Random matrix approach to multivariate correlations: Some limiting cases

M. S. Santhanam and Holger Kantz
Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Strasse 38, Dresden 01187, Germany
(Received 25 June 2003; revised manuscript received 23 January 2004; published 4 May 2004)

Recent advances have shown that the empirical correlation matrices of dynamical systems can be modeled as random matrices, for most part, chosen from an appropriate ensemble of the random matrix theory (RMT). In this work, we study certain limiting cases where this approach could potentially break down. Using a combination of analytical and numerical tools, we especially study the eigenvalue density and its spacing distribution. We show that the correlation matrices obtained from multivariate spatiotemporal timeseries, in a regime of spatiotemporal chaos, lead to strong deviations from RMT. We illustrate the results with time-series data drawn from coupled map lattices. We also explore the transition to the RMT regime from the limiting cases.

DOI: 10.1103/PhysRevE.69.056102 PACS number(s): 02.50.Sk, 05.45.Tp, 05.40.−a, 05.45.Ra

I. INTRODUCTION

The study of spatial correlations in spatially extended, complex systems such as economics, weather, social sciences, etc., provides us with deeper insight into information and structure generated by their dynamics. However, in many of the most interesting cases *ab initio* models are currently not established, so that theoretical studies of correlations cannot be performed. The alternative is to investigate the properties of observed time-series data, including numerical estimates of spatial correlations. The evaluation of such numerical results requires the existence of a paradigm, which gives the framework for their interpretation. Such a paradigm is supplied by random matrix theory (RMT) [1], which was originally developed for the interpretation of nuclear spectra and later on gained new relevance for the understanding of quantum systems whose classical counterpart is chaotic. Recently, it was reported from several empirical studies that the spectra of correlation matrices, constructed from multivariate time-series data, can be modeled as random matrices chosen from an appropriate ensemble of the random matrix theory. Such results were demonstrated to hold in several interesting practical cases such as stock-market fluctuations, ECG data, atmospheric time series, and internet interconnections [2]. These findings testify the universal nature of spectral fluctuations, which underlie the RMT line of thinking, in the correlation matrix spectra.

The interpretation of the spectra of empirical correlation matrices requires more care than of corresponding matrices from, e.g., quantum chaos. In both cases, we would like to distinguish between system specific signatures and universal features. The former express themselves in the smoothed level density, whereas the latter are represented by the fluctuations around this smooth curve. In time-series analysis, the matrix elements are subject to uncertainty such as measurement noise on the time-series data, but also statistical fluctuations due to finite sample effects. When characterizing time-series data in terms of RMT, we do not want to characterize these trivial sources of fluctuations which are present on every data set, but we aim at the extraction of features which would be shared by an infinite amount of data without measurement noise. The eigenfunctions of the correlation matrices constructed from such empirical time series carry the information contained in the original time-series data in a graded manner and provide a compact representation for it. Thus, by applying an RMT based approach, we can identify nonrandom components of the correlation matrix spectra as deviations from RMT predictions [3]. In terms of applications, RMT provides another criterion to distinguish between signal and noise in correlation matrix spectra.

Even though the previous studies of empirical time series confirm the RMT-like universal fluctuations in correlation matrix spectra, there are evident and relevant limiting cases that are not yet explored in this context. For instance, a multivariate data set that is perfectly correlated will not display RMT-like spectral fluctuations. This question can be treated more broadly in terms of eigenvalue spectral fluctuation properties. Within the class of real symmetric matrices, the eigenvalues of correlation matrices can display one of the following properties: (i) perfectly correlated eigenvalues, leading to equally spaced eigenvalues, (ii) completely uncorrelated eigenvalues and (iii) eigenvalues displaying various degrees of correlation, called the "level repulsion" in RMT language. All the studies reported so far of empirical time-series data displaying RMT-like spectral fluctuations, fall in the category (iii) above. The first two cases [(i) and (ii)] could be thought of as limiting cases where RMT based arguments could break down. It is in fact known from quantum chaos literature that the distribution of random level spacing, coming from completely uncorrelated eigenvalues, cannot be described by the standard Wigner-Dyson RMT approach but is known to be Poisson distributed from semiclassical considerations [4]. As some properties of the time-series data change, it is possible to go from a limiting case to the RMT domain characterized by level repulsion or vice versa.

In this paper, we will examine the limiting case corresponding to perfectly correlated eigenvalues of the correlation matrix and explore under what circumstances transition to RMT-like spectral fluctuations take place. In answering this question, we are also led to another interesting problem of the spectral signature of spatiotemporal chaos. In all the previous works that consider spectra of correlation matrices from empirical time series [2], it is not known with certainty if those data are chaotic or not. Thus, the spectral manifes-

tations of spatiotemporal chaos remain relatively unexplored. To study both these questions, we assume simple models for correlation matrices which can be manipulated analytically to a certain extent. Further, we use coupled map lattices (CML) that have emerged as one of the paradigms of spatiotemporal dynamics, to generate multivariate time series with desired properties. Thus, this effort represents an attempt to understand the conditions under which RMT models for correlation matrix spectra will break down and to look for any alternative ways to understand such spectra. On the other hand, various dynamical regimes of CMLs might be distinguishable in terms of the properties of their correlation matrix spectra. This will contribute to the understanding the spatiotemporal dynamical features in terms of its correlation matrix spectra. In the following section, we will introduce the coupled map lattice and the correlation matrix formalism.

II. CORRELATIONS IN COUPLED MAP LATTICES

A. Correlation matrix

Consider a time series of the form $z'(x,t)$, where $x=1,2,\ldots,n$ and $t=1,2,\ldots,p$ denote the discretized space and time, respectively. In this, the time series at every spatial point is treated as a different variable. We define the normalized variable as

$$z(x,t)=\frac{z'(x,t)-\langle z'(x)\rangle}{\sigma(x)}, \tag{1}$$

where the brackets $\langle\cdot\rangle$ represent temporal averages and $\sigma(x)$ the standard deviation of z' at position x. Then, the equal-time cross-correlation matrix that represents the spatial correlations can be written as

$$S_{x,x'}=\langle z(x,t)z(x',t)\rangle, \quad x,x'=1,2,\ldots,n. \tag{2}$$

The correlation matrix is symmetric by construction. In addition, a large class of processes are translationally invariant and the correlation matrix can contain that additional symmetry, too. We will use this property for our correlation models in the context of CML. In time-series analysis, the averages $\langle\cdot\rangle$ have to be replaced by estimates obtained from finite samples. As usual, we will use the maximum likelihood estimates, $\langle a(t)\rangle\approx 1/p\sum_{t=1}^{p}a(t)$. These estimates contain statistical uncertainty, which disappears for $p\to\infty$. Ideally we require $p\gg n$ to have reasonably correct correlation estimates.

B. Coupled map lattices

The concept of CML was introduced as a simple model capable of displaying complex dynamical behavior generic to many spatiotemporal systems and has been extensively studied in the last 20 years [5]. CMLs have discrete time and space, labeled by t and i, respectively, in Eq. (3), but a continuous state space. Many spatiotemporal phenomena could be modeled by CML dynamics [6]. In this work, we use CMLs as a source of multivariate spatiotemporal data with required properties. They can be easily created numerically by simple iteration. By a change of system parameters we can tune the dynamics for desired spatial correlation properties, many of which have already been studied and reported [7]. We consider the class of diffusively coupled map lattices in one dimension with sites $i=1,2,\ldots,n'$ of the form

$$y_{t+1}^{i}=(1-\epsilon)f(y_t^i)+\frac{\epsilon}{2}[f(y_t^{i+1})+f(y_t^{i-1})], \tag{3}$$

where $f(y)=1-ay^2$ is the local logistic map controlled by the parameter a. The parameter ϵ is a measure of coupling between nearest-neighbor lattice sites. We choose periodic boundary conditions, $x(n+1)=x(1)$. In order to eliminate boundary effects from our correlation matrices, we construct them from the sublattice $x\in[20,n'-20]$. If the spatial two-point correlations decay very fast, a lattice point 20 sites away from the boundary can be supposed to be insensitive to the boundary conditions. For the numerical computations reported in this paper, the CML with $n'=500$ and hence a sublattice size $n=460$ was chosen and iterated, starting from random initial conditions, for $p=5\times10^7$ time steps, after discarding 10^5 transient iterates.

As the parameters a and ϵ are varied, the spatiotemporal map displays various dynamical features such as frozen random patterns, pattern selection, space-time intermittency, and spatiotemporal chaos [7]. Since we intend to study limiting cases of strongly diagonally dominant correlation matrices, the interesting CML dynamics is found in the regime of spatiotemporal chaos, where correlations are known to decay rather quickly as a function lattice sites.

III. CORRELATION MATRIX MODEL

A. Tridiagonal matrix model

By definition of uncorrelatedness, the correlation matrix of a sequence of uncorrelated random variables is a diagonal matrix [8]. Here, we consider dynamics of CML corresponding to spatiotemporal chaotic regime where the spatial correlation decays fairly sharply ensuring that the correlation matrix is strongly diagonally dominant. In this context, we get for the empirical correlation matrix, in the limit as $t\to\infty$, a symmetric banded matrix whose bandwidth depends on how quickly the spatial correlations decay to zero.

First, we consider the simplest tridiagonal matrix model corresponding to extremely fast decay of correlations. Hence, the correlation matrix of order $n-1$ has the following form

$$\mathbf{X}=\begin{pmatrix} 1 & b & 0 & 0 & \cdots \\ b & 1 & b & 0 & \cdots \\ 0 & b & 1 & b & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}. \tag{4}$$

Since a correlation matrix is a positive semidefinite matrix, it is required that $b<1/2$ in order for $\mathbf{X}$ to be a proper model of a correlation matrix.

The eigenvalues of $\mathbf{X}$ are given by

$$\lambda_j^o=1+2b\cos(j\pi/n), \quad j=1,2,\ldots,(n-1). \tag{5}$$

The eigenvectors are

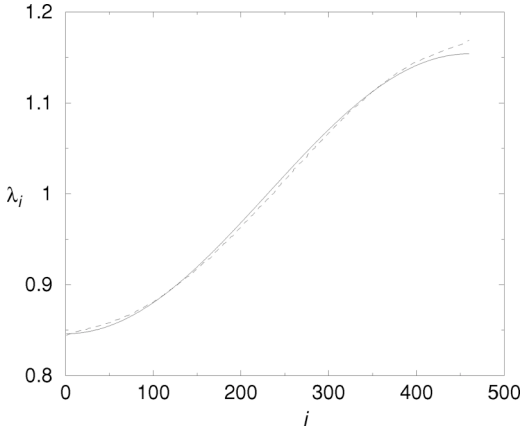

FIG. 1. Numerically determined eigenvalues of the correlation matrix for CML time-series data with $a=2.00$ and $\epsilon=0.08$ (dashed curve). The solid curve is the analytical result of the tridiagonal model, Eq. (5) with b taken as the average of the first off-diagonal elements of the CML matrix; $b=0.022$.

$$v_j^k = \sin(kj\pi/n), \quad k = 1,2, \ldots, (n-1),$$

where k labels the eigenvector components. As shown in Fig. 1, the spectrum of a full correlation matrix from CML dynamics for very small coupling strength ϵ and large parameter a can be approximately modeled by this relation. This figure illustrates that the full correlation matrix is effectively tridiagonal in nature, where the particular structure of $\mathbf{X}$ reflects the translational invariance of the CML dynamics. The disagreement between the numerical CML eigenvalues and Eq. (5) has two sources, namely, the finite-time series length leading to statistical fluctuations of the CML matrix elements, and the fact that the correlation matrix from CML is not exactly tridiagonal.

B. Perturbed matrix model

In order to analyze a more realistic situation, we intend to allow more off-diagonal entries. We hence perturb the tridiagonal matrix of order $n-1$ by adding four more off diagonals, such that with $1 \geqslant b, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4 \geqslant 0$:

$$\mathbf{X^p} = \begin{pmatrix} 1 & b & \epsilon_1 & \epsilon_2 & \epsilon_3 & \epsilon_4 & 0 & \cdots \\ b & 1 & b & \epsilon_1 & \epsilon_2 & \epsilon_3 & \epsilon_4 & \cdots \\ \epsilon_1 & b & 1 & b & \epsilon_1 & \epsilon_2 & \epsilon_3 & \cdots \\ \epsilon_2 & \epsilon_1 & b & 1 & b & \cdots & \cdots & \cdots \\ \epsilon_3 & \epsilon_2 & \epsilon_1 & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}. \tag{6}$$

This is just the banded version of the well known symmetric Toeplitz matrix, which arises as the covariance matrix in the study of stationary and translationally invariant stochastic processes such as moving average processes, autoregressive processes, and harmonic processes [8]. This matrix structure is also reminiscent of a variant of the one dimensional tight binding lattice models (TBM) without disorder in condensed matter physics [9]. In typical tight binding models, often nearest neighbor hopping is considered. This is equivalent to the tridiagonal model, essentially a symmetric tridiagonal Toeplitz matrix, given in Eq. (4). But, $\mathbf{X^p}$ is the Hamiltonian for TBMs without disorder that account for longer range site to site interactions. Later we will see that the spectral fluctuation properties of $\mathbf{X^p}$ strongly resemble those of one-dimentional (1D) quantum systems. To the best of our knowledge, no closed form solutions are known for the eigenvalues of symmetric Toeplitz matrix, beyond those with tridiagonal structure [10]. To estimate the eigenvalues, we use the standard Rayleigh-Schrödinger perturbation theory in quantum mechanics [11] in a straightforward way to obtain the perturbed eigenvalues. The tridiagonal matrix $\mathbf{X}$ is taken as the unperturbed "system" whose spectrum is exactly known. The perturbation matrix $\mathbf{V}$ is similar to $\mathbf{X^p}$ but with its principal tridiagonal entries set to zero, such that $\mathbf{X^p}=\mathbf{X}+\mathbf{V}$. Then the perturbed eigenvalues, to first order, can be obtained from

$$\lambda_j^1 = \frac{\langle v_j^k | \mathbf{V} | v_j^k \rangle}{C_j}, \tag{7}$$

where C_j is the normalization constant

$$C_j = \frac{n-1}{2} - \frac{\cos j\pi \sin[(n-1)u]}{2 \sin u}. \tag{8}$$

Here and in the following we use $(u=j\pi/n)$ and $\cos j\pi = (-1)^j$. The perturbed eigenvalues, to first order, are given by

$$\begin{aligned} \lambda_j^1 = \lambda_j^o &+ \frac{\epsilon_1}{C_j}\left[(n-3)\cos 2u - \frac{(-i)^j \sin[(n-3)u]}{\sin u}\right] \\ &+ \frac{\epsilon_2}{C_j}\left[(n-4)\cos 3u - \frac{(-i)^j \sin(n-4)u}{\sin u}\right] \\ &+ \frac{\epsilon_3}{C_j}\left[(n-5)\cos 4u - \frac{(-i)^j \sin(n-5)u}{\sin u}\right] \\ &+ \frac{\epsilon_4}{C_j}\left[(n-6)\cos 5u - \frac{(-i)^j \sin(n-6)u}{\sin u}\right]. \end{aligned} \tag{9}$$

Note that as $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4 \to 0$, we recover the correct unperturbed eigenvalues. Another important condition is that $b, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4 \ll 1$, in order for the matrix to be in the perturbative regime. Second, there are further conditions on these parameters needed to guarantee the positive semidefiniteness of the correlation matrix. The fact that the jth perturbed eigenvalue depends on a single integer j, apart from other quantities independent of j, gives the clue that the system represented by the matrix in Eq. (6) is, in a sense, similar to one-dimensional problems like the harmonic oscillator in quantum mechanics, as far as the spectral properties are concerned. Later we will see that it helps to keep this analogy in mind since the spectral fluctuation properties of $\mathbf{X^p}$ strongly resemble those of 1D quantum systems.

It is straightforward to obtain the asymptotic form for the perturbed eigenvalues. They are, for $n \gg 1$,

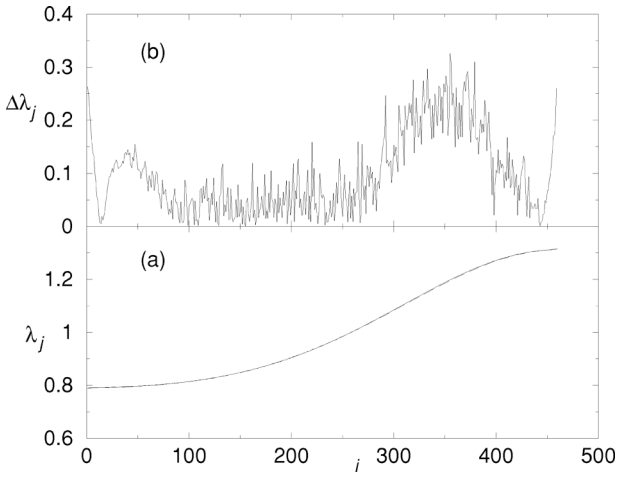

FIG. 2. (a) Numerically obtained eigenvalues of a correlation matrix for a CML with $a=1.98$ and $\epsilon=0.08$ (dashed line, almost invisible). Solid curve: analytical eigenvalues in Eq. (10) with $b=0.127, \epsilon_1=0.025, \epsilon_2=0.0023$ of the matrix $\mathbf{X}^p$, where the numerical values of the parameters were chosen according to the CML matrix. (b) Percentage error, $\Delta\lambda_j=\delta\lambda_j\times 100/\lambda_j$ between numerical and analytical eigenvalues.

$$\lambda_j^1=\lambda_j^o+2\sum_{k=1}^{4}\epsilon_k\cos(k+1)u. \tag{10}$$

In the regime of strong spatiotemporal chaos of CML, this relation provides an estimate for the eigenvalues of large correlation matrices. In Figs. 2 and 3(a), we show how well the analytical result in Eq. (10) fits the numerically calculated eigenvalues of full correlation matrix with specified CML parameters. We see that because of the effectively banded nature of the correlation matrix, the agreement with Eq. (10) is very good. The percentage error, shown in Figs. 2 and 3(b), between numerical eigenvalues and those given by Eq. (10) shows that the matrix in Eq. (6) is a fairly good approximation to the effectively banded correlation matrix, as far as the eigenvalues are concerned.

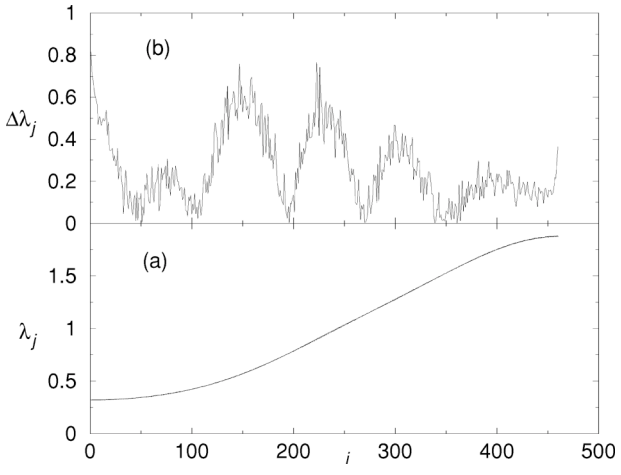

FIG. 3. The same as Fig. 2 for different CML parameters, $a=1.95$ and $\epsilon=0.2$.

In many cases where correlation matrices are constructed from experimental time series, the spectrum is characterized by a few dominant and a large number of less-dominant eigenvalues. In contrast to that, the eigenvalue curves in Figs. 2 and 3 do not have this feature, indicating that the spectral properties could be different from that of a generic correlation matrix. It is important to note that even though Eq. (10) is strictly valid for correlations as $t\to\infty$, we see that it is reasonably accurate even for correlations estimated from finite-time CML dynamics. However, as we will see later, the finite-time estimate of correlations lead to transitions in spectral statistics, especially in case of spacing distributions.

IV. EIGENVALUE DENSITY

We will look at two quantities that have been studied in the context of applying RMT methods to time-series correlations. The first of them is the eigenvalue (level) density, the second the level spacing distribution.

Let $N(\lambda)$ be the integrated eigenvalue density which gives the number of eigenvalues less than a given value λ. Then, the eigenvalue or level density is given by

$$d(\lambda)=\frac{dN(\lambda)}{d\lambda}. \tag{11}$$

This can be obtained assuming random correlation matrix [14] and is found to be in good agreement with the empirical time-series data from stock market fluctuations [13].

A. Tridiagonal case

In this case, the eigenvalue density can be exactly worked out as follows. From Eq. (5) we obtain

$$N(\lambda)=\frac{n}{\pi}\cos^{-1}\left(\frac{\lambda-1}{2b}\right). \tag{12}$$

Then, the normalized level density is given by

$$d_t(\lambda)=\frac{1}{2\pi b\sqrt{1-\left(\frac{\lambda-1}{2b}\right)^2}}, \tag{13}$$

where $1-2b<\lambda<1+2b$, where the full interval is only explored if the rank of matrix $n\to\infty$ [see Eq. (5)]. Note that the level density has peaks of equal magnitude at both the ends of the spectrum. Figure 4 shows that the analytical result in Eq. (13) is a fairly good approximation to the numerically obtained eigenvalue density of the full correlation matrix for CML data. The discrepancy between the two can be attributed to the effect of nonzero elements in the full correlation matrix, as opposed to tridiagonal matrix assumed by the theory, and to the finite-time estimate of correlations. The maximal (λ_+) and minimal (λ_-) eigenvalue, $\lambda_\pm\approx 1\pm 2b$, for large n, agree well with the range of eigenvalues which we find for the chosen CML data $b=0.0775$, as one can visually verify from Fig. 4.

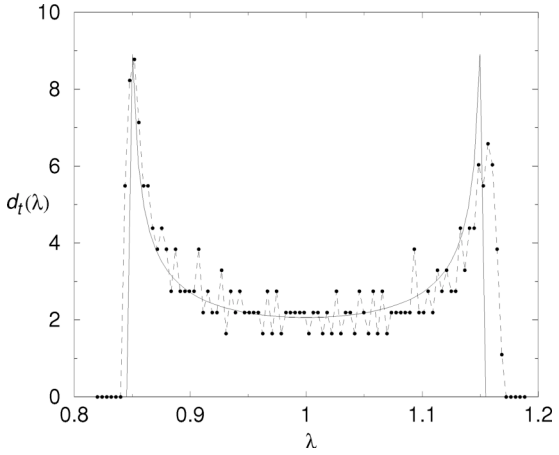

FIG. 4. Level density of spectra from the full correlation matrix for CML data with $a=2.0$ and $\epsilon=0.05$. The dotted curve is the numerical result, while the solid curve represents Eq. (13).

B. Beyond tridiagonal: Perturbed matrix

For a general perturbed matrix, it is not straightforward to obtain the integrated level density exactly. However, under certain constraints and approximations, we can obtain a closed formula. Further in this direction, we treat the special case, namely, geometric decay of correlations away from the principal diagonal of the correlation matrix.

1. Geometric decay of correlations

We start with Eq. (10) and assume a geometric decay of correlations so that

$$\epsilon_m \approx b^{m+1}, \quad m = 1, \ldots, 4. \tag{14}$$

By the transformation $b^{m+1}=e^{\log\, b^{m+1}}$, we see that this represents an exponentially decaying case. Invoking the fact that $\epsilon_m \ll 1$, we approximate λ_j from Eq. (10) in following way

$$\lambda_j \approx 1 + 2\sum_{m=1}^{\infty} b^m \cos mu. \tag{15}$$

Note that, for convenience, we have dropped the superscript on λ. The summation on right-hand side yields [12]

$$\lambda_j \approx \frac{1-b^2}{1-2b\,\cos u + b^2}. \tag{16}$$

Then, under this approximation, the integrated level density is given by

$$N(\lambda) = \frac{n}{\pi}\,\cos^{-1}\left(\frac{\lambda(b^2+1)+b^2-1}{2b\lambda}\right). \tag{17}$$

Now the normalized density of states can be determined to be

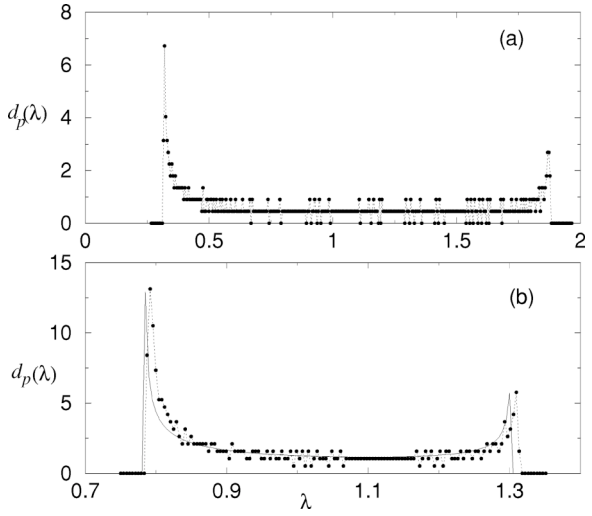

FIG. 5. Level density of the correlation matrix for CML with (a) $a=1.95$ and $\epsilon=0.2$ (b) $a=1.98$ and $\epsilon=0.08$. The solid curve in (b) is the analytical level density obtained from Eq. (18).

$$d_p(\lambda) = \frac{1}{2\pi\, b\, \lambda^2} \quad \frac{1-b^2}{\sqrt{1-\dfrac{(-1+b^2+(1+b^2)\,\lambda)^2}{4\,b^2\,\lambda^2}}}. \tag{18}$$

We compare this result with the numerical level density from the full correlation matrix spectra of CML for $a=1.98$, $\epsilon=0.08$. For this case, we have $b=0.128, \epsilon_1=0.025, \epsilon_2=0.0023, \epsilon_3=0.0008$, which roughly follow the assumption made in Eq. (14). As Fig. 5(b) shows, the agreement is fairly good, justifying the assumptions made in obtaining this relation.

We notice that level density and is peaked at the edges of the spectrum though the magnitudes of peaks at both the ends of the spectrum are not equal. The numerical result displayed in Fig. 5(a) for CML data with larger nearest-neighbor coupling has a qualitatively similar level density as the one predicted by Eq. (18). A quantitative comparison with the analytical result is not possible because the correlation coefficients do not decay in any obvious geometric sense, i.e., grossly violate the assumption in Eq. (14). It is clear that as the coupling strength in CML is increased, the lattice sites get strongly coupled which is reflected in the correlation coefficients that do not decay fast enough. Thus, the effective banded nature of correlation matrix is lost. Since numerically we have only the finite-time estimates of correlations, we tend, at least qualitatively, towards the limit of random matrix level density. This is evident from the fact that the asymmetry of the peaks in the level density gets further pronounced as coupling strength is increased (compare the Figs. 4 and 5. We recall that the random matrix level density for $p>n$ has a pronounced peak only at the lower end of the spectrum [14].

2. Eigenvalue bounds

From the integrated level density in Eq. (17), we can draw an important, though approximate, conclusion on the bounds of the eigenvalues of the correlation matrix under consideration. For $n \gg 1$, when the condition

$$\left(\frac{\lambda(b^2+1)+b^2-1}{2b\lambda}\right) \approx \pm 1 \tag{19}$$

is met, then approximately the integrated level density attains extremal values. We obtain the following relation for maximum (λ_+) and minimum (λ_-) eigenvalue

$$\lambda_\pm = \frac{1-b^2}{(1\pm b)^2}, \tag{20}$$

which implies the interesting result that

$$\lambda_+ \approx \frac{1}{\lambda_-}. \tag{21}$$

The correlation matrix spectra for CML lattice data with $a=1.98$ and $\epsilon=0.08$ have maximal and minimal eigenvalues to be 1.3097 and 0.7904, approximately fulfilling the above relation. This can also be verified in Fig. 5(b). This line of analysis allows us to understand the level density of correlation matrices with rapidly decaying correlations. Unlike the spacing distribution to be studied in the following section, the eigenvalue density is not, in general, expected to display universal behavior. But we expect our approach and the eigenvalue density relations to hold good for the case of rapidly falling correlations, which is a general feature of spatiotemporal chaotic systems.

V. SPACING DISTRIBUTION

In this section we investigate the nearest-neighbor eigenvalue spacing distribution, widely studied in the context of random matrix theory [1]. The eigenvalue spacings s are defined as the differences between the successive eigenvalues, $s_i=E_{i+1}-E_i$, where E_i is the unfolded eigenvalue. The unfolding procedure expresses the spacings in terms of local mean spacing. The spectrum is unfolded in order to have a uniform scale leading to mean spacing of unity so that the spacing distribution from various systems can be compared on an equal footing.

To unfold the spectrum, we note that the integrated level density is a sum of average part plus an oscillating part, written as, $N(\lambda)\equiv N_{avg}(\lambda)+N_{osc}(\lambda)$. A similar expression can be written for the level density $d(\lambda)$, too. Then, the unfolded levels are obtained as $E_i=N_{avg}(\lambda_i)$. In fact, for a general spectrum which is a function of a single integer ("quantum number") and $\lambda \geqslant 0$, the exact average level density for the nondegenerate case can be formally written down as [15]

$$d_{avg}(\lambda) = |N'(\lambda)|. \tag{22}$$

Thus, we note that Eq. (17) actually is the average part of the integrated level density. Since we have the analytical form of the average integrated level density $N(\lambda)$, we can unfold the spectrum analytically.

Our analytical framework does not account for the finite-time estimate of correlations. The statistical fluctuations break the translational invariance, to smaller or larger extents, in practice. Second, inspite of spatiotemporal chaos, the spatial correlations in CML do not monotonously decay to zero. Thus, the numerical spectra always contain such effects that arise from dynamics of "real" systems. Our work with CML indicates, at a qualitative level, that the spacing distribution is more sensitive to eigenvalue perturbations than the level density. Hence, we compare the spacing distribution results with those determined numerically from CML dynamics for (i) a truncated correlation matrix in which certain effective number of off-diagonal elements are retained and the rest set to zero; (ii) the full correlation matrix. The spectra of the truncated correlation matrices will allow us to verify the ideally expected spacing distribution while the full matrix will illustrate the real scenario.

A. Tridiagonal matrix

Taking the cue from the analogy of the correlation matrix with one-dimensional quantum systems, we expect the spectrum to display constant level spacing and its distribution to be strongly peaked at $s=1$. In our approximate scheme, this result arises easily, since the average part of integrated level density is, $N_{avg}(\lambda_j^o)=N(\lambda_j^o)$. The unfolded spectrum then follows from Eq. (12),

$$E_j = N_{avg}(\lambda_j^o) = N(\lambda_j^o) = j, \quad j=1,2,\ldots,(n-1). \tag{23}$$

Thus, the spacings are uniformly unity and the spectrum will have spacing distribution peaked at $s=1$. The system does not show any fluctuation in eigenvalue spacings. When numerically performed with CML data, we obtain a pronounced peak at $s=1.0$.

B. Perturbed matrix

1. Truncated correlation matrix

In this case, we are not able to perform the unfolding analytically for a general case. However, for the special assumption made in Eq. (14), the unfolding procedure is carried out by substituting the eigenvalue λ in Eq. (17) from Eq. (16). Then, we obtain from simple manipulation

$$E_j = N(\lambda_j) = \frac{nu}{\pi} = j. \tag{24}$$

Thus, for the case of perturbed matrix that follow the assumption in Eq. (14), the spacing distribution is peaked at $s=1$. The numerical results indicate that this is true in general, irrespective of the nature of correlations decay. First, we consider the CML correlation matrix spectra for $a=1.98$ and $\epsilon=0.08$, where we have retained five principal off-diagonals and set the rest to zero since the magnitude of these elements was less than 10^{-5}. This truncation is done to enforce "ideal" conditions, that closely approximate the assumptions made to obtain theoretical results. For the empirical correlation matrix from CML, we unfold its spectrum and determine the spacing distribution numerically.

As shown in Fig. 6, the spectrum in this case displays a strong peak at $s=1$ quite in agreement with the analysis done above. We also note that the effect of finite-time estimation of correlations manifests itself in the form of spread of spacing distribution around $s=1$. Thus, this spacing distribution

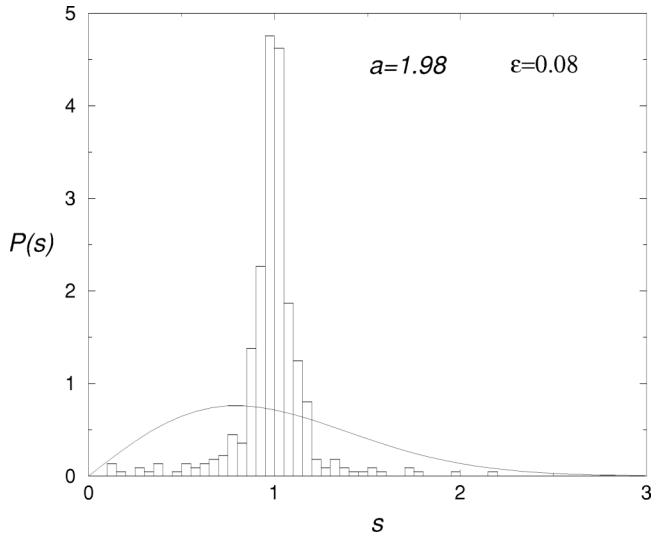

FIG. 6. Spacing distribution for the truncated correlation matrix spectra for CML. The solid curve is the Gaussian orthogonal ensemble (GOE) spacing distribution.

could be thought of as the counterpart of "integrable" limit for correlation matrices. This is also characteristic of the spectra of one dimensional quantum systems that are extremely rigid. This is known to occur in a many body system of bosons described by *k*-body interactions and is understood in the framework of embedded Gaussian ensembles [16].

C. Full correlation matrix

The CML correlation matrix estimated from finite-time dynamics can also be thought of as a case of translational invariance being broken. Thus, the spacing distribution of full correlation matrix could behave differently from the ideal cases we consider to obtain analytical results. We show from numerical spacing distributions for full correlation matrix from CML that the effect of finite-time correlation estimate still preserves the pronounced peak at $s=1$ but often leads to strong spread in spacings around $s=1$. This is shown in Fig. 8 for CML with $a=1.98$ and $\epsilon=0.08$. In Fig. 7 we show the spacing distribution for a particular choice of CML parameter values which, as pointed out earlier, is not amenable to any analytical treatment in the present framework. The spread of spacings in the vicinity of $s=1$ can be attributed to finite-time estimate of correlations. Our numerical experiments with CML indicate that the magnitude of fluctuations due to finite-time estimate of correlation coefficients may not be related in a straightforward way to the CML parameters. Thus, the extent of spread around $s=1$ in Figs 7 and 8 arises from different magnitude of fluctuations in these two cases.

VI. TRANSITION TO GOE

The analytical results in the previous sections indicate that the exact translational invariance and hence the symmetric banded Toeplitz structure for the correlation matrix leads to a spacing distribution with a peak at $s=1$. Numerical results

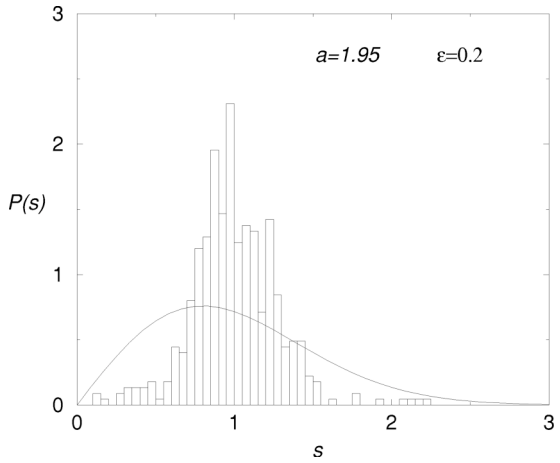

FIG. 7. Spacing distribution of the full correlation matrix spectra for CML. The solid curve is the GOE spacing distribution.

obtained from full correlation matrices show that the deviations from the expected spacing distribution are caused by the statistical fluctuations of the matrix elements due to the finite-time estimate of correlations which effectively break the translational symmetry. Such deviations are especially striking in Figs. 7 and 8. As anticipated by Eq. (24) for geometric decay in correlations, the spacing distribution peaks in the close vicinity of $s=1$. But the numerical distribution is rather broad. Thus, in practice, the results indicate a tendency to move towards a Gaussian orthogonal ensemble-like distribution under certain conditions, which we explore below.

We look at this transition within the scope of the present study, namely, the diagonally dominant correlation matrix and the effects arising from inevitable finite-time estimate of correlations. Under these conditions, we show below that transition to predominantly GOE like statistics takes place depending upon the extent to which the translational symmetry in the correlation matrix is broken. For this purpose, we construct a correlation matrix from multivariate Gaussian distributed random numbers with a zero mean and a given covariance $\boldsymbol{\sigma}$, $\boldsymbol{G}(\mathbf{0},\boldsymbol{\sigma})$. To ensure that the correlation matrix

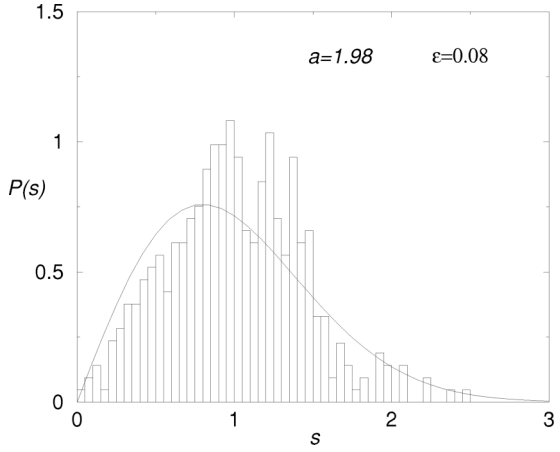

FIG. 8. Spacing distribution of the full correlation matrix spectra for CML. The solid curve is the GOE spacing distribution.

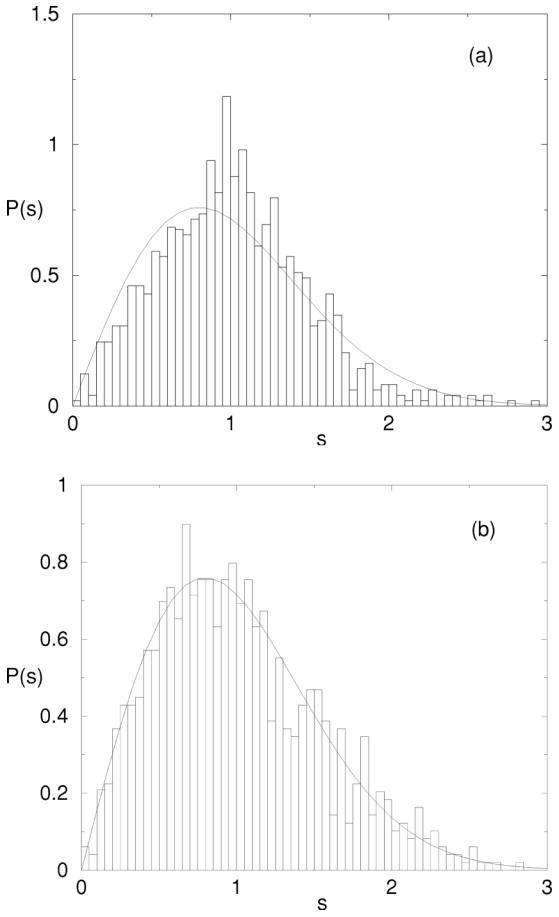

FIG. 9. Spacing distribution of the full correlation matrix spectra from multivariate random variables sampled (a) 6×10^5 times and (b) 4000 times. See text for details. The solid curve is the GOE spacing distribution.

remains effectively banded, we take $\sigma_{ij}=\exp(-r|i-j|)$, where r is used to control the extent of the correlation tails. In this experiment, we take $r=0.2$ and generate random numbers from $n=1000$ random variables sampled up to $p=6\times10^5$ times, leading to correlation matrix of order 1000. Figure 9 shows the spacing distribution of the random correlation matrix obtained by drawing (a) 300000 samples and (b) 4000 samples. If large number of samples ($p\gg n$) are drawn from $\boldsymbol{G}(\mathbf{0},\boldsymbol{\sigma})$, then the estimated correlations closely approximate correct correlation values. Thus, the correlation matrix tends to satisfy the translational invariance more closely. If we view the finite sampling errors as noise, then we might say that the amount of noise is less. Thus, in Fig. 9(a) we see that the spacing distribution is still peaked in the vicinity of $s=1$. We also note that it deviates from GOE curve too. On the other hand, if the correlation matrix is obtained from fewer samples ($p<n$ or $p\sim n$), then the estimated covariances could be noisy. For such cases, as the Fig. 9(b) shows, the spacing distribution closely approximates the GOE distribution from RMT. This is confirmed quantitatively by a Kolmogorov-Smirnov (KS) test [17] for goodness of fit with cumulative GOE distribution. At 15% significance level, a KS test could not reject the hypothesis that the empirical distribution in Fig. 9(b) is drawn from GOE. In the case of empirical distribution in Fig. 9(a), this hypothesis is rejected.

The transition to GOE-like spacing distribution can be understood along the following lines. In general, any correlation matrix arising from a practical multivariate time-series data, set, such as in stock market fluctuation, atmospheric anomalies, medical data, etc., contain signal and a noise component. If the signal components are removed from such data sets by subtracting out the dominant principal components, then the data set would contain essentially noise components. The spectra of such purely noise components would display qualitative behavior similar to the stochastic processes and hence would have spacing distribution with a pronounced peak at $s=1$. As nontrivial, system-specific correlations, that go to make up the signal component, are added to the "noisy" system, the eigenvalues change their position and the spectrum makes a transition towards GOE-like distribution. This argument, however, does not contradict earlier works where noisy components of large correlations display GOE statistics [2]. Thus, the transition to GOE-like spectral statistics arises from breaking the Toeplitz structure of the correlation matrix. This has certain practical implications where the length of the available time series is limited. In time series data from physical systems that have at least approximate translational symmetry, the finite-time estimate of correlations leads to Toeplitz structure being broken. We notice from Figs. 5(b) and 8 that the spacing distribution is more sensitive to matrix perturbations than the eigenvalue density. This, then, could lead to incorrect conclusions about the nature of spacing distribution in practical cases. Hence, it is important to ascertain if translational symmetry is present in empirical time series to avoid such pitfalls or alternatively to obtain very reliable correlation estimates.

VII. DISCUSSIONS AND CONCLUSIONS

We have studied the statistical properties of the correlation matrix spectra and focussed on the limiting cases that cannot be described by the standard Wigner-Dyson ensembles of random matrix theory. We study correlation matrices with entries rapidly decaying away from the diagonal. From the spectral point of view, this limit corresponds to that of equally spaced eigenvalues. This is in contrast with the standard RMT spectral signature of level repulsion. We assume simple models for correlations in a multivariate setting and obtain results for the eigenvalue density and nearest-neighbor spacing distribution. For both these quantities, the qualitative behaviour is different from predictions based on RMT. These results hold exactly for stationary and translationally invariant stochastic processes whose correlation matrices have Toeplitz form. All along we compare the theoretical results with the full correlation matrix spectra constructed from CML dynamics in its spatio temporal chaotic regime. The agreement is fairly good though corrupted by effects of finite-time estimates. In fact, these effects are more pronounced in the case of spacing distribution than the eigenvalue density. On the other hand, these anomalies also provide the connection with the empirical correlation matrices arising from time-series data. We find that GOE-like spectral statistics arises from breaking the Toeplitz structure

of correlation matrix. We also point out certain pitfalls in determining the spacing distribution for correlation matrix spectra from empirical time series. This work also sheds light on the spectral statistics of space-time chaos. Quite ironically, the regime of spatiotemporal chaos in multivariate data is associated with nonRMT behavior.

This work does not provide an RMT based explanation nor do the results represent an average over some appropriate ensemble. Essentially we study systematic deviations from RMT and the treatment is outside the framework of RMT. At this point, an analogy with the quantum systems could be called for. The diagonally dominant Hermitian matrices as operators in quantum mechanics, under certain conditions, generically show Poisson level spacing distribution. This result has no counterpart in RMT but has been derived based on semiclassical analysis [4], which is akin to accounting for system specific features that remain outside the purview of RMT. This work represents a similar effort in the context of spectral statistics of correlation matrix. In recent years, an ensemble of power-law banded random matrices (BRM) [18] and their variants [19] such as the diagonally dominant ones have been considered in the context of metal-insulator transitions in condensed matter systems. This is a class of GOE matrices whose variance depends on the distance from the diagonal while the correlation matrices belong to a class of Laguerre ensemble. Another motivation for BRMs is to interpolate between GOE and poisson level statistics. However, the suitability of BRMs for describing the spectral properties of correlation matrices have not yet been explored. Such an approach may lead to interesting results within the RMT approach.

To extend the analogy with quantum systems, the one dimensional systems like the power law potentials of the form $V(x)=x^n$, where $n>0$ is an integer, and certain variant of tight binding models for crystal lattices without disorder display delta peaked spacing distribution. After unfolding, these systems display equispaced spectrum like a picket fence. In a sense, such a spectrum is not considered interesting because it is devoid of any fluctuations. This work shows that such a spectrum appears in the correlation matrix of translationally invariant stochastic systems, under certain conditions. Numerical results from CML show that such a spectral limit is also exhibited by realistic correlation matrices. As pointed out, certain many body bosonic systems with *k*—body interactions carry this spectral signature [16]. However, an interesting point is that the correlation matrix spectra is an instance where transition from picket fence spectrum to GOE-like level repulsion occurs without any associated change in the dimensions of the system. For example, to see any transition away from picket fence spectrum, say, in a 1D quantum system, we need to increase the system dimensionality in terms of space or time. On the other hand, the physically interesting TBMs feature spectral fluctuation transition between Poisson and GOE.

We believe we have obtained new results on the limiting behavior of the correlation matrix spectra vis-a-vis random matrix theory. The assumed symmetric Toeplitz structure is true strictly for a class of translationally invariant stochastic systems. However, for real complex systems, an appropriate approach would be to use random perturbations to Toeplitz structure. However, at present, random perturbation methods for the Toeplitz operators does not provide perturbed eigenvalues in closed form yet. Such an approach might provide useful extension to the results obtained in this work. Another avenue not completely pursued in this work is the nature of transition to GOE statistics. In the case of banded Hermitian matrices [20], the transition from Poisson to GOE is known to be parameterized by $h=b^2/N$ where b is the bandwidth and N is the size of the matrix. It might be interesting to look for such scaling in transition statistics.

ACKNOWLEDGMENTS

One of us (M.S.S.) thanks Henning Schomerus for some useful discussions. M.S.S. also thanks William Trench and Kenneth Driessel for pointing out certain aspects of solutions for Toeplitz matrix eigenvalue problem.

[1] M. L. Mehta, *Random Matrices* (Academic, New York, (1991); T. Guhr, A. Muller-Groeling, and H. A. Weidenmuller, Phys. Rep. **299**, 189 (1999).
[2] L. Laloux , Phys. Rev. Lett. **83**, 1467 (1999); V. Plerou , *ibid.* **83**, 1471 (1999); M. S. Santhanam and Prabir K. Patra, Phys. Rev. E **64**, 016102 (2001); M. Barthēlemy , *ibid.* **66**, 056110 (2002).
[3] P. Gopikrishnan , Phys. Rev. E **64**, 035106(R) (2001).
[4] M. V. Berry and M. Tabor, Proc. R. Soc. London, Ser. A **356**, 375 (1977).
[5] *Theory and Applications of Coupled Map Lattices*, edited by K. Kaneko (Wiley, New York, 1993).
[6] R. Kapral, in *Theory and Applications of Coupled Map Lattices* (Ref. [5]).
[7] K. Kaneko, Physica D **34**, 1 (1989).
[8] J. L. Doob, *Stochastic Processes* (Wiley, New York, 1990).
[9] N. W. Ashcroft and D. N. Mermin, *Solid State Physics* (Harcourt Brace College Publishing, Fort Worth, 1976); see also p. 1156 of Ref. [11].
[10] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications* (Chelsea Publising Company, New York 1984); W. Trench and K. Driessel (private communication).
[11] C. Cohen-Tannoudji, B. Diu, and F. Laloe, *Quantum Mechanics* (Wiley, New York, 1997).
[12] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1994).
[13] V. Plerou , Phys. Rev. E **65**, 066126 (2002).
[14] A. M. Sengupta and P. P. Mitra, Phys. Rev. E **60**, 3389 (1999).
[15] Matthias Brack and R. K. Bhaduri, *Semiclassical Physics* (Addison-Wesley, Reading, MA, 1997).
[16] T. Asaga , Europhys. Lett. **56**, 340 (2001).
[17] For a brief explanation and practical implementation in Fortran, see William Press , *Numerical Recipes in Fortran: The Art of Scientific Computing* (Cambridge University Press,

Cambridge, 1992).
[18] A. D. Mirlin , Phys. Rev. E **54**, 3221 (1996); E. Cuevas, V. Gasparian and M. Ortuno, Phys. Rev. Lett. **87**, 056601 (2001); I. Varga and D. Braun, Phys. Rev. B **61**, R11859 (2000).
[19] O. Yevtushenko and V. E. Kravtsov, J. Phys. A **36**, 8265 (2003); e-print cond-mat/0309548.
[20] G. Casati, L. Molinari, and F. Izrailev, Phys. Rev. Lett. **64**, 1851 (1990).