Physical meaning of the time-correlation length obtained in a computer simulation

Juan J. Morales and María J. Nuevo

Departamento de Física, Facultad de Ciencias, Universidad de Extremadura, 06071 Badajoz, Spain

(Received 2 November 1992)

A comparison is made of three statistical methods of determining the autocorrelation coefficients and correlation length in a time-data series obtained by a computer simulation. Three points of state were chosen where the fluctuations and correlations play different roles: in the liquid, in the solid near melting, and in the melting zone. Taking account of the physical meaning of the results, it is found that there is always a limiting value of the lag at which the series must be cut off.

I. INTRODUCTION

A common question that arises when a mathematical physics topic is to be treated is how to choose the most suitable mathematical method for solving or understanding the physical problem. Once chosen, one must take into account whether the method can be applied fully or with limitations given by physical reasoning. For example, in statistical simulations there is a rule which shows directly the nature of the n computed data points: if the variance of a calculated physical property x is n times the variance of the mean $\bar{x}$, the data are uncorrelated, if not, they are correlated. The strength of the correlation will be reflected in how different those variances are. Mathematically [1]

$$\mathrm{var}(\bar{x}) = \frac{\mathrm{var}(x)}{n}\left[1 + 2\sum_{k=1}^{n-1}\left[1-\frac{k}{n}\right]r_k\right] . \tag{1}$$

The problem arises in the way in which the autocorrelation coefficient (ACC) at lag k, r_k, is defined and therefore which approximation is to be used for the correlating length (CL), τ, of the series, the summation in Eq. (1). The approximation should be justifiable physically (for example, there should not really be any infinite lag correlations).

The statistical methods used are briefly described in Sec. II, Sec. III presents the results for the chosen system, with the conclusions in Sec. IV.

II. THE STATISTICAL METHODS

We shall consider the common situation in which n is a finite number of points and K is a limit which separates the correlated ($r_k \neq 0$ for $k < K$) and uncorrelated ($r_k = 0$ for $k > K$) data points. Our ACC and CL are given, respectively, by

$$r_k = \frac{1}{n-k}\sum_{i-k+1}^{n}\frac{(x_i-\bar{x})(x_{i-k}-\bar{x})}{\mathrm{var}(x_i)} \tag{2}$$

and

$$\tau_M = \sum_{k=1}^{K}\left[1-\frac{k}{n}\right]r_k = \sum_{k=1}^{K}r_k' , \tag{3}$$

where

$$r_k' = \left[1-\frac{k}{n}\right]r_k . \tag{4}$$

Straatsma, Berendsen, and Stam [2] consider the practical application of the case $n \gg K$ when the k/n term in Eqs. (3) and (4) is neglected and the new correlation length τ_S reduces to the summation of r_k, instead of r_k'.

In a recent publication Dietrich and Dette [3] perform the summation for τ_M and τ_S from $k=1$ to $n-1$ instead of assuming the existence of a limiting lag K. They found that the resulting τ_S varies strongly while τ_M is constant (-0.5). They therefore use a different expression for r_k in order to give a better estimate statistically of the correlation [4]

$$\hat{r}_k = \frac{\displaystyle\sum_{i=1}^{n-k}(x_i-\bar{x}_1)(x_{i+k}-\bar{x}_2)}{\left[\displaystyle\sum_{i=1}^{n-k}(x_i-\bar{x}_1)^2\sum_{i=1}^{n-k}(x_{i+k}-\bar{x}_2)^2\right]^{1/2}} , \tag{5}$$

where $\bar{x}_1$ and $\bar{x}_2$ are the means of the first $n-k$ and last k observations of the series $x_1,\dots,x_{n-k},x_{n-k+1},\dots,x_n$, respectively, and the denominators are proportional to the respective standard deviations, σ_1 and σ_2.

According to the case, this estimate may be modified to be a function of the mean of the whole series, giving

$$\tilde{r}_k = \frac{\displaystyle\sum_{i=1}^{n-k}(x_i-\bar{x})(x_{i+k}-\bar{x})}{\left[\displaystyle\sum_{i=1}^{n-k}(x_i-\bar{x})^2\sum_{i=1}^{n-k}(x_{i+k}-\bar{x})^2\right]^{1/2}} . \tag{6}$$

If the data are uncorrelated, $\bar{x}_1=\bar{x}_2=\bar{x}$ so that Eq. (5) reduces to Eq. (6), and in addition, as $\mathrm{var}(x_i)=\sigma^2$, and $\sigma^2=\sigma_1^2+\sigma_2^2$, Eq. (6) becomes Eq. (4). Therefore

$$\hat{r}_k = \tilde{r}_k = r_k' . \tag{7}$$

The expression corresponding to Eq. (3) for the CL's for Eqs. (5) and (6) is

$$(\hat{\tau},\tilde{\tau}) = \sum_{k=1}^{n-1}\left[1-\frac{k}{n}\right](\hat{r}_k,\tilde{r}_k) \tag{8}$$

and

$$\hat{\tau} = \tilde{\tau} = \tau_M \qquad (9)$$

when K in Eq. (3) takes the value $n-1$. If the data are correlated, the ACC and the CL should be different and the magnitude of the difference will depend on the strength of the correlation.

The Dietrich-Dette assumption implies that the correlation has a long-range tail, the points being correlated however large n is.

We will show from our simulation data that this hypothesis, which is only justified mathematically, leads to results not nearly as good as those which are based on physical arguments of a limit to long-term correlations.

III. RESULTS

Computer simulations were performed for a two-dimensional Lennard-Jones (12-6) fluid using the Toxvaerd algorithm [5] and the usual molecular-dynamics simulation of the microcanonical ensemble. Three points of state were chosen: the first in the stable liquid far from any transition, where the fluctuations of the thermodynamic variables are very small and hence one may presume that the data are uncorrelated; the second, in the solid near melting, where spontaneous fluctuations drive the system back and forth between two points of state [6], and finite-range correlations are expected; the third, in the melting zone, with a long-term fluctuation [7], where the correlation is expected to be long range.

The study was focused on comparing our method with the two proposed by Dietrich and Dette [3]. For simplicity, the method of Straatsma, Berendsen, and Stam [2] was omitted because it is a particular case of ours. Further details of the computer simulation and final results can be found in Refs. [6] and [7]. For all the cases, the temperature was the thermodynamic variable chosen to study the correlations. Each temperature value was taken as a subaverage over $800h$, the time step being $h = 0.005(m\sigma^2/\varepsilon)^{1/2}$, where m, σ, and ε are the units of mass, length, and energy in the Lennard-Jones model.

The correlations were calculated without restricting k, i.e., from $k=1$ to $n-1$, as in Ref. [3]. Figures 1, 3, and 5 show the ACC's r_k' [Eq. (4)], $\hat{r}_k$ [Eq. (5)], and $\tilde{r}_k$ [Eq. (6)], for the temperature in the liquid, solid, and melting, respectively, and Figs. 2, 4, and 6 show the corresponding CL's τ_M [Eq. (3)], $\hat{\tau}$, and $\tilde{\tau}$ [Eq. (8)] for the same systems.

From Fig. 1 one can see how, from the beginning, the fluctuations of the ACC are regular, with peaks above and below the zero value, meaning that the experimental data can be considered uncorrelated as was to be expected in principle for the liquid system. The ACC's obtained from the three methods are equal, verifying Eq. (7), but only up to a lag $k=25$ (half the total points). Beyond this value, r_k' converges towards zero whereas the fluctuations of $\hat{r}_k$ and $\tilde{r}_k$ increase, that corresponding to $\hat{r}_k$ being considerably greater. The behavior of the three CL's is quite similar (Fig. 2): they are zero for the first few values of k and then descend to the value -0.5. This constant value is the limit of the series in Eq. (3) when $k=n-1$, as deduced mathematically in Ref. [3]. But it is also the limit for the series in Eq. (8) if the data are uncorrelated, verifying the identities of Eq. (9) with the
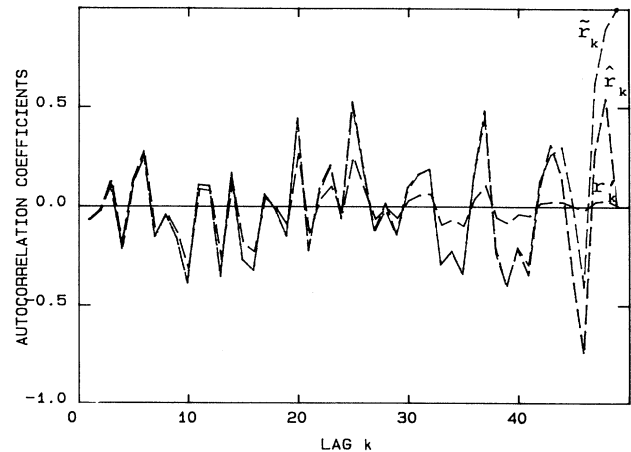

FIG. 1. Autocorrelation coefficients r_k', $\hat{r}_k$, and $\tilde{r}_k$ for the liquid system ($kT/\varepsilon = 1.0$ and $\rho\sigma^2 = 0.7937$) vs lag k.

same limitation $k=25$ as for Eq. (7). Because of the clearness of the results and the similarity of the curves, it was not necessary to run the system any longer.

The results for the correlation in the solid near melting are quite different. It was found that the system spontaneously melts and resolidifies twice, at $10\,000h$ and $40\,000h$. In the solid the ACC's (Fig. 3) start from a maximum value and fall sharply within the few first values of k, with no difference between the three methods. Then there appears a second smooth maximum for $k=14$, where the methods begin to separate. This maximum corresponds to the first time at which the solid system melts and resolidifies. After that, the methods behave differently: r_k' remains negative and very close to zero with small fluctuations, $\hat{r}_k$ take positive and negative values with large fluctuations, and $\tilde{r}_k$ breaks down, giving meaningless large negative values. The second time the system melts, at about $k=55$, is reflected by the higher peaks in the fluctuations for all three methods but none of them manifest any true correlation. The behavior of the CL's shown in Fig. 4 is very clear: τ_M and $\tilde{\tau}$ reach a
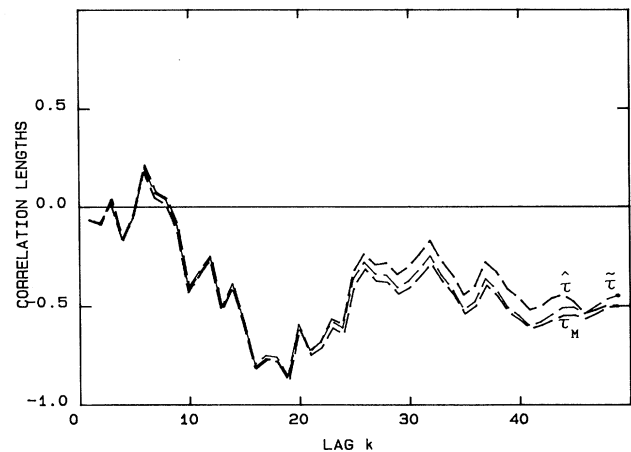

FIG. 2. Correlation lengths τ_M, $\hat{\tau}$, and $\tilde{\tau}$ vs lag k in the liquid system.

FIG. 3. The same as Fig. 1 for the solid system ($kT/\varepsilon = 1.0$ and $\rho\sigma^2 = 0.9048$).
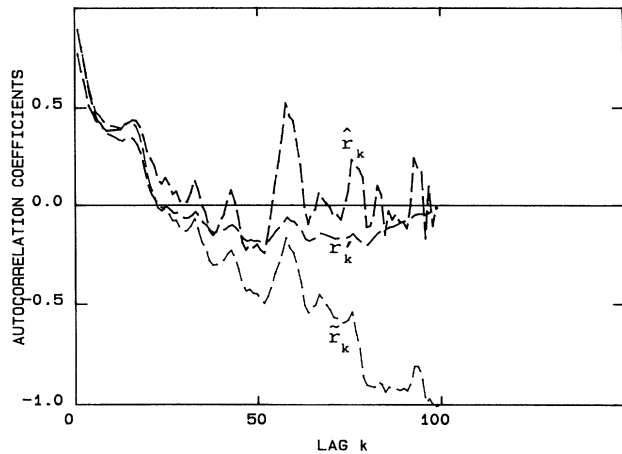

FIG. 5. The same as Fig. 1 for the melting system ($kT/\varepsilon = 0.7$ and $\rho\sigma^2 = 0.84$).

maximum at about $k = 23$, the value at which r_k' and $\tilde{r}_k$ pass through zero in Fig. 3, and then fall becoming negative towards the end, $\tilde{\tau}$ being more negative than τ_M. The other CL $\hat{\tau}$ behaves differently, reaching a higher and later maximum and maintaining a plateau at this value.

The correlations for the melted system have yet another kind of behavior. In the melting zone, the time evolution of the temperature of the system shows large periodic fluctuations. The ACC's for this system (Fig. 5) reflect this with all three methods behaving alike until $k = 40$ (1/3.5 of the total number of points). There is then a second part until $k = 80$ during which they show a similar behavior with slightly different values. But from this point to the end, the fluctuations of r_k' tend smoothly to zero, whereas the fluctuations of $\hat{r}_k$ and $\tilde{r}_k$ around zero become much larger, those corresponding to $\hat{r}_k$ being greater as was also the case in the liquid system. The CL's for this system are shown in Fig. 6. The minimum at around $k = 20$ corresponds to where the ACC's become negative in Fig. 5. The behavior of the three methods is similar until $k = 75$, after which $\hat{\tau}$ climbs to a positive plateau at about 2.0. The other two CL's stay
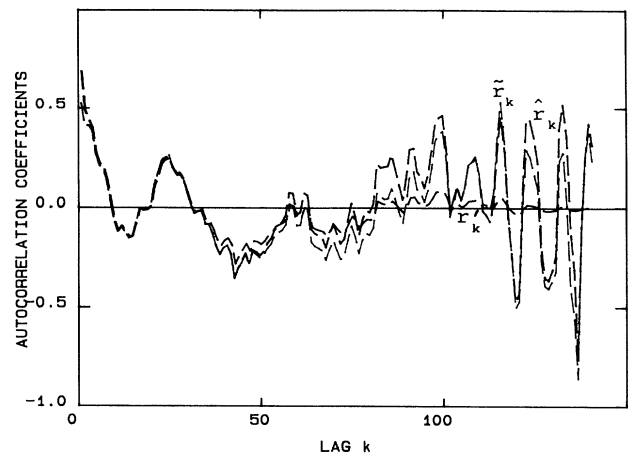
very close to each other, $\tilde{\tau}$ approaching zero and τ_M -0.5, as was expected.

IV. CONCLUSIONS

The three physical cases studied clearly contradict Dietrich and Dette's hypothesis that there exists no maximum lag k. When there is apparently no maximum in some given case, this is not because the data are all of them correlated, but because they are uncorrelated from the beginning (hence $K = 0$) as was seen in the liquid (Figs. 1 and 2). When the system has regular fluctuations, but with some spontaneous jumps, as in the solid, all the three methods show a short-range correlation before the method proposed for $\tilde{r}_k$ breaks down (Fig. 3) and its large fluctuations give rise to a large fictitious permanent correlation length (Fig. 4). When the system shows oscillating fluctuations, as in the melting zone, there exist long trend correlations which, for τ_M and $\tilde{\tau}$, die out far before reaching the last value of $k = n - 1$ (Figs. 5 and 6), and for the $\hat{\tau}$ reach a plateau of about 2.0 denoting again a smaller, but false, correlation.
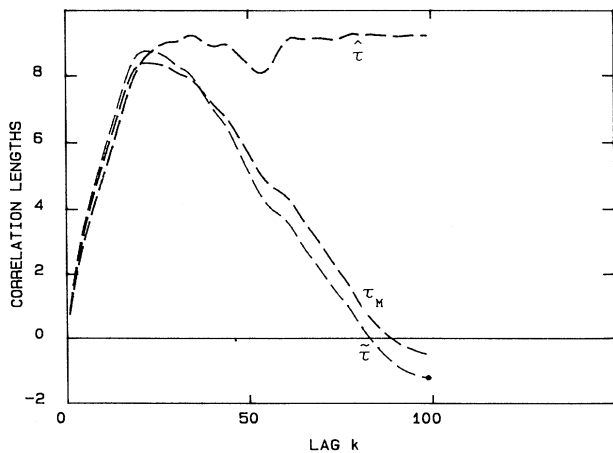

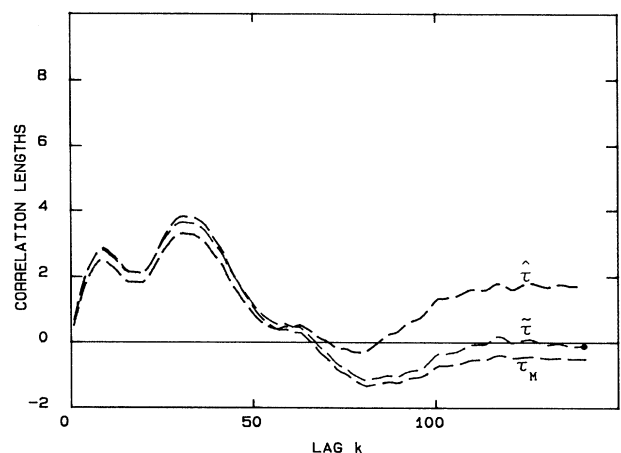FIG. 4. The same as Fig. 2 for the solid system.

FIG. 6. The same as Fig. 2 for the melting system.

In the method proposed by Dietrich and Dette, one fact is clear: beyond a certain value of k (depending on the system), the fluctuations in the ACC become larger as k increases and may therefore give not only a false value for the CL, but a fictitious permanent correlation in the data. The explanation is to be found in the way in which the summations in Eqs. (5), (6), and (8) are carried out. For $k = 1$ the ACC's have $n - 1$ terms and the CL one term; for $k = 2$, the ACC's have $n - 2$ terms and the CL two terms, etc. That is, for each unit increase in k, the number of terms which determine the ACC's decrease by one. The corresponding statistical noise thus also increases. When finally $k = n - 1$, there is only one term in the ACC (and $n - 1$ in the CL) which means physically that the first experimental value is correlated with the last, omitting the other $n - 1$ points, independently of how large n is. The CL's, which are the sum of the ACC's, collect all these noises and can be completely masked. Therefore one must eliminate the noise in the ACC and CL by choosing an appropriate cutoff in the lag k. This can be done by simple observation of a plot of the ACC or CL versus lag k, as in Figs. 1–6.

ACKNOWLEDGMENT

[1] J. J. Morales, M. J. Nuevo, and L. F. Rull, J. Comput. Phys. **89**, 432 (1990).

[2] T. P. Straatsma, H. J. C. Berendsen, and A. J. Stam, Mol. Phys. **57**, 89 (1986).

[3] S. Dietrich and H. Dette, J. Comput. Phys. **101**, 224 (1992).

[4] M. G. Kendal, *Time Series* (Griffin, London, 1973); G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications* (Holden-Day, San Francisco, 1968).

[5] S. Toxvaerd, J. Comput. Phys. **47**, 444 (1982).

[6] L. F. Rull, J. J. Morales, and F. Cuadros, Phys. Rev. B **32**, 6050 (1985).

[7] J. J. Morales, F. Cuadros, and L. F. Rull, J. Chem. Phys. **86**, 2960 (1987).