KYBERNETIKA — VOLUME 18 (1982), NUMBER 6

A LOCAL STRUCTURE OF STATIONARY PERFECTLY NOISELESS CODES BETWEEN STATIONARY NON-ERGODIC SOURCES

II: Applications

ŠTEFAN ŠUJAN

The general results from Part I are applied in order to determine a complete set of invariants for the class of all conditionally Bernoulli sources, with respect to both metric and finitary isomorphisms. The results extend Ornstein and Keane-Smorodinsky isomorphism theorems. A generalization of Krieger's finite generator theorem is proved in a form which exhibits a close connection between generator problems of ergodic theory and noiseless source coding. A finitary version of it gives an extension of an almost topological generator theorem of Denker and Keane. A recent result of Kieffer on zero-error transmission of ergodic sources over stationary channels is generalized to transmission of aperiodic non-ergodic sources. A sufficient condition for ε-transmissibility is derived which results in a new interpretation of ε-rates. Several related problems are investigated.

INTRODUCTION

This paper is a continuation of the first part under the same title, hereafter referred to as [I]. We assume that the reader is familiar with notations, definitions, and results presented there. Throughout the paper references to Part I are indicated by writing, e.g., Lemma I.1, formula (I.15), while single numbering, e.g., formula (10), refers to the present text. References [1–19] are given in Part I, references [20–38] are listed at the end of this paper.

The whole paper is devoted to applications of general results obtained in [I] to various problems of ergodic and information theories, including those ones described in Examples I.1 through I.3.

1. PERFECTLY NOISELESS CODING BETWEEN CONDITIONALLY BERNOULLI SOURCES

Let $[A, \mu]$ be a stationary source over a countable discrete alphabet A. We let $\mathscr{E}_n(\varepsilon)$ denote the set of all block length n code books with error probability less than

ε; $n \in N$, $0 < \varepsilon < 1$. That is, $\Delta_n \in \mathscr{E}_n(\varepsilon)$ if $\Delta_n \subset A^n$ and

$$\mu\{u \in A^Z : u^n \in \Delta_n\} > 1 - \varepsilon .$$

The *rate* of a code book Δ_n is defined to be the number $n^{-1} \log |\Delta_n|$ ($\log = \log_2$, and we let $\exp = \exp_2$ denote the corresponding exponential). Put

$$L_n(\varepsilon, \mu) = \min \{|\Delta_n| : \Delta_n \in \mathscr{E}_n(\varepsilon)\} ;$$

it is clear that $L_n(\varepsilon, \mu) < \infty$ for any $\varepsilon > 0$. The limit (if it exists)

$$H_\varepsilon(\mu) = \lim_{n \to \infty} n^{-1} \log L_n(\varepsilon, \mu) \tag{1}$$

is called the *epsilon-rate* of the source $[A, \mu]$ (see [32], where also related source coding theorems are proved, and [34] for a motivation of that concept). The limit

$$H^*(\mu) = \lim_{\varepsilon \to 0+} H_\varepsilon(\mu) \tag{2}$$

was introduced in [37] and called the *asymptotic rate*. It was generalized in the spirit of Kolmogorov-Sinai invariant and applied to generator problems for general measure theoretic dynamical systems (cf. [19, 33, 35]).

Let us recall some basic properties of these quantities. Recall that the *entropy* of a stationary source $[A, \mu]$ is defined as the limit

$$H(\mu) = - \lim_{n \to \infty} n^{-1} \int \log \mu\{u' : (u')^n = u^n\} \, \mu(\mathrm{d}u) . \tag{3}$$

Let

$$d^A(t) = \mu\{u \in R_A : H(\mu_u) \leqq t\} , \quad t \geqq 0 \tag{4}$$

and

$$c^A(\delta) = \inf\{t : d^A(t) \geqq \delta\} , \quad 0 < \delta < 1 . \tag{5}$$

As shown in [18], if $\mu \in \mathbf{E}(A)$, then the limit in (1) exists for all $\varepsilon \in (0, 1)$, and $H_\varepsilon(\mu) = H(\mu)$. If $\mu \in \mathbf{M}(A) \setminus \mathbf{E}(A)$ then, in general, the limits in (1) exist for all but a countable set of values $\varepsilon \in (0, 1)$. A more detailed description of this exceptional set is given in [32], where it is shown that $H_\varepsilon(\mu)$ exists and

$$H_\varepsilon(\mu) = c^A(1 - \varepsilon) \tag{6}$$

if and only if $1 - \varepsilon$ is a continuity point of c^A (observe that (6) is in fact a coding theorem and its converse formulated in case of a fixed level of error probability). In particular, (6) takes place also for $\varepsilon = 0$ and this gives the formula

$$H^*(\mu) = c^A(1) = \text{ess. sup}\{H(\mu_u) : u \in R_A \bmod \mu\} \tag{7}$$

first obtained by Winkelbauer in [37].

Let $[B, \varkappa]$ be another stationary source over a countable alphabet B. We let $H(\varkappa)$, $H_\varepsilon(\varkappa)$, $H^*(\varkappa)$, d^B and c^B denote the above introduced quantities for this source.

Theorem 1. Let $[A, \mu]$ and $[B, \varkappa]$ be two mod 0 isomorphic aperiodic and stationary sources over countable discrete alphabets. Then we have that

$$d^A(t) = d^B(t)\,, \quad t \geqq 0\,. \tag{8}$$

Proof. Let $\bar{\varphi} : A^Z \to B^Z$ be the corresponding perfectly noiseless code. If $u \in R_A$ then $\bar{\varphi}(u) \in R_B$ and $\mu_u \bar{\varphi}^{-1} = \mu_{\bar{\varphi}(u)}$ (cf. Lemmas I.2 and I.3). By Theorem I.1, for m_0^A-almost all $\xi \in Q^A$, there is a mod 0 isomorphism $\bar{\varphi}_\xi$ between $[A, m_\xi]$ and $[B, m_\eta]$, $\eta = \bar{\varphi}(\xi)$. But if $\xi = R_A(u)$ then $\eta = R_B(\bar{\varphi}(u)) = R_B(\bar{\varphi}_\xi(u))$ so that $H(\mu_u) = H(\mu_{\bar{\varphi}(u)})$, because entropy is an isomorphism invariant [1]. Consequently, the latter relation takes place for μ-almost all $u \in R_A$. Hence, the relations

$$\{u \in R_A : H(\mu_u) \leqq t\} = \{u \in R_A : H(\mu_{\bar{\varphi}(u)}) \leqq t\} = \bar{\varphi}^{-1}\{v \in R_B : H(\mu_v) \leqq t\}$$

are valid mod 0. Since $\mu \bar{\varphi}^{-1} = \varkappa$, (8) follows. □

Corollary 1. If $[A, \mu]$ and $[B, \varkappa]$ are finitarily isomorphic then $d^A(t) = d^B(t)$ for all $t \geqq 0$.

In other words, the distribution function of entropy of ergodic components is an *invariant* with respect to both mod 0 and finitary isomorphisms. A combination of (5), (6) and (8) shows that isomorphic sources exhibit identical behaviour of rates of block codes for all admissible levels of error probability.

Of course, the most interesting question is when the distribution function of entropy is a *complete invariant*. In our language, we ask if there is a class of stationary sources for which identical behaviour of rates of block codes implies that the sources are perfectly noiseless codings of each other.

If $\mu \in \mathbf{E}(A)$ is arbitrary then $d^A(t)$ is concentrated at $t = H(\mu)$ (cf. (4)). Consequently for ergodic sources the new invariant established in Theorem 1 reduces to entropy. This leads to the following observation. Suppose $\mathbf{M} \subset \mathbf{M}(A)$ and $\mathbf{N} \subset \mathbf{M}(B)$ are sets of stationary sources for which the converse of Theorem 1 holds true:

if $\mu \in \mathbf{M}$, $\varkappa \in \mathbf{N}$, and $d^A(t) = d^B(t)$ for all $t \geqq 0$, then the sources $[A, \mu]$ and $[B, \varkappa]$ are isomorphic.

By Theorem I.1 the corresponding perfectly noiseless code $\bar{\varphi} : A^Z \to B^Z$ splits into a family $(\bar{\varphi}_\xi;\ \xi \in Q^A)$ of local codes. Since entropy is an isomorphism invariant, $H(m_\xi) = H(m_\eta)$, where $\eta \in Q^B$ corresponds to ξ by Theorem I.1 (a). In order the above property be true, we must be able to construct the local isomorphisms $\bar{\varphi}_\xi$ by knowing merely that $d^A(t) = d^B(t)$ for all $t \geqq 0$. But this gives us only knowledge of entropies of the ergodic components so that a necessary condition for the above formulated converse is that entropy be a complete invariant for ergodic components of sources from $\mathbf{M}$ and $\mathbf{N}$. In light of Ornstein's isomorphism theorem [11, 12] we are led to the following characterization of classes $\mathbf{M}$ and $\mathbf{N}$.

Let A be a finite set with $|A|$ elements. We assume that $|A| \geqq 2$ and let S_A denote the set of all probability vectors $p = (p_1, \ldots, p_{|A|})$ with at least two positive entries. Each $p \in S_A$ induces a memoryless source $[A', m_p]$ ($A' \subset A$, $A' = A$ if all entries of p are positive). Clearly any such source is aperiodic. Let W be an arbitrary probability measure on Borel subsets of S_A (with respect to the usual Euclidean topology). Then put

$$\mu_W(E) = \int_{S_A} m_p(E)\, W(\mathrm{d}p)\,, \quad E \in \mathscr{A}^Z\,. \tag{9}$$

Then μ_W is *exchangeable*; that is, invariant under the group of all permutations of Z which leave all but finitely many elements of Z fixed (cf. [27, 28]). Conversely, any exchangeable probability μ on $(A^Z, \mathscr{A}^Z)$ is of the form μ_W for some W, as follows from a generalization of de Finetti's theorem [28].

Alternatively, one can imagine an exchangeable source as *conditionally memoryless*. Indeed, if $[A, \mu, U]$ is such then there exists a random variable $\tilde{U}$ such that, conditioned on $\tilde{U}$, the random variables U_i, $i \in Z$, are independent and identically distributed [28].

As μ_W is also T_A-invariant, it admits also an ergodic decomposition of the form (I.16). However, the σ-fields $\mathscr{I}(A)$ (see (I.4)) and that of all exchangeable events give rise to isomorphic measure algebras under any exchangeable probability [27]. Thus, (9) is merely a reparametrization of (I.16). We let denote

$$\tilde{d}^A(t) = W\{p \in S_A : H(p) \leqq t\}\,, \quad t \geqq 0\,;$$

$$\tilde{c}^A(\delta) = \inf\{t : \tilde{d}^A(t) \geqq \delta\}\,, \qquad 0 < \delta < 1\,,$$

where

$$H(p) = H(m_p) = -\sum_{i=1}^{|A|} p_i \log p_i$$

with the usual convention that $0.\log 0 = 0$. Also, let $[B, \varkappa]$ be defined by

$$\varkappa(F) = \int_{S_B} m_{\tilde{p}}(F)\, \tilde{W}(\mathrm{d}\tilde{p})\,, \quad F \in \mathscr{B}^Z$$

and let $\tilde{d}^B$ and $\tilde{c}^B$ be defined as above.

Theorem 2. Let $[A, \mu]$ and $[B, \varkappa]$ be two conditionally memoryless sources over finite alphabets such that $d^A(t) = d^B(t)$ for all $t \geqq 0$. Then there exists a perfectly noiseless code $\bar{\varphi} : A^Z \to B^Z$ such that $\varkappa = \mu\bar{\varphi}^{-1}$.

Proof. First of all observe that the two sources are aperiodic. For, if there were periodic trajectories then they had to meet two or more of mutually disjoint sets supporting different ergodic components, and this is impossible. We claim the existence of sets $E_0 \in \mathscr{A}^Z$ and $F_0 \in \mathscr{B}^Z$ such that $\mu(E_0) = \varkappa(F_0) = 1$, and for any $u \in E_0$ ($v \in F_0$) there is a unique $v \in F_0$ ($u \in E_0$) such that the sources $[A, \mu_u]$ and $[B, \mu_v]$ are mod 0 isomorphic (of course, uniqueness is again understood mod the partitions ϱ_A and ϱ_B; see (I.19)). To this end, put $R_A(p) = \{u \in R_A : \mu_u = m_p\}$, $p \in S_A$. Then $m_q(R_A(p)) = 1$ if $p = q$, and $=0$ if $p \neq q$. Since $\tilde{d}^A = \tilde{d}^B$, for W-almost all $p \in S_A$ (for $\tilde{W}$-almost all $\tilde{p} \in S_B$) we can find a $\tilde{p} \in S_B$ (a $p \in S_A$) such that $H(p) = H(\tilde{p})$. By Ornstein's isomorphism theorem [11] the sources $[A, m_p]$ and $[B, m_{\tilde{p}}]$ are isomorphic. At this stage it should be clear how to define a family of local isomorphisms required in Theorem I.2. The details are left to the reader. □

Corollary 2. The conclusion of Theorem 2 remains valid also in the following cases:

(a) the ergodic components of sources $[A, \mu]$ and $[B, \varkappa]$ are Bernoulli sources, and
(b) the alphabets are countably infinite and the components with infinite entropies have the same weights.

Proof. Part (a) follows from the fact that Ornstein's isomorphism theorem is valid also for Bernoulli sources [12]. Part (b) follows from the fact that any two generalized Bernoulli shifts with the same, possibly infinite, entropy are isomorphic (see [12] or [17]). □

Corrollary 3. If $[A, \mu]$ and $[B, \varkappa]$ are stationary sources over finite alphabets such that all ergodic components are either memoryless or mixing multistep Markov processes, then we can find a finitary code $\bar{\varphi}$ satisfying the conclusion of Theorem 2.

Proof. This follows, on account of Theorems I.3 and I.4, from the fact that entropy is a finitary isomorphism invariant which is complete for indicated classes of ergodic sources; see [7] and [23]. □

If a Bernoulli source is not of Markov type then it seems necessary that it must be a sequential coding of some memoryless source such that the coding length is infinite with positive probability. Hence, one cannot expect that Corollary 3 extends to general conditionally Bernoulli sources.

2. FINITE GENERATORS AND NOISELESS SOURCE CODING

Since the problem of finite generators is discussed in detail elsewhere (cf. [35]), we shall here point out several related results connected with source coding problems. As explained in [35], the core of Krieger's argument is a method of reduction of the alphabet size. Thus, we shall deal only with stationary sources. In this particular case, Krieger's theorem [10] says the following: if $[A, \mu]$ is an ergodic source over a countable alphabet and with finite entropy $H(\mu)$ (see (3)) then there exists an ergodic source $[B, \varkappa]$ and a perfectly noiseless code $\bar{\varphi} : A^Z \to B^Z$ such that $\varkappa = \mu\bar{\varphi}^{-1}$ and

$$|B| \leqq \text{INT}\,[\exp H(\mu)] + 1\,. \tag{10}$$

Here and in the sequel, INT (t) denotes the integer part of $t \geqq 0$. Winkelbauer [19] used the idea described in Section I.4 in order to prove the existence of B and $\bar{\varphi}$ for aperiodic stationary sources $[A, \mu]$ with finite asymptotic rate $H^*(\mu)$ (see (2)), however, without obtaining the bound (10). In [35] it is proved that under quite general circumstance the method of Section I.4 gives actually also the desired bound

$$|B| \leqq \text{INT}\,[\exp H^*(\mu)] + 1\,. \tag{11}$$

The weak topology on $\mathbf{M}(A)$ can be metrized, e.g., using the distance

$$d_w(\mu_1, \mu_2) = \sum_{n=1}^{\infty} \sum_{\boldsymbol{u} \in A^n} |\mu_1^n(\boldsymbol{u}) - \mu_2^n(\boldsymbol{u})|\,, \tag{12}$$

where $\mu^n(\boldsymbol{u}) = \mu\{u \in A^Z : u^n = \boldsymbol{u}\}$. The following result was announced in [36]:

Theorem 3. Let $[A, \mu, U]$ be an aperiodic ergodic source over a countable discrete alphabet A such that $H(\mu)$ is finite. Let $[B, \tau]$ be any ergodic source over a finite alphabet B such that $H(\tau) > H(\mu)$. For any $\delta > 0$ there exists a source $[B, \lambda, V]$ and a perfectly noiseless code $\bar{\varphi} : A^Z \to B^Z$ such that $V = \bar{\varphi}U$ and $d_w(\lambda, \tau) < \delta$.

As pointed out in [36] (see also the end of the present section), Krieger's theorem is a particular case of Theorem 3. The role of condition that $H(\tau) > H(\mu)$ is explained in detail in [16] and in Chapter IX, pp. 54–56 of [17]. The assumption of aperiodicity of the source $[A, \mu, U]$ is required by our method of the proof which employs Ornstein's coding technique based on Rohlin's lemma. But this lemma is valid only for aperiodic sources (see, e.g., [16] and [17]). However, our main goal is to extend Theorem 3 to aperiodic non-ergodic sources. As well-known, almost all ergodic components of a stationary aperiodic source are aperiodic (cf. [33], Lemma 4.1 or [4]).

Proof of Theorem 3. Let $[A, \mu, U]$, $[B, \tau]$, and $\delta > 0$ be as in the theorem. If A is finite then Theorem 3 is but Theorem 1 of [8] specialized to the case of zero-error transmission over a noiseless channel.

Hence suppose A is countably infinite. We can and shall assume that $A = N = \{1, 2, \ldots\}$. Let $A(k) = \{1, \ldots, k + 1\}$ for $k \geqq 1$, and put

$$(\tau_k u)_i = \begin{cases} u_i & \text{if } u_i \leqq k\,, \\ k + 1 & \text{if } u_i > k\,. \end{cases}$$

Then $\tau_k : A^Z \to A(k)^Z$ is a stationary code, the induced source $[A(k), \mu\tau_k^{-1}]$ is stationary and ergodic, and

$$H(\mu\tau_k^{-1}) \uparrow H(\mu) \quad \text{as} \quad k \to \infty$$

(see [31] for a systematic account of alphabet quantizers). Of course, the sources $[A(k), \mu\tau_k^{-1}]$ need not be aperiodic. We claim, however, that there is a $k_0 \in N$ such that for each $k \geqq k_0$ the source $[A(k), \mu\tau_k^{-1}]$ is also aperiodic. To this end observe that a source $[A, \mu]$ is aperiodic if and only if the measure μ is non-atomic. In particular, an ergodic source (being indecomposable in the convex set of all stationary sources) can be either periodic or aperiodic. It follows that $H(\mu) > 0$. If all sources $[A(k), \mu\tau_k^{-1}]$ were periodic, we would have $H(\mu\tau_k^{-1}) = 0$ for all $k \in N$, a contradiction. Hence, there is a $k_0 \in N$ satisfying the claim.

Pick a sequence $(\alpha_i)_{i \in N}$ of positive numbers such that

(a) $$\sum_{i=1}^{\infty} \alpha_i < \infty .$$

For each $i \in N$ find $k_i \geqq k_0$ such that

$$\text{Prob}\,[U_0 \neq U(k_i)_0] \leqq \alpha_i ,$$

where $\text{dist}\,(U(k_i)) = \mu\tau_{k_i}^{-1}$. We can and do assume that $(k_i)_{i \in N}$ is increasing. Put $\hat{A}(i) = A(k_i)$, $\hat{\mu}(i) = \mu\tau_{k_i}^{-1}$ and $\hat{U}(i) = U(k_i)$, $i \in N$. Thus, we have a sequence of ever finer quantizations $([\hat{A}(i), \hat{\mu}(i), \hat{U}(i)])_{i \in N}$ of the original source $[A, \hat{\mu}, U]$ such that, for each $i \in N$, $[\hat{A}(i), \hat{\mu}(i), \hat{U}(i)]$ is ergodic, aperiodic, and satisfies

$$\text{Prob}\,[U_0 \neq \hat{U}(i)_0] \leqq \alpha_i .$$

Next pick sequences of positive numbers $(\varepsilon_i)_{i \in N}$, $(\delta_i)_{i \in N}$ such that

(b) $$\delta_i < (56)^{-4}/2 , \quad \varepsilon_i < \tfrac{1}{2} , \quad \sum_i (\varepsilon_i + \delta_i^{1/4}) < \infty$$

and

(c) $$\lim_{i \to \infty} 48\delta_i^{1/4} \log\,[\text{card}\,(\hat{A}(i))] = 0 .$$

Define the sequence $(\eta_i)_{i \in N}$ by $\eta_i = \sum_{k=i}^{\infty} (\varepsilon_k + 56\delta_k^{1/4} + \alpha_k)$. It follows from (a) and (b) that $\eta_i \to 0$ as $i \to \infty$. Next we choose weakly open neighborhoods $\mathcal{O}$, $\mathcal{O}'$ of τ in $\mathbf{E}(B)$ so that the following holds:

(d) if $(\lambda_n)_{n \in N} \subset \mathcal{O}'$, $\lambda \in \mathbf{E}(B)$, and $\lambda_n \to \lambda$ weakly, then $\lambda \in \mathcal{O}$ and $d_w(\lambda, \tau) < \delta$ (cf. (12)).

On the first step we use Lemmas 10 and 11 of [8] (in the simplified form for a noiseless channel) in order to find positive integers m_1, N_1, sets $S(1)$, $T(1)$, $F(1)$, $W(1)$, functions $\varphi(1)$, $\psi(1)$, and a process $X(1)$ over alphabet B so that

(e) $\hat{U}(1)$, $X(1)$ are jointly ergodic and $\lambda(1) = \text{dist}\,(X(1))$ is in $\mathcal{O}'$;

(f) $m_1 > N_1$, $F(1) \subset \hat{A}(1)^{2m_1+1}$ and $\{u \in \hat{A}(1)^Z : u_{-m_1}^{2m_1+1} \in F(1)\} = \tilde{F}$ is an N_1-set (that is, the sets $\tilde{F}$, $\tilde{T}_A\tilde{F}, \ldots, T_A^{N_1-1}\tilde{F}$ are pairwise disjoint); $W(1) \subset B^{N_1}$ and the set $\{x \in B^Z : x^{2N_1} \in W(1) \times W(1)\}$ is an N_1-set;

(g) $S(1) \subset T(1) \subset \hat{A}(1)^{N_1}$, and
$N_1 \operatorname{Prob}[\hat{U}(1)_{-m_1}^{2m_1+1} \in F(1), \hat{U}(1)_{N_1-m_1}^{2m_1+1} \in F(1), \hat{U}(1)^{2N_1} \in S(1) \times S(1)] > 1 - 2\varepsilon_1$,
$N_1 \operatorname{Prob}[\hat{U}(1)_{-m_1}^{2m_1+1} \in F(1), \hat{U}(1)_{N_1-m_1}^{2m_1+1} \in F(1)\ \hat{U}(1)^{2N_1} \in T(1) \times T(1)] > 1 - \delta_1$;

(h) $\varphi(1) : S(1) \to W(1)$, $\psi(1): W(1) \to T(1)$; with probability one if $\hat{U}(1)_{-m_1}^{2m_1+1} \in F(1)$, $\hat{U}(1)^{N_1} \in T(1)$ then $X(1)^{N_1} \in W(1)$ and $\hat{U}(1)^{N_1} = \psi(1)[X(1)^{N_1}]$, and if $\hat{U}(1)_{-m_1}^{2m_1+1} \in F(1)$ and $\hat{U}(1)^{N_1} \in S(1)$, then $X(1)^{N_1} = \varphi(1)[\hat{U}(1)^{N_1}]$; and

(i) $H(X(1)) > H(\hat{U}(1)) + q_B(8\delta_1^{1/4}) + q_{\hat{A}(1)}(24\delta_1^{1/4})$, where $q_C(\varepsilon) = -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon) + 2\varepsilon \log[\operatorname{card}(C)]$.

In the original setup of [8], $[A, \mu, U]$ itself was a finite alphabet source so that it was possible to construct an initial coding $X(1)$ of U satisfying (e)–(i). Using Ornstein's technique of constructing very good codes from good ones (see [8], Lemmas 8–10) Kieffer constructed a sequence $X(i)$ of subsequent improvements on $X(1)$. The most important feature of this technique is that it allows to define the coding functions $\varphi(i)$ (and $\psi(i)$) "nearly" consistently (see (j') below). We shall see that we can do almost the same, however, we must control both the fitness of codes and the quantization error. This will be done by starting with $\hat{U}(i)$ on the i-th step. In order to avoid overcomplicated formulae we describe only the transition from step 1 to step 2. First observe that

$$\operatorname{Prob}[U_0 \in \hat{A}(2) \setminus \hat{A}(1)] = \operatorname{Prob}[U_0 = \hat{U}(2)_0, U_0 \neq \hat{U}(1)_0] \leqq$$
$$\leqq \operatorname{Prob}[U_0 \neq \hat{U}(1)_0] \leqq \alpha_1 .$$

As in [8], p. 125 we can find positive intergers m_2, N_2, sets $S(2)$, $T(2)$, $F(2)$, $W(2)$, functions $\varphi(2)$, $\psi(2)$, and a process $X(2)$ over alphabet B so that the analogues of properties (e)–(i) above hold true and

(j') if $u \in S(2)$, for at least $N_2 N_1^{-1}(1 - 2\varepsilon_1 - 56\delta_1^{1/4})$ of the integers $j \in \{m_1 + 1, \ldots \ldots, N_2 - N_1 - m_1\}$ one has $u_{j-m_1}^{2m_1+1} \in F(1)$, $u_{j+N_1-m_1}^{2m_1+1} \in F(1)$, $u_j^{2N_1} \in S(1) \times S(1)$, and $(\varphi(2)\, u)_j^{N_1} = \varphi(1)\, u_j^{N_1}$.

When coding $\hat{U}(2)$ instead of $\hat{U}(1)$, then there are at most $N_2 N_1^{-1}\alpha_1$ integers j for which the assertion in (j') can be falsified due to quantization errors occurring from the use of different quantization levels. Thus, we get

(j) if $u \in S(2)$, for at least $N_2 N_1^{-1}(1 - 2\varepsilon_1 - 56\delta_1^{1/4} - \alpha_1)$ of the integers $j \in \{m_1 + 1, \ldots, N_2 - N_1 - m_1\}$ the conclusion of (j') holds true.

For the sake of simplicity we do not write down the properties (e)–(j) for general $i < 1$. It follows from the definition of the numbers η_i that (j) entails

(k) if $k > i$ and $u \in S(k)$, then for at least $N_k N_i^{-1}(1 - \eta_i)$ of the integers $j \in \in \{m_i + 1, \ldots, N_k - N_i - m_i\}$ one has $u_{j-m_i}^{2m_i+1} \in F(i)$, $u_{j+N_i-m_i}^{2m_i+1} \in F(i)$, $u_j^{2N_i} \in \in S(i) \times S(i)$, and $(\varphi(k)u)_j^{N_i} = \varphi(i)(u_j^{N_i})$.

Next we use the block coding functions $\varphi(i)$ and $\psi(i)$ to define, for each $i \geqq 1$, stationary codes $\bar{\varphi}(i) : \hat{A}(i)^Z \to B^Z$ and $\bar{\psi}(i) : B^Z \to \hat{A}(i)^Z$:

(l) $[\bar{\varphi}(i)\, u]_j^{N_i} = \varphi(i)\, u_j^{N_i}$ if $u_{j-m_i}^{2m_i+1} \in F(i)$, $u_j^{N_i} \in S(i)$;
$[\bar{\psi}(i)\, x]_j^{N_i} = \psi(i)\, x_j^{N_i}$ if $x_j^{2N_i} \in W(i) \times W(i)$.

Assertions (k), (l), and (h) written for a general i entail that

$$\text{Prob}\,[X(k)_0 \neq X(i)_0] \leqq \eta_i + 1 - N_k \,\text{Prob}\,[\hat{U}(k)_{-m_k}^{2m_k+1} \in F(k),\ \hat{U}(k)^{N_k} \in S(k)]$$

whenever $k > i$. If $i \to \infty$, then $\eta_i \to 0$ and $k \to \infty$. But if $k \to \infty$ then the latter summand approaches one so that we can find a process V over alphabet B for which

$$\lim_{k \to \infty} \text{Prob}\,[X(k)_0 \neq V_0] = 0\,.$$

Let $\lambda = \text{dist}\,(V)$. By construction, $\text{dist}\,(X(k)) \in \mathcal{O}'$ for each $k \in N$, and $\text{dist}\,(X(k))$ weakly converges to λ, whence (d) applies and gives

$$d_w(\lambda, \tau) < \delta\,.$$

It remains to prove that the sources $[A, \mu, U]$ and $[B, \lambda, V]$ are isomorphic. We have constructed codes $\bar{\psi}(i)$ and processes $X(i)$ jointly ergodic with $\hat{U}(i)$ $(i \in N)$ so that

$$\text{Prob}\,[(\bar{\psi}(i)\, X(i))_0 \neq \hat{U}(i)_0] \to 0\,.$$

Moreover, we have chosen the processes $\hat{U}(i)$ and $X(i)$ so that

$$\text{Prob}\,[U_0 \neq \hat{U}(i)_0] \leqq \alpha_i \to 0\,,$$
$$\text{Prob}\,[V_0 \neq X(i)_0] \to 0$$

as $i \to \infty$. If $V_0 = X(i)_0$ with high probability then we have, with high probability, that $(\bar{\psi}(i)\, V)_0 = (\bar{\psi}(i)\, X(i))_0$. Let us sketch the idea of the proof. By our construction, $\bar{\psi}(i)$ has been obtained from a block coding function so that it depends only on a finite number of coordinates. As pointed out by Gray (see [21], Lemma 3.2) such codes have a continuity property to the effect that the range of error propagation is bounded. Thus, they are insensitive with respect to rare errors. However, if i is large enough then the error $V_0 \neq X(i)_0$ is rare. It follows that

$$\text{Prob}\,[(\bar{\psi}(i)\, V)_0 \neq U_0] \leqq \text{Prob}\,[(\bar{\psi}(i)\, V)_0 \neq (\bar{\psi}(i)\, X(i))_0] +$$
$$+ \text{Prob}\,[(\bar{\psi}(i)\, X(i))_0 \neq \hat{U}(i)_0] + \text{Prob}\,[\hat{U}(i)_0 \neq U_0]$$

so that the probability on the left hand side approaches zero as $i \to \infty$. A similar reasoning applies to the encoders $\bar{\varphi}(i)$, and this completes the proof. □

Of course, one may ask whether weak approximation is the best we can do. Known results (see a survey in [4]) give various approximations to finite dimensional distributions. Here we have an approximation to the process distribution. If, for example, approximation in d_w was replaced by approximation in $\bar{d}$ (see [29] and [12] for definitions of $\bar{d}$-distance), and if $[B, \tau]$ was chosen as a memoryless source, then this would force $[A, \mu]$ to be a Bernoulli source. This follows easily from the fact that the class of Bernoulli sources is $\bar{d}$-closed and that, for Bernoulli sources, closeness in finite dimensional distributions and closeness in entropy imply closeness in $\bar{d}$ (in fact, this is the contents of Ornstein's characterization of Bernoulli process by the property of being finitely determined [12]).

Now we shall extend Theorem 3 to the non-ergodic case. First let us prove the following simple assertion.

Lemma 1. Let $[B, \lambda]$ be stationary, let $[B, \tau]$ be ergodic, both over the same alphabet B. Let

$$\lambda = \int_{R_B} \lambda_v \lambda\,(\mathrm{d}v)$$

be the ergodic decomposition of $[B, \lambda]$ (cf. (I.16)). Then

$$d_w(\lambda, \tau) \leqq \int_{R_B} d_w(\lambda_v, \tau)\, \lambda(\mathrm{d}v)\,.$$

Proof. Let

$$v_n(\lambda, \tau) = \tfrac{1}{2} \sum_{\mathbf{v} \in B^n} |\lambda^n(\mathbf{v}) - \tau^n(\mathbf{v})|\,, \quad n = 1, 2, \ldots$$

denote the n-th order variational distance. It follows from (12) that

$$d_w(\lambda, \tau) = \sum_{n=1}^{\infty} 2^{-n+1}\, v_n(\lambda, \tau)\,. \tag{13}$$

Now,

$$2v_n(\lambda, \tau) \leqq \sum_{\mathbf{v} \in B^n} \int_{R_B} |\lambda_u^n(\mathbf{v}) - \tau^n(\mathbf{v})|\, \lambda(\mathrm{d}u) = \int_{R_B} 2v_n(\lambda_u, \tau)\, \lambda(\mathrm{d}u)\,.$$

Since the series in (13) is absolutely convergent, the result follows. □

Theorem 4. Let $[A, \mu, U]$ be an aperiodic stationary source over a countable alphabet A such that $H^*(\mu)$ is finite. Let $[B, \tau]$ be an ergodic source over a finite alphabet such that $H(\tau) > H^*(\mu)$. For any $\delta > 0$ there exists a source $[B, \lambda, V]$ and a perfectly noiseless code $\bar{\varphi} : A^Z \to B^Z$ such that $V = \bar{\varphi} U$ and $d_w(\lambda, \tau) < \delta$.

Proof. By (7), our assumption entails that

$$\mu\{u \in R_A : H(\mu_u) < H(\tau)\} = 1\,.$$

Use Theorem 3 in order to construct local representations $[B, \lambda_u, V^u]$ of sources $[A, \mu_u, U^u]$ using stationary codes $\bar{\varphi}_u : A^Z \to B^Z$, i.e.,

$$V^u = \bar{\varphi}_u U^u \quad \text{and} \quad d_w(\lambda_u, \tau) < \delta .$$

These local isomorphisms give rise, via the construction of Section I.4, to a perfectly noiseless code $\bar{\varphi} : A^Z \to B^Z$ such that the sources $[A, \mu]$ and $[B, \mu\bar{\varphi}^{-1}]$ are isomorphic. Put $\lambda = \mu\bar{\varphi}^{-1}$. Then

$$\lambda(F) = \mu\bar{\varphi}^{-1}(F) = \int_{R_A} \mu_u\bar{\varphi}^{-1}(F)\, \mu(\mathrm{d}u) = \int_{R_A} \mu_u(\bar{\varphi}_u^{-1} F \cap R_A(u))\, \mu(\mathrm{d}u) =$$

$$= \int_{R_B} \mu_v(F \cap R_B(v))\, \mu\bar{\varphi}^{-1}(\mathrm{d}v)\,, \quad F \in \mathscr{B}^Z .$$

Since the ergodic components are mod 0 uniquely determined, it follows that the above representing measures λ_u satisfy

$$\lambda_u = \mu_{\bar{\varphi}(u)} \bmod 0 .$$

The rest follows from Lemma 1. □

Theorem 4 (and also Theorem 3) has the following information theoretic interpretation. Let

$$K = \min \{k \in N : \log k > H^*(\mu)\} = \mathrm{INT}\,[\exp H^*(\mu)] + 1 .$$

Put $B = \{1, 2, \ldots, K\}$ and pick $[B, \tau]$ as an equiprobable memoryless source over B. Then $H(\tau) = \log K > H^*(\mu)$ so that Theorem 4 applies. Accordingly, there is a perfectly noiseless code $\bar{\varphi} : A^Z \to B^Z$ such that the encoded process $V = \bar{\varphi}U$ is as close as we please to an equiprobable memoryless source. Considering B as the alphabet of a noiseless channel we see that $\bar{\varphi}$ achieves the goal of noiseless source coding – the redundancy removal [21].

Our next aim is to consider *almost topological* generators, i.e., generators such that the related perfectly noiseless codes (see Section I.1) are finitary [5]. For topological reasons it is convenient to deal with compact alphabets. Hence, we suppose that A is either finite or a one-point compactification of N. In the latter case we assume that $\mathbf{M}(A)$ contains only measures supported by subsets of N^Z.

Following [5], a *shift dynamical system* is a quadruple $(X, X \cap \mathscr{A}^Z, \mu, T_A)$, where X is a closed invariant subset of A^Z and μ is an invariant probability measure vanishing outside X (if A is a one-point compactification of N, we require that $\mu(X \cap N^Z) = 1$). A *partition* is a finite or countable sequence $\mathscr{P} = (P_0, P_1, \ldots)$ of open subsets of X such that $P_i \cap P_j = \emptyset$ if $i \neq j$, $\mu(\partial P_i) = 0$ for all i (∂P = boundary of P), $\mu(\cup_i P_i) = 1$, and $X = (\cup_i P_i)^-$ (P^- = closure of P). Let $X_1 \subset X$ be an arbitrary invariant residual set of full measure. Given a partition $\mathscr{P}$ we define

$$X_{\mathscr{P}} = \bigcap_{i \in Z} T_A^i(\cup_j P_j \cap X_1) .$$

Let B denote the index set of $\mathscr{P}$ or its one-point compactification if $\mathscr{P}$ is infinite. We define a map $\bar{\varphi}_{\mathscr{P}} : X_{\mathscr{P}} \to B^Z$ by the property that

$$\bar{\varphi}_{\mathscr{P}}(x) = (v_i; i \in Z) \quad \text{if} \quad T_A^i x \in P_{v_i}, \quad i \in Z .$$

If $Y_{\mathscr{P}} = (\bar{\varphi}_{\mathscr{P}} X_{\mathscr{P}})^-$ and $\varkappa_{\mathscr{P}} = \mu \bar{\varphi}_{\mathscr{P}}^{-1}$, we obtain a shift dynamical system $(Y_{\mathscr{P}}, Y_{\mathscr{P}} \cap \mathscr{B}^Z, \varkappa_{\mathscr{P}}, T_B)$ such that $\bar{\varphi}_{\mathscr{P}}$ is a finitary stationary code. The partition $\mathscr{P}$ is said to be a *generator*, if there exists a subset $X_2 \subset X_{\mathscr{P}}$ (residual, and of full measure) such that $\bar{\varphi}_{\mathscr{P}}$ restricted to X_2 is a finitary isomorphism. Denker and Keane [5, Theorem 20] proved that any ergodic shift dynamical system $(X, X \cap \mathscr{A}^Z, \mu, T_A)$ with finite entropy $H(\mu)$ has a finite generator $\mathscr{P}$ with at most INT $[\exp H(\mu)] + 1$ atoms. Using the results of [1] for finitary codes we get the following assertion:

Theorem 5. Let $[A, \mu, U]$ be a stationary aperiodic source whose alphabet A is finite or a one-point compactification of N; in the latter case suppose that $\mu(N^Z) = 1$ and $H^*(\mu) < \infty$. Suppose $[B, \tau]$ is an equiprobable memoryless source over a finite alphabet B with INT $[\exp H^*(\mu)] + 1$ letters. For any $\delta > 0$ there exists a source $[B, \lambda, V]$ and a perfectly noiseless and *finitary* code $\bar{\varphi} : A^Z \to B^Z$ such that $V = \bar{\varphi} U$ ano $d_w(\lambda, \tau) < \delta$.

In other words, we can accomplish the goal of noiseless source coding by a stationary sequential coding such that the coding length is finite with probability one.

3. ZERO-ERROR TRANSMISSION OF STATIONARY NON-ERGODIC SOURCES

Throughout this section we assume, unless otherwise stated, that $[A, \mu, U]$ is an aperiodic stationary source over a finite alphabet. Let $[B, \nu, C]$ be a *stationary channel*, wheıe the alphabets B and C are also assumed finite. That is, $\nu = \{\nu_x : x \in B^Z\}$ is a family of probability measures on $(C^Z, \mathscr{C}^Z)$ such that the map $x \mapsto \nu_x(F)$ from B^Z into $[0, 1]$ is $\mathscr{B}^Z$-measurable for each event $F \in \mathscr{C}^Z$, and

$$(14) \qquad \nu_{T_B x}(T_C F) = \nu_x(F) ; \quad x \in B^Z , \quad F \in \mathscr{C}^Z .$$

If $\lambda \in \mathbf{P}(B)$, we let $\lambda\nu$, denote the double source (= the joint input/output distribution). $\lambda\nu$ is the measure in $\mathbf{P}(B \times C)$ uniquely determined by the properties that

$$(15) \qquad \lambda\nu(E \times F) = \int_E \nu_x(F)\, \lambda(\mathrm{d}x) ; \quad E \in \mathscr{B}^Z , \quad F \in \mathscr{C}^Z .$$

(We make the obvious identification between the spaces $B^Z \times C^Z$ and $(B \times C)^Z$ without particular comments.) The channel $[B, \nu, C]$ is sais to be *ergodic* if $\lambda\nu \in \mathbf{E}(B \times C)$ whenever $\lambda \in \mathbf{E}(B)$. The channel $[B, \nu, C]$ is said to be *weakly continuous*, if the map

$$\lambda \to \lambda\nu : \mathbf{E}(B) \to \mathbf{E}(B \times C)$$

is continuous at each $\lambda \in \mathbf{E}(B)$. Weakly continuous channels contain all reasonable families of stationary channels (memoryless, finite memory ones, $\bar{d}$-continuous channels) and, as pointed out by Kieffer [25, 26], they are the most general channels for which one may reasonably ask for a coding theorem.

The first result on zero-error stationary coding for transmission of stationary sources over stationary channels was obtained by Gray, Ornstein, and Dobrushin [22] who supposed that $[A, \mu]$ was a Bernoulli source and $[B, \nu, C]$ a $\bar{d}$-continuous and totally ergodic channel. For our purposes the most important thing is an observation in [22] formulated below as Lemma 2. But first let us recall concepts from [22] and [8].

A stationary source $[A, \mu]$ is *zero-error transmissible* over the stationary channel $[B, \nu, C]$ if there exist stationary codes $\bar{f}: A^{\mathbf{Z}} \to B^{\mathbf{Z}}$, $\bar{g}: C^{\mathbf{Z}} \to A^{\mathbf{Z}}$, and a Markov chain U, X, Y such that $\operatorname{dist}(U) = \mu$, $\operatorname{dist}(Y|X) = \nu$, $X = \bar{f}U$, and $U = \bar{g}Y \bmod 0$.

We say that a stationary source $[B, \lambda]$ is *ν-invulnerable* if there are process X and Y such that $\operatorname{dist}(X) = \lambda$, $\operatorname{dist}(Y|X) = \nu$, and X is a stationary perfectly noiseless coding of Y, say $X = \bar{h}Y$. Thus, from the point of view of transmission of the process X, the channel behaves as perfectly noiseless. The proof of the following elementary lemma is left to the reades.

Lemma 2. A source $[A, \mu]$ is zero-error transmissible over a channel $[B, \nu, C]$ if and only if there exists a ν-invulnerable source $[B, \lambda]$ isomorphic to $[A, \mu]$.

Next we introduce the concepts of ε-transmissibility and ε-invulnerability. A source $[A, \mu]$ is said to be *ε-transmissible* over the channel $[B, \nu, C]$ if there exist $\bar{f}$, $\bar{g}$, U, X, and Y as above such that

$$\text{(15)} \qquad \operatorname{Prob}[U_0 \neq (\bar{g}Y)_0] \leqq \varepsilon .$$

A source $[B, \lambda]$ is *ε-invulnerable* if there exist X, Y, and $\bar{h}$ as above so that

$$\text{(16)} \qquad \operatorname{Prob}[X_0 \neq (\bar{h}Y)_0] \leqq \varepsilon .$$

Lemma 3. Suppose there exists an ε-invulnerable source $[B, \lambda]$ isomorphic to $[A, \mu]$. Then the stationary source $[A, \mu]$ is ε-transmissible over the stationary channel $[B, \nu, C]$.

Proof. Let $[B, \lambda, X]$ be ε-invulnerable, and let $f^*: A^{\mathbf{Z}} \to B^{\mathbf{Z}}$ be a perfectly noiseless code with $\lambda = \mu(f^*)^{-1}$. Since f^* is perfectly noiseless, $\operatorname{Prob}[X_0 \neq (f^*U)_0] = 0$. Put $\bar{f} = f^*$ and $\bar{g} = (f^*)^{-1} \circ \bar{h}$, where $\bar{h}$ appears in (17). Then

$$\operatorname{Prob}[U_0 \neq (\bar{g}Y)_0] = \operatorname{Prob}[U_0 \neq (((f^*)^{-1} \circ \bar{h})\, Y)_0] =$$
$$= \operatorname{Prob}[(f^*U)_0 \neq (\bar{h}Y)_0] \leqq \operatorname{Prob}[(f^*U)_0 \neq X_0] +$$
$$+ \operatorname{Prob}[X_0 \neq (\bar{h}Y)_0] \leqq \varepsilon . \qquad \square$$

We do not know whether the converse of Lemma 3 is true. But it does not seem very likely, for this would imply that any encoder/decoder pair $(\bar{f}, \bar{g})$ satisfying (16) could be modified into a pair (f^*, g^*) such that g^* is responsible for all errors.

Next we shall relate the concept of ν-invulnerability with ergodic decompositions. The main result of this section will be a simple consequence of the following two lemmas.

Lemma 4. Let $[A, \mu, U]$ be an aperiodic stationary source and $[B, \nu, C]$ a stationary channel. Suppose there exists a ν-invulnerable source $[B, \lambda]$ isomorphic to $[A, \mu]$, and let

$$\lambda = \int_{R_B} \lambda_x \, \lambda(\mathrm{d}x)$$

be its ergodic decomposition. Then

$$\lambda\{x \in R_B : [B, \lambda_x] \text{ is } \nu\text{-invulnerable}\} = 1 . \tag{18}$$

Proof. By Lemma 2, $[A, \mu, U]$ is zero-error transmissible over $[B, \nu, C]$ so that there exist stationary codes $\bar{f} : A^{\mathbf{Z}} \to B^{\mathbf{Z}}$, $\bar{g} : C^{\mathbf{Z}} \to A^{\mathbf{Z}}$, and a Markov chain U, X, Y with $\operatorname{dist}(U) = \mu$, $X = \bar{f}U$, $\operatorname{dist}(Y \mid X) = \nu$, and

$$\operatorname{Prob}[U_0 \neq (\bar{g}Y)_0] = \int \nu_{\bar{f}(u)}(g^{-1}(A \setminus \{u_0\}))\, \mu(\mathrm{d}u) = 0 ,$$

where $g : C^{\mathbf{Z}} \to A$ is the map which corresponds to $\bar{g}$ according to (I.9). If $u \in R_A$, let U^u denote the process with $\operatorname{dist}(U^u) = \mu_u$, $X^u = \bar{f}U^u$, and let Y^u denote the output process of the channel $[B, \nu, C]$ when the input process was X^u. By the above formula we see that

$$\mu\{u \in R_A : \nu_{\bar{f}(u)}(g^{-1}(A \setminus \{u_0\})) = 0\} = 1 .$$

Using the ergodic decomposition formula (I.16) we conclude from this relation that

$$\mu\{w \in R_A : \mu_w\{u \in R_A : \nu_{\bar{f}(u)}(g^{-1}(A \setminus \{u_0\})) = 0\} = 1\} = 1 ;$$

that is,

$$\mu\{w \in R_A : \operatorname{Prob}[U_0^w \neq (\bar{g}Y^w)_0] = 0\} = 1 . \tag{19}$$

It remains to relate (19) with (18). Let $f^* : A^{\mathbf{Z}} \to B^{\mathbf{Z}}$ be a perfectly noiseless code such that $\lambda = \mu(f^*)^{-1}$; f^* exists by assumption. By Theorem I.1, for μ-almost all $u \in R_A$ (for λ-almost all $x \in R_B$) we can find $x \in R_B$ ($u \in R_A$) so that $\lambda_x = \mu_u(f^*)^{-1}$, and these points u and x are unique modulo the partitions ϱ_A and ϱ_B (see (I.19)). At the same time, almost all ergodic components $[A, \mu_u, U^u]$ of $[A, \mu, U]$ are zero-error transmissible. Use Lemma 2 in order to find ν-invulnerable sources $[B, \lambda_u]$ and perfectly noiseless codes $\bar{f}_u : A^{\mathbf{Z}} \to B^{\mathbf{Z}}$ such that $\lambda_u = \mu_u \bar{f}_u^{-1}$.

For any $u \in R_A$ such that $[A, \mu_u]$ is zero-error transmissible put $\xi = R_A(u)$ and $\eta = R_B(f^*(u))$. Theorem I.1 applies and shows that f^* splits into a family $(f_\xi^*; \xi \in Q^A)$

of local isomorphisms; the f_ξ^*'s are defined for m_0^A-almost all ergodic fibres $\xi \in Q^A$. It follows that

$$m_0^A\{\xi \in Q^A : u \in \xi, f_\xi^* = \tilde{f}_u\} = 1 .$$

Indeed, $f_\xi^* = \tilde{f}_u$ for m_ξ-almost all $u \in \xi$, and hence the latter conclusion follows from the canonical decomposition (I.21). Returning back to the original parametrization we get that

$$\mu\{u \in R_A : \lambda_u = \mu_u(f^*)^{-1}\} = 1$$

But $\mu_u(f^*)^{-1} = \lambda_x$ so that the sources $[B, \lambda_x]$ are ν-invulnerable, being identical with the corresponding ν-invulnearble sources $[B, \lambda_u]$. Since $\lambda(f^*R_A) = \mu(f^*)^{-1} . (f^*R_A) = \mu(R_A) = 1$, the desired conclusion (18) follows. □

The converse of Lemma 4 requires ergodicity of the channel in order we can construct a global decoder from local decoders based on ergodic components of the channel output process.

Lemma 5. Let $[A, \mu]$ be a stationary aperiodic source and $[B, \nu, C]$ a stationary and ergodic channel. Suppose that for μ-almost all ergodic components $[A, \mu_u]$ there exist ν-invulnerable sources $[B, \lambda_u]$ isomorphic to $[A, \mu_u]$. Then the source $[B, \lambda]$ is ν-invulnerable, where

$$\lambda = \int_{R_A} \lambda_u \, \lambda(\mathrm{d}u) .$$

Proof. Use our assumptions and Lemma 2. Accordingly, for μ-almost all $u \in R_A$, we can find codes $\bar{f}_u : A^Z \to B^Z$, $\bar{g}_u : C^Z \to A^Z$, and processes U^u, $X^u = \bar{f}_u U^u$ and Y^u so that

$$\text{Prob}\,[U_0^u \neq (\bar{g}_u Y^u)_0] = 0 .$$

By definition of error probability it follows that

$$\mu\{u \in R_A : \mu_u\{w \in R_A : \nu_{\bar{f}_u(w)}(g_u^{-1}(A \setminus \{w_0\}))\} = 0\} = 1\} = 1 .$$

If $u \notin R_A(u')$, then $R_A(u) \cap R_A(u') = \emptyset$. Since any $\bar{f}_u$ is defined on a subset of $R_A(u)$, we can define a code $\bar{f} : A^Z \to B^Z$ by

$$\bar{f}(w) = \bar{f}_u(w) \quad \text{if} \quad w \in R_A(u) , \quad u \in R_A .$$

This defines $\bar{f}$ μ-almost everywhere. Since U^u is ergodic, X^u is ergodic so that the joint input/output process (X^u, Y^u) is ergodic, as follows from ergodicity of the channel $[B, \nu, C]$. Hence we see that Y^u must be ergodic, too. Consequently, the code $\bar{g}_u$ is also defined on a subset of $R_C(y)$ for some $y \in R_C$. Therefore, we can define a code $\bar{g} : C^Z \to A^Z$ in the same way as $\bar{f}$ was defined. But then

$$\mu\{u \in R_A : \mu_u\{w \in R_A : \nu_{\bar{f}(w)}(g^{-1}(A \setminus \{w_0\})) = 0\} = 1\} = 1 .$$

Using again the formula (I.16) we see that the source $[A, \mu]$ must be zero-error transmissible over $[B, \nu, C]$. By Lemma 2, there exists a ν-invulnerable source $[B, \lambda^*]$

and a perfectly noiseless code $f^* : A^Z \to B^Z$ such that $\lambda^* = \mu(f^*)^{-1}$. By Theorem I.1, f^* splits into a family $(f_\xi^*; \xi \in Q^A)$ of local codes. It follows that

$$f_\xi^*(u') = \bar{f}_u(u') \quad \text{if} \quad u, u' \in \xi ,$$

and this holds true for m_ξ-almost all $u, u' \in \xi$ and for m_0^A-almost all $\xi \in Q^A$. Using the canonical decomposition formula (I.21) and passing back to the original parametrization we conclude that

$$\mu\{u \in R_A : \bar{f}(u) \neq f^*(u)\} = 0$$

Consequently, $\lambda = \lambda^* = \mu(f^*)^{-1}$. By construction, $[B, \lambda^*]$ is ν-invulnerable so that $[B, \lambda]$ is ν-invulnerable as asserted. □

Let $C(\nu)$ denote the *Shannon capacity* of the channel $[B, \nu, C]$, that is,

(20) $$C(\nu) = \sup \{I(\lambda\nu) : \lambda \in \mathbf{E}(B)\} ,$$

where $I(\lambda\nu)$ is the average mutual information of the double source (see, e.g., [38]). Our next theorem extends the zero-error transmission theorem of Kieffer [8].

Theorem 6. Let $[A, \mu]$ be a stationary aperiodic source and $[B, \nu, C]$ an ergodic and weakly continuous channel, where the alphabets A, B, and C are finite. Then

(a) $[A, \mu]$ is zero-error transmissible over $[B, \nu, C]$ if $H^*(\mu) < C(\nu)$, and

(b) $[A, \mu]$ is not zero-error transmissible if $H^*(\mu) > C(\nu)$.

Proof. Suppose that $H^*(\mu) > C(\nu)$. Then (7) implies that

$$\mu\{u \in R_A : H(\mu_u) > C(\nu)\} > 0 .$$

By Theorem 10 of [24], if u is from the latter set, the source $[A, \mu_u]$ is not zero-error transmissible. If $[A, \mu]$ itself was zero-error transmissible, then Lemma 4 would imply a contradiction. This proves part (b).

Conversely, let $H^*(\mu) < C(\nu)$. This means that

$$\mu\{u \in R_A : H(\mu_u) < C(\nu)\} = 1$$

By Kieffer's transmission theorem [8], each $[A, \mu_u]$ is zero-error transmissible. Combine Lemma 2 with Lemma 5. This proves part (a). □

Let us comment briefly on Theorem 6 (see [8] and [22] for a detailed discussion). Theorem 6 says that the Shannon capacity $C(\nu)$ of an ergodic and weakly continuous channel $[B, \nu, C]$ can be expressed as the supremum of asymptotic rates of stationary aperiodic sources which are zero-error transmissible over the channel.

This differs from the original Shannon's concept of zero-error capacity which was defined as the maximum rate below which zero-error transmission is possible using *block* encoders and decoders. Of course, on may ask how to implement stationary codes which require the knowledge of whole infinite sequences. The point is that one can approximate such infinite codes arbitrarily well by sliding-block codes of a sufficiently large but finite order (see Theorem 3.1 in [21]). In this way we get the usual ε-formulation of the coding theorem for transmission over a channel. The important additional knowledge is that, conversely, these approximations converge in a very precise manner to infinite zero-error codes. This convergence assertion has no counterpart within the traditional block coding approach.

4. EPSILON-TRANSMISSIBILITY

Winkelbauer [38] proved a coding theorem for decomposable channels which, by definition, are non-ergodic, and found that the asymptotic behaviour of channel codes for fixed values of probability of decoding error is governed by so-called ε-capacities, a channel analogue of ε-rates introduced above. An inspection of his proof shows that the proved actually a little bit more – a transmission theorem for (block) transmission of ergodic sources over decomposable channels. Our result is in a sense dual in that we consider an ergodic channel but transmit a non-ergodic source.

Theorem 7. Let $[A, \mu]$ be a stationary aperiodic source, and let $[B, \nu, C]$ be an ergodic and weakly continuous channel. Suppose that all alphabets are finite and that

(21) $$H^*(\mu) < \log |B| .$$

If

$$\mu\{u \in R_A : H(\mu_u) < C(\nu)\} \geqq 1 - \varepsilon ,$$

then $[A, \mu]$ is ε-transmissible over $[B, \nu, C]$.

Observe that, owing to (6), Theorem 7 says the following. If

(22) $$C(\nu) > H_\varepsilon(\mu)$$

then the source $[A, \mu]$ is ε-transmissible over $[B, \nu, C]$. This makes possible to define the ε-rates by the formula

(23) $$H_\varepsilon(\mu) = \inf \{C(\nu) : \mu \text{ is } \varepsilon\text{-transmissible over } [B, \nu, C]\} .$$

This formula has the advantage of being much more intuitive than previous formulas in terms of the quantities $L_n(\varepsilon, \mu)$ and c^A. An interesting open problem is to clarify the situation concerning a converse of Theorem 7. The difficulty here is of the same kind as discussed above in connection with Lemma 3.

Proof of Theorem 7. Let us decompose the set R_A into two disjoint subsets E_1 and E_2, where

$$E_1 = \{u \in R_A : H(\mu_u) < C(\nu)\} ,$$
$$E_2 = \{u \in R_A : C(\nu) \leqq H(\mu_u) < \log |B|\} .$$

(By (21), the union of these two sets exhausts almost all of R_A.) If $u \in E_1$, then $[A, \mu_u]$ is zero-error transmissible over $[B, \nu, C]$. By Lemma 2, we find a ν-invulnerable source $[B, \lambda_u]$ isomorphic to $[A, \mu_u]$. Let $\bar{f}_u : A^Z \to B^Z$ denote the corresponding perfectly noiseless code, and let $\bar{h}_u : C^Z \to B^Z$ be a code such that

$$\text{Prob}\left[X_0^u \neq (\bar{h}_u Y^u)_0\right] = 0 ;$$

$\bar{h}_u$ exists by ν-invulnerability of the source $[B, \lambda_u, X^u] = [B, \mu_u \bar{f}_u^{-1}, \bar{f}_u U^u]$. If $u \in E_2$ then we use the fact that $H(\mu_u) < \log |B|$ in order to find a perfectly noiseless code $\bar{f}_u : A^Z \to B^Z$ and a source $[B, \lambda_u, X^u]$ such that $\lambda_u = \mu_u \bar{f}_u^{-1}$ and $X^u = \bar{f}_u U^u$. This can be done by Krieger's theorem, and is the only place where we use assumption (21). Observe that for an arbitrary stationary code $\bar{h}_u : C^Z \to B^Z$ we have that

$$\text{Prob}\left[X_0^u \neq (\bar{h}_u Y^u)_0\right] \leqq 1 , \quad u \in E_2 .$$

However, we shall choose $\bar{h}_u$ for $u \in E_2$ with a little bit more care below. Using Theorem I.2 we compose the codes $\bar{f}_u$, $u \in R_A$, into a single perfectly noiseless code $\bar{f} : A^Z \to B^Z$. This will be our encoder. In order to define a decoder $\bar{g} : C^Z \to A^Z$ we proceed as follows. If $u \in E_1$, put $\bar{g}_u(y) = \bar{f}_u^{-1}(\bar{h}_u(y))$. By definition,

$$\text{Prob}\left[U_0^u \neq (\bar{g}_u Y^u)_0\right] = 0 , \quad u \in E_1 .$$

For $u \in E_2$, we choose the codes $\bar{h}_u$ acting only on corresponding ergodic fibres. Indeed, if $u \in R_A$ then the channel input process $X^u = \bar{f}_u U^u$ has its trajectories in $R_B(\bar{f}_u(u))$. Since $[B, \nu, C]$ is ergodic, the channel output process Y^u corresponding to the input X^u is ergodic and so has its trajectories in some set $R_C(y)$, $y = y(u)$. We define $\bar{h}_u$ as an arbitrary stationary code from $R_C(y)$ into $R_B(\bar{f}_u(u))$. Then we proceed as in the first case. In this way we get codes $\bar{g}_u$. Clearly,

$$\text{Prob}\left[U_0^u \neq (g_u Y^u)_0\right] \leqq 1 , \quad u \in E_2 .$$

Let $\bar{g} : C^Z \to A^Z$ be composed of the codes $\bar{g}_u$ as in the proof of Lemma 5. It remains to prove that

$$\text{Prob}\left[U_0 \neq (\bar{g} Y)_0\right] \leqq \varepsilon .$$

To see this observe that

$$\text{Prob}\left[U_0 \neq (\bar{g}Y)_0\right] = \int_{R_A} \text{Prob}\left[U_0 \neq (\bar{g}Y)_0 \mid U = U^u\right] \text{Prob}\left[U = U^u\right] \mu(\mathrm{d}u) .$$

If $U = U^u$ then $\bar{g}Y = \bar{g}Y^u = \bar{f}_u^{-1}(\bar{h}_u Y)$ so that

$$\text{Prob}\left[U_0 \neq (\bar{g}Y)_0 \mid U = U^u\right] = \text{Prob}\left[U_0^u \neq (\bar{f}_u^{-1}(\bar{h}_u Y))_0\right] .$$

But if $u \in E_1$ then the latter probability is zero, while if $u \in E_2$, it is at most one. Consequently,

$$\text{Prob}\left[U_0 \neq (\bar{g}Y)_0\right] \leqq 0 \,.\, \mu(E_1) + 1 \,.\, \int_{E_2} \text{Prob}\left[U = U^u\right] \mu(\mathrm{d}u) = \mu(E_2) < \varepsilon . \ \square$$

5. CONCLUDING REMARKS

Let us discuss briefly some related problems. First of all, the results of Section 1 can be used in order to find necessary and sufficient conditions for metric isomorphisms between two Axiom A diffeomorphisms (see [2] and [30]).

So let $f : M \to M$ be a diffeomorphism of a Riemannian C^∞-manifold M satisfying Smale's axiom A. By Smale spectral decomposition theorem, the set $\Omega(f)$ of all nonwandering points can be written as a finite disjoint union of closed sets $\Omega_1, \ldots, \Omega_s$ such that $f \mid \Omega_i$ is topologically transitive (by passing to some finite power of f we can assume that $f \mid \Omega_i$ is even topologically mixing). Bowen [2] proved that in the mixing case for any Hölder continuous function $\varphi : \Omega_i \to (-\infty, \infty)$ there exists a unique f-invariant measure μ_φ on Ω_i such that the expression

$$H(\mu_\varphi) + \int_{\Omega_i} \varphi \, \mathrm{d}\mu_\varphi$$

attains its maximum. The measure μ_φ is Bernoulli so that has an isomorphic representation as a Bernoulli source. Thus, reasonable invariant measures on $\Omega(f)$ are of the form

$$\mu = \sum_{i=1}^{s} \alpha_i \, \mu_{\varphi_i}(\alpha_i \geqq 0, \textstyle\sum \alpha_i = 1)$$

where $(\varphi_1, \ldots, \varphi_s)$ is a vector of Hölder continuous functions $\varphi_i : \Omega_i \to (-\infty, \infty)$. We do not formulate the corresponding results, for at present we do not know a reasonable physical interpretation for such invariant measures.

Let $I = (I_1, I_2, \ldots)$ be a finite or countable sequence of functions on regular points. Let $\boldsymbol{S}$ be a class of ergodic sources for which I is a complete invariant. In our considerations $I = I_1$, where $I_1(u) = H(\mu_u)$ and $\boldsymbol{S}$ was the class of all Bernoulli sources. An interesting open problem is whether the distribution function

$$d^A(t_1, t_2, \ldots) = \mu\{u \in R_A : I_i(u) \leqq t_i, \, i \in N\}$$

is a complete invariant for all mixtures of sources from $\boldsymbol{S}$. The problem is motivated by recent results of Alder and Marcus [20] who proved, in particular, that the pair (entropy, ergodic period) forms a complete set of invariants, with respect to finitary isomorphisms, for the class of all irreducible subshifts of finite type.

Any channel can be considered as a stochastic code. If $[B, \lambda]$ is a ν-invulnerable source, and if $\bar{h} : C^Z \to B^Z$ is the corresponding stationary code, then the pair $(\nu, \bar{h})$ can be considered as a perfectly noiseless random encoder/deterministic decoder pair. Does our theory apply to such source encoder/decoder pairs? Of course, we have no ergodic decomposition for stationary channels, however, we can start with various types of decomposable channels.

ACKNOWLEDGMENT

The author would like to express sincere thanks to Karel Winkelbauer and John Kieffer for stimulating discussions concerning the subject of the present paper.

(Received January 25, 1982.)

REFERENCES

[1] Š. Šujan: A local structure of stationary perfectly noiseless codes between stationary non-ergodic sources. Kybernetika *18* (1982), 5, 361—375.

[20] R. L. Adler and B. Marcus: Topological entropy and equivalence of dynamical systems. Mem. Amer. Math. Soc. *219* (1979).

[21] R. M. Gray: Sliding block source coding. IEEE Trans. Inform. Theory *IT-21* (1975), 357 to 368.

[22] R. M. Gray, D. S. Ornstein, and R. L. Dobrushin: Block synchronization, sliding-block coding, invulnerable sources, and zero-error codes for discrete noisy channels. Ann. Probab. *8* (1980), 639—674.

[23] M. Keane and M. Smorodinsky: The finitary isomorphism theorem for Markov shifts. Bull. Amer. Math. Soc. (New Series) *1* (1979), 346—438.

[24] J. C. Kieffer: On the transmission of Bernoulli sources over stationary channels. Ann. Probab. *8* (1980), 942—961.

[25] J. C. Kieffer: Block coding for a stationary channel satisfying a weak continuity condition (to appear).

[26] J. C. Kieffer: Sliding-block coding for weakly continuous channels. IEEE Trans. Inform. Theory *IT-28* (1982), 2—10.

[27] R. Olshen: The coincidence of measure algebras under an exchangeable probability. Z. Wahrsch. verw. Gebiete *18* (1971), 153—158.

[28] R. Olshen: A note on exchangeable sequences. Z. Wahrsch. verw. Gebiete *28* (1974), 317—321.

[29] D. S. Ornstein: An application of ergodic theory to probability theory. Ann. Probab. *1* (1973), 43—58.

[30] S. Smale: Differentiable dynamical systems. Bull. Amer. Math. Soc. 73 (1967), 747—817.

[31] Š. Šujan: A generalized coding problem for discrete information sources. Kybernetika *13* (1977), Supplement, 95 pp.

[32] Š. Šujan: Epsilon-rates, epsilon-quantiles, and group coding theorems for finitely additive information sources. Kybernetika *16* (1980), 105—119.

[33] Š. Šujan: Generators of an abelian group of invertible measure-preserving transformations. Monatsh. Math. *90* (1980), 67—79.

[34] Š. Šujan: Epsilon-rates and noiseless fixed-rate block coding for stationary non-ergodic sources (submitted).

[35] Š. Šujan: Generators for amenable group actions (submitted).

[36] Š. Šujan: Codes in ergodic theory and information: some examples (submitted).

[37] K. Winkelbauer: On discrete information sources. Trans. Third Prague Conf. Inform. Theory etc., NČSAV, Prague 1964, 765—830.

[38] K. Winkelbauer: On the coding theorem for decomposable discrete information channels I, II. Kybernetika *7* (1971), 109—123, 230—255.

RNDr. Štefan Šujan, CSc., Ústav merania a meracej techniky CEFV SAV (Institute of Measurement and Measuring Technique, Electro-Physical Research Centre — Slovak Academy of Sciences) 842 19 Bratislava, Czechoslovakia.