Statistics of correlation functions from molecular dynamics

I. Bitsanis, M. Tirrell, and H. Ted Davis
University of Minnesota, Chemical Engineering and Materials Science Department, Minneapolis, Minnesota 55455
(Received 9 March 1987)

The statistical uncertainty in the calculation of velocity autocorrelation functions and self-diffusion coefficients from molecular dynamics is empirically determined from the spread of the molecular-dynamics results for an ensemble of macroscopically identical systems. The "experimental" uncertainties of the velocity-autocorrelation-function values at equilibrium and in the presence of flow agree well with theoretical predictions. The uncertainty of the self-diffusion coefficient is found to decrease as the inverse square root of the averaging time.

The calculation of transport coefficients from molecular dynamics (MD) simulations typically employs the formulas of linear-response theory[1] that relate these quantities to the time integrals of the corresponding current correlation functions, the so-called Green-Kubo relations. It is, therefore, essential to know how accurately the correlation functions can be determined from a simulation of specified duration. The time correlation function of a dynamical quantity $A(t)$ is defined as

$$C_\infty(\tau)=\langle A(t)A(t+\tau)\rangle \ . \tag{1}$$

The average denoted by $\langle\ \rangle$ is an ensemble average over an equilibrium distribution. For ergodic systems the ensemble average is equal to an infinite time average of a single system, namely,

$$C_\infty(\tau)=\lim_{T\to\infty} C_T(\tau) \ , \tag{2}$$

where

$$C_T(\tau)=T^{-1}\int_0^T dt\ A(t)A(t+\tau) \ . \tag{3}$$

In the language of stochastic processes the function $\phi(\tau)=A(t)A(t+\tau)$ is a sample function of a stochastic process $\phi(\tau)$ and has an implicit dependence on the initial point in the phase space. For stationary processes $\phi(\tau)$ is independent of t. We see, therefore, that $C_T(\tau)$ and $C_\infty(\tau)$ are not actually numbers but random variables. The ergodic character (ergodicity in the mean) essentially states that the variance of the random variable $C_\infty(\tau)$ vanishes.[2]

What one calculates in MD, however, is $C_T(\tau)$ rather than $C_\infty(\tau)$ and the question is how good an estimate of $C_\infty(\tau)$ is the finite time average $C_T(\tau)$. The answer to this question requires the calculation of the variance of $C_T(\tau)$ which, unlike $C_\infty(\tau)$, does not vanish even for ergodic processes. This calculation in general involves the fourth-order moments of the dynamical variable $A(t)$ about which almost no information can be extracted from the simulation itself.

The first and most useful study of this question was presented by Zwanzig and Ailawadi.[3] Using the assumption that $A(t)$ is a Gaussian random variable (which is often but not always true), they expressed the fourth-order moments of $A(t)$ in terms of second-order moments. By defining a relaxation time for $C_\infty(\tau)$,

$$t_r=2\int_0^\infty d\tau[C_\infty(\tau)/C_\infty(0)]^2 \ , \tag{4}$$

they arrived at the result

$$\langle\Delta(\tau_1)\Delta(\tau_2)\rangle\approx(2t_r/T)[C_\infty(0)]^2 \ , \tag{5}$$

where $\Delta(\tau)=C_T(\tau)-C_\infty(\tau)$. Frequently in MD simulations one calculates

$$R_T(\tau)=C_T(\tau)/C_T(0) \ . \tag{6}$$

Zwanzig and Ailawadi's estimate for the standard deviation of this random variable [uncertainty of $R_T(\tau)$] is

$$\sigma_R=(2t_r/T)^{1/2}[1-R_\infty(\tau)] \ , \tag{7}$$

where $R_\infty(\tau)=C_\infty(\tau)/C_\infty(0)$.

A more general analysis that does not employ the Gaussian assumption for $A(t)$ was presented by Picinbono.[4] Similar estimates of the uncertainty σ_R have been worked out for the correlation functions of dynamical variables that are Poisson processes[5] and for orientational correlation functions.[6]

Since its publication Eq. (7) has been used by many authors to estimate the uncertainty of the time correlation functions calculated from MD simulations. Surprisingly, however, no detailed comparison of the predictions of Eq. (7) with simulation data has appeared up to date, except from an early, rather limited one.[7]

The purpose of this communication is to present such a detailed comparison. The data from an equilibrium MD simulation of saturated liquid Ar at 120 K have been used for this purpose. The simulation, which was performed in another context,[8] covered a real time interval of 0.6 nsec (60 000 time steps). The trajectory was divided into six parts of duration 0.1 nsec each and the autocorrelation functions of the x and y components of the velocity were calculated for the last five subintervals. In this way ten velocity autocorrelation functions (VACF's) were available. From the sample of ten values $C_{T,i}(\tau)$ an empirical standard deviation was determined,

$$\sigma_R^e = \left[\frac{1}{N-1} \sum_{i=1}^{N} [R_{T,i}(\tau) - \bar{R}_T(\tau)]^2 \right]^{1/2}, \qquad \bar{R}_T(\tau) = \frac{1}{N} \sum_{i=1}^{N} R_{T,i}(\tau), \quad N = 10 . \tag{8}$$

The empirical standard deviation σ_R^e must be compared with $\sigma_R / n^{1/2}$ [see Eq. (7)], where n is the number of particles in the simulation ($n = 108$ in our case). This is so because the VACF's of the system are the averages of the VACF's of the individual particles.[3] As we can see from Fig. 1 the theoretical prediction (solid line) is a good order-of-magnitude estimate of the actual uncertainties, especially if one allows for the poor statistics of the experimental uncertainty estimate. The statistical error of the experimental uncertainties was tentatively determined from the difference between the experimental uncertainties that result if we use only the x or only the y VACF's in Eqs. (8). The dotted lines in Fig. 1 show the error limits. The theoretical uncertainty lies almost everywhere between these limits. In particular, the agreement is very good for short times. For $\tau > 1.2$ psec the magnitude of the VACF itself is similar or smaller than the uncertainty. Therefore, a 0.1 nsec simulation is unable to provide any quantitative information for the long-time decay of the VACF's.

As it is customary in error analysis we replaced $C_\infty(\tau)$ in Eqs. (4) and (7) by

$$\bar{C}_T(\tau) = \frac{1}{N} \sum_{i=1}^{N} C_{T,i}(\tau) , \tag{9}$$

which is determined from simulation data. In this way we calculated the theoretical curve (solid line) in Fig. 1. Normally, however, only one sample curve is available and the usefulness of the theoretical prediction consists in the fact that it remains a good estimate of the actual uncertainty even if one of the $C_{T,i}(\tau)$, instead of the average $\bar{C}_T(\tau)$, is used as an input in Eqs. (4) and (7). To demonstrate this point the theoretical curve that results from Eqs. (4) and (7) if we use as input the sample function $C_{T,j}(\tau)$ that deviates the most from the average VACF $\bar{C}_T(\tau)$ was also calculated. This curve is not shown in Fig. 1 since it is practically identical with the theoretical curve (solid line) which results if $\bar{C}_T(\tau)$ is used in Eqs. (4) and (7).

In Fig. 2 a similar comparison is presented for the same system. The system now is not at thermodynamic equilibrium but undergoes Couette flow[8] in the x direction, which destroys the symmetry on the x,y plane. As shown in Ref. 8, however, for the shear rate of the simulation no systematic difference between the x and y VACF's was found. For this reason we consider the x and y VACF's to be identical. The x component of the velocity is of course the peculiar velocity in the x direction, i.e., the difference of the actual particle velocity from the flow velocity at the location of the particle. The data from the 12 subintervals of a 1.2 nsec trajectory were used in Eqs. (4), (7), and (8) in order to calculate the theoretical and experimental uncertainties of the x and y VACF's. Thus 24 sample VACF's were available in this case. For this nonequilibrium system the distribution of the x and y velocities is no longer Gaussian, although the velocity relative to the mean-flow velocity does not deviate much from the Gaussian form. We see from Fig. 2 that the agreement is as good as in the equilibrium case which suggests that the success of the theory is not overly sensitive to the Gaussian assumption.

We finally report the observed dependence of the uncertainty of the diffusivity on the averaging time T. Since the diffusivity in the x or y direction is related to the VACF's via[1]

$$D_{x_i} = \int_0^\infty C(\tau) d\tau, \quad x_i = x \text{ or } y \tag{10}$$

the predictions of Refs. 3 and 4 are not directly applic-

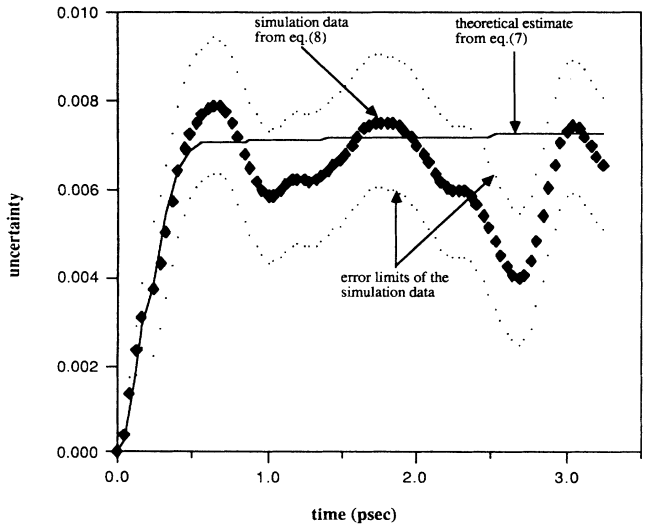

FIG. 1. Theoretical and experimental uncertainties of the VACF's for the equilibrium system.

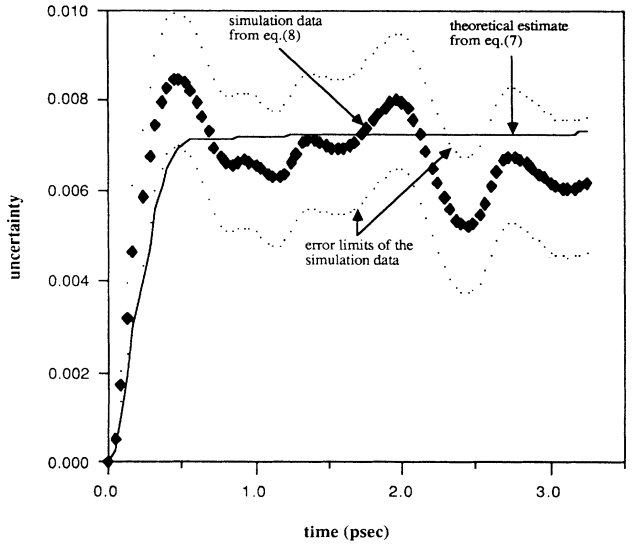

FIG. 2. Theoretical and experimental uncertainties of the VACF's for the Couette-flow system.

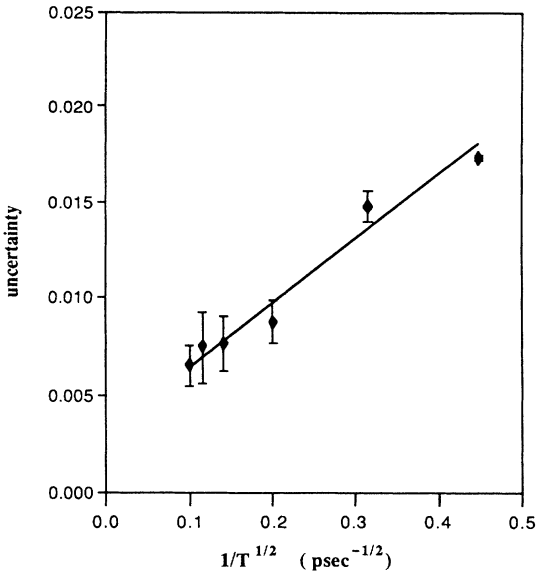

FIG. 3. Diffusivity uncertainty for the equilibrium system.

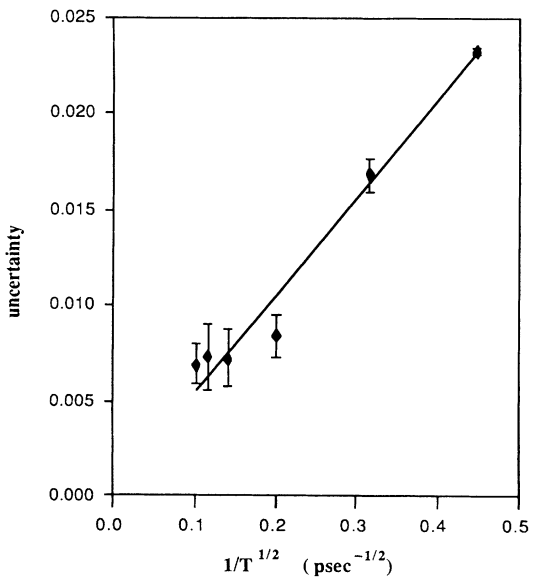

FIG. 4. Diffusivity uncertainty for the Couette-flow system.

able. A rigorous prediction is available[6] for systems of particles that undergo Brownian (diffusive) motion instead of deterministic (classical) motion. According to Ref. 6 the uncertainty of the translational diffusivity σ_D goes as

$$\sigma_D \sim T^{-1/2} . \tag{11}$$

In Refs. 3, 4, and 5 a similar result was found for the uncertainty of the normalized VACF's, namely,

$$\sigma_R \sim T^{-1/2} . \tag{12}$$

Although Eq. (12) was initially expected to be valid for Gaussian dynamical variables[3] its validity was proven for other classes of dynamical variables also.[4,5] The findings in Refs. 3–6 suggest that Eq. (11) might be true for deterministic systems as well, although exceptions are known to exist (e.g., for hard disks or hard square systems[3,9,10]). To test Eq. (11) the diffusivities from the five subintervals of the equilibrium simulation and the twelve subintervals of the Couette-flow simulation were calculated using six different averaging times T_l. These times were $T_1=5$ psec, $T_2=10$ psec, $T_3=25$ psec, $T_4=50$ psec, $T_5=75$ psec, $T_6=100$ psec. For each T_l, 10 equilibrium diffusivities and 24 flow diffusivities were available. An experimental standard deviation was determined for each T_l,

$$\sigma_{D_l}=\left[\frac{1}{N-1}\sum_{i=1}^{N}(D_{i,l}-\bar{D}_l)^2\right]^{1/2},$$
$$\bar{D}_l=\frac{1}{N}\sum_{i=1}^{N}D_{i,l}, \tag{13}$$

where $N=10$ for the equilibrium system and $N=24$ for the system undergoing Couette flow and i refers to the subinterval of the simulation trajectory. Consequently the six data points σ_{D_l} versus $T_l^{-1/2}$ were fitted to a straight line. The result is shown in Fig. 3 for the equilibrium system and in Fig. 4 for the system undergoing flow. The data follow Eq. (11) reasonably closely (the correlation coefficients of the linear fits are 0.985 and 0.984, respectively). The error limits were again determined from the difference between the uncertainty values that result if only the x or only the y diffusivities are used in Eq. (13).

A direct test of Eq. (12) was also made. Although we do not present the detailed results here we mention that the data fit closer a straight line for most of the 80 different τ examined than the data for the diffusivities do.

From the comparisons we presented we conclude that the theory of Zwanzig and Ailawadi[3] offers a very good order-of-magnitude estimate of the correlation functions uncertainty when these are calculated via the finite time-averaging procedure employed in MD simulations. This is true for both systems having a Gaussian velocity distribution and a slightly different non-Gaussian velocity distribution. Furthermore, our results provide strong evidence that the uncertainty of the diffusivity reduces as the inverse square root of the averaging time in these simulations.

This research was supported in part by grants from the National Science Foundation (NSF-DMR-8115733), the Department of Energy, and the University of Minnesota Supercomputer Institute.

[1]R. Zwanzig, Ann. Rev. Phys. Chem. **16**, 67 (1965).
[2]A. Papoulis, *Probability Random Variables and Stochastic Processes* (McGraw-Hill, New York, 1965).
[3]R. Zwanzig and N. K. Ailawadi, Phys. Rev. **182**, 280 (1969).
[4]B. Picinbono, Phys. Rev. A **16**, 2174 (1977).
[5]N. Go and F. Kano, J. Chem. Phys. **75**, 4166 (1981).
[6]S. Lee and M. Karplus, J. Chem. Phys. **81**, 6106 (1984).
[7]N. K. Ailawadi, Phys. Rev. A **5**, 1968 (1972).
[8]I. Bitsanis, J. J. Magda, M. Tirrell, and H. T. Davis, J. Chem. Phys. (to be published).
[9]B. J. Alder and T. E. Wainwright, Phys. Rev. A **1**, 18 (1970).
[10]C. Carlier and H. L. Frisch, Phys. Rev. A **6**, 1153 (1972).