

## Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems

Aviad Cohen and Itamar Procaccia

*Department of Chemical Physics, Weizmann Institute of Science, Rehovot 76100, Israel*

(Received 27 June 1984)

The extraction of the Kolmogorov (metric) entropy from an experimental time signal is discussed. Theoretically we stress the concept of generators and that the existence of an expansive constant guarantees that a finite-time series would be sufficient for the calculation of the metric entropy. On the basis of the theory we attempt to propose optimal algorithms which are tested on a number of examples. The approach is applicable to both dissipative and conservative dynamical systems.

### I. INTRODUCTION

The experimental investigation of the transition to chaos and its development in physical systems entails probing the time dependence of a variable that can be easily measured in the laboratory. Taking fluid systems as an example, some investigators prefer to measure velocity fields, whereas others choose temperature or temperature gradients. Evidently, any macroscopic system can be studied by probing a variety of variables. It becomes apparent, therefore, that in order to have a valid description and characterization of the dynamical aspects of physical systems it is mandatory to use quantities that are invariant to the choice of the measured variable.<sup>1,2</sup> This is one of the reasons for the intense recent interest in the measurement of the (fractal) dimensions of strange attractors that arise in chaotic dissipative systems.<sup>3-7</sup> These dimensions are invariant to a smooth change of coordinates and therefore serve usefully to characterize the chaotic motion. The subject of this paper is another set of invariants, i.e., the entropies, and in particular the Kolmogorov (or metric) entropy  $K$ .<sup>8,9</sup> The extraction of the Kolmogorov entropy from an experimental signal is of particular interest since this quantity quantifies "how chaotic" a signal is. A regular trajectory has  $K=0$ . A purely random signal has  $K=\infty$ . A deterministic chaotic signal is characterized by a finite  $K$ , and  $K$  is related to the "predictability time"  $T$  by  $K \sim T^{-1}$  where  $T$  is the average time into the future for which knowledge of the state of the system can be used to predict its evolution.<sup>10</sup> Another relation of interest of  $K$  is to the sum of positive Lyapunov exponents<sup>11,12</sup> (which measure the instability of nearby trajectories).

Intuitively speaking, the entropies arise when one considers measurements of finite accuracy. Suppose that one can measure the state of the system (i.e., position in phase space) with accuracy  $\epsilon$  and time with accuracy  $\tau$ . Then a "trajectory" is given as a time sequence of boxes  $\{i_1, i_2, \dots, i_d\}$  which means that the system is in the box  $i_1$  (of size  $\epsilon^F$  where  $F$  is the number of degrees of freedom) at time  $\tau$ , in box  $i_2$  at time  $2\tau$ , etc. Since the state of the system is given only to finite accuracy, there is a finite probability of finding the system in various boxes  $i_2$

at time  $2\tau$ , various boxes  $i_3$  at time  $3\tau$ , etc. Considering the joint probability  $P(i_1, i_2, \dots, i_d)$  we suggested previously to consider the  $q$ -order entropies<sup>13</sup>

$$K_q = - \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{d\tau} \frac{1}{q-1} \times \log_2 \sum_{i_1, \dots, i_d} P^q(i_1, \dots, i_d) \quad (1.1)$$

and argued that  $\lim_{q \rightarrow 0} K_q$  yields the so-called topological entropy, whereas  $\lim_{q \rightarrow 1^+} K_q$  is the metric or Kolmogorov entropy  $K$ ,

$$K = - \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{d\tau} \sum_{i_1, \dots, i_d} P(i_1, \dots, i_d) \times \log_2 P(i_1, \dots, i_d). \quad (1.2)$$

(A more precise definition, which entails a supremum over all partitions, not necessarily uniform as considered here, is given in Sec. II.) Since one can show that  $K_q \geq K_{q'}$  for any  $q' > q$ , we proposed in previous work algorithms for the estimate of  $K$  by the quantity  $K_2$ .<sup>13</sup> In this paper algorithms for the extraction of  $K$  itself from a time signal will be proposed and tested on model systems, both dissipative and conservative.

The development of optimal algorithms calls for a review of the theoretical background of the Kolmogorov entropy. Such a review is offered in Sec. II where emphasis is put on the concept of a generator, which is a partition of phase space whose entropy with respect to the dynamical system equals the metric entropy and whose further refinement does not increase the computed value of the entropy. The importance of the location of generators is clear. In any experimental application the number of points in a time series is limited (and often quite modest). The finer the needed partition is, the longer is the sequence that has to be used for the estimate of  $K$ . Therefore, the existence of a generator guarantees that a good estimate of  $K$  can be obtained with a finite sequence.

From the theoretical point of view we stress that the existence of generators is guaranteed by the existence of an “expansive constant” for the dynamics. From the practical point of view we propose how to locate generators via the application of the algorithm.

In Sec. III we present examples, both conservative and dissipative, with the aim of clarifying the usefulness and pitfalls of the proposed algorithms. Readers who are mostly interested in the application might wish to skip the theoretical background and go directly to Sec. III where a summary of the procedure is given. Section IV offers concluding remarks.

## II. THE THEORETICAL BACKGROUND

### A. Entropy and generators

In the following we consider discrete dynamical systems. In the case of flows we refer to their Poincaré or one-time maps. Thus we assume having a map  $f$ , operating on a manifold  $M$ , equipped with an invariant probability measure  $\mu$ . A *partition* of  $M$  is a collection of measurable disjoint sets whose union is  $M$ . Given two finite partitions

$$\Phi = \{\phi_1, \dots, \phi_m\}, \quad \Psi = \{\psi_1, \dots, \psi_n\}, \quad (2.1)$$

we denote by  $\Phi \vee \Psi$  the partition

$$\{\Phi_1 \cap \psi_1, \dots, \phi_m \cap \psi_n\}. \quad (2.2)$$

The entropy of the partition  $\Phi$  is

$$K(\Phi) = - \sum_i \mu(\phi_i) \log_2[\mu(\phi_i)]. \quad (2.3)$$

For experimental applications  $\mu(\phi_i)$  can be estimated in principle by measuring a long-time series of  $N$  points and counting the number of points  $N_i$  which fall in  $\phi_i$ . Then  $\mu(\phi_i) \simeq N_i/N$ . In practice such “box-counting” algorithms are not efficient and the algorithms discussed below circumvent this step. Consider now the partitions that arise by operating the map  $f^{-1}$  on  $\phi$  and the sequence

$$\frac{1}{n} K \left[ \bigvee_{i=0}^{n-1} f^{-i}(\Phi) \right]. \quad (2.4)$$

It can be shown<sup>14</sup> that this sequence is decreasing and converges to a limit. The limit

$$K(f, \Phi) = \lim_{n \rightarrow \infty} \frac{1}{n} K \left[ \bigvee_{i=0}^{n-1} f^{-i}(\Phi) \right] \quad (2.5)$$

is called the *entropy of  $\Phi$  with respect to  $f$* . The *metric entropy of  $f$*  is<sup>8,9,14</sup>

$$K(f) = \sup_{\Phi} K(f, \Phi), \quad (2.6)$$

clearly, this definition which entails a supremum on all partitions seems discouraging for experimental applications. It becomes therefore essential to be able to eliminate this requirement of considering all partitions. One can do so with the help of the concept of *generators*.

A precise mathematical definition of a generator can be

found in Ref. 14. Intuitively speaking, a finite partition  $\Pi$  is a generator if the partition

$$\bigvee_{i=-n}^n f^{-i}[\Pi]$$

becomes infinitely fine as  $n \rightarrow \infty$ .

It can be shown that finite generators exist in most interesting cases.<sup>14</sup> The importance of a generator lies in the fact that<sup>15</sup>

$$K[f, \Pi] = K[f]. \quad (2.7)$$

That is, the entropy of the generator with respect to  $f$  is equal to the metric entropy.

Any refinement of a generator is also a generator. Since in many cases it seems more than plausible that a generator can be refined into a uniform partition with  $\epsilon$  small enough we may expect intuitively the existence of a uniform partition which is a generator.

The idea can be made more exact in the following way: define the map  $f$  to be *expansive* if there exists a  $\delta$  such that if  $X \neq Y$  then

$$d(f^n(X), f^n(Y)) > \delta \quad (2.8)$$

for some integer  $n$ , where  $d(\cdot, \cdot)$  is a distance between two points defined with some convenient norm. The quantity  $\delta$  is said to be an expansive constant for  $f$ . We observe that expansiveness includes contractions. Axiom-A systems are examples of expansive maps.<sup>15</sup> For an expansive homeomorphism defined on a compact space we have the following theorem.<sup>14</sup>

*Theorem.* If  $\delta$  is an expansive constant for  $f$  any partition with a diameter less than  $\delta$  is a generator.

Intuitively speaking, if we have such a partition, two trajectories emanating from distinct points must ultimately separate, ending up in different boxes. Therefore, this partition reflects the divergence properties of  $f$ .

The numerical extraction of  $K$  from either an experimental or computer-generated sequence can be crucially facilitated by the knowledge of the quantity  $\delta$ . The existence of reasonably large  $\delta$  guarantees that a realistic number of points in a time series is all that is needed for an accurate estimate of  $K$ . To further clarify this point we offer in Appendix A an example of the calculation of an expansive constant  $\delta$  for the baker transformation. Later numerical tests (Sec. III) corroborate the assertion that uniform partitions with  $\epsilon < \delta$  yield  $K$  as their entropy.

### B. Convergence rates

Given are two finite partitions of  $M$ :

$$\Phi = \{\phi_i\}_{i=1}^m, \quad \Psi = \{\psi_k\}_{k=1}^n.$$

If every element of  $\Phi$  is a subset of an element of  $\Psi$  then  $\Phi$  is a *refinement* of  $\Psi$  and we write

$$\Psi \leq \Phi.$$

The conditional entropy of  $\Phi$  given  $\Psi$  is defined as

$$K(\Phi/\Psi) = - \sum_{k=1}^n \mu(\psi_k) \sum_{i=1}^M \frac{\mu(\psi_k \cap \phi_i)}{\mu(\psi_k)} \log_2 \frac{\mu(\psi_k \cap \phi_i)}{\mu(\psi_k)} \tag{2.9}$$

whereas the quantities  $k_n$  are defined by

$$k_n(f, \Phi) = K \left[ \bigvee_{i=0}^{n-1} f^{-i}(\Phi) \right]. \tag{2.10}$$

It can be shown by induction, using the properties listed in Appendix B, that

$$k_n(f, \Phi) = K(\Phi) + \sum_{i=1}^{n-1} K \left[ \Phi / \bigvee_{k=1}^i f^{-1}(\Phi) \right]. \tag{2.11}$$

Let us define now the quantities:

$$\begin{aligned} \delta_n(f, \Phi) &= k_n(f, \Phi) - k_{n-1}(f, \Phi) \\ &= K \left[ \Phi / \bigvee_{i=1}^{n-1} f^{-1}(\Phi) \right], \quad n > 2. \end{aligned} \tag{2.12}$$

The quantity  $\delta_n$  can be assigned the following physical interpretation.  $k_n$  is the information obtained by performing  $n$  consecutive experiments associated with the partition  $\Phi$ . Therefore,  $\delta_n$  is the *information gain* of the  $n$ th experiment given the results of the first  $n-1$  experiments.

Adding  $\delta_1 = K(\Phi)$  to the sequence  $\delta_n$ ,  $n > 2$  one can show that  $\delta_1 \geq \delta_2$  and that for  $m, n \geq 2$  we have  $\delta_m \geq \delta_n$  for any  $m \leq n$ . The consequence of these observations is that the sequence  $\delta_n$  is a positive nonincreasing sequence and hence converges to a limit  $L$ .

Consider now the sequence

$$B_n = k_n/n. \tag{2.13}$$

One notices that since  $B_n = n^{-1} \sum_{i=1}^n \delta_i$ ,  $B_n$  is the Cesaro sequence of the sequence  $\delta_n$ . Therefore  $B_n$  converges to the same limit  $L$ . In Appendix B we show that  $B_n$  is also a nonincreasing sequence. The main statement proved there, however, is that *the sequence  $\delta_n$  converges to  $L$  more quickly than the sequence  $B_n$ .*

Thus although the definition of the entropy of  $\Phi$  with respect to  $f$  was given in Eq. (2.5) as  $K(f, \Phi) = \lim_{n \rightarrow \infty} B_n$ , we conclude that the numerical computations of the partition entropy with respect to  $f$  should be based on the calculation of  $\delta_n$  due to its superior convergence rate. It will be shown below, however, that the sequence  $B_n$ , which converges more slowly, will prove to be useful for the location of generators.

### C. Computation of the entropy of a uniform partition

Suppose now that our  $m$ -dimensional phase space is partitioned to a uniform partition of boxes of size  $\epsilon$ , and that we are given  $N$  points in a time sequence  $\{X_n\}_{n=1}^N$ . Let  $N_i$  denote the number of points of the sequence falling within box  $i$ . The points  $X_n$  are relabeled by  $X_i^k$  where  $i$  is the index of the box to which  $X_n$  belongs and  $1 < k < N_i$  is the index of  $X_n$  in the box. Given two points  $X, Y$  in  $R^F$  we recall that the square metric  $\rho$  is given by

$$\rho(X, Y) = \max_{i=1}^F |X_i - Y_i|. \tag{2.14}$$

Given a point  $Z$  on the attractor, we define the individual correlation function of  $Z$  by

$$C_Z(\epsilon) = \frac{N_Z(\epsilon)}{N}, \tag{2.15}$$

where  $N_Z(\epsilon)$  is the number of points in the generating sequence whose square distance from  $Z$  is less than  $\epsilon$ .

To calculate the partition entropy of  $\Phi$  we recall from Sec. II A that if  $p_i$  is the invariant probability measure associated with box  $i$  then  $P_i \simeq N_i/N$  provided  $N$  is large enough. Hence

$$K(\Phi) \simeq - \sum_{i=1}^I \frac{N_i}{N} \log_2 \frac{N_i}{N}. \tag{2.16}$$

As is well known,<sup>16</sup> direct determination of  $N_i$  (box counting) is impractical in general. To bypass this difficulty we note that if  $Z_i$  denotes the center of box  $i$  then

$$\frac{N_i}{N} = C_{Z_i} \left[ \frac{\epsilon}{2} \right], \tag{2.17}$$

we make the *assumption*

$$\frac{1}{N_i} \sum_{k=1}^{N_i} \log_2 \left[ C_{X_i^k} \left[ \frac{\epsilon}{2} \right] \right] = \log_2 \left[ C_{Z_i} \left[ \frac{\epsilon}{2} \right] \right]. \tag{2.18}$$

This means that the individual correlation function of the center of the box is the geometric mean of the individual correlation functions of all the points of the sequence that fall within box  $i$ . This assumption is analogous to a mean-value theorem (but not a consequence of such). For a smoothly varying probability density it seems plausible that the values of properties at the center of the  $i$ th box will be close to their average over the box.

Using Eqs. (2.16)–(2.18) we get

$$\begin{aligned} K(\Phi) &\simeq - \sum_{i=1}^I \frac{N_i}{N} \log_2 \frac{N_i}{N} \\ &= - \frac{1}{N} \sum_{i=1}^I N_i \log_2 \left[ C_{Z_i} \left[ \frac{\epsilon}{2} \right] \right] \\ &= - \frac{1}{N} \sum_{i=1}^I \sum_{k=1}^{N_i} \log_2 \left[ C_{X_i^k} \left[ \frac{\epsilon}{2} \right] \right] \\ &= - \frac{1}{N} \sum_{n=1}^N \log_2 \left[ C_{X_n} \left[ \frac{\epsilon}{2} \right] \right]. \end{aligned} \tag{2.19}$$

Hence

$$K(\Phi) \simeq - \left\langle \log_2 \left[ C \left[ \frac{\epsilon}{2} \right] \right] \right\rangle; \tag{2.20}$$

we obtain the result that the partition entropy is minus the average  $\log_2$  of individual correlation functions over all the points of the generating sequence.

#### D. Computing the partition entropy and the location of generators

In this section we continue the discussion of the previous one. Given a point  $X_p$  of the generating sequence we define the  $d$ -order individual correlation function:

$$C_{X_p^d}(\epsilon) = \frac{N_{X_p^d}(\epsilon)}{N}, \quad (2.21)$$

where  $N_{X_p^d}(\epsilon)$  is the number of points  $X_r$  of the generating sequence satisfying

$$\rho[\langle X_p, X_{p+1}, \dots, X_{p+d-1} \rangle, \langle X_r, X_{r+1}, \dots, X_{r+d-1} \rangle] < \epsilon. \quad (2.22)$$

In analogy with Sec. III C we get

$$K \left[ \bigvee_{i=0}^{d-1} f^{-i}(\Phi) \right] \simeq - \left\langle \log_2 \left[ C^d \left[ \frac{\epsilon}{2} \right] \right] \right\rangle. \quad (2.23)$$

From (2.12) and (2.23) it follows that the sequence  $\delta_n$  is given by

$$\delta_d \simeq \left\langle \log_2 \left[ C^{d-1} \left[ \frac{\epsilon}{2} \right] \right] \right\rangle - \left\langle \log_2 \left[ C^d \left[ \frac{\epsilon}{2} \right] \right] \right\rangle. \quad (2.24)$$

Hence the partition entropy with respect to  $f$  is given by

$$K(f, \Phi) \simeq \lim_{d \rightarrow \infty} \left\{ \left\langle \log_2 \left[ C^{d-1} \left[ \frac{\epsilon}{2} \right] \right] \right\rangle - \left\langle \log_2 \left[ C^d \left[ \frac{\epsilon}{2} \right] \right] \right\rangle \right\}. \quad (2.25)$$

The sequence  $B_d$  is given by

$$B_d \simeq - \frac{1}{d} \left\langle \log_2 \left[ C^d \left[ \frac{\epsilon}{2} \right] \right] \right\rangle. \quad (2.26)$$

We note in passing that the information dimension<sup>2,17</sup> can be calculated by

$$D_1 \cong \lim_{\epsilon \rightarrow 0} \frac{\langle \log_2 [C^d(\epsilon)] \rangle}{\log_2 \epsilon}. \quad (2.27)$$

We thus see that we can get both the  $\delta_d$  and the  $B_d$  sequences. The idea is now that by plotting  $B_d$  versus  $d$  for various  $\epsilon$  we get a family of monotone decreasing curves with the higher curves belonging to the smaller values of  $\epsilon$ . If there exists a curve such that all the curves above it converge to it from above we call the corresponding partition a *relative generator*. The entropy of the relative generator with respect to  $f$  is our best estimate for the metric entropy.

### III. EXAMPLES

In this section the methods discussed in Sec. II are applied to a number of examples. Sections III A and III B are devoted to dissipative systems which are more suitable as test models since the contraction in phase space reduces statistical problems.

As a first example we chose the baker transformation. Since the baker transformation is locally linear or the composition of a linear map and a translation it is amenable to a theoretical analysis and analytical formulas for the metric entropy and information dimension can be derived.<sup>17-19</sup> Thus we could compare the results of the computation with the theoretical values. The second dissipative system is the Hénon map. Although no analytical formulas are available this map has been subject to an extensive numerical study and numerical quantities are available for comparison.

In conservative dynamical systems we find two important groups.

(a) Axiom- $A$  systems.

(b) Systems satisfying the conditions of the KAM theorem. These are not axiom- $A$  systems since they have elliptic periodic points.<sup>20</sup>

As an example of the first kind we picked the Arnold cat map. Like the baker transformation this map has a simple form, being locally linear. Therefore one can derive analytic formulas for the desired quantities and use them to check the results of the computation. As a model of a Kolmogorov-Arnold-Moser (KAM) system we pick the Froeschlé map.<sup>21</sup> This map has a phase portrait which contains the essential features of a Poincaré map of a Hamiltonian system. In particular, there are both regular and irregular regions in phase space.

Before presenting the results we reiterate the steps taken in all cases, and which are the practical conclusions of Sec. II. These are as follows.

(1) We start with a point in the basin of attraction and iterate it a couple of thousand times to obtain a point  $X_0$  which virtually lies on the attractor.

(2) We iterate the map  $f$  on  $X_0$  a large number of iterations  $N$  to create the generating sequence.

(3) The individual correlation function of order  $d$  of a given point is calculated using the same algorithm employed in Ref. 22. Appendix A.

(4) We average the individual correlation function. The averaging is not carried over all points of the generating sequence but over a smaller number of points chosen randomly. This number is determined by starting with a small number of points and increasing it until no significant change in the average value is observed.

(5) The information dimension is obtained by plotting  $\langle \log_2 [C^d(\epsilon)] \rangle$  versus  $\log_2 \epsilon$  and computing the slope of the curve at low values of  $\epsilon$  (This is done for various values of  $d$ ).

(6) The partition entropies with respect to  $f$  are calculated by plotting  $\delta_d$  [calculated from formula (2.24)] versus  $d$ .

(7) The relative generator is located as described in Sec. II D. All logarithms in the numerical work are calculated to base 2.

### A. Baker transformation

The baker transformation is

$$\begin{aligned} X_{n+1} &= \begin{cases} \lambda_a X_n, & Y_n \leq \alpha \\ \frac{1}{2} + \lambda_b X_n, & Y_n > \alpha \end{cases} \\ Y_{n+1} &= \begin{cases} \frac{1}{\alpha} Y_n, & Y_n \leq \alpha \\ \frac{1}{1-\alpha} (Y_n - \alpha), & Y_n > \alpha \end{cases} \end{aligned} \quad (3.1)$$

where

$$0 \leq X_n, Y_n \leq 1, \quad 0 \leq \lambda_a, \lambda_b \leq \frac{1}{2}, \quad 0 < \alpha < 1.$$

We study the case  $\lambda_a = \lambda_b = \alpha = \frac{1}{4}$ . Owing to the locally simple form of the transformation, analytic formulas for the information dimension and Lyapunov exponents can be derived.<sup>17-19</sup> The information dimension is given by

$$D_1 = 1 + \frac{H(\alpha)}{\alpha \log_2 \frac{1}{\lambda_a} + (1-\alpha) \log_2 \frac{1}{\lambda_b}}, \quad (3.2)$$

where

$$H(\alpha) = \alpha \log_2 \frac{1}{\alpha} + (1-\alpha) \log_2 \frac{1}{1-\alpha}. \quad (3.3)$$

The maximum Lyapunov exponent is given by  $\lambda_1 = H(\alpha)$ . For the map under investigation we obtain

$$D_1 = 1.406 \quad (3.4)$$

and assuming  $K(f) = \lambda_1$ ,

$$K(f) = 0.811. \quad (3.5)$$

In the numerical computations we used 100 000 iterations and carried the averaging over 500 points. The results were stable up to 0.01 (in the information dimension) upon increasing the number of iterations up to 400 000 and the averaging number to 1000. The results could be reproduced with 50 000 iterations but with more statistical fluctuations. In general, results were found to be more sensitive to the number of iterations than to the averaging number. We used the initial conditions  $x=0.5$ ,  $y=0.5$ , and 5000 iterations to converge on the attractor.

Figure 1 shows plots of  $\langle \log_2 C^d(\epsilon) \rangle$  versus  $\log_2 \epsilon$  for  $d=1-15$ . We expect the following.

- (i) The curves should attain a constant slope (the same for all curves) as  $\epsilon \rightarrow 0$ .
- (ii) The slope should be equal to the information dimension.

As the size of the partition  $\epsilon$  decreases, more iterations are needed to yield the invariant measure of the boxes of the partition. Therefore we further expect the following.

- (1) The constancy of the slopes will break down for  $\epsilon$  smaller than some value  $\epsilon_c$ .
- (2) As we increase the embedding dimension the distances between points increase. Therefore  $\epsilon_c$  should increase with the embedding dimension and linear portion

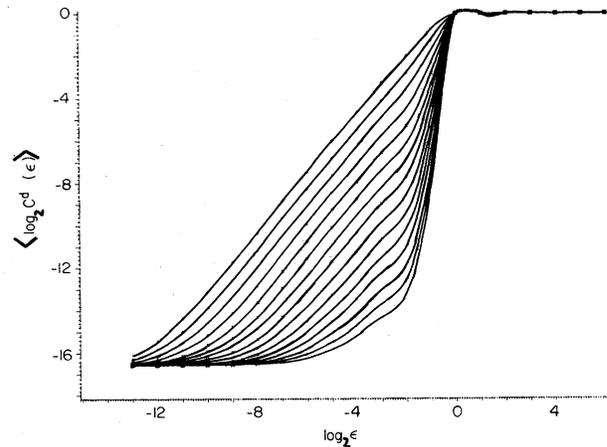


FIG. 1.  $\langle \log_2 C^d(\epsilon) \rangle$  vs  $\log_2 \epsilon$  for the baker transformation for embedding dimensions  $d=1-15$ . ( $d=1$  is the upper most curve.) Average slope is 1.41. For a discussion of this figure see the text.

of the curve should shift to higher values of  $\epsilon$ .

(3) Linearity should break down for  $\epsilon$  large enough due to saturation.

(4) If we increase the number of iterates of the generating sequence, the linear portion of the curves should extend to smaller  $\epsilon$ .

Figure 1 confirms all these expectations save the last one, which we confirmed independently. We calculated the slopes of the curves by linear regression using points which fall within the linear parts of Fig. 1. The slopes have the same value up to the experimental error. The average value  $\mathcal{S}_{av}$  is

$$\mathcal{S}_{av} = 1.41 \pm 0.01. \quad (3.6)$$

The error has been estimated by checking the variations in results under change of parameters (see above) and should be treated with caution. It is seen that  $\mathcal{S}_{av}$  is equal to theoretical value of the information dimension and the linear fit is good.

To avoid statistical problems in the computation of the metric entropy we chose partitions that fall within the linear part of the curves of Fig. 1 for all embedding dimensions used. Figure 2 shows a plot of  $\frac{1}{3} [\langle \log_2 C^d(\epsilon) \rangle - \langle \log_2 C^{d+3}(\epsilon) \rangle]$  versus  $d$  for  $\epsilon=2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}$ . This quantity settles down to a constant value already at the low dimensions confirming the property of quick convergence which was argued in Sec. II C. The constant value is roughly the same for all partitions with  $\epsilon < \frac{1}{4}$  suggesting that these are generators. This agrees also with the theoretical result of Appendix A that  $\frac{1}{4}$  is an expansive constant for  $f$ . For low embedding dimensions the plotted quantities decrease with dimension as can be predicted theoretically (see Appendix B) but for higher dimensions this tendency is disrupted indicating that statistical problems set in.

Following the above we pick the minimum values of the curves corresponding to  $\epsilon < \frac{1}{4}$  as our estimate of the entropy. We get

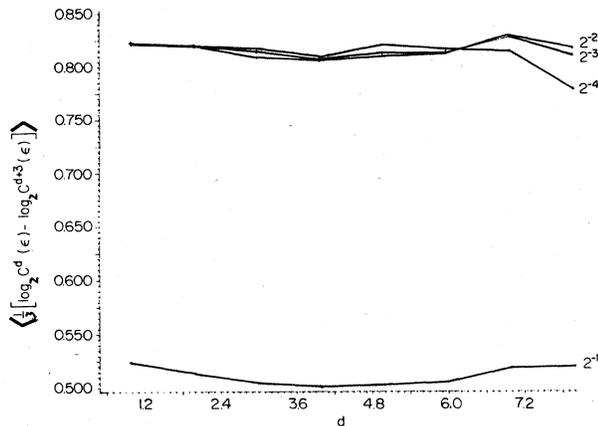


FIG. 2.  $\frac{1}{3}[\langle \log_2 C^d(\epsilon) \rangle - \langle \log_2 C^{d+3}(\epsilon) \rangle]$  vs  $d$  for the baker transformation for  $\epsilon = 2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}$ . The convergence for  $\epsilon < \frac{1}{4}$  supports the theoretical result that partitions with  $\epsilon < \frac{1}{4}$  are generators.

$$H(f) \approx 0.81. \tag{3.7}$$

This value is the same as the theoretical value but we cannot give an estimate of the numerical error.

Figure 3 shows plots of  $-d^{-1} \langle \log_2 C^d(\epsilon) \rangle$  versus  $d$  for  $\epsilon = 2^{-1}, \dots, 2^{-13}$ . The curves decrease with  $d$ . The smaller the radius of the partition the higher the curve. All the curves of  $\epsilon < \frac{1}{4}$  appear to converge together, far removed from the lower curve, supporting the surmise that they belong to generators.

B. Hénon map

The Hénon map is given by

$$X_{n+1} = 1 - aX_n^2 + Y_n, \tag{3.8}$$

$$Y_{n+1} = bX_n.$$

We picked the frequently used values  $a = 1.40, b = 0.3$ . We tested the above procedure for this map both on the basis of the original phase points and with phase points reconstructed from the  $X$  coordinate. The results are similar and we therefore display here the results based on the original phase points, which are marginally better. We again used 100 000 iterations and carried the averaging over 500 points. The results were stable up to 0.01 (in the information dimension) upon increasing the number of iterations up to 400 000 and the averaging number to 1000. The results could be again reproduced with 50 000 iterations but with more statistical fluctuations. In general, results were found to be more sensitive to the number of iterations than to averaging number. We used the initial conditions  $x = 0.639, y = 0.189$ , and 5000 iterations to converge on the attractor.

Figure 4 shows plots of  $\langle \log_2 C^d(\epsilon) \rangle$  versus  $\log_2 \epsilon$  for  $d = 1-15$ . The curves display the same features as those of the baker transformation (Fig. 1) and the same remarks apply here. The slopes are the same for all embedding di-

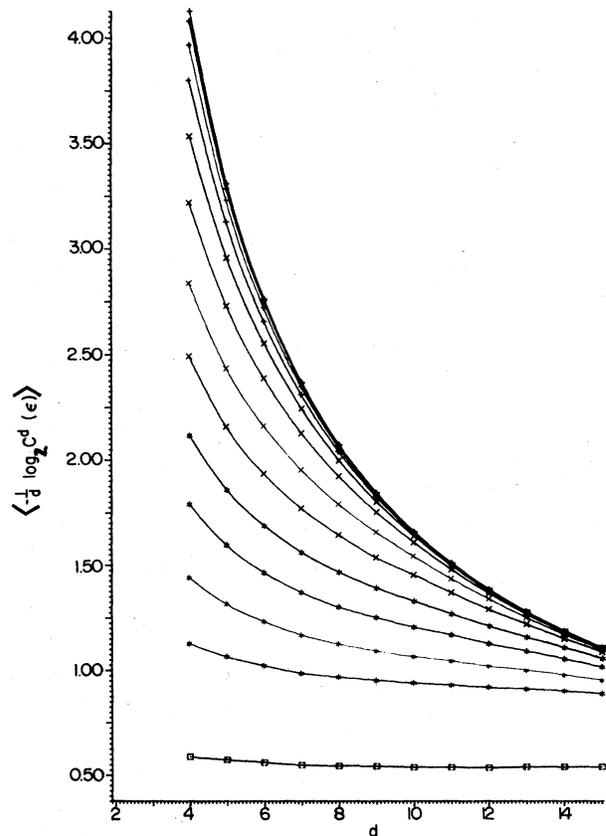


FIG. 3.  $-d^{-1} \langle \log_2 C^d(\epsilon) \rangle$  vs  $d$  for the baker transformation for  $\epsilon = 2^{-1}, \dots, 2^{-13}$ . All curves with  $\epsilon < \frac{1}{4}$  appear to converge together, but with much slower convergence rate as argued theoretically.

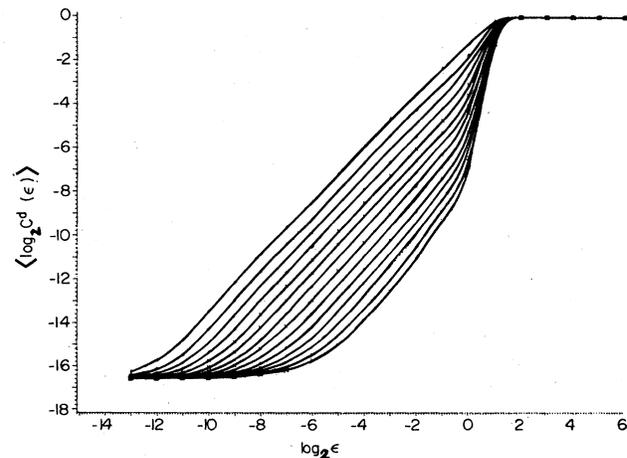


FIG. 4. Same as Fig. 1 but for the Hénon map. Average slope is 1.26.

mensions. The linear fit as attested by the correlation coefficients is good. The average slope is

$$\mathcal{S}_{av} = 1.26 \pm 0.01. \tag{3.9}$$

The same value has been obtained previously by various methods.<sup>19</sup>

Figure 5 shows plots of  $\frac{1}{3}[\langle \log_2 C^d(\epsilon) \rangle - \langle \log_2 C^{d+3}(\epsilon) \rangle]$  versus  $d$  for  $\epsilon = 2^0, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}, 2^{-5}$ . These quantities settle down to a roughly constant value at low dimensions. The curves are decreasing as we expect, the exception being partitions of small radius at high embedding dimension. This is ascribed to the statistical problems that arise if the number of iterates used is not large enough to produce a sufficient number of pairs with small distances. From the figure it is seen that the curves corresponding  $\epsilon = 2^{-4}$  and  $2^{-5}$  coincide (up to the precision with which we could hope to extract the metric entropy) and we conjecture that the corresponding partitions are generators. As our estimate of the entropy we take the lowest point on the curve corresponding to  $\epsilon = 2^{-5}$ . We have

$$H(f) \approx 0.61. \tag{3.10}$$

This value is in accordance<sup>23</sup> with the value obtained in Ref. 19.

Figure 6 shows plots of  $-d^{-1} \langle \log_2 C^d(\epsilon) \rangle$  for partitions with  $\epsilon = 2^0, \dots, 2^{-11}$ . The curves corresponding to smaller  $\epsilon$  are higher and the curves are decreasing as expected. The convergence is slow as compared to Fig. 5. Figure 6 lends some support to our assumption that the partition corresponding to  $\epsilon = 2^{-5}$  is a generator. The curves above it appear to converge to it at least up to the accuracy that we could hope to achieve in such numerical experiments.

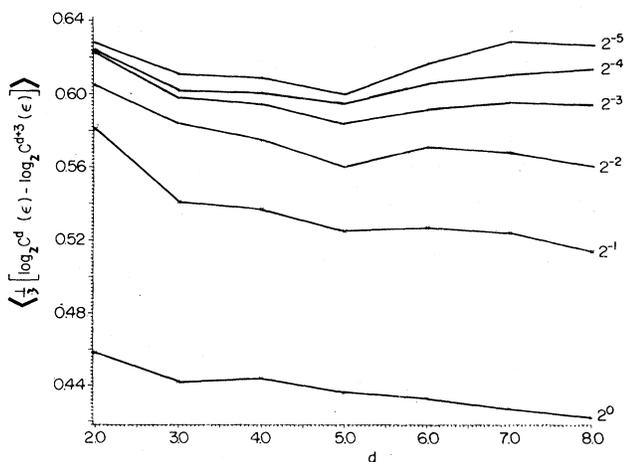


FIG. 5. Same as Fig. 2 but for the Hénon map, and for  $\epsilon = 2^0, 2^{-1}, 2^{-2}, \dots, 2^{-5}$ . The rise for large  $d$  in the curves with  $\epsilon < 2^{-2}$  is ascribed to statistical uncertainties. Within the numerical uncertainties it appears that partitions with  $\epsilon < 2^{-4}$  are generators.

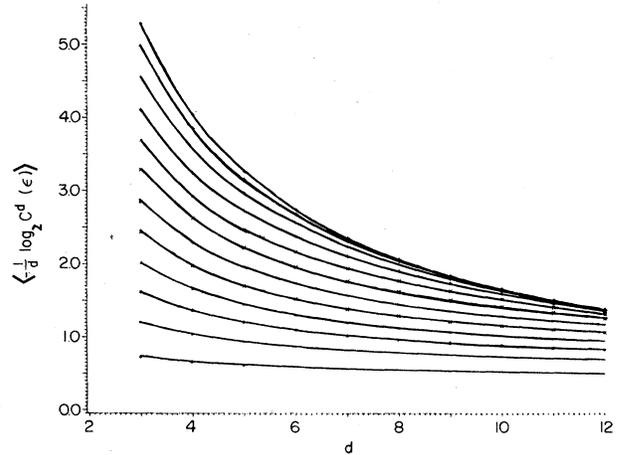


FIG. 6. Same as Fig. 3 but for the Hénon map with partitions  $\epsilon = 2^0, \dots, 2^{-11}$ . The convergence is again slower than in Fig. 5, but supports the surmise that the partition with  $\epsilon = 2^{-5}$  is a generator.

### C. Cat map

Let  $M$  be the torus

$$M = \{(x, y) \mid (x, y) \pmod{1}\}.$$

Let  $\mu$  be the ordinary Lebesgue measure. We study the map

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \pmod{1} \tag{3.11}$$

where  $a, b, c, d$  are integers satisfying  $ad - bc = 1$ . This map is an axiom- $A$  system and its attractor consists of the whole torus  $M$ . For the immediate purpose of this section we cite the following result.<sup>24</sup>

*Theorem.* The metric entropy of  $f$  is given by  $H(f) = \log_2 |\lambda_1|$  where  $\lambda_1$  is the proper value whose modulus is greater than 1 of the matrix  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ . In this work we examine the map

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \pmod{1}. \tag{3.12}$$

Hence

$$H(f) = \log_2 \frac{3 + \sqrt{5}}{2} = 1.388. \tag{3.13}$$

Since  $f$  is axiom  $A$  it possesses a unique invariant measure with respect to which it is ergodic on  $M$ . Since  $f$  is conservative and mixing this must be the Lebesgue measure. Hence the information dimension of the attractor  $M$  is

$$D_1(M) = 2. \tag{3.14}$$

We found numerically that because  $f$  is conservative, a very large number of iterations is required to obtain the invariant probability measure for partitions of radius much smaller than 1. This fact introduces great difficulties in the implementation of our algorithm.

To get reasonable results for the cat map 1 500 000 iterations were required. There was no sensitivity, however, to the number of points used in the averaging and we reduce the number to 200 to cut computation time. We used reconstructed phase points for the calculations, since this procedure has advantages when statistical problems prevail.<sup>25</sup>

Figure 7 shows plots of  $\langle \log_2 C^d(\epsilon) \rangle$  versus  $\log_2 \epsilon$  for  $d=2-10$ . The curves display the same features as those of the baker transformation (Fig. 1) but the range of linearity is considerably smaller and linearity is destroyed much faster as we increase the embedding dimension. This forewarns us that less dimensions will be available for the metric entropy computation. The average slope is 2, but it is difficult to estimate the error.

Figure 8 shows plots of  $\frac{1}{2}[\langle \log_2 C^d(\epsilon) \rangle - \langle \log_2 C^{d+2}(\epsilon) \rangle]$  versus  $d$  for  $\epsilon=2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}, 2^{-5}$ . These quantities settle down to a roughly constant value at low dimensions. For  $\epsilon < 2^{-2}$  all the curves are very close. This suggests that all these curves belong to generators. These curves should be decreasing and the fact that they begin to rise in high dimensions is a further evidence for statistical troubles there. The metric entropy is extracted from the minima of the curves. We get

$$K(f) \simeq 1.38 . \tag{3.15}$$

The theoretical value is 1.39 but again we have no apparent way to estimate the error in the numerical value.

Figure 9 shows plots of  $-d^{-1} \langle \log_2 C^d(\epsilon) \rangle$  for partitions with  $\epsilon=2^{-1}, \dots, 2^{-12}$ . The curves corresponding to smaller  $\epsilon$  are higher and the curves are decreasing as expected. The convergence is slow as compared to Fig. 8. Figure 9 lends some support to our assumption that partitions corresponding to  $\epsilon < 2^{-2}$  are generators. Convergence, however, is far from complete and no conclusive statements can be made.

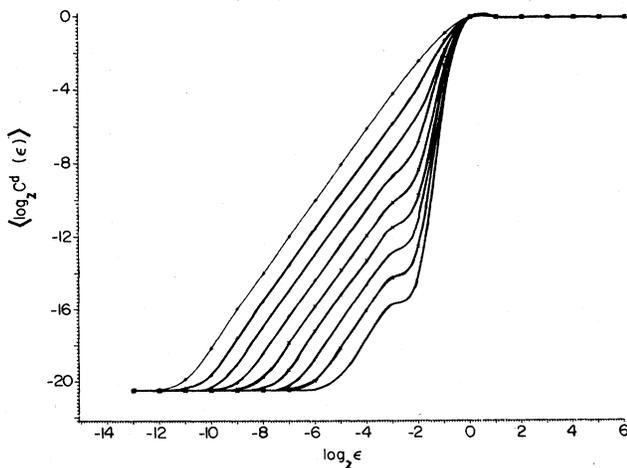


FIG. 7. Same as Fig. 1 for the cat map for  $d=2-10$ . Average slope is 2.

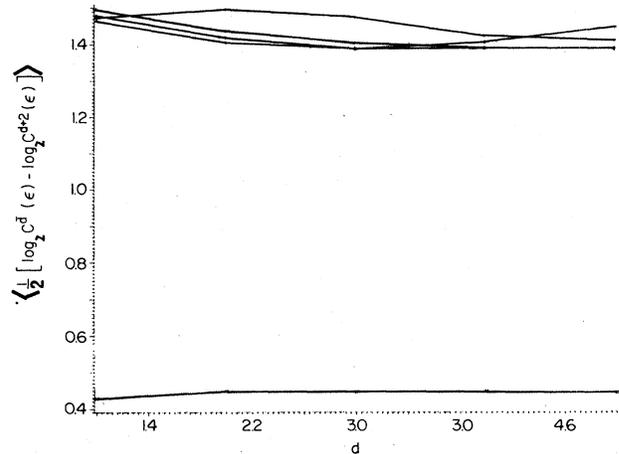


FIG. 8. Same as Fig. 2 for the cat map with  $\epsilon=2^{-1}, 2^{-1}, \dots, 2^{-5}$ . The convergence suggests that all partitions with  $\epsilon < 2^{-2}$  are generators.

#### D. Froeschlé map

Our last example is the map<sup>21</sup>

$$X_{n+1} = X_n + A \sin Y_n \pmod{2\pi} \tag{3.16}$$

$$Y_{n+1} = X_n + Y_n + A \sin Y_n \pmod{2\pi}$$

with  $A=1.3$ . This is a conservative system which unlike the cat map displays some of the characteristics of Poincaré maps of Hamiltonian systems. The fixed points of the map are given by

$$\begin{aligned} X &= 0, \\ Y &= k\pi, \quad k=0,1. \end{aligned} \tag{3.17}$$

A fixed point of the map is elliptic if  $\cos Y = -1$  and hyperbolic if  $\cos Y = 1$ . The phase portrait of the map consists of a lattice of islands of invariant curves surrounding

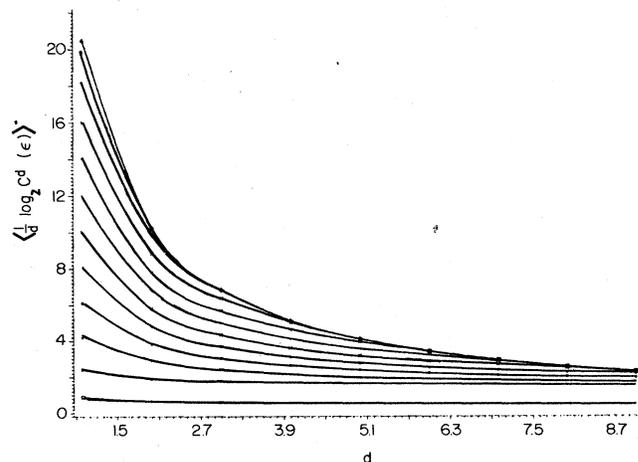


FIG. 9. Same as Fig. 3 for the cat map for partitions with  $\epsilon=2^{-1}, \dots, 2^{-12}$ . The convergence is again slower than in Fig. 8 but supports the assertion about generators.

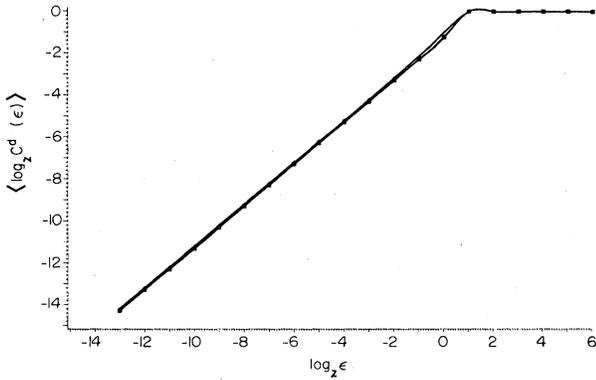


FIG. 10. Same as Fig. 1 but for the Froeschle map starting from initial conditions that lead to regular behavior. Shown are curves for embedding in  $d=2-19$ . All the curves converge because  $K=0$ . Average slope is 1.

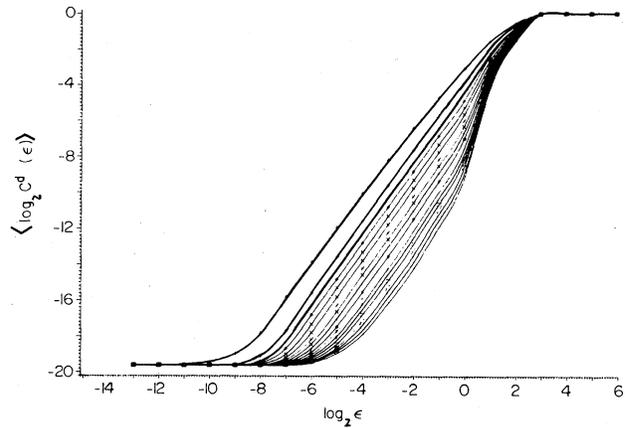


FIG. 12. Same as Fig. 10 but with initial conditions that lead to chaotic behavior. Average slope is 2.

elliptic fixed points, embedded in a connected region of stochasticity.<sup>21</sup>

**Results.** Here we confined our attention to the distinction between regular and irregular trajectories. We used 800 000 iterations and the averaging was carried over 200 points. For the regular zone we chose  $x=0, y=2.5$ . For the irregular zone we chose  $x=2, y=0$ .

**Regular zone.** Figure 10 shows plots of  $\langle \log_2 C^d(\epsilon) \rangle$  versus  $\log_2 \epsilon$  for  $d=2-19$ . The curves merge together with slope 1.

Figure 11 shows plots of  $\frac{1}{3}[\langle \log_2 C^d(\epsilon) \rangle - \langle \log_2 C^{d+3}(\epsilon) \rangle]$  versus  $d$  for partitions with various radii. There is a sharp decrease to very small values.

**Irregular zone.** Figure 12 shows plots of  $\langle \log_2 C^d(\epsilon) \rangle$  versus  $\log_2 \epsilon$  for  $d=2-19$ . The picture is similar to that obtained for the dissipative systems. The slope is constant and is equal to 2. We observe that the situation here is better than in the case of the cat map although less iterations were used.

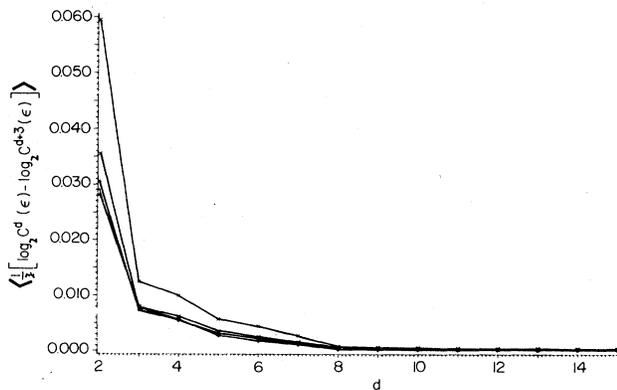


FIG. 11. Same as Fig. 3 for the Froeschle map, with initial conditions that lead to regular behavior. The sharp decrease to very small values indicates that  $K=0$ .

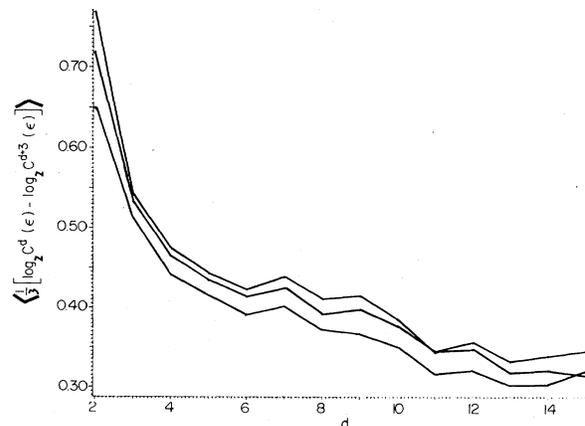


FIG. 13. Same as Fig. 11 but with initial conditions that lead to chaotic behavior. Curves saturate at a finite value indicating positive  $K$  entropy.

Figure 13 shows plots of  $\frac{1}{3}[\langle \log_2 C^d(\epsilon) \rangle - \langle \log_2 C^{d+3}(\epsilon) \rangle]$  versus  $d$  for partitions with various radii. There is a clear difference between this case and the regular case. Although more iterations are probably needed to obtain exact values of the metric entropy, there is no sharp decrease to very small values.

#### IV. CONCLUDING REMARKS

We presented methods for extracting the Kolmogorov entropy from the time signal with some emphasis on the details of the theory and the proposed algorithms. Generally speaking it seems that the proposed algorithms yield reliable values in all cases, but that dissipative systems are easier to deal with than conservative ones. It seems that the algorithms are applicable to experimental signals in the same way that the algorithms for computing dimensions have proved useful. We hope that in the near future experimental values of Kolmogorov entropy will begin to appear.

**APPENDIX A: MAXIMAL EXPANSIVE CONSTANT FOR THE BAKER TRANSFORMATION**

As an example of the calculation of an expansive constant we pick the map  $f$  defined by

$$X_{n+1} = \begin{cases} \frac{1}{4}X_n, & Y_n \leq \frac{1}{4} \\ \frac{1}{2} + \frac{1}{4}X_n, & Y_n > \frac{1}{4} \end{cases} \tag{A1}$$

$$Y_{n+1} = \begin{cases} 4Y_n & Y_n \leq \frac{1}{4} \\ \frac{4}{3}(Y_n - \frac{1}{4}), & Y_n > \frac{1}{4} \end{cases} \tag{A2}$$

$f$  is defined on the unit square  $M$ . We shall establish that  $f$  has an attracting set  $A$ , and that the restriction of  $f$  to  $A$  is expansive with a maximal expansive constant  $\frac{1}{4}$ .

Given a point  $(x, y)$  in  $M$  one can write

$$X = \sum_{k=1}^{\infty} \frac{a_k}{4^k}, \tag{A3}$$

where  $a_k$  can assume the values 0,1,2,3. The sequence  $a_1, a_2, \dots, a_n, \dots$  is the symbolic sequence of  $X$ . The action of  $f$  on  $X$  is most conveniently expressed in terms of its action on the symbolic sequence:  $f(a_1, a_2, \dots, a_n, \dots) = T(y), a_1, a_2, \dots$  where here  $T(y)$  is the function

$$T(y) = \begin{cases} 0, & y \leq \frac{1}{4} \\ 2, & y > \frac{1}{4} \end{cases} \tag{A4}$$

Define  $A$  to be the set of points of  $M$  with  $X$  coordinates whose symbolic sequences satisfy

$$a_k = 0 \text{ or } a_k = 2 \text{ for all } k$$

or

$$a_k = 0 \text{ or } a_k = 2, \quad k \leq n \tag{A5}$$

and

$$a_k = 3, \quad k > n$$

where  $n \geq 1$ .  $A$  is a compact attracting set whose domain of attraction is  $M$ . The restriction of  $f$  to  $A$ ,  $\hat{f}$ , is invertible. The action of  $\hat{f}^{-1}$  on the symbolic sequence of the  $X$  coordinate is

$$\hat{f}^{-1}(a_1, a_2, \dots, a_n) = a_2, a_3, \dots, a_n \tag{A6}$$

while its action on  $Y$  is given by

$$\hat{f}^{-1}(Y) = \begin{cases} \frac{1}{4}Y, & a_1 = 0 \\ \frac{3}{4}Y, & a_1 = 2 \end{cases} \tag{A7}$$

We can now show that  $\frac{1}{4}$  is an expansive constant for  $\hat{f}$ . We note that given two distinct points with different  $y$  coordinates their symbolic sequences must differ at some

place, say, the  $k$ th place. Iterating  $f$   $k$  times we end up with  $x$  coordinates differing in the first place in their symbolic sequences. Therefore the  $x$  coordinates differ by at least  $\frac{1}{4}$ . If the  $y$  coordinates are the same then the  $x$  coordinates must differ in some place in their symbolic sequence say the  $k$ th place. Iterating  $f^{-1}$   $k - 1$  times we obtain  $x$  coordinates differing in their first place. Again this implies a difference of at least  $\frac{1}{4}$ . Thus  $\frac{1}{4}$  is an expansive constant.

**APPENDIX B; CONVERGENCE RATES OF SERIES OF THE METRIC ENTROPY**

The arguments given here follow the presentation of Ref. 14. Given two partitions where one is a refinement of the other, we define the conditional entropy as in Eq. (2.9). We list here a few basic properties of partitions and entropies<sup>14</sup> (it is assumed that  $\phi, \psi$ , and  $\theta$  are finite partitions of  $M$ ).

- (1)  $K(\phi/\psi) > 0$ .
- (2)  $K(\phi \wedge \psi) = K(\psi) + K(\phi/\psi)$ .
- (3)  $K(\phi) > K(\phi/\psi)$ .
- (4)  $\psi \leq \phi$  implies  $K(\psi) \leq K(\phi)$ .
- (5)  $\psi \leq \phi$  implies  $K(\theta/\psi) \geq K(\theta/\phi)$ .
- (6)  $K(f^{-1}(\phi)/f^{-1}(\psi)) = K(\phi/\psi)$ .
- (7)  $f^{-n}(\phi \wedge \psi) = f^{-n}(\phi) \wedge f^{-n}(\psi)$ .
- (8)  $K(f^{-1}(\phi)) = K(\phi)$ .

Using properties (2), (7), and (8) it can be shown<sup>14</sup> that

$$K \left[ \bigvee_{i=0}^{n-1} f^{-i}(\phi) \right] = H(\phi) + \sum_{i=1}^{n-1} K \left[ \phi / \bigvee_{k=1}^i f^{-k}(\phi) \right]. \tag{B1}$$

From properties (1), (3), and (5) it can be seen that the sequence

$$A_i = K(\phi), \tag{B2}$$

$$A_n = K \left[ \bigvee_{i=0}^{n-1} f^{-i}(\phi) \right] - K \left[ \bigvee_{i=0}^{n-2} f^{-i}(\phi) \right] \tag{B3}$$

is non-negative and nonincreasing and hence converges to a limit  $L$ . The entropy of  $f$  with respect to  $\phi$  is defined as

$$K(f, \phi) = \lim_{n \rightarrow \infty} B_n, \tag{B4}$$

where

$$B_n = \frac{1}{n} K \left[ \bigvee_{i=0}^{n-1} f^{-i}(\phi) \right]. \tag{B5}$$

The existence is proved by noting that

$$B_n = \frac{1}{n} \sum_{k=1}^n A_k \tag{B6}$$

and hence

$$\lim_{n \rightarrow \infty} B_n = \lim_{n \rightarrow \infty} A_n = L. \quad (\text{B7})$$

For computations it is observed that the sequence  $A_n$  converges more quickly to the limit  $L$ . This can be seen as

follows. Since  $A_n$  is nonincreasing  $A_n \geq L$ . Furthermore,

$$B_n = \frac{1}{n} \sum_{k=1}^n A_k \geq A_n. \quad (\text{B8})$$

Hence

$$|A_n - L| \leq |B_n - L|. \quad (\text{B9})$$

- <sup>1</sup>F. Takens, in Proceedings of the 13th Brazilian Colloquium on Mathematics (unpublished); in *Proceedings of the Symposium on Dynamical Systems and Turbulence, Warwick 1979–1980*, edited by D. A. Rand and L. S. Young (Springer, Berlin, 1981); L. S. Young, *Ergodic Theor. and Dyn. Syst.* **2**, 109 (1982); M. Brin and A. B. Katok (unpublished).
- <sup>2</sup>I. Procaccia, Proceedings of the Nobel Symposium on the Physics of Chaos and Related Topics [Phys. Scr. (in press)].
- <sup>3</sup>P. Grassberger and I. Procaccia, *Phys. Rev. Lett.* **50**, 346 (1983).
- <sup>4</sup>B. Malraison, P. Atten, P. Bergé, and M. Dubois, *J. Phys. (Paris) Lett.* **44**, (1983).
- <sup>5</sup>J. Guckenheimer and G. Buzyna, *Phys. Rev. Lett.* **51**, 1438 (1983).
- <sup>6</sup>A. Brandstater, J. Swift, H. L. Swinney, A. Wolf, J. D. Farmer, E. Jen, and P. J. Crutchfield, *Phys. Rev. Lett.* **51**, 1442 (1983).
- <sup>7</sup>Some further experimental examples can be found in the Proceedings of the Nobel Symposium on the Physics of Chaos and Related Topics [Phys. Scr. (in press)].
- <sup>8</sup>A. N. Kolmogorov, *Dokl. Akad. Nauk SSSR* **124**, 754 (1959).
- <sup>9</sup>Ya. G. Sinai, *Dokl. Akad. Nauk SSSR* **124**, 768 (1959).
- <sup>10</sup>R. Shaw, *Z. Naturforsch.* **36A**, 80 (1981).
- <sup>11</sup>D. Ruelle, in *Bifurcation Theory and its Application in Scientific Disciplines* (New York Academy of Science, New York, 1979), Vol. 316.
- <sup>12</sup>J. B. Pesin, *Russ. Math. Surveys* **32**, 455 (1977).
- <sup>13</sup>P. Grassberger and I. Procaccia, *Phys. Rev. A* **28**, 2591 (1983).
- <sup>14</sup>P. Walter, *Ergodic Theory—Introductory Lecture Notes* (Springer, Berlin, 1975).
- <sup>15</sup>M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic Theory on Compact Spaces*, Vol. 527 of *Lecture Notes in Mathematics* (Springer, Berlin, 1976), p. 107.
- <sup>16</sup>H. S. Greenside, A. Wolf, J. Swift, and T. Pignataro, *Phys. Rev. A* **25**, 3453 (1982).
- <sup>17</sup>J. D. Farmer, E. Ott, and J. A. Yorke, *Physica (Utrecht)* **7D**, 153 (1983).
- <sup>18</sup>H. G. E. Hentschel and I. Procaccia, *Physica (Utrecht)* **8D**, 435 (1983).
- <sup>19</sup>P. Grassberger and I. Procaccia, *Physica (Utrecht)* **D** (in press).
- <sup>20</sup>Recall that in an axiom- $A$  system every nonwandering point must be hyperbolic.
- <sup>21</sup>C. Froeschlé, *Astron. Astrophys.* **9**, 15 (1970).
- <sup>22</sup>P. Grassberger and I. Procaccia, *Physica (Utrecht)* **9D**, 189 (1983).
- <sup>23</sup>Notice that here logarithms are calculated to base 2.
- <sup>24</sup>V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, Reading, Mass. 1968).
- <sup>25</sup>To see this point assume that pairs of consecutive iterates  $(x_n, x_{n+1})$  constitute an embedding of the true trajectory in  $R^2$ . Then the triple  $(x_n, x_{n+1}, x_{n+2})$  is a pair of consecutive points in the reconstructed trajectory. Since we are using the square metric (see Sec. II C) the distance between two pairs of consecutive reconstructed points is the distance between two triples. On the other hand, in the original trajectory the distance between two pairs of consecutive points is the distance between two quadruples. Thus distances between consecutive sequences of reconstructed points are smaller and the statistical problems are reduced.