Mayer Sampling: Calculation of Cluster Integrals using Free-Energy Perturbation Methods

Jayant K. Singh and David A. Kofke
Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, New York 14260-4200, USA

(Received 24 December 2003; published 2 June 2004)

Free-energy simulation methods are applied toward the calculation of cluster integrals that appear in diagrammatic methods of statistical mechanics. In this approach, Monte Carlo sampling is performed on a number of molecules equal to the order of the integral, and configurations are weighted according to the absolute value of the integrand. An umbrella-sampling average yields the value of the cluster integral in reference to a known integral. Virial coefficients, up to the sixth for the Lennard-Jones model and the fifth for the SPCE model of water, are calculated as a demonstration.

DOI: 10.1103/PhysRevLett.92.220601 PACS numbers: 05.20.Jj, 05.10.Ln, 51.30.+i

Statistical mechanical theories of fluids aim to describe the behavior of the bulk phase beginning from a model for the intermolecular interactions. The most successful of these theories, such as the virial treatment of gases and the Percus-Yevick theory for condensed phases, are built upon an enumeration of "cluster integrals" taken over configurations of a fixed set of molecules [1]. One of the fundamental obstacles to the systematic improvement (and in some cases even the application) of theories based on cluster integrals is the inability to compute the integrals they represent for any but the simplest cases. Thus it has proved feasible to compute only up to the fifth virial coefficient for simple spherically symmetric potentials such as the Lennard-Jones (LJ) model [2–5], and the fourth virial has just been calculated for a cylindrically symmetric molecule [6], while only the third virial coefficient has been computed for slightly more complex potentials, such as water [7]. Up to the eighth virial coefficient has been computed for hard spheres [8], but this is an exceptional case that benefits greatly from the simple form and short range of the hard-sphere potential. Many of the successes in calculating difficult cluster integrals have required much effort to examine the nature of the integral, exploiting all possible symmetries, and applying analytical treatments such as expansions in Legendre polynomials. On the other hand, second virial coefficients can be calculated routinely for almost any intermolecular potential, even large molecules such as alkanes or potentials described through *ab initio* treatments. The second virial coefficient is one of the very few routinely computable quantities that rigorously connects a molecular model to bulk-phase behavior, and consequently it plays an important role in developing and assessing molecular models.

In this Letter, we present a very simple but general method to calculate cluster integrals. The approach is suitable for integrals of any order, and can be applied to potentials of any complexity. The method has the character of a molecular simulation performed on the molecules represented in the integral, and there are two key ideas in its application to cluster integrals. First, we generate the molecule configurations using Metropolis Monte Carlo (MC) [9] with importance sampling of configurations based on the magnitude of the interactions represented in the cluster integral. Second, we aim to evaluate the ratio of the desired cluster integral to a known reference integral, and do not attempt to evaluate the cluster-integral directly. In taking this approach we can then bring to bear a powerful array of techniques that have previously been developed for the calculation of free-energy differences by molecular simulation [9]. We propose "Mayer sampling" as the name given to this technique of computing cluster integrals using importance-sampling and free-energy methods. There are many variations and applications that can be pursued from this basic approach. To fix ideas, in the present work we focus on one particular problem: namely, the evaluation of cluster integrals needed to calculate virial coefficients from a molecular model [10]. We demonstrate by calculating the sixth virial coefficient for the LJ model, and the fourth and fifth virial coefficients for the SPCE model of water [11], none of which have been reported previously.

The cluster integral that gives the classical second virial coefficient is [1,10]

$$B_2(T) = -\frac{1}{2}\int f_{12} d\mathbf{r}_2, \tag{1}$$

and for pairwise-additive potentials the third virial coefficient is

$$B_3(T) = -\frac{1}{3}\iint f_{12} f_{13} f_{23} d\mathbf{r}_2 d\mathbf{r}_3. \tag{2}$$

In the integrals $f_{ij} = [\exp(-\beta u_{ij}) - 1]$ is the Mayer function, where u_{ij} is the pair potential between molecules labeled i and j, and $\beta = (k_B T)^{-1}$ is the reciprocal temperature in energy units. The formulas indicate integration over only the positions of each molecule, but in general integration must be performed also over all rotational and internal degrees of freedom available to each one. Higher-order coefficients are sums of cluster integrals. For instance, the fourth virial coefficient $B_4(T)$ is a sum of three types of 4-molecule cluster integrals, while the fifth virial coefficient $B_5(T)$ is the sum of ten 5-molecule integrals.

Free-energy perturbation formulas are well suited for evaluation of such integrals. The umbrella-sampling method provides one such formula [9]

$$\Gamma(T) = \Gamma_o \frac{\Gamma}{\Gamma_o} = \Gamma_o \frac{\langle \gamma/\pi \rangle_\pi}{\langle \gamma_o/\pi \rangle_\pi}. \tag{3}$$

We use $\Gamma(T)$ as the notation for a general cluster integral or sum of integrals, with integrand (or sum of integrands) $\gamma(\mathbf{r}^n; T)$; for example, if Γ is the third virial coefficient B_3, then $\gamma \equiv f_{12} f_{13} f_{23}$. The angle brackets indicate the "ensemble-average" integral over all configurations and orientations of the n molecules, and the subscript π indicates that sampling is governed by the (normalized) probability distribution $\pi(\mathbf{r}^n; T)$: $\langle M \rangle_\pi \equiv \int d\mathbf{r}^n M\pi / \int d\mathbf{r}^n \pi$. The subscript "o" indicates a quantity for the reference system. This idea generalizes an approach that has emerged recently [12] for the evaluation of cluster integrals in the bridge function.

There is considerable latitude available in the selection of both the reference system and the sampling distribution $\pi(\mathbf{r}^n)$. We expect that one (or a combination) of two choices commonly will be employed in selecting a reference system. In the first choice, the intermolecular potential or the temperature of the reference system differs from that of the target system, but otherwise they correspond to the same cluster integral (or sum of integrals). Typically the reference system would be defined by the hard-sphere potential, for which the cluster integral is temperature independent and in many cases is known. In the second choice, the intermolecular potential for the target and the reference are the same, but each system is given by different products of Mayer functions—they correspond to different cluster integrals. This approach can be used to relate more complicated, highly connected cluster integrals to ones that are simpler and which can be evaluated by other means.

Regarding the distribution π, the importance-sampling approach suggests that it be given by the magnitude of the target-system integrand, $\pi = |\gamma(\mathbf{r}^n; T)|$ (the absolute value must be used because γ is negative for some configurations). We have investigated several variations on this theme, and found that a particularly convenient formulation is to select π as the absolute value of the *sum of all the cluster integrands in the desired virial coefficient, with their appropriate weights.* If we further select γ itself as the same cluster-integral sum, Eq. (3) *yields directly the virial coefficient.* Moreover, the numerator in Eq. (3) averages just the sign of the cluster sum, and contributions to it are all ± 1. This feature is helpful to understanding the performance of the calculation. The cluster-integral sum that gives the virial coefficients is known to suffer from large cancellations between the different terms, which causes difficulty in determining the virial sum precisely. If implemented as described here, the present approach *does not require the addition of (large) positive and negative averages* to obtain their (small) sum. Instead we have found this numerical problem to manifest itself in a different way: at the conditions where this cancellation is most problematic, the single average $\langle \mathrm{sgn}(\gamma) \rangle_{|\gamma|}$ as suggested here becomes very small. A simple probabilistic argument indicates that the required amount of sampling will scale inversely with the square of this average, so from the value of this average we can get an indication of whether we have performed sufficient sampling. This feature also presents a clear target for the focus of future work that aims to improve the performance of these calculations.

We have applied the methodology described here to evaluate the second to the sixth virial coefficients for the LJ model, and the second to fifth for the SPCE potential for water (all classical). The calculations were conducted as follows. MC sampling was performed for a number of molecules equal to the order of the virial coefficient being computed. Trials were attempted to displace and (if appropriate) rotate the molecules. We found it helpful to select a random number of molecules to perturb in one trial (so, e.g., with equal likelihood, sometimes one molecule is moved, sometimes two, etc., and sometimes all of them are). Each trial was accepted with probability $\min(1, \pi^{\mathrm{new}}/\pi^{\mathrm{old}})$, where π is defined as the absolute value of the weighted sum of the cluster integrands contributing to the calculated virial coefficient. The resulting configuration contributed to the two averages in Eq. (3). For the denominator in this equation, we defined γ_o to be the ring cluster (of same order) for the hard-sphere potential, for which Γ_o is known; we examined several values of the hard-sphere diameter near the LJ σ (for the oxygen atom in the case of the water calculations) and did not observe a significant effect on the results. The value of a cluster for a given configuration was determined using all unique permutations of the labeling of the molecules. Simulations sampled $M = 10^7$ to 10^9 configurations, depending on the rate of convergence of the averages. Step sizes for the trials were adjusted in a short "equilibration" period, before accumulating averages, to achieve a 50% acceptance rate for trial moves; it is important that this step size not be adjusted once averaging is begun.

Results are presented in Tables I and II, respectively, where they are compared to available literature data. The coefficients B_6 for LJ and B_4 and B_5 for water have not previously been reported. Overall the data show very good precision and agree well with the literature values, although some discrepancies are present. Figure 1 presents the scaled standard deviation $M^{1/2}\sigma_B/|B|$ describing independent measurements of the virial coefficients using this method. For large M this quantity should be independent of M, so from these data one can estimate

TABLE I. Virial coefficients for the Lennard-Jones model as calculated using the Mayer-sampling method.[a]

	T*										
	0.625	0.75	1.0	1.2	1.3	1.4	1.5	2	2.5	5	10
B_2/b	-5.7578_3	-4.1757_3	-2.5381_2	-1.8359_1	-1.5842_1	-1.3759_1	-1.2010_1	-0.6275_1	-0.3126_1	$0.243\,32_4$	$0.460\,87_3$
Lit[b]	-5.7578	-4.1759	-2.5381	-1.8359	-1.5841	-1.3758	-1.2009	-0.6276	-0.3126	0.2433	0.4609
B_3/b^2	-8.237_2	-1.7923_7	0.4299_2	0.5922_4	0.5881_2	0.5682_1	0.5433_1	$0.437\,03_8$	$0.381\,00_4$	$0.315\,07_9$	0.2860_1
Lit[b]	-8.2355	-1.7915	0.4297	0.5924	0.5882	0.5683	0.5434	0.4371	0.3810	0.3151	0.2861
B_4/b^3	-120.82_{20}	-18.77_3	-0.2697_{20}	0.3385_5	0.3168_5	0.2701_4	0.2256_3	$0.122\,79_7$	0.1131_1	0.1341_1	0.1156_2
Lit[b]	-121.08	-18.84	-0.2769	0.3359	0.3157	0.2695	0.2250	0.1228	0.1131	0.1341	0.1156
Lit[c]			-0.265	0.3386_6	0.3169_6	0.2700_5	0.2253_3	0.1233_3		$0.134\,15_1$	$0.115\,59_1$
Lit[d]	$-120.558\,87$	$-18.744\,73$	-0.26810		$0.316\,93$	$0.270\,17$	$0.225\,48$	$0.122\,85$	$0.113\,14$	$0.134\,12$	$0.115\,59$
B_5/b^4	-2129_{10}	-182.68_{80}	-2.77_3	0.015_8	0.0375_{60}	-0.0064_{90}	-0.033_3	-0.011_1	0.0366_1	0.0630_1	$0.038\,94_5$
Lit[b]	-2136.2	-185.9	-2.860	0.0150	0.0361	-0.0022	-0.0303	-0.0100	0.0365	0.0629	0.0390
Lit[e]			-2.767	0.0236	0.0379	0.0018	-0.0306	-0.0101	0.0363	0.0628	0.0390
B_6/b^5	$-18\,800_{3000}$	-1360_{200}	-15_1	-0.35_{20}	-0.219_{40}	-0.174_{60}	-0.096_{60}	0.0187_{80}	0.0528_{40}	0.0244_5	0.0089_1

[a] Values are reduced by the value of the hard-sphere second virial coefficient $b = 2\pi\sigma^3/3$, and temperatures are given in LJ units, $T^* = kT/\varepsilon$, where σ and ε are the LJ size and energy parameters, respectively. Reported values represent averages from at least 4 independent calculations, each sampling at least 10^8 configurations. Subscripted digits represent the confidence limits (standard error of the mean) for the rightmost digits of the value.
[b] Reference [2].
[c] Reference [3]; average of compressibility and pressure values, with error given as half their difference.
[d] Reference [5].
[e] Reference [4].

the amount of sampling required to achieve a given fractional precision in the measurement of a virial coefficient. These results will provide a useful basis for comparison of the effectiveness of variants of the proposed method as they are developed—a lower value is of course better. The figure and tables indicate that increased sampling is required to get good precision when calculating higher-order coefficients at lower temperatures. It is also interesting to find that the anomalies of the critical point are manifested in the calculation of these few-particle averages, particularly for the Lennard-Jones system, where we find a significant jump in the scaled standard deviation near the critical temperature T_c.

We find that our results for water differ consistently from the available literature values. The difference is in a direction that would indicate overrepresentation of low-energy (attractive) configurations in the literature data, but we do not have any independent means to resolve the discrepancy in this direction. We can report that the difference is not diminished by added sampling in our calculations.

There are several points worth noting in respect to this method. First, it is very important that the sampling distribution π be selected such that it does not have small values for configurations in which γ or γ_o are non-negligible. If this happens the averages seen in Eq. (3) will have large contributions from configurations that are rarely sampled. This problem is well known in free-energy perturbation calculations [13] (e.g., the failure of particle deletion), and it is easily avoided. As a general

TABLE II. Virial coefficients for the SPCE model of water [11] as calculated using the Mayer-sampling method.[a]

	T(K)								
	373	423	473	523	573	623	673	723	773
B_2 (cm^3/mol)	-1804.9_4	-886.5_4	-518.9_1	-339.54_4	-239.14_{10}	-177.15_4	-135.99_2	-107.14_{20}	-85.96_5
Literature[b]	-1856	-907	-529	-345	-242	-179	-137	-108	-87
B_3 (10^4 cm^6/mol^2)	-1029_8	-104.5_2	-13.73_3	-0.874_6	1.076_1	1.165_1	0.9530_6	0.7380_3	0.5674_3
Literature[b]	-1070	-108	-14	-0.866	1.11	1.19	0.974	0.752	0.578
B_4 (10^7 cm^9/mol^3)	$-24\,300_{2000}$	-285_{10}	3.35_{10}	3.369_{40}	1.121_{30}	0.358_3	0.115_1	0.0377_5	0.0095_2
B_5 (10^{10} cm^{12}/mol^4)	$-13\,700_{13000}$	-859_{500}	-32.9_{150}	-2.32_{30}	0.0620_{100}	0.0920_{20}	0.0398_{10}	0.0184_6	0.00820_{20}

[a] Values are reported using the format described in Table I.
[b] Reference [7].

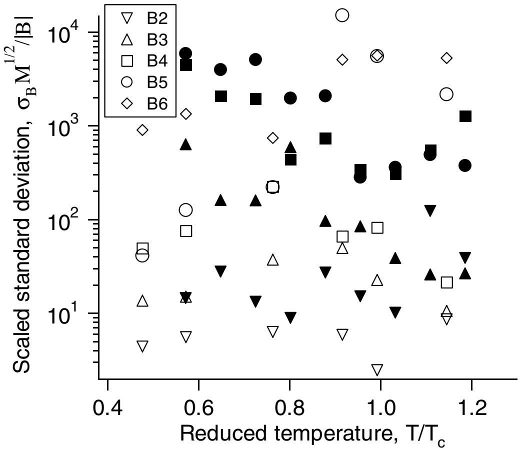

FIG. 1. Standard deviation of the measured virial coefficient across several (4 to 10) independent runs each sampling M configurations. Values are multiplied by $M^{1/2}$, and divided by the absolute value of the measured virial coefficient itself. Abscissa is the temperature reduced by the critical-point temperature for each model. Open symbols describe data for the LJ model, and filled symbols are for SPCE water.

rule, π should not have f bonds between molecules that are not bonded in γ and γ_o.

Second, the Mayer-sampling calculation differs from a typical MC molecular simulation in a number of ways. The simulation is conducted in free space, with no volume imposed, no periodic boundaries, and no artificial truncation of the potential. Moreover, the molecules can move right through each other, thereby eliminating the main source of ergodic traps (steric repulsion) seen in standard MC simulation.

Third, the technique is easily extended to other types of systems, such as mixtures, or systems described by non–pairwise-additive potentials. Application to mixtures is appealing because the dependence of the virial coefficients on composition is known exactly, so the treatment provides a route to thermodynamics that does not rely on an empirical "mixing rule" for the composition dependence. With respect to applications to other potentials, we anticipate that some day it will be feasible to apply this technique, and extensions of it, using *ab initio* potentials computed on the fly, thereby permitting very good estimation of many-body properties such as the critical point from first-principles calculations.

Fourth, the calculations described here are perfectly parallelizable. Independent MC simulations can be performed for the same system, and the averages combined to obtain a single result. There is no prospect or need for more sophisticated parallelization approaches, such as domain decomposition, because there are so few particles sampled in a simulation. In this sense the calculation has most assuredly been broken into its most elementary computational unit, and moreover the combination of elementary single-processor calculations is quite trivial.

Finally, the particular methodology presented here adapts just one of a range of well-established methods for computing free energies by molecular simulation. The formula given by Eq. (3) is an umbrella-sampling calculation. Other approaches, such as overlap sampling [14], might prove worthwhile as an alternative for the calculation of cluster integrals. Density-of-states methods [9] could also be considered, but we should point out that the structure of the cluster integrands is such that the adoption of these ideas for this purpose is not straightforward. Additionally, advanced sampling methods, such as parallel tempering (PT) [9], could be helpful in this application. Although we did not pursue it here, it should be straightforward to measure cluster integrals over a continuous range of temperatures with a single simulation, or set of PT simulations. Clearly, there are many avenues and applications to explore. The combination of Mayer sampling with integral equation methods is a particularly appealing prospect.

This work is supported by the U.S. Department of Energy. Computational resources were provided by the University at Buffalo Center for Computational Research.

---

[1] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, London, 1986).
[2] J. A. Barker, P. J. Leonard, and A. Pompe, J. Chem. Phys. **44**, 4206 (1966).
[3] D. Henderson and L. Oden, Mol. Phys. **10**, 405 (1966).
[4] S. Kim, D. Henderson, and L. Oden, Trans. Faraday Soc. **65**, 2308 (1969).
[5] K. M. Dyer, J. S. Perkyns, and B. M. Pettitt, Theor. Chem. Acc. **105**, 244 (2001).
[6] L. G. MacDowell, C. Menduina, C. Vega, and E. de Miguel, J. Chem. Phys. **119**, 11 367 (2003).
[7] P. G. Kusalik, F. Liden, and I. M. Svishchev, J. Chem. Phys. **103**, 10 169 (1995).
[8] F. R. Ree and W. G. Hoover, J. Chem. Phys. **40**, 939 (1964); **46**, 4181 (1967); E. J. Janse van Rensburg, J. Phys. A **26**, 4805 (1993); A. Y. Vlasov, X. M. You, and A. J. Masters, Mol. Phys. **100**, 3313 (2002); N. Clisby and B. M. McCoy, J. Stat. Phys. **114**, 1361 (2004).
[9] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, San Diego, 2002).
[10] E. A. Mason and T. H. Spurling, *The Virial Equation of State* (Pergamon Press, Oxford, 1969).
[11] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[12] S. Rast, P. H. Fries, and H. Krienke, Mol. Phys. **96**, 1543 (1999); S. Labik, H. Gabrielova, J. Kolafa, and A. Malijevsky, Mol. Phys. **101**, 1139 (2003).
[13] N. D. Lu and D. A. Kofke, J. Chem. Phys. **115**, 6866 (2001).
[14] N. D. Lu, J. K. Singh, and D. A. Kofke, J. Chem. Phys. **118**, 2977 (2003).