Applying algorithmic complexity to define chaos in the motion of complex systems

Andrea Crisanti and Massimo Falcioni*
Dipartimento di Fisica, Università "La Sapienza," Piazzale Aldo Moro 2, I-00185 Roma, Italy

Giorgio Mantica†
Dipartimento di Matematica, Università di Milano a Como, Via Lucini 3, I-22100 Como, Italy

Angelo Vulpiani*
Dipartimento di Fisica, Università "La Sapienza", Piazzale Aldo Moro 2, I-00185 Roma, Italy
(Received 3 December 1993)

We define chaotic motion for dynamical systems acting in finite, discrete spaces via the deterministic randomness of their trajectories. The theory of algorithmic complexity is used to provide the meaning of randomness for symbolic sequences derived from these trajectories, and a practical test of randomness is devised on the basis of an ideal, physically motivated, model of a computer. Two examples—a discretized standard map, and a fully connected neural network—are studied analytically and numerically.

PACS number(s): 05.45.+b

I. INTRODUCTION

In research papers, seminars, and even popular magazines one encounters with increasing frequency the sentence "this motion (dynamics, pattern, state of things, ... is chaotic." What exact meaning should we give to this statement, which would render in a satisfactory way our understanding of chaos, and could be applied to the largest class of dynamical situations? At the present moment, we have a fairly good idea of what chaos is in—say—low-dimensional Hamiltonian systems. Yet, the extension of this concept to infinite-dimensional systems governed by partial differential equations, or to dynamics in spaces with a finite but arbitrarily large number of states, is far less than obvious. In this paper, we focus our attention on this latter class, which includes cellular automata [1], certain neural networks [2], etc.; these systems are of utmost importance, since they seem to be well suited for the description of the phenomena occurring in complex systems. Ranging from turbulence to DNA coding, these phenomena mark the frontier of today's dynamics.

Chaos is a multifaceted entity, and its essence can be grasped at different levels of sophistication. It is usually associated with fast decay of correlation functions—whether temporal or spatial, but more often it is identified as *sensitive dependence on initial conditions* [3]. This feature is present whenever an evolution law $\boldsymbol{x}(0) \to \boldsymbol{x}(t) := S^t\boldsymbol{x}(0)$—the operator S^t can be induced by a map, a differential equation, or any other means—is such that close initial vectors $\boldsymbol{x}(0)$ give rise to trajectories parting exponentially fast in time. This concept is rendered precise in the definition of Lyapunov exponents, whose positive value is often taken as a definition of chaos. While this definition catches one of the main features of what we would like to call chaos, it is incomplete in some respects. In fact, on the one hand it leads sometimes to unsatisfactory answers, on the other hand it cannot always be applied, as in the case of systems with discrete state space.

The first drawback appears in infinite-dimensional systems, like that describing in fluid dynamics a scalar field $\theta(\boldsymbol{x},t)$ (e.g., the local concentration of a pollutant) driven by a given velocity field $\boldsymbol{v}(\boldsymbol{x},t)$ (e.g., water flow in porous rocks). The evolution of $\theta(\boldsymbol{x},t)$ is governed by the equation

$$\partial_t\theta(\boldsymbol{x},t) + [\boldsymbol{v}(\boldsymbol{x},t)\cdot\nabla]\theta(\boldsymbol{x},t) = D\Delta\theta(\boldsymbol{x},t), \qquad (1)$$

where D is the bare diffusion constant. To detect if sensitive dependence on the initial condition $\theta(\boldsymbol{x},0)$ is present, one can ideally change this function by a small perturbation $\delta\theta(\boldsymbol{x},0)$, and measure, as time evolves, the difference $\delta\theta(\boldsymbol{x},t)$ between the true and the perturbed solutions of Eq. (1). A simple computation shows that

$$\int |\delta\theta(\boldsymbol{x},t)|^2 d\boldsymbol{x} \le \int |\delta\theta(\boldsymbol{x},0)|^2 d\boldsymbol{x}. \qquad (2)$$

Errors do not grow in time, at least in the L^2 metric which seems natural to this problem, and sensitive dependence on initial conditions is totally absent. Nonetheless, the evolution of θ can be chaotic, in some respect [4]. For simplicity, consider the diffusionless case $D = 0$. The solution of Eq. (1) can be written

$$\theta(\boldsymbol{x},t) = \theta(S^{-t}\boldsymbol{x},0), \qquad (3)$$

where S^t is the formal solution of the stream equation

*Also at INFN, Sezione di Roma, Italy.
†Electronic address: MANTICA@MILANO.INFN.IT

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{v}(\boldsymbol{x}, t). \tag{4}$$

The motion induced by (4) in the usual three-dimensional space can be completely chaotic, and the "time record" $F(t) = \theta(\boldsymbol{x}, t)$ of the pollutant concentration read at fixed position $\boldsymbol{x}$ can be a complicated function as well, even if (2) holds. Other well-known indicators of stochasticity can be used to characterize this complexity. For instance, one might compute the Kolmogorov entropy [5] associated to the record $F(t)$, via the Grassberger-Procaccia method and a suitable embedding technique [6]: a positive value will summon to trial the sequence—and its generator, the dynamical system (1)—on chaoticity charges.

As a matter of fact, the paradox presented in the last example is identically present *ceteris paribus* in the Liouville description of classical mechanics, where a linear operator determines the motion of probability distributions in phase space. This motion is not sensitively dependent on initial conditions (in the metric of a proper functional space) even for completely chaotic systems like the Arnol'd cat [7]. This paradox is resolved noting—as done, e.g., in [8]—that the function-theoretical description of this evolution demands an exponentially enlarging vector basis set. Chaos therefore reveals itself in the form of a large information burden.

All that shows, in a clear way, the remarkable difference between the complexity of the dynamical systems, like the one defined by (1), which do not suffer from an exponential growth of the errors, and that of their trajectories, as given by $F(t)$.

Sensitive dependence on initial conditions fails to be effective also in the large class of systems with discrete states, discussed in this work. Discreteness renders obviously impossible the determination of trajectories with arbitrarily close initial conditions, as required in the definition of Lyapunov exponents. As a partial remedy, a metric with some physical significance can usually be defined, and initial states with minimal distance can be identified and followed under time evolution. Increase of the distance of nearby states goes under the name of *damage spreading* [9].

Recall now that positivity of Lyapunov exponents alone is not sufficient for chaos: together with *stretching*, one needs *folding*. This latter aspect seems to be connate to finite systems, so we will not spend any more time on it beyond that required to note that the net result of stretching and folding is to provide bounded and unpredictable orbits. In fact, under a suitable partitioning of phase space these orbits can be translated into sequences of symbols which display significant stochastic aspects. This reduction is the main tool of symbolic dynamics. We shall lay our definition of chaos on the randomness of the symbolic sequences representing the time evolution of a system, a common ground for discrete and continuous systems alike. Randomness will be defined according to Kolmogorov's theory of algorithmic complexity [10,11], following an idea first employed in dynamics by Alekseev and Jacobson [12] and Brudno [13], and later developed by Ford and co-workers [14–16] into the notion of *deterministic randomness*.

In discrete, finite systems, no artifice is required to translate motions into symbolic sequences: associating a letter (or a binary word) to each space point does the job, and trajectories become strings of these letters. Following Ford, we will call *deterministically random-chaotic* those trajectories which are informationally incompressible, that is, those for which no algorithm can be found to "encode" their associated sequence while compressing it. Clearly, for any dynamical system, a possible scheme to encode a trajectory is the motion-defining algorithm itself, endowed with a specification of the initial condition; these two things together can be thought of as a computer program which outputs a precise dynamical sequence. Whenever the computer program is significantly shorter than its output, the motion will be defined compressible, or regular. To focus ideas, let us introduce two examples to be studied in detail in this paper.

First, we consider a discretized standard map, obtained by considering lattice points in the torus $[0, 2\pi]^2$ of the form (x, y) with $x = 2\pi q/L$, $y = 2\pi p/L$, and with q, p integers between 1 and L, also an integer. The action of the map is defined as follows:

$$\begin{aligned} q_{n+1} &= \left[q_n + k \tfrac{L}{2\pi} \sin(p_n \tfrac{2\pi}{L}) \right] \bmod L, \\ p_{n+1} &= (p_n + q_{n+1}) \bmod L, \end{aligned} \tag{5}$$

where [] means the integer part, where k is an external parameter (the "kick strength"), and where the index n labels discrete time. When $k \gg 1$ the motion of the continuum standard map is chaotic; will the discretized motion be the same? This example draws additional importance from the debated problem of *quantum chaos*: quantum mechanics can be regarded as a discretized version of classical mechanics, acting on suitable lattices in phase space. In analogy with [17,18,8], we can think of L in the above equation as the inverse of Planck's constant $\hbar$.

The second example is a fully connected, deterministic neural network. The system is composed of L spins, $\sigma_i = -1, 1$, which evolve according to a simple rule

$$\sigma_i(n+1) = \mathrm{sgn}\left(\sum_{j=1}^{L} J_{ij} \sigma_j(n) \right), \quad i = 1, \ldots, L \tag{6}$$

where J_{ij} is a given random quenched coupling matrix, see Sec. IV for details. The large cardinality of the number of possible states of this system, 2^L, provides us with a good model for complex systems. Moreover, a recent investigation [19] shows that these states are visited in a rather irregular fashion: even if eventually periodic (as in all discrete, finite systems) the motion is characterized by exponentially long transients, and "local" complexity. In fact, the time sequence of σ_i, at fixed site i, possesses many statistical properties, including positive Shannon and topological entropies: an ideal patient indeed for our surgical table.

This paper develops according to the following scheme. In Sec. II we briefly recall some basic concepts of algorithmic complexity theory and we draw a formalism for dynamical systems in finite spaces. Section III is the

heart of the paper: we introduce the concept of a realistic computing machine as a means of estimating algorithmic complexity and detecting chaos or deterministic randomness in motions in finite spaces. Section IV is devoted to the study of the systems defined above, by means of numerical computations and probabilistic approximations. In Sec. V the reader may find some further remarks and the conclusions.

II. DETERMINISTIC RANDOMNESS OF MOTIONS IN DISCRETE, FINITE SPACES

A dynamical system can be interpreted as a computer, which for us is nothing else than a Turing machine capable of taking a binary sequence as input and producing a binary sequence as output. This approach is certainly not new, see e.g., [20], but it is particularly suited to treat systems with finite, discrete state space. Let us first see how motions in these spaces can be described in terms of binary sequences.

A proper numbering of all available states reduces dynamics to a deterministic map over the set of the first M numbers:

$$F\colon\ m \to F(m),\quad F:\ \{1,\dots,M\} \to \{1,\dots,M\}. \tag{7}$$

A single state m is a sequence of (at most) $\log_2(M)$ binary digits, and its "trajectory" is given by $F^{(j)}(m)$, for "times" $j = 0, 1, \dots, N-1$. These data can be translated into a binary sequence $\underline{\Sigma}$ of length $l_{\underline{\Sigma}}(N, M)$,

$$l_{\underline{\Sigma}}(N, M) = N \log_2(M). \tag{8}$$

We wish to characterize the chaoticity of a dynamical system according to the complexity of its trajectories. Note that this is precisely the main scope of symbolic dynamics, where (7) is obtained by a suitable partition of phase space, and therefore is a stochastic map. In our case, no continuous phase space is hiding behind the numbering of the states, but algorithmic complexity gives us a clue and a means to catch the essence of the problem. In fact, suppose one would need to send the sequence $\underline{\Sigma}$ to a friend living on an outer planet, and that long-distance companies charge a dramatic price per bit transmitted. It will be natural for this unfortunate researcher to look for a convenient *coding* of $\underline{\Sigma}$: an encrypting program of length less than $l_{\underline{\Sigma}}$ which will be air-mailed in place of the full sequence. Algorithmic complexity theory defines complexity of a sequence the length of the shortest computer program programmable on a universal machine, capable of outputting the sequence and stopping afterwards [11,10]. In other words, this number (complexity) can be characterized as the length of the shortest possible definition of the sequence itself. Seen from this perspective, and remembering some renowned paradoxes of logics, it is not surprising that the actual complexity of a given sequence cannot be computed, but only estimated, counting, for instance, the bit length of a particular, though not optimal, encoding program.

Without any doubt, a coding which will immediately occur to the aforementioned scientist, and which is always available *a priori* for the sequence $\underline{\Sigma}$, consists of the record composed of the initial state m, whose length is $\log_2(M)$, the time span of the evolution, N, codable in $\log_2(N)$ bits, and the specification of the rule F which generates the dynamics. Let us call l_F the binary length of the algorithm for F alone. An upper bound to the optimal l_F can be easily obtained: since F consists of a possible way to associate a new state m' to any original state m, it can be coded by the table $m \to m'$, with the result that

$$l_F \le 2M \log_2(M). \tag{9}$$

The estimate (9) will indeed be appropriate for most (in probabilistic sense) maps F, and will be largely abundant only for a restricted set of maps, which, on the other hand, may contain important physical examples. To sum up, the coding length $l_c(N, M)$ can be estimated as

$$\begin{aligned} l_c(N, M) &= l_F + \log_2(M) + \log_2(N) + D \\ &\le (2M+1)\log_2(M) + \log_2(N) + D, \end{aligned} \tag{10}$$

where D is a constant independent of N and M.

On the basis of these considerations, we can introduce the following.

Definition: We shall call regular or compressible those sequences that can be encoded by a computer program whose length $l_c(N, M)$ increases less than linearly in N, when N ranges over a physically significant interval, at fixed, physically significant values of M. The remaining set will be termed chaotic, or incompressible.

As shown by Alekseev and Brudno, the typical trajectory of a classical unstable system meets this definition, no matter how fine is the partition which defines symbolic dynamics. This guarantees that our definition agrees with the most important characterization of chaos (in our opinion) holding for continuous systems. The proviso about a finite interval of values of N is crucial when dealing with finite systems, as we shall see.

The definition given above comprises a crucial idea in the word "increases": for us, the increase of $l_c(N, M)$ shall be its *leading scaling behavior* as a function of N, at physically reasonable values of N and M. Indeed, this definition also implies that, for instance, logarithmic corrections to a linear behavior of $l_c(N, M)$ are unimportant, and cannot change its chaotic properties. To explain the reasons behind this choice, we must seek help for our physical intuition.

Comparing the "growth rate" of $l_c(N, M)$ to N is quite analogous to what is done, for instance, in the similar but independent theory of computational complexity [21,22], which ranges different problems by the *scaling*, with respect to problem size, of the number of operations needed by an optimal solution algorithm. We have problems which are linear, quadratic, polynomial in system size, and also *NP complete problems*, for which no polynomial solution procedure exists. From this perspective, scaling relations for the number of computer operations are more fundamental than their absolute values, because they discriminate between problems which can be technologically mastered by a reasonable increase (polynomial) of resources and those which cannot.

The same ideal setup is also applicable to algorith-

mic complexity. In fact, quite in the same way as NP complete problems quickly blow out of solvability range, because a linear increase in problem size turns into an exponentially larger demand in computer speed, an incompressible dynamics produces a set of possible motions (words) whose cardinality increases exponentially in N, and therefore requires a literal specification (transmission, in the example above) of each, integral, sequence $\underline{\Sigma}$, without any significant compression. An important consequence of this fact, worth mentioning at this point, is the lack of *long term predictability* of these sequences, implied by the richness of the set of dynamical words.

Escorted by the penetrating analysis of algorithmic complexity, we can now reconsider the dynamics given by the rule (7), and its daughter sequences $\underline{\Sigma}$. Equations (8) and (10) show that, when N grows indefinitely and M is kept constant, the N dependence of $l_c(N, M)$ is logarithmically bounded and hence the sequence is compressible. Yet, note that the estimate (10) may *not* be optimal in a large range of parameter values, possibly including the physical M values. This happens notably when studying complex systems, where the observation time is typically much smaller than the cardinality of state space, $N \ll M$.

Let us henceforth assume to be in this latter case, and let us also assume that l_F is maximal (that is, of the order of $M \log_2 M$). For short sequences $\underline{\Sigma}$ the trivial code (*verbatim* transmission of $\underline{\Sigma}$) will surely beat the dynamical coding: $l_{\underline{\Sigma}} < l_c$. We denote by $\bar{N}$ the break-even point, when the two codings are roughly equivalent. One can easily find that

$$\bar{N} \sim M, \tag{11}$$

where the symbol $\sim$ indicates here and in the following the leading order contribution. Formula (11) means that only after the "time" $\bar{N}$ does the information contained in the sequence $\underline{\Sigma}$ become compressible. At the same time, we can also expect that the sequence becomes predictable. Indeed, it happens that either during the time $\bar{N}$ we have visited all states, and constructed the full iteration table (we had enough time to do it), or the dynamics has taken place within a smaller subset, which has been also fully explored and understood. In a sense, this regularity comes *after the fact*, because M (and hence $\bar{N}$) is (in our hypothesis) much larger than the physically significant times over which a system is observed. Therefore most trajectories generated under these conditions have maximal complexity through all the observation time, and will be termed chaotic.

In all other cases, when a better coding of F is available, the increase of the break-even point with M is slower than linear. These sequences will be termed regular. At the extreme end lies the case when the formula for F is so concise as to be logarithmically compressible:

$$l_F \sim \log_2 M. \tag{12}$$

The break-even point is then roughly independent of M, and the *rule* is so simple to constrain immediately the complexity of $\underline{\Sigma}$. A notable example of this case is provided by the discrete linear homeomorphisms of the torus studied by Percival and Vivaldi [23]: elegant algebraic concepts translate the simplicity of the dynamical evolution of these systems.

The analysis just concluded was based on the idea that the map (7) could be performed in exact arithmetic. To broaden the scope of our investigation, we now consider a more realistic representation of a dynamical system, as a map F where errors can occur due to lack of precision in carrying out arithmetical operations.

III. DYNAMICAL SYSTEMS AS REALISTIC COMPUTING MACHINES

The evolution equation (7) can be carried out, in principle, in exact arithmetic. Yet, in practice, maps of the type (7) involve some kind of real number computations, or physical measurements on analog devices, followed by suitable truncations, as in our examples (5), (6). Due to finite precision, errors may arise in these computations. Clearly, these errors prevent us from coding a trajectory over arbitrarily long times just by knowing the dynamical rule, and the initial condition. This demands a refinement of the theory exposed so far, which amounts to taking into account the "stability" of the rule (7).

We now regard the action of F as a computer program which receives as input the discrete state m, and a vector of parameters $\boldsymbol{C}$, with C components, needed to define the action of F. Since we suppose that these parameters are all known within precision 2^{-q}, the binary length of the bit translation of $\boldsymbol{C}$ is qC. The output sequence $\underline{\Sigma}$ is still determined by repeated applications of a map F, but *errors* are now possible due to the incomplete knowledge of the true values, C_i^T, of the parameters $\boldsymbol{C}$. We render this unavoidable ambiguity in the determination of $\boldsymbol{C}$ by adding a random round-off error ω_i to the unknown digits of each component C_i:

$$C_i = C_i^T + \omega_i. \tag{13}$$

The random variables ω_i are uniformly distributed in $[-2^{-q-1}, 2^{-q-1}]$. Within these assumptions, the coding length l_F becomes

$$l_F \sim Cq + \log_2 M, \tag{14}$$

but we are not assured that this information (augmented by the specification of the initial state) is enough to define the sequence $\underline{\Sigma}$ up to any time N, because errors can intervene.

This uncertainty leads to the following problem: given two identical initial conditions, and the same vector of true coefficients $\boldsymbol{C}$, what is the time over which two different "physical" computations, characterized by different realizations of the sequence of random variables ω_i, produce the same dynamics? Obviously, the answer depends on the realization of the random variables, and the initial condition of the motion. Averaging over both, we define the *first error time* $\tilde{N}$; this is also the average length of sequences $\underline{\Sigma}$ which can be effectively coded by the realistic computer F. For later convenience, let us

call ϵ the arithmetic precision: $\epsilon = 2^{-q}$. The "first-error time" $\widetilde{N}$ then depends, for a fixed map F, on ϵ alone and we can envisage two main forms of this dependence. We can have, for instance,

$$\widetilde{N} \sim \frac{1}{\epsilon^{\alpha}}. \tag{15}$$

This case is the analog of polynomial error propagation, since a linear decrease in ϵ, which we can think to be associated to a linear price, spent either in laboratory instrumentation or in hardware or software for floating point operations, leads to a power-law increase of the "codable" time. The second case includes logarithmic behaviors, for which

$$\widetilde{N} \sim \log_2 \left(\frac{1}{\epsilon} \right). \tag{16}$$

Here, the increase of one order of magnitude in computing capabilities results only in a linear increase in the coding time.

Without much surprise, we identify logarithmic dependence with chaos. All other behaviors, including power law, correspond to regular motion. In fact, in the latter case a trajectory of length $\widetilde{N}$ can be coded by a program of length l_c,

$$l_c \sim \log_2 M + \frac{D_1}{\alpha} \log_2 \widetilde{N}, \tag{17}$$

while in the realm of chaos—Eq. (16)—l_c scales exactly as $\widetilde{N}$:

$$l_c \sim \log_2 M + D_2 \widetilde{N}, \tag{18}$$

where D_1 and D_2 are two suitable constants.

The one just presented will be our working definition of chaos for discrete dynamical systems. The attentive reader may wonder if this is equivalent to the definition proposed in Sec. II. Once the forcefully nonrigorous nature of our arguments is forgiven, the answer is affirmative. In fact, computing errors can be regarded as an artifice to obtain a careful bound on the complexity of a dynamical sequence $\underline{\Sigma}$.

To see why this is the case, let us fix the time N and the parameters $\boldsymbol{C}$ to their q-digit representations, without any indeterminacy. Suppose also that numerical operations are carried out exactly. In this case, $l_{\underline{\Sigma}}$ and l_F are still given by Eqs. (8) and (14), respectively, and $l_c = l_F + \log_2 N$ is an upper bound on the complexity of $\underline{\Sigma}$. Can it be improved, possibly reducing the "cost" l_F? If $\underline{\Sigma}$ is insensitive (up to time N) to some of the digits of the parameters C_j (that is to say, parameters have been overdetermined), reduction is immediate. An optimal bound can therefore be obtained by "tuning" the number of bits q—hence, ϵ—to the minimum value required to have a correct determination of the sequence $\underline{\Sigma}$ up to time N. This is how $\widetilde{N}(\epsilon)$ is defined. Using the scaling behavior of this latter quantity, the reader can finally derive the equivalence of the two definitions.

We are now ready to apply this theory to the two dynamical systems introduced in Sec. I and to investigate their complexity properties.

IV. TWO EXAMPLES: DISCRETE HAMILTONIAN SYSTEMS, AND FULLY CONNECTED NEURAL NETWORKS

The arguments exposed in the preceding section allow us to put bounds on the complexity of a dynamical evolution just by studying the dependence of the first error time $\widetilde{N}(\epsilon)$ on the precision ϵ. This will be done in this section, performing suitable computer experiments and analytical estimates.

Let us first consider the case of maps in discretized phase space, represented here by the standard map (5). The number of possible states of this system, which we call V_L, is L^2. As a consequence, unlike that of continuous maps, the long time behavior of (5) is periodic. The typical period T_L is, nevertheless, very long; a reasonable estimate is $T_L \sim V_L^{1/2}$ [24,25]. The relation $T_L \sim L^{D/2}$, with D the correlation dimension, was proven for hyperbolic systems [26]. By taking L large enough we are assured that periodic motions will recur only over sufficiently large times, which increase on average as L. We shall consider the characteristics of the motion in the transients before periodicities set in.

In Eq. (5) the only external parameter is the kick strength k: the vector $\boldsymbol{C}$ consists of the sole component $C_1 = k$. Numerical determination of the first error time $\widetilde{N}(\epsilon)$ is performed as follows: we take two trajectories starting from the same initial condition, but with slightly different evolution laws. The first is a reference trajectory obtained from (5) with a given k, while the second corresponds to $k' = k + \omega$, where ω is a random variable uniformly distributed in the range $[-\epsilon, \epsilon]$. In general, $\widetilde{N}(\epsilon)$ depends on the initial condition (x_0, y_0). For any given k and L we compute the average time $\langle \widetilde{N}(\epsilon) \rangle$ and the typical time $\exp\langle \ln \widetilde{N}(\epsilon) \rangle$; angle brackets indicate averages taken over the random variable ω, and initial conditions or, equivalently, the time evolution along a trajectory. Experimental data obtained by this procedure are reported in Fig. 1.

Numerical evidence supports the following picture: at fixed L, the first-error time is roughly constant for large values of the error ϵ, while it shows a power-law behavior with exponent very close to one for smaller errors ϵ. The transition between the two regimes occurs at a critical value $\epsilon_c(L)$ which scales as $\epsilon_c \sim 1/L$. In formulas:

$$\langle \widetilde{N}(\epsilon) \rangle \sim \begin{cases} O(1) & \text{for } \epsilon > \epsilon_c(L) \\ (1/\epsilon)^{\alpha} & \text{for } \epsilon < \epsilon_c(L), \end{cases} \tag{19}$$

with exponent $\alpha \simeq 1$.

It is rather easy to give analytical estimates supporting the numerical evidence. While elementary, they help us to clarify the reasons behind the very narrow regions of chaoticity of the system (5). Because the unique parameter defining the model is k, it is immediate that the value of y_{n+1} is correct, if x_{n+1}, y_n are such. Errors can therefore affect only x_{n+1}. Now, letting $k = k^T + \omega$,

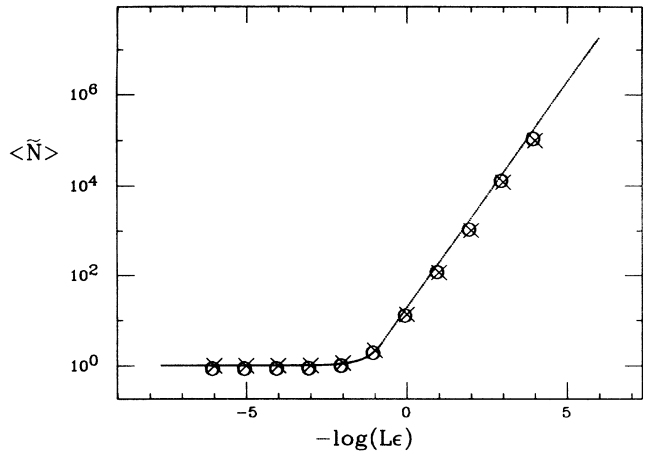

FIG. 1. $\langle\widetilde{N}(\epsilon)\rangle$ as a function of $-\log_{10}(\epsilon L)$, for the discretized standard map with $k = 7.3$ for $L = 10^5$ ($\circ$) and $L = 10^7$ ($\times$). The full line is the curve $P(\epsilon)^{-1}$ with $P(\epsilon)$ given by Eq. (23).

$z = x_n + k\sin y_n$, $z^T = x_n + k^T \sin y_n$, we see that an error will occur when z and z^T lie in different "cells" of size $l := \frac{2\pi}{L}$ in the (periodic) circle. We can safely assume that the values of z^T are uniformly distributed over the circle, and therefore the cells are visited with equal probability. The random variable $\xi = \omega \sin y_n$ takes values in the interval $[-\eta, \eta]$, where $\eta = \epsilon \sin y_n$ for $0 \le y_n < \pi$. The case $\pi \le y_n < 2\pi$ is similar.

An error will then be formally defined as the following random event:

$$\mathrm{mod}\,(z^T + \xi, l) \neq \mathrm{mod}\,(z^T, l). \tag{20}$$

Since z^T and ξ are uncorrelated, the error probability $p(\eta)$ can be computed exactly, with the result

$$p(\eta) = \begin{cases} \frac{\eta}{2l} & \text{for } \eta < l \\ 1 - \frac{l}{2\eta} & \text{for } \eta > l. \end{cases} \tag{21}$$

The case $\pi \le y_n < 2\pi$ leads to an identical result.

The simple average (21) must now be completed taking account that the y_n are also uniformly distributed in the unit cell, and can be considered uncorrelated from the x_n, as far as k is large, and taking integrals is our only task. With this in mind, we can compute the global error probability $P(\epsilon)$,

$$P(\epsilon) = \frac{1}{\pi}\int_0^\pi p(\epsilon \sin y)\, dy. \tag{22}$$

Performing the integral leads to

$$P(\epsilon) = \begin{cases} \frac{\epsilon}{\pi l} & \text{for } \epsilon < l \\ \frac{\epsilon}{\pi l}(1 - \cos\theta) + 1 - \frac{2\theta}{\pi} + \frac{l}{\pi\epsilon}\ln\tan(\frac{\theta}{2}) & \text{for } \epsilon > l. \end{cases} \tag{23}$$

where the angle θ is defined via $\sin\theta = l/\epsilon$, and we remind the reader that $l := \frac{2\pi}{L}$. Finally, we can assume that an error at time n happens independently of the past dynamical history, so that the average waiting time for the occurrence of an error is the inverse of its probability, $\widetilde{N}(\epsilon) = P(\epsilon)^{-1}$. Numerical simulations show that the behavior proposed in Eq. (19) holds. The curve $P(\epsilon)^{-1}$ is plotted in Fig. 1, in comparison with the experimental data.

To have a comparison with the continuum dynamics of the usual standard map, we have computed the time $N_\Delta(\epsilon)$ required for two trajectories to reach lattice points farther than a fixed distance Δ in the discrete phase space of Eq. (5). Since Δ is defined by rescaling with $1/L$ the Euclidean distances on the integer square lattice of side L, it is appropriate to the large-L limit. We found

$$\langle N_\Delta(\epsilon)\rangle \sim \begin{cases} \ln(\Delta/\epsilon) & \text{for } \epsilon > \epsilon_c(L) \\ (1/\epsilon)^\alpha & \text{for } \epsilon < \epsilon_c(L), \end{cases} \tag{24}$$

again with $\alpha \simeq 1$. In the above, as usual, we just report the scaling behavior of the quantity under observation. We remark that when $\epsilon < \epsilon_c(L)$, $N_\Delta(\epsilon)$ is roughly independent of Δ [which contributes only a logarithmic correction to Eq. (24)].

The detail of the behavior of $N_\Delta(\epsilon)$ for large ϵ is shown in Fig. 2. Since the typical time and the average exit time from a ball of radius Δ are very close, intermittency is negligible (if present) and we can estimate the quantity $\langle 1/N_\Delta\rangle$ versus ϵ from the data of the figure. This investigation reveals the Lyapunov exponent λ of the continuous standard map. In fact, if ϵ is large, the first error is almost always instantaneous, and two "parted" trajectories immediately experience the typical instability of the system, so that

$$\langle N_\Delta(\epsilon)\rangle \simeq \frac{1}{\lambda}\ln\left(\frac{\Delta}{\epsilon}\right). \tag{25}$$

This is just another verification of the similarity of the effect of a small disturbance on the equations of motion

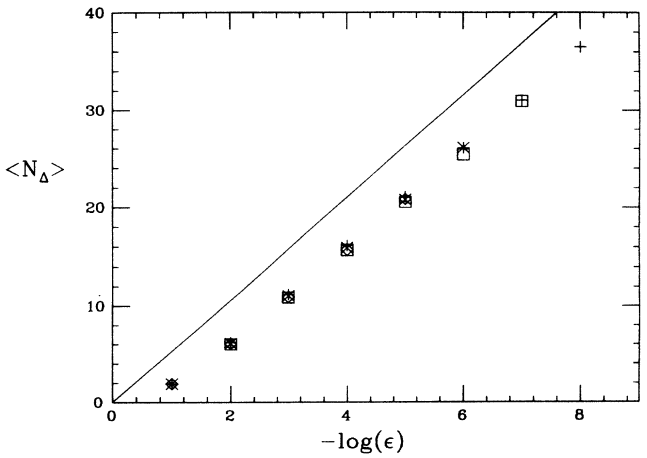

FIG. 2. $\langle N_\Delta(\epsilon)\rangle$ as a function of $-\log_{10}(\epsilon)$, for the discretized standard map with $k = 2$ for $\Delta = 0.05$ and $L = 10^5$ ($\Diamond$), $L = 10^6$ ($\times$), $L = 10^7$ ($\Box$), and $L = 10^8$ ($+$). The full line is $N_\Delta = (1/\lambda)\ln(1/\epsilon)$ where λ is the maximal Lyapunov exponent of the continuous standard map.

and of a small error in the initial conditions for a dynamical evolution, discussed in Ref. [25].

These results unveil the nature of the dynamics of this discrete system: its trajectories are incompressible and therefore chaotic only for large values of ϵ, the cutoff value decreasing as $1/L$. This helps us also to understand the extent to which the dynamics of the discrete standard map eq. (5) is equivalent to its continuum counterpart. When k is large, and $\epsilon > \epsilon_c$, the two systems possess chaotic trajectories. Yet, the maximum amount of complexity available in this regime to the discrete system is attained at the break-even point $\bar{N}$ between the trivial and the dynamical code. Simple calculations show that, to the leading order, $\bar{N} \sim \log_2(L)$. Up to this time, trajectories are informationally complete, hence chaotic. Elapsing this brief instant of absolute freedom, the discrete system appears "regular," i.e., compressible and predictable. Therefore continuous and discrete systems are similar (as far as chaos is concerned) only over logarithmically short times.

For longer times, one has that

$$l_{\Sigma} \sim \exp(l_c). \tag{26}$$

This logarithmic simplicity is well explained by the simple calculations reported above: errors occur here because "missing" of a lattice point takes place; precision in phase space targeting is proportional to ϵ^{-1}, and spacing of phase space points scales as L^{-1}.

It is important to stress that the system appears "regular" on time scales much smaller than the typical period $T_L \sim L^{D/2}$.

It is now clear why the very same features are also characteristic of the quantum dynamics of the Arnol'd cat map, and most likely of the generic low-dimensional quantum system. In fact, in the quantum version of a classically chaotic system, a semiclassical wave function chosen as initial condition follows a dynamics which is approximately classical up to a time t_c given by

$$t_c \sim \frac{1}{\lambda} \ln\left(\frac{I}{\hbar}\right), \tag{27}$$

where λ is the Lyapunov exponent of the classical motion, and I is a typical action. Over this time, the system develops all its complexity. For longer times the behavior becomes regular, and predictable. The existence of this cutoff time can be determined by complexity arguments analogous to those exposed in this work [27,28].

A quite similar state of things is found when examining the dynamics of neural networks, Eq. (6). The parameters defining the model are the matrix entries J_{ij}. They can be decomposed into a symmetric and an antisymmetric part:

$$J_{ij} = J^s_{ij} + \gamma J^a_{ij}, \tag{28}$$

$$J^s_{ji} = J^s_{ij}, \qquad J^a_{ji} = -J^a_{ij}, \qquad J_{ii} = 0. \tag{29}$$

In Eq. (29) we have also requested the diagonal elements to be null. The parameter γ determines the degree of asymmetry of the couplings. The model (6), (28), (29) is obtained by a determination of the $C = L(L-1)+1$ parameters of the model: the matrix elements of J and the parameter γ. The former can be arbitrarily specified. Indeed, they will be taken to be a realization of a sequence of quenched Gaussian variables of zero mean value and variance $1/(L-1)$. The scaling of the variance with L ensures that each of the quantities $\sum_j J_{ij}\sigma_j$ in (6) is of order $O(1)$ for any L. This model derives its relevance from the existence of a transition from a dynamics characterized by short periodic motions to one with exponentially long transients, occurring at $\gamma_c \simeq 0.5$ [29,19].

Numerical simulations are performed at a fixed value of γ larger than γ_c, comparing a reference dynamics generated by the coupling J_{ij} with that induced by the couplings $J'_{ij} = J'^s_{ij} + \gamma J'^a_{ij}$, where $J'^{(a,s)}_{ij} = J^{(a,s)}_{ij} + \epsilon \delta J^{(a,s)}_{ij}$, and the random quenched variables $\delta J^{(a,s)}_{ij}$ have the same statistical properties of $J^{(a,s)}_{ij}$. The results reported in Fig. 3 show that

$$\langle \tilde{N}(\epsilon) \rangle \simeq \exp\langle \ln\, \tilde{N}(\epsilon) \rangle \sim \begin{cases} O(1) & \text{for } \epsilon > \epsilon_c \\ (1/\epsilon)^{\alpha} & \text{for } \epsilon < \epsilon_c, \end{cases} \tag{30}$$

with a power-law exponent $\alpha \simeq 1$. A transition from chaotic to compressible dynamics is observed also here, with a cutoff ϵ_c which scales as L^{-1}.

Also in this case, we can provide a simple argument to explain the scaling of $\langle \tilde{N}(\epsilon) \rangle$ as a function of ϵ. Suppose that, at time t, one has $\sigma'_i = \sigma_i$, for all i. In order for $\sigma'_i(t+1) = -\sigma_i(t+1)$ for some i at time $t+1$ it is necessary and sufficient that

$$h_i(t) h'_i(t) < 0, \tag{31}$$

where $h_i(t) = \sum_j J_{ij}\sigma_j(t)$ and $h'_i(t) = \sum_j J'_{ij}\sigma_j(t)$. By introducing the variables

$$x = \sum_j J_{ij}\sigma_j(t), \qquad y = \sum_j \delta J_{ij}\sigma_j(t), \tag{32}$$

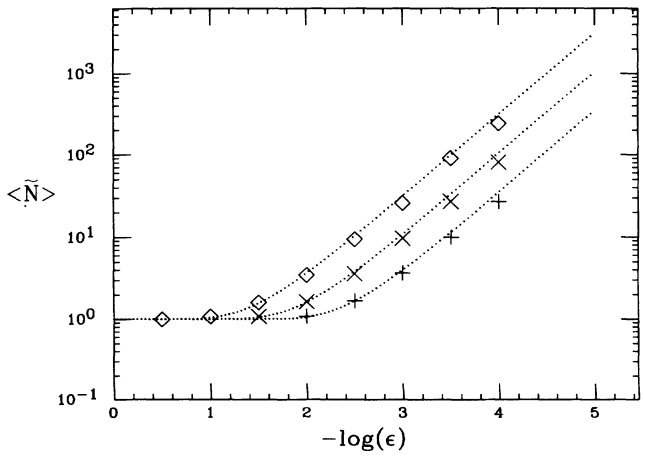

FIG. 3. $\langle \tilde{N}(\epsilon) \rangle$ as a function of $-\log_{10}(\epsilon)$, for the fully connected neural network with $\gamma = 0.9$ for $L = 100$ ($\diamond$), $L = 300$ ($\times$), $L = 900$ ($+$). The dotted lines indicate $P(\epsilon)^{-1}$ with $P(\epsilon)$ given by Eq. (35).

one has $h_i(t) = x$ and $h_i'(t) = x+\epsilon y$, so that the condition (31) is satisfied in the region in the x-y plane delimited by the y axis and the line $y = -x/\epsilon$. We denote by ϕ the angle of this angular sector:

$$\phi = \frac{\pi}{2} - \arctan\left(\frac{1}{\epsilon}\right). \tag{33}$$

For large values of L, x and y are Gaussian variables with zero mean and unit variance. Moreover, $\langle xy\rangle = \sum_j J_{ij}\delta J_{ij}$. For a typical δJ_{ij} one has $\langle xy\rangle \sim L^{-1/2}$. Therefore the probability that condition (31) for the spin i be satisfied is

$$P_i = \int\int_A P(x,y)dxdy = \frac{\phi}{\pi}, \tag{34}$$

where A is the angular sector defined above, and $P(x,y) = (1/2\pi)\exp[-(x^2+y^2)/2]$. For $\epsilon \ll 1$, $P_i \sim \epsilon$ and, for large ϵ, $P_i \simeq 1/2$. Since h_i is independent of h_j for $j \neq i$ one obtains that the probability of the event $\sigma_i(t+1) \neq \sigma_i'(t+1)$, for some i, is given by

$$\begin{aligned} P(\epsilon) &= 1 - [1 - P_i(\epsilon)]^L \\ &= 1 - \left[1 - \frac{\pi}{2} + \arctan\left(\frac{1}{\epsilon}\right)\right]^L \\ &\sim \begin{cases} L\epsilon & \text{for } \epsilon \ll \frac{1}{L} \\ 1 & \text{for } \epsilon \gg \frac{1}{L}. \end{cases} \end{aligned} \tag{35}$$

We can therefore derive the large-L behavior

$$\langle \widetilde{N}(\epsilon)\rangle = P(\epsilon)^{-1} \sim \begin{cases} \frac{1}{L\epsilon} & \text{for } \epsilon \ll \frac{1}{L} \\ 1 & \text{for } \epsilon \gg \frac{1}{L}, \end{cases} \tag{36}$$

an estimate which is in good agreement with numerical computations.

To complete the panorama of stochastic tests on the motion of our neural network, we have computed the rate of damage spreading. To do this, we consider two trajectories $\underline{\sigma}$ and $\underline{\widetilde{\sigma}}$, generated by means of the rule (6), with the same couplings $\{J_{ij}\}$, but with two minimally different initial conditions: $\sigma_1(0) = -\widetilde{\sigma_1}(0)$ and $\sigma_i(0) = \widetilde{\sigma}_i(0)$ for $i \neq 1$. This minimal difference is obviously related to a particular metric, the so-called bit-counting or Hamming distance of information theory:

$$d(\underline{\sigma},\underline{\widetilde{\sigma}}) = \sum_{i=1}^{L} |\sigma_i - \widetilde{\sigma}_i|. \tag{37}$$

An exponential increase in time of the Hamming distance is somehow analogous to positivity of the maximal Lyapunov exponent, although—together with remarks already presented—it must be observed that this metric may not always have a clear physical significance.

We have studied the time evolution of the so-called damage, $D(n)$, which is simply defined in terms of the Hamming distance:

$$D(n) = 1 - \frac{1}{L}\sum_{i=1}^{L} \sigma_i(n)\widetilde{\sigma}_i(n) = \frac{d(n)}{2L}. \tag{38}$$

Numerical simulations are shown in Fig. 4: the linear growth of damage spreading is evident.

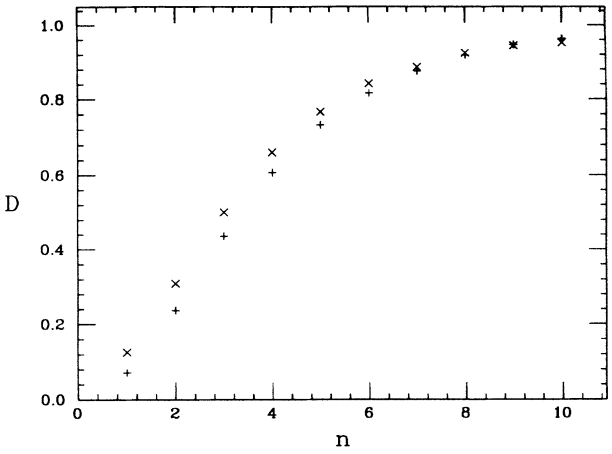

FIG. 4. $D(n)$ as function of n for the fully connected neural network, with $\gamma = 0.9$, for $L = 100$ (×), $L = 300$ (+).

V. CONCLUSIONS: DETERMINISTIC RANDOMNESS AND IRREGULAR BEHAVIOR

It is generally believed that standard theoretical concepts are enough to catch the essence of chaos, even in complex systems. On the basis of this belief, one might think that the heuristic complexity arguments presented in this paper do not add any new piece of information to what can be described, e.g., via the computation of Lyapunov exponents and/or the analysis of time sequences with the embedding technique. We have shown that the situation is, on the contrary, more intriguing: there exist compressible systems, i.e., "regular," which at the same time show certain behaviors typical of "chaotic" systems. In addition, a transition from incompressible to compressible dynamics takes place in these systems, when either the "bit precision" ϵ of a computer or the "length scale" $\Delta x \sim 1/L$ of a dynamical system are varied, keeping the other parameters constant.

These facts occur in two significant representatives of important and distinct classes of dynamical systems in finite spaces: discrete Hamiltonian flows, and neural networks. In the former, the lattice standard map of eq. (5), time sequences "seems" irregular, correlations decay to zero very quickly with increasing time delay, and yet we have shown the logarithmic simplicity of these trajectories. Distinct stochastic aspects are also present in our second example, the neural network model of Eqs. (6), (28), (29), when $\gamma > \gamma_c \simeq 0.5$: time correlations decay exponentially fast to zero and Shannon entropy of the sequences $\sigma_i(0),\ldots,\sigma_i(n)$ is positive. Yet at the same time we have shown that trajectories are compressible and hence regular. The additional detail that spreading of a "small" error occurs at a linear rate adds the final touch to the picture. Similar contrasting details have been recently found by Politi *et al.* in chains of coupled stable maps. In spite of a negative maximum Lyapunov exponent they find an irregular time behavior with exponential decay of correlation functions [30].

In the presence of this clouded horizon of stochastic tests, providing ambiguous evidence, but each possessing

its piece of truth, we have adopted deterministic randomness as the basis for our definition of chaos in systems with finite number of states, following Alekseev and Ford. We believe that this perspective deserves investigation in many respects. First, our present attempt at the description of complex systems can surely be improved, and more detailed analyses performed. Algorithmic complexity is a theory of general purpose, and its generality is such as to allow various applications in the physical sciences: all it requires are symbolic sequences clearly defined and motivated on physical grounds. Moreover, this investigation can lead to the development of new coding and decoding machines just by using familiar dynamical systems: many applications of neural networks and cellular automata have already been proposed for this scope (for a simple example of a coding dynamical system, which also includes a complexity analysis, see [31]).

Generality is another desirable property present in this approach: since we have started from Alekseev's treatment of traditional low-dimensional classical systems, deterministic randomness as defined herein becomes a unifying definition of chaos in a vast class of physical systems. This class can also include quantum mechanical systems. In fact, many of the features observed in this work are also present in the transition from quantum to semiclassical mechanics, and affect the deceitful appearance of *quantum chaos* [27].

Finally, observe that in Sec. III we have introduced our working definition of chaos (via the heuristic concept of realistic computer) as a means of verifying the conditions of the rigorous definition of Sec. II. Yet, following purely algorithmic considerations, we have arrived at a "practical" definition which encompasses the idea of sensitive dependence on the parameters of a model. This natural substitute of sensitive dependence on initial conditions for systems with discrete states is indeed an appreciated gift from our precursors.

ACKNOWLEDGMENTS

We thank P. Grassberger for the careful reading of the paper and for the interesting discussions on the complexity in dynamics. G.M. was supported under Grant No. 58/43/87 from ENEA, Area Energia de Innovazione, Divisione Calcolo. M.F. and A.V. are supported by INFN "iniziativa specifica Meccanica Statistica."

[1] S. Wolfram, Commun. Math. Phys. **95**, 15 (1984); Phys. Rev. Lett. **54**, 735 (1984).
[2] D.J. Amit, *Modeling Brain Function* (Cambridge University Press, Cambridge, England, 1989).
[3] Robert L. Devaney, *An Introduction to Chaotic Dynamical Systems* (The Benjamin/Cummings Publishing Co., Inc., Menlo Park, CA, 1986).
[4] H. Aref, J. Fluid Mech. **143**, 1 (1984); A. Crisanti, M. Falcioni, G. Paladin, and A. Vulpiani, Riv. Nuovo Cimento **14**, 1 (1991).
[5] Ya. Sinai, *Introduction to Ergodic Theory* (Princeton University Press, Princeton, NJ, 1976).
[6] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983); G. Paladin and A. Vulpiani, J. Phys. A **19**, 997 (1986).
[7] V.I. Arnol'd and A. Avez, *Ergodic Problems of Classical Mechanics* (W.A. Benjamin, Inc., New York, 1968).
[8] J. Ford, G. Mantica, and G.H. Ristow, Physica D **50**, 493 (1991).
[9] S.A. Kauffman, J. Theor. Biol. **22**, 437 (1969); B. Derrida and A. Stauffer, Europhys. Lett. **2**, 739 (1986).
[10] A.K. Zvonkin and L.A. Levin, Usp. Mat. Nauk. **25**, 85 (1970).
[11] G. Chaitin, *Information, Randomness & Incompleteness* (World Scientific, Singapore, 1987).
[12] V.M. Alekseev and M.V. Jacobson, Phys. Rep. **75**, 287 (1981).
[13] A. Brudno, Russ. Math. Surv. **33**, 197 (1978).
[14] J. Ford, Phys. Today **36**, (4), 40 (1983).
[15] B. Eckhardt, J. Ford, and F. Vivaldi, Physica D **13**, 339 (1984).
[16] J. Ford, in *Chaotic Dynamics and Fractals*, edited by M.F. Barnsley and S. Demko (Academic Press, New York, 1986).
[17] B.V. Chirikov, F.M. Izrailev, and D.L. Shepelyansky, Sov. Sci. Rev. C **2**, 209 (1981).
[18] J.H. Hannay and M.V. Berry, Physica D **1**, 267 (1980).
[19] A. Crisanti, M. Falcioni, and A. Vulpiani, J. Phys. A **26**, 3441 (1993).
[20] E. Fredkin and T. Toffoli, Int. J. Theor. Phys. **21**, 219 (1982).
[21] H. Rogers, Jr., *Theory of Recursive Functions and Effective Computability* (McGraw-Hill, New York, 1967).
[22] H. Woźniakowski, Annu. Rev. Comput. Sci. **1**, 319 (1986).
[23] I. Percival and F. Vivaldi, Physica D **25**, 105 (1987); **27**, 373 (1987).
[24] B. Harris, Ann. Math. Stat. **31**, 1045 (1960); F. Rannou, Astron. Astrophys. **31**, 289 (1974); Y.E. Levy, Phys. Lett. **88**, 1 (1982); C. Grebogi, E. Ott, and J.A. Yorke, Phys. Rev. A **38**, 3688 (1988).
[25] A. Crisanti, M. Falcioni, and A. Vulpiani, Physica A **160**, 482 (1989).
[26] C. Beck, Phys. Lett. A **136**, 121 (1989).
[27] G. Mantica and J. Ford, in *Quantum Chaos — Quantum Measurement*, Vol. 358 of *NATO Advanced Study Institute, Series C*, edited by P. Cvitanović, I. Percival, and A. Wirzba (Kluwer, Dordrecht, 1992).
[28] J. Ford and G. Mantica, Am. J. Phys **60**, 1086 (1992).
[29] K. Nützel, J. Phys. A **21**, L151 (1991).
[30] A. Politi, R. Livi, G.-L. Oppo, and R. Kapral, Europhys. Lett. **22**, 571 (1993).
[31] G. Mantica and A. Sloan, Complex Sys. **3**, 37 (1989); G. Mantica and B.G. Giraud, Chaos **2**, 225 (1992).