INFORMATION AND CONTROL 46, 93–107 (1980)

Some Universal Noiseless Multiterminal Source Coding Theorems

JOHN C. KIEFFER*

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

Fixed and variable-rate block and sliding-block weighted universal noiseless coding theorems are obtained which extend the Slepian-Wolf theorem for a single multiterminal source to a family of finite-alphabet, stationary, ergodic multiterminal sources.

I. INTRODUCTION

Suppose we are given a multiterminal source consisting of the finite-state processes $(X^{(1)},..., X^{(n)})$, which we assume to have a stationary and ergodic joint distribution P. Slepian and Wolf (1973) and Cover (1975) determined the rate region $\mathscr{R}(P)$ of all vectors $(R_1,..., R_n)$ such that each subsource $X^{(i)}$ can be block encoded at rate R_i into a process $\hat{X}^{(i)}$, and then $(X^{(1)},..., X^{(n)})$ can be recovered with almost zero probability of block error by applying some block decoder to $(\hat{X}^{(1)},..., \hat{X}^{(n)})$. Suppose the distribution P is not known precisely, but is known to lie in some family of distributions Λ. Ideally, for a given rate vector $(R_1,..., R_n)$, one would like to find universal block encoders achieving the rates $(R_1,..., R_n)$ and a universal block decoder achieving small probability of error for *every* $P \in \Lambda$. Clearly, a necessary condition on the rate vector so that this is possible is that it lie in $\mathscr{R}(P)$ for every $P \in \Lambda$. This condition is not sufficient unless the family Λ is compact in an appropriate sense. However, in this paper, we will show the condition is sufficient in the weaker sense that weighted universal coders can be found which universally code $(X^{(1)},..., X^{(n)})$ for "almost all" distributions in Λ (with respect to some a priori weight distribution on Λ). A variable-rate version of this result is also obtained, where $(R_1,..., R_n)$ is allowed to depend on $P \in \Lambda$. In that case, the rate of the ith universal block encoder (as measured by the expected code word length per unit time for a fixed variable-length noiseless coder applied to $\hat{X}^{(i)}$) is desired to be $R_i = R_i(P)$ for almost every $P \in \Lambda$. For the variable-rate weighted universal coders to exist, it is necessary to impose the additional requirement that each R_i depend on P only through

* Research of author supported by NSF Grants ENG-76 02276 and MCS-78 21335 and by the Joint Services Electronics Program under Contract N00014-79-C-0424.

0019-9958/80/080093–15$02.00/0

the marginal distribution of $X^{(i)}$ under P. We then consider the case where sliding-block coders are used instead of block coders. Precise statements of these results are given in the next section.

II. Statement of Main Results

Notation. If $X_1, \ldots, X_n$ are measurable functions defined on a measurable space Ω and taking their values in the measurable spaces $S_1, \ldots, S_n$, respectively, $(X_1, \ldots, X_n)$ denotes the map from $\Omega \to S_1 \times \cdots \times S_n$ such that $(X_1, \ldots, X_n)(\omega) = (X_1(\omega), \ldots, X_n(\omega))$, $\omega \in \Omega$.

If $(\Omega, \mathscr{F}, P)$ is a probability space, Ω_1 is a measurable space and X is a random variable defined on Ω with values in Ω_1, P^X denotes the distribution of X; that is, the probability measure on Ω_1 such that

$$P^X(E) = P(X \in E), \qquad E \text{ a measurable subset of } \Omega_1.$$

Let Z be the set of integers. If a symbol S denotes a finite set, the corresponding script letter $\mathscr{S}$ will denote the set of all subsets of S, and $(S^\infty, \mathscr{S}^\infty)$ will denote the measurable space consisting of S^∞, the set of all bilateral sequences $x = (x_i \colon i \in Z)$ from S and $\mathscr{S}^\infty$, the usual product σ-field of subsets of S^∞. If $x \in S^\infty$ and $i \in Z$, x_i denotes the ith coordinate of x and if $j \geqslant i$, x_i^j or $[x]_i^j$ denotes the $(j - i + 1)$-tuple $(x_i, \ldots, x_j)$. x^n or $[x]^n$ denotes $(x_1, \ldots, x_n)$. Similarly if $S_1, \ldots, S_n$ are finite sets and $(x_1, \ldots, x_n) \in S_1^\infty \times \cdots \times S_n^\infty$, then $(x_1, \ldots, x_n)_i^j$ denotes $([x_1]_i^j, \ldots, [x_n]_i^j)$ and $(x_1, \ldots, x_n)^N$ denotes $([x_1]^N, \ldots, [x_n]^N)$. By a finite state process X (with state space S), we mean that for some measurable space Ω and finite set S, X is a measurable map from $\Omega \to S^\infty$. For each $i \in Z$, X_i denotes the map from $\Omega \to S$ such that $X_i(\omega) = X(\omega)_i$, $\omega \in \Omega$. X_i^j or $[X]_i^j$ denotes the random vector $(X_i, \ldots, X_j)$. X^n or $[X]^n$ denotes $(X_1, \ldots, X_n)$. If $X^{(1)}, \ldots, X^{(n)}$ are finite-state processes, $(X^{(1)}, \ldots, X^{(n)})_i^j$ denotes $([X^{(1)}]_i^j, \ldots, [X^{(n)}]_i^j)$ and $(X^{(1)}, \ldots, X^{(n)})^N$ denotes $([X^{(1)}]^N, \ldots, [X^{(n)}]^N)$.

If X, Y are random variables, $H(X)$ and $H(X \mid Y)$ denote entropy and conditional entropy, respectively. If X is a finite-state N-stationary process for some $N = 1, 2, \ldots$, $\bar{H}(X)$ denotes the entropy of the process:

$$\bar{H}(X) = \lim_{n \to \infty} n^{-1} H(X_1, \ldots, X_n).$$

If (X, Y) are jointly N-stationary, $\bar{H}(X \mid Y)$ denotes the conditional entropy

$$\bar{H}(X \mid Y) = \lim_{n \to \infty} n^{-1} H(X_1, \ldots, X_n \mid Y_1, \ldots, Y_n).$$

We write $\bar{H}_P(X)$, $\bar{H}_P(X \mid Y)$ when it is necessary to emphasize the underlying probability measure P.

If A is a finite set, $T_A\colon A^\infty \to A^\infty$ denotes the shift transformation. If $A_1, \ldots, A_n$ are finite sets, $T_{A_1,\ldots,A_n}\colon A_1^\infty \times \cdots \times A_n^\infty \to A_1^\infty \times \cdots \times A_n^\infty$ denotes the transformation

$$T_{A_1,\ldots,A_n}(x_1, \ldots, x_n) \equiv (T_{A_1}x_1, \ldots, T_{A_n}x_n).$$

If $A_1, \ldots, A_n$ are finite sets let $\mathscr{E}(A_1, \ldots, A_n)$ denote the set of all probability measures on $\mathcal{O}\!\ell_1^\infty \times \cdots \times \mathcal{O}\!\ell_n^\infty$ stationary and ergodic with respect to $T_{A_1,\ldots,A_n}$. We make $\mathscr{E}(A_1, \ldots, A_n)$ a measurable space by adjoining the smallest σ-field of subsets of $\mathscr{E}(A_1, \ldots, A_n)$ such that for each $E \in \mathcal{O}\!\ell_1^\infty \times \cdots \times \mathcal{O}\!\ell_n^\infty$, the map $P \to P(E)$ from $\mathscr{E}(A_1, \ldots, A_n) \to [0, 1]$ is measurable.

If $(\Lambda, \mathscr{M})$ and $(\Omega, \mathscr{F})$ are two measurable spaces, we call a family $\{P_\theta\colon \theta \in \Lambda\}$ of probability measures on $\mathscr{F}$ measurable if for each $E \in \mathscr{F}$, the map $\theta \to P_\theta(E)$ from $\Lambda \to [0, 1]$ is $\mathscr{M}$-measurable.

Codes. If $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ are finite sets, $\varphi\colon A_1^\infty \times \cdots \times A_n^\infty \to B_1^\infty \times \cdots \times B_n^\infty$ is called a block code of order N if there exists $\varphi'\colon A_1^N \times \cdots \times A_n^N \to B_1^N \times \cdots \times B_n^N$ such that

$$\varphi(x_1, \ldots, x_n)_{iN+1}^{iN+N} \equiv \varphi'[(x_1, \ldots, x_n)_{iN+1}^{iN+N}], \qquad i \in Z.$$

If $\varphi\colon A^\infty \to B^\infty$ is a block code of order N, the rate $r(\varphi)$ of φ is defined to be $N^{-1} \log|\{\varphi(x)_1^N\colon x \in A^\infty\}|$, where if S is a finite set, $|S|$ denotes the cardinality of S. (All logarithms in this paper are to base 2.)

A map $\psi\colon A_1^\infty \times \cdots \times A_n^\infty \to B_1^\infty \times \cdots \times B_n^\infty$ is called a stationary code if $\psi(T_{A_1,\ldots,A_n}(x_1, \ldots, x_n)) \equiv T_{B_1,\ldots,B_n}\psi(x_1, \ldots, x_n)$. It is called a sliding-block code if it is stationary and for some M, $\psi(x_1, \ldots, x_n) = \psi(y_1, \ldots, y_n)$ if $(x_1, \ldots, x_n)_{-M}^M = (y_1, \ldots, y_n)_{-M}^M$. The rate $r(\psi)$ of a sliding-block code $\psi\colon A^\infty \to B^\infty$ is

$$\lim_{N\to\infty} N^{-1} \log |\{\psi(x)_1^N\colon x \in A^\infty\}|.$$

Let $\{0, 1\}^*$ be the set of all finite sequences of zeroes and ones. If A is a finite set a map $\tau\colon A \to \{0, 1\}^*$ is called a noiseless variable-length code if τ is one-to-one and $\tau(A)$ satisfies the prefix condition.

Multiterminal sources. Let n be a positive integer. By a n-parameter multiterminal source we mean a pair $[(X^{(1)}, \ldots, X^{(n)}), P]$, where the $X^{(i)}$ are finite state processes defined on a common measurable space $(\Omega, \mathscr{F})$ and P is a probability measure on $\mathscr{F}$. If the processes $\{X^{(i)}\}$ are jointly stationary (ergodic) with respect to P $[(X^{(1)}, \ldots, X^{(n)}), P]$ is called a stationary (ergodic) source.

Let E^n be the set of all n-tuples of real numbers. If $[(X^{(1)}, \ldots, X^{(n)}), P]$ is a multiterminal source and the $\{X^{(i)}\}$ are jointly N-stationary with respect to P for some N, define $\mathscr{R}[(X^{(1)}, \ldots, X^{(n)}), P]$ to be the set of all $R = (R_1, \ldots, R_n) \in E^n$ such that

$$\bar{H}((X^{(j)}\colon j \in S) \mid (X^{(j)}\colon j \notin S)) \leqslant \sum_{j\in S} R_j,$$

for every nonempty subset S of $\{1, 2,..., n\}$. (In the preceding, if each $X^{(i)}$ has state space A_i, we interpret a variable $(X^{(j)}: j \in T)$ as a process with state space $\prod_{j\in T} A_j$ rather than its customary interpretation as a function with values in the space $\prod_{j\in T} A_j^{\infty}$. We also interpret $\bar{H}((X^{(j)}: j \in T)|(X^{(j)}: j \notin T))$ to be $\bar{H}((X^{(j)}: j \in T))$ if there exists no $j \notin T$.) We note that if $[(X^{(1)},..., X^{(n)}), P]$ is stationary and ergodic then $\mathscr{R}[(X^{(1)},..., X^{(n)}), P]$ is the rate region for noiseless coding of that source (Cover, 1975).

Fixed and variable rate specifications. Let $X^{(1)},..., X^{(n)}$ be processes on $(\Omega, \mathscr{F})$ with state spaces $A_1,..., A_n$, respectively. Let $\{P_\theta: \theta \in \Lambda\}$ be a family of probability measures on $\mathscr{F}$. We suppose $[(X^{(1)},..., X^{(n)}), P_\theta]$ is a stationary, ergodic source, $\theta \in \Lambda$. We say that $\{R(\theta): \theta \in \Lambda\} \subset E^n$ is a variable-rate specification for the family of sources $\{[(X^{(1)},..., X^{(n)}), P_\theta]: \theta \in \Lambda\}$ if for each i there is a bounded measurable map $F_i: \mathscr{E}(A_i) \to [0, \infty]$ such that

(a) $R_i(\theta) = F_i(P_\theta^{X^{(i)}}), i = 1,..., n; \theta \in \Lambda,$

(b) $R(\theta) \in \mathscr{R}[(X^{(1)},..., X^{(n)}), P_\theta], \theta \in \Lambda.$

We say $R \in E^n$ is a fixed-rate specification for the family $\{[(X^{(1)},..., X^{(n)}), P_\theta]: \theta \in \Lambda\}$ if

$$R \in \mathscr{R}[(X^{(1)},..., X^{(n)}), P_\theta], \theta \in \Lambda.$$

Weighted universal coding. We state here the main results, to be proved in subsequent sections. The results are weighted universal coling theorems for noiseless coding of a family of ergodic multiterminal sources. In particular, they imply the coding theorem of Cover (1975) for a single multiterminal stationary, ergodic source, which was an extension of a result of Slepian and Wolf (1973). As a simple corollary to these theorems, which we leave to the reader, one can delineate the rate regions in E^n for noiseless coding of a stationary perhaps non-ergodic source with respect to each of the following four types of coding: fixed-rate block coding, variable-rate block-coding, fixed-rate sliding-block coding, variable-rate sliding-block coding. The rate region for fixed-rate block coding will coincide with the rate region for fixed-rate sliding-block coding. Also the rate region for variable rate block coding will coincide with the rate region for variable-rate sliding-block coding. The fixed-rate region is a subset of the variable-rate region, and may be a proper subset, unless the stationary source is ergodic, in which case the regions coincide.

The following notation is used in the statement of the theorems to follow. $(\Lambda, \mathscr{M}, \lambda)$ is a probability space and $(\Omega, \mathscr{F})$ is a measurable space. $\{P_\theta: \theta \in \Lambda\}$ is a measurable family of probability measures on $\mathscr{F}$. $X^{(1)},..., X^{(n)}$ are finite-state processes defined on Ω with state spaces $A_1,..., A_n$, respectively. For each $\theta \in \Lambda$, we assume the multiterminal source $[(X^{(1)},..., X^{(n)}), P_\theta]$ is stationary and ergodic.

THEOREM 1. *Let $\{R(\theta): \theta \in \Lambda\} \subset E^n$ be a variable-rate specification for the* *family of stationary, ergodic multiterminal sources* $\{[(X^{(1)},..., X^{(n)}), P_\theta]: \theta \in \Lambda\}$. *Then, given* $\epsilon > 0$, *there exists a positive integer* N, *block codes* $\varphi_i: A_i^\infty \to A_i^\infty$ $(i = 1,..., n)$ *of order* N, *a block code* $\delta: A_1^\infty \times \cdots \times A_n^\infty \to A_1^\infty \times \cdots \times A_n^\infty$ *of order* N, *noiseless variable-length codes* $\tau_i: A_i^N \to \{0, 1\}^*$ $(i = 1,..., n)$, *and a set* $W \subset \Lambda$ *with* $\lambda(W) > 1 - \epsilon$ *such that for each* $\theta \in W$,

(a) $P_\theta[(X^{(1)},..., X^{(n)})^N \neq \delta(\varphi_1(X^{(1)}),..., \varphi_n(X^{(n)}))^N] < \epsilon$.

(b) $N^{-1}E_{P_\theta}\ell[\tau_i(\varphi_i(X^{(i)})^N)] \leqslant R_i(\theta) + \epsilon, \qquad i = 1,..., n$.

(*Note.* In the preceding, ℓ denotes length, and E_{P_θ} denotes expectation with respect to P_θ.)

THEOREM 2. *Let* $R \in E^n$ *be a fixed-rate specification for the family of stationary, ergodic sources* $\{[(X^{(1)},..., X^{(n)}), P_\theta]\}$. *Then given* $\epsilon > 0$, *there exists a positive integer* N, *block codes* $\varphi_i: A_i^\infty \to A_i^\infty$ $(i = 1,..., n)$ *of order* N, *a block code* $\delta: A_1^\infty \times \cdots \times A_n^\infty \to A_1^\infty \times \cdots \times A_n^\infty$ *of order* N, *and a set* $W \subset \Lambda$ *with* $\lambda(W) > 1 - \epsilon$ *such that*

(a) $r(\varphi_i) < R_i + \epsilon$, $i = 1,..., n$.

(b) $P_\theta[(X^{(1)},..., X^{(n)})^N \neq \delta(\varphi_1(X^{(1)}),..., \varphi_n(X^{(n)}))^N] < \epsilon$, $\theta \in W$.

THEOREM 3. *Let* $\{R(\theta): \theta \in \Lambda\}$ *be a variable-rate specification for the family of stationary, ergodic sources* $\{[(X^{(1)},..., X^{(n)}), P_\theta]\}$. *Then, given* $\epsilon > 0$, *there exist sliding-block codes* $\psi_i: A_i^\infty \to A_i^\infty$ $(i = 1,..., n)$, *a sliding-block code* $\delta: A_1^\infty \times \cdots \times A_n^\infty \to A_1^\infty \times \cdots \times A_n^\infty$, *noiseless variable-length codes* $\tau_i: A_i^M \to \{0, 1\}^*$ $(i = 1,..., n)$ *for some* M, *and a set* $W \subset \Lambda$ *with* $\lambda(W) > 1 - \epsilon$ *such that for each* $\theta \in W$

(a) $P_\theta[(X^{(1)},..., X^{(n)})_0 \neq \delta(\psi_1(X^{(1)}),..., \psi_n(X^{(n)}))_0] < \epsilon$.

(b) $M^{-1}E_{P_\theta}\ell[\tau_i(\psi_i(X^{(i)})^M)] \leqslant R_i(\theta) + \epsilon$, $i = 1,..., n$.

THEOREM 4. *Let* R *be a fixed-rate specification for the stationary, ergodic sources* $\{[(X^{(1)},..., X^{(n)}), P_\theta]\}$. *Given* $\epsilon > 0$, *there exist sliding-block codes* $\psi_i: A_i^\infty \to A_i^\infty$ $(i = 1,..., n)$, *a sliding-block code* $\delta: A_1^\infty \times \cdots \times A_n^\infty \to A_1^\infty \times \cdots \times A_n^\infty$, *and a set* $W \subset \Lambda$ *w th* $\lambda(W) > 1 - \epsilon$ *such that*

(a) $r(\psi_i) < R_i + \epsilon$, $i = 1,..., n$.

(b) $P_\theta[(X^{(1)},..., X^{(n)})_0 \neq \delta(\psi_1(X^{(1)}),..., \psi_n(X^{(n)}))_0] < \epsilon$, $\theta \in W$.

III. BUILDING A GOOD BLOCK CODE

If X is a discrete random variable on a probability space $(\Omega, \mathscr{F}, P)$, let $P(X)$ denote the function from Ω to $[0, 1]$ such that

$$P(X)(\omega) = P[X = X(\omega)], \qquad \omega \in \Omega.$$

If Y is another discrete random variable, let $P(X \mid Y)$ denote the function

$$P(X \mid Y) = P(X, Y)/P(Y), \qquad P(Y) > 0$$
$$= 0 \qquad , \qquad \text{elsewhere.}$$

The following coding lemma allows us to give an easy proof of Theorems 1 and 2. The proof uses a type of random coding argument due to Cover (1975).

LEMMA 1. *Let $A_1, \ldots, A_n$ be finite sets. Let X_i $(i = 1, \ldots, n)$ be the projection of $A_1 \times \cdots \times A_n$ onto A_i. For each i, let a map $f_i\colon A_i \to [0, \infty)$ be given. Let P be a probability measure on $A_1 \times \cdots \times A_n$. Given $c > 0$, there exist maps $\varphi_i\colon A_i \to A_i$ $(i = 1, \ldots, n)$, a map $\delta\colon A_1 \times \cdots \times A_n \to A_1 \times \cdots \times A_n$, and noiseless variable-length codes $\tau_i\colon A_i \to \{0, 1\}^*$ $(i = 1, \ldots, n)$ such that*

$$\text{(a)} \quad P[(X_1, \ldots, X_n) \neq \delta(\varphi_1(X_1), \ldots, \varphi_n(X_n))] \leqslant 2^{n-c}$$
$$+ \sum_{\substack{S \subset \{1, \ldots, n\} \\ S \neq \varnothing}} P\left[P((X_j : j \in S) \mid (X_j : j \notin S)) < \prod_{j \in S} 2^{-f_j(X_j)}\right];$$

$$\text{(b)} \quad \ell[\tau_i(\varphi_i(X_i))] \leqslant \log |f_i(A_i)| + f_i(X_i) + c + 1, \; i = 1, \ldots, n;$$

$$\text{(c)} \quad \log |\varphi_i(A_i)| \leqslant \log |f_i(A_i)| + \max f_i(A_i) + c + 1, \; i = 1, \ldots, n.$$

(*Note.* In (a), by $P((X_j : j \in S) | (X_j : j \notin S))$ we mean the function $P((X_j : j \in S))$ if $S = \{1, \ldots, n\}$.)

Proof. If S is a finite set, we will call a map $\sigma\colon S \to \{1, 2, \ldots\}$ a length function if $\sum_{y \in S} 2^{-\sigma(y)} \leqslant 1$. From Gallager (1968, Chapter 3), if $\tau\colon S \to \{0, 1\}^*$ is a noiseless variable-length code then the formula $\sigma(y) = \ell[\tau(y)]$ defines a length function on S; conversely, given a length function σ on S, there is a noiseless variable-length code $\tau\colon S \to \{0, 1\}^*$ such that $\sigma(y) = \ell[\tau(y)]$, $y \in S$. Thus, to prove Lemma 1, all we need to find are maps $\varphi_i\colon A_i \to A_i$ $(i = 1, \ldots, n)$, a map $\delta\colon A_1 \times \cdots \times A_n \to A_1 \times \cdots \times A_n$, and length functions $\sigma_i\colon A_i \to \{1, 2, \ldots\}$ $(i = 1, \ldots, n)$ such that (a), (c) hold and

$$\text{(b')} \quad \sigma_i(\varphi_i(X_i)) \leqslant \log |f_i(A_i)| + f_i(X_i) + c + 1, \qquad i = 1, \ldots, n.$$

Let $C_i = f_i(A_i)$, $i = 1, \ldots, n$. Let $T = \{(i, x, y) : i = 1, \ldots, n;\ x \in A_i;\ y \in C_i\}$. Let $D_{i,x,y} = \{1, \ldots, \lceil 2^{y+c} \rceil\}$, $(i, x, y) \in T$. (If r is a real number, $\lceil r \rceil$ denotes the smallest integer $\geqslant r$.) Let $D = \prod_{(i,x,y) \in T} D_{i,x,y}$. For each i, let $B_i = \bigcup_{y \in C_i} \{1, \ldots, \lceil 2^{y+c} \rceil\} \times \{y\}$. For each $i = 1, \ldots, n$, and $z = (z_{j,x,y} : (j, x, y) \in T) \in D$, let $\varphi_i^z\colon A_i \to B_i$ be the map

$$\varphi_i^z(x) = (z_{i,x,f_i(x)}, f_i(x)), \qquad x \in A_i.$$

Let $\sigma_i\colon B_i \to \{1, 2, \ldots\}$ be the length function such that

$$\sigma(k, y) = \log | C_i | + \lceil y + c \rceil, \qquad (k, y) \in B_i.$$

Let E be the subset of $A_1 \times \cdots \times A_n$ such that

$$E = \bigcap_{\substack{S \subset \{1,\ldots,n\} \\ S \neq \varnothing}} \left\{ P((X_j : j \in S) \mid (X_j : j \notin S)) \geqslant \prod_{j \in S} 2^{-f_j(X_j)} \right\}$$

For each $z \in D$, let $\delta_z \colon B_1 \times \cdots \times B_n \to A_1 \times \cdots \times A_n$ be a map such that if $(k_1, y_1) \in B_1, \ldots, (k_n, y_n) \in B_n$ then $\delta_z((k_1, y_1), \ldots, (k_n, y_n)) = (x_1, \ldots, x_n)$ if $(x_1, \ldots, x_n)$ is the only element of E such that

(d) $f_i(x_i) = y_i, \quad i = 1, \ldots, n,$

(e) $z_{i, x_i, y_i} = k_i, \quad i = 1, \ldots, n.$

On some probability space $(\Omega, \mathscr{F}, \lambda)$ we may define random variables $X_1', \ldots, X_n'$, $\{Z_{i,x,y} \colon (i, x, y) \in T\}$ such that

(f) each X_i' is A_i-valued and the distribution of $(X_1', \ldots, X_n')$ is P;

(g) for each $(i, x, y) \in T$, $Z_{i,x,y}$ is uniformly distributed over $\{1, \ldots, \lceil 2^{y+c} \rceil\}$;

(h) $\{Z_{i,x,y} \colon (i, x, y) \in T\}$ are independent;

(i) $(X_1', \ldots, X_n')$ and the D-valued random variable $Z = (Z_{i,x,y} \colon (i, x, y) \in T)$ are independent.

Let Q denote the quantity on the right-hand side of the inequality in (a). If we can show that

(j) $\lambda[(X_1', \ldots, X_n') \neq \delta_Z(\varphi_1^Z(X_1'), \ldots, \varphi_n^Z(X_n'))] \leqslant Q,$

then because of (i), we will have for some $z \in D$ that

(k) $P[(X_1, \ldots, X_n) \neq \delta_z(\varphi_1^z(X_1), \ldots, \varphi_n^z(X_n))] \leqslant Q.$

We now try to derive (j). The left-hand side of (j) is no bigger than

(l) $P[(X_1, \ldots, X_n) \notin E] + \sum_{(y_1, \ldots, y_n) \in C_1 \times \cdots \times C_n} \lambda[f_i(X_i') = y_i \quad (i = 1, \ldots, n),$

and there exists in E a $(x_1, \ldots, x_n) \neq (X_1', \ldots, X_n')$ such that

$f_i(x_i) = y_i$ and $Z_{i, x_i, y_i} = Z_{i, X_i', y_i}$ for all i].

For each $(y_1, \ldots, y_n) \in C_1 \times \cdots \times C_n$, the summand in (l) is no bigger than

(m) $\sum_{x'} P(x') \sum_{S} \sum_{x \in E_S} \lambda[Z_{j, x_j, y_j} = Z_{j, x_j', y_j}, j \in S],$

where the outermost sum is over all $x' = (x_1', \ldots, x_n') \in A_1 \times \cdots \times A_n$ such that $P(x') > 0$ and $f_i(x_i') = y_i$ for all i, the middle sum is over all nonempty subsets

S of $\{1,\ldots, n\}$, and in the innermost sum E_S represents the set of all $x \in \prod_{j\in S} A_j$ such that $x_j \neq x_j'$, $j \in S$, and

$$P[(X_j: j\in S) = x \mid X_j = x_j', j \notin S] \geqslant \prod_{j\in S} 2^{-y_j}.$$

(The middle sum arises by observing that if $x, x' \in A_1 \times \cdots \times A_n$ and $x \neq x'$ then for some nonempty $S \subset \{1,\ldots, n\}$, we have $x_j \neq x_j'$ if and only if $j \in S$.) Now

$$\lambda[Z_{j,x_j,y_j} = Z_{j,x_j',y_j}, j \in S] = \left(\prod_{j\in S} \lceil 2^{y_j+c}\rceil\right)^{-1},$$

since all the variables involved are independent and $x_j \neq x_j'$, $j \in S$. Calculating the innermost sum in (m) we get $|E_S|(\prod_{j\in S}\lceil 2^{y_j+c}\rceil)^{-1}$. Since each $x \in E_S$ has a probability lower bounded by $\prod_{j\in S} 2^{-y_j}$, we must have $|E_S| \leqslant \prod_{j\in S} 2^{y_j}$. We can now observe that (j) will follow. Thus we may fix $z \in D$ such that (k) holds. Setting $\varphi_i = \varphi_i^z$ and $\delta = \delta_z$, we get (a), (c), (b′). Since for each i, $|\varphi_i(A_i)| \leqslant |A_i|$, we can assume $B_i = A_i$, $i = 1,\ldots, n$.

Proof of Theorems 1 *and* 2. As shown in the proof of Theorem 4 of Kieffer (1980a), we can assume without loss of generality that $A = \Omega = A_1^\infty \times \cdots \times A_n^\infty$, that each $X^{(i)}$ is the projection from $A_1^\infty \times \cdots \times A_n^\infty \to A_i^\infty$, and that the measures $\{P_\theta: \theta \in \Omega\}$ are the ergodic components of the measure λ. More precisely, we assume each $P_\theta \in \mathscr{E}(A_1,\ldots, A_n)$ and that

(a) $P_\theta(E) = \lim_{k\to\infty} k^{-1}\sum_{i=0}^{k-1} I_E(T^i_{A_1,\ldots,A_n}\theta)$, for λ-almost all $\theta \in \Omega$, where I_E denotes the indicator function of the set $E \in \mathcal{A}_1^\infty \times \cdots \times \mathcal{A}_n^\infty$,

(b) $P\{\theta: P_\theta = P\} = 1$, $P \in \mathscr{E}(A_1,\ldots, A_n)$,

(c) $\lambda(E) = \int_\Omega P_\theta(E)\, d\lambda(\theta)$, $E \in \mathcal{A}_1^\infty \times \cdots \times \mathcal{A}_n^\infty$.

Let $\{R(\theta): \theta \in \Omega\}$ be a variable-rate specification. For each $i = 1,\ldots, n$, let $R_i : \Omega \to [0, \infty)$ be the function such that $R_i(\theta)$ is the ith component of $R(\theta)$, $\theta \in \Omega$. Now $R_i(\theta)$ depends on θ through $P_\theta^{X^{(i)}}$, and by (a), $P_\theta^{X^{(i)}}$ depends on θ through $X^{(i)}(\theta)$. Hence, given $\delta > 0$, there is a finite set $C_i \subset [0, \infty)$ and for each N a function $F_i^N: A_i^N \to C_i$ such that the functions $\{F_i^N([X^{(i)}]^N)\}$ converge almost surely with respect to λ as $N \to \infty$, and

$$R_i + \delta \leqslant \lim_{N\to\infty} F_i^N([X^{(i)}]^N) \leqslant R_i + 2\delta \qquad \text{a.s. } [\lambda]. \tag{3.1}$$

By a result of Parthasarathy (1963), if S is a nonempty subset of $\{1,\ldots, n\}$, for λ-almost all θ

$$\lim_{N\to\infty} - N^{-1}\log \lambda(([X^{(j)}]^N: j \in S) \mid ([X^{(j)}]^N: j \notin S))(\theta)$$

$$= \bar{H}_{P_\theta}((X^{(j)}: j \in S) \mid (X^{(j)}: j \notin S)) \leqslant \sum_{j\in S} R_j(\theta).$$

Therefore,

$$\lim_{N\to\infty} \lambda \left[\lambda(([X^{(j)}]^N: j \in S) \mid ([X^{(j)}]^N: j \notin S)) < \prod_{j\in S} 2^{-NF_j{}^N([X^{(j)}]^N)}\right] = 0. \quad (3.2)$$

Applying Lemma 1, for N sufficiently large there exist block codes $\varphi_i: A_i{}^\infty \to A_i{}^\infty$ $(i = 1,..., N)$ of order N, a block code $\delta: A_1{}^\infty \times \cdots \times A_n{}^\infty \to A_1{}^\infty \times \cdots \times A_n{}^\infty$ of order N, and noiseless variable-length codes $\tau_i: A_i{}^N \to \{0, 1\}^*$ such that

(d) $N^{-1}\ell[\tau_i(\varphi_i(X^{(i)})^N)] \leqslant \delta + F_i{}^N([X^{(i)}]^N)$,

(e) $\lambda[(X^{(1)},..., X^{(n)})^N \neq \delta(\varphi_1(X^{(1)}),..., \varphi_n(X^{(n)}))^N] \to 0$.

From (3.1) and (d), we obtain

(f) $\limsup_{N\to\infty} N^{-1}\ell[\tau_i(\varphi_i(X^{(i)})^N)] \leqslant 3\delta + R_i$ a.s. $[\lambda]$.

Taking a conditional expectation, since $R_i = R_i(\theta)$ a.s. $[P_\theta]$, (f), (e) give

(g) $P_\theta[\limsup_{N\to\infty} N^{-1}\ell[\tau_i(\varphi_i(X^{(i)})^N)] \leqslant 3\delta + R_i(\theta)] = 1$, a.s. $[\lambda]$.

(h) $P_\theta[(X^{(1)},..., X^{(n)})^N \neq \delta(\varphi_1(X^{(1)}),..., \varphi_n(X^{(n)}))^N] \to 0$ stochastically with respect to λ.

Theorem 1 follows from (g), (h) by a simple application of Egoroff's theorem (Ash, 1972, p. 94), provided we take δ to be small enough relative to ϵ. If $(R_1 ,..., R_n)$ is a fixed-rate specification, note that (3.2) holds with $F_j{}^N([X^{(j)}]^N)$ replaced by $R_j + \delta$. One now applies part (c) of Lemma 1.

IV. Building a Good Sliding-Block Code

In this section we prove Lemma 2 which will allow us to build a good sliding-block code from a good block code, and thereby enable us to prove Theorems 3 and 4. Before proceeding with the Lemma, we need to introduce some more notation.

Let $A_1 ,..., A_n$ be finite sets. For $N = 1, 2,...$, let $\mathscr{P}_N(A_1 ,..., A_n)$ denote the set of all probability measures on $\mathcal{O}\!\ell_1{}^\infty \times \cdots \times \mathcal{O}\!\ell_n{}^\infty$ stationary with respect to $T^N_{A_1,...,A_n}$. Let $\mathscr{P}_\infty(A_1 ,..., A_n) = \bigcup_{N=1}^\infty \mathscr{P}_N(A_1 ,..., A_n)$. We define $f: A_1{}^\infty \times \cdots \times A_n{}^\infty \to [0, \infty)$ to be finite-dimensional (f.d.) if for some positive integer M,

$$f(x_1 ,..., x_n) = f(y_1 ,..., y_n) \qquad \text{if } [x_i]_{-M}^M = [y_i]_{-M}^M , \qquad i = 1,..., n.$$

If $\{\mu_k: k = 1, 2,...\} \cup \{\mu\} \subset \mathscr{P}_1(A_1 ,..., A_n)$ we say $\mu_k \to \mu$ weakly if $E_{\mu_k} f \to E_\mu f$ for every f.d. $f: A_1{}^\infty \times \cdots \times A_n{}^\infty \to [0, \infty)$. The weak topology on $\mathscr{P}_1(A_1 ,..., A_n)$ is the unique metric topology with this convergence (see Parthasarathy, 1967).

Fix finite sets A, B and let $X: A^\infty \times B^\infty \to A^\infty$ and $Y: A^\infty \times B^\infty \to B^\infty$ be the maps such that $X(x, y) = x$, $Y(x, y) = y$.

We call $F: \mathscr{P}_\infty(A, B) \to [0, \infty)$ a nice function if

(a) F is affine on the convex set $\mathscr{P}_\infty(A, B)$; that is, if $\mu, \nu \in \mathscr{P}_\infty(A, B)$ and $0 < \alpha < 1$, then $F(\alpha\mu + (1-\alpha)\nu) = \alpha F(\mu) + (1-\alpha)F(\nu)$.

(b) F is uppersemicontinuous on $\mathscr{P}_1(A, B)$ relative to the weak topology.

(c) If $\mu \in \mathscr{P}_\infty(A, B)$ and $\hat{Y}$ is a process with state space B which is a stationary or block coding of H satisfying $\bar{H}_\mu(Y \mid \hat{Y}) = 0$, then $F(\mu) = F(\mu^{(X \mid \hat{Y})})$.

(d) $F(\mu) = F(\mu \cdot T_{A,B}^{-1})$, $\mu \in P_\infty(A, B)$.

As an example of a nice function, we cite the map $\mu \to \bar{H}_\mu(X \mid Y)$.

A channel is a triple $[A, \tau, B]$ where A, B are finite sets and $\tau = \{\tau_x: x \in A^\infty\}$ is a measurable family of probability measures on B^∞.

We call a sequence $x \in A^\infty$ periodic if for some n $T_A{}^n x = x$. If x is periodic, define the period of x to be the smallest n such that $T_A{}^n x = x$.

If S_1, S_2 are subsets of some common set, define $S_1 - S_2 = \{\omega \in S_1: \omega \notin S_2\}$.

LEMMA 2. *Let $(\Omega, \mathscr{F})$ be a measurable space. Let $(\Lambda, \mathscr{M}, \lambda)$ be a probability space. Let $\{P_\theta: \theta \in \Lambda\}$ be a measurable family of probability measures on $\mathscr{F}$. Let C, D be finite sets. Let U, V be processes defined on Ω with state spaces C, D, respectively. We suppose $\{U, V\}$ are jointly stationary and ergodic under each P_θ, $\theta \in \Lambda$. Let $\mathscr{C}$ be a finite collection of nice functions from $\mathscr{P}_\infty(C, D) \to [0, \infty]$. Let $\varphi: D^\infty \to D^\infty$ and $\delta: C^\infty \times D^\infty \to D^\infty$ be block codes of order N. Given $\epsilon > 0$, there exist sliding-block codes $\hat{\varphi}: D^\infty \to D^\infty$ and $\hat{\delta}: C^\infty \times D^\infty \to D^\infty$, and a subset W of Λ with $\lambda(\Lambda - W) < \epsilon$ such that if $\theta \in W$*

(a) $F(P_\theta^{(U, \hat{\varphi}(V))}) \leqslant F(P_\theta^{(U, \varphi(V))}) + \epsilon$, $F \in \mathscr{C}$,

(b) $P_\theta[V_0 \neq \hat{\delta}(U, \hat{\varphi}(V))_0] \leqslant P_\theta[V^N \neq \delta(U, \varphi(V))^N] + \epsilon$.

Proof. By Theorem 3.1 of Gray (1975), it suffices to find stationary codes $\hat{\varphi}$, $\hat{\delta}$ for which (a), (b) hold. If $r(\varphi) = \log |D|$, then (a), (b) hold with $\hat{\varphi}$ the identity map, $\hat{\delta}(u, y) \equiv y$. So we can assume $r(\varphi) < \log |D|$. From the theory of ergodic processes, given $\theta \in \Lambda$, the process V is either aperiodic under P_θ (which means that $P_\theta(V = v) = 0$, $v \in D^\infty$), or is periodic under V (which means that for some n, there is a periodic $v \in D^\infty$ with period n such that $P_\theta(V = T_D{}^i v) = n^{-1}$, $0 \leqslant i \leqslant n-1$). Let $W_0 = \{\theta \in \Lambda: V \text{ is aperiodic under } P_\theta\}$, $W_1 = \{\theta \in \Lambda: V \text{ is periodic under } P_\theta\}$. Choose k a multiple of N and $W_2 \subset W_1$ so that $\lambda(W_1 - W_2) < \epsilon/3$ and for every $\theta \in W_2$

$$P_\theta(V \text{ is periodic with period} \leqslant k) = 1.$$

Since $r(\varphi) < \log |D|$, there exists for some multiple L of k a $b \in D^L$ such that $b \notin \{\varphi(v)^L: v \in D^\infty\}$ and the sequence $\tilde{x}$ in D^∞ such that $\tilde{x}_{iL+1}^{iL+L} = b$ $(i \in Z)$ has period L. For each multiple j of L such that $j > 2L$, define $\varphi_j\colon D^\infty \to D^\infty$ to be the block code of order $j + 2L$ such that

$$\varphi_j(x)_{i+1}^{i+2L} = (b, b) \qquad \text{if } i \equiv 0 \bmod j + 2L,$$
$$\varphi_j(x)_s = \varphi(x)_s \qquad \text{for all other coordinates } s.$$

Define $\delta_j\colon C^\infty \times D^\infty \to D^\infty$ to be a sliding-block code such that

(c) $\delta_j(u, y) = T_{C,D}^{-s}\delta(T_C{}^s u, T_D{}^s y)$ if $\{i \in Z\colon y_{i+1}^{i+2L} = (b, b)\} = \{i \in Z\colon i \equiv s \bmod j + 2L\}$ for some $0 \leqslant s \leqslant j + 2L - 1$.

(d) $\delta_j(u, y) = y$ if y is periodic with period $\leqslant k$.

Fix $\bar{U}$, $\bar{V}$, $\bar{Y}$ to be the processes defined on $C^\infty \times D^\infty \times D^\infty$ with respective state spaces C, D, D such that $\bar{U}(u, v, y) = u$, $\bar{V}(u, v, y) = v$, $\bar{Y}(u, v, y) = y$. If P is a probability measure on $C^\infty \times D^\infty$, and $[D, \nu, D]$ is a channel, let $P\nu$ be the probability measure on $C^\infty \times D^\infty \times D^\infty$ such that under $P\nu$, $\bar{U}$, $\bar{V}$, $\bar{Y}$ form a Markov chain, the distribution of $(\bar{U}, \bar{V})$ is P, and the distribution of $\bar{Y}$ conditioned on $\bar{V}$ is given by ν. Let $[D, \tau, D]$, $[D, \tau_j, D]$ be the channels such that for each $x \in D^\infty$, τ_x is equidistributed over $\{T_D^{-i}(\varphi(T_D{}^i x))\colon 0 \leqslant i \leqslant N - 1\}$ and $(\tau_j)_x$ is equidistributed over $\{T_D^{-i}(\varphi_j(T_D{}^i x))\colon 0 \leqslant i \leqslant j + 2L - 1\}$. It can be seen that for all $\theta \in \Lambda$,

(e) $P_\theta \tau_j \to P_\theta \tau$ weakly

(f) $\lim_{j\to\infty} P_\theta \tau_j[\bar{V}_0 \neq \delta_j(\bar{U}, \bar{Y})_0] \leqslant P_\theta[V^N \neq \delta(U, \varphi(V))^N]$.

By (e), for each $\theta \in \Lambda$ and each $F \in \mathscr{C}$,

(g) $\limsup_{j\to\infty} F(P_\theta \tau_j^{(\bar{U},\bar{Y})}) \leqslant F(P_\theta \tau^{(\bar{U},\bar{Y})}) = F(P_\theta^{(U,\varphi(V))})$.

Hence by Egoroff's theorem, there is $W_3 \subset W_0$ with $\lambda(W_0 - W_3) < \epsilon/3$ and j so large that setting $\hat{\tau} = \tau_j$, $\hat{\delta} = \delta_j$, we have for $\theta \in W_3$ that

(h) $P_\theta \hat{\tau}[\bar{V}_0 \neq \hat{\delta}(\bar{U}, \bar{Y})_0] \leqslant P_\theta[V^N \neq \delta(U, \varphi(V))^N] + \epsilon/2$.

(i) $F(P_\theta \hat{\tau}^{(\bar{U},\bar{Y})}) \leqslant F(P_\theta^{(U,\varphi(V))}) + \epsilon/2, F \in \mathscr{C}$.

By Lemma 6 of Kieffer (1980b) and Theorem 2 of Kieffer and Rahe (1981), there is a sequence $\{\psi_j\}$ of sliding-block codes from $D^\infty \to D^\infty$ such that $P_\theta \hat{\tau}^{(\bar{V},\psi_j(\bar{V}))} \to P_\theta \hat{\tau}^{(\bar{V},\bar{Y})}$ weakly, for every $\theta \in W_0$. By Lemma 5 of the Appendix, $P_\theta^{(U,V,\psi_j(V))} = P_\theta \hat{\tau}^{(\bar{U},\bar{V},\psi_j(\bar{V}))} \to P_\theta \hat{\tau}$, for every $\theta \in W_0$. Applying Egoroff's theorem again, we obtain $W_4 \subset W_3$ with $\lambda(W_3 - W_4) < \epsilon/3$ and j so large that setting $\psi = \psi_j$ we have for every $\theta \in W_4$,

(j) $P_\theta[V_0 \neq \hat{\delta}(U, \varphi(V))_0] \leqslant P_\theta[V^N \neq \delta(U, \varphi(V))^N] + \epsilon$,

(k) $F(P_\theta^{(U,\psi(V))}) \leqslant F(P_\theta^{(U,\varphi(V))}) + \epsilon, F \in \mathscr{C}$.

Define $\hat{\varphi}\colon D^\infty \to D^\infty$ to be the stationary code such that $\hat{\varphi}(x) = x$, if x is periodic; $\varphi = \psi$, otherwise. Set $W = W_4 \cup W_2$.

In the following, let $\mathbf{1}$ denote the n-vector $(1, 1, \ldots, 1)$, and let $h(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$, $0 < \alpha \leqslant 1/2$.

LEMMA 3. *Let the notation preceding Theorem* 1 *prevail. Let* $\{R(\theta)\colon \theta \in \Lambda\}$ *be a variable-rate specification for the family of stationary, ergodic sources* $\{[(X^{(1)}, \ldots, X^{(n)}), P_\theta]\colon \theta \in \Lambda\}$. *Given* $\epsilon > 0$ *there exists a process* U *with state space* A_1 *which is sliding-block coding of* $X^{(1)}$, *a process* $\tilde{X}^{(1)}$ *with state space* A_1 *which is a sliding-block coding of* $(U, X^{(2)}, \ldots, X^{(n)})$, *and a set* $W \subset \Lambda$ *with* $\lambda(W) > 1 - \epsilon$ *such that*:

(a) $\{R(\theta) + \epsilon\mathbf{1}\colon \theta \in W\}$ *is a variable-rate specification for* $\{[(U, X^{(2)}, \ldots, X^{(n)}), P_\theta]\colon \theta \in W\}$.

(b) $P_\theta(X_0^{(1)} \neq \tilde{X}_0^{(1)}) < \epsilon$, $\theta \in W$.

(c) $\bar{H}_{P_\theta}(U) \leqslant R_1(\theta) + \epsilon$, $\theta \in W$.

Proof. Let $M = \max_i \log |A_i|$. Choose $\alpha > 0$ so small that $\alpha + h(\alpha) + M\alpha < \epsilon/2$, $2\alpha < \epsilon$, $\alpha < 1/2$. By Theorem 1, there exists a positive integer N, block codes $\varphi_i\colon A_1^\infty \to A_1^\infty$ of order $N (i = 1, \ldots, n)$, a block code $f\colon A_1^\infty \times \cdots \times A_n^\infty \to A_1^\infty \times \cdots \times A_n^\infty$ of order N, and a set $W_1 \subset \Lambda$ with $\lambda(W_1) > 1 - \epsilon/2$ such that for $\theta \in W_1$,

(d) $P_\theta[(X^{(1)}, \ldots, X^{(n)})^N \neq f(\varphi_1(X^{(1)}), \ldots, \varphi_n(X^{(n)}))^N] < \alpha$.

(e) $\bar{H}_{P_\theta}(\varphi_i(X^{(i)})) \leqslant R_i(\theta) + \alpha$, $i = 1, \ldots, n$.

Because of (d), there exists a block code $g\colon A_1^\infty \times \cdots \times A_n^\infty \to A_1^\infty \times \cdots \times A_n^\infty$ of order N such that for all $\theta \in W_1$,

(f) $P_\theta[(\varphi_1(X^{(1)}), X^{(2)}, \ldots, X^{(n)})^N \neq g(\varphi_1(X^{(1)}), \varphi_2(X^{(2)}), \ldots, \varphi_n(X^{(n)}))^N] < \alpha$.

Applying (e), (f) and Lemma 4 of the Appendix, we see that

(g) $R(\theta) + (\epsilon/2)\mathbf{1} \in \mathscr{R}[(\varphi_1(X^{(1)}), X^{(2)}, \ldots, X^{(n)}), P_\theta]$, $\theta \in W_1$.

Also, because of (d) there exists a block code $h\colon A_1^\infty \times \cdots \times A_n^\infty \to A_1^\infty \times \cdots \times A_n^\infty$ of order N such that

(h) $P_\theta[(X^{(1)}, \ldots, X^{(n)})^N \neq h(\varphi_1(X^{(1)}), X^{(2)}, \ldots, X^{(n)})^N] < \epsilon/2$, $\theta \in W_1$.

Applying Lemma 2, we see from the statements (e), (g), (h) that there must exist $W \subset W_1$ with $\lambda(W_1 - W) < \epsilon/2$, a sliding-block coding U of $X^{(1)}$ and a sliding-block coding $\tilde{X}^{(1)}$ of $(U, X^{(2)}, \ldots, X^{(n)})$ such that (a)–(c) hold.

Proof of Theorems 3 *and* 4. Let $\{R(\theta)\colon \theta \in \Lambda\}$ be a variable-rate specification for the family of multiterminal sources given in Theorem 3. We note that in place of (b) of Theorem 3, we need only show that for $\theta \in W$ we have

(b′) $\bar{H}_{P\theta}(\psi_i(X^i)) \leqslant R_i(\theta) + \epsilon/2$, $i = 1, \ldots, n$.

For, by a weak universal noiseless coding theorem (Kieffer, 1978, Theorem 1), (b′) implies that (b) holds for some M and some noiseless variable-length code

τ_i: $A_i^M \to \{0, 1\}^*$, provided we reduce W by a λ-small amount. To get condition (b') above and condition (a) of Theorem 3 to hold, apply Lemma 3 n times. Therefore Theorem 3 follows, and then Theorem 4 follows from Theorem 3. For, if U is a finite-state process ergodic with respect to each P_θ, and $\bar{H}_{P_\theta}(U) < K$ for each θ, by (Ziv, 1972, Theorem 4) and (Kieffer, 1980a, Theorem 1) there exists for each $\epsilon > 0$ a sequence of sliding-block codes $\{\varphi_n\}$ such that $r(\varphi_n) < K + \epsilon$ for all n and for every θ, $P_\theta(U_0 \neq \varphi_n(U)_0) \to 0$.

Appendix

LEMMA 4. *Let $X^{(1)},..., X^{(n)}$ be processes defined on $(\Omega, \mathscr{F})$ with finite state spaces $A_1,..., A_n$. Let P be a probability measure on $\mathscr{F}$ with respect to which $\{X^{(1)},..., X^{(n)}\}$ are jointly N-stationary. Let φ_i: $A_i^\infty \to A_i^\infty$ $(i = 1,..., n)$ and δ: $A_1^\infty \times \cdots \times A_n^\infty \to A_1^\infty \times \cdots \times A_n^\infty$ be block codes of order N such that*

$$P[(X^{(1)},..., X^{(n)})^N \neq \delta(\varphi_1(X^{(1)}),..., \varphi_n(X^{(n)}))^N] \leqslant \epsilon \leqslant 1/2.$$

Then

$$R \in \mathscr{R}[(X^{(1)},..., X^{(n)}), P],$$

where

$$R_i = N^{-1}H(\varphi_i(X^{(i)})^N) + h(\epsilon) + \epsilon \log |A_i|, \qquad i = 1,..., n.$$

Proof. Let S be a nonempty subset of $\{1,..., n\}$. Let $U = (X^{(j)}: j \notin S)$, $V = (X^{(j)}: j \in S)$, $\hat{V} = (\varphi_j(X^{(j)}): j \in S)$, $C = \prod_{j \notin S} A_j$, $D = \prod_{j \in S} A_j$. We regard U, V, $\hat{V}$ as processes with state space C, D, D, respectively. It is easy to see that there is a block code δ': $C^\infty \times D^\infty \to D^\infty$ such that

$$P[V^N \neq \delta'(U, \hat{V})^N] \leqslant \epsilon.$$

By Fano's inequality (Ash, 1965, p. 80)

$$\begin{aligned} \bar{H}((X^{(j)}: j \in S) \mid (X^{(j)}: j \notin S)) = \bar{H}(V \mid U) &\leqslant N^{-1}(V^N \mid U^N) \\ &\leqslant N^{-1}H(\hat{V}^N) + N^{-1}H(V^N \mid \hat{V}^N, U^N) \\ &\leqslant \sum_{j \in S} N^{-1}H(\varphi_j(X^{(j)})^N) + h(\epsilon) + \epsilon \log |D| \\ &\leqslant \sum_{j \in S} R_j . \end{aligned}$$

LEMMA 5. *Let U, X, Y be processes defined on the probability space $(\Omega, \mathscr{F}, P)$ with state spaces A, B, C, respectively. Suppose that with respect to P these processes are jointly stationary and form a Markov chain (in the indicated order). Let $\{\varphi_n\}$*

be a sequence of sliding-block codes from $B^\infty \to C^\infty$ *such that* $P^{(X,\varphi_n(X))} \to P^{(X,Y)}$ *weakly. Then,* $P^{(U,X,\varphi_n(X))} \to P^{(U,X,Y)}$ *weakly.*

Proof. We have to show that

$$E[f(U)\,g(X)h(\varphi_n(X))] \to E[f(U)\,g(X)h(Y)],$$

for f.d. functions taking their values in [0, 1]. Using the Markov property, we see that

$$E[f(U)\,g(X)h(Y)] = E[E[f(U) \mid X]\,g(X)h(Y)]$$
$$E[f(U)\,g(X)h(\varphi_n(X))] = E[E[f(U) \mid X]\,g(X)h(\varphi_n(X))].$$

Fix $\epsilon > 0$. Find a f.d. function F such that

$$E[\,|F(X) - E[f(U) \mid X]|] < \epsilon/3.$$

Then,

$$\begin{aligned} &|\,E[f(U)\,g(X)h(Y)] - E[f(U)\,g(X)h(\varphi_n(X))]| \\ &\quad\leqslant |\,E[F(X)\,g(X)h(Y)] - E[F(X)\,g(X)h(\varphi_n(X))]| + 2\epsilon/3 < \epsilon, \end{aligned}$$

for n sufficiently large.

RECEIVED: May 5, 1979; REVISED: June 19, 1980.

REFERENCES

ASH, R. B. (1965), "Information Theory," Interscience–Wiley, New York.

ASH, R. B. (1972), "Real Analysis and Probability," Academic Press, New York.

COVER, T. M. (1975), A proof of the data-compression theorem of Slepian and Wolf for ergodic sources, *IEEE Trans. Inform. Theory* **21**, 226–228.

GALLAGER, R. G. (1968), "Information Theory and Reliable Communication," Wiley,, New York.

GRAY, R. M. (1975), Sliding-block source coding, *IEEE Trans. Inform. Theory* **21**, 357–368.

KIEFFER, J. C. (1978), A unified approach to weak universal source coding, *IEEE Trans. Inform. Theory* **24**, 674–682.

KIEFFER, J. C. (1980a), Extension of source coding theorems for block codes to sliding-block codes, *IEEE Trans. Inform. Theory*, in press.

KIEFFER, J. C. (1980b), On the transmission of Bernoulli sources over stationary channels, *Annals of Probability*, to appear.

KIEFFER, J. C., AND RAHE, M. (1981), Selecting universal partitions in ergodic theory, *Annals of Probability*, to appear.

PARTHASARATHY, K. R. (1963), Effective entropy rate and transmission of information through channels with additive random noise, *Sankhya Ser. A* **25**, 75–84.

PARTHASARATHY, K. R. (1967), "Probability Measures on Metric Spaces," Academic Press, New York.

SLEPIAN, D., AND WOLF, J. (1973), Noiseless coding of correlated information sources, *IEEE Trans. Inform. Theory* **19**, 471–480.

ZIV, J. (1972), Coding of sources with unknown statistics—Part I: Probability of encoding error, *IEEE Trans. Inform. Theory* **18**, 384–389.