Physica 7D (1983) 201–223
North-Holland Publishing Company

SYMBOLIC DYNAMICS OF NOISY CHAOS

J.P. CRUTCHFIELD and N.H. PACKARD
Physics Board of Studies, University of California, Santa Cruz, California, USA

One model of randomness observed in physical systems is that low-dimensional deterministic chaotic attractors underly the observations. A phenomenological theory of chaotic dynamics requires an accounting of the information flow from the observed system to the observer, the amount of information available in observations, and just how this information affects predictions of the system's future behavior. In an effort to develop such a description, we discuss the information theory of highly discretized observations of random behavior. Metric entropy and topological entropy are well-defined invariant measures of such an attractor's "level of chaos", and are computable using symbolic dynamics. Real physical systems that display low dimensional dynamics are, however, inevitably coupled to high-dimensional randomness, e.g. thermal noise. We investigate the effects of such fluctuations coupled to deterministic chaotic systems, in particular, the metric entropy's response to the fluctuations. We find that the entropy increases with a power law in the noise level, and that the convergence of the entropy and the effect of fluctuations can be cast as a scaling theory. We also argue that in addition to the metric entropy, there is a second scaling invariant quantity that characterizes a deterministic system with added fluctuations: I_0, the maximum average information obtainable about the initial condition that produces a particular sequence of measurements (or symbols).

1. The role of fluctuations in dynamical systems modeling

The work of Lorenz [1] and Ruelle and Takens [2] has led to the idea that randomness observed in physical systems may in some cases be modeled by low-dimensional chaotic attractors. A growing body of experimental evidence now supports this view [3]. This data also demonstrates that any purely deterministic model is incomplete, since the dynamics of physical systems is inevitably coupled to some source of fluctuations. We shall refer to these fluctuations as *external fluctuations**. Another attribute that must be incorporated into an accurate model for observed randomness is fluctuations of the measuring instrument, these we will call *observational noise*. Observational noise differs markedly from heat bath fluctuations in that it does not affect the temporal evolution of the system being observed (assuming a classical measurement process); rather, it directly limits what may be inferred about the system under study. We will be concerned only with the effects of external fluctuations here; for further discussion of this classification of noise types see ref. 4.

Incorporation of any kind of fluctuation into a dynamical description implies that observables become average quantities, the average being taken over all possible fluctuations†. For the case of a chaotic deterministic dynamical system, we are led to the idea that observables are average quantities, where the average is taken with respect to the asymptotic probability distribution. When fluctuations are added to such a system, they produce a new asymptotic probability distribution. A formal expression of how this distribution arises will be presented below. In referring to a probability distribution $P(x)$ we will find it convenient to

* We can give an unambiguous definition of this in terms of the ideas presented in this paper: External fluctuations may be regarded as a second dynamical system (coupled to the system of interest) with sufficiently high entropy h_μ so that all the information I from a measurement is lost after the typical time τ used for sampling the first system. In other words, $I/\tau \ll h_\mu$, where I/τ is the information acquisition rate. This allows for an operational definition of a non-deterministic source of random behavior as a deterministic system whose entropy is sufficiently large to preclude an observer's geometric reconstruction of the source's dynamics. All of our information quantities will be measured in bits and so, in particular, all logarithms will be taken to the base 2.

† We will assume fluctuations to be drawn from a stationary ensemble at each time.

0167-2789/83/0000–0000/$03.00

also speak of the associated *measure* μ defined by

$$\mu(A) = \int_A P(x)\,\mathrm{d}x,$$

where A is some set.

We will be concerned with the effect of fluctuations on measurements of randomness, in particular, their effect on the metric entropy. As we shall see from numerical computations, the metric entropy is relatively insensitive to observational noise, but is strongly dependent on external fluctuations coupled to the dynamics; and so we will concentrate mostly on this latter case. The prototypical chaotic systems we shall use are iterated maps of the unit interval I onto itself: $x_{n+1} = f(x_n)$, where f is some nonlinear function. These will also be referred to as one-dimensional maps. We will model the effects of external fluctuations with a stochastic difference equation of the form

$$x_{n+1} = f(x_n) + \xi_n, \tag{1}$$

where ξ_n is a delta-correlated random variable. Numerical experiments [5] indicate that the response of the metric entropy to the added fluctuations ξ is insensitive to the details of their probability distribution. We will assume ξ to have zero mean, and to be evenly distributed over some finite interval, with a standard deviation, or *noise level*, σ.

For dynamical systems with added fluctuations there are not many rigorous results. Kifer [6] has proven that for hyperbolic attractors* the invariant measure converges weakly to the correct zero noise limit. Boyarsky [7] proved that for one-dimensional maps that have slope everywhere greater than one, there exists some noise level for which the invariant measure of the system with fluctuations approximates the zero noise invariant measure with arbitrary accuracy (strong convergence), and that *all* initial conditions have time averages that correspond to averages with respect to the invariant measure.

We begin by reviewing entropy measurement techniques for deterministic systems. We will then investigate the effects of external noise on the symbolic dynamics, and discover that the amount of information $I(n)$ about the initial condition that produces a symbol sequence of length n reaches a limit I_0 at some particular length n_c that is dependent on and scales with the noise level. Furthermore, the added noise produces entropy convergence features that also obey scaling laws. After describing the scaling features of the entropy, we will discuss numerical experiments in which we compute the scaling exponents for many different systems. We then describe an alternate entropy-like quantity similar in spirit to the Lyapunov characteristic exponent, conjecturing equality with the symbolic dynamics metric entropy. We conclude with an overview and a brief discussion of some experimental applications.

2. Symbolic dynamics and entropy for deterministic systems

We must first review the case of observing a deterministic dynamical system. We will consider time to be discrete, and the dynamical system to be a map f from a space of states M into itself, $f: M \to M$. We assume f has some ergodic invariant measure $\bar{\mu}$. If f has an attractor, we will restrict our attention to the attractor, and assume that almost all (with respect to Lebesque measure m) initial conditions approach the attractor and have trajectories that are asymptotically described by the measure $\bar{\mu}$ on the attractor; i.e. that for almost all points the measure

$$\mu_N(x) = \frac{1}{N}\sum_{n=1}^{N} \delta_{f^n x}$$

* In the context of one-dimensional maps, this means that the absolute value of the slope of the map must be greater than one everywhere on the attractor.

converges weakly to $\bar{\mu}$. Oono and Osikawa [8] refer to this assumption as the "condition for observable chaos".

This may seem like an amazing assumption from a mathematical viewpoint, but it is proven rigorously for axiom-A systems, where $\bar{\mu}$ is the Bowen–Ruelle measure. For maps of the unit interval this assumption will hold for all maps that have an invariant measure that is absolutely continuous with respect to Lebesque measure. Consideration of noise added to the dynamics also makes this assumption plausible for most physical contexts, as will be discussed in following section.

The behavior of a dynamical system f: $M \to M$ can have many symbolic representations, each obtained by using a *measurement partition*, $P = \{P_1, \ldots, P_q\}$, to divide the state space M into a finite number of sets each of which is labeled with a symbol $s_i \in \{1, \ldots, q\} \equiv S$. The time evolution $(x_0, x_1, x_2, \ldots)$ of the dynamical system f: $M \to M$ is then translated into a sequence of symbols labeling the partition elements visited by an orbit

$$s = \{s_0, s_1, s_2, \ldots\}$$

and f itself is replaced by a *shift* operator σ which re-indexes a symbol sequence; that is,

$$\sigma(s) = s',$$

where for each symbol in the sequence s',

$$s'_i = (\sigma(s))_i = s_{i+1}.$$

Thus the shift operator σ merely moves the time origin of a symbol sequence one place to the right.

In the space of all possible symbol sequences

$$\Sigma \equiv \{s = (s_0, s_1, \ldots)\},$$

the *observed* or *admissable sequences* are those which satisfy

$$x_i = f^i(x_0) \in P_{s_i}.$$

The set of admissable sequences Σ_f along with the shift σ is called a *subshift*. (Σ_f, σ) is the symbolic dynamical system induced by f using the measurement partition P.

The symbol sequences of Σ_f are a coding for the orbits of f: $M \to M$. A finite sequence of symbols $(s_0^n, \ldots, s_{n-1}^n)$ defines an *n-cylinder* $s^n \equiv \{s : s_i = s_i^n, i = 0, \ldots, n-1\}$ which is a subset of Σ_f consisting of all sequences whose first n elements match with those of s_i^n. An n-cylinder s^n corresponds to a set of orbits that are "close" to one another in that their initial conditions and first $n-1$ iterates fall in the same respective partition elements. Since these orbits must follow each other for at least $n-1$ iterations, they must all have initial conditions that are close, belonging to some set $U \subset M$. We thus have a map Δ from n-cylinders to subsets of M:

$$\Delta(s^n) = \{x \,|\, f^i(x) \in P_{s_i}, \quad \text{for } i = 0, \ldots, n-1\}.$$

To a different n-cylinder will correspond a different set of orbits whose initial conditions are contained in some other set $U' \subset M$. M will become partitioned into as many subsets as there are n-cylinders. As n is increased, this *n-cylinder partition* will become increasingly refined. The refinement caused by taking an increasing number of symbols is illustrated in fig. 1, where M is the unit interval

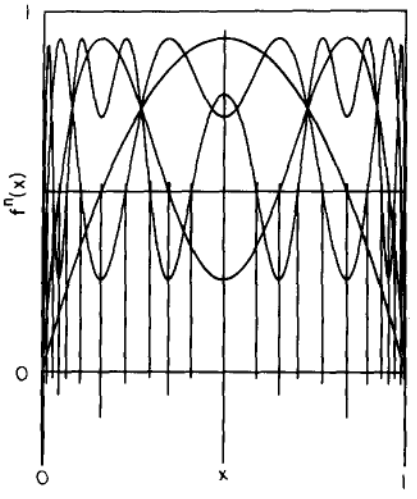

Fig. 1. Construction of the partition induced by taking n symbols (i.e. specifying an n-cylinder) with the measurement partition $\{[0, 0.5], (0.5, 0]\}$. The 1-cylinder, 2-cylinder, 3-cylinder, and 4-cylinder partitions are shown with successively shorter tic marks below the x-axis.

[0, 1], and f is the quadratic logistic equation, $f(x) = rx(1-x)$, with $r = 3.7$. We have used the measurement partition formed by cutting the interval in half at $d = 0.5$, the *critical point* of f where the slope vanishes. We will label the left subinterval "0" and the right "1". We see from the figure that the dividing points for the n-cylinder induced partition are simply the collection

$$\{d, f^{-1}(d), f^{-2}(d), \ldots, f^{-(n-1)}(d) \ldots\}$$

whenever the specified inverse images exist. If the map is not everywhere two onto one (i.e. $r < 4$), some of the inverse images will not exist, corresponding to the fact that some n-cylinders are non-admissable. Changing the measurement partition clearly generates a different set of admissable sequences, just as it generates different n-cylinder partitions.

The usefulness of symbolic dynamics as a representation for the orbits of f can be captured in the following commutative diagram:

$$\begin{array}{ccc} \Sigma_f & \xrightarrow{\sigma} & \Sigma_f \\ \pi \downarrow & & \downarrow \pi \\ M & \xrightarrow[f]{} & M \end{array}$$

with the projection operator

$$\pi(s_0, s_1, \ldots) = \bigcap_{i=0}^{\infty} f^{-i}(P_{s_i}).$$

One can then study the simpler, albeit abstract, symbolic dynamical system in order to answer various questions about the original dynamical system. Within this construction, every point on the attractor will have at least one symbol sequence representation. There are a few ambiguities in the labeling of orbits by symbol sequences that prevent π from being invertible, but our discussion of the entropy will prove to be insensitive to the ambiguities*.

The space of one-sided symbol sequences can easily be metrized by mapping each symbol sequence to a power series

$$\Phi(x) = \sum_{i=1}^{\infty} \frac{S(f^i x)}{q^i},$$

where $s(x)$ is the symbol labeling the measurement partition element containing x (the denominator is 2^i only if the partition has two elements). For the case of a binary partition, this map identifies every sequence with a binary fraction whose value lies in [0, 1]. We will conveniently confuse s^n with its binary fraction representation unless the distinction is necessary†.

A Cantor set structure in the symbol sequences of the chaotic logistic equation is revealed in fig. 2 by a sequence of probability distributions for n-cylinder binary fractions: with the increase in length of the n-cylinder the distributions show successively more, although narrower, peaks. Another demonstration of the Cantor set structure of

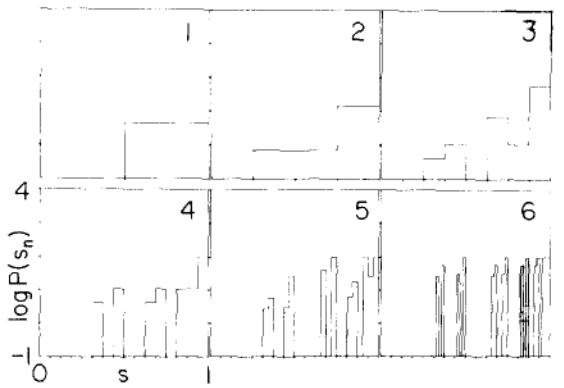

Fig. 2. The Cantor set structure of the subshift (Σ_f, σ) is shown in this sequence of probability distributions for n-cylinders: $n = 1, 2, 3, 4, 5$, and 6. Each n-cylinder has been mapped onto the unit interval by using its binary fraction. In this example f is the quadratic logistic equation with $r = 3.7$.

* An example of one such ambituity is that there can be two symbol sequences that are nowhere the same, but label the same point on the interval: e.g. 100000 . . . and 011111 . . . both label the same point $x = 0.5$ in the limit of infinite length.

† Milnor and Thurston [9] show how to form a slightly more sophisticated "invariant coordinate" which is monotonic. Our entropy calculations do not require this feature, so we use the computationally simpler binary fraction.

Σ_f is the graph of the distribution of symbols s (truncated to a finite n-cylinder with $n = 12$ and mapped onto the unit interval using its binary fraction) versus position x, illustrated in fig. 3.

We will now embark on the task of characterizing the chaotic behavior in a dynamical system using topological and metric entropies, in that order. After giving their definitions, we will show how these quantities may be computed numerically using the symbol sequence representation of orbits. Our analysis follows Shannon [10].

Heuristically, the topological entropy of a dynamical system measures the asymptotic growth rate of the number of resolvable orbits (using a given measurement partition) whose initial conditions are all close. Equivalently, the topological entropy quantifies the average time-rate h of spreading a subset over nearby subsets. This process is most easily illustrated by considering a collection of subsets which form a "cover" of the state space M. The dynamic f spreads a single cover element over other elements after some time t. The number of new cover elements $N(t)$ visited by points in the original cover element can be written,

$$N(t) \sim e^{ht},$$

where $h > 0$ for chaotic dynamical systems. With this geometric motivation, we will now consider a more formal definition of the topological entropy h [11].

For a compact topological space M, with an open cover U, let $N(U)$ be the number of sets in a subcover of minimal cardinality. Two covers U and V may be "combined" to form a refinement W by

$$W = U \vee V$$
$$= \{A \cap B \mid A \in U \quad \text{and} \quad B \in V\}$$

Now if f: $M \to M$ is a continuous map, the *topological entropy* of f with respect to the cover U is defined as

$$h(f, U) = \lim_{n \to \infty} \log \frac{N(U^n)}{n},$$

where

$$U^n = U \vee f^{-1}U \vee \ldots \vee f^{1-n}U.$$

The topological entropy $h(f)$ of the map itself is then the supremum of $h(f, U)$ over all open covers U.

The supremum is obtained only if the measurement partition is "good" in that there is an unambiguous correspondence between orbits of f and symbol sequences. Only with such a good partition is the topological entropy of Σ_f obtained using partition P exactly $h(f)$, the topological entropy of f. There is no general procedure for finding such a good partition, but we will give numerical evidence that such partitions are easily found for simple piecewise monotone maps of the unit interval. Given such a partition, however, we have a readily computable algorithm for $h(f)$:

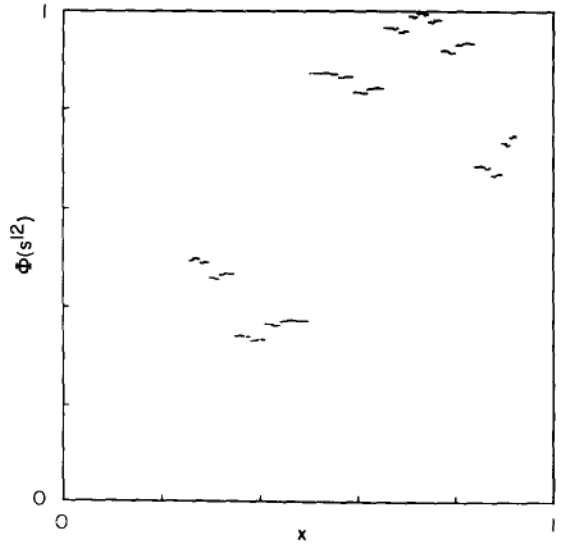

Fig. 3. 2000 iterations of the 'logistic equation' with $r = 3.7$, showing the Cantor set structure of the distribution of sequences in Σ_f. Graphed is $\Phi(s^{12})$ against position x, where s^{12} is the sequence obtained from the initial condition x. The density of points on the x-axis is the asymptotic distribution of f on the unit interval; the density of points on the y-axis is the Cantor distribution illustrated in the previous figure.

simply counting the number of n-cylinders. Note that in the space of symbol sequences Σ_f, each n-cylinder s^n is an open set, and the class of all n-cylinders is an open cover. Thus the topological entropy of the system (Σ_f, σ_f) is given by

$$\lim_{n\to\infty} \frac{\log N(n)}{n} \to h(\sigma_f),$$

where $N(n)$ is the number of admissable n-cylinders*. $N(n)$ is readily obtainable numerically, so this formula presents us with a computable algorithm for the topological entropy†.

In presenting the topological entropy before the metric entropy we have purposely reversed their historical order because there is a sense in which the metric entropy is a generalization of the topological entropy: the metric entropy also measures the asymptotic growth rate of the number of resolvable orbits (using a given measurement partition) having close initial conditions, but weights each orbit with its probability of occurrence.

The definition of metric entropy for the dynamical system (M, f) requires an invariant measure $\bar{\mu}$ and a sigma-algebra of measurable subsets of M: more structure than needed for the definition of topological entropy.

If $P = \{P_i\}$ is a finite measurable partition of M with p elements, we define the entropy of P as

$$H_{\bar{\mu}}(P) = -\sum_{i=1}^{p} \bar{\mu}(P_i) \log(\bar{\mu}(P_i)).$$

Given two partitions P and Q, their refinement is

$$P \vee Q = \{P_i \cap Q_j \,|\, P_i \in P \quad \text{and} \quad Q_j \in Q\}.$$

The metric entropy of f with respect to the partition P is defined by

$$h_{\bar{\mu}}(f, P) = \lim_{n\to\infty} \frac{1}{n} H_{\bar{\mu}}(P^n),$$

where

$$P^n = P \vee f^{-1}P \vee \ldots \vee f^{1-n}P.$$

Finally, the metric entropy of f itself is

$$h_{\bar{\mu}} = \sup_P h(f, P),$$

where the supremum is taken over all partitions P.

As for the topological entropy, the supremum is obtained only for special partitions; Kolmogorov‡ proved that the desired requirement is that the partition be *generating*. This is the case if the smallest sigma-algebra containing $\Delta(s^n)$ for all $n > 0$ coincides with the sigma-algebra of measurable subsets in M. In simpler terms, a partition is generating if, as the length of all sequences becomes large, the sequences label individual points. Thus, only if P is a generating partition we have

$$h_{\bar{\mu}}(f) = h_{\bar{\mu}}(f, P).$$

Again, if we label the elements of the partition P with symbols, the entropy of $h_{\mu}(\sigma_f)$ is exactly $h_{\bar{\mu}}(f, P)$, with

$$\mu(s^n) = \int_{\Delta(s^n)} d\bar{\mu} = \bar{\mu}(\Delta(s^n)).$$

Note that the entropy h_{μ} of (Σ_f, σ) is equal to $h_{\bar{\mu}}(f)$ *only* if the measurement partition is generating. For arbitrary measurement partitions,

$$h_{\mu}(\sigma_f) \leqslant h_{\bar{\mu}}(f).$$

* For the case of symbolic dynamics, this formula for the topological entropy was first introduced by Parry [12], but is essentially the same as the "channel capacity" introduced by Shannon [10].

† Crutchfield and Shaw [13] have developed other algorithms to compute the topological entropy of a map f based on representing the dynamics as a branching process with a deterministic transition matrix. For certain cases, these techniques allow one to analytically calculate the topological entropy and so to study, for example, the convergence of the topological entropy directly (c.f. ref. 14). These techniques are related to the kneading calculus of Milnor and Thurston [9].

‡ This theorem as well as the original definition of metric entropy are presented in Kolmogorov [15].

Assuming a generating measurement partition, the identification between n-cylinders and elements of the refinement P^n allows us to estimate the measure of each element of P^n by accumulating a frequency histogram for the observed n-cylinders. (Note that P^n is exactly the n-cylinder partition illustrated in fig. 1 for $n = 4$.) We may then obtain an n-symbol estimate for the topological entropy from either

$$h(n) = \frac{\log N(n)}{n}$$

or

$$h(n) = \log N(n) - \log N(n-1),$$

and estimates for the metric entropy from

$$h_\mu = \frac{H_\mu(n)}{n}$$

or

$$h_\mu(n) = H_\mu(n) - H_\mu(n-1).$$

It is easily shown that the latter estimate for h_μ converges more quickly than the former [10], so all of our numerical computations of $h_\mu(n)$ will use this expression. Fig. 4 illustrates an example computation of $h(n)$ and $h_\mu(n)$ for the logistic equation, $f(x) = rx(1-x)$ at a typical parameter value, $r = 3.7$.

In order to illustrate the dependence of the entropy on the measurement partition used, we have computed $h(13)$ and $h_\mu(13)$ for a range of binary (two-element) measurement partitions; the results are illustrated in fig. 5. We call the location $x = d$ at which we decide whether a point x on an orbit is either a "0" or a "1" the *decision point*. For two values of the decision point, $d = 0.5$ and $d = 0.839\ldots$ (an inverse image of the critical point), $h_\mu(13)$ is maximized, giving evidence that these values of d yield a generating partition. Note that $h_\mu(13)$ is greater than the Lyapunov characteristic exponent (to be discussed in more detail shortly) because the metric entropy has not converged by thirteen symbols (cf. fig. 4).

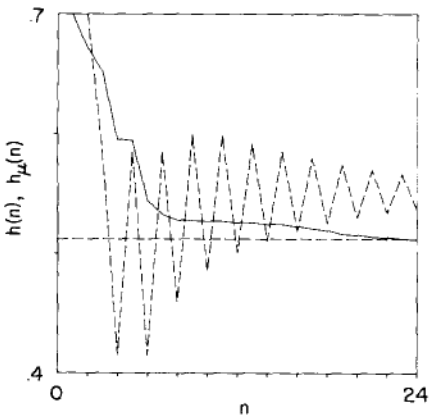

Fig. 4. Entropy convergence for the logistic equation $f(x) = rx(1-x)$, with $r = 3.7$; the solid line represents $h_\mu(n)$ and the oscillating dashed line represents $h(n)$. 2×10^8 iterations were used. The horizontal dashed line is the Lyapunov characteristic exponent.

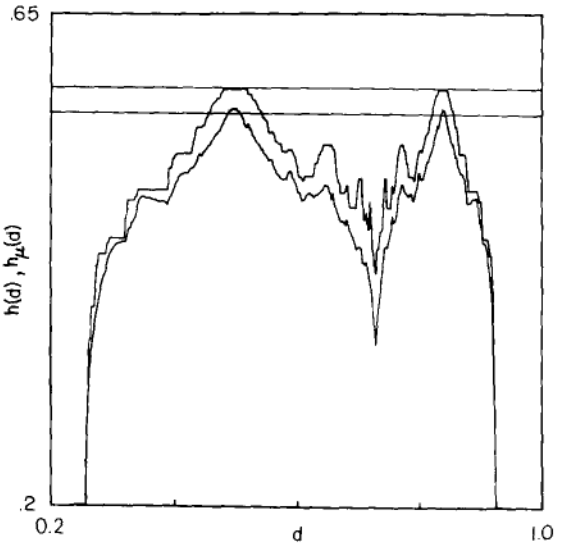

Fig. 5. $h(13)$ (upper curve) and $h_\mu(13)$ for the logistic equation with $r = 3.7$, using different measurement partitions obtained by varying the decision point d. $h(13)$ is actually an average of $h(6), \ldots, h(13)$ to eliminate the oscillitory effects. The upper horizontal line is the topological entropy calculated to one part in 10^6 with the kneading determinant [13, 14]. The lower horizontal line is the Lyapunov characteristic exponent calculated to within 0.1%.

From the above definition of the metric entropy,

it is easy to see that $h \geqslant h_\mu$, since $h(f, P^n)$ is maximized when each element of P^n is equally probable (i.e. $\mu(P_i^n) = 1/N(n)$ for all i). In this case, the formula for the metric entropy reduces to that for the topological entropy. This is also evident from a theorem due to Goodwyn [16] and Dinaburg [17], which states that

$$h = \sup_\mu h_\mu,$$

where the supremum is taken over all invariant measures μ.

One of the primary roles of entropy in dynamical systems theory is that it is an invariant [15], which is to say that any two dynamical systems (M, f, μ) and (M', f', μ') have the same metric entropy if they are related by a isomorphism that preserves measure. We will not use this fact at all in our entropy calculations for deterministic systems, but when noise is added to the dynamics, we will address the question of how the invariance of the entropy is affected.

We now introduce Lyapunov characteristic exponents as another measure of chaos, and discuss their relationship to the entropies described above. The Lyapunov characteristic exponents measure the average asymptotic divergence rate of nearby trajectories in different directions of a system's state space [18, 19]. For our one dimensional examples, f: $I \to I$, there is only one characteristic exponent λ. It can be easily calculated since the divergence of nearby trajectories is simply proportional to the derivative of f [19]:

$$\lambda = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \log|f'(x_n)|.$$

Or equivalently, if a continuous ergodic invariant measure $\bar{\mu}$ exists, then the characteristic exponent is given by

$$\lambda = \int_0^1 \log|f'(x)| \, \mathrm{d}\bar{\mu}.$$

If M is an axiom-A attractor, there is a prescription for constructing a partition which is generating, and the equality of the metric entropy h_μ and the sum of the positive Lyapunov characteristic exponents can be proven [20]. In fact, whenever an absolutely continuous invariant measure exists, a theorem due to Piesin [21] shows that the metric entropy of a diffeomorphism is equal to the sum of the positive exponents*. Ruelle [22] proved that for any C^2 map that has an absolutely continuous invariant measure

$$h_\mu \leqslant \sum_i \lambda_i^+,$$

where the λ_i^+ are all the positive Lyapunov characteristic exponents, and he has conjectured that equality holds. For a wide class of maps of the unit interval, Ledrappier [23] has shown that an ergodic measure having positive metric entropy is absolutely continuous with respect to Lebesque measure if and only if the metric entropy is equal to the Lyapunov exponent. Shimada [24] obtained good numerical agreement between the characteristic exponent and the metric entropy for the Lorenz attractor and its induced symbolic dynamics using only 9 symbols, and Curry [25] has computed a metric entropy slightly lower than the positive characteristic exponent for a two-dimensional diffeomorphism (Hénon's map)†. Our numerical results for several maps of the unit interval (including the logistic equation) indicate that the metric entropy is indeed equal to the Lyapunov exponent, supporting Ruelle's conjecture and indicating the existence of an absolutely continuous invariant measure whose probability distribution is well approximated by a frequency histogram.

* In the general case, the exponents are a function of initial condition, so the sum must be integrated over the attractor, but we will consider only the case of an ergodic attractor where the exponents are constant almost everywhere with respect to the asymptotic invariant measure.

† Curry's underestimate of the entropy is probably due to the fact that the partition he chose was not generating.

3. Symbolic dynamics and entropy in the presence of noise

One of the reasons that there are so few results on the response of the metric entropy to added fluctuations is that there are problems with the definition of metric entropy (as well as its computation) in the presence of fluctuations. There are also problems with the definition and computation of Lyapunov characteristic exponents for systems with added fluctuations. Some of the problems associated with the metric entropy are:

(1) There is no clear definition of a generating partition for a deterministic system with added noise. Increasingly long sequences of measurements can no longer isolate the system into an arbitrarily fine partition element (where for fineness we mean to use Lebesque measure on the unit interval).

(2) A related problem is that the entropy with respect to a particular partition diverges as the partition is made increasingly fine [26], rendering problematic the definition of a "true" entropy that is independent of partition.

(3) Even using a coarse (e.g. binary) partition, a fixed point with added noise will have nonzero entropy if a partition divider is placed on the fixed point*. (This entropy would then give an estimate of the external noise in the system.)

(4) The effect of adding noise will depend on what coordinate system the noise is added to. One might hope that the response of the metric entropy to noise should be independent in the limit of small noise, but this point is not yet clear from the theory†.

In spite of these problems, we may take a well-defined operational approach to the measurement of metric entropy in the presence of noise: the algorithm embodied in the definitions and estimates yields an unambiguous value of the metric entropy with respect to a particular measurement partition. Any sequence of measurements on a physical system will produce a string of observed symbols; our operational approach will give a measure of the predictability of this string. The measurement partition we will use will be of the same form as that used for the deterministic one-dimensional maps, namely a binary partition of the form $\{[0, d), [d, 1]\}$ where $0 < d < 1$. Given this kind of binary partition, one may again ask if there is a value of d that maximizes the entropy, and we find empirically in fig. 6 that $d = 0.5$ gives a maximum value just as it does for the deterministic case illustrated in fig. 5. This is partial

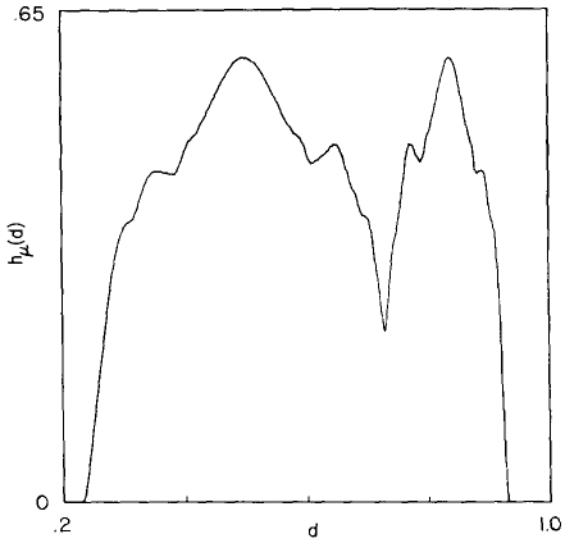

Fig. 6. Entropy h_μ with respect to a binary measurement partition $\{[0, d), [d, 1]\}$ as a function of the decision point d, for the logistic equation with $r = 3.7$ and added noise of width $\sigma = 2^{-7}$. Compare fig. 5.

* This example is due to Doyne Farmer.

† The fact that the observed asymptotic probability distributions will depend on the coordinate system used suggests that if one has some a priori reason for believing the noise to have a particular distribution (e.g. Gaussian), one should, in principle, be able to adjust the coordinate system used to observe the system until the noise displays the correct distribution. An experimentalist's model would thus include the specification of a *physically preferred coordinate system* in which the noise was added. Most systems may be too complicated to give any clue about the "correct" noise distribution, however. For example, in fluid systems with some underlying low-dimensional chaotic attractor, even if we assume that the fluid is being driven by thermal noise, it is not a priori clear what form will be taken by the noise terms added to the equations of motion on the attractor, since the thermal noise will undoubtedly be filtered by many dynamical effects.

justification for our use of this particular measurement partition, but because we are considering only binary partitions, we have not escaped points (1) and (2) above. We must again stress that the metric entropy of a noisy process like eq. (1) depends on the measurement partition used, but we will take the liberty of referring to the "metric entropy" of such a system as that computed using the measurement partition $\{[0, 0.5), [0.5, 1]\}$ unless otherwise noted. Though the entropy h_μ diverges as the measurement partition becomes fine [26], we may still conjecture that for measurement partitions with coarser resolution than the noise level*, our computations give an invariant, well-defined value for h_μ†. To begin the discussion of our numerical computations of the entropy in the presence of noise we will first examine a few properties of the asymptotic probability distributions both of the noisy map on the interval and of the shift on the space of observed sequences Σ_f.

Before considering entropy computation, we will first remark on a few features of the invariant measure, which will in turn have certain implications for entropy measurements. In the deterministic case f: $M \to M$, the asymptotic invariant distribution function $\bar{P}(x)$ is the fixed point of the Frobenius–Perron operator L_f given by

$$(L_f P)(x) = \sum_{y = f^{-1}(x)} \frac{P(y)}{|f'(y)|}.$$

This operator may be written as a Fredholm equation

$$(L_f P)(x) = \int \delta(f(y) - x) P(y)\, dy, \tag{2}$$

where the equivalence is established by integrating the right-hand side using a change of variables $y' = f(y)$.

If noise ξ (with a distribution $P_\sigma(\xi)$ having zero mean and width σ) is added to the deterministic map, forming the noisy map

$$x_{n+1} = f_\xi(x_n) = f(x_n) + \xi_n,$$

an additional average must take place with respect to the noise:

$$\begin{aligned}(L_{f_\xi} P)(x) &= \int \delta(f(y) + \xi - x) P_\sigma(\xi) P(y)\, dy \\ &= \int P_\sigma(f(y) - x) P(y)\, dy. \end{aligned} \tag{3}$$

Thus we see that the deterministic Frobenius–Perron operator is generalized to include the effects of fluctuations by simply replacing the delta function in eq. (2) by the noise distribution function. This formalism has been used by Schraiman, Wayne, and Martin [27], as well as Haken and Meyer-Kress [28], Takahashi [29], and Feigenbaum and Hasslacher [30]. The asymptotic probability distribution for the noisy map is in principle numerically computable using eq. (3). We have not used this expression to compute the distribution (our entropy computations are based on frequency histograms instead), but we may use eq. (3) to infer at least one qualitative property of the asymptotic distribution $\bar{P}(x)$ on the unit interval: Since the distribution must be invariant under the noisy Frobenius–Perron operator, which includes a convolution of the noise distribution, the asymptotic distribution $\bar{P}(x)$ will have no structure on length scales less than the noise level σ.

The primary difference between the symbolic dynamics of a purely deterministic system and that of a deterministic system with added noise is the nature of the identification between a particular symbol sequence and the set of initial conditions that might have produced that sequence. For the deterministic case there is a direct correspondence between symbol sequences and sub-intervals of the

* We mean here that the size of the smallest partition element must be larger than the induced noise level. The induced noise level is obtained from the width of the distribution of the added noise by multiplying this width by the map's maximum slope.

† The well-defined value must still be obtained using a supremum over partitions of a given resolution similar to the supremum illustrated in fig. 6.

unit interval, with the sub-intervals becoming increasingly small as the symbol sequences get longer (cf. the construction of the n-cylinder partition illustrated in fig. 1). When noise is added, instead of there being a sub-interval, every point of which produces a particular sequence, there is a set of points which have some probability of producing a particular sequence. We will label the probability distribution of finding the sequence s^n for an initial condition x as $P_{s^n}(x)$.

For the deterministic case, we have

$$P_{s^n}(x) = X_{\Delta(s^n)},$$

where $X_{[a,b]}$ is the characteristic function over the interval $[a, b]$:

$$X_{\Delta(s^n)}(x) = \begin{cases} 1, & \text{for } x \in \Delta(s^n), \\ 0, & \text{for } x \notin \Delta(s^n), \end{cases}$$

and where $\Delta(s^n)$ is the set of initial conditions that can produce s^n for the deterministic case. We may then use the Frobenius–Perron equation to find $P_{s^{n+1}}(x)$ from $P_{s^n}(x)$. First, for the deterministic case, this gives

$$X_{\Delta(s^n 1)}(x) = \int_{y \in \Delta(s^n)} \delta(f(y) - x)\bar{P}(y)\,\mathrm{d}y.$$

When noise is added, $P_{s^n}(x) = X_{\Delta(s^n)}$ becomes smeared because we must use the noisy Frobenius–Perron operator, which includes a convolution of the noise distribution P_σ:

$$P_{s^{n+1}}(x) = \int P_\sigma(f(y) - x)P_{s^n}(y)\,\mathrm{d}y. \tag{4}$$

The smearing of the partition boundaries, or dividers, that takes place with each application of this operator decreases with successive applications*. The effective width of a partition element increases by $\sigma/|f'(y_i)|^i$, where the y_i are the appropriate inverse images of the deterministic divider. Another way of phrasing this observation is that averaging over fluctuations of width σ at each of n iterations is equivalent, for the purposes of constructing $P_{s^n}(x)$, to averaging over n sets of fluctuations of the initial condition each having a magnitude $\sigma \approx |f'(y_i)|^{-i}$. The convergence of the $P_{s^n}(x)$ to a distribution of a fixed width is illustrated for the logistic equation ($r = 3.7$) in fig. 7a. Note that $\log P_{s^n}(x)$ appears parabolic for large n in the semi-log plots of fig. 7, indicating that $P_{s^n}(x)$ is Gaussian, as might be expected from the repeated convolution of eq. (4).

We see, then, that the picture of bins (elements of an n-cylinder partition) being split into sub-bins (elements of an $(n+1)$-cylinder partition) for the purely deterministic map (cf. fig. 1) is replaced by probability distributions splitting into daughter probability distributions for a deterministic map with added noise. Consider the situation when the width of the distribution $P_{s^n}(x)$ is large compared to the size of the deterministic bin (i.e. the length of $\Delta(s^n)$): Because $P_{s^n}(x)$ converges to a distribution of fixed width for large enough n, daughter distributions have nearly the same width (and in fact nearly coincide), as illustrated in fig. 8. Since the probability of s^n is given by

$$\mu(s^n) = \int P_{s^n}(x)\bar{P}(x)\,\mathrm{d}x,$$

we see that for large enough n†

$$\mu(s^n 1) \approx \mu(s^n 0). \tag{5}$$

* Actually, the width decreases only when the slope evaluated at the appropriate inverse image of the deterministic divider is greater than one. It appears, however (as illustrated in the following figure) that even when there are occasional contributions of slopes less than one, as for the logistic equation, the fact that the "average asymptotic slope" is greater than one causes the width of the distributions $P_{s^n}(x)$ to approach a limit. If the "average asymptotic slope" (this quantity is really well defined only for purely deterministic systems) is less than one (i.e. when the attractor is a periodic orbit) $P_{s^n}(x)$ diverges to cover the entire interval, since in this case all initial conditions end up giving the same periodic symbol sequence.

† We are also using the fact that $\bar{P}(x)$ does not change much over the width of $P_{s^n}(x)$.

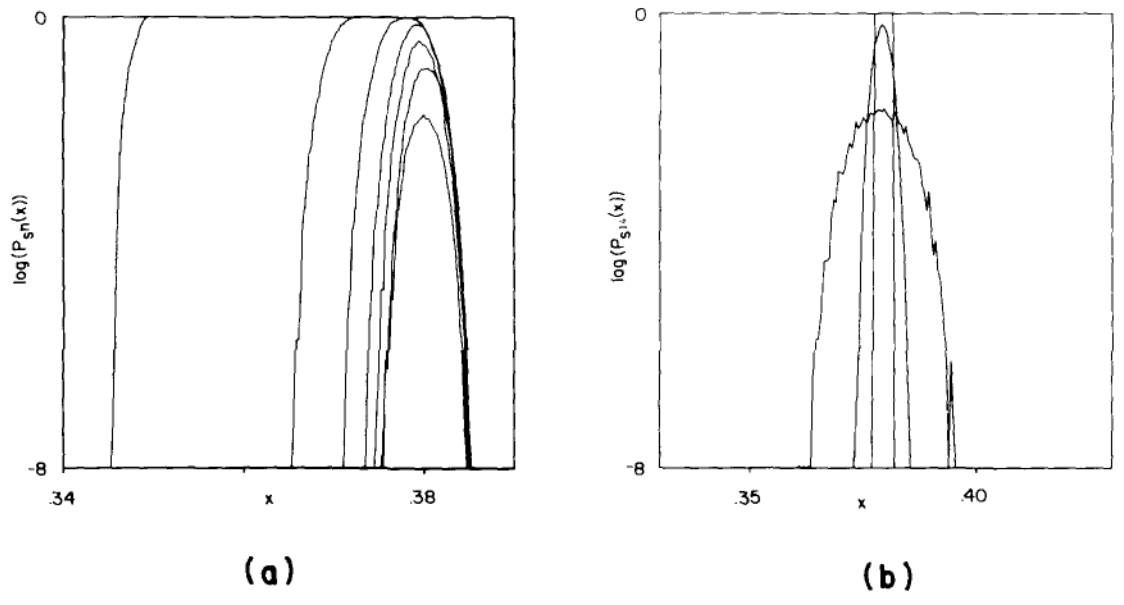

Fig. 7. (a) For a fixed noise level, shown is $P_{s^n}(x)$ for $n = 8$, 9, 11, 13, 15, 16, and 18 (the values of n corresponding to splitting of this particular series of bins in the deterministic case). (b) Fixing $n = 14$, shown is $P_{s^{14}}(x)$ for the deterministic case and for two noise levels $\sigma = 2^{-10}$ and $\sigma = 2^{-7}$. The sequence used was $s^{18} = (010101110111111010)$ (the shorter sequences are truncations: $s^8 = (01010111)$, etc.).

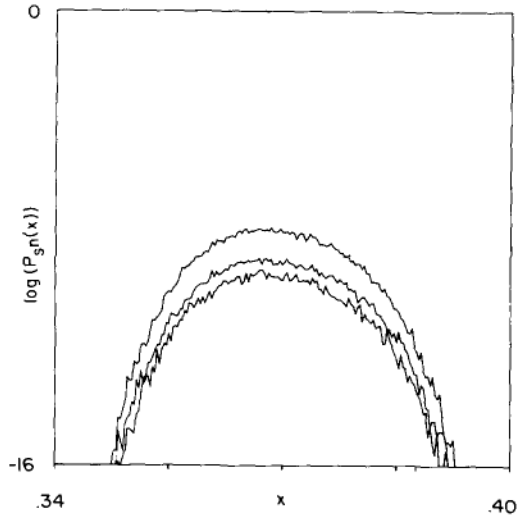

Fig. 8. Splitting of $P_{s^{17}}(x)$ into two daughter distributions $P_{s^{171}}$ and $P_{s^{170}}$. The top distribution is $P_{s^{17}}(x)$, the second distribution is $P_{s^{171}}$, and the third is $P_{s^{170}}$. s^{17} is the same as used in the previous figure.

This condition has some interesting implications that we will now discuss.

$P_{s^n}(x)$ is the distribution of initial conditions that produce the sequence s^n. We may then ask how much information about the initial condition is obtained by observing the sequence s^n, given the asymptotic distribution on the unit interval $\bar{P}(x)$. The appropriate informational measure turns out to be [31, 32]

$$I(s^n) = \int P_{s^n}(x) \log \frac{P_{s^n}(x)}{\bar{P}(x)} \, dx. \tag{6}$$

Then the average information obtained by specifying n symbols is

$$I(n) = \sum_{s^n} \mu(s^n) I(s^n). \tag{7}$$

For n large enough so that the width of $P_{s^n}(x)$ has reached its noisy asymptotic value we may use the conditions $\mu(s^n 1) \approx \mu(s^n 0)$ and $P_{s^n 0}(x) \approx P_{s^n 1}(x)$ to deduce that $I(n) \approx I(n+1)$, which means that for large enough n, observation of additional symbols

gives *no additional information* about the initial condition. Stated another way, in the presence of noise, the *attainable information* about the initial condition reaches some maximum value I_0, which clearly depends on the noise level. The situation is illustrated schematically in fig. 9. For any given noise level, we may augment our conjecture that a well-defined metric entropy exists, and conjecture the existence of *two* well-defined invariant quantities that characterize a deterministic system with noise: h_μ and I_0*.

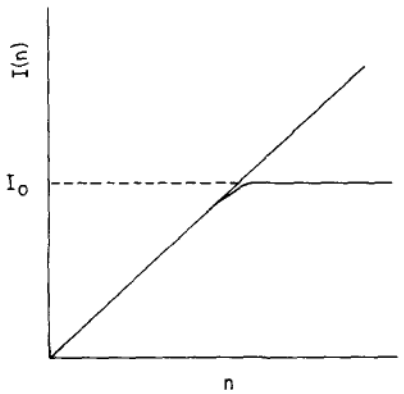

Fig. 9. Schematic illustration of the effect of noise on the attainable information, given by eq. (7), from a sequence of n measurements, or symbols. The straight line corresponds to the deterministic case.

We may also use the condition $\mu(s^n0) \approx \mu(s^n1)$ for n large enough to deduce entropy convergence properties. Convergence of the entropy for finite length n symbol sequences is exactly the condition that the symbolic dynamics be equivalent to an m-state Markov process, where m is the least integer that produces a distribution $P_{s^n}(x)$ satisfying

$$\text{Var } P_{s^m}(x) > \text{Var } X_{\Delta(s^m)}(x).$$

The finite state Markov property insures that the entropy reaches its converged value for $n \gtrsim m$; we will call this phenomenon the *noise floor*, and say that the *convergence knee* occurs at $n = m$. Fig. 10 shows that as the noise level is increased, the convergence knee occurs for smaller values of m. The following section shows how these effects may be described in terms of a scaling theory.

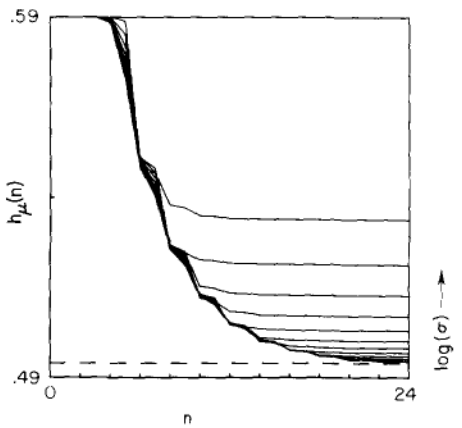

Fig. 10. Entropy convergence of the logistic equation at the parameter value where two bands join to one, $r = 3.67857\ldots$, for increasing noise levels $\sigma = 2^{-18}, \ldots, 2^{-7}$. The Lyapunov characteristic exponent is shown by the dashed line.

4. Scaling properties of entropy measurement

Considering the entropy as a function of both $N = 2^n$ and σ, we may define the normalized *excess entropy*† as

$$\bar{h}_\mu(N, \sigma) = \frac{h_\mu(N, \sigma) - h_\mu(\infty, 0)}{h_\mu(\infty, 0)}.$$

We then find that the data illustrated in fig. 10 displays power law behavior in N:

$$\bar{h}_\mu(N, 0) \sim N^{-\gamma},$$

and power law behavior in σ:

$$\bar{h}_\mu(\infty, \sigma) \sim \sigma^{\beta},$$

* Of course we must still include the proviso that the measurement partition has coarser resolution than the noise level as long as we use our algorithm to compute h_μ. Rob Shaw [3] makes a similar conjecture for an entropy-like quantity computed using a measurement partition with resolution finer than the noise. We will discuss the relationship between these ideas in section 6.

† This can be considered as the *topological pressure* for finite symbol sequences in the presence of noise because we may assume that $h_\mu(\infty, 0) = \lambda$ (cf. discussion above), and in our numerical computations we actually use λ for the value of $h_\mu(\infty, 0)$.

(where ∞ has been approximated by $N = 2^{24}$) [5]. The least squares fits used to estimate the *convergence exponent* γ and the *noise exponent* β are quite good (cf. table I). The scaling in N is visible in the zero noise curve of fig. 10, and the power law increase with noise is illustrated in fig. 11.

Our observation that the metric entropy increases with a power law in response to added fluctuations is reminiscent of some results concerning the response of the Lyapunov characteristic exponent to added fluctuations. It has been shown for maps with a quadratic maximum that at the asymptotic limit of a band merging cascade (i.e. at the onset of chaos) noise added to the dynamics cause a power law increase in the Lyapunov characteristic exponent [4, 27, 33]. The cause of this power law must, however, be fundamentally different from the cause of the power law reported here. Their derivation of power law behavior of the Lyapunov characteristic exponent at band merging cascades (as well as similar results near tangent bifurcations) relies on the change in the attractor's geometry (i.e. the structure of the attractor on the unit interval) as noise is added. Furthermore, only the nearness to crucial bifurcation parameter values allows the change in the attractor's geometry to be systematically described using renormalization group techniques. The power law behavior we describe here appears to hold more generally, including parameter values away from bifurcation cascades, where the geometry of the attractor changes very little with added noise.

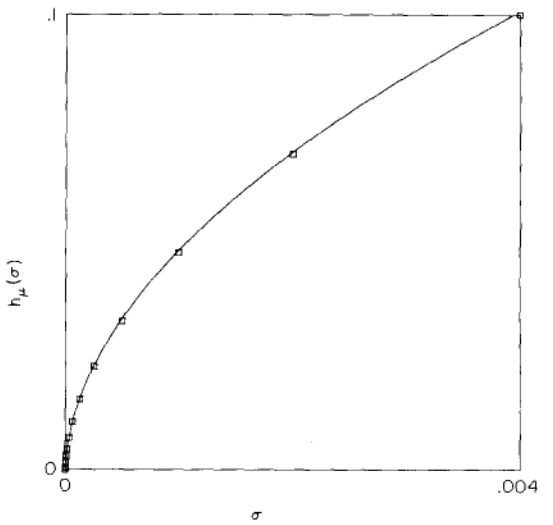

Fig. 11. Power law increase of $\hat{h}_\mu(r = 24, \sigma)$ for the logistic map at $r = 3.67857\ldots$ where two bands join to one.

Table I
Numerical calculations of scaling exponents

System	γ	β	ω
Logistic $r = 3.9$	0.48 ± 0.2	0.56 ± 0.05 (1.0)	0.86
Logistic $r = 3.7$	0.4 ± 0.2	0.53 ± 0.05 (1.3)	0.76
Logistic 2→1 bands	0.38 ± 0.02	0.52 ± 0.01 (0.9)	0.73
Logistic 4→2 bands	—	0.51 ± 0.02 (1.5)	—
Logistic r_c	—	0.345 ± 0.01 (1.0)	—
Logistic 2→1 bands (function space perturbation)	0.38 ± 0.02	0.53 ± 0.01 (1.0)	0.72
Collet and Eckmann map (2→1 band)	0.41 ± 0.1 (0.80)	0.62 ± 0.1	0.66
Tent, $s = 1.43$	0.55 ± 0.1 (0.86)	1.01 ± 0.01 (0.95)	0.55
Tent, 2→1 bands	0.50 ± 0.02 (0.82)	1.06 ± 0.08 (35)	0.47
Tent, 2→1 bands (function space perturbation)	0.51 ± 0.1	1.05 ± 0.08	0.49
Cusp map	—	1.04 ± 0.05 (4.3)	—
Random walk	—	0.92 ± 0.05 (0.05)	—
Toral automorphism	—	0.9 ± 0.02 (0.5)	—

Note: numbers in parentheses are constants of proportionality.

The two numerically observed power laws in N and σ lead us to posit the *scaling hypothesis* that $\bar{h}_\mu(N, \sigma)$ is a homogeneous function of N and σ, namely, that

$$\bar{h}_\mu(\lambda^\gamma N, \lambda^{-\beta}\sigma) = \lambda \bar{h}_\mu(N, \sigma),$$

where λ is an arbitrary change in scale. This sort of scaling hypothesis has been studied extensively in critical phenomena [34], and it is easily shown that the homogeneity of $\bar{h}_\mu(N, \sigma)$ in both variables implies that $\bar{h}_\mu(N, \sigma)$ may be written as a function of a single scaling variable multiplied by a power law. This reduction may be accomplished in two different ways:

$$\bar{h}_\mu(N, \sigma) = \sigma^\beta H(N\sigma^{\beta/\gamma}) \tag{8}$$

or

$$\bar{h}_\mu(N, \sigma) = N^{-\gamma} H'(\sigma N^{-\gamma/\beta}).$$

Since we are interested primarily in the response of $\bar{h}_\mu(N, \sigma)$ to noise, we will concentrate on the first scaling representation of $\bar{h}_\mu(N, \sigma)$.

The scaling hypothesis may be tested empirically (i.e. using data from a numerical simulation) by graphing $\sigma^{-\beta}\bar{h}_\mu(N, \sigma)$ as a function of the scaling variable $N\sigma^{\beta/\gamma}$, and observing whether or not the data lie on a well-defined function $H(N\sigma^{\beta/\gamma})$. The results of this procedure applied to the data shown in fig. 10 are displayed in fig. 12, where we see a convincing numerical verification of the scaling hypothesis. It is interesting to note that while in critical phenomena, scaling often occurs only asymptotically ("asymptotically" means $\sigma \to 0$ or $N \to \infty$ in this context) for this dynamical system we see scaling for *all* σ and N.

We see in fig. 12 that all of the convergence knees are mapped to a single knee of $H(N\sigma^{\beta/\gamma})$. This signals another scaling relation describing the convergence knee, since this implies that the set of (N, σ) for which a convergence knee occurs must satisfy

$$N\sigma^{-\gamma/\beta} = \text{constant}.$$

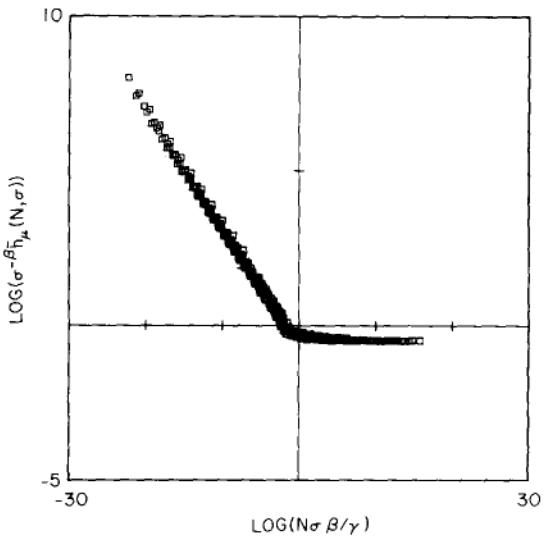

Fig. 12. All the data shown in fig. 10 are replotted here using the homogeneous function representation of eq. (8). The fact that all the data points lie on a well-defined function is verification of the scaling hypothesis.

Either N or σ may be regarded as dependent variables, $N_c(\sigma)$ or $\sigma_c(N)$, in this relation that defines the condition for the occurrence of a convergence knee. And at the convergence knee we may write either

$$N_c(\sigma) \sim \sigma^\omega \quad \text{or} \quad \sigma_c(N) \sim N^{1/\omega},$$

where we find the convergence knee exponent is given by

$$\omega = \gamma/\beta. \tag{9}$$

The same result may be obtained from an eigenvalue equation

$$\frac{\sigma_c(N)}{\sigma_c(2N))} \sim \frac{N_c(2\sigma)}{N_c(\sigma)} \sim \kappa,$$

where $\omega = \log \kappa$.

We have, then, two equivalent interpretations of the eigenvalue κ: First, it is the factor by which the converence knee noise level σ_c must decrease if we are to observe convergence using symbol sequences

of length $n+1=\log(2N)$ rather than of length $n=\log(N)$. Second, if we decrease the noise level by a factor of 2, then κ gives the relative increase in the length of symbol sequence at which we will find convergence. κ will probably be easier to directly measure than the exponents and, at least, it provides a simple way to summarize the net effect of noise on entropy convergence. Values of ω derived from β and γ are tabulated in table I.

The type of relationship between different scaling exponents exemplified by eq. (9) is quite common in the study of critical phenomena. The scaling exponents may be viewed as parameters that describe a surface $\bar{h}_\mu(N,\sigma)$ over the (N,σ) plane. This surface is shown in fig. 13, which illustrates the geometrical significance of the scaling exponents γ, β, and ω.

5. Further numerical experiments

We will now discuss the results obtained from simulating several different systems. For each system we have computed the convergence and noise exponents; the results are tabulated in table I. We will first describe each of the systems studied, then discuss the numerical results.

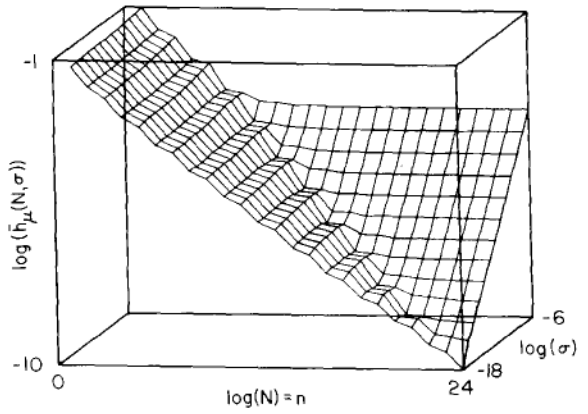

Fig. 13. All the data shown in fig. 10 are replotted as a three dimensional surface. The slope of the line along the front face is $-\gamma$, and the slope of the line along the right face is β. The intersection of these two surfaces defines a line of slope ω in the proper projection onto the scaling variable $N\sigma^\gamma$.

We study four different one-dimensional maps of the unit interval:

(i) the logistic equation, $f(x)=rx(1-x)$ for five parameter values, $r=3.9$, $r=3.7$ (i.e. two "typical" chaotic parameter values), $r=3.67857\ldots$ (where two bands merge into one), $r=3.59257\ldots$ (where four bands merge into two), and $r=3.5699456\ldots$ (the onset of chaos);

(ii) the tent map:

$$f(x)=\begin{cases} sx, & \text{for } 0<x\leqslant 0.5,\\ s(1-x), & \text{for } 0.5<x<1,\end{cases}$$

with $s=1.43$ (a "typical" chaotic parameter value with topological entropy approximately equal to that of the logistic equation at $r=3.7$) and $s=\sqrt{2}$ (the parameter value where two bands join to one);

(iii) Collet and Eckmann's map:

$$f(x)=\begin{cases} 2x, & \text{for } 0<x<0.5-\delta,\\ 1-\dfrac{(x-0.5)^2}{\delta}, & \text{for } 0.5-\delta\leqslant x\leqslant 0.5+\delta\\ 2(1-x), & \text{for } 0.5+\delta<x<1,\end{cases}$$

with $\delta=1/6$ (the parameter value where two bands join to one); and

(iv) the cusp map:

$$f(x)=a(1-|2x-1|^{1+\epsilon}),$$

with $\epsilon=-0.05$ and $a=0.66445776\ldots$ (one of the parameter values where two bands join to one for this ϵ).

We include one two-dimensional system, the linear toral automorphism whose matrix is

$$T=\begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix}.$$

There is a general procedure for constructing a Markov partition (that is generating) for such a map [35], and we use this as the measurement partition.

Another system we study has one of the simplest possible deterministic parts; a random walk on the circle (we have identified the ends of the unit interval to prevent escape of the orbit):

$$f_\xi(x) = x + \xi \text{ [mod 1]},$$

where the circle is coordinatized by the unit interval (with 0 and 1 identified, and the usual measurement partition $\{[0, 0.5), [0.5, 1)\}$ is used to produce the symbol seqquences. The deterministic part of this map is simply the identity, $f(x) = x$, which has no attractor, and which clearly has zero entropy (the symbol sequence is periodic). As noise is added, however, every sequence becomes possible, through for small noise levels, long sequences of 0's and 1's will be most probable. As we have done for maps of the unit interval, we may, for this system, numerically accumulate probability histograms for n-cylinders and compute h_μ as before*. We find that $h_\mu(n)$ converges almost immediately to h_μ, i.e. $h_\mu(2) \approx h_\mu(\infty)$†.

Before discussing the other numerical results contained in table I, we will consider a question concerning the nature of the fluctuations, and how they are coupled to the deterministic system. Eq. (1) represents a very specific model for external fluctuations, namely additive noise. There are, however, many alternatives to perturbing the deterministic function by simply adding noise; one example is multiplicative noise (for the logistic equation this is equivalent to adding noise to the parameter‡). A natural question then arises: which method of adding noise correctly models external fluctuations in a physical system?

Perturbations of a physical system may best be thought of as a perturbation of the dynamics, and not simply a perturbation of the trajectory. A "correct" model for perturbations of a deterministic function $f \in F(M) = \{f\colon M \to M\}$ would choose a function at each time step from an ensemble of functions, with the ensemble centered about the deterministic zero noise limit f. Additive noise simply represents a choice from an ensemble that extends along a one-parameter family of functions $q \to f_q\colon M \to M\colon x \to f(x) + q$. We have modeled the more general case by expanding the function f in a Taylor series (for convenience we will now consider a map on the unit interval: $M = [0, 1]$), and perturb each coefficient separately:

$$f_\xi(x) = (a_0 + \xi_0) + (a_1 + \xi_1)x + (a_2 + \xi_2)x^2 \ldots, \quad (10)$$

where each ξ_i is an independent random variable with zero mean, and where $\{a_i\}$ represent the Taylor coefficients of the deterministic function. For example, when we take the deterministic function to be the logistic map, $f(x) = rx(1 - x)$, the deterministic coefficients are $\{a_0 = 0, \quad a_1 = r, \quad a_2 = -r, \; a_i = 0 \text{ for all } i > 2\}$.

The entries in table I labeled "function space perturbation" represent noise added as in eq. (10) up to sixth order. Comparing the noise exponents for these systems with the noise exponents obtained from simple additive fluctuations, we see agreement to within numerical error. This result gives some confidence that models using additive fluctuations may reflect behavior of physical systems with external fluctuations quite well.

We will now summarize a few interesting aspects of the results listed in table I. Some of the results may be coincidentally similar and lead to erroneous extrapolations. Conjectures based on these results must be verified with further numerical work as well as theoretical progress. The largest error in most of these computations is due to inaccuracy in the estimation of h_μ in the absence of noise; we have assumed the conjecture $h_\mu = \lambda$ (supported by our numerical evidence) and so

* The probabilities of the n-cylinders (and hence h_μ itself) are analytically computable using techniques from the theory of random walks; this calculation will be presented in a future paper.

† This result is similar to the entropy convergence of the toral automorphism.

‡ Crutchfield, Farmer, and Huberman [4] have shown for the logistic equation that for any ensemble of additive fluctuations $\{\xi\}$ there is an equivalent ensemble $\{\xi'\}$ of parametric fluctuations (with a different distribution than that for ξ, in general) that will yield the same time averages over trajectories.

estimate h_μ by the Lyapunov characteristic exponent λ computed using 10^7 iterations (giving an accuracy of $\approx 0.1\%$).

The convergence exponents show no discernable features. For the logistic equation, the convergence exponent decreases from $\gamma = 0.48$ at $r = 3.9$ as the parameter is lowered; at r_c there is no power law convergence. In fact, it is easy to show that at r_c

$$h_\mu(n) \approx \frac{\log n}{n}.$$

The fact that all of the maps at $2 \to 1$ band joining parameter values do not agree in their convergence exponents reveals that the convergence exponent is not constant under topological conjugacy (for all such maps $h = 0.5$). There is no known general technique to compute the convergence exponents, but for special cases (e.g. tent maps at band joinings) γ can be computed exactly to be $\gamma = 0.5$ for $2 \to 1$ band merging [26]. This value agrees extremely well with the numerical value quoted in table I.

Both the random walk and the toral automorphism have $\beta \approx 0.9$ (we see no particular theoretical reason for such a close match). For the logistic map, β decreases from ≈ 0.56 at $r = 3.9$ to ≈ 0.34 at the onset of chaos*, $r_c = 3.5699456\ldots$. For Collet and Eckmann's map at band joinings, we find $\beta \approx 0.6$, indicating that the noise exponent is neither a topological invariant nor universal for quadratic maps.

For all the tent maps simulated, we have $\beta \approx 1.0$, the same value of β as obtained for the cusp map. This leads us to the conjecture: everywhere expanding maps† have a noise exponent $\beta = 1$. There are other reasons for such a conjecture besides the numerical results listed in table I, for instance, the structure of the asymptotic probability distribution on the unit interval. Maps with a critical point (where the slope vanishes) have distributions with infinite singularities, expanding maps do not.

For maps with critical points, these singularities lead to a very non-uniform probability distribution of symbol sequences. In this case, the highly probable sequences are less affected by noise, and do not readily yield new observable sequences. Consequently, the entropy increases more slowly with noise level for maps with critical points. The first class of seven examples in table I with low noise exponents consists of maps with critical points; whereas the maps in the second class of six examples listed in the table have relatively high noise exponents, but no critical points.

6. Lyapunov characteristic exponents and other measures of chaos in the presence of fluctuations

In the context of deterministic systems, we have seen that for an attractor on the unit interval with one positive Lyapunov characteristic exponent λ and an absolutely continuous invariant measure μ [22],

$$h_\mu \leqslant \lambda,$$

and we have presented numerical evidence for equality. The purpose of this section is to see how this kind of result may be generalized to include systems with added fluctuations.

Just as the definition of metric entropy is problematic for systems with added fluctuations, so is the definition of Lyapunov characteristic exponents. For one-dimensional maps, the Lyapunov characteristic exponent can no longer be defined as the average slope of the map because the derivative of the noisy map is not defined. Two approaches to this problem have appeared in the literature. The first technique is to compute Lyapunov characteristic exponents numerically by using the deterministic slope of the map along a noisy trajectory [4, 33, 36]. These computations give quite good results at the asymptotic limit of band merging cascades, where the numerical re-

* This value for β agrees with the power law increase of the Lyapunov characteristic exponent at r_c for the logistic equation [4, 27, 33].

† A map f: $I \to I$ is everywhere expanding if $|f'| > 1$ for all points on the attractor.

sults can be checked against theoretical predictions [27, 33]. This may seem surprising, but the numerical results are probably good for the same reason that the theoretical predictions can be made: at band merging cascades, the response of the Lyapunov characteristic exponent is dominated by the change in the geometrical structure of the attractor* when noise is added.

The second definition of Lyapunov characteristic exponents in the presence of noise is due to Schraiman, Wayne, and Martin [27]:

$$\lambda = \lim_{n\to\infty,\epsilon\to 0} \frac{1}{n} \cdot \log \left| \frac{\langle f_\xi^n(x)\rangle - \langle f_\xi^n(x+\epsilon)\rangle}{\epsilon} \right|,$$

where iteration of the noisy map is given by eq. (1), and where $\langle \ldots \rangle$ denotes an average over the ensemble of noise fluctuations. The noise amplitude must be small enough, and the limits taken carefully for this definition to make sense. When thought of as a measure of the initial spreading rate of two noise distributions whose means are separated by ϵ, this expression for λ is close to a third formulation of Lyapunov characteristic exponents in the presence of noise which we will now discuss (equivalence may eventually be proven).

We have defined h_μ in terms of symbolic dynamics (with a generating measurement partition), but there is another important alternate measure of a system's information generation in terms of the average initial spreading rate of narrow probability distributions†. This formulation has been discussed by Shaw [38]; Farmer, Crutchfield, Froehling, Packard, and Shaw [39]; and Farmer [32]. The spreading rate of sharp distributions is close in spirit to the definition of Lyapunov characteristic exponents (since the spreading of very sharp distributions is governed by the slope of the map) and the correspondence can be made exact for sufficiently simple maps (e.g. piecewise expanding maps). The main reason for discussing this spreading rate here is that it generalizes quite naturally to systems with added fluctuations, and such a measure may in fact be the most appropriate generalization of Lyapunov characteristic exponents for such systems. We will now define the spreading rate and discuss a few qualitative features for different examples, then outline some conjectures relating this picture to the symbolic dynamics quantities already discussed.

As we have noted previously, a one-dimensional map f: $I \to I$ has an associated Frobenius–Perron operator on the space of probability distributions on I given by

$$(L_f P)(x) = \int \delta(f(y) - x) P(y)\, dy.$$

If f has an asymptotic ergodic invariant measure $\bar{\mu}$, then its distribution function $\bar{P}(x)$ must be a fixed point of the operator L_f. Non-equilibrium distribution functions $P(x)$ approach $\bar{P}(x)$ under successive iterations of L_f. The essential idea is to formulate an informational measure of the *rate* that $P(x)$ approaches $\bar{P}(x)$.

To begin, the measure of the amount of information contained in $P(x)$ relative to $\bar{P}(x)$ is

$$\int P(x) \log \frac{P(x)}{\bar{P}(x)}\, dx.$$

Now consider how much information is obtained by making a measurement using a measurement partition $A = \{A_i\}$. If the system is found in the ith partition element, the amount of information obtained is

$$I_i = \int X_{A_i}(x) \log \frac{X_{A_i}(x)}{\bar{P}(x)}\, dx$$
$$= -\log \bar{\mu}(A_i). \qquad (11)$$

* By "geometrical structure," we mean the band like structure of the attractor near r_c [37].

† Here we are identifying a narrow probability distribution with the ensemble of states the system may be in after a (precise but finite) measurement. The time evolution of a sharp distribution is obtained by application of the Frobenius–Perron operator as in eq. (2) (see eq. (3) for the case of added noise). The evolution of a sharp probability distribution is illustrated quite graphically in the movie "Mixing Properties of Strange Attractors," made by Doyne Farmer.

As time passes, if the system is chaotic, the information obtained by the measurement is lost because the distribution X_{A_i} spreads:

$$I_i(t) = \int L_f^t X_{A_i}(x) \log \frac{L_f^t X_{A_i}(x)}{\bar{P}(x)} \, dx .$$

Note that for $t = 0$, this equation reduces to eq. (11). We may now ask for the average information loss after a measurement, where the average is to be taken over all possible initial measurements*

$$\bar{I}(t) = \sum_i \bar{\mu}(A_i) I_i(t).$$

Farmer [32] has given an alternate (equivalent) expression for this quantity:

$$I(t) = \sum_{i,j} \bar{\mu}(f(A_i) \cap A_j) \log \frac{\bar{\mu}(f(A_i) \cap A_j)}{\bar{\mu}(f(A_i)) \bar{\mu}(A_j)}.$$

For a deterministic system we then have the situation illustrated in fig. 14a. A sharp distribution containing a significant amount of information $\bar{I}(0)$ gradually relaxes to the asymptotic distribution, at which point $\bar{I}(t) = 0$ for large enough t†. The slope of $\bar{I}(t)$ well before it goes to zero is then a measure of the loss rate of initial information, which we shall call k_μ. k_μ has been conjectured to be equal to h_μ (Shaw [38]; Farmer, Crutchfield, Froehling, Packard, Shaw [39]; Farmer [32]) for deterministic systems‡.

For contrast, consider the case when the measurement partition is used simply to sample a white noise process. In this case, the probability of any measurement outcome is independent of all previous outcomes, so $\bar{I}(t)$ goes to zero after the first time step, as illustrated in fig. 14b.

When noise is added to a deterministic system, and a measurement partition finer than the noise level is used, we expect $\bar{I}(t)$ to behave something like fig. 14c. Much of the information obtained from an initial measurement using a measurement partition with a typical partition element size smaller than the noise level σ is immediately lost as the sharp probability distribution X_{A_i} spreads out on the first time step into a distribution of width $\approx \sigma$. $\bar{I}_0 \equiv \bar{I}(1)$ then represents the true amount of information that can be obtained from a measurement; using any finer measurement partition can give no more information about the future behavior of the system.

We are now in a position to phrase the conjectures relating this picture to the measurements of chaos using symbolic dynamics: (i) $k_\mu = h_\mu$ and (ii) $\bar{I}_0 = I_0$. The noise level must, of course be small enough so that there is some time interval for which $\bar{I}(t)$ displays a well defined constant slope. Numerical experiments are underway to check these conjectures.

7. Concluding comments

The effects of fluctuations added to chaotic deterministic dynamical systems reveal the concept of "infinitely precise points" as invalid in many contexts. A new mathematical foundation of classical mechanics is needed; one that uses primitives derived from noise processes. Ruelle [42] has made significant progress in this direction. Though the inclusion of fluctuations in a dynamical model adds many analytical complications to a subject already incompletely understood, there is hope, based on physical observations and numerical computations, that there may be several rewarding

* Note that $\bar{I}(t)$ must be distinguished from $I(n)$ defined in eq. (7); $I(n)$ is the rate that information (with respect to the previous $n-1$ symbols) is acquired with the observation of new symbols, and $I(t)$ is the average rate that information contained in an initial condition (using a particular measurement partition) is lost.

† This is actually a crude picture with details which may change for different systems; e.g.: (i) Phase coherent attractors have $I(t) > 0$ as $t \to \infty$ (cf. Farmer et al. [39]); (ii) Rob Shaw [3] has pointed out that for maps with a critical point the initial slope of $I(t)$ will be larger than h_μ, and then decrease to h_μ.

‡ Goldstein and Penrose [40] have introduced a similar information loss rate which, for certain systems, Goldstein [41] proved to be equal to the metric entropy.

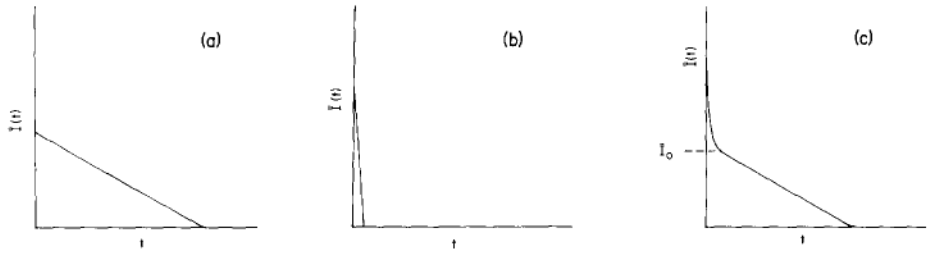

Fig. 14. (a) Schematic representation of $\bar{I}(t)$ for a deterministic system. For certain systems (e.g. one dimensional maps with a critical point) the slope of $I(t)$ will be greater than k_μ at $t = 1$. (b) Schematic representation of $\bar{I}(t)$ for measurements of a white noise process. (c) Schematic representation of $I(t)$ for a deterministic system with added noise, where a measurement partition finer than the noise level has been assumed.

simplifications lurking in the theory. Assuming such a theory may be formulated, most of the numerical results presented here should be consequences of the theory, so they will hopefully point the direction for some future theoretical developments. We will now review our results in this light.

For a chaotic deterministic system, successive measurements (using a "good" measurement partition) pinpoint the initial condition whose orbit produced the observations with arbitrary accuracy (i.e. an arbitrarily large amount of information about the initial condition may be obtained from an arbitrarily long sequence of measurements). When noise is added to the deterministic dynamics, we have observed that the initial condition may be specified only to within some uncertainty, even with an arbitrarily long sequence of measurements. This has led to the proposal that a chaotic system with added fluctuations is characterized by *two* invariant quantities: (i) I_0: the maximum average information (about the initial condition) obtainable from a sequence of measurements; and (ii) h_μ: the average information generation rate (simply the metric entropy in the case of a deterministic system)*.

* We have also conjectured these two quantities to be equal to $\bar{I}_0$ and k_μ, the maximum amount of information that can be stored in an initial condition, and the average loss rate of information after a measurement, respectively.

I_0 has not been computed numerically yet, but the information production rate h_μ (with respect to a given measurement partition) is easily computed using the same algorithms used to compute h_μ for deterministic systems. Upon pursuing the question of how h_μ depends on the fluctuations added to the deterministic dynamics, we find that h_μ increases with a power law in the noise level σ: $h_\mu \approx \sigma^\beta$. We have found that this power law increases seems to happen very generally (for all systems studied here). The exact value of the noise exponent β varies with the system under study, though our numerical experiments have led to the conjecture that a wide class of systems (those reducible to a one-dimensional map f: $I \to I$ with $|f'| > 1$) has a noise exponent $\beta = 1$. We have combined the power law response of h_μ with the power law convergence of the entropy as a function of the number of symbols observed, to form a homogeneous function description of entropy measurement. In this context, a scaling hypothesis has been verified numerically.

The power law increase in the metric entropy may be regarded as the discovery of a new phenomenon, an observable feature of the information production properties of any physical system that can successfully be modeled by a low dimensional chaotic dynamical system coupled to external fluctuations. There is a growing body of very good experimental evidence that supports such a model; convincing one-dimensional return maps have

been obtained for fluid systems and for chemical systems*.

The noise exponent should be measurable, given reasonable experimental accuracy, though we have no prediction for its value if the one-dimensional map that underlies the observed behavior has a critical point. So far, all the return maps constructed from experimental data appear to have a critical point (or several critical points). There are, however, many physical systems that should be describable by a one-dimensional return map whose slope (absolute value) is always greater than one. One example would be a Benard convection fluid system constrained to excite only those modes described by the Lorenz equations, which have a cusp-like one-dimensional return map. For these systems, we might expect a noise exponent of $\beta = 1$.

Fluctuations are now generally recognized as the source of much of the diverse complexity we see in the world around us (especially in the biosphere). It has been hypothesized (by R. Shaw [46], for example) that what we call "diverse complexity" is a result of intrinsic dynamical properties of some (complicated) dynamical system, in particular, of the system's information generating properties. The informational properties of most of the dynamical systems underlying and producing this complexity are, however, poorly understood. One example of how the current picture of information generation in chaotic dynamical systems must be generalized, is that unlike the chaotic systems studied here, the information generated by the dynamics of complicated evolving systems like the biosphere is stored in physical structures, which then serve as the base for even more complicated evolution. There are many other similar problems to be faced, but the results presented here will hopefully serve as a starting point for the study of the role fluctuations will play in the context of these more complicated systems.

* Cited here are "non-trivial" physical systems in which one might not naively expect to see low-dimensional chaos because of the many degrees of freedom that could potentially participate in the dynamics. Return maps have, of course, been successfully constructed for much simpler physical systems (e.g. electrical oscillator circuits) in which low-dimensional chaos is expected (cf. Crutchfield [43]; Packard, Crutchfield, Farmer, Shaw [44]; Gollub, Romer, and Socolar [45]) because of the few degrees of freedom involved.

Acknowledgements

The authors wish to thank Doyne Farmer, David Fried, John Gukenheimer, and Rob Shaw, for helpful discussions, and the Center for Nonlinear Studies, Los Alamos National Laboratory, for the warm hospitality offered during the time this work was written up. This work was supported in part by NSF grant 443150-21299.

References

[1] E.N. Lorenz, J. Atmos. Sci. 20 (1963) 130.
[2] D. Ruelle and F. Takens, Comm. Math. Phys. 20 (1974) 167.
[3] For example: J.L. Hudson and J.C. Manken, J. Chem. Phys. 74 (1981) 6171. J.C. Roux, A. Rossi, S. Bachelart and C. Vidal, Phys. Lett. 77A (1980) 391; R. Shaw "The Dripping Faucet as a Model Chaotic System", UCSC Preprint (1982). A. Libchaber (this volume); and H. Hauke and Y. Maeno (this volume).
[4] J.P. Crutchfield and B.A. Huberman, Phys. Lett. 77A (1980) 407; J.P. Crutchfield, J.D. Farmer and B.A. Huberman, "Fluctuations and Simple Chaotic Dynamics", Physics Reports 92 (1982) 45.
[5] J.P. Crutchfield and N.H. Packard, Int. J. Theo. Phys. **21** (1982) 433; and J.P. Crutchfield and N.H. Packard, in Evolution of Ordered and Chaotic Patterns, H. Haken, ed. (Springer, Berlin, 1982).
[6] Ju. I. Kifer, USSR Izvestija 8 (1974) 1083.
[7] M. Boyarsky, Trans. Am. Math. Soc. 257 (1980) 350.
[8] Y. Oono and M. Osikawa, Prog. Theo. Phys. 64 (1980) 54.
[9] J. Milnor and W. Thurston, "On Iterated Maps of the Interval, I and II", Princeton University preprint (1977).
[10] C.E. Shannon and W.E. Weaver, The Mathematical Theory of Communication, (Univ. of Illinois Press, Urbana, Illinois, 1949).
[11] R. L. Adler, A.G. Konheim and M.H. McAndrew, Trans. Am. Math. Soc. 114 (1965) 309.
[12] W. Parry, Trans. Am. Math. Soc. 112 (1964) 55.
[13] J.P. Crutchfield and R. Shaw (unpublished). See J.P. Crutchfield, UCSC Ph.D. dissertation (1983).
[14] P. Collet, J.P. Crutchfield, J.-P. Eckmann, "On Computing the Topological Entropy of Maps", to appear Comm. Math. Phys. (1983).

[15] A.N. Kolmogorov, Dokl. Akad. Nauk. SSSR 119 (1958) 861; Dokl. Akad. Nauk. SSSR 124 (1959) 754.
[16] L.W. Goodwyn, Proc. Am. Math. Soc. 23 (1969) 697.
[17] E.I. Dinaburg, Sov. Math. Dokl. 11 (1970) 13.
[18] G. Benettin, L. Galgani, A. Giorgilli and J.-M. Strelcyn, Meccanica (1980) 21, and C.R. Acad. Sc. Paris 286 (1978) 431. I. Shimada and T. Nagashima, Prog. Theo. Phys. 61 (1979) 1605.
[19] R. Shaw, Zeit. fur Naturfor. 36a (1981) 80.
[20] R. Bowen and D. Ruelle, Inven. Math. 29 (1975) 181.
[21] Ya. B. Piesin, Uspek. Math. Nauk. 32 (1977) 55.
[22] D. Ruelle, Boll. Soc. Brasil. Math. **9** (1978) 331; see also D. Ruelle, Comm. Math. Phys. 55 (1977) 47.
[23] F. Ledrappier, Ergod. Theo. Dyn. Sys. 1 (1981) 77.
[24] I. Shimada, Prog. Theo. Phys. 62 (1979) 61.
[25] J.H. Curry, "On Computing the Entropy of the Henon Attractor", Institut des Hautes Etudes Scientifique, Burr-sur-Yvette, preprint (1981).
[26] N.H. Packard, UCSC Ph.D. Dissertation (1982).
[27] B. Schraiman, C.E. Wayne and P.C. Martin, Phys. Rev. Lett. 46 (1981) 935.
[28] H. Haken and G. Mayer-Kress, Z. Physik B **43** (1981) 185; Phys. Lett. 84A (1981) 159.
[29] Y. Takahashi, "Observable Chaos and Variational Principle for One-Dimensional Maps", Dept. Math., Coll. Gen. Educ., University of Tokyo, preprint (1981).
[30] M. Feigenbaum and B. Hasslacher, "Irrational Decimations and Path Integrals for External Noise", Phys. Rev. Lett. 49 (1982) 605.
[31] F. Schlogl, Z. Physik 243 (1971) 303.
[32] J.D. Farmer, "Information Dimension and the Probabalistic Nature of Chaos", Zeit. fur Naturfor. 37a (1982) 1304. See also J.D. Farmer, UCSC Ph.D. Dissertation (1981).
[33] J.P. Crutchfield, M. Nauenberg and J. Rudnick, Phys. Rev. Lett. 46 (1981) 933.
[34] H.E. Stanley, Introduction to Phase Transitions and Critical Phenomena (Oxford Univ. Press, New York, 1971).
[35] R.L. Adler and B. Weiss, Mem. Am. Math. Soc. 98 (1970) 1.
[36] H. Haken and G. Mayer-Kress, J. Stat. Phys. 26 (1981) 149.
[37] E.N. Lorenz, New York Acad. Sci. Annals 357 (1980) 282. J.P. Crutchfield, J.D. Farmer, N.H. Packard, R. Shaw, R.J. Donnelly and G. Jones, Phys. Lett. 76A (1980) 1.
[38] R. Shaw, UCSC Ph.D. Dissertation (1980).
[39] J.D. Farmer, J.P. Crutchfield, H. Froehling, N.H. Packard and R. Shaw, New York Acad. Sci. Annals 357 (1980) 453.
[40] S. Goldstein and O. Penrose, J. Stat. Phys. 24 (1981) 325.
[41] S. Goldstein, Israel. J. Math. 38 (1981) 241.
[42] D. Ruelle, Comm. Math. Phys. 82 (1981) 137.
[43] J.P. Crutchfield, UCSC Senior Thesis (1979).
[44] N.H. Packard, J.P. Crutchfield, J.D. Farmer and R. Shaw, Phys. Rev. Lett. 45 (1980) 712.
[45] J.P. Gollub, E.J. Romer and J.E. Socolar, J. Stat. Phys. 23 (1980) 321.
[46] R. Shaw, cover letter for Louis Jacot Prize Competition, Paris (1977).