t \rightarrow \infty$ in the preceding we obtain, since x is a regular point,

$$(5.11) \quad \begin{aligned} & |\mu\nu^\psi[X_{-K-M}^{K+M} = a, Y_{-K}^K = b] - \hat{\nu}_a(b)\mu[\tilde{X}_{-K-M}^{K+M} = a]| \\ & \leq 3\delta + \mu[F^c \cap \{x: x_{-K-M}^{K+M} = a, \psi(x)_{-K}^K = b\}] \\ & \quad + \hat{\nu}_a(b)\mu[F^c \cap \{\tilde{X}_{-K-M}^{K+M} = a\}]. \end{aligned}$$

Substituting $\mu\nu^\psi[X_{-K-M}^{K+M} = a, Y_{-K}^K = b]$ for $\hat{\nu}_a(b)\mu[X_{-K-M}^{K+M} = a]$ in (5.11), and summing over a, b , we obtain by (5.3) that

$$\begin{aligned} & \sum_{a \in A^{2M+2K+1}, b \in B^{2K+1}} |\mu\nu^\psi[X_{-K-M}^{K+M} = a, Y_{-K}^K = b] \\ & \quad - \mu\nu[X_{-K-M}^{K+M} = a, Y_{-K}^K = b]| \leq \varepsilon + 2\mu(F^c). \end{aligned}$$

(5.2) follows from this.

We remark that P3, P4 easily follows from Theorem 2.

VI. Families of sources concentrated on coding sets. In this section we characterize when a family $\mathfrak{Q} \subset \mathfrak{N}$ is concentrated on coding sets (Theorem 4) or uniformly concentrated on coding sets (Theorem 3).

LEMMA 4. $\mathfrak{D} \subset \mathfrak{M}$ is concentrated on coding sets if and only if there exists a sequence of f.d. sets $\{F_N\}_{N=1}^\infty$ such that

- (a) $\lim_{N \rightarrow \infty} \mu(F_N) = 0, \mu \in \mathfrak{M}_p.$
- (b) $\lim_{N \rightarrow \infty} \mu(F_N) = 1, \mu \in \mathfrak{D}.$

Proof. Let \mathfrak{D} be concentrated on the coding sets $\{F_N\}$ where F_N has order N . Let $\mu \in \mathfrak{M}_p$ be of order N' . Let $N \geq N'$. Then F_N does not contain any periodic sequence of period $\leq N'$ and so $\mu(F_N) = 0$. Thus $\lim_{n \rightarrow \infty} \mu(F_n) = 0$ for all $\mu \in \mathfrak{M}_p$ which have finite order. If $\mu \in \mathfrak{M}_p$, and $\varepsilon > 0$, we show $\limsup_{n \rightarrow \infty} \mu(F_n) \leq \varepsilon$. Let $E_m = \{x \in A^\infty: x \text{ is periodic with period } \leq m\}$. Then $\mu(E_m) \rightarrow 1$ as $m \rightarrow \infty$. Fix m so large that $\mu(E_m) \geq 1 - \varepsilon$. Let μ' be the measure such that $\mu'(E) = \mu(E \cap E_m) / \mu(E_m), E \in \mathcal{Q}^\infty$. Then $\mu' \in \mathfrak{M}_p$ and is of order m , so $\lim_{n \rightarrow \infty} \mu'(F_n) = 0$. But $\mu(F_n) \leq \mu'(F_n) + \varepsilon$, and so $\limsup_{n \rightarrow \infty} \mu(F_n) \leq \varepsilon$. Thus (a) holds — (b) automatically holds.

Conversely, let $\{F_n\}$ be a sequence of f.d. sets such that (a)–(b) hold. Let $x \in A^\infty$ be periodic with period N . Let μ be the periodic measure which places probability N^{-1} on each of $x, Tx, \dots, T^{N-1}x$. Since $\mu(F_n) \rightarrow 0$, we must have $\mu(F_n) = 0$ for n sufficiently large. Hence F_n cannot contain x , for n sufficiently large. Since x was an arbitrary periodic sequence of period N , and N was arbitrary, there is a subsequence $\{F_{n_k}\}_{k=1}^\infty$ of $\{F_n\}$ such that for each k, F_{n_k} is a coding act of order k . By definition, \mathfrak{D} is concentrated on the coding sets $\{F_{n_k}\}$.

DEFINITION. Recall that $S_p \subset A^\infty$ is the set of periodic points. For each $\delta > 0$, define $\mathfrak{M}_a^\delta = \{\mu \in \mathfrak{M}: \mu(S_p) \leq \delta\}$.

Note the following: if $\mu \in \mathfrak{M}$, then there exists $\mu_0 \in \mathfrak{M}_p, \mu_1 \in \mathfrak{M}_a$ such that $\mu = (1 - \alpha)\mu_0 + \alpha\mu_1$, where $\alpha = \mu(S_p^c)$. If $\mu(S_p) = 0$ or 1, this is trivial. If $0 < \mu(S_p) < 1$, define μ_0, μ_1 so that

$$\mu_0(E) = \frac{\mu(E \cap S_p)}{\mu(S_p)}, \quad E \in \mathcal{Q}^\infty.$$

$$\mu_1(E) = \frac{\mu(E \cap S_p^c)}{\mu(S_p^c)}, \quad E \in \mathcal{Q}^\infty.$$

LEMMA 5. Let \mathcal{K} be a closed subset of \mathfrak{M} . Let $\delta > 0$. Assume $\mathcal{K} \subset \mathfrak{M}_a^\delta$. For any positive integer N , there exists a coding set F of order N such that $\mu(F) \geq 1 - 4\delta, \mu \in \mathcal{K}$.

Proof. Fix N . Let $E = \{\tilde{X}^{N!} \neq \tilde{X}_{N!+1}^{2N!}\}$. Since E contains no periodic sequence of period $\leq N$,

$$(6.1) \quad \mu(E) = 0, \quad \mu \in \mathfrak{N}_p \text{ of order } \leq N.$$

If $\mu \in \mathfrak{N}$ and $\mu(E) = 0$, then μ is periodic of order $\leq N!$. Hence

$$(6.2) \quad \mu(E) > 0, \quad \mu \in \mathfrak{N}_a.$$

Let $\tilde{E} = \{x \in E: T^i x \in E \text{ for infinitely many } i > 0\}$. Let $\phi: \tilde{E} \rightarrow \{1, 2, \dots\}$ be the time of first return to E . That is,

$$\phi(x) = \inf\{i > 0: T^i x \in E\}, \quad x \in \tilde{E}.$$

By Kac's Recurrence Theorem [4]

$$(6.3) \quad \int_{\tilde{E}} \phi \, d\mu = 1, \quad \mu \in \mathfrak{E}, \mu(E) > 0.$$

Now by ergodic decomposition theory [17], given $\mu \in \mathfrak{N}$ there exists a family $\{\mu_\omega: \omega \in A^\infty\} \subset \mathfrak{E}$ such that

(a) For each $F \in \mathcal{Q}^\infty$, the map $\omega \rightarrow \mu_\omega(F)$ from $A^\infty \rightarrow [0, 1]$ is \mathcal{Q}^∞ -measurable.

(b) $\mu(F) = \int_{A^\infty} \mu_\omega(F) \, d\mu(\omega)$, $F \in \mathcal{Q}^\infty$.

Let $\mu \in \mathfrak{N}_a$. Since $\mu(S_p) = 0$, if $\{\mu_\omega\}$ is the decomposition of μ satisfying (a)–(b), we must have $\mu_\omega(S_p) = 0$ for μ -almost all $\omega \in A^\infty$. Hence we may assume that the “ergodic components” $\{\mu_\omega\}$ of $\mu \in \mathfrak{N}_a$ are all in \mathfrak{N}_a . Thus we must have

$$(6.4) \quad \int_{\tilde{E}} \phi \, d\mu = 1, \quad \mu \in \mathfrak{N}_a,$$

since this already holds for all $\mu \in \mathfrak{E} \cap \mathfrak{N}_a$ by (6.2), (6.3). Now if $\mu \in \mathfrak{N}$, by the remarks preceding this lemma we may decompose μ in the form $\mu = (1 - \alpha)\mu_0 + \alpha\mu_1$, where $\mu_0 \in \mathfrak{N}_p$, $\mu_1 \in \mathfrak{N}_a$, and $\alpha = \mu(S_p^c)$. Hence by (6.4) we have for all $\mu \in \mathfrak{N}$ that $\int_{\tilde{E}} \phi \, d\mu \geq \alpha \int_{\tilde{E}} \phi \, d\mu_1 = \alpha = \mu(S_p^c)$. Since $\mu(S_p^c) \geq 1 - \delta$ for $\mu \in \mathfrak{K}$, this gives

$$(6.5) \quad \lim_{n \rightarrow \infty} \min \left[\int_{\tilde{E} \cap \{\phi \leq n\}} \phi \, d\mu, 1 - \delta \right] = 1 - \delta, \quad \mu \in \mathfrak{K}.$$

Now $\mu(\tilde{E}) = \mu(E)$ for all $\mu \in \mathfrak{N}$ buy Poincaré's Recurrence Theorem [4]. Thus, for $n = 1, 2, \dots$, letting f_n by the f.d. function

$$f_n = I_E \sum_{i=1}^n i \left[\prod_{j=1}^{i-1} I_{T^{-j}E^c} \right] I_{T^{-i}E},$$

we have

$$\int_{A^\infty} f_n d\mu = \int_{\bar{E} \cap \{\phi \leq n\}} \phi d\mu, \quad \mu \in \mathfrak{M}.$$

This implies that in (6.5), the pointwise limit of an increasing sequence of continuous functions on the compact set \mathcal{K} is being taken. Thus, by Dini's Theorem [1, p. 181], the convergence in (6.5) is uniform on \mathcal{K} . Therefore we may fix n so that

$$(6.6) \quad \int f_n d\mu > 1 - 3\delta/2, \quad \mu \in \mathcal{K}.$$

Now by (6.3),

$$(6.7) \quad 0 \leq \int f_n d\mu \leq \int \phi d\mu \leq 1, \quad \mu \in \mathfrak{M}.$$

By (6.1)

$$(6.8) \quad \int f_n d\mu = 0, \quad \mu \in \mathfrak{M}_p \text{ of order } \leq N.$$

Choose $\beta > 0$ so that $2\beta < \min(N^{-1}, \delta/2)$. Using Lemma A1 of the Appendix and (6.6)–(6.8) we obtain a f.d. g such that

$$(c) \quad -\beta \leq g \leq 1 + \beta$$

$$(d) \quad \int g d\mu > 1 - 3\delta/2, \quad \mu \in \mathcal{K}$$

$$(e) \quad \int g d\mu = 0, \quad \mu \in \mathfrak{M}_p \text{ of order } \leq N.$$

Let $\hat{g} = (g + \beta)/(1 + 2\beta)$. Then

$$(f) \quad 0 \leq \hat{g} \leq 1$$

$$(g) \quad \int \hat{g} d\mu \geq (1 + 2\beta)^{-1}(1 - 3\delta/2) \geq 1 - 2\beta - 3\delta/2 \geq 1 - 2\delta, \quad \mu \in \mathcal{K}$$

$$(h) \quad \int \hat{g} d\mu = \beta(1 + 2\beta)^{-1} < \beta, \quad \mu \in \mathfrak{M}_p \text{ of order } \leq N.$$

Set $F = I_{\{\hat{g} \geq 1/2\}}$. By Chebyshev's Inequality and (h),

$$(6.9) \quad \mu(F) \leq 2 \int \hat{g} d\mu < 2\beta, \quad \mu \in \mathfrak{M}_p \text{ of order } \leq N.$$

Let x be a periodic sequence of period $j \leq N$. Let μ be the periodic source which assigns probability j^{-1} to each of $x, Tx, \dots, T^{j-1}x$. Now $\mu(F)$ is either 0 or is at least j^{-1} . But $j^{-1} \geq N^{-1} > 2\beta$, so the latter possibility is ruled out by (6.9). Hence F cannot contain any periodic sequence of period $\leq N$, which means F is a coding set of order N . Using Chebyshev's Inequality and (f), (g),

$$1 - \mu(F) = \mu[1 - \hat{g} \geq 1/2] \leq 2 \int (1 - \hat{g}) d\mu \leq 4\delta, \quad \mu \in \mathcal{K}.$$

Hence, $\mu(F) \geq 1 - 4\delta, \mu \in \mathcal{K}$.

THEOREM 3. *Let $\mathfrak{D} \subset \mathfrak{M}$. Then \mathfrak{D} is uniformly concentrated on coding sets if and only if $\overline{\mathfrak{D}} \subset \mathfrak{M}_a$, where $\overline{\mathfrak{D}}$ denotes the closure of \mathfrak{D} in the weak topology on \mathfrak{M} .*

Proof. By Lemma 5, if $\overline{\mathfrak{D}} \subset \mathfrak{M}_a$ then \mathfrak{D} is uniformly concentrated on coding sets. Conversely, let us suppose \mathfrak{D} is uniformly concentrated on coding sets. Pick coding sets $\{F_N\}_{N=1}^\infty$, where F_N is of order N , and $\mu(F_N) \geq 1 - N^{-1}$, $\mu \in \mathfrak{D}$, $N = 1, 2, \dots$. Let

$$\mathfrak{G} = \bigcap_{N=1}^\infty \{\mu \in \mathfrak{M} : \mu(F_N) \geq 1 - N^{-1}\}.$$

Then $\mathfrak{D} \subset \mathfrak{G}$ and \mathfrak{G} is a closed subset of \mathfrak{M} . All that remains to be shown is that $\mathfrak{G} \subset \mathfrak{M}_a$. Let $\mu \in \mathfrak{G}$. Let $\varepsilon > 0$. Decompose μ in the form $\mu = (1 - \alpha)\mu_0 + \alpha\mu_1$, where $\mu_0 \in \mathfrak{M}_p$, $\mu_1 \in \mathfrak{M}_a$, and $\alpha = \mu(S_p^c)$. As shown in the proof of Lemma 4, $\mu_0(F_N) \rightarrow 0$ since $\mu_0 \in \mathfrak{M}_p$. Hence

$$1 = \lim_{N \rightarrow \infty} \mu(F_N) = (1 - \alpha) \lim_{N \rightarrow \infty} \mu_0(F_N) + \alpha \lim_{N \rightarrow \infty} \mu_1(F_N) \leq \alpha.$$

Consequently $\mu(S_p^c) = 1$ and so $\mu \in \mathfrak{M}_a$.

THEOREM 4. *Let $\mathfrak{D} \subset \mathfrak{M}$. Then \mathfrak{D} is concentrated on coding sets if and only if for any $\varepsilon > 0$ there exists a σ -compact subset \mathfrak{F}_ε of \mathfrak{M} such that $\mathfrak{D} \subset \mathfrak{F}_\varepsilon \subset \mathfrak{M}_a^\varepsilon$.*

Proof. Suppose for any $\varepsilon > 0$, there is a σ -compact set \mathfrak{F}_ε such that $\mathfrak{D} \subset \mathfrak{F}_\varepsilon \subset \mathfrak{M}_a^\varepsilon$. For each $j = 1, 2, \dots$, write $\mathfrak{F}_{j-1} = \bigcup_{i=1}^\infty \mathcal{K}_i^{(j)}$, where each $\mathcal{K}_i^{(j)}$ is compact and $\mathcal{K}_i^{(j)} \subset \mathcal{K}_{i+1}^{(j)}$, $i = 1, 2, \dots$. Applying Lemma 5, for each $\mathcal{K}_i^{(j)}$ we may find a coding set $F_i^{(j)}$ of order i such that

$$(6.10) \quad \mu(F_i^{(j)}) \geq 1 - 4j^{-1}, \quad \mu \in \mathcal{K}_i^{(j)}.$$

Let $F_i = \bigcup_{j=1}^i F_i^{(j)}$, $i = 1, 2, \dots$. Then each F_i is a coding set of order i and

$$(6.11) \quad \mu(F_i) \geq \mu(F_i^{(j)}), \quad j \leq i.$$

We show \mathfrak{D} is concentrated on the coding sets $\{F_i\}$. Let $\mu \in \mathfrak{D}$, and suppose $\{\mu(F_i)\}$ does not converge to 1. Pick integers $\{i_j\}_{j=1}^\infty$ such that $\mu \in \mathcal{K}_{i_j}^{(j)}$, $j = 1, 2, \dots$. Pick $\beta < 1$ and integers $\{m_j\}_{j=1}^\infty$ such that

- (a) $\mu(F_{m_j}) \leq \beta, j = 1, 2, \dots$,
- (b) $m_j \geq \max(j, i_j), j = 1, 2, \dots$.

Then for each j , $\mu \in \mathcal{K}_{m_j}^{(j)}$ and $\mu(F_{m_j}) \geq \mu(F_{m_j}^{(j)})$ by (6.11). By (6.10), $\mu(F_{m_j}^{(j)}) \geq 1 - 4j^{-1}$. Thus, $\lim_{j \rightarrow \infty} \mu(F_{m_j}) = 1$, which contradicts (a).

Conversely, suppose \mathfrak{D} is concentrated on coding sets. By Lemma 4, pick f.d. sets $\{F_n\}$ such that $\mu(F_n) \rightarrow 0$ for $\mu \in \mathfrak{N}_p$ and $\mu(F_n) \rightarrow 1$ for $\mu \in \mathfrak{D}$. Let $\varepsilon > 0$. Let

$$\mathfrak{F}_\varepsilon = \bigcup_{k=1}^\infty \bigcap_{n=k}^\infty \{ \mu \in \mathfrak{N} : \mu(F_n) \geq 1 - \varepsilon \}.$$

Then \mathfrak{F}_ε is σ -compact and $\mathfrak{D} \subset \mathfrak{F}_\varepsilon$. To complete the proof we need to show that $\mathfrak{F}_\varepsilon \subset \mathfrak{N}_a^\varepsilon$. Let $\mu \in \mathfrak{F}_\varepsilon$. Write $\mu = (1 - \alpha)\mu_0 + \alpha\mu_1$, where $\mu_0 \in \mathfrak{N}_p$, $\mu_1 \in \mathfrak{N}_a$, and $\alpha = \mu(S_p^c)$. Since $(1 - \alpha)\mu_0(F_n) + \alpha \geq \mu(F_n) \geq 1 - \varepsilon$ for n sufficiently large, and $\mu_0(F_n) \rightarrow 0$, we get $\alpha \geq 1 - \varepsilon$. Thus $\mu(S_p) \leq \varepsilon$ and so $\mu \in \mathfrak{N}_a^\varepsilon$ by definition of $\mathfrak{N}_a^\varepsilon$.

Examples. (A) Any finite subset of \mathfrak{N}_a is closed and thus is uniformly concentrated on coding sets by Theorem 3.

(B) If $\varepsilon > 0$ the set \mathfrak{D} of all product measures (“memoryless sources”) μ in \mathfrak{N} such that $\min_{a \in A} \mu(\tilde{X}_1 = a) \geq \varepsilon$ is uniformly concentrated on coding sets by Theorem 3.

(C) If $\varepsilon > 0$, the set \mathfrak{D} of all Markov measures μ in \mathfrak{N} whose matrix of transition probabilities $\pi^\mu = \{ \pi_{i,j}^\mu : i, j \in A \}$ satisfies $\min_{i,j \in A} \pi_{i,j}^\mu \geq \varepsilon$, is uniformly concentrated on coding sets by Theorem 3.

(D) The set \mathfrak{D} of all measures in \mathfrak{N}_a relative to which T is a mixing transformation is concentrated on coding sets. For, if $\mu \in \mathfrak{D}$,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mu[\tilde{X}^n = \tilde{X}_{n+1}^{2n}] &\leq \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \mu[\tilde{X}^k = \tilde{X}_{n+1}^{n+k}] \\ &= \lim_{k \rightarrow \infty} \sum_{a \in A^k} \{ \mu[\tilde{X}^k = a] \}^2 \leq \lim_{k \rightarrow \infty} \max_{a \in A^k} \mu[\tilde{X}^k = a] = 0. \end{aligned}$$

Hence, setting $F_n = \{ \tilde{X}_1^{n!} \neq \tilde{X}_{n!+1}^{2n!} \}$, F_n is a coding set of order n and $\mu(F_n) \rightarrow 1$ for $\mu \in \mathfrak{D}$.

(E) The set \mathfrak{D} of all $\mu \in \mathfrak{G}$ with positive entropy is concentrated on coding sets. To see this, set

$$F_n = \bigcup_{i=0}^{n!-2} \{ \tilde{X}_{in!+1}^{(i+1)n!} \neq \tilde{X}_{(i+1)n!+1}^{(i+2)n!} \}.$$

Then F_n is a coding set of order n . There is a subset G_n of $A^{(n!)^2}$ such that $F_n = \{ X^{(n!)^2} \in G_n \}$, and $(n!)^{-2} \log |G_n^c| \rightarrow 0$ as $n \rightarrow \infty$. Hence, by the Shannon-McMillan Theorem [6], if $\mu \in \mathfrak{D}$, $\mu(F_n^c) \rightarrow 0$.

(F) \mathfrak{N}_a is *not* concentrated on coding sets. To see this, we first observe that \mathfrak{N}_a is a \mathfrak{G}_δ subset of \mathfrak{N} ; that is, \mathfrak{N}_a is a countable intersection of open subsets of \mathfrak{N} . (For,

$$fM_a = \bigcap_{x \in S_p} \bigcap_{n=1}^\infty \{ \mu \in \mathfrak{N} : \mu(x) < n^{-1} \}.)$$

If \mathfrak{N}_a were concentrated on coding sets, by Theorem 4 there would exist a σ -compact subset \mathfrak{F} of \mathfrak{N} such that $\mathfrak{N}_a \subset \mathfrak{F}$ and $\mathfrak{F} \cap \mathfrak{N}_p = \emptyset$. We can write $\mathfrak{F} = \bigcup_{i=1}^\infty \mathcal{K}_i$, where each \mathcal{K}_i is compact. Since \mathfrak{N}_a is a \mathfrak{G}_δ subset of the compact metrizable space \mathfrak{N} , the relative topology on \mathfrak{N}_a is metrizable with a complete metric [1, Thm. A9.9]. Thus \mathfrak{N}_a is a Baire space [1, Thm. A9.2], and so the \mathfrak{N}_a -closure of some $\mathcal{K}_1 \cap \mathfrak{N}_a$ must contain a non-empty \mathfrak{N}_a -open set. Therefore there is an open subset \emptyset of \mathfrak{N} such that $\emptyset \cap \mathfrak{N}_a \subset \mathcal{K}_i$ for a certain i . Now \mathfrak{N}_p is dense in \mathfrak{N} [18]. Thus we may pick $\mu \in \emptyset \cap \mathfrak{N}_p$. Since \mathfrak{N}_a is also dense in \mathfrak{N} [18], we may pick $\{\mu_n\} \subset \mathfrak{N}_a$ such that $\mu_n \rightarrow \mu$. Then $\mu_n \in \emptyset \cap \mathfrak{N}_a$, eventually. This would force $\mu \in \mathcal{K}_i$, since \mathcal{K}_i is closed. But $\mathcal{K}_i \cap \mathfrak{N}_p = \emptyset$, so this is a contradiction.

To conclude this section we point out how P5 can be shown to follow from P1 and P3. Given $\varepsilon > 0$, $\nu \in \mathfrak{S}$, $\mu_1, \dots, \mu_n \in \mathfrak{N}_a$, and f.d. sets $E_1, \dots, E_k \subset A^\infty \times B^\infty$, we need to find $\nu' \in \mathfrak{S}_{df}$ such that

$$(6.12) \quad |\mu_i \nu'(E_j) - \mu_i \nu(E_j)| < \varepsilon, \quad i = 1, \dots, n; j = 1, \dots, k.$$

First, by P1 find $\hat{\nu} \in \mathfrak{S}_f$ such that

$$(6.13) \quad |\mu_i \hat{\nu}(E_j) - \mu_i \nu(E_j)| < \varepsilon/2, \quad i = 1, \dots, n; j = 1, \dots, k.$$

Then, as observed in Example (A), $\{\mu_1, \dots, \mu_n\}$ is uniformly concentrated on coding sets; thus, by P3 there must exist $\nu' \in \mathfrak{S}_{df}$ such that

$$(6.14) \quad |\mu_i \nu'(E_j) - \mu_i \hat{\nu}(E_j)| < \varepsilon/2, \quad \text{for all } i, j.$$

From (6.13), (6.14), we obtain (6.12).

APPENDIX

The following is Lemma A3 of [8], in which paper a proof is given.

LEMMA A1. *Let $f: A^\infty \rightarrow R$ be a f.d. function. Suppose $\int_{A^\infty} f d\mu \geq 0$ for all $\mu \in \mathfrak{N}$. Then, given $\varepsilon > 0$, there exists a positive integer N such that for all $n \geq N$,*

$$n^{-1} \sum_{i=0}^{n-1} f \cdot T^i \geq -\varepsilon \quad \text{throughout } A^\infty.$$

LEMMA A2. *Let $\nu \in \mathfrak{S}$. Let k be a positive integer. Let ν' be a channel with input space A^∞ and output space B^k , such that for each $b \in B^k$, the map $x \rightarrow \nu'_x(b)$ from $A^\infty \rightarrow [0, 1]$ is f.d. Then there exists $\hat{\nu} \in \mathfrak{S}_f$ such that*

the following holds for every $\mu \in \mathfrak{N}$, $1 \leq j \leq k$, $a \in A^j$, $b \in B^j$:

$$(A.1) \quad \left| \mu\nu[X^j = a, Y^j = b] - \mu\hat{\nu}[X^j = a, Y^j = b] \right| \leq 2j/k + \sum_{a' \in A^k, b' \in B^k} \left| \int_{\{\tilde{X}^k = a'\}} [\nu_x(\tilde{Y}^k = b') - \nu'_x(b')] d\mu(x) \right|.$$

Proof. Let $\tilde{\nu}$ be the k -stationary channel with input space A^∞ and output space B^∞ such that

$$\tilde{\nu}_x \left(\bigtimes_{i=-\infty}^{\infty} \{ \tilde{Y}_{ik+1}^{ik+k} = b_i \} \right) = \prod_{i=-\infty}^{\infty} \nu'_{T^i x}(b_i),$$

for all $x \in A^\infty$ and $\{b_i\}_{i=-\infty}^{\infty} \subset B^k$. Then, for $\mu \in \mathfrak{N}$, $a \in A^k$, $b \in B^k$,

$$\begin{aligned} & \left| \mu\tilde{\nu}[X^k = a, Y^k = b] - \mu\nu[X^k = a, Y^k = b] \right| \\ &= \left| \int_{\{\tilde{X}^k = a\}} \tilde{\nu}_x(\tilde{Y}^k = b) d\mu(x) - \int_{\{\tilde{X}^k = a\}} \nu_x(\tilde{Y}^k = b) d\mu(x) \right| \\ &= \left| \int_{\{\tilde{X}^k = a\}} [\nu'_x(b) - \nu_x(\tilde{Y}^k = b)] d\mu(x) \right|. \end{aligned}$$

It follows from this that if $f: A^k \times B^k \rightarrow [-1, 1]$, then

$$(A.2) \quad \left| E_{\mu\tilde{\nu}}f(X^k, Y^k) - E_{\mu\nu}f(X^k, Y^k) \right| \leq \sum_{a \in A^k, b \in B^k} \left| \int_{\{\tilde{X}^k = a\}} [\nu'_x(b) - \nu_x(\tilde{Y}^k = b)] d\mu(x) \right|.$$

Let $\hat{\nu} \in \mathfrak{S}$ be the channel such that

$$\hat{\nu}_x(E) = k^{-1} \sum_{i=0}^{k-1} \tilde{\nu}_{T^i x}(T^i E), \quad x \in A^\infty, E \in \mathfrak{B}^\infty.$$

It is easily checked that $\hat{\nu} \in \mathfrak{S}_f$. Now for all $\mu \in \mathfrak{N}$, $1 \leq j \leq k$, $a \in A^j$, $b \in B^j$,

$$(A.3) \quad \begin{aligned} & \left| \mu\hat{\nu}[X^j = a, Y^j = b] - \mu\nu[X^j = a, Y^j = b] \right| \\ &= \left| E_{\mu\tilde{\nu}} \left[k^{-1} \sum_{i=0}^{k-1} I_{\{X_{i+1}^j = a, Y_{i+1}^j = b\}} \right] - E_{\mu\nu} \left[k^{-1} \sum_{i=0}^{k-1} I_{\{X_{i+1}^j = a, Y_{i+1}^j = b\}} \right] \right| \\ &\leq | E_{\mu\tilde{\nu}}f(X^k, Y^k) - E_{\mu\nu}f(X^k, Y^k) | + 2j/k, \end{aligned}$$

where $f: A^k \times B^k \rightarrow [-1, 1]$ is the function such that

$$f(X^k, Y^k) = (k - j + 1)^{-1} \sum_{i=0}^{k-j} I_{\{X_{i+j}^+ = a, Y_{i+j}^+ = b\}}.$$

Coupling (A.3) with (A.2), we get (A.1).

For our next lemma we need to present some notation. Let k be a positive integer. Let \mathfrak{P}^k be the set of all \mathcal{Q}^∞ -measurable ordered partitions $P = \{P^{(1)}, \dots, P^{(k)}\}$ of A^∞ consisting of k sets. Let $\tilde{\mathfrak{P}}^k$ be the subset of \mathfrak{P}^k consisting of those partitions whose elements are f.d. sets. If $\mu \in \mathfrak{M}$ and $P_1, P_2 \in \mathfrak{P}^k$, then the μ -distance between P_1 and P_2 , written $|P_1 - P_2|_\mu$, is defined to be $\sum_{j=1}^k \mu(P_1 \Delta P_2^j)$.

Recall that \mathfrak{E} is the set of all ergodic sources in \mathfrak{M} . Consider the measurable space $(\mathfrak{E}, \mathfrak{F}(\mathfrak{E}))$, where $\mathfrak{F}(\mathfrak{E})$ is the smallest σ -field of subsets of \mathfrak{E} such that for each $E \in \mathcal{E}^\infty$, the map $\mu \rightarrow \mu(E)$ from $\mathfrak{E} \rightarrow [0, 1]$ is measurable.

The following is a consequence of Theorem 1 of [12].

LEMMA A3. Let $\mathfrak{D} \in \mathfrak{F}(\mathfrak{E})$. Let X be a set. A family $\{\Phi_\mu: \mu \in \mathfrak{D}\}$ of maps from $\mathfrak{P}^k \rightarrow X$ is given such that

(a) $\Phi_\mu(P_1) = \Phi_\mu(P_2)$ if $|P_1 - P_2|_\mu = 0$, $P_1, P_2 \in \mathfrak{P}^k$, $\mu \in \mathfrak{D}$.

Let E be a subset of X such that

(b) $\{\mu \in \mathfrak{D}: \Phi_\mu(P) \in E\} \in \mathfrak{F}(\mathfrak{E})$, $P \in \mathfrak{P}^k$.

(c) $\Phi_\mu^{-1}(E) \cap \tilde{\mathfrak{P}}^k \neq \emptyset$, $\mu \in \mathfrak{D}$.

Then

$$\bigcap_{\mu \in \mathfrak{D}} \Phi_\mu^{-1}(E) \neq \emptyset.$$

Using the preceding lemma along with P5, we now are able to show that P6 follows. Suppose we are given $\nu \in \mathfrak{S}, \varepsilon > 0$, and f.d. sets $E_1, \dots, E_m \subset A^\infty \times B^\infty$. To establish P6, we need to find a stationary code $\psi: A^\infty \rightarrow B^\infty$ such that

$$(A.4) \quad \sup_{1 \leq j \leq m} \sup_{\mu \in \mathfrak{N}_a} |\mu\nu^\psi(E_j) - \mu\nu(E_j)| < \varepsilon.$$

Let $|B| = k$ and $B = \{b_1, \dots, b_k\}$. Let \mathcal{C} be the set of all stationary codes $\psi: A^\infty \rightarrow B^\infty$. There is a one-to-one map τ of \mathfrak{P}^k onto \mathcal{C} defined as follows: If $P = \{P^{(1)}, \dots, P^{(k)}\} \in \mathfrak{P}^k$, define $\tau(P) = \psi$ to be the stationary code $\psi: A^\infty \rightarrow B^\infty$ such that $\{x \in A^\infty: \psi(x)_0 = b_j\} = P^{(j)}, j = 1, \dots, k$. In Lemma

A3, take $\mathfrak{D} = \mathfrak{N}_a$, $X = R$, and for each $\mu \in \mathfrak{N}_a$, let $\Phi_\mu: \mathfrak{P}^k \rightarrow R$ be the map such that

$$\Phi_\mu(P) = \sup_{i \leq j \leq m} |\mu\nu^{\tau(P)}(E_j) - \mu\nu(E_j)|, \quad P \in \mathfrak{P}^k.$$

Let $E = \{x \in R: x \leq \varepsilon/2\}$. Assumptions (a)–(b) of Lemma A3 are easily seen to hold. Assumption (c) is a consequence of P5 since τ maps $\tilde{\mathfrak{P}}^k$ onto the set of all finite stationary codes. Applying Lemma A3, we obtain $P \in \mathfrak{P}^k$ such that $\Phi_\mu(P) \in E$, $\mu \in \mathfrak{N}_a$. Letting $\psi = \tau(P)$, this means

$$\sup_{1 \leq j \leq m} |\mu\nu^\psi(E_j) - \mu\nu(E_j)| \leq \varepsilon/2, \quad \mu \in \mathfrak{N}_a.$$

This implies (A.4).

LEMMA A4. *Let S be a finite set. Let $\pi = \{\pi_{ij}: i, j \in S\}$ be a stochastic matrix. Let $(\Omega, \mathfrak{F}, P)$ be a probability space. Let $X, Y: \Omega \rightarrow S$ be measurable functions such that $P[X = i, Y = j] = P[X = i]\pi_{ij}$, $i, j \in S$. Let $f: S \rightarrow [0, 1]$. Then*

$$|\text{COV}(f(X), f(Y))| \leq \min(\alpha(\pi), 1).$$

Proof. Let $p_i = P[X = i]$, $i \in S$; let $\mu_1 = E(X)$, $\mu_2 = E(Y)$. Then

$$\text{COV}(f(X), f(Y)) = \sum_{i,j \in S} (f(i) - \mu_1)(f(j) - \mu_2)p_i\pi_{ij}.$$

Since

$$|f(i) - \mu_1||f(j) - \mu_2| \leq 1, \quad i, j \in S,$$

it is clear that $|\text{COV}(f(X), f(Y))| \leq 1$. Fix $\hat{i} \in S$. Then

$$\begin{aligned} & \sum_{i,j \in S} (f(i) - \mu_1)(f(j) - \mu_2)p_i\pi_{i\hat{j}} \\ &= \left(\sum_i (f(i) - \mu_1)p_i \right) \left(\sum_j (f(j) - \mu_2)\pi_{i\hat{j}} \right) = 0. \end{aligned}$$

Thus

$$\begin{aligned} & |\text{COV}(f(X), f(Y))| \\ &= \left| \text{COV}(f(X), f(Y)) - \sum_{i,j} (f(i) - \mu_1)(f(j) - \mu_2)p_i\pi_{i\hat{j}} \right| \\ &\leq \sum_{i,j} p_i |\pi_{ij} - \pi_{i\hat{j}}| \leq \alpha(\pi). \end{aligned}$$

REFERENCES

1. R. B. Ash, *Real Analysis and Probability*, Academic Press, 1972.
2. R. M. Gray, D. L. Neuhoff and D. S. Ornstein, *Nonblock source coding with a fidelity criterion*, *Ann. Probability*, **3** (1975), 478–491.
3. R. M. Gray and D. S. Ornstein, *Block coding for discrete stationary \bar{d} -continuous noisy channels*, *IEEE Trans. Information Theory*, **25** (1979), 292–306.
4. M. Kac, *On the notion of recurrence in discrete stochastic processes*, *Bull. Amer. Math. Soc.*, **53** (1947), 1002–1010.
5. S. Kakutani, *Induced measure-preserving transformations*, *Proc. Imp. Acad. Sci. Tokyo*, **19** (1943), 635–641.
6. A. I. Khinchin, *Mathematical Foundations of Information Theory*, Dover Publications, 1957.
7. J. C. Kieffer, *On the transmission of Bernoulli sources over stationary channels*, *Ann. Probability*, **8** (1980), 962–973.
8. ———, *Extension of source coding theorems for block codes to sliding-block codes*, *IEEE Trans. Information Theory*, **26** (1980), 679–692.
9. ———, *Zero-error stationary coding over stationary channels*, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **56** (1981), 113–126.
10. ———, *Block coding for weakly continuous channels*, *IEEE Trans. Information Theory*, **27** (1981), 721–727.
11. ———, *Sliding-block coding for weakly continuous channels*, *IEEE Trans. Information Theory*, **28** (1982), 2–10.
12. J. C. Kieffer and M. Rahe, *Selecting universal partitions in ergodic theory*, *Ann. Probability*, **9** (1981), 705–709.
13. P. A. Meyer, *Probability and Potentials*, Blaisdell, 1966.
14. D. L. Neuhoff and P. C. Shields, *Channels with almost finite memory*, *IEEE Trans. Information Theory*, **25** (1979), 440–447.
15. ———, *Indecomposable finite state channels and primitive approximation*, *IEEE Trans. Information Theory*, **28** (1982), 11–18.
16. ———, *Channel entropy and primitive approximation*, *Ann. Probability*, **10** (1982), 188–198.
17. J. C. Oxtoby, *Ergodic sets*, *Bull. Amer. Math. Soc.*, **58** (1952), 116–136.
18. K. R. Parthasarathy, *On the category of ergodic measures*, *Illinois J. Math.*, **5**, (1961), 648–656.
19. ———, *Probability Measures on Metric Spaces*, Academic Press, 1967.
20. P. C. Shields and D. L. Neuhoff, *Block and sliding-block source coding*, *IEEE Trans. Information Theory*, **23** (1977), 211–215.
21. J. Wolfowitz, *Products of indecomposable aperiodic stochastic matrices*, *Proc. Amer. Math. Soc.*, **14** (1963), 733–737.

Received December 29, 1978 and in revised form May 25, 1982. Research of author supported NSF Grants ENG-76-02276 and ECS-7821335.

UNIVERSITY OF MISSOURI-ROLLA
ROLLA, MO 65401

