Self-similar sequences and universal scaling of dynamical entropies

Jan Freund,* Werner Ebeling, and Katja Rateitschak
Institute of Physics, Humboldt-University Berlin, Invalidenstraße 110, D-10115 Berlin, Germany

(Received 9 February 1996)

Symbol sequences play a prominent role in the context of symbolic dynamics. Important features of a dynamical system are reflected by related statistics of subsequences. A dynamical behavior giving rise to a self-similar attractor and universal scaling relations, expressed by critical exponents, will lead to self-similar statistics of subsequences. In the present paper we show how self-similar distributions of subsequences, i.e., temporal self-similarity, can be connected with a scaling relation for dynamical entropies. Moreover, the effect of slightly perturbing perfectly self-similar sequences by contaminating them with noise is investigated. The achieved results are of importance for physical processes marking the borderline between order and chaos. [S1063-651X(96)11110-7]

PACS number(s): 87.10.+e, 02.50.−r, 05.45.+b, 05.70.Fh

I. INTRODUCTION

The statistical analysis of chaotic or stochastic systems usually employs standard tools such as correlation function, power spectrum, or time-dependent standard deviation. The method of symbolic dynamics [1] provides the background for more sophisticated techniques, especially those related to informational measures, e.g., n-block entropies [2], Rényi entropies [3] or transinformation [4]. Moreover, these *dynamical entropies* can be applied to symbol sequences which are not directly related to dynamical systems but which are found instead as the result of evolutionary processes, e.g., the DNA, music, or texts. By defining a unique mapping of symbols onto real numbers correlation function [5], power spectrum [6,7] and time-dependent standard deviation [8] can be translated but the freedom of choosing the mapping function introduces some arbitrariness [9]. Throughout this paper we will concentrate on the concept of n-block entropies which essentially is based on Shannon's measure of syntactical information assigned to messages [10].

Let $A=\{a_1, \ldots, a_\lambda\}$ be a set of symbols which we will also call the *alphabet* In the context of symbolic dynamics subsequences $(i_1, \ldots, i_n) \in A^n$ of length n, also named *n-words*, have to be regarded as sampled trajectory segments of length $n\tau$ (stroboscopic observation with a time window τ). Assuming the sample sequence to be produced by a stationary and ergodic source [11] —which is a direct consequence of stationary and ergodic dynamics —we can do simple word counting, hence, achieving a relative frequency distribution for n-words. The relative frequencies are used to estimate the underlying n-word probabilities, in the following denoted $p(i_1, \ldots, i_n)$.

We note in passing that in practice the finite length L of the sample string poses a severe problem for this estimation. The reason is a combinatorial explosion of the number of generally possible words λ^n (respectively the effective number of words λ^{hn} [11]) which has to be outperformed by the number $[\mathcal{O}(L)]$ of n-words excerpted from the sample string. Correction formulas designed to cure this disease [12,13] do not increase the range of reliable estimation by orders. However, since we will always use an analytical approach to derive probability distributions, we do not have to face this problem here.

The prediction of an n-word is connected with an average uncertainty quantified by the Shannon measure and which will be named the n-block entropy H_n,

$$H_n := -\sum_{(i_1, \ldots, i_n) \in A^n} p(i_1, \ldots, i_n) \log_\lambda p(i_1, \ldots, i_n). \tag{1}$$

We choose $\log_2 \lambda$ *bits* as the unit of information. This choice is favorable since then the inequality $0 \leq H_n \leq n$ holds.

A quantity derived from the n-block entropies is the (n-)average uncertainty per symbol [10], denoted by $H(n) := H_n/n$. Its limit for $n \to \infty$ is named the *entropy of the source* [14],

$$h := \lim_{n\to\infty} H(n). \tag{2}$$

This quantity is closely related to the *Kolmogorov-Sinai entropy* (KS) [15,16] of a dynamical system. For a wide class of systems [17] the KS coincides with the sum of positive Lyapunov exponents according to a famous theorem by Pesin.

Furthermore, we can define the conditional entropies, denoted by h_n,

$$h_n := H_{n+1} - H_n \quad (n=1,2, \ldots). \tag{3}$$

We supplement this definition by $h_0 := H_1$. The h_n are interpreted as the average uncertainty linked with the prediction of the symbol i_{n+1} given knowledge of n preceding states $i_1, \ldots, i_n$. Correlations existing between the subsequence $(i_1, \ldots, i_n)$ and the symbol i_{n+1} will reduce this average uncertainty, hence, the series of h_n will decline when increasing the range n of noticed prehistory. It can be shown that the limit of h_n for $n \to \infty$ coincides with the limit in (2), i.e., $\lim_{n\to\infty} h_n = h$.

*Electronic address: janf@summa.physik.hu-berlin.de

The decline of the series of h_n approaching h is a characteristic feature of symbol sequences and dynamical systems, respectively, which produce them. For a *Markov source of order m* [18]—which is defined by the following property of conditional probabilities:

$$p(i_{n+1}|i_1,\ldots,i_n)=p(i_{n+1}|i_{n-m},\ldots,i_n) \quad (\forall n>m) \tag{4}$$

valid for all $(i_1,\ldots,i_{n+1})\in A^{n+1}$ —it can be proven [18] that $h_n=H_{m+1}-H_m=h$ holds true for all $n\geqslant m$. This motivates the definition of m as the *memory of the source*. Szépfalusy and co-workers [19] shared this attitude.

The scaling behavior of (h_n-h) is directly related to the *effective measure complexity* ($C_{\rm EM}$), defined by Grassberger [20],

$$C_{\rm EM}:=\sum_{n=0}^{\infty}(h_n-h). \tag{5}$$

As was noticed by Grassberger [20] and Szépfalusy [21,22], an exponential decay of the (h_n-h) is the typical case, i.e., $(h_n-h)\sim\exp(-\gamma n)$. The corresponding decay rate γ can be expressed in terms of a special Rényi entropy, namely, $\gamma=2K_3$ (for piecewise analytic 1D maps [21], see also [22]). For such systems a Markovian approximation of sufficiently high order m (4) is rather effective. Then it is reasonable to define the inverse decay rate γ^{-1} as an effective memory.

An exceptional behavior is found for systems which mark the borderline between order and chaos. They play an important role in the framework of *dynamical phase transitions* [23]. Comparatively strong correlations — which are not as trivial as in the case of a periodic sequence — exist over a wide range and cause the conditional entropies to decay slower than exponentially ($K_3=0$!). Usually an infinite memory is assigned to systems of that kind. Typical processes giving rise to such a behavior are of the intermittent type [23]. As will be shown below, a subexponential scaling of (h_n-h) can be derived for systems exhibiting the phenomenon of temporal self-similarity, e.g., for the logistic map at the period doubling accumulation point. It was conjectured by Ebeling and Nicolis [2,24] that an infinite memory generally is encountered in processes related to information processing systems.

II. CONSTRUCTING SELF-SIMILAR SEQUENCES

The starting point for our considerations is the celebrated logistic map at the accumulation point $r_\infty=3.569\,945\,67\ldots$,

$$x_{n+1}=f(x_n)=r_\infty x_n(1-x_n). \tag{6}$$

The attractor possesses a self-similar structure [25]. Choosing the bipartition $[0,\frac{1}{2}]\rightarrow$ ''0'' and $(\frac{1}{2},1]\rightarrow$ ''1'' the symbolic dynamics yields the *binary Feigenbaum sequence*. A first entropy analysis of such a sequence was performed by Grassberger. For word lengths $n=2^k$ he derived a formula for the n-block entropies [20] which reads

$$H_{n=2^k}=\log_2\left(\frac{3n}{2}\right). \tag{7}$$

This formula implies $h=0$ and, by virtue of the Pesin theorem, is in agreement with the zero Lyapunov exponent.

Ebeling and Nicolis [2] supplemented Grassberger's result for all values of n in between two such word lengths. The basis for their result were empirical rules derived by observing typical sequences. Numerical simulations [26] confirmed their theoretical formulas. The fundamental observation was an overall decay of the h_n which is slower than exponential, namely, according to a $1/n$ law. Hence, this sequence not only possesses an infinite memory but, moreover, the $C_{\rm EM}$ yields the value infinity.

A quite similar observation was made by Gramss [27], who investigated a binary sequence which can be produced applying the method of symbolic dynamics to the critical circle map [28,29]. His results were achieved making use of an equivalent grammatical composition rule for the sequence:

$$\begin{aligned} b_0 &= \text{``1''}, \\ b_1 &= \text{``1''}, \\ b_{z+1} &= b_z\circ b_{z-1} \quad (z=1,2,\ldots). \end{aligned} \tag{8}$$

Here, z is a recursion index and $b_z\circ b_{z-1}$ symbolizes the concatenation of the symbol strings b_z and b_{z-1}. It is a simple but important observation that the length of the sequence b_z equals the Fibonacci number F_{z+1}. The sequence z_∞ clearly exhibits a self-similar structure: performing the replacement

$$\text{``0''}\rightarrow\text{``1'' and ``1''}\rightarrow\text{``10''} \tag{9}$$

exactly yields the same sequence. The replacement scheme (9) motivated the name *rabbit sequence* [30].

Gramss' derivation showed that the overall scaling behavior of the H_n again was of a $\log_{2n}$ type which leads to $h_n\sim n^{-1}$. This similarity arouses suspicion that the binary Feigenbaum sequence should be reproducible in close analogy. In fact, we could specify the following composition rule

$$\begin{aligned} a_0 &= \text{``1''} \\ a_1, &= \text{``10''}, \\ a_{z+1} &= a_z\circ a_{z-1}\circ a_{z-1} \quad (z=1,2,\ldots). \end{aligned} \tag{10}$$

Here, the length of the sequence a_z equals the values 2^z. Performing the two replacement schemes

$$\text{``0''}\rightarrow\text{``11'' and ``1''}\rightarrow\text{``10''} \tag{11}$$

or

$$\text{``0''}\rightarrow\text{``11'' and ``1''}\rightarrow\text{``01''} \tag{12}$$

again leaves the infinite sequence a_∞ invariant. These replacement rules can be regarded as special types of the so-called context-free Lindenmayer systems [31] which have been investigated as a model for spatial $1/f$ spectra in [5].

We point out that composition rules analogous to (8) and (10) do not necessarily guarantee a $1/n$ behavior of the h_n. An example demonstrating this clearly is given by

$$\begin{aligned} a_0 &= \text{``0''}, \\ a_1 &= \text{``1''}, \\ a_{z+1} &= a_{z-1}\circ a_z\circ a_{z-1}. \end{aligned} \tag{13}$$

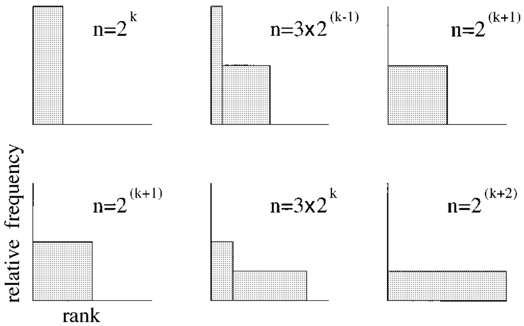

FIG. 1. A schematic plot of self-similar rank-ordered word distributions of the binary Feigenbaum sequence for different values of n. The self-similarity relates distributions for word lengths $2^k \leftrightarrow 2^{k+1}$, $3\times 2^{k-1} \leftrightarrow 3\times 2^k$, $2^{k+1} \leftrightarrow 2^{k+2}$.

This composition rule produces the periodic string $(\text{``01''})^\infty$ which yields $h_0=1$ and $\{h_n\}_{n=1,2,\ldots}=0$.

The empirical rules characterizing the binary Feigenbaum sequence formulated by Ebeling and Nicolis [2] can be understood from the grammatical composition rules (10) [32]; this connection basically rests on three lemmata (analogous to two lemmata found by Gramss [27]). In particular, the rank-ordered n-word distributions can be derived explicitly. The basic structure of the series of these distributions is sketched in Fig. 1. We see that self-similarity relates distributions for word lengths $2^k \leftrightarrow 2^{k+1}$, $3\times 2^{k-1} \leftrightarrow 3\times 2^k$, $2^{k+1} \leftrightarrow 2^{k+2}$. The transition between two such related distributions is achieved simply by rescaling the horizontal and vertical axis by a factor 2 and $\frac{1}{2}$, respectively (and vice versa). This observation is at the heart of the scaling relation for the n-block entropies and will be elaborated in a generalized approach in the next section.

III. GENERAL SELF-SIMILAR DISTRIBUTIONS

In order to generalize the notion of self-similar rank-ordered word distribution we list the following basic properties.

(i) We have a series of word lengths, denoted by $\{n_k\}_{k=k_0}^\infty$, which is defined by a function of the index k, i.e., $n_k = g(k)$.

(ii) For every *root word* of length n_k with probability $p(i_1,\ldots,i_{n_k})$ there exist m *offsprings* of length n_{k+1}, denoted $(j_1^1,\ldots,j_{n_{k+1}}^1),\ldots,(j_1^m,\ldots,j_{n_{k+1}}^m)$, with equal probability $p(i_1,\ldots,i_{n_k})/m$. Notice that now the probability distributions are not restricted to the case of piecewise constant functions as in the case of the binary Feigenbaum sequence. A schematic sketch of such general self-similar distributions is depicted in Fig. 2.

We realize that the root word's probability is equally distributed among its m offsprings. It is important to notice that the terms *root word* and *offsprings* are not necessarily meant in the context of a grammatical root-branching relation. The reason is that entropies never care about a permutation of labels which allows for the rank-ordering. In fact, the relation between a root word and its m offsprings is governed by m replacement operations, e.g., (11,12), leaving the infinite sequence invariant.

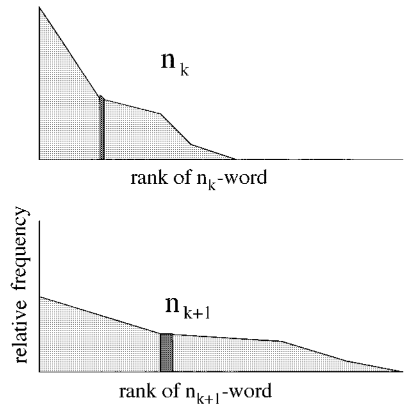

FIG. 2. Schematic sketch of general self-similar rank-ordered word distributions; dark shaded regions: root word (top), m offsprings (bottom).

A few lines of calculation yields the following relation for the n_k-block entropies:

$$H_{n_k} = H_{n_{k_0}} + (k-k_0)\log_\lambda m. \tag{14}$$

This is exactly the type of relation for the H_{n_k} already derived for the binary Feigenbaum sequence by Grassberger [20]: Namely, we only have to realize that in this special case $m=2$, $n_k=g(k)=2^k$, $k_0=1$, and $\lambda=2$, thus arriving at

$$H_{n_k=2^k} = H_2 + (k-1)\log_2 2 \tag{15}$$

$$= \log_2 3 + \log_2 \frac{2^k}{2} \tag{16}$$

$$= \log_2\left(\frac{3n_k}{2}\right), \tag{17}$$

i.e., we perfectly recover (7). We point out that the logarithmic dependence of the n_k-block entropies is a direct consequence of the fact that the n_k, relating self-similar word distributions, stretch according to an exponential law, i.e., $n_k = g(k) \sim 2^k \Rightarrow k = g^{-1}(n_k) \sim \log_2 n_k$. For the binary Feigenbaum sequence this function corresponds to $g(k)=2^k$ and for the rabbit sequence the series of Fibonacci numbers can equivalently be represented by the function $g(k) = 1/\sqrt{5}(\gamma^{-k} - \gamma^k)$ with $\gamma = (\sqrt{5}-1)/2$. This explains the infinite memory, $(h_n - h) \sim n^{-1}$, found in both cases and elucidates its universal character.

We can go further and ask under which conditions we will find a scaling law of the type $H_{n_k} \sim \sqrt{n_k}$; this square root type growth law has been achieved first by Hilberg [33] and subsequently by Ebeling *et al.* [2,4], analyzing realistic texts. The constructive answer is that we have to inquire the possibility of composing sequences which yield self-similar rank-ordered word distributions obeying $n_k \sim k^2$. However, no example has been constructed yet and it seems very likely that creating sequences in a fashion analogous to (8) or (10) mostly [see above exception (13)] results in exponentially growing n_k, hence, yielding logarithmic scaling laws for the H_{n_k}.

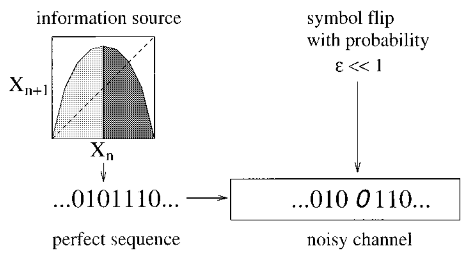

FIG. 3. After its production by the unperturbed source the perfectly self-similar sequence is fed into the transmission channel. External white noise ($\epsilon \ll 1$) flips symbols of the message independently, thus ''contaminating'' the sequence.

In passing, we mention that parallels can be established between self-similarity found in the rank-ordered word distributions and in the field of critical phenomena [34]. The replacement procedures for the symbols, e.g., (11,12), should be regarded in analogy to the *block–spin transformation* [35]. The relation between block entropies (14) resembles the *Kadanoff transformation* [34], connecting Hamiltonians formulated at different scales of resolution. The precise formulation of these analogies should be done in the context of the *thermodynamic formalism* [36], where the H_n can be regarded as *generalized free energies* of an n-particle system.

IV. THE EFFECT OF NOISE

The above discussed self-similarity of rank-ordered word distributions is compatible with $h=0$ only. The question arises, what happens when noise affects the symbol sequence? Denoting the strength of noise by some parameter ϵ we expect the entropy of the source h^ϵ to rise slightly above the value zero, i.e., $h^\epsilon = f(\epsilon) > 0$. But whether or not the infinite memory, i.e., the characteristic scaling behavior of the H_n, will survive this perturbation is not clear in advance.

In order to clarify this problem we introduce noise in the following way: At first the perfectly self-similar symbol sequence is created. In a second sweep we randomly flip symbols. For a physical motivation we might think of an information source which emits a message. This message is fed into a channel, by a transmitter, it passes the channel and is received at its other end. During this transmission process random noise affects the signal. This noise is assumed to be white, which means the chance to flip a symbol is independent of what happens to all other symbols. The probability to flip a single symbol is parametrized by ϵ. For an illustration see Fig. 3.

In order to develop the consequences most clearly, again, we consider the binary Feigenbaum sequence and discuss the case $n=2^k$. As can be seen from Fig. 1, the rank-ordered 2^k-word distribution simply is a step function. The influence of noise will change this unaffected word distribution by creating previously forbidden words; thus all possible words can be found. Nevertheless, we will find a natural grouping of words into classes essentially connected to the number of flips performed with respect to an original word (Hamming distance). In the limit of very small noise intensity, $\epsilon \ll 1$, the group with the largest probabilities will be formed by the original words, i.e., words found in the unaffected sequence. Their number is denoted by N_0 and their probabilities will be changed from $p_0 := N_0^{-1}$ to $p_0(1-\epsilon)^n$. The next group will comprise all words which are created by flipping only one symbol. Their probabilities will be $(p_0)\epsilon(1-\epsilon)^{n-1}$. The probability for each member of this group essentially is diminished by a factor ϵ. The number of words which are created in this way will be denoted by N_1. The line of argument proceeds in an analogous manner: Words created by j flips will have a probability given by $(p_0)\epsilon^j(1-\epsilon)^{(n-j)}$ and their number is denoted by N_j. This yields a rank-ordered word distribution which has a staircase shape. The heights of neighboring plateaus vary by a factor ϵ and their lengths are given by the numbers N_j.

At this point we mention that the true picture might slightly deviate since there is a chance that a selected word will be *multiply addressed* because of it originating from different unperturbed words and due to a different total number of flips. This effect, however, will be suppressed whenever the effective number of words λ^{nh} (for $n \to \infty$) [11] is much less than the number of all possible words λ^n; in this case different flips most likely will create different words. So the above idealization, ignoring multiple addressing, should become increasingly better for systems with $h \ll 1$ — which is guaranteed for self-similar sequences — and for increasing word lengths n. Figure 4 illustrates these explanations.

Inserting the idealized rank-ordered word distribution into the Shannon functional yields, after some steps of calculation, the following result:

$$H_n^\epsilon \leq H_n^0 + n h(\epsilon), \tag{18}$$

where

$$h(\epsilon) = -\left(1 - \frac{\langle j \rangle}{n}\right)\log_2(1-\epsilon) - \frac{\langle j \rangle}{n}\log_2(\epsilon) \tag{19}$$

and the average number of flips explicitly reads

$$\langle j \rangle = \sum_{j=0}^{n} j \frac{N_j}{N_0} \epsilon^j (1-\epsilon)^{(n-j)}. \tag{20}$$

The $<$ sign in (18) is due to multiple addressing which tends to sharpen the distribution as explained above. Generally, $N_j \leq \binom{n}{j}$ holds true, again, because of multiple addressing; hence, $\langle j \rangle \leq n\epsilon$. Inserting $n\epsilon$ into formula (19) yields

$$h(\epsilon) \leq -(1-\epsilon)\log_2(1-\epsilon) - \epsilon\log_2(\epsilon). \tag{21}$$

For the upper bound character on the right side, the closer it approaches an equality relation, the less frequent will be the multiple addressing, i.e., the smaller h^0 and the larger n will be.

We see that the H_n^ϵ and h_n^ϵ are shifted by an amount $nh(\epsilon)$ and $h(\epsilon)$, respectively, with respect to the related values of the unperturbed source. Figures 5 and 6 depict the theoretical (upper bound) H_n^ϵ and h_n^ϵ, respectively, in comparison with numerical data for $\epsilon = 0.01$ and $\epsilon = 0.1$. Whereas for $\epsilon = 0.1$ the upper bound character clearly is vis-

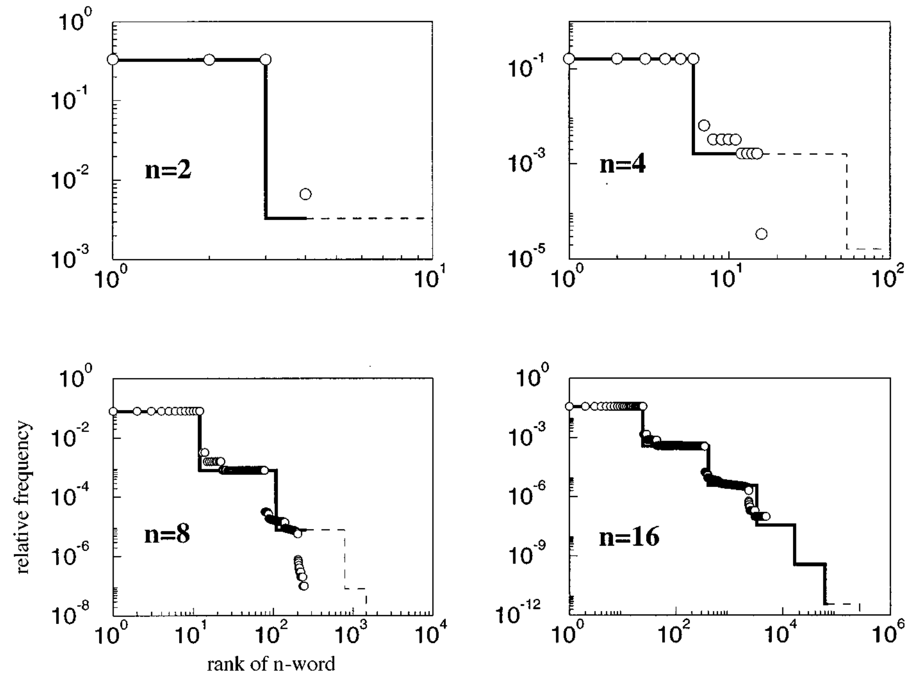

FIG. 4. The rank-ordered word distributions of the ($\epsilon=0.01$) noise contaminated binary Feigenbaum sequence for $n=2,4,8,16$; circles: numerical data, lines: plateau structure ruled by $N_j(k)=\binom{n}{j}$ (up to 2^n fat, beyond dashed). Multiple addressing of words gives rise to substructures superposed onto the dominant staircase structure. The finite sample length ($L=10^7$) is responsible for the collapse of smallest probabilities at the right end.

ible for $\epsilon=0.01$ the agreement becomes almost perfect. Notice that $\epsilon=0$ corresponds to the unaffected binary Feigenbaum sequence.

The most important result is the insight that the scaling relation for dynamical entropies survives the prescribed perturbation, at least for sufficiently small ϵ. Moreover, since the above mentioned arguments remain valid for an arbitrary system, provided its entropy h^0 is sufficiently small, we can extend the result to sequences which obey sufficiently restrictive selection rules, e.g., periodic sequences. Then the entropy of the ''noise contaminated'' source reads $h^{\epsilon}=h^0+h(\epsilon)$.

In this section we considered binary sequences only. A generalization to sequences constructed from an alphabet containing λ letters is given in the Appendix.

V. CONCLUSION

We have exemplified the creation of self-similar sequences by a symbolic dynamics of critical systems. An equivalent grammatical composition rule allowed for a generalized construction scheme. The self-similar features of the sequence itself can be derived employing the last mentioned approach. Moreover, the self-similarity of rank-ordered word distributions can be traced back to the composition rule. A very general relation between these self-similar distributions

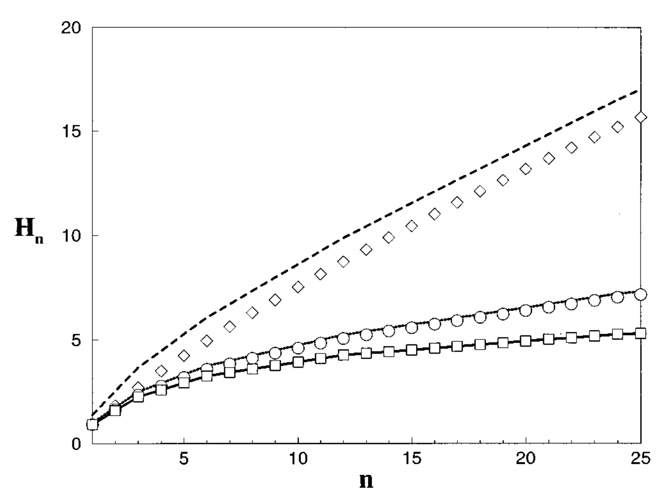

FIG. 5. The n-block entropies H_n^{ϵ} for the noise contaminated binary Feigenbaum sequence: $\epsilon=0.1$ with diamonds (numerical) and dashed line (theoretical upper bound); $\epsilon=0.01$ with circles (numerical) and dotted line (theoretical upper bound). The unaffected binary Feigenbaum sequence corresponds to $\epsilon=0$ with squares (numerical) and solid line (theoretical).

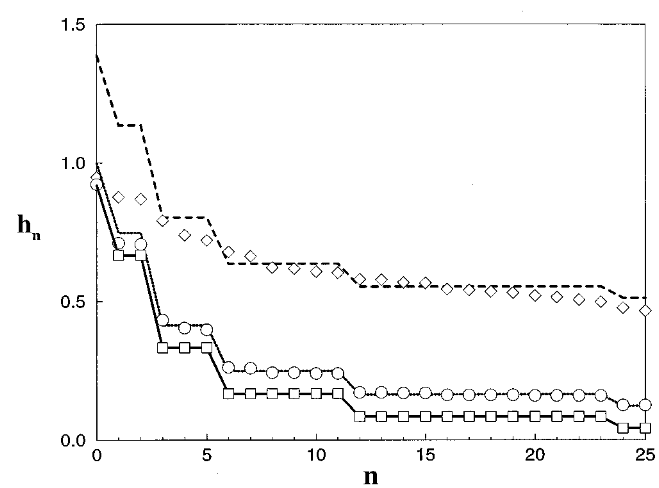

FIG. 6. The conditional entropies h_n^{ϵ} for the noise contaminated binary Feigenbaum sequence: $\epsilon=0.1$ with diamonds (numerical) and dashed line (theoretical); $\epsilon=0.01$ with circles (numerical) and dotted line (theoretical). The unaffected binary Feigenbaum sequence corresponds to $\epsilon=0$ with squares (numerical) and solid line (theoretical).

and universal scaling relations for dynamical entropies could be established. This insight might help to solve open questions related to natural sequences, e.g., DNA, music, or texts. Self-similar dynamical behavior can be regarded in analogy to the scale invariance of critical systems. A perturbation of perfectly self-similar sequences, contaminating them by noise-induced symbol flips, is reflected by a rising entropy of the source h^{ϵ} which could be quantified. The infinite memory survives the considered perturbation in the limit of small noise, i.e., for $\epsilon \ll 1$. This statement remains valid for arbitrary systems provided $h \ll 1$.

ACKNOWLEDGMENTS

We thank Hanspaper Herzel for bringing to our attention two important publications. Encouragement and support by Thorsten Pöschel is gratefully acknowledged.

APPENDIX: GENERALIZATION TO A LARGER ALPHABET

In the preceding section we considered the effect of contaminating binary sequences related to $h^0 \ll 1$ by independently flipping symbols with a probability ϵ. A generalization to the case of a larger alphabet, consisting of λ letters, means at each position of the unaffected sequence replacing the original symbol by one of the $(\lambda-1)$ complementary ones with probability ϵ. The above reasoning for the binary alphabet essentially is identical and the analogous calculations are straightforward.

Equation (18) remains the same but Eq. (19) changes to

$$h(\epsilon) = -\left(1-\frac{\langle j\rangle}{n}\right)\log_{\lambda}(1-\epsilon) - \frac{\langle j\rangle}{n}\log_{\lambda}\left(\frac{\epsilon}{\lambda-1}\right) \quad \text{(A1)}$$

and the average number of flips generalizes to

$$\langle j\rangle = \sum_{j=0}^{n} j\,\frac{N_j}{N_0}\,\frac{\epsilon^j}{\lambda-1}(1-\epsilon)^{(n-j)}. \quad \text{(A2)}$$

Again, we can find an upper bound to the source entropy shift,

$$h(\epsilon) \leqslant -(1-\epsilon)\log_{\lambda}(1-\epsilon) - \epsilon\log_{\lambda}\left(\frac{\epsilon}{\lambda-1}\right). \quad \text{(A3)}$$

This upper bound already was derived by intuitive reasoning for a related problem applying to DNA sequences, i.e., for $\lambda=4$ (measuring information in bits rather than in λits) [37].

[1] V. M. Alekseev and M. V. Yakobson, Phys. Rep. **75**, 287 (1981).
[2] W. Ebeling and G. Nicolis, Chaos, Solitons Fractals **2**, 635 (1992).
[3] A. Rényi, *Probability Theory* (North-Holland, Amsterdam, 1970).
[4] W. Ebeling and T. Pöschel, Europhys. Lett. **26**, 241 (1994).
[5] W. Li, Physica A **43**, 5240 (1991).
[6] R. F. Voss and J. Clarke, J. Accoust. Soc. **63**, 258 (1978).
[7] V. S. Anishchenko, W. Ebeling, and A. B. Neiman, Chaos, Solitons Fractals **4**, 69 (1994).
[8] H. E. Stanley, S.V. Buldyrev, A. L. Goldberger, Z. D. Goldberger, S. Havlin, R. N. Mantegna, S. M. Ossadnik, C.-K. Peng, and M. Simons, Physica A **205**, 214 (1994).
[9] H. Herzel and I. Groβe, Physica A **216**, 518 (1995).
[10] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[11] A. I. Khinchin, *Mathematical Foundations of Information Theory* (Dover Publ., New York, 1957).
[12] A. O. Schmitt, H. Herzel, and W. Ebeling, Europhys. Lett. **23**, 303 (1993).
[13] T. Pöschel, W. Ebeling, and H. Rosé, J. Stat. Phys. **80**, 1443 (1995).
[14] B. McMillan, Ann. Math. Stat. **24**, 196 (1953).
[15] A. Kolmogorov, Dokl. Akad. Nauk SSSR **119**, 861 (1958).
[16] Ya. G. Sinai, Dokl. Akad. Nauk SSSR **124**, 768 (1959).
[17] F. Ledrappier and L.-S. Young, Ann. Math. **122**, 509 (1985); **122**, 540 (1985).
[18] L. Gatlin, *Information Theory and the Living System* (Columbia Univ. Press, New York, 1972).
[19] G. Győrgyi and P. Szépfalusy, Phys. Rev. A **31**, 3477 (1985).
[20] P. Grassberger, Int. J. Theor. Phys. **25**, 907 (1986).
[21] P. Szépfalusy and G. Győrgyi, Phys. Rev. A **33**, 2825 (1986).
[22] P. Szépfalusy, Phys. Scr. **T25**, 226 (1989).
[23] A. Csordás, G. Győrgyi, P. Szépfalusy, and T. Tél, Chaos **3**, 31 (1993).
[24] W. Ebeling and G. Nicolis, Europhys. Lett. **14**, 191 (1991).
[25] M. J. Feigenbaum, J. Stat. Phys. **19**, 25 (1978).
[26] W. Ebeling, J. Freund, and K. Rateitschak, Int. J. Bifurcation Chaos **6**, 611 (1996).
[27] T. Gramss, Phys. Rev. E **50**, 2616 (1994).
[28] I. Procaccia, S. Thomae, and C. Tressor, Phys. Rev. A **35**, 1884 (1987).
[29] M. R. Schroeder, *Fractals, Chaos, Power Laws* (Freeman, New York, 1991).
[30] M. R. Schroeder, *Number Theory in Science and Communication*, Springer Series in Information Sciences, 2nd enlarged ed. (Springer, Berlin, 1986).
[31] A. Lindenmayer, J. Theor. Biol. **18**, 280 (1968).
[32] K. Rateitschak, J. Freund, and W. Ebeling, in *Entropy and Entropy Generation: Fundamentals and Applications*, edited by J. S. Shiner (Kluwer, Dordrecht, 1996).
[33] W. Hilberg, Frequenz **44**, 243 (1990).
[34] S. K. Ma, *Modern Theory of Critical Phenomena* (Benjamin, New York, 1976).
[35] K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).
[36] C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems* (Cambridge University Press, Cambridge, 1993).
[37] H. Herzel, W. Ebeling, and A. O. Schmitt, Phys. Rev. E **50**, 5061 (1994).