FLUID PHASE EQUILIBRIA

ELSEVIER Fluid Phase Equilibria 150–151 (1998) 41–49

Precision and accuracy of staged free-energy perturbation methods for computing the chemical potential by molecular simulation

David A. Kofke [a,*], Peter T. Cummings [b,c]

[a] *Department of Chemical Engineering, State University of New York, Buffalo, NY 14260-4200, USA*
[b] *Department of Chemical Engineering, University of Tennessee, Knoxville, TN 37996-2200, USA*
[c] *Chemical Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6268, USA*

Abstract

We examine the performance of free-energy perturbation methods when applied to compute the chemical potential of the Lennard–Jones model by Monte Carlo simulation. We emphasize the accuracy and precision of various implementations of the methodology, particularly in the context of the relative effectiveness of ‘insertion’ vs. ‘deletion’ approaches. The study is limited to a single state point and system size. In accord with recent arguments made in the context of the hard-sphere model, we find that any single- or multi-stage approach that incorporates a ‘deletion’ component shows greatly diminished accuracy and precision when compared to its ‘insertion’ counterpart. We also confirm our earlier conjecture that the entropy rather than the free energy is the important quantity to examine when formulating optimal multistage free-energy perturbation schemes. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Chemical potential; Method of calculation; Molecular simulation

1. Introduction

The chemical potential is a key quantity in the calculation of phase and chemical equilibria. Therefore its evaluation (explicitly or implicitly) is a high priority in simulations aiming to study these phenomena. Often this measurement is very difficult to perform with any precision. The origin of the difficulty is typically explained by pointing out that molecular configurations that contribute most to the measurement are not well sampled during a conventional simulation. This explanation tells only half of the story. It must also be noted that the poor sampling can come about from one of two

---

[*] Corresponding author. Tel.: +1-716-645-2911; fax: +1-716-645-3822; e-mail: kofke@eng.buffalo.edu

PII: S0378-3812(98)00274-X

fundamentally different mechanisms, one entropic and the other energetic [1]. The efficient and successful application of an algorithm for computing the chemical potential requires an appreciation of these effects. A lack of understanding here can (and has) lead to a misapplication of methodology and, more seriously, a misinterpretation of results. The purpose of this paper is to demonstrate these points through application of staged free-energy perturbation methods to the calculation of the chemical potential of the Lennard–Jones model system.

Available methods for computing the chemical potential by molecular simulation are of four types. In all cases the goal is to evaluate the free-energy difference between systems differing in the presence of a single molecule. This difference is related in a trivial way to the chemical potential. The (overlapping) categories are as follows [1].

Free-energy perturbation. This method permits computation of the free-energy difference between two systems while simulating only one of them. It is the focus of this paper.

Expanded ensembles [2–5]. In this approach, the system wanders (with bias) among different thermodynamic states, or ensembles, and the frequency with which the states are visited permits evaluation of the free-energy differences.

Thermodynamic integration. In simplest form this involves the straightforward numerical integration of the fundamental thermodynamic equation of a chosen ensemble. The most useful extensions formulate non-traditional thermodynamic integration pathways.

Histogram/distribution methods [6–9]. These approaches gather the greatest possible amount of information from a (or each) simulation, perhaps add some understanding from theory, and thereby yield a value for the chemical potential. We have identified [1] these methods as the most likely route to advance the chemical-potential measurement methodology, but we do not consider them further in this report.

2. Free-energy perturbation methods

2.1. Single-stage methods

The working equation for free-energy perturbation connects the Helmholtz free energies A for two systems (subscript 0 and 1, respectively) that differ in some specific way [10]

$$\exp[-\beta(A_1 - A_0)] = \langle \exp[-\beta(\Phi_1 - \Phi_0)] \rangle_0 \tag{1}$$

where $\beta = 1/kT$ with T the absolute temperature and k Boltzmann's constant; Φ is the Hamiltonian. The angle-brackets describe a canonical-ensemble average, and the '0' subscript on it indicates that configurations in the ensemble are weighted according to the potential Φ_0. Obviously, Eq. (1) can be written with the '0' and '1' subscripts interchanged, so in principle either system may serve as the 'reference'. In many instances Φ is defined as the intermolecular potential for a system of N or $N-1$ particles, U_N or U_{N-1}, respectively.

Taking the '0' system in Eq. (1) as having $N-1$ particles, and the '1' system as having N, $A_1 - A_0$ is the chemical potential and we obtain Widom's method [11,12]

$$\exp[-\beta\mu_{\mathrm{r}}] = \langle e_t \rangle_{N-1} \tag{2}$$

where $e_t = \exp(-\beta u_t)$ is the Boltzmann factor of the energy $u_t \equiv U_N - U_{N-1}$ of the Nth 'test' particle. The method works well if sufficiently many configurations find no test-particle overlap, i.e., if configurations important to the N-particle system are well sampled. Alternatively, if we take the '0' system having N particles and the '1' system comprising $N-1$, we derive the 'test-particle removal' formula [7] (and references therein)

$$\exp[+\beta\mu_r] = \langle 1/e_t \rangle_N \tag{3}$$

The test particle may be any of the N (interacting) spheres. In practice a simulation based on this formula is completely unreliable, as the average contains an enormous contribution from configurations that are never sampled. The outcome is a consistent, practically systematic, underestimation of the chemical potential.

Superficially, test-particle insertion and removal are symmetric operations, so it is perhaps surprising to see that measurement methods based on each have drastically different levels of reliability. In the context of Widom insertion/deletion this outcome is well known [7,13–17] and well (if not widely) understood. A diagram is helpful in conveying the nature of the asymmetry. Fig. 1 presents a highly schematic depiction of phase space—the $3N$-dimensional space of particle configurations. The dark oval represents all configurations that are important to the system of $N-1$ particles, and the white square within it represents all configurations important also to the N-particle system. Here 'important' means those configurations that contribute significantly to the partition function. The significant feature is that the N-important region lies within the $N-1$ important region, and may represent a very small fraction of it (say of the order of 10^{-7} or smaller). This situation is rigorously true for hard potentials, and it is likely to be a highly accurate depiction of systems of soft particles and even systems for which attractive forces are important (say, a fluid near its triple point), particularly if the intermolecular potential is isotropic.

In Widom's insertion method, the system samples the $(N-1)$-important region, and finds contributions to the ensemble average of Eq. (2) in those (sometimes rare) instances in which the N-important region is encountered by chance. Here the barrier to sampling is *entropic*. The probability of sampling the significant region is in proportion to its fractional representation, e.g., it is a probability of order 10^{-7}. In contrast, the particle deletion method samples the N-important region

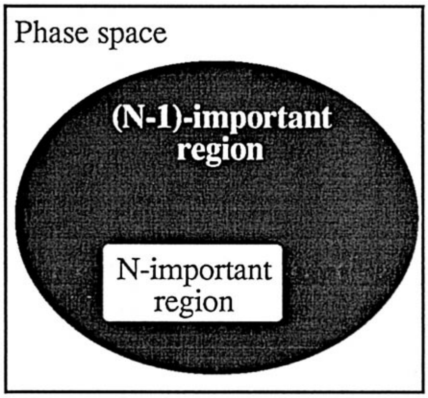

Fig. 1. Schematic depiction of phase space, and the relation between regions important to systems of $N-1$ and N particles, respectively.

while the greatest contribution to the ensemble average of Eq. (3) lies in those $(N-1)$-important configurations outside of this region. The significant configurations are not sampled because they are of a prohibitively high energy, i.e., the barrier to sampling is *energetic*. The probability of sampling the significant region is in proportion to the Boltzmann factor, which may be an *extremely* tiny probability, e.g., 10^{-50} or smaller. The contribution to the average is inversely proportional to the likelihood of sampling, so missing these low-probability configurations has a disastrous consequence for the accuracy of the measurement. The problem is especially insidious because the error statistics for the measurement do not betray the trouble, which is to say the method returns a result that is precisely wrong.

2.2. Multi-stage methods

Multi-stage methods are based on the idea of constructing the free-energy difference of interest $(A_N - A_{N-1})$ via $(A_N - A_W) + (A_W - A_{N-1})$, where A_W is the free energy for an intermediate system that is formulated to facilitate the chemical-potential calculation. This intermediate is defined by the potential Φ_W, which in turn is defined via a weight function W: $\Phi_W = U_{N-1} - \ln W$. Four approaches may be constructed depending on how one chooses to compute the component free energy differences $(A_N - A_W)$ and $(A_W - A_{N-1})$ using Eq. (1). These are summarized in Table 1. The working formulas there specify the ensemble averages needed in the systems of N particles, $N-1$ particles, or the intermediate W, as indicated.

Umbrella sampling is perhaps the best known of the methods [18,19]. It has the advantage of giving the chemical potential by a single simulation in which two averages are recorded while sampling the W system. Bennett's method [6] is in some sense the opposite: it requires simulations of both the $N-1$ and N-particle systems. Because the weighting function W has no bearing on the sampling of configurations, this method is amenable to analytic optimization. Bennett completed this, and his name is usually associated with the staging approach together with the optimized prescription for W. Staged insertion and staged deletion are less widely considered as techniques for measuring the chemical potential, although they are routinely practiced in the computational chemistry community [20].

The considerations of entropic vs. energetic barriers to sampling apply—albeit to a lesser degree—when free-energy perturbation methods are extended to these multistage forms. In particular the multistage methods suffer to the extent that they contain a 'deletion' staging component, defined

Table 1
Four approaches to splicing single-stage free-energy perturbations to construct multistage forms

Staging approach	Name	Formula $\exp[-\beta\mu_r] =$
$(N-1) \leftarrow W \rightarrow N$	Umbrella sampling	$\langle e_t/W\rangle_W \langle 1/W\rangle_W^{-1}$
$(N-1) \rightarrow W \leftarrow N$	Bennett's method	$\langle W/e_t\rangle_N^{-1} \langle W\rangle_{N-1}$
$(N-1) \rightarrow W \rightarrow N$	Staged insertion	$\langle e_t/W\rangle_W \langle W\rangle_{N-1}$
$(N-1) \leftarrow W \leftarrow N$	Staged deletion	$\langle W/e_t\rangle_N^{-1} \langle 1/W\rangle_W^{-1}$

The arrows point from the system in which the sampling is being performed (the '0' system of Eq. (1)), to the perturbation system (the '1' system).

such that the barrier to good sampling is *energetic* in origin. By tinkering with the weighting function, umbrella sampling and Bennett's method can overcome this handicap to some degree, but for umbrella sampling this involves a delicate balancing, while with Bennett's optimization the method reverts to something that is little different than a single-stage insertion. The only one of the four approaches that has no 'deletion' component is staged insertion, and indeed our study of these methods in the context of the hard-sphere model found this approach to be highly effective [1].

3. Method

3.1. Definition of intermediate stage

We consider two-stage free-energy perturbation methods in which the intermediate stage has the Nth 'test' particle interacting with the remaining $N-1$ particles as a simple hard sphere of diameter α. Thus $W=0$ if the center of the test particle lies within a distance α of another particle center, and $W=1$ otherwise. This choice of an intermediate stage immediately precludes application of the umbrella-sampling and the staged-deletion methods of Table 1, because both of these methods will always yield the value zero for the residual free energy difference $A_W - A_{N-1}$. This particular selection of W was recently advocated by Parsonage [21] in a Bennett's-form staging scheme, and it is very similar to one proposed earlier by Han et al. [16] (again in a Bennett's-form context).

3.2. Description of simulations

We conducted canonical-ensemble Monte Carlo (MC) simulations of 107 Lennard–Jones (LJ) particles and one hard sphere (HS) at a density $\rho\sigma^3 = 0.9$ and temperature $kT/\varepsilon = 1.2$. The LJ potential was truncated at a separation of 2.5σ and no long-range correction was applied. Each simulation sampled 10,000 cycles, where one cycle is defined as one attempted MC trial per particle. Among the trials were moves that attempted to swap the HS with a LJ particle, and attempts to re-insert the HS at a random position. At the end of each cycle the energy change associated with transforming each LJ particle individually into a HS was measured and kept as the running average of $\langle W/e_t \rangle_N$; the energy change associated with the transformation of the HS to a LJ particle was measured and kept as the running average of $\langle e_t/W \rangle_W$. Also at the end of each cycle one Widom insertion was performed to measure both $\langle W \rangle_{N-1}$ and $\langle e_t \rangle_{N-1}$ (respectively, HS and LJ insertion averages). Several values of the HS diameter α were examined between the values 0.9 and 0.5. For each value of α, 50 completely independent 10,000-cycle simulations were conducted, each starting from its own, independently equilibrated initial configuration.

It is clear that for the purpose of economizing the calculations we have taken some liberty in measuring the averages $\langle W/e_t \rangle_N$, $\langle W \rangle_{N-1}$, and $\langle e_t \rangle_{N-1}$. Strictly, the system being simulated is the intermediate W system, so only averages of the form $\langle \ldots \rangle_W$ are obtainable. The error incurred with this approximation is not likely to be much worse than other finite-size effects that are known to be significant for $N = 108$ at this density. Our aim in the present study is not to determine a good infinite-system value for the chemical potential, but instead to demonstrate some basic points about how staging calculations should be performed. We will see that the consequences of bad staging overwhelm any inaccuracies attributable to the finite size of the simulated system.

4. Results and discussion

The results are listed in Table 2. We report the mean of the 50 independent ensemble averages gathered for each value of the HS diameter α, thus each value represents the average from 500,000 simulation cycles. In different instances we present the average with its associated variance from the 50 simulations, or with its standard error. These two statistics in the present case are related by the formula (standard error) = (variance/50)$^{1/2}$. For each α, the chemical potential is reported twice: once as computed via single-stage Widom insertion ($\langle e_t \rangle_{N-1}$), and once as computed via the two-stage insertion using the HS intermediate ($\langle W \rangle_{N-1} \langle e_t/W \rangle_W$). We note that all two-stage values of the chemical potential are mutually consistent—considering their standard errors—and in particular there seems to be no trend with α. This gives us confidence that we have not introduced a serious error by taking the averages $\langle \ldots \rangle_N$ and $\langle \ldots \rangle_{N-1}$ during simulations of the W system. The single-stage values are not consistent with the two-stage results, even considering the standard errors. This is a point we will return to shortly.

In principle, the averages for the second stage HS → LJ 'insertion' transformation ($\langle e_t/W \rangle_W$) and its 'deletion' opposite LJ → HS ($\langle W/e_t \rangle_N$) should be equal in magnitude and opposite in sign. It is clear from Table 2 that this is not the case. The deletion value is consistently larger in magnitude, which translates into a consistently smaller value for the chemical potential. This outcome is evident in the results of Han et al. [16] and Parsonage [21]—both studies reported chemical potentials consistently smaller than the straight single-stage Widom insertion values. Clearly our second-stage deletion–insertion discrepancy must be attributed to inadequate sampling in the deletion calculation, which (not surprisingly) returns a result that is virtually (and incorrectly) independent of α. Parsonage noticed this breakdown in his studies, but did not consider it to be a significant problem until α is taken somewhere less than 0.875. We see problems in evidence even for α as large as 0.9. First we note the contrast between the variances of the insertion and deletion methods: the variance associated with HS → LJ is two orders of magnitude smaller than that with LJ → HS. A particularly worrisome feature of the LJ → HS calculation is demonstrated in Fig. 2. The bias in the stochastic error for the LJ → HS calculation is striking. Only three of the 50 10,000-cycle simulations return an average that is below the correct value; all others are above. When $\langle W/e_t \rangle_N$ (as opposed to the logarithm of it

Table 2
Simulation results for the various staging elements

HS	Partial insertion/deletion averages			Chemical potential, $\beta\mu_r$	
diameter, α	LJ → HS $-\ln\langle W/e_t \rangle_N$	HS → LJ $-\ln\langle e_t/W \rangle_W$	HS insertion $-\ln\langle W \rangle_{N-1}$	One stage $-\ln\langle e_t \rangle_{N-1}$	Two stages $-\ln[\langle W \rangle_{N-1} \langle e_t/W \rangle_W]$
0.90	6.30 (1.6)	−6.21 (0.02)	8.73 (0.41)	2.7_4	2.52_9
0.88	5.77 (4.8)	−5.47 (0.04)	7.98 (0.22)	1.4_5	2.51_7
0.85	5.57 (5.7)	−4.38 (0.07)	6.96 (0.10)	4.9_3	2.58_6
0.82	5.00 (20)	−3.43 (0.16)	5.97 (0.03)	4.6_4	2.54_6
0.80	3.43 (19)	−2.78 (0.29)	5.38 (0.02)	1.8_7	2.60_7
0.70	4.70 (19)	−0.59 (1.7)	2.93 (0.002)	2.6_5	2.34_{18}
0.50	5.72 (7.8)	+1.66 (5.0)	0.63 (0.000)	3.3_4	2.29_{30}

Values inside parentheses are the variance associated with the tabled value, while subscripted values describe the standard error in the last digit of the tabled value.

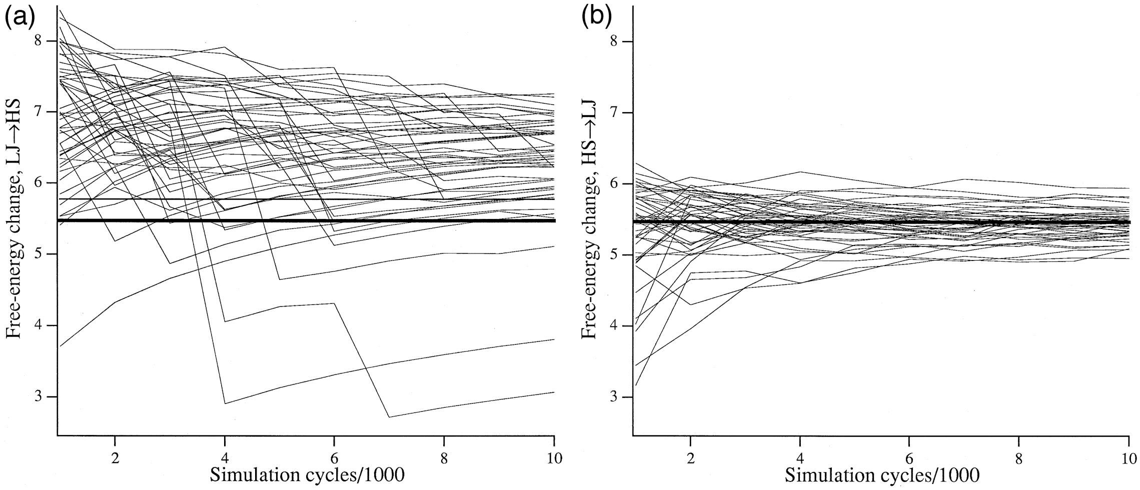

Fig. 2. Running averages for the second stage of a two-stage free-energy perturbation employing a hard-sphere intermediate of diameter $\alpha = 0.88$. Thin lines represent each of fifty independent 10,000-cycle simulations, and the thick horizonal line at 5.47 describes the best estimate of the correct value for the 108-particle system. (a) 'deletion' calculation LJ → HS; (b) 'insertion' calculation HS → LJ.

presented in the figure) is averaged over all 50 runs, these three are sufficient to bring the overall chemical potential reasonably close to the correct result, although it remains too high (see Table 2). In sharp contrast the HS → LJ calculation yields a result that is not biased in either direction while exhibiting a much smaller variance.

It is of interest to examine the effectiveness of the staged-insertion technique as a function of the intermediate-stage HS diameter α. We are especially interested in how the effectiveness is associated with the difference between the entropy change for the first stage and the entropy change for the second stage, $\Delta(\Delta S) = \Delta S_{\mathrm{HS} \to \mathrm{LJ}} - \Delta S_{\mathrm{HS}}$. It has been advocated [22,23] that this difference be formed for the *free energy* changes rather than the entropy changes, and that optimum staging occurs when the difference is zero. The qualitative arguments presented in the introduction (and in more detail in our study using the hard-sphere model [1]) have led us to speculate that the entropy, not the free energy, is the appropriate quantity to consider to optimize the staging. The entropy differences are easily computed from the simulation data. For an infinite system there is no energy change associated with the insertion of HS, so $\Delta S_{\mathrm{HS}}/k = -\beta \Delta A_{\mathrm{HS}} = \ln\langle W \rangle_{N-1}$. The energy change associated with converting the HS to a LJ particle is just twice the average energy per particle, which has been reported by Parsonage [21] as $U/N = -5.36$. Thus $\Delta S_{\mathrm{HS} \to \mathrm{LJ}}/k = \beta(-\Delta A + \Delta U)_{\mathrm{HS} \to \mathrm{LJ}} = \ln\langle e_t/W \rangle_W - 8.93$. The entropy-change difference and the free-energy-change difference are displayed in Fig. 3. In this figure we also present the reduced variance $M\sigma^2$, where $M = 10{,}000$ is number of free-energy perturbation measurements per simulation; this group is expected [1] to be asymptotically independent of M for $M \to \infty$. Fig. 3 confirms that the entropy, and not the free-energy, is the appropriate quantity to consider when defining an intermediate stage that minimizes the overall error in the calculation. The minimum variance occurs at slightly larger α than predicted by the entropy rule. This discrepancy has to do with the measurement of the energy component of the total free-energy change, and can be explained quantitatively (to be described in future work).

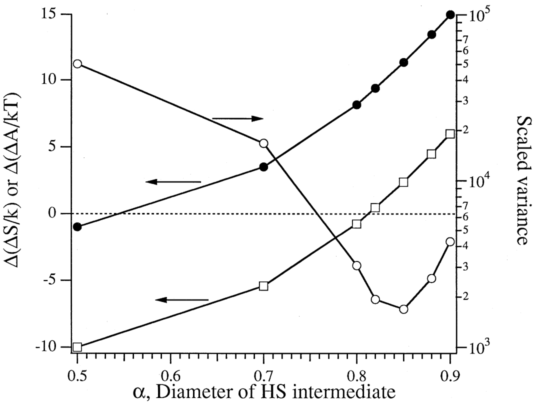

Fig. 3. Difference $\Delta(\Delta S/k)$ in entropy change between the stages of a two-stage free-energy perturbation calculation (open squares). Corresponding quantity for free energy is presented also (filled circles). Intersection with the dashed line indicates the zero. Scaled variance $M\sigma^2$ is plotted on the right ordinate (open circles). All results are computed from MC simulations described in the text.

Before closing, we should note that the free-energy perturbation calculation in the 'insertion' direction is not without its pitfalls. A careful examination of Fig. 3b (and the single-stage insertion results of Table 2) finds that the fluctuations in a given 10,000-cycle simulation are misleading, and would indicate an error estimate that is smaller than actually warranted. Nevertheless the approach is far superior to the equivalent 'deletion' calculation (where the same situation applies) because (1) the true error is in fact smaller; (2) there is little or no systematic bias to the error; and (3) it is possible, by analyzing the entropy change and applying probabilistic arguments similar to those presented in our hard-sphere study [1], to know how much sampling is needed to obtain credible results. One does not have to rely on the simulation statistics alone to apply confidence limits to the free-energy change. A similar analysis cannot reliably be performed for the deletion analog.

5. Conclusions

We have demonstrated the substantial asymmetry of free-energy perturbation calculations in application to a realistic molecular model. It is critical always to apply this methodology in the 'insertion' direction, i.e., the ensemble governing the sampling should be of greater entropy than the perturbation ensemble. We have confirmed our earlier conjecture that the entropy difference is the important quantity to consider when optimizing staged free-energy perturbation calculations. The lessons learned by studying these methods in the context of hard spheres have been shown to possess broader applicability.

Acknowledgements

The research effort of PTC was supported by the Division of Chemical Sciences, Office of Basic Energy Sciences, U.S. Department of Energy (DOE). The research effort of DAK was performed as part of a sabbatical leave at the University of Tennessee (UT), partially supported by Oak Ridge National Laboratory (ORNL) through funding provided to UT to support the UT/ORNL Distinguished Scientist Program. ORNL is managed by Lockheed Martin Energy Research for the DOE under Contract No. DE-AC05-96OR22464.

References

[1] D.A. Kofke, P.T. Cummings, Mol. Phys. 92 (1997) 973.
[2] I. Nezbeda, J. Kolafa, Mol. Sim. 5 (1991) 391–403.
[3] A.P. Lyubartsev, A.A. Marsinovski, S.V. Shevkunov, P.N. Vorontsov-Velyaminov, J. Chem. Phys. 96 (1992) 1776–1783.
[4] R.D. Kaminsky, J. Chem. Phys. 101 (1994) 4986–4994.
[5] P. Attard, J. Chem. Phys. 98 (1993) 2225–2231.
[6] C.H. Bennett, J. Comp. Phys. 22 (1976) 245–268.
[7] K.S. Shing, K.E. Gubbins, Mol. Phys. 46 (1982) 1109–1128.
[8] A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett. 61 (1988) 2635–2638.
[9] A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett. 63 (1989) 1195–1198.
[10] R.W. Zwanzig, J. Chem. Phys. 22 (1954) 1420–1425.
[11] B. Widom, J. Chem. Phys. 39 (1963) 2808–2812.
[12] J.L. Jackson, L.S. Klein, Phys. Flu. 7 (1964) 228–231.
[13] J.G. Powles, W.A.B. Evans, N. Quirke, Mol. Phys. 46 (1982) 1347–1370.
[14] J.G. Powles, Chem. Phys. Lett. 86 (1982) 335–339.
[15] B. Guillot, Y. Guissani, Mol. Phys. 54 (1985) 455–465.
[16] K.-K. Han, J.H. Cushman, D.J. Diestler, J. Chem. Phys. 93 (1990) 5167–5171.
[17] N. Parsonage, J. Chem. Soc., Faraday Trans. 92 (1996) 1129–1134.
[18] J.P. Valleau, D.N. Card, J. Chem. Phys. 57 (1972) 5457–5462.
[19] G.M. Torrie, J.P. Valleau, J. Comp. Phys. 23 (1977) 187–199.
[20] P. Kollman, Chem. Rev. 32 (1993) 2395–2417.
[21] N.G. Parsonage, Mol. Phys. 89 (1996) 1133–1144.
[22] D.A. Pearlman, P.A. Kollman, J. Chem. Phys. 90 (1989) 2460–2470.
[23] T.P. Straatsma, J.A. McCammon, J. Chem. Phys. 95 (1991) 1175–1188.