Exit-Time Approach to ϵ-Entropy

M. Abel,[1] L. Biferale,[2] M. Cencini,[1] M. Falcioni,[1] D. Vergni,[1] and A. Vulpiani[1]

[1]*Department of Physics and INFM, University of Rome "La Sapienza," Piazzale Aldo Moro 2, I-00185 Roma, Italy*

[2]*Department of Physics and INFM, University of Rome "Tor Vergata," via della Ricerca Scientifica 1, I-00133 Roma, Italy*

(Received 17 November 1999)

An efficient approach to the calculation of the ϵ-entropy is proposed. The method is based on the idea of looking at the information content of a string of data, by analyzing the signal only at the instants when the fluctuations are larger than a certain threshold ϵ, i.e., by looking at the exit-time statistics. The practical and theoretical advantages of our method with respect to the usual one are shown by the examples of a deterministic map and a self-affine stochastic process.

PACS numbers: 05.45.Tp, 89.70.+c

The problem of quantifying the degree of complexity of an evolving system is ubiquitous in natural science (for a nice review, see Ref. [1]). The typical questions may range from the aim to distinguish between stochastic or chaotic systems to the more pragmatic goal of determining the degree of complexity (read predictability) at varying the resolution in phase-space and time [2,3].

From the pioneering works of Shannon and Weaver [4] and Kolmogorov [5], a proper mathematical tool, the Kolmogorov-Sinai (KS) entropy, h_{KS}, has been developed to address the above question quantitatively and unambiguously (in principle). The main idea is very natural: we must look at the information contained in the time evolution as a probe of the underlying dynamics. This is realized by studying the symbolic dynamics obtained by assigning different symbols to different cells of a finite partition of the phase-space. The probability distribution of allowed sequences (words) is determined by the dynamical evolution. The average information gain is defined by comparing sequences of length m and $m + 1$, in the limit of large m. By allowing the length of the words, m, to infinity and going to infinitely refined partition, one obtains the KS entropy, which is a measure of the degree of complexity of the trajectory. The KS entropy also determines the rate of information transmission necessary to unambiguously reconstruct the signal.

Unfortunately, only in simple, low-dimensional, dynamical systems can such a procedure be properly carried out with conventional methods [2,3,6]. The reason is that for high-dimensional systems the computational resources are not sufficient to cope with the very high resolution and extremely long time series required. Moreover, in many systems, such as in turbulence, the existence of nontrivial fluctuations on different time and spatial scales cannot be captured by the KS entropy. This calls for a more general tool to quantify the degree of predictability which depends on the analyzed range of scales and frequencies. This was the aim that led Shannon and Weaver [4] and Kolmogorov [7] to introduce the so-called ϵ-entropy, later generalized to the (ϵ, τ)-entropy [2,6]. Conceptually it corresponds to the rate of information transmission necessary to reconstruct a signal with a finite accuracy ϵ, and with a sampling time interval τ. The naive (ϵ, τ) computation is usually performed by looking at the Shannon entropy of the coarse grained dynamics on an (ϵ, τ)-grid in the phase-space and time. This method suffers from so many computational drawbacks that it is almost useless for many realistic time series [3]. Different attempts in this direction might be the study of the finite size Lyapunov exponent [8] or using more refined techniques to extract spatial-time fluctuations based on wavelets analysis [9].

The aim of this Letter is to introduce an alternative approach for the determination of the (ϵ, τ)-entropy, based on the analysis of exit times. In a few words, the idea consists of looking at the information content of a string of data, without analyzing the signal at any fixed time, τ, but only when the fluctuations are larger than some fixed threshold, ϵ. This simple observation allows for a remarkable improvement of the computational possibility to measure the (ϵ, τ)-entropy, as will be discussed in detail later.

We believe that the approach presented here is unavoidable in all of those cases when either the high dimensionality of the underlying phase-space or the necessity to disentangle nontrivial correlations at different analyzed time scales leads to the failure of standard methods.

Let us just briefly recall the conventional way to calculate the (ϵ, τ)-entropy for the case of a time-continuous signal $x(t) \in \mathbb{R}$, recorded during a (long) time interval T. One defines an ϵ-grid on the phase-space and a τ-grid on time. If the motion is bounded, the trajectory visits only a finite number of cells; therefore to each subsequent length $n\tau$ from $x(t)$, one can associate a word of length n, out of a finite alphabet: $W_t^n(\epsilon, \tau) = (S_t, S_{t+\tau}, \ldots, S_{t+(n-1)\tau})$, where S_t labels the cell containing $x(t)$. From the probability distribution of the above words one calculates the block entropies $H_n(\epsilon, \tau)$:

$$H_n(\epsilon, \tau) = -\sum_{\{W^n(\epsilon,\tau)\}} P\big(W^n(\epsilon, \tau)\big) \ln P\big(W^n(\epsilon, \tau)\big)\,. \quad (1)$$

the (ϵ, τ)-entropy per unit time; $h(\epsilon, \tau)$ is finally defined as

$$h_n(\epsilon, \tau) = \frac{1}{\tau}\left[H_{n+1}(\epsilon, \tau) - H_n(\epsilon, \tau)\right], \quad (2)$$

 0031-9007/00/84(26)/6002(4)$15.00

$$h(\epsilon,\tau) = \lim_{n\to\infty} h_n(\epsilon,\tau) = \frac{1}{\tau}\lim_{n\to\infty}\frac{1}{n}H_n(\epsilon,\tau), \qquad (3)$$

where for practical reasons the dependence on the details of the partition is ignored, while the rigorous definition is given in terms of the infimum over all possible partitions with elements of diameter smaller than ϵ [2]. Notice that the above-defined (ϵ,τ)-entropy is nothing but the Shannon entropy of the sequence of symbols $(S_t, S_{t+\tau},\ldots)$ associated with the given signal. The Kolmogorov-Sinai entropy, h_{KS}, is obtained by taking the limit $(\epsilon,\tau)\to 0$:

$$h_{\mathrm{KS}} = \lim_{\tau\to 0}\lim_{\epsilon\to 0} h(\epsilon,\tau). \qquad (4)$$

In systems with both time and signal continuous variables, the entropy $h(\epsilon,\tau)$ is the same for any choice of τ. Thus one obtains an ϵ-entropy $h(\epsilon)$ per unit time. In particular, in a pure deterministic flow, one can put $h(\epsilon) = h(\epsilon,\tau=1)$ [10]. Also for discrete-time systems, one defines $h(\epsilon)\equiv h(\epsilon,\tau=1)$. When $h(\epsilon)$ exists, $h_{\mathrm{KS}} = \lim_{\epsilon\to 0} h(\epsilon)$.

Let us recall that, for a genuine deterministic chaotic system, one has $0 < h_{\mathrm{KS}} < \infty$ ($h_{\mathrm{KS}} = 0$ for a regular motion) while, for a continuous random process, $h_{\mathrm{KS}} = \infty$. Therefore, in order to distinguish between a purely deterministic system and a stochastic system it is necessary to perform the limit $\epsilon\to 0$ in (4). Obviously, from a physical and/or numerical point of view this is impossible. Nevertheless, by looking at the behavior of the (ϵ,τ)-entropy at varying ϵ, one can have some qualitative and quantitative insights into the chaotic or stochastic nature of the process. For most of the usual stochastic processes, one can explicitly give an estimate of the entropy scaling behavior when $\epsilon\to 0$ [2]. For instance, in the case of a stationary Gaussian process with spectrum $S(\omega)\propto\omega^{-2}$, one has [2] $h(\epsilon)\sim 1/\epsilon^2$ for small ϵ. Let us now introduce the main point of this Letter by discussing in detail the difficulties that may arise in measuring the ϵ-entropy for the following nontrivial example of a chaotic-diffusive map [11]:

$$x_{t+1} = x_t + p\sin 2\pi x_t. \qquad (5)$$

When $p > 0.7326\ldots$, this map produces a diffusive behavior on large scales, so one expects $h(\epsilon)\simeq\lambda$ for $\epsilon < 1$ and $h(\epsilon)\propto\frac{D}{\epsilon^2}$ for $\epsilon > 1$, where λ is the Lyapunov exponent and D is the diffusion coefficient, defined by $D = \lim_{t\to\infty}\langle(x_t - x_0)^2\rangle/2t$. As far as we know, it is difficult to find the dependence of D on p. The numerical computation of $h(\epsilon)$, using the standard codification, is already highly nontrivial in this simple system. This can be seen by looking at Fig. 1, where the diffusive behavior is roughly obtained only by considering the envelope of $h_n(\epsilon,\tau)$ evaluated for different values of τ; while looking at any single (small) value of τ (one would like to put $\tau = 1$), one obtains a rather inconclusive result. This is due to the fact that one has to consider very large block lengths n when computing $h(\epsilon,\tau)$, in order to obtain a good convergence for $H_n(\epsilon,\tau) - H_{n-1}(\epsilon,\tau)$ in (3). Indeed, in the diffusive regime, a simple dimensional argument shows that the characteristic time of the system is $T_\epsilon\approx\epsilon^2/D$. If we consider, for example, $\epsilon = 10$ and typical values of the diffusion coefficient, $D = O(10^{-1})$ (as numerically obtained for values of p near 1), T_ϵ comes out much larger than the elementary sampling time $\tau = 1$.

Our approach to calculating $h(\epsilon)$ differs from the usual procedure which constructs the coding sequence of the signal at a given level of accuracy. Specifically, we use a different way to sample the time, i.e., instead of using a constant time interval τ, we sample according to the exit time $t(\epsilon)$ on an alternating grid of cell size ϵ. We consider the original continuous-time record $x(t)$ and a reference starting time $t = t_0$; the subsequent exit time t_1 is then defined as the first time necessary to have an absolute variation equal to $\epsilon/2$ in $x(t)$, i.e., $|x(t_0 + t_1) - x(t_0)| \geq \epsilon/2$. This is the time the signal takes to exit the cell of size ϵ. We then restart from t_1 to look for the next exit time t_2, i.e., the first time such that $|x(t_0 + t_1 + t_2) - x(t_0 + t_1)| \geq \epsilon/2$, and so on. Notice that, with this definition, the coarse graining grid is not fixed, but it is always centered in the last exit position. In this way we obtain a sequence of exit times $\{t_i(\epsilon)\}$. To distinguish the direction of the exit (up or down out of a cell), we introduce the label $k_i = \pm 1$, depending on whether the signal is exiting above or below. By doing so, the trajectory is univocally coded with the required accuracy, by the sequence $\big((t_1,k_1),(t_2,k_2),\ldots,(t_M,k_M)\big)$, where M is the total number of exit-time events observed during the total time T. Correspondingly, an "exit-time word" of length n is a sequence of a couple of symbols $\Omega_i^n(\epsilon) = \big((t_i,k_i),(t_{i+1},k_{i+1}),\ldots,(t_{i+n-1},k_{i+n-1})\big)$. From these words, one first calculates the block entropies,

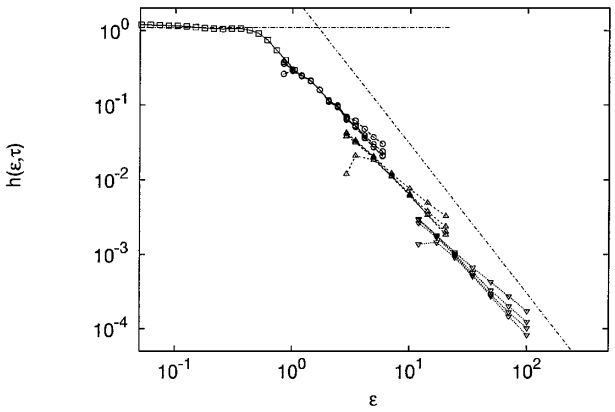

FIG. 1. Numerically evaluated (ϵ,τ)-entropy for the map (5) with $p = 0.8$ computed in the usual way [6] at $\tau = 1$ ($\circ$), $\tau = 10$ ($\triangle$), and $\tau = 100$ ($\triangledown$), and different block lengths [$n = 4$ (above), $n = 8$, 12, 20 (below)]. The boxes ($\square$) give the entropy computed with $\tau = 1$ by using periodic boundary condition over 40 cells. The latter is necessary in order to compute the Lyapunov exponent $\lambda = h_{\mathrm{KS}} = 1.15$. The straight lines correspond to the two asymptotic behaviors, $h(\epsilon) = h_{\mathrm{KS}}$ and $h(\epsilon)\sim\epsilon^{-2}$.

$H_n^\Omega(\epsilon)$, and then the exit-time ϵ-entropies: $h^\Omega(\epsilon) \equiv \lim_{n\to\infty} H_{n+1}^\Omega(\epsilon) - H_n^\Omega(\epsilon)$. Let us notice that $h^\Omega(\epsilon)$ is an ϵ-entropy *per exit* and that $M = T/\langle t(\epsilon)\rangle$. To find a relation between $h(\epsilon)$ and $h^\Omega(\epsilon)$ we can argue as follows: The exit-time coding is a faithful reconstruction, within the accuracy ϵ, of the original signal and there is a one-to-one correspondence between the (exit-time) histories and the (ϵ,τ) histories (for $\tau \to 0$) originating from a given ϵ-cell. The Shannon-McMillan theorem [12] ensures that the number of the typical (ϵ,τ) histories of length N, $\mathcal{N}(\epsilon,N)$, is such that $\ln\mathcal{N}(\epsilon,N) \simeq h(\epsilon)N\tau = h(\epsilon)T$. For the number of typical (exit-time) histories of length M, $\mathcal{M}(\epsilon,M)$, we have $\ln\mathcal{M}(\epsilon,M) \simeq h^\Omega(\epsilon)M$. If we consider $T = M\langle t(\epsilon)\rangle$ we must obtain the same number of (very long) histories. This implies that

$$h(\epsilon) = Mh^\Omega(\epsilon)/T = \frac{h^\Omega(\epsilon)}{\langle t(\epsilon)\rangle}. \tag{6}$$

Now we are left with the determination of $h^\Omega(\epsilon)$. This implies a discretization, τ_e, of the exit times. The exit-time entropy, $h^\Omega(\epsilon)$, becomes an exit-time (ϵ,τ_e)-entropy, $h^\Omega(\epsilon,\tau_e)$, obtained from the sequence $\{\eta_i,k_i\}$, where η_i identifies the exit-time cell containing t_i. Equation (6) now becomes

$$h(\epsilon) = \lim_{\tau_e\to 0} h^\Omega(\epsilon,\tau_e)/\langle t(\epsilon)\rangle \simeq h^\Omega(\epsilon,\tau_e)/\langle t(\epsilon)\rangle, \tag{7}$$

the latter relation being valid for small enough τ_e (in all practical situations there exists a minimum τ_e given by the highest acquisition frequency; i.e., the limit $\tau_e \to 0$ cannot be reached). At this point it is important to stress that in most cases the leading ϵ contribution to $h(\epsilon)$ in (7) is given by the mean exit-time $\langle t(\epsilon)\rangle$ and not by $h^\Omega(\epsilon,\tau_e)$. At any rate, the computation of $h^\Omega(\epsilon,\tau_e)$ is compulsory in order to recover a zero entropy for regular (e.g., periodic) signals. It is easy to obtain the following bounds for $h^\Omega(\epsilon,\tau_e) = h^\Omega(\{\eta_i,k_i\})$:

$$h^\Omega(\{k_i\}) \le h^\Omega(\{\eta_i,k_i\}),$$

$$h^\Omega(\{\eta_i,k_i\}) \le h^\Omega(\{\eta_i\}) + h^\Omega(\{k_i\}) \le h^\Omega(\{k_i\}) + H_1^\Omega(\{\eta_i\}),$$

where $h^\Omega(\{k_i\})$ is the Shannon entropy of the sequence $\{k_i\}$ and $H_1^\Omega(\{\eta_i\})$ is the one-symbol entropy of the $\{\eta_i\}$. Therefore we have

$$\frac{h^\Omega(\{k_i\})}{\langle t(\epsilon)\rangle} \le h(\epsilon) \le \frac{h^\Omega(\{k_i\}) + c(\epsilon) + \ln(\langle t(\epsilon)\rangle/\tau_e)}{\langle t(\epsilon)\rangle}, \tag{8}$$

where $c(\epsilon) = -\int p(z)\ln p(z)\,dz$, and $p(z)$ is the probability distribution function of the rescaled exit time $z(\epsilon) = t(\epsilon)/\langle t(\epsilon)\rangle$.

We shall see in the following that the above bounds are rather good, and typically $\langle t(\epsilon)\rangle$ shows the same scaling behavior as $h(\epsilon)$. One could wonder why the exit-time approach is better than the usual one. The reason is simple (and somehow deep): in the exit-time approach it is not necessary to use a very large block size since, at fixed ϵ, the typical time at that scale is automatically given by $\langle t(\epsilon)\rangle$. This fact is particularly clear in the case of Brownian motion. In such a case $\langle t(\epsilon)\rangle \propto \epsilon^2/D$, where D is the diffusion coefficient. As previously discussed, the computation of the $h(\epsilon)$ with the standard methods implies the use of very large block sizes, of order ϵ^2/D.

With our method the $\langle t(\epsilon)\rangle$ captures the correct scaling behavior and the exit-time entropy introduces, at worst, a subleading logarithmic contribution $h^\Omega(\epsilon,\tau_e) \sim \ln(\langle t(\epsilon)\rangle/\tau_e)$. This is because $c(\epsilon)$ is $O(1)$ and independent of ϵ for a self-affine signal, and the $h^\Omega(\{k_i\}) \le \ln 2$ term is small compared with $\ln(\langle t(\epsilon)\rangle/\tau_e)$ (for not too small ϵ), so that, neglecting the logarithmic corrections, $h(\epsilon) \simeq 1/\langle t(\epsilon)\rangle \propto D\epsilon^{-2}$. In order to clarify this point, we plot in Fig. 2 the calculation of the (ϵ,τ)-entropy *via* the exit-time approach for the previously discussed diffusive map. Figure 2 must be compared with Fig. 1, where the usual approach has been used. While in Fig. 1, the expected ϵ-entropy scaling is roughly recovered as an envelope over many different τ; in our case the predicted behavior is easily recovered for any ϵ, with a remarkable improvement in the quality of the result.

Let us now briefly comment on the limit $\epsilon \to 0$ for a discrete-time system (e.g., maps). In this limit the exit-time approach coincides with the usual one: we just have to observe that the exit times practically always coincide with the minimum sampling time and we have to consider the possibility of having jumps over more than one cell, i.e., the k_i symbols may take values $\pm 1, \pm 2, \ldots$.

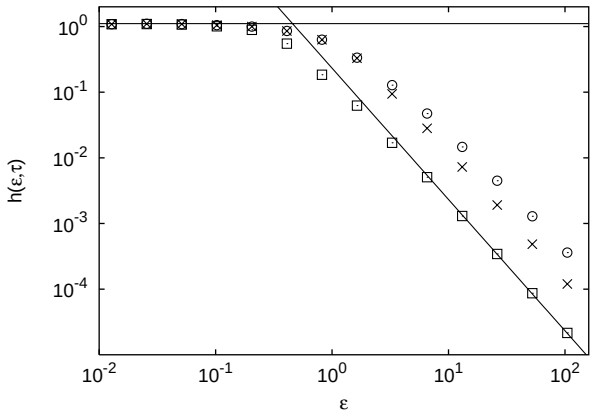

FIG. 2. Numerically computed lower bound (□) and upper bound (with $\tau_e = 1$) (∘) of $h(\epsilon)$ (8) for the same map of Fig. 1. The two straight (solid) lines correspond to the asymptotic behaviors as in Fig. 1. We also present the (ϵ,τ_e)-entropy $h^\Omega(\epsilon,\tau_e)/\langle t(\epsilon)\rangle$ with $\tau_e = 0.1\langle t(\epsilon)\rangle$ (×).

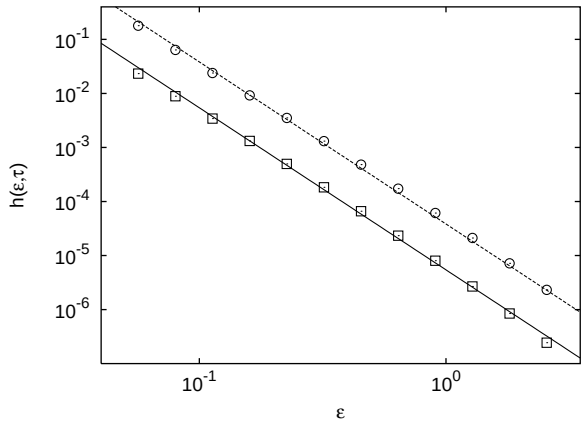

FIG. 3. Numerically computed lower bound (□) and upper bound [with $\tau_e = 0.1\langle t(\epsilon)\rangle$] (○) for the (ϵ, τ_e)-entropy in the case of a self-affine signal with $\xi = 1/3$ evaluated by using the exit-time approach. The two straight lines show the scaling ϵ^{-3}.

As another example, we present the calculation of ϵ-entropy for a self-affine stochastic signal with Hölder exponent $\xi = 1/3$, i.e., $|x(t) - x(t + \Delta t)| \sim (\Delta t)^{1/3}$. Such a signal can be seen as a stochastic surrogate of a turbulent signal (ignoring intermittency) and can be constructed in different ways (see Ref. [13], and references therein). A simple dimensional estimate, which is rigorous for Gaussian processes [7], tells us that the leading contribution to the ϵ-entropy scaling is given by $h(\epsilon) \sim \epsilon^{-3}$. To generate the self-affine signal we use a recently proposed algorithm [13], where $x(t)$ is obtained by using many Langevin processes. In Fig. 3 we show the bounds (8) for (ϵ, τ_e)-entropy calculated *via* the exit-time approach. We observe an extended region of well-defined scaling, which is the same for $1/\langle t(\epsilon)\rangle \sim \epsilon^{-3}$. The usual approach (not shown) gives a poor estimate for the scaling as the envelope of $h(\epsilon, \tau)$ computed for various τ (see, for example, Figs. 15–18 in [2]), as in the case of Fig. 1.

In conclusion, we have introduced an efficient method to calculate the ϵ-entropy based on the analysis of the exit-time statistics. It is able to disentangle in a more proper way the leading contributions to $h(\epsilon)$ at the scale ϵ, compared with the standard way which is based on a coarse grained dynamics on a fixed (ϵ, τ)-grid. We have presented the application to two examples, a chaotic diffusive map and a stochastic self-affine signal. More applications to chaotic systems, stochastic multiaffine processes, and experimental turbulent signals will be reported elsewhere. The entropy approach used to evaluate the degree of complexity of a time series also allows one to attack much more sophisticated problems often encountered in dynamical system theory. We just mention, e.g., the problem to disentangle correlations between time series obtained by measuring different observables of the same system. This might be addressed by using the conditional entropy. Another interesting problem is the determination of the Renyi entropies [1]—reproposed by Tsallis in a statistical mechanics context [14]—which characterizes the nontypical sequences for finite times.

We acknowledge useful discussions with G. Boffetta and A. Celani. This work has been partially supported by INFM (PRA-TURBO) and by the European Network *Intermittency in Turbulent Systems* (Contract No. FMRX-CT98-0175). M. A. is supported by the European Network *Intermittency in Turbulent Systems.*

[1] R. Badii and A. Politi, *Complexity. Hierarchical Structures and Scaling in Physics* (Cambridge University Press, Cambridge, England, 1997).
[2] P. Gaspard and X.-J. Wang, Phys. Rep. **235**, 291 (1993).
[3] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, England, 1997).
[4] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, Illinois, 1949).
[5] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **119**, 861 (1958).
[6] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983); A. Cohen and I. Procaccia, Phys. Rev. A **31**, 1872 (1985).
[7] A. N. Kolmogorov, IRE Trans. Inf. Theory **1**, 102 (1956); I. M. Gelfand, A. N. Kolmogorov and A. M. Yaglom, in *Selected Works of A.N. Kolmogorov,* edited by A. N. Shiryeyev (Kluwer Academic, Dordrecht, 1993), Vol. III, p. 33.
[8] E. Aurell, G. Boffetta, A. Crisanti, G. Paladin, and A. Vulpiani, Phys. Rev. Lett. **77**, 1262 (1996); J. Phys. A **30**, 1 (1997).
[9] M. Farge, Annu. Rev. Fluid Mech. **72**, 395 (1992).
[10] J. P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985); P. Billingsley, *Ergodic Theory and Information* (Wiley, New York, 1965).
[11] M. Schell, S. Fraser, and R. Kapral, Phys. Rev. A **26**, 504 (1982).
[12] A. I. Khinchin, *Mathematical Foundations of Information Theory* (Dover, New York, 1957).
[13] L. Biferale, G. Boffetta, A. Celani, A. Crisanti, and A. Vulpiani, Phys. Rev. E **57**, R6261 (1998).
[14] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).