Quantitative comparison and optimization of methods for evaluating the chemical potential by molecular simulation

By DAVID A. KOFKE† and PETER T. CUMMINGS‡§

† Department of Chemical Engineering, State University of New York, Buffalo, New York 14260-4200, USA
‡ Department of Chemical Engineering, University of Tennessee, Knoxville, Tennessee 37996-2200, USA
§ Chemical Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6268, USA

(Received 10 April 1997; accepted 2 June 1997)

The precision of several methods for computing the chemical potential by molecular simulation is investigated. The study does not apply molecular simulation to the analysis but instead works with models of the simulation process. These models enable the variance of the chemical potential to be computed accurately and very quickly and thereby permits the methods (free-energy perturbation, expanded ensembles, thermodynamic integration, and histogram-distribution methods) to be optimized and compared over a range of densities. The study focuses exclusively on the hard-sphere model. This model is simple and well characterized; yet it exhibits the essential features that make the chemical potential calculation difficult; arguments are presented to support the broader applicability of the study. The severe asymmetry of particle insertion against particle deletion is highlighted, and it is shown that any staged free-energy perturbation method with a ‘deletion’ component is highly prone to systematic error. More generally this implies that such methods should always be staged in the direction of decreasing entropy. Other findings show that uniform sampling is not optimal for umbrella-sampling and expanded-ensemble applications, although it remains a good rule of thumb for tuning these approaches. Among the techniques we study, optimally staged insertion and the distribution-histogram methods are the most efficient and precise. The latter is effective only when used in an interpolative fashion, and we identify it as the most likely route to further progress in the field.

1. Introduction

A knowledge of the chemical potential is indispensable when calculating thermodynamic phase and reaction equilibria. Methods for computing the chemical potential by molecular simulation have a long history, and it remains a subject of continuing interest. The reason is that chemical potential calculations are among the most troublesome encountered routinely in simulation. Admittedly, they can always be completed if one expends sufficient computational effort. However, the ease and reliability of the chemical potential determination often decide whether a calculation of greater intrinsic interest (e.g., a phase diagram) is completed well or is even attempted at all. The introduction of a new method for computing the chemical potential is often accompanied by an application that shows it is capable of yielding a correct result, usually in comparison with established data. Researchers vary in the degree to which they examine whether the method is more efficient at producing its correct result. Some work [1–4] has addressed in a quantitative fashion the efficiency of various techniques, but such studies are scarce and none is very broad in scope. The primary aim of the present effort is to address this deficiency. We are able to go well beyond previous endeavours because we work with models of the simulation process, sidestepping the tricky and computationally expensive problem of relying on simulations to provide their own error statistics. Simulations are used in only a few instances, to verify the models. The only prior study to take this approach is that of Nezbeda and Kolafa [2] which has features common to the present work but is narrower in scope.

Computing the chemical potential is a special case of the more general problem of computing a free-energy difference, and the inherent difficulty of this calculation is well understood [5–7] An abstraction of the problem is presented in figure 1(a). In general, one wishes to know the difference in free energy $A_1 - A_0$ between two ‘states’ (indicated by the subscripts 0 and 1) that
differ in some way; temperature, intermolecular potential, a molecular conformation (i.e. trans as against gauche) are typical examples. For the chemical potential calculation the two states differ in the presence of a single molecule. The large square represents the entire $3N$ (or greater)-dimensional space of configurations, and the two enclosed regions represent those configurations important to the 0 and 1 states respectively (‘important’ meaning that they contribute significantly to the corresponding partition function, e.g. 99% of the contribution is found within). Calculation of $A_1 - A_0$ requires not only that both 0- and 1-important regions be well sampled, but also that the region between them be examined (or otherwise understood). The problems introduced by this requirement are diminished to the extent that the regions overlap. Lost in this simple rendition is the complexity of the shape of the 0- and 1-important regions (a wispy, barely connected grey-scale picture would provide a better conception than the shapes of figure 1(a)) and the concomitant difficulty of identifying (or defining) the region ‘between’ them.

This paper is focused in two ways. First, we restrict our attention to the chemical potential rather than to the more general problem of computing free-energy differences. Second, we conduct our study entirely within the context of the hard-sphere model. At first glance this choice may seem to limit severely the general applicability of our results, but in fact it is not a serious restriction. On the contrary, the hard-sphere model captures the essential difficulty of the chemical potential calculation and, because so much is known about the hard-sphere system (in particular its equation of state and its structure), we can formulate quantitative models of the calculation process itself. With these models we can systematically optimize, compare and ultimately better understand the known methods for computing the chemical potential.

We can begin by pointing out that all non-overlap configurations of hard spheres contribute equally to the partition function. So what happens when the $N$th sphere is added to a system of $N-1$ spheres, as needed to complete the chemical potential calculation? The picture is as in figure 1(b). The configurations important to the $N$-particle system form a wholly contained subset of those important to $N-1$ particles. There is complete overlap of the two regions, and the problem is not in bridging them but in focusing sufficient attention on the $N$-important region, which often represents an extremely small fraction of the $(N-1)$-important region (for hard spheres at freezing the fraction is $10^{-7}$). Why is this observation significant? First, it highlights the severe asymmetry of the relation between the two states, a point which is lost if they are conceived as in figure 1(a). Second, it brings up the notion of entropic as opposed to energetic origins of poor sampling. The $(N-1)$-particle system might not sample well the $N$-important region simply because it cannot find it, whereas the $N$-particle system does not sample the

---

**Figure 1.** Schematic representation of configuration space, and the regions important to free-energy calculations. In each figure the large square represents all the configuration space for $N-1$ particles. (a) Situation in which important regions have no substantial overlap and are of comparable size. (b) Situation characteristic of systems differing in the presence of a single molecule. The darkened $N$-important region is wholly within (and part of) the $(N-1)$-important region. (c) Same as (b), except describing how anisotropic attractive interactions can give rise to $N$-important configurations that are not important to the $(N-1)$-particle system.
(\(N - 1\))-important region because of a strong energetic prohibition. Indeed, this prohibition must be very strong if it can so greatly contract the configuration space of the \(N\)-particle system (\(\Delta U/kT \approx 20\) for a \(10^{-7}\) contraction). Such energetics are seen only in the repulsive portion of the intermolecular potential, that is the part that is distilled into the hard-sphere model. Finally, we note that (for the hard-sphere model), if one observes only the energy landscape, nothing is learned by simulating exclusively the \(N\)-important region, because within its confines there are absolutely no features to be probed. Instead one must examine it from the outside, in the context of the \(N - 1\) region, to which it represents the only interesting feature. Alternatively, one can look for features in the configuration landscape, which has a much richer and more intuitive texture.

How does this picture change when the intermolecular potential evolves to more realistic forms? First, the addition of more segments to the sphere (forming molecules) increases the magnitude of the difficulty, but much of the qualitative character remains. Remarkably effective methods have been developed [8–13] specifically for molecular systems, and our present focus prevents us from studying them. Nevertheless what we learn from studying atomic systems is relevant to these approaches. Second, softening the potential blurs the outline of the \(N\)-important region but again introduces no qualitative changes to the problem. It remains (almost) always possible to form a cavity that accommodates a new sphere with squeezing the other molecules into the high-energy configurations found outside the \((N - 1)\)-important region. Third, for isotropic attractions, such as dispersion forces, nearly the same can be said. However, such interactions will add texture to the energy landscape of the \(N\)-important region and thereby enhance the value of strategies that involve sampling it exclusively. This is an element of the general problem that is not captured by a hard-sphere-based study. On the other hand, the energy differences involved are much smaller than those associated with the repulsive interactions; so probing these features is a simpler and less important problem than the sampling needed to characterize the \(N\)-important region in the context of the \((N - 1)\)-important region.

The situation is more complex if strong orientation-dependent attractions (such as hydrogen bonding) are involved. Even so, insertion of a simple isotropic (e.g. non-polar) solute into a hydrogen-bonding solvent is much as already described. A structured solvent can rearrange itself to accommodate such a solute without incurring a prohibitive energy penalty; that this seems unlikely is only a restatement of the entropy effect mentioned above. A qualitative difference may finally arise if the solute hydrogen bonds as well. Then the \(N\)-impor-
tant region would include configurations where some of the solvent molecules are oriented to bond with the solute. Such configurations (with the solute removed) might not contribute significantly to the \((N - 1)\) system, as shown in figure 1 (c). Nevertheless an important feature remains, namely that the \(N\)-important region occupies a much smaller portion of configuration space than does the \((N - 1)\)-important region. Poor sampling of the \(N\)-important region by the \(N - 1\) system may now have both energetic and entropic origins, while poor sampling of the \(N - 1\) region by the \(N\)-particle system is still due to strongly unfavourable energetics. Their relation is still very asymmetric.

We make these points to reinforce the contention that the hard-sphere model exhibits the most important features of the problem of computing the chemical potential by molecular simulation. Their consequences are discussed in the next section, where we review the established methods for measurement of the chemical potential and describe in particular how they are applied to hard spheres. This lays out the formalism for the measurement process. In section 3 we discuss the matter of formulating a weighting function that focuses sampling on the \(N\)-important region sufficiently but not excessively. In section 4 we describe our means for modelling the (Markov) process of measuring the hard-sphere chemical potential, and we apply our models to the methods of section 2. We close with concluding remarks in section 5.

2. Background

Methods for evaluating the chemical potential have been reviewed many times [1, 5–7, 14–16] but it is worthwhile to present them again to facilitate the discourse that follows. First, let us point out the quantity of interest from a simulation is the residual chemical potential \(\mu_r\), that is the value above that of the ideal gas at the same density \(\rho\) (henceforth all quantities are presented in units such that the hard-sphere diameter \(d\) and the temperature \(k_B T\) are unity):

\[
\mu_r = \mu - (\mu_0 + \ln \rho).
\] (1)

Unless otherwise specified, ideal-gas contributions will not be included in anything that follows.

We group the available methods for chemical potential measurement into four broad classes, which are not mutually exclusive: free-energy perturbation methods; expanded ensembles; thermodynamic integration; distribution–histogram methods. In every case the goal is to connect a system of \(N - 1\) spheres with one of \(N\) spheres, because the difference in the respective free energies is exactly the desired chemical potential.
2.1. Free-energy perturbation

Methods in the first group are based on the identity [17]

\[
\exp\left[-(A_1 - A_0)\right] = \frac{Z_1}{Z_0} = \left\langle \exp\left[-(\varphi_1 - \varphi_0)\right]\right\rangle_0.
\]  

(2)

Here, \(A\) and \(Z\) are the Helmholtz free energy and the canonical configurational integral, respectively, and \(\varphi\) is the (configuration-space) Hamiltonian. The subscripts 0 and 1 refer to systems that differ in how the potential \(\varphi\) depends on the particle coordinates \(q^N\) (the systems may differ in more general ways, but this interpretation serves our purposes). The angular brackets describe a canonical ensemble average, and the subscript 0 on them indicates that configurations in the ensemble are weighted according to the (effective) potential \(\varphi_0\), that is \(\langle X\rangle_0 = \frac{1}{Z_0} \int X \exp(-\varphi_0) \, dq^N\). Obviously, equation (2) may be written with the subscripts 0 and 1 interchanged; so in principle either system may serve as the ‘reference’. In many instances, \(\varphi\) is defined as the intermolecular potential for a system of \(N\) or \(N - 1\) particles, \(U_N(q^N)\) or \(U_{N-1}(q^N)\) respectively; in the latter case the coordinate of the \(N\)th molecule has no effect on the potential.

2.1.1. Single-stage methods

The simplest means for measuring the chemical potential is associated with the name of Widom [18] and derives directly from equation (2). Taking the 0 system as having \(N - 1\) particles, and the 1 system as having \(N\), \(A_1 - A_0\) is the chemical potential and we obtain [18, 19]

\[
\exp(-\mu_t) = \left\langle e_i\right\rangle_{N-1},
\]  

(3)

where \(e_i = \exp(-u_i)\) is the Boltzmann factor of the energy \(u_i = U_N - U_{N-1}\) of the \(N\)th ‘test’ particle. For hard spheres, \(e_i\) is zero when the test particle overlaps one of the other \(N - 1\) particles, and it is unity otherwise. Thus the ensemble average is the fraction of all configurations of the \(N\) particles (selected to prohibit overlap among \(N - 1\) of them) in which the test particle overlaps none of the others. In practice the average is measured by occasionally inserting the test particle into the simulation volume, measuring \(e_i\) and then removing it before continuing the simulation; consequently this method is known as ‘test-particle insertion’. It works well if sufficiently many configurations find no test-particle overlap, that is if configurations important to the \(N\)-particle system are well sampled. We quantify ‘sufficiently many’ in section 4.

Alternatively, if we take the 0 system having \(N\) particles and the 1 system consisting of \(N - 1\), we derive the ‘test-particle removal’ formula (see [14] and references therein)

\[
\exp(+\mu_t) = \left\langle 1/e_i\right\rangle_N.
\]  

(4)

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 infinite contribution from configurations that are never sampled. One always recovers the nonsense result \(\mu_t = 0\), with no stochastic error. Equation (4) per se is not incorrect. Rather the flaw is its implementation, which has wholly inadequate sampling of configurations important to the \((N - 1)\)-particle system.

The test-removal catastrophe is not a pathology of the hard-sphere model, although its starkness in this instance provides a caricature of the general problem. It has been described and explained (at least in part) several times [14, 20–24] in connection with the Lennard-Jones model, but the nature and severity of the problem are still not universally appreciated. Referring to figure 1(b), in the general case (beyond hard spheres) equation (4) prescribes sampling with a very strong bias toward the \(N\)-important region; the probability of leaving the region vanishes as (for example) \(\exp(-1/r^{12})\), where \(r\) approaches zero as one moves farther from the \(N\)-important region (in fact the probability is significantly smaller, owing to contributions from multiple test-particle overlaps). A computer powerful enough to sample against this bias will never be built. This is a serious problem because the contribution from these unreachable regions is growing just fast enough to offset the bias against them; so their contribution to the ensemble average is significant. The difficulty that we describe here is qualitatively different from that associated with test-particle insertion. There the problem is entropic in origin (one is searching for the important region rather than fighting in vain to enter it), and it is feasible (although perhaps not desirable) to expend sufficient computation to complete the search by brute force.

2.1.2. Multistage methods

More sophisticated and more flexible methods can be developed by applying equation (2) in stages. The \((N - 1)\)- and \(N\)-particle systems can be joined by one or more intermediates that are constructed to facilitate the chemical potential calculation. Our initial discussion will focus on methods employing only one such intermediate, as the introduction of additional stages follows easily. To maintain consistency with the notation of previous work, we shall define the intermediate by the weighting function \(W(q^N)\) such that its potential \(\varphi_W(q^N) = U_{N-1} - \ln W\), and its free energy is denoted \(A_W\). The idea is to construct \((A_N - A_{N-1})\) via \((A_N - A_W) + (A_W - A_{N-1})\), with each difference given as a separate simulation average.

The first choice that one confronts is which of these three systems \((N - 1, W\) and \(N\)) will for each stage form the ‘reference’ and which will define the perturbation measured in the simulation (i.e. which is 0 and which
is 1, in the notation of equation (3)). Four possibilities arise, and they are summarized in table 1. If the $W$ system is selected as the reference and the $U_{N-1}$ and $U_N$ systems are given as separate perturbations, the ‘umbrella-sampling’ formula of Valleau and co-workers [25, 26] is obtained:

$$\exp(-\mu_t) = \frac{\langle e_t / W \rangle_W}{\langle 1/W \rangle_W}. \quad (5)$$

An important feature of this choice is that both averages (the numerator and the denominator) can be measured in a single simulation sampling on $\exp(-e_W)$. The numerator has its only contributions from configurations important to the $N$-particle system (because of the $e_t$ term), while the largest contributions to the denominator occur when the test particle overlaps another sphere. The weight $W$ must be delicately balanced to avoid tipping the sampling too much towards one or the other average. Surprisingly few applications of this formula to chemical potential calculations have been described [27, 28] Additional staging of the umbrella-sampling formula results in the ‘double-wide sampling’ technique of Jorgensen and Ravimohan [29] Compared with other multistage methods it has the apparent advantage of characterizing about twice as many intermediate stages as the number of simulations performed.

One might instead take the $(N-1)$- and $N$-particle systems as separate references, and the $W$ system as defining the perturbation on each. Then [30]

$$\exp(-\mu_t) = \frac{\langle W \rangle_{N-1}}{\langle W / e_t \rangle_N}. \quad (6)$$

An algorithm based on this formula differs from one based on equation (5) in several important ways. The numerator and denominator in equation (6) cannot both be computed in a single simulation; so two are needed. However, in neither simulation does the weight function $W$ influence the sampling of configurations. Consequently, it is easy to prescribe an optimum $W$, one that minimizes the variance in the measured chemical potential. This was done by Bennett [30] as part of the most insightful work ever published on this topic, and the corresponding method bears his name. Unfortunately in the present application, namely determining the chemical potential of hard spheres, the approach collapses and gives a result that is formally identical with test-particle insertion. This outcome is implicit in Bennett’s discussion and does not detract from his work because he was concerned with the more general problem of computing free-energy differences by molecular simulation. Bennett’s method (equation (6), but without optimizing the weighting functions) forms the basis of two nearly identical chemical potential calculation methods that have been recently proposed [23, 31] and a third that is very similar [32].

The failure of Bennett’s method in the present situation arises because of its partial reliance on test-particle removal. In particular the denominator in equation (6) does not sample all representative configurations important to the $W$ system. This insight points us to the only viable alternative to equation (2.5), a method that we shall call ‘staged insertion’:

$$\exp(-\mu_t) = \frac{\langle e_t / W \rangle_W}{\langle W \rangle_{N-1}}. \quad (7)$$

This formula retains the apparent disadvantage of specifying two separate simulations to determine the needed averages. Nevertheless, it may be hoped that equation (7) prescribes a superior approach because the perturbations described by both its averages are formulated in the most advantageous direction; that is, configurations important to the perturbation system are a subset of the configurations sampled. A good result does not rely on a chance excursion from the targeted sample space; rather it relies on a chance encounter within this space. The multistage extension of this method [33] is particularly useful, and its working equation is

$$\exp(-\mu_t) = \frac{\langle e_t / W \rangle_{w_1} \langle W_{n-1} / W_{n-2} \rangle \cdots \langle W_2 / W_1 \rangle_{w_{n-1}}}{\langle W / e_t \rangle_{N-1}}. \quad (8)$$

where $W_i$ is the weighting function for the $i$th stage.

The final method is the opposite of staged insertion, in the same sense that equation (4) describes a method that is opposite to equation (3). It might be termed ‘staged deletion’ or ‘annihilation’:

$$\exp(-\mu_t) = \frac{1}{\langle W / e_t \rangle_{N-1} \langle 1/W \rangle_W}. \quad (9)$$

It is not a good approach, and we shall not consider it in any detail.

The lessons taken from examination of the single-stage methods can be applied to understand the strengths and weaknesses of their multistage extensions.

Table 1. Four approaches to splicing single-stage free-energy perturbations to construct multistage forms. The arrows point from the system in which the sampling is being performed (the 0 system of equation (2)), to the perturbation system (the 1 system).

<table>
<thead>
<tr>
<th>Staging approach</th>
<th>Name</th>
<th>Working equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N - 1 \leftarrow W \rightarrow N$</td>
<td>Umbrella sampling</td>
<td>5</td>
</tr>
<tr>
<td>$N - 1 \rightarrow W \leftarrow N$</td>
<td>Bennett’s method</td>
<td>6</td>
</tr>
<tr>
<td>$N - 1 \rightarrow W \rightarrow N$</td>
<td>Staged insertion</td>
<td>7</td>
</tr>
<tr>
<td>$N - 1 \leftarrow W \leftarrow N$</td>
<td>Staged deletion</td>
<td>9</td>
</tr>
</tbody>
</table>
Figure 2 provides some depiction. Properly implemented, staged insertion simply homes in on the important region; we shall show that this is one of the most effective ways to compute the chemical potential. An ineffective implementation (not shown in figure 2) of staged insertion would be one that uniformly decreases the importance of configurations outside the \(N\)-important region, for example if the test-particle is turned on by increasing its energy parameter rather than its diameter. Other multistage methods rely on particle removal in some degree and are weekend by it. Staged deletion is hopelessly bad for hard spheres. It can be made to work for more realistic potentials, but then it is very inefficient and highly prone to systematic error, points that deserve broader recognition than is currently accorded. Bennett’s method and umbrella sampling both rely half on particle removal. Bennett’s method suffers more because it prescribes sampling of the \(N\)-particle system to characterize the \(W\) system. Thus it fails if the \(W\)-important configurations lie at all outside the \(N\)-important region. Bennett’s optimization minimizes its losses by taking the two regions as identical, thereby rendering the method equivalent to particle insertion. Umbrella sampling leaves open the question of how to weight the sampling of configuration space. A sufficiently gentle \(W\) can be selected so that the method is not defeated by its need to examine a large region that is trying not to sample. However, a careless choice for \(W\) leaves the method ineffective. A prime example is provided by application of the double-wide sampling technique [29] to compute solvation free energies. There \(W\) is typically defined via a simple change in particle size (and perhaps energy and/or charge). Unless very many stages are used, the outcome is systematically incorrect results, which always fall between the results of pure staged insertion and pure staged deletion. The different outcomes from these two paths (known as the hysteresis) have been ascribed equally to systematic errors in both and, by splitting the difference, double-wide sampling has been taken as the superior technique [34, 35]. We argue that this interpretation is completely incorrect, and that the staged-insertion data are the only data to be trusted. It can produce much more reliable results using far fewer stages.

We hasten to add that these considerations apply mainly to sampling problems of the type shown in figure 1(b). All multiple-staging approaches are viable (perhaps not equally so) for the situation depicted in figure 1(a).

2.2. Expanded ensembles

The second group of measurement methods is based on the idea of an expanded [12, 36](or augmented [37]) ensemble, an approach also suggested by Attard [38] and by Nezbeda and Kolafa [2], the latter pre-dating all other applications. Attard [38] and Nezbeda and Kolafa [2] considered the approach in the context of chemical potential calculations. A simulation in an expanded ensemble samples configurations (as usual) but in addition permits variation of a parameter that is not just a dummy variable of the partition function integral. It samples, for instance, the temperature or (of more relevance here) a parameter used to define the potential \(U_N\). For example (as in Attard’s study) the diameter \(d_t\) of a test sphere may fluctuate, as part of the Monte Carlo Markov process, between -1 (where its collision diameter is zero and it has no interactions with the other spheres) and 1 (or greater); rejected are any attempts to
increase the diameter that result in overlap with another sphere. Also, the random walk is biased by an applied weighting function $W(d_t)$ which ensures that the larger diameters are well sampled. A histogram is taken and the frequencies with which $d_t$ is -1 and 1, $h_0$ and $h_1$ respectively returns the chemical potential

$$\exp(-\mu_r) = \frac{h_1/W_1}{h_0/W_0},$$

(10)

where $W_0$ and $W_1$ are the corresponding values of the weight function. Note that in this approach configurations in which $d_t$ is neither -1 or 1 do not contribute explicitly to the calculation of the chemical potential. Nevertheless such values are important because they provide a path for $d_t$ to transition between the two limits, thereby ensuring that both are sampled well. As with the test-particle insertion (equation (3)) and umbrella sampling (equation (5)), in principle only a single simulation is needed to collect all required averages.

The notion of an expanded ensemble has also been applied in ways that are not focused on the chemical potential calculation. Liu and Berne [39] proposed it (independently) to aid sampling in molecular simulations. The concept provides a natural means for inserting and removing molecules from a system. Accordingly Kaminsky [37] and Escobedo and de Pablo [13, 40] have advocated its use in simulations requiring such moves, such as grand canonical and Gibbs ensemble calculations [41, 42] and a variant [43] of the Gibbs–Duham integration method [44–46]. Although we do not analyse the method in this context our results may be helpful to such applications.

2.3. Distribution–histogram methods

An important class of methods relies on the tabulation of distributions that characterize how the simulation samples configuration space. These approaches have their seeds in Bennett’s [30] seminal work and they were popularized through the efforts of Shing and Gubbins [14] who focused their application to the computation of the chemical potential. Extensions [33, 47, 48] and reformulations [49] have followed. Histogram reweighting methods [50–54] represent a significant and non-trivial generalization of the approach and are gaining in popularity [55]. The key quantities are commonly denoted $f(x)$ and $g(x)$, where $x$ is a (possibly multidimensional) projection or parametrization of the configuration space: $x = x(q)$. In previous work (applied to the chemical potential), $x$ has always been defined as the test-particle energy $u_t$, but it need not be restricted to this choice. However, it must be possible to determine uniquely the test-particle energy (or, more to the point, $e_t$) for any $x$. Then $f(x)$ is the probability density for $x$ in the system where the test particle does not interact with the other particles, and $g(x)$ is the corresponding distribution when the test particle does interact; thus

$$f(x) = \langle \delta(x - \hat{x}) \rangle_{N-1},$$

$$g(x) = \langle \delta(x - \hat{x}) \rangle_N,$$

(11)

where $\delta$ is the Dirac delta. Following Allen’s [56] presentation, $f$ and $g$ can be related as follows:

$$g(x) = \frac{Q_N}{Q_N} \langle \delta(x - \hat{x}) \rangle_{N-1} e_l(\hat{x})$$

$$= \frac{Q_N}{Q_N} \langle \delta(x - \hat{x}) \rangle_{N-1} e_l(x)$$

$$= \exp(\mu_r) f(x) e_l(x);$$

so

$$\mu_r = \ln \left[ g(x) + u_t(x) - \ln f(x) \right].$$

(12)

With the distributions $f$ and $g$ estimated from a simulation, one may apply this formula in one of two ways. Almost all applications have attempted to fit the right-hand side of equation (12) to a straight line of zero slope, yielding $\mu_r$ as the intercept (simple variants of this procedure are used also). In practice the distributions $f$ and $g$ often do not overlap sufficiently well for them both to be evaluated reliably for any $x$, and the approach fails. In this regard the method provides an advantage over simple insertion averaging only to the extent that the $g$-distribution samples outside the $N$-important region, which is to say it provides little advantage. Perhaps another benefit accrues if a weighted regression is applied; so the best data have the most influence in determining $\mu_r$. The second way to apply equation (12) is discussed but rarely practised. One may plot $\ln \left[ g(x) + u_t(x) \right]$ and $\ln f(x)$ as separate curves on the same graph and regress the value $\mu_r$ by which one must be shifted so that they extrapolate smoothly into one another. This procedure may be the most powerful means of implementing this formalism, as it can provide a result absent any overlap between the distributions. A good outcome can be ensured if an understanding of the expected functional forms of $f(x)$ and $g(x)$ is available. Then the process of bridging any gap between the distributions is (as pointed out by Bennett [30] akin to interpolation and can be accomplished very reliably.

The $f$-and-$g$ sampling method is always explained in the context of particle insertion and deletion: $g$ is to $f$ as equation (4) is to equation (3). This viewpoint is appropriate if the data are analysed via the first method above, because in this instance the $g$ sampling has the task of exploring the boundaries of the $N$-important region.
applied using the second analysis method, a more appropriate concept invokes multistage insertion: \( g \) is to \( f \)
as the first ensemble average is to the second in equation (7). The \( g \) sampling is most effective in its reconnaissance of
the \( N \)-important region, and this is all that it is asked to do by the second method.

Multistage versions of the distribution method may be formulated [33]. For example, if we define \( h(x) \) as the probability distribution of \( x \) for a biased ensemble according to
\[
h(x) = \langle \delta(x - \hat{x}) \rangle_w,
\]
then following the steps outlined before equation (12) we can show that
\[
A_N - A_W = \ln \left[ g(x) \right] + u_t(x) - \left\{ \ln \left[ h(x) \right] - \ln \left[ W(x) \right] \right\},
\]
\[
A_W - A_{N-1} = \ln \left[ h(x) \right] - \ln \left[ W(x) \right] - \ln \left[ f(x) \right].
\]
The two free-energy differences on the right-hand sides of these equations can be evaluated using the analysis outlined above, and their sum gives \( \mu_r \). Alternatively, one of the differences may be evaluated using the distribution method, and the remainder with the perturbation methods of section 2.1. If the first free-energy difference of equations (14) is obtained by another means, \( x \) need determine only \( W \) uniquely; if instead the second difference is obtained independently, then it is sufficient that \( x \) uniquely determines only the difference \( u_t - (- \ln W) \).

To examine the distribution method in application to hard spheres, we choose \( x = r_{NN} \), with \( r_{NN} \) the distance of the test particle to its nearest neighbour. If \( r_{NN} < 1 \), \( e_t = 0 \); otherwise it is unity. This choice cannot be applied to other potentials because in general \( r_{NN} \) is not sufficient to determine \( e_t \). Nevertheless variations of this approach can be applied generally, a prospect that we discuss in section 4.

2.4. Thermodynamic integration

The final class of methods encompasses thermodynamic integration. This is perhaps the most reliable of all approaches, and it is usually used for particularly difficult chemical potential measurements (such as for solids). It requires that a series of simulations be performed along a path that joins the system of interest to a reference system of known chemical potential. The principal drawback of this approach is its perceived inefficiency (simulations at many ‘uninteresting’ states are required) and one must also take care that the integration path does not traverse a phase transition. However, a properly formulated and executed integration need not be inherently less efficient at delivering the chemical potential than the methods just described. Integration along a path that gradually couples a test particle to the others is very unlikely to incite a phase change and

in fact samples configurations very similar to those encountered in the methods described above. Nevertheless in many cases the simulations at the uninteresting integration states are not easily applied to the measurement of other thermodynamic quantities (e.g. the pressure) at the state of interest, and this poses a significant disadvantage when compared with test-particle insertion (for example), but not when compared with methods that suffer the same deficiency (such as expanded ensembles). Another concern regards the behaviour of the integrand in the limit of vanishing test particle, as a carelessly chosen integration path can introduce an integrable singularity in this limit. The problem is easily avoided [57–59].

Integration can be cast as an approximation to an exact multistage free-energy perturbation (section 2.1.2) that follows exactly the same path as the integral; yet integration differs in several important ways. It makes assumptions about the smoothness of the integrand, which may be interpreted as statements regarding the smallness of higher-order central moments of the energy distribution compared with the integration step size taken to the corresponding power. Integration is accurate to the extent that these assumptions are valid. More significantly, the linearization inherent in thermodynamic integration diminishes the impact of collective fluctuations on the free-energy measurement process. As a result the integrand can be evaluated much more reliably in a simulation than can the exponential of equation (3) (for example). On the other hand, the integrand involves averaging a derivative of the energy, which often has its greatest contributions where the energy is large (and thus less well sampled). Perhaps the most important consequence of the linearization is a symmetrization of the free-energy calculation; the precision of the result is independent of the direction of the integration path (in sharp contrast with the methods of section 2.1). This symmetry is an important contributor to the robustness of the method, as it permits one to apply the technique without thinking about the considerations depicted in figure 1. Unfortunately many of these features are not captured well by integrations involving the hard-sphere model on which this study focuses. With this limitation in mind we nevertheless include thermodynamic integration in our study.

2.5. Other methods

There exist other techniques that we do not consider here. Consistent with our focus on hard spheres, we do not concern ourselves greatly with methods specific to mixtures or to chain molecules [8–13, 60] except in the sense that these methods are largely (non-trivial) extensions of the methods we do examine. We do not examine thermodynamic integration or free-energy perturbation.
Thesecondterminthesquarebracketsiszeroifthetestparticleoverlapsanothersphere,buttahmainviablemethod(e.g.forsolids)orareotherwisemostappropriate(e.g.formappingarangeofstates).Alsowedonotexaminemethodsthatrelyontheimpositionofinhomogeneitiesinthesimulatedsystemmethods[63]thatputthesystemofinterestinphysicalcontact(andpresumablythermodynamic equilibrium)withanotherthatismeasleystudied.Withtheseschemes,oneisalways leftwonderinghowmucherrorisintroducedbyslow equilibrium,thefacesandotherfinite-sizeeffects thatarelikelytobecrelarge.Finally,weexclude fromconsiderationmethodsthataregrandcanonical innaturemethods[5,16]methods‘measure’thechemical potentialbyimposingitandaveragingthedensitybysimulation.Manynofthemes techniquesthatwexaminecanbeformulatedinthisfashion,andoingsomaybe advantageous.Examplesincludestudiesofadsorption,orphasiequilibriaingeneralviatheGibbesensemble[41,42].

3. The problem of choosing a weighting function


\[ W(q^N) = \prod_{i \in nbr} \exp(-w_i) \prod_{i \notin nbr} \exp(-w_{i(non-nbr)}). \]  

Typicallythenon-neighbourinteraction\( w_{i(non-nbr)} \) is takenaszeronointeraction)orthatoffullhard sphere. Technicallywisastruepairpotential,asfortheparametersthenearliestherspherechansifadiffersphereisbroughtcloser;thisfeaturepresentsnoproblemsinprincipleorinpractice.

Itisworthpointingoutthedistinctionbetweena‘hard’and‘soft’formofW,analogoustohard andsoftintermolecularpotentials.AhardWtakes onlythecases0or1,correspondingtosomemodiﬁed overlapcriterionforthetestparticle.Examplesincludea Wthathasatestparticleinteractingsasa sphereofreduceddiameter,orreathathasatest particlebehavesasahearthedicroathallbutitsnearestneighbour,withwhichithasnointeractionatall. Algorithmassthatemployahardinteractionwiththenearlestsphereatall.Algorithmsthatareestablished,thistheseachaswiththefunctionalformoftheoverlappingvariableW\( \times vw_1 \times vq^N \times \exp(2\mu_v) \times e_1(q^N) \]^\{1/2\}. (15)

The second term in the square brackets is zero if the test particle overlaps another sphere, but it may be very large when no overlap occurs (upon taking the square root, \( 10^7 \) at the freezing density). Thus an algorithm based on this weighting function must wait for the very infrequent even in which the test particle overlaps no other and, once such a configuration is found, the simulation remains in this state for an equally long period. The fact that transitions into the overlap state are as rare as the opposite renders this method inferior even to test-particle insertion. For sufficiently long simulations the method will succeed in giving equal time to configurations important to both the (\( N-1 \)) - and \( N \)-particle systems, which shows only that compliance with this intuitively appealing guideline is no guarantee of success.

The problem of selecting \( W \) has two facets, one qualitative, the other quantitative. The first concerns the functional form of the dependence of \( W \) on the configuration-space variables \( q^N \). It seems that this choice must be based on the investigator’s intuition. Clearly, taking \( W \) in terms of the test-particle energy is a poor choice for hard spheres (and thus, we would argue, poor in general). Many workers limit their choice to forms that are based on the intermolecular potential of the system being studied, defining \( W \) by altering one or more of the potential parameters. Although reasonable, this is an unnecessarily restrictive class. We shall consider choices that work for hard spheres while remaining appropriate for more general forms of the intermolecular potential. Our approach takes \( W \) as a function of the set of scalar distances from some of or all the other spheres: \( W = W(r^N) \); this can be done by defining a spherically symmetric pair potential \( w(r) \) that governs the interaction of the test particle with the others. The approach can be varied via the form of \( w(r) \); also \( w(r) \) can be selected differently for the nearest-neighbour (or first two nearest-neighbour, etc.) sphere than it is for the others. Thus

\[ W(q^N) = \prod_{i \in nbr} \exp(-w_i) \prod_{i \notin nbr} \exp(-w_{i(non-nbr)}). \]  

The quantitative aspect of selecting \( W \) addresses how to distribute the weighting. Once the qualitative features of \( W \) are established, this question can be addressed by a systematic analysis. It is inevitable that any prescri-
tion for ‘the best’ weight function can be made only with the assumption that much of what is to be computed from the simulation is known a priori. A clear example is seen in equation (15) where the weight function is given explicitly in terms of $\mu_r$, whose calculation is of course the goal of the simulation. This circularity cannot be avoided, and it is a significant hindrance to non-Boltzmann methods. It is important to keep in mind that the output from the simulation is in principle independent of $W$; so imperfect knowledge of these ‘circular quantities’ does not diminish the accuracy (as distinct from precision) of the result; the problem is analogous to needing an initial guess before implementing a root-finding algorithm. In practice several things may be done: one may perform short simulations on small systems to estimate the unknowns; one may refine the weight function as the simulation proceeds until satisfactory convergence is observed [66, 67] (one must be very careful when practising this, as it can lead to violation of the fundamental assumptions of the Markov process underlying the Monte Carlo simulation [68]; one may have theoretical or semiempirical knowledge that can be brought to bear, for example approximate relations describing homologous series (of alkanes, say). The question that has never been addressed, and which we attempt to uncover here, is: how should one devise a weight function if all such quantities can be assumed known?

4. Modelling and analysis

To optimize and compare the methods presented in section 2 it is very helpful to work with a model of the calculation process. Ultimately the optimization and comparison are based on some suitably defined variance. Measuring the variance by simulation (rather than a model thereof) requires careful attention to correlations in the data, and consequently many (or long) independent simulations. Even then the outcome can be fooled by statistical anomalies or non-ergodicities of the type that we are attempting to defeat. In contrast a good model of the simulation gives a very accurate description of the variance in the averages, and it does so in a tiny fraction of the time needed by even a single simulation. However, to formulate a model, one needs a large amount of information about the system of interest, and this is one reason that we choose the hard-sphere fluid for our study.

4.1. Test-particle insertion

The simplest case to model is test-particle insertion. We assume that the outcome of each attempt to insert a test sphere is statistically independent of those from other attempts. One might approximate this in practice by performing one insertion trial at the end of each simulation cycle (defined as $N$ elementary Monte Carlo trials); more or fewer intervening cycles may be required, depending on the density, and we shall not concern ourselves with this (important) detail. If a total of $M$ insertion trials is attempted in a simulation, the probability $p_M$ that exactly $m$ succeed (do not encounter overlap) is described by the binomial distribution

$$p_M(m; p_1) = \binom{M}{m} p_1^m (1 - p_1)^{M-m},$$

where $p_1$ is the probability that any single insertion attempt is successful. The mean of this distribution is $M p_1$, and its variance is $M p_1 (1 - p_1)$. According to equation (3) the residual chemical potential from a simulation would be $-\ln(M/M)$. The nonlinear dependence on $M$ means that we cannot evaluate the variance in $\mu$ knowing only the variance in $m$, unless perhaps $M$ is large. The non-zero probability that $m = 0$ particularly troubles the statistics. As a remedy we shall work directly with $\exp(-\mu)$ and take our variance thus (cf. equation (3), suppressing the ‘$N$ - 1’ subscript for clarity of notation):

$$\sigma^2 = \frac{\langle e_1 \rangle - \langle e_1 \rangle^2}{\langle e_1 \rangle^2} = \frac{1 - p_1}{M p_1},$$

where we use an overbar to indicate an average over many independent simulations (or, equivalently, many independent blocks within a single simulation). The latter equality follows from $\langle e_1 \rangle = m/M$ and the statistics quoted above for $p_M$. Our measure is interpreted as a fractional square error in $\exp(-\mu)$ although it is equivalent to what standard propagation-of-error formulae would give for the absolute square error in $\mu$ via equation (3).

The modelling is completed by specifying $p_1$. The Carnahan–Starling [69] equation of state for hard-sphere fluids is accurate up to the freezing transition. The residual chemical potential derived from it provides $p_1$ as a function of density $\rho$:

$$-\ln[p_1(\rho)] = \mu_r(\rho) = \frac{\eta(8 - 9\eta + 3\eta^2)}{(1 - \eta)^3},$$

where $\eta = (\pi/6)\rho$.

The variance according to the model is presented as a function of density in figure 3. We examine densities from $\rho = 0.05$ (at which the insertion probability $p_1$ equals 0.8 according to equation (19)) to $\rho = 0.95$ (for which $p_1 = 5 \times 10^{-8}$), which is slightly above the freezing density of 0.94. The appropriate variance measure is the product $M \sigma^2$, which according to equation (18) is independent of the number $M$ of measurements. For sufficiently large $M$ this independence applies to all methods; so we take the group $M \sigma^2$ as our measure of quality, smaller being better. Note that, if we take a ‘reasonable length’ for a simulation to be 10 000
cycles, and the chemical potential to be ‘sufficiently well’ determined if it is known to within ±0.1 (0.6% of the value at freezing), then the value 100 forms the important threshold in $M\sigma^2$. A method exhibiting a scaled variance below this value is viable at that density.

We verified our model of test-particle insertion by conducting a series of Monte Carlo simulations. Fifty simulations each attempting 10 000 insertions were conducted for densities of 0.5, 0.7 and 0.9 respectively. One insertion was tried at the end of each simulation cycle. Each simulation was begun from its own previously equilibrated initial configuration; so they may all be taken as independent. For all simulations $N = 108$ spheres were used. This number is sufficiently small to raise concern for finite-size effects. We applied the adjustment of Siepmann et al. [70] which adds a density-dependent term proportional to $1/N$ to the right-hand side of equation (19). The simulation data and the model are compared in table 2, and they agree very well.

### 4.2. Expanded ensemble

#### 4.2.1. Basic model

We now turn to modelling chemical potential calculations in which a weighting function is applied to influence the sampling process. The easiest case to model is the expanded ensemble described in section 2.2, in which the diameter $d_t$ of a test sphere is allowed to take on a range of values from -1 to 1 (collision diameter from zero to unity). More specifically, $d_t$ fluctuates among a predetermined set of $n$ discrete values $\{d_{t0}, d_{t1}, \ldots, d_{tn}\}$. Our modelling approach entails setting up an approximate single-step transition-probability matrix (of size $n \times n$) that describes how the test-particle samples from these diameters. The associated Markov process can be analysed rigorously to extract statistics regarding the frequency with which each diameter is sampled. Propagation of error then yields a measure for the variance in the chemical potential.

The effective transition-probability matrix will be designated $\Pi$, and its elements $\pi_{ij}$. Each transition probability is given as the product of three terms:

$$\pi_{ij} = T_{ij} \times \min\left(1, \frac{\Omega_j}{\Omega_i}\right) \times \min(1, \chi_{ij}). \quad (20)$$

The first term, $T_{ij}$, is the probability that the trial $i \rightarrow j$ will be attempted. We allow the state to move in a single step as far as $k_{\text{max}}$ neighbours on either side of $i$, and the neighbour $j$ is selected with uniform probability from among those eligible; thus

$$T_{ij} = \begin{cases} \min(k_{\text{max}}, |i - 1|), & |i - j| \leq k_{\text{max}}, \\ \min(k_{\text{max}}, n - i), & |i - j| > k_{\text{max}} \text{ or } i = j. \end{cases} \quad (21)$$

The second term in equation (20) characterizes the probability that the diameter change will not result in overlap. The quantity $\Omega_j/\Omega_i$ is a ratio of partition functions for the corresponding systems. The non-overlap
probability is unity if the diameter is decreased \((j < i)\); otherwise it is approximated by the ratio. This ratio is related to the chemical potentials of spheres of diameters \(d_{ij}\) and \(d_{ij'}\) in an otherwise pure fluid of hard spheres of unit diameter:

\[
\frac{\Omega_j}{\Omega_i} = \exp\left[\left(\mu_i(d_{ij}) - \mu_i(d_{ij'})\right)\right]. \tag{22}
\]

The needed chemical potentials may be obtained from the Mansoori–Carnahan–Starling–Leland (MCSL) \([71]\) equation of state for hard-sphere mixtures; the appropriate formula is reproduced in appendix 3. The final term in equation \((20)\) is the acceptance probability for the transition. It is constructed to ensure microscopic reversibility in the molecular simulation, which requires that

\[
\chi_{ij} = \frac{T_{ij}W_i}{T_{ij}W_j}, \tag{23}
\]

where \(W_i\) is the imposed weighting associated with test-particle diameter \(d_{ij}\).

An elementary step in this Markov model of the molecular simulation corresponds to one attempted diameter-change trial. The model assumes that the outcome of each such trial is independent of those attempted immediately before and after it. This means that the simulation should attempt a sufficient number of its elementary steps (particle displacements) between each attempted diameter-change trial. Moves that attempt to swap the positions of the test particle and a real particle, or to reinsert the test particle at a random position, are very helpful in establishing a new environment for the test particle \([37, 56]\).

### 4.2.2. Covariance matrix from transition probabilities

The transition-probability matrix derived in section 4.2.1 is small enough to permit application of rigorous formulæ for its occupancy statistics. In general these statistics will depend on the initial occupancy state (test-particle diameter) of the process, and they depend in a complicated way on the number of diameter-change trials \(M\). However, in the limit \(M \to \infty\) the dependence on the initial state vanishes, and the dependence on \(M\) becomes simple. We shall work exclusively with this limiting case.

The limiting distribution matrix \(\Pi\) is defined as \(\lim_{M \to \infty} \Pi^M\). It is simply a matrix of identical rows set to the limiting distribution. In the present application \(\phi\) is known \textit{a priori} from the construction of the transition probabilities: \(\phi_{ij} = \pi_j = W_i/\Omega_j\), where we designate \(\pi_j\) as the limiting probability that state \(j\) will be occupied.

The second moments of the occupation frequency also can be computed very efficiently in the limit of large \(M\), and from them the variances (including covariances) are obtained \([72]\):

\[
M\sigma_{ij}^2 = \pi_i \pi_j + \pi_i s_{ij} + \pi_j s_{ji} - \delta_{ij}\pi_i,
\tag{24}
\]

where \(\delta_{ij}\) is the Kronecker delta and \(s_{ij}\) are the elements of the matrix

\[
S = (I - \Pi + \phi)^{-1} - \phi,
\tag{25}
\]

where \(I\) is the identity matrix and the \(-1\) superscript indicates a matrix inverse. Note that the right-hand side of equation \((24)\) does not depend on \(M\).

### 4.2.3. Error measure; verification of model

The right-hand side of equation \((10)\) for the chemical potential is the ratio of two simulation averages (i.e. the histogram values). As discussed in section 4.1, this nonlinear dependence on the averages complicates the analysis of the variance; particularly troublesome is the event that the simulation returns zero for the average in the denominator. A well constructed simulation algorithm run for a sufficiently long time will not encounter this problem, but a poor or short algorithm may, and this possibility interferes with the comparison and optimization. This problem is encountered as well in the umbrella-sampling method; so we shall continue our discussion in general terms. For the chemical potential given in terms of the ratio of two simulation averages \(\langle n \rangle\) and \(\langle d \rangle\), that is

\[
\exp(-\mu_i) = \frac{\langle n \rangle}{\langle d \rangle},
\tag{26}
\]

a robust variance measure can be constructed thus:

\[
\sigma^2 = \frac{1}{\langle n^2 \rangle} \left( \langle n \rangle - \frac{\langle n \rangle}{\langle d \rangle} \right)^2.
\tag{27}
\]

(Equation \((18)\) results as a special case when \(\langle d \rangle = 1\).) We emphasize that the angular brackets indicate an average taken over a simulation of a finite length characterized by \(M\), and the overbar indicates an average from an infinite number of independent simulations. Finally, we write this variance measure in terms of the (co)variances of the numerator and denominator averages, which in the present application are merely the values of the first and last bin of the histogram:

\[
\sigma^2 = \frac{\sigma_{nn}^2}{\langle n \rangle^2} - 2 \frac{\sigma_{nd}^2}{\langle n \rangle \langle d \rangle} + \frac{\sigma_{dd}^2}{\langle d \rangle^2}.
\tag{28}
\]

We note again that this variance measure is comparable with the variance in the residual chemical potential itself if the simulations are sufficiently long \((M \to \infty)\).

The model was verified by several series of independent Monte Carlo simulations that were conducted as described in section 4.1 except that, instead of per-
forming a trial insertion at the end of each simulation cycle, a diameter-change trial was performed. Representative results are included in Table 2, where again excellent agreement is found between the model and the observed variances. Examination of test-particle diameter distributions (not shown) finds that they too are in agreement with the model.

4.2.4. Optimization

The weighting function $W_i$ provides a degree of freedom with which we can improve the performance of the expanded-ensemble methodology. Moreover, we can adjust the discrete values of the diameter that the test-particle samples, as well as the trial-step parameter $k_{\text{max}}$ defined via equation (21). We fixed $k_{\text{max}}$ as one does in an actual simulation, that is by adjusting it to achieve a fixed rate of acceptance of the trial diameter changes. This acceptance rate can be computed from the model as $1 - \sum \pi_i \pi_{ij}$. We found that the variances predicted by the model were insensitive to the value of the acceptance-rate target; we chose 35%.

We consider first the effect of the number of bins $n$ on the variance, and results are displayed in Figure 4 for a system of density $\rho = 0.95$. Four cases are examined, considering whether the weighting function is selected to yield a uniform distribution of diameters or is optimized to minimize the variance, and whether the bins represent a uniform distribution of diameters or a distribution that makes uniform the difference in chemical potentials between adjacent diameters (i.e. $\mu(\sigma_i) - \mu(\sigma_{i-1}) = \text{constant}$). The uniform distribution is important because it is the ‘rule of thumb’ that one usually aims for when adjusting weights in an expanded-ensemble or umbrella-sampling application. For all cases significant reduction in the variance is observed when bins are added, up to about $n = 10$ where a shallow minimum is observed. Choosing diameters uniform in $\Delta \mu$ improves the performance, as of course does optimizing the weights (completed numerically). However, in neither case is the improvement dramatic, a factor of two at most. To provide perspective, we note that the use of no weighting at all ($W_i = 1$) moves the variance five orders of magnitude above the curves displayed in the figure.

Representative distributions that correspond to optimal weightings are displayed in Figure 5. The optimum is not uniform, particularly for smaller $n$. Interestingly, the non-uniformity is such that the tails of the distribution are attenuated. It is certainly not unreasonable to have expected exactly the opposite, as these bins are the two used directly in the formula for the chemical potential, equation (10)). It is as if the act of getting to the tails and out is more important than being in them. The effect is particularly dramatic for the largest diameter, where the histogram value is too small even to register on the figure. This feature makes sense when one realizes that absolutely no information can be gained by having the test sphere at its full size. When there it can only become smaller, which of course never registers overlap. This observation suggests that an improved algorithm could be formulated by never permitting the test sphere to take on the largest diameter. Instead, it samples among all the other available diameters and measures the difference between the penultimate diameter and the full-size diameter via a test-growth method. Modelling this approach is a straightforward extension of the methods already described, and the result is included in Figure 4. As expected, the variance is decreased by the modification. The optimal distribution still exhibits a slight attenuation of sampling of the penultimate diameter (Figure 5). In all cases, switching to diameters spaced uniformly in chemical potential differences causes the optimal distribution to become more uniform, although the occupancy fraction for the largest diameter is still very small.

The density dependence is included in Figure 3, with $n$ selected at its optimal value of 10. The method becomes
superior to test insertion only when applied at sufficiently high density (greater than 0.6), which is the region of greatest interest. However, the method is ‘viable’ (as defined in section 4.1) only over those densities for which test-particle insertion also applies (and also works better).

We finish by noting that the \( n = 2 \) case with uniform sampling is decidedly worse than simple test-particle insertion; at \( M \sigma^2 = 7 \times 10^7 \) it is off the scale of figure 4 and about four times larger than the Widom insertion; the situation holds at all densities. This is a clear illustration of the point made in connection with equation (15), namely that when optimizing the weights it is imperative to consider the effect that they have on the sampling process. The bicanonical ensemble of Swope and Andersen [68] is also an instance of the \( n = 2 \) case of an expanded ensemble, although their application does not suffer as much as we describe here because they employ a biased insertion–deletion algorithm very much like the excluded volume map sampling of Deitrick et al. [1].

4.3. Staged insertion

4.3.1. Diameter staging

The staged-insertion technique was introduced in section 2.1.2 and employs equation (7) as its working equation. If we restrict our attention to ‘hard’ weighting functions \( W \) (as discussed in section 3) we can analyse and optimize the method without resorting to the transition-matrix methods introduced in section 4.2. The analysis proceeds easily for an arbitrary number of
intermediate stages (equation (8)); so we shall work with this general case.

To choose a specific instance we define the intermediate-stage weighting functions in terms of a set of symbols $p_i, d_{i1}, d_{i2}, \ldots, d_{in}$ such that $W_j$ is zero if the test particle of diameter $d_{i1}$ overlaps another sphere; otherwise it is unity. According to equation (8), the first simulation is conducted for a system of $N - 1$ spheres, and the fraction $p_1$ that a sphere of diameter $d_{i1}$ encounters no overlap upon test insertion is recorded. Then in the second simulation the test particle of diameter $d_{i1}$ interacts with the other spheres, and one records the fraction $p_2$ that a trial increase to diameter $d_{i2}$ encounters no overlap. The third, fourth, etc., simulation stages, if used, proceed in a like manner. The chemical potential exp($-\mu_r$) is the product of all these measured probabilities. Its variance is the sum of the variances of the $n$ independent measurements, each of which is given by a formula analogous to equation (18); thus

$$M\sigma^2 = M \sum_{i=1}^{n} \frac{1 - p_i}{m_i p_i},$$

(29)

where $m_i$ is the number of measurements performed during stage $i$. To provide a fair comparison with other approaches we require $\sum m_i = M$, which renders the right-hand side of equation (29) independent of $M$ (for fixed $m_i / M$).

The performance of the multistage insertion method may be optimized with respect to the number $m_i$ of measurements devoted to each stage, the diameters $d_{i1}$ and the number $n$ of stages. Considering first the optimization with respect to $m_i$, application of Lagrange's method finds that

$$m_i^{opt} = M \left( \frac{1 - p_i}{p_i} \right)^{1/2} \left/ \sum_{j} \left( \frac{1 - p_j}{p_j} \right)^{1/2} \right.,$$

(30)

for which

$$M\sigma_{opt}^2 = \left[ \sum \left( \frac{1 - p_i}{p_i} \right)^{1/2} \right]^2.$$

(31)

Optimization with respect to the diameters is best conducted by working directly with the ‘insertion probabilities’ $p_i$. The minimization of the variance of equation (31) is subject to the constraint $\prod p_i = \exp(-\mu_r)$. Lagrange's method applied to the logarithms finds that the optimum value of $p_i$ is independent of $i$; so we conclude that the optimal $p_i$ are all equal to one another, and

$$p_i^{opt} = \exp \left( -\frac{\mu_r}{n} \right).$$

(32)

The variance takes a simple form in this case

$$M\sigma_{opt}^2(p_i) = \left( \frac{1}{\exp(-\mu_r/n)} - 1 \right).$$

(33)

A slightly less general version of this result was obtained by Nezbeda and Kolafa [2]. In figure 6 the efficiency is plotted as a function of the number $n$ of stages for several hard-sphere densities (which sets the value of the single-stage test-particle insertion probability $p_1 = \exp(-\mu_r)$ via equation (19)). For high densities, it is advantageous to use more stages, but there is a limit beyond which additional stages degrade performance. For hard spheres at freezing, ten stages are optimal. For insertion probabilities of 10% or greater (hard-sphere densities below about 0.35), a single stage is best. In all instances the greatest relative improvement comes with the introduction of the first intermediate stage. Subsequent stages provide increasingly diminished returns until the optimum is reached. The density dependence of the variance (upon full optimization of the method) is presented in figure 7 (and in figure 3). A curve also is presented in figure 7 to demonstrate the penalty for not optimizing the number of stages (i.e. always using exactly one intermediate, $n = 2$).

4.3.2. Neighbour staging: insert and expel

We examine next an approach based on a single intermediate stage in which the test particle has full hard-sphere interactions with all other particles except for the particle nearest to it, and with this nearest neighbour it has no interaction at all. This is the first step on a path

Figure 6. Scaled variance as a function of number of stages for the staged-insertion technique. The lines are plots of equation (33), where the insertion probabilities are obtained via the Carnahan–Starling equation (equations (19)) for the densities indicated. Note that $n = 1$ is equivalent to simple test-particle insertion.
that in the next section will lead us to some conclusions regarding the viability of the umbrella-sampling method characterized by equation (5). Thus, in the first stage, one measures the fraction \( p_1 \) of test-particle insertions in which the randomly inserted sphere overlaps with at most one other sphere. In the second simulation the test particle moves about with at most one overlapped, and the fraction \( p_2 \) of the time that it has zero overlaps (and thus has expelled the nearest neighbour) is recorded. Modelling this process requires knowledge of the hard-sphere cavity–cavity correlation function \( y^{(2)}(r) \). Accurate formulae are available, and we have employed the formula proposed by Boulik [73] which is reproduced in appendix 3. In terms of this function, \( p_1 \) is

\[
 p_1 = \left( 1 + \int_0^1 \rho y^{(2)}(r) 4\pi r^2 \, dr \right) \exp(-\mu_r). \tag{34}
\]

The unit term in the large parentheses corresponds to insertion with no overlap, and the integral term is insertion with exactly one overlap. The chemical potential \( \mu_r \) is obtained by equation (19). The easiest route to \( p_2 \) is via equation (7), which in the present case gives \( \exp(-\mu_r) = p_1 p_2 \). Equation (30) for the variance in the chemical potential still applies, with \( p_1 \) and \( p_2 \) properly reinterpreted.

The only freedom available to minimize the variance is with the numbers \( m_1 \) and \( m_2 \) of measurements taken in each simulation stage. Choosing these in the optimal manner prescribed by equation (30), the variance associated with this method is presented in figure 7. It is instructive to examine the two components of the variance, namely that associated with inserting the test particle, and that arising when the nearest neighbour is expelled from the overlap volume of the test particle. Figure 7 shows that they are out of balance at all densities, that is a much greater contribution is associated with the nearest-neighbour expulsion (it is interesting that the optimization analysis shows that this cannot be remedied by shifting more simulation cycles to the expulsion measurement). The imbalance is a manifestation of the fact that \( p_1 \) and \( p_2 \) are not equal (\( p_1 \) is about two orders of magnitude greater than \( p_2 \) at the 0.9 density), as they would be in the optimum case. One might consider variants in which the test particle can overlap as many as two spheres (or three, or four, etc.) rather than one. Estimation of the probability of insertion with exactly two overlaps requires knowledge of the three-body cavity correlation function. However, it is clear that such a modification would only exacerbate the present problem, as it would put the insertion and expulsion probabilities even further out of balance.

4.4. Umbrella sampling

We turn now to the method known as umbrella sampling, which has equation (5) as its working equation. We consider first the application of the method in conjunction with the ‘insert-and-expel’ staged-insertion technique just discussed. There it was found that the expulsion component was the most problematic to the chemical potential calculation; so we shall consider application of umbrella sampling to improve just this part. In particular, we introduce an additional stage, intermediate between the \( W \) and \( N \) systems of section 4.3.2. In this stage the test particle interacts with its nearest neighbour according to a pseudopotential \( w(r) \), rather than not at all. The discussion of section 2.1 shows that this function cannot be given by any hard potential; so instead we shall attempt to formulate it to optimize performance; interaction of the test particle with all other particles remains the full hard-sphere potential (this is a feature of both the \( W \) and the \( N \) systems; so the hard interaction presents no problem). For simplicity we shall consider forms that have \( w(r) = 0 \) for \( r > 1 \). Thus the chemical potential is computed as

\[
 \exp(-\mu_r) = \left( W_1 \right)_{N=1} \frac{\langle e_1/W_2 \rangle_{W_2}}{\langle 1/W_2 \rangle_{W_2}}, \tag{35}
\]

where \( W_1 \) is the weighting function of section 4.3.2, and \( \ln W_2 = - \sum_{i=1}^{N-1} w(r_{ii}) \). In principle, two simulations are needed to obtain the right-hand side of equation
but we shall be generous to the method and assume that all necessary averages can be computed via a single simulation of length $M$ (this has the effect of shifting the I&E: Insert curve of figure 7 down by about one order of magnitude at the highest density).

The pseudopotential $w$ has a marked effect on the sampling for the average $\langle e_i/W \rangle_w$, and the full Markov-chain modelling of section 4.2 must be applied. To do so we discretize the nearest-neighbour separation into $n-1$ bins for $0 \leq r \leq 1$ (uniformly separated in $r$), and one bin for $r > 1$. The transition probabilities between the bins are a product of the three terms presented in equation (20). In the simulation being modelled, transitions from one separation to another arise from two types of elementary Monte Carlo moves: displacement of each sphere about a cube of fixed size centred on its present location, and random displacement of the test sphere to any point within the simulation volume. This is difficult to model; so we take $T_{ij}$ so that transitions occur only to neighbouring bins, and for them $T_{ij} = V_j/(V_{i-1} + V_i + V_{i+1})$, where $V_j$ is the volume of bin $j$; more sophisticated forms were examined and were not found to work significantly better. The overlap probabilities $\Omega_i$ in equation (20) were estimated (within a multiplicative constant) as $V_i^{-1} \int \rho^{(2)}(r)4\pi r^2 dr$, where the integral extends over separations appropriate to bin $i$; with this scaling it is appropriate to set $\Omega_n = 1$. This model then assumes that overlap does not occur if the nearest-neighbour sphere is moved towards the test sphere. This too is untrue for the simulation, but better modelling would require accounting for three-body correlations. Finally the acceptance probability was given by $\chi_{ij} = W_j/W_i = \exp[-(w_j - w_i)]$. The limiting distribution is $V \sum V_i W_i$.

The optimal sampling distributions for three densities are displayed in figure 8. The bars represent the probability that the nearest neighbour is found within the bin centred on the abscissa value (remembering that the rightmost bin represents all conditions in which the test particle has no overlap). The values in the figure have been divided by the volume of the shell represented by the bin so that they have the character of a radial distribution function (except for the rightmost bin, which is simply the fraction of the time that the test particle has no overlaps). Also presented on the figures are the optimal weighting functions. At the higher densities the optimal distributions are slightly peaked, with the peak moving towards the origin as the density is increased. Very little sampling of the non-overlap configuration is prescribed. At lower densities the optimal distribution (and the optimal weight) is nearly uniform, with about 8% of configurations having no overlap. The outcome is very similar to that seen in the expanded-ensemble study of section 4.2. For the highest-density system we have verified by Monte Carlo simulation the distribution predicted by the model. This is included in the figure.

The chemical potential variance associated with the procedure is displayed as a function of density in figure 3. In table 2 the variance from a set of 50 independent 10 000 cycle Monte Carlo simulations is compared with the model variance at $\rho = 0.90$. The agreement is less satisfactory than seen for the other methods. The full variance is obtained by adding the variance associated with the first stage of the calculation, inserting a test particle and observing the fraction that find zero or one overlap; in our present analysis this contributes little to the total variance. Application of the umbrella-sampling method has reduced by two orders of
magnitude the variance of the second stage of the calculation. Nevertheless, the variance is consistently greater than that observed for the much simpler method of optimized multistage insertion of section 4.3.1. More important, the situation is not going to be improved by working with a more obvious implementation of umbrella sampling, namely one in which the test particle interacts with all neighbours according to the pseudopotential \( \psi(r) \). This step would eliminate the need for the first stage, while compounding the difficulty of implementing the second. Umbrella sampling does perform reasonably well overall but, given the difficulty and importance of formulating an effective weighting function, we do not see it as competitive with the other methods.

4.5. Thermodynamic integration

We shall consider an integration path analogous to the expanded ensemble discussed earlier, that is we integrate the change in free energy as a test sphere diameter (not the collision diameter) goes from zero to unity. The hard-sphere integrand differs qualitatively from that used to integrate a soft potential along an analogous pathway, but it is hoped that our analysis can provide valid quantitative insight regarding the efficiency of thermodynamic integration relative to the other methods. With this approach the chemical potential is given by

\[
\mu_r = \mu_r(d_t = 0) + \int_0^1 \Gamma(\hat{d}) \, d\hat{d}.
\]

We assume that the residual potential for the ‘point’ sphere of zero diameter is evaluated by another means, presumably test-particle insertion. For hard spheres this quantity is given exactly in terms of the packing fraction \( \eta \):

\[
\mu_r(d_t = 0) = -\ln (1 - \eta),
\]

but we shall conduct the analysis to mimic the more general case in which the term must be measured in a simulation. Regardless, the contribution to \( M \sigma^2 \) is negligible, adding about 1-0 at the freezing density. The integrand of equation (36) is \( \Gamma = \partial \mu / \partial d \). For the hard-sphere potential this derivative is not given by an analytic form, and instead it must be measured by a limiting process. This measurement is one of the few that we must model in a way that is very particular to the hard-sphere model.

To evaluate \( \Gamma \), one tabulates a histogram of probabilities (frequencies) \( p_i \) that the test particle encounters no overlap when its diameter is test increased to \( d_i + \Delta \varepsilon \) (thus \( p_i = p(i \varepsilon; d_t) \)). The difference between the chemical potentials of spheres of diameter \( d_i + \Delta \varepsilon \) and \( d_i \) is given by \( \Delta \mu_i = -\ln p_i \); thus \( \Gamma(d_t) \) may be obtained by the limiting \((i \rightarrow 0)\) slope of a regression of \( \Delta \mu_i \) against \( i \varepsilon \). Much care is required when propagating the uncertain-

ties in the \( p_i \) to yield a variance for \( \Gamma \), as fluctuations in these probabilities are highly correlated within a simulation. We leave most of the details to appendix 1. We note, however, that our regression employed a weighted fit to a cubic polynomial in \( i \varepsilon \) with zero intercept, which is exactly the dependence prescribed by the MCSL equation of state [71] (see appendix 3). We feel that our analysis is somewhat generous to the methodology, and that we are likely to underestimate the variance associated with the method (as applied to hard spheres).

We apply a Simpson rule quadrature and obtain a variance in \( \mu_r \) via simple propagation of error of the uncorrelated variances in the \( \Gamma \) values. We examined the effect of variation in the number of quadrature points (i.e. the number of simulations used to join the \((N - 1)\)- and \(N\)-particle systems), keeping the total number \( M \) of simulation cycles constant and divided equally among the simulations (a simulation cycle is presumed to provide one independent measurement of the \( p_i \) histogram). We find that with this constraint the overall variance in \( \mu_r \) is virtually independent of the number of quadrature points or simulations performed. The loss in precision in the measurement of \( \Gamma \) that results when each simulation is shortened to accommodate additional simulations is compensated exactly by the diminished contribution that each \( \Gamma \) has to the Simpson rule sum. This means that a ‘quasistatic’ integration or ‘slow-growth’ method [7] in which \( d_t \) is very slowly increased through many short simulations can yield a result as precise as a method that increases \( d_t \) in jumps between a few long simulations. While remaining equally precise, the quasistatic approach would have the added advantage of being more accurate, because it would virtually eliminate the error associated with the finite-step approximation to the integral. It also offers the efficiency of not requiring ‘relaxation’ cycles between subsequent simulations, assuming that the traversal of the integration path is not too rapid [74].

There is room to optimize the integration procedure by choosing how to distribute the \( M \) total simulation cycles among the quadrature points. The optimization proceeds exactly as detailed in section 4.3.1 for \( m_i^{\text{opt}} \). The precision of the integration can be improved slightly (about 30%) by optimizing in this fashion. One might also exercise the freedom to select the quadrature points. This optimization is not as easy as that encountered in optimizing staged insertion, and we did not make a serious attempt to do this. However, we did examine the variance when the quadrature points are chosen to yield uniform chemical potential differences between each pair, and we observed only a very small (2%) improvement. The precision seems to be insensitive to the selection of the quadrature points.
Thermodynamic integration is compared with the other methods in figure 3. It is clear that this is not the best approach if efficiency is the primary consideration. In particular the superficial advantage that integration enjoys over free-energy perturbation is not enough to offset the lower precision associated with the measurement of its integrand. This conclusion is only tentative, of course, given the unique nature of the integrand calculation for the hard-sphere model (where collective fluctuations still play an important role). We note that our result contrasts sharply with the conclusions made recently by Beutler et al. [4] who based their arguments on the ensemble dependence of the methods.

4.6. Distribution methods

The fourth and final class of methods reviewed in section 2 are those based on the tabulation of a frequency distribution in some suitable variable \( x \). The common choice \( x = u_i \) is untenable for hard spheres, which leads us to surmise that it is a poor choice in general. In our study we take \( x = r_{NN} \), the distance of the test sphere to its nearest neighbour. For \( r_{NN} < 1 \), \( e_1 = 0 \) while, for \( r_{NN} \geq 1 \), \( e_1 = 1 \). Thus \( r_{NN} \) uniquely determines \( e_1 \) and the method is viable. For realistic potentials this is not the case, as the full set of neighbour distances is needed to determine \( e_1 \). A variation of our present method may be applied instead. One could work with the intermediate stage presented in section 4.3.2, in which the test particle interacts with all but its nearest neighbour. The easier calculation, namely the insertion, could be computed via test-particle insertion, while the more difficult component, namely the expulsion, could be treated by a distribution method (which requires only that \( u_t \) \( = - \ln W \) be determined by \( x \); this difference is just the energy of interaction of the test particle with its nearest neighbour). An extension of this idea that focuses on two (or more) nearest neighbours could also be applied. In this case \( x = \{ r_{NN,1}, r_{NN,2} \} \) is multivariate. The approach is not as unwieldy as it may seem at first glance, and it is worth consideration although beyond the scope of the present study.

For reasons to be presented shortly, we shall apply the distribution method to determine the difference in free energy when a particle of zero diameter (rather than zero collision diameter) is transformed into a full sphere. The chemical potential is recovered once the work of inserting the zero-diameter particle is evaluated (cf. equation (37)). We otherwise restrict our attention to a single-stage method.

If we take \( f \) and \( g \) as the distribution of the distance to the test-particle’s nearest neighbour given that no spheres may lie within the distance \( r_{NN} \) (0.5 or 1.0 respectively), then each distribution can be expressed as a product:

\[
\text{probability that no spheres are within } r_{NN}, \text{ given that none is within } r_{NN}^{\text{min}} \times \text{probability density that a sphere is at } r_{NN}, \text{ given that none is within } r_{NN}^{\text{min}}.
\] (38)

Scaled-particle theory (SPT) [75, 76] tells us that the first term is of the form \( \exp \left[ a_3 r_{NN}^3 + a_2 r_{NN}^2 + a_1 r_{NN} + a_0 (g) + O(r_{NN}^2) \right] \) while the second is \( 3a_3 r_{NN}^2 + 2a_2 r_{NN} + a_1 + O(r_{NN}^2) \). The latter is independent of \( r_{NN}^{\text{min}} \) and thus is the same function for the \( f \) and \( g \) distributions. The former contains the \( f \)- and \( g \)-dependent coefficients \( a_1 \) and \( a_0 \) respectively which, according to equation (12), yield the chemical potential upon taking their difference. If we define \( f \) and \( g \) in the manner at present being considered, we would require fitting the simulation data to a form that is nonlinear in the coefficients \( a_i \). This feature is inconvenient, less precise than an analytic regression and greatly complicates the modelling of the calculation. Instead we define the \( f \) and \( g \) distributions using just the first term of equation (38).

The chemical potential measurement procedure can be summarized as follows. One simulation is performed. In it the test particle interacts as a sphere of collision diameter 0.5 (diameter zero). The simulation tabulates as a function of \( r \) the fraction of the time that no sphere lies within a distance \( r \) of the test sphere, yielding the distribution that we shall designate \( f \). The corresponding distribution is tabulated for each full (real) sphere, and the average over all spheres is the distribution we shall designate \( g \). The data are fitted to the forms

\[
\ln \left[ f(r) \right] = b_3 x^3 + b_2 x^2 + b_1 x, \\
\ln \left[ g(r) \right] = b_3 x^3 + b_2 x^2 + b_1 x + \triangle \mu,
\] (39)

where \( x = 2r - 1 \); the fit is weighted as described in appendix 1. The \( f \) distribution is defined for a test particle of non-zero collision diameter because SPT indicates that a different functional form applies for \( r < 0.5 \) (non-analyticities occur also at \( r = 1/3^{1/2} \), etc., but they are successively weaker). The point-particle chemical potential (equation (37)) is added to the regressed value of the coefficient \( \triangle \mu \) of equation (39) to yield the full hard-sphere chemical potential. Note that the method benefits from our decision to permit both \( f \) and \( g \) to be sampled from the same simulation. We see the presence of the test particle as a finite-size effect in the \( g \)-tabulation, an effect not necessarily worse than other finite-size effects that are less easily identified and removed. Offsetting this is a decision to model the \( g \) measurement on the assumption that only one full sphere were examined; in practice, tabulating distributions for all \( N - 1 \) real spheres does not add much overhead to the calculation and should be done. We take the
measured chemical potential as pertaining to the density of the \((N-1)\)-particle system.

This calculation is easily modelled, given good estimates for the expected \(f\) and \(g\) distributions. We encountered these functions (as we have now chosen to define them) before, in section 4.5 where they were used to determine the integrand for the thermodynamic integration method. There they were denoted \(p(\delta t, d)\) and represented the probability that a particle of diameter \(d_\ell\) can be grown to \(d_\ell + \delta t\) without experiencing overlap. The connection to \(f\) and \(g\) is easily made by realizing that such non-overlap implies that no spheres are within \(r = (1 + d_\ell + \delta t)/2\). Thus \(f\) and \(g\) can be estimated via a formula for the chemical potential of a solute of arbitrary diameter in an otherwise pure solvent. Further details of the variance calculation are reserved for appendix 2.

In figure 9 we present the \(f\) and \(g\) distributions predicted by the model, together with simulation data (10 000 cycles; 108 particles) for the same. The error bars on the model data indicate the variance that the model expects in the simulation data and is in very good accord with the noise level observed in the actual data (the \(g\) data are somewhat less noisy than the model describes, for the reason discussed above). Also shown in the figure are the regression weights used for fitting the data to equation (39). The variance that the model predicts for the chemical potential calculation is plotted in figure 3 as a function of density. The approach is best of all those considered. Perhaps the only surprise is that it is not even better, given the assistance provided to it by SPT. We performed 50 Monte Carlo simulations of 10 000 cycles to verify the model's estimate for the variance. The comparison is satisfactory and is included in table 2.

The method that we describe here has some features in common with the ‘SP-MC’ method recently proposed by Labik and Smith [77, 78]. Their method records the insertion probabilities for a hard sphere as a function of its diameter and extrapolates to unit diameter while accounting for certain constraints obtained from the SPT. However, they do not collect the equivalent of our \(g\) distribution to aid the extrapolation but instead apply the result from (an additional) extrapolation of the radial distribution function to contact. Also, we do not see an easy route by which this particular method could be extended to non-hard potentials. Nevertheless, their approach represents one of the only (if not the first) theoretically based interpolative means for evaluating the chemical potential.

5. Concluding remarks

Often a given chemical potential calculation method can be made more efficient by introducing certain sampling tricks (for lack of a better term). For example, the excluded-volume map sampling used by Deitrick et al. [1] or the rotational bias used by Cracknell et al. [79] can be employed to enhance the likelihood of a successful test-particle insertion. For methods in which the test particle interacts it is possible to perform trial displacements more often for molecules near the test particle [38, 80] Force-biasing techniques [5] could also be used to good effect in many of the methods that we studied. Still, tricks that look like they should be helpful are sometimes found to provide no benefit [81]. Our work has not considered such efficiencies, partly because it is difficult to gauge the computational cost of implementing them and partly to keep our study to a manageable size. In general these techniques are valuable and can have significant impact on the performance of a calculation, sufficient to affect any conclusions that one might draw from the present study.

An important component of this work is the application of quantitative models to molecular simulations so that their performance may be characterized quickly and accurately. This step has permitted us to compare and optimize various techniques for evaluating the chemical potential, and the approach may find uses in understanding other problems in molecular simulation. An important element of the analysis is a standard (but not widely known) formula (equation (24)) for the limiting covariances of the state occupancies in a Markov
process. The transient form of this equation is available [72] and may also be useful in further studies.

The problematic features of the chemical potential calculation arise in other free-energy calculations, and in these instances the lessons of the present study should be applied. The hallmark trait is a configuration space important to one of the states being much smaller than the space important to the other, and perhaps wholly contained within it (figure 1 (b)). It is feasible to formulate strategies that overcome this entropy barrier, whereas strategies based on introducing and sampling against a confinement bias yield poor results and are prone to systematic error. Thus in general a staged free-energy perturbation calculation should proceed in the ‘insertion’ direction, which is the direction of decreasing entropy. Ideally the stages will be defined so that the entropy difference from one stage to the next is constant over the entire path (cf. section 4.3.1). This contrasts with the established rule of thumb that aims for constant free-energy differences [82, 83]. Of course, the hard-sphere model does not distinguish entropy and free energy; so this statement is something of a conjecture. However, it is not too difficult to contrive examples in which two states have the same entropy but very different free energies, and which could be easily treated in a free-energy perturbation calculation, or examples of two states with very different entropies but the same free energy, and not amenable to calculation. Moreover, this conclusion is in agreement with the analysis of Wood et al. [84] who found that the magnitude of the energy change is not an important determinant of systematic errors in free-energy perturbation calculations. Unfortunately, their analysis is flawed in its assumed form for the wings of the energy distribution; it takes them as decaying faster than \( \exp(-\Delta E/kT) \) and thus would grossly underestimate the systematic error associated with a staged-deletion calculation.

A consequence of what we say is that many (not all, and perhaps not even most) free-energy perturbation calculations are much easier to conduct with precision than is commonly accepted in some circles. The current perception stems from a lack of appreciation of the asymmetry of the staging process. The results obtained from following the path in either direction are not equally reliable, and that they show some difference does not mean that they are both wrong.

The problem of adequately sampling the \((N-1)\)-particle system has been suppressed in this study; we refer here to the matter of moving the system through configuration space efficiently, avoiding non-ergodicities and navigating bottlenecks. Sometimes it is a serious problem, but often it is not [7]. Connected to this issue is the question of correlation ‘times’, which sets the correspondence between the \(M\) independent chemical potential measurements and the length of the actual simulation. To a first approximation there is no reason to believe that these correlation times will vary between the chemical potential measurement methods that we have considered; so this matter should not affect our conclusions about their relative efficiencies. The sampling problem is one that looms over all methodological issues of molecular simulation. Before it can be seriously addressed, it is necessary to have an adequate understanding of the factors that most directly impact the precision of free-energy measurements; promotion of this understanding is the primary aim of this paper. Otherwise we have no basis to distinguish failures due to inadequate sampling from those that are more easily treated, such as staging in the wrong direction.

Our study shows that the best methods for calculation of the chemical potential are among the simplest: staged insertion and distribution methods (f- and g-sampling). These are simplest in the sense that it is not required to formulate a delicately balanced weighting function to make them work. It is suggested that thermodynamic integration is not a most efficient technique, but our hard-sphere focus prevents us from concluding this generally. Efficient use of staged insertion requires some idea of the entropy (or maybe free-energy) differences to be expected along the path, so that they may be distributed evenly along it. Otherwise the approach is simple to implement and works well even for a crudely constructed ‘hard’ weighting function \(W\) (as defined in section 3). Nevertheless, it seems that there are limits to the degree to which staged insertion can yield a good result, and that the best prospect for future progress lies in the distribution methods. Their interpolative nature leverages their ability to minimize the amount of sampling required to produce a result of given precision. However, significant development is needed to make the method generally useful and reliable. In particular the proper functional forms on which to base the interpolation must be identified for the various systems of interest to practicing simulators. Most probably a multi-dimensional implementation of the method will be required.

D.A.K. would like to acknowledge valuable conversations with Hank Cochran and Sanat Kumar. The research effort of P.T.C. was supported by the Division of Chemical Sciences, Office of Basic Energy Sciences, US Department of Energy (DOE). The research effort of D.A.K. 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
Appendix 1. Variance associated with thermodynamic integration

The integrand for the thermodynamic integration pathway examined in section 4.5 describes the instantaneous change in free energy as the hard-sphere diameter is increased by a small amount. This quantity is not given by an analytic form that can be used in a simulation but must instead be evaluated by what is essentially numerical differentiation. One tabulates in a simulation the quantities \( y_i = -\ln p_i \), where \( p_i \) is the probability that the test particle encounters no overlap when its diameter is test increased to \( d_i + \epsilon i \). A weighted least-squares fit of \( y_i \) to a zero-intercept cubic in \( \epsilon \) minimizes the objective function \( \Omega \):

\[
\Omega = \sum_{i=1}^{n_i} \left[ y_i - \left( a_1 \epsilon_i + a_2 \epsilon_i^2 + a_3 \epsilon_i^3 \right) \right]^2 \omega_i^2
\]

(A.1.1)

with respect to the coefficients \( a_i \); the desired derivative \( r \) of equation (36) is simply \( a_1 \). The regression is simple enough to permit working with the analytic solution for \( a_1 \):

\[
a_1 = \frac{1}{D} \left[ (E_6 E_4 - E_3^2) Y_1 + (E_5 E_4 - E_6 E_3) Y_2 + (E_5 E_3 - E_2^2) Y_3 \right],
\]

(A.1.2)

where \( E_n = \sum_j \epsilon_i^j \omega_i^j \), \( Y_n = \sum_j y_i \epsilon_i^j \omega_i^j \) and \( D = E_6 E_4 E_2 - E_6 E_3^2 + 2 E_5 E_4 E_3 - E_5 E_2^2 - E_3^2 \). The quantities subject to stochastic error enter only through the \( Y_n \) terms. The \( p_i \) and thus the \( y_i \) are not statistically independent, but instead we expect independence among the conditional probabilities \( q_i \), where \( q_i \) is the probability that the test particle encounters no overlap on test growing to the diameter \( d_i + \epsilon_i \), given that it encounters no overlap upon growing to \( d_i + \epsilon_{i-1} \). From elementary probability we have

\[
p_i = p_{i-1} g_i,
\]

(A.1.3)

with \( p_0 = 1 \). The \( q_i \) are modelled using the MCHS hard-sphere mixture equation of state [71] via their relation to the chemical potential (cf. equation (19)): \( q_i = \exp\left[-\mu_i(d_i + \epsilon_i) + \mu_i(d_i + \epsilon_{i-1})\right] \). Each \( q_i \) describes its own little insertion average and will be distributed according to a binomial distribution (cf. equation (17)), except that the effective number of attempts to sample \( q_i \) is attenuated by the number of times that overlap is encountered upon growing to \( d_i + \epsilon_{i-1} \). Thus

\[
(\delta q_{n})^2 = \frac{q_n (1 - q_n)}{M p_{n-1}}
\]

(A.1.4)

(we performed several Monte Carlo simulations to verify this formula and the statistical independence of the \( q \)). The independent variances in the \( q \) are propagated to the coefficient \( a_1 \) in the standard way. Combination of all the relations results in a compact formula for the variance in the coefficient (the limiting slope):

\[
M(\delta a_1)^2 = \sum_{j=1}^{n_0} \left( \sum_{j=1}^{n_0} \frac{\partial q_j}{\partial y_i} \right)^2 \frac{1 - q_j}{q_j p_{j-1}} = \sum_{j=1}^{n_0} \left( \sum_{j=1}^{n_0} \frac{\partial q_j}{\partial y_i} \right)^2 \frac{1 - q_j}{q_j p_{j-1}},
\]

(A.1.5)

where the partial derivatives \( \partial q_j / \partial y_i \) are obtained from equation (A.2) but with each \( Y_n \) replaced by \( \epsilon_i \omega_i^j \). Finally, we take the weights \( \omega_i^j \) as the reciprocal of the expected variance in the corresponding \( y \). We note that the lack of independence among the \( p_i \) means that the weighted least-squares objective function of equation (A.1.1) does not provide the ‘maximum-likelihood’ estimate of the fitting coefficients. We conducted a more sophisticated regression [85] that recognizes the correlation of the \( p_i \) but found only marginal improvement to the precision; so we do not describe the analysis here.

We found that our results for the variance in the thermodynamic integration chemical potential were very insensitive to the choice of \( \epsilon \) within the range 0.001–0.02. Our calculations used \( \epsilon = 0.005 \) and \( n_0 = 200 \). The fitted form used in equation (A.1.1) is no more approximate than the models being used to characterize the hard spheres; so it provides a very accurate value for the limiting slope (i.e. the analysis sees no systematic error from the numerical differentiation).

Appendix 2. Regression and variance for \( f\text{-and } g\text{-sampling}

The regression that yields the chemical potential according to the \( f\text{-and } g\text{-sampling} \) method, implemented as described in section 4.6, is based on minimizing the objective

\[
\Omega = \sum_{i=1}^{n_0} \left[ \ln f_i - (b_1 x_i + b_2 x_i^2 + b_3 x_i^3) \right]^2 \omega_{i,f}^2
\]

\[
+ \sum_{i=1}^{n_0} \left[ \ln g_i - (\Delta \mu + b_1 x_i + b_2 x_i^2 + b_3 x_i^3) \right]^2 \omega_{i,g}^2,
\]

(A.2.1)

where \( f_i \) and \( g_i \) are the fraction of all configurations where the nearest-neighbour sphere to the test particle of diameter \( d_i \) is beyond \( r_i = (0.5 + x_i) (1 + d_i + \epsilon) / 2 \). The variance in \( \Delta \mu \) is determined using the analysis presented for the variance of the thermodynamic integration integrand. In the present case the explicit formula for \( \Delta \mu \) analogous to equation (A.1.2) is too unwieldy to reproduce here, but it is easily obtained and coded in Fortran using the Maple symbolic mathematics package. The weights for the fit are determined from the variances in \( \ln f \) and \( \ln g \), as described in appendix 1.
Appendix 3. Semiempirical formulae for hard spheres

The models used in this work required knowledge of some thermodynamic and structural properties of hard spheres; in particular we needed the chemical potential of a component in a mixture, and the cavity correlation function. As a matter of convenience we reproduce here the well established semiempirical formulae that we applied to the simulation models.

We already presented (equation (19)) the formula for the chemical potential of pure hard spheres. The MCSP [71] equation of state is a simple and accurate generalization of the Carnahan–Starling [69] treatment to mixtures. From it, one can derive a chemical potential for a mixture component as a cubic polynomial in its diameter \(d\); thus

\[
\mu = - \ln (1 - \xi) + \psi_1 d + \psi_2 d^2 + \psi_3 d^3, \quad (A 3.1)
\]

where the coefficients \(\psi_i\) depend on the moments of the distribution of diameters:

\[
\psi_1 = \frac{3 \xi}{1 - \xi},
\]

\[
\psi_2 = \frac{3 \xi^2}{1 - \xi} + \frac{3 \xi^2}{\xi_0(1 - \xi)} + \frac{3 \xi^2}{\xi^2(1 - \xi)} \ln (1 - \xi),
\]

\[
\psi_3 = \left(1 - \frac{\xi^3}{\xi_0^3} + \frac{3 \xi^2}{\xi_0(1 - \xi)} \right) + \frac{3 \xi^2}{1 - \xi} + \frac{3 \xi^2}{(1 - \xi)^2} \ln (1 - \xi), \quad (A 3.2)
\]

where \(\xi_0 = (\pi/6) \rho \sum x_i d_i^n\), and \(x_i\) is the mole fraction of the component of diameter \(d_i\) and the sum is over all species. We considered only systems at infinite dilution in an otherwise pure fluid of hard spheres of diameter \(d = 1\). In this case the coefficients simplify thus:

\[
\psi_1 = \frac{3 \eta}{1 - \eta},
\]

\[
\psi_2 = \frac{3 \eta(2 - \eta)}{(1 - \eta)^2} + 3 \ln (1 - \eta),
\]

\[
\psi_3 = - \frac{\eta(3 \eta^2 - 6 \eta + 1)}{(1 - \eta)^3} - 2 \ln (1 - \eta), \quad (A 3.3)
\]

where \(\eta = (\pi/6) \rho\).

Boublík [73] exploited the relation between the cavity correlation function \(\gamma(r)\) and the chemical potential of a hard-dumbbell molecule of bond length \(r\) in a system of hard spheres. The formula that he derived is as follows:

\[
\ln [\gamma(r)] = 2 \mu_{HS} - \mu_{HD} \quad (A 3.4)
\]

with

\[
\mu_{HS} = - \ln (1 - \eta) + \frac{7 \eta}{1 - \eta} + \frac{15 \eta^2}{2(1 - \eta)^2} + \frac{\eta^3(2 - \eta/3)}{(1 - \eta)^3},
\]

\[
\mu_{HD} = - \ln (1 - \eta) + (3 R + 3 S + V) \frac{\eta}{1 - \eta}
\]

\[
+ (3 R^2 + 6 S^2 + 6 V) \frac{\eta^2}{2(1 - \eta)^2} + V \frac{\eta^3(2 - \eta/3)}{(1 - \eta)^3}, \quad (A 3.5)
\]

where

\[
R = 1 + \frac{1}{2} r,
\]

\[
S = 1 + r,
\]

\[
V = 1 + \frac{3}{2} r - \frac{1}{2} r^3.
\]

References