KYBERNETIKA — VOLUME 7 (1971), NUMBER 3

On the Coding Theorem for Decomposable Discrete Information Channels II

KAREL WINKELBAUER

In this Part II all the theorems are proved that were stated in Part I of this paper (cf. the preceding issue of this journal), and that were used in the proof of the main theorem on ε-capacity as established in Part I. A special paper is to follow in the next issue of this journal on the regularity condition for decomposable channels where some other facts used sometimes in Part I without explicit proofs, will be established.

3. PROOF OF THE FIRST INEQUALITY

We have seen in Section 2 that the proof of the sufficiency of the regularity condition for the validity of the assertion given in the main theorem is based on two fundamental inequalities, viz. the first inequality as stated in Theorem 1, and the second inequality as stated in Theorem 3. The final purpose of this section is to establish the validity of the first inequality given in Theorem 1; however, another purpose of this section is to prepare some of the auxiliary tools that are needed in the proofs of the subsequent sections.

Let us assume first that we are given a finite non-empty set M; the following convention may be sometimes useful for the concepts defined in the space M^I: if necessary, we shall write the subscript M in parentheses, i.e.

$$(3.1) \qquad \mathbf{K}(M)\,,\quad \mathbf{F}(M)\,,\quad T(M)\,,\quad \mathcal{M}(M)\,,\quad \mathcal{M}^*(M)\,,\quad R(M)$$

instead of the subscripted forms (cf. (1.2), (1.3), (1.9)).

We shall make use of the following terminology: a probability measure μ defined on the σ-algebra $\mathbf{F}(M)$ will be said to be *n-invariant* ($n = 1, 2, \ldots$) if $\mu(T^n E) = \mu(E)$ for every $E \in \mathbf{F}(M)$, where $T = T(M)$; μ is said to be *n-ergodic* (cf. the nomenclature employed in [12]) if it is n-invariant and not non-trivially decomposable into n-invariant measures; hence,

$$(3.2) \qquad \mathcal{M}_M = \mathcal{M}(M) = \{\mu : \mu \text{ a 1-invariant measure on } \mathbf{F}_M\}\,,$$
$$\mathcal{M}^*_M = \mathcal{M}^*(M) = \{\mu : \mu \in \mathcal{M}(M),\ \mu \text{ 1-ergodic}\}\,;$$

especially, the concept of 1-ergodicity coincides with that of indecomposability (i.e. ergodicity with respect to T_M).

Lemma 3.1. *If μ is an n-invariant measure, then the measure $\tilde{\mu}$ defined by*

$$\tilde{\mu}(E) = (1/n)\sum_{i=0}^{n-1} \mu(T^i E) \quad \text{for every} \quad E \in \mathbf{F}_M (T = T_M) \tag{3.3}$$

is 1-invariant, i.e. $\tilde{\mu} \in \mathscr{M}_M$; *if μ is n-ergodic, then $\tilde{\mu}$ is 1-ergodic,* i.e. $\tilde{\mu} \in \mathscr{M}_M^*$.

The facts stated in the preceding lemma are proved in [6] (cf. Theorem 5.2, and Theorem 7.1; cf. also [12]).

In what follows we shall say that a probability measure μ defined on the class $\mathbf{F}_M$ is *n-independent* if

$$\mu(T^{kn}[z] \cap [z']) = \mu(T^{kn}[z])\, \mu[z'] \quad \text{for every} \quad z, z' \in M^{kn},$$

and

$$\mu_{kn}\{(z_1, \ldots, z_k)\} = \prod_{i=1}^{k} \mu_n\{z_i\} = \mu(T^{kn}[(z_1, \ldots, z_k)])$$

$$\text{for every} \quad (z_1, \ldots, z_k) \in M^{kn}(z_i \in M^n,\ i = 1, \ldots, k), \quad k = 1, 2, \ldots \quad (T = T_M);$$

cf. the definitions (1.4), (1.5). The class of all n-independent measures in the space M^I will be denoted by $\mathscr{M}_n(M)$, $n = 1, 2, \ldots$ It is easy to see that an n-independent measure is n-invariant; a standard reasoning yields the fact that any n-independent measure is, at the same time, n-ergodic (cf. [11], Chapter 1). Applying Lemma 3.1, we obtain that, for any $\mu \in \mathscr{M}_n(M)$, the measure $\tilde{\mu}$ defined by (3.3) is ergodic. We shall set

$$\tilde{\mathscr{M}}_n(M) = \{\tilde{\mu} : \mu \in \mathscr{M}_n(M)\} \subset \mathscr{M}^*(M), \tag{3.4}$$

$$\mathscr{M}_n(M) = \{\mu : \mu \text{ an } n\text{-independent measure}\}.$$

In words, $\tilde{\mathscr{M}}_n(M)$ consists of all those (ergodic) measures μ' such that $\mu' = \tilde{\mu}$ for some n-independent measure μ, where $\tilde{\mu}$ is defined by (3.3).

In the next section we shall make use of the following notation: if $\varkappa$ is a mapping of the set M^n into the set M^{m+n} (cf. (1.4)), then the symbol $\tau[\varkappa]$ will designate the (measurable) transformation of the space M^I into itself defined by the relation

$$\{(\tau[\varkappa]\, z)_{-m+k(m+n)+i}\}_{0 \leq i < m+n} = \varkappa(\{z_{kn+j}\}_{0 \leq j < n}), \quad k \in I \tag{3.5}$$

for every z in M^I ($n > 0$, $m \geqq 0$). It is shown in [6] (cf. Theorem 7.2) that, for any probability measure defined on $\mathbf{F}_M$,

$$\textit{if } \mu \textit{ is n-ergodic, then } \mu\tau^{-1} \textit{ is } (m+n)\textit{-ergodic for } \tau = \tau[\varkappa], \tag{3.6}$$

where

$$\mu\tau^{-1}(E) = \mu(\tau^{-1}E) \quad \text{for every} \quad E \in \mathbf{F}_M.$$

Let us remind the fact stated in Section 1 that the set R_M of all regular points in the space M^I satisfies the relation

$$(3.7) \qquad \mu(R_M) = 1 \quad \text{for any} \quad \mu \in \mathcal{M}_M\,; \quad \text{moreover}\,,$$
$$\mathcal{M}_M^* = \{\mu_z : z \in R_M\}\,,$$
$$R_M = R(M) = \{z : z \in M^I,\ z \text{ regular}\}\,;$$

cf. (1.9), (3.1), (3.2). Let us mention in addition that the results of ergodic theory that we shall make use of in the sequel, are collected in [9], Sec. 2 (references included).

As to the entropy rate, let us point out that $\mathcal{H}$ maps the class $\mathcal{M}_M^*$ of indecomposable measures (cf. (3.2)) onto the entire closed interval $[0, \log \pi(M)]$; in more detail,

$$(3.8) \qquad 0 \leqq \mathcal{H}(\mu) \leqq \log \pi(M) \quad \text{for} \quad \mu \in \mathcal{M}_M\,,$$
$$\pi(M) = \text{the number of elements in } M\,, \quad \log = \log_2\,;$$

for any t, $0 \leqq t \leqq \log \pi(M)$, *there is* $\mu \in \mathcal{M}_1(M)$ *such that* $\mathcal{H}(\mu) = t$

(i.e. μ is 1-independent; cf. (3.4)); the proof may be found, for instance, in [6], Sec. 6, Lemma 7. Let us emphasize that the notion of entropy rate as defined by (1.7) makes sense for any n-invariant measure μ; it is shown in [6] (cf. Theorem 6.7) that

$$(3.9) \qquad \mathcal{H}(\tilde{\mu}) = \mathcal{H}(\mu)\,, \ \mu \ n\text{-invariant}, \ \tilde{\mu} \text{ defined by } (3.3) \text{ for } \mu\,.$$

The following lemma was proved in [6] as a part of Theorem 6.8.

Lemma 3.2. *If μ is an n-invariant measure, and if $\varkappa$ is a one-to-one mapping of M^n into M^{m+n} (m, n non-negative integers, $n > 0$), then* (cf. (3.5))

$$(m + n)\,\mathcal{H}(\mu\tau^{-1}) = n\,\mathcal{H}(\mu) \quad \text{for} \quad \tau = \tau[\varkappa]\,.$$

As another useful fact, of which we shall make use in this section, we state that

$$(3.10) \quad \lim_n \int_{R(M)} |(1/n) \log \mu[z_0, z_1, \ldots, z_{n-1}] + \mathcal{H}(\mu_z)|\, \mathrm{d}\mu(z) = 0 \quad \text{for any} \quad \mu \in \mathcal{M}_M\,,$$
$$[z_0, z_1, \ldots, z_{n-1}] = [(z_0, z_1, \ldots, z_{n-1})] \quad \text{for} \quad z \in M^I \quad (\text{cf. } (1.4))\,;$$

the proof may be found in [9], Sec. 2, Theorem 2 (cf. also [4]); in words, the sequence $-(1/n) \log \mu[z_0, z_1, \ldots, z_{n-1}]$ converges in the mean (with respect to μ) to the entropy rate of the ergodic component μ_z (cf. (1.9), (3.7)). It was shown by Pathasarathy in [13] that

$$(3.11) \qquad \mathcal{H}(\mu) = \int_{R(M)} \mathcal{H}(\mu_z)\, \mathrm{d}\mu(z) \quad \text{for every} \quad \mu \in \mathcal{M}_M\,.$$

If we define the quantity $L_n(\varepsilon, \mu)$ as the minimum (1.6), then it is known (cf. [1], Part I, and [7], Theorem 9.2, for the general case) that, for $0 < \varepsilon < 1$,

$$(3.12) \qquad \lim_n \frac{1}{n} \log L_n(\varepsilon, \mu) = \mathscr{H}(\mu) \quad \text{for} \quad \mu \in \mathscr{M}_M^* \quad (\text{cf. } (3.2)),$$

i.e. the limit exists for any indecomposable measure μ independently of ε and equals precisely the entropy rate; the latter fact will be used several times in the sequel.

Given the alphabets A, B, let us add some remarks as to the information rate associated with measures in $\mathscr{M}(AB)$. It is well-known that

$$(3.13) \qquad 0 \leqq I(\omega) = \mathscr{H}(\omega^A) + \mathscr{H}(\omega^B) - \mathscr{H}(\omega) \leqq \log \pi(B)$$

for any $\omega \in \mathscr{M}(AB)$, where $\mathscr{H}(\omega^A)$, $\mathscr{H}(\omega^B)$ are the entropy rates of the marginal measures associated with ω; cf. (1.11), (1.12), and (3.8). The equality in (3.13) shows that the notion of information rate makes sense for any n-invariant measure ω $(n = 1, 2, \ldots)$. Putting

$$(3.14) \qquad \mathscr{R}_n(\omega) = H_n(\omega^A) + H_n(\omega^B) - H_n(\omega), \quad n = 1, 2, \ldots,$$

(cf. (1.8)), we obtain the relation

$$(3.15) \qquad I(\omega) = \lim_{m \to \infty} \frac{1}{m} \mathscr{R}_m(\omega)$$

for any n-invariant measure ω on $\mathbf{F}(AB)$.

Let us mention that in Section 2 we have made use of the following shorthand notation

$$(3.16) \qquad I_z = I(\omega_z) \quad \text{for} \quad z \in R_{AB},$$

where ω_z is the ergodic measure associated with the regular point z according to (1.9). As an easy consequence of (3.11) and (3.13) we find that

$$(3.17) \qquad I(\omega) = \int_{R(AB)} I(\omega_z) \, d\omega(z) \quad \text{for any} \quad \omega \in \mathscr{M}(AB);$$

it is because $x \in R(A)$, $y \in R(B)$, $(\omega_{xy})^A = \mu_x$, and $(\omega_{xy})^B = \mu_y$ for any $xy \in R(AB)$ (cf. (1.10)) as immediately follows from the definition of the set $R(AB)$; cf. (3.7), (1.9). Let us remind that the measurability of I_z is an immediate consequence of the measurability of ω_z and of the definition (1.12) or relation (3.15); moreover, I_z is an invariant function (with respect to $T = T(AB)$), i.e. $I_{Tz} = I_z$, since $\omega_{Tz} = \omega_z$ for $z \in R(AB)$.

To simplify our notations, we shall define two auxiliary quantities $I^*(\omega)$ and $I_*(\omega)$ as the essential extrema

$$(3.18) \qquad \begin{aligned} I^*(\omega) &= \text{ess.sup} \{I_z : z \in R_{AB}[\omega]\}, \\ I_*(\omega) &= \text{ess.inf} \{I_z : z \in R_{AB}[\omega]\} \quad \text{for} \quad \omega \in \mathscr{M}_{AB}. \end{aligned}$$

Let us denote by $c^*(\theta, \omega)$ and $c_*(\theta, \omega)$ the upper and lower θ-quantile of the random variable I_z taken with respect to a probability measure ω on $\mathbf{F}(AB)$, i.e.

$$(3.19) \qquad c^*(\theta, \omega) = \sup \{r : \omega\{I_z \geqq r\} \geqq 1 - \theta\} \quad \text{for} \quad 0 \leqq \theta < 1 ,$$

$$c_*(\theta, \omega) = \inf \{r : \omega\{I_z \leqq r\} \geqq \theta\} \qquad \text{for} \quad 0 < \theta \leqq 1 .$$

We have

$$(3.20) \qquad I_*(\omega) = c^*(0, \omega) , \quad I^*(\omega) = c_*(1, \omega) \quad \text{for} \quad \omega \in \mathcal{M}_{AB} .$$

Lemma 3.3. *If ω is a probability measure defined on $\mathbf{F}(AB)$, then*

(1) *both $c^*(\theta, \omega)$, $c_*(\theta, \omega)$ are monotonically increasing as functions of θ, and $c_*(\theta, \omega) \leqq c^*(\theta, \omega)$ for $0 < \theta < 1$, $c^*(\theta_1, \omega) \leqq c_*(\theta_2, \omega)$ for $0 \leqq \theta_1 < \theta_2 \leqq 1$;*

(2) *$c_*(\theta, \omega)$ is continuous from the left at every positive $\theta \leqq 1$, $c^*(\theta, \omega)$ is continuous from the right at every non-negative $\theta < 1$, and*

$$c^*(\theta - 0, \omega) = c_*(\theta, \omega) , \quad c_*(\theta + 0, \omega) = c^*(\theta, \omega) ;$$

especially,

$$\lim_{\theta \to 1} c_*(\theta, \omega) = \lim_{\theta \to 1} c^*(\theta, \omega) = I^*(\omega) , \quad \lim_{\theta \to 0} c_*(\theta, \omega) = \lim_{\theta \to 0} c^*(\theta, \omega) = I_*(\omega) ;$$

(3) *the equality $c_*(\theta, \omega) = c^*(\theta, \omega)$ holds if and only if θ is not a discontinuity point of $c^*(\theta, \omega)$, or equivalently, if and only if θ is not a discontinuity point of $c_*(\theta, \omega)$.*

Proof. Denoting by F the probability distribution function of the random variable I_z, e.g. $F(t) = \omega\{I_z \leqq t\}$, t real, we may rewrite the definitions (3.19) in the form

$$c^*(\theta, \omega) = \sup \{r : F(r) \leqq \theta\} = \inf \{r : F(r) > \theta\} ,$$

$$c_*(\theta, \omega) = \sup \{r : F(r) < \theta\} = \inf \{r : F(r) \geqq \theta\} .$$

The latter equalities immediately imply the first assertion of the lemma. Since

$$\bigcup_n \{r : F(r) > \theta + 2^{-n}\} = \{r : F(r) > \theta\} , \quad \bigcup_n \{r : F(r) < \theta - 2^{-n}\} = \{r : F(r) < \theta\} ,$$

we easily obtain from the same equalities that $c^*(\theta + 0, \omega) = c^*(\theta, \omega)$ for $0 \leqq \theta < 1$, and $c_*(\theta - 0, \omega) = c_*(\theta, \omega)$ for $0 > \theta \leqq 1$. This together with the first part of the lemma implies the assertion of the second part. The third part is a direct consequence of the validity of the assertions stated in the first two parts of the lemma.

To be more in accordance with the notations employed in [8], we shall use the symbol $\mathcal{N}_{st}$ as an alternative notation for the class $\mathcal{N}(A \mid B)$ of all stationary channels. Similarly, we shall denote the class of all stationary inputs by $\mathcal{M}_{st}$, and that of all ergodic inputs by $\mathcal{M}_{erg}$; summarized in symbols:

$$(3.21) \qquad \mathcal{M}_{st} = \mathcal{M}(B) , \quad \mathcal{M}_{erg} = \mathcal{M}^*(B) , \quad \mathcal{N}_{st} = \mathcal{N}(A \mid B) , \quad \text{i.e.}$$

$\mathcal{N}_{st} = \{\nu : \nu$ a stationary channel, A its output alphabet, B its input alphabet$\}$;
cf. (3.1), (3.2).

Lemma 3.4. *If* $\nu \in \mathcal{N}_{st}$, $\mu \in \mathcal{M}_{erg}$, $G \in \mathbf{F}(AB)$, *and* $TG = G$ *for* $T = T(AB)$, i.e. G *is* T_{AB}*-invariant, then* (cf. (1.15))

$$\nu_{Ty}(G_{Ty}) = \nu_y(G_y) \text{ for } T = T_B, \quad \mu\{y : y \in R_B, \nu_y(G_y) = \nu\,\mu(G)\} = 1,$$

i.e. $\nu_y(G_y)$ *is* T_B*-invariant, and* $\nu_y(G_y) = \nu\,\mu(G)$ a.s. $[\mu]$.

Proof. Since $(T_A x)(T_B y) = T_{AB}(xy)$, it follows from the stationarity of ν that

$$\nu(\{x : x(T_B y) \in G\} \mid T_B y) = \nu(\{x : T_{AB}(xy) \in G\} \mid y).$$

Then the T_{AB}-invariance of G guarantees the T_B-invariance of $\nu_y(G_y)$. The second part of the lemma follows from (1.15) because of the almost sure constancy of an invariant function relatively to an ergodic measure; cf. also (3.7).

We shall see in what follows that we cannot restrict ourselves in our analysis to channels with finite duration of past history: we must make use of such quantities in our considerations that make sense for any stationary channel; so is another pair of auxiliary quantities, viz.

$$(3.22)\quad C^*(\nu) = \sup\{I^*(\nu\mu) : \mu \in \mathcal{M}_{erg}\}, \quad C_*(\nu) = \sup\{I_*(\nu\mu) : \mu \in \mathcal{M}_{erg}\} \text{ for } \nu \in \mathcal{N}_{st}.$$

Let us remark that the extremality properties of I^*, I_* (cf. (3.20)) enable us to assert that

$$(3.23)\qquad C_*(\nu) \leqq C^*(\nu) \leqq \log \pi(B),$$
$$C^*(\nu) = \sup\{I^*(\nu\mu) : \mu \in \mathcal{M}_{st}\}, \quad C_*(\nu) = \sup\{I_*(\nu\mu) : \mu \in \mathcal{M}_{st}\}.$$

The inequalites follow directly from the definitions (3.22) and (3.18), and from the inequality $I_z \leqq \log \pi(B)$ valid for any $z \in R_{AB}$ (cf. (3.13), (3.16)); the equalities are simple consequences of the relations

$$\nu\,\mu\{I_z \leqq r\} = \int_{R(B)} \nu\mu_y\{I_z \leqq r\}\,\mathrm{d}\mu(y),$$
$$\nu\,\mu\{I_z \geqq r\} = \int_{R(B)} \nu\mu_y\{I_z \geqq r\}\,\mathrm{d}\mu(y), \quad r \text{ real}, \quad \mu \in \mathcal{M}_{st}$$

which follow from (1.15) and from the second part of Lemma 4 * stated in Sec. 2

* The evident misprint concerning the formula standing in the second part of Lemma 4 given in [9] is to be corrected as follows: the formula is to state that

$$\int f\,\mathrm{d}\mu = \int_R \left[\int f\,\mathrm{d}\mu_z\right] \mathrm{d}\mu(z).$$

 of [9]. It is because the latter relations imply that

$$I^*(\nu\mu) \leqq C^*(\nu)\,, \quad I_*(\nu\mu) \leqq C_*(\nu) \quad \text{for any} \quad \mu \in \mathscr{M}_{\text{st}}\,,$$

which gives the desired result.

Applying formula (3.17) to the information rate $I(\nu\mu)$, we obtain that

$$I_*(\nu\mu) \leqq I(\nu\mu) \leqq I^*(\nu\mu) \quad \text{for any} \quad \mu \in \mathscr{M}_{\text{st}}\,;$$

hence the inequalities

$$C_*(\nu) \leqq \mathscr{C}(\nu) \leqq C^*(\nu)\,, \quad \nu \in \mathscr{N}_{\text{st}}\,, \tag{3.24}$$

hold for the transmission-rate capacity $\mathscr{C}(\nu)$; cf. (1.22). If ν is an ergodic channel, then

$$I_*(\nu\mu) = I^*(\nu\mu) = I(\nu\mu) \quad \text{for any} \quad \mu \in \mathscr{M}_{\text{erg}}\,;$$

it is because then the measure $\nu\mu$ is ergodic so that $\nu\mu\{I_z \leqq r\}$ equals 0 or 1 according to $I(\nu\mu) > r$ or $I(\nu\mu) \leqq r$ (cf. Lemma 3 in [9]). Consequently,

$$C_*(\nu) = C^*(\nu) = \mathscr{C}(\nu) \quad \text{for any} \quad \nu \in \mathscr{N}_{\text{erg}}\,; \tag{3.25}$$

cf. (1.21).

In our analysis an important role plays the following pair of auxiliary quantities (based on the concepts of θ-quantiles as defined by (3.19)) given by

$$\begin{aligned} c^*(\theta, \nu) &= \sup\{c^*(\theta, \nu\mu) : \mu \in \mathscr{M}_{\text{erg}}\} \quad \text{for} \quad 0 \leqq \theta < 1\,, \\ c^*(\theta, \nu) &= \sup\{c_*(\theta, \nu\mu) : \mu \in \mathscr{M}_{\text{erg}}\} \quad \text{for} \quad 0 > \theta \leqq 1\,,\ \nu \in \mathscr{N}_{\text{st}}\,; \end{aligned} \tag{3.26}$$

the first definition, of course, coincides with that given in (2.2); cf. (3.1), (3.21). We have (cf. (3.20))

$$C_*(\nu) = c^*(0, \nu)\,, \quad C^*(\nu) = c_*(1, \nu) \quad \text{for any} \quad \nu \in \mathscr{N}_{\text{st}}\,. \tag{3.27}$$

Lemma 3.5. *If $\nu \in \mathscr{N}_{\text{st}}$ then*

(1) *both $c^*(\theta, \nu)$, $c_*(\theta, \nu)$ are monotonically increasing functions of θ, and*

$$c_*(\theta, \nu) \leqq c^*(\theta, \nu) \text{ for } 0 < \theta < 1\,, \quad c^*(\theta_1, \nu) \leqq c_*(\theta_2, \nu) \text{ for } 0 \leqq \theta_1 < \theta_2 \leqq 1\,;$$

(2) *$c_*(\theta, \nu)$ is continuous from the left at any point θ such that $0 < \theta \leqq 1$, and $c^*(\theta - 0, \nu) = c_*(\theta, \nu)$; especially,*

$$\lim_{\theta \to 1} c_*(\theta, \nu) = \lim_{\theta \to 1} c^*(\theta, \nu) = C^*(\nu)\,;$$

(3) *$\theta \neq 0$ is a discontinuity point of $c^*(\theta, \nu)$ if and only if θ is a discontinuity point of $c_*(\theta, \nu)$; if θ is not a discontinuity point of $c^*(\theta, \nu)$ (or $c_*(\theta, \nu)$) then*

$$c_*(\theta, \nu) = c^*(\theta, \nu)\,.$$

Remark. In general, $c^*(\theta + 0, \nu) \geqq c^*(\theta, \nu)$, especially,

$$C_*(\nu) \leqq \lim_{\theta \to 0} c^*(\theta, \nu) = \lim_{\theta \to 0} c_*(\theta, \nu) \tag{3.28}$$

(cf. (3.27); it is because

$$c^*(\theta + 0, \nu) = \inf_{\varepsilon > 0} \sup_\mu c^*(\theta + \varepsilon, \nu\mu) \geqq$$
$$\geqq \sup_\mu \inf_{\varepsilon > 0} c^*(\theta + \varepsilon, \nu\mu) = c^*(\theta, \nu)$$

as follows from the second part of Lemma 3.3.

Proof. The first and the second part of the lemma immediately follow from the corresponding parts of Lemma 3.3. As to the third part, assumming, for instance, that θ $(0 < \theta < 1)$ is not a discontinuity point of $c^*(\theta, \nu)$, and applying the relations stated in the preceding parts of the lemma, we get that

$$c^*(\theta - 0, \nu) = c_*(\theta, \nu) \leqq c^*(\theta, \nu) \leqq$$
$$\leqq c_*(\theta + 0, \nu) = c^*(\theta + 0, \nu) = c^*(\theta - 0, \nu) = c_*(\theta - 0, \nu)$$

which shows that c_* is continuous at θ, and that $c_*(\theta, \nu) = c^*(\theta, \nu)$, Q.E.D.

If ν is a stationary channel, then the information rate $I(\nu\mu)$, called here also the transmission rate (of the source μ relatively to ν), as defined for the measure $\omega = \nu\mu$ by (1.12), makes sense for any n-invariant measure μ defined on the class $\mathbf{F}_B$ ($n = 1, 2, \ldots$); it is because the measure $\nu\mu$ as defined by (1.15) is n-invariant on $\mathbf{F}_{AB}$ for μ n-invariant on $\mathbf{F}_B$, $\nu \in \mathcal{N}_{st}$ arbitrary: more generally,

$$(\nu\mu)\, T^i_{AB} = \nu(\mu T^i_B) \quad \text{for any} \quad \nu \in \mathcal{N}_{st}\,,$$

μ an arbitrary measure on $\mathbf{F}_B$ $(i \in I)$. Making use of the latter relation, and applying the equality (3.9) to measures μ, $\nu\mu$, and $(\nu\mu)^A$, we obtain the equality

$$I(\nu\tilde{\mu}) = I(\nu\mu)\,,\ \mu\ n\text{-invariant, } \tilde{\mu} \text{ defined by (3.3) } (\nu \in \mathcal{N}_{st})\,, \tag{3.29}$$

as follows from (3.13) since, written symbolically,

$$\widetilde{\nu\mu} = \nu\tilde{\mu}\,, \quad \widetilde{((\nu\mu)^A)} = (\nu\tilde{\mu})^A\,.$$

Given $\nu \in \mathcal{N}_{st}$, and μ an n-invariant measure, we shall denote by $\mathcal{H}(\mu \mid \nu)$ the equivocation of μ with respect to ν, i.e.

$$\mathcal{H}(\mu \mid \nu) = \mathcal{H}(\nu\mu) - \mathcal{H}((\nu\mu)^A)\,. \tag{3.30}$$

Putting

$$H_n(\mu \mid \nu) = H_n(\nu\mu) - H_n((\nu\mu)^A)\,, \quad n = 1, 2, \ldots, \tag{3.31}$$

 (cf. (1.8)), we obtain the relation

$$(3.32)\qquad \mathscr{H}(\mu \mid \nu) = \lim_{m} \frac{1}{m} H_m(\mu \mid \nu)$$

(cf. (1.7), (3.8)); on the other hand, we have

$$(3.33)\qquad \begin{aligned} I(\nu\mu) &= \mathscr{H}(\mu) - \mathscr{H}(\mu \mid \nu), \\ \mathscr{R}_n(\nu\mu) &= H_n(\mu) - H_n(\mu \mid \nu) \end{aligned}$$

(cf. (3.13), (3.14)). The inequality

$$(3.34)\qquad H_{kn}(\mu \mid \nu) \leqq k\, H_n(\mu \mid \nu), \quad \mu \ n\text{-invariant},$$

which follows from the well-known properties of conditional entropies (cf. [11], Chapter 2, Sec. 6, or [6], Sec. 2, esp. Theorem 3, and Theorem 4), together with (3.32) yields the relation

$$(3.35)\qquad \mathscr{H}(\mu \mid \nu) \leqq \frac{1}{n} H_n(\mu \mid \nu), \quad \mu \ n\text{-invariant} \quad (\nu \in \mathscr{N}_{st}).$$

Lemma 3.6. *If μ is an n-independent measure,* i.e. $\mu \in \mathscr{M}_n(B)$ *(n is a natural number*; cf. (3.4)), *and if ν is a stationary channel,* i.e. $\nu \in \mathscr{N}_{st}$ (cf. (3.21)), *then*

$$k\, \mathscr{R}_n(\nu\mu) \leqq \mathscr{R}_{kn}(\nu\mu) \quad \text{for} \quad k = 1, 2, \ldots$$

Proof. The n-independence of measure μ implies the validity of the relation

$$H_{kn}(\mu) = k\, H_n(\mu), \quad k = 1, 2, \ldots$$

Making use (3.33) and (3.34), we easily deduce from the latter equality the desired inequality, Q.E.D.

Let us recall that if ν is a stationary channel with finite past history, then $m(\nu)$ means its duration, i.e.

$$(3.36)\qquad m(\nu) = \min \{m : \nu \text{ satisfies (1.16) for } m\}, \quad \nu \in \mathscr{N}_{\text{f.past}}$$

(cf. (1.21)). It is well-known that the space B^I is a compact metric space relatively to, for example, the distance function

$$(3.37)\qquad \max \{(1 + |i|)^{-1} : y_i \neq y'_i (i \in I)\}, \quad y, y' \in B^I (y \neq y'),$$

and that $\boldsymbol{F}_B$ coincides with the class of all Borel sets in the metric space B^I. As an immediate consequence of the finite-past-history condition (1.16) we obtain that

$$(3.38)\qquad \nu_y(E), \quad E \in \mathbf{K}_B, \text{ as a function of } y \text{ is continuous}$$

on the metric space B^I with respect to the distance function (3.37); i.e. the assertion (3.38) is valid for any finite-dimensional cylinder E lying in the space B^I provided that $\nu \in \mathscr{N}_{\mathrm{f.past}}$.

Given a non-negative real number r, we shall associate with any channel $\nu \in \mathscr{N}_{\mathrm{f.past}}$ a pair of channels $\bar{\nu}$, $\tilde{\nu}$ (in general, both with infinite duration of past history) by definitions

$$(3.39) \qquad \bar{\nu}_y(E) = (1 - \bar{q}_y)^{-1}\, \nu_y(E \cap \bar{G}_y) \quad \text{for} \quad \bar{q}_y < 1\,,$$

$$\bar{\nu}_y(E) = \nu_y(E) \quad \text{for} \quad \bar{q}_y = 1\,,$$

$$\bar{q}_y = 1 - \nu_y(\bar{G}_y)\,, \quad \bar{G} = \{z : z \in R_{AB},\, I_z \geqq r\}\,;$$

$$(3.40) \qquad \tilde{\nu}_y(E) = q_y^{-1}\, \nu_y(E \cap \tilde{G}_y) \quad \text{for} \quad q_y > 0\,,$$

$$\tilde{\nu}_y(E) = \nu_y(E) \quad \text{for} \quad q_y = 0\,,$$

$$q_y = \nu_y(\tilde{G}_y)\,, \quad \tilde{G}_y = \{z : z \in R_{AB},\, I_z \leqq r\}\,; \quad E \in \mathbf{F}_A\,, \quad y \in B^I$$

(cf. (1.13)); we assert that

$$(3.41) \qquad \bar{\nu} \in \mathscr{N}_{\mathrm{st}}\,, \quad \text{and} \quad \tilde{\nu} \in \mathscr{N}_{\mathrm{st}}\,.$$

We shall show that, for instance, $\tilde{\nu}$ is stationary. Owing to Lemma 3.4 we have that $q_{Ty} = q_y$ for $T = T_B$; on the other hand,

$$\nu(T_A E \cap \{x : x(T_B y) \in \tilde{G}\} \mid T_B y) = \nu(E \cap \{x : T_{AB}(xy) \in \tilde{G}\} \mid y)$$

since ν is stationary so that the stationarity od $\tilde{\nu}$ follows from the T_{AB}-invariance of $\tilde{G}$.

Now we are able to proceed and prepare some lemmas which are needed in the proof of Theorem 1. Let us define the following auxiliary quantity $I_{n,m}(z; \omega)$ for any $z \in (A \times B)^I$, $\omega \in \mathscr{M}_{AB}$, n, m non-negative integers ($n > 0$), by

$$(3.42) \qquad I_{n,m}(xy; \omega) =$$

$$= \frac{1}{n} \log \frac{\omega\{x'y' : \{x'_i\}_{0 \leqq i < n} = \{x_i\}_{0 \leqq i < n}, \{y'_i\}_{-m \leqq i < n} = \{y_i\}_{-m \leqq i < n}\}}{\omega^A\{x' : \{x'_i\}_{0 \leqq i < n} = \{x_i\}_{0 \leqq i < n}\}\, \omega^B\{y' : \{y'_i\}_{-m \leqq i < n} = \{y_i\}_{-m \leqq i < n}\}}$$

$$\text{for} \quad x \in A^I\,, \quad y \in B^I \quad \left(\log = \log_2, \frac{0}{0} = 0\right).$$

Lemma 3.7. *For any integer $m \geqq 0$, and for any $\omega \in \mathscr{M}_{AB}$,*

$$\lim_n \int_{R(AB)} |I_{n,m}(z; \omega) - I(\omega_z)|\, \mathrm{d}\omega(z) = 0\,.$$

Proof. Since $x \in R_A$, $y \in R_B$ (cf. (3.7), (1.9)), and

$$I(\omega_{xy}) = \mathscr{H}(\mu_x) + \mathscr{H}(\mu_y) - \mathscr{H}(\omega_{xy}) \quad \text{for} \quad xy \in R_{AB}$$

 so that

$$|I_{n,m}(xy;\omega) - I_{xy}| \leqq |-(1/n)\log \omega^A[x_0, \ldots, x_{n-1}] - \mathscr{H}(\mu_x)| +$$
$$+ \left|\frac{m+n}{n}\left(\frac{-1}{m+n}\right)\log \omega^B(T_B^m[y_{-m}, \ldots, y_{n-1}]) - \mathscr{H}(\mu_y)\right| +$$
$$+ |(1/n)\log \omega\{x'y' : x' \in [x_0, \ldots, x_{n-1}],\ y' \in T_B^m[y_{-m}, \ldots, y_{n-1}]\} - \mathscr{H}(\omega_{xy})|\,,$$

it easy to deduce from the latter inequality the desired result by applying (3.10) to ω^A, ω^B, ω because of the inequalities

$$\omega(T_{AB}^m[(x_{-m}, y_{-m}), \ldots, (x_{n-1}, y_{n-1})]) \leqq$$
$$\leqq \omega\{x'y' : x' \in [x_0, \ldots, x_{n-1}],\ y' \in T_B^m[y_{-m}, \ldots, y_{n-1}]\} \leqq$$
$$\leqq \omega[(x_0, y_0), \ldots, (x_{n-1}, y_{n-1})]\,.$$

In the proof of Theorem 1 we shall make use of another group of auxiliary quantities, viz (cf. (1.4), (1.5), (1.3), (3.8), (1.1))

(3.43) $$\omega_{n,m}(x, y) = \omega\{x'y' : x' \in [x],\ y' \in T_B^m[y]\}\,,$$

$$\omega_{n,m}(x \mid y) = \omega_{n,m}(x, y)/\omega_{m+n}^B\{y\}\,;\quad x \in A^n\,,\quad y \in B^{m+n}\left(\frac{0}{0} = 0\right);$$

$$\omega_{n,m}(W) = \sum_{(x,y)\in W}\omega_{n,m}(x, y)\,,\quad \omega_{n,m}(E \mid y) = \sum_{x\in E}\omega_{n,m}(x \mid y)\,;$$
$$W \subset A^n \times B^{m+n}\,,\quad E \subset A^n\,,\quad y \in B^{m+n}\,;$$

(3.44) $$S_{n,m}(\psi; \varepsilon, \omega) = \pi\{y : y \in B^{m+n},\ \omega_{n,m}(\psi^{-1}\{y\} \mid y) > 1 - \varepsilon\}\,,$$

$$S_{n,m}(\varepsilon, \omega) = \max\{S_n(\psi; \varepsilon, \omega) : \psi \in (B^{m+n})^{A^n}\}\,;$$
$$\psi : A^n \to B^{m+n}\,,\quad 0 < \varepsilon < 1\,;$$

(3.45) $$I_{n,m}(x, y; \omega) = \frac{1}{n}\log\frac{\omega_{n,m}(x \mid y)}{\omega_n^A\{x\}}\,;\quad x \in A^n\,,\quad y \in B^{m+n}\,;$$

n, m non-negative integers, $n > 0$, $\omega \in \mathscr{M}_{AB}$; cf. also (1.18), (1.19).

Lemma 3.8. *If* $\omega \in \mathscr{M}_{AB}$, $0 < \varepsilon < 1$, t *positive, and*

$$\omega_{n,m}\{(x, y) : I_{n,m}(x, y; \omega) > t\} > 1 - \tfrac{1}{2}\varepsilon$$

then $S_{n,m}(\varepsilon, \omega) > (\tfrac{1}{2}\varepsilon)\,2^{nt}$.

Proof. Let us set

$$E(y) = \{x : I_{n,m}(x, y; \omega) > t\}\,,\quad y \in B^{m+n}\,.$$

Let us construct a finite sequence $y^1, y^2, \ldots, y^S$ of points in B^{m+n} such that

$$\omega_{n,m}(A_j \mid y^j) > 1 - \varepsilon\,, \quad A_j = E(y^j) - \bigcup_{i=1}^{j-1} E(y^i)\,; \quad j = 1, \ldots, S\,;$$

$$\omega_{n,m}(E(y) - \bigcup_{j=1}^{S} E(y^j) \mid y) \leqq 1 - \varepsilon \quad \text{for every} \quad y \in B^{m+n}\,;$$

the possibility of the construction performed by induction follows from the inequality

$$\sum_{y \in B^{m+n}} \omega^B_{m+n}\{y\}\ \omega_{n,m}(E(y) \mid y) > 1 - (\varepsilon/2)$$

which coincides with that given in the assumptions of the lemma, and which shows that $\omega_{n,m}(E(y) \mid y) > 1 - \varepsilon$ holds for at least one y. According to (3.45)

$$\omega_{n,m}(x \mid y) > 2^{nt} \omega^A_n\{x\} \quad \text{for} \quad x \in E(y)\,, \quad \omega^A_n\{x\} > 0\,.$$

The rest of the proof is based upon the latter inequality and follows the lines of the proof of Feinstein's lemma given in [1] (cf. also the proof of Theorem 12.1 in [6]). By the method just mentioned we deduce the inequality $S > (\varepsilon/2)\, 2^{nt}$. If ψ is a mapping of A^n into B^{m+n} such that $\psi x = y^j$ for $x \in A_j$, then (cf. (3.44))

$$S_{n,m}(\varepsilon, \omega) \geqq S_{n,m}(\psi; \varepsilon, \omega) \geqq S > (\tfrac{1}{2}\varepsilon)\, 2^{nt}$$

which is the desired result.

Lemma 3.9. *If r is a given non-negative real number, if $\bar{\nu}$ is the channel associated with a given $\nu \in \mathcal{N}_{\text{f.past}}$ by definition (3.39), and if $\mu \in \mathcal{M}_{\text{erg}}$ then*

$$q < \varepsilon < 1 \quad \text{implies} \quad S_n(\varepsilon, \nu) \geqq S_{n,m}(\varepsilon - q, \bar{\nu}\mu)\,, \quad n = 1, 2, \ldots,$$

where $q = \nu\mu\{I_z < r\}$, and $m = m(\nu)$.

Proof. Making use of Lemma 3.4 and putting $q = 1 - \nu\mu(\bar{G})$, where $\bar{G}$ is defined in (3.39), we obtain that $\mu\{y : q_y = q\} = 1$ so that, owing to (1.15), we find that

$$\bar{\nu}\mu(G) = (1 - q)^{-1} \int \nu_y(G_y \cap \bar{G}_y)\, \mathrm{d}\mu(y) \leqq (1 - q)^{-1}\, \nu\mu(G)\,, \quad G \in \mathbf{F}_{AB}(q < 1)\,.$$

From here and from (1.17) and (3.43) we deduce that

$$\nu_n(E \mid y) = \frac{(\nu\mu)_{n,m}\{(x, y) : x \in E\}}{\mu_{m+n}\{y\}} \geqq (1 - q)\, \omega_{n,m}(E \mid y)$$

for $E \subset A^n$, $y \in B^{m+n}$, where we have set $\omega = \bar{\nu}\mu$ (hence $\omega^B = \mu$); consequently,

$$\omega_{n,m}(E \mid y) > 1 - \varepsilon \quad \text{implies} \quad \nu_n(E \mid y) > 1 - (q + \varepsilon)$$

for $0 < \varepsilon < 1 - q$ $(q < 1)$. Now we easily deduce from the latter implication by making use of definitions (1.18) and (3.44) that

$$S_n(\psi; q + \varepsilon, \nu) \geqq S_{n,m}(\psi; \varepsilon, \omega), \quad \psi : A^n \to B^{m+n}, \quad q + \varepsilon < 1$$

which together with (1.19) yields the desired result.

Before stating Theorem 1 we recall that $c(\varepsilon, \nu)$ is defined as the lower limit in (2.1) for any $\nu \in \mathcal{N}_{\text{f.past}}$, and that $c^*(\theta, \nu)$ is defined by (3.26) or (2.2), respectively.

Theorem 1. *If ν is a stationary channel with finite past history, then*

$$0 \leqq \theta < \varepsilon < 1 \quad \text{implies} \quad c^*(\theta, \nu) \leqq c(\varepsilon, \nu).$$

Proof. Let us assume that we are given θ, ε such that $0 \leqq \theta < \varepsilon < 1$, $c^*(\theta, \nu) > 0$; the case $c^*(\theta, \nu) = 0$ follows from the inequality $c(\varepsilon, \nu) \geqq 0$ universally valid. It is sufficient to prove that, for any positive $r < c^*(\theta, \nu)$, the inequality $c(\varepsilon, \nu) \geqq r$ holds.

In the remainder of the proof we shall assume that we are given a positive real number r such that $r < c^*(\theta, \nu)$. Let $\mu \in \mathcal{M}_{\text{erg}}$ be such that $r \leqq c^*(\theta, \nu\mu)$; the existence of such μ follows from definition (3.26). Let $\bar{\nu}$ be the channel that corresponds to the given channel ν and the given r according to definition (3.39). Let us set

$$(1) \qquad \omega = \bar{\nu}\mu, \quad 1 - q = \nu\mu(\bar{G}), \quad m = m(\nu);$$

cf. (3.39), (3.36). Then $\omega \in \mathcal{M}(AB)$ as follows from (3.41). Since the sequence $I_{n,m}(z; \omega)$ converges in the mean (with respect to ω) to the information rate I_z of the ergodic component ω_z according to Lemma 3.7, it converges to the same limit in probability. Let us choose λ such that $0 < \lambda < r$. Then there is $n_0 = n_0(\varepsilon')$ with the property that

$$(2) \qquad \omega\{z : z \in R_{AB}, I_{n,m}(z; \omega) > I_z - \lambda\} > 1 - \tfrac{1}{2}\varepsilon' \quad \text{for} \quad n \geqq n_0, \quad \varepsilon' \text{ given } (0 < \varepsilon' < 1).$$

It follows from the definition of $\bar{\nu}$ that $\bar{\nu}\,\mu(\bar{G}) = \omega\{I_z \geqq r\} = 1$; then we can deduce from (2) that

$$(3) \qquad \omega\{z : I_{n,m}(z; \omega) > t\} > 1 - \tfrac{1}{2}\varepsilon' \quad \text{for} \quad n \geqq n_0,$$

where we have set $t = r - \lambda$ (hence $t > 0$). Rewriting (3) with the aid of (3.43) and (3.45) (cf. (3.42)), we obtain that

$$\omega_{n,m}\{(x, y) : I_{n,m}(x, y; \omega) > t\} > 1 - \tfrac{1}{2}\varepsilon' \quad \text{for} \quad n \geqq n_0.$$

Since the latter relation shows that the assumptions of Lemma 3.8 are satisfied for $n \geqq n_0$, we shall find that

$$(4) \qquad S_{n,m}(\varepsilon', \omega) > \tfrac{1}{2}\varepsilon' \,.\, 2^{nt} \quad \text{for} \quad n \geqq n_0(\varepsilon').$$

On the other hand the assumption that $r \leqq c^*(\theta, \nu\mu)$ implies that (cf. (3.19) and (1)) $q \leqq \theta$, i.e. $q < \varepsilon$ so that Lemma 3.9 may be applied: we obtain that

$$S_n(\varepsilon, \nu) \geqq S_{n,m}(\varepsilon', \omega) \quad \text{for} \quad \varepsilon' = \varepsilon - q \quad (n = 1, 2, \ldots) .$$

The latter inequality together with the inequality (4) imply that

$$c(\varepsilon, \nu) = \liminf_n \frac{1}{n} \log S_n(\varepsilon, \nu) \geqq t = r - \lambda$$

for any λ, $0 < \lambda < r$ because λ was chosen arbitrarily; hence it follows the desired result that $c(\varepsilon, \nu) \geqq r$ which proves the validity of the theorem.

4. FORMULA FOR CAPACITY

This section will be devoted to the proof of Theorem 2. Here we have to work with a group of other auxiliary quantities which are connected with the concept of the probability of error. First we define

$$\text{(4.1)} \qquad \mathrm{e}_n(\omega) = 1 - \sum_{x \in A^n} \max_{y \in B^n} \omega[xy]$$

for any probability measure ω on $\mathbf{F}_{AB}$ $(n = 1, 2, \ldots)$. Then we may assert that (cf. (3.31)), for any $\nu \in \mathcal{N}_{\text{st}}$, μ n-invariant,

$$\text{(4.2)} \qquad H_n(\mu \mid \nu) \leqq n\, \mathrm{e}_n(\nu\mu) \log \pi(B) + 1 \quad (\log = \log_2) ;$$

the latter inequality corresponds to a well-known lemma of Feinstein (cf. [1], Part II, Sec. 2, Lemma 2.3, or [6], Theorem 3.1).

In accordance with [8] let us set

$$\text{(4.3)} \qquad \mathrm{e}_n(\mu, \nu, \tau) = 1 - \sum_{x \in A^n} \max_{z \in B^n} \int_{[z]} \nu_{\tau\zeta}[x] \, \mathrm{d}\mu(\zeta)$$

for $\nu \in \mathcal{N}_{\text{st}}$ and for any probability measure μ on $\mathbf{F}_B$; τ is a measurable mapping of B^I into itself. The following lemma will be used in the proof of Theorem 2.

Lemma 4.1. *If* $\nu \in \mathcal{N}_{\text{st}}$, μ *a probability measure on* $\mathbf{F}_B$, n, m *integers* $(n > 0$, $m \geqq 0)$, *and if* $\varkappa$ *is a one-to-one mapping of* B^n *into* B^{m+n}, *then*

$$\mathrm{e}_{m+n}(\nu(\mu\tau^{-1}T_B^m)) \leqq \mathrm{e}_n(\mu, \nu, \tau) \quad \text{for} \quad \tau = \tau[\varkappa]$$

(cf. (3.5)).

Proof. Let δ be a mapping of A^n into B^n such that

$$\text{(4.4)} \qquad \int_{[\delta x]} \nu_{\tau\zeta}[x] \, \mathrm{d}\mu(\zeta) = \max_{z \in B^n} \int_{[z]} \nu_{\tau\zeta}[x] \, \mathrm{d}\mu(\zeta) , \quad x \in A^n ;$$

then the probability of error (4.3) may be expressed in the form

$$\text{(4.5)} \qquad \mathrm{e}_n(\mu, \nu, \tau) = 1 - \textstyle\sum_{z\in B^n} \displaystyle\int_{[z]} \nu([\delta^{-1}\{z\}] \mid \tau\zeta) \, \mathrm{d}\mu(\zeta) \,,$$

where we have set (cf. (2.4) and (2.5) in [8])

$$[E] = \bigcup\{[x] : x \in E\} \,, \quad E \subset A^n \,.$$

It follows from the assumption of the lemma that $[\varkappa^{-1}y] = \tau^{-1} T_B^m[y]$ for $y \in B^{m+n}$ since $\tau = \tau[\varkappa]$ as defined by (3.5) so that from (4.5) we obtain that, because of $\varkappa$ being one-to-one,

$$\text{(4.6)} \qquad 1 - \mathrm{e}_n(\mu, \nu, \tau) = \textstyle\sum_{y\in B^{m+n}} \displaystyle\int_{[\varkappa^{-1}y]} \nu([\delta^{-1}\{\varkappa^{-1}y\}] \mid \tau\zeta) \, \mathrm{d}\mu(\zeta) =$$

$$= \textstyle\sum_{y\in B^{m+n}} \displaystyle\int_{T_B{}^m[y]} \nu([\delta^{-1}\{\varkappa^{-1}y\}] \mid \eta) \, \mathrm{d}\mu\tau^{-1}(\eta).$$

Let us define $\psi : A^{m+n} \to B^{m+n}$ by

$$\psi x = \varkappa(\delta(x_{m+1}, \ldots, x_{m+n})) \,, \quad x \in A^{m+n} \,.$$

Then it follows from the stationarity of ν (cf. (1.14)) and from (4.6) that

$$1 - \mathrm{e}_n(\mu, \nu, \tau) = \textstyle\sum_{y\in B^{m+n}} \displaystyle\int_{[y]} \nu([\psi^{-1}\{y\}] \mid \eta) \, \mathrm{d}\mu\tau^{-1} T_B^m(\eta) \,.$$

Then latter relation together with the formula

$$\mathrm{e}_{m+n}(\nu(\mu\tau^{-1}T_B^m)) = 1 - \textstyle\sum_{x\in A^{m+n}} \max_{y\in B^{m+n}} \displaystyle\int_{[y]} \nu_\eta[x] \, \mathrm{d}\mu\tau^{-1}T_B^m(\eta)$$

imply the desired inequality.

Lemma 4.2. *If* $\nu \in \mathcal{N}_{\mathrm{st}}$, $m = m(\nu)$, *n a positive integer*, $\varkappa$ *a one-to-one mapping of* B^n *into* B^{m+n}, $\tau = \tau[\varkappa]$ (cf. (3.5)), *and if* $\tilde{\nu}$ *is the channel associated with* ν *by definition* (3.40) *for* $r = c_*(\theta, \nu)$ [cf. (3.22)], *then*

$$\theta \mathrm{e}_n(\mu, \tilde{\nu}, \tau) \leqq \mathrm{e}_n(\mu, \nu, \tau) \quad \text{for any } \mu \text{ } n\text{-ergodic} \quad (0 < \theta \leqq 1) \,.$$

Proof. If $\delta : A^n \to B^n$ possesses property (4.4), then it follows from (4.6) and from definitions (3.40) that

$$1 - \mathrm{e}_n(\mu, \nu, \tau) \leqq \textstyle\sum_{y\in B^{m+n}} \displaystyle\int_{T_B{}^m[y]} (q_\eta \, \tilde{\nu}_\eta[\delta^{-1}\{\varkappa^{-1}y\}] + (1 - q_\eta)) \, \mathrm{d}\mu \, \tau^{-1}(\eta) \,.$$

Let us define the measure μ' by

$$\mu' = (1/(m+n)) \sum_{i=0}^{m+n-1} \mu\tau^{-1} T_B^i . \tag{4.7}$$

Since by assumption μ is n-ergodic, $\mu\tau^{-1}$ is $(m+n)$-ergodic according to (3.6) so that Lemma 3.1 applied to $\mu\tau^{-1}$ yields that μ' is 1-ergodic, i.e. $\mu' \in \mathscr{M}_{\text{erg}}$. Making use of the relation

$$\nu\mu' = (1/(m+n)) \sum_{i=0}^{m+n-1} \nu(\mu\tau^{-1})\, T_{AB}^i ,$$

and putting $q = \nu(\mu\tau^{-1})(\tilde{G})$, we obtain from the T_{AB}-invariance of $\tilde{G}$ (cf. (3.40)) that $\nu\,\mu'(\tilde{G}) = q$. Applying Lemma 3.4 we get $\mu'\{\eta : q_\eta = q\} = 1$, where $q_\eta = \nu_\eta(\tilde{G}_\eta)$ by definition (3.40); since q_η is T_B-invariant so that the set $\{\eta : q_\eta = q\}$ is T_B-invariant as well, it follows from (4.7) that

$$\mu\,\tau^{-1}\{\eta : q_\eta = q\} = \mu'\{\eta : q_\eta = q\} = 1 .$$

From here we obtain the inequality

$$1 - \mathrm{e}_n(\mu, \nu, \tau) \leqq q \textstyle\sum_{y \in B^{m+n}} \displaystyle\int_{T_B{}^m[y]} \tilde{\nu}_\eta[\delta^{-1}\{\varkappa^{-1} y\}]\, \mathrm{d}\mu\tau^{-1}(\eta) + (1-q) ;$$

according to (4.6) we have

$$\mathrm{e}_n(\mu, \nu, \tau) \geqq q\, \mathrm{e}_n(\mu, \tilde{\nu}, \tau) .$$

On the other hand, it follows from the assumption that $r = c_*(\theta, \nu)$, the inequality $r \geqq c_*(\theta, \nu\mu')$ so that $q = \nu\mu'\{I_z \leqq r\} \geqq \theta$ which implies the desired result.

In what follows we shall set in accordance with [8] (cf. (1.5), (1.17))

$$\begin{gathered} \mathrm{e}_n'(\mu, \nu, \varkappa, \delta) = 1 - \textstyle\sum_{z \in B^n} \nu_n(\delta^{-1}\{z\} \mid \varkappa z)\, \mu_n\{z\} \\ \text{for} \quad \nu \in \mathscr{N}_{\text{f.past}}, \quad \mu \in \mathscr{M}_{\text{st}} \quad (\text{cf. } (3.21), (1.21)) , \\ \varkappa : B^n \to B^{m+n}\ (m = m(\nu)) , \quad \delta : A^n \to B^n\ (n = 1, 2, \ldots) . \end{gathered} \tag{4.8}$$

It is easy to see that (cf. (3.5))

$$\mathrm{e}_n(\mu, \nu, \tau[\varkappa]) \leqq \mathrm{e}_n'(\mu, \nu, \varkappa, \delta) \quad \text{for any} \quad \delta \in (B^n)^{A^n} ; \tag{4.9}$$

see [8], formula (2.9). Repeating word by word the proof of Lemma 2.1 from [7] in our notations, we immediately obtain the following

Lemma 4.3. *If* $\nu \in \mathscr{N}_{\text{f.past}}$, $\mu \in \mathscr{M}_{\text{st}}$, $0 < \varepsilon < 1$, $0 < \varepsilon' < 1$, *then the inequality* (cf. (1.6), (1.19))

$$L_n(\varepsilon', \mu) \leqq S_n(\varepsilon, \nu)$$

 implies that there are a one-to-one mapping $\varkappa : B^n \to B^{m+n}$ $(m = m(\nu))$, *and a mapping* $\delta : A^n \to B^n$ *such that*

$$e'_n(\mu, \nu, \varkappa, \delta) < \varepsilon + \varepsilon' .$$

In the next lemma which will be used in the subsequent section, we have set (cf. (3.5))

$$\text{(4.10)} \qquad e_n(\mu, \nu) = \min \{e_n(\mu, \nu, \tau[\varkappa]) : \varkappa \in (B^{m+n})^{B^n}\} .$$

Lemma 4.4. *If* $\nu \in \mathcal{N}^*_{\text{erg}}$, $\mu \in \mathcal{M}_{\text{erg}}$, *and* $\mathcal{C}(\nu) < \mathcal{H}(\mu)$ *then*

$$\liminf_n e_n(\mu, \nu) = \lim_n e_n(\mu, \nu) = 1 .$$

The proof of the lemma is the same as that of Lemma 10.1 stated in [7] if use is made of (3.12) and the strong-stability condition (1.23).

Now we are prepared to derive a basic lemma which constitutes the main tool in proving Theorem 2.

Lemma 4.5. *If* ν *is a stationary channel with finite past history, then the inequalities* $0 < \varepsilon < \theta \leqq 1$ *imply the inequality* (cf. (3.26), (2.1))

$$c_*(\theta, \nu) + \frac{\varepsilon}{\theta} \log \pi(B) \geqq c(\varepsilon, \nu) .$$

Proof. Let us assume that we are given ε, θ such that $0 < \varepsilon < \theta \leqq 1$. Excluding trivial cases we may assume that $c(\varepsilon, \nu) > 0$, and $k = \log \pi(B) > 0$.

Choose λ such that $0 < \lambda < 1$. Owing to (3.8), there is $\mu \in \mathcal{M}_1(B)$, i.e. μ 1-independent, having the property that

$$\text{(1)} \qquad c(\varepsilon, \nu) - \tfrac{1}{4}\lambda \leqq \mathcal{H}(\mu) < c(\varepsilon, \nu) .$$

Choosing $\varepsilon' > 0$ such that

$$\text{(2)} \qquad \varepsilon' \leqq \frac{\theta\lambda}{4k} ,$$

and making use of (3.12) and the definition of $c(\varepsilon, \nu)$ (cf. (2.1)), we may conclude from (1) that there is n_1 with the property that

$$\text{(3)} \qquad L_n(\varepsilon', \mu) \leqq S_n(\varepsilon, \nu) \quad \text{for all} \quad n \geqq n_1 .$$

Putting $m = m(\nu)$, and

$$\text{(4)} \qquad n_2 = 4(m + 1)\, k\lambda^{-1} ,$$

and taking a natural number $n \geqq \max(n_1, n_2)$ fixed, we deduce from (3) and Lemma 4.3 that there are a one-to-one mapping $\varkappa : B^n \to B^{m+n}$, and a mapping $\delta : A^n \to B^n$ such that

$$(5) \qquad e_n'(\mu, \nu, \varkappa, \delta) < \varepsilon + \varepsilon' .$$

Let us set (cf. (3.5))

$$(6) \qquad \tau = \tau[\varkappa], \quad \mu^* = \mu\tau^{-1}T_B^m, \quad r = c_*(\theta, \nu) .$$

Let $\tilde{\nu}$ be the channel associated with ν by definition (3.40) for r defined in (6). Since μ is 1-independent, it is n-independent; hence μ is n-ergodic, $\mu\tau^{-1}$ is $(m + n)$-ergodic by (3.6) so that μ^* is $(m + n)$-ergodic as at once follows from the definition of μ^*. Consequently, $I(\tilde{\nu}\mu^*)$ as given by (3.33) makes sense.

Making use of Lemma 4.1 and Lemma 4.2, and of (4.9) and (5) yields the inequality (cf. (4.1), (4.3))

$$e_{m+n}(\tilde{\nu}\mu^*) < \frac{\varepsilon + \varepsilon'}{\theta} .$$

Then it follows from (3.35) and (4.2) that

$$(7) \qquad \mathscr{H}(\mu^* \mid \nu) \leqq \frac{\varepsilon + \varepsilon'}{\theta} k + \frac{1}{m + n} .$$

On the other hand, $\mathscr{H}(\mu^*) = \mathscr{H}(\mu\tau^{-1})$ as follows directly from definition (1.7), and Lemma 3.2 implies that (cf. (3.8))

$$\mathscr{H}(\mu\tau^{-1}) = \mathscr{H}(\mu) - \frac{m}{n} \mathscr{H}(\mu) \geqq \mathscr{H}(\mu) - \frac{m}{n} k .$$

Combining the latter inequality with (7), we deduce from (3.33) that

$$(8) \qquad I(\tilde{\nu}\mu^*) \geqq \mathscr{H}(\mu) - \frac{\varepsilon}{\theta} k - \frac{m}{n} k - \frac{\varepsilon'}{\theta} k - \frac{1}{m + n} .$$

Let μ' be defined by (4.7) so that $\mu' = \tilde{\mu}^*$ in the sense of (3.3). Then by making use of (3.29) we find according to (3.40) and (3.17) that

$$I(\tilde{\nu}\mu^*) = I(\tilde{\nu}\mu') = \int_{R(AB)} I_z \, d\tilde{\nu}\mu'(z) \leqq r$$

because $\tilde{\nu}\mu'\{I_z \leqq r\} = 1$. Since $n \geqq n_2$ we obtain from the latter relations and from (8), (1), (2), and (4) that

$$r = c_*(\theta, \nu) \geqq c(\varepsilon, \nu) - \frac{\varepsilon}{\theta} k - \lambda$$

for any $\lambda > 0$ arbitrarily small; this implies the desired result.

Let us remind that the *capacity* $C(v)$ of a stationary channel v with finite past history was defined in Section 2 by the relation

$$C(v) = \lim_{\varepsilon \to 0} c(\varepsilon, v) = c(0+, v) ; \tag{4.11}$$

the limit exists because $S_n(\varepsilon, v)$ is monotonically increasing in ε. It was shown in [6] that capacity may be defined equivalently by the expression where the lower limit is substituted by the upper limit (cf. (2.1)); viz. it holds that

$$C(v) = \lim_{\varepsilon \to 0} \bar{c}(\varepsilon, v) = \bar{c}(0+, v) .$$

Theorem 2. *If v is a stationary channel with finite past history, then its capacity $C(v)$ may be expressed in the form:*

$$C(v) = \lim_{\theta \to 0} c^*(\theta, v) = \lim_{\theta \to 0} c_*(\theta, v) .$$

Proof. Theorem 1 yields the inequality (cf. (4.11))

$$c^*(0+, v) = \lim_{\theta \to 0} c^*(\theta, v) \leqq \lim_{\varepsilon \to 0} c(\varepsilon, v) = C(v) .$$

On the other hand, Lemma 4.5 implies that

$$C(v) = \lim_{\varepsilon \to 0} c(\varepsilon, v) \leqq c_*(\theta, v) \quad \text{for all} \quad \theta > 0$$

so that

$$C(v) \leqq \lim_{\theta \to 0} c_*(\theta, v) .$$

The assertion of Theorem 2 is a consequence of the above inequalities and of Lemma 3.5.

We may define the concept of *dual capacity*, denoted by $\bar{C}(v)$ for $v \in \mathcal{N}_{\mathrm{f.past}}$, by

$$\bar{C}(v) = \lim_{\varepsilon \to 1} \bar{c}(\varepsilon, v) = \bar{c}(1-, v) . \tag{4.12}$$

It follows from Theorem 1, (3.28), and (3.24) that (cf. (3.22), (3.23))

$$C_*(v) \leqq C(v) \leqq \mathscr{C}(v) \leqq C^*(v) \leqq \bar{C}(v) \quad \text{for} \quad v \in \mathcal{N}_{\mathrm{f.past}} ,$$

because Theorem 21.1 in [6] establishes the inequality $C(v) \leqq \mathscr{C}(v)$. According to the latter inequalities, or according to Theorem 21.3 in [6] (establishing the equality $C(v) = \mathscr{C}(v)$ for $v \in \mathcal{N}_{\mathrm{erg}}$) we obtain from (3.25) that, for ergodic channels,

$$C_*(v) = C(v) = \mathscr{C}(v) = C^*(v) \quad \text{for} \quad v \in \mathcal{N}_{\mathrm{erg}} ,$$

and that, under the condition of strong stability moreover,

$$C^*(\nu) = \bar{C}(\nu) = \mathscr{C}(\nu) \quad \text{for} \quad \nu \in \mathscr{N}^*_{\text{erg}} .$$

The latter considerations lead to the conjecture that ergodicity implies strong stability.

5. SOME PROPERTIES OF DECOMPOSABLE CHANNELS

Throughout the entire section $\{\nu^\alpha\}_{\alpha\in\mathscr{A}}$ is supposed to be a measurable family of channels with parameters α in a measurable space $(\mathscr{A}, \mathbf{A})$ such that $\nu^\alpha \in \mathscr{N}_{\text{f.past}}$ for every $\alpha \in \mathscr{A}$; then (cf. (3.36))

(5.1) $\qquad m(\nu^\alpha)$ is a measurable function of parameter α ;

it is because $\nu^\alpha_\eta[x]$ is a continuous function of $\eta \in B^I$ as follows from (3.38), and because

$$\{\alpha : m(\nu^\alpha) \leqq m\} = \bigcap_n \bigcap_{x\in A^n} \bigcap_{y\in B^{m+n}} \bigcap_{\eta\in D} \{\alpha : \nu^\alpha_\eta[x] = \nu^\alpha([x] \mid u_{n,m}(y))\}$$

where $u_{n,m}$ is the uniquely determined mapping of B^{m+n} into B^I such that satisfies the relation

(5.2) $\qquad u_{n,m}(y) \in T^m_B[y]\,, \quad (u_{n,m}(y))_j = b_0\,, \quad j < -m \text{ or } j \geqq n\,,$

$y \in B^{m+n}$, $b_0 \in B$ fixed, and where D is a countable set dense in the (compact) metric space B^I (under the distance function given by (3.37)).

Let us assume that $m(\nu^\alpha) \leqq m$ for all $\alpha \in \mathscr{A}$ and for some integer m; then

(5.3) $\qquad \nu^\alpha_y(E)$ is a measurable function of (α, y)

on the space $(\mathscr{A} \times B^I, \mathbf{A} \times \mathbf{F}_B)$ for every $E \in \mathbf{F}_A$. The measurability of $\nu^\alpha_y[x]$ follows from the relation (for $x \in A^n$, n natural)

$$\{(\alpha, \eta) : \nu^\alpha_y[x] < t,\ m(\nu^\alpha) \leqq m\} =$$
$$= \bigcup_{y\in B^{m+n}}(\{\alpha : \nu^\alpha_n(x \mid y) < t\} \times T^m_B[y])$$

since the measurability of $\nu^\alpha_n(x \mid y)$ in α is guaranteed by assumption: viz. (cf. (5.2)) $\nu^\alpha_n(x \mid y) = \nu^\alpha([x] \mid u_{n,m}(y))$. The latter fact together with stationarity of ν^α and measurability of T_B imply the assertion (5.3).

We shall assume in the whole section that we are given a channel distribution ξ of the family $\{\nu^\alpha\}_{\alpha\in\mathscr{A}}$, i.e. a probability measure on the space $(\mathscr{A}, \mathbf{A})$. Throughout this section the symbol ν is reserved for the mixture of channels ν^α with respect to ξ, i.e.

(5.4) $$\nu = \int \nu^\alpha \, \mathrm{d}\xi(\alpha)$$

 (cf. (1.27)); in other words, ν is the decomposable channel with components ν^α and channel distribution ξ.

Remark. In case the assumption $m(\nu^\alpha) \leqq m$ is replaced by the assumption $m(\nu^\alpha) \leqq m$ a.s. $[\xi]$, $\nu_y^\alpha(E)$ is (α,y)-measurable almost surely so that in the following considerations the space $\mathscr{A}$ is to be substituted by its (measurable) subspace (cf. (5.1)) $\{\alpha : m(\nu^\alpha) \leqq m\}$ which is of probability one with respect to ξ.

Owing to (5.3) we may apply Fubini's theorem to the iterated integral

$$\nu\mu(G) = \iint \nu_y^\alpha(G_y)\, d\xi(\alpha)\, d\mu(y)$$

which yields the formula

$$\nu\mu(G) = \int \nu^\alpha\mu(G)\, d\xi(\alpha)\,, \tag{5.5}$$

$$\nu^\alpha\mu(G) \quad \text{is } \alpha\text{-measurable for} \quad G \in \boldsymbol{F}_{AB}\,.$$

Making use of (5.5) and (4.5) (cf. (4.4), (4.3)), we get that

$$\mathrm{e}_n(\mu, \nu, \tau) \geqq \int \mathrm{e}_n(\mu, \nu^\alpha, \tau)\, d\xi(\alpha) \tag{5.6}$$

for τ measurable, μ a probability measure on $\boldsymbol{F}_B$.

Now we shall add the assumption that $\nu \in \mathscr{N}^*_{\text{erg}}$ (cf. (1.21)) for every $\alpha \in \mathscr{A}$. Then the measurability of $S_n(\psi; \varepsilon, \nu)$ (cf. (1.18)) follows from the relation (cf. (5.1))

$$\{\alpha : m(\nu^\alpha) = m,\ S_n(\psi; \varepsilon, \nu^\alpha) \leqq k\} =$$
$$= \bigcup \{ \bigcap_{y \in F} \{\alpha : \nu_n^\alpha(\psi^{-1}\{y\} \mid y) \leqq 1 - \varepsilon\} : F \subset B^{m+n},\ \pi(B^{m+n} - F) \leqq k\}\,.$$

The latter fact guarantees measurability of $\mathscr{C}^\alpha$ (cf. (2.6)) according to (1.19) and (1.23).

Let us remark that for the lower and upper θ-quantiles $\tilde{c}(\theta, \nu)$, and $\tilde{c}'(\theta, \nu)$ of the random variable $\mathscr{C}^\alpha$ as defined by (1.28) and (2.5), the same assertions are valid as for the θ-quantiles $c_*(\theta, \omega)$, $c^*(\theta, \omega)$ as established in Lemma 3.3; summarized:

(5.7) *$\tilde{c}$ is continuous from the left at $\theta \leqq 1$ $(\theta > 0)$,*
$\tilde{c}'$ is continuous from the right at $\theta \geqq 0$ $(\theta < 1)$,
$\tilde{c}(\theta, \nu) = \tilde{c}'(\theta, \nu)$ if and only if $\tilde{c}$ is continuous at θ, or equivalently, if and only if $\tilde{c}'$ is continuous at θ.

Let us remark that (3.23) together with (3.25) implies that

$$\tilde{c}(1, \nu) = \operatorname{ess\,sup}\{\mathscr{C}^\alpha : \alpha \in \mathscr{A}[\xi]\} \leqq \log \pi(B)\,.$$

According to (5.5), $I(\nu^{\alpha}\mu)$ is a random variable which has the same probability distribution (with respect to ξ) as the random variable I_z taken with respect to $\nu\mu$ as seen from the equality

(5.8) $$\nu\,\mu\{I_z \leqq r\} = \xi\{I(\nu^{\alpha}\mu) \leqq r\} \quad (r \text{ real})$$

established in Section 2.

Theorem 3. *If ν is a decomposable channel with strongly stable ergodic components, then*

$$0 < \varepsilon < \theta \leqq 1 \quad \text{implies} \quad \bar{c}(\varepsilon, \nu) \leqq \bar{c}(\theta, \nu)\,.$$

Proof. Assume the contrary that $\bar{c}(\varepsilon, \nu) > \bar{c}(\theta, \nu)$ for some ε, θ such that $0 < \varepsilon < < \theta \leqq 1$. By (3.8) there is $\mu \in \mathcal{M}_{\text{erg}}$ such that

(1) $$\bar{c}(\theta, \nu) < \mathcal{H}(\mu) < \bar{c}(\varepsilon, \nu)\,.$$

Let us set

(2) $$r = \bar{c}(\theta, \nu)\,, \quad \Lambda = \{\alpha : \mathscr{C}(\nu^{\alpha}) < \mathcal{H}(\mu)\}\,.$$

From Lemma 4.4 we conclude that

$$\alpha \in \Lambda \quad \text{implies} \quad \lim_n \mathrm{e}_n(\mu, \nu^{\alpha}) = 1$$

so that

(3) $$\{\alpha : \lim_n \mathrm{e}_n(\mu, \nu^{\alpha}) = 1\} \supset \Lambda \supset \{\alpha : \mathscr{C}^{\alpha} \leqq r\}$$

let us mention that measurability of $\mathrm{e}_n(\mu, \nu^{\alpha})$ is guaranteed by (5.5) according to (4.10). Since $\xi\{\mathscr{C}^{\alpha} \leqq r\} \geqq \theta$ as follows from (2), we obtain from (3) that

(4) $$\xi\{\alpha : \lim_n \mathrm{e}_n(\mu, \nu^{\alpha}) = 1\} \geqq \theta\,.$$

Choose $\varepsilon' > 0$ such that

(5) $$\varepsilon + 3\varepsilon' < \theta\,.$$

Then we deduce from (4) that there is a subscript n_0 such that

$$\xi\{\alpha : \mathrm{e}_n(\mu, \nu^{\alpha}) \geqq 1 - \varepsilon'\} \geqq \theta - \varepsilon' \quad \text{for all} \quad n \geqq n_0\,.$$

By using the inequality (5.6) we find that

$$\mathrm{e}_n(\mu, \nu, \tau[\varkappa] \geqq \int_{\{\alpha : e_n(\mu, \nu^{\alpha}) \geqq 1 - \varepsilon'\}} \mathrm{e}_n(\mu, \nu^{\alpha}, \tau[\varkappa])\, \mathrm{d}\xi(\alpha) \geqq (1 - \varepsilon')(\theta - \varepsilon')$$

 for every $\varkappa : B^n \to B^{m+n}$ so that

(6) $$e_n(\mu, \nu) > \theta - 2\varepsilon' \quad \text{for all} \quad n \geqq n_0 .$$

On the other hand, it follows from (1) according to (3.12) and definition of $\bar{c}(\varepsilon, \nu)$ that

$$L_n(\varepsilon', \mu) < S_n(\varepsilon, \nu) \quad \text{for some} \quad n \geqq n_0 .$$

From here and from Lemma 4.3 we deduce that the inequality (cf. (4.8), (4.9)) $e_n(\mu, \nu) < \varepsilon + \varepsilon'$ is valid for some $n \geqq n_0$. We have by (6) and (5) for the n:

$$e_n(\mu, \nu) < \varepsilon + \varepsilon' < \theta - 2\varepsilon' < e_n(\mu, \nu)$$

which gives the desired contradiction.

In the proof of Theorem 4 we shall make use of the following auxiliary quantities:

(5.9) $$\mathscr{C}^*(\nu^\alpha) = \limsup_n \left(\frac{1}{n} \sup\nolimits_{\mu \in \mathscr{M}_n(B)} \mathscr{R}_n(\nu^\alpha \mu)\right), \quad \alpha \in \mathscr{A} ,$$

where $\mathscr{M}_n(B)$ represents the class of all n-independent measures; cf. (3.4), (3.14). In what follows we shall put

(5.10) $$\mathscr{M}_n^*(B) = \{\mu : \mu \in \mathscr{M}_n(B),\ \mu_n\{y\} \text{ is rational for all } y \in B^n\} .$$

Then it is easy to see that

(5.11) $$\sup \{\mathscr{R}_n(\nu^\alpha \mu) : \mu \in \mathscr{M}_n^*(B)\} = \sup \{\mathscr{R}_n(\nu^\alpha \mu) : \mu \in \mathscr{M}_n(B)\} .$$

The latter relation shows that $\mathscr{C}^*(\nu^\alpha)$ is a measurable function of parameter α. It follows from definition (5.9) that

(5.12) $$\mathscr{C}^*(\nu^\alpha) \geqq \mathscr{C}(\nu^\alpha) , \quad \alpha \in \mathscr{A} .$$

Lemma: $C^*(\nu) \geqq \text{ess.sup} \{\mathscr{C}^*(\nu^\alpha) : \alpha \in \mathscr{A}[\xi]\}$.

Proof. From the definition of $C^* = C^*(\nu)$ it follows that

$$\xi\{\alpha : I(\nu^\alpha \mu) \leqq C^*\} = 1 \quad \text{for every} \quad \mu \in \mathscr{M}_{\text{erg}} .$$

Putting (cf. (3.4))

$$\tilde{\mathscr{M}}_n^* = \{\tilde{\mu} : \mu \in \mathscr{M}_n^*(B)\} \subset \tilde{\mathscr{M}}_n(B) , \quad \tilde{\mathscr{M}}^* = \bigcup_{n=1}^{\infty} \tilde{\mathscr{M}}_n^* ,$$

$$\Lambda = \bigcap\nolimits_{\tilde{\mu} \in \tilde{\mathscr{M}}^*} \{\alpha : I(\nu^\alpha \tilde{\mu}) \leqq C^*\} ,$$

we obtain from the preceding relation that ($\tilde{\mathscr{M}}^*$ countable)

(1) $$\xi(\Lambda) = 1 .$$

Choose $\lambda > 0$ arbitrarily. Given $\alpha \in A$, $\mu \in \mathcal{M}_n^*(B)$, n natural, Lemma 3.6 and (3.15) yield the relations

$$\frac{1}{n}\mathcal{R}_n(v^\alpha\mu) \leqq \frac{1}{kn}\mathcal{R}_{kn}(v^\alpha\mu) < I(v^\alpha\mu) + \lambda \leqq C^* + \lambda$$

for some k. Hence we deduce according to (5.11) that

$$\mathscr{C}^*(v^\alpha) \leqq C^* + \lambda \quad \text{for every} \quad \lambda > 0\,, \quad \alpha \in A\,,$$

i.e. $\mathscr{C}^*(v^\alpha) \leqq C^*$ for every $\alpha \in A$ which together with (1) gives the desired result.

Theorem 4. *If v is a decomposable channel with strongly stable ergodic components, then*

$$\lim_{\theta\to 1} c^*(\theta, v) = \lim_{\theta\to 1} c_*(\theta, v) = \lim_{\varepsilon\to 1} c(\varepsilon, v) = \lim_{\varepsilon\to 1} \bar{c}(\varepsilon, v) = C^*(v) = \bar{C}(v) =$$
$$= \text{ess.sup}\,\{\mathscr{C}(v^\alpha) : \alpha \in \mathscr{A}[\xi]\}\,.$$

Proof. Using the preceding lemma and inequality (5.12) we obtain the relation

$$C^*(v) \geqq \text{ess.sup}\,\mathscr{C}^\alpha\,.$$

On the other hand, Lemma 2.1, Lemma 3.5, and (5.7) yield the relations

$$C^*(v) = \lim_{\theta\to 1} c^*(\theta, v) \leqq \lim_{\theta\to 1} \bar{c}(\theta, v) = \bar{c}(1, v) = \text{ess.sup}\,\mathscr{C}^\alpha\,.$$

The two inequalities just derived together with Theorem 1 and Theorem 3 imply the assertion of the theorem.

Theorem 5. *Any channel decomposable into components with additive ergodic noise is regular.*

Proof. Assume that the noise distribution of channel v^α is $\mu^\alpha \in \mathcal{M}_{\text{erg}}$. Then we deduce according to (1.32) that the composed channel given by (5.4) has additive noise with noise distribution $\bar{\mu}$ expressed by, written symbolically,

$$\bar{\mu} = \int \mu^\alpha\, \mathrm{d}\xi(\alpha)\,.$$

Theorem 3.1 stated in [4] yields the inequalities

$$\log \pi(A) - h^*(\varepsilon) \leqq c(\varepsilon, v) \leqq \bar{c}(\varepsilon, v) \leqq \log \pi(A) - h_*(\varepsilon)\,,$$

where

$$h^*(\theta) = \sup\,\{r : \bar{\mu}\{\mathscr{H}(\mu_z) \geqq r\} \geqq 1 - \theta\}\,,$$
$$h_*(\theta) = \inf\,\{r : \bar{\mu}\{\mathscr{H}(\mu_z) \leqq r\} \geqq \theta\}\,.$$

 It is easy to derive the equalities

$$\mathscr{C}(\nu^{\alpha}) = \log \pi(A) - \mathscr{H}(\mu^{\alpha}),$$
$$\bar{\mu}\{\mathscr{H}(\mu_z) \leqq r\} = \xi\{\alpha : \mathscr{H}(\mu^{\alpha}) \leqq r\}$$

which imply the relations

$$\tilde{c}(\theta, \nu) = \log \pi(A) - h^*(\theta), \quad \tilde{c}'(\theta, \nu) = \log \pi(A) - h_*(\theta).$$

The latter relations together with the preceding ones show that $c(\varepsilon, \nu) = \bar{c}(\varepsilon, \nu) = \tilde{c}(\varepsilon, \nu)$ except a countable set of ε's. Our considerations are valid without any change for every channel $\nu^{[r]}$ as defined by (1.29). From here and from the main theorem established on the basis of Theorems 1 to 4 we conclude that the channel ν, i.e. the channel distribution ξ must be regular, Q.E.D.

(Received December 28, 1970.)

REFERENCES

To both Parts:

[1] A. I. Khinchin: Mathematical foundations of information theory. Dover Publications, New York 1957.

[2] J. Nedoma: Capacity of a discrete channel. Transact. First Prague Conf. on Inform.Theory etc. Prague 1957, 143—181.

[3] J. Nedoma: On non-ergodic channels. Transact. Second Prague Conf. on Inform. Theory etc. Prague 1960, 363—395.

[4] K. R. Parthasarathy: Effective entropy rate and transmission of information through channels with additive random noise. Sankhya *A 25* (1963), 75—84.

[5] V. A. Rohlin: New progress in the theory of transformation with invariant measure (in Russian). Usp. Mat. Nauk *15* (1960), 3—26.

[6] K. Winkelbauer: Channels with finite past history. Transact. Second Prague Conf. on Inform. Theory etc. Prague 1960, 685—831.

[7] K. Winkelbauer: On discrete information sources. Transact. Third Prague Conf. on Inform. Theory etc. Prague 1964, 765—830.

[8] K. Winkelbauer: Axiomatic definition of channel capacity and entropy rate. Transact. Fourth Prague Conf. on Inform. Theory etc., Prague 1967, 661—705.

[9] K. Winkelbauer: On the asymptotic rate of non-ergodic information sources. Kybernetika *6* (1970), 2, 127—148.

[10] J. Wolfowitz: Coding theorems of information theory. Second ed. Berlin 1964.

To Part II only:

[11] P. Billingsley: Ergodic Theory and Information. New York 1965.

[12] J. Nedoma: Über die Ergodizität und *r*-Ergodizität stationärer Wahrscheinlichkeitsmasse. Z. Wahrscheinlichkeitstheorie *2* (1963), 90—97.

[13] K. R. Parthasarathy: On the Integral Representation of the Rate of Transmission of a Stationary Channel. Illinois Journ. Math. *2* (1961), 299—305.

K větě o kódování pro rozložitelné diskrétní sdělovací kanály II

Karel Winkelbauer

Tato druhá část článku je zcela věnována důkazům teorémů, jež byly použity v první části při důkazu hlavní věty o existenci ε-kapacity (srovn. první část článku v minulém čísle časopisu).

Doc. RNDr. Mg. Mat. Karel Winkelbauer, DrSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Vyšehradská 49, Praha 2.