

Advanced Monte Carlo Techniques

D. Frenkel

FOM Institute for Atomic and Molecular Physics

Kruislaan 407

1098 SJ Amsterdam

The Netherlands

ABSTRACT. This Chapter starts with a description of Monte Carlo techniques to study many-body systems in the canonical (constant- NVT), isobaric-isothermal (constant- NPT) and grand-canonical (constant- μVT) ensembles. Subsequently, I give a unified description of two families of 'smart' Monte Carlo schemes, namely Swendsen-Wang-style cluster moves, and configurational-bias MC techniques for collective moves of pre-selected particles. Next, I briefly discuss free-energy calculations in the context of first-order phase transitions. I conclude with a discussion of recent advances in grand-canonical simulations of (hybrid) lattice models.

1. Introduction

In this Chapter, I will assume that the basic idea behind the Monte Carlo technique of Metropolis *et al.* [1], as described in Chapter 1, is known. Nevertheless, I briefly recall those aspects of the Metropolis method that are essential for what follows. The Metropolis method provides us with a recipe to perform a random walk through configuration space, in such a way that the probability of visiting a given configuration i (i.e. a given set of coordinates $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$), is proportional to the Boltzmann weight g_i of that configuration. In order to construct such a random walk, one should specify the matrix π_{ij} that determines the probability to make a transition from state i to another state j . The matrix π_{ij} should satisfy the important condition that, if we prepare an ensemble of systems in equilibrium, then after one Monte Carlo step, this ensemble is still in equilibrium. One way to achieve this is to impose that π_{ij} satisfies the detailed-balance condition

$$g_i \pi_{ij} = g_j \pi_{ji} . \quad (1)$$

There are many possible forms of the transition matrix π_{ij} that satisfy eqn (1). Let us consider how π_{ij} is constructed in practice. We recall that

a Monte Carlo move consists of two stages: first we perform a trial move from state i to state j . We denote the transition matrix that determines the probability to perform a trial move from i to j by α_{ij} . α is usually referred to as the underlying matrix of the Markov chain. The next stage is the decision to either accept or reject this trial move. Let us denote the probability of accepting a trial move from i to j by P_{ij} . Clearly, $\pi_{ij} = \alpha_{ij}P_{ij}$. In the original Metropolis scheme, α is chosen to be a symmetric matrix. However, in later sections we shall see several examples where α is *not* symmetric. If α is symmetric, we can rewrite eqn (1) in terms of the P_{ij} only :

$$\varrho_i P_{ij} = \varrho_j P_{ji} . \quad (2)$$

From eqn (2) follows:

$$P_{ij}/P_{ji} = \varrho_j/\varrho_i = \exp(-\beta(\mathcal{V}_j - \mathcal{V}_i)) . \quad (3)$$

Again, there are many possible choices for P_{ij} that satisfy this condition (and the obvious condition that the probability P_{ij} cannot exceed 1). The choice of Metropolis *et al.* was

$$P_{ij} = \min(\varrho_j/\varrho_i, 1) . \quad (4)$$

Note that we have not specified the matrix α , but for the fact that it must be symmetric. This reflects the fact that there is a considerable freedom in the choice of our trial moves. We will come back to this point in subsequent sections. Thus far, we have not mentioned another important condition that π_{ij} should satisfy, namely that it is *ergodic*, i.e. that every accessible point in configuration space can be reached in a finite number of Monte Carlo steps from any other point. Although there are simple MC schemes that are guaranteed to be ergodic, these are often not the most efficient schemes. Conversely, there are many efficient Monte Carlo schemes that have either not been proven to be ergodic or, worse, been proven to be non-ergodic. If you are worried that purists may dislike your favorite non-ergodic scheme, the solution is to mix that scheme with a less efficient ergodic scheme. Then everybody is happy.

Before proceeding with a discussion of Monte-Carlo sampling methods, it is important to stress one limitation of the Metropolis-style MC method. The MC method *can* be used to estimate configuration-space averages of structural properties of a many-body system, i.e.

$$\langle \mathcal{A} \rangle_{NVT} = \frac{\int d\mathbf{q}^N \mathcal{A}(\mathbf{q}^N) \exp(-\beta\mathcal{V}(\mathbf{q}^N))}{Z_{NVT}} , \quad (5)$$

with $Z_{NVT} = \int d\mathbf{q}^N \exp(-\beta\mathcal{V}(\mathbf{q}^N))$. However, the MC method cannot be used to compute the configurational integral Z_{NVT} , as it depends on the accessible *volume* in configuration space. Knowledge of the configurational integral is required to compute "thermal" properties, such as the Helmholtz free

energy F , the Gibbs free energy G , and the chemical potential μ . Techniques to compute the latter quantities (that play a crucial role in the calculation of phase equilibria), will be discussed in considerable detail in subsequent sections.

2. Other ensembles

In the previous section, I have briefly reviewed the Metropolis Monte Carlo simulation method. This technique was designed to compute the average value of an arbitrary function $\mathcal{A}(\mathbf{q}^N)$ of the coordinates of a system of N particles at constant temperature T and volume V (eqn (5)):

$$\langle \mathcal{A} \rangle_{NVT} = \frac{\int d\mathbf{q}^N \mathcal{A}(\mathbf{q}^N) \exp(-\beta\mathcal{V}(\mathbf{q}^N))}{Z_{NVT}}.$$

This average is usually referred to as the ‘canonical’ ensemble average. To be more precise, the true canonical ensemble average includes the particle momenta. However, in all cases of practical interest, the integration over momenta can be carried out analytically, and therefore does not require numerical sampling.

A large fraction of all Monte Carlo calculations to date have been performed in the constant- NVT ensemble. However, it is sometimes advisable to perform a simulation where thermodynamic variables other than N, V or T are kept fixed. In the present section we discuss such simulations. I will limit this discussion to those ensembles that have found most widespread use in the simulation of classical many-body systems, such as the constant- NPT and μVT ensembles. I refer readers interested in *microcanonical* MC simulations to the original paper by Creutz [2] and the book by Binder [3].

2.1. ISOBARIC-ISOTHERMAL ENSEMBLE

Apart from the canonical ensemble, the isobaric-isothermal (constant- NPT) ensemble is probably the one most widely used in Monte Carlo simulations. This is not surprising because most real experiments are also carried out under conditions of controlled pressure and temperature. Moreover, constant- NPT simulations can be used to measure the equation of state of a model system even if the virial expression for the pressure cannot be readily evaluated. This may, for instance, be the case for certain models of non-spherical hard-core molecules. But also for the increasingly important class of models where the (non-pairwise additive) potential energy function is computed *numerically* for each new configuration. Finally, it is often convenient to use constant- NPT MC to simulate systems in the vicinity of a first-order phase transition, because at constant pressure the system is free (given enough time, of course) to transform completely into the state of lowest (Gibbs)

free energy, whereas in a constant- NVT simulation the system may be kept at a density where it would like to phase-separate into two bulk phases of different density, but is prevented from doing so by finite size effects.

Constant-pressure Monte Carlo simulations were first described by Wood [4] in the context of a simulation study of two-dimensional hard disks. Although the method introduced by Wood is very elegant, it is not readily applied to systems with arbitrary continuous potentials. McDonald [5] was the first to apply constant- NPT simulations to a system with continuous intermolecular forces (a Lennard-Jones mixture), and it is the constant-pressure method of McDonald that is now being used almost universally, and that is discussed below.

I will derive the basic equations of constant-pressure Monte Carlo in a way that may appear unnecessarily complicated. However, this derivation has the advantage that the same framework can be used to introduce some of the other non- NVT Monte Carlo methods to be discussed below. For the sake of convenience we shall initially assume that we are dealing with a system of N identical atoms. The partition function for this system is given by:

$$Q(N, V, T) = \frac{1}{\Lambda^{3N} N!} \int_0^L \cdots \int_0^L d\mathbf{q}^N \exp(-\beta\mathcal{V}(\mathbf{q}^N)), \quad (6)$$

where $\Lambda = \sqrt{h^2/2\pi mk_B T}$ is the thermal de Broglie wavelength. It is convenient to rewrite eqn (6) in a slightly different way. We have assumed that the system is contained in a cubic box with diameter $L = V^{1/3}$. We now define scaled coordinates \mathbf{s}^N , by:

$$\mathbf{q}_i = L\mathbf{s}_i$$

for $i = 1, 2, \dots, N$. If we now insert these scaled coordinates in eqn (6) we obtain:

$$Q(N, V, T) = \frac{V^N}{\Lambda^{3N} N!} \int_0^1 \cdots \int_0^1 d\mathbf{s}^N \exp(-\beta\mathcal{V}(\mathbf{s}^N; L)). \quad (7)$$

In eqn (7), we have written $\mathcal{V}(\mathbf{s}^N; L)$ to indicate that \mathcal{V} depends on the *real* rather than the *scaled* distances between the particles. The expression for the Helmholtz free energy of the system is:

$$\begin{aligned} F(N, V, T) &= -k_B T \ln Q \\ &= -k_B T \ln \left(\frac{V^N}{\Lambda^{3N} N!} \right) - k_B T \ln \left(\int d\mathbf{s}^N \exp(-\beta\mathcal{V}(\mathbf{s}^N; L)) \right) \\ &= F_{\text{id}}(N, V, T) + F_{\text{ex}}(N, V, T). \end{aligned} \quad (8)$$

In the last line of the above equation we have identified the two contributions to the Helmholtz free energy on the previous line as the ideal-gas expression

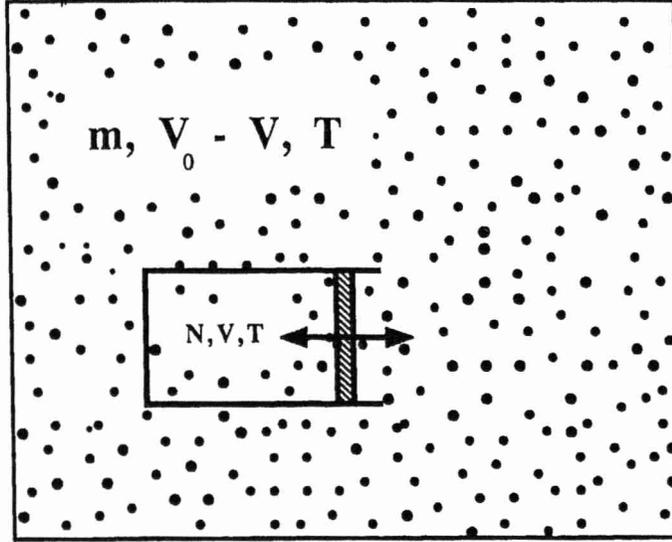


Fig. 1. Ideal gas (m particles, volume $V_0 - V$) can exchange volume with an N -particle system (volume V)

plus an excess part. Let us now assume that the system is separated by a 'piston' [6] from an ideal gas reservoir (see Fig. 1). The total volume of the system plus reservoir is fixed at a value V_0 . The total number of particles is M . Hence the volume accessible to the $m \equiv M - N$ ideal gas molecules is $V_0 - V$. The partition function of the total system is simply the product of the partition functions of the constituent sub-systems:

$$Q(N, m, V, V_0, T) = \frac{V^N (V_0 - V)^m}{\Lambda^{3M} N! m!} \int ds^m \int ds^N \exp(-\beta \mathcal{V}(s^N; L)). \quad (9)$$

Note that the integral over the s^m scaled coordinates of the ideal gas yields simply 1. For the sake of compactness we have assumed that the thermal wavelength of the ideal gas molecules is also equal to Λ . The total free energy of this combined system is $F_{\text{tot}} = -k_B T \ln Q(N, m, V, V_0, T)$. Now let us assume that the 'piston' between the two subsystems is free to move, so that the volume V of the N -particle subsystem can fluctuate. Of course, the most probable value of V will be the one that minimizes the free energy of the combined system. The probability density $\mathcal{P}(V)$ that the N -particle

subsystem has a volume V is given by:

$$\mathcal{P}(V) = \frac{V^N (V_0 - V)^m \int ds^N \exp(-\beta \mathcal{V}(s^N; L))}{\int_0^{V_0} dV' V'^N (V_0 - V')^m \int ds^N \exp(-\beta \mathcal{V}(s^N; L'))} \quad (10)$$

We now consider the limit that the size of the reservoir tends to infinity ($V_0 \rightarrow \infty, M \rightarrow \infty, m/V_0 \rightarrow \rho$). In that limit, a small volume change of the small system does not change the pressure P of the large system. In other words, the large system works as a manostat for the small system. In that case we can simplify eqns (9) and (10). Note that in the limit $V/V_0 \rightarrow 0$ we can write:

$$(V_0 - V)^m = V_0^m (1 - (V/V_0))^m \rightarrow V_0^m \exp(-mV/V_0).$$

In the limit $m \rightarrow \infty$, $\exp(-mV/V_0) \rightarrow \exp(-\rho V)$. But, as the reservoir contains an ideal gas, ρ can be written as βP . With these substitutions, eqn (10) becomes:

$$\mathcal{P}(V) = \frac{V^N \exp(-\beta PV) \int ds^N \exp(-\beta \mathcal{V}(s^N; L))}{\int_0^{V_0} dV' V'^N \exp(-\beta PV') \int ds^N \exp(-\beta \mathcal{V}(s^N; L'))} \quad (11)$$

In the same limit, the difference in free energy between the combined system and the ideal-gas system in the absence of the N -particle subsystem is the well-known Gibbs free energy G :

$$G(N, P, T) = -k_B T \ln \int dV \frac{V^N \exp(-\beta PV)}{\Lambda^{3N} N!} \int ds^N \exp(-\beta \mathcal{V}(s^N; L)) \quad (12)$$

Equation (11) is the starting point for constant- NPT Monte Carlo simulations. The idea is that the probability density to find the small system in a particular configuration of the N atoms (as specified by s^N) at a given volume V is given by:

$$\mathcal{P}(V; s^N) = \frac{V^N \exp(-\beta PV) \exp(-\beta \mathcal{V}(s^N; L))}{\int_0^{V_0} dV' V'^N \exp(-\beta PV') \int ds^N \exp(-\beta \mathcal{V}(s^N; L'))}$$

We can now carry out Metropolis sampling on the reduced coordinates s^N and the volume V , with a weight-function $\rho(s^N, V)$ proportional to $\exp(-\beta\{\mathcal{V}(s^N, V) + PV - N\beta^{-1} \ln V\})$. In the constant- NPT MC method, V is simply treated as an additional coordinate, and trial moves in V must satisfy the same rules as trial moves in q (in particular, we should maintain the reversibility of the underlying Markov chain). Let us assume that our trial moves consist of an attempted change of the volume from V to $V' = V + \Delta V$, where ΔV is a random number, uniformly distributed over the interval $[-\Delta V_{\max}, +\Delta V_{\max}]$. In the Metropolis scheme such a random, volume-changing move will be accepted if:

$$\exp(-\beta[\mathcal{V}(s^N, V') - \mathcal{V}(s^N, V) + P(V' - V) - N\beta^{-1} \ln(V'/V)]) > \mathcal{R}, \quad (13)$$

where \mathcal{R} is a random number, uniformly distributed over the interval $[0, 1]$. Instead of attempting random changes in the volume itself, one might construct trial moves in the box-length L [5], or in the logarithm of the volume [7]. Such trial moves are equally legitimate, as long as the reversibility of the underlying Markov chain is maintained. However, such alternative schemes result in a slightly different form for eqn (13). The isobaric-isothermal MC method has been extended to simulate systems at constant stress. This problem was first in the context of molecular dynamics simulations by Parrinello and Rahman [8, 9] who developed an extension of the constant-pressure molecular dynamics technique introduced by Andersen [10]. The extension of the Parrinello-Rahman method to Monte Carlo simulations is straightforward [11].

2.2. GRAND CANONICAL ENSEMBLE

The Metropolis sampling scheme was introduced as a method to compute thermal averages of functions $\mathcal{A}(\mathbf{q}^N)$, that depend explicitly on the coordinates of the molecules in the N -body system under study. Examples of such ‘mechanical’ properties are the potential energy or the virial contribution to the pressure. However, the Metropolis method cannot be used to determine the configurational integral

$$Z_{NVT} \equiv \int d\mathbf{q}^N \exp(-\beta\mathcal{V}(\mathbf{q}^N))$$

itself. The latter quantity measures the effective volume in configuration space that is accessible to the system. Hence the original Metropolis scheme cannot be used to determine those thermodynamic properties of a system that depend explicitly on the configurational integral. Examples of such ‘thermal’ properties are: the Helmholtz free energy

$$F = -k_B T \ln Q_{NVT} ,$$

the entropy S and the Gibbs free energy G . However, although the Metropolis method cannot be used to measure, for instance, free energies directly, it can be used to measure the *difference* in free energy between two possible states of an N -body system. This fact is exploited in the grand-canonical Monte Carlo method (GCMC) first implemented for classical fluids by Norman and Filinov [12], and later extended and improved by a number of other groups [13–21]. The basic idea of the GCMC method is explained below.

In order to understand the statistical mechanical basis of the GCMC technique, let us return to eqn (9) of section 2.1. This equation gives the partition function of a combined system of N interacting particles in volume V and $m = M - N$ ideal gas molecules in volume $V_0 - V$:

$$Q(N, m, V, V_0, T) = \frac{V^N (V_0 - V)^m}{\Lambda^{3M} N! m!} \int ds^m \int ds^N \exp(-\beta\mathcal{V}(s^N)) . \quad (9)$$

Now, instead of allowing the two systems exchange volume, let us see what happens if the systems can exchange particles (see Fig. 2). To be more

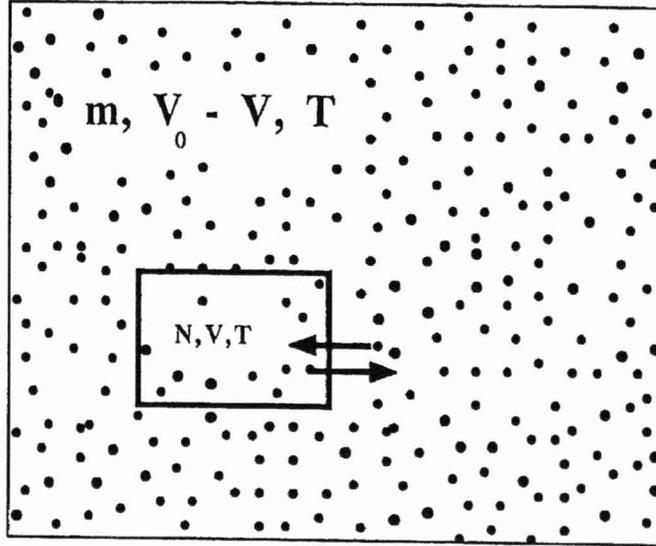


Fig. 2. Ideal gas (m particles, volume $V_0 - V$) can exchange particles with an N -particle system (volume V)

precise, we assume that the molecules in the two sub-volumes are actually identical particles. The only difference is that when they find themselves in volume V they interact and when they are in volume $V_0 - V$, they do not. If we transfer a molecule i from a reduced coordinate s_i in the volume $V_0 - V$ to the same reduced coordinate in volume V , then the potential energy function \mathcal{V} changes from $\mathcal{V}(s^N)$ to $\mathcal{V}(s^{N+1})$. The expression for the total partition function of the system, including all possible distributions of the M particles over the two sub-volumes is:

$$Q(M, V, V_0, T) = \sum_{N=0}^M \frac{V^N (V_0 - V)^{M-N} \int ds^{M-N} \int ds^N \exp(-\beta \mathcal{V}(s^N))}{\Lambda^{3M} N! (M - N)!} . \quad (14)$$

Following the approach of section 2.1, we now write the probability density to find a system with $M - N$ particles at reduced coordinates s^{M-N} in volume $V' \equiv V_0 - V$ and N particles at reduced coordinates s^N in volume V :

$$\mathcal{P}(s^M; N) = \frac{V^N V'^{M-N}}{Q(M, V, V', T) \Lambda^{3M} N! (M - N)!} \exp(-\beta \mathcal{V}(s^N)) . \quad (15)$$

Let us now consider a trial move in which a particle is transferred from V' to the same scaled coordinate in V . First we should make sure that we construct an underlying Markov chain that is symmetric. Symmetry, in this case, implies that the *a priori* probability to move a particle from V' to V should be equal to the *a priori* probability of the reverse move. The probability of acceptance of a trial move in which we move a particle *to* or *from* volume V is determined by the ratio of the corresponding probability densities (eqn (15)):

$$P_{N \rightarrow N+1} = \frac{V(M-N)}{V'(N+1)} \exp(-\beta(\mathcal{V}(\mathbf{s}^{N+1}) - \mathcal{V}(\mathbf{s}^N))) \quad (16)$$

$$P_{N+1 \rightarrow N} = \frac{V'(N+1)}{V(M-N)} \exp(-\beta(\mathcal{V}(\mathbf{s}^N) - \mathcal{V}(\mathbf{s}^{N+1}))). \quad (17)$$

Now let us consider the limit that the ideal gas system is very much larger than the interacting system: $M \rightarrow \infty$, $V' \rightarrow \infty$, $(M/V') \rightarrow \rho$. Note that for an ideal gas the chemical potential μ is related to the particle-density ρ by: $\mu = k_B T \ln \Lambda^3 \rho$. Therefore, in the limit $(M/N) \rightarrow \infty$, eqns (16),(17) become:

$$P_{N \rightarrow N+1} = \frac{V}{N+1} \Lambda^{-3} \exp(\beta\mu) \exp(-\beta(\mathcal{V}(\mathbf{s}^{N+1}) - \mathcal{V}(\mathbf{s}^N))) \quad (18)$$

$$P_{N+1 \rightarrow N} = \frac{(N+1)}{V} \Lambda^3 \exp(-\beta\mu) \exp(-\beta(\mathcal{V}(\mathbf{s}^N) - \mathcal{V}(\mathbf{s}^{N+1}))). \quad (19)$$

In the Metropolis scheme, an attempted particle addition to (removal from) volume V is accepted if $P_{N \rightarrow N+1} > \mathcal{R}$ ($P_{N+1 \rightarrow N} > \mathcal{R}$), where \mathcal{R} is a random number uniformly distributed in the interval $[0, 1]$.

Equations (18), (19) are the basic equations for grand canonical Monte Carlo simulations. Acceptable trial moves are: (i) the random insertion of an additional particle at any point in volume V ; (ii) the random removal of any of the N particles in volume V . In addition, the particles in volume V can sample the accessible configuration space by conventional Monte Carlo sampling. Note that in eqns (18), (19) all explicit reference to the ideal-gas system has disappeared. The rules only refer to the system of volume V .

The most salient feature of the GCMC technique is that in such simulations the chemical potential μ is *imposed*, while the number of particles N is a fluctuating quantity. During the simulation we may measure other thermodynamic quantities, such as the pressure P , the average density $\langle \rho \rangle$ or the internal energy $\langle \mathcal{V} \rangle$. As we know μ , we can derive all other 'thermal' properties, such as the Helmholtz free energy or the entropy. This may seem surprising. After all, in section 1 we stated that Metropolis sampling *cannot* be used to sample absolute free energies and related quantities. Yet, with grand canonical MC we appear to be doing precisely that. The answer is that, in fact, we do not. What we measure is not an absolute but a relative

free energy. In GCMC, we are equating the chemical potential of a molecule in an ideal gas at density ρ (for the ideal-gas case we know how to compute μ) and the chemical potential of the same species in an interacting system at density ρ' .

Grand canonical MC works best if the acceptance of trial moves by which particles are added or removed is not too low. For atomic fluids this condition effectively limits the maximum density at which the method can be used to about twice the critical density. Special tricks are needed to extend GCMC to somewhat higher densities [19]. GCMC is easily implemented for mixtures and inhomogeneous systems, such as fluids near interfaces. In fact, some of the most useful applications of the GCMC method are precisely in these areas of research. Although GCMC can be applied to simple models of non-spherical molecules, it converges very poorly for all but the smallest polyatomic molecules. However, in section 4.1, we shall discuss techniques that can be used to generalize GCMC simulations to chain molecules. For more details on the more conventional grand canonical Monte Carlo method, the reader is referred to the book by Allen and Tildesley [22] and a review paper by Frenkel [23].

2.3. GIBBS ENSEMBLE

One of the prime reasons why we are interested in simulations at constant P or constant μ , is that the condition for coexistence of two or more phases I, II, ... is that the pressure of all coexisting phases must be equal ($P_I = P_{II} = \dots = P$), as must be the temperature ($T_I = T_{II} = \dots = T$) and the chemical potentials of all species ($\mu_I^\alpha = \mu_{II}^\alpha = \dots = \mu^\alpha$). Hence one might be inclined to think that the best ensemble to study would be the 'constant- μPT ensemble'. The quotation marks around the name of this 'ensemble' are intentional, because, strictly speaking, no such ensemble exists. The reason is simple: if we specify only intensive parameters, such as P, T and μ , the extensive variables (such as V) are unbounded. Another way to say the same thing is that the set P, T, μ is linearly dependent. In order to get a decent ensemble, we must fix at least one extensive variable. In the case of constant-pressure MC this variable is the number of particles N , while in grand canonical Monte Carlo, the volume V of the system is fixed.

A few years ago, however, Panagiotopoulos [24] developed a new Monte Carlo scheme, called the 'Gibbs method', that comes very close to achieving the impossible: namely simulating phase equilibria under conditions where the pressure, temperature and chemical potential(s) of the coexisting phases are equal. The reason why this method can work is that, although the *difference* between chemical potentials in different phases is fixed (namely, at $\Delta\mu = 0$), the absolute values are still free to change. The Gibbs-ensemble method has developed into one of the most important techniques to study

phase-equilibria (in particular, fluid-fluid equilibria) by Monte Carlo simulation. In view of the great importance of this method, Chapter 6 in this volume has been devoted to the ins and outs of the Gibbs ensemble method.

2.4. SEMI-GRAND ENSEMBLE

At present, the Gibbs ensemble method has become the technique *par excellence* to study fluid-fluid phase equilibria. However, like the grand-canonical ensemble scheme, the method does rely on a reasonable number of successful particle insertions to achieve compositional equilibrium. As a consequence, the Gibbs ensemble method cannot be used to study phase equilibria involving very dense phase or crystalline solids. Although there exist, to my knowledge, no numerical schemes that allow *direct* simulation of, say, solid-liquid phase coexistence in a multicomponent system, there exists a technique that greatly facilitates the numerical study of such phase equilibria. This is the so-called 'semi-grand ensemble method' of Kofke and Glandt [25]. In this section, I shall present a simple description of this method.

In order to understand why the semi-grand ensemble can be successful where other schemes fail, it is useful to borrow a result that will be derived in section 3.1. In that section, we show that the excess chemical potential of a molecule in a fluid is related to the average Boltzmann factor associated with the random addition of such a molecule to a fluid with N particles present (eqn (40)):

$$\mu_{ex} = -k_B T \ln \left\langle \int ds_{N+1} \exp(-\beta \Delta \mathcal{V}) \right\rangle_N .$$

Let us assume that we wish to study phase equilibria in a binary mixture. In that case, we must compute the Gibbs free energy per mole of the mixture, as a function of the composition:

$$G(X_A) = X_A \mu_A + X_B \mu_B , \quad (20)$$

where $X_A (= 1 - X_B)$ denotes the mole fraction of species A , and μ_A (μ_B) denote the chemical potentials of the two components in the mixture. Let us assume that we have somehow succeeded in computing the Gibbs free energy of one of the pure phases (for instance, by one of the thermodynamic integration methods described in section 5). At first sight it might seem that, in order to compute G as a function of X_A , we have to repeat such a thermodynamic integration for a large number of X_A values. Fortunately, this is usually not the case. Rather than recomputing $G(X)$ for a number of compositions, we can study the *change* in $G(X)$ with X . In other words, we need to have a 'microscopic' expression for

$$\left(\frac{\partial G(X)}{\partial X} \right)_{P,T,N} = \mu_A - \mu_B \quad (21)$$

$$= (\mu_A - \mu_B)_{\text{id}} + (\mu_A - \mu_B)_{\text{ex}} .$$

In the first line of eqn (21), we have used the Gibbs-Duhem relation. I assume that the ideal-gas contributions to the chemical potential of both A and B are known. The quantity that we must compute is $\Delta\mu_{\text{ex}} \equiv (\mu_A - \mu_B)_{\text{ex}}$. Naively, one might try to measure this quantity by using the particle insertion method to obtain μ_{ex} of species A and B separately and then subtracting the result. Although such an approach would be correct *in principle*, it is time consuming and not very accurate. Fortunately, $\Delta\mu_{\text{ex}}$ can be obtained much more directly by measuring the Boltzmann factor associated with a virtual trial move where a randomly selected particle of type B is transformed into a particle of type A [26–28]. I leave it as an exercise to the reader to derive that the resulting expression for $\Delta\mu_{\text{ex}}$ at constant pressure is:

$$\Delta\mu_{\text{ex}} = -k_{\text{B}}T \ln \left\langle \frac{N_B}{N_A + 1} \exp(-\beta\Delta\mathcal{V}^{+-}) \right\rangle \quad (22)$$

where $\Delta\mathcal{V}^{+-}$ denotes the change in potential energy of the system if one particle of type B is changed into type A . $-k_{\text{B}}T \ln(N_B/[N_A + 1])$ is simply the ideal mixing contribution to the chemical potential. The point to note about eqn (22) is that for a perfect mixture (i.e. A and B have the same intermolecular interactions), $\ln\langle \exp(-\beta\Delta\mathcal{V}^{+-}) \rangle$ is identically equal to zero. In other words, we may obtain very good statistics on $\Delta\mu_{\text{ex}}$ even if the direct measurement of the excess chemical potential of the individual species would yield poor statistics.

The aim of this introduction to the semi-grand ensemble is twofold: first of all, eqn (22) shows that the Boltzmann factor associated with the change of identity of a particle in a mixture is related to the difference in excess chemical potential of the two species that take part in the interchange. And secondly, I have made it plausible that we can get good statistics on $\Delta\mu_{\text{ex}}$ even when the particle insertion method to measure the excess chemical potential of the individual species fails, for instance in a crystalline solid [29]. I recall that the grand-canonical Monte Carlo method has about the same range of applicability as the particle insertion method. It is therefore logical to infer that it should be possible to construct a simulation scheme based on particle *interchanges* that should work under conditions where the GCMC scheme fails. The semi-grand canonical Monte Carlo (SGCMC) method is such a scheme.

How does the SGCMC method work? Let us first consider the expression for Ξ , the grand-canonical partition function for an n -component mixture:

$$\Xi = \sum_{N_1, N_2, \dots, N_n} \prod_{i=1}^n \frac{q_i^{N_i} \exp(\beta\mu_i N_i)}{N_i!} V^N \int ds^N \exp(-\beta\mathcal{V}(s^N)) , \quad (23)$$

where $N \equiv \sum_i N_i$, $\mathcal{V}(\mathbf{s}^N)$ denotes the potential energy function of the n -component mixture and q_i is the 'kinetic' contribution to the partition function due to species i . Next, we consider a related partition function Ξ' , identical to Ξ , but for the fact that we have imposed the constraint that $N = \sum_i N_i$ is fixed. If N is fixed, we can eliminate one of the N_i , for instance N_1 , from the sum in eqn (23) and we obtain:

$$\Xi' = \sum_{N_2, \dots, N_n} q_1^N \exp(\beta \mu_1 N) \prod_{i=1}^n \left(\frac{q_i}{q_1} \right)^{N_i} \frac{\exp(\beta(\mu_i - \mu_1)N_i)}{N_i!} \times \left(V^N \int d\mathbf{s}^N \exp(-\beta \mathcal{V}(\mathbf{s}^N)) \right). \quad (24)$$

We now multiply this equation on both sides by $\exp(-\beta \mu_1 N)$ and we define a new partition function $\mathcal{Y} \equiv \Xi' \exp(-\beta \mu_1 N)$:

$$\mathcal{Y} = \sum_{N_2, \dots, N_n} q_1^N \prod_{i=1}^n \left(\frac{q_i}{q_1} \right)^{N_i} \frac{\exp(\beta(\mu_i - \mu_1)N_i)}{N_i!} \times \left(V^N \int d\mathbf{s}^N \exp(-\beta \mathcal{V}(\mathbf{s}^N)) \right). \quad (25)$$

The next step is subtle. We shall reinterpret the sum over all N_i in eqns (23)–(25). In these equations, we had assumed that to every composition N_1, N_2, \dots, N_n , there corresponds *one* term in the sum. Let us now take a somewhat different point of view: we assume that all these different species are all manifestations of one and the same 'particle'. This sounds strange, so I shall use an analogy to explain what I mean. Let us consider that we have a group of 100 people, made up of 3 groups: eaters, drinkers and sleepers. In fact, we want to consider all possible combinations of these groups, with the constraint that the total number is fixed. One such combination would be: 30 eaters, 30 drinkers and 40 sleepers. Now we make a discovery: one and the same person can be an eater, a drinker or a sleeper, but not simultaneously. Now our sum over all combinations becomes different: we have 100 'persons' who can all take on any one out of the three possible identities. In that case, we have many more ways in which we can make a group of 30 eaters, 30 drinkers and 40 sleepers, namely $100!/(30!30!40!)$. If we wish to have the same total number of terms in our sum as before, we have to divide by this factor.

Let us now translate this example back to the sum over particles in eqn (25). We replace the sum over numbers of particles of species i , by a sum over all possible identities of all particles. But then we must correct for double counting by dividing by $N!/(N_1! \times \dots \times N_n!)$. If we do that,

eqn (25) becomes:

$$\mathcal{Y} = \sum_{\text{identities}} \frac{q_1^N}{N!} \prod_{i=1}^n \left(\frac{q_i}{q_1} \right)^{N_i} \exp(\beta(\mu_i - \mu_1)N_i) \times \left(V^N \int ds^N \exp(-\beta\mathcal{V}(s^N)) \right). \quad (26)$$

Finally, it turns out that it is more convenient to consider the corresponding ensemble at constant pressure. In that case, the partition function changes to

$$\mathcal{Y}' = \beta P \int dV \exp(-\beta PV) \frac{(V q_1)^N}{N!} \times \left(\sum_{\text{identities}} \prod_{i=1}^n \left(\frac{q_i}{q_1} \right)^{N_i} \exp(\beta(\mu_i - \mu_1)N_i) \int ds^N \exp(-\beta\mathcal{V}(s^N)) \right). \quad (27)$$

For cosmetic reasons, we rewrite eqn (27) in terms of the *fugacities* f_i , rather than the chemical potentials μ_i . Recall that the fugacity of a species i is defined by the expression:

$$\mu_i(P, T, \{X_i\}) \equiv \mu_i^0(T) + k_B T \ln f_i, \quad (28)$$

where $\mu_i^0(T)$ is the chemical potential of the ideal gas reference state ($P = 1$) of species i . Using the expression for the chemical potential of an ideal gas at pressure P , and recalling that for an ideal gas $f = P$, it is easy to show that:

$$\mu_i^0(T) = -k_B T \ln(k_B T q_i). \quad (29)$$

Inserting eqn (29) in eqn (27), we obtain:

$$\mathcal{Y}' = \beta P \int dV \exp(-\beta PV) \frac{(V q_1)^N}{N!} \sum_{\text{identities}} \prod_{i=1}^n \left(\frac{f_i}{f_1} \right)^{N_i} \times \left(\int ds^N \exp(-\beta\mathcal{V}(s^N)) \right). \quad (30)$$

By now the reader is probably thoroughly confused by all the algebra. In order to see what we have done, it is instructive to look at the characteristic thermodynamic function associated with the successive partition functions. We started with a grand-canonical ensemble. The link with thermodynamics is given by:

$$\beta PV = \ln \Xi(V, T, \{\mu_i\}). \quad (31)$$

The transformation to \mathcal{Y} (eqn (25)), corresponds to a change to the thermodynamic variable

$$\beta(PV - \mu_1 N) = \ln \mathcal{Y}(V, T, N, \{\mu_i | i \neq 1\}). \quad (32)$$

Finally, we change to constant pressure, which means that the characteristic thermodynamic function becomes:

$$\beta\mu_1 N = -\ln \mathcal{Y}'(P, T, \{\ln(f_i/f_1) | i \neq 1\}) . \quad (33)$$

Rather than use $\ln(f_i/f_1)$ as the independent variable, it is more convenient to follow Ref. [25] and use the *fugacity fraction* ξ_i , defined as

$$\xi_i \equiv \frac{f_i}{\sum_{j=1}^n f_j} . \quad (34)$$

The advantage is that, while $\ln(f_i/f_1)$ varies between $-\infty$ and $+\infty$ as we go from pure 1 to pure i , ξ_i varies between 0 and 1. Clearly,

$$\beta\mu_1 N = -\ln \left(\beta P \int dV \exp(-\beta PV) \frac{(Vq_1)^N}{N!} \times \sum_{\text{identities } i=1}^n \prod \left(\frac{\xi_i}{\xi_1} \right)^{N_i} \int ds^N \exp(-\beta \mathcal{V}(s^N)) \right) . \quad (35)$$

How does μ_1 , the chemical potential of the reference species, change with the fugacity fractions of the other species? To see this we consider the derivative of eqn (35) with respect to ξ_i :

$$\left(\frac{\partial \beta\mu_1 N}{\partial \xi_i} \right)_{N, P, T, \{\xi_j | j \neq i\}} = - \left\langle \frac{N_i}{\xi_i} + \frac{N - N_1}{\xi_1} \right\rangle , \quad (36)$$

where we have used the fact that $d\xi_1 = -\sum_{j \neq 1} d\xi_j$. Equation (36) tells us how we can measure μ_1 changes as we change the chemical potential difference between species 1 and the other species. Let us consider the application to phase coexistence in a binary mixture. In that case we only vary ξ_2 . First we measure the excess chemical potential in phase I consisting of pure 1 and in phase II consisting of pure species 2 (for instance by thermodynamic integration). Next we compute the change in μ_1 in phase I as we increase ξ_2 from 0 and the corresponding change in μ_1 in phase II as we lower ξ_2 from 1. The point where $\mu_1^{(I)}(\xi_2) = \mu_1^{(II)}(\xi_2)$ is the coexistence point, because at that point $f_1^{(I)} = f_1^{(II)}$ and $f_2^{(I)} = f_2^{(II)}$. Note that we have not specified the nature of phase I and II. They could be liquid, solid or liquid crystalline.

The only practical problem that remains is the Monte Carlo sampling of $-\langle N_i/\xi_i + (N - N_1)/\xi_1 \rangle$ (eqn (36)). Note that the N_i are the dependent variables. The ξ_i are imposed during a given simulation. In addition to the usual particle moves and volume changes we must now also consider a move where a particle changes identity. To this end, we select one of the N particles at random and with equal probability assign it one of the n possible identities. The probability of accepting such a trial move is

$$P_{\text{acc}} = \min \left(1, \frac{\xi_{i'}}{\xi_i} \exp(-\beta \Delta \mathcal{V}(s^N)) \right) , \quad (37)$$

where $\Delta\mathcal{V}(s^N)$ denotes the change in potential energy of the system if we change the identity of a randomly selected particle from i into i' .

I conclude this discussion of the semi-grand ensemble with three comments. First of all, SGCMC is very well suited to study phase equilibria in multicomponent systems that are also in chemical equilibrium. The reason is that every chemical equilibrium simply imposes a relation between the fugacities of the reacting species. Hence, the only effect of a chemical equilibrium is that the number of *independent* fugacities is reduced by one. In practice, the method cannot be used to study *any* chemical equilibrium: the total number of molecules should be conserved in the reaction.

Secondly, the SGCMC scheme can be used to simulate phase equilibria in truly polydisperse systems, including polydisperse solids. And finally, it can be quite advantageous to combine the SGCMC method with the Gibbs ensemble method for mixtures. In that case the fugacity ratios in both simulation boxes are kept the same. In other words, we allow particles in either box to change identity while remaining in the same box. But in addition we allow trial moves where we attempt to move a particle of the reference species 1 from one box to the other. Now the selection of the particle to be swapped goes as follows: first select box I or box II with equal probability. Next, select any molecule of type 1 in the selected box and try to insert it in the other box (see Chapter 6). (The implementation that I suggest here is slightly different from the one advocated in Ref. [25] and closer to the approach of Stapleton *et al.* [30]). The natural choice for the reference species 1 is clearly that species that can be swapped most efficiently, e.g. the smallest molecule in the system.

3. Virtual moves

In the previous sections we have been discussing several Monte Carlo schemes that all aimed to sample equilibrium averages in one ensemble or another. The procedure by which this was achieved was by some kind of random walk through configuration space, where the probability of acceptance of a trial move from one configuration to another was determined by the ratio of the probabilities of these configurations. In the present section we discuss sampling schemes in which certain trial moves are attempted *but never accepted!*

3.1. PARTICLE INSERTION METHOD

As an example of a Monte Carlo technique that employ virtual moves, we consider the so-called 'particle-insertion' method (often referred to as the Widom method [31]) to measure the chemical potential μ of a species in a pure fluid or in a mixture. The statistical mechanics that is the basis for this

method is quite simple. Consider the definition of the chemical potential μ_α of a species α . From thermodynamics we know that μ is defined as:

$$\mu = \left(\frac{\partial G}{\partial N} \right)_{PT} = \left(\frac{\partial F}{\partial N} \right)_{VT} = -T \left(\frac{\partial S}{\partial N} \right)_{VE}, \quad (38)$$

where G , F and S are the Gibbs free energy, the Helmholtz free energy and the entropy, respectively. Here, and in the next few paragraphs, we focus on a one-component system, and hence we drop the subscript α . If we express the Helmholtz free energy of an N -particle system in terms of the partition function Q_N (eqn (8)), then it is obvious from eqn (38) that, for sufficiently large N the chemical potential is given by: $\mu = -k_B T \ln(Q_{N+1}/Q_N)$. If we use the explicit form (eqn (8)) for Q_N , we find:

$$\begin{aligned} \mu &= -k_B T \ln(Q_{N+1}/Q_N) \\ &= -k_B T \ln \left(\frac{V}{\Lambda^3(N+1)} \right) - k_B T \ln \left(\frac{\int ds^{N+1} \exp(-\beta \mathcal{V}(s^{N+1}))}{\int ds^N \exp(-\beta \mathcal{V}(s^N))} \right) \\ &= \mu_{\text{id}}(V) + \mu_{\text{ex}}. \end{aligned} \quad (39)$$

In the last line of eqn (39), we have indicated the separation into the ideal-gas contribution to the chemical potential, and the excess part. As $\mu_{\text{id}}(V)$ can be evaluated analytically, we focus on μ_{ex} . We now separate the potential energy of the $(N+1)$ -particle system into the potential energy function of the N -particle system, $\mathcal{V}(s^N)$, and the interaction energy of the $(N+1)^{\text{th}}$ particle with the rest: $\Delta \mathcal{V} \equiv \mathcal{V}(s^{N+1}) - \mathcal{V}(s^N)$. Using this separation, we can write μ_{ex} as:

$$\mu_{\text{ex}} = -k_B T \ln \left\langle \int ds_{N+1} \exp(-\beta \Delta \mathcal{V}) \right\rangle_N, \quad (40)$$

where $\langle \dots \rangle_N$ denotes canonical ensemble averaging over the configuration space of the N -particle system. The important point to note is that eqn (40) expresses μ_{ex} as an ensemble average that can be sampled by the conventional Metropolis scheme. There is only one aspect of this equation that makes it different from the averages that we considered before, namely the fact that we compute the average of an *integral* over the position of particle $N+1$. This last integral can be sampled by brute-force (unweighted) Monte Carlo sampling. In practice the procedure is as follows: we carry out a perfectly normal constant NVT Monte Carlo simulation on the system of N particles. At frequent intervals during this simulation (for instance, after every MC trial move) we randomly generate a coordinate s_{N+1} , uniformly over the unit cube. With this value of s_{N+1} , we then compute $\exp(-\beta \Delta \mathcal{V})$. By averaging the latter quantity over all generated trial positions, we obtain the average that appears in eqn (40). So, in effect, we are computing the average of the Boltzmann factor associated with the random insertion

of an additional particle in an N -particle system, *but we never accept any such trial insertions*, because then we would no longer be sampling the average needed in eqn (40). The Widom method is a very powerful method to compute the chemical potential of (not too dense) atomic and simple molecular liquids. Its main advantage is its great simplicity, and the fact that it can be added on to an existing constant- NVT MC program, without any modifications to the original sampling scheme: we are simply computing one more thermal average. There is something else about the Widom method that makes it appealing, but that has nothing to do with computational efficiency: it really provides an insight into the *meaning* of the chemical potential, an insight that is often hard to extract from most text-books on statistical thermodynamics.

Extensive numerical studies of the excess chemical potential of finite periodic systems, have shown that this quantity is rather strongly dependent on the size of the system [32, 33]. Computer simulations are typically carried out for periodic systems in which the fundamental cell contains of order 10^2 to 10^3 particles, and the correction needed to give the infinite-system result can therefore be large. Of course, it is possible to estimate the finite-size correction empirically by carrying out simulations for different values of N (the number of particles), but this is very time consuming. It would clearly be much more convenient if the finite-size correction could be estimated directly, since it would then be possible to estimate the chemical potential in the thermodynamic limit on the basis of results obtained from simulations of relatively small systems. In fact, Siepmann, McDonald and Frenkel [34] have recently derived an expression for the leading ($\mathcal{O}(1/N)$) system-size dependence of the excess chemical potential:

$$\Delta\mu_{\text{ex}}(N) = \frac{1}{2N} \left(\frac{\partial P}{\partial \rho} \right) \left[1 - k_{\text{B}}T \left(\frac{\partial \rho}{\partial P} \right) - \rho k_{\text{B}}T \frac{(\partial^2 P / \partial \rho^2)}{(\partial P / \partial \rho)^2} \right]. \quad (41)$$

As shown in Ref. [34], this expression agrees with the exact result for hard rods in one dimension, and is in excellent agreement with numerical results for hard-disks in two dimensions.

The particle insertion method can be modified to measure the *difference* in chemical potential between two species α and β in a mixture. In this case a trial move consists of an attempt to transform a particle of species α into species β (without, of course, ever accepting such trial moves). For more details, the reader is referred to Refs. [35, 28]. Let us finally mention that the particle insertion and swapping techniques are not limited to the measurement of chemical potentials. In fact, a wide class of partial molar quantities (such as, for example the partial molar enthalpy h_α or the partial molar volume v_α) can be measured in this way. For details, see Refs. [28, 36].

The particle insertion method fails under the same conditions where grand canonical Monte Carlo simulations become inefficient, and for the same reason: namely that the probability of 'accepting' a trial insertion becomes very small. One consequence is that the 'brute-force' particle insertion method is less suited for molecular than for atomic systems. Recently, however, Siepmann [37] has developed a scheme to measure the chemical potential of flexible molecules with a finite number of conformations. A similar approach, in the context of a lattice model, had been proposed earlier by Harris and Rice [38]. The next section discusses the basic idea behind these schemes.

3.2. CHEMICAL POTENTIAL OF MACRO-MOLECULES WITH DISCRETE CONFORMATIONS

In order to understand the methods that have been devised to calculate the chemical potential of chain molecules, it is instructive to first consider how we would compute μ_{ex} of a chain molecule with the Widom technique. To this end, I introduce the following notation: the position of the first segment of the chain molecule is denoted by \mathbf{q} and the conformation of the molecule is described by Γ . The configurational part of the partition function of a system of chain molecules can be written as

$$Q_{\text{chain}}(N, V, T) = \frac{1}{N!} \int d\mathbf{q}^N \sum_{\Gamma_1, \dots, \Gamma_n} \exp(-\beta\mathcal{V}(\mathbf{q}^N, \Gamma^N)). \quad (42)$$

The analogy with the previous sections suggests that the excess chemical potential of a chain molecule is obtained by considering the ratio

$$Q(N+1, V, T) / [Q(N, V, T) Q_{\text{non-interacting}}(1, V, T)],$$

where $Q(N+1, V, T)$ is the (configurational part of) the partition function of a system of $N+1$ interacting chain molecules and $Q(N, V, T) \times Q_{\text{non-interacting}}(1, V, T)$ is the partition function for a system consisting of N interacting chains and one chain that does not interact with the others. The latter chain plays the role of the ideal gas molecule in the previous sections. Note, however, that although this molecule does not interact with any of the other molecules it *does* interact with itself, both through bonded and through non-bonded interactions. Unfortunately, this is not a particularly useful reference state, as we do not, in general, know the partition function of an isolated self-avoiding chain.

We therefore use another reference state, namely that of the isolated non-selfavoiding chain. To be specific, let us consider the case of a molecule that consists of ℓ segments. Starting from segment 1, we can add segment 2 in b_2 equivalent directions, and so on. Clearly, the total number of non-selfavoiding conformations is $\Omega_{\text{id}} = \prod_{i=1}^{\ell} b_i$. For convenience, I have assumed

that for a given i , all b_i directions are equally likely (i.e. I ignore *gauche-trans* potential energy differences and I even allow the ideal chain to fold back on itself). These limitations are not essential but they simplify the notation. Finally, I assume that all b_i are the same. Hence, for the simple model that we consider, $\Omega_{id} = b^l$. If we use such an ideal chain as our reference system, the expression for the excess chemical potential becomes

$$\begin{aligned} \beta\mu_{ex} &= -k_B T \ln \left(\frac{Q_{\text{chain}}(N+1, V, T)}{Q(N, V, T)Q_{\text{ideal}}(1, V, T)} \right) \\ &= -k_B T \ln \left\langle \exp[-\beta\Delta\mathcal{V}(\mathbf{q}^N, \Gamma^N; \mathbf{q}_{N+1}, \Gamma_{N+1})] \right\rangle, \end{aligned} \quad (43)$$

where $\Delta\mathcal{V}$ denotes the interaction of the test chain with the N chains that are already present in the system *and with itself*, while $\langle \dots \rangle$ indicates averaging over all starting positions and all ideal-chain conformations of a randomly inserted chain.

The problem with the Widom approach to eqn (43) is that almost all randomly inserted ideal chain conformations will overlap either with particles already present in the system, or internally. The most important contributions to μ_{ex} will come from the extremely rare cases where the trial chain happens to be in just the right conformation to fit into the available space in the fluid. Clearly, it would be desirable if we could restrict our sampling to those conformations that satisfy this condition. If we do that, we introduce a bias in our computation of the insertion probability and we must somehow correct for that bias. The scheme developed by Harris and Rice and, independently, by Siepmann shows how that can be done.

In order to explain this scheme, I will first describe how the method works, and then show that it does indeed lead to the desired answer. The approach used in Refs. [37, 38] consists of two steps: in the first step a chain conformation is generated in such a way that 'acceptable' conformations are created with a high probability. The next step corrects for this bias by multiplying with a weight factor. A scheme that generates 'acceptable' chain conformations with a high probability was developed by Rosenbluth and Rosenbluth in the early fifties [39]. In the Rosenbluth scheme, a conformation of a chain molecule is constructed segment-by-segment. For every segment, we have a choice of b possible directions. In the Rosenbluth scheme, this choice is not random but favors the direction with the largest Boltzmann factor. To be specific, the probability (P) to generate a polymer with a conformation Γ using the Rosenbluth algorithm is given by

$$P_{\Gamma} = \prod_{i=1}^l \frac{\exp[-\beta v^{(i)}(\Gamma_i)]}{Z_i}, \quad (44)$$

where $v^{(i)}(\Gamma_i)$ denotes the energy of segment i of the chain with conformation Γ (note that this energy excludes the contributions of segments $i+1$ to l , so

the total energy of the chain is given by: $\mathcal{V}_\Gamma = \sum_{i=1}^{\ell} v^{(i)}(\Gamma_i)$. Z_i in eqn (44) is shorthand for

$$Z_i \equiv \sum_{j=1}^b \exp \left[-\beta v^{(i)}(\Gamma_j) \right] .$$

where j enumerates all possible orientations from which the i^{th} segment of the chain can be chosen and $v^{(i)}(\Gamma_j)$ denotes the potential energy of the i^{th} segment in orientation j . An important property of the probability given by eqn (44) is that it is normalized, i.e.

$$\sum_{\Gamma} P_{\Gamma} = 1 .$$

The Rosenbluth weight factor that corrects for the bias in the selection of conformation Γ is given by

$$W_{\Gamma} \equiv \prod_{i=1}^{\ell} \frac{Z_i}{b} . \quad (45)$$

Now let us assume that we use the Rosenbluth scheme to generate a large number of chain conformations while keeping the coordinates of all other particles in the system fixed. For this set of conformations, we compute the average of the Rosenbluth weight factor W , \overline{W} . If we also perform an ensemble average over all coordinates and conformations of the N particles in the system, we obtain

$$\langle W \rangle = \left\langle \sum_{\Gamma} P_{\Gamma}(\mathbf{q}^N, \Gamma^N) W_{\Gamma}(\mathbf{q}^N, \Gamma^N) \right\rangle \quad (46)$$

where the angular brackets denote the ensemble average over all configurations $\{\mathbf{q}^N\}$ of the 'solvent'. Note that the test polymer does not form part of the N -particle system. Therefore the probability to find the remaining particles in a configuration $\{\mathbf{q}^N, \Gamma^N\}$ does not depend on the conformation Γ of the polymer.

In order to simplify the expression for the average in eqn (46), we first consider the average of the Rosenbluth factor for a given configuration $\{\mathbf{q}^N, \Gamma^N\}$ of the solvent

$$\overline{W}(\{\mathbf{q}^N, \Gamma^N\}) = \sum_{\Gamma} P_{\Gamma}(\mathbf{q}^N) W_{\Gamma}(\{\mathbf{q}^N, \Gamma^N\}) . \quad (47)$$

Substitution of eqns (44) and (45) yields

$$\overline{W} = \sum_{\Gamma} \left[\prod_{i=1}^{\ell} \frac{\exp \left[-\beta v^{(i)}(\Gamma_i) \right]}{Z_i} \right] \left[\prod_{i=1}^{\ell} \frac{Z_i}{b} \right]$$

$$\begin{aligned}
&= \sum_{\Gamma} \prod_{i=1}^{\ell} \frac{1}{b} \exp[-\beta v^{(i)}(\Gamma_i)] \\
&= \sum_{\Gamma} \frac{1}{b^{\ell}} \exp[-\beta \mathcal{V}_{\Gamma}] ,
\end{aligned} \tag{48}$$

where we have dropped all explicit reference to the solvent coordinates $\{\mathbf{q}^N, \Gamma^N\}$. Note that eqn (48) can be interpreted as an average over all *ideal* chain conformations of the Boltzmann factor $\exp[-\beta \mathcal{V}_{\Gamma}]$. If we now substitute eqn (48) in eqn (46) we obtain

$$\langle W \rangle = \frac{\sum_{\Gamma} \langle \exp[-\beta \Delta \mathcal{V}(\mathbf{q}^N, \Gamma^N; \mathbf{q}_{N+1}, \Gamma_{N+1})] \rangle}{\sum_{\Gamma}} . \tag{49}$$

If we compare eqn (49) with eqn (43), we see that the ensemble average of the Rosenbluth factor is directly related to the excess chemical potential of the chain molecule.

$$\beta \mu_{\text{ex}} = -k_{\text{B}} T \ln \langle W \rangle . \tag{50}$$

This completes our proof that a measurement of the average Rosenbluth factor of a trial chain can indeed be used to estimate the excess chemical potential of a polymer in a dense fluid.

Siepmann has compared this method with a brute-force application of the Widom method for a simple two-dimensional model of chain molecules on a lattice [37], and found the 'Rosenbluth' scheme to be very much superior, in particular at high densities. I should stress that the above method to measure the chemical potential is in no way limited to chain molecules on a lattice. What is essential is that the number of possible directions for each segment (b) relative to the previous one is finite. The special case of chain molecules on a lattice was studied by Mooij and Frenkel [40, 41] who found that, in that case, the calculation of the chemical potential can even be made 2-3 orders of magnitude more efficient than the original scheme of Refs. [37, 38].

3.3. EXTENSION TO CONTINUOUSLY DEFORMABLE MOLECULES

The numerical computation of the (excess) chemical potential of a flexible chain (with or without elastic forces that counteract bending), is rather different from the corresponding calculation for a chain molecule that has a large but fixed number of undeformable conformations.

Below, I shall consider the case of a flexible molecule *with* internal energy. The fully flexible case then follows trivially. Consider a 'worm-like' chain of ℓ linear segments. The potential energy of a given conformation has two contributions:

- (i) The internal potential energy \mathcal{V}_{int} is equal to the sum of the contributions of the individual joints. A joint between segments i and $i + 1$ (say) has a potential energy $v(\theta_i)$ that depends on the angle θ_i between the successive segments. For instance, $v(\theta_i)$ could be of the form $v(\theta) = \alpha\theta^2$. For realistic models for poly-atomic molecules, \mathcal{V}_{int} would account for all local internal potential energy changes due to bending and torsion.
- (ii) The 'external' potential energy \mathcal{V}_{ext} . This energy accounts for all interactions with other molecules and for the non-bonded intra-molecular interactions. In addition, interactions with any external field that may be present are also included in \mathcal{V}_{ext} .

In what follows I shall refer to the chain in the absence of the 'external' interactions as the *ideal* chain. Clearly, the conformational partition function of the ideal chain is equal to

$$Z_{\text{id}} = c \int \cdots \int d\Gamma_1 \cdots d\Gamma_\ell \prod_{i=1}^{\ell} \exp(-\beta v_{\text{id}}(\theta_i)) \quad (51)$$

where c is a numerical constant. Our aim is to compute the effect of the external interactions on the conformational partition function. Hence, we wish to evaluate Z/Z_{id} , where Z denotes the partition function of the interacting chain. The excess chemical potential of the interacting chain is given by

$$\mu_{\text{ex}} = -k_{\text{B}}T \ln(Z/Z_{\text{id}}).$$

Below, I shall sketch an efficient procedure to evaluate this ratio by Monte Carlo sampling. However, before considering the 'smart' approach, let us briefly review two not-so-smart methods.

The simplest (and most stupid) way to compute the excess chemical potential of the interacting chain is to generate a very large number of completely random conformations of the freely jointed chain. For every conformation we compute both $\exp(-\beta\mathcal{V}_{\text{int}})$ and $\exp(-\beta[\mathcal{V}_{\text{int}} + \mathcal{V}_{\text{ext}}])$. The average of the former quantity is proportional to Z_{id} , while the average of the latter Boltzmann factor is proportional to Z . The ratio of these two averages should therefore yield Z/Z_{id} .

The problem with this approach is that the overwhelming majority of randomly generated conformations correspond to worm-like chains with a very high internal energy (and therefore a very small Boltzmann weight). Hence the statistical accuracy of this sampling scheme will be very poor.

The second scheme is designed to alleviate this problem. Rather than generating conformations of a freely jointed chain, we now sample the internal angles in the chain in such a way that the probability of finding a given angle θ_i is given by the Boltzmann weight

$$P(\theta_i) = \frac{\exp(-\beta v(\theta_i))}{\int d\Gamma_i \exp(-\beta v(\theta_i))}.$$

Here, and in what follows, I use the symbol Γ_i to denote the unit vector that specifies the orientation of the i^{th} segment of the chain molecule. For every conformation thus generated, we compute the Boltzmann factor $\exp(-\beta\mathcal{V}_{\text{ext}})$. The average of this Boltzmann weight is then equal to

$$\begin{aligned} \langle \exp(-\beta\mathcal{V}_{\text{ext}}) \rangle &= \frac{\int \prod d\Gamma \exp(-\beta[\mathcal{V}_{\text{id}} + \mathcal{V}_{\text{ext}}])}{\int \prod d\Gamma \exp(-\beta\mathcal{V}_{\text{id}})} \\ &= Z/Z_{\text{id}}. \end{aligned} \quad (52)$$

This approach is obviously superior to the first scheme. However, in many practical situations it will still yield poor statistics, because most ideal chain conformations will not correspond to energetically favourable situations for the interacting chain. Hence the Boltzmann weights will, again, be small for most conformations and the statistical accuracy will not be very good.

The problem with both schemes described above, is that neither allows us to focus on those conformations that should contribute most to Z , namely the ones for which both the internal and external potential energy are not much larger than a few $k_{\text{B}}T$ per degree of freedom. It would clearly be desirable to *bias* the sampling towards such favourable conformations. Let us now consider one possible solution to the sampling problem sketched above. The procedure is similar to the scheme to compute the excess chemical potential of a chain molecule with many fixed conformations (see section 3.2). Yet, there is an important difference precisely because the number of conformations is now, in principle, infinite. We can never hope to sample over *all* possible orientations of a new segment as we grow a chain. Hence, we generate a random sample of possible segment directions and use these in a modified Rosenbluth scheme. Below I shall first sketch how the sampling scheme works. Next I shall show that it does indeed yield the desired average.

To compute μ_{ex} , we apply the following 'recipe' to construct a conformation of a chain of ℓ segments. The construction of chain conformations proceeds segment by segment. Let us consider the addition of one such segment. To be specific, let us assume that we have already grown i segments, and that we are trying to add segment $i + 1$. This is done as follows:

- (i) Generate a fixed number (say b) trial segments. The orientations of the trial segments are distributed according to the Boltzmann weight associated with the internal energy $v(\theta)$. We denote the different trial segment by indices $1, 2, \dots, b$.
- (ii) For all b trial segments, we compute the 'external' Boltzmann factor $\exp(-\beta v_{\text{ext}}(j))$.
- (iii) Select one of the trial segments, say j , with a probability

$$P_j = \exp(-\beta v_{\text{ext}}(j)) / Z_i,$$

where we have defined

$$Z_i \equiv \sum_{j'=1}^b \exp(-\beta v_{\text{ext}}(j')) .$$

(iv) Add this segment as segment $i + 1$ to the chain and store the corresponding partial Rosenbluth weight $w_i = Z_i/b$.

The desired ratio Z/Z_{id} is then equal to the average value (over many trial chains) of the product of the partial Rosenbluth weights:

$$Z/Z_{\text{id}} = \left\langle \prod_{i=1}^{\ell} w_i \right\rangle . \quad (53)$$

The advantage of this scheme is that step (iii) biases the sampling towards energetically favourable conformations. However, it still remains to be shown that eqn (53) is, in fact, correct. This, I shall now proceed to do. To this end let us consider the probability with which we generate a given chain conformation. This probability is the product of a number of factors. Let us first consider these factors for one segment, and then later extend the result to the complete chain. The probability to generate a given set of b trial segments with orientations Γ_1 through Γ_b is

$$P_{\text{id}}(\Gamma_1) \cdot P_{\text{id}}(\Gamma_2) \dots P_{\text{id}}(\Gamma_b) d\Gamma_1 \dots d\Gamma_b .$$

The probability of selecting any one of these trial segments, say segment j , is

$$\exp(-\beta v_{\text{ext}}(j))/Z_i .$$

We wish to compute the average of a quantity, say w , over all possible sets of trial segments and all possible choices of the segment. To this end, we must sum over all j and integrate over all orientations $\prod_{j'=1}^b d\Gamma_{j'}$ (i.e. we average over the normalized probability distribution for the orientation of segment $i + 1$):

$$\langle w \rangle = \int \left[\prod_{j'=1}^b d\Gamma_{j'} P_{\text{id}}(\Gamma_{j'}) \right] \sum_{j''=1}^b \frac{\exp(-\beta v_{\text{ext}}(j''))}{Z_i} w(1, 2, \dots, b) . \quad (54)$$

Now we make use of the fact that $w_i(1, 2, \dots, b)$ is equal to Z_i/b (see step (iv) of the 'recipe' above). Inserting this expression in eqn (54), we obtain:

$$\langle w \rangle = \int \left[\prod_{j'=1}^b d\Gamma_{j'} P_{\text{id}}(\Gamma_{j'}) \right] \sum_{j''=1}^b \frac{\exp(-\beta v_{\text{ext}}(j''))}{b} . \quad (55)$$

Now, as the labeling of the trial segments is arbitrary, all b terms in the sum in eqn (55) yield the same contribution, and eqn (55) simplifies to

$$\langle w \rangle = \int d\Gamma P_{id}(\Gamma) \exp(-\beta v_{ext}(\Gamma)) \quad (56)$$

$$= \frac{\int d\Gamma \exp(-\beta[v_{id}(\Gamma) + v_{ext}(\Gamma)])}{\int d\Gamma \exp(-\beta v_{id}(\Gamma))} \quad (57)$$

$$= \frac{Z^{(1)}}{Z_{id}^{(1)}}, \quad (58)$$

which is indeed the desired result, but for the fact that the expression in eqn (58) refers to 1 segment (as indicated by the superscript in $Z^{(1)}$). The extension to a chain of ℓ segments is straightforward, albeit that the intermediate expressions become a little unwieldy.

The above scheme to compute the chemical potential of continuously deformable molecules has recently been tested by Frenkel and Smit [42] and by de Pablo *et al.* [43]. The results of Refs. [42, 43] indicate that the method can be applied directly to chain molecules of 20–30 segments. However, there is some evidence that, for longer chain lengths, the scheme runs in to sampling problems. In that case, there is an alternative scheme to compute the chemical potential of chain molecules, due to Kumar *et al.* [44]. This scheme is more robust, but also more expensive than the method described above, because it computes the *incremental* change in the chemical potential of a chain molecule upon sequential addition of segments. In principle, the computation of the chemical potential of a chain molecule consisting of ℓ segments, would require $\ell + 1$ simulations in the Kumar scheme. However, Kumar *et al.* argue that, in practice, reasonable estimates of the chemical potential may be obtained using far fewer runs. The idea behind the latter approach is that ‘incremental’ excess chemical potential can be related to the excess chemical potential of the polymer as a whole, when the latter increases linearly with the number of segments.

This assumption has been tested for an isolated chain by Smit *et al.* [45], using the present, rigorous scheme. The work of Ref. [45] suggests that, at least in this case, the assumption of a linear length dependence of the excess chemical potential of a chain molecule is, in fact, an oversimplification. It seems likely that a hybrid of the techniques of Refs. [42] and [44] will be preferable to either technique.

3.4. OVERLAPPING DISTRIBUTION METHOD

The reader may wonder why, in the previous sections, we have only been discussing a trial move that attempts to *add* a particle to the system, and not the reverse move. After all, the excess chemical potential can also be

written as:

$$\begin{aligned}\mu &= k_{\text{B}}T \ln(Q_N/Q_{N+1}) \\ &= \mu_{\text{id}} + k_{\text{B}}T \ln \langle \exp(+\beta\Delta\mathcal{V}) \rangle_{N+1},\end{aligned}\quad (59)$$

where $\Delta\mathcal{V}$ denotes the interaction energy of particle $N+1$ with the remaining N particles. It would seem that eqn (59) can be sampled by straightforward Metropolis Monte Carlo. In general, however, this is not true. The reason is that the function $\exp(\beta\Delta\mathcal{V})$ is, in principle not bounded. It can become arbitrarily large, as $\Delta\mathcal{V}$ grows. (Incidentally, this is not true for $\exp(-\beta\Delta\mathcal{V})$, because one of the conditions that a system must satisfy in order to be describable by classical statistical mechanics is that its potential energy function has a lower bound). The problem with eqn (59) is that very large values of the integrand coincide with very small values ($\mathcal{O}(\exp(-\beta\Delta\mathcal{V}))$) of the Boltzmann factor (that determines how often a configuration is sampled during a Monte Carlo run). As a consequence, an appreciable contribution to the average in eqn (59) comes from a part of configuration space that is hardly ever, or indeed never, sampled during a run. Hard spheres offer a good illustration. As the potential energy function of non-overlapping hard spheres is always zero, a simple Monte Carlo sampling of eqn (59) for a dense fluid of hard spheres would always yield the nonsensical estimate $\mu_{\text{ex}} = 0$ (whereas, in fact, at freezing, $\mu_{\text{ex}}/k_{\text{B}}T \sim 15$). The correct way to obtain chemical potentials from simulations involving both particle insertions and particle removals has been indicated by Shing and Gubbins [46, 47]. However, I find it convenient to discuss this problem in the context of a more general technique to measure free energy differences, first introduced by Bennett [48], called the overlapping distribution method.

Consider two N -particle systems, labeled 0 and 1 with partition functions Q_0 and Q_1 . For convenience we assume that both systems have the same volume V , but this is not essential. From eqn (8) it follows that the free energy difference $\Delta F = F_1 - F_0$ can be written as:

$$\begin{aligned}\Delta F &= -k_{\text{B}}T \ln(Q_1/Q_0) \\ &= -k_{\text{B}}T \ln \left(\frac{\int ds^N \exp(-\beta\mathcal{V}_1(s^N))}{\int ds^N \exp(-\beta\mathcal{V}_0(s^N))} \right).\end{aligned}\quad (60)$$

Suppose that we are carrying out a (Metropolis) sampling of the configuration space of system 1. For every configuration visited during this sampling of system 1 we can compute the potential energy of system 0 ($\mathcal{V}_0(s^N)$) for the same configuration, and hence the potential energy difference $\Delta\mathcal{V} = \mathcal{V}_1(s^N) - \mathcal{V}_0(s^N)$ (see Fig. 3). We use this information to construct a histogram that measures the probability density for the potential energy difference $\Delta\mathcal{V}$. Let us denote this probability density by $p_1(\Delta\mathcal{V})$. In the canonical

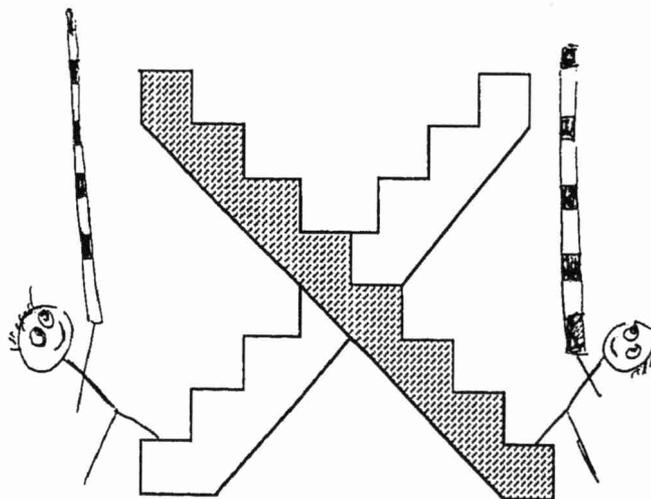


Fig. 3. Example of measurement of histogram of potential energy differences (here represented by height differences) between two systems 0 and 1. Note that the two histograms will, in general, be different.

ensemble, $p_1(\Delta\mathcal{V})$ is given by:

$$p_1(\Delta\mathcal{V}) = \frac{\int ds^N \exp(-\beta\mathcal{V}_1) \delta(\mathcal{V}_1 - \mathcal{V}_0 - \Delta\mathcal{V})}{q_1}, \quad (61)$$

where we have denoted the scaled, configurational part of the partition function by a lower-case q (e.g. $q_1 = \int ds^N \exp(-\beta\mathcal{V}_1(s^N))$). The δ -function in eqn (61) means that we can substitute $\mathcal{V}_0 + \Delta\mathcal{V}$ for \mathcal{V}_1 in the Boltzmann factor, hence:

$$\begin{aligned} p_1(\Delta\mathcal{V}) &= \frac{\int ds^N \exp(-\beta(\mathcal{V}_0 + \Delta\mathcal{V})) \delta(\mathcal{V}_1 - \mathcal{V}_0 - \Delta\mathcal{V})}{q_1} \\ &= \frac{q_0}{q_1} \exp(-\beta\Delta\mathcal{V}) \frac{\int ds^N \exp(-\beta\mathcal{V}_0) \delta(\mathcal{V}_1 - \mathcal{V}_0 - \Delta\mathcal{V})}{q_0} \\ &= \frac{q_0}{q_1} \exp(-\beta\Delta\mathcal{V}) p_0(\Delta\mathcal{V}), \end{aligned} \quad (62)$$

where $p_0(\Delta\mathcal{V})$ is the probability density to find a potential energy difference $\Delta\mathcal{V}$ between systems 1 and 0, while Boltzmann-sampling the available configurations of system 0. As the free energy difference between systems 1 and

0 is simply $\Delta F = -k_{\text{B}}T \ln(q_1/q_0)$, we find from eqn (62):

$$\ln p_1(\Delta\mathcal{V}) = \beta(\Delta F - \Delta\mathcal{V}) + \ln p_0(\Delta\mathcal{V}). \quad (63)$$

In order to obtain ΔF from eqn (63) in practical cases, it is convenient to define two functions f_0 and f_1 by:

$$f_0(\Delta\mathcal{V}) = \ln p_0(\Delta\mathcal{V}) - \frac{\beta\Delta\mathcal{V}}{2}$$

and

$$f_1(\Delta\mathcal{V}) = \ln p_1(\Delta\mathcal{V}) + \frac{\beta\Delta\mathcal{V}}{2}$$

such that

$$f_1(\Delta\mathcal{V}) = f_0(\Delta\mathcal{V}) + \beta\Delta F.$$

Suppose that we have measured f_0 and f_1 in two separate simulations (one sampling system 0, the other system 1). We can then obtain ΔF by fitting the functions f_0 and f_1 to two polynomials in $\Delta\mathcal{V}$ that are identical but for the constant term. The constant offset between the two polynomials yields our estimate for ΔF . Note that, in order to perform such a fit, it is not even necessary that there exists a range of $\Delta\mathcal{V}$ where both f_0 and f_1 can be measured. However, in the absence of such a range of overlap, the statistical accuracy of the method is usually poor.

Now consider the particle insertion/removal problem. Let us assume that system 1 is a system with N interacting particles, while system 0 contains $N - 1$ interacting particles and 1 ideal-gas particle. The difference in free energy between these two systems is obviously equal to μ_{ex} . Applying eqn (63) to this particular case, we find:

$$\Delta\mu_{\text{ex}} = f_1(\Delta\mathcal{V}) - f_0(\Delta\mathcal{V}). \quad (64)$$

Equation (64) is equivalent to the result obtained by Shing and Gubbins. Using the overlapping distribution method it is possible to combine the results of simulations with trial insertions and trial removals to arrive at a more accurate estimate for the chemical potential. A generalization of the Shing-and-Gubbins method to chain molecules, has been proposed by de Pablo *et al.* [43].

3.5. UMBRELLA SAMPLING

In the previous section, we introduced the distribution functions $p_0(\Delta\mathcal{V})$ and $p_1(\Delta\mathcal{V})$ that measure the probability of finding system 0 (1) in an equilibrium configuration s^N where its potential energy difference with system 1 (0) equals $\Delta\mathcal{V}$. At first sight it would seem that knowledge of either p_0 or

p_1 is sufficient to estimate the free energy difference between systems 0 and 1. After all, eqn (63) states that

$$p_1(\Delta\mathcal{V}) = p_0(\Delta\mathcal{V}) \exp(\beta(\Delta F - \Delta\mathcal{V})) .$$

If we integrate over $\Delta\mathcal{V}$ on both sides of this equation, we obtain:

$$\begin{aligned} \int_{-\infty}^{\infty} p_1(\Delta\mathcal{V}) d\Delta\mathcal{V} &= \exp(\beta\Delta F) \int_{-\infty}^{\infty} p_0(\Delta\mathcal{V}) \exp(-\beta\Delta\mathcal{V}) d\Delta\mathcal{V} \\ 1 &= \exp(\beta\Delta F) \langle \exp(-\beta\Delta\mathcal{V}) \rangle_0 , \end{aligned} \quad (65)$$

or

$$\exp(-\beta\Delta F) = \langle \exp(-\beta\Delta\mathcal{V}) \rangle_0$$

Although eqn (65) is very useful to estimate free energy differences between two systems that are not too dissimilar, its applicability is limited. The problem is that, in many cases of practical interest, the largest contributions to the average $\langle \exp(-\beta\Delta\mathcal{V}) \rangle_0$ come from regions of configuration space where $p_0(\Delta\mathcal{V})$ is very small, while $\exp(-\beta\Delta\mathcal{V})$ is very large. As a result, the statistical error in ΔF is large.

One method to achieve a more accurate estimate of ΔF is the 'umbrella sampling' scheme suggested by Torrie and Valleau [49]. The basic idea behind umbrella sampling is that, in order to obtain an accurate estimate of the free energy difference between two systems (0 and 1), one should sample both the part of configuration space that is accessible to system 1 and the part that is accessible to 0. In order to achieve this in a single simulation, one should modify the Markov chain that governs the sampling of configuration space. This is achieved by multiplying the Boltzmann factor of system 0 by a (positive) weighting function w . As a result, the probability of visiting a point \mathbf{q}^N in configuration space is now proportional to $\exp(-\beta\mathcal{V}_0(\mathbf{q}^N))w(\mathbf{q}^N)$. The expression for $\langle \exp(-\beta\Delta\mathcal{V}) \rangle_0$ now becomes:

$$\begin{aligned} \langle \exp(-\beta\Delta\mathcal{V}) \rangle_0 &= \\ &= \frac{\int d\mathbf{q}^N \exp(-\beta\mathcal{V}_0(\mathbf{q}^N))w(\mathbf{q}^N) [\exp(-\beta\Delta\mathcal{V}(\mathbf{q}^N))/w(\mathbf{q}^N)]}{\int d\mathbf{q}^N \exp(-\beta\mathcal{V}_0(\mathbf{q}^N))w(\mathbf{q}^N)w^{-1}(\mathbf{q}^N)} , \end{aligned} \quad (66)$$

or, introducing the notation $\langle \dots \rangle_w$ to denote an average over a probability distribution proportional to $\exp(-\beta\mathcal{V}_0(\mathbf{q}^N))w(\mathbf{q}^N)$:

$$\begin{aligned} \langle \exp(-\beta\Delta\mathcal{V}) \rangle_0 &= \frac{\langle \exp(-\beta\Delta\mathcal{V})/w \rangle_w}{\langle 1/w \rangle_w} \\ &= \langle \exp(-\beta\Delta\mathcal{V})/w \rangle_w \langle w \rangle_0 . \end{aligned} \quad (67)$$

The second line of eqn (67) shows that, in order for both terms on the right hand of this equation to be non-zero, $p_0(\Delta\mathcal{V})w$ should have an appreciable overlap with $p_1(\Delta\mathcal{V})$, while w itself should overlap with $p_0(\Delta\mathcal{V})$. This 'bridging' property of w is responsible for the name 'umbrella' sampling.

Although umbrella sampling is, in principle, a powerful technique, one drawback is that the function w is not known *a priori*. Rather, it must be constructed using the available information about the functions p_0 and p_1 . This requires some skill. A second (though related) problem is that an unfortunate choice of w may result in estimates for ΔF that appear reasonable but are, in fact, wrong. Only by systematic testing is it possible to get a feeling for the statistical errors in an umbrella sampling calculation. At first sight, it might seem advantageous to refine the computation of w in such a way that *all* relevant configurations can be sampled in one run. Surprisingly, this is not the case. Rather, it is advantageous to perform several umbrella sampling runs in (partially overlapping) ‘windows’. To see this, let us define an ‘order parameter’ Φ that is a measure for the ‘location’ of a given configuration between systems 0 and 1. For instance, Φ might be the potential energy, or the (in constant-pressure simulations) the density. Let us denote the average value of Φ in system 1 by Φ_{\max} and the value in 0, by Φ_{\min} . Let us assume that we sample an interval $\Phi_{\max} - \Phi_{\min} \equiv \Delta\Phi$ in n umbrella-sampling simulations. The optimum choice of n is clearly the one that samples the complete Φ -interval in the minimum computing time. In order to estimate this time, let us assume that the system performs a random walk in Φ -space within the window $\Delta\Phi/n$. Associated with the random walk in Φ -space is a ‘diffusion constant’ D_Φ . The characteristic time needed to sample one interval $\Delta\Phi/n$ is

$$\tau_n = \frac{(\Delta\Phi/n)^2}{D_\Phi}.$$

Clearly, the total time to sample all n windows is

$$\tau_{\text{tot}} = n\tau_n = \frac{(\Delta\Phi)^2}{nD_\Phi}.$$

The important point to note is that the computing time *decreases* with increasing n . It would, however, be incorrect to assume that n should be chosen as large as possible. The actual equilibration time of a run in one of the Φ -windows also depends on the rate at which all coordinates ‘orthogonal’ to Φ are sampled. Let us denote this time by τ_\perp . Clearly, once τ_\perp becomes appreciably larger than τ_n , the total computation will scale as $n \times \tau_\perp$. This suggests that the optimum choice of n is the one for which $\tau_n \approx \tau_\perp$. For a more detailed discussion, see Ref. [50]. A recent example of multi-stage umbrella sampling applied to the problem of homogeneous nucleation can be found in Ref. [51].

It has been suggested [52] that there is, in fact, a unique optimum choice for w that minimizes the statistical error in the measurement of the free-energy difference between two systems in a single simulation. In order to

achieve this, the sampling of configuration space should be carried out with an overall probability density W , given by

$$W = \text{const.} \times \sqrt{\frac{\exp(-2\beta\mathcal{V}_0)}{Q_0^2} + \frac{\exp(-2\beta\mathcal{V}_1)}{Q_1^2}}. \quad (68)$$

Although the above equation is, in principle, correct, its practical use may be limited by two factors. One is not serious, the other one usually is. The not-serious problem is that we do not know Q_0/Q_1 *a priori* (if we did, there would be little need for umbrella sampling). In practice, the simulation will be done using a reasonable guess for this ratio. Any error in this guess will worsen the statistics of our simulation somewhat, but it will not systematically affect our estimate for the free-energy difference. The more serious restriction of eqn (68) is that it will only work if systems 0 and 1 are sufficiently close in configuration space that W is not vanishingly small in between. We can interpret $-k_B T \ln W$ as an effective free-energy surface probed in the simulation. If the basins in configuration space belonging to systems 0 and 1 are separated by a region where W is very small, then there is a high 'free energy barrier' between the two systems. Our sampling scheme will then converge poorly. In fact, the equilibration time will diverge as $\exp(\beta F_{\max})$, where $F_{\max} = -k_B T \ln W_{\min}$ is the height of the free-energy barrier separating the two regions in configuration space. As before, this problem may be alleviated by replacing a single run by a series of runs restricted to successive 'windows' that 'tile' the path in configuration space between 0 and 1.

4. Beyond single-particle moves

Thus far, I have described Monte Carlo procedures that were based on the original Metropolis method to sample the configurations of the system under study. In particular, I have assumed that a trial move consisted of an attempted displacement of a single, randomly selected, particle. In fact, there have been many attempts to construct Monte Carlo schemes that utilize 'collective' many-particle moves. Special techniques are required to construct such collective Monte Carlo moves, because a scheme where we would simply attempt a random displacement of all N particles simultaneously, would be *less* rather than *more* efficient than a sequence of single-particle moves, because a very small step-size is required to get a reasonable acceptance. In order to construct larger collective trial moves, several authors have constructed hybrid techniques that are, in a sense, intermediate between Monte Carlo and molecular dynamics (see e.g. Ref. [53] and the book of Allen and Tildesley [22]). The idea behind the earlier 'smart' MC schemes was to construct trial moves that were not completely random, but utilized information about the forces acting on the particles in the system. Recently,

these techniques have been considerably extended [54] in such a way that an MD algorithm can be converted into a ‘collective’ MC algorithm. The difference is that, whereas for MD, we should use time increments that are sufficiently small to ensure energy conservation (see Chapter 2), for MC we should use large time steps, and accept or reject the position thus generated using (say) the Metropolis criterion. It should be stressed that this scheme can only be applied to MD algorithms that are time-reversible or, more precisely, that conserve volume in phase space (otherwise there would be a bias in the sampling for which we should correct). For more details, the reader is referred to Ref. [54]. In the remainder of this section, we shall discuss other ‘collective’ MC schemes of a completely different nature. The reason to include this discussion is that, during the past few years, much progress has been made in the development of Monte-Carlo schemes where a group of particles is displaced in a trial move. In fact, we should distinguish between two rather different schemes:

- (i) The first scheme, or actually family of schemes, is based on the cluster algorithm of Swendsen and Wang [55]. In this method, stochastic rules are used to construct clusters, and these clusters are subsequently moved (or, in the case of spins, flipped) as a single unit.
- (ii) In the second scheme, that I denote by the name ‘Configurational Bias Monte Carlo’, we know in advance which particles we shall attempt to move e.g. some or all particles belonging to a given chain molecule). The scheme provides a set of rules that make it possible to perform large-scale collective moves with a reasonable acceptance probability.

In the following sections, I describe first the Configurational-Bias MC (CBMC) method (section 4.1) and then the “Swendsen-Wang” method (section 4.2). In the latter case, my presentation will be a bit different from the one given in the recent review by Swendsen, Wang and Ferrenberg in the book by Binder [3]. The latter review emphasizes application to lattice systems. In contrast, I try to present the scheme in a way that facilitates extension to off-lattice problems. Both the SW and the CBMC methods are examples of Monte-Carlo schemes in which the *a priori* probability to make a transition from state i to state j (i.e. the matrix α_{ij} in section 1) is *not* symmetric. In this respect, both schemes are related to the earlier ‘smart’ or ‘force-bias’ Monte-Carlo schemes (see Refs. [22, 53]).

4.1. CONFIGURATIONAL-BIAS MONTE CARLO SCHEME

The configurational-bias MC method was originally developed in the context of simulations of chain molecules. The central idea is that the Rosenbluth trial insertion scheme described in sections 3.2 and 3.3, can be used as a starting point for a Monte Carlo scheme to sample equilibrium configurations of systems consisting of chain molecules [38, 56, 57]. At first sight, this

may not appear to be a new result but a very old one. After all, the original Rosenbluth scheme itself was designed as a method to sample polymer conformations. However, the Rosenbluth scheme suffers from the drawback that it generates an unrepresentative sample of all polymer conformations: i.e. the probability to generate a particular conformation Γ using the Rosenbluth scheme, is *not* proportional to the Boltzmann weight of that conformation. The Rosenbluth weight W , discussed in the section 3.2, was introduced to correct for this bias in the computation of thermal averages. However, such a correction procedure only works for relatively short chains. This drawback of the Rosenbluth sampling scheme is, in fact, well known (see, e.g. Ref. [58]). The solution of this problem is to bias the Rosenbluth sampling in such a way that the correct (Boltzmann) distribution of chain conformations is generated in a Monte Carlo sequence. In the configurational bias scheme to be discussed below, the Rosenbluth weight is used to bias the *acceptance* of trial conformations that are generated with the Rosenbluth procedure. As a consequence, all conformations are generated with their *correct* Boltzmann weight. This removes the main drawback of the original Rosenbluth scheme. Moreover, I shall show that the configurational bias scheme can also be extended to continuously deformable molecules, such as worm-like chains.

4.1.1. Sampling discrete conformations. Consider a trial move from a chain-conformation Γ_1 with Rosenbluth weight w_1 to some other conformation Γ_2 with Rosenbluth weight w_2 . Let us denote the probability of finding the chain in conformation Γ_1 by P_1 and the corresponding probability for Γ_2 by P_2 . The number of chains in conformation 1(2), N_1 (N_2) is obviously proportional P_1 (P_2). We wish to perform a Monte-Carlo move such that detailed balance is satisfied, i.e. the rate K_{12} with which conformations of type 1 are transformed into type 2 equals the reverse rate:

$$K_{12} = K_{21}. \quad (69)$$

Let us try to compute K_{12} and K_{21} . K_{12} is equal to N_1 times $P_{\text{gen}}(\Gamma_2)$, the probability of generating chain 2 by Rosenbluth sampling, times the acceptance probability P_{acc} . We assume that P_{acc} is an, as yet unspecified function of the ratio w_2/w_1 , $g(w_2/w_1)$. $P_{\text{gen}}(\Gamma_2)$ is equal to

$$P_{\text{gen}}(\Gamma_2) = \prod_{i=1}^l \frac{\exp(-\beta v^{(i)}(\Gamma_2(i)))}{Z_i}, \quad (70)$$

where, as before, we use the notation $Z_i \equiv \sum_{j=1}^b \exp(-\beta v^{(i)}(\Gamma_2(j)))$. The sum in Z_i runs over all b possible directions of the i^{th} segment of the polymer. With this notation, eqn (69) becomes

$$N_1 P_{\text{gen}}(\Gamma_2) g(w_2/w_1) = N_2 P_{\text{gen}}(\Gamma_1) g(w_1/w_2). \quad (71)$$

Equation (71) can be simplified by using the result of the section 3.2 that the Rosenbluth weight w_2 times the generating probability $P_{\text{gen}}(\Gamma_2)$ is proportional to N_2 (and similarly for 1). Hence multiplying eqn (71) on the left by w_2/w_2 and on the right by w_1/w_1 yields

$$N_1 N_2 \frac{g(w_2/w_1)}{w_2} = N_2 N_1 \frac{g(w_1/w_2)}{w_1} \quad (72)$$

or

$$\frac{g(w_2/w_1)}{g(w_1/w_2)} = \frac{w_2}{w_1}. \quad (73)$$

There are many choices of g that satisfy this condition. One obvious choice is the Metropolis form [1]

$$g(x) = \min(x, 1). \quad (74)$$

In words, the Configurational Bias Monte Carlo (CBMC) scheme works as follows:

- (i) Generate a trial conformation by using the Rosenbluth scheme (i.e. eqn (70) to regrow the entire molecule, or part thereof.
- (ii) Compute the Rosenbluth weights w_{trial} and w_{old} of the trial conformation and of the old conformation.
- (iii) Accept the trial move with a probability $\min(w_{\text{trial}}/w_{\text{old}}, 1)$.

4.1.2. Continuously deformable chain. Next consider Configurational Bias Monte Carlo sampling of worm-like chains or, for that matter, any chain with flexible joints with or without internal bending energy. Again, we first consider the probability to generate a trial configuration Γ_j using the (extended) Rosenbluth scheme. As before, we only consider the expression for one of the ℓ segments, to keep the equations simple:

$$P_{\text{gen}}(j) = d\Gamma_j P_{\text{id}}(\Gamma_j) \left[\prod_{j' \neq j}^b d\Gamma_{j'} P_{\text{id}}(\Gamma_{j'}) \right] \frac{\exp(-\beta v_{\text{ext}}(j))}{\sum_{j'=1}^b \exp(-\beta v_{\text{ext}}(j'))}. \quad (75)$$

We now generate around the original conformation i the same set of additional directions (i.e. those conformations in eqn (75) with $j' \neq j$). Next we compute the ratio of the Rosenbluth weights $w_j^{(t)}/w_i^{(o)}$, where

$$w_j^{(t)} = \frac{\exp(-\beta v_{\text{ext}}^{(t)}(j)) + \sum_{j' \neq j}^b \exp(-\beta v_{\text{ext}}^{(t)}(j'))}{b}$$

and

$$w_i^{(o)} = \frac{\exp(-\beta v_{\text{ext}}^{(o)}(i)) + \sum_{j' \neq j}^b \exp(-\beta v_{\text{ext}}^{(o)}(j'))}{b}$$

The superscript (t) and (o) distinguish the trial conformation from the old conformation. Again, we shall devise a Metropolis criterion to determine the acceptance probability of a trial move from i to j . As before, this acceptance probability is determined by the ratio $x \equiv w_j^{(t)}/w_i^{(o)}$ (actually, for a molecule of l segments, we should compute a product of such factors). Let us assume that $w_j^{(t)} < w_i^{(o)}$. In that case, $P_{\text{acc}}(i \rightarrow j) = x$ while $P_{\text{acc}}(j \rightarrow i) = 1$. Next, let us check whether detailed balance is satisfied. To do so, we write down the explicit expressions for K_{ij} and K_{ji} .

$$K_{ij} = N_i P_{\text{gen}}(j) w_j^{(t)} / w_i^{(o)}$$

and

$$K_{ji} = N_j P_{\text{gen}}(i) 1.$$

Here it should be stressed that there are many different routes to go from i to j and vice versa, depending on the selection of the trial orientations j' . What we show is that, as for *any* specific choice of this set, K_{ij} and K_{ji} are equal ('super-detailed balance'), this equality must also hold for the sum over all possible sets j' . In addition, we use the fact that

$$P_{\text{gen}}(j) w_j \sim N_j \prod_{j' \neq j}^b d\Gamma_{j'} P_{\text{id}}(\Gamma_{j'})$$

and

$$P_{\text{gen}}(i) w_i \sim N_i \prod_{j' \neq i}^b d\Gamma_{j'} P_{\text{id}}(\Gamma_{j'}).$$

For a given set j' , we can therefore write:

$$K_{ij} = N_i P_{\text{gen}}(j) w_j / w_i = \text{constant} \times N_i N_j \prod_{j' \neq j}^b d\Gamma_{j'} P_{\text{id}}(\Gamma_{j'}) \frac{1}{w_i}$$

while

$$K_{ji} = N_j P_{\text{gen}}(i) 1 = \text{constant} \times N_j N_i \prod_{j' \neq i}^b d\Gamma_{j'} P_{\text{id}}(\Gamma_{j'}) \frac{1}{w_i}.$$

Hence, K_{ij} is indeed equal to K_{ji} . This completes the proof that the above Configurational Bias MC scheme satisfies detailed balance.

The CBMC scheme for completely flexible molecules has been applied by Frenkel *et al.* [59] to study conformational changes of chain molecules in solution. The method was found to be particularly efficient in generating large conformational changes of such a molecule. More interestingly, the CBMC scheme makes it possible to perform direct simulations of phase equilibria in polymer mixtures, using the Gibbs-ensemble method [60, 61].

4.1.3. *From polymers back to atoms.* Up to this point, I have introduced the configurational-bias MC scheme as a method to sample conformations of linear chain molecules. However, the method is applicable to branched polymers and ring polymers. In the latter case, the condition that the N^{th} segment is connected to the first, is accounted for by biasing the generation of trial directions. The *a priori* probability to generate a given trial direction Γ_i as the m^{th} segment, is proportional to the number of ideal (i.e. non-(self)-avoiding) conformations that connect this segment to the first segment in $N - m$ steps.

Above, the configurational-bias scheme has been presented exclusively as a method to generate *polymer* conformations. In fact, the method is more general than that. It can be used as a scheme to perform collective rearrangements of *any* set of labeled coordinates. Such a scheme could, for instance, be used to carry out Monte Carlo moves to swap n small particles within a volume ΔV with one large particle that occupies the same (excluded) volume. Similarly, one could envisage a Monte Carlo move in which part of a polymer *plus the surrounding solvent* is regrown. Such an approach may be useful in the study of molecules of biological interest.

4.2. THE SWENDSEN-WANG SAMPLING SCHEME

The central idea behind the Swendsen-Wang (SW) scheme [55] and subsequent extensions and modifications (see Ref. [3]), is to generate trial configurations with a probability that is proportional to the Boltzmann weight of that configuration. As a result, the subsequent trial moves can be accepted with 100% probability. The easiest way to describe the SW scheme is to consider the expression for detailed balance in a Monte Carlo scheme to simulate a discrete spin model. Consider an ‘old’ configuration (labeled by a superscript o) and a ‘new’ configuration, denoted by a superscript n . Detailed balance is satisfied if the following equality holds:

$$\begin{aligned} P_{\text{B}}^o P_{\text{Gen}}^o(\{\text{cluster}\}) P_{\text{sel}}^{\{\text{cl}\}}(\text{flip}) P_{\text{acc}}(o \rightarrow n) = \\ P_{\text{B}}^n P_{\text{Gen}}^n(\{\text{cluster}\}) P_{\text{sel}}^{\{\text{cl}\}}(\text{flip}) P_{\text{acc}}(n \rightarrow o) \end{aligned} \quad (76)$$

where $P_{\text{B}}^{o(n)}$ is the Boltzmann weight of old (new) configuration, $P_{\text{Gen}}(\{\text{cluster}\})$ denotes the probability of breaking the system up into a specified set of clusters, each of which can be subsequently flipped at random, $P_{\text{sel}}^{\{\text{cl}\}}(\text{flip})$ is the probability to select, from a set of clusters $\{1, 2, \dots, p\}$ a given subset (e.g. $\{1, 5, \dots, p - 1\}$) that is to be flipped, and finally $P_{\text{acc}}(o \rightarrow n)$ is the acceptance probability of a given trial move. We can simplify eqn (76) in two ways: first of all, we note that the probability $P_{\text{sel}}^{\{\text{cl}\}}(\text{flip})$ is the same for the forward and reverse moves (i.e. we consider the *same* subset of the *same* set of clusters). Moreover, we wish to impose $P_{\text{acc}} = 1$ for both forward

and reverse moves. This may not always be feasible. However, for most of the cases that I discuss below, this is indeed possible. The detailed balance equation then becomes:

$$P_B^o P_{\text{Gen}}^o(\{\text{cluster}\}) = P_B^n P_{\text{Gen}}^n(\{\text{cluster}\}) \quad (77)$$

or

$$\frac{P_{\text{Gen}}^o(\{\text{cluster}\})}{P_{\text{Gen}}^n(\{\text{cluster}\})} = \frac{P_B^n}{P_B^o} = \exp(-\beta \Delta E), \quad (78)$$

where ΔE is the difference in energy between the new and the old configurations. The trick is then to find a recipe for cluster generation that will satisfy eqn (78). In order to illustrate how this works, I consider two simple examples, namely the Ising model and the xy -model. The extension to many other models is straightforward.

4.2.1. Ising model. For the construction of the Ising SW algorithm, the dimensionality of the model is irrelevant. Consider a given configuration of the spin system, with N_p spin-pairs parallel and N_a spin-pairs anti-parallel. The total energy of that configuration is:

$$E = (N_a - N_p)J$$

where J denotes the strength of the nearest-neighbor interaction. The Boltzmann weight of that configuration is

$$P_B^o = \exp(-\beta J(N_a - N_p))/Z$$

where Z is the partition function of the system. In general, Z is unknown, but that is unimportant. We only use the fact that Z is a constant. Next, we construct clusters by creating bonds between spin pairs according to the following recipe:

- If nearest neighbors are *anti-parallel*, they are not connected.
- If nearest neighbors are *parallel*, they are connected with probability p and disconnected with probability $(1 - p)$.

Here, I have assumed that J is positive. If J is negative (anti-ferromagnetic interaction), parallel spins are not connected, while anti-parallel spins are connected with a probability p .

In the case that we consider, there are N_p parallel spin pairs. The probability that n_c of these are connected and $n_b = N_p - n_c$ are 'broken' is:

$$P_{\text{Gen}}^o(\{\text{cluster}\}) = p^{n_c} (1 - p)^{n_b} .$$

Note that this is the probability of connecting (*c.q.* breaking) a *specified* subset of all links between parallel spins. Once the connected bonds have

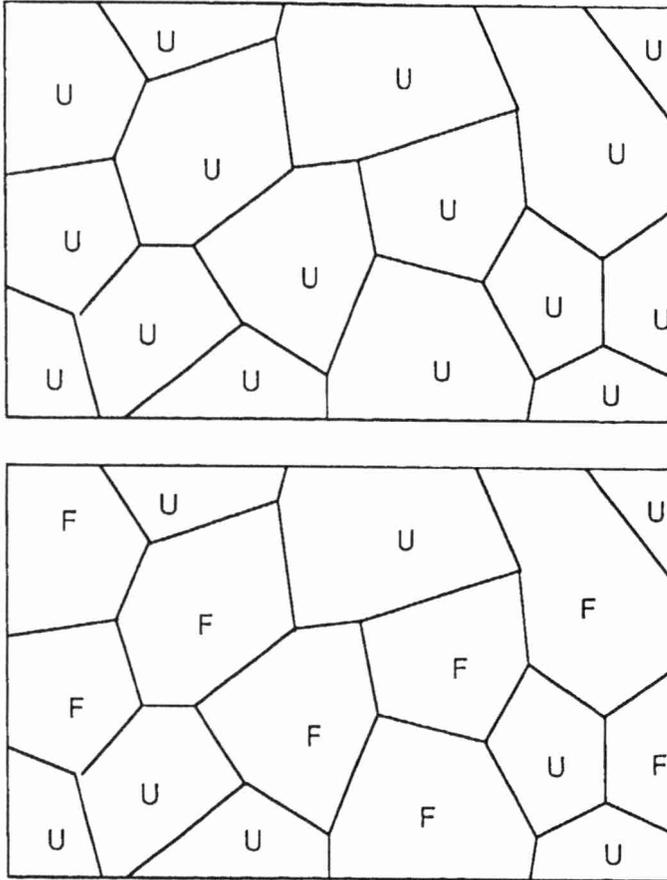


Fig. 4. Swendsen-Wang algorithm. Top figure: the system is divided into clusters. **U** denotes the “unflipped” state of the cluster. In the bottom figure, some clusters have been flipped (**F**). Note, however, that the cluster *structure* is unchanged.

been selected, we can define the clusters in the system. A cluster is a set of spins that are at least singly connected by bonds. We now choose our subset of clusters to flip (see Fig. 4). After the cluster flipping, the number of parallel and anti-parallel spin pairs will have changed, e.g.

$$N_p^{\text{new}} = N_p^{\text{old}} + \Delta$$

and (hence)

$$N_a^{\text{new}} = N_a^{\text{old}} - \Delta.$$

Therefore, the total energy of the system will have changed by an amount $-2J\Delta$:

$$E_{\text{new}} = E_{\text{old}} - 2J\Delta .$$

Let us now consider the probability to make the *reverse* move. In order to do this, we should generate the *same* cluster structure, but now starting from a situation where there are $N_p + \Delta$ parallel spin pairs and $N_a - \Delta$ anti-parallel pairs. As before, the bonds between anti-parallel pairs are assumed to be broken (this is compatible with the same cluster structure). We also know that the new number of connected bonds, n'_c must be equal to n_c , *because the same number of connected bonds is required to generate the same cluster structure*. The difference appears when we consider how many of the bonds between parallel spins in the new configuration should be broken (n'_b). Using

$$N_p^{\text{new}} = n'_c + n'_b = n_c + n'_b = N_p^{\text{old}} + \Delta = n_c + n_b + \Delta$$

we see that

$$n'_b = n_b + \Delta$$

If we insert this in eqn (78), we obtain

$$\frac{P_{\text{Gen}}^o(\{\text{cluster}\})}{P_{\text{Gen}}^n(\{\text{cluster}\})} = \frac{p^{n_c}(1-p)^{n_b}}{p^{n_c}(1-p)^{n_b+\Delta}} = (1-p)^{-\Delta} = \frac{P_B^n}{P_B^o} = \exp(2\beta J\Delta) . (79)$$

In order to satisfy this equation, we must have:

$$1 - p = \exp(-2\beta J)$$

or

$$p = 1 - \exp(-2\beta J) ,$$

which is the Swendsen-Wang rule.

4.2.2. xy-model. In this section, I describe one application to a 'continuous' model, namely the planar Heisenberg model (*xy*-model). In this model, spins on neighboring lattice sites have an interaction

$$v_{ij} = -J\mathbf{u}_i \cdot \mathbf{u}_j$$

where \mathbf{u}_i denotes the unit vector that specifies the direction of a spin on site i , while J measures the strength of the interaction. The first step is to define the cluster moves. The essential requirement is that the relative orientation of particles within one cluster is not changed. One way to achieve this is to reflect the orientation of all particles within the selected cluster with respect

to a *randomly* chosen plane. For convenience, I use coordinates such that this plane is perpendicular to the x -axis. Let us consider what happens to the interaction energy of two neighboring spins if one is reflected, but not the other. Initially, the interaction energy was

$$v_{ij}^{\text{old}} = -J(u_i^x u_j^x + u_i^y u_j^y)$$

If only one spin is reflected (either one), the product $u_i^x u_j^x$ changes sign. Therefore,

$$v_{ij}^{\text{new}} = v_{ij}^{\text{old}} + 2J u_i^x u_j^x \equiv v_{ij}^{\text{old}} + \Delta v_{ij}.$$

Note that Δv_{ij} may be either positive or negative. Now we construct clusters by linking neighboring spins according to the following recipe:

- If $\Delta v_{ij} \leq 0$, i and j are always *disconnected*. We call these spin pairs ‘anti-parallel’, by analogy to the Ising case.
- If $\Delta v_{ij} > 0$, i and j are *connected* with a probability p_{ij} and disconnected with a probability $(1 - p_{ij})$. We call these spin pairs ‘parallel’, by analogy to the Ising case.

Note that the probability to break a bond between ‘parallel’ spins depends on the magnitude of Δv_{ij} .

Let us next consider how we can impose detailed balance. The equivalent of eqn (77) is

$$P_B^o \prod_{ij}^{n_c} p_{ij} \prod_{kl}^{n_b} (1 - p_{kl}) = P_B^n \prod_{ij}^{n'_c} p_{ij} \prod_{kl}^{n'_b} (1 - p_{kl}) \quad (80)$$

where the set of pairs $\{ij\}$ and $\{kl\}$ denote all pairs of neighboring spins that are ‘parallel’. Note that, as before, the set $\{ij\}$ is the *same* for the old and new configuration, because we are considering the same cluster structure. As a consequence, $n_c = n'_c$. However, the number of disconnected parallel spin pairs will, in general differ in the old and new situations. We can write the change in energy between the old and new configurations due to cluster flips as

$$\Delta E = \sum_{ij} \Delta v_{ij}^+ + \sum_{kl} \Delta v_{kl}^-$$

where Δv^+ denotes a positive energy change (i.e. where a pair of spins goes from ‘parallel’ to ‘anti-parallel’) and Δv^- denotes a negative energy change. Note that, for a given pair ij , Δ is either positive or negative (or zero). Hence, the sets $\{ij\}$ and $\{kl\}$ that give non-zero contributions are *non-overlapping*. We can now re-express our condition for detailed balance

(eqn (80)) as:

$$P_B^o \prod_{ij}^{n_c} p_{ij} \prod_{kl}^{n_b} (1 - p_{kl}) = P_B^o \exp \left[-\beta \left(\sum_{\alpha\beta} \Delta v_{\alpha\beta}^+ + \sum_{\gamma\delta} \Delta v_{\gamma\delta}^- \right) \right] \prod_{ij}^{n'_c} p_{ij} \prod_{kl}^{n'_b} (1 - p_{kl}). \quad (81)$$

Clearly, as the same set of bonds are connected in the old and new configuration, we can simplify the above equation

$$\prod_{kl}^{n_b} (1 - p_{kl}) = \exp \left[-\beta \left(\sum_{\alpha\beta} \Delta v_{\alpha\beta}^+ + \sum_{\gamma\delta} \Delta v_{\gamma\delta}^- \right) \right] \prod_{kl}^{n'_b} (1 - p_{kl}). \quad (82)$$

Let us denote the number of bonds between parallel spins that are broken both in the old *and* in the new configuration by N_b . We now write $n_b = N_b + \Delta n_b^o$ and $n'_b = N_b + \Delta n_b^n$. Using this definition, we can rewrite our detailed balance condition as

$$\prod_{ij}^{\Delta n_b^o} (1 - p_{ij}) \exp \left[\beta \left(\sum_{ij} \Delta v_{ij}^+ \right) \right] = \exp \left[-\beta \left(\sum_{kl} \Delta v_{kl}^- \right) \right] \prod_{kl}^{\Delta n_b^n} (1 - p_{kl}). \quad (83)$$

It is important to note that the pairs with a positive (negative) Δv are precisely the ones that are parallel in the old (new) configuration but not in the new (old) one. The above equality can be satisfied trivially if we impose that

$$1 - p_{ij} = \exp(-\beta \Delta v_{ij}^+)$$

or

$$p_{ij} = 1 - \exp(-\beta \Delta v_{ij}^+).$$

This, then, is the 'Swendsen-Wang' recipe for the xy -model. In fact, it will work for a wide class of d -dimensional spin models. It contains the Ising model as a special case.

4.2.3. Translationally disordered systems. Here the problem is more complex. Consider two 'molecules' i and j that either belong to the same cluster, or move independently. We should now distinguish three cases:

- (i) The relative coordinates of i and j are unchanged (they belong to the same cluster). This yields $\Delta v_{ij}=0$.

(ii) i moves, but not j . This yields $\Delta v_{ij} \equiv v_1$.

(iii) j moves, but not i . This yields $\Delta v_{ij} \equiv v_2$.

In general, $v_1 \neq v_2$. There are probably many ways to define clusters in this situation. I suggest the following, scheme. Grow a *single cluster* starting from a randomly chosen site (say i). An atom k in the cluster is linked to another atom (l) with an, as yet unspecified, probability p_{kl} . We terminate the growth of the cluster as soon as we have exhausted all possibilities to form bonds to atoms outside the cluster. We then try to move this cluster over a distance Δ . The trial move is accepted with an (as yet unspecified) probability $P_{\text{acc}}(0 \rightarrow 1)$, where 0 denotes the old position, and 1 the new position. The problem arises when we consider the reverse move: i.e. we attempt to move atom i over a distance $-\Delta$. For any atom j , moving i over a distance $-\Delta$ and keeping j fixed, is equivalent to moving j over $+\Delta$ and keeping i fixed. As a consequence, Δv_{ij} is now equal to v_2 , rather than v_1 . Clearly, whatever we do, the probability of forming a bond in a cluster should not depend on the sign of Δ . One solution would be to have p_{ij} depend on the current distance of i and j only (such an approach has been followed successfully by Wu, Chandler and Smit [62]). Alternatively, one may construct p_{ij} in such a way that it depends on Δ , but in such a way that it is symmetric in v_1 and v_2 . A possible expression would be:

$$p_{ij} = 1 - \frac{\min(1, \exp(-\beta v_1)) + \min(1, \exp(-\beta v_2))}{2}$$

This expression reduces to the previous one for the Ising and xy models. The difference with these earlier examples is that, in those cases, once a cluster (or set of clusters) had been generated, the cluster moves were accepted with unit probability. In contrast, we must now explicitly compute the acceptance probability of a trial move. If we denote the potential energy of the old (new) configuration by \mathcal{V}_0 (\mathcal{V}_1), the detailed balance condition requires that

$$\exp(-\beta \mathcal{V}_0) \prod_{kl} (1 - p_{kl}^f) P_{\text{acc}}(0 \rightarrow 1) = \exp(-\beta \mathcal{V}_1) \prod_{kl} (1 - p_{kl}^r) P_{\text{acc}}(1 \rightarrow 0) \quad (84)$$

where k denotes a particle *on* the cluster, and l a particle outside it. The superscripts f and r denote forward and reverse moves. In writing eqn (84), we have used the fact that the probability to form intra-cluster bonds is the same for forward and reverse moves. Note, however, that for particles on different clusters, the probability to form a bond on the forward and reverse moves is not the same. From eqn (84), we derive an expression for the ratio of the acceptance probabilities:

$$R \equiv \frac{P_{\text{acc}}(0 \rightarrow 1)}{P_{\text{acc}}(1 \rightarrow 0)} = \exp(-\beta(\mathcal{V}_1 - \mathcal{V}_0)) \prod_{kl} \frac{(1 - p_{kl}^r)}{(1 - p_{kl}^f)}. \quad (85)$$

Note that all quantities on the right hand of eqn (85) can be evaluated once the cluster around i has been generated. As before, it is probably best to use the Metropolis rule

$$P_{\text{acc}}(0 \rightarrow 1) = \min(R, 1).$$

Clearly, many choices for p_{kl} are possible and there is, as yet, little experience with such schemes.

4.2.4. An alternative for 'Metropolis' sampling?. Finally, I wish to point out that the concept underlying the Swendsen-Wang scheme suggests an alternative to the Metropolis sampling for single-particle and collective moves in conventional Monte Carlo simulations. Before discussing *how* this can be achieved, I should first explain *why* such an alternative scheme is, in fact, desirable. To see this, consider the often quoted rule-of-thumb which states that it is advisable to choose the magnitude of a trial move such that the acceptance probability of such moves is approximately 50%. Of course, this is not a rigorous rule, but for systems that interact through a continuous potential, it is reasonable. In contrast, if we consider systems with hard-core interactions, it is better to work with much lower acceptance probabilities. The reason is the following: a trial move of particle i is accepted with a probability $P_{\text{acc}}(i) = \exp(-\beta\Delta\mathcal{V}(i))$, where $\Delta\mathcal{V}(i)$ represents the change in the potential energy of the system due to the displacement of particle i . In what follows, I shall assume that all interactions are pairwise additive, i.e.

$$\Delta\mathcal{V} = \sum_{j \neq i} \Delta v_{ij}.$$

In systems with continuous intermolecular interactions, the usual procedure to determine whether or not a trial move is accepted, is to compute Δv_{ij} for all j and then, at the end, compare $\exp(-\beta \sum \Delta v_{ij})$ with a random number between 0 and 1. However, for hard-core systems, we can compute $\exp(-\beta \Delta v_{ij})$ for every j and reject the move *as soon as we find an overlap*. As a consequence, rejected moves are cheaper than accepted moves in a hard-core system, but not in a system with continuous interactions. We can now balance the amount of computational work spent on accepted and rejected moves in the hard core system, by increasing the magnitude of the trial move (and thereby reducing the acceptance probability). I shall now show that we can use the concept of 'bonds' to construct a MC scheme for continuous potentials that will allow us to reject *at an early stage* moves that are likely to be 'unfavorable'. For hard-core particles, this scheme reduces to the conventional Metropolis scheme (see, however, below).

As before, I consider the trial move of a single particle i . Particle i interacts with 'neighboring' particles $\{j\}$. Due to the trial displacement the

interaction energy will change by an amount Δv_{ij} . I now construct *bonds* between particle i and its neighbors $\{j\}$ according to the following rule:

- (i) If $\Delta v_{ij} \leq 0$, i and j are not connected.
- (ii) If $\Delta v_{ij} > 0$, i and j are connected with a probability $p_{ij} = 1 - \exp(-\beta \Delta v_{ij})$ and disconnected otherwise.

The trial move is carried out if particle i is not connected with any other particle j . More importantly, the trial move is rejected *as soon as a bond is formed between i and any j* . Note that this scheme is different from the Metropolis scheme, because it will result in the rejection of some moves that *lower* the potential energy. The scheme does, however, satisfy detailed balance. To see this, consider the probability of accepting a trial move from the 'old' position of particle i to the new position:

$$P_{\text{acc}}(i \rightarrow i') = \prod_{j \neq i} (1 - p_{ij}) = \exp \left[-\beta \sum_{j \neq i} \Delta v_{ij}^+ \right], \quad (86)$$

where the sum over j enumerates all interaction energies that *increase* as a result of the trial move. Next, consider the reverse move. I will assume that the underlying Markov chain of the Monte Carlo scheme is symmetric, i.e. the *a priori* probabilities of going from i to i' and from i' to i are equal. The acceptance probability for the trial move from i' to i is

$$P_{\text{acc}}(i' \rightarrow i) = \prod_{j \neq i} (1 - p_{ij}) = \exp \left[-\beta \sum_{j \neq i} \Delta v_{ij}^- \right]. \quad (87)$$

But now the sum over j includes only those pairs $\{ij\}$ for which the energy increases when going from the trial position to the original one. The ratio of these acceptance probabilities is

$$\frac{P_{\text{acc}}(i \rightarrow i')}{P_{\text{acc}}(i' \rightarrow i)} = \exp \left[-\beta \sum_{j \neq i} \Delta v_{ij} \right]. \quad (88)$$

In eqn (88), the sum over j is unrestricted! Hence our scheme satisfies detailed balance.

A few remarks are in place. First of all, one should bear in mind that there is a trade-off between the possibility of rejecting a move at an early stage and the computation of a large number of quasi-random numbers. In practice, it will probably be best to use the bond-generating MC scheme only for those interactions that are likely to change appreciably in a trial move, i.e. the interactions with near neighbors. If this part of the test is passed, the change in long-range interactions can be summed in the usual way and the normal Metropolis rule can be used to test if the overall move will be accepted.

Secondly, the scheme described here, is in no way limited to single-particle moves. In fact, it is probably most useful to apply it to complex many-particle moves that have a very low acceptance. Examples are volume changes in constant pressure simulations of systems with strong, short-ranged attractions (e.g. the attractive, hard-core Yukawa system). A second interesting application is the *staging* approach to path-integral Monte Carlo in systems with *attractive* interactions. A third, useful application is the sampling of *configurational-bias* moves in the simulation of systems of chain molecules (see section 4.1). Here too, the average acceptance probability of trial moves may be very small. This suggests that we may often find that, after 'regrowing' only a small fraction of the chain molecule, the ratio of the Rosenbluth weights that we are considering becomes very small. This implies that, most likely, the move will be rejected. In the conventional CBMC scheme, we must nevertheless follow the regrowth process to the very end. In contrast, the present scheme allows us to terminate a 'doomed' trial move at an earlier stage. It should be noted that, in this application, the bond-generating MC scheme should be superior to the conventional 'Metropolis' scheme, *even for hard-core molecules*.

In this brief outline, I have only sketched a few obvious applications. However, there are clearly many others. In general, one may expect the present scheme to be most efficient in situations where conventional Monte Carlo moves (i) are expensive and (ii) have a very low acceptance probability. The present scheme allows one to quickly eliminate a large fraction of the 'doomed' trial moves. It should be stressed that we are free to decompose the contributions to the acceptance probability of trial moves in any way that we consider convenient (as long as this is done consistently). In particular, this means that there is no need to generate a quasi-random number for *every* interaction that we compute. However, as the best way of implementing the algorithm clearly depends on the system under consideration, I shall not discuss this point here.

5. Thermodynamic integration

In section 3.1 it was mentioned that knowledge of 'thermal' quantities, such as μ , is usually necessary to locate the coexistence line for a first-order phase-transition. At first sight knowledge of μ may appear superfluous. After all, a computer simulation mimics the behaviour of a real solid or liquid. If the simulation is ergodic it should spontaneously transform to whatever phase is thermodynamically most stable, and then we would know all there is to know. Unfortunately, this approach does not work. At least, not for phase transitions involving three-dimensional solids. At a solid-solid or solid-liquid phase transition very strong hysteresis effects are usually observed in a simulation (see Fig. 5). In fact, it is very difficult to nucleate a crystal

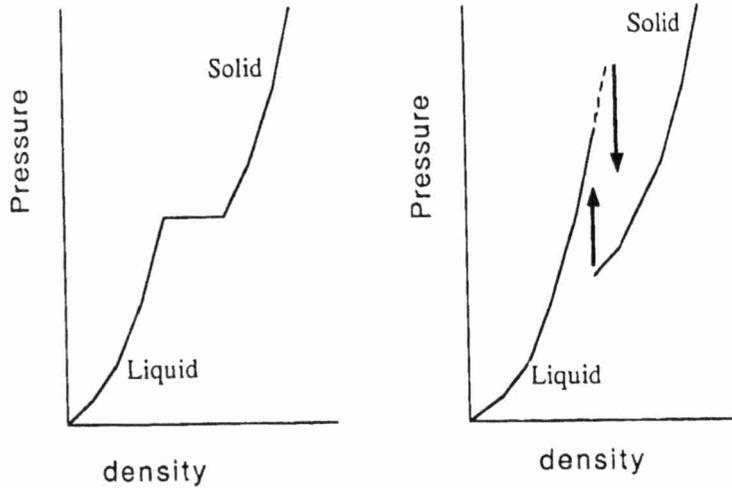


Fig. 5. First-order phase transitions in small model systems with periodic boundary conditions (right) tend to exhibit much stronger hysteresis effects than are usually observed in the real world (left)

from a liquid during a simulation. Hence, to locate the point where two phases coexist, we must compute the chemical potential of the homogeneous phases at the same temperature and pressure and find the point where the two μ 's are equal. For not-too-dense fluids, one might use the Gibbs method (see Chapter 6) to allow the computer to find this point. However, the latter method does not work for phase transitions involving solids. In such cases we really must compute μ . General methods to compute μ and related thermal quantities by computer simulation are discussed in Ref. [23]. In the present section I wish to describe a simple, and usually reliable, method to compute the free energy F (and hence μ).

Let us first recall how free energies are measured in real experiments. In the real world free energies cannot be obtained from a single measurement either. What can be measured, however, is the derivative of the free energy with respect to volume V and temperature T :

$$\left(\frac{\partial F}{\partial V}\right)_{NT} = -P \quad (89)$$

and

$$\left(\frac{\partial F/T}{\partial 1/T}\right)_{VT} = E. \quad (90)$$

Here P is the pressure and E the energy of the system under consideration.

The trick is now to find a reversible path that links the state under consideration to a state of known free energy. The change in F along that path can then simply be evaluated by integration of eqns (89) and (90). There are only very few thermodynamic states where the free energy of a substance is known. One state is the ideal gas phase, the other is the perfectly ordered ground state at $T = 0K$.

In computer simulations, the situation is quite similar. In order to compute the free energy of a dense liquid, one may construct a reversible path to the very dilute gas phase. It is not really necessary to go all the way to the ideal gas. But at least one should reach a state that is sufficiently dilute that the free energy can be computed accurately, either from knowledge of the first few terms in the virial expansion of the compressibility factor $PV/Nk_B T$, or that the chemical potential can be computed by other means (see section 3.1 and 3.4). For the solid, the ideal gas reference state is less useful (although techniques have been developed to construct a reversible path from a dense solid to a dilute (lattice-) gas [63]). The obvious reference state for solids is the harmonic lattice. Computing the absolute free energy of a harmonic solid is relatively straightforward, at least for atomic and simple molecular solids. However, not all solid phases can be reached by a reversible route from a harmonic reference state. For instance, in molecular systems it is quite common to find a strongly anharmonic plastic phase just below the melting line. This plastic phase is not (meta-) stable at low temperatures.

Fortunately, in computer simulations we do not have to rely on the presence of a 'natural' reversible path between the phase under study and a reference state of known free energy. If such a path does not exist, we can construct an artificial path. This is in fact a standard trick in statistical mechanics (see e.g. Ref. [64]). It works as follows: Consider a case where we need to know the free energy $F(V, T)$ of a system with a potential energy function \mathcal{V}_1 , where \mathcal{V}_1 is such that no 'natural' reversible path exists to a state of known free energy. Suppose now that we can find another model system with a potential energy function \mathcal{V}_0 for which the free energy *can* be computed exactly. Now let us define a generalized potential energy function $\mathcal{V}(\lambda)$, such that $\mathcal{V}(\lambda = 0) = \mathcal{V}_0$ and $\mathcal{V}(\lambda = 1) = \mathcal{V}_1$. The free energy of a system with this generalized potential is denoted by $F(\lambda)$. Although $F(\lambda)$ itself cannot be measured directly in a simulation, we can measure its derivative with respect to λ :

$$\left(\frac{\partial F}{\partial \lambda}\right)_{NVT\lambda} = \left\langle \frac{\partial \mathcal{V}(\lambda)}{\partial \lambda} \right\rangle_{NVT\lambda} \quad (91)$$

If the path from $\lambda = 0$ to $\lambda = 1$ is reversible, we can use eqn (91) to compute the desired $F(V, T)$. We simply measure $\langle \partial \mathcal{V} / \partial \lambda \rangle$ for a number of values of λ between 0 and 1. Typically, 10 quadrature points will be sufficient to get the absolute free energy per particle accurate to within $0.01k_B T$.

It is however important to select a reasonable reference system. One of the safest approaches appears to be to choose as a reference system an Einstein crystal with the same structure as the phase under study [65]. This choice of reference system makes it extremely improbable that the path connecting $\lambda = 0$ and $\lambda = 1$ will cross an (irreversible) first-order phase transition from the initial structure to another, only to go back to its original structure for still larger values of λ . Nevertheless, it is important that the parametrization of $\mathcal{V}(\lambda)$ be chosen carefully. Usually, a linear parametrization (i.e. $\mathcal{V}(\lambda) = \lambda\mathcal{V}_1 + (1-\lambda)\mathcal{V}_0$) is quite satisfactory. But occasionally such a parametrization may lead to weak (and relatively harmless) singularities in eqn (91) for $\lambda \rightarrow 0$. More details about such free energy computations can be found in Ref. [23].

Similar techniques can be used to locate first-order phase transitions involving phases with partial order (e.g. liquid crystals). For details, the reader is referred to Refs. [66–68].

6. The grand canonical miracle

In this final section I discuss entropy-driven phase transitions. The main aim of the examples discussed here is to show that, in some cases, the grand-canonical ensemble greatly simplifies the description of solvent effects on the structure and phase behavior of complex fluids.

6.1. BINARY MIXTURES

Phase separation in binary mixtures is the example that is used in many textbooks to illustrate the competition between energy and entropy in a phase transformation. For a mixture at constant total volume V , the Helmholtz free energy F , should be minimal. As a first approximation, the entropy of mixing of a mixture of two species A and B , is replaced by the entropy of mixing of an ideal mixture

$$S_{\text{id}}(X) = -Nk_{\text{B}}[X \ln X + (1 - X) \ln(1 - X)] \quad (92)$$

where X denotes the mole fraction of one component (say A): $X = N_A/(N_A + N_B)$. The entropy of mixing given by eqn (92) is a convex function of X . As a consequence, $S_{\text{id}}(X)$ will always decrease if phase separation takes place. This implies that phase separation can only take place if the resulting decrease in energy E outweighs the increase in $-TS_{\text{id}}$. In a hard-core mixture, there is no energy change upon mixing. Hence, if eqn (92) were exact, we should never observe phase separation in a hard-core mixture. And, even though eqn (92) is known to be only an approximation, it was commonly thought that a fluid mixture of dissimilar hard spheres would not phase separate. This belief was, at least partly, based on the work of Lebowitz and

Rowlinson [69], who studied the phase-behavior of such binary mixtures of dissimilar hard spheres, using a more accurate approximation for the entropy of mixing, based on the Percus-Yevick (PY) integral equation. Lebowitz and Rowlinson found that, at least within that approximation, hard spheres of arbitrary size ratio will mix in all proportions in the fluid phase. The implication of this result was that entropic effects are not enough to cause a miscibility gap in a simple fluid mixture. Recently, however, Biben and Hansen [70] have applied a more sophisticated analytical theory for dense fluid mixtures to the same problem. Unlike the PY approximation, this theory is found to predict that an asymmetric binary hard-sphere mixture should phase separate if the sizes of the two spheres are sufficiently dissimilar (typically, if the size ratio is less than 0.2). However, as in the case of Ref. [69], it is based on an approximate theory for the hard-sphere mixture. Hence, one may wonder to what extent the result found in Ref. [70] depends on the approximations that are used to compute the equation-of-state of the mixture. Clearly, it would be highly desirable to have a model system of a binary mixture for which the existence of a purely entropic demixing transition can be proven exactly.

It turns out that, in a special case [71], this can indeed be done by constructing a simple lattice model of binary hard-core mixture. By transforming to a grand-canonical ensemble, this binary hard-core mixture can be mapped onto a one-component lattice gas with attractive nearest neighbor interactions. This mapping is of interest because the one-component lattice gases can, in turn, be mapped onto Ising-like models for which, in some cases, the phase behavior is known exactly. To illustrate our approach, first consider a trivial model-system, namely a square lattice with at most one particle allowed per square. Apart from the fact that no two particles can occupy the same square face, there is no interaction between the particles. For a lattice of N sites, the grand-canonical partition function is:

$$\Xi = \sum_{\{n_i\}} \exp \left[\beta \mu \sum_i n_i \right]. \quad (93)$$

The sum is over all allowed sets of occupation numbers $\{n_i\}$ and μ is the chemical potential. Next, we include "small" hard particles that are allowed to sit on the links of the lattice. (see Fig. 6). These small particles are excluded from the edges of a square that is occupied by a large particle.

For a given configuration $\{n_i\}$ of the large particles, one can then exactly calculate the grand canonical partition function of the small particles. Let $M = M(\{n_i\})$ be the number of free spaces accessible to the small particles. Then clearly:

$$\Xi_{\text{small}}(\{n_i\}) = \sum_{l=0}^M \frac{M! z_s^l}{l!(M-l)!} = (1 + z_s)^{M(\{n_i\})}, \quad (94)$$

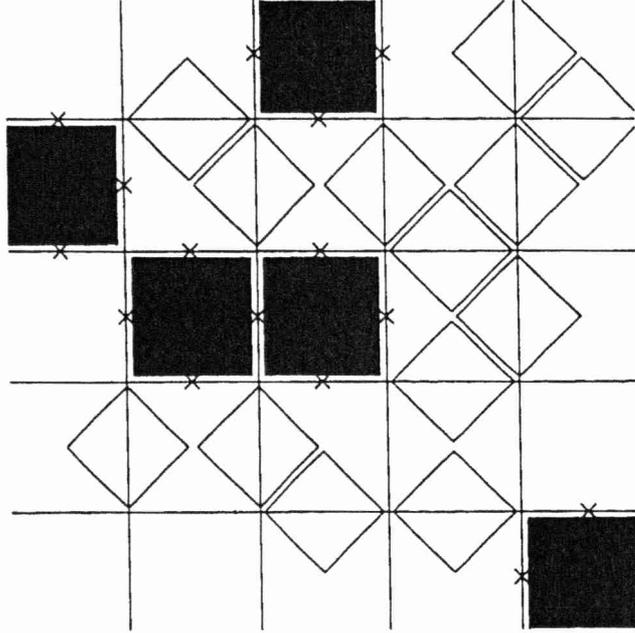


Fig. 6. Exactly solvable lattice-gas model for a mixture of large (black squares) and small (white squares) hard particles. The crosses indicate which small-particle sites are excluded by the presence of large particles.

where z_s is the fugacity of the small particles. M can be written as

$$M(\{n_i\}) = 2N - 4 \sum_i n_i + \sum_{\langle ij \rangle} n_i n_j, \quad (95)$$

where $2N$ is the number of links on the lattice and the second sum is over nearest-neighbor pairs, and comes from the fact that when two large particles touch, one link is doubly excluded [72]. The total partition function for the "mixture" is:

$$\Xi_{\text{mixture}} = \sum_{\{n_i\}} \exp \left\{ [\beta\mu - 4 \log(1 + z_s)] \sum_i n_i + [\log(1 + z_s)] \sum_{\langle ij \rangle} n_i n_j \right\}, \quad (96)$$

where we have omitted a constant factor $(1 + z_s)^{2N}$. Equation (96) is simply the expression for the grand-canonical partition function of a one-component lattice-gas Ising model with a shifted chemical potential and an effective

nearest neighbor attraction with an interaction strength $\log(1+z_s)/\beta$. As is well known, this lattice model can again be transformed to a 2-D Ising spin model that can be solved in the zero field case [73, 74]. In the language of our mixture model, no external magnetic field means:

$$(1+z_s)^2 = z_1, \quad (97)$$

where $z_1 = \exp \beta\mu$, the large particle fugacity. The order-disorder transition in the Ising model then corresponds to phase separation in the language of our model. This demixing is purely entropic, just like the transition predicted [70] for the hard sphere mixture. In fact, the mapping described above can also be carried through when energetic interactions between the large particles are included. However, for the sake of clarity, we will restrict ourselves to athermal hard-core mixtures.

Of course, there is a wide variety of lattice models for hard-core mixtures that can be mapped onto one-component systems with effective attraction. The model discussed above is only special in the sense that it can be mapped onto a model that is exactly solvable. In particular, from the known results for the 2D-Ising model (see e.g. Ref. [75]), it is straightforward to evaluate the composition of the coexisting phases along the entire coexistence line. This coexistence curve is shown in Fig. 7. One important question that is raised by the work of Biben and Hansen [70] is whether the demixing transition is of the fluid-fluid or the fluid-solid type. The phase transition in the square-lattice model that we discussed above provides no answer to this question, as there is no distinction between 'liquid' and 'solid' in a lattice gas on a square, or simple-cubic lattice. Yet, from the above mapping, it is immediately obvious that a fluid-fluid transition does, in fact, occur in another hard-core mixture, namely that of a mixture of hard-core monomers and polymers on a lattice. To this end, we consider, once again our model of a mixture of large and small hard squares (cubes, in 3-D) on a square (cubic) lattice, i.e. the model that could be mapped onto the 1-component lattice gas with nearest neighbor interactions. We now construct 'polymers' by connecting N large squares (cubes), while the solvent is represented by the small particles. The simplest polymers are made by connecting N large particles by nearest neighbor bonds only. A grand-canonical summation over all configurations of the small particles, yields a very simple expression for the (canonical) partition function of the polymers, namely

$$Z_{\text{polymers}} = \sum_{\{n_i\}} \exp \left[\frac{J}{k_B T} \sum_{\langle ij \rangle} n_i n_j \right] \quad (98)$$

with $J \equiv k_B T \log(1+z_s)$ and where the sum is over all acceptable (i.e. non-overlapping) configurations of the hard-core polymers. Equation (98) is precisely the expression for the partition function of the Flory-Huggins lattice

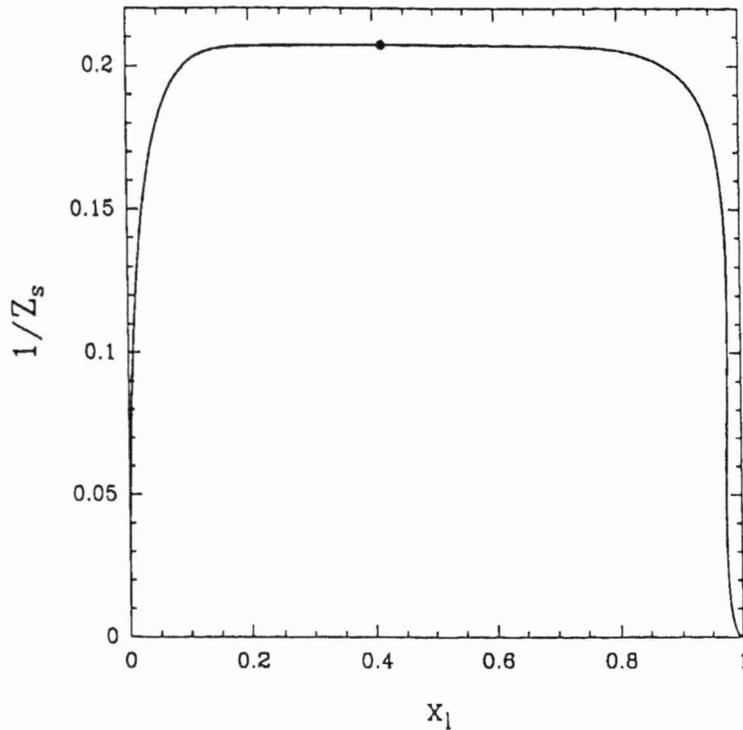


Fig. 7. Phase diagram of the model system shown in Fig. 6. X_1 denotes the molar fraction of large squares, z_s is the fugacity of small squares. The drawn curve indicates the composition of coexisting phases. Note that this curve is asymmetric. In particular, the critical point (indicated by a black dot) is located at $X_1 = 2^{1/2} - 1$. Along the coexistence curve the fugacities of the large and small particles are related through eqn (97).

model [76]. This model has been studied extensively both using approximate analytical theories, in particular the well-known Flory-Huggins theory and modifications thereof (for a critical review see [77]) and, more recently, by direct numerical simulations (see e.g. Refs. [58, 78]). Traditionally, the coupling constant J has been interpreted as a purely *energetic* interaction. In contrast, in the context of the model that we consider, J is of completely entropic origin. We can now translate the results that have been obtained for the Flory-Huggins lattice-gas model directly to our hard-core mixture. This means that many effects that in the past were interpreted as originating from energetic attractions, can also be obtained from purely entropic effects. In particular, the existence of a first-order fluid-fluid phase separation in this model provides direct proof that a purely entropic demixing transition ex-

ists in a hard-core polymer-solvent mixture. Our analysis also yields a very simple ‘entropic’ interpretation of other, closely related phenomena, such as polymer collapse in a ‘poor’ solvent. It is, perhaps, worth pointing out that in our model, the solvent molecules are smaller than the monomeric units of the polymer. In contrast, in the original Flory-Huggins lattice model, there is no such size-asymmetry. We can, however, construct a hard-core lattice model where solvent and monomeric units occupy the same volume. It can be shown that the solvent-induced polymer collapse will also occur in such a ‘symmetric’ model system.

Finally, we note that the Flory-Huggins theory for polymer solutions yields the following (approximate) expression for the free-energy of a polymer solution:

$$\frac{F_{\text{mix}}}{k_{\text{B}}T} = \frac{\phi}{N} \log \phi + (1 - \phi) \log(1 - \phi) + \chi \phi(1 - \phi) \quad (99)$$

where ϕ is the fraction of the volume occupied by polymer, while the parameter χ is related to the coupling constant J of the original lattice model, by

$$\chi = \frac{1}{2} \frac{J}{k_{\text{B}}T} n_{\text{b}} = \frac{1}{2} n_{\text{b}} \ln(1 + z_{\text{s}}), \quad (100)$$

where n_{b} is the coordination number of the lattice. If, as was assumed in the original Flory-Huggins theory, J is due to energetic interactions, then χ should vary as $1/T$. However, in the present (extreme) interpretation of the same lattice-gas model, the parameter χ would be *completely independent of temperature*. There is, in fact, a large body of experimental data that shows that, for many polymer solutions, χ has a large, if not dominant, entropic (i.e. temperature-independent) part.

6.2. DEPLETION FLOCCULATION

Let us next consider a slightly more complex example of an entropy-driven phase separation in a binary mixture, namely polymer-induced flocculation of colloids. Experimentally, it is well known that the addition of a small amount of free, non-adsorbing polymer to a colloidal suspension induces an effective attraction between the colloidal particles and may even lead to coagulation. This effect has been studied extensively [79] and is well understood, at least qualitatively. As in the example discussed above, the polymer-induced attraction between colloids is an *entropic* effect: when the colloidal particles are close together, the total number of accessible polymer conformations is larger than when the colloidal particles are far apart. However, although the physical mechanism responsible for polymer-induced coagulation is understood *qualitatively*, a quantitative description of this phenomenon

is difficult. This is so because the polymer-induced attraction between the colloidal particles is non-pairwise additive. Moreover, it depends both on the osmotic pressure of the polymer and on the concentration of the colloid. Yet, in the theoretical description of polymer-induced clustering [80], the effect of the polymer is usually replaced by an effective, density-independent, pair-wise additive attraction between the colloidal particles. However, in this case, no analytical evaluation of the (grand-canonical) partition function is possible, even when one considers only the very simplest model, *viz.* that of a mixture of hard-core colloidal particles with ideal chain molecules with conformations that are restricted to a lattice. In this case, it would clearly be desirable to carry out 'exact' numerical simulations to investigate the phase behavior. Yet, the computational problems are still formidable. What is required is a numerical scheme that samples the positions of the colloidal particles while averaging over all possible conformations of a large (and fluctuating) number of chain molecules. The 'conventional' Monte Carlo schemes to simulate lattice models of polymer systems [58] would be woefully inadequate for such a calculation.

Fortunately, it is possible to construct a rigorous and efficient Monte Carlo scheme to study this model. Our approach relies on the fact that we can recursively compute the partition function of an ideal (non-self avoiding) chain on a lattice in an arbitrary external potential [81, 82]. This is most easily seen by considering a chain of length $\ell - 1$ on a lattice. For convenience, we assume that the external potential is either zero or infinite. The total number of accessible ideal chain conformations that terminate on lattice site i is denoted by $\omega_{\ell-1}(i)$. The total partition function $\Omega_{\ell-1}$ is equal to $\sum_i \omega_{\ell-1}(i)$, where the sum runs over all lattice sites. The total number of chains of length ℓ that terminate on site i is clearly equal to the sum of the total number of chains of length $\ell - 1$ that terminate on any of the neighbors of i , multiplied by the Boltzmann factor associated with site i . Using such a recursive scheme, we can compute *exactly* Ω_ℓ , the partition function of a single ideal polymer of arbitrary length ℓ on a lattice, in an arbitrary external potential. This scheme can be used as a starting point to study self-avoiding polymers [83, 84], but for the present purpose we limit ourselves to ideal polymers. Up to this point we have not specified the nature of the 'external' potential. We now assume that this potential is due to the presence of N hard, spherical colloidal particles each of which occupies many lattice sites. The polymer partition function clearly depends on the coordinates \mathbf{r}^N of the colloidal particles: $\Omega_\ell(\mathbf{r}^N)$. The configurational part of the partition function of the system of N colloids plus one polymer of length ℓ in volume V is then given by:

$$Z(V, N, 1) = \int_V d\mathbf{r}^N e^{-\nu_h \cdot (\mathbf{r}^N)} \Omega_\ell(\mathbf{r}^N), \quad (101)$$

where $\mathcal{V}_{\text{hs}}(\mathbf{r}^N)$ denotes the hard-sphere interaction. Next, we make use of the fact that we are considering *ideal* polymers. In that case we can immediately write down the corresponding partition function for N colloids and M ideal polymers:

$$Z(V, N, M) = \int_{\mathcal{V}} d\mathbf{r}^N e^{-\mathcal{V}_{\text{hs}}(\mathbf{r}^N)} [\Omega_{\ell}(\mathbf{r}^N)]^M / M!, \quad (102)$$

where the factor $1/M!$ accounts for the fact that the polymers are indistinguishable. Using eqn (102) it is straightforward to transform to an ensemble where the polymer chemical potential (i.e. the osmotic pressure) is kept fixed. The corresponding grand-canonical partition function is given by:

$$\begin{aligned} \Xi(V, N, \mu) &= \sum_{M=0}^{\infty} e^{M\mu_{\text{ch}}} Z(V, N, M) / M! \\ &= \int_{\mathcal{V}} d\mathbf{r}^N e^{-\mathcal{V}_{\text{hs}}(\mathbf{r}^N)} \sum_{M=0}^{\infty} e^{M\mu_{\text{ch}}} [\Omega_{\ell}(\mathbf{r}^N)]^M / M! \\ &= \int_{\mathcal{V}} d\mathbf{r}^N e^{-\mathcal{V}_{\text{hs}}(\mathbf{r}^N)} e^{z\Omega_{\ell}(\mathbf{r}^N)}. \end{aligned} \quad (103)$$

In the last line of eqn (103), we have introduced the polymer fugacity $z \equiv \exp(\mu_{\text{ch}})$, where μ_{ch} denotes the chemical potential of the chain molecules.

The important point to note is that eqn (103) allows us to evaluate the properties of the colloidal particles in osmotic equilibrium with a polymer reservoir. In particular, it shows that we can perform Monte Carlo sampling of the colloidal particles. The polymers only affect $\mathcal{V}_{\text{eff}}(\mathbf{r}^N)$, the effective interaction between the colloidal particles:

$$\mathcal{V}_{\text{eff}}(\mathbf{r}^N) \equiv \mathcal{V}_{\text{hs}}(\mathbf{r}^N) - z \Omega_{\ell}(\mathbf{r}^N). \quad (104)$$

$z\Omega_{\ell}(\mathbf{r}^N)$ measures the entropic interaction between the colloids due to *all* possible polymer conformations. This entropic interaction is, in principle, not pairwise additive. In fact, it is shown in Ref. [82], that for all but the shortest chain molecules, this non-additivity of the polymer-induced interaction between the colloids, has a pronounced effect on the structure and stability of the mixture. For more details, we refer the reader to Ref. [82].

6.3. FROM SOAP TO SATURN?

Finally, I wish to discuss a more speculative, and somewhat exotic, example of an entropy-driven phase transition, namely the demixing of a system of ideal point fermions (“non-interacting electrons”) and classical fluid particles. The analogy between this system and the previous ones rests on the fact that, in Feynman’s path-integral formulation of quantum mechanics, there is an isomorphism between the partition-function of a d -dimensional

quantum-mechanical system, and that of a $(d + 1)$ -dimensional ‘polymer’ system. Ideal quantum particles in d dimensions can be described by ideal, directed polymers in $d+1$ dimensions. This suggests that we may expect that mixtures of ideal quantum particles and atoms (effectively, hard spheres) behave much like a mixture of ideal polymers and colloids. In fact, the situation in the quantum case is complicated by the fact that one must take exchange into account. This is relatively straightforward in the case of numerical simulations of bosons [85], but it creates serious problems for fermions (the ‘sign’-problem). However, in Ref. [86], it is shown how, at least for a mixture of atoms and ideal fermions, these problems may be overcome. Alavi and Frenkel [86] show that it is, in fact, possible to evaluate numerically (but exactly) the grand-canonical partition function of ideal fermions in the presence of arbitrary “obstacles” (in particular, atoms). This makes it possible to perform Monte Carlo sampling of the atomic positions in the presence of the fermion ‘bath’. Here, I will not discuss the technical aspects of these simulations. Rather, I will summarize the main conclusion, which is that there is indeed strong evidence for a demixing transition as the chemical potential of the fermions is increased beyond a certain critical value. Again, there are no attractive forces in this system. Hence, this fermion-induced phase separation appears to be yet another example of “attraction through repulsion”. For more details, the reader is referred to Ref. [86]. It is amusing to note that the phase-separation in a mixture of ideal fermions and hard-spherical “atoms” may have an intriguing counterpart in planetary astronomy, namely in the structure of the interior of Saturn. It is thought that the core of Saturn, which is a hydrogen-helium planet, consists of nearly pure helium [87], whilst at the conditions relevant to the interior of Saturn ($T \approx 10^4\text{K}$, $P \approx 1 - 10\text{Mbar}$) the hydrogen probably is metallic. If we view the helium as consisting of ‘hard spheres’, and the metallic hydrogen as a nearly free electron gas (in which the hydrogen nuclei play the role of a neutralising background), we might expect on the basis of the calculations in Ref. [86] that such a mixture would phase-separate. Clearly, this example is far removed from most other topics discussed during this school, but it does provide a nice illustration of the unity of physics.

Acknowledgements

The work of the FOM Institute is part of the research program of FOM and is supported by the ‘Nederlandse Organisatie voor Wetenschappelijk Onderzoek’(NWO).

References

- [1] N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, and E. TELLER, *J. Chem. Phys.* **21**, 1087 (1953).

- [2] M. CREUTZ, *Phys. Rev. Lett.* **50**, 1411 (1983).
- [3] K. BINDER, *The Monte Carlo Method in Condensed Matter Physics*, Springer, Berlin, 1992.
- [4] W. W. WOOD, *J. Chem. Phys.* **48**, 415 (1968).
- [5] I. R. McDONALD, *Molec. Phys.* **23**, 41 (1972).
- [6] Actually, there is no need to assume a real piston. The systems with volume V and $V_0 - V$ may both be isolated systems subject to their individual (periodic) boundary conditions. The only constraint that we impose is that the sum of the volumes of the two systems equals V_0 .
- [7] R. EPPENGA and D. FRENKEL, *Molec. Phys.* **52**, 1303 (1984).
- [8] M. PARRINELLO and A. RAHMAN, *Phys. Rev. Lett.* **45**, 1196 (1980).
- [9] M. PARRINELLO and A. RAHMAN, *J. Appl. Phys.* **52**, 7182 (1981).
- [10] H. C. ANDERSEN, *J. Chem. Phys.* **72**, 2384 (1980).
- [11] R. NAJAFABADI and S. YIP, *Scripta Metall.* **17**, 1199 (1983).
- [12] G. E. NORMAN and V. S. FILINOV, *High Temp. Res. USSR* **7**, 216 (1969).
- [13] D. J. ADAMS, *Molec. Phys.* **28**, 1241 (1974).
- [14] D. J. ADAMS, *Molec. Phys.* **29**, 307 (1975).
- [15] D. J. ADAMS, *Molec. Phys.* **32**, 647 (1976).
- [16] D. J. ADAMS, *Molec. Phys.* **37**, 211 (1979).
- [17] L. A. ROWLEY, D. NICHOLSON, and N. G. PARSONAGE, *J. Comput. Phys.* **17**, 401 (1975).
- [18] J. YAO, R. A. GREENKORN, and K. C. CHAO, *Molec. Phys.* **46**, 587 (1982).
- [19] M. MEZEI, *Molec. Phys.* **40**, 901 (1980).
- [20] J. P. VALLEAU and K. L. COHEN, *J. Chem. Phys.* **72**, 3935 (1980).
- [21] W. VAN MEEGEN and I. SNOOK, *J. Chem. Phys.* **73**, 4656 (1980).
- [22] M. P. ALLEN and D. J. TILDESLEY, *Computer Simulation of Liquids*, Clarendon, Oxford, 1987.
- [23] D. FRENKEL, Free-energy computation and first-order phase transitions, in *Molecular Dynamics Simulations of Statistical Mechanical Systems*, edited by G. CICCOTTI and W. G. HOOVER, pp. 151-188, North-Holland, Amsterdam, 1986.
- [24] A. Z. PANAGIOTOPOULOS, *Molec. Phys.* **61**, 813 (1987).
- [25] D. A. KOFKE and E. D. GLANDT, *Molec. Phys.* **64**, 1105 (1988).
- [26] K. S. SHING, *Chem. Phys. Lett.* **119**, 149 (1985).
- [27] K. S. SHING, *J. Chem. Phys.* **85**, 1105 (1986).
- [28] P. SINDZINGRE, G. CICCOTTI, C. MASSOBRIO, and D. FRENKEL, *Chem. Phys. Lett.* **136**, 35 (1987).
- [29] W. G. T. KRANENDONK and D. FRENKEL, *Molec. Phys.* **72**, 699 (1991).
- [30] M. R. STAPLETON, D. J. TILDESLEY, and N. QUIRKE, *J. Chem. Phys.* **92**, 4456 (1990).
- [31] B. WIDOM, *J. Chem. Phys.* **39**, 2808 (1963).
- [32] D. J. ADAMS, *Molec. Phys.* **28**, 1252 (1974).
- [33] U. HEINBUCH and J. FISCHER, *Molec. Simul.* **1**, 109 (1987).
- [34] J. I. SIEPMANN, I. R. McDONALD, and D. FRENKEL, *J. Phys. Cond. Mat.* **4**, 679 (1992).

- [35] K. S. SHING and S. T. CHUNG, *J. Phys. Chem.* **91**, 1674 (1987).
- [36] P. SINDZINGRE, C. MASSOBRIO, G. CICCOTTI, and D. FRENKEL, *Chem. Phys.* **129**, 213 (1989).
- [37] J. I. SIEPMANN, *Molec. Phys.* **70**, 1145 (1990).
- [38] J. HARRIS and S. A. RICE, *J. Chem. Phys.* **88**, 1298 (1988).
- [39] M. N. ROSENBLUTH and A. W. ROSENBLUTH, *J. Chem. Phys.* **23**, 356 (1955).
- [40] D. FRENKEL, *J. Phys. Cond. Mat.* **2(SA)**, 265 (1990).
- [41] G. C. A. M. MOOIJ and D. FRENKEL, *Molec. Phys.* **74**, 41 (1991).
- [42] D. FRENKEL and B. SMIT, *Molec. Phys.* **75**, 983 (1992).
- [43] J. J. DE PABLO, M. LASO, and U. W. SUTER, *J. Chem. Phys.* **96**, 6157 (1992).
- [44] S. K. KUMAR, I. SZLEIFER, and A. Z. PANAGIOTOPOULOS, *Phys. Rev. Lett.* **66**, 2935 (1991).
- [45] B. SMIT, G. C. A. M. MOOIJ, and D. FRENKEL, *Phys. Rev. Lett.* **68**, 3657 (1992), Comment.
- [46] K. S. SHING and K. E. GUBBINS, *Molec. Phys.* **46**, 1109 (1982).
- [47] K. S. SHING and K. E. GUBBINS, *Molec. Phys.* **49**, 1121 (1983).
- [48] C. H. BENNETT, *J. Comput. Phys.* **22**, 245 (1976).
- [49] G. M. TORRIE and J. P. VALLEAU, *J. Comput. Phys.* **23**, 187 (1977).
- [50] D. CHANDLER, *Introduction to Modern Statistical Mechanics*, Oxford University Press, New York, 1987.
- [51] J. S. VAN DUJNEVELDT and D. FRENKEL, *J. Chem. Phys.* **96**, 4655 (1992).
- [52] K.-K. HAN, *Phys. Lett. A* **165**, 28 (1992).
- [53] P. J. ROSSKY, J. D. DOLL, and H. L. FRIEDMAN, *J. Chem. Phys.* **69**, 4628 (1978).
- [54] S. DUANE, A. D. STEVENSON, B. J. PENDLETON, and D. ROWETH, *Phys. Lett. B* **195**, 216 (1987).
- [55] R. H. SWENDSEN and J.-S. WANG, *Phys. Rev. Lett.* **58**, 86 (1987).
- [56] J. I. SIEPMANN and D. FRENKEL, *Molec. Phys.* **75**, 59 (1992).
- [57] J. J. DE PABLO, M. LASO, and U. W. SUTER, *J. Chem. Phys.* **96**, 2394 (1992).
- [58] K. KREMER and K. BINDER, *Comput. Phys. Rep.* **7**, 259 (1988).
- [59] G. C. A. M. MOOIJ, D. FRENKEL, and B. SMIT, *J. Phys. Cond. Mat.* **4**, 3053 (1992).
- [60] G. C. A. M. MOOIJ, D. FRENKEL, and B. SMIT, *J. Phys. Cond. Mat.* **4**, L255 (1992).
- [61] M. LASO, J. J. DE PABLO, and U. W. SUTER, *J. Chem. Phys.* **97**, 2817 (1992).
- [62] D. WU, D. CHANDLER, and B. SMIT, *J. Phys. Chem.* **96**, 4077 (1992).
- [63] W. G. HOOVER and F. H. REE, *J. Chem. Phys.* **47**, 4873 (1967).
- [64] J. P. HANSEN and I. R. McDONALD, *Theory of Simple Liquids*, Academic Press, London, 2nd edition, 1986.
- [65] D. FRENKEL and A. J. C. LADD, *J. Chem. Phys.* **81**, 3188 (1984).
- [66] D. FRENKEL and B. M. MULDER, *Molec. Phys.* **55**, 1171 (1985).
- [67] A. STROOBANTS, H. N. W. LEKKERKERKER, and D. FRENKEL, *Phys. Rev. A* **36**, 2929 (1987).
- [68] D. FRENKEL, H. N. W. LEKKERKERKER, and A. STROOBANTS, *Nature* **332**, 822 (1988).

- [69] J. L. LEBOWITZ and J. S. ROWLINSON, *J. Chem. Phys.* **41**, 133 (1964).
- [70] T. BIBEN and J. P. HANSEN, *Phys. Rev. Lett.* **66**, 2215 (1991).
- [71] D. FRENKEL and A. A. LOUIS, *Phys. Rev. Lett.* **68**, 3363 (1992).
- [72] In fact, we can perform the same mapping for more extreme size-ratios, as long as the edges of the large particles exclude an integer number of small-particle sites.
- [73] T. D. LEE and C. N. YANG, *Phys. Rev.* **87**, 410 (1952).
- [74] L. ONSAGER, *Phys. Rev.* **65**, 117 (1944).
- [75] L. E. REICHL, *A modern course in Statistical Physics*, University of Texas Press, Austin, 1980.
- [76] P. J. FLORY, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, 1953.
- [77] K. F. FREED and M. G. BAWENDI, *J. Chem. Phys.* **93**, 2194 (1989).
- [78] A. BAUMGÄRTNER, in *Application of the Monte Carlo method in Statistical Physics*, edited by K. BINDER, pp. 145–180, Springer, Berlin, 1984.
- [79] D. H. NAPPER, *Polymeric Stabilization of Colloidal Dispersions*, Academic Press, 1983.
- [80] B. VINCENT, J. EDWARDS, S. EMMETT, and R. GROOT, *Coll. Surf.* **31**, 267 (1988).
- [81] D. FRENKEL, *J. Phys. Cond. Mat.* **2**, SA285 (1990).
- [82] E. J. MEIJER and D. FRENKEL, *Phys. Rev. Lett.* **67**, 1110 (1991).
- [83] G. C. A. M. MOOIJ and D. FRENKEL, *Molec. Phys.* **74**, 41 (1991).
- [84] D. FRENKEL, *Physica A* **176**, 54 (1991).
- [85] D. M. CEPERLEY and E. L. POLLOCK, *Phys. Rev. Lett.* **56**, 351 (1986).
- [86] A. ALAVI and D. FRENKEL, *J. Chem. Phys.* **97**, 9249 (1992).
- [87] D. J. STEVENSON and E. E. SALPETER, *The Astrophysical Journal Supplement Series* **35**, 221 (1977).