Information content of signals using correlation function expansions of the entropy

Phil Attard
School of Chemistry F11, University of Sydney, Sydney, New South Wales 2006, Australia

Owen G. Jepps and Stjepan Marčelja
Department of Applied Mathematics, Research School of Physical Sciences and Engineering, Australian National University, Canberra, Australian Capital Territory 0200, Australia

(Received 16 June 1997)

Formally exact series expressions are derived for the entropy (information content) of a time series or signal by making systematic expansions for the higher-order correlation functions using generalized Kirkwood and Markov superpositions. Termination of the series after two or three terms provides tractable and accurate approximations for calculating the entropy. Signals generated by a Gaussian random process are simulated using Lorentzian and Gaussian spectral densities (exponential and Gaussian covariance functions) and the entropy is calculated as a function of the correlation length. The validity of the truncated Kirkwood expansion is restricted to weakly correlated signals, whereas the truncated Markov expansion is uniformly accurate; the leading two terms yield the entropy exactly in the limits of both weak and strong correlations. The concept of entropy for a continuous signal is explored in detail and it is shown that it depends upon the level of digitization and the frequency of sampling. The limiting forms are analyzed for a continuous signal with exponentially decaying covariance, for which explicit results can be obtained. Explicit results are also obtained for the binary discrete case that is isomorphic to the Ising spin lattice model. [S1063-651X(97)09210-6]

PACS number(s): 05.50.+q, 89.70.+c, 02.50.Ga, 02.50.Cw

INTRODUCTION

A quantitative analysis of the communication or storage of data necessarily involves measuring the amount of information involved. Compression, which in general replaces frequently occurring data strings by shorter code strings, ideally entails no loss of information and hence the efficiency and limitation of the process may be gauged by measuring the latter. It was Shannon [1] who formulated the mathematical theory of communication and founded it upon a particular measure of information that is closely connected with the thermodynamic quantity entropy; both measure order and predictability. Shannon's formula already occurred in Gibbs's treatment of statistical mechanics and even earlier the particular example given by Boltzmann had provided the microscopic basis for thermodynamics. The equivalence of information and entropy is epitomized in Jaynes's maximum-entropy formulation of statistics and statistical mechanics [2].

We formally consider a signal to be a sequence of N symbols that come in L types. For a continuous signal this is the result of sampling at N nodes and digitizing into L levels. Hence there are L^N possible distinct signals and we suppose that the probability of each is $\wp_i$. That is, we imagine the signal to be drawn from an ensemble of similar signals with certain common characteristics, each signal in the ensemble being replicated according to the distribution $\wp_i$. Shannon defined the entropy of a signal as $S=-\Sigma_i \wp_i \ln \wp_i$ and the entropy per node as $s=S/N$.

As in statistical mechanics, this formal definition is of limited practical use since for large N the number of possible distinct signals is astronomical and collecting statistics for each one rapidly becomes prohibitive. As an alternative, Shannon provided a prescriptive formula that uses the probabilities of the subsequences of symbols that appear in a given signal. That is, from a frequency histogram of the signal, one calculates $\wp_i^{(n)}$ for each of the L^n possible subsequences of n consecutive symbols. This is the probability that an arbitrarily chosen run of n symbols of the signal will be the ith subsequence. The signal has entropy per node $s=-\lim_{n\to\infty} n^{-1}\Sigma_i \wp_i^{(n)} \ln \wp_i^{(n)}$.

Whether or not it is actually feasible to use this recipe to calculate the information content of a particular signal depends upon the degree to which the signal is correlated. For almost random signals, in which successive symbols are almost uncorrelated, the formula rapidly converges. But for highly correlated signals, in which successive symbols are predetermined and can be accurately predicted, large cluster sizes are required before the entropy expression converges. In this case the number of possible clusters can actually exceed the length of the signal and the statistics are insufficient to evaluate the entropy.

A meaningful signal is intermediate between a completely regular signal, which has zero entropy, and a purely random one, $\wp_i=1/L^N$, in which case the entropy is given by Boltzmann's expression $S=N\ln L$. Information reflects correlations between successive symbols of a signal, which in turn are a manifestation of the redundancy of the language; in general, for meaningful natural communication $S\leqslant N\ln L$. It is these correlations that make compression and decryption feasible. A signal encoded for brevity removes this redundancy and conveys the same amount of information with fewer symbols: The optimally abbreviated signal would be of length $\mathcal{N}=S/\ln L\leqslant N$ (i.e., for the number of encoded symbols actually sent, it would appear to have Boltzmann's maximum entropy).

The problem with estimating the entropy from the prob-

ability of the occurrence of subsequences is that it treats each subsequence as independent. It is not until the length of the subsequences exceeds the correlation length of the signal that this becomes a good approximation. As discussed above, this limits its usefulness to weakly correlated signals, which unfortunately are those that have the least potential for compression. What is desirable is a way of calculating the information entropy that includes successively higher-order correlations in a systematic fashion.

In this paper we explore formally exact expansions for the information entropy in terms of correlation functions. Similar expansions are well known in liquid-state statistical mechanics, and here we show that they can be developed for information theory applications. We show that in some cases the expansions are rapidly converging and that it is feasible to calculate the first few terms. We test the expansions for Gaussian random signals and show that the particular Markov expansion that we develop works well for both uncorrelated and highly correlated signals.

In setting out the paper we have deferred the derivation of the entropy expansions until after a discussion of the generation of a correlated random signal, which immediately follows. In Sec. I we also discuss the two-state leveling that was used and the simulation method. In Sec. II we derive expansions for the entropy from a Kirkwood and from a Markov superposition formula. We discuss these expansions in the limits of extreme correlations and we test them against the simulation results for exponential and for Gaussian covariance functions. Section III is concerned with the entropy of a continuous signal and its dependence upon the sampling and the digitization. A uniformly sampled signal with exponential covariance is shown to have Markov continuum correlations, and an exact analytic result for the entropy is obtained and analyzed in the infinite sampling and digitization limits. A similar analysis is carried out for a binary model in which the digital correlations are Markovian, which turns out to be just the Ising spin lattice model of statistical mechanics. We conclude with a summary of the main results and a discussion of the prospects for generalizing the Markovian approximation to more than one dimension.

I. CORRELATED RANDOM SIGNAL

We have in mind a general treatment of communication and data storage, but to be specific we shall formulate the problem as if it were a one-dimensional signal or time series. Moreover, we imagine that the signal is already sampled so that it comprises N data. We shall also digitize the signal into L levels; the total number of possible distinct signals is then L^N. We shall speak of the probability of the occurrence of a signal or of a sequence of data. By this we mean two things. First, we can imagine that the signals are taken from a large collection of signals that share similar characteristics, and the probability of a particular signal refers to the frequency with which that signal occurs in the collection, and similarly for a particular data sequence. Second, we can imagine that the signal is very long and that we can measure the frequency of the occurrence of particular data sequences along its length, assuming that the signal is homogeneous in a statistical sense. These two interpretations of the probability of data sequences essentially correspond to the ergodic hypothesis of statistical mechanics, namely, that ensemble averages and time averages are equivalent. In the present context this equivalence implies that in a statistical sense the signal is stationary or homogeneous in time.

A. Gaussian random signal

Suppose we sample a continuous signal at points $r_1, r_2, \ldots, r_N$ and that each measurement yields $s_i \in (-\infty, \infty)$. We focus on the probability density, which is proportional to the probability that at r_1 the signal is between s_1 and $s_1 + ds_1$, at r_2 the signal is between s_2 and $s_2 + ds_2$, $\ldots$, and at r_N the signal is between s_N and $s_N + ds_N$. The signal represents a Gaussian or normal random process if the probability density is of the form

$$\omega^{(N)}(s^N; r^N) = (2\pi)^{-N/2} |K_N|^{-1/2} \exp\left[\frac{-1}{2} \mathbf{s}^{NT} K_N^{-1} \mathbf{s}^N\right], \tag{1.1}$$

where $\mathbf{s}^{NT} = (s_1, s_2, \ldots, s_N)$ is the transpose of the column vector $\mathbf{s}^N$ and K_N is the N-dimensional covariance matrix. As usual we define reduced probability densities $\omega^{(n)}(s^n; r^n)$, which are obtained by integrating out the remaining $N-n$ samples. This corresponds to projecting the covariance matrix onto an n-dimensional subspace and hence $\omega^{(n)}$ is also Gaussian with covariance matrix K_n. The elements of the covariance matrix are in general

$$\{K\}_{ij} \equiv \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle = \int_{-\infty}^{\infty} ds_i \int_{-\infty}^{\infty} ds_j (s_i - \langle s_i \rangle)(s_j - \langle s_j \rangle) \omega^{(2)}(s_i, s_j; r_i, r_j) \tag{1.2}$$

and they take the form $\{K\}_{ij} = K(r_i, r_j) = K(r_{ij})$, $r_{ij} = |r_i - r_j|$, for a stationary process. For the Gaussian process the signal is symmetric about its mean, which without loss of generality we take to be zero,

$$\langle s_i \rangle = \int_{-\infty}^{\infty} ds_i s_i \omega^{(1)}(s_i; r_i) = 0, \tag{1.3}$$

and following convention we scale the signal such that

$$\{K\}_{ii} = 1. \tag{1.4}$$

One way to generate a normal distribution is from the superposition of random waves

$$s(r) = \sqrt{\frac{2}{M}} \sum_{n=1}^{M} \cos(k_n r - \phi_n). \tag{1.5}$$

Here each phase ϕ_n is randomly selected from a uniform distribution on $[0, 2\pi]$ and each wave vector k_n is randomly selected from a specified spectral distribution $f(k)$, which we assume to be normalized. Hence the average of a function that depends upon the $s(r_i)$ is

$$\langle a\rangle = \frac{1}{(2\pi)^M}\int_0^{2\pi} d\phi_1\cdots d\phi_M \times \int_{-\infty}^{\infty} dk_1\cdots dk_M f(k_1)\cdots f(k_M)a. \quad (1.6)$$

In particular,

$$\langle s(r)\rangle = 0, \quad (1.7)$$

$$\langle s(r)^2\rangle = \frac{1}{(2\pi)^M}\frac{2}{M}\sum_{n=1}^{M}\sum_{m=1}^{M}\int_0^{2\pi} d\phi_1\cdots d\phi_M \times \int_{-\infty}^{\infty} dk_1\cdots dk_M f(k_1)\cdots f(k_M) \times\cos(k_n r-\phi_n)\cos(k_m r-\phi_m) = \frac{1}{\pi}\int_{-\infty}^{\infty} dk\ f(k)\int_0^{2\pi} d\phi\ \cos^2(kr-\phi) = 1, \quad (1.8)$$

and

$$K(r_1,r_2)=\langle s(r_1)s(r_2)\rangle = \frac{1}{(2\pi)^M}\frac{2}{M}\sum_{n=1}^{M}\sum_{m=1}^{M} \times\int_0^{2\pi} d\phi_1\cdots d\phi_M \times\int_{-\infty}^{\infty} dk_1\cdots dk_M f(k_1)\cdots f(k_M) \times\cos(k_n r_1-\phi_n)\cos(k_m r_2-\phi_m) = \int_{-\infty}^{\infty} dk\ f(k)\cos k(r_1-r_2). \quad (1.9)$$

We see that the covariance or height-height correlation function is just 2π times the inverse Fourier transform of the spectral distribution of the waves. Finally, the signal generated by this superposition of random waves is indeed Gaussian,

$$\omega_{(M)}^{(1)}(t;r) = \langle \delta(t-s(r))\rangle_M = \int_{-\infty}^{\infty} dt'\,\omega_{(M-1)}^{(1)}(t';r)\,\omega_{(1)}^{(1)}(t-t';r), \quad (1.10)$$

where we have exhibited the dependence on the number of waves explicitly. Taking the Fourier transform of the convolution integral, we obtain

$$\hat{\omega}_{(M)}^{(1)}(q;r) = \hat{\omega}_{(M-1)}^{(1)}(q;r)\,\hat{\omega}_{(1)}^{(1)}(q;r) = \hat{\omega}_{(1)}^{(1)}(q;r)^M. \quad (1.11)$$

Now

$$\omega_{(1)}^{(1)}(t;r) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ f(k)\int_0^{2\pi} d\phi\ \delta(t-\sqrt{2/M}\cos(kr-\phi)) = \frac{1/\pi}{\sqrt{2/M-t^2}}, \quad |t|<\sqrt{2/M} \quad (1.12)$$

and hence

$$\hat{\omega}_{(M)}^{(1)}(q;r) = [J_0(q\sqrt{2/M})]^M \sim\left(1-\frac{2q^2}{4M}\right)^M \sim e^{-q^2/2},\ \ M\to\infty. \quad (1.13)$$

Inverting this we have

$$\omega^{(1)}(s,r) = \frac{1}{\sqrt{2\pi}}e^{-s^2/2}, \quad (1.14)$$

which is the desired normal form. This argument can be generalized to show that $\omega^{(N)}$ itself is Gaussian.

B. Binary digitization

The continuous Gaussian random signal is digitized by assigning the measured value at a node to a cell indexed by an integer. That is, the probability densities are converted to probabilities by integrating them over the level widths. For the numerical tests we consider only the binary problem $L=2$, in which the states correspond to whether the signal at a particular time is positive or negative. The singlet probability for a positive signal is

$$\wp^{(1)}(+;r)\equiv\rho = \int_0^{\infty} ds\ \omega^{(1)}(s;r) = \frac{1}{\sqrt{2\pi}}\int_0^{\infty} ds\ e^{-s^2/2} = 1/2, \quad (1.15)$$

which is independent of r because we are dealing with homogeneous signals.

The pair probability depends upon the spectral distribution of the signal via $K(r)$. Inverting the covariance matrix we have

$$\omega^{(2)}(s_1,s_2;r) = \frac{1}{2\pi\sqrt{1-K(r)^2}} \times\exp\left[-\frac{s_1^2-2K(r)s_1s_2+s_2^2}{2[1-K(r)^2]}\right]. \quad (1.16)$$

The probability that two measurements separated by r are both positive is [3,4]

$$\wp^{(2)}(+,+;r) = \int_0^{\infty} ds_1\int_0^{\infty} ds_2\,\omega^{(2)}(s_1,s_2;r) = \left[1+\frac{2}{\pi}\arcsin K(r)\right]\Big/ 4 \quad (1.17)$$

and the pair-correlation function is

$$g^{(2)}(\sigma_1,\sigma_2;r)\equiv\wp^{(2)}(\sigma_1,\sigma_2;r)/\rho(\sigma_1)\rho(\sigma_2)$$
$$=1+\sigma_1\sigma_2\frac{2}{\pi}\text{arcsin}K(r), \tag{1.18}$$

where $\sigma_i=\pm 1$. It is also possible to give an analytic expression for the triplet correlation function [3,4]

$$g^{(3)}(\sigma_1,\sigma_2,\sigma_3;r_{12},r_{23},r_{31})$$
$$=1+\frac{2}{\pi}[\sigma_1\sigma_2\text{arcsin}K(r_{12})$$
$$+\sigma_2\sigma_3\text{arcsin}K(r_{23})+\sigma_3\sigma_1\text{arcsin}K(r_{31})]. \tag{1.19}$$

C. Simulation

To obtain benchmarks to test the entropy expansions we simulated correlated Gaussian signals by the superposition of random waves. The details of the simulation and the estimation of the entropy have been described by Marčelja [5]. We used a Lorentzian spectral distribution

$$f(k)=\frac{\gamma/\pi}{\gamma^2+k^2}, \tag{1.20}$$

which has exponential covariance function

$$K(r)=e^{-\gamma|r|}. \tag{1.21}$$

We also used a Gaussian spectral distribution

$$f(k)=\frac{\sigma}{\sqrt{2\pi}}e^{-\sigma^2k^2/2}, \tag{1.22}$$

which has covariance function

$$K(r)=e^{-r^2/2\sigma^2}. \tag{1.23}$$

Here γ^{-1} and σ are the correlation lengths of the respective models. Both the Gaussian and the Lorentzian can be used to represent naturally occurring signals. The difference between them is that the Lorentzian model contains much-higher-frequency components than the Gaussian model.

In the case of the Lorentzian we used the discrete Fourier transform for the simulations. For sampling on a grid of N points of uniform spacing Δ_r such that $r_n=n\Delta_r$,

$$f_n=\frac{N^{-1}\sinh\gamma\Delta_r}{\cosh\gamma\Delta_r-\cos k_n\Delta_r}, \tag{1.24}$$

where $k_n=2\pi n/N\Delta_r$. For the Gaussian model we approximated its discrete Fourier transform by evaluating the continuous function on the grid points, $f_n=f(k_n)$.

Figure 1 shows typical random Gaussian signals obtained from the simulation. The most noticeable difference between the Lorentzian and Gaussian models are the high-frequency components that are present in the former, due to the slow k^{-2} decay of the spectral density. The low-frequency components of the two signals are rather similar and in both cases one can visualize similarly sized domains. This similarity is particularly clear in the digitized, sampled signals.

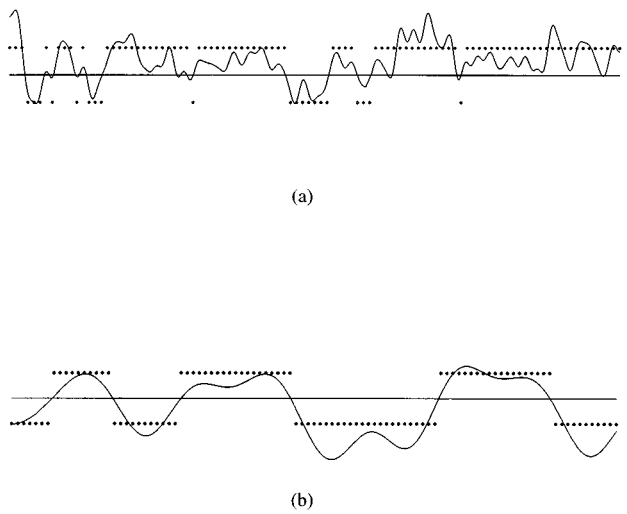

FIG. 1. Correlated random signals with (a) Lorentzian and (b) Gaussian spectral densities. The result of binary digitization and regular sampling of the signal is shown as symbols. The correlation length is five nodes in both cases.

For the simulations the Gaussian random process is constructed in Fourier space and then the fast Fourier transform is used to give $s(r_i)$. The number of sample points was typically $N=2^{16}$–2^{18}. The larger number of points was required for signals with longer correlation lengths. One random wave was generated for each of the discrete wave vectors $\{k_n\}$ (thus $M=N$) by choosing the phases $\{\phi_n\}$ randomly from a uniform distribution on $[0,2\pi]$ and by assigning the wave an amplitude in proportion to the spectral distribution. The number of independent signals used to collect the statistics was typically 2000. It was always checked that the results were independent of the simulation system size and the number of realizations.

The entropy was calculated in a fashion similar to that used by Shannon to estimate the redundancy of the English language [1]. First the signal is leveled at each sampling point so that it has a binary value. Clusters of m sites are analyzed and the probability of each of the 2^m possibilities is evaluated from the number of times it recurs in the process. The entropy of clusters of this size is simply $S^{(m)}=\Sigma_{i=1}^{2^m}\wp_i\ln\wp_i$. To save memory we exploit both the up-down symmetry of the spins, by storing a sequence and its complement together, and the time-reversal symmetry of the signal, by storing a sequence together with its reversal. The entropy per site $S^{(m)}/m$ is then plotted against $1/m$ and the infinite limit result is estimated by extrapolating to $1/m=0$.

An example of the extrapolation procedure used for the estimation of the entropy in the simulations is shown in Fig. 2. The entropy per node is plotted against the reciprocal of the number of nodes used in the subsequences. This example, which is a Lorentzian signal with relatively short-ranged correlations $\gamma=0.5$, is typical of the results obtained here. From the figure we estimate the simulated entropy per node for this signal as $s=0.6008$.

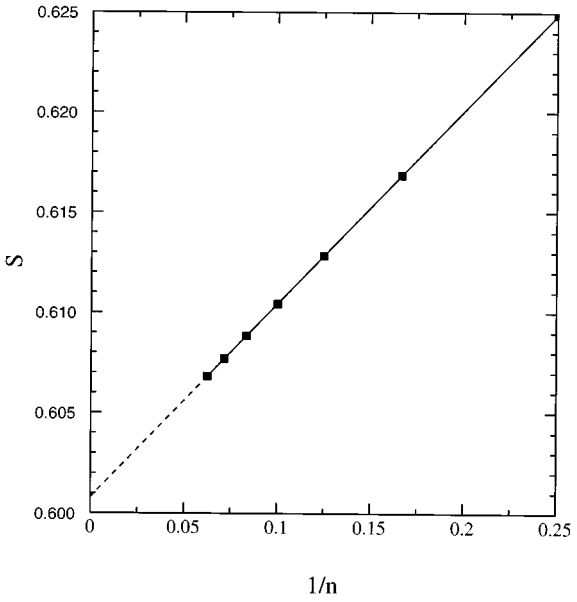

FIG. 2. Entropy per site obtained from the simulations by extrapolation, using the probability distribution of subsequences of size n. This is an example of the procedure for the Lorentzian model with $\gamma\Delta_r=0.5$.

II. ENTROPY EXPANSIONS

A. Generalized Kirkwood superposition expansion

We cast the digitized, sampled signal as a spin lattice model in statistical mechanics. We consider a lattice of N sites, each site occupied by a spin variable with L possible levels: $\sigma_i=1,2,\ldots,L,\ i=1,2,\ldots,N$. Each configuration of the spins corresponds to a particular signal. There is no constraint on the value of the total spin: In the language of the lattice gas this is an $(L-1)$-component open system. We will denote a configuration of a subset of n sites by $(\sigma^n;r^n)$, which means σ_1 at r_1, σ_2 at r_2, ..., and σ_n at r_n, where r_i is a particular lattice site. We let $\wp^{(N)}(\sigma^N;r^N)$ be the probability of a particular configuration of the system occurring. Shannon [1] showed that the information entropy is

$$S=-k\sum_{\sigma^N}\wp^{(N)}(\sigma^N;r^N)\ln\wp^{(N)}(\sigma^N;r^N). \tag{2.1}$$

The sum is over the L^N possible configurations of the system. Henceforth we shall set the constant $k=1$. Although formally exact, this expression cannot be used in this form because in general one does not know the probability of all of the configurations of the system. If all configurations were equally likely then $\wp^{(N)}=L^{-N}$ and $S=N\ln L$. Correlations imply that some configurations are more likely than others, which makes the system predictable to some extent and consequently reduces the entropy, $S\leqslant N\ln L$.

We will now formally expand the entropy in terms of correlation functions. First we define the n-site probability, which is the probability that a particular subset of n sites will have a specific set of spins, irrespective of the spins at the remaining sites,

$$\wp^{(n)}(\tau^n;r^n)=\sum_{\sigma^N}\wp^{(N)}(\sigma^N;r^N)\prod_{i=1}^{n}\delta(\tau_i,\sigma_i), \tag{2.2}$$

where the Kronecker delta appears. Perhaps most important is the singlet probability $\wp^{(1)}(\tau;r)$; for a homogeneous system this is independent of the site r, in which case we denote it by $\rho(\tau)$. We shall also define a correlation function $g^{(n)}$, which is just the probability normalized by the singlet probabilities,

$$g^{(n)}(\tau^n;r^n)=\wp^{(n)}(\tau^n;r^n)\Big/\prod_{i=1}^{n}\wp^{(1)}(\tau_i;r_i). \tag{2.3}$$

The pair-correlation function is the most familiar of these; for a homogeneous system this only depends upon the (lattice) separation of the two sites $g^{(2)}(\tau_1,\tau_2;r_{12})$. The departure of $g^{(n)}$ from unity measures the correlation between the subset of sites; if the sites are independent of each other then $g^{(n)}=1$. We shall be concerned with systems with finite-ranged correlations, which means in the limit that one site is far from the rest; we have $g^{(n)}(\tau^n;r^n)\to g^{(n-1)}(\tau^{n-1};r^{n-1})$, $r_n\to\infty$. If all the sites are well separated, $g^{(n)}(\tau^n;r^n)\to 1$, $r^n\to\infty$. Note that because we are dealing with an open system the asymptotic limit is precisely unity.

We can use the asymptotic behavior of the correlation functions to define a generalized Kirkwood superposition expansion. For $n\geqslant 3$ we define

$$g^{(n)}(\sigma^n;r^n)\equiv\Delta^{(n)}(\sigma^n;r^n)\prod_{s=2}^{n-1}\times\left[\prod_{\{i^s\}}g^{(s)}(\sigma_{i_1},\ldots\sigma_{i_s};r_{i_1},\ldots r_{i_s})\right]^{(-1)^{s+n-1}}, \tag{2.4}$$

where the inner product is over the ${}^nC_s$ different ways of choosing s sites from the n sites. This expression formally defines the remainder $\Delta^{(n)}$; the generalized Kirkwood superposition approximation is to take $\Delta^{(n)}=1$ (cf. Ref. [6]). Reiss [7] derived the generalized superposition expression using a variational argument, and it can be shown to be the only superposition approximation that is consistent with the asymptotic behavior of the correlation functions [8]. For the triplet correlation function we have (using an obvious abbreviated notation)

$$g^{(3)}(1,2,3)=g^{(2)}(1,2)g^{(2)}(2,3)g^{(2)}(3,1)\Delta^{(3)}(1,2,3). \tag{2.5}$$

Setting $\Delta^{(3)}=1$ yields the Kirkwood superposition approximation [9]. For the quadruplet we obtain

$$g^{(4)}(1,2,3,4)=\frac{g^{(3)}(1,2,3)g^{(3)}(1,2,4)g^{(3)}(1,3,4)g^{(3)}(2,3,4)}{g^{(2)}(1,2)g^{(2)}(1,3)g^{(2)}(1,4)g^{(2)}(2,3)g^{(2)}(2,4)g^{(2)}(3,4)}\Delta^{(4)}(1,2,3,4)$$
$$=g^{(2)}(1,2)g^{(2)}(1,3)g^{(2)}(1,4)g^{(2)}(2,3)g^{(2)}(2,4)g^{(2)}(3,4)\Delta^{(3)}(1,2,3)\Delta^{(3)}(1,2,4)$$
$$\times\Delta^{(3)}(1,3,4)\Delta^{(3)}(2,3,4)\Delta^{(4)}(1,2,3,4). \tag{2.6}$$

Setting $\Delta^{(4)}=1$ yields the superposition approximation given by Fisher and Kopeliovich [10]. By induction it follows that

$$g^{(N)}(1,2,\ldots,N)=g^{(2)}(1,2)g^{(2)}(1,3)\ldots g^{(2)}(N-1,N)$$
$$\times\Delta^{(3)}(1,2,3)\ldots\Delta^{(N)}(1,2,\ldots,N). \tag{2.7}$$

We can now use this formal expression for the N-particle correlation function in the expression for the entropy. We write

$$\begin{aligned}\ln\wp^{(N)}(\sigma^N;r^N)&=\ln[\wp^{(1)}(\sigma_1;r_1)\cdots\wp^{(1)}(\sigma_N;r_N)g^{(N)}(\sigma^N;r^N)]\\&=\sum_{i=1}^{N}\ln\wp^{(1)}(\sigma_i;r_i)+\ln[g^{(2)}(1,2)\cdots g^{(2)}(N-1,N)\\&\quad\times\Delta^{(3)}(1,2,3)\cdots\Delta^{(N)}(1,2,\ldots,N)]\\&=\sum_{i=1}^{N}\ln\wp^{(1)}(\sigma_i;r_i)+\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\ln g^{(2)}(\sigma_i,\sigma_j;r_i,r_j)\\&\quad+\sum_{i=1}^{N-2}\sum_{j=i+1}^{N-1}\sum_{k=j+1}^{N}\ln\Delta^{(3)}(\sigma_i,\sigma_j,\sigma_k;r_i,r_j,r_k)+\cdots.\end{aligned} \tag{2.8}$$

Using the definition of the reduced probability functions, the entropy becomes

$$\begin{aligned}S&=-\sum_{\sigma^N}\wp^{(N)}(\sigma^N;r^N)\ln\wp^{(N)}(\sigma^N;r^N)\\&=-\sum_{i=1}^{N}\sum_{\sigma_i=1}^{L}\wp^{(1)}(\sigma_i;r_i)\ln\wp^{(1)}(\sigma_i;r_i)\\&\quad-\sum_{i<j}\sum_{\sigma_i,\sigma_j}\wp^{(2)}(\sigma_i,\sigma_j;r_i,r_j)\ln g^{(2)}(\sigma_i,\sigma_j;r_i,r_j)\\&\quad-\sum_{i<j<k}\sum_{\sigma_i,\sigma_j,\sigma_k}\wp^{(3)}(\sigma_i,\sigma_j,\sigma_k;r_i,r_j,r_k)\\&\quad\times\ln\Delta^{(3)}(\sigma_i,\sigma_j,\sigma_k;r_i,r_j,r_k)\cdots.\end{aligned} \tag{2.9}$$

For the homogeneous problem and in the limit $N\to\infty$, this becomes

$$\begin{aligned}S/N&=-\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma)\\&\quad-\frac{1}{2!}\sum_{j=1}^{N}\sum_{\sigma_i,\sigma_j}\wp^{(2)}(\sigma_i,\sigma_j;r_{ij})\ln g^{(2)}(\sigma_i,\sigma_j;r_{ij})\\&\quad-\frac{1}{3!}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\sigma_i,\sigma_j,\sigma_k}\\&\quad\times\wp^{(3)}(\sigma_i,\sigma_j,\sigma_k;r_{ij},r_{jk},r_{ki})\\&\quad\times\ln\Delta^{(3)}(\sigma_i,\sigma_j,\sigma_k;r_{ij},r_{jk},r_{ki})-\cdots.\end{aligned} \tag{2.10}$$

Note that these are distinct site probabilities, so that $\wp^{(n)}(\sigma^n;r^n)=0$ if any $r_i=r_j$. This then is a formally exact expansion for the entropy in terms of successively higher-order correlation functions $S/N=s_1+s_2+\cdots$. An approximation for the entropy can be obtained by neglecting s_n and higher-order terms, which corresponds to setting $\Delta^{(n)}=1$. Such correlation function expansions for the entropy are well known in liquid-state theory (see Ref. [11] and references therein); the present derivation follows that of Wallace [12].

B. Generalized Markovian expansion

The superposition approximation given above is generally applicable. But for a one-dimensional lattice it is possible to do better by exploiting the order inherent in the geometry. Specifically, the correlation between three sites i, j, and k, ordered along the line such that $r_i<r_j<r_k$, is largely determined by the correlation of the terminal sites with the central site,

$$g^{(3)}(\sigma_i,\sigma_j,\sigma_k;r_i,r_j,r_k)\approx g^{(2)}(\sigma_i,\sigma_j;r_i,r_j)g^{(2)}(\sigma_j,\sigma_k;r_j,r_k). \tag{2.11}$$

This Markovian approximation for the triplet correlation function has been previously exploited in theories for chain polymers [8,13]. It has the correct asymptotic behavior as $r_k\to\infty$, namely, only the correlation between sites i and j remains. For the quadruplet case we can generalize this by taking

$$g^{(4)}(\sigma_i,\sigma_j,\sigma_k,\sigma_l;r_i,r_j,r_k,r_l) \approx \frac{g^{(3)}(\sigma_i,\sigma_j,\sigma_k;r_i,r_j,r_k)g^{(3)}(\sigma_j,\sigma_k,\sigma_l;r_j,r_k,r_l)}{g^{(2)}(\sigma_j,\sigma_k;r_j,r_k)}, \tag{2.12}$$

where $r_i<r_j<r_k<r_l$. The denominator corrects for the double counting of the shared pair. This may be seen since as $r_l \to \infty$, $g^{(3)}(\sigma_j,\sigma_k,\sigma_l;r_j,r_k,r_l) \to g^{(2)}(\sigma_j,\sigma_k;r_j,r_k)$, which cancels with the denominator leaving $g^{(4)}(\sigma_i,\sigma_j,\sigma_k,\sigma_l;r_i,r_j,r_k,r_l) \to g^{(3)}(\sigma_i,\sigma_j,\sigma_k;r_i,r_j,r_k)$, as required. Extending this to the general case, we define $\overline{\Delta^{(n)}}$ as the ratio of the n-site correlation function to its generalized Markovian superposition approximation and we have the formal definition

$$g^{(n)}(\sigma_{i_1}\cdots\sigma_{i_n};r_{i_1}\cdots r_{i_n}) \equiv \frac{g^{(n-1)}(\sigma_{i_1}\cdots\sigma_{i_{n-1}};r_{i_1}\cdots r_{i_{n-1}})g^{(n-1)}(\sigma_{i_2}\cdots\sigma_{i_n};r_{i_2}\cdots r_{i_n})}{g^{(n-2)}(\sigma_{i_2}\cdots\sigma_{i_n};r_{i_2}\cdots r_{i_n})}\overline{\Delta^{(n)}}(\sigma_{i_1}\cdots\sigma_{i_n};r_{i_1}\cdots r_{i_n}), \tag{2.13}$$

where $r_{i_j}<r_{i_k}$ if $j<k$.

The entropy is a functional of $g^{(N)}$, which because of the ordering involves all N sites consecutively. We denote consecutive site probabilities by $P^{(n)}$ and consecutive site correlations by $G^{(n)}$ and exhibit the location of only the first site. Hence we have

$$\begin{aligned} &G^{(N)}(\sigma_1,\sigma_2,\ldots,\sigma_N;r_1) \\ &= \frac{G^{(N-1)}(\sigma_1,\ldots,\sigma_{N-1};r_1)G^{(N-1)}(\sigma_2,\ldots,\sigma_N;r_2)}{G^{(N-2)}(\sigma_2,\ldots,\sigma_{N-1};r_2)}\overline{\Delta^{(N)}}(\sigma_1,\sigma_2,\ldots,\sigma_N;r_1,r_2,\cdots r_N) \\ &= G^{(2)}(\sigma_1,\sigma_2;r_1)G^{(2)}(\sigma_2,\sigma_3;r_2)\cdots G^{(2)}(\sigma_{N-1},\sigma_N;r_{N-1}) \\ &\quad\times\overline{\Delta^{(3)}}(\sigma_1,\sigma_2,\sigma_3;r_1,r_2,r_3)\cdots\overline{\Delta^{(3)}}(\sigma_{N-2},\sigma_{N-1},\sigma_N;r_{N-2},r_{N-1},r_N)\cdots\overline{\Delta^{(N)}}(\sigma_1,\sigma_2,\ldots,\sigma_N;r_1,r_2,\ldots,r_N) \end{aligned} \tag{2.14}$$

and consequently the entropy is

$$\begin{aligned} S=&-\sum_{i=1}^{N}\sum_{\sigma_i=1}^{L}\wp^{(1)}(\sigma_i;r_i)\ln\wp^{(1)}(\sigma_i;r_i) \\ &-\sum_{i=1}^{N-1}\sum_{\sigma_i,\sigma_{i+1}}P^{(2)}(\sigma_i,\sigma_{i+1};r_i)\ln G^{(2)}(\sigma_i,\sigma_{i+1};r_i) \\ &-\sum_{i=1}^{N-2}\sum_{\sigma^3}P^{(3)}(\sigma_i,\sigma_{i+1},\sigma_{i+2};r_i) \\ &\times\ln\overline{\Delta^{(3)}}(\sigma_i,\sigma_{i+1},\sigma_{i+2};r_i,r_{i+1},r_{i+2})-\cdots. \end{aligned} \tag{2.15}$$

In the limit $N\to\infty$ (so that $N-1$ can be replaced by N, etc.) and for a homogeneous system $[\wp(\sigma,r_i)=\wp(\sigma)\equiv\rho(\sigma)]$, the entropy per site is

$$\begin{aligned} S/N=&-\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma)-\sum_{\sigma,\tau}\rho(\sigma)\rho(\tau)G^{(2)}(\sigma,\tau)\ln G^{(2)} \\ &\times(\sigma,\tau)-\sum_{\sigma,\tau,\lambda}\rho(\sigma)\rho(\tau)\rho(\lambda)G^{(3)}(\sigma,\tau,\lambda) \\ &\times\ln\overline{\Delta^{(3)}}(\sigma,\tau,\lambda)-\cdots. \end{aligned} \tag{2.16}$$

Notice that compared to the generalized Kirkwood superposition approximation, which sums correlations over all sites with an appropriate combinatorial factor, this expression only involves correlations between consecutive sites.

C. Limit of large and small correlations

A guide to the accuracy of the approximations obtained by truncating the expansions is given by examining the limits of extreme correlations. In the low-correlation limit we have

$$\wp^{(N)}(\sigma^N;r^N)\sim\prod_{i=1}^{N}\rho(\sigma_i) \tag{2.17}$$

and the entropy is

$$S/N=-\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma). \tag{2.18}$$

In this limit all the correlation functions are unity, $g^{(n)}(\sigma^n;r^n)=1$, $n=2,\ldots,N$. Consequently, both the Kirkwood and the Markov generalized superposition approximations are exact, $\Delta^{(n)}(\sigma^n;r^n)=\overline{\Delta^{(n)}}(\sigma^n;r^n)=1$, $n=3,\ldots,N$, and only the first term contributes to the entropy expansion, giving the exact result in both cases.

In the opposite fully correlated regime, where there is a distribution of the spin among the different systems of the ensemble, but the spins of any one system in the ensemble are identical, we have

$$\wp^{(N)}(\sigma^N;r^N)\sim\rho(\sigma)\prod_{i=1}^{N}\delta(\sigma_i,\sigma). \tag{2.19}$$

The entropy is

$$S=-\sum_{\sigma^N}\wp^{(N)}(\sigma^N;r^N)\ln\wp^{(N)}(\sigma^N;r^N)=-\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma), \tag{2.20}$$

which no longer scales with the size of the system. In this limit the probability is $\wp^{(n)}(\sigma^n;r^n)=\rho(\sigma)\Pi_{i=1}^{n}\delta(\sigma_i,\sigma)$, and hence the correlation functions for identical spins are $g^{(n)}(\sigma\cdots\sigma;r^n)=\rho(\sigma)^{1-n}$, $n=2,\ldots,N$. Now in the case of the generalized Kirkwood superposition approximation the remainder is

$$\Delta^{(n)}(\sigma\cdots\sigma;r^n)=\rho(\sigma)^{-(n-1)}\rho(\sigma)^{(n-2)^nC_{n-1}}\times\rho(\sigma)^{-(n-3)^nC_{n-2}}\cdots\rho(\sigma)^{{}^nC_2(-1)^{n+1}}. \tag{2.21}$$

The exponent is

$$\begin{aligned}\exp&=(1-n)[{}^nC_n-{}^nC_{n-1}+{}^nC_{n-2}-\cdots+(-1)^{n+2n}C_2]\\&\quad-[{}^nC_{n-1}-2\,{}^nC_{n-2}+3\,{}^nC_{n-3}\\&\quad-\cdots+(-1)^{n+1}(n-2)^nC_2]\\&=(1-n)[{}^nC_1-{}^nC_0](-1)^n\\&\quad-[(n-1)^nC_1-n^nC_0](-1)^{n+1}\\&=(-1)^{n+1},\end{aligned} \tag{2.22}$$

which gives

$$\Delta^{(n)}(\sigma\cdots\sigma;r^n)=\rho(\sigma)^{(-1)^{n+1}}. \tag{2.23}$$

Using this the generalized Kirkwood superposition approximation for the entropy yields (replacing the sums over particle positions by $N-1$, $N-2$, etc.)

$$\begin{aligned}S&=-N\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma)-\frac{N(N-1)}{2!}\sum_{\sigma=1}^{L}\rho(\sigma)\ln g^{(2)}(\sigma,\sigma)\\&\quad-\frac{N(N-1)(N-2)}{3!}\sum_{\sigma=1}^{L}\rho(\sigma)\ln\Delta^{(3)}(\sigma,\sigma,\sigma)-\cdots\\&=\left[-1+\left(1-N+\frac{N(N-1)}{2!}-\frac{N(N-1)(N-2)}{3!}+\cdots\right)\right]\\&\quad\times\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma)\\&=[-1+(1-1)^N]\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma)\\&=-\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma).\end{aligned} \tag{2.24}$$

This is certainly the exact result, but it required the precise cancellation of all the terms. In general, if the expansion is truncated after a few terms $n\lesssim N/2$, an error of order N^n will be made. This completely dominates the exact entropy, which is of order unity. We conclude that the generalized Kirkwood superposition approximation is not a good expansion for highly correlated systems.

TABLE I. Entropy per site for the Lorentzian model as a function of inverse correlation length.

		Kirkwood		Markov	
γ	Simulation	s_1+s_2	$s_1+s_2+s_3$	s_1+s_2	$s_1+s_2+s_3$
1.5	0.6827	0.6823	0.6828	0.6829	0.6828
1.0	0.6636	0.6598	0.6641	0.6641	0.6636
0.6	0.6205	0.5960	0.6299	0.6231	0.6209
0.5	0.6008	0.5592	0.6216	0.6044	0.6011
0.25	0.511	0.3108	0.7688	0.5216	0.5130
0.1	0.383	-0.5088	3.6089	0.4049	0.3898
0.05	0.2964	-1.9202	16.724	0.3247	0.3069
0.025	0.222	-4.7733	74.680	0.2559	0.2375
0.01	0.148	-13.373	502.8	0.1833	0.1664
0.005	0.108	-27.727	2064	0.1410	0.1260
0.0025	0.078	-56.452	8368	0.1076	0.0949
0.001	0.053	-142.64	52716	0.0747	0.0648

The generalized Markov superposition approximation is much better behaved in this regard since the remainder is

$$\overline{\Delta^{(n)}}(\sigma\cdots\sigma;r^n)=\frac{G^{(n)}(\sigma\cdots\sigma)G^{(n-2)}(\sigma\cdots\sigma)}{G^{(n-1)}(\sigma\cdots\sigma)G^{(n-1)}(\sigma\cdots\sigma)}=\frac{\rho(\sigma)^{1-n}\rho(\sigma)^{3-n}}{\rho(\sigma)^{2-n}\rho(\sigma)^{2-n}}=1. \tag{2.25}$$

Accordingly, only the first two terms in the entropy expansion are nonzero,

$$\begin{aligned}S&=-N\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma)-(N-1)\sum_{\sigma=1}^{L}\rho(\sigma)\ln G^{(2)}(\sigma,\sigma)\\&=-[N-(N-1)]\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma)=-\sum_{\sigma=1}^{L}\rho(\sigma)\ln\rho(\sigma),\end{aligned} \tag{2.26}$$

which is the exact result. In this case we see that any truncation of the Markovian expansion for $n\geqslant 3$ will still yield the exact entropy in the completely correlated limit.

D. Numerical results

Results for the Lorentzian and for the Gaussian models are given in Tables I and II, respectively. For both the generalized Kirkwood superposition approximation and the Markov superposition approximation we give the two- and three-term expansions. It can be seen that in the latter case two terms are already good and that the third term gives a minor but consistent improvement. In general, the entropy per node is a maximum for the random signal (small correlation length) and goes to zero as the correlation length increases.

Figure 3 graphically tests the truncated expansion based upon the generalized Kirkwood superposition expression against the simulation result for the Lorentzian model. For weak correlations $\gamma\gtrsim 0.5$, the two- and three-term expan-

TABLE II. Entropy per site for the Gaussian model as a function of correlation length.

		Kirkwood		Markov	
σ	Simulation	s_1+s_2	$s_1+s_2+s_3$	s_1+s_2	$s_1+s_2+s_3$
800	0.004	-105.992	16 794	0.0035	0.0035
400	0.007	-85.8003	10 880	0.0065	0.0065
200	0.011	-58.6195	4676.509	0.0118	0.0118
100	0.022	-31.7868	1249.406	0.0215	0.0215
50	0.040	-15.418	296.5258	0.0385	0.0385
20	0.080	-5.5511	40.6511	0.0817	0.0815
12	0.120	-2.9237	12.3082	0.1224	0.1218
5	0.230	-0.6361	1.2667	0.2364	0.2339
2	0.4225	0.31755	0.42066	0.43274	0.42520
1	0.5972	0.60066	0.59786	0.60443	0.59888
0.7	0.6569	0.66528	0.66422	0.66533	0.66423
0.5	0.6815	0.68941	0.68938	0.68941	0.68938
0.36788	0.68945	0.69302	0.69302	0.69302	0.69302
0.25	0.69234	0.69315	0.69315	0.69315	0.69315
0.125	0.69310	0.69315	0.69315	0.69315	0.69315
0.0625	0.69314	0.69315	0.69315	0.69315	0.69315
0.02	0.69315	0.69315	0.69315	0.69315	0.69315

sions agree with each other and with the simulation result. In this regime the entropy per site is just $S/N \approx s_1 = \ln 2$. As the correlation length is increased the entropy drops and the two approximations bracket the simulation result. In fact, they begin to diverge as the third term in the expansion overcorrects the error in the first two terms. The approximation is useless for signals with correlations extending more than a few nodes.

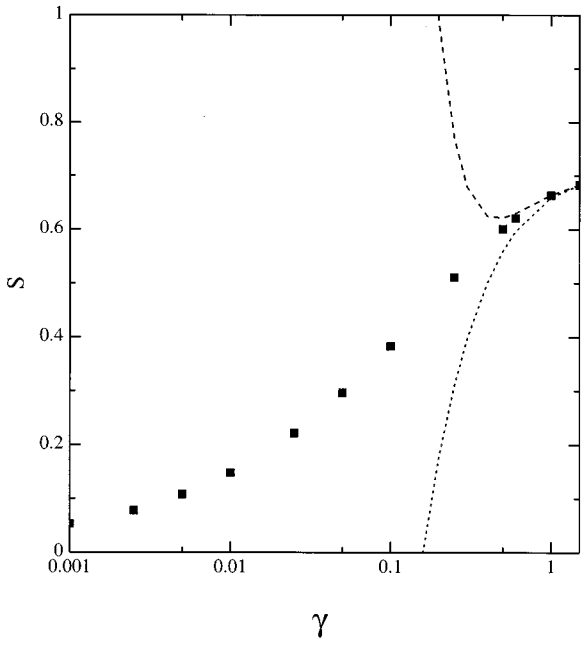

FIG. 3. Entropy per site as a function of inverse correlation length (in units of the node spacing) for the binary digitized Lorentzian model. The symbols represent the simulation results and the curves represent the generalized Kirkwood superposition approximation (the dotted curve uses two terms in the series and the dashed curve uses three).

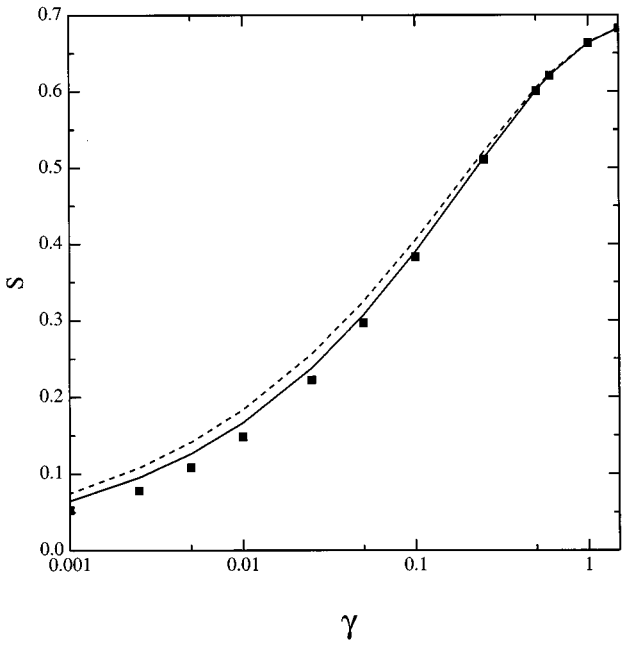

FIG. 4. Test of the generalized Markov superposition approximation for the Lorentzian model. The symbols are the simulation data, the dashed curve is the two-term series, and the full curve includes the first three terms.

In contrast, the truncated expansion based upon the generalized Markov superposition expression is uniformally accurate for all correlation lengths, as can be seen in Fig. 4. The two-term approximation is already quite good and including the third term gives even better agreement with the simulation data.

Figure 5 tests the expansions for the Gaussian signal and the conclusions are similar to those drawn for the Lorentzian: The validity of the Kirkwood approximation is restricted to weak correlations and the Markov expression is uniformly accurate over the whole regime.

The sigmoidal character of the entropy as a function of the inverse correlation length is more apparent for the Gaussian signal than for the Lorentzian (cf. Fig. 4). Although both asymptotically approach $s \sim \ln 2$ for correlations not extending beyond nearest neighbors, they approach the opposite highly correlated limit $s = 0$ at different rates. The Gaussian already appears asymptotic by the time the correlation length is 100 nodes or so, whereas the Lorentzian entropy is still noticeably nonzero even with a correlation length of 1000. This is no doubt due to the influence of the high-frequency spectral components in the latter model.

The Markov approximation will work well for a broad class of signals. We also carried out tests for periodic signals by using a model with sinusoidal covariance and exponentially decaying amplitude. The performance of the Markov approximation was similar to the above, except for signals with long-period oscillations and even longer correlation length. The entropy for these highly correlated signals tended to be overestimated by both the simulations and the Markov approximation; the deterministic changes of sign were perceived as random by both. We have not shown these results

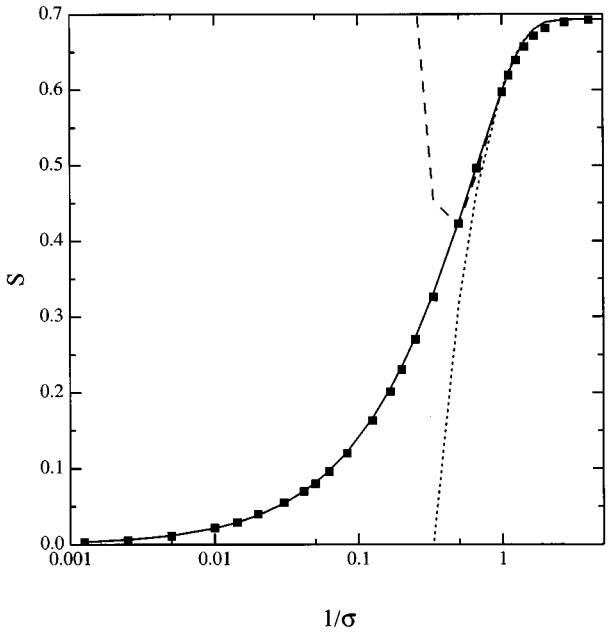

FIG. 5. Gaussian model. The symbols are the simulation data, the dotted and dashed curves are the two- and three-term generalized Kirkwood superposition approximation, respectively, and the full curve is the three-term generalized Markov superposition approximation, which on this scale is indistinguishable from the two-term version.

because the cluster sizes required to estimate reliably the entropy from the simulations was intractably large in this regime.

III. ENTROPY OF A CONTINUOUS SIGNAL

The analysis and results that were obtained above applied to a signal that had already been sampled and digitized. We now address the question of whether it is meaningful to speak of the entropy of a continuous signal and we explore the relationship between the probability and the probability density and between the corresponding digital and the analogue entropies.

In Sec. III A we distinguish between the formally exact expression for the discrete entropy and an approximate expression that is applicable to continuous signals and we give the criterion that sets the regime of validity of the latter (Sec. III A 1). In Sec. III B we explore the example of an exponentially covariant continuous Gaussian random signal, which yields to exact analysis. In Sec. III B 1 it is shown that the correlation functions of the model are Markovian, which allows us to obtain explicitly the results for the approximate entropy. The validity of the digitization approximation is derived in Sec. III B 2, and the entropy in the limits of infinite sampling and digitization is obtained explicitly in the general case (Sec. III B 3) and for binary digitization (Sec. III B 4). The results of this section make clear the sense in which the entropy of a continuous signal depends upon the degree of sampling and digitization. In Sec. III C 1 we give an example of a binary digitized signal that has Markovian correlations, which means that the entropy can be calculated exactly. This turns out to be equivalent to the one-dimensional Ising spin lattice model with nearest-neighbor interactions.

A. Exact and approximate entropies

In practice, a continuous signal has to be sampled and digitized, by measuring it at regular intervals and by assigning each measurement one of a discrete set of values. This is appropriate and consistent with using entropy to measure information content because entropy itself is only defined for discrete probabilities. The question we address here is the dependence of the value of the entropy on the process of sampling and digitization.

We imagine that the continuous signal $s(r)$ has been sampled at regular intervals, so that $s_n = s(r_n)$ and $r_n = n\Delta_r$, where $s \in (-\infty, \infty)$ is real variable, and we now wish to digitize it and to convert the probability densities $\Omega^{(n)}$ to probabilities $P^{(n)}$. We define an integer variable σ that indexes the cell in which the measured signal falls. For simplicity we take the cells to be of uniform width Δ_s. [It is straightforward to replace this constant width by a function $\Delta_s(s)$, if desired.] The discretization is accomplished by means of a coarse-grained δ function,

$$\Theta_\Delta(s,t) \equiv \theta(s-t+\Delta_s/2)\,\theta(t-s+\Delta_s/2), \tag{3.1}$$

where the Heaviside step function is $\theta(x)=1$, $x>0$, and $\theta(x)=0$, $x<0$. This is more closely related to the Kronecker δ rather than the Dirac δ, which is important because we shall use the property $\Theta \ln \Theta = 0$. With this the discrete probability for n consecutive sites is defined to be

$$P^{(n)}(\sigma^n) \equiv \int ds_1 \cdots ds_n \Omega^{(n)}(s^n) \prod_{i=1}^{n} \Theta_\Delta(\sigma_i \Delta_s, s_i). \tag{3.2}$$

The entropy is the functional of the discrete probability that has been used throughout. Accordingly, this definition of the discrete probability determines uniquely the entropy of the continuous signal

$$S(\Delta_s, \Delta_r) \equiv -\sum_{\sigma^N} P^{(N)}(\sigma^N) \ln P^{(N)}(\sigma^N). \tag{3.3}$$

Here we have explicitly indicated that the value of the entropy of the continuous signal depends upon the degree of sampling and upon the level of digitization; it is not possible to speak of an entropy independent of these. It is emphasized that this defines *the* entropy of the continuous signal. This definition and that for the probability are formally exact and as such they are always valid.

It is possible to make a useful approximation to these exact expressions that becomes increasingly valid as the level of digitization is refined. We approximate the integral by a trapezoidal sum to obtain an approximation for the consecutive site probability $\tilde{P}^{(n)} \approx P^{(n)}$, where

$$\tilde{P}^{(n)}(\sigma^n) \equiv \Delta_s^n \Omega^{(n)}(\Delta_s \sigma_1, \ldots, \Delta_s \sigma_n). \tag{3.4}$$

This is a valid approximation for the discrete probability when the integrand is slowly varying over the range of the discrete cells. We obtain the criteria for the validity of the approximation below.

We may also define an entropy that is an integral of the probability densities

$$\widetilde{S}(\Delta_s,\Delta_r)\equiv-\int_{-\infty}^{\infty}ds_1\cdots ds_N\Omega^{(N)}(s^N)\ln\Delta_s^N\Omega^{(N)}(s^N). \tag{3.5}$$

The reason for choosing this particular definition is that within the validity of the digitization approximation,

$$\begin{aligned}\widetilde{S}(\Delta_s,\Delta_r)&\approx-\sum_{\sigma^N}\Delta_s^N\Omega^{(N)}(\sigma_1\Delta_s,\ldots,\sigma_N\Delta_s)\\&\quad\times\ln\Delta_s^N\Omega^{(N)}(\sigma_1\Delta_s,\ldots,\sigma_N\Delta_s)\\&\approx-\sum_{\sigma^N}\widetilde{P}^{(N)}(\sigma^N)\ln\widetilde{P}^{(N)}(\sigma^N).\end{aligned} \tag{3.6}$$

Hence this has the same appearance as the formal definition of entropy and insofar as $\widetilde{P}^{(n)}\approx P^{(n)}$, then $\widetilde{S}\approx S$. There is no way of avoiding the appearance of the cell size Δ_s in the defining expression for $\widetilde{S}$, as dimensional considerations make clear. In many applications entropy differences with the same level cuts are the main concern, in which case as an additive constant the width of the levels does not contribute.

Validity of the digitization approximation

In practice, a guide to the permitted size of the level spacing Δ_s for the digitization approximation to be valid may be obtained by analyzing the singlet and neighbor correlation function. For Gaussian signals we expect that the width of the levels should be small compared to the variance and indeed this is essentially the criterion that we derive. Perhaps the most straightforward way to proceed is to note that if the probability density is slowly varying over the cell width, then the probability of adjacent levels must be similar, $\widetilde{P}^{(1)}(\sigma\pm1)\approx\widetilde{P}^{(1)}(\sigma)$ or $\Omega^{(1)}(s\pm\Delta_s)\approx\Omega^{(1)}(s)$. Furthermore, errors in the tail of the distribution are unimportant, so that we only need to apply this criterion in the region of maximum probability

$$|\Omega^{(1)}(\Delta_s)-\Omega^{(1)}(0)|\ll\Omega^{(1)}(0). \tag{3.7}$$

This alone is not sufficient because we are also concerned with effects of correlations on the digitization. The most important correlation is between neighbor nodes. Again, we insist upon a gradual variation in the neighbor probability $\Omega^{(2)}(s_1,s_2\pm\Delta_s)\approx\Omega^{(2)}(s_1,s_2)$. The worst case scenario is clearly $s_1=s_2=0$ and the criterion becomes

$$|\Omega^{(2)}(0,\Delta_s)-\Omega^{(2)}(0,0)|\ll\Omega^{(2)}(0,0). \tag{3.8}$$

B. Example: Exponential covariance

1. Markovian factorization of the correlation densities

The aim of this section is to solve a specific continuum model in which the approximate entropy can be obtained analytically and to exhibit the continuum limits of this model explicitly. We consider a Gaussian random signal with exponential covariance (this is the same as the model with Lorentzian spectral distribution considered numerically above)

$$\langle s(r)s(t)\rangle=e^{-\gamma|r-t|}, \tag{3.9}$$

where $s(r)\in(-\infty,\infty)$. If the signal is sampled at regular intervals, so that $s_n=s(r_n)$ and $r_n=n\Delta_r$, then the covariance is

$$\langle s_is_j\rangle=x^{|i-j|}, \tag{3.10}$$

where $x\equiv e^{-\gamma\Delta_r}$.

We now focus upon consecutive nodes. The covariance matrix for n consecutive nodes is

$$Q_n=\begin{pmatrix}1&x&x^2&\cdots&x^{n-1}\\x&1&x&\cdots&x^{n-2}\\x^2&x&1&\cdots&x^{n-3}\\\vdots&&&\ddots&\vdots\\x^{n-1}&x^{n-2}&x^{n-3}&\cdots&1\end{pmatrix}. \tag{3.11}$$

By induction it follows that the inverse is the tridiagonal matrix

$$Q_n^{-1}=\frac{1}{1-x^2}\begin{pmatrix}1&-x&0&\cdots&0\\-x&1+x^2&-x&\cdots&0\\0&-x&1+x^2&\cdots&0\\\vdots&&&\ddots&\vdots\\0&0&0&\cdots&1\end{pmatrix}. \tag{3.12}$$

That is, entries on the main diagonal equal $1+x^2$, except for the first and the last, which equal 1, and entries immediately above and below the main diagonal equal $-x$. Again, by induction we can show that the determinant of the consecutive covariance matrix is

$$|Q_n|=(1-x^2)^{n-1}. \tag{3.13}$$

The usual Gaussian probability density for n consecutive sites

$$\Omega^{(n)}(s^n)=\frac{\exp[-s^{nT}Q_n^{-1}s^n/2]}{(2\pi)^{n/2}(1-x^2)^{(n-1)/2}} \tag{3.14}$$

has an exponent that in this case simplifies to

$$\begin{aligned}\frac{-s^{nT}Q_n^{-1}s^n}{2}&=\frac{-1}{2(1-x^2)}\Bigg[(1+x^2)\sum_{i=1}^{n}s_i^2-2x\sum_{i=1}^{n-1}s_{i+1}s_i\\&\quad-x^2(s_1^2+s_n^2)\Bigg].\end{aligned} \tag{3.15}$$

We introduce the consecutive site correlation function

$$\Gamma^{(n)}(s^n)=\Omega^{(n)}(s^n)\Big/\prod_{i=1}^{n}\Omega^{(1)}(s_i), \tag{3.16}$$

where the single site probability density is, as usual,

$$\Omega^{(1)}(s)=\frac{1}{\sqrt{2\pi}}e^{-s^2/2} \tag{3.17}$$

and the neighbor probability density is

$$\Omega^{(2)}(s_1,s_2)=\frac{1}{2\pi\sqrt{1-x^2}}\exp\left[\frac{-1}{2(1-x^2)}(s_1^2-2s_1s_2x+s_2^2)\right]. \tag{3.18}$$

The exponential covariance function is important because it gives a Markovian consecutive site correlation function. We exploit this fact in evaluating the approximate entropy, whose correlation function expansion terminates after two terms. The Markovian character follows because the exponent of the n-site probability density may be written

$$-\frac{s^{nT}Q_n^{-1}s^n}{2}=-\frac{s^{(n-1)T}Q_{n-1}^{-1}s^{n-1}}{2}-\frac{1}{2(1-x^2)}[s_n^2+x^2s_{n-1}^2-2xs_ns_{n-1}] \tag{3.19}$$

and consequently

$$\begin{aligned}\Omega^{(n)}(s^n)&=\Omega^{(n-1)}(s^{n-1})\frac{1}{\sqrt{2\pi(1-x^2)}}\\&\quad\times\exp\left[\frac{-1}{2(1-x^2)}(s_n^2+x^2s_{n-1}^2-2xs_ns_{n-1})\right]\\&=\Omega^{(n-1)}(s^{n-1})\Omega^{(2)}(s_{n-1},s_n)/\Omega^{(1)}(s_{n-1})\\&=\Omega^{(1)}(s_1)\prod_{i=1}^{n-1}\Omega^{(2)}(s_i,s_{i+1})/\Omega^{(1)}(s_i).\end{aligned} \tag{3.20}$$

Alternatively,

$$\Gamma^{(n)}(s^n)=\prod_{i=1}^{n-1}\Gamma^{(2)}(s_i,s_{i+1}), \tag{3.21}$$

which is the Markovian factorization. This exact result for the Gaussian probability density is peculiar to an exponential covariance function and equally spaced sampling points.

The exact $P^{(n)}$ does not preserve the Markovian character of the consecutive site correlations. However, the approximation $\widetilde{P}^{(n)}=\Delta_s^n\Omega^{(n)}$ remains Markovian and hence the approximate entropy can be calculated directly from the first two terms of the Markovian expansion,

$$\begin{aligned}\widetilde{S}(\Delta_s,\Delta_r)&=-\int_{-\infty}^{\infty}ds_1\cdots ds_N\Omega^{(N)}(s^N)\ln\Delta_s^N\Omega^{(N)}(s^N)\\&=-\int_{-\infty}^{\infty}ds_1\cdots ds_N\Omega^{(N)}(s^N)\\&\quad\times\ln\Delta_s^N\Omega^{(1)}(s_1)\cdots\Omega^{(1)}(s_N)\\&\quad\times\Gamma^{(2)}(s_1,s_2)\cdots\Gamma^{(2)}(s_{N-1},s_N)\\&=-N\int_{-\infty}^{\infty}ds_1\Omega^{(1)}(s_1)\ln\Delta_s\Omega^{(1)}(s_1)\\&\quad-N\int_{-\infty}^{\infty}ds_1\int_{-\infty}^{\infty}ds_2\Omega^{(2)}(s_1,s_2)\ln\Gamma^{(2)}(s_1,s_2)\\&=\frac{N}{2}\left[1+\ln\frac{2\pi(1-x^2)}{\Delta_s^2}\right],\end{aligned} \tag{3.22}$$

assuming that $N-1$ can be replaced by N.

This result is a particular example of the general result for a Gaussian probability density obtained by Shannon [1]. In the present language the general result would be written

$$\widetilde{S}(\Delta_s,\Delta_r)=\frac{N}{2}+\ln[(2\pi)^{N/2}\Delta_s^{-N}\sqrt{|K_N|}]. \tag{3.23}$$

The advantage of the present analysis is that it makes clear the relationship between the exact and the approximate entropies; Shannon would have considered his expression for $\widetilde{S}$ to be the entropy of the signal, not just an approximation. In addition, the factor arising from the width of the levels Δ_s^{-N} would have been neglected and the entropy would have appeared to be independent of the sampling and of the digitization. The present model allows all steps of the derivation to be exhibited and a final analytic expression for the entropy because the determinant of the covariance matrix has been obtained explicitly. We now use these explicit results to evaluate the continuum limits.

2. Regime for digitization

We now need to digitize the continuous signal to obtain a result for the exact entropy. The formal expressions for $P^{(n)}$ and $S(\Delta_s,\Delta_r)$ were given above, but the present signal does not yield analytic results for them. Analytic results were obtained for $\widetilde{P}^{(n)}$ and $\widetilde{S}(\Delta_s,\Delta_r)$ and here we evaluate the criteria for the validity of these approximations.

The first criterion was based on the singlet probability density $|\Omega^{(1)}(\Delta_s)-\Omega^{(1)}(0)|\ll\Omega^{(1)}(0)$. For the present normal distribution this reduces to

$$\Delta_s\ll\sqrt{2}. \tag{3.24}$$

The criterion based on neighbor correlations was $|\Omega^{(2)}(0,\Delta_s)-\Omega^{(2)}(0,0)|\ll\Omega^{(2)}(0,0)$. In the present case the digitization is valid when

$$\Delta_s\ll\sqrt{2(1-x^2)}. \tag{3.25}$$

Since in general $x<1$, this supersedes the bound established by the singlet probability. Notice that these bounds limit the width of the levels to be much less than the variance of the signal.

3. Continuum limits

The expression for the entropy of the digitized, sampled, continuous signal $S(\Delta_s,\Delta_r)=\Sigma_{\sigma^N}P^{(N)}(\sigma^N)\ln P^{(N)}(\sigma^N)$ is formally exact and we want to explore its behavior in the continuum limits $\Delta_r\to 0$ and $\Delta_s\to 0$. The expression for the digitized entropy $\tilde{S}(\Delta_s,\Delta_r)=\int ds^N\Omega^{(N)}(s^N)\ln\Delta_s^N\Omega^{(N)}(s^N)$ is an approximation that is valid in the limit $\Delta_s\to 0$. For the present exponentially covariant signal, $\tilde{S}(\Delta_s,\Delta_r)=N[1+\ln 2\pi(1-x^2)/\Delta_s^2]/2$, where $x=\exp[-\gamma\Delta_r]$. We have that $\tilde{S}=S$ when $\Delta_s\ll\sqrt{2(1-x^2)}$. The entropy is also a function of the number of sample points N and of the number of levels L, and in addition to the limits $\Delta_r\to 0$ and $\Delta_s\to 0$, we shall consider the limits $N\to\infty$ and $L\to\infty$. Of these 16 possible combinations of limits the three most important will be the one with $N\Delta_r$ fixed, the one with $L\Delta_s$ fixed, and the combination of these two.

We begin with continuous digitization with fixed sampling $\Delta_s\to 0$ and $L\to\infty$, with $L\Delta_s$, N, and Δ_r fixed. This is the limit where the digitization approximation is valid and

$$S(\Delta_s,\Delta_r)=\tilde{S}(\Delta_s,\Delta_r)\sim -N\ln\Delta_s\sim N\ln L, \tag{3.26}$$

which is just the uncorrelated limit. This is expected since if the variance of successive samples is large compared to the spacing between the levels, then the continuum nature of the signal is lost and it is indistinguishable from a random signal. The quantity $S-N\ln L$ will approach a finite limit as $\Delta_s\to 0$, $L\to\infty$ and hence may be regarded as an effective continuum entropy that may be used as a practical indicator of the sufficiency of the digitization.

A related limit is $L\to\infty$, with Δ_s fixed. In this case the extra levels beyond $\Delta_s L^*=s^*$ add nothing because $\Omega^{(1)}(s)=0$, $s\gtrsim s^*$. Hence, in this particular limit $S\to$ const. Henceforth we shall assume that the limits are taken with $\Delta_s L$ fixed.

The infinite sampling limit is $\gamma\Delta_r\to 0$ or $x\to 1$. We cannot simply insert this into the expression for the digitized entropy because that would give $\tilde{S}\to-\infty$, which cannot be correct since the entropy should be bounded below by zero. The problem is the violation of the digitization criterion and we must simultaneously take $\Delta_s\ll\sqrt{2(1-x^2)}\sim\sqrt{4\gamma\Delta_r}$, and $L\to\infty$ with $L\Delta_s$ fixed. In view of this we can define the continuum limit of the entropy as

$$S(\Delta_s,\Delta_r)=\tilde{S}(\Delta_s,\Delta_r)\to\frac{N}{2}+N\ln\frac{\sqrt{2\pi 4\gamma\Delta_r}}{\Delta_s},$$

$$\Delta_r\to 0,\ \Delta_s\to 0,\ \frac{\Delta_s^2}{\Delta_r}=\text{const}\ll 4\gamma. \tag{3.27}$$

This result holds whether $N\to\infty$ or whether it is fixed.

Now consider the infinite sampling limit $\Delta_r\to 0$, with Δ_s fixed. Obviously $\Delta_s\gg\sqrt{4\gamma\Delta_r}$ and the digitization approximation is not valid. In this case we cannot use $\tilde{S}$, since it is not a valid approximation to the true entropy, and instead we must calculate S and P directly. The first and easiest case is with fixed N, so that $N\Delta_r\to 0$. For a continuous signal $s(r+\Delta_r)=s(r)+O(\Delta_r)$ and hence almost everywhere $P^{(2)}(\sigma_1,\sigma_2)=P^{(1)}(\sigma_1)\delta(\sigma_1,\sigma_2)$, $\Delta_r\to 0$. It follows that

$$P^{(n)}(\sigma^n)\sim P^{(1)}(\sigma_1)\prod_{i=2}^{n}\delta(\sigma_1,\sigma_i),\quad n\gamma\Delta_r\ll 1. \tag{3.28}$$

The entropy in this limit reduces to

$$S(\Delta_s,\Delta_r)\to-\sum_{\sigma}P^{(1)}(\sigma)\ln P^{(1)}(\sigma),$$

$$\Delta_r\to 0,\ N,\Delta_s\ \text{fixed}. \tag{3.29}$$

The limiting result is independent of x and so the criterion for the validity of the digitization approximation now depends only upon $\tilde{P}^{(1)}(\sigma)$, namely, $\Delta_s\ll\sqrt{2}$. If this holds we can explicitly evaluate the entropy

$$\begin{aligned}S(\Delta_s,\Delta_r)&=\tilde{S}(\Delta_s,\Delta_r)\\&=-\int_{-\infty}^{\infty}ds\ \Omega^{(1)}(s)\ln\Delta_s\Omega^{(1)}(s)\\&=\frac{1}{2}+\ln\frac{\sqrt{2\pi}}{\Delta_s},\quad \Delta_r\to 0,\ \Delta_s\ll\sqrt{2},N\ \text{fixed}.\end{aligned} \tag{3.30}$$

In both cases we see that in the infinite sampling limit with fixed level spacing the entropy no longer scales with the number of sample nodes. In essence, as the nodes get closer together they become so highly correlated that consecutive nodes almost always are in the same state.

The second, more realistic, infinite sampling limit has $\Delta_r\to 0$, $N\to\infty$, with $N\Delta_r$ fixed and Δ_s fixed. In this case we obtain the functional form of the limiting result by considering correlations at the pair level. For a continuous signal we expect that successive closely spaced nodes will almost certainly be in the same state, so that $P^{(2)}(\sigma,\sigma)=[1-\epsilon]P^{(1)}(\sigma)$, where $\epsilon\ll 1$, and $P^{(2)}(\sigma,\sigma+1)=\epsilon P^{(1)}(\sigma)/2$. There is no chance of the states of consecutive nodes being separated by more than one level for Δ_s fixed, $\Delta_r\to 0$. The dependence of ϵ on Δ_r follows from

$$\begin{aligned}&P^{(2)}(\sigma,\sigma+1)\\&=\int_{\sigma\Delta_s-\Delta_s/2}^{\sigma\Delta_s+\Delta_s/2}ds_1\int_{\sigma\Delta_s+\Delta_s/2}^{\sigma\Delta_s+3\Delta_s/2}ds_2\frac{1}{2\pi\sqrt{1-x^2}}\\&\quad\times\exp\left[\frac{-1}{2(1-x^2)}(s_1^2-2xs_1s_2+s_2^2)\right]\\&=\frac{e^{-a^2/(1+x)}}{2\pi\sqrt{1-x^2}}\int_0^{\Delta_s}dt_1\int_0^{\Delta_s}dt_2\\&\quad\times\exp\left[\frac{-[t_1^2+2xt_1t_2+t_2^2]}{2(1-x^2)}\right]\exp\left[\frac{2a(t_1-t_2)}{2(1+x)}\right],\end{aligned} \tag{3.31}$$

where $a\equiv\sigma\Delta_s+\Delta_s/2$. As $\Delta_r\to 0$, $x\to 1$ and the integrand is dominated by the region $t_1\lesssim\sqrt{1-x^2}$ and $t_2\lesssim\sqrt{1-x^2}$. In this region the integrand is of order unity and hence the double

TABLE III. Continuum limits of the entropy of the exponentially covariant continuous signal.

Limit	Fixed	$S(\Delta_s,\Delta_r)$
$L\to\infty$	Δ_s, Δ_r, N	$S\to$ const
$\Delta_s\to 0$, $L\to\infty$	$L\Delta_s$, Δ_r, N	$S\sim N\ln L$
$\Delta_s\to 0$, $\Delta_r\to 0$	Δ_s^2/Δ_r, L, N	$S\sim N[1+\ln 8\pi\gamma\Delta_r/\Delta_s^2]/2$
$\Delta_r\to 0$	$\Delta_s\ll\sqrt{2}$, L, N	$S\sim 1/2+\ln\sqrt{2\pi}/\Delta_s$
$\Delta_r\to 0$, $N\to\infty$	$N\Delta_r$, Δ_s, L	$S\sim aN^{1/2}\ln N$

integral has value $O(1-x^2)$. Since the numerator of the prefactor is proportional to $P^{(1)}(\sigma)$, we conclude that

$$\epsilon=O(\sqrt{1-x^2})=O(\sqrt{2\gamma\Delta_r}), \tag{3.32}$$

which confirms that changes of the level are increasingly rare as $\Delta_r\to 0$. On average the number of level changes is $N\Sigma_\sigma P^{(1)}(\sigma)\epsilon=O(N\sqrt{2\gamma\Delta_r})$.

Now consider how the entropy changes if we double the number of nodes by inserting a new node between each existing pair of nodes. If the signal at the nodes on either side of the new one has the same digitized value σ, then the new node almost certainly has the same value (the probability is $1-\epsilon'$, where $\epsilon'\ll\epsilon$). If the new node is inserted between two nodes where the level changes from σ to $\sigma+1$, then it will definitely have one of these values, each with probability 1/2. In other words, the additional uncertainty that arises due to doubling the number of nodes and halving their spacing only depends upon the number of changes of level and each contributes just ln2,

$$S_{2N}-S_N=O(N\sqrt{2\gamma\Delta_r}\ln 2). \tag{3.33}$$

Since in this particular limit $\Delta_r\propto 1/N$, we conclude that

$$S(\Delta_s,\Delta_r)\sim O(N^{1/2}\ln N),$$
$$N\to\infty,\quad \Delta_r\to 0,\ N\Delta_r,\Delta_s \text{ fixed}. \tag{3.34}$$

The reason that we cannot give the numerical value of the constant prefactor is that we did not explicitly evaluate the double integral above. In addition we neglected ϵ', the calculation of which would involve the correlation function for three consecutive sites. Nevertheless, we have obtained the functional form of the entropy and in this continuum sampling limit we conclude that the entropy per node goes to zero. This is because the additional sites provide little new information. The results for the entropy in the various continuum limits are summarized in Table III.

4. Binary digitization

The last continuum limit above ($N\to\infty$, $N\Delta_r$, Δ_s fixed) can be explicitly confirmed for the binary leveled wave with exponential covariance function. Using an asymptotic expansion for small $\gamma\Delta_r$, we have

$$\arcsin K(\Delta_r)=\arcsin[\exp(-\gamma\Delta_r)]\sim\arcsin[1-\gamma\Delta_r]$$
$$\sim\frac{\pi}{2}-\sqrt{2\gamma\Delta_r}+O(\gamma\Delta_r). \tag{3.35}$$

To linear order in $y\equiv\sqrt{2\gamma\Delta_r}/\pi$, the neighbor pair-correlation function is

$$G^{(2)}(\sigma,\tau)\sim 1+\sigma\tau-2\sigma\tau y \tag{3.36}$$

and the neighbor triplet-correlation function is

$$G^{(3)}(\sigma,\tau,\lambda)\sim 1+\sigma\tau+\tau\lambda+\lambda\sigma-2y[\sigma\tau+\tau\lambda+\lambda\sigma\sqrt{2}]. \tag{3.37}$$

Keeping only the first three terms, the Markov entropy expansion is

$$\begin{aligned}S/N=&-\sum_\sigma\rho(\sigma)\ln\rho(\sigma)-\sum_{\sigma,\tau}\rho(\sigma)\rho(\tau)G^{(2)}(\sigma,\tau)\\&\times\ln G^{(2)}(\sigma,\tau)-\sum_{\sigma,\tau,\lambda}\rho(\sigma)\rho(\tau)\rho(\lambda)\\&\times G^{(3)}(\sigma,\tau,\lambda)\ln\frac{G^{(3)}(\sigma,\tau,\lambda)}{G^{(2)}(\sigma,\tau)G^{(2)}(\tau,\lambda)}\\=&\ln 2+\frac{1}{4}\sum_{\sigma,\tau}G^{(2)}(\sigma,\tau)\ln G^{(2)}(\sigma,\tau)\\&-\frac{1}{8}\sum_{\sigma,\tau,\lambda}G^{(3)}(\sigma,\tau,\lambda)\ln G^{(3)}(\sigma,\tau,\lambda).\end{aligned} \tag{3.38}$$

Expanding the pair term to linear order we obtain

$$s_2/N\sim\frac{2}{4}[(2-2y)\ln(2-2y)+2y\ln 2y]\sim\ln 2+y\ln y-y. \tag{3.39}$$

Similarly, the triplet term becomes

$$\begin{aligned}s_3/N\sim&\frac{-2}{8}(4-2y[2+\sqrt{2}])\ln(4-2y[2+\sqrt{2}])\\&+4\sqrt{2}y\ln 2\sqrt{2}y+(4y-2\sqrt{2}y)\ln(4y-2\sqrt{2}y)\\\sim&-2\ln 2+\frac{2+\sqrt{2}}{2}y(1+2\ln 2)-\sqrt{2}y\ln 2\sqrt{2}\\&-\frac{2-\sqrt{2}}{2}y\ln 2(2-\sqrt{2})-\frac{2+\sqrt{2}}{2}y\ln y.\end{aligned} \tag{3.40}$$

Accordingly,

$$\begin{aligned}S(\Delta_s,\Delta_r)\sim&N[1-\ln y+\sqrt{2}\ln 2-(\sqrt{2}-1)\ln(2-\sqrt{2})]\frac{y}{\sqrt{2}}\\\sim&-N\sqrt{\gamma\Delta_r}[a\ln\gamma\Delta_r-b]+O(\gamma\Delta_r),\quad\gamma\Delta_r\to 0.\end{aligned} \tag{3.41}$$

The neglected higher-order terms in the Markovian entropy expansion contribute the numerical value of the constants a and b. We note that if the total sampling time is fixed, then $\Delta_r\propto N^{-1}$ and the entropy of the signal goes like $S\sim N^{1/2}\ln N$, $N\to\infty$, which agrees with the analysis given above for a multilevel signal. That the entropy per sampling node should go to zero is quite reasonable since we can

increasingly predict the state of the next sample the closer it is to the present measurement.

C. Example: Markovian digital correlations

The main virtue of the exponentially covariant model that was solved above was that the correlation densities were Markovian, which meant that it was possible to obtain explicit analytic results for the approximate entropy $\tilde{S}$. The drawback was that this Markovian character was not preserved by the digitization process, and hence the exact entropy could only be evaluated in certain continuum limits, namely, $\Delta_s \to 0$. Hence, in the binary digitized signal we could not obtain an explicit expression for the exact entropy because the Markov expansion does not terminate, as can be seen from the fact that $s_3 \neq 0$. The digitization destroys the Markov character of the original continuous signal, as can be seen already at the three-site level. Recall that the pair-correlation function is

$$g^{(2)}(\sigma_1,\sigma_2;r) \equiv \wp^{(2)}(\sigma_1,\sigma_2;r)/\rho(\sigma_1)\rho(\sigma_2) = 1+\sigma_1\sigma_2\frac{2}{\pi}\arcsin K(r), \tag{3.42}$$

where $\sigma_i = \pm 1$, and the triplet correlation function is

$$g^{(3)}(\sigma_1,\sigma_2,\sigma_3;r_{12},r_{23},r_{31}) = 1+\frac{2}{\pi}[\sigma_1\sigma_2\arcsin K(r_{12}) + \sigma_2\sigma_3\arcsin K(r_{23}) + \sigma_3\sigma_1\arcsin K(r_{31})]. \tag{3.43}$$

Evidently this can only be expressed as a Markov superposition for consecutive equally spaced sites if the covariance function obeys certain properties, namely,

$$[\arcsin K(\Delta_r)]^2 = \frac{\pi}{2}\arcsin K(2\Delta_r). \tag{3.44}$$

The expression holds to first order for exponential covariance as $\gamma \to 0$ (highly correlated limit) and also for a Gaussian covariance function to second order in the same limit. Even if this expression is satisfied, there is no guarantee that Markovian correlations will occur for $n > 3$.

Binary digitization

Rather than digitizing a known continuous signal, in this section we seek the digitized covariance function that will give a Markovian digital signal. For a binary digitized signal $\sigma_i = \pm 1$, a Gaussian probability for consecutive sites in Markov form is

$$P^{(n)}(\sigma^n) = A_n \exp\left[z\sum_{i=1}^{n-1} \sigma_i\sigma_{i+1} \right]. \tag{3.45}$$

No pure quadratic terms are included here because $\sigma_i^2 = 1$ for all configurations. It is evident that this is just the one-dimensional Ising model with nearest-neighbor coupling parameter $-zk_BT$. The normalization constant follows from the reduction formula

$$P^{(n-1)}(\sigma^{n-1}) = \sum_{\sigma_n=\pm 1} P^{(n)}(\sigma^n) = A_n \exp\left[z\sum_{i=1}^{n-2} \sigma_i\sigma_{i+1} \right] \sum_{\sigma_n=\pm 1} e^{z\sigma_{n-1}\sigma_n} = \frac{A_n}{A_{n-1}} P^{(n-1)}(\sigma^{n-1}) 2\cosh z. \tag{3.46}$$

Hence $A_n = A_1 A^{1-n}$, where $A = 2\cosh z$, and

$$A_1 = P^{(1)}(\sigma) = 1/2. \tag{3.47}$$

By construction, the consecutive site correlation function has Markovian factorization

$$G^{(n)}(\sigma^n) \equiv P^{(n)}(\sigma^n)/A_1^n = \prod_{i=1}^{n-1} G^{(2)}(\sigma_i,\sigma_{i+1}). \tag{3.48}$$

What is the digital covariance function that corresponds to the Markovian probability distribution? We have

$$K(n\Delta_r) \equiv \langle \sigma_1\sigma_n \rangle = \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_n=\pm 1} \sigma_1\sigma_n P^{(n)}(\sigma^n) = A^{-1} \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_{n-1}=\pm 1} \times P^{(n-1)}(\sigma^{n-1}) \sigma_1 \sum_{\sigma_n=\pm 1} \sigma_n e^{z\sigma_n\sigma_{n-1}}. \tag{3.49}$$

Now if $\sigma_{n-1} = +1$ the final sum is $2\sinh z$ and if $\sigma_{n-1} = -1$ the final sum is $-2\sinh z$, which gives

$$\langle \sigma_1\sigma_n \rangle = \tanh z \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_{n-1}=\pm 1} P^{(n-1)}(\sigma^{n-1}) \sigma_1\sigma_{n-1} = [\tanh z]\langle \sigma_1\sigma_{n-1} \rangle = [\tanh z]^{n-1}. \tag{3.50}$$

Hence an exponential digital covariance function $K(n\Delta_r) = \exp[-\gamma\Delta_r|n|]$ will give a Markovian consecutive site correlation function for a binary digitized signal, where the decay length is related to the Markovian decay parameter z by

$$e^{-\gamma\Delta_r} = \tanh z. \tag{3.51}$$

The entropy is

$$S = -\sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_N=\pm 1} P^{(N)}(\sigma^N) \ln P^{(N)}(\sigma^N) = -N \sum_{\sigma=\pm 1} P^{(1)}(\sigma) \ln P^{(1)}(\sigma) - (N-1) \sum_{\sigma_1=\pm 1} \sum_{\sigma_2=\pm 1} P^{(2)}(\sigma_1,\sigma_2) \ln G^{(2)}(\sigma_1,\sigma_2)$$

$$= N\ln 2 - N \sum_{\sigma_1=\pm 1} \sum_{\sigma_2=\pm 1} \frac{1}{2A} e^{z\sigma_1\sigma_2} \ln \frac{4}{2A} e^{z\sigma_1\sigma_2}$$

$$= N\ln 2 - N\ln \frac{2}{A} - Nz\langle \sigma_1\sigma_2 \rangle = N\ln[2\cosh z] - Nz\tanh z. \tag{3.52}$$

Note that in the continuum limit $\gamma\Delta_r \to 0$, $z \to -\ln\sqrt{\gamma\Delta_r} \to \infty$. In this limit

$$S/N \sim z\gamma\Delta_r = -\frac{\gamma\Delta_r}{2}\ln\gamma\Delta_r, \tag{3.53}$$

or $S \sim N^{1/2}\ln N$, $N \propto 1/\Delta_r$. This agrees with the limits found at the ends of Secs. III B 3 and III B 4, with the constant prefactor being obtained explicitly in this example. Similar analysis could be carried out beyond the binary level for more general models in which the digital correlations are Markovian (e.g., the Potts and the related spin lattice models of statistical mechanics).

CONCLUSION

We have addressed two main issues in this paper: the value of the entropy of continuous signals and the use of expansions for the information entropy. For a continuous signal we pointed out that sampling and digitization are fundamentally intertwined with the measurement process, which is consistent with the essential dependence of the information entropy on discrete probabilities. The nuance of our analysis was to distinguish between, on the one hand, the formally exact entropy sum that uses the discrete probabilities of a sampled, digitized signal and, on the other hand, the approximation that sets the probability simply proportional to the probability density and approximates the entropy as an integral of the latter. Although the latter procedure generally yields to analytic evaluation and is increasingly accurate as the widths of the digitization levels are decreased, it must be stressed that it is an approximation to the true entropy of the signal. In both cases the information entropy is a function of the sampling and the digitization and it is not possible to speak of *the* information content of a continuous signal independent of the value of these. The analysis of the continuum limit (infinite sampling and digitization) was carried out for the exponentially covariant Gaussian random signal, but the results are likely qualitatively applicable in general.

By expanding the many-site correlation functions as a product of lower-order functions and a correction factor, we obtained formally exact expansions for the information entropy. When we set the remainders beyond a certain order to unity we obtained a superposition approximation for the higher-order correlation functions, which in turn truncated the expansion for the entropy at the same order. We explored two superposition approximations: a generalization of the Kirkwood superposition approximation that is common in liquid-state statistical mechanics and a Markov superposition that is ideally suited for one-dimensional signals and time series. We compared with Monte Carlo simulations of a binary-leveled correlated random signal, using exponential and Gaussian covariance, and we showed that the Markov approximation was superior and only two or three terms of the entropy expansion sufficed for accurate results over the whole regime from low to highly correlated signals. The Markov approximation will work well for a broad class of signals, including periodic or quasiperiodic ones, except possibly in the case of long-period, highly correlated signals.

The success of the Markov approximation is likely due in large part to the one-dimensional nature of the signals that we examined here since it is formulated to take full advantage of the order inherent in this particular geometry. In considering higher-dimensional problems, such as image processing or tomography, it is possible in principle to treat them as a one-dimensional problem resulting from serial sampling and the Markov expansion could simply be applied. Such an approach is unlikely to yield good results because the mapping to one dimension induces long-range correlations with period equal to the length of the scan line and it would be inaccurate to apply the superposition approximation to correlations of lower order than this. We conclude that the generalized Kirkwood superposition expansion will have an advantage for higher-dimensional systems. However, as we pointed out in the text, the generalized Kirkwood superposition approximation fails for highly correlated systems because it overcounts the correlations. What we require is a Markov-type approximation that generalizes the notion of order to higher dimensions. For example, one could approximate the three-site correlation function as the product of the two most highly correlated pair functions, which in practice means the two closest pairs of sites, and analogously for the higher-order correlation functions. The utility of this or other approximations in higher dimensions remain to be explored.

[1] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

[2] E. T. Jaynes, in *Papers on Probability, Statistics, and Statistical Physics*, edited by R. D. Rosenkrantz (Reidel, Dordrecht, 1983).

[3] W. F. Sheppard, Philos. Trans. R. Soc. London, Ser. A **192**, 101 (1898); Proc. R. Soc. London **62**, 170 (1898).

[4] M. Teubner, Europhys. Lett. **14**, 403 (1991).

[5] S. Marčelja, Physica A **231**, 168 (1996).

[6] G. Stell, in *The Equilibrium Theory of Classical Fluids*, edited by H. Frisch and J. L. Lebowitz (Benjamin, New York, 1964).

[7] H. Reiss, J. Stat. Phys. **6**, 39 (1972).

[8] P. Attard, J. Chem. Phys. **102**, 5411 (1995).

[9] J. G. Kirkwood, J. Chem. Phys. **10**, 394 (1942).

[10] I. Z. Fisher and B. L. Kopeliovich, Dokl. Akad. Nauk SSSR **133**, 81 (1960) [Sov. Phys. Dokl. **5**, 761 (1961)].

[11] A. Baranyai and D. J. Evans, Phys. Rev. A **40**, 3817 (1989).

[12] D. C. Wallace, J. Chem. Phys. **87**, 2282 (1987).

[13] S. G. Whittington and L. G. Dunfield, J. Phys. A **6**, 484 (1973).