The Annals of Mathematical Statistics
1970, Vol. 41, No. 6, 2093–2104

THE CAPACITY AND AMBIGUITY OF A TRANSDUCER[1]

BY WILLIAM M. CONNER

University of Pittsburgh

A particular noiseless, discrete channel with memory (called a transducer) is made to correspond to a function in the unit square by associating the infinite sequences of symbols of the transducer with the expansions of points in the unit interval. It is shown that the Hausdorff dimension of the set of points received over the transducer is equal to the transducer capacity. A definition of ambiguity is given which has a geometric interpretation in the square, and it is shown that the transducer has a homogeneity property by proving that the ambiguity is almost everywhere the same.

1. Introduction. Besicovitch [1], Eggleston [7], [8] and others have calculated the Hausdorff dimension of subsets of the unit interval defined by placing certain restrictions on the digits of expansions of numbers. For example, let $M(p)$, $0 \leqq p \leqq 1$, be the set of points in the unit interval containing 1 in their dyadic expansions in the proportion p, i.e., $x = .x_1 x_2 \cdots$ belongs to $M(p)$ if and only if $\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} x_k = p$. Eggleston [7] has shown that the dimension of $M(p)$ is $-p \log_2 p - (1-p) \log_2(1-p)$. (A simplified proof is obtained by using a general theorem due to Billingsley [5] page 142.)

Observing that this value for the dimension is the entropy of an information source, Kinney [14] and Billingsley [2], [3], [4] sought a connection between dimension theory and information theory by making the infinite sequences of symbols from the source correspond to the expansions of points in the unit interval. Theorem 1 of Kinney's paper asserts the existence of a set of measure one whose dimension is the entropy of a Markov source. Dym [6] recently extended this theorem to general stationary, ergodic sources. Theorem 2 of Kinney's paper is concerned with noiseless coding and shows that the dimension of a certain set corresponding to the coded messages is equal to the capacity of the noiseless channel. Smorodinsky [16] recently extended this theorem to very general noiseless channels.

As yet little work has been done on examining the correspondence between a discrete, noisy channel with memory and the unit square. Such an investigation is begun in this paper, although not in this generality. We introduce the element of memory but not noise, and examine the following particular type of noiseless channel with memory.

Let $S = \{0, 1, \cdots, b-1\}$ where $b \geqq 2$ is an integer and let m be a positive integer. Let $\mathscr{F}_m$ be the set of all functions f^* from $S^m = \underbrace{S \times S \times \cdots \times S}_{m \text{ factors}}$ into S. We note that $\operatorname{crd} \mathscr{F}_m = b^{b^m}$ where crd denotes the cardinal number of a set.

Received January 7, 1970.

[1] This paper is taken from the author's doctoral dissertation written at Michigan State University under the direction of Professor John R. Kinney, to whom the author wishes to express his appreciation.

Let $f^* \in \mathscr{F}_m$ and let n be a positive integer. We define a function f_n from S^{n+m-1} into S^n as follows. Let $x = x_1 \cdots x_{n+m-1} \in S^{n+m-1}$ and let $y_i = f^*(x_i \cdots x_{i+m-1})$, $i = 1, \cdots, n$. Then the sequence $y = y_1 \cdots y_n \in S^n$, and we define $f_n(x) = y$. Note that $f_1 = f^*$.

Now let $x \in (0, 1]$ and let $\sum_{i=1}^{\infty} x_i b^{-i}$ be the (unique) nonterminating base b expansion of x. We define a function f from (0, 1] into [0, 1] by $f(x) = \sum_{i=1}^{\infty} y_i b^{-i}$ where $y_i = f^*(x_i \cdots x_{i+m-1})$, $i = 1, 2, \cdots$. In the terminology of Shannon [15], the function f is an example of a transducer of memory m. In the terminology of Feinstein [9], Billingsley [5], Khinchin [13], and others f is a noiseless, discrete channel (or code) of memory m. We will call $\{x_i\}$ the input sequence and $\{y_i\}$ the output (or received) sequence; and, following Shannon, we will call f a transducer of memory m.

We now introduce some notation and definitions. Let $x \in (0, 1]$ and let $x = \sum_{i=1}^{\infty} x_i b^{-i}$ be the nonterminating base b expansion of x. Define $b_i(x) = x_i$ for all i, i.e., $b_i(x)$ is the ith digit of the nonterminating base b expansion of x. A set of the form $\{x : b_i(x) = s_i, i = 1, \cdots, n\}$, where $s_i \in S$, is denoted by $[s_1, \cdots, s_n]$ and is called a cylinder of length b^{-n}. Note that $[s_1, \cdots, s_n]$ is a half-open (open on the left) b-adic interval of length b^{-n} (i.e., an interval of the form $(j/b^n, (j+1)/b^n]$ for some j, $0 \leqq j \leqq b^n - 1$).

The following definition of dimension in the unit interval, which extends Hausdorff's original definition, is given by Billingsley [5]. Let $M \subset (0, 1]$, let α and ρ be positive real numbers, and let μ be a probability measure on the Borel sets of (0, 1]. Define $\mu_\alpha(M, \rho) = \inf \sum_i \mu(v_i)^\alpha$, where the infimum is taken over all μ-ρ-coverings of M, a μ-ρ-covering being a covering by cylinders v_i with $\mu(v_i) < \rho$. It is clear that $\mu_\alpha(M, \rho) \leqq \mu_\alpha(M, \rho')$ for $\rho' < \rho$, so the limit

$$\mu_\alpha(M) = \lim_{\rho \to 0} \mu_\alpha(M, \rho)$$

exists (but may be infinite). It can be shown that for fixed M there is an α_0 such that $\mu_\alpha(M) = \infty$ for $\alpha < \alpha_0$ and $\mu_\alpha(M) = 0$ for $\alpha > \alpha_0$. The number α_0 is called the (Hausdorff) dimension of M with respect to μ and is denoted by $\dim_\mu M$. We will denote Lebesgue measure by L and will write $\dim M$ instead of $\dim_L M$.

Let $\mathscr{B}$ be the Borel sets in (0, 1] and define a transformation T on (0, 1] by $Tx = \sum_{i=1}^{\infty} b_{i+1}(x) b^{-i}$ for $x \in (0, 1]$; T is a left shift on the digits of the base b expansion of x. Let μ be a probability measure on $\mathscr{B}$; then if $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathscr{B}$, T is said to be measure preserving. In such a case we will say that μ is stationary. T is called ergodic under μ if for each $B \in \mathscr{B}$ such that $T^{-1}B = B$, $\mu(B)$ is either zero or one. In such a case we will say that μ is ergodic, omitting reference to T. We will denote by $\mathscr{M}$ the set of all probability measures on $\mathscr{B}$ which are stationary and ergodic.

Entropy was introduced into information theory by Shannon [15]; Kolmogorov and Sinai have extended the notion of entropy to general measure preserving transformations (see Billingsley [5]). The entropy of a stationary probability measure μ is defined as follows. For each positive integer n define

$$H_n(\mu) = -\sum \mu([s_1, \cdots, s_n]) \log \mu([s_1, \cdots, s_n]),$$

where the summation is taken over all $s_1 \cdots s_n \in S^n$, and where we take $0 \log 0$ to be zero. It can be shown (see, for example, Feinstein [9] page 85) that the limit $H(\mu) = \lim_{n\to\infty} H_n(\mu)/n$ exists. We call $H(\mu)$ the entropy of μ.

Similar definitions can be given for the unit square. Define a transformation T_1 on $(0, 1] \times (0, 1]$ by $T_1(x, y) = (Tx, Ty)$ for each $(x, y) \in (0, 1] \times (0, 1]$. As before a probability measure P on the Borel sets of $(0, 1] \times (0, 1]$ will be called stationary if T_1 preserves P. For stationary P we define

$$H_n(P) = -\sum P([s_1, \cdots, s_n, r_1, \cdots, r_n]) \log P([s_1, \cdots, s_n, r_1, \cdots, r_n]),$$

where $[s_1, \cdots, s_n, r_1, \cdots, r_n] = \{(x, y): b_i(x) = s_i,\ b_i(y) = r_i,\ 1 \leqq i \leqq n\}$, and where the summation is taken over all $s_1 \cdots s_n,\ r_1 \cdots r_n \in S^n$. Again it can be shown (Feinstein [9] page 87) that the limit $H(P) = \lim_{n\to\infty} H_n(P)/n$, called the entropy of P, exists.

In Section 2 we give two definitions of the capacity of the transducer and show that they are equivalent. It is shown in Section 3 that the dimension of the set of all received sequences is equal to the capacity. Finally, in Section 4 we define and examine the ambiguity of the transducer. It will be shown that the transducer has a homogeneity property by proving that the ambiguity is almost everywhere the same.

All logarithms throughout this paper are to the base b.

2. The capacity of the transducer. Let $f^* \in \mathscr{F}_m$ and let f be the corresponding transducer of memory m. Let n be a positive integer and let $N(n) = \operatorname{crd} f_n(S^{n+m-1})$, i.e., $N(n)$ is the number of distinct output sequences of length n which correspond to the b^{n+m-1} input sequences of length $n+m-1$. Following Shannon's terminology [15] we call

$$C = \lim_{n\to\infty} n^{-1} \log N(n)$$

the capacity of the transducer.

THEOREM 2.1. *The limit C exists and satisfies* $0 \leqq C \leqq 1$.

PROOF. Let k and n be positive integers. There are $N(k)$ different ways in which a received sequence of length $k+n$ may begin, and at most $N(n)$ different ways in which it may end. Hence $N(k+n) \leqq N(k)N(n)$. Also it is obvious that $N(k) \leqq N(n)$ for $k \leqq n$.

We now follow a well-known procedure (see, for example, Feinstein [9] page 85) to show that C exists. Let $a = \inf_n n^{-1} \log N(n)$ and let $\varepsilon > 0$ be given. There exists an integer r such that $\log N(r)/r \leqq a+\varepsilon$. For any integer $n \geqq r$ define k_n by $(k_n - 1)r \leqq n < k_n r$. By the two inequalities above we have $\log N(n) \leqq \log N(k_n r) \leqq k_n \log N(r)$, and thus

$$\frac{\log N(n)}{n} \leqq \frac{k_n r}{n} \frac{\log N(r)}{r} \leqq \frac{k_n r}{(k_n - 1)r}(a+\varepsilon) = \frac{k_n}{k_n - 1}(a+\varepsilon).$$

As n approaches ∞, k_n approaches ∞ and hence $k_n/(k_n - 1)$ approaches 1.

It follows that $\limsup_n n^{-1} \log N(n) \leqq a+\varepsilon$. Since ε was arbitrary, we have $\limsup_n n^{-1} \log N(n) \leqq a$, and since $n^{-1} \log N(n) \geqq a$ for all n, we have $\liminf_n n^{-1} \log N(n) \geqq a$. Thus C exists (and is equal to a).

Clearly $C \geqq 0$ since $n^{-1} \log N(n) \geqq 0$ for all n. As for $C \leqq 1$, we note that $n^{-1} \log N(n) \leqq n^{-1} \log b^n = 1$ for all n. □

We now apply another definition of capacity, introduced by Shannon [15] for noisy channels, to our transducer.

For each $x \in (0, 1]$ define a probability measure ν_x on $\mathscr{B}$ by letting ν_x assign unit mass to the point $f(x)$. Let $\mu \in \mathscr{M}$ and, for $M \in \mathscr{C}$, where $\mathscr{C}$ is the class of Borel sets in $(0, 1] \times (0, 1]$, define

$$P(M) = \int_{(0,1]} \nu_x(\{y : (x, y) \in M\})\, \mu(dx). \tag{2.2}$$

It is easily seen that f is measurable with respect to $\mathscr{B}$, and using this fact, standard arguments show that the integrand in (2.2) is measurable. Thus the integral (2.2) is defined.

It is easily verified that $P(M) = \mu(\mathrm{proj}_x \{M \cap \text{graph of } f\})$, where proj_x denotes the projection on the x-axis, and that P is a probability measure. Also the set function λ defined, for $B \in \mathscr{B}$, by

$$\lambda(B) = P((0, 1] \times B) \tag{2.3}$$

is easily seen to be a probability measure. We note that $\lambda(B) = \mu(\{x : f(x) \in B\})$.

THEOREM 2.4. *P is stationary and $\lambda \in \mathscr{M}$.*

PROOF. It is clear that f and T commute, and thus we have

$$\{x : f(x) \in T^{-1}B\} = T^{-1}\{x : f(x) \in B\}.$$

Now for any $B \in \mathscr{B}$,

$$\begin{aligned} \lambda(T^{-1}B) &= \mu(\{x : f(x) \in T^{-1}B\}) \\ &= \mu(T^{-1}\{x : f(x) \in B\}) \\ &= \mu(\{x : f(x) \in B\}) \\ &= \lambda(B), \end{aligned}$$

where the next to the last equality follows by the stationarity of μ. Thus λ is stationary.

If $B \in \mathscr{B}$ is such that $T^{-1}B = B$, then we have

$$\begin{aligned} T^{-1}\{x : f(x) \in B\} &= \{x : f(x) \in T^{-1}B\} \\ &= \{x : f(x) \in B\}. \end{aligned}$$

It follows by the ergodicity of μ that $\lambda(B)$ (which is equal to $\mu(\{x : f(x) \in B\})$) is either zero or one. Thus λ is ergodic.

To see that P is stationary it suffices to prove that $P(T_1^{-1}(B \times C)) = P(B \times C)$ for all $B, C \in \mathscr{B}$ (Billingsley [5] page 4). We see that

$$\begin{aligned} P(T_1^{-1}(B \times C)) &= P(T^{-1}B \times T^{-1}C) \\ &= \mu(\text{proj}_x\{(T^{-1}B \times T^{-1}C) \cap \text{ graph of } f\}) \\ &= \mu(\{x : f(x) \in T^{-1}C\} \cap T^{-1}B) \\ &= \mu(T^{-1}\{x : f(x) \in C\} \cap T^{-1}B) \\ &= \mu(T^{-1}(\{x : f(x) \in C\} \cap B)) \\ &= \mu(\{x : f(x) \in C\} \cap B) \\ &= P(B \times C). \quad \square \end{aligned}$$

Now for $\mu \in \mathscr{M}$, all three measures μ, λ, and P are stationary so their entropies are defined. We let

$$R_\mu = H(\mu) + H(\lambda) - H(P) \tag{2.5}$$

and $C' = \sup_{\mu \in \mathscr{M}} R_\mu$.

We show that C' can be called the transducer capacity by proving the following theorem.

THEOREM 2.6. $C' = C$.

PROOF. For any integer $n \geqq 1$, the triple (S^{n+m-1}, S^n, f_n) forms a discrete, noiseless, memoryless channel, where we think of a transmitted sequence $x \in S^{n+m-1}$ being received as the sequence $f_n(x) \in S^n$. The capacity C_n of this channel is easily computed to be $\log N(n)$ (see Feinstein [9] for the definition of capacity of a memoryless channel). Feinstein [10] has shown that $\lim_{n\to\infty} n^{-1}C_n = C'$. But since $\lim_{n\to\infty} n^{-1}C_n = \lim_{n\to\infty} n^{-1}\log N(n) = C$, we have $C = C'$. $\square$

3. The dimension of the received set. Let $f^* \in \mathscr{F}_m$ and let f be the corresponding transducer of memory m. Let $Y = f((0, 1])$ be the range of f (Y is the collection of all possible received sequences). In this section we show that the Hausdorff dimension of Y is C, the transducer capacity.

LEMMA 3.1. *The expression* (2.5) *for the rate of transmission* R_μ *of the transducer reduces to* $R_\mu = H(\lambda)$.

PROOF. Since by (2.5), $R_\mu = H(\mu) + H(\lambda) - H(P)$, we must show that $H(\mu) - H(P) = 0$. We will make use of the following two forms of the same inequality: if p, q, and r are positive real numbers then

$$(p+q)\log(p+q) \geqq p\log p + q\log q \tag{3.2}$$

and

$$(p+q)\log(p+q) + r\log r \leqq (p+q+r)\log(p+q+r). \tag{3.3}$$

The basic inequality follows from the monotonicity of the logarithm.

Now

$$(3.4)\quad \begin{aligned} H_n(P) &= -\sum P([x_1,\cdots,x_n,y_1,\cdots,y_n])\log P([x_1,\cdots,x_n,y_1,\cdots,y_n]) \\ &= -\sum \mu([x_1,\cdots,x_n]\cap\{x: f(x)\in[y_1,\cdots,y_n]\}) \\ &\qquad\qquad \cdot\log\mu([x_1,\cdots,x_n]\cap\{x: f(x)\in[y_1,\cdots,y_n]\}). \end{aligned}$$

The set $[x_1,\cdots,x_n]\cap\{x:f(x)\in[y_1,\cdots,y_n]\}$ is either empty or is the disjoint union of cylinders (b-adic intervals) of length $b^{-(n+m-1)}$. Then by using (3.2) and (3.3) on the summation (3.4) it is seen that $H_n(\mu)\leqq H_n(P)\leqq H_{n+m-1}(\mu)$. Thus

$$\begin{aligned} H(\mu) &= \lim_{n\to\infty}\frac{H_n(\mu)}{n}\leqq\lim_{n\to\infty}\frac{H_n(P)}{n} = H(P) \\ &\leqq \lim_{n\to\infty}\frac{H_{n+m-1}(\mu)}{n} = \lim_{n\to\infty}\frac{H_{n+m-1}(\mu)}{n+m-1} = H(\mu). \end{aligned}$$

Hence $H(\mu)=H(P)$. □

THEOREM 3.5. $\dim Y = C$.

PROOF. Clearly, for each positive integer n, $N(n)$ b-adic intervals of length b^{-n} will cover Y. Let $\rho>0$ and $\varepsilon>0$ be given, and choose a positive integer k such that $b^{-k}<\rho$ and $C+\varepsilon>k^{-1}\log N(k)$. Then

$$\begin{aligned} L_{C+\varepsilon}(Y,\rho) &\leqq N(k)b^{-k(C+\varepsilon)} \\ &< N(k)b^{-\log N(k)} = 1. \end{aligned}$$

Since ρ was arbitrary, it follows that $L_{C+\varepsilon}(Y)\leqq 1$, and thus $\dim Y\leqq C+\varepsilon$. Since ε was arbitrary, we have $\dim Y\leqq C$.

We now show that $\dim Y\geqq C$. Let $\varepsilon>0$ be given and choose $\mu\in\mathcal{M}$ so that $R_\mu>C'-\varepsilon$. Then by Theorem 2.6 and Lemma 3.1 we have $H(\lambda)>C-\varepsilon$. Now $\lambda\in\mathcal{M}$ by Theorem 2.4 and thus the Shannon–McMillan–Breiman theorem (see Billingsley [5]) shows that

$$(3.6)\qquad \lim\nolimits_{n\to\infty} -1/n\log\lambda([b_1(y),\cdots,b_n(y)]) = H(\lambda)\quad \text{a.e. } [\lambda].$$

Let M be the set of y's for which (3.6) holds. Then $M\cap Y\subset M$ and by a general theorem due to Billingsley [5] page 141, we have $\dim M\cap Y = H(\lambda)\dim_\lambda M\cap Y$. Now $\lambda(M)=1$ and $\lambda(Y)=\mu(\{x:f(x)\in Y\})=\mu((0,1])=1$ so $\lambda(M\cap Y)=1$. Thus $\dim_\lambda M\cap Y=1$, so $\dim M\cap Y = H(\lambda)>C-\varepsilon$. Since ε was arbitrary, we have $\dim M\cap Y\geqq C$, and since $M\cap Y\subset Y$, we have $\dim Y\geqq\dim M\cap Y\geqq C$. □

We remark that $\dim M\cap Y = H(\lambda)$ can also be shown by using a theorem of Dym [6], Theorem 2. Dym provides a direct proof not involving Billingsley's general theorem.

We also note that both definitions of capacity given in Section 2 were used in proving Theorem 3.5. The C definition was used in showing $\dim Y\leqq C$ and the C' definition was used in showing $\dim Y\geqq C$.

4. The ambiguity of the transducer. Let $f^* \in \mathscr{F}_m$, $m \geqq 1$, and let f be the corresponding transducer of memory m. Let $y \in (0, 1]$ and for $n \geqq 1$, let

$$M_n(y) = \operatorname{crd} f_n^{-1}(b_1(y) \cdots b_n(y)),$$

i.e., $M_n(y)$ is the number of input sequences of length $n+m-1$ which map into the output sequence $b_1(y) \cdots b_n(y)$. In this section we examine the quantity $\lim_{n \to \infty} n^{-1} \log M_n(y)$, called the ambiguity of the transducer at the point y, and we examine the dimension of the set $M(y) = f^{-1}(y) = \{x : f(x) = y\}$. We will call $M(y)$ the ambiguity set of y.

When y is a b-adic point, the set $M(y)$ consists of two disjoint parts, namely,

(4.1) $A = \{x : f^*(b_i(x) \cdots b_{i+m-1}(x))$ is the ith digit of the nonterminating expansion of y, for all $i \geqq 1\}$ and

$B = \{x : f^*(b_i(x) \cdots b_{i+m-1}(x))$ is the ith digit of the terminating expansion of y, for all $i \geqq 1\}$.

In this case we wish to consider only the set A so, for y a b-adic point, we redefine $M(y)$ to be the set A.

For the moment we restrict our attention to the case $m \geqq 2$. We begin by defining a set of matrices and stating a theorem which allow computation of $M_n(y)$. These definitions and the theorem appear in a report by Hedlund [12].[2]

Let r and s be integers satisfying $0 \leqq r \leqq b^{m-1}-1$, $0 \leqq s \leqq b^{m-1}-1$, and let $r_1 r_2 \cdots r_{m-1}$ and $s_1 s_2 \cdots s_{m-1}$ be the b-adic representations of r and s respectively, i.e., $r = r_1 b^{m-2} + \cdots + r_{m-2} b + r_{m-1}$ and $s = s_1 b^{m-2} + \cdots + s_{m-2} b + s_{m-1}$. A sequence $x = x_1 x_2 \cdots x_k \in S^k$ is said to begin with r and end with s provided the initial $(m-1)$-sequence of x is $r_1 \cdots r_{m-1}$ and the terminal $(m-1)$-sequence of x is $s_1 \cdots s_{m-1}$.

Let $A^i = (a^i_{rs})$, $0 \leqq i \leqq b-1$, be the square matrices of order b^{m-1} defined as follows. For $0 \leqq r, s \leqq b^{m-1}-1$, a^i_{rs} is the number of members of $f_1^{-1}(i) = (f^*)^{-1}(i)$ which begin with r and end with s.

THEOREM 4.2. *Let $y = y_1 \cdots y_k \in S^k$ and let $W(y) = W(y_1 \cdots y_k) = A^{y_1} A^{y_2} \cdots A^{y_k}$. Then w_{rs}, the r, s entry of $W(y)$, $0 \leqq r$, $s \leqq b^{m-1}-1$, is the number of members of $f_k^{-1}(y_1 \cdots y_k)$ which begin with r and end with s.*

PROOF. The proof is by induction on the length k of the sequences. See Hedlund [12]. □

By the weight of a matrix A, denoted by $|A|$, we will mean the sum of all the entries of A.

COROLLARY 4.3 *If $y \in (0, 1]$, then $M_n(y) = |W(b_1(y) \cdots b_n(y))|$.*

We now show that the ambiguity exists and has the same value for almost all y. We begin with a lemma.

---

[2] The author is grateful to Professor G. A. Hedlund of Yale University for sending a copy of his report.

LEMMA 4.4. *Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $n \times n$ matrices with nonnegative entries. Then $|AB| \leqq |A|\,|B|$.*

PROOF. We see that

$$|AB| = \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ik} b_{kj} \tag{4.5}$$ and

$$|A|\,|B| = (\textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij})(\sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}). \tag{4.6}$$

It is clear that each term in the summation (4.5) appears on the right-hand side of (4.6), and since all terms are nonnegative we have $|AB| \leqq |A|\,|B|$. □

THEOREM 4.7. *Let f be a transducer of memory $m \geqq 2$, let $\mu \in \mathscr{M}$, and let λ be defined by* (2.3). *Then $\lim_{n\to\infty} n^{-1} \log M_n(y)$ exists and has the same value for almost all $y[\lambda]$.*

PROOF. We define a stochastic process $Y_1, Y_2, \cdots$ with domain $(0, 1]$ and values in the set of $b^{m-1} \times b^{m-1}$ matrices as follows. For $y \in (0, 1]$ define $Y_n(y) = A_T^{b_n(y)}$, where the subscript T denotes the transpose operation. By Theorem 2.4 we have $\lambda \in \mathscr{M}$. Then defining the norm of a matrix $A = (a_{ij})$ by $\|A\| = \max_i \sum_j |a_{ij}|$, we have by a theorem of Furstenberg and Kesten [11], Theorem 2 that

$$\lim_{n\to\infty} n^{-1} \log \|Y_n(y) \cdots Y_1(y)\| \tag{4.8}$$

exists and has the same value for almost all $y[\lambda]$. We may rewrite (4.8) as

$$\lim_{n\to\infty} n^{-1} \log \|(A^{b_1(y)} \cdots A^{b_n(y)})_T\|$$ or

$$\lim_{n\to\infty} n^{-1} \log \|W_T(b_1(y) \cdots b_n(y))\|. \tag{4.9}$$

The only property of the norm used by Furstenberg and Kesten in their proofs is that $\|AB\| \leqq \|A\|\,\|B\|$ for any two matrices A and B. Since the matrices A^i are all nonnegative, we have the same inequality when the norm is replaced by the weight (Lemma 4.4). Hence we may use the weight instead of the norm in (4.9). Then, noting that

$$|W_T(b_1(y) \cdots b_n(y))| = |W(b_1(y) \cdots b_n(y))|,$$

and using Corollary 4.3, we have $\lim_{n\to\infty} n^{-1} \log M_n(y)$ exists and has the same value for almost all $y\ [\lambda]$. □

If we let D be the limit in Theorem 4.7, we have the following theorem.

THEOREM 4.10. *Let f be a transducer of memory $m \geqq 2$, let $\mu \in \mathscr{M}$, and let λ be defined by* (2.3). *Then $\dim M(y) \leqq D$ for almost all $y[\lambda]$.*

PROOF. Let $E = \{y: \lim_{n\to\infty} n^{-1} \log M_n(y) = D\}$. By Theorem 4.7, $\lambda(E) = 1$. Let $y \in E$. Clearly, for each positive integer n, $M_n(y)$ b-adic intervals of length $b^{-(n+m-1)}$ will cover $M(y)$. (It is here that we use the fact that for y a b-adic point, $M(y)$ is defined by the set (4.1). See Example 4.12). Now let $\rho > 0$ and $\varepsilon > 0$ be given, and choose a positive integer k such that $b^{-(k+m-1)} < \rho$ and $D+\varepsilon > k^{-1} \log M_k(y)$. Then $L_{D+\varepsilon}(M(y), \rho) \leqq M_k(y) b^{-(k+m-1)(D+\varepsilon)} < M_k(y)\, b^{-\log M_k(y)} b^{-(m-1)(D+\varepsilon)} \leqq 1$.

Since ρ was arbitrary, it follows that $L_{D+\varepsilon}(M(y)) \leqq 1$, and thus $\dim M(y) \leqq D+\varepsilon$. Since ε was arbitrary, we have

$$\dim M(y) \leqq D. \tag{4.11}$$

Since (4.11) holds for all $y \in E$, the proof is completed. □

EXAMPLE 4.12. We demonstrate in this example why for y a b-adic point, $M(y)$ is defined by the set (4.1). Let $b = 2$, $m = 2$, and let $f^* \in \mathscr{F}_2$ be defined by $f^*(00) = f^*(11) = 0$ and $f^*(01) = f^*(10) = 1$. Let y be the dyadic point $.0111\cdots$, and note that $.1000\cdots$ is also equal to y. Now $M_1(y) = \operatorname{crd}(f^*)^{-1}(0) = \operatorname{crd}\{00, 11\} = 2$. The two intervals $(0, \frac{1}{4}]$ and $(\frac{3}{4}, 1]$ represented by the two sequences 00 and 11 do not cover $f^{-1}(y)$ since the point $x = .0111\cdots$ belongs to $f^{-1}(y)$ but does not belong to either $(0, \frac{1}{4}]$ or $(\frac{3}{4}, 1]$. However, if $M(y)$ is defined to be the set (4.1) and not the set $f^{-1}(y)$, then it is clear that for all $n \geqq 1$, the $M_n(y)$ intervals represented by the set of sequences $f_n^{-1}(b_1(y)\cdots b_n(y))$ will cover $M(y)$, a property which is essential for the proof of Theorem 4.10.

The discussion so far in this section has been for transducers of memory $m \geqq 2$, since the matrices A^i are not defined for $m = 1$. We now examine the case $m = 1$ separately.

Let $f^* \in \mathscr{F}_1$ with f the corresponding transducer of memory 1. We represent f^* by a $b \times b$ matrix $D = (d_{ij})$, $0 \leqq i, j \leqq b-1$, as follows. The entry d_{ij} is one if $f^*(i) = j$ and is zero if $f^*(i) \neq j$. Thus each row of D contains a single entry of one and all other entries of the row are zero. Define $l_j = \sum_{i=0}^{b-1} d_{ij}$, $0 \leqq j \leqq b-1$; l_j is the number of elements of S which map to j under f^*.

Let $\mu \in \mathscr{M}$ and define $p_i = \mu([i])$, $0 \leqq i \leqq b-1$ and $q_i = \sum_{j=0}^{b-1} d_{ji} p_j$, $0 \leqq i \leqq b-1$. We note that λ defined by (2.3) is such that $\lambda([i]) = q_i$, $0 \leqq i \leqq b-1$, since $\lambda([i]) = \mu(\{x : f(x) \in [i]\}) = \sum' \mu([j]) = q_i$, where $\sum'$ denotes the summation taken over those j for which $f^*(j) = i$.

We now prove the equivalent of Theorem 4.7 for the case $m = 1$. The proof is similar to that of Theorem 4.7, using l_i instead of A^i and the ergodic theorem instead of the Furstenberg and Kesten theorem.

THEOREM 4.13. *Let f be a transducer of memory* 1, *let $\mu \in \mathscr{M}$, and let λ be defined by* (2.3). *Then the ambiguity at the point y, $\lim_{n\to\infty} n^{-1} \log M_n(y)$, exists and has the same value* (*namely*, $q_0 \log l_0 + \cdots + q_{b-1} \log l_{b-1}$) *for almost all y* [λ].

PROOF. Let $h_n^i(y)$ be the number of occurrences of i, $0 \leqq i \leqq b-1$, among the first n digits of the nonterminating base b expansion of y. By Theorem 2.4, $\lambda \in \mathscr{M}$ and we may apply the pointwise ergodic theorem (see Billingsley [5] page 13) to conclude that for each i, $0 \leqq i \leqq b-1$,

$$\lim_{n\to\infty} n^{-1} h_n^i(y) = q_i \tag{4.14}$$

for almost all $y[\lambda]$. Let D_i, $0 \leqq i \leqq b-1$, be the set of y's for which (4.14) holds. Then for $y \in F = \bigcap_{i=0}^{b-1} D_i$, $\lim_{n\to\infty} n^{-1} h_n^i(y) = q_i$ for all i, and $\lambda(F) = 1$ since $\lambda(D_i) = 1, 0 \leqq i \leqq b-1$.

Now since for each $n \geqq 1$ we obviously have $M_n(y) = l_0^{h_n^0(y)} \cdots l_{b-1}^{h_n^{b-1}(y)}$, then

$$\frac{\log M_n(y)}{n} = \frac{h_n^0(y)}{n}\log l_0 + \cdots + \frac{h_n^{b-1}(y)}{n}\log l_{b-1},$$

where we take 0^0 to be one and 0 log 0 to be zero. Hence if $y \in F$, we have $\lim_{n\to\infty} n^{-1}\log M_n(y)$ exists and is equal to $q_0 \log l_0 + \cdots + q_{b-1}\log l_{b-1}$. Since $\lambda(F) = 1$, the proof is completed. □

We now prove the equivalent of Theorem 4.10 for the case $m = 1$. In this case we are able to obtain an equality for the dimension of the ambiguity set rather than just an upper bound. We begin with a lemma.

LEMMA 4.15. *Let $y \in (0, 1]$ and let $E_n = \bigcup'[x_1, \cdots, x_n]$ where $\bigcup'$ denotes the union over those $x_1, \cdots, x_n$ such that $f_n(x_1 \cdots x_n) = b_1(y) \cdots b_n(y)$ (there are clearly $M_n(y)$ such sequences $x_1, \cdots, x_n$). Then*

$$\bigcap_{n=1}^{\infty} E_n = M(y).$$

PROOF. Let $x \in M(y)$. Then $f^*(b(x_i))$ equals the ith digit of the nonterminating expansion of y, i.e., $f^*(b(x_i)) = b_i(y)$. Thus $f_n(b_1(x) \cdots b_n(x)) = b_1(y) \cdots b_n(y)$ for all n. Hence $x \in E_n$ so $M(y) \subset E_n$ for all n. Therefore

$$M(y) \subset \bigcap_{n=1}^{\infty} E_n. \tag{4.16}$$

Let $x \in \bigcap_{n=1}^{\infty} E_n$. Then $f_n(b_1(x) \cdots b_n(x)) = b_1(y) \cdots b_n(y)$ for all n so $f^*(b_n(x)) = b_n(y)$ for all n. Thus $x \in M(y)$ so

$$\bigcap_{n=1}^{\infty} E_n \subset M(y). \tag{4.17}$$

Inclusions (4.16) and (4.17) give the desired result. □

THEOREM 4.18. *Let f be a transducer of memory 1, let $\mu \in \mathcal{M}$, and let λ be defined by (2.3). Then*

$$\dim M(y) = \textstyle\sum_{i=0}^{b-1} q_i \log l_i$$

for almost all $y[\lambda]$.

PROOF. Let $y \in (0, 1]$ and let $y_i = b_i(y)$ for all $i \geqq 1$. For each integer $k \geqq 1$ and each sequence $x_1 \cdots x_k \in S^k$, define a function p_k on S^k by

$$\begin{aligned} p_k(x_1 \cdots x_k) &= 1/M_k(y) = 1/l_{y_1} \cdots l_{y_k} \quad &&\text{if } f_k(x_1 \cdots x_k) = y_1 \cdots y_k; \\ &= 0 &&\text{otherwise.} \end{aligned}$$

It is clear that

$$p_k(x_1 \cdots x_k) \geqq 0 \qquad \text{for all } k. \tag{4.19}$$

Also we have

$$\textstyle\sum_{i \in S} p_1(i) = l_{y_1}/l_{y_1} = 1. \tag{4.20}$$

We show next that

$$\sum_{i\in S} p_{k+1}(x_1\cdots x_k i) = p_k(x_1\cdots x_k). \tag{4.21}$$

If $p_k(x_1\cdots x_k) = 0$ then $f_k(x_1\cdots x_k) \neq y_1\cdots y_k$ so $f_{k+1}(x_1\cdots x_k i) \neq y_1\cdots y_k y_{k+1}$ for all $i\in S$. Hence $p_{k+1}(x_1\cdots x_k i) = 0$ for all $i\in S$ so (4.21) is true in this case. If $p_k(x_1\cdots x_k) = 1/l_{y_1}\cdots l_{y_n}$ then $f_{k+1}(x_1\cdots x_k i) = y_1\cdots y_k y_{k+1}$ for the $l_{y_{k+1}}$ values of i for which $f^*(i) = y_{k+1}$. Hence $p_{k+1}(x_1\cdots x_k i)$ is $1/l_{y_1}\cdots l_{y_k} l_{y_{k+1}}$ for those $l_{y_{k+1}}$ values of i and is zero for the remaining values of i. Therefore

$$\sum_{i\in S} p_{k+1}(x_1\cdots x_k i) = l_{y_{k+1}}/l_{y_1}\cdots l_{y_{k+1}} = p_k(x_1\cdots x_k)$$

so (4.21) is also true in this case.

Finally we see that for any sequence $x_1,\cdots,x_k$ of elements of S, we have

$$\lim_{n\to\infty} p_{k+n}(x_1\cdots x_k \underbrace{0\cdots 0}_{n\ 0\text{'s}}) = 0. \tag{4.22}$$

It follows from (4.19), (4.20), (4.21) and (4.22) that there exists a probability measure ν_y on the unit interval such that

$$\nu_y([x_1,\cdots,x_k]) = p_k(x_1\cdots x_k),$$

where $x_1\cdots x_k \in S^k$ (see Billingsley [5] page 35).

Now let F be the set of y's for which Theorem 4.13 holds, and let $y\in F$. If $x\in M(y)$ and if we set $x_i = b_i(x)$ and $y_i = b_i(y)$, then

$$\lim_{n\to\infty} -\frac{\log \nu_y([x_1,\cdots,x_n])}{n} = \lim_{n\to\infty} -\frac{\log(1/l_{y_1}\cdots l_{y_n})}{n}$$
$$= \lim_{n\to\infty}\frac{\log l_{y_1}\cdots l_{y_n}}{n} = \lim_{n\to\infty}\frac{\log M_n(y)}{n} = \sum_{i=0}^{b-1} q_i \log l_i.$$

The next to the last equality follows from the fact that we obviously have $M_n(y) = l_{y_1}\cdots l_{y_n}$, and the last equality follows from Theorem 4.13. By a theorem of Billingsley [5] page 141, we now have

$$\dim M(y) = \dim_{\nu_y} M(y) \sum_{i=0}^{b-1} q_i \log l_i \tag{4.23}$$

for all $y\in F$.

If E_n is defined as in Lemma 4.15, it is seen that $\nu_y(E_n) = 1$. Since $\bigcap_{n=1}^{\infty} E_n = M(y)$ by Lemma 4.15 and since $\{E_n\}$ is a decreasing sequence, we have $\nu_y(M(y)) = 1$. Hence $\dim_{\nu_y} M(y) = 1$ and thus from (4.23) we have

$$\dim M(y) = \sum_{i=0}^{b-1} q_i \log l_i$$

for all $y\in F$. Since $\lambda(F) = 1$ by Theorem 4.13, the proof is completed. □

For the case $m = 1$ we were able to calculate the dimension of the ambiguity set $M(y)$ (Theorem 4.18), whereas in the case $m \geqq 2$, we were able to obtain only an upper bound on $\dim M(y)$ (Theorem 4.10). Conjecture is that for the case $m \geqq 2$, $\dim M(y)$ is actually equal to this upper bound. However, the method of proof of Theorem 4.18 cannot be used to prove this conjecture, since the consistency condition (4.21) may not hold for $m \geqq 2$. The following example demonstrates this fact.

EXAMPLE 4.24. Let $b = 2$, $m = 2$, and let $f^* \in \mathscr{F}_2$ be defined by $f^*(00) = 0$ and $f^*(01) = f^*(10) = f^*(11) = 1$. The function f_2 is then as follows:

Domain Value	Functional Value
0 0 0	0 0
0 0 1	0 1
0 1 0	1 1
0 1 1	1 1
1 0 0	1 0
1 0 1	1 1
1 1 0	1 1
1 1 1	1 1

Let $y \in (\frac{3}{4}, 1]$; we see that $M_1(y) = \text{crd}\{01, 10, 11\} = 3$ and $M_2(y) = \text{crd}\{010, 011, 101, 110, 111\} = 5$. If we define (as in the proof of Theorem 4.18) $p_2(00) = 0$, $p_2(01) = p_2(10) = p_2(11) = \frac{1}{3}$ and $p_3(000) = p_3(001) = p_3(100) = 0$, $p(010) = p(011) = p(101) = p(110) = p(111) = \frac{1}{5}$, then it is clear that the consistency condition (4.21) does not hold. Thus for any $y \in (\frac{3}{4}, 1]$, we cannot define a measure v_y as we did in Theorem 4.18.

REFERENCES

[1] BESICOVITCH, A. S. (1935). On the sum of digits of real numbers represented in the dyadic system. *Math. Ann.* **110** 321–330.

[2] BILLINGSLEY, P. (1960). Hausdorff dimension in probability theory. *Illinois J. Math.* **4** 187–209.

[3] BILLINGSLEY, P. (1961a). Hausdorff dimension in probability theory II. *Illinois J. Math.* **5** 291–298.

[4] BILLINGSLEY, P. (1961b). On the coding theorem for the noiseless channel. *Ann. Math. Statist.* **32** 594–601.

[5] BILLINGSLEY, P. (1965). *Ergodic Theory and Information.* Wiley, New York.

[6] DYM, H. (1968). On a class of monotone functions generated by ergodic sequences. *Amer. Math. Monthly* **75** 594–601.

[7] EGGLESTON, H. G. (1949). The fractional dimension of a set defined by decimal properties. *Quart. J. Math. Oxford Ser.* (2) **20** 31–36.

[8] EGGLESTON, H. G. (1952). Sets of fractional dimensions which occur in some problems of number theory. *Proc. London Math. Soc.* **54** 42–93.

[9] FEINSTEIN, A. (1958). *Foundations of Information Theory.* McGraw-Hill, New York.

[10] FEINSTEIN, A. (1959). On the coding theorem and its converse for finite-memory channels. *Information and Control* **2** 25–44.

[11] FURSTENBERG, H. and KESTEN, H. (1960). Products of random matrices. *Ann. Math. Statist.* **31** 457–469.

[12] HEDLUND, G. A. (1961). Mappings on sequence spaces. Communications Research Division Technical Report No. 1, von Neumann Hall.

[13] KHINCHIN, A. I. (1957). *Mathematical Foundations of Information Theory.* Dover, New York.

[14] KINNEY, J. R. (1958). Singular functions associated with Markov chains. *Proc. Amer. Math. Soc.* **9** 603–608.

[15] SHANNON, C. E. and WEAVER, W. (1964). *The Mathematical Theory of Communication.* Univ. of Illinois. (Reprinted from *Bell System Tech. J.* **27** (1948) 379–423, 623–656.)

[16] SMORODINSKY, M. (1968). The capacity of a general noiseless channel and its connection with Hausdorff dimension. *Proc. Amer. Math. Soc.* **19** 1247–1254.