Criteria for disorder in aperiodic sequences

A. Yu. Turygin and V. R. Chechetkin

Troitsk Institute of Innovative and Thermonuclear Research, 142092 Troitsk, Moscow Region, Russia
(Submitted 18 November 1993)
Zh. Eksp. Teor. Fiz. **106**, 335–354 (July 1994)

The spectral representation of a sequence of symbols is used to derive various criteria for disorder. Expressions are found for statistical distributions of the spectral harmonics and criteria for the heights of the harmonics to be statistically meaningful are introduced for the structure factors of the sequence, along with the degree of correlation of the various symbols. The results are applied to analyze substitution sequences and symbolic dynamics.

1. INTRODUCTION

In many physical problems it is necessary to deal with formal sequences of symbols. These physical applications are related mainly to the theory of quasicrystals and substitution sequences,[1–5] and also to the study of the behavior of dynamical systems using the methods of symbolic dynamics.[6–10] In Refs. 11–14 physical techniques for analyzing sequences were applied to the fundamental problem of the structural investigation of genome DNA sequences. In practice one usually does not know *a priori* anything about the algorithm used to generate the sequence, and the analysis begins by asking the simplest question: Is the sequence random or regular? If the sequence contains both random and regular aspects, then the question arises of how to distinguish them. Moreover, it is desirable to have quantitative criteria for the degree of complexity and disorder of a sequence, and also the correlations between symbols.

The best-established approach to analyzing the complexity of sequences is based on the general theory of information[15,16] (for applications see Refs. 6–10, 17, and 18). The underlying information in this case is the probability $P(\{S_n\})$ of encountering the various combinations of symbols forming a sequence of length n for larger and larger values of n. The degree of complexity is quantified by the metric entropy,

$$K=-\lim_{n\to\infty}\frac{1}{n}\sum_{\{S_n\}}P(\{S_n\})\log P(\{S_n\}). \tag{1.1}$$

For l different symbols the evaluation of $P(\{S_n\})$ requires analysis of l^n combinations. Actual calculations, e.g., for the logistic mapping[10] revealed that a satisfactory estimate of K requires searching through more than 2^{24} combinations and sequences with an overall length greater than $4\cdot 10^8$ symbols. Accordingly, the need to search through an exponentially large number of symbol combinations and the need to use sequences of extreme length interferes with the application of the information-theoretical approach. In fact, the quantity K can often be calculated only when the original algorithm used to generate the sequence is known, which drastically reduces the number of possible combinations. A second limitation is related to the fact that for finite n (in the actual analysis these values are generally small) the entropy K characterizes only local correlations of the symbols, and long-range correlations must be studied separately by other means (see, e.g., Refs. 11–14).

In the present work we consider an alternative approach using the spectral representation of the sequence. This technique is a slight modification of the usual spectral technique for analyzing substitution sequences.[2–5] The basic idea of the work consists of studying the statistical properties of the harmonics of the Fourier representation for symbolic sequences. The existence of appropriate well-developed analytical and numerical techniques[19] makes this approach quite convenient for practical applications. As shown by the results of test simulations and consideration of specific examples, the proposed criteria are applicable for relatively short sequences ($\gtrsim 10^3$ symbols). It is assumed that the original sequence belongs to a general form. For scale-invariant sequences the structural entropy, introduced in Sec. 4.4, permits one to carry out additional classification of spectra having identical multifractal properties.

Fourier transformations are also widely used in analyzing DNA sequences.[11,13,20–26] Their use in this case serves two purposes: first, in order to identify hidden periodicities[20,22–24] and long-range correlations;[11,13] secondly, to compare genome sequences of DNA from different organisms with one another.[21,25,26] For two random sequences the cross-correlation coefficient would be of order $\sim M^{-1/2}$ (where M is the length of the sequences being compared), i.e., small for $M\gg 1$. Hence a large nonzero correlation coefficient would imply that the two sequences are similar. This argument, however, is inapplicable in connection with cross-correlations of symbols within a single sequence, since excluded-volume effects give rise to correlations of order unity even for random sequences. In Sec. 4.3 we give a quantitative necessary criterion for this case. The use of Fourier transformations to analyze DNA sequences is dictated by the further need for rapid processing of a large quantity of information and the fact that the fast Fourier transform is the most efficient of the algorithms currently known. Application of the results of Sec. 4 yields simple and convenient criteria for analyzing such sequences and permits one to obtain interesting data about the relative structural properties of DNA from different organisms.[27] We will not go into a more detailed discussion here of this important problem.

This paper is organized as follows. In Sec. 2 a general

 1063-7761/94/070186-11$10.00

formulation of the problem is given. Expressions for the characteristic function and also for the probability distribution of the spectral harmonics are given in Sec. 3. In Sec. 4 they are used to present specific criteria for the degree of order in a sequence. The application of the results to analyzing substitution sequences and symbolic dynamics is illustrated in Sec. 5. The concluding section, Sec. 6, contains some remarks about possible generalizations of the results.

2. SPECTRAL REPRESENTATION OF A SEQUENCE

2.1. The structure factor of a sequence

Completely random sequences of symbols are the most disordered and algorithmically complicated. Relative properties for a sequence of general form can be obtained by comparing them with the corresponding structural properties of random sequences with the same whole numbers of symbols as in the sample sequence. Specific criteria are given in Sec. 4. There and in the following sections we introduce the fundamental quantities which are studied and present results of the general theory for random sequences.

Consider a sequence $\{A_k\}$ of length M consisting of l different symbols. It can be specified by using the position function:

$$\rho_{m,\alpha}=\begin{cases}1, & \text{if the } \alpha\text{th symbol occupies the } m\text{th position,}\\ 0, & \text{otherwise,}\end{cases}$$
$$\alpha\in(A_1,\ldots,A_l),\quad m=1,\ldots,M. \tag{2.1}$$

Then the Fourier harmonics corresponding to the subsequence made up of the symbols of type α is defined by

$$\rho_\alpha(q_n)=M^{-1/2}\sum_{m=1}^{M}\rho_{m,\alpha}e^{-iq_nm},$$
$$q_n=2\pi n/M,\ n=0,1,\ldots,M-1, \tag{2.2}$$

and the inverse transformation takes the form

$$\rho_{m,\alpha}=M^{-1/2}\sum_{n=0}^{M-1}\rho_\alpha(q_n)e^{iq_nm},\quad m=1,\ldots,M. \tag{2.3}$$

The Fourier harmonic of order zero contains no information about the distribution of the symbols and depends only on their total number N_α,

$$\rho_\alpha(0)=N_\alpha/M^{1/2}. \tag{2.4}$$

The requirement that $\rho_{m,\alpha}$ be real yields the condition

$$\rho_\alpha^*(q_n)=\rho_\alpha(2\pi-q_n) \tag{2.5}$$

(here and below the asterisk denotes complex conjugation).

Below we express the basic characteristics using the elements of the $l\times l$ matrix structure factor of the sequence:

$$F_{\alpha\beta}(q_n)=\rho_\alpha(q_n)\rho_\beta^*(q_n). \tag{2.6}$$

From (2.5) we find

$$F_{\alpha\beta}(q_n)=F_{\beta\alpha}(2\pi-q_n), \tag{2.7}$$

in particular, the diagonal elements $F_{\alpha\alpha}(q_n)$ are symmetric functions with center of symmetry $q_n=\pi$.

The structure function of the sequence can be related to the pair correlations of the different symbols. In order to show this, we introduce the cyclic pair correlations of the function:

$$K^0_{\alpha\beta}(m_0)=M^{-1}\sum_{m=1}^{M}\tilde\rho_{m,\alpha}\tilde\rho_{m+m_0,\beta}, \tag{2.8}$$

$$\tilde\rho_{m,\alpha}=\begin{cases}\rho_{m,\alpha}, & 1\leqslant m\leqslant M,\\ \rho_{m-M,\alpha}, & M+1\leqslant m\leqslant 2M-1,\end{cases} \tag{2.9}$$

where $1\leqslant m_0\leqslant M-1$. Then from Eqs. (2.1)–(2.9) we find that the two properties are related by the Wiener–Khinchin relation,

$$K^c_{\alpha\beta}(m_0)=M^{-1}\sum_{n=0}^{M-1}F_{\alpha\beta}(q_n)e^{-iq_nm_0}. \tag{2.10}$$

From the definitions (2.8) and (2.9) it follows that

$$K^c_{\alpha\beta}(m_0)=K^c_{\beta\alpha}(M-m_0). \tag{2.11}$$

Higher products of the Fourier harmonics can be related to higher-order correlation functions.

2.2. Sum rules

The statistical criteria are expressed using a set of universal (independent of the specific symbol distribution) quantities determined by exact sum rules. The first relation is derived directly from Eqs. (2.1)–(2.7) and takes the form

$$\sum_{n=0}^{M-1}F_{\alpha\beta}(q_n)=\sum_{m=1}^{M}\rho_{m,\alpha}\rho_{m,\beta}=\delta_{\alpha\beta}N_\alpha, \tag{2.12}$$

where $\delta_{\alpha\beta}$ if the Kronecker symbol. Making use of expression (2.4) for the zeroth-order harmonics we find

$$\bar F_{\alpha\beta}=\left(\sum_{n=1}^{M-1}F_{\alpha\beta}(q_n)\right)/(M-1)$$
$$=(\delta_{\alpha\beta}N_\alpha-N_\alpha N_\beta/M)/(M-1). \tag{2.13}$$

Similarly we can derive a more general sum rule:

$$\sum_{\substack{0\leqslant q_n\leqslant 2\pi(M-1)/M\\ (q_{n_1}+\cdots+q_{n_r})\bmod 2\pi=0}}\rho_{\alpha_1}(q_{n_1})\cdots\rho_{\alpha_r}(q_{n_r})$$
$$=M^{(r-2)/2}\sum_{m=1}^{M}\rho_{m,\alpha_1}\cdots\rho_{m,\alpha_r}$$
$$=\delta_{\alpha_1\alpha_2}\ldots\delta_{\alpha_1\alpha_r}M^{(r-2)/2}N_{\alpha_1}. \tag{2.14}$$

From Eq. (2.10) we find a sum rule which relates the reciprocal deviations of the cyclic correlation functions and the elements of the structure factor from their corresponding average values:

$$\sum_{m_0=1}^{M-1}(K^c_{\alpha\beta}(m_0)-\bar{K}^c_{\alpha\beta})(K^c_{\gamma\delta}(m_0)-\bar{K}^c_{\gamma\delta})$$

$$=M^{-1}\sum_{n=1}^{M-1}(F_{\alpha\beta}(q_n)-\bar{F}_{\alpha\beta})(F^*_{\gamma\delta}(q_n)-\bar{F}^*_{\gamma\delta}), \quad (2.15)$$

$$\bar{K}^c_{\alpha\beta}(m_0)=(M-1)^{-1}\sum_{m_0=1}^{M-1}K^c_{\alpha\beta}(m_0)$$

$$=(N_\alpha N_\beta-\delta_{\alpha\beta}N_\beta)/M(M-1). \quad (2.16)$$

2.3. Excluded-volume effects

Each position in the sequence is occupied by just one symbol, and the sequence itself contains no gaps. These conditions impose further restrictions on the position function (2.1),

$$\sum_{\alpha=1}^{l}\rho_{m,\alpha}=1, \quad (2.17)$$

i.e., the locations of the subsequences of $l-1$ different symbols uniquely determine the locations of the remaining subsequence. In terms of the Fourier harmonics this condition can be formulated as follows:

$$\sum_{\alpha=1}^{l}\rho_\alpha(q_n)=0 \quad (n\neq 0). \quad (2.18)$$

Relation (2.17) gives rise to specific correlations (conventionally referred to as "excluded-volume effects") even for random sequences. Furthermore, Eqs. (2.17) and (2.18) reduce the number of independent correlation functions and elements of the structure factor.

3. STATISTICAL PROPERTIES OF RANDOM SEQUENCES

3.1. Characteristic functions

The statistical distribution of Fourier harmonics can be found by averaging the characteristic function (see, e.g., Refs. 28 and 29),

$$Z=\exp\left(i\sum_{\alpha=1}^{l}\sum_{n=1}^{M-1}u_\alpha(q_n)\rho_\alpha(q_n)\right), \quad (3.1)$$

over an ensemble of random realizations of sequences $\{N_\alpha\}$ with fixed total numbers of symbols. It is convenient to impose the same condition on the auxiliary variables $u_\alpha(q_n)$ as in (2.5),

$$u^*_\alpha(q_n)=u_\alpha(2\pi-q_n). \quad (3.2)$$

Different products of Fourier harmonics are obtained by differentiating Z with respect to the auxiliary variables $u_\alpha(q_n)$ and then equating $u_\alpha(q_n)$ to zero, e.g.,

$$\rho_\alpha(q_n)=\partial Z/i\partial u_\alpha(q_n)\Big|_{\{u_\alpha\}=0}, \quad (3.3)$$

etc. Using the definitions (2.1) and (2.2) we rewrite Z in the form

$$Z=\prod_{\alpha=1}^{l}\prod_{m=1}^{M}(1+\rho_{m,\alpha}z_{m,\alpha}), \quad (3.4)$$

$$z_{m,\alpha}=\exp\left(iM^{-1/2}\sum_{n=1}^{M-1}u_\alpha(q_n)e^{-iq_nm}\right)-1. \quad (3.5)$$

Thus, the problem reduces to averaging the different products $\rho_{m,\alpha}$.

The average is determined by simple combinatoric considerations, and the result is equal to

$$\langle\underbrace{\rho_{m_1,A_1}\cdots\rho_{m_{L_1},A_1}}_{L_1}\cdots\underbrace{\rho_{m_{L_1+\dots+L_{l-1}+1},A_l}\cdots\rho_{m_{L_1+\dots+L_l},A_l}}_{L_l}\rangle$$

$$=C^{M-L_1-\dots-L_l}_{N_1-L_1,\dots,N_l-L_l}/C^M_{N_1,\dots,N_l}, \quad (3.6)$$

$$C^m_{n_1,\dots,n_l}=m!/n_1!\dots n_l!,$$

$$n_1+\dots+n_l=m, \quad 0!=1. \quad (3.7)$$

Here the angle brackets denote an average over the ensemble of random realizations, L_k is the total number of position functions corresponding to the symbols A_k, and N_k is the total number of symbols A_k in the sequence of length M. The right-hand side of Eq. (3.6) is equal to the ratio of the two combinatoric factors $C^M_{N_1,\dots,N_l}$, the total number of different random realizations, and $C^{M-L_1-\dots-L_l}_{N_1-L_1,\dots,N_l-L_l}$, the total number of realizations under the condition that $L_1,\dots,L_l$ positions of the different symbols are fixed [these are the same positions which enter into the left-hand side of Eq. (3.6)]. Averaging (3.4) in accordance with (3.6) and symmetrizing the expressions with respect to the positions of identical symbols, we find

$$\langle Z\rangle=1/C^M_{N_1,\dots,N_l}\left(\sum_{L_1,\dots,L_l=0}^{L_1\leq N_1,\dots,L_l\leq N_l}\;\sum'^{M}_{m_1,\dots,m_{L_1+\dots+L_l}=1}\right. \quad (3.8)$$

$$\left.\times C^{M-L_1-\dots-L_l}_{N_1-L_1,\dots,N_l-L_l}/L_1!\dots L_l!z_{m_1,A_1}\dots z_{m_{L_1+\dots+L_l},A_l}\right).$$

The prime on the summation sign over the different symbol positions means that all terms with two (or more) identical subscripts $\{m_k\}$ must be excluded from the summation.

Direct evaluation using the definition (3.3) yields

$$\langle\rho_\alpha(q_n)\rangle=0 \quad (q_n\neq 0), \quad (3.9)$$

$$\langle\rho_\alpha(q_n)\rho_\beta(q_{n'})\rangle=\begin{cases}\bar{F}_{\alpha\beta}, & q_n+q_{n'}=2\pi\\ 0, & q_n+q_{n'}\neq 2\pi\end{cases} \quad (3.10)$$

$$\langle F_{\alpha\alpha}(q_n)F_{\beta\beta}(q_n)\rangle-\langle F_{\alpha\alpha}(q_n)\rangle\langle F_{\beta\beta}(q_n)\rangle$$

$$=N_\alpha(N_\alpha-1)N_\beta(N_\beta-1)/M(M-1)^2$$

$$\times(M-2), \quad (\alpha\neq\beta), \quad (3.11)$$

$$\langle F_{\alpha\alpha}^2(q_n)\rangle - \langle F_{\alpha\alpha}(q_n)\rangle^2 = N_\alpha(N_\alpha - 1)(M - N_\alpha) \times (M - N_\alpha - 1)/M(M-1)^2(M-2), \tag{3.12}$$

$$\langle F_{\alpha\beta}(q_{n'})F_{\gamma\delta}(q_n)\rangle - \langle F_{\alpha\beta}(q_{n'})\rangle\langle F_{\gamma\delta}(q_n)\rangle \propto 1/M, \tag{3.13}$$

$$(q_n \neq q_{n'},\ q_n + q_{n'} \neq 2\pi).$$

As can be seen from Eq. (3.13), the weaker correlations of harmonics with different wave numbers q_n for $M \gg 1$ imply that the relative fluctuations of different spectral sums are typically $\sim M^{-1/2}$.

In the average over the ensemble of random realizations only those products of Fourier harmonics $\rho_\alpha(q_n)$ for which the sum of the wave numbers of all factors is a multiple of 2π will be nonzero. All such products enter in the sum rules (2.14). In the limit $N_\alpha \gg 1$, $M \gg 1$ the following asymptotic cumulant expansion is valid for (3.8) (to leading order in $\sim 1/N_\alpha$, $1/M$):

$$\ln\langle Z\rangle \approx \sum_{\{r\}} \sum_{\{q_{n_r}\}} i^{r_1+\ldots+r_l}/r_1!\ldots r_l! \times \langle\!\langle \underbrace{\rho_{A_1}(q_{n_1})\cdots\rho_{A_1}(q_{n_{r_1}})}_{r_1}\cdots\underbrace{\rho_{A_l}(q_{n_{r_1+\ldots+r_{l-1}+1}})\cdots\rho_{A_l}(q_{n_{r_1+\ldots+r_l}})}_{r_l}\rangle\!\rangle \times u_{A_1}(q_{n_1})\cdots u_{A_1}(q_{n_{r_1}})\cdots u_{A_l}(q_{n_{r_1+\ldots+r_{l-1}+1}})\cdots u_{A_l}(q_{n_{r_1+\ldots+r_l}}). \tag{3.14}$$

In the expansion (3.14) the summation is bounded by the conditions $0 \leqslant r_k \ll M$, $r_1+\ldots+r_l \ll M$, $2\pi/M \leqslant q_{n_r} \leqslant 2\pi(M-1)/M$. Furthermore, in each cumulant $\langle\!\langle \rho_{A_1}(q_{n_1})\ldots\rho_{A_l}(q_{n_{r_1+\ldots+r_l}})\rangle\!\rangle$, no partial sum over an arbitrary set of wave numbers other than the complete sum over all wave numbers can be a multiple $2\pi p$ (where p is a whole number), i.e.,

$$\left(\sum q_{n_r}\right) \bmod 2\pi \neq 0,\quad q_{n_r} \in (q_{n_1},\ldots,q_{n_{r_1+\ldots+r_l}}), \tag{3.15a}$$

$$(q_{n_1} + \ldots + q_{n_{r_1+\ldots+r_l}}) \bmod 2\pi = 0. \tag{3.15b}$$

From (2.18) we find a condition which holds for all the cumulants:

$$\sum_{\alpha=1}^{l} \langle\!\langle \ldots\rho_\alpha(q_n)\ldots\rangle\!\rangle = 0. \tag{3.16}$$

Exact sum rules of the type studied in Sec. 2.2 determine that there is a quasi-ergodic equivalence between ensemble averaging and averaging over the spectrum [cf. Eq. (3.10)]. Consequently, the simplest explicit expression for the cumulant can be found by recursively identifying the corresponding contributions from (2.14), e.g.,

$$\langle\!\langle \rho_\alpha(q_n)\rho_\beta(q_{n'})\rangle\!\rangle = \bar{F}_{\alpha\beta}, \tag{3.17}$$

$$\langle\!\langle \rho_\alpha(q_n)\rho_\beta(q_{n'})\rho_\gamma(q_{n''})\rangle\!\rangle = [\delta_{\alpha\gamma}\delta_{\alpha\beta}M^{1/2}N_\alpha - N_\alpha N_\beta N_\gamma/M^{3/2} - (M-1) \times (N_\alpha\bar{F}_{\beta\gamma} + N_\beta\bar{F}_{\alpha\gamma} + N_\gamma\bar{F}_{\alpha\beta})/M^{1/2}]/(M-1)(M-2), \tag{3.18}$$

and so on.

3.2. Probability distribution function

In what follows we restrict ourselves to the statistical distribution of harmonics with identical wave numbers q_n (along with the complex conjugate quantity). To leading order in $M^{-1/2}$ we can use for this purpose the characteristic function

$$\langle Z\rangle_n = \exp\left(-\sum_{\alpha,\beta=1}^{l} \bar{F}_{\alpha\beta}u_\alpha u_\beta^*\right). \tag{3.19}$$

We recall that according to (2.18) $l-1$ harmonics for different symbols uniquely determine the remaining one. Inverting (3.19) with respect to the $l-1$ variables (and equating the remaining one to zero), we find the joint probability distribution function for $l-1$ harmonics:

$$P_{l-1}(|\rho_{\alpha_1}|,\ldots,|\rho_{\alpha_{l-1}}|;\varphi_{\alpha_1},\ldots,\varphi_{\alpha_{l-1}}) = (\det\|\bar{F}_{\alpha\beta}^{(l-1)}\|)^{-1}\exp\left(-\sum_{\alpha,\beta=1}^{l-1}\bar{R}_{\alpha\beta}^{(l-1)}\rho_\alpha\rho_\beta^*\right), \tag{3.20}$$

$$\sum_{\beta=1}^{l-1}\bar{R}_{\alpha\beta}^{(l-1)}\bar{F}_{\beta\gamma}^{(l-1)} = \sum_{\beta=1}^{l-1}\bar{F}_{\alpha\beta}^{(l-1)}\bar{R}_{\beta\gamma}^{(l-1)} = \delta_{\alpha\gamma} = \begin{cases}1, & \alpha=\gamma,\\ 0, & \alpha\neq\gamma.\end{cases} \tag{3.21}$$

Here $\bar{F}_{\alpha\beta}^{(l-1)}$ is a $(l-1)\times(l-1)$ submatrix of $\bar{F}_{\alpha\beta}$ [see Eq. (2.13)] for the specified $l-1$ symbols, $\det\|\bar{F}_{\alpha\beta}^{(l-1)}\|$ is its determinant, and the complex Fourier harmonics ρ_α are described in terms of the modulus $|\rho_\alpha|$ and phase φ_α,

$$\rho_\alpha = |\rho_\alpha|e^{i\varphi_\alpha},\quad \rho_\alpha^* = |\rho_\alpha|e^{-i\varphi_\alpha}, \tag{3.22}$$

while an arbitrary function $\Phi^{(l-1)}$ depending on the variables $|\rho_\alpha|$ and φ_α is averaged according to

$$\langle\Phi^{(l-1)}\rangle=\int\Phi^{(l-1)}P_{l-1}d\Omega_{l-1},\tag{3.23}$$

$$d\Omega_{l-1}=|\rho_{\alpha_1}|d|\rho_{\alpha_1}|d\varphi_{\alpha_1}/\pi\dots|\rho_{\alpha_{l-1}}|d|\rho_{\alpha_{l-1}}|d\varphi_{\alpha_{l-1}}/\pi\tag{3.24}$$

with the range of integration $0\leqslant|\rho_\alpha|\leqslant\infty$, $0<\varphi_\alpha<2\pi$. Note that the $l\times l$ matrix $\bar{F}_{\alpha\beta}$ itself is degenerate by virtue of conditions (3.16) and (3.17).

Integrating (3.20) by parts we can derive reduced distribution functions for a smaller set of symbols. In particular, the one-symbol harmonic distribution function is equal to

$$P_1(F_{\alpha\alpha})=\exp(-F_{\alpha\alpha}/\bar{F}_{\alpha\alpha})/\bar{F}_{\alpha\alpha}\tag{3.25}$$

and agrees with the familiar Rayleigh distribution.[29] Rayleigh treated the problem of the distribution of the sum of the amplitudes of harmonic oscillations with random phases. The formal equivalence of the two problems (and hence the identity of the distribution) is completely obvious [see Eq. (2.2)].

3.3. Variational principle

We can derive the distribution function (3.20) by maximizing the entropy functional:

$$H_{l-1}=-\int P_{l-1}\ln P_{l-1}d\Omega_{l-1},\tag{3.26}$$

under the additional conditions

$$\int P_{l-1}d\Omega_{l-1}=1,\tag{3.27}$$

$$\int\rho_\alpha\rho_\beta^*P_{l-1}d\Omega_{l-1}=\bar{F}_{\alpha\beta}^{(l-1)}.\tag{3.28}$$

The functional (3.26) reaches its maximum for P_{l-1} defined by Eq. (3.20), and is equal to

$$H_{\max,l-1}=(l-1)+\ln\det\|\bar{F}_{\alpha\beta}^{(l-1)}\|.\tag{3.29}$$

The quantity $\det\|\bar{F}_{\alpha\beta}^{(l-1)}\|$ is invariant with respect to the choice of the $l-1$ symbols. The spectral entropy (3.26) is the analog of the metric entropy (1.1), and also characterizes the complexity of the sequence as a whole. Unfortunately, calculations of the distribution function P_{l-1} are quite involved in practice, and it is necessary in applications to restrict ourselves to somewhat cruder criteria, given in the following section.

4. DISORDER CRITERIA IN APERIODIC SEQUENCES

4.1. Distribution of the amplitude harmonics

The general results obtained in the preceding section can be used to derive various specific disorder criteria. We begin with the amplitude distribution of the diagonal elements $F_{\alpha\alpha}(q_n)$.

By virtue of the symmetry conditions (2.7), only half the harmonics $F_{\alpha\alpha}(q_n)$ can be regarded as approximately independent. To be specific we restrict ourselves to the left half of the spectrum, $0<q_n\leqslant\pi$. From Eq. (3.25) we find that the probability that the amplitude of a harmonic exceeds some specified value $F_{\alpha\alpha}^{(0)}$ is given by the expression

$$\mathrm{Prob}\{F_{\alpha\alpha}>F_{\alpha\alpha}^{(0)}\}=\int_{F_{\alpha\alpha}^{(0)}}^{\infty}dF'_{\alpha\alpha}P_1(F'_{\alpha\alpha})$$

$$=\exp(-F_{\alpha\alpha}^{(0)}/\bar{F}_{\alpha\alpha}).\tag{4.1a}$$

This also implies that the average number of harmonics greater than $F_{\alpha\alpha}^{(0)}$ is equal to

$$\langle n_\alpha\rangle=(M/2)\exp(-F_{\alpha\alpha}^{(0)}/\bar{F}_{\alpha\alpha}).\tag{4.1b}$$

The condition $\langle n_\alpha\rangle=1$ determines the typical value of sharp amplitude spikes in random spectra:

$$F_{\alpha\alpha,\max}=\bar{F}_{\alpha\alpha}\ln(M/2).\tag{4.2}$$

The probability that all $M/2$ harmonics simultaneously have amplitudes less than $F_{\alpha\alpha}^{(0)}$, is approximately equal to $(1-\exp(-F_{\alpha\alpha}^{(0)}/\bar{F}_{\alpha\alpha}))^{M/2}$. Consequently, the probability that at least one of the $M/2$ harmonics is greater than $F_{\alpha\alpha}^{(0)}$ takes the form

$$\mathrm{Prob}\{F_{\alpha\alpha}>F_{\alpha\alpha}^{(0)};\ M/2\}$$

$$=1-[1-\exp(-F_{\alpha\alpha}^{(0)}/F_{\alpha\alpha})]^{M/2}$$

$$\approx1-\exp(-\exp[-(F_{\alpha\alpha}^{(0)}-F_{\alpha\alpha,\max})/\bar{F}_{\alpha\alpha})]\tag{4.3}$$

and corresponds to a type-I distribution in radio noise statistics.[30]

We can also find the exact probability that at least one of the $M/2$ harmonics has an amplitude less than $F_{\alpha\alpha}^{(0)}$:

$$\mathrm{Prob}\{F_{\alpha\alpha}<F_{\alpha\alpha}^{(0)};\ M/2\}=1-\exp(-F_{\alpha\alpha}^{(0)}/F_{\alpha\alpha,\min}),\tag{4.4a}$$

$$F_{\alpha\alpha,\min}=\bar{F}_{\alpha\alpha}/(M/2).\tag{4.4b}$$

The quantities $F_{\alpha\alpha,\max}$ and $F_{\alpha\alpha,\min}$ determine the effects of mesoscopic fluctuations associated with specific random realizations.

4.2. Smoothed spectra

In many cases order and disorder coexist. There are many techniques for distinguishing the regular component of a spectrum.[19] The simplest technique consists of smoothing a spectrum over s neighboring harmonics,

$$\tilde{F}_{\alpha\alpha}(q_n)=(2s+1)^{-1}\sum_{n'=n-s}^{n+s}F_{\alpha\alpha}(q_{n'}).\tag{4.5}$$

If $s\ll M$ holds, then correlations of harmonics with different q_n can be disregarded to lowest order [see Eq. (3.13)], and the distribution $\tilde{F}_{\alpha\alpha}(q_n)$ is determined by the Nakagami function:[29]

$$\tilde{P}_s(\tilde{F}_{\alpha\alpha}) = \int_0^\infty \cdots \int_0^\infty \delta\left(\tilde{F}_{\alpha\alpha} - (2s+1)^{-1} \sum_{n'=n-s}^{n+s} F_{\alpha\alpha}(q_{n'})\right) \times P_1(F_{\alpha\alpha}(q_{n-s}))\ldots P_1(F_{\alpha\alpha}(q_{n+s}))dF_{\alpha\alpha} \times (q_{n-s})\ldots dF_{\alpha\alpha}(q_{n+s}) = ((2s+1)/\bar{F}_{\alpha\alpha})^{2s+1}\tilde{F}^{2s}_{\alpha\alpha}/ \Gamma(2s+1)\exp(-(2s+1)\tilde{F}_{\alpha\alpha}/\bar{F}_{\alpha\alpha}), \quad (4.6)$$

where $P_1(F_{\alpha\alpha}(q_n))$ is given by (3.25) and $\Gamma(2s+1)$ is a gamma function. The average value of the amplitude and the dispersion for (4.6) are equal to

$$\langle \tilde{F}_{\alpha\alpha}\rangle = \bar{F}_{\alpha\alpha}, \quad \sigma^2(\tilde{F}_{\alpha\alpha}) = \bar{F}^2_{\alpha\alpha}/(2s+1). \quad (4.7)$$

In the limit $M \gg s \gg 1$ the distribution (4.6) goes over to a Gaussian distribution with mean and dispersion given by (4.7).

4.3. Cross-correlations

The correlations in the positions of different symbols are characterized by the cross-correlation coefficient:[28,29]

$$k(F_{\alpha\beta}|F_{\gamma\delta}) = \sum_{n=1}^{M-1} (F_{\alpha\beta}(q_n) - \bar{F}_{\alpha\beta})(F^*_{\gamma\delta}(q_n) - \bar{F}^*_{\gamma\delta})/(M-1)\sigma(F_{\alpha\beta})\sigma(F_{\gamma\delta}), \quad (4.8)$$

$$\sigma^2(F_{\alpha\beta}) = \sum_{n=1}^{M-1} (F_{\alpha\beta}(q_n) - \bar{F}_{\alpha\beta}) \times (F^*_{\alpha\beta}(q_n) - \bar{F}^*_{\alpha\beta})/(M-1). \quad (4.9)$$

If the value of k approaches unity, then the positions of the symbols are completely correlated, whereas if we have $k \approx 0$, then there are no correlations. These characteristics can be calculated using the cyclic correlation functions [see (2.8)–(2.10), (2.15), and (2.16)]:

$$\sigma(K^c_{\alpha\beta}) = \sigma(F_{\alpha\beta})/M^{1/2}, \quad (4.10)$$

$$k(K^c_{\alpha\beta}|K^c_{\gamma\delta}) = k(F_{\alpha\beta}|F_{\gamma\delta}). \quad (4.11)$$

Using the fact that the relative fluctuations of the spectral sums are of order $\sim M^{-1/2}$ and passing from the average over the spectrum to the ensemble average, we find using (3.19) and (3.20)

$$\sigma^2(F_{\alpha\beta}) = \bar{F}_{\alpha\alpha}\bar{F}_{\beta\beta}, \quad (4.12)$$

$$k(F_{\alpha\beta}|F_{\gamma\delta}) = \bar{F}_{\alpha\gamma}\bar{F}_{\delta\beta}/(\bar{F}_{\alpha\alpha}\bar{F}_{\beta\beta}\bar{F}_{\gamma\gamma}\bar{F}_{\delta\delta})^{1/2}. \quad (4.13)$$

In particular, for $\alpha \neq \beta$ we have

$$k(F_{\alpha\alpha}|F_{\beta\beta}) \equiv k_{\alpha\beta} = N_\alpha N_\beta/(M-N_\alpha)(M-N_\beta). \quad (4.14)$$

Equation (4.14) has a simple physical interpretation. The correlation coefficient $k_{\alpha\beta}$ is equal to the probability of simultaneously finding α symbols in locations which are free of β symbols, and vice versa.

The correlations can also be described in terms of the cross-information,[15]

$$I_{\alpha\beta} = \int P_2(|\rho_\alpha|,|\rho_\beta|;\varphi_\alpha,\varphi_\beta)\ln(P_2(|\rho_\alpha|,|\rho_\beta|;\varphi_\alpha,\varphi_\beta)/ P_1(|\rho_\alpha|)P_1(|\rho_\beta|))d\Omega_2. \quad (4.15)$$

Taking into account Eqs. (3.20), (3.22), (3.24), (3.25), and (4.14), we can convert the explicit expression for random sequences to the form (see Refs. 15 and 16)

$$I_{\alpha\beta} = -\ln(1-k_{\alpha\beta}). \quad (4.16)$$

The higher the cross-correlations, the greater the cross-information.

To conclude this section we introduce an estimate for the correlations between two uncorrelated random sequences (1 and 2) of the same length M (but in general, with different symbols):

$$k(F^{(1)}_{\alpha\beta}|F^{(2)}_{\gamma\delta}) = \sum_{n=1}^{M-1} (F^{(1)}_{\alpha\beta}(q_n) - \bar{F}^{(1)}_{\alpha\beta})(F^{*(2)}_{\gamma\delta}(q_n) - \bar{F}^{*(2)}_{\gamma\delta})/(M-1)\sigma(F^{(1)}_{\alpha\beta})\sigma(F^{(2)}_{\gamma\delta}).$$

After averaging independently over the ensemble we find

$$\langle k(F^{(1)}_{\alpha\beta}|F^{(2)}_{\gamma\delta})\rangle \approx 0,$$

$$\langle k^2(F^{(1)}_{\alpha\beta}|F^{(2)}_{\gamma\delta})\rangle \approx 1/M(\bar{F}^{(1)^2}_{\alpha\beta}\bar{F}^{(2)^2}_{\gamma\delta}/\bar{F}^{(1)}_{\alpha\alpha}\bar{F}^{(1)}_{\beta\beta}\bar{F}^{(2)}_{\delta\delta}\bar{F}^{(2)}_{\gamma\gamma}+1).$$

Thus, the mesoscopic cross-correlations of two uncorrelated random sequences are found to be of order $\sim M^{-1/2}$.

4.4. Structural entropy of a sequence

The cross-correlation coefficients described in the previous section only characterize the relative positions of sequences of symbols, not the degree of disorder. Specifically, two almost identical random sequences would be strongly correlated, but would still be random. Below we present some criteria for disorder.

Consider a function $f(x)$ with monotonic first derivative and smooth second derivative. Hence the equation $f'(x) = \text{const}$ has a unique solution. Next we consider the spectral sum

$$S_\alpha = \sum_{n=1}^{M-1} f(F_{\alpha\alpha}(q_n)/\bar{F}_{\alpha\alpha}) \quad (4.17)$$

with the auxiliary condition

$$\sum_{n=1}^{M-1} F_{\alpha\alpha}(q_n)/\bar{F}_{\alpha\alpha} = \text{const} \quad (4.18)$$

[corresponding to the sum rule (2.13)]. Using the standard technique of Lagrange multipliers we find that a local extremum S_α is attained for a strictly uniform distribution of amplitudes according to the spectrum

$$F_{\alpha\alpha}(q_n) = \bar{F}_{\alpha\alpha}. \quad (4.19)$$

Since for random sequences the harmonics are distributed more uniformly over the spectrum than for ordered sequences (cf. the spectra for crystals and quasicrystals with sharp Bragg peaks), S_α can be used as an approximate structural entropy. A more precise quantitative criterion can be found by averaging S_α using the distribution function (3.25),

$$\langle S_\alpha \rangle = (M-1)\langle f(F_{\alpha\alpha}/\bar{F}_{\alpha\alpha})\rangle, \tag{4.20}$$

$$\langle f(F_{\alpha\alpha}(q_n)/\bar{F}_{\alpha\alpha})\rangle = \int_0^\infty f(F_{\alpha\alpha}(q_n)/\bar{F}_{\alpha\alpha}) P_1(F_{\alpha\alpha}) dF_{\alpha\alpha}. \tag{4.21}$$

Typical choices of $f(F_{\alpha\alpha}(q_n)/\bar{F}_{\alpha\alpha})$ correspond to

$$f(F_{\alpha\alpha}(q_n)/\bar{F}_{\alpha\alpha}) = \begin{cases} \ln(F_{\alpha\alpha}/\bar{F}_{\alpha\alpha}), & (4.22a) \\ -(F_{\alpha\alpha}/\bar{F}_{\alpha\alpha})\ln(F_{\alpha\alpha}/\bar{F}_{\alpha\alpha}), & (4.22b) \\ (F_{\alpha\alpha}/\bar{F}_{\alpha\alpha})^r, & (4.22c) \end{cases}$$

and the corresponding averages are equal to

$$\langle f(F_{\alpha\alpha}/\bar{F}_{\alpha\alpha})\rangle = \begin{cases} -\mathbf{C}, & (4.23a) \\ -(1-\mathbf{C}), & (4.23b) \\ \Gamma(r+1), & (4.23c) \end{cases}$$

where $C=0.577215...$ is the Euler constant. The choice (4.22a) corresponds to the spectral definition of the information entropy[19] (other definitions are equivalent only for quasi-Gaussian statistics). With this choice, however, S_α diverges for an arbitrary sequence with hidden periodicity, when some of the harmonics vanish (this choice does not distinguish, e.g., between a doublerandom sequence and a sequence with period two). Consequently in what follows we primarily use the definition (4.22b). The functions (4.22a) and (4.22b) correspond to the local maximum of the structural entropy (4.17), while the power law (4.22c) yields a maximum in the range of exponents $0<r<1$ and a minimum for $r>1$.

For scale-invariant sequences we can also use the local spectral probabilities introduced in Ref. 4,

$$p_\alpha(q_n) = F_{\alpha\alpha}(q_n)/(M-1)\bar{F}_{\alpha\alpha}. \tag{4.24}$$

The properties of such sequences are described using the multidimensional multifractal spectrum (see Refs. 4 and 31)

$$\sum_{n=1}^{M-1} p_{\alpha_1}^{r_1}(q_n) \ldots p_{\alpha_{l-1}}^{r_{l-1}}(q_n) \sim (M-1)^{-\tau(r_1,\ldots,r_{l-1})}. \tag{4.25}$$

In the special case of the subsequence consisting of the α symbols we find

$$\sum_{n=1}^{M-1} p_\alpha^r(q_n) = C_{r,\alpha}(M-1)^{-(r-1)D_{r,\alpha}}, \tag{4.26}$$

where $D_{r,\alpha}$ is the Renyi dimensionality and $C_{r,\alpha}$ is a constant. It follows from the considerations at the beginning of this section that if the inequality $D_r^{(1)} > D_r^{(2)} (r>0)$ holds for two scale-invariant subsequences of symbols, then the first subsequence can be regarded as more disordered than the second. If two subsequences belong to a single universality class and have the same Renyi dimensionality D_r, where $C_r^{(1)} > C_r^{(2)} (0<r<1)$ and $C_r^{(1)} < C_r^{(2)} (r>1)$, then this conclusion remains valid with respect to the two subsequences from a single universality class.

It can be seen from the results of Sec. 4.1 and Eq. (4.23c) that for random sequences with the choice (4.23c)

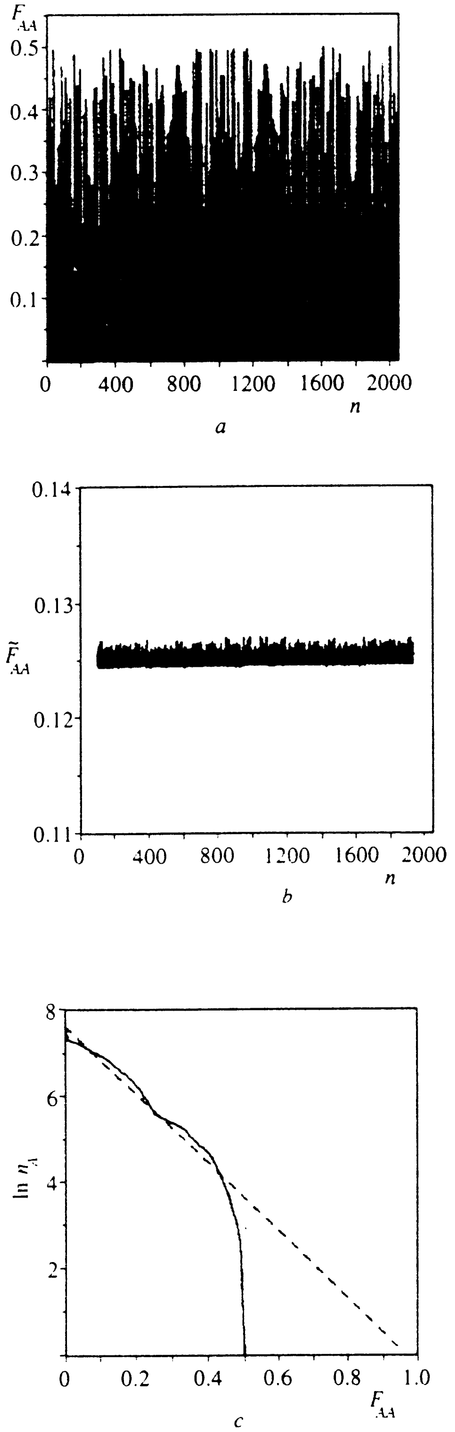

FIG. 1. Spectral characteristics for the Rudin–Shapiro sequence after $p=10$ iterations, beginning with $ABCD$ ($M=4096$); a) harmonics for the diagonal elements of the structure factor $F_{AA}(q_n)$, $I \leqslant n \leqslant M/2-I$; b) smoothed spectrum with $s=100$ [Eq. (4.5)]; c) logarithm of the number of harmonics greater than a fixed value of F_{AA} (solid trace). The dashed line corresponds to the theoretical prediction for a random sequence [Eqs. (4.1b) and (5.5)].

the mesoscopic fluctuations dominate in the range of indices $r<-1$ and $r \gtrsim \ln M/\ln\ln M$. In these cases individual harmonics with magnitudes of order $F_{\alpha\alpha,\min}$ and $F_{\alpha\alpha,\max}$ dominate in the sum (4.17). The situation in which analytical continuation into the region $r<0$ is impossible because of mesoscopic fluctuations is typical for many scale-invariant sequences.[4] The requirement that S_α be as insensitive as possible to mesoscopic fluctuations in practice determines the optimum choice of the function $f(F_{\alpha\alpha}/\bar{F}_{\alpha\alpha})$. However, the exact criterion requires that correlations of harmonics with different q_n be taken into account, and it is difficult to derive.

5. APPLICATIONS OF THE RESULTS

5.1. Substitution sequences

The application of the above criteria can be illustrated in some specific examples. For comparison we will consider two cases, in which one of the sequences is nearly random and the other is far from random, but has a fairly broad spectrum. We begin with substitution sequences whose growth is determined by the iterative substitutions

$$A_1 \to \sigma_1(A_1,...,A_l),...,A_l \to \sigma_l(A_1,...,A_l), \tag{5.1}$$

where $\sigma_k(A_1,...,A_l)$ is some combination of the symbols $A_1,...,A_l$. The substitution rules can be either deterministic [as in Eq. (5.1)] or probabilistic (see, e.g., Ref. 32).

In Fig. 1 typical results are shown for the Rudin–Shapiro sequence

$$A \to AC,\quad B \to DC,\quad C \to AB,\quad D \to DB, \tag{5.2}$$

after ten iterations, beginning with the initial sequence $ABCD$ ($M=4096$). It is well known[2,4] that Rudin–Shapiro substitution is very strongly randomizing. It can be verified that in the process of successive iterations the symbols A and D always remain in odd positions, while B and C only occupy even positions. This leads to sharp coherent Bragg peaks for $q_n=\pi$:

$$\rho_A(0)=\rho_B(0)=\rho_C(0)=\rho_D(0)=-\rho_A(\pi)=\rho_B(\pi)$$
$$=\rho_C(\pi)=-\rho_D(\pi)=N/M^{1/2}, \tag{5.3}$$

$$N=2^p,\quad M=2^{p+2},\quad p\geqslant 1, \tag{5.4}$$

where p is the number of iterations. Consequently, the Rudin–Shapiro sequence can be represented either as the result of intersite merging of two binary sequences (A–D and B–C) or else as partial destruction of an exact period-2 sequence. The coherent harmonics with $q_n=\pi$ should be subtracted from the various spectral sums and the values of the average harmonics redefined:

$$\bar{F}_{\alpha\alpha}=2\left(\sum_{n=1}^{M/2-1} F_{\alpha\alpha}(q_n)\right)/(M-2)=2^{p-2}/(2^{p+1}-1). \tag{5.5}$$

The relative positions of the symbols A and D, and also of B and C, are completely correlated, $k_{AD}=k_{BC}=1$ [as they should be for binary sequences (see Eqs. (2.18) and (4.14))], while at the same time the correlations between the symbols in the even and the odd positions are practically zero, $k_{AB}=k_{AC}=k_{DB}=k_{DC}=1.47\times 10^{-3}$ for $p=10$. To three significant figures the values of the dispersion coincide with the average values of the harmonics (5.5) [compare Eq. (4.12)]. The magnitudes of the structural entropies for different symbols (4.17) and (4.22b) (disregarding the contribution from $q_n=\pi$) are equal to $2.20\cdot 10^3$ for $p=10$ [compared with $1.73\cdot 10^3$ according to (4.20) and (4.23b) when $(M-1)$ is replaced by $(M-2)$]. In Fig. 1c noticeable deviations are observed only for $\sim$30–40 harmonics out of 2047 (and these are toward the smaller amplitudes). All these results confirm the ideas about randomization in Rudin–Shapiro substitution (while preserving, however, the partially broken period-2 property).

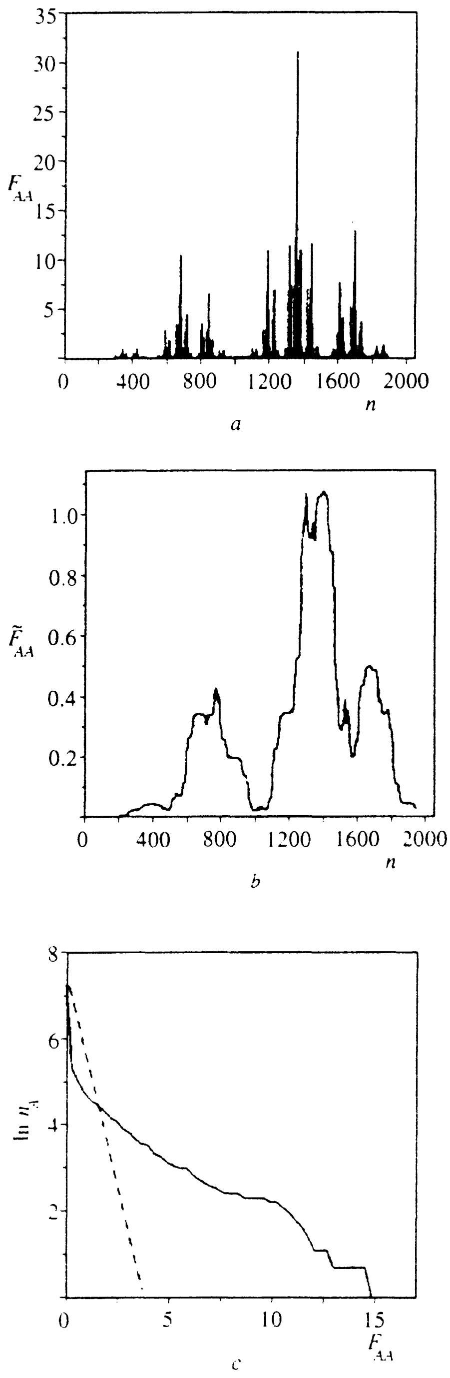

FIG. 2. Spectral characteristics for the Toohey-Morse sequence after $p=12$ iterations, beginning with A ($M=4096$): a) harmonics for $F_{AA}(q_n)=F_{BB}(q_n)$, $1\leqslant n\leqslant M/2$; b) smoothed spectrum with $s=100$; c) logarithm of the number of harmonics greater than a fixed value F_{AA} (solid trace). The dashed line corresponds to the theoretical prediction for a doubled random sequence with the same total number of symbols.

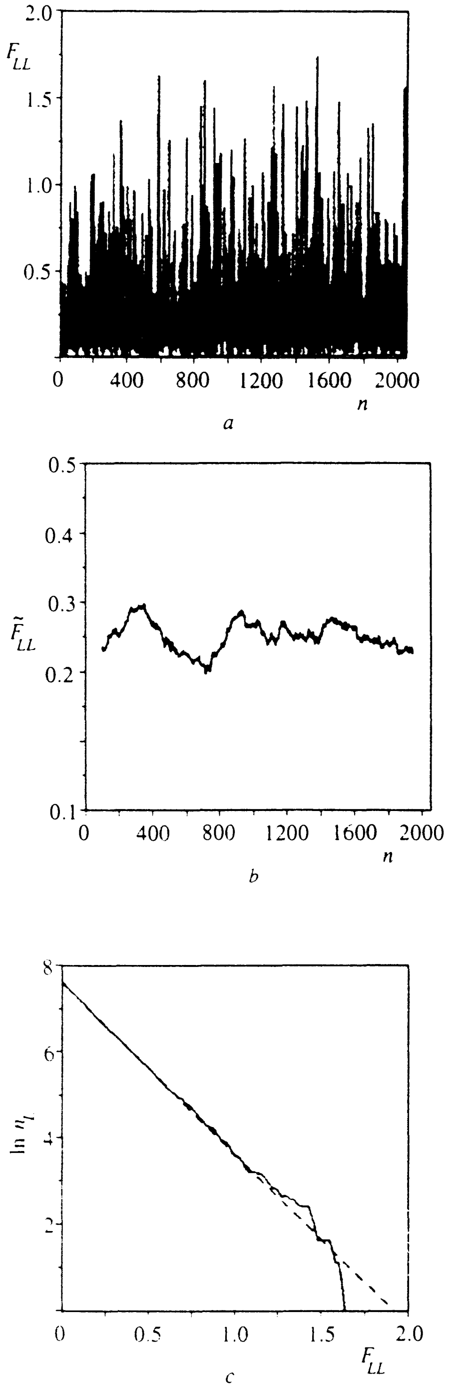

FIG. 3. Spectral characteristics for a sequence of symbols generated by the logistic mapping (5.7) for $r=4$ and $x_1=0.4$ after $p=4095$ iterations ($M=4096$); a) harmonics for $F_{LL}(q_n)=F_{RR}(q_n)$ with $1\leqslant n\leqslant M/2$; b) smoothed spectrum with $s=100$; c) logarithm of the number of harmonics greater than a given value F_{LL} (solid trace). The dashed line corresponds to the theoretical prediction (4.1b) for a random sequence with $\bar{F}_{LL}$, defined by Eq. (5.8).

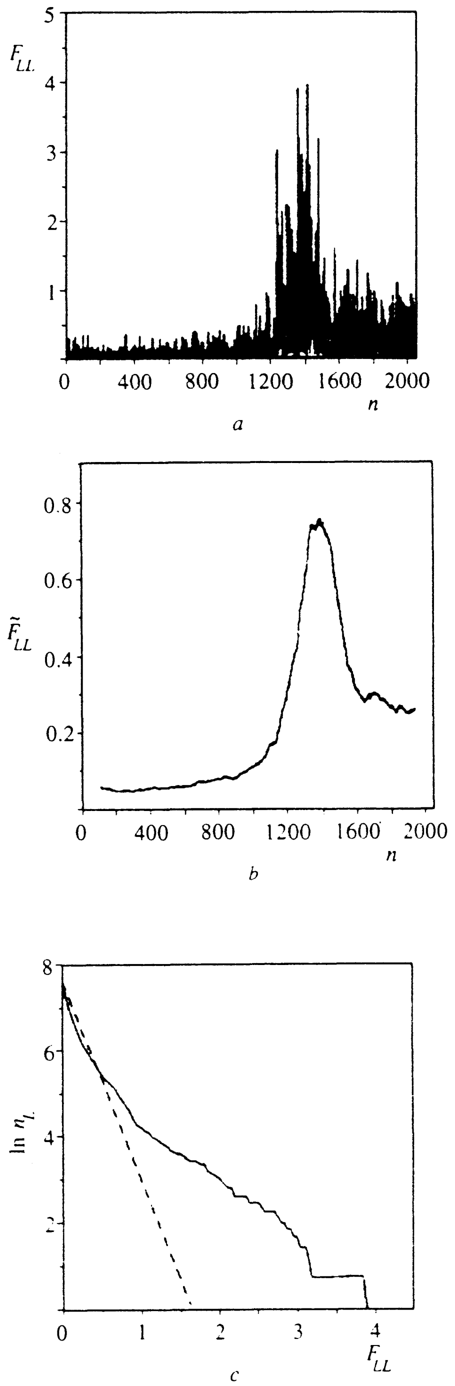

FIG. 4. Structural properties for a sequence generated by the logistic mapping with the same parameters as in Fig. 3 except $r_c-r=0.002$; $r_c=1+\sqrt{8}$.

In Fig. 2 a second example is shown for the Toohey–Morse binary sequence:

$$A\rightarrow AB, \quad B\rightarrow BA, \tag{5.6}$$

after twelve iterations, beginning with the initializing symbol A ($M=4096$). For this sequence all even harmonics vanish.[4,5] Hence we will compare it with the doubled random sequence (with zero odd harmonics). In what follows all averages and sums are evaluated only for nonzero harmonics. Then $\bar{F}=1/2$, holds, and the calculated values of the dispersion and the structure entropy (4.17), (4.22b) for $p=12$ are equal to 1.74 and $3.94\cdot 10^3$, compared with the values 0.5 and $8.66\cdot 10^2$ for the corresponding doubled random sequence with the same total number of symbols. The maximum height of the $n=1365$ harmonic is equal to $F_{AA,\max}=31.0(F_{\max}/\bar{F}=62.0)$, and is considerably greater than the spikes in the random sequences [Eq. (4.2)]. Thus, the Toohey–Morse sequence is in fact far from random, in accordance with the rigorous theory.

5.2. Symbolic dynamics

A rough description of the behavior of dynamical systems can be expressed in terms of symbolic dynamics.[6–8] For this purpose all of phase space is broken up into nonoverlapping cells, each of which is assigned a specific symbol. The evolution of the point describing the system in phase space generates a sequence of symbols. A natural form of subdivision exists only for one-dimensional mappings. As an illustration we take the thoroughly studied logistic mapping,[33]

$$x_{n+1}=rx_n(1-x_n), \quad 1<r\leqslant 4. \tag{5.7}$$

The interval [0,1] is determined by the zero of the product $rx(1-x)$, and the sequence is specified as follows. If $0<x_n<1/2$ holds, then the nth position is filled with the symbol L, but if $1/2<x_n<1$ holds, then the symbol R is inserted. The average values of the harmonics are equal:

$$\bar{F}_{LL}=\bar{F}_{RR}=N_LN_R/M(M-1), \tag{5.8}$$

where N_L and N_R are the total numbers of the symbols L and R, and $N_L+N_R=M$ is the total length of the sequence.

Figures 3 and 4 display the results for completely chaotic evolution ($r=4$) and the approximately period-3 dynamics with random alternation ($r_c-r=0.002$, $r_c=1+\sqrt{8}$) (Refs. 33 and 34). The numerical values of the various parameters for $x_1=0.4$ after $p=4095$ iterations ($M=4096$) are equal to: 1) $r=4$; $N_L=2040$, $N_R=2056$; $\bar{F}_{LL}=\bar{F}_{RR}=0.250$; $\sigma(F_{LL})=\sigma(F_{RR})=0.251$; the magnitude of the structural entropy defined according to (4.17) and (4.22b) is $1.75\cdot 10^3$; the theoretical values of the entropy for random sequences [Eqs. (4.20) and (4.23b)] is $1.73\cdot 10^3$; 2) $r_c-r=0.002$, $r_c=1+\sqrt{8}$; $N_L=1263$, $N_R=2833$; $\bar{F}_{LL}=\bar{F}_{RR}=0.213$; $\sigma(F_{LL})=\sigma(F_{RR})=0.370$; the value of the structural entropy [Eqs. (4.17) and (4.22b)] is $3.39\cdot 10^3$ when the theoretical value of the random sequences is the same, $1.73\cdot 10^3$. These results display the fairly high sensitivity of the spectral criteria.

6. CONCLUSION

Our results show that the spectral representation gives rise to simple and convenient criteria for the correlation of different symbols and the degree of disorder in a sequence. In this work we have restricted ourselves to comparisons with purely random sequences. It is not difficult to understand qualitatively how the spectral properties change for sequences with a finite memory. For systems with a finite memory (including the case of damping p-periodicity) we can use the following approximate expressions for pair correlation functions [Eq. (2.8) with $m_0<M/2$]:

$$\Delta K_{\alpha\alpha}(m_0)=K_{\alpha\alpha}(m_0)-\bar{K}_{\alpha\alpha}\propto\exp(-m_0/r_c), \tag{6.1a}$$

$$\Delta K_{\alpha\alpha}(m_0)\propto\cos(2\pi m_0/p)\exp(-m_0/r_c). \tag{6.1b}$$

By inverting the Wiener–Khinchin relation (2.10) we find that in the presence of additional correlations ($\Delta K>0$) the spectrum in the region of harmonics with small wave numbers ($q_n\lesssim 1/r_c$) will be enriched with harmonics having larger amplitudes in comparison with the remaining part of the spectrum, while in the case of any correlations ($\Delta K<0$) the amplitudes in this part of the spectrum will be somewhat reduced. Damping of the periodicity [Eq. (6.1b)] leads to spearing-out and to a characteristic Lorentz form of the Bragg peaks with width $\Delta q\sim 1/r_c$. These spectra enable us to roughly identify finite correlations, but the related quantitative criteria require additional consideration. Spectral analysis also enables us to exhibit long-range correlations in the system at distances comparable with the total length of the sequence.[11,13,27] Thus, the method covers essentially the whole range of investigation of the structural properties of sequences.

We express our gratitude to A. A. Vedenov, A. A. Ezhov, V. A. Kutvitskii, and A. L. Chernyakov for useful discussions.

We are also obliged to the Soros Fund for partial financial support of the work.

[1] T. Janssen and A. Janner, Adv. Phys. **36**, 519 (1987).
[2] M. Queffelec, *Substitution Dynamical Systems: Spectral Analysis* (Lecture Notes in Mathematics, Vol. **1294**), Springer, Berlin (1987).
[3] J. M. Perez-Moto, F. J. Zuniga, and G. Madariaga (eds.), *Methods of Structural Analysis of Modulated Structures and Quasicrystals*, World Scientific, Singapore (1991).
[4] C. Godreche and J. M. Luck, J. Phys. A: Math. Gen. **23**, 3769 (1990).
[5] Z. Cheng, R. Savit, and R. Merlin, Phys. Rev. B **37**, 4375 (1988); Z. Cheng and R. Savit, Phys. Rev. A **44**, 6379 (1991).
[6] R. Bowen, *Methods of Symbolic Dynamics* [Russian translation], Mir, Moscow (1979).
[7] V. M. Alexeev and M. V. Yakobson, Phys. Rep. **75**, 290 (1981).
[8] B.-L. Hao, *Elementary Symbolic Dynamics and Chaos in Dissipative Systems*, World Scientific, Singapore (1989); Physica D **51**, 161 (1991).
[9] P. Grassberger, R. Badii, and A. Politi, J. Stat. Phys. **51**, 135 (1988).
[10] A. Politi, Phys. Lett. A **136**, 374 (1989).
[11] W. Li and K. Kaneko, Europhys. Lett. **17**, 655 (1992).
[12] W. Li, Int. J. Bifurcation and Chaos **2**, 137 (1992).
[13] R. F. Voss, Phys. Rev. Lett. **68**, 3805 (1992).
[14] C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, Nature **356**, 168 (1992).
[15] C. E. Shannon, in *Theory of Electrical Signal Transmission in the Presence of Noise* [Russian translation], IL, Moscow (1953).
[16] A. N. Kolmogorov, *Information Theory and the Theory of Algorithms* [in Russian], Nauka, Moscow (1987).
[17] B. L. Burrows and K. W. Sulston, J. Phys. A: Math. Gen. **24**, 3979 (1991).
[18] R. Kluiving, H. W. Capel, and R. A. Pasmanter, Physica A **183**, 67, 96 (1992); **186**, 405 (1992).
[19] S. L. Marple, Jr., *Digital Spectral Analysis*, Prentice-Hall, Englewood Cliffs, NJ (1987).
[20] A. D. McLachlan and M. Stewart, J. Mol. Biol. **103**, 271 (1976).
[21] J. Felsenstein, S. Sawyer, and R. Kochin, Nucleic Acids Res. **10**, 133 (1982).
[22] B. D. Silverman and R. Linsker, J. Theor. Biol. **118**, 295 (1986).
[23] S. Tavare and B. W. Giddings, in *Mathematical Methods for DNA Sequences*, M. S. Waterman (ed.), CRC Press, Boca Raton, FL (1989).
[24] A. A. Deev, I. M. Edintsov, G. R. Ivanitskiĭ, and A. S. Kuniskiĭ, Biofiz. **34**, 564 (1989).
[25] D. C. Benson, Nucleic Acids Res. **18**, 3001, 6305 (1990).
[26] E. A. Cheever, G. C. Overton, and D. B. Searls, Comput. Appl. Biosci. **7**, 143 (1991).
[27] V. R. Chechetkin, L. A. Knizhnikova, and A. Yu. Turygin, J. Biomol. Struct. Dyn. **12**, No. 4 (1994).
[28] N. G. Van Kampen, *Stochastic Processes in Chemistry and Physics*, revised edition, North-Holland, Amsterdam (1992).

[29] S. M. Rytov, *Principles of Statistical Radiophysics*, 2nd ed., Springer, New York (1987).
[30] M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, Springer, Berlin (1983).
[31] C. Meneveau, K. R. Sreenivasan, P. Kailasnath, and M. S. Fan, Phys. Rev. A **41**, 894 (1990).
[32] W. Li, Phys. Rev. A **43**, 5240 (1991).
[33] H. G. Schuster, *Deterministic Chaos*, Physik-Verlag, Deerfield Park, FL (1984).
[34] J. E. Hirsch, B. E. Huberman, and D. Scalapino, Phys. Rev. A **25**, 519 (1981).

Translated by David L. Book