Independent coordinates for strange attractors from mutual information

Andrew M. Fraser and Harry L. Swinney
Department of Physics, University of Texas, Austin, Texas 78712
(Received 22 July 1985)

The mutual information I is examined for a model dynamical system and for chaotic data from an experiment on the Belousov-Zhabotinskii reaction. An $N \log N$ algorithm for calculating I is presented. As proposed by Shaw, a minimum in I is found to be a good criterion for the choice of time delay in phase-portrait reconstruction from time-series data. This criterion is shown to be far superior to choosing a zero of the autocorrelation function.

In recent years much progress has been made in understanding low-dimensional chaos in fields as diverse as hydrodynamics,[1] epidemiology,[2] chemistry,[3] and solid-state physics.[4] Insight into chaotic behavior has been achieved by quantitatively characterizing the post-transient phase-space orbits ("strange attractors") in terms of fractal dimension, metric entropy, and the spectrum of Lyapunov exponents. A crucial development responsible for much of the recent progress was the realization that *multi*dimensional phase portraits could be constructed from measurements of a *single* scalar time series.[5,6] Portraits are constructed by expanding a scalar time series $s(t)$ into a vector time series $\mathbf{X}(t)$ using time delays T: $\mathbf{X}(t)=\{x_0(t),x_1(t),\ldots,x_n(t),\ldots\}$, where $x_n(t)=s(t+nT)$. For an infinite amount of noise-free data, the time delay T can in principle be chosen almost arbitrarily.[6] However, experiments[3] show that the quality of the portraits depends on the value chosen for T, and experimenters[7] and theorists[8] note that there are no criteria for choosing T in the literature. In this paper we present mutual-information calculations for data from strange attractors in the Belousov-Zhabotinskii reaction and the Rössler system[9] and explain why such calculations provide an excellent criterion for choosing T in most systems. Also, we propose that mutual information could provide a quantitative characterization of chaotic spatial patterns.

If delay coordinates are used with a very small T for a two-dimensional reconstruction in the presence of experimental noise, $x_0(t)$ and $x_1(t)$ will be indistinguishable, and all trajectories will appear to lie on the line $x_0=x_1$. To avoid this, the chosen T should make x_0 and x_1 independent. A naive choice of T, which is equivalent to requiring linear independence, is the value for which the autocorrelation function first passes through zero.

Mutual information, to be defined precisely later, measures the *general* dependence of two variables; therefore, it provides a better criterion for the choice of T than the autocorrelation function, which only measures *linear* dependence. Shaw has suggested[10] that the value of T that produces the first local minimum of mutual information be used for phase portraits. The two criteria are compared in Fig. 1. The phase portrait on the left, constructed for T corresponding to the first zero in the autocorrelation function, has high mutual information because most of the probability lies in the dark line along the upper left edge. Much of the stretching and folding occurs in this region where the trajectories are experimentally indistinguishable; hence, it is difficult to deduce quantitative information about the dynamics from this phase portrait. In contrast, in the phase portrait on the right, which corresponds to a minimum of mutual information, the beginning of a fold can be seen at the bottom of the central band; the chaotic dynamics can be qualitatively and quantitatively deduced from this portrait. The figure demonstrates that mutual information can be very different from autocorrelation, and that mutual information selects superior T values.

We now consider a model system, the Rössler attractor,[9] because the original phase space is available and the data are precise and easy to obtain. Its representation in the original phase space will be compared to delay reconstructions from simulated experimental measurements of a single variable. The system is defined by

$$\begin{aligned} \dot{x} &= -z-y \ , \\ \dot{y} &= x+ay \ , \\ \dot{z} &= b+z(x-c) \ . \end{aligned} \tag{1}$$

We examine this model for parameter values $a=0.15$, $b=0.20$, and $c=10.0$, considering a trajectory generated using the Runge-Kutta method with a fixed time step of $\pi/100$. The trajectory was integrated from an initial condition of (10,0,0); after discarding the first 1000 steps to allow the trajectory to fall to the attractor, a file of 1 048 576 points was recorded. Figure 2(a) shows part of this trajectory. Over the entire file the average time for one orbit was found to be 193.3 integration steps. Henceforth, time for this model will be given in terms of this average orbital time. Experimental measurements were simulated by adding noise to the x values of the file each time they were read. The noise was flat over ± 0.1 while the rms deviation of x was 11.7, yielding a signal-to-noise ratio of 46 dB. In Fig. 2 the plots in the first row are derived from the original x and y coordinates, while the plots in the second and third rows are derived from simulated measurements for different time delays.

An isolated x measurement constrains the system to a stripe of points in phase space, as illustrated in Fig. 2(b). A good second measurement is one that provides new in-

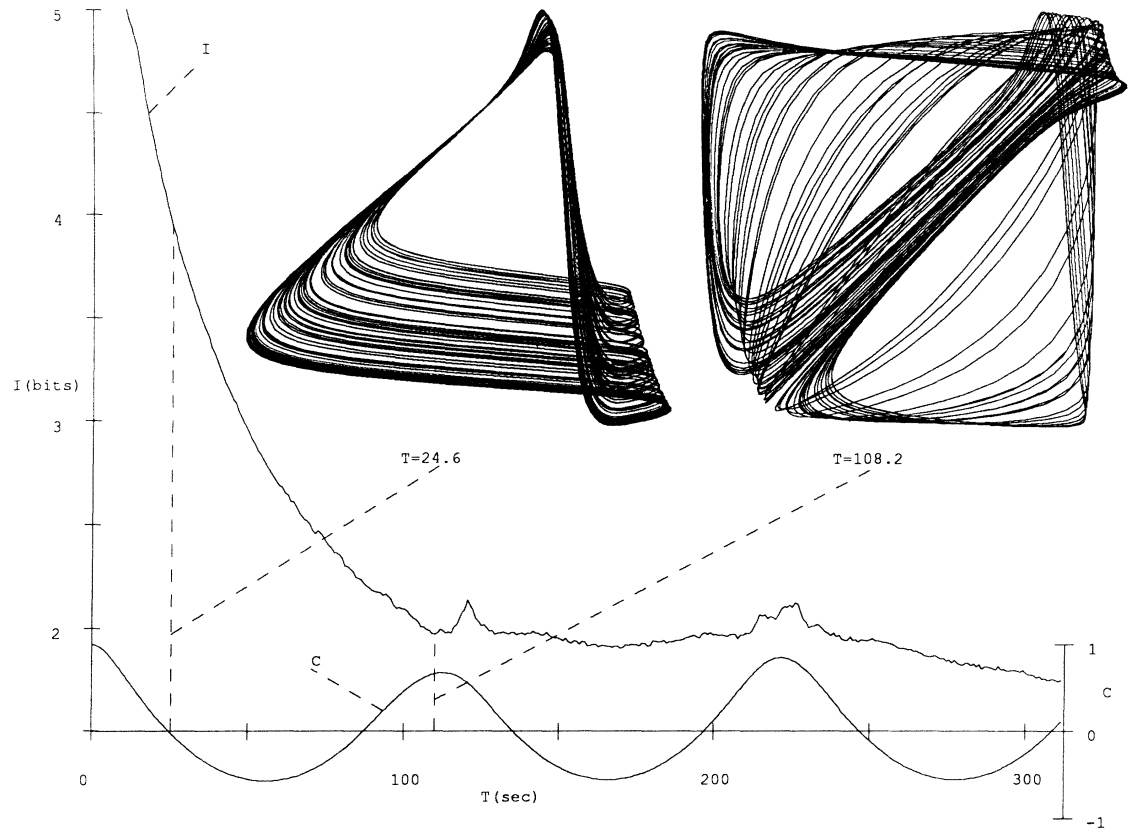

FIG. 1. Phase portraits of the Roux attractor (Ref. 3) in the Belousov-Zhabotinskii reaction. The dependence of the mutual information I and the autocorrelation function C on T are shown for calculations over 32 768 points. The coordinates used in constructing the portrait on the left are linearly independent (zero autocorrelation), while the coordinates used in the portrait on the right are more generally independent (local minimum of mutual information).

formation about the system state by dividing this stripe into as many parts as possible. Figure 2(c) indicates that $T=0.23$ orbits roughly corresponds to a coordinate rotation of 80°. Such a rotation provides a good second measurement. Figure 2(e) shows that a later measurement at $T=3.26$ orbits is also good at subdividing the stripe; however, the folds in the later measurement can never be resolved for data of finite resolution.

The goal is to choose a good delay without looking at the original phase space. The key to such a choice lies in the histograms of the second row of Fig. 2. Good delays correspond to flat histograms, which in turn roughly correspond to small values of the mutual information I. In a chaotic system (one with positive metric entropy) any measurement stripe will eventually spread back to the invariant measure. This is accomplished by stretching and folding like that in Fig. 2(e). To avoid this type of spreading, earlier flat histograms are preferred to later ones; hence, the *first* local minimum of I is preferred to later minima. It is possible that for some systems the spreading could be fast enough to dominate the rotation, defeating our method.

Shannon's information theory[11] provides a formalism for quantifying the concepts of spreading and new information. For information theory to apply, the probabilities of the messages considered must exist, and to use the theory the probabilities must be accessible. We are applying information theory to strange attractors, and the messages we are considering are the values that measurements of the attractors might take. Strange attractors are ergodic and have well-defined asymptotic probability distributions. Thus the probabilities of the messages we consider exist, and long-time averages converge to the probabilities.

Information theory is usually discussed in terms of a signaling system. Consider a process in which messages are sent to an experimenter across the channel of his instruments. Let S denote the whole system which consists of a set of possible messages $s_1, s_2, \ldots, s_n$, and the associated probabilities $P_s(s_1), P_s(s_2), \ldots, P_s(s_n)$. P_s maps messages to probabilities. The subscript is necessary because more than one such function will be considered at a time. If the possible messages are continuous, S denotes the system, s denotes a possible message, and $P_s(s)$ is the probability density at s.

The average amount of information gained from a measurement that specifies s is the entropy H of a system,

$$H(S) = -\sum_i P_s(s_i) \log P_s(s_i) \ . \tag{2}$$

$H(S)$ is the quantity of surprise you should feel upon reading the result of a measurement. If the log is taken to the base two, H is in units of bits. In Fig. 2(a) the proba-

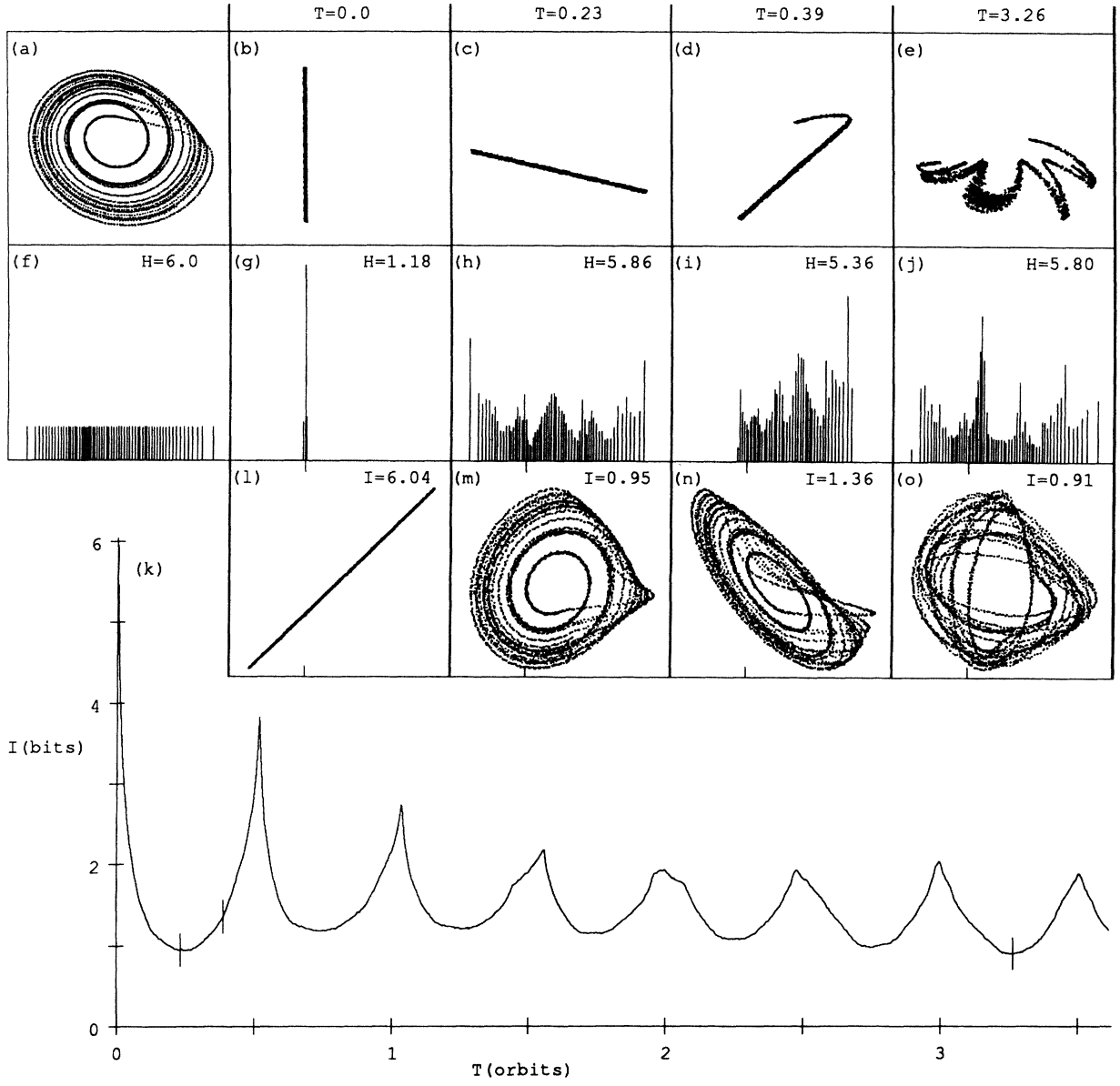

FIG. 2. Rotation and spreading in the Rössler attractor. The entire attractor projected on the original x,y coordinates is shown in (a), and (f) shows a histogram of x measurements for this attractor for 64 equiprobable bins. The measurements corresponding to bin 20 are shown in (b), and the corresponding histogram is (g); figures (c), (d), and (e) show these points at later times, and the corresponding histograms are (h), (i), and (j), respectively. For short times the points are essentially only rotated, as (c) illustrates, while for long times there is spreading as well as rotation, as (e) illustrates. Phase portraits constructed by the time delay method for the delays in (b)—(e) are shown in (l)—(o), respectively, where the tick marks indicate bin 20. [Histograms (g)—(j) are given by the probability densities between the tick marks.] The mutual information (calculated over 65 536 points) is shown as a function of T in (k).

bility of an isolated measurement being in any one of the 64 bins is $\frac{1}{64}$ [shown in Fig. 2(f)] and H is $-\log_2(\frac{1}{64})=6$ bits. If the grid consisted of 128 equiprobable bins, H would be 7 bits. This dependence of H on grid is related to the fact that if s were a continuous variable, the value of H would depend on the coordinates. While the limit of the sum for the discrete case diverges as the partition becomes finer, the integral

$$H(S)=-\int P_s(s)\log P_s(s)ds \tag{3}$$

does not diverge, but its value depends on the coordinates chosen. The reason is that the argument of the log function in (3) has the same units as $1/ds$. Individual entropies of continuous systems depend on coordinates, but we are interested in a coordinate-independent difference of entropies. So we will continue our development with the discrete case and then consider the continuous limit of our result.

We are interested in measuring how dependent the values of $x(t+T)$ are on the values of $x(t)$. By making

the assignment $[s,q]=[x(t),x(t+T)]$, we can consider a general coupled system (S,Q) and ask, "Given that s has been measured and found to be s_i, what uncertainty is there in a measurement of q?". The answer is

$$H(Q\mid s_i)=-\sum_j P_{q\mid s}(q_j\mid s_i)\log[P_{q\mid s}(q_j\mid s_i)] \tag{4}$$

$$=-\sum_j[P_{sq}(s_i,q_j)/P_s(s_i)]\times\log[P_{sq}(s_i,q_j)/P_s(s_i)]\ , \tag{5}$$

where $P_{q\mid s}(q_j\mid s_i)$ is the probability that a measurement of q will yield q_j, given that the measured value of s is s_i. Figure 2(h) shows the probability distribution of x in the Rössler system after a time of 0.23 orbits, given that x began in bin 20. The conditional entropy is $H(Q\mid s_{20})=5.86$ bits.

The next question is, "Given that x has been measured at time t, what is the average uncertainty in a measurement of x at time $t+T$?". The answer is given by averaging $H(Q\mid s_i)$ over s_i, which yields

$$H(Q\mid S)=\sum_i P_s(s_i)H(Q\mid s_i)$$

$$=-\sum_{i,j}P_{sq}(s_i,q_j)\log[P_{sq}(s_i,q_j)/P_s(s_i)]$$

$$=H(S,Q)-H(S)\ , \tag{6}$$

where

$$H(S,Q)=-\sum_{i,j}P_{sq}(s_i,q_j)\log[P_{sq}(s_i,q_j)]\ . \tag{7}$$

$H(Q)$ is the uncertainty of q in isolation, and $H(Q\mid S)$ is the uncertainty of q given a measurement of s. So the amount that a measurement of s reduces the uncertainty of q is

$$I(Q,S)=H(Q)-H(Q\mid S)$$

$$=H(Q)+H(S)-H(S,Q)=I(S,Q)\ . \tag{8}$$

This is the *mutual information*. It is the answer to the question, "Given a measurement of s, how many bits on the average can be predicted about q?". If S and Q are continuous,

$$I(S,Q)=\int P_{sq}(s,q)\log[P_{sq}(s,q)/P_s(s)P_q(q)]\,ds\,dq\ . \tag{9}$$

The argument of the log in (9) is dimensionless so the integral is *independent* of the coordinates chosen. It is important to remember that mutual information is not a function of the variables s and q, but that it is a functional of the joint probability distribution P_{sq}. It is a global measure of the bumpiness of P_{sq}. If s and q are the same to within the noise, then $I(S,Q)$ specifies the relative accuracy of the measurements in bits, i.e., how much information one measurement gives about a second measurement of the same variable.

If Q is a delayed image of S, then a delay phase portrait gives the estimated joint distribution P_{sq}, and I is a statistic calculated on the portrait that evaluates how redundant the second axis is. The values of I for four phase portraits are shown in Fig. 2. The first minimum in I yields the best choice of time delay, $T=0.23$ orbits.

The principal difficulty in calculating mutual information from experimental data is in estimating P_{sq} from histograms. If a box in the (s,q) plane of size $\Delta s\Delta q$ has N_{sq} points in it, we estimate P_{sq} to be $N_{sq}/N_{\text{total}}\Delta s\Delta q$ uniformly across the box. Choosing any box size has advantages and liabilities. For a given number of data, larger boxes have more points, and hence the estimate of the average probability is more accurate, but the estimates of P_{sq} are too flat, underestimating $I(S,Q)$. Smaller boxes let one follow changes in P_{sq} over short distances, but allow the fluctuations that are due to small sample size to be interpreted as small-scale structure in P_{sq}, overestimating $I(S,Q)$. No single box size is best over the whole (s,q) plane.

We have developed an algorithm that covers the (s,q) plane with a partition in which the size of each element is tailored to the local situation. We define a sequence of partitions of the (s,q) plane, $G_0,G_1,G_2,\ldots,G_m,\ldots$, such that each partition is a rectangular grid of 4^m elements generated by dividing each axis into 2^m equiprobable segments. $R_m(K_m)$ denotes an element of G_m, and K_m is an index that takes one of 4^m possible values. Applying such a partition to a continuous probability density function P_{sq} produces a discrete probability mass function $P_{sq}(R_m(K_m))$. Since the elements are rectangular it makes sense to also define $P_s(R_m(K_m))$ and $P_q(R_m(K_m))$. Associated with the sequence of partitions is a sequence $i_0,i_1,i_2,\ldots$, which converges to $I(S,Q)$, where

$$i_m=\sum_{K_m}P_{sq}(R_m(K_m))\times\log[P_{sq}(R_m(K_m))/P_s(R_m(K_m))P_q(R_m(K_m))]\ . \tag{10}$$

One representation of K_m is ordered pairs $K_m=(i,j)$, where i indicates a range of s values, and j indicates a range of q values; see Fig. 3. With this notation and

$$P_s(R_m(i,j))=P_q(R_m(i,j))=(\tfrac{1}{2})^m\ ,$$

$$i_m=m\log(4)+\sum_{i,j=0}P_{sq}(R_m(i,j))\log[P_{sq}(R_m(i,j))]\ . \tag{11}$$

The algorithm is recursive and requires a recursive approach to the definition of $I(S,Q)$, so we organize G_m as a tree rather than as a matrix. To get G_1 from G_0, $R_0(K_0)$ is divided into four subelements $R_1(0)$, $R_1(1)$, $R_1(2)$, and $R_1(3)$. To get G_2, each element of G_1 is divided into four subelements $\{R_2(k_1,k_2)$: $k_1,k_2\in\{0,1,2,3\}\}$; see Fig. 3. In this fashion each element of G_m can be specified by $R_m(k_1,k_2,\ldots,k_m)$. The elements of the partition are the same as those in the matrix organization, but the 4^m values of K_m are represented as m tuples rather than as Cartesian pairs.

We now look at what happens to the element $R_m(K_m)$ of G_m in going from m to $m+1$. Here K_m is a particular m tuple $(k_1,k_2,\ldots,k_m)$. Define C_T as the contribution to i_m of $R_m(K_m)$ and C_j as the contribution to i_{m+1} of the subelement $R_{m+1}(K_m,j)$. Then we have

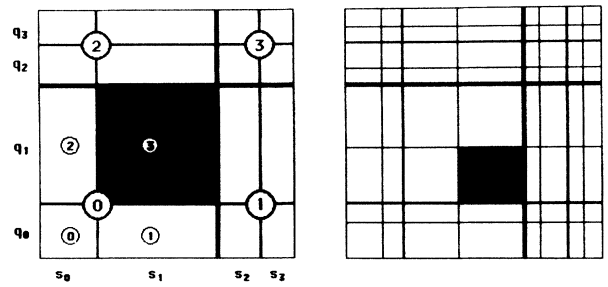

FIG. 3. Two steps in the sequence of partitions. The shaded element of G_2 is labeled $R_2(1,1)$ in the matrix notation and $R_2(0,3)$ in the tree notation. In G_3 the shaded element is labeled $R_3(3,2)$ in the matrix notation and $R_3(0,3,1)$ in the tree notation.

$$C_T = P_T m \log(4) + P_T \log(P_T) , \tag{12}$$

where $P_T = P_{sq}(R_m(K_m))$. Similarly,

$$C_j = P_j(m+1)\log(4) + P_j \log(P_j) , \tag{13}$$

where $P_j = P_{sq}(R_{m+1}(K_m,j))$. The contribution of the area covered by $R_m(K_m)$ to i_{m+1} is

$$\sum_{j=0}^{3} C_j = P_T(m+1)\log(4) + \sum_{j=0}^{3} P_j \log(P_j) . \tag{14}$$

Notice that to go from level m to level $m+1$ [from (12) to (14)] a term $P_T \log(4)$ is added and the term $P_T \log(P_T)$ is replaced with $\sum P_j \log(P_j)$. This is the basic recursion step. Also notice that if P_{sq} is flat over $R_m(K_m)$, then $P_0 = P_1 = P_2 = P_3 = P_T/4$ and $C_T = C_0 + C_1 + C_2 + C_3$; thus there is no point in subdividing the element. Now we can write $I(S,Q)$ in terms of a recursive function,

$$I(S,Q) = F(R_0(K_0)) , \tag{15}$$

where if P_{sq} is uniform over $R_m(K_m)$,

$$F(R_m(K_m)) = P_{sq}(R_m(K_m))\log[P_{sq}(R_m(K_m))] , \tag{16a}$$

and if P_{sq} is nonuniform over $R_m(K_m)$,

$$F(R_m(K_m)) = P_{sq}(R_m(K_m))\log(4) + \sum_{j=0}^{3} F(R_{m+1}(K_m,j)) . \tag{16b}$$

This results in a mixed partition being used to calculate $I(S,Q)$. The recursion goes deeper in areas where P_{sq} has finer structure, yielding smaller partition elements where they are needed.

For experimental data $P_{sq}(R_m(K_m))$ is estimated by $N(R_m(K_m))/N_0$, where $N(R_m(K_m))$ is the number of events observed in partition element $R_m(K_m)$, and N_0 is the total number of events observed. Substituting these estimates into (12) and (14) produces

$$C_T = (1/N_0)\{N(R_m(K_m))m \log(4) - N(R_m(K_m))\log(N_0) + N(R_m(K_m))\log[N(R_m(K_m))]\} , \tag{17}$$

$$\sum_{j=0}^{3} C_j = (1/N_0)\left[N(R_m(K_m))m \log(4) - N(R_m(K_m))\log(N_0) + N(R_m(K_m))\log(4) + \sum_{j=0}^{3} N(R_{m+1}(K_m,j))\log[N(R_{m+1}(K_m,j))] \right] . \tag{18}$$

Thus the recursion step consists of replacing the term

$$N(R_m(K_m))\log[N(R_m(K_m))]$$

by

$$N(R_m(K_m))\log(4) + \sum_j N(R_{m+1}(K_m,j))\log[N(R_{m+1}(K_m,j))] .$$

This step only operates on numbers of events and is independent of normalization, so we can write

$$I(S,Q) = (1/N_0)F(R_0(K_0)) - \log(N_0) , \tag{19}$$

where if there is no substructure in $R_m(K_m)$,

$$F(R_m(K_m)) = N(R_m(K_m))\log[N(R_m(K_m))] , \tag{20a}$$

and if there is substructure in $R_m(K_m)$,

$$F(R_m(K_m)) = N(R_m(K_m))\log(4) + \sum_{j=0}^{3} F(R_{m+1}(K_m,j)) . \tag{20b}$$

We use a χ-square test to check for substructure in $R_m(K_m)$. The null hypothesis is that P_{sq} is flat over $R_m(K_m)$, which implies that the distributions for $\{N(R_{m+1}(K_m,i)): 0 \le i \le 3\}$ and $\{N(R_{m+2}(K_m,i,j)): 0 \le i \le 3 \text{ and } 0 \le j \le 3\}$ are flat multinomials. If we let

$$N \equiv N(R_m(K_m)) ,$$

$$a_i \equiv N(R_{m+1}(K_m,i)) ,$$

and

$$b_{ij} \equiv N(R_{m+2}(K_m,i,j)) ,$$

the reduced χ-square statistics and 20% confidence levels are

$$\chi_3^2=\left[\tfrac{16}{9}(1/N)\sum_{i=0}^{3}(a_i-N/4)^2\right]<1.547\ , \tag{21}$$

$$\chi_{15}^2=\left[\tfrac{256}{225}(1/N)\sum_{i,j=0}^{3}(b_{ij}-N/16)^2\right]<1.287\ . \tag{22}$$

If either of these inequalities fails we conclude that there is substructure in $R_m(K_m)$.

Our algorithm operates on a pair of sequences of numbers whose lengths are a power of 2: $\{[x(t),y(t)]: 0\le t<2^n\}$, where t is an integer. If a phase portrait is being evaluated, y will be a delayed version of x. Each input sequence usually consists of floating point numbers in which the range and the density over intervals of the range is unknown. We change variables to obtain sequences of known range and uniform density. The variable change goes from floating point (x,y) to integer (s,q) representation in a fashion that preserves orderings, with the constraints that if $x(t_1)<x(t_2)$, then $s(t_1)<s(t_2)$ and if $y(t_1)<y(t_2)$, then $q(t_1)<q(t_2)$ and $0\le s(t)<2^n$ and $0\le q(t)<2^n$. This requires $\{s(t)\}$ and $\{q(t)\}$ to be permutations of the sequence $\{0$ to $2^n-1\}$. The change of variable is done using a quick-sort algorithm. If the binary representations of $s(t)$ and $q(t)$ are available after the change of variable, then $[s(t),q(t)]$ can be assigned to the appropriate element of any partition G_m by inspection. The final step is to invoke (19) to evaluate $I(S,Q)$ using (21) and (22) to test for substructure in $R_m(K_m)$.

If the number of events is N_0, the time the algorithm takes for the slowest case is proportional to $N_0\log N_0$. The change of variable is done by an $N\log N$ sort. The other time-consuming operation is dividing up the events at each branch of the tree. The worst case for this step is if $s(t)=q(t)$ for all t. In this case the events are divided into two groups at each branch, and the number of levels goes as $\log N_0$. Since all the events must be dealt with at each level, the time for the algorithm goes as $N_0\log N_0$. We have implemented the algorithm in the C language on a Digital Equipment Corporation VAX computer and have found that for each value of T it takes 97 sec to calculate $I(X(t),X(T+t))$ over 16 384 points.

There are refinements, extensions, and applications of the algorithm that can be pursued in future work. We would like to develop a more sophisticated statistical test for flatness. The present test is crude and the 20% confidence level was chosen arbitrarily. The recursion stops when the data imply that the parent distribution is flat, or when the data become too sparse to reveal any structure, without indicating which. If it stops because the data are sparse, the algorithm could grossly underestimate I. A test that calculated error bars for I would be better.

We have found the present algorithm, which operates on two-dimensional sets of data, helpful for choosing delays for higher-dimensional reconstructions, but it would still be useful to generalize it to higher dimensions. Recall that

$$I(X_0,X_1)=H(X_0)+H(X_1)-H(X_0,X_1)\ .$$

One can generalize by defining

$$I_n(X_0,X_1,\ldots,X_n)=\sum_j[H(X_j)-H(X_0,X_1,\ldots,X_n)]\ .$$

I_n is the number of bits that are redundant in a vector measurement (I_1 is the old mutual information). The redundancy occurs because knowledge of some components of a vector measurement can be used to predict something about the other components. In an algorithm that calculated I_n, elements would be subdivided into 2^{n+1} parts at each level instead of just 4 and the statistical test would be altered, but otherwise the algorithm could be the same as the present version. If the vector were a time-delay reconstruction, plots of $I_n(T)$, like those in Figs. 1 and 2, could be used to choose delays for higher-dimensional reconstructions. If n were large enough (larger than the number of degrees of freedom), I_n should be a monotonically decreasing function of T, and the slope of I_n/n should give an estimate of the metric entropy. Shaw[12] has used I_1 to estimate the metric entropy of a one-dimensional map and has discussed the technical problems.

The mutual information provides a key to quantifying *spatial* coherence or chaos, a subject of current interest.[13–15] Our method could be applied directly to a scalar field to analyze the space analog of a two-variable delay reconstruction, $\mathbf{X}=[x(r_0),x(r_0+\Delta)]$. If the method were extended to multiple spatial delays Δ, the slope of a plot of the generalized mutual information $I_n(\Delta)/n$ would give the spatial analog of the metric entropy.

In conclusion, we have presented a recursive method of calculating mutual information. Using this method we have demonstrated for two systems that the first minimum in the mutual information provides the best available systematic criterion for choosing time delays for phase portraits. We have also successfully used the method on other systems. We hope that the method will be widely used, and that it will yield new insights into the temporal and spatial behavior of nonequilibrium systems.[16]

We thank K. Coffman for providing the data used in Fig. 1. This work was supported by NSF Fluid Mechanics program Grant No. MSM8206889 and the U.S. Department of Energy, Office of Basic Energy Sciences, Contract No. DE-AS05-84ER13147.

[1]A. Brandstater, J. Swift, H. L. Swinney, A. Wolf, D. Farmer, E. Jen, and J. Crutchfield, Phys. Rev. Lett. **51**, 1442 (1983).

[2]W. M. Schaffer and M. Kot, J. Theor. Biol. **112**, 403 (1985).

[3]J.-C. Roux, R. H. Simoyi, and H. L. Swinney, Physica **8D**, 257 (1983).

[4]S. W. Teitsworth and R. M. Westervelt, Phys. Rev. Lett. **53**, 2587 (1984).

[5]N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).

[6]F. Takens, in *Dynamical Systems and Turbulence, Warwick,*

1980, Vol. 898 of *Lecture Notes in Mathematics,* edited by D. A. Rand and L. S. Young (Springer, Berlin, 1981), p. 366.

[7]P. Atten, J. G. Caputo, B. Malraison, and Y. Gagne, J. Mec. Theor. Appl., Special Issue, "Bifurcation et Comportement Chaotique," edited by G. Iooss (1984), p. 133.

[8]J. P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).

[9]O. E. Rössler, Phys. Lett. **57A**, 397 (1976).

[10]R. S. Shaw (private communication).

[11]C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

[12]R. S. Shaw, *The Dripping Faucet as a Model Chaotic System* (Aerial Press, Santa Cruz, CA, 1985).

[13]J. L. Lumley, in *Transition and Turbulence,* edited by R. E. Meyer (Academic, New York, 1981), p. 215.

[14]S. Wolfram, Rev. Mod. Phys. **55**, 601 (1983).

[15]Y. Sawada, M. Matsushita, M. Yamazaki, and H. Kondo, Phys. Scr. **T9**, 130 (1985).

[16]On request we will provide our code as a listing, or by ARPAnet mail. Our ARPAnet address is phib412@ut-ngp.