PHYSICAL REVIEW LETTERS

VOLUME 63 | 10 JULY 1989 | NUMBER 2

Inferring Statistical Complexity

James P. Crutchfield[(a)] and Karl Young[(b)]

Physics Department, University of California, Berkeley, California 94720

(Received 13 December 1988)

Statistical mechanics is used to describe the observed information processing complexity of nonlinear dynamical systems. We introduce a measure of complexity distinct from and dual to the information-theoretic entropies and dimensions. A technique is presented that directly reconstructs minimal equations of motion from the recursive structure of measurement sequences. Application to the period-doubling cascade demonstrates a form of superuniversality that refers only to the entropy and complexity of a data stream.

PACS numbers: 05.20.−y, 03.65.Bz, 05.45.+b, 05.70.Fh

Prior to the discovery of chaos, physical processes were broadly described in terms of two extreme models of behavior: periodic and random. Both are simple, but in distinct senses: the former since the behavior is temporally repetitive; the latter since it affords a compact statistical description. The existence of chaos demonstrates that there is a rich spectrum of unpredictability spanning these two extremes. Information-theoretic descriptions of this spectrum, (say) using the dynamical entropies, measure the raw diversity of temporal patterns: Periodic behavior has low information content; random, high content. What this misses, however, is the statistical simplicity of random behavior. Said more directly, the dynamical entropies do not capture the computational effort required in modeling complex behavior.[1] For example, periodic processes are readily forecast by using a stored finite pattern; stochastic processes by using a source of random numbers. Behavior between these extremes, while of intermediate information content, is more complex in that the most concise description is an amalgam of regular and stochastic processes.

To address the deficiency of conventional entropies this Letter introduces a measure of complexity that quantifies the intrinsic computation of a physical process. The basis of our approach is an abstract notion of complexity: the information contained in the minimum number of equivalence classes obtained by reducing a data set modulo a symmetry. In the following this is given concrete form for the case of data sets produced by nonlinear dynamical systems and reduced by time translation invariance, the symmetry appropriate to forecasting. Central to this is a procedure for reconstructing computationally equivalent machines, whose properties lead to quantitative estimates of complexity and entropy. This is developed using the statistical mechanics of orbit ensembles, rather than focusing on the computational complexity of individual orbits.[2-5] As an application example, we find that the period-doubling route to chaos exhibits a λ-like phase transition with a second order component and that at the accumulation point the behavior has infinite complexity, but is not computation universal.

The first step is obtaining a data stream. The (unknowable) exact states of an observed system are translated into a sequence of symbols via a measurement channel.[6] This process is described by a parametrized partition M_ε of the state space, consisting of cells of size ε that are sampled every τ time units. A measurement sequence consists of the successive elements of M_ε visited over time by the system's state. Using the instrument $\{M_\varepsilon, \tau\}$, a sequence of states $\{x_{\bar{n}}\}$ is mapped into a sequence of symbols $\{s_n : s_n \in A\}$, where $A = \{0, \ldots, k-1\}$ is the alphabet of labels for the k $(\approx \varepsilon^{-m_{\text{bed}}})$ partition elements and m_{bed} is the data set's embedding dimension.[1] A common example, to which we shall return, is the logistic map of the interval, $x_{n+1} = r x_n (1 - x_n)$, observed with the binary generating partition $M_{1/2} = \{[0,0.5),[0.5,1]\}$ whose elements are labeled with $A = \{0,1\}$.[7] We shall refer to the computational models reconstructed from such data as *ε-machines*, in order to emphasize their dependence on the measuring instrument.

Given the data stream in the form of a long measurement sequence, the first step in machine reconstruction is the construction of a tree. A *tree* $T=\{\mathbf{n},\mathbf{l}\}$ consists of *nodes* $\mathbf{n}=\{n_i\}$ and directed, labeled *links* $\mathbf{l}=\{l_i\}$ connecting them in a hierarchical structure with no closed paths. An L-level *subtree* T_n^L at node n is a tree that starts at node n and contains all nodes that can be reached within L links. To construct a tree from a measurement sequence we simply parse the latter for all length-L sequences (L-cylinders) and from this construct the tree with links up to level L that are labeled with individual symbols up to that time. We add probabilistic structure to the tree by recording for each node n_i the number $N_i(L)$ of occurrences of the associated L-cylinder relative to the total number $N(L)$ observed, $p_{n_i}^T(L)=n_i(L)/N(L)$. This gives a hierarchical approximation of the measure in orbit space. Tree representations of data streams are closely related to the hierarchical algorithm used for measuring dynamical entropies.[6,7]

The ε-machines are represented by a class of labeled, directed multigraphs, or *l-diagraphs*.[8] They are related to the Shannon graphs of informative theory,[9] to Weiss's sofic systems[10] in symbolic dynamics, to discrete finite automata (DFA)[11] in computation theory, and to regular languages in Chomsky's hierarchy.[12] In the most general formulation, we are concerned with probabilistic versions of these. Their topological structure is described by an l-digraph $G=\{\mathbf{V},\mathbf{E}\}$ that consists of *vertices* $\mathbf{V}=\{v_i\}$ and directed *edges* $\mathbf{E}=\{e_i\}$ connecting them, each of which is labeled by a symbol $s\in A$.

To reconstruct a topological ε machine we define an equivalence relation, *subtree similarity*, denoted $\approx$, on the nodes of the tree by the condition that the L-subtrees are identical: $n_i\approx n_j$ if and only if (iff) $T_{n_i}^L=T_{n_j}^L$. Subtree equivalence means that the link structure is identical. This equivalence relation induces a set of equivalence classes $\{\mathbf{C}_l^L\colon l=1,\ldots,K\}$ given by $\mathbf{C}_l^L=\{n\in\mathbf{n}\colon n_i\in C_l^L \text{ and } n_j\in C_l^L \text{ iff } n_i\approx n_j\}$. We refer to the archetypal subtree link structure for each class as a *morph*. A graph G_L is then constructed by associating a vertex to each tree node L-level equivalence class. Two vertices v_k and v_l are connected by a directed edge if the transition exists in the tree between nodes in the equivalence classes, $n_i\to n_j\colon n_i\in\mathbf{C}_k^L,\ n_j\in\mathbf{C}_l^L$. The corresponding edge is labeled by the symbol(s) associated with the tree links connecting the tree nodes in the two equivalence classes.

In this way, ε-machine reconstruction deduces from the diversity of individual temporal patterns "generalized states," associated with the graph vertices, that are optimal for forecasting. The topological ε-machines so reconstructed capture the essential computational aspects of the data stream by virtue of the following instantiation of Occam's razor.

Theorem—.Topological reconstruction of G_L produces the minimal machine recognizing the language and the generalized states specified up to L-cylinders by the measurement sequence.

The generalization to reconstructing metric ε-machines that contain the probabilistic structure of the data stream follows by a straightforward extension of subtree similarity. Two L-subtrees are δ similar if they are topologically similar and their corresponding links individually are equally probable within some $\delta\geq 0$. There is also a motivating theorem: Metric reconstruction yields minimal metric ε-machines.

In order to reconstruct an ε-machine it is necessary to have a measure of the "goodness of fit" for determining ε, τ, δ, and the level L of subtree approximation. This is given by the *graph indeterminacy*, which measures the degree of ambiguity in transitions between graph vertices. The indeterminacy[6] I_G of a labeled digraph G is defined as the weighted conditional entropy

$$I_G=\sum_{v\in\mathbf{V}}p_v\sum_{s\in A}p(s\,|\,v)\sum_{v'\in\mathbf{V}}p(v'\,|\,v;s)\log p(v'\,|\,v;s)\,,$$

where $p(v'\,|\,v;s)$ is the transition probability from vertex v to v' along an edge labeled with symbol s, $p(s\,|\,v)$ is the probability that s is emitted on leaving v, and p_v is the probability of vertex v; logarithms are to the base 2. An ε-machine is reconstructable from L-level equivalence classes if I_{G_L} vanishes. Finite indeterminacy, at some given L, ε, τ, and δ, indicates a residual amount of extrinsic noise at the level of approximation.

We now turn to the statistical-mechanical description of ε-machines, the central result of which is a natural definition of *complexity* that is dual to the dynamical entropies and dimensions. We define the partition function for n-cylinders $\{s^n\}$ as

$$Z_\alpha(n)=\sum_{\{s^n\}}e^{\alpha\log p(s^n)}\,,$$

where α is a formal parameter related to the inverse temperature of statistical mechanics. The α-order total Renyi information,[6] or *free information*, in the n-cylinders is given by $H_\alpha(n)=(1-\alpha)^{-1}\log Z_\alpha(n)$. The dynamical α-order entropies are given by the thermodynamic (large orbit space volume) limit

$$h_\alpha=\lim_{n\to\infty}(1-\alpha)^{-1}n^{-1}\log Z_\alpha(n)\,.$$

An ε-machine is described by a set $\{T^{(s)}\colon s\in A\}$ of transition matrices, one for each symbol $s\in A$, where $T^{(s)}=\{p_{s,ij}\}$ with $p_{s,ij}=p(v_j\,|\,v_i;s)$. We define the probabilistic connection matrix as $T_\alpha=\{t_{ij}\}=\sum_{\{s\in A\}}T_\alpha^{(s)}$, where the parametrized matrix $T_\alpha^{(s)}=\{p_{s,ij}^\alpha\}$. T defines a Green's function on the tree of measurement sequences. The eigenvalue spectrum $S[T_\alpha]$ will be denoted $\{\lambda_i(\alpha)\}$. The largest eigenvalue λ_α of T_α is real for ε-machines. The associated eigenvector $\hat{p}_\alpha=\{p_v^\alpha\colon v\in\mathbf{V}\}$ has nonnegative elements that give the asymptotic vertex probabilities.

As a measure of the information processing capacity of an ε-machine, we define the α-order graph complexity as the Renyi entropy of $\hat{p}_\alpha$,

$$C_\alpha = (1-\alpha)^{-1} \log \sum_{v \in \mathbf{V}} p_v^\alpha .$$

This is an intensive thermodynamic quantity that measures the average amount of α information contained in the morphs. It also quantifies the informational fluctuations in the data stream. Fluctuations in the free information are measured via the total excess (α) entropy for L-cylinders[13-15]

$$F_\alpha(L) = H_\alpha(L) - h_\alpha L = \sum_{n=1}^{L} [H_\alpha(n) - H_\alpha(n-1) - h_\alpha] .$$

It is useful in its own right since if a machine is finite, i.e., $F_1(L)$ is finite, the process is weak Bernoulli.[14] $F_\alpha(L)$ is a measure of fluctuations of finite cylinder set statistics from asymptotic. In the thermodynamic limit, $L \to \infty$, the α-order complexity is simply proportional to F_α. Heuristically speaking, the complexity measures the average amount of mathematical work required to produce a fluctuation.

Developing the partition function in terms of eigenvalues of T_α, the α-order total entropies can also be expressed in terms of the ε-machine eigenvalue λ_α: $H_\alpha(n) \approx (1-\alpha)^{-1} \log \lambda_\alpha^n$, $n \to \infty$. The Renyi entropy spectrum is then $h_\alpha = (1-\alpha)^{-1} \log \lambda_\alpha$. For completeness, we note that the spectrum d_α of Renyi dimensions can be similarly related to the reconstructed ε-machine via the ε scaling of H_α.

There are two important cases for the α-order graph complexity that we now explicitly consider. The first is the topological case, when $\alpha = 0$. T_0 is the digraph's connection matrix, the Renyi entropy $h_0 = \log \lambda_0$ is the topological entropy h, and the graph complexity is the probabilistic algorithmic complexity, $C_0(G) = \log |\mathbf{V}|$. In the case of cellular automata and assuming one has the equations of motion, this quantity has been referred to as the algorithmic complexity.[16] It is important, however, to distinguish C_0 from the Chaitin-Kolmogorov complexity, which goes under the same name and describes the complexity of individual measurement sequences.[2-5] The quantities that we have defined here are *probabilistic* and referred to Turing machines with a random internal register.

The second (metric) case of interest is the "high-temperature limit," when $\alpha = 1$. Here, h_α is the metric entropy: $h_\mu = \lim_{\alpha \to 1} h_\alpha = -d\lambda_\alpha / d\alpha$. C_α is the graph complexity, $C_1 = -\sum_{\{v \in \mathbf{V}\}} p_v \log p_v$, that is, the Shannon entropy of $\hat{p}_1$.[15,17]

We demonstrate the practical implementation of ε-machine reconstruction and the foregoing formalism on the well-studied period-doubling cascade, as exhibited by the logistic map. Figure 1 displays topological ε-machines for several notable parameter values. Figure 2 shows the graph complexity $C_\alpha(L)$ versus the specific entropy $L^{-1} H_\alpha(L)$ for the metric case, $\alpha = 1$, at parameter values associated with period-doubling cascades of various periodicities. Of particular interest is the appearance of a phase transition as a function of specific entropy. The observed divergence in C_1 indicates a form of "superuniversality" for the transition, since we are only considering the complexity's dependence on the information-theoretic characterization, viz, H_α, of the measurement sequences, and not the dependence on parameter value. The lower bound on C_1, attained for periodic behavior and at band mergings, indicates a λ-like phase transition with a second-order component. The remaining chaotic data are always significantly above the lower bound.

This example shows that C_α is a measure of complexity distinct from and dual to standard information mea-

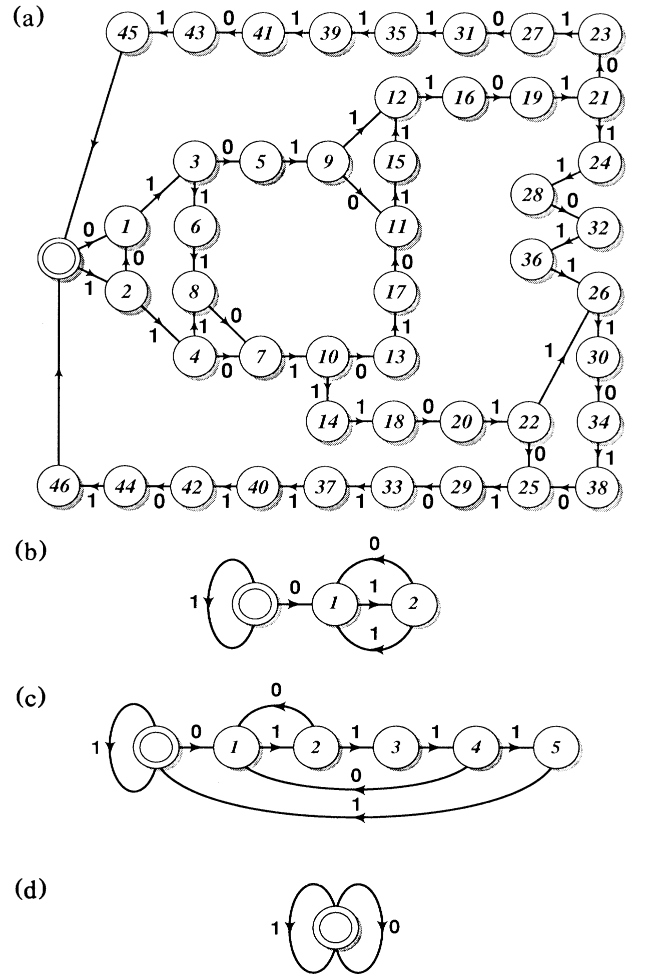

FIG. 1. Topological ε-machine l-digraphs for the logistic map at (a) the first period-doubling accumulation $r_c = 3.569\,945\,671\ldots$, (b) the band merging $r_{2B \to 1B} = 3.678\,59\ldots$, (c) the "typical" chaotic value $r = 37$, and (d) the most chaotic value $r = 4$. (a) and (c) show approximations of infinite ε-machines. The start vertex is indicated by a double circle; all states are accepting; otherwise, see Ref. 11.

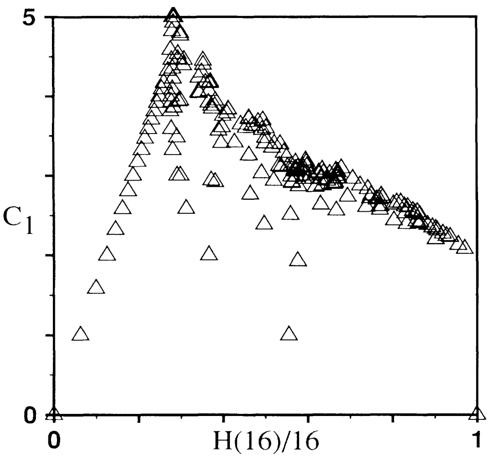

FIG. 2. Graph complexity C_1 vs specific entropy $H_1(16)/16$, using the binary, generating partition $\{[0,0.5),[0.5,1]\}$, for the logistic map at 193 parameter values $r \in [3,4]$ associated with various period-doubling cascades. For most, the underlying tree was constructed from 32-cylinders and machines from 16-cylinders. From high-entropy data sets smaller cylinders were used as determined by storage. Note the phase transition (divergence) at $H^* \approx 0.28$. Below H^* behavior is periodic and $C_\alpha = H_\alpha = \log(\text{period})$. Above H^*, the data are chaotic. The lower bound $C_\alpha = \log(B)$ is attained at $B \rightarrow B/2$ band mergings.

sures of dimension d_α and entropy h_α. The latter, given by the ε-machines's eigenvalue, simply measures the diversity of observed patterns in a data stream: the more random the source, the more patterns, and so the higher the information content. The graph complexity, however, given by the ε-machine's eigenvector, measures the computational resources required to reproduce a data stream.[18] It vanishes for trivially periodic and for purely random data sets. A reconstructed ε-machine reflects a balanced utilization of deterministic and random information processing. We claim this model basis is the proper one for describing the computational complexity of physical processes, since the latter always have some residual extrinsic fluctuations.

The extension of ε-machines beyond regular languages to context-free languages and higher levels in Chomsky's hierarchy[12] is the next major step. It can be shown, for example, that despite its infinite graph complexity period-doubling accumulation has a finite-stack machine description and so is not computation universal. The application of machine reconstruction using spatial translation symmetry and the associated statistical mechanics to estimate spatiotemporal machines is straightforward.[1] Reconstructing language hierarchies for data sets and minimal machines for spatial systems will be reported elsewhere.

We have argued for a chaotic measurement theory that exploits the intimate relationship between information, computation, and forecasting.[1,6,7] This Letter establishes the connections constructively and within the framework of statistical mechanics. With a direct measure of an ε-machine's complexity, the theory gives a computation-theoretic foundation to the notions of model optimality and, most importantly, a measure of the computational complexity of estimated models.[1] The generality of ε-machine reconstruction and its ability to infer generalized states from a data stream suggest that it captures essential aspects of learning. That it can be put into a statistical mechanical framework, therefore, suggests the existence of a complete theoretical basis for "artificial science": the fully automatic deduction of optimal models of physical processes.[1]

This work was funded in part by ONR Contract No. N00014-86-K-0154 and by the 1988 Cray-Sponsored University Research and Development Grant Program. The authors thank Professor Carson Jeffries for his continuing support.

[a]Internet address: chaos @ gojira, berkeley. edu.

[b]Permanent address: Physics Board of Studies, University of California, Santa Cruz, CA 95064. Electronic address: karl @ gojira, berkeley. edu.

[1]J. P. Crutchfield and B. S. McNamara, Complex Syst. **1**, 417 (1987).

[2]G. Chaitin, J. Assoc. Comput. Mach. **13**, 145 (1966).

[3]A. N. Kolomogorov, Prob. Inf. Transm. **1**, 1 (1965).

[4]J. Ford, in *Chaotic Dynamics and Fractals,* edited by S. G. Demko, (Academic, Orlando, 1986), 1.

[5]A. A. Brudno, Trans. Moscow Math. Soc. **44**, 127 (1983).

[6]J. P. Crutchfield, Ph.D. dissertation, University of California, Santa Cruz, 1983 (University Microfilms International, Minnesota, 1984).

[7]J. P. Crutchfield and N. H. Packard, Int. J. Theor. Phys. **21**,433 (1982).

[8]D. M. Cvetkovic, M. Doob, and H. Sachs, *Spectra of Graphs* (Academic, New York, 1980).

[9]C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Univ. Illinois Press, Champaign-Urbana, 1962).

[10]R. Fischer, Monatsh. Math **80**, 179 (1975).

[11]J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, Reading MA, 1979)

[12]N. Chomsky, IRE Trans. Inf. Theory **2**, 113 (1956).

[13]J. P. Crutchfield and N. H. Packard, Physica (Amsterdam) **7D**, 201 (1983).

[14]N. H. Packard, Ph.D. dissertation, University of California, Santa Cruz, 1982 (unpublished).

[15]Compare "EMC" and "SC" in P. Grassberger, Int. J. Theor. Phys. **25**, 907 (1986).

[16]S. Wolfram, Commun. Math. Phys. **96**, 15 (1984).

[17]Compare "diversity" of undirected, unlabeled trees in C. P. Bachas and B. A. Huberman, Phys. Rev. Lett. **57**, 1965 (1986).

[18]Compare "thermodynamic depth" in S. Lloyd and H. Pagels, Ann. Phys. (N.Y.) **188**, 186 (1988), and "logical depth" in C. H. Bennett, in *Emerging Syntheses in the Sciences,* edited by D. Pines (Addison-Wesley, Redwood City, 1988).