Direct entropy calculation from computer simulation of liquids

Andras Baranyai* and Denis J. Evans
Australian National University, Research School of Chemistry, GPO Box 4, Canberra, Australian Capital Territory 2601, Australia
(Received 8 June 1989)

We develop and test a direct method for computing two-particle excess entropies in the canonical ensemble. The method is based on a systematic expansion of the entropy into one-body, two-body, three-body, etc., contributions. Unlike earlier canonical ensemble methods the resulting expressions for the entropy are *local*. It is shown that, at liquid densities, the three-body (and higher) terms are small but not negligible. The largest discrepancies are found at intermediate densities.

INTRODUCTION

Since the introduction of computer simulation, one of the most challenging tasks has been the calculation of entropic quantities. There are a number of different ways that have been proposed to do this.[1] Surprisingly, the most systematic and direct approach to the problem has not been attempted.

In 1952, H. S. Green[2] used Kirkwood's factorization[3] of the N-particle distribution function to derive an expansion for the entropy in terms of the partial N-particle distribution functions. A similar approach was subsequently explored by Nettleton and M. S. Green[4] in 1958 using graphical techniques, and by Raveché[5] in 1971 using a generating function involving isothermal activity derivatives of correlation functionals. Mountain and Raveché[6] tested their expressions for the hard-sphere fluid using the Percus-Yevick equation, and for liquid neon and rubidium using the experimentally measured pair-correlation functions of the substances.

Recently, Wallace,[7] apparently unaware of Raveché and Mountain's work, rederived the entropy expansion and calculated the entropy of liquid sodium near to its melting point using experimental structure factor data. He found that his expansion, terminated at the pair level, was accurate to within estimated uncertainties of 2%.

An application of the entropy expansion using computer simulation has recently been carried out by Evans[8] who studied the nonequilibrium entropy of soft disks undergoing thermostatted planar Couette flow.

The aim of the present work is to test the applicability of the method to computer simulations of three-dimensional fluids at equilibrium. In the first part of this paper we summarize and discuss some of the main elements of the derivation. This will clarify the way in which entropy calculations should be performed in canonical or microcanonical computer simulations. In the second part of the paper we present numerical results for the Lennard-Jones system.

THEORY

Because of the well-known difficulties, particularly at high densities, of performing computer simulations in the grand canonical ensemble, we focus our attention on the canonical ensemble. We begin with Wallace's expression for the entropy of a canonical ensemble of systems:[7]

$$S_N=-\frac{Nk}{\rho}\int f_N^{(1)}(\mathbf{p})\ln[h^3 f_N^{(1)}(\mathbf{p})]d\mathbf{p}-\frac{1}{2}k\rho^2\int\int g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_2)\ln[g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_2)]d\mathbf{r}_1 d\mathbf{r}_2$$
$$-\frac{1}{3!}k\rho^3\int\int\int g_N^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)\ln[\delta g_N^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)]d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3-\cdots . \quad (1)$$

The one-particle distribution is

$$f_N^{(1)}(\mathbf{p})=\rho(2\pi m k_B T)^{-3/2}\exp\left[\frac{\mathbf{p}^2}{2mk_B T}\right] \quad (2)$$

and k_B is the Boltzmann constant, ρ is the number density, h is the Planck constant, $\mathbf{p}$ is the momentum, m is the mass of the particles, and T is the absolute temperature. $g_N^{(1)}(\mathbf{r}_1,\mathbf{r}_2)$ and $g_N^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$ are the *canonical* pair and triplet correlation functions, respectively. The definition of $\delta g_N^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$ is

$$g_N^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)=g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_2)g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_3)$$
$$\times g_N^{(2)}(\mathbf{r}_2,\mathbf{r}_3)\delta g_N^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_2) . \quad (3)$$

The terms in the expansion (1), are identified as the one-, two-, and three-particle entropies, respectively. In each of the configurational terms one integral can be performed explicitly to give a factor, N, and thus (1) can be rewritten in a more convenient form using intensive quantities

$$s = S_N/Nk = s_1 + s_c \ . \tag{4}$$

Here s_1 is the one-particle term which can be explicitly evaluated as

$$s_1 = \tfrac{3}{2} - \ln(\rho\Lambda^3) \ , \tag{5}$$

where Λ is the de Broglie wavelength

$$\Lambda = h/(2\pi m k_B T)^{1/2} \ . \tag{6}$$

The term s_c in (2) will be referred to in the following as the configurational entropy. The total entropy per particle of a perfect gas, $s_{\rm PG}$, is

$$s_{\rm PG} = \tfrac{5}{2} - \ln(\rho\Lambda^3) \ . \tag{7}$$

Comparing (7) and (5) Wallace[7] came to the conclusion that the gas and the dense fluid occupy different regions of phase space. The erroneous remark resulted from the fact that, in spite of his noting the importance of the correct long-range behavior, he did not actually calculate the higher-order terms for the perfect gas.

The value of $g_N^{(n)}$ for a perfect gas is

$$g_N^{(n)} = \frac{N(N-1)\cdots(N-n+1)}{N^n} \ . \tag{8}$$

These values are clearly the limiting values of the correlation functions when r goes to infinity. Substituting (8) into the configurational terms of (1)—for the omitted terms see Ref. 7—the integrals can be carried out explicitly. The results are shown in Table I.

As can be seen from Table I the *configurational* contribution to the entropy of the perfect gas in *canonical* ensemble is exactly 1. Summarizing the results for a perfect gas in the canonical ensemble we see that the single-particle entropy is given by (5) but, perhaps surprisingly, the perfect gas entropy is given by (7). The difference is due to summing all the configurational contributions to the perfect gas entropy. The reason why the perfect gas entropy contains configurational contributions to arbitrary order in N is because in the canonical ensemble N is fixed and the partial distribution functions do not go to unity at long range. They are given by (8) instead. If we are interested in computing the entropy of a dense fluid we will clearly have to truncate our evaluation of the entropy expansion at some relatively low level—often at the two-particle level. If we simply ignore the higher-order configurational terms, then just as in the case of a perfect gas, we will derive an incorrect value for the entropy. In any calculation of the entropy we must explicitly evaluate the low-order contributions to the entropy and then add these values to the sum of the perfect gas configurational components for the neglected high-order terms. (Note, if the pair-correlation function only is taken into account, the contribution from the neglected terms is $\frac{1}{2}$.)

These conclusions are only valid if the whole system is considered. This means that in molecular dynamics (MD) simulations, the distribution functions should be calculated for the *entire* simulation cell. Clearly, the images cannot be involved in the calculation because their contribution to the entropy is zero. However, because the simulation cell usually has a cubic shape, care is required in calculating the configurational contributions correctly.

Because we know the exact integrals of the various canonical ensemble correlation functions we can now rewrite Eq. (1) in an alternative way. By adding zero to the right-hand side of (1) we can write

$$\begin{aligned} s &= s_1 - \tfrac{1}{2}\rho\int g_N^{(2)}\ln g_N^{(2)}d\mathbf{r} + \tfrac{1}{2} + \tfrac{1}{2}\rho\int (g_N^{(2)}-1)d\mathbf{r} \\ &\quad - \tfrac{1}{6}\rho^2\int\int g_N^{(3)}\ln(\delta g_N^{(3)})d\mathbf{r}^2 + \tfrac{1}{6} + \tfrac{1}{6}\rho^2\int\int (g_N^{(3)} - 3g_N^{(2)}g_N^{(2)} + 3g_N^{(2)} - 1)d\mathbf{r}^2 - \cdots \\ &= s_{\rm PG} - \tfrac{1}{2}\rho\int g_N^{(2)}\ln(g_N^{(2)})d\mathbf{r} + \tfrac{1}{2}\rho\int (g_N^{(2)}-1)d\mathbf{r} \\ &\quad - \tfrac{1}{6}\rho^2\int\int g_N^{(3)}\ln(\delta g_N^{(3)})d\mathbf{r}^2 + \tfrac{1}{6}\rho^2\int\int (g_N^{(3)} - 3g_N^{(2)}g_N^{(2)} + 3g_N^{(2)} - 1)d\mathbf{r}^2 - \cdots \ . \end{aligned} \tag{9}$$

As we shall see, this form is more accurate for numerical computations of the canonical ensemble entropy. The reason why it is more accurate is related to the fact that Eq. (9) is an ensemble invariant form for the entropy. Equation (9) was derived by Nettleton and M. S. Green[4] and also by Raveché,[5] as an expression for the entropy for the *grand canonical ensemble*. That an ensemble invariant form for the entropy exists, is obvious once we

TABLE I. Perfect gas configurational entropies.

n	Contribution to $s_{\rm PG}(N)$	$s_{\rm PG}(\infty)$
2	$-\frac{1}{2}(N-1)\ln[(N-1)/N]$	$\frac{1}{2}$
3	$-\frac{1}{6}(N-1)(N-2)\ln[N(N-2)/(N-1)^2]$	$\frac{1}{6}$
4	$-\frac{1}{24}(N-1)(N-2)(N-3)\ln[(N-1)^3(N-3)/N(N-2)^3]$	$\frac{1}{12}$
5	$-\frac{1}{120}(N-1)(N-2)(N-3)(N-4)\ln[N(N-2)^6(N-4)/(N-3)^4(N-1)^4]$	$\frac{1}{20}$
	...	
$\sum_n$	$-(1/N)\ln(N!/N^N)$	1

consider the fact that a grand canonical ensemble can be constructed by considering a fixed relative volume of a much larger canonical ensemble and then taking the thermodynamic limit. Thus, if a *local* expression for the entropy exists, it must be an ensemble invariant. Equation (9) is such a local form.

We should point out that for a grand canonical perfect gas, $s_{PG}=s_1$ and $s_{ex}=0$. Furthermore for this system, each of the terms on the right-hand side of (9), vanishes. This is because for the grand canonical ensemble the distribution functions have no $O(1/N)$, long-ranged corrections. In the canonical ensemble the distribution functions are long ranged but the sums of the integrals appearing in (9) are not long ranged.

RESULTS AND DISCUSSION

In order to test the usefulness of Eq. (1) and (9) for numerical calculations of the entropy, we performed a series of molecular dynamics simulations of the Lennard-Jones fluid. The simulations were performed using Gaussian isokinetic dynamics. Most of the calculations were performed for 500 particle systems. The potential was truncated at 2.5σ.

Using Eq. (1), the integrals must go over the entire volume of the minimum image cell. Therefore the pair-correlation function must be determined without any kind of cutoff within the simulation cell. If $r<L/2$, where L is the length of the cubic simulation cell, the size of the volume element at distance r is $4\pi[r^3-(r-\Delta r)^3]/3$, where r measures the pair separation and Δr is the thickness of the shell used to form histograms. For $L/2<r<3^{1/2}L/2$, the volume element can be written as $4\pi A(r)[r^3-(r-\Delta r)^3]/3$, where $0<A(r)<1$. The A function measures the relative volume of a spherical shell cut by a cube.

To avoid a cumbersome analytic determination of $A(r)$, a simple Monte Carlo algorithm has been applied. The resulting function can be seen on Fig. 1. For $r<L/2$, the function should be exactly 1. This is consistent with our Monte Carlo results. In our entropy calculations, however, we used the exact value of $A(r)$ for $r<L/2$. For the sake of numerical consistency, the $A(r)$ function was calculated in exactly the same manner as the corresponding $g(r)$. We used exactly the same r grid for the calculation of $A(r)$ and $g(r)$. The largest numerical relative error in our estimation of $A(r)$, was less than 0.003. The relative error in the volume integral of $A(r)4\pi r^2 dr/3$ was less than 0.0001.

The pair-correlation functions were calculated from equilibrated molecular dynamics runs with the reduced time step of 0.004. After every 20 time steps a configuration sample was used to increment the $g(r)$ histograms. Each post equilibrated run had a length of 20 000 time steps.

In the two-particle case the integral can be written as a sum over the minimum image cube

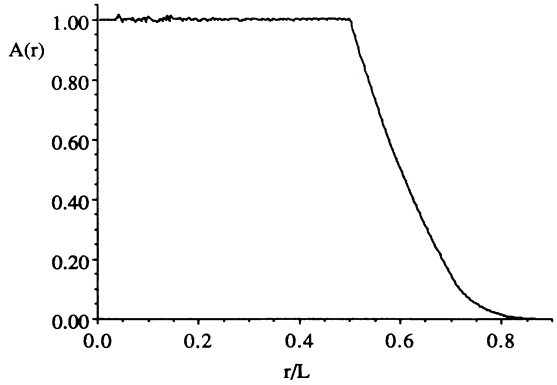

FIG. 1. Shows a Monte Carlo estimate of the function $A(r)$. L is the length of the simulation cell. This function is required for the computation of the two-body excess entropy using the nonlocal form (1).

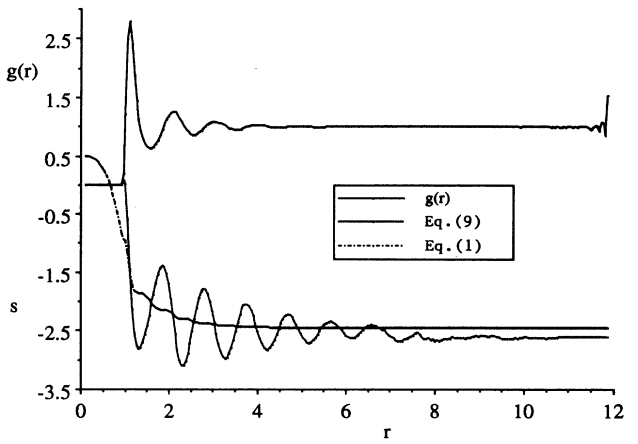

FIG. 2. Shows the radial distribution function $g(r)$ and two estimates of the excess two-body entropy for a Lennard-Jones fluid ($N=2048$, $T=0.75$, $\rho=0.8$). The $g(r)$ is computed into the corners of the simulation cell using the $A(r)$ function shown in Fig. 1. The local estimate of the entropy (9), is more accurate than that computed using Eq. (1).

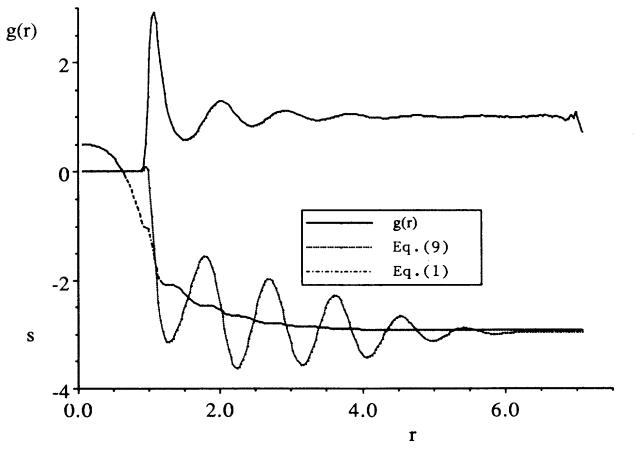

FIG. 3. Shows similar results as Fig. 2 except the state point is ($N=500$, $T=1.15$, $\rho=0.92$).

TABLE II. A^{ex} and U^{ex} are the excess Helmholtz and internal energies. s^{ex} and s_2 are the excess and two-body entropies per particle. U^{ex} is corrected for the Lennard-Jones cutoff at 2.5σ. s_2 is the two-particle excess entropy per particle. It contains the perfect gas configurational terms for all the higher-order terms. A^{ex} and s^{ex} were calculated using thermodynamic integration in Ref. 9. For these values, $T=1.15$ and $N=500$ were used.

ρ	A^{ex}/N	U^{ex}/N	s^{ex}	s_2	$s^{ex}-s_2$
0.92	−1.784	−5.953	−3.625	−3.43	−0.195
0.85	−2.043	−5.665	−3.150	−2.83	−0.320
0.75	−2.169	−5.108	−2.556	−2.18	−0.376
0.65	−2.115	−4.458	−2.037	−1.72	−0.317
0.60	−2.043	−4.130	−1.815	−1.54	−0.275
0.50	−1.823	−3.499	−1.457	−1.25	−0.207

TABLE III. Entropies in the triple-point region. p is the reduced hydrostatic pressure.

ρ	T	U^{ex}/N	p	s^{ex}	s_2	$s^{ex}-s_2$
0.8	0.75	−5.772	−0.294	−3.226	−2.98	−0.286
0.7	0.75	−5.076	−0.812	−2.595	−2.31	−0.285
0.835	0.903	−5.833	1.097	−3.248	−3.02	−0.228

$$-\frac{\rho}{2}\sum_r \frac{N(r)}{\Delta V(r)\rho}\ln\left[\frac{N(r)}{\Delta V(r)\rho}\right]\Delta V(r) = -\tfrac{1}{2}\sum_r N(r)\ln\left[\frac{N(r)}{\Delta V(r)\rho}\right], \quad (10)$$

where $N(r)$ is the average number of particles in the shell $A(r)4\pi r^2 dr$.

Some calculations were performed for different system sizes in order to investigate the number dependence of the results. The largest system contained 2048 particles. It is easy to see that for such a large system, accurate calculations of the entropy require increasingly accurate determinations of $A(r)$. This is because, if $\sum\Delta V(r)=V$ has an error of δV, the resulting error in the integral will be $\sim\delta V N/\rho$. Thus this source of error for the entropy per particle is extensive.

This form of error is eliminated when the calculation is done using (9). This is a clear advantage of the ensemble invariant formulation of the entropy. Figures 2 and 3 show the computed entropies calculated as running integrals for two different state points and system sizes, ($T=0.75$, $\rho=0.8$, $N=2048$) and ($T=1.15$, $\rho=0.92$, $N=500$), respectively. All variables are given in the usual reduced forms—reduced by the Lennard-Jones σ, ε parameters and the particle mass m.

The upper curve is the pair-correlation function that is computed into the corners of the simulation cell. As we move towards the corners of the simulation cell, the $g(r)$'s become noisy due to the relatively small numbers of pairs of particles with appropriate configurations. Below, we see, as a function of r, the integrals for the entropy per particle computed using Eq. (1) and (9). The Eq. (1) data exhibit an oscillatory approach to the excess entropy of the system. These oscillations are due to the oscillations in $g(r)$. The oscillations in the entropy are longer ranged than they are for $g(r)$ itself because they are enhanced by the quadratic growth of the volume element.

The local form for the entropy, Eq. (9), converges much more rapidly and smoothly towards its asymptotic value. At $r=3^{1/2}L/2$, the entropy computed from Eq. (1) and (9) should converge to the same value. For the $N=500$ system the agreement is good. For $N=2048$ the disagreement is due to the numerical uncertainties in computing $A(r)$, as required for Eq. (1). The major source of error comes from the region where r is a little greater than $L/2$. Errors in determining $A(r)$ in this region have a large effect on the entropy as calculated from Eq. (1). It should be noted that the seemingly large statistical fluctuations in $g(r)$ at $r=3^{1/2}L/2$ (i.e., in the corners of the box) have practically no effect on the computed entropy.

By comparing our results with previous equations of state data,[9] we can form some idea of the relative importance of the one- and two-body contributions to the excess entropy. This comparison is given in Tables II–IV and Figs. 4 and 5. The statistical error in our computed

TABLE IV. Two-particle excess entropy. For these values, $T=1.15$ and $N=500$ were used.

ρ	U^{ex}/N	p	s_2
0.9	−5.510	6.377	−2.871
0.8	−5.121	3.282	−2.234
0.7	−4.579	1.591	−1.757
0.6	−3.957	0.779	−1.401
0.5	−3.320	0.418	−1.115
0.4	−2.695	0.281	−0.876
0.3	−2.067	0.227	−0.660
0.2	−1.440	0.183	−0.466
0.1	−0.737	0.117	−0.240

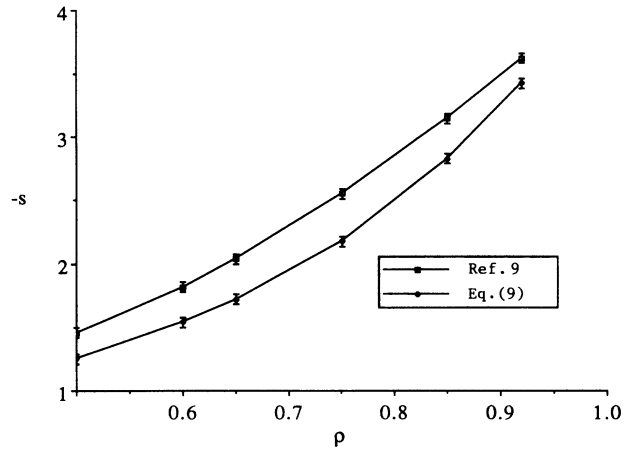

FIG. 4. Shows a comparison of the excess entropy computed by thermodynamic integration (Ref. 9), with the two-body excess entropy computed using (9). The data is for the $T=1.15$ isotherm. The data is from Table II.

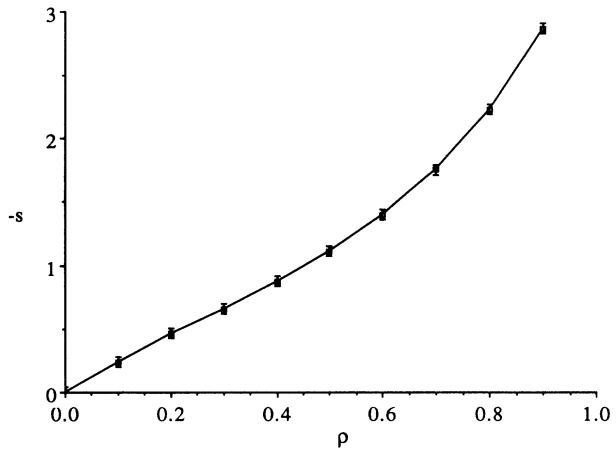

FIG. 5. Shows the two-body excess entropy computed using (9), for the $T=1.5$ isotherms. The data is from Table IV.

two-particle excess entropies is about ± 0.02. This error could be reduced much further by performing longer runs with more particles. Our statistical uncertainties are, however, of the same order as those reported in Refs. 1 and 9.

The data are given in the usual reduced quantities per particle. The ex superscript refers to the difference from the perfect gas value (excess). A^{ex} and s^{ex} are taken from Ref. 9 and were calculated by *thermodynamic integration.* U^{ex} is the corrected excess internal energy per particle as calculated in this paper. s_2 is the two-particle excess entropy per particle, calculated using the *local form for the two-particle entropy,* (9). All of the new results given in Tables II–IV refer to 500 particle simulations.

From Tables II and III we see that the two-particle excess entropy contributes more than 85% of the excess entropy for all the thermodynamic states considered. In the triple-point region the two-body excess entropy contributes as much as 95%. This could be due to partial cancellation of the omitted higher-order terms. (The three- and four-body terms may partially cancel each other.)

Figures 4 and 5 give graphs of the excess entropies in Tables II and IV, respectively. In Fig. 4, the upper curve is from Ref. 9, while the lower curve is from the present paper. One can see quite clearly that the two-body excess entropy agrees with the result from thermodynamic integration at the density extremes. The agreement is poorest at intermediate densities.

Table IV gives our results for the two-particle excess entropy, s_2, calculated along the $T=1.5$ isotherm.

It can be seen from the results that the two-particle entropy contributes at least 85% of the total excess entropy. Surprisingly, the agreement towards the triple point is better than at intermediate densities. In this respect our results are in agreement with those of Mountain and Raveché[6] for PY hard spheres.

The determination of pair entropy from experimentally measured correlation functions is much less accurate than the value determined from simulation. This is because of the relatively larger error present in the measured structure factors.

DISCUSSION

We have presented a new method of calculating the pair contribution to the entropy. It is based on a local, ensemble-independent expression for the entropy [Eq. (9)]. Our numerical results show the clear superiority of this method compared to the older method based on H. S. Green's nonlocal expression [Eq. (1)]. The calculations of the pair entropy are, by modern standards, quite easy to perform. No simulation runs reported in this paper were longer than 20 000 time steps.

The main problem that remains to be solved, if this technique is to develop into a standard method for computing entropies and free energies, is the calculation of higher-order terms. It seems clear that local versions of the three-body entropy will be far easier to compute than their nonlocal analogs based on Eq. (1). The necessary three-dimensional histograms would involve nontrivial computing resources but since it seems likely that they only contribute $<20\%$ to the total excess triple-point entropy, the three-body terms will not need to be evaluated with the same relative accuracy as that required for the two-body terms. It is also likely that the local entropy expansion will permit us to more easily use polynomial expansions of $g^{(3)}$. Expansion methods for the higher-order entropy contributions are extremely complex if the nonlocal form (1) is used.

*Permanent address: Eötvös University, Laboratory of Theoretical Chemistry, Budapest, Múzeum krt 6-8., 1088-Hungary.

[1]M. Mezei, Mol. Phys. **61**, 565 (1987).
[2]H. S. Green, *The Molecular Theory of Fluids* (North-Holland, Amsterdam, 1952).
[3]J. G. Kirkwood, J. Chem. Phys. **10**, 394 (1942).
[4]R. E. Nettleton and M. S. Green, J. Chem. Phys. **29**, 1365 (1958).
[5]H. J. Raveché, J. Chem. Phys. **55**, 2242 (1971).
[6]R. D. Mountain and H. J. Raveché, J. Chem. Phys. **29**, 2250 (1971).
[7]D. C. Wallace, J. Chem. Phys. **87**, 2282 (1987).
[8]D. J. Evans, J. Stat. Phys. (to be published).
[9]J. P. Hansen and L. Verlet, Phys. Rev. **184**, 151 (1969).