PHYSICAL REVIEW E VOLUME 56, NUMBER 4 OCTOBER 1997

Gaussian statistics of the hard-sphere fluid

Gavin E. Crooks and David Chandler
Department of Chemistry, University of California, Berkeley, California 94720

(Received 30 May 1997)

By computer simulation of the hard-sphere fluid, we have determined the probabilities of observing N molecular centers within molecular sized volumes of the fluid. These probability distributions are found to be almost exactly Gaussian at medium densities. A maximum entropy prediction constructed from knowledge of the first two moments of the distribution and a prior distribution consistent with ideal gas behavior quantitatively predicts the occupation distribution for low and medium densities. [S1063-651X(97)07110-9]

PACS number(s): 61.20.Gy, 61.20.Ja, 61.20.Ne, 89.70.+c

I. INTRODUCTION

In a novel application of a computer simulation of liquid water, Hummer *et al.* [1] examined the probabilities of observing N water molecules within molecular sized volumes They showed that these distributions are almost exactly Gaussian. Thus they were able to use information theory [2–4] and the oxygen-oxygen radial distribution function to predict the probability of cavity formation in water. Their approach provides a quantitative theory for hydrophobic chemical potentials and the solute-solute potentials of mean force from experimentally accessible information.

The approach is closely related to scaled particle theory [5–8], and to the Gaussian field models of fluids [9]. The Pratt–Chandler theory of hydrophobicity [10] is based upon the Gaussian model.

In this paper, we carry out an analysis similar to that of Ref. [1], but for the case of a hard-sphere fluid — the ideal fluid that provides a useful reference system for statistical perturbation theories [11]. We show that away from very low and very high fluid densities, the particle occupation distributions for hard-spheres in molecular sized volumes are nearly Gaussian, consistent with the finding in Ref. [1] for liquid water. Therefore, to a good approximation, for both water and simple hard-sphere fluids the effective Hamiltonians are bilinear in density fields. As such, approximate normal modes of these fluids are linear functionals of these fields.

In the next section, we present our computer simulation results. Then, in Sec. III we review aspects of information theory and its application to the current problem. Deviations from Gaussian statistics at low densities can be accounted for by employing a prior distribution that is consistent with ideal gas behavior. This correction allows quantitative predictions of the occupation distributions, and therefore chemical potentials, at low to medium densities using the average density and pair distribution function of the fluid.

II. COMPUTER SIMULATIONS

By Monte Carlo simulation, we have computed the probability, $P(N;v)$, that N hard-spheres reside in a volume v. In general, one could consider volumes of arbitrary shape and size. Here, we have limited our considerations to spherical volumes of various diameters as large as $d=4\sigma$, where σ is the diameter of the hard-sphere solvent particles. A cavity of diameter d can hold a solute particle of diameter $d-\sigma$. Data were collected for reduced densities, $0.1\leqslant\rho^*=\rho\sigma^3\leqslant0.9$, where ρ is the average particle density in the simulation box. To avoid the possibility of finite size effects, we simulated a reasonable large system—a periodic box with side length 32σ containing $32\,768\rho^*$ particles. This box is approximately 1000 times larger than the largest volume considered. Approximately 150×10^6 data points were collected for each volume at each density.

Data from the simulation are shown in Figs. 1 and 2. The points are plotted directly from the simulation data. For reference, the solid lines are pure Gaussians with the same mean and standard deviations, which are defined by summation over positive N (rather than integration over $\pm\infty$). Figure 1 shows the occupation probabilities for spherical volumes ranging from $d=2.0\sigma$ to $d=4.0\sigma$ for a density $\rho^*=0.50$. It is evident that at this density the distributions are very nearly Gaussian for all volume sizes considered. The smallest volume, $d=2.0\sigma$, corresponds to the exclusion volume of the hard-sphere fluid itself. The largest, $d=4.0\sigma$, has a volume 64 times greater than that of a hard-sphere particle.

In contrast, Fig. 2 shows the occupation probabilities for a

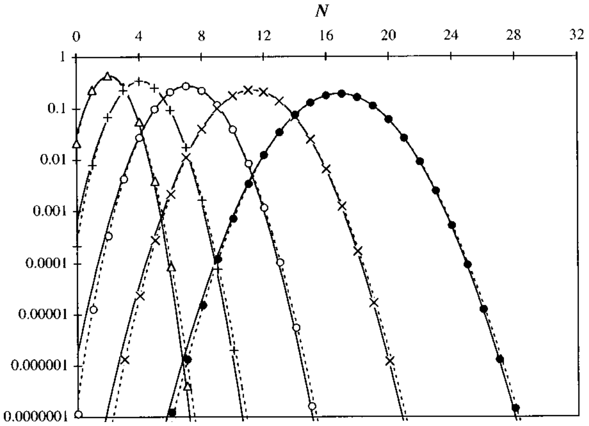

FIG. 1. Probabilities $P(N;v)$ of observing N hard-sphere (diameter σ) centers in a spherical volume of diameter 2.0σ (△), 2.5σ (+), 3.0σ (○), 3.5σ (×), or 4.0σ (●) at a density of $\rho^*=0.50$. Points are from computer simulation. Lines are the two-moment maxent fit to the data using either an uninformative (—) or ideal gas (- - -) prior.

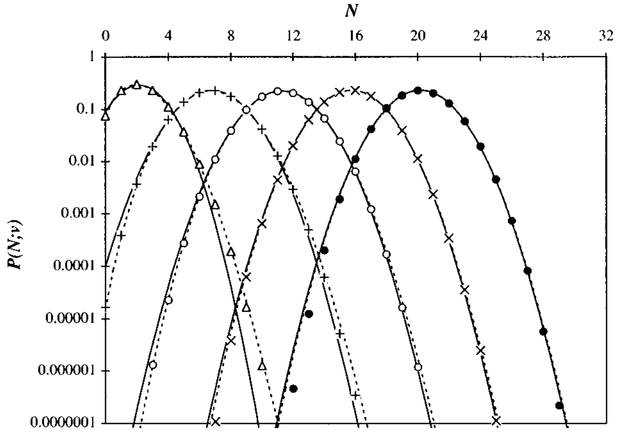

FIG. 2. Probabilities $P(N;v)$ of observing N hard-sphere (diameter σ) centers in a spherical volume of diameter 3.5σ and at densities of $\rho^*=$ 0.1 (△), 0.3 (+), 0.5 (○), 0.7 (×), or 0.9 (●). Points are from computer simulation. Lines are the two-moment maxent fit to the data using either an uninformative (—) or ideal gas (- - -) prior.

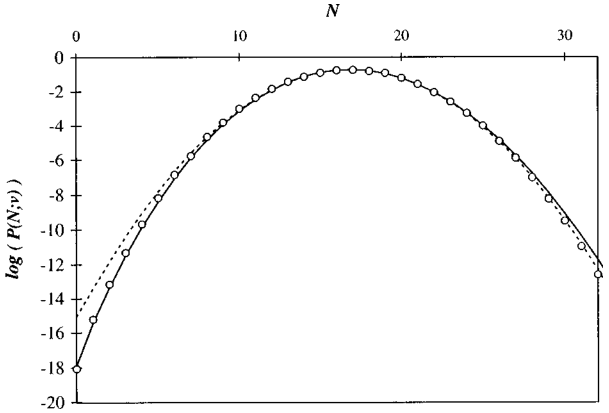

FIG. 3. Logarithm (base 10) of the probabilities $P(N;v)$ of observing N hard-sphere (diameter σ) centers in a spherical volume of diameter 4.0σ and at a density of $\rho^*=0.5$. Points (○) are from computer simulations using an umbrella sampling technique. Lines are the two-moment maxent fit to the data using either an uninformative (—) or ideal gas (- - -) prior.

volume of fixed diameter ($d=3.5\sigma$) at densities ranging from $\rho^*=0.10$ to 0.90. At medium densities the distributions are almost exactly Gaussian. At very high densities, $\rho^*=0.90$, deviations from pure Gaussian behavior are evident. This is to be expected, since at slightly larger densities, $\rho^*=0.94$, the hard-sphere fluid undergoes an order–disorder phase transition [12,13]. At low densities deviations from simple Gaussian behavior are also evident. At these densities the fluid behaves more like an ideal gas, which, for small volumes, is not well described by Gaussian occupation distributions.

Straightforward simulations can only provide data for small fluctuations, as large fluctuations are too improbable to be observed. To gather statistics for the wings of the distribution, we have used umbrella sampling [14] with a periodic simulation box of side length 16σ containing $4096\rho^*$ particles. An area of the simulation is selected as our volume, and constrained to contain only N or $N+1$ particle centers, where N is varied over the relevant occupation range. As the ratio, $P(N;v)/P(N+1;v)$, is the same for the constrained and unconstrained system, we can reconstruct the true distribution from these probability ratios.

Figure 3 shows the probability distribution calculated using umbrella sampling for a reduced density of 0.5, and a volume diameter of 4σ. The solid line is a Gaussian, which closely matches the simulation data for small fluctuations. At large fluctuations, deviations from Gaussian behavior are evident at both high and low occupation numbers.

The central moments

$$B_m(v)=\langle(N-\langle N\rangle_v)^m\rangle_v \qquad (1)$$

were calculated directly from the simulation data. The notation $\langle\cdots\rangle_v$ indicated an average with the distribution $P(N;v)$, the probability of finding N molecules in the volume v. In Fig. 4 we can see that the variance, $B_2(v)$, is largely independent of the density. At high densities small oscillations in the variance exhibit the granularity of the fluid. The third central moment, $B_3(v)$, is shown in Fig. 4(b). The oscillations with volume size are evident. For a Gaussian model, $B_3(v)=0$, and therefore such a model cannot accommodate these molecular level details.

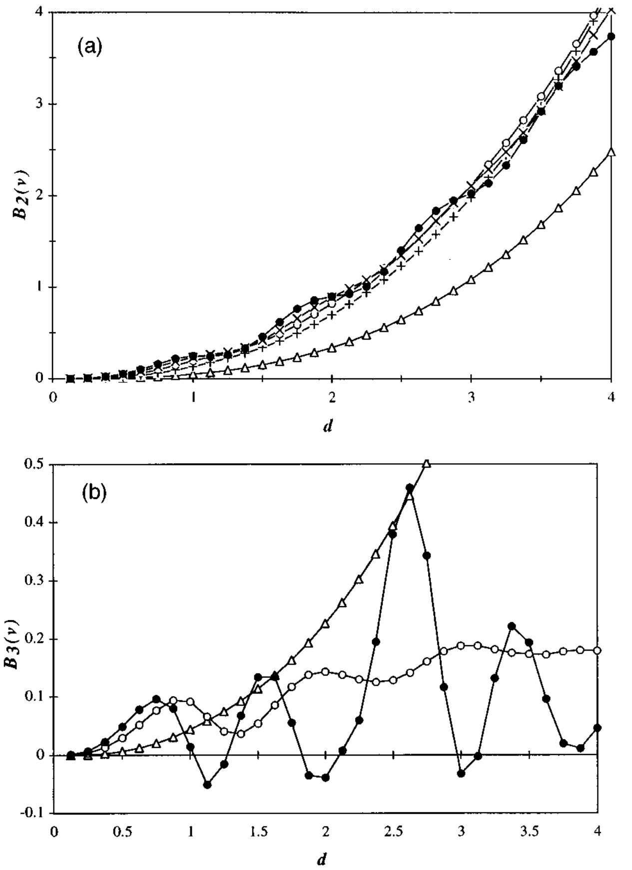

FIG. 4. (a) Variance, $B_2(v)$, for spherical volumes, $v=\pi d^3/6$, at five reduced densities: ρ^* = 0.1 (△), 0.3 (+), 0.5 (○), 0.7 (×), 0.9 (●). (b) Third central moment $B_3(v)$ for spherical volumes, $v=\pi d^3/6$, at three reduced densities: ρ^* = 0.1 (△), 0.5 (○), 0.9 (●). Lines are guides to the eye.

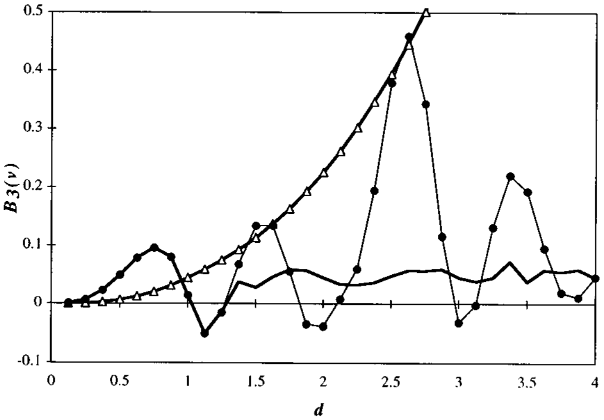

FIG. 5. Third central moment, $B_3(v)$, for spherical volumes, $v=\pi d^3/6$, at two reduced densities: $\rho^*=0.1$ (△) and 0.9 (●). The heavy black lines are calculated from the ideal gas prior two-moment maxent fit. The light line is a guide to the eye for the simulation data at $\rho^*=0.9$.

III. INFORMATION THEORY

For a hard-sphere system, the effect of introducing a solute particle into the fluid is to exclude the solvent from some volume v. The excess chemical potential of the solute, μ^{ex}, is the work required to form this cavity in the solution. The excess solute chemical potential is directly related to the probability that this volume is devoid of particles via

$$\beta\mu^{\mathrm{ex}}=-\ln P(0;v), \tag{2}$$

where $\beta=1/k_BT$, k_B is Boltzmann's constant and T is the temperature.

The moments of this probability distribution are related to the n-particle molecular correlation functions, $g^{(n)}(\mathbf{r}_1,\ldots,\mathbf{r}_n)$ [15]. Specifically,

$$\alpha_k=\langle N(N-1)\cdots(N-k+1)\rangle_v \tag{3}$$

$$=\rho^k\int_v\cdots\int_v g^{(k)}(\mathbf{r}_1\cdots\mathbf{r}_k)d\mathbf{r}_1\cdots d\mathbf{r}_k. \tag{4}$$

Probability theory provides the following inversion formula to obtain the individual probabilities [16,17]:

$$P(N;v)=\sum_{k=N}^{\infty}(-1)^{k-N}\frac{\alpha_k}{N!(k-N)!}. \tag{5}$$

Knowledge of the average density, ρ, and the pair distribution function, $g^{(2)}(|\mathbf{r}_1-\mathbf{r}_2|)$, allows the calculation of the first two moments, $\langle N\rangle_v$ and $\langle N^2\rangle_v$. The use of the inversion formula, Eq. (5), requires the knowledge of all moments of the distribution. The almost Gaussian behavior seen in the simulation data, however, suggests that the first two moments are sufficient to provide a good guess at the true distribution. This approach provides a semi-empirical theory of solvation since, for real fluids, average density and pair distributions can be determined experimentally. Indeed, Pratt and co-workers have stressed this point in their work on hydrophobicity [1,18].

Information theory [2–4] provides a systematic method of reconstructing a least biased best guess for the probability distribution given the available information. To do so we maximize the entropy

$$S=-\sum_N P(N;v)\ln P(N;v) \tag{6}$$

of the distribution with respect to the known constraints on the system, $\Sigma_N P(N;v)=1$ and $\Sigma_N N^m P(N;v)=\langle N^m\rangle_v$. If the first through mth moments are known we are led to the functional form

$$P(N;v)=\exp(\lambda_0+\lambda_1 N+\lambda_2 N^2+\cdots+\lambda_m N^m), \tag{7}$$

where the λ_i's are Lagrange multipliers to be determined from the constraints. The maximum entropy (maxent) prediction of $\beta\mu^{\mathrm{ex}}$, given the available data, is $-\lambda_0$.

Alternatively, we can use a relative or cross entropy [3,4],

$$\eta=-\sum_N P(N;v)\ln\left(\frac{P(N;v)}{\hat{P}(N;v)}\right), \tag{8}$$

which is a measure of the information gain on changing from the prior probabilities, $\hat{P}(N;v)$, to the posterior probabilities $P(N;v)$. The prior probabilities allow us to explicitly include our expectations about the system's behavior. In the absence of any new information the posterior probabilities are equal to the prior probabilities and the relative information is zero. In the limit of gaining complete information about the system, such as all moments of the probability distribution, the prior probabilities become irrelevant.

Maximizing the entropy is equivalent to maximizing the cross entropy using a uniform, or uninformative prior. In the system under consideration this would be $\hat{P}(N;v)=1/N_{\max}$, where $N_{\max}$ is the maximum number of particles that can fit in the volume of interest. This is an appropriate prior when we have an enumeration of all the possible states of the system. In practice, the exact value of $N_{\max}$ for a given v is not important. The resulting maxent distribution depends only weakly on this value.

An examination of the simulation data in Fig. 2 shows that the largest deviations from the two-moment uninformative prior maxent prediction occur at low densities. The behavior of fluid systems at low enough densities should be identical with that of an ideal gas. The probability that each particle is located within the volume of interest is small, and approximately independent of the location of any other particle. As such, a Poisson probability distribution

$$\hat{P}(N;v)=\frac{\langle N\rangle_v^N\exp(-\langle N\rangle_v)}{N!} \tag{9}$$

is a more appropriate prior than a uniform distribution. This distribution is distinctly non-Gaussian for small $\langle N\rangle_v$ and N.

Maximizing the cross entropy using this prior we conclude that

$$P(N;v)=\frac{1}{N!}\exp(\lambda_0+\lambda_1 N+\lambda_2 N^2+\cdots+\lambda_m N^m). \tag{10}$$

All constants (those terms that do not depend on N) have been absorbed into the λ_i's. Comparing with the uniform prior maxent fit Eq. (7), we see that the only difference is the addition of an $N!$ divisor.

We have used Eq. (10), truncated at the quadratic term, to provide an improved estimate of the true distribution. Figures 1 and 2 compare the simulation data (points) to both the uniform (solid lines) and ideal gas (dashed lines) prior two-moment predictions. We can see that ideal gas prior results in a dramatically improved fit at low densities. There is some improvement at medium densities and almost none at high densities.

It is not apparent from the standard Monte Carlo simulations whether the maxent predictions remain valid for large volumes and large density fluctuations. Figure 3 shows the probability distribution for a reduced density of 0.5, and a volume diameter of 4σ, calculated from the Monte Carlo simulation with umbrella sampling. The Gaussian predictions of the two-moment fit are accurate for all spontaneous fluctuations that occur in practice. The deviations at low occupations are well described by the ideal gas prior two-moment fit. Thus we can quantitatively predict the excess chemical potential of a solute of diameter 3σ. At high occupations qualitatively similar deviations are presumably due to the tight, solidlike packing of solvent into this volume. Such tight packing is extremely rare at this density, so that these deviations do not affect the maxent predictions at low and medium occupations.

Figure 5 shows that at low densities, the ideal gas prior, two-moment fit accurately predicts the third central moment. The ideal gas prior does not provide an accurate theory for the errors in Gaussian statistics at high densities. For $d<2/\sqrt{3}$ the two-moment fit is exact at any density because no more than two particles can fit in the volume.

When the probability of observing a zero occupancy void is not too small, we can obtain the solute chemical potential, $\beta\mu^{\mathrm{ex}}$, directly from the simulation using Eq. (2). This chemical potential is compared in Fig. 6 with values calculated via the maxent formalism. The uniform prior predictions, the solid lines, are in serious error at all but the smallest densities and solute volumes. Use of the ideal gas prior (dashed line) improves the predictions at all densities. At reduced densities, $\rho^* \leqslant 0.5$, this fit is able to give quantitatively correct answers, apparently irrespective of the size of the cavity. At higher densities the improvements are not enough to provide quantitative agreement.

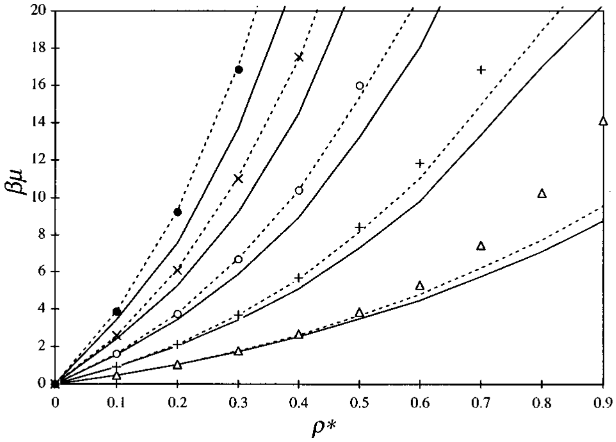

FIG. 6. The excess chemical potential, $\beta\mu^{\mathrm{ex}}$, for spherical solutes in a hard-sphere solute as a function of reduced solvent density $\rho^* = \sigma^3 N/V$, where σ is the diameter of a solvent sphere. Points are from computer simulation, and the lines are maximum entropy fits using the mean and second moment and an uninformative (—) or an ideal gas (- - -) prior. Diameter of solute is 1.0σ (△), 1.5σ (+), 2.0 σ (○), 2.5σ (×), or 3.0σ (●).

IV. DISCUSSION

It is apparent from our simulations and analysis that, for the volumes and densities considered, density fluctuations in the hard-sphere fluid are well described by Gaussian statistics. The question arises as to whether the small anharmonicities at large fluctuations can be predicted from knowledge of higher moments of the distribution. We can use the simulation data to perform a three-moment maxent fit. With a uniform prior this led to small improvements, as compared to the uniform two-moment fit, for reduced densities below 0.7, but almost no improvement at higher densities. The ideal gas prior, three-moment fit is only marginally better than the ideal gas prior two-moment fit.

In conclusion, to correct and improve upon the two-moment information theory prediction we need a physical perspective that allows the construction of a physically reasonable prior. This prior provides information about all the moments of the distribution. To develop a quantitatively accurate theory for high densities may require information about the solid phases of the hard-sphere fluid.

ACKNOWLEDGMENTS

This research has been supported in part by a grant from the National Science Foundation.

---

[1] G. Hummer, S. Garde, A. E. García, A. Pohorille, and L. R. Pratt, Proc. Natl. Acad. Sci. **93**, 8951 (1996).
[2] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[3] E. T. Jaynes, in *E. T. Jaynes: Papers on Probability, Statistics, and Statistical Physics*, edited by R. D. Rosenkrantz (Reidel, Dordrecht, 1983).
[4] J. E. Shore and R. W. Johnson, IEEE Trans. Inf. Theory **IT-26**, 26 (1990).
[5] H. Reiss, H. L. Frisch, and J. L. Lebowitz, J. Chem. Phys. **31**, 369 (1959).
[6] F. H. Stillinger, J. Solution Chem. **2**, 141 (1973).
[7] H. Reiss, in *Statistical Mechanics and Statistical Methods in Theory and Application*, edited by U. Landman (Plenum Press, New York, 1977), pp. 99–140.
[8] B. J. Berne, Proc. Natl. Acad. Sci. USA **93**, 8800 (1996).
[9] D. Chandler, Phys. Rev. E **48**, 2989 (1993).
[10] L. R. Pratt and D. Chandler, J. Chem. Phys. **67**, 3683 (1977).
[11] D. Chandler, J. D. Weeks, and H. C. Andersen, Science **220**, 787 (1983).

[12] B. J. Alder and T. E. Wainwright, J. Chem. Phys. **27**, 1208 (1957).

[13] W. W. Wood and J. D. Jacobson, J. Chem. Phys. **27**, 1207 (1957).

[14] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, San Diego, 1986).

[15] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic Press, London, 1986).

[16] P. A. P. Moran, *An Introduction to Probability Theory* (Oxford University Press, London, 1984).

[17] A. Papoulis, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill Book Company, New York, 1965).

[18] S. Garde, G. Hummer, A. E. García, M. E. Paulaitis, and L. R. Pratt, Phys. Rev. Lett. **77**, 4966 (1996).