

[F1] H. Freudenthal, *Neuaufbau der Endenitheorie*, Ann. of Math. vol. 43, no. 2 (1942) pp. 261-279.

[F2] ———, *Über die Enden diskreter Räume und Gruppen*, Comment. Math. Helv. vol. 17 (1944-1945) pp. 1-38.

[HW] W. Hurewicz and H. Wallman, *Dimension theory*, Princeton, N. J., Princeton University Press, 1948.

[K] J. L. Kelley, *General topology*, New York, Van Nostrand, 1955.

[Z] L. Zippin, *On semicompact spaces*, Amer. J. Math. vol. 57 (1935) pp. 327-341.

UNIVERSITY OF WISCONSIN, MILWAUKEE

ERGODIC AND MIXING PROPERTIES OF INFINITE MEMORY CHANNELS¹

ROY L. ADLER

1. Introduction. A. I. Khinchin [5] states that if an ergodic message space is fed into a channel with finite memory then the output message space is ergodic along with the compound message space of the input with the output. However, Khinchin's notion of finite memory, definition (1) below, is apparently insufficient to yield this result. K. Takano [6] has been able to establish it by strengthening the definition of finite memory to include both (1) and (2) below. The essential requirement for this theorem, however, is really (3) of which (2) is a special case. Condition (3) expresses that the output of a channel be asymptotically independent from the remote past of the input. The method of proof is an application of a functional form of the notion of ergodicity involving Cesàro convergence of a certain sequence of integrals. In addition this technique can be used to discuss some of the mixing and ergodic properties of the output with respect to the input and the channel.

2. Notation and Definitions.² Let (X, \mathfrak{X}) be a measurable space with X a space of points and \mathfrak{X} a sigma-field of measurable subsets of X . Usually in information theory X is referred to as an alphabet and is a finite set of points. We shall make no such restriction here. Con-

Received by the editors August 11, 1960 and, in revised form, December 9, 1960.

¹ Part of this work was supported by Air Force Contract No. SAR AF-49(638)-224 while the author was at Yale University.

² For complete definitions of concepts and standard results of measure theory and ergodic theory consult [2; 3] respectively.

sider the two-sided infinite product measurable space $(X^*, \mathfrak{X}^*) = \prod_{i=-\infty}^{\infty} (X_i, \mathfrak{X}_i)$ where $X_i = X$, $\mathfrak{X}_i = \mathfrak{X}$, $i = 0, \pm 1, \pm 2, \dots$, and let λ be a probability measure on \mathfrak{X}^* , i.e., λ is a countably additive measure on \mathfrak{X}^* with $\lambda(X^*) = 1$. We shall call the measure space $(X^*, \mathfrak{X}^*, \lambda)$ a *message space*. Let σ denote the shift transformation on any two-sided infinite product space: i.e., if $x^* = (\dots, x_{-1}, x_0, x_1, \dots)$ then $\sigma x^* = (\dots, x'_{-1}, x'_0, x'_1, \dots)$ where $x'_i = x_{i+1}$, $i = 0, \pm 1, \pm 2, \dots$. A message space $(X^*, \mathfrak{X}^*, \lambda)$ is *stationary* if σ is a λ -measuring preserving transformation or, equivalently stated, λ is a σ -invariant measure. We also sometimes say that the measure λ is stationary. A stationary message space $(X^*, \mathfrak{X}^*, \lambda)$ is *ergodic* if σ is an ergodic λ -measure preserving transformation, and in this case we sometimes say that λ is an ergodic measure. Consider a message space $(X^*, \mathfrak{X}^*, \lambda)$ along with another two-sided infinite product measurable space $(Y^*, \mathfrak{Y}^*) = \prod_{i=-\infty}^{\infty} (Y_i, \mathfrak{Y}_i)$ where $Y_i = Y$ is a set of points and $\mathfrak{Y}_i = \mathfrak{Y}$ is a sigma-field of measurable subsets of Y , $i = 0, \pm 1, \pm 2, \dots$. A *channel* is defined to be a triple $((X^*, \mathfrak{X}^*, \lambda), (Y^*, \mathfrak{Y}^*), \nu(\cdot, \cdot))$ where $\nu(\cdot, \cdot)$ is a function on $X^* \times \mathfrak{Y}^*$ which satisfies

- (i) $\nu(x^*, \cdot)$ is a probability measure on \mathfrak{Y}^* for λ -almost all x^* .
- (ii) $\nu(\cdot, E^*)$ is a measurable function of x^* for each fixed set $E^* \in \mathfrak{Y}^*$.

We can refer to $(X^*, \mathfrak{X}^*, \lambda)$ as the *input* to the channel and (Y^*, \mathfrak{Y}^*) as the *output* and shall abbreviate $((X^*, \mathfrak{X}^*, \lambda), (Y^*, \mathfrak{Y}^*), \nu(\cdot, \cdot))$ to (X^*, Y^*, ν) . A channel is *stationary* if $\nu(\sigma x^*, \sigma E^*) = \nu(x^*, E^*)$ for λ -almost all x^* for each $E^* \in \mathfrak{Y}^*$. An input message space $(X^*, \mathfrak{X}^*, \lambda)$ and a channel (X^*, Y^*, ν) define a *compound message space* $(X^* \times Y^*, \mathfrak{X}^* \times \mathfrak{Y}^*, \omega)$ where ω is a probability measure on $\mathfrak{X}^* \times \mathfrak{Y}^*$ uniquely determined by its value

$$\omega(A^{**}) = \int_{C^*} \nu(x^*, E^*) \lambda(dx^*)$$

for cylinder sets $A^{**} = C^* \times E^*$, $C^* \in \mathfrak{X}^*$, $E^* \in \mathfrak{Y}^*$. Next we can form the output message space $(Y^*, \mathfrak{Y}^*, \mu)$ where μ is given by

$$\mu(E^*) = \omega(X^* \times E^*) = \int_{X^*} \nu(x^*, E^*) \lambda(dx^*)$$

for $E^* \in \mathfrak{Y}^*$. Finally, it is clear what the shift transformation σ is on $X^* \times Y^*$: it is the direct product transformations of shifts on each space X^* and Y^* ; i.e. $\sigma(x^*, y^*) = (\sigma x^*, \sigma y^*)$.

* This is sometimes referred to as the *source-channel*, the channel itself being $((X^*, \mathfrak{X}^*), (Y^*, \mathfrak{Y}^*), \nu(\cdot, \cdot))$ with ν satisfying (i) above for all x^* .

For the sake of completeness we next state and prove the standard proposition.

PROPOSITION. *If $(X^*, \mathfrak{X}^*, \lambda)$ is a stationary message space and (X^*, Y^*, ν) is a stationary channel then both the compound message space and the output message space are stationary.*

PROOF. Let $A^{**} = C^* \times E^*$, $C^* \in \mathfrak{X}^*$, $E^* \in \mathfrak{Y}^*$. $\omega(\sigma A^{**}) = \omega(\sigma C^* \times \sigma E^*) = \int_{\sigma C^*} \nu(x^*, \sigma E^*) \lambda(dx^*)$. Let $z^* = \sigma^{-1}x^*$. Changing variables we get $\omega(\sigma A^{**}) = \int_{C^*} \nu(\sigma z^*, \sigma E^*) \lambda(\sigma dz^*)$. Then by stationarity of channel and σ -invariance of λ we have

$$\omega(\sigma A^{**}) = \int_{C^*} \nu(z^*, E^*) \lambda(dz^*) = \omega(A^{**}).$$

From this we can conclude $\omega(\sigma A^{**}) = \omega(A^{**})$ for arbitrary $A^{**} \in \mathfrak{X}^* \times \mathfrak{Y}^*$ which is the desired result.

3. Properties of channels and main theorem.

(1) **DEFINITION.** A channel (X^*, Y^*, ν) is said to have *finite memory* provided there exists a fixed positive integer m such that if E^* is any cylinder set in \mathfrak{Y}^* of the form $E^* = \dots \times Y \times E_t \times E_{t+1} \times \dots \times E_{t+n-1} \times Y \times \dots$, $E_t, E_{t+1}, \dots, E_{t+n-1} \in \mathfrak{Y}$ and if $x_i = x'_i$ for $i = t - m, \dots, t + n - 1$, where x_i is the i th coordinate of x^* and x'_i is the i th coordinate of $x^{*'}$, t an arbitrary integer and n non-negative, then

$$\nu(x^*, E^*) = \nu(x^{*'}, E^*).$$

The smallest such m is called the *length of the memory*. If no such m exists the channel is said to have *infinite memory*.

(2) **DEFINITION.** A channel (X^*, Y^*, ν) is *independent from the remote past* provided there exists a positive integer m such that if E^* and F^* are two cylinder sets in \mathfrak{Y}^* of the form $E^* = \dots \times Y \times E_i \times E_{i+1} \times \dots \times E_j \times Y \times \dots$, $F^* = \dots \times Y \times F_l \times F_{l+1} \times \dots \times F_k \times Y \times \dots$ where $E_i, E_{i+1}, \dots, E_j, F_l, F_{l+1}, \dots, F_k \in \mathfrak{Y}^*$ and i, j, l, k , are integers satisfying $i \leq j, l \leq k, j + m < l$ then

$$\nu(x^*, E^* \cap F^*) = \nu(x^*, E^*) \cdot \nu(x^*, F^*)$$

for λ -almost all x^* . The smallest such m gives the *order of remoteness*.

(3) **DEFINITION.** We shall say a channel (X^*, Y^*, ν) is *asymptotically independent from the remote past* if for any two cylinder sets E^*, F^* in \mathfrak{Y}^*

$$\lim_{n \rightarrow \infty} [\nu(x^*, \sigma^n E^* \cap F^*) - \nu(x^*, \sigma^n E^*) \nu(x^*, F^*)] = 0$$

for λ -almost all x^* . Clearly (2) \Rightarrow (3). We are now in a position to give a proof of Khinchin's statement for a general class of channels obeying the asymptotic independence property above.

THEOREM. *Let $(X^*, \mathfrak{X}^*, \lambda)$ be an ergodic stationary message space and (X^*, Y^*, ν) a stationary channel asymptotically independent from the remote past. Then the compound message space $(X^* \times Y^*, \mathfrak{X}^* \times \mathfrak{Y}^*, \omega)$ as well as the output message space $(Y^*, \mathfrak{Y}^*, \mu)$ are ergodic.*

PROOF. For ergodicity it suffices to prove

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \omega(\sigma^n A^{**} \cap B^{**}) = \omega(A^{**})\omega(B^{**})$$

where $A^{**} = C^* \times E^*$, $B^{**} = D^* \times F^*$, with C^* , D^* cylinder sets in \mathfrak{X}^* and E^* , F^* cylinder sets in \mathfrak{Y}^* . Let $\chi_{C^*}(x^*)$ and $\chi_{D^*}(x^*)$ denote the characteristic function of C^* and D^* respectively. By stationarity of the channel we have the second equality below, and from the functional form of ergodicity of σ on X^* with respect to the measure λ we obtain the limit (see [2, p. 36]).

$$\begin{aligned} & \frac{1}{N} \sum_{n=0}^{N-1} \int_{\sigma^n C^* \cap D^*} \nu(x^*, \sigma^n E^*) \nu(x^*, F^*) \lambda(dx^*) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \int_{X^*} \nu(x^*, \sigma^n E^*) \chi_{C^*}(\sigma^{-n} x^*) \nu(x^*, F^*) \chi_{D^*}(x^*) \lambda(dx^*) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \int_{X^*} \nu(\sigma^{-n} x^*, E^*) \chi_{C^*}(\sigma^{-n} x^*) \nu(x^*, F^*) \chi_{D^*}(x^*) \lambda(dx^*) \\ &\rightarrow \int_{X^*} \nu(x^*, E^*) \chi_{C^*}(x^*) \lambda(dx^*) \cdot \int_{X^*} \nu(x^*, F^*) \chi_{D^*}(x^*) \lambda(dx^*) \\ &= \omega(A^{**})\omega(B^{**}), \qquad \text{as } N \rightarrow \infty. \end{aligned}$$

By the asymptotic independence from the remote past we have as $n \rightarrow \infty$

$$[\nu(x^*, \sigma^n E^* \cap F^*) - \nu(x^*, \sigma^n E^*)\nu(x^*, F^*)] \rightarrow 0$$

for λ -almost all x^* . Then even though the set over which we integrate varies with n , we can still obtain from dominated convergence the limit

$$\int_{\sigma^n C^* \cap D^*} [\nu(x^*, \sigma^n E^* \cap F^*) - \nu(x^*, \sigma^n E^*)\nu(x^*, F^*)] \lambda(dx^*) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Since ordinary convergence implies Cesàro convergence,

$$\frac{1}{N} \sum_{n=0}^{N-1} \int_{\sigma^n C^* \cap D^*} [\nu(x^*, \sigma^n E^* \cap F^*) - \nu(x^*, \sigma^n E^*)\nu(x^*, F^*)] \lambda(dx^*) \rightarrow 0,$$

as $N \rightarrow \infty$.

Combining the above limits we get

$$\begin{aligned} & \left[\frac{1}{N} \sum_{n=0}^{N-1} \omega(\sigma^n A^{**} \cap B^{**}) - \omega(A^{**})\omega(B^{**}) \right] \\ &= \left[\frac{1}{N} \sum_{n=0}^{N-1} \int_{\sigma^n C^* \cap D^*} \nu(x^*, \sigma^n E^* \cap F^*) (dx^*) \right. \\ & \quad \left. - \frac{1}{N} \sum_{n=0}^{N-1} \int_{\sigma^n C^* \cap D^*} \nu(x^*, \sigma^n E^*)\nu(x^*, F^*) \lambda(dx^*) \right] \\ & \quad + \left[\frac{1}{N} \sum_{n=0}^{N-1} \int_{\sigma^n C^* \cap D^*} \nu(x^*, \sigma^n E^*)\nu(x^*, F^*) \lambda(dx^*) - \omega(A^{**})\omega(B^{**}) \right] \\ & \rightarrow 0, \text{ as } N \rightarrow \infty. \end{aligned} \tag{Q.E.D.}$$

4. Some ergodic and mixing properties. For what follows measures and channels will always be assumed stationary. The above theorem can also be proved using a slightly weaker notion than asymptotic independence from the past. In view of the close resemblance of (3) to the notion of strong mixing for measure preserving transformations (see [2, p. 37]), let us agree to also call a channel satisfying (3) *strongly mixing* (SM). We shall say a channel is *weakly mixing* (WM) if for any two cylinder sets E^*, F^* in \mathfrak{Y}^* there exists a sequence of integers J of density zero such that for n restricted to be outside J the sequence

$$[\nu(x^*, \sigma^n E^* \cap F^*) - \nu(x^*, \sigma^n E^*)\nu(x^*, F^*)]$$

converges to zero for λ -almost all x^* . This is equivalent to the condition

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} |\nu(x^*, \sigma^n E^* \cap F^*) - \nu(x^*, \sigma^n E^*)\nu(x^*, F^*)| = 0$$

for λ -almost all x^* (see [2, p. 38]). Continuing in this manner we call a channel *ergodic* (ERG) if for any $E^*, F^* \in \mathfrak{Y}^*$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} [\nu(x^*, \sigma^n E^* \cap F^*) - \nu(x^*, \sigma^n E^*)\nu(x^*, F^*)] = 0$$

for λ -almost all x^* .

Furthermore the input measure λ is *strongly mixing* (SM) if for any sets $C^*, D^* \in \mathfrak{X}^*$

$$\lim_{n \rightarrow \infty} [\lambda(\sigma^n C^* \cap D^*) - \lambda(C^*)\lambda(D^*)] = 0;$$

it is *weakly mixing* (WM) if for any $C^*, D^* \in \mathfrak{X}^*$,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} |\lambda(\sigma^n C^* \cap D^*) - \lambda(C^*)\lambda(D^*)| = 0;$$

or equivalently there exists a sequence of integers J of density zero such that for n restricted to be outside J we have

$$\lim_{n \rightarrow \infty} \lambda(\sigma^n C^* \cap D^*) = \lambda(C^*)\lambda(D^*);$$

it is *ergodic* (ERG) if for $C^*, D^* \in \mathfrak{X}^*$,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} [\lambda(\sigma^n C^* \cap D^*) - \lambda(C^*)\lambda(D^*)] = 0.$$

Similar statements apply to the measure ω . Furthermore it is clear for both measures and channels that ERG is a weaker property than WM which is in turn weaker than SM.

We have proved in the main theorem

$$\lambda\text{-ERG and channel-SM} \Rightarrow \omega\text{-ERG}.$$

With merely minor modifications in the argument of this theorem we can establish the following.

THEOREM. *If the input measure λ is P and the channel is Q then the compound measure ω is the weaker of P and Q where P is any of the properties SM, WM, or ERG and Q is either SM or WM.*

These facts are tabulated below:

		channel		
		SM	WM	ERG
Input measure	SM	SM	WM	(*)
	WM	WM	WM	(*)
	ERG	ERG	ERG	(**)

where inside the table are read off the properties of ω determined by

the input and channel properties listed on the side and top. Furthermore these are the strongest statements which are possible in general. This becomes evident by considering $\nu(x^*, E^*) = \mu(E^*)$ for all x^* . Then ω reduces to the direct product measure $\lambda \times \mu$ for which it is easy to verify our assertion. In fact in this case (*) = ERG while (**) may or may not be ERG (necessary and sufficient conditions for (**) = ERG have been provided by Y. Kawada [4]). However, in the more general case of channels, the technique of this paper breaks down; neither a proof nor a counter example—which is the more likely—for (*) = ERG has yet been found; and no necessary and sufficient conditions generalizing Kawada's results for (**) = ERG are known.

BIBLIOGRAPHY

1. A. Feinstein, *Foundations of information theory*, New York, McGraw-Hill Book Co., Inc., 1958.
2. P. R. Halmos, *Lectures in ergodic theory*, Tokyo, The Mathematics Society of Japan, 1956.
3. ———, *Measure theory*, Princeton, D. Van Nostrand Co., Inc., 1950.
4. Y. Kawada, *Über die massstreuen Abbildungen in Producträumen*, Proceedings of the Imperial Academy of Tokyo vol. 19 (1943) pp. 525–527.
5. A. I. Khinchin, *Mathematical foundations of information theory*, English translation, New York, Dover Publications, Inc., 1958.
6. K. Takano, *On the basic theorems of information theory*, Ann. Inst. Statist. Math. Tokyo vol. 9 (1958) pp. 53–77.

YALE UNIVERSITY AND

INTERNATIONAL BUSINESS MACHINES CORPORATION