PHYSICAL REVIEW E, VOLUME 63, 065203(R)

Multiscale analysis of complex systems

P.-M. Binder* and Jorge A. Plazas
Departamento de Física, Universidad de Los Andes, Apartado Aéreo 4976, Bogotá, Colombia
(Received 7 November 2000; revised manuscript received 28 February 2001; published 29 May 2001)

We calculate block information versus size profiles for two-symbol strings generated by several dynamical processes: random, periodic, regular language, and substitutive. The profiles provide a good diagnostic of the complexity of the strings.

DOI: 10.1103/PhysRevE.63.065203 PACS number(s): 05.45.−a, 05.10.−a, 89.75.−k, 89.70.+c

Despite considerable effort, there is no consensus on an ideal measure of complexity, or even on what exactly a complex system is (see [1,2] for extensive recent reviews). Candidate complex systems include neural networks, finite-alphabet strings generated by diverse computational processes, cellular automata, genetic algorithms and other adaptive systems, living organisms, evolutive systems, and social organizations [3]. The approaches to quantifying complexity include symbolic dynamics, information and ergodic theory, thermodynamics, generalized dimensions and entropies, theory of computation, logical depth and sophistication, forecasting measures, topological exponents, and hierarchical scaling [2,4–6,16]. All the above help us to understand a rich, important, and yet elusive concept.

In this Rapid Communication we apply a multiscaling approach to several sequences of zeros and ones generated by a variety of processes: periodic, random, deterministic finite automata, and substitutive languages, all of interest as dynamical systems. By plotting Shannon information obtained with different block sizes, we obtain information-scale profiles that allow us to distinguish between all the above classes of strings. We observe a variety of functional forms that help us integrate several of the previous approaches, in particular topological exponents, effective measure and statistical complexities [5,6], and the von Mises definition of randomness [7].

One among the early candidate measures of complexity, Shannon entropy was soon rejected because it is more a measure of randomness than of the structure inherent in complexity (see Ref. [8]). Simple functions of entropy have also been proposed, and have drawn criticism [9]. Our contribution is to measure Shannon entropy as a function of (length or time) scales, as will be explained below.

Our approach has similarities to and differences from the complexity profiles of Ref. [1]. In that work, Bar-Yam considers the entropy $S(L)=\log\Omega(L)$, where $\Omega(L)$ is the number of microstates of a physical system discernible at a given level of resolution L, which can be space or time. In this view, it is possible to discern *more* microstates in a system observed at small L, so that S is a *decreasing* function of L. Instead, we consider *single* two-symbol experimental time sequences to which one has access. We look at these sequences in increasingly wide windows of sizes $n=1,2,\ldots,$ and calculate the probability p_b of each of the possible 2^n blocks. With these, we compute the quantity $I(n)=-\Sigma p_b \log p_b$, where log is understood to be base 2. $I(n)$ is also known in the literature as H_n, the block entropy. It is a measure of the available information we obtain about the structure of the string by looking at it at scale n. For small enough n, it increases monotonically with n, since wider windows must give at least as much information as narrower ones.

Next, we describe the processes by which we have generated strings. We show the first 40 symbols of a typical string within each class. (a) Random. These are uniformly random sequences of zeros and ones, denoted by R, generated with a recently proposed yet well-tested random number generator [10]. Having good generators is of great importance in physics simulations [11]:

$$R=0001101101100010100011001101000001000010\ldots. \tag{1}$$

(b) Periodic. These are sequences with lowest period $N=4,8,16,\ldots,$ denoted by $P4,P8,\ldots$ respectively. These were generated by repeating subsequences of $N-1$ zeros and one 1. For example,

$$P4=0001000100010001000100010001000100010001\ldots. \tag{2}$$

(c) Regular languages. These can also be described by regular expressions [2]. They can be generated by finite-state directed graphs, or deterministic finite automata. New words are generated by adding symbols at the end of allowed shorter words. Regular languages in many cases represent chaotic processes [6,12,13].

We use the following three examples. $D1$ is an eight-state automaton that describes the symbolic dynamics of the logistic equation (with $r=3.7$), known to be in the same universality class as several systems of physical interest (e.g., lasers and fluids). The automaton appears in [6]:

$$D1=1111101011101011111011010101111101011101\ldots. \tag{3}$$

$D2$ appears in Fig. 2 of Ref. [12] and has four states:

$$D2=1111101111110111101010111101111101010111\ldots. \tag{4}$$

*Author to whom correspondence should be addressed. Electronic address: p@faoa.uniandes.edu.co

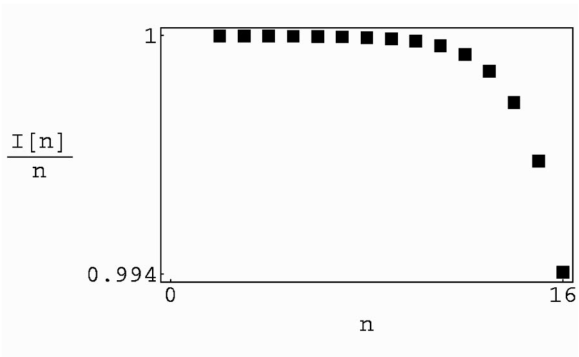

FIG. 1. $I(n)/n$ profiles for random sequence R.

$D3=(101)(00+1)^*(101)$, where + denotes or, and can also be produced by the ten-state automaton shown in Fig. 9.2 of Ref. [2]:

$$D3=110110111101010101111010111010110110111\ldots . \tag{5}$$

Finally, (d) substitutions (S). These are string-generation processes in which a substring is replaced by a longer substring, following an allowed set of rules. Unlike the previous cases, the string can grow at any place, not just the end. In particular, we will consider Morse-Thue sequences [14], associated with the period-doubling transition from order to chaos in certain dissipative systems:

$$S=101011110010111010111101011110101011001 0\ldots . \tag{6}$$

The above processes are typical of a wide variety of behaviors of dynamical systems. Random sequences help simulate noise (for example, thermal), periodic sequences represent oscillating systems, regular sequences describe some chaotic processes, and substitutions provide the simplest description of the order-chaos transition in the period-doubling route. Moreover, a parallel with computational languages classes [15] exists: the substitutive process we study is in the context-free class, while the others are regular languages.

Before presenting our information-block size profiles, we discuss briefly the hierarchical definition of complexity introduced by Badii and Politi [2,16]. Their characterization relies on topological entropy, defined for the set of $N(m)$ allowed strings of length m, as $\lim_{m\to\infty}(1/m)\log N(m)$. The first exponent $C^{(1)}$ is the topological entropy of the set of of allowed words. Hence, it is a measure of the cardinality of the language. For higher exponents, one must find the set of irreducible forbidden words of the language, and the topological entropy of this set yields $C^{(2)}$. An irreducible word is one that cannot be decomposed into subwords strictly shorter than itself. $C^{(2)}$ is a measure of the difficulty of approximating the original language through subshifts of finite type with increasing memory (see [2], pp. 255-260). Next, one finds the topological entropy of the irreducible forbidden words in the set of irreducible forbidden words, which yields $C^{(3)}$, and so on. Each additional exponent gives a measure of the difficulty of further compressing the rules in the language.

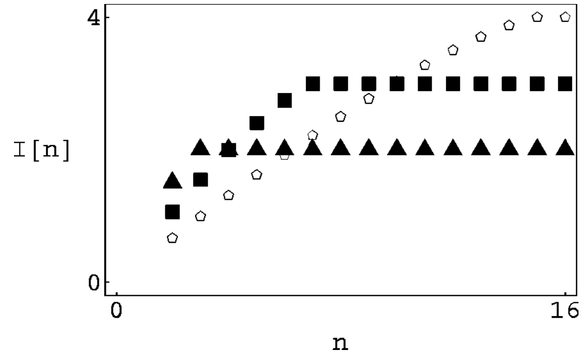

FIG. 2. $I(n)$ profile for periodic sequences $P4$ (triangles), $P8$ (squares), and $P16$ (open circles).

It is expected that $C^{(k+1)}\leqslant C^{(k)}$, and that eventually the topological exponents for a language become zero starting with some given integer k. This procedure yields $C^{(1)}=0$ for all periodic sequences, and $C^{(2)}=0$ for all random sequences. Generally, regular languages and expressions have nonzero $C^{(1)}$ but zero $C^{(2)}$, although exceptions are known such as the $D3$ language introduced above. The reasons for the exceptions are usually well understood [2]: in this particular case, the initial and terminal sequences (101) do not allow for total compression of irreducible forbidden words. Higher-level languages such as substitutions often have $C^{(2)}\neq 0$.

Now we present results of information-size profiles obtained for the processes described above. Several strings of length 2^{15} were studied for each process. Error bars are about the size of the symbols. Depending on the case, we plot either $I(n)$ or $I(n)/n$ versus n.

Figure 1 shows $I(n)/n$ profiles for the R sequence. A perfectly random series, according to the von Mises definition [7], would have perfectly equidistributed blocks of all sizes, and hence $I(n)/n=1$. In this figure we observe that this is almost the case for $n\leqslant 10$, with subtle yet statistically significant deviations for larger n. Hence, multiscale analysis provides a useful, first-principles random number generator test.

Figure 2 shows $I(n)$ profiles for three P sequences, of periods $N=4,8,16$, respectively. We see that information saturates near $n=N$. This is because only N block types are

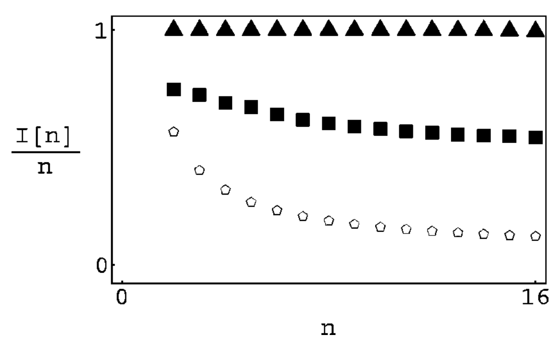

FIG. 3. $I(n)/n$ profiles for regular expressions $D1$ (squares) and $D2$ (open circles). Sequence R (triangles) is also shown for comparison.

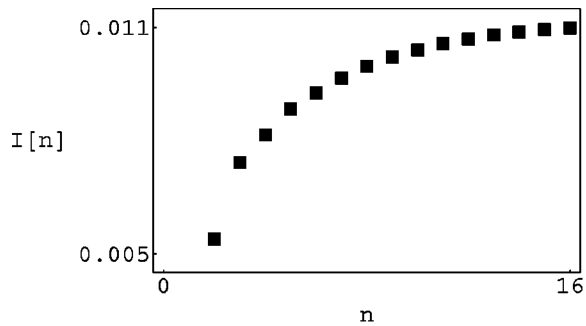

FIG. 4. $I(n)$ profile for regular expression $D3$.

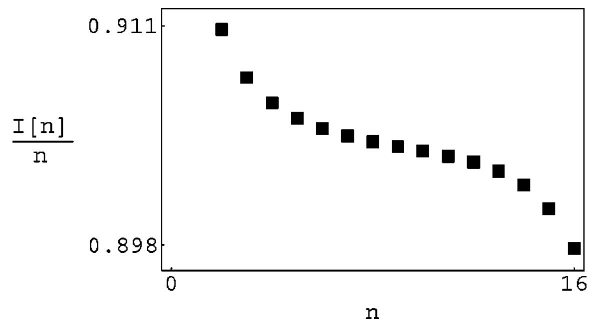

FIG. 5. $I(n)/n$ profile for Morse-Thue sequence *S*.

sampled for all sizes larger than *N*. Clearly, $I(n)/n \to 0$ for large *n*, consistent with the low cardinality of blocks, which implies $C^{(1)} = 0$.

Figure 3 shows $I(n)/n$ profiles for sequences $D1$ and $D2$, along with *R* for comparison. We recall that for these $C^{(1)} \neq 0$ (the allowed words tend to a constant fraction), as discussed by Badii and Politi [2,16]. This is reflected in the profiles, in which $I(n)/n$ tends to a constant value. As discussed by Grassberger [5], and also in [2], pp. 112 and 254, this saturation indicates a finite effective complexity, which is the average information needed to specify the $(n+1)$th symbol, given the previous *n* symbols. This convergence is typical of regular languages. Note, however, that language $D2$ has a transition from the starting state to itself (transient), which causes a slow convergence to the final value of $I(n)/n$.

Figure 4 shows the $I(n)$ profile for process $D3$. One should observe asymptotically a straight line, which up to $n = 15$ has not appeared. This is consistent with the finding that $C^{(2)} \neq 0$, as discussed above.

Last, Fig. 5 shows an $I(n)/n$ profile for the Morse-Thue substitutive language associated with period-doubling accumulation points. For this system $C^{(1)} \sim 10/3$ (see [2], p. 83), and $C^{(2)} = 0$. In a consistent fashion, the complexity profile shows a higher degree of structure for the values of *n* studied than the other sequences, while it decays for long *n* consistently with the zero value for the second topological exponent. The inflection point near $n = 8$ appears to be new. A second example, the Fibonacci system studied in [2], also shows nonmonotonic behavior of $I(n)/n$.

In summary, we find that information-block size profiles capture the properties of strings generated by dynamical processes of widely different degrees of complexity. While our profiles do not take explicitly into account computer theoretical quantities such as irreducible forbidden words, they seem to accurately reflect the properties of at least the first two topological exponents in the hierarchy proposed by Badii and Politi [16]. The profiles also seem to adequately recognize the existence of typical regular languages, in particular the convergence of the slope of $I(n)$ versus *n*, related to the effective measure complexity of these languages. Exceptional regular languages such as $D3$ can also be discerned, as can strings in higher language classes such as *S*, both characterized by profile functions with a higher degree of structure. Finally, the profiles serve as a random number generator test, which coincides with an eighty-year-old foundational definition of randomness, and which has identified subtle deviations from ideality in a top-quality random number generator. Figure 1 shows that from blocks of length $n > 10$ the $n+1$th bit starts to become slightly predictable.

In our profiles we follow the ideas of Bar-Yam [1] of an alternative to a definition of complexity, which consists not of one (as in effective measure complexity or statistical complexity, [5,13]), or a set of two or three numbers (as in the topological exponents), but rather a complete description of the system across scales. Our choice of information (or block entropy) is one of the simplest ones possible, and is particularly appropriate for finite-alphabet strings generated by specific processes and obtainable through a single time series measurement, as entropy [1] was found to be appropriate for physical objects. We have chosen processes in several well-known classes of dynamical systems, and shown that the power of our profiles to distinguish processes is quite high.

We conclude with three remarks (1) In this paper we have used complexity profiles to characterize temporal sequences generated by dynamical systems. An obvious extension would be to consider spatio-temporal sequences from systems out of equilibrium [17]. (2) Another obvious direction would be to extend the analysis from finite-alphabet strings to time series of real-valued measurements. (3) So far our approach is phenomenological, and similar in spirit, for example, to the $f(\alpha)$ description of a multifractal [18]. Ideally, one should use the results as a stepping stone to understand the underlying dynamical mechanisms that generate the sequence. In this respect, Kullback-Leibler entropy or Fisher information [19] may be appropriate tools.

The authors thank R. Badii, Y. Bar-Yam, and P. L'Ecuyer for useful discussions and Colciencias for for financial support (Contract No. RC-45-2000).

[1] Y. Bar-Yam, *Dynamics of Complex Systems* (Addison-Wesley, Reading, MA, 1997).
[2] R. Badii and A. Politi, *Complexity: Hierarchical Structure and Scaling in Physics* (Cambridge University Press, Cambridge, UK, 1997).
[3] See D.J. Amit, *Modeling Brain Function: The World of Attrac-*

tor Neural Networks (Cambridge University Press, Cambridge, UK, 1989); V.M. Alekseev and M.V. Yakobson, Phys. Rep. **75**, 287 (1981); *Cellular Automata: Theory and Experiment*, edited by H.A. Gutowitz (MIT Press, Cambridge, MA, 1991); M. Mitchell, *An Introduction to Genetic Algorithms* (MIT Press, Cambridge, MA, 1996); K. Sigmund, *Games of Life* (Oxford University Press, Oxford, UK, 1993); S.A. Kauffman, *The Origins of Order: Self Organization and Selection in Evolution* (Oxford University Press, New York, 1993); H. Mintzberg, *The Structuring of Organizations* (Prentice-Hall, Englewood Cliffs, NJ, 1979).

[4] See V.M. Alekseev and M.V. Yakobson, Phys. Rep. **75**, 287 (1981); C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems* (Cambridge University Press, Cambridge, UK, 1993); G. D'Alessandro and A. Politi, Phys. Rev. Lett. **64**, 1609 (1990); S. Lloyd and H. Pagels, Ann. Phys. (San Diego) **188**, 186 (1988); C.H. Bennett, in *Complexity, Entropy and the Physics of Information* (Addison-Wesley, Redwood City, CA, 1990), p. 137; M. Koppel and H. Atlan, Inf. Sci. (N.Y.) **56**, 23 (1991).

[5] P. Grassberger, Int. J. Theor. Phys. **25**, 907 (1986); P. Grassberger, Z. Naturforsch., A: Phys. Sci. **43**, 671 (1988).

[6] J.P. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1989).

[7] R. von Mises, Math. Z. **5**, 52 (1919); M. van Lambalgen, Ph.D. thesis, University of Amsterdam, 1987 (unpublished); A. Compagner, Am. J. Phys. **59**, 700 (1991).

[8] J.P. Crutchfield, Physica D **75**, 11 (1994).

[9] J.S. Shiner, M. Davison, and P.T. Landsberg, Phys. Rev. E **59**, 1459 (1999); J.P. Crutchfield, D.P. Feldman, and C.R. Shalizi, *ibid.* **62**, 2996 (2000); P.-M. Binder and N. Perry, *ibid.* **62**, 2998 (2000); J.S. Shiner, M. Davison, and P.T. Landsberg, *ibid.* **62**, 3000 (2000).

[10] P. L'Ecuyer, Oper. Res. **47**, 159 (1999).

[11] A.M. Ferrenberg, D.P. Landau, and Y.J. Wong, Phys. Rev. Lett. **69**, 3382 (1992); I. Vattulainen, T. Ala-Nissila, and K. Kankaala, *ibid.* **73**, 2513 (1994); J. Maddox, Nature (London) **372**, 403 (1994); A. Compagner, Phys. Rev. E **52**, 5634 (1995).

[12] N. Perry and P.-M. Binder, Phys. Rev. E **60**, 459 (1999).

[13] J.P. Crutchfield and C.R. Shalizi, Phys. Rev. E **59**, 275 (1999).

[14] I. Procaccia, S. Thomae, and C. Tresser, Phys. Rev. A **35**, 1884 (1987).

[15] J.E. Hopcroft and J.D. Ullman, *Formal Languages and Their Relation to Automata* (Addison-Wesley, Reading, MA, 1969).

[16] R. Badii and A. Politi, Phys. Rev. Lett. **78**, 444 (1997).

[17] M.C. Cross and P.C. Hohenberg, Rev. Mod. Phys. **65**, 2 (1993).

[18] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B.I. Shraiman, Phys. Rev. A **33**, 1141 (1986).

[19] R. Badii, Europhys. Lett. **13**, 599 (1990); P.-M. Binder, Phys. Rev. E **61**, R3303 (2000).